



DEPARTMENT OF COMPUTER SCIENCE  
FACULTY OF SCIENCE, KASETSART UNIVERSITY

# COMPUTATIONAL SCIENCE & BIG DATA FOR SUSTAINABLE AND INNOVATIVE FUTURE

August 2-5, 2017

Thailand Science Park, Pathum Thani, Thailand

# ANSCSE 21

The 21<sup>st</sup> International Annual Symposium on Computational Science and Engineering



e-Proceeding

ANSCSE 21

COMPUTATIONAL SCIENCE & BIG DATA FOR SUSTAINABLE AND INNOVATIVE FUTURE

**e-Proceeding**

**ANSCSE 21**

The 21<sup>st</sup> International Annual Symposium on  
Computational Science and Engineering

August 2-5, 2017

Thailand Science Park, Pathum Thani, Thailand

## CONTENT

	Page
WELCOME MESSAGES	i
COMMITTEE	v
PLENARY LECTURES	xiii
INVITED SPEAKERS	xv

	Page
<b>Effects of Incorporating Genetic Models into a Genetic Programming Tree Ensemble for Genetic Association Studies</b> <i>P. Aksornsingchai, K. Jutawongcharoen, W. Wongseree, N. Chaiyaratana, and D. Setsirichok</i>	1
<b>Incorporating a Genetic Model into a Logistic Regression Model Improves SNP Selection by Lasso for Genetic Association Studies</b> <i>D. Setsirichok, W. Wongseree, and N. Chaiyaratana</i>	9
<b>Backward Bifurcation of SEIR Epidemic Model with Treatment Function</b> <i>Sasiporn Rattanasupha, Settapat Chinviriyasit</i>	16
<b>Molecular Docking Investigation of WR99210 Analogues as Novel <i>M.tuberculosis</i> DHFR Inhibitors</b> <i>Nattaya Mekjirawat, Patcharapon Prasertth, Atcharapun Chockborvornpaisan, Suriyawut Kulatee, and Luckhana Lawtrakul</i>	22
<b>Enantioselectivity and Enzyme-Ligand Docking Studies of pfDHFR and Cycloguanil compounds</b> <i>S. Kulatee, P. Toochinda, L. Lawtrakul</i>	28
<b>Computational Docking Study of 4-oxo-N-(2-(piperidin-1-yl) ethyl)-4H-chromene-2-carboxamide as Acetylcholinesterase and Butyrylcholinesterase Inhibitors</b> <i>P. Suwanhom, V. S. Lee, T. Nualnoi, L. Lomlim</i>	38
<b>Molecular Modelling of 5-Lipoxygenase Enzyme with Antiasthmatic Substances from Plai</b> <i>K. Jitapunkul, O. Poachanukoon, S. Hannongbua, P. Toochinda, and L. Lawtrakul</i>	44

	<b>Page</b>
<b>A Theoretical Study of Small Schiff Base Complexes with Heavy Metal</b> <i>Kunanon Chattrairat , Piyawan Leepheng, Songwut Suramitr and Darinee Phromyothin</i>	48
<b>Influence of Metal Species in Porphyrin-based Metal-organic Frameworks on Carbon Dioxide Capture</b> <i>Panchanit Piyakeeratikul, Bundet Boekfa, Sippakorn Wannakao, Taweesak Pila and Kanokwan Kongpatpanich</i>	55
<b>The Numerical Solution of Fractional Angiogenesis Problem by Meshless Local Petrov-Galerkin Method</b> <i>K. Phramrung, A. Luadsong, and N. Ascharyaphotha</i>	60
<b>Automatic Control for Water Droplet on Metal Surface</b> <i>Prasarnphun Saisin, Santi Wannippanto, Yongyut Pattanapong</i>	70
<b>The Meshless Local Petrov-Galerkin Method for Solving the New Black-Scholes-Schrodinger Model</b> <i>N. Nualsaard, A. Luadsong, and N. Ascharyaphotha</i>	78
<b>The Development of a VM Auto-Scaling Software for OpenStack Cloud</b> <i>Krailerk Manopattanagorn and Putchong Uthayopas</i>	87
<b>Application of Intensified Current Search to Design Optimal PIDD<sup>2</sup> Controller for BLDC Motor Speed Control with Back EMF Detection</b> <i>Danupon Kumpanya, and Chookiat Kiree</i>	93
<b>A Simple Numerical Model of an Active Magnetic Regenerator in Operation</b> <i>Pairat Tulyaprawat, Ratchatee Techapiesancharoenkij, Yuranan Hanlumyuang</i>	102
<b>System Tuning for Energy Efficient Big Data Infrastructure</b> <i>Soratouch Pornmaneerattanatri and Putchong Uthayopas</i>	110
<b>Express Lane Services on Software-Defined Networks</b> <i>Peeranon Wattanapong, Kittipat Jutawongcharoen, and Vara Varavithya</i>	115



## WELCOME MESSAGE

**Dr. Wannee Chinsirikul**

*Executive Director,  
National Nanotechnology Center (NANOTEC),  
National Science and Technology  
Development Agency (NSTDA), Thailand*

Welcome to Thailand!

It is my great pleasure to welcome you to the 21<sup>st</sup> International Annual Symposium on Computational Science and Engineering (ANSCSE21). ANSCSE serves as an annual international forum for computational scientists and engineers and is one of the largest gathering of Thai computational research community.

In late 2016, the Thai government promoted “Thailand 4.0” as the new economic model aimed at pulling Thailand out of the middle-income trap, and push the country in the high-income range. The model focuses on creating a Value-based Economy supported by three elements: High Income Nation, Inclusive Society, and Sustainable growth and development. Science, technology, and innovation will play a key role in driving the success of this model.

As you are aware, digital innovation is one of the fastest growing technology and business in this decade and one which has considerable impact on the world economy. For this reason, it is without a doubt that ANSCSE21 theme of “Computational Science and Big Data for Sustainable and Innovative Future” ties well with the emerging digital era and direction of Thailand 4.0 model. In so doing, this symposium will provide opportunities for researchers and scholars from private sectors, academia, and research institutes in the field of computational science and engineering to exchange technical information, stimulate ideas and foster collaboration for driving sustainable innovation and economy.

Again, Welcome and best wishes for a successful ANSCSE21, and thank you for your participation.

Dr. Wannee Chinsirikul

Executive Director of National Nanotechnology Center (NANOTEC)  
National Science and Technology Development Agency (NSTDA), Thailand



## WELCOME MESSAGE

**Asst. Prof. Dr. Putchong Uthayopas**

*Acting President of Computational Science and  
Engineering Association (CSEA),  
Thailand*

Dear Friends and Colleagues,

It is a great pleasure and an honor to extend to you a warm invitation to attend the ANSCSE21, the 21st International Annual Symposium on Computational Science and Engineering, to be held August 2-4, 2017. This year the symposium is organized by Nanoscale Simulation Laboratory of the National Nanotechnology Center (NANOTEC), Computational Science and Engineering Association (CSEA) and National e-Science Infrastructure Consortium. ANSCSE21 has always been one of the greatest gatherings of computational scientists, computer science, and engineering researchers. After 21 years, we have seen many signs of progress and so many interesting researches being conducted in this area. In this digital age, rapid progress has been driven by artificial intelligence, big data, and much higher computing power enabled by new technology such as GPU, FPGA. Thus, the vital role that computational science plays in human social development becomes clearer and clearer every day.

One of the great spirits of ANSCSE is the live discussion among fellow international researchers. After a few days of intense discussion on our works, the organizer kindly arranges an excursion to the ancient city of Ayutthaya, the UNESCO world heritage site. I am certain that everyone will enjoy the talk along with the rich history and the beauty of Ayutthaya.

Finally, I look forward to meeting all of you. Thank you for sharing your thoughts and ideas in ANSCSE21.

Best Wishes,

A handwritten signature in black ink, appearing to be 'Putchong Uthayopas', written in a cursive style.

Dr. Putchong Uthayopas, Kasetsart University

Acting President of Computational Science and Engineering Association (Thailand)



## WELCOME MESSAGE

### **Dr. Supawadee Namuangruk**

*Head of Nanoscale Simulation Laboratory,  
National Nanotechnology Center (NANOTEC),  
National Science and Technology Development  
Agency (NSTDA), Thailand*

Dear Colleagues,

On behalf of the organizing committee, I am honored and delighted to welcome you to the 21<sup>st</sup> International Annual Symposium on Computational Science and Engineering (ANSCSE21). It is a great honor for NANOTEC to be the host of ANSCSE21. Our co-hosts are Computer Science and Engineering Association, National e-Science Infrastructure Consortium and Department of Computer Science, Kasetsart University.

Over twenty years, ANSCSE has a long history of gathering researchers who are in the field of computational science and engineering to cross-fertilize ideas and to strengthen both local and international networks. The theme of this year is “Computational Science and Big Data for Sustainable and Innovative Future”. Under this theme, ANSCSE21 covers various disciplines of computational science and engineering including fields of Biology, Chemistry, Physics, Fluid Dynamics, Solid Mechanics, High Performance Computing, Cloud Computing, and Computer Science and Engineering.

There are 5 plenary talks, 38 invited talks and about 84 oral and poster presentations. This year, the scientific programs are accompanied with the Big Data Analytics workshop. This conference aims to provide an exciting venue for scientists to present and exchange ideas, as well as to renew existing collaborations and developing new ones. ANSCSE21 also offers the great opportunity for academics and industrial people to meet and exchange ideas and information.

As a conference chair of ANSCSE21, I would like to express my appreciation to the steering committee, the honorary chairs, the international advisory board, the scientific committee chair, the program chairs, the scientific committee, the reviewers, our sponsors and the organizing team. Last but not the least; recognition should also go to the local organizing committee members who have worked extremely hard in planning and organizing the technical programs and supporting social arrangements.

Finally, ANSCSE21 truly serves the venue for networking and knowledge sharing as a consequence of the comprehensive presentations as well as high-level plenary and panel sessions. We hope you will take the utmost advantage of this event to explore new innovation.

Sincerely yours,

A handwritten signature in blue ink, appearing to read 'Dr. Supawadee Namuangruk', with a stylized flourish at the end.

Dr. Supawadee Namuangruk

Conference Chair,  
Nanoscale Simulation Laboratory  
National Nanotechnology Center (NANOTEC)  
National Science and Technology Development Agency (NSTDA), Thailand

## COMMITTEE

### Steering Committee Chair

Asst. Prof. Dr. Putchong Uthayopas

*Kasetsart University and Acting President of Computational Science and Engineering Association (CSEA), Thailand*

Assoc. Prof. Dr. Vudhichai Parasuk

*Chulalongkorn University and President-elect of Computational Science and Engineering Association (CSEA), Thailand*

Prof. Dr. Supa Hannongbua

*Dean of Faculty of Science, Kasetsart University, Thailand*

### Honorary Chair

Dr. Wannee Chinsirikul

*Executive Director, National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Thailand*

Dr. Uracha Ruktanonchai

*Deputy Executive Director Research & Development, National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Thailand*

Dr. Kajornsak Faungnawakij

*Director of Nanomaterials and Nanosystems Engineering Research Unit, National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Thailand*

### Conference Organizers

National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Thailand

Computational Science and Engineering Association (CSEA), Thailand

National E-Science Infrastructure Consortium, Thailand

Department of Computer Science, Faculty of Science, Kasetsart University, Thailand

### **Conference Organizing Committee Chair**

Dr. Supawadee Namuangruk

*Head of Nanoscale Simulation Laboratory, National Nanotechnology Center (NANOTEC),  
National Science and Technology Development Agency (NSTDA), Thailand*

### **International Advisory Board**

Prof. Dr. Peter Wolschann

*University of Vienna, Austria*

Prof. Dr. Siegfried Fritzsche

*University of Leipzig, Germany*

Prof. Dr. Masahiro Ehara

*Institute for Molecular Science, Japan*

Prof. Dr. Tetsuya Taketsugu

*Hokkaido University, Japan*

Dr. Ito Chao

*Academia Sinica, Taiwan*

Prof. Dr. Aiichiro Nakano

*University of Southern California, USA*

### **Scientific Committee Chair**

Assoc. Prof. Dr. Siriporn Jungsuttiwong

*Ubon Ratchathani University, Thailand*

### **Technical Program Committee Chair**

#### **Computational Chemistry Chair:**

Assoc. Prof. Dr. Pornpan Pungpo

*Ubon Ratchathani University, Thailand*

#### **Computational Physics Chair:**

Assoc. Prof. Dr. Anucha Yangthaisong  
*Ubon Ratchathani University, Thailand*

Asst. Prof. Dr. Worasak Sukkabot  
*Ubon Ratchathani University, Thailand*

**Computational Biology and Bioinformatics Chair:**

Asst. Prof. Dr. Thanyada Rungrotmongkol  
*Chulalongkorn University, Thailand*

**Computational Fluid Dynamics and Solid Mechanics Chair:**

Assoc. Prof. Dr. Vejapong Juttijudata  
*Kasetsart University, Thailand*

**High Performance Computing and Cloud Computing Chair:**

Asst. Prof. Dr. Putchong Uthayopas  
*Kasetsart University, Thailand*

**Computer Science and Engineering Chair:**

Assoc. Prof. Dr. Vara Varavithya  
*King Mongkut's University of Technology North Bangkok, Thailand*

Dr. Supakit Prueksaaron  
*Thammasat University, Thailand*

**International Scientific committee**

Prof. Dr. Shinji Saito  
*Institute for Molecular Science, Japan*

Assoc. Prof. Dr. Hisashi Okumura  
*Institute for Molecular Science, Japan*

Asst. Prof. Dr. Satoru Itoh  
*Institute for Molecular Science, Japan*

Prof. Dr. Jun-ya Hasegawa  
*Hokkaido University, Japan*

Assoc. Prof. Dr. Akira Nakayama  
*Hokkaido University, Japan*

Asst. Prof. Dr. Min Gao  
*Hokkaido University, Japan*

Dr. Manussada Ratanasak  
*Hokkaido University, Japan*

Assoc. Prof. Dr. Norio Yoshida  
*Kyushu University, Japan*

Assoc. Prof. Dr. Vannajan Sanghiran Lee  
*University of Malaya, Malaysia*

Prof. Dr. Jen-Shiang K. Yu  
*National Chiao Tung University, Taiwan*

Dr. Cherri Chao-Ping Hsu  
*Academia Sinica, Taiwan*

Dr. Kaito Takahashi  
*Academia Sinica, Taiwan*

Prof. Dr. Jyh-Chiang Jiang  
*National Taiwan University of Science and Technology, Taiwan*

Prof. Dr. Lam K. Huynh  
*Vietnam National University, Vietnam*

Prof. Dr. Ras B. Pandey  
*The University of Southern Mississippi, USA*

### **Scientific committee**

Prof. Dr. Supa Hannongbua  
*Kasetsart University, Thailand*

Dr. Chalee Vorakulpipat  
*National Electronics and Computer Technology Center (NECTEC), NSTDA, Thailand*

Asst. Prof. Dr. Duangkamol Tantanak  
*King Mongkut's Institute of Technology Ladkrabang, Thailand*

Asst. Prof. Dr. Nawee Kungwan  
*Chiang Mai University, Thailand*

Asst. Prof. Dr. Bundet Boekfa  
*Kasetsart University, Thailand*

Asst. Prof. Dr. Jakkapan Sirijaraensre  
*Kasetsart University, Thailand*

Asst. Prof. Dr. Patchreenart Saparpakorn  
*Kasetsart University, Thailand*

Asst. Prof. Dr. Songwut Suramitr  
*Kasetsart University, Thailand*

Asst. Prof. Dr. Chan Inntam

*Ubon Ratchathani University, Thailand*

Dr. Piti Treesukul

*Kasetsart University, Thailand*

Asst. Prof. Dr. Supaporn Dokmaisrijan

*Walailak University, Thailand*

Dr. Pornthip Boonsri

*Srinakharinwirot University, Thailand*

Dr. Potjaman Poolmee

*Kasetsart University Kamphaengsaen Campus, Thailand*

Dr. Malinee Promkatkaew

*Kasetsart University Si Racha Campus, Thailand*

Assoc. Prof. Dr. Duangkamol Gleeson

*King Mongkut's Institute of technology Ladkrabang, Thailand*

Assoc. Prof. Dr. PhonThep Somphonphisut

*Chulalongkorn University, Thailand*

Asst. Prof. Dr. Panida Kongsawadworakul

*Mahidol University, Thailand*

Asst. Prof. Dr. Khatcharin Siritwong

*Khon Kaen University, Thailand*

Asst. Prof. Dr. Somsak Pianwanit

*Chulalongkorn University, Thailand*

Asst. Prof. Dr. Nadtanet Nunthaboot

*Maharakham University, Thailand*

Dr. Varomyalin Tipmanee

*Prince of Songkla University, Thailand*

Assoc. Prof. Dr. Udomsilp Pinsook

*Chulalongkorn University, Thailand*

Dr. Sittipong Komin

*Ubon Ratchathani University, Thailand*

Asst. Prof. Dr. Kanchana Savalertporn

*Ubon Ratchathani University, Thailand*

Assoc. Prof. Dr. Teparksorn Pengpan

*Prince of Songkla University, Thailand*

Dr. Pruet Kalasuwan

*Prince of Songkla University, Thailand*

Dr. Jessada Chureemart

*Maharakham University, Thailand*

Dr. Sujin Suwana

*Mahidol University, Thailand*

Dr. Naruemon Pratanwanich

*Chulalongkorn University, Thailand*

Uthumpon Arsawang

*University of Leipzig, Germany*

Dr. Teeraphong Phongpattanacharoen

*Chulalongkorn University, Thailand*

Dr. Arthitaya Meeprasert

*Chulalongkorn University, Thailand*

Dr. Saranya Maneeroj

*Chulalongkorn University, Thailand*

Dr. Supawadee Namuangruk

*NANOTEC, Thailand*

Dr. Thana Maihom

*Kasetsart University, Thailand*

Dr. Manaschai Kunaseth

*NANOTEC, Thailand*

Dr. Yuranan Hanlumyuang

*Kasetsart University, Thailand*

Dr. Chompoonut Rungnim

*NANOTEC, Thailand*

Dr. Tanin Nanok

*Kasetsart University, Thailand*

Dr. Rathawat Daengngern

*King Mongkut's Institute of Technology Ladkrabang, Thailand*

Jittima Meeprasert

*NANOTEC, Thailand*

Dr. Damrongrit Setsirichok

*King Mongkut's University of Technology North Bangkok, Thailand*

Dr. Komsilp Kotmool

*Mahidol Wittayanusorn School, Thailand*

**Organizing committee**

Dr. Supawadee Namuangruk

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Chalee Vorakulpipat

*National Electronics and Computer Technology Center (NECTEC), NSTDA, Thailand*

Asst. Prof. Nawe Kungwan

*Chiang Mai University, Thailand*

Asst. Prof. Phornphimon Maitarad

*Shanghai University, China*

Tortrakul Poolsopha

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Sasinee Watcharaput

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Benyapa Suwan

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Wipaporn Ekamornthanakul

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Channarong Prommakan

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Sukunya Phramkhao

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Chayuda Karaphol

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Yaninee Khantapat

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Pongsit Rattanakonvit

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Jariyamas Amornprapaipis

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Amornpan Koysungnoen

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Panee Santiveeraphan

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Worawut Khoonkrong

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Chompoonut Rungnim

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Pussana Hirunsit

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Manaschai Kunaseth

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Monrudee Liangruksa

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Dr. Anchalee Junkaew

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Jittima Meeprasert

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

Chirawat Chitpakdee

*National Nanotechnology Center (NANOTEC), NSTDA, Thailand*

## PLENARY LECTURES



### **PL-1: Utilizing IBM Power Systems to Optimize Genomic Applications Compute**

**Denise Ruffner**

*Business Development Executive, Worldwide Genomic Solutions, IBM, USA*



### **PL-2: Structure and Catalytic Activity of Nanocluster Catalysts**

**Prof. Dr. Masahiro Ehara**

*Institute for Molecular Science, Japan  
SOKENDAI, the Graduate University for Advanced Studies,  
Japan*

*Element Strategy Initiative for Catalysts and Batteries, Kyoto  
University, Japan*



### **PL-3: Extreme-scale Atomistic Simulations of Nanomaterials**

**Prof. Dr. Aiichiro Nakano**

*University of Southern California, Los Angeles, CA, USA*



**PL-4: Chemistry and Our Common Future**

**Dr. Ito Chao**

*Institute of Chemistry, Academia Sinica, Taipei, Taiwan*



**PL-5: Designing a Massive Coil to Shield JUNO from the Earth's Magnetic Field**

**Prof. Dr. Sukit Limpijumnong**

*Suranaree University of Technology, Thailand*

## INVITED SPEAKERS

### Computational Physics and Fluid Dynamics

PFD-I-1	Arthit	Vongachariya	Siam Cement Chemicals	Thailand
PFD-I-2	Somboon	Otarawanna	MTEC, NSTDA	Thailand
PFD-I-3	Siegfried	Fritzsche	University of Leipzig	Germany
PFD-I-4	Sittipong	Komin	Ubon Ratchathani University	Thailand
PFD-I-5	Jessada	CHUREEMART	Maharakham University	Thailand
PFD-I-6	Phanwadee	Chureemart	Maharakham University	Thailand
PFD-I-7	Norraphat	SRIMANOBHAS	Chulalongkorn University	Thailand
PFD-I-8	Teparksorn	Pengpan	Prince of Songkla University	Thailand
PFD-I-9	Komsilp	Kotmool	Mahidol Wittayanusorn School	Thailand
PFD-I-10	Piti	Ongmongkolkul	Mahidol University International College	Thailand

### Computational Chemistry

CHE-I-1	Tetsuya	Taketsugu	Hokkaido University	Japan
CHE-I-2	Panida	Surawatanawong	Mahidol University	Thailand
CHE-I-3	Minh Tho	Nguyen	KU Leuven	Belgium
CHE-I-4	Jun-ya	Hasegawa	Hokkaido University	Japan
CHE-I-5	Kaito	Takahashi	Academia Sinica	Taiwan
CHE-I-6	Jyh-Chiang	Jiang	National Taiwan University of Science and Technology	Taiwan
CHE-I-7	Arkira	Nakayama	Hokkaido University	Japan
CHE-I-8	Chao-Ping	Hsu	Academia Sinica	Taiwan
CHE-I-9	Vinich	Promarak	Vidyasirimedhi Institute of Science and Technology	Thailand
CHE-I-10	Lam	Hunynh	Vietnam National University	Vietnam
CHE-I-11	Tim	Kowalczyk	Western Washington University	USA
CHE-I-12	Min	Gao	Hokkaido University	Japan
CHE-I-13	Manussada	Ratanasak	Hokkaido University	Japan

### Computational Biology and Bioinformatics

BIO-I-1	Ras	Pandey	The University of Southern Mississippi	USA
BIO-I-2	Norio	Yoshida	Kyushu University	Japan
BIO-I-3	Hisashi	Okumura	Institute for Molecular Science	Japan
BIO-I-4	Kiattawee	Choowongkomon	Kasetsart University	Thailand
BIO-I-5	Jen-Shiang	Yu	National Chiao Tung University	Taiwan
BIO-I-6	Satoru	Itoh	Institute for Molecular Science	Japan
BIO-I-7	Shinji	Saito	Institute for Molecular Science	Japan
BIO-I-8	Varomyalin	Tipmanee	Prince of Songkla University	Thailand
BIO-I-9	Peter	Wolschann	University of Vienna	Austria
BIO-I-10	Vannajan	Lee	University of Malaya	Malaysia
BIO-I-11	Syed Sikander	Azam	Quaid-i-Azam University Islamabad	Pakistan
BIO-I-12	Sissades	Tongsima	BIOTEC, NSTDA	Thailand

### High Performance Computing, Cloud Computing, Computer Science and Engineering

CSE-I-1	Zong-Yao	Chen	ACER	Taiwan
CSE-I-2	Ekasit	Kijsipongse	NECTEC, NSTDA	Thailand
CSE-I-3	Waranyu	Wongseree	King Mongkut University of Technology North Bangkok	Thailand
CSE-I-4	Chantana	Chantrapornchai	Kasetsart University	Thailand



# Effects of Incorporating Genetic Models into a Genetic Programming Tree Ensemble for Genetic Association Studies

P. Aksornsingchai<sup>1</sup>, K. Jutawongcharoen<sup>1</sup>, W. Wongseree<sup>1</sup>, N. Chaiyaratana<sup>1,2</sup>, and D. Setsirichok<sup>1,C</sup>

<sup>1</sup> *Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

<sup>2</sup> *Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*

<sup>C</sup> **E-mail:** d.setsirichok@gmail.com; **Fax:** +66 2 585 7350; **Tel.** +66 2 555 2000 x8421

## ABSTRACT

This article presents the effects of incorporating genetic models into a tree ensemble which was used for identifying informative single nucleotide polymorphisms (SNPs) in genetic association studies. Each tree in the ensemble was evolved by genetic programming (GP) where the function set consisted of genetic models, genetic-Boolean models and Boolean functions while the terminal set consisted of SNPs. A bootstrap aggregating or bagging approach was employed during the tree construction and classification accuracy evaluation. Therefore, it is possible to measure variable importance of each SNP. The genetic programming tree ensemble (GPTE) was benchmarked against a random forest (RF) in various simulation settings involving jointly recessive, single-locus recessive, semi-pure epistasis and pure epistasis models. The results indicated that GPTE outperformed RF in terms of the ability to identify causative SNPs when the difference between penetrances of disease-predisposing and protective genotypes is small. This suggests that incorporating genetic models into a tree ensemble is beneficial since the aforementioned scenario is close to the real scenario in genetic association studies.

**Keywords:** Case-Control Study, Ensemble of Classifiers, Genetic Association Study, Genetic Programming, Random Forest, Single Nucleotide Polymorphism.

## 1. INTRODUCTION

With the completion of the Human Genome Project, it is now possible to locate over 3,000,000 single nucleotide polymorphisms (SNPs), which are common genetic markers, in the human genome [1, 2]. This poses one of the biggest challenges for genetic epidemiologists, which is to identify SNPs that are associated with complex diseases [3]. Nowadays, a genetic association study is treated as the first step to identify the causes of a complex disease of interest. In a population-based study, case-control samples are drawn from a population where SNP information is then extracted through genotyping. The difference between genotypic distribution among case samples and that among control samples usually signifies that SNPs are associated with the disease. In other words, the study aims to identify informative SNPs among available SNPs that lead to case-control classification [4].

From a pattern recognition viewpoint, the identification of informative SNPs can be treated as an attribute selection problem. This means that all three attribute selection approaches—filter, wrapper and embedded approaches—are suitable for the task [5]. The filter approach interests in detecting the genotypic distribution difference through a statistical or mathematical measure. The wrapper approach searches for the best subset of available SNP set that produces the highest classification accuracy according to a chosen classifier. The embedded approach integrates attribute selection into the classifier construction.

Among the embedded techniques, a technique which is proven to be highly capable of identifying informative SNPs is a random forest (RF). RF is essentially an ensemble of decision trees. Traditionally, a decision tree is deterministically constructed by successively choosing an attribute which provides the best means for classifying samples at each node. However, each tree in RF is randomly constructed where the attribute for each node is stochastically selected from available attributes. A permutation of attribute values and the observation of its effect on the classification accuracy of RF can then be carried out for attribute selection [6]. RF was successfully applied to a number of genetic association studies which cover the presence of epistasis or gene-gene interactions and genetic heterogeneity where the same disease is caused by multiple independent genetic factors [7, 8].

Although RF is a suitable technique for genetic association studies, it is impossible to directly incorporate genetic models into RF. Many traditional classifiers including logistic regression are capable of exploiting genetic models [9]. Therefore, it is interesting to study the effects of incorporating genetic models into a tree ensemble. One possibility is to use genetic programming (GP) to generate each tree in the ensemble. GP is a variant of genetic algorithm that is specifically designed to evolve a tree [10]. Each GP tree therefore inherits an ability to perform stochastic attribute selection. A genetic model, a genetic-Boolean model or a Boolean function, which can be chosen from a predefined set, is assigned to each non-leaf node in a GP tree. A benchmark experiment in which both RF and GP tree ensemble (GPTE) are used as classifiers where selected attributes are subsequently compared would provide an answer to the question regarding the necessity of genetic models.

## 2. METHODS

### 2.1. Disease model

The benchmark experiment was carried out using a simulation. Causative or informative SNPs in each simulated dataset were governed by a two-locus disease model. The model is a penetrance-based model and defines the probability that an individual with a specific genotype has the disease [11]. Consider a two-locus disease model consisting of loci A and B.  $A$  and  $a$  respectively denote the major and minor alleles at locus A. Similarly,  $B$  and  $b$  respectively denote the major and minor alleles at locus B. At each locus, 0, 1 and 2 respectively represent a homozygous wild-type genotype ( $AA$  and  $BB$ ), a heterozygous genotype ( $Aa$  and  $Bb$ ) and a homozygous variant genotype ( $aa$  and  $bb$ ). The penetrance of genotype  $i$  at locus A and genotype  $j$  at locus B is defined by  $f_{ij} \in [0,1]$ . Therefore, the marginal penetrance  $M_{Ai}$  for genotype  $i$  at locus A is

$$M_{Ai} = p_B^2 f_{i0} + 2p_B(1-p_B)f_{i1} + (1-p_B)^2 f_{i2}, i \in \{0,1,2\} \quad (1)$$

and the marginal penetrance  $M_{Bj}$  for genotype  $j$  at locus B is

$$M_{Bj} = p_A^2 f_{0j} + 2p_A(1-p_A)f_{1j} + (1-p_A)^2 f_{2j}, j \in \{0,1,2\} \quad (2)$$

where  $p_A$  and  $p_B$  are the major allele frequencies. Penetrances and marginal penetrances for a two-locus disease model are shown in Table 1. Four models for the simulation—M1, M7, M97 and M170—were taken from the collection of two-locus disease models described in Li and Reich [12]. Penetrances for these models are given in Table 2. The description of these models follows.

**Table 1.** Penetrances and marginal penetrances for a two-locus disease model.

Genotype	Penetrance of Genotype			Marginal Penetrance
	$BB$	$Bb$	$bb$	
$AA$	$f_{00}$	$f_{01}$	$f_{02}$	$M_{A0}$
$Aa$	$f_{10}$	$f_{11}$	$f_{12}$	$M_{A1}$
$aa$	$f_{20}$	$f_{21}$	$f_{22}$	$M_{A2}$
Marginal Penetrance	$M_{B0}$	$M_{B1}$	$M_{B2}$	

**Table 2.** Penetrances for disease models M1, M7, M97 and M170.

Genotype	Penetrance for Disease Model			
	M1	M7	M97	M170
<i>AABB</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$
<i>AABb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$\alpha$
<i>AAbb</i>	$1 - \alpha$	$1 - \alpha$	$\alpha$	$1 - \alpha$
<i>AaBB</i>	$1 - \alpha$	$1 - \alpha$	$\alpha$	$\alpha$
<i>AaBb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$
<i>Aabb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$\alpha$
<i>aaBB</i>	$1 - \alpha$	$\alpha$	$1 - \alpha$	$1 - \alpha$
<i>aaBb</i>	$1 - \alpha$	$\alpha$	$1 - \alpha$	$\alpha$
<i>aabb</i>	$\alpha$	$\alpha$	$\alpha$	$1 - \alpha$

$\alpha \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$ . Genotypes with penetrance =  $\alpha$  are disease-predisposing genotypes while genotypes with penetrance =  $1 - \alpha$  are protective genotypes.

M1 is a jointly recessive model. In other words, an individual needs to have homozygous variant genotypes at both loci to have an increased risk of being an affected individual. M7 is a single-locus recessive model. Allele *a* is the only disease allele. Therefore, locus B does not contribute towards disease susceptibility. M97 is a symmetric model with respect to the permutation of alleles *A* and *a*. The marginal penetrances  $M_{A_0}$ ,  $M_{A_1}$  and  $M_{A_2}$  for genotypes *AA*, *Aa* and *aa*, respectively are equal when the major allele frequency  $p_B = 0.5$ . As a result, the SNP at locus A exhibits no marginal single-locus effect. This makes it impossible to detect the causative SNP at locus A by single-locus analysis. M97 is an example of semi-pure epistasis models [13]. M170 is an interference model. Essentially, it is neither a dominant nor recessive model. The marginal penetrances  $M_{A_0}$ ,  $M_{A_1}$ ,  $M_{A_2}$ ,  $M_{B_0}$ ,  $M_{B_1}$  and  $M_{B_2}$  for genotypes *AA*, *Aa*, *aa*, *BB*, *Bb* and *bb*, respectively are equal when the major allele frequencies  $p_A = p_B = 0.5$ . As a result, the SNPs at both loci A and B exhibit no marginal single-locus effects and cannot be detected by single-locus analysis. M170 is an example of pure epistasis models [11].

## 2.2. Random forest

RF is an ensemble of decision trees [6]. Each tree undergoes a top-down construction. This begins with the root node where an attribute is selected as the test. Descendant nodes are then created according to the value of attribute used as the test. Subsequently, samples are sent to the appropriate descendant nodes. The process of descendant node construction and sending samples to descendant nodes is repeated until the tree size is maximised. Unlike most decision trees that attribute selection for the test at each node is deterministic, an attribute for each node in a tree in RF is selected for its suitability for use as the test from a group of randomly picked attributes. Empirical studies showed that  $\lceil \sqrt{\text{total number of attributes}} \rceil$  attributes are sufficient for the group. This leads to diversity among trees. As a result, the class decision is dictated by the majority vote among trees. Since the class decision is obtained in this manner, allowing each tree to grow to its maximum size does not lead to data over-fitting.

Classification accuracy of RF is generally evaluated using a bootstrap aggregating or bagging approach. A bootstrap sample set, which has the same size as the original sample set, is created by sampling from the original sample set with replacement. Approximately 36.8% of bootstrap samples are duplicates while the remaining samples are unique. Original samples that are absent from bootstrap samples are out-of-bag samples. Bootstrap samples are used to construct a tree while out-of-bag samples are used to evaluate classification accuracy. A new bootstrap sample set is created for each tree. Therefore, the votes from trees that the interesting sample is one of their out-of-bag samples are counted during classification accuracy evaluation.

The evaluation of classification accuracy using a bagging approach also leads to a means to evaluate attribute importance or variable importance. It is achieved through a permutation approach. Basically, the value of the interesting attribute is randomly permuted. The permuted attribute and the remaining non-permuted attributes are then used as inputs of RF where the classification accuracy is re-evaluated. The disruption of correlation between the interesting attribute and the class is subsequently determined. If the interesting attribute is highly correlated with the class or is important for the classification, then the disruption makes the classification accuracy decrease. The average difference between the classification accuracy obtained for the original samples and that obtained for the samples with one permuted attribute from all trees is the variable importance [7]. It is defined as

$$I_T(Attr) = \frac{1}{T} \sum_{j=1}^T \frac{1}{N_j} \sum_{i=1}^N [1(V_j(s_i) = c_i) - 1(V_j(s_i^{Attr,j}) = c_i)] t_{ij}. \quad (3)$$

$I_T(Attr)$  is the variable importance of attribute  $Attr$ .  $s_i$  is the original sample  $i$ .  $s_i^{Attr,j}$  is the sample  $i$  with the value of attribute  $Attr$  randomly permuted among the out-of-bag samples for tree  $j$ .  $V_j$  is the vote from tree  $j$ .  $c_i$  is the true class for sample  $i$ .  $1(C)$  denotes the indicator function that returns value 1 when the condition  $C$  is true and returns value 0 when the condition  $C$  is false.  $t_{ij}$  is the indicator that has value 1 when sample  $i$  is an out-of-bag sample for tree  $j$  and has value 0 when sample  $i$  is a bootstrap sample for tree  $j$ .  $N_j$  is the number of out-of-bag samples for tree  $j$ ,  $N$  is the number of samples and  $T$  is the number of trees. The variable importance follows a normal distribution. Therefore, the variable importance can be standardised by dividing it with its standard error, which is derived from the variance of the index calculated from each tree. The standardised variable importance  $Z_T(Attr)$  of attribute  $Attr$  is defined as

$$Z_T(Attr) = I_T(Attr) / \sqrt{\sigma_T^2(Attr)/T} \quad (4)$$

where  $\sigma_T^2(Attr)$  is the variance of variable importance  $I_T(Attr)$ .

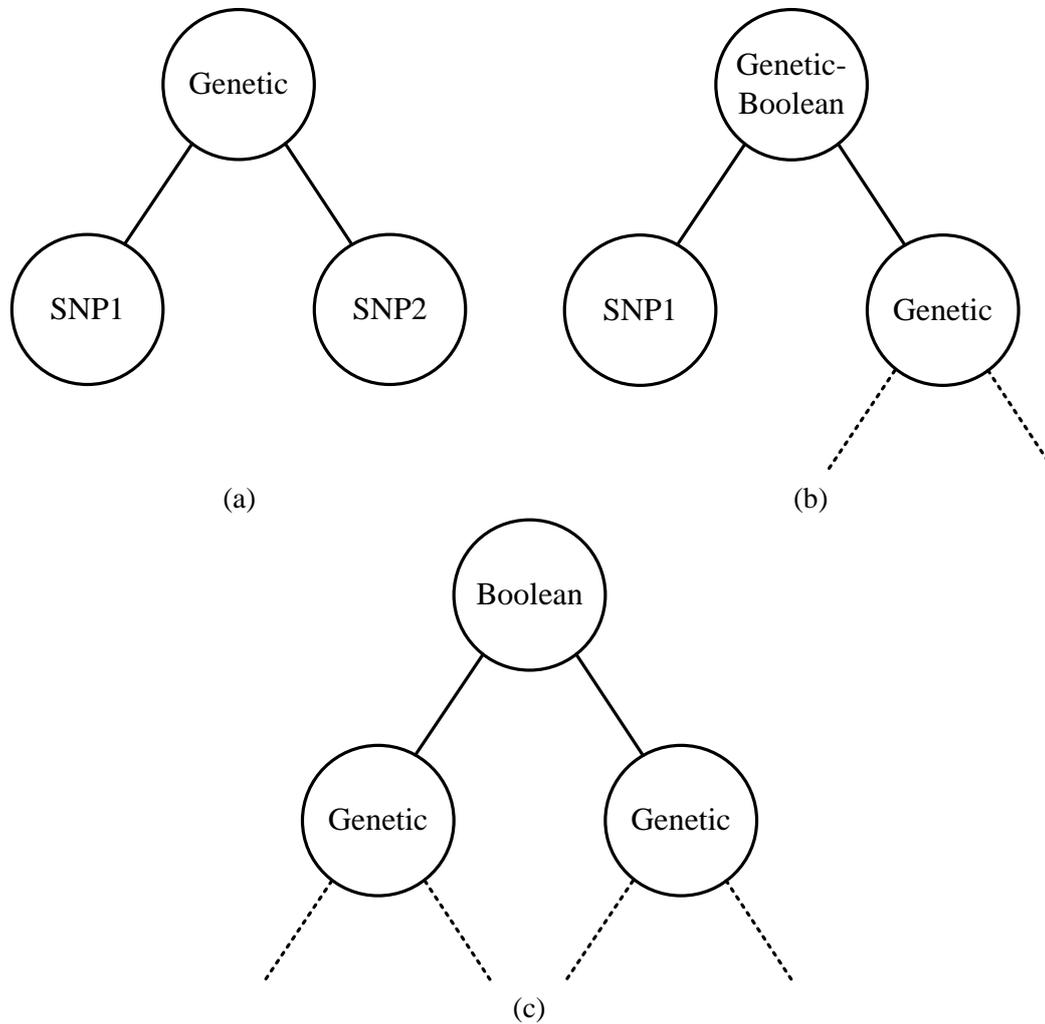
Strobl et al. [14] suggested that informative attributes can be selected according to their variable importance as follows. It was observed that the variable importance of irrelevant or random attributes varies around zero. As a result, attributes with positive variable importance in which the index magnitude is comparable to that of attributes with negative variable importance are not informative. Hence, all attributes with positive variable importance that exceeds the aforementioned range of magnitude should be considered as being informative. In this study, the above attribute selection criterion was applied to standardised variable importance. The number of trees in RF was set to 100.

### 2.3. Genetic programming tree ensemble

GP constructs a tree by means of evolutionary computation [10]. This begins by randomly initialising a population of trees. The fitness of each tree is evaluated according to its suitability for the task at hand. For classification, fitness is generally based on classification accuracy. Each tree is then selected for use as a parent according to its fitness. Offspring are generated by crossing over through exchanging subtrees between parents and mutating nodes or subtrees. After a population of next-generation trees is created, the process of fitness evaluation, selection, crossover and mutation can be repeated until the predefined number of generations is reached.

In this study, a GP tree is a binary tree; the ensemble of trees was then used as a classifier. Consider a binary tree in Figure 1a. The tree takes two inputs, which are SNPs. Its output is either 0 or 1 depending on the genotypic inputs and the mapping function. The mapping function converts two inputs where each input has one of three possible values into a binary output. Therefore, there are  $2^9 = 512$  mapping functions that cover all possible two-locus genetic models. The binary tree in Figure 1b takes two inputs: one input is a SNP and the other is the output of its right subtree. The mapping function thus converts two inputs—one input with three possible values and the other with two possible values—into a binary output.

Therefore, there are  $2^6 = 64$  mapping functions that cover all possible genetic-Boolean models. The binary tree in Figure 1c takes two inputs, which are the outputs of its subtrees. The mapping function converts two binary inputs into a binary output. Therefore, there are  $2^4 = 16$  mapping functions that cover all possible Boolean functions.



**Figure 1.** Mapping function. (a) Genetic model. (b) Genetic-Boolean model. (c) Boolean function.

**Table 3.** Parameter setting for GPTE.

Parameter	Value
Number of trees in the ensemble	100
Maximum tree height	2
Minimum tree height	1
Selection technique	Stochastic universal sampling
Number of elitist individuals	2
Crossover probability	0.7
Mutation probability	0.1
Size of the function set	$512 + 64 + 16 = 592$
Size of the terminal set	20
Population initialisation method	Ramped half-and-half method
Population size	200
Number of generations	25

Similar to RF, classification accuracy of GPTE can be evaluated using a bagging approach. This means that the evolution of each tree in the ensemble has to take place independently. A bootstrap sample set and an out-of-bag sample set were first created. Then a population of trees was constructed from the bootstrap sample set and was evolved through evolutionary computation. The fittest tree was subsequently selected for the ensemble. Naturally, different bootstrap and out-of-bag sample sets were created for different trees in the ensemble. Since the bagging approach was used for classification accuracy evaluation, variable importance can be estimated. The attribute selection criterion described in section 2.2 was also employed. The parameter setting for GPTE is given in Table 3.

### 3. RESULTS AND DISCUSSION

GPTE was benchmarked against RF in a simulation trial involving disease models described in section 2.1. The interesting differences between penetrances of disease-predisposing and protective genotypes were 0.2, 0.4, 0.6, 0.8 and 1.0. Each simulated dataset consisted of 20 unlinked SNPs where a disease model was present. The allele frequencies of both causative SNPs were 0.5 while the minor allele frequencies (MAFs) of the remaining SNPs were between 0.05 and 0.5. They conform to the MAFs of SNPs interested by the International HapMap Project [15]. The dataset consisted of 400 balanced case-control samples. All SNPs in control samples were in Hardy-Weinberg equilibrium. One hundred independent datasets for each simulation setting were generated by genomeSIM [16]. It was shown in Wongseree et al. [13] and Setsirichok et al. [17] that an efficient algorithm reports a low number of output SNPs and a high number of correctly-identified causative SNPs. Therefore, the number of output SNPs and the number of correctly-identified causative SNPs were used as the performance indicators. A paired *t*-test was applied to results obtained for each simulation setting to assess the significance of difference in algorithm performance.

The results from all simulation settings are given in Tables 4–7. The number of output SNPs reported by GPTE was not different from that reported by RF in the majority of simulation settings. There were six simulation settings with small difference between penetrances of disease-predisposing and protective genotypes where GPTE reported more output SNPs than RF ( $p < 0.05$ ). Similarly, the number of correctly-identified causative SNPs

**Table 4.** Performance of RF and GPTE in the datasets with disease model M1.

$\alpha$	Number of Output SNPs			Number of Correctly-Identified Causative SNPs		
	RF	GPTE	<i>p</i> -value	RF	GPTE	<i>p</i> -value
0.6	1.52 (2.34)	2.79 (1.27)	< 0.0001	0.52 (0.59)	2.00 (0.00)	< 0.0001
0.7	1.93 (2.37)	2.58 (1.88)	0.0328	0.68 (0.65)	0.96 (0.76)	0.0058
0.8	3.13 (2.40)	3.33 (1.76)	0.5035	1.71 (0.54)	1.83 (0.38)	0.0127
0.9	3.17 (1.85)	3.11 (1.58)	0.8055	2.00 (0.00)	2.00 (0.00)	-
1.0	3.34 (1.83)	2.90 (1.23)	0.0476	2.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 5.** Performance of RF and GPTE in the datasets with disease model M7.

$\alpha$	Number of Output SNPs			Number of Correctly-Identified Causative SNPs		
	RF	GPTE	<i>p</i> -value	RF	GPTE	<i>p</i> -value
0.6	1.96 (2.38)	1.98 (1.28)	0.9411	0.66 (0.48)	1.00 (0.00)	< 0.0001
0.7	2.26 (1.69)	2.62 (1.30)	0.0940	0.99 (0.10)	1.00 (0.00)	0.3197
0.8	2.10 (1.92)	2.25 (1.38)	0.5267	1.00 (0.00)	1.00 (0.00)	-
0.9	2.41 (2.11)	2.04 (1.20)	0.1290	1.00 (0.00)	1.00 (0.00)	-
1.0	2.25 (1.91)	2.15 (1.34)	0.6683	1.00 (0.00)	1.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 6.** Performance of RF and GPTE in the datasets with disease model M97.

$\alpha$	Number of Output SNPs			Number of Correctly-Identified Causative SNPs		
	RF	GPTE	<i>p</i> -value	RF	GPTE	<i>p</i> -value
0.6	1.58 (1.91)	3.13 (1.53)	< 0.0001	0.60 (0.62)	2.00 (0.00)	< 0.0001
0.7	2.74 (2.04)	3.47 (1.53)	0.0048	1.34 (0.61)	1.94 (0.24)	< 0.0001
0.8	3.08 (1.77)	3.26 (1.27)	0.4101	1.98 (0.14)	2.00 (0.00)	0.1583
0.9	3.23 (1.80)	2.98 (1.45)	0.2812	2.00 (0.00)	2.00 (0.00)	-
1.0	3.06 (1.45)	3.00 (1.26)	0.7546	2.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 7.** Performance of RF and GPTE in the datasets with disease model M170.

$\alpha$	Number of Output SNPs			Number of Correctly-Identified Causative SNPs		
	RF	GPTE	<i>p</i> -value	RF	GPTE	<i>p</i> -value
0.6	1.37 (2.30)	3.78 (2.44)	< 0.0001	0.48 (0.61)	2.00 (0.00)	< 0.0001
0.7	2.47 (2.77)	3.43 (1.40)	0.0023	0.91 (0.73)	2.00 (0.00)	< 0.0001
0.8	3.21 (1.95)	3.30 (1.43)	0.7104	1.93 (0.26)	2.00 (0.00)	0.0075
0.9	3.21 (1.98)	2.85 (1.12)	0.1159	2.00 (0.00)	2.00 (0.00)	-
1.0	3.03 (1.60)	3.08 (1.39)	0.8137	2.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

reported by GPTE was not different from that reported by RF when there is a large difference between penetrances of disease-predisposing and protective genotypes. However, GPTE was capable of reporting more correctly-identified causative SNPs when the difference between penetrances of disease-predisposing and protective genotypes is small ( $p < 0.05$ ). The simulation results can be further interpreted as follows.

Marginal single-locus effects play an important role in the ability to detect causative SNPs. It is possible to detect causative SNPs by single-locus analysis when they exhibit large marginal single-locus effects. This implies that the detection task is not difficult. On the other hand, multi-locus analysis is required when causative SNPs exhibit small or no marginal single-effects. This is because detecting each causative SNP independently is either difficult or impossible. The marginal single-locus effects of both SNPs in disease model M1, the SNP at locus A in disease model M7 and the SNP at locus B in disease model M97 depend on the value of  $\alpha$ . Decreasing  $\alpha$  leads to the decreasing of marginal single-locus effects at these loci. It was observed that GPTE detected more causative SNPs than RF when the values of  $\alpha$  are 0.6 and 0.7. This suggests that incorporating genetic models into a tree ensemble helps improving the ability to detect causative SNPs when marginal single-locus effects are small. Nonetheless, this comes at a price of increasing the possibility of reporting more erroneous SNPs in the output.

#### 4. CONCLUSION

In this article, the effects of incorporating genetic models into a tree ensemble for detecting informative SNPs in genetic association studies were investigated. The tree ensemble was evolved by GP. Genetic models, genetic-Boolean models and Boolean functions served as members of the function set while SNPs served as members of the terminal set in GP. GPTE was benchmarked against RF, which does not incorporate genetic models, in various simulation settings involving different two-locus disease models. The results indicated that the ability to detect causative SNPs in simulated datasets of GPTE and RF was similar when causative SNPs exhibit large marginal single-locus effects. However, GPTE detected more causative SNPs than RF when causative SNPs exhibit small marginal

single-locus effects. This suggests that incorporating genetic models into a tree ensemble is required for the latter scenario, which closely resembles to that in genetic association studies. Setsirichok et al. [18] showed that incorporating a genetic model, which covers mean, additive, dominant and epistatic interaction effects, into a logistic regression model also improves the detection of informative SNPs. However, the logistic regression model with an incorporated genetic model covers more erroneous SNPs than GPTE when causative SNPs exhibit small marginal single-locus effects. Further experiments for comparing the performance of GPTE and that of a logistic regression model with an incorporated genetic model are therefore recommended.

## REFERENCES

1. Bentley, D. R., *Med. Res. Rev.*, 2000, **20**(3), 189–96.
2. Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W. C., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., Busam, D. A., Beeson, K. Y., McIntosh, T. C., Remington, K. A., Abril, J. F., Gill, J., Borman, J., Rogers, Y.-H., Frazier, M. E., Scherer, S. W., Strausberg, R. L., and Venter, J. C., *PLoS Biol.*, 2007, **5**(10), e254.
3. Heidema, A. G., Boer, J. M. A., Nagelkerke, N., Mariman, E. C. M., van der A, D. L., and Feskens, E. J. M., *BMC Genet.*, 2006, **7**, 23.
4. Ziegler, A. and König, I. R., *A Statistical Approach to Genetic Epidemiology*, 2<sup>nd</sup> ed., Wiley-Blackwell, Weinheim, 2010.
5. Saeys, Y., Inza, I., and Larrañaga, P., *Bioinformatics*, 2007, **23**(19), 2507–17.
6. Breiman, L., *Mach. Learn.*, 2001, **45**(1), 5–32.
7. Lunetta, K. L., Hayward, L. B., Segal, J., and van Eerdewegh, P., *BMC Genet.*, 2004, **5**, 32.
8. Meng, Y. A., Yu, Y., Cupples, L. A., Farrer, L. A., and Lunetta, K. L., *BMC Bioinformatics*, 2009, **10**, 78.
9. Cordell, H. J., *Hum. Mol. Genet.*, 2002, **11**(20), 2463–8.
10. Koza, J., *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, 1992.
11. Culverhouse, R., Suarez, B. K., Lin, J., and Reich, T., *Am. J. Hum. Genet.*, 2002, **70**(2), 461–71.
12. Li, W. and Reich, J., *Hum. Hered.*, 2000, **50**(6), 334–49.
13. Wongseree, W., Assawamakin, A., Piroonratana, T., Sinsomros, S., Limwongse, C., and Chaiyaratana, N., *BMC Bioinformatics*, 2009, **10**, 294.
14. Strobl, C., Malley, J., and Tutz, G., *Psychol. Methods*, 2009, **14**(4), 323–48.
15. The International HapMap Consortium, *Nature*, 2005, **437**(7063), 1299–320.
16. Dudek, S. M., Motsinger, A. A., Velez, D. R., Williams, S. M., and Ritchie, M. D., *Proceedings of the Pacific Symposium on Biocomputing 2006*, Edited by Altman, R. B., Dunker, A. K., Hunter, L., Murray, T., and Klein, T. E., World Scientific, Singapore, 2006, 499–510.
17. Setsirichok, D., Tienboon, P., Jaroonruang, N., Kittichaijaroen, S., Wongseree, W., Piroonratana, T., Usavanarong, T., Limwongse, C., Apornthewan, C., Phadoongsidhi, M., and Chaiyaratana, N., *SpringerPlus*, 2013, **2**, 230.
18. Setsirichok, D., Wongseree, W., and Chaiyaratana, N., *The 21<sup>st</sup> International Annual Symposium on Computational Science and Engineering*, National Science and Technology Development Agency, Pathum Thani, Thailand, 2017, in press.

## ACKNOWLEDGMENTS

The authors are extremely grateful to two anonymous reviewers for their valuable comments and suggestions, which have contributed a lot towards improving the content and presentation of this article.

# Incorporating a Genetic Model into a Logistic Regression Model Improves SNP Selection by Lasso for Genetic Association Studies

D. Setsirichok<sup>1,C</sup>, W. Wongseree<sup>1</sup>, and N. Chaiyaratana<sup>1,2</sup>

<sup>1</sup> *Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

<sup>2</sup> *Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*

<sup>C</sup> **E-mail:** d.setsirichok@gmail.com; **Fax:** +66 2 585 7350; **Tel.** +66 2 555 2000 x8421

## ABSTRACT

This article presents a comparison between a logistic regression model with an incorporated genetic model and a standard logistic regression model in a genetic association simulation. A least absolute shrinkage and selection operator (lasso) was used to obtain the logistic regression models' parameters, which provide information about single nucleotide polymorphisms (SNPs) selected for the logistic regression models. The incorporated genetic model covered mean, additive, dominant and epistatic interaction effects. Jointly recessive, single-locus recessive, semi-pure epistasis and pure epistasis models were the disease models explored in the simulation. The results indicated that the logistic regression model with an incorporated genetic model was better than the standard logistic regression model in terms of the number of correctly-identified causative SNPs in the logistic regression model. This suggests that incorporating a genetic model into a logistic regression model is important for detecting informative SNPs in genetic association studies.

**Keywords:** Case-Control Study, Genetic Association Study, Lasso, Logistic Regression, Single Nucleotide Polymorphism.

## 1. INTRODUCTION

Genetic association studies focus on the identification of genetic markers that are associated with complex diseases. Although there are many types of genetic marker, the most commonly found genetic marker is a single nucleotide polymorphism (SNP). The Human Genome Project reveals that the number of SNPs in a human genome exceeds 3,000,000 [1, 2]. This certainly poses a great challenge on genetic epidemiologists who attempt to locate informative SNPs that are associated with a complex disease of interest.

Both family-based and population-based studies are viable for genetic association studies [3]. Family-based and population-based studies aim to identify common disease alleles among related and unrelated individuals, respectively. From a classification viewpoint, a population-based study is essentially an attribute selection problem. In other words, the study attempts to select SNPs that can be used to classify case-control samples from affected and unaffected individuals. Many regression analysis with attribute selection methods are suitable for the task. They include a least absolute shrinkage and selection operator (lasso) [4] and an elastic net (EN) [5]. Lasso exploits an  $l_1$  penalty during the optimisation of a logistic regression model's coefficients, which leads to the attribute selection capability. In contrast, EN exploits a combined  $l_1/l_2$  penalty. Lasso is suitable when attributes are uncorrelated while EN is suitable when attributes are correlated.

The logistic regression model's parameters that represent the coefficients for SNPs should be zero if these SNPs are not associated with the disease. Nonetheless, the possibility of

irrelevant SNPs having non-zero coefficients exists. Hence, it is necessary to determine whether each non-zero coefficient is crucial for the classification capability of the logistic regression model. Recently, a method for determining statistical significance of coefficients in a logistic regression model through a covariance test and monitoring coefficients that should be selected for the logistic regression model was proposed [6].

With the availability of regression analysis with attribute selection methods and how to determine the statistical significance of and monitor a logistic regression model's coefficients, it is possible to apply the aforementioned methodology to population-based association studies. There are two possible strategies: direct application of the methodology and incorporating a genetic model into the logistic regression model [7] prior to the application of the methodology. These two strategies are the interest of this article. Since the candidate gene approach to genetic association studies mainly focuses on selection of unlinked SNPs, lasso was chosen for this study.

## 2. METHODS

### 2.1. Disease model

The study was carried out using a simulation. Each simulated dataset consisted of two causative SNPs. Both causative SNPs were governed by a penetrance-based disease model. Consider a disease model consisting of loci A and B.  $A$  and  $a$  respectively represent the major and minor alleles at locus A while  $B$  and  $b$  respectively represent the major and minor alleles at locus B. Let 0, 1 and 2 respectively represent a homozygous wild-type genotype ( $AA$  and  $BB$ ), a heterozygous genotype ( $Aa$  and  $Bb$ ) and a homozygous variant genotype ( $aa$  and  $bb$ ) at each locus. In addition, let  $f_{ij} \in [0,1]$  be the penetrance of genotype  $i$  at locus A and genotype  $j$  at locus B, and  $p_A$  and  $p_B$  be the major allele frequencies. The marginal penetrance  $M_{Ai}$  for genotype  $i$  at locus A is therefore defined by

$$M_{Ai} = p_B^2 f_{i0} + 2p_B(1-p_B)f_{i1} + (1-p_B)^2 f_{i2}, i \in \{0,1,2\} \quad (1)$$

while the marginal penetrance  $M_{Bj}$  for genotype  $j$  at locus B is defined by

$$M_{Bj} = p_A^2 f_{0j} + 2p_A(1-p_A)f_{1j} + (1-p_A)^2 f_{2j}, j \in \{0,1,2\}. \quad (2)$$

The disease models M1, M7, M97 and M170 [8] as shown in Table 1 were the interesting models in the simulation. M1 is a jointly recessive model. M7 is a single-locus disease model in which allele  $a$  is the disease allele. M97 is a symmetric model with respect to the permutation of alleles  $A$  and  $a$ . The marginal penetrances  $M_{A0}$ ,  $M_{A1}$  and  $M_{A2}$  for genotypes  $AA$ ,  $Aa$  and  $aa$ , respectively become equal when the major allele frequency  $p_B$  equals the minor allele frequency (MAF)  $p_b$ . The causative SNP A thus exhibits no marginal single-locus effect. As a result, it cannot be detected by single-locus analysis. M97 is a semi-pure epistasis model in this situation. M170 is an inference model, which is neither a dominant nor recessive model. The marginal penetrances  $M_{A0}$ ,  $M_{A1}$ ,  $M_{A2}$ ,  $M_{B0}$ ,  $M_{B1}$  and  $M_{B2}$  for genotypes  $AA$ ,  $Aa$ ,  $aa$ ,  $BB$ ,  $Bb$  and  $bb$ , respectively become equal when the major allele frequency  $p_A$  equals the MAF  $p_a$  and the major allele frequency  $p_B$  equals the MAF  $p_b$ . Both causative SNPs A and B thus exhibit no marginal single-locus effects. M170 is a pure epistasis model in this situation.

### 2.2. Least absolute shrinkage and selection operator

A least absolute shrinkage and selection operator (lasso) is a regression analysis with attribute selection method which is suitable for both linear and logistic regression [4]. The technique can identify uncorrelated informative attribute because its penalty function exploits an  $l_1$  penalty. Since this article interests in case-control studies, the following description of lasso focuses on logistic regression.

**Table 1.** Penetrances for disease models M1, M7, M97 and M170.

Genotype	Penetrance for Disease Model			
	M1	M7	M97	M170
<i>AABB</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$
<i>AABb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$\alpha$
<i>AAbb</i>	$1 - \alpha$	$1 - \alpha$	$\alpha$	$1 - \alpha$
<i>AaBB</i>	$1 - \alpha$	$1 - \alpha$	$\alpha$	$\alpha$
<i>AaBb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$
<i>Aabb</i>	$1 - \alpha$	$1 - \alpha$	$1 - \alpha$	$\alpha$
<i>aaBB</i>	$1 - \alpha$	$\alpha$	$1 - \alpha$	$1 - \alpha$
<i>aaBb</i>	$1 - \alpha$	$\alpha$	$1 - \alpha$	$\alpha$
<i>aabb</i>	$\alpha$	$\alpha$	$\alpha$	$1 - \alpha$

$\alpha \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$ . Genotypes with penetrance =  $\alpha$  are disease-predisposing genotypes while genotypes with penetrance =  $1 - \alpha$  are protective genotypes.

Let  $G$  be a response variable that takes an arbitrarily chosen value of 1 or 2. In addition, let  $s$  be a sample which is composed of  $P$  attributes. The class-conditional probabilities dictated by a logistic regression model are defined by

$$\Pr(G = 1|s) = \frac{1}{1 + e^{-(\beta_0 + s^T \beta)}} \quad (3)$$

and

$$\begin{aligned} \Pr(G = 2|s) &= \frac{1}{1 + e^{(\beta_0 + s^T \beta)}} \\ &= 1 - \Pr(G = 1|s). \end{aligned} \quad (4)$$

$\beta_0$  is the intercept and  $\beta$  represents the coefficients for the attributes. Equations (3) and (4) yield

$$\ln \frac{\Pr(G = 1|s)}{\Pr(G = 2|s)} = \beta_0 + s^T \beta. \quad (5)$$

Define  $p(s_i) = \Pr(G = 1|s_i)$  as the probability described in equation (3) for the sample  $s_i$  from a set of  $N$  samples at a value of  $\beta_0$  and  $\beta$ . In addition, define  $y_i$  as the binary class for the sample  $s_i$ . Lasso maximises the penalised log-likelihood

$$\begin{aligned} &\max_{(\beta_0, \beta) \in \mathbf{R}^{P+1}} \left[ \frac{1}{N} \sum_{i=1}^N \{y_i \ln p(s_i) + (1 - y_i) \ln(1 - p(s_i))\} - \lambda \|\beta\|_1 \right] \\ &= \max_{(\beta_0, \beta) \in \mathbf{R}^{P+1}} \left[ \frac{1}{N} \sum_{i=1}^N \{y_i (\beta_0 + s_i^T \beta) - \ln(1 + e^{(\beta_0 + s_i^T \beta)})\} - \lambda \|\beta\|_1 \right] \end{aligned} \quad (6)$$

where  $\lambda$  is the regularisation parameter. A Taylor series expansion at fixed parameters  $\tilde{\beta}_0$  and  $\tilde{\beta}$  provides a quadratic approximation to the log-likelihood part of equation (6). This yields

$$l_Q(\beta_0, \beta) = -\frac{1}{2N} \sum_{i=1}^N w_i (y_i' - \beta_0 - s_i^T \beta)^2 + C(\tilde{\beta}_0, \tilde{\beta})^2 \quad (7)$$

where  $y_i'$  is the working response and is defined by

$$y_i' = \tilde{\beta}_0 + s_i^T \tilde{\beta} + \frac{y_i - \tilde{p}(s_i)}{\tilde{p}(s_i)(1 - \tilde{p}(s_i))}, \quad (8)$$

$w_i$  is the weight associated with the sample  $s_i$  and is defined by

$$w_i = \tilde{p}(s_i)(1 - \tilde{p}(s_i)), \quad (9)$$

$$\tilde{p}(s_i) = \frac{1}{1 + e^{-(\tilde{\beta}_0 + s_i^T \tilde{\beta})}} \quad (10)$$

and  $C(\tilde{\beta}_0, \tilde{\beta})^2$  is a constant.

The penalised weighted least squares problem

$$\min_{(\beta_0, \beta) \in \mathbf{R}^{p+1}} \left[ -l_Q(\beta_0, \beta) + \lambda \|\beta\|_1 \right] \quad (11)$$

can be solved by least angle regression (LARS) [9]. Essentially, the regularisation parameter  $\lambda$  is initially set to infinity. As a result, the coefficients  $\beta$  for all attributes are zero. The regularisation parameter then decreases until the first non-zero coefficient for an attribute enters the logistic regression model. The decreasing of the regularisation parameter and the entering of a new non-zero coefficient continue until the regularisation parameter is zero. The coefficient profiles generated by LARS are piecewise smooth where each smooth section ends when a new non-zero coefficient enters the model [10].

Each time that a new non-zero coefficient enters the logistic regression model, it is possible to test the statistical significance of this coefficient. A covariance test statistic, which follows an Exp(1) asymptotic distribution, was proven to be suitable for the task [6]. The coefficient is said to be statistically significant if its  $p$ -value is less than 0.05. The monitoring of coefficients begins when the first coefficient with a significant  $p$ -value enters the model. The monitoring process continues as long as the entering coefficient has a significant  $p$ -value and terminates when a coefficient with a  $p$ -value  $\geq 0.05$  enters the model immediately after a coefficient with a significant  $p$ -value. The last coefficient with a significant  $p$ -value that enters the model and all coefficients that enter the model prior to this coefficient mark the presence of informative attributes that should be selected.

### 2.3. Genetic model

A logistic regression model can be directly applied to case-control datasets. For a dataset containing  $P$  SNPs, the logistic regression model accommodates  $P$  coefficients and one intercept. However, a two-locus genetic model can be incorporated into the logistic regression model. A two-locus genetic model is of interest because it was proven that multiple causative SNPs can be detected through two-locus analysis [11, 12]. Cordell [7] explained that a logistic regression model that incorporates a two-locus genetic model is defined by

$$\Pr(G = 1 | \langle x_1, x_2, z_1, z_2 \rangle) = \frac{1}{1 + e^{-(\beta_0 + a_1 x_1 + d_1 z_1 + a_2 x_2 + d_2 z_2 + i_{aa} x_1 x_2 + i_{ad} x_1 z_2 + i_{da} z_1 x_2 + i_{dd} z_1 z_2)}} \quad (12)$$

and

$$\begin{aligned} \Pr(G = 2 | \langle x_1, x_2, z_1, z_2 \rangle) &= \frac{1}{1 + e^{(\beta_0 + a_1 x_1 + d_1 z_1 + a_2 x_2 + d_2 z_2 + i_{aa} x_1 x_2 + i_{ad} x_1 z_2 + i_{da} z_1 x_2 + i_{dd} z_1 z_2)}} \\ &= 1 - \Pr(G = 1 | \langle x_1, x_2, z_1, z_2 \rangle). \end{aligned} \quad (13)$$

$x_i$  and  $z_i$  are dummy variables representing the genotype at locus  $i$ . Generally,  $x_i = 1$  and  $z_i = -0.5$  for a homozygous wild-type genotype,  $x_i = 0$  and  $z_i = 0.5$  for a heterozygous genotype, and  $x_i = -1$  and  $z_i = -0.5$  for a homozygous variant genotype. The intercept  $\beta_0$  represents the mean effect while the coefficients  $a_i$  and  $d_i$  represent the additive and dominant effects at loci  $i$ , respectively. Therefore, the coefficients  $i_{aa}$ ,  $i_{ad}$ ,  $i_{da}$  and  $i_{dd}$  represent four possible epistatic interaction effects. Once the genetic model expands to cover  $P$  loci, the numbers of coefficients for additive, dominant and epistatic interaction effects are  $P$ ,  $P$  and  $4 \times \binom{P}{2}$ , respectively. This means that the number of coefficients in a logistic regression model with an incorporated genetic model is a quadratic function of the number of SNPs.

## 3. RESULTS AND DISCUSSION

A logistic regression model with an incorporated genetic model was compared with a standard logistic model in a simulation trial involving disease models described in section 2.1.

Lasso was used to obtain the logistic regression models' parameters. Each simulated dataset consisted of 20 unlinked SNPs where two SNPs were causative SNPs. Both causative SNPs were governed by one of four disease models. The differences between penetrances of disease-predisposing and protective genotypes in each disease model were 0.2, 0.4, 0.6, 0.8 and 1.0. The MAFs of irrelevant SNPs were between 0.05 and 0.5 while the allele frequencies of both causative SNPs were 0.5. They conform to the MAFs of SNPs targeted by the International HapMap Project [13]. The dataset consisted of 400 balanced case-control samples in which all causative SNPs in control samples were in Hardy-Weinberg equilibrium. genomeSIM [14] was used to generate 100 independent datasets for each simulation setting. The interesting performance indicators were the number of selected SNPs and the number of correctly-identified causative SNPs in the logistic regression model. These performance indicators were proven to be suitable when datasets contain two or more causative SNPs and causative SNP detection is performed through two-locus analysis [11, 12]. If coefficients, whether they represent additive, dominant or epistatic interaction effects, were selected for the logistic regression model with an incorporated genetic model, then all SNPs corresponding to those coefficients were selected. The significance of difference between the numbers of SNPs in the logistic regression models was assessed through a paired *t*-test.

The results from all simulation settings are given in Tables 2–5. The number of selected SNPs in the logistic regression model with an incorporated genetic model was either similar to or higher than that in the standard logistic regression model in every simulation setting. On the other hand, the numbers of correctly-identified causative SNPs in the logistic regression model with an incorporated genetic model and standard logistic regression model were similar in two scenarios. These are the scenarios when the disease model is M1 while the difference

**Table 2.** Number of selected SNPs and number of correctly-identified causative SNPs in the standard logistic regression model and the logistic regression model with an incorporated genetic model for the datasets with disease model M1.

$\alpha$	Number of Selected SNPs			Number of Correctly-Identified Causative SNPs		
	Standard	Genetic	<i>p</i> -value	Standard	Genetic	<i>p</i> -value
0.6	0.11 (0.35)	12.61 (3.72)	< 0.0001	0.00 (0.00)	1.47 (0.67)	< 0.0001
0.7	0.25 (0.52)	11.83 (4.29)	< 0.0001	0.20 (0.43)	1.78 (0.48)	< 0.0001
0.8	0.69 (0.71)	7.19 (5.27)	< 0.0001	0.68 (0.68)	1.91 (0.29)	< 0.0001
0.9	1.94 (0.42)	2.47 (1.77)	0.0039	1.88 (0.33)	1.91 (0.29)	0.4914
1.0	2.00 (0.00)	2.00 (0.00)	-	2.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 3.** Number of selected SNPs and number of correctly-identified causative SNPs in the standard logistic regression model and the logistic regression model with an incorporated genetic model for the datasets with disease model M7.

$\alpha$	Number of Selected SNPs			Number of Correctly-Identified Causative SNPs		
	Standard	Genetic	<i>p</i> -value	Standard	Genetic	<i>p</i> -value
0.6	0.50 (0.61)	11.12 (4.96)	< 0.0001	0.44 (0.50)	0.97 (0.17)	< 0.0001
0.7	1.08 (0.27)	2.21 (3.47)	0.0014	1.00 (0.00)	1.00 (0.00)	-
0.8	1.07 (0.26)	1.07 (0.33)	1.0000	1.00 (0.00)	1.00 (0.00)	-
0.9	1.07 (0.26)	1.02 (0.14)	0.0889	1.00 (0.00)	1.00 (0.00)	-
1.0	1.00 (0.00)	1.00 (0.00)	-	1.00 (0.00)	1.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 4.** Number of selected SNPs and number of correctly-identified causative SNPs in the standard logistic regression model and the logistic regression model with an incorporated genetic model for the datasets with disease model M97.

$\alpha$	Number of Selected SNPs			Number of Correctly-Identified Causative SNPs		
	Standard	Genetic	<i>p</i> -value	Standard	Genetic	<i>p</i> -value
0.6	0.09 (0.35)	12.24 (3.90)	< 0.0001	0.00 (0.00)	1.79 (0.46)	< 0.0001
0.7	0.13 (0.37)	2.40 (2.19)	< 0.0001	0.02 (0.14)	1.96 (0.20)	< 0.0001
0.8	0.10 (0.30)	2.07 (0.33)	< 0.0001	0.00 (0.00)	2.00 (0.00)	-
0.9	0.05 (0.22)	2.07 (0.33)	< 0.0001	0.00 (0.00)	2.00 (0.00)	-
1.0	0.08 (0.30)	2.00 (0.00)	< 0.0001	0.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

**Table 5.** Number of selected SNPs and number of correctly-identified causative SNPs in the standard logistic regression model and the logistic regression model with an incorporated genetic model for the datasets with disease model M170.

$\alpha$	Number of Selected SNPs			Number of Correctly-Identified Causative SNPs		
	Standard	Genetic	<i>p</i> -value	Standard	Genetic	<i>p</i> -value
0.6	0.13 (0.37)	7.97 (6.20)	< 0.0001	0.00 (0.00)	1.92 (0.37)	< 0.0001
0.7	0.10 (0.33)	2.06 (0.31)	< 0.0001	0.00 (0.00)	2.00 (0.00)	-
0.8	0.12 (0.38)	2.03 (0.22)	< 0.0001	0.01 (0.10)	2.00 (0.00)	< 0.0001
0.9	0.07 (0.26)	2.03 (0.22)	< 0.0001	0.01 (0.10)	2.00 (0.00)	< 0.0001
1.0	0.09 (0.29)	2.00 (0.00)	< 0.0001	0.00 (0.00)	2.00 (0.00)	-

The results were averaged over 100 independent simulations. The numbers in brackets are standard deviation.

between penetrances of disease-predisposing and protective genotypes is large and when the disease model is M7 while the difference between penetrances of disease-predisposing and protective genotypes is moderate or large. The logistic regression model with an incorporated genetic model contained more correctly-identified causative SNPs in the remaining scenarios ( $p < 0.0001$ ). The simulation results can be further interpreted as follows.

Causative SNPs governed by disease models M1 and M7 are not difficult to detect. This is because the only disease-predisposing genotype in disease model M1 is *aabb* while the disease-predisposing genotypes in disease model M7 must contain two copies of allele *a*. It means that a standard logistic regression model is sufficient for describing the association between these causative SNPs and a disease of interest. Nonetheless, the decreasing of marginal single-locus effect at both loci of disease model M1 and at locus A of disease model M7 as a result of the decreasing of  $\alpha$  makes the detection of causative SNPs becomes more difficult when using a standard logistic regression model. In contrast, causative SNPs governed by disease models M97 and M170 are more difficult to detect. This is because causative SNP A in disease model M97 and causative SNPs A and B in disease model M170 exhibit no marginal single-locus effects. As a result, it is impossible to detect causative SNPs in disease models M97 and M170 using a standard logistic regression model. This suggests that incorporating a genetic model into a logistic regression model is beneficial for detecting causative SNPs. Nonetheless, this also leads to the increased probability of including more erroneous SNPs in the logistic regression model. In addition, the computational time required by the logistic regression model with an incorporated genetic model was approximately a square of that required by the standard logistic regression model. This is because the incorporation of a genetic model introduces more coefficients to a logistic regression model.

#### 4. CONCLUSION

In this article, a logistic regression model with an incorporated genetic model was compared with a standard logistic regression model in a simulation trial designed for genetic association studies. Lasso was used to obtain the logistic regression models' parameters, which provide information about selected SNPs. The incorporated genetic model covered mean, additive, dominant and epistatic interaction effects. The interesting simulation settings involved different two-locus disease models. The results indicated that the number of correctly-identified causative SNPs in the logistic regression model with an incorporated genetic model was either similar to or higher than that in the standard logistic regression model in every simulation setting. This suggests that incorporating a genetic model into a logistic regression model is crucial for causative SNP detection.

#### REFERENCES

1. Bentley, D. R., *Med. Res. Rev.*, 2000, **20**(3), 189–96.
2. Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W. C., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., Busam, D. A., Beeson, K. Y., McIntosh, T. C., Remington, K. A., Abril, J. F., Gill, J., Borman, J., Rogers, Y.-H., Frazier, M. E., Scherer, S. W., Strausberg, R. L., and Venter, J. C., *PLoS Biol.*, 2007, **5**(10), e254.
3. Ziegler, A. and König, I. R., *A Statistical Approach to Genetic Epidemiology*, 2<sup>nd</sup> ed., Wiley-Blackwell, Weinheim, 2010.
4. Tibshirani, R., *J. R. Stat. Soc. Ser. B-Methodol.*, 1996, **58**(1), 267–88.
5. Zou, H. and Hastie, T., *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 2005, **67**(2), 301–20.
6. Lockhart, R., Taylor, J., Tibshirani, R. J., and Tibshirani, R., *Ann. Stat.*, 2014, **42**(2), 413–68.
7. Cordell, H. J., *Hum. Mol. Genet.*, 2002, **11**(20), 2463–8.
8. Li, W. and Reich, J., *Hum. Hered.*, 2000, **50**(6), 334–49.
9. Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., *Ann Stat.*, 2004, **32**(2), 407–99.
10. Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2<sup>nd</sup> ed., Springer, New York, 2009.
11. Wongseree, W., Assawamakin, A., Piroonratana, T., Sinsomros, S., Limwongse, C., and Chaiyaratana, N., *BMC Bioinformatics*, 2009, **10**, 294.
12. Setsirichok, D., Tienboon, P., Jaroonruang, N., Kittichaijaroen, S., Wongseree, W., Piroonratana, T., Usavanarong, T., Limwongse, C., Aporntewan, C., Phadoongsidhi, M., and Chaiyaratana, N., *SpringerPlus*, 2013, **2**, 230.
13. The International HapMap Consortium, *Nature*, 2005, **437**(7063), 1299–320.
14. Dudek, S. M., Motsinger, A. A., Velez, D. R., Williams, S. M., and Ritchie, M. D., *Proceedings of the Pacific Symposium on Biocomputing 2006*, Edited by Altman, R. B., Dunker, A. K., Hunter, L., Murray, T., and Klein, T. E., World Scientific, Singapore, 2006, 499–510.

#### ACKNOWLEDGMENTS

The authors are extremely grateful to two anonymous reviewers for their valuable comments and suggestions, which have contributed a lot towards improving the content and presentation of this article.

# Backward Bifurcation of SEIR Epidemic Model with Treatment Function

Sasiporn Rattanasupha<sup>1C</sup>, Settapat Chinviriyasit<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

<sup>2</sup> Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

<sup>C</sup> E-mail: sasipornoh22@gmail.com; Tel. +66 8 1845 0841

## ABSTRACT

SEIR epidemic model with treatment function is proposed to investigate the effect for treatment of infectious on the disease spread. It is found that when the effect of the infected being delayed for treatment becomes stronger than some level the backward bifurcation will take place. That is to say, driving the basic reproduction number below one is not enough to control and eradicate the disease.

**Keywords:** Backward Bifurcation, SEIR Epidemic Model, Treatment Function.

## 1. INTRODUCTION

Mathematical model of epidemiology is the modelling of an infectious diseases which has been used to predict the transmission dynamics of the infectious diseases in the host population [1-4]. In epidemic model, the treatment is an important method to control or decreasing the spread of diseases such as measles, flu and tuberculosis. Recently some author [5-8] have found that epidemic models leading to backward bifurcation, since the requirement of the basic reproduction number is less than unity to be insufficient for disease elimination. Thus, it is important to identify backward bifurcations to obtain thresholds for the control of diseases.

The purpose of this paper is to study the backward bifurcation of SEIR epidemic model with treatment function. The model compounds with 4 individuals: Susceptible ( $S$ ), Exposed ( $E$ ), Infected ( $I$ ) and Recovered ( $R$ ) as shown in the following system of equations:

$$\begin{aligned}\frac{dS}{dt} &= A - \beta SI - \mu S \\ \frac{dE}{dt} &= \beta SI - (\mu + \varepsilon)E \\ \frac{dI}{dt} &= \varepsilon E - (\mu + \gamma + d)I - \frac{rI}{1 + \alpha I} \\ \frac{dR}{dt} &= \gamma I - \mu R + \frac{rI}{1 + \alpha I}\end{aligned}\tag{1}$$

Where  $A$  is the recruitment rate,  $\beta$  is the infection rate,  $\mu$  is the natural death rate,  $\varepsilon$  is the progression rate to symptom development (the rate at which an infected individual becomes infectious per unit time),  $\gamma$  is the removal rate (the rate at which an infectious individual recovers per unit time),  $d$  is the disease-related death and  $T(I)$  is the treatment

rate function. The treatment function is defined by  $T(I) = \frac{rI}{1+\alpha I}$  where  $r$  is the cure rate,  $\alpha$  is present to measure the extent of the effect of there being a delay in the treatment of infection.

Since the first three equations in system (1) do not depend on the fourth equation, and therefore, this equation can be omitted without loss of generality. Hence, system (1) can be reduced as

$$\begin{aligned}\frac{dS}{dt} &= A - \beta SI - \mu S \\ \frac{dE}{dt} &= \beta SI - (\mu + \varepsilon)E \\ \frac{dI}{dt} &= \varepsilon E - (\mu + \gamma + d)I - \frac{rI}{1 + \alpha I}\end{aligned}\quad (2)$$

The system (2) leads to

$$\begin{aligned}\frac{dS}{dt} + \frac{dE}{dt} + \frac{dI}{dt} &= A - \mu(S + E + I) - (\gamma + d)I - \frac{rI}{1 + \alpha I} \\ &\leq A - \mu(S + E + I)\end{aligned}$$

Then  $\limsup_{t \rightarrow \infty} (S + E + I) \leq \frac{A}{\mu}$ . Thus, the system (2) is studied in the close set the feasible region

$$D = \left\{ (S + E + I) \in \mathbb{R}_+^3 \mid S + E + I \leq \frac{A}{\mu} \right\} \quad (3)$$

which is positively invariant with respect to the system (2).

The organization of this paper is as follows: The existence of equilibria and backward bifurcation are analyzed in Section 2. The numerical solutions are simulated in Section 3. Finally, conclusion are given in Section 4.

## 2. THE EXISTENCE OF EQUILIBRIA AND BACKWARD BIFURCATION

### 2.1 Disease-free Equilibrium (DFE) and the Basic reproduction number

The system (2) has a Disease-free Equilibrium (DFE), obtained by setting the right-hand side of the system (2) to zero, given by  $X_0 = (S^0, E^0, I^0) = \left( \frac{A}{\mu}, 0, 0 \right)$

We introduce the threshold parameter  $R_0$ . By using the next generation method [9], it is convenient to define

$$R_0 = \frac{\varepsilon \beta A}{\mu(\mu + \varepsilon)(\mu + \gamma + d + r)}$$

which is the Basic reproduction number.

## 2.2 Endemic Equilibrium (EE)

The endemic equilibrium of the system (2), denoted by  $X^* = (S^*, E^*, I^*)$

$$S^* = \frac{A}{\beta I^* + \mu} \quad , \quad E^* = \frac{\beta A I^*}{(\mu + \varepsilon)(\beta I^* + \mu)}$$

and  $I^*$  is the positive solution of the following equation

$$aI^{*2} + bI^* + c = 0 \quad (4)$$

where

$$\begin{aligned} a &= \alpha\beta(\mu + \varepsilon)(\mu + \gamma + d) \\ b &= \beta(\mu + \varepsilon)(\mu + \gamma + d + r) + \alpha\mu(\mu + \varepsilon)(\mu + \gamma + d) - \varepsilon\beta\alpha A \\ c &= \mu(\mu + \varepsilon)(\mu + \gamma + d + r)(1 - R_0) \end{aligned}$$

The backward bifurcation is investigated in this research. This is indicated that the system (2) has backward bifurcation at  $R_0 = 1$ , if and only if  $\alpha > 0, b < 0$ . The explicit criterion of  $\alpha$  in terms of the parameter  $\beta, \mu, \gamma, d, r$  for the existence of a backward bifurcation at  $R_0 = 1$  are given. When  $R_0 = 1, c = 0$ , these yield

$$\varepsilon\beta A = \mu(\mu + \varepsilon)(\mu + \gamma + d + r). \quad (5)$$

The condition  $b < 0$  is equivalent to

$$\beta(\mu + \varepsilon)(\mu + \gamma + d + r) + \alpha\mu(\mu + \varepsilon)(\mu + \gamma + d) < \varepsilon\beta\alpha A \quad (6)$$

From (5) and (6), we get

$$\beta(\mu + \varepsilon)(\mu + \gamma + d + r) + \alpha\mu(\mu + \varepsilon)(\mu + \gamma + d) < \alpha\mu(\mu + \varepsilon)(\mu + \gamma + d + r)$$

which reduces to

$$\alpha > \frac{\beta(\mu + \gamma + d + r)}{\mu r} \equiv \alpha_0. \quad (7)$$

The following theorem is established.

**Theorem 1** When  $0 < \alpha_0 < \alpha$  then the system (2) has a backward bifurcation at  $R_0 = 1$ ,

$$\text{where } \alpha_0 = \frac{\beta(\mu + \gamma + d + r)}{\mu r}$$

So a backward bifurcation occurs at  $R_0 = 1$  and  $\alpha > \alpha_0$ . Furthermore, we can show that when the effect of the infected being delayed for treatment becomes stronger than some level, the backward bifurcation will take place. Thus the effect of the infected being delayed for treatment,  $\alpha$  is one of the factors which lead to the backward bifurcation.

Under the condition of Theorem 1, the system (2) have two endemic equilibria for an

interval of values of  $R_0$  from a critical value,  $R_c$  to  $R_0 = 1$ . To calculate  $R_c$ , defined by  $b^2 - 4ac = 0$  given by

$$R_c = 1 - \frac{b^2}{4a\mu(\mu + \varepsilon)(\mu + \gamma + d + r)} \quad (8)$$

However, if  $\alpha > 0, b < 0$ , the system (2) have two different endemic equilibria when  $R_c < R_0 < 1$ .

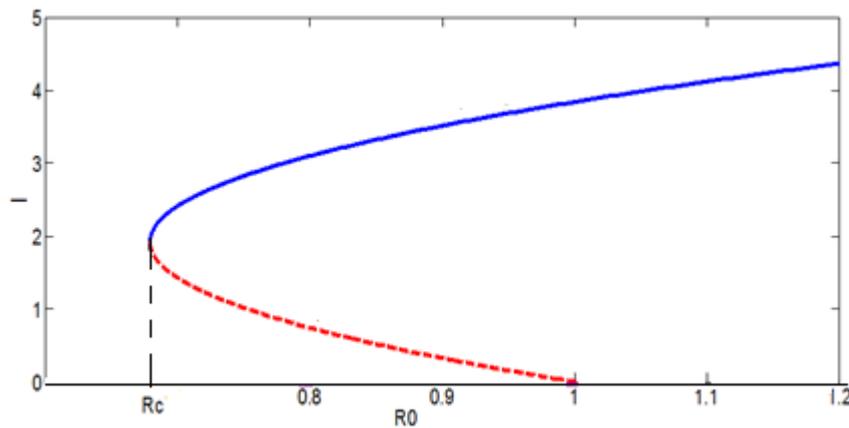
### 3. NUMERICAL SIMULATIONS

In order to investigate the dynamic behaviors of the system (2). The following set of parameters [5] are given as

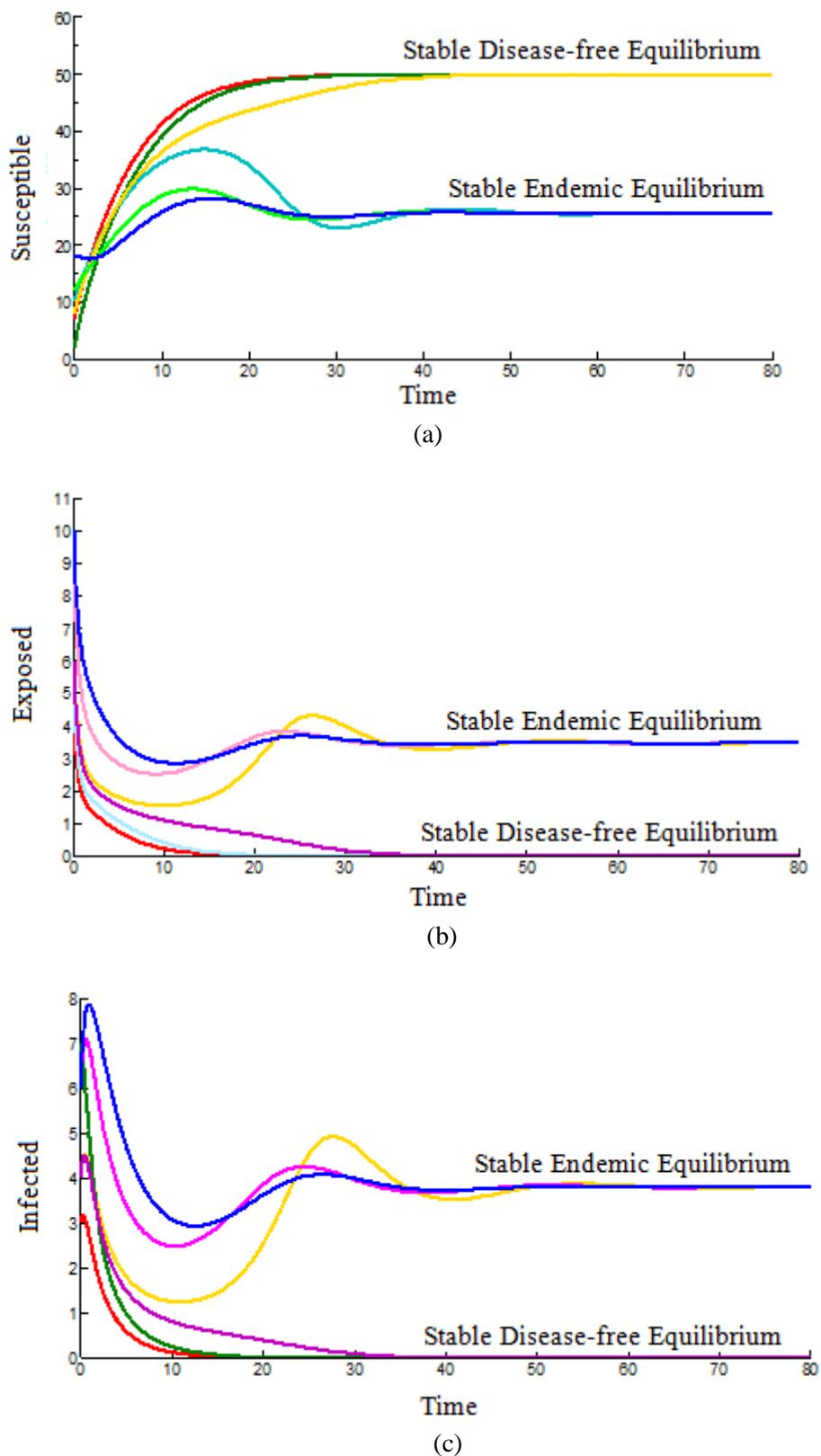
$$A = 10, \varepsilon = 1.2, \gamma = 0.4, d = 0.2$$

whereas  $\beta, \mu, r$  and  $\alpha$  are carried out.

For parameter values  $A = 10, \beta = 0.05, \mu = 0.2, \varepsilon = 1.2, \gamma = 0.4, d = 0.2$  and  $r = 2, \alpha = 1.5$ , it is seen that  $R_0 = 0.7653 < 1$ ,  $R_c = 0.4859 < R_0 < 1$ ,  $b = -0.3680 < 0$  and  $\alpha_0 = 0.3500 < \alpha$ . From theorem 1, we know that the system (2) has a backward bifurcation with two endemic equilibria when  $R_c < R_0 < 1$  shown in Fig.1 and the system (2) has the bistable equilibria: a disease-free equilibrium  $X_0 = (50, 0, 0)$  and an endemic equilibrium  $X_1^* = (25.6235, 3.4823, 3.8053)$ . And the other endemic equilibrium  $X_2^* = (43.7097, 0.8986, 0.5756)$  is unstable shown in Fig.2 (a)-(c).



**Figure 1.** Backward bifurcation of the model with two endemic equilibria when  $A = 10, \beta = 0.05, \mu = 0.2, \varepsilon = 1.2, \gamma = 0.4, d = 0.2$ ,  $r = 2, \alpha = 1.5$  and  $R_c < R_0 < 1$



**Figure 2.** (a) – (c) The system has the bistable equilibria: a disease-free equilibrium  $X_0 = (50, 0, 0)$  and an endemic equilibrium  $X_1^* = (25.6235, 3.4823, 3.8053)$ . And the other endemic equilibrium  $X_2^* = (43.7097, 0.8986, 0.5756)$  is unstable. In this case  $A = 10, \beta = 0.05, \mu = 0.2, \varepsilon = 1.2, \gamma = 0.4, d = 0.2$  and  $r = 2, \alpha = 1.5$

#### 4. CONCLUSION

In this paper, we consider backward bifurcation of SEIR epidemic model with treatment function. We found that, the basic reproduction number below unity is not enough to eradicate the disease when the effect of the infected being delayed for treatment becomes stronger than some level, the backward bifurcation will take place when  $0 < \alpha_0 < \alpha$  and here is a critical value  $R_c$  at the turning point. Thus the effect of the infected being delayed for treatment,  $\alpha$  is one of the factors which lead to the backward bifurcation. And consider, when  $R_c < R_0 < 1$ , the system (2) has the bistable equilibria: a disease-free equilibrium given by  $X_0$  and an endemic equilibrium  $X_1^*$ . And the other endemic equilibrium  $X_2^*$  is unstable.

#### REFERENCES

1. Arino, J., Brauer, F., van den Driessche, P., Watmough, J., and Wu, J., *J. Thero.Biol.*, 2008, **253**(1), 118-130.
2. Coburn, B. J., Wagner, B. G., and Blower, S., *BMC Medicine*, 2009, **7**(30), 1-8.
3. Tchuente, J., Khamis, S., Agosto, F., and Mpeshe, S., *Acta Biotheor.*, 2011, 59(1), 1-28.
4. Ruan, S., Wang, W., *J. Differ. Equations*, 2003, **188**, 135-163.
5. Zhou, X., Cui, J., *Commun. Nonlinear Sci. Numer. Simulat.*, 2011, **16**, 4438-4450.
6. Zhang, X., Liu, X., *J. Math. Anal. Appl.*, 2008, **348**, 433-443.
7. Hu, Z., Liu, S., Wang, H., *Nonlinear Anal. Real World Appl.*, 2003, **9**, 2302-2312.
8. Zhang, J., Jia, J., Song, X., *Scientific World J.*, 2014, 1-11.
9. Van den driessche, P., and Watmough, J., *Math. Biosci.*, 2002, **180**, 29-48.

# Molecular Docking Investigation of WR99210 Analogues as Novel *M.tuberculosis* DHFR Inhibitors

**Nattaya Mekjirawat<sup>1</sup>, Patcharapon Prasertth<sup>1</sup>, Atcharapun Chockborvornpaisan<sup>1</sup>,**

**Suriyawut Kulatee<sup>1</sup>, and Luckhana Lawtrakul<sup>1,C</sup>**

<sup>1</sup>*School of Bio-Chemical Engineering and Technology,*

*Sirindhorn International Institute of Technology, Thammasat University*

*P.O. Box 22, Thammasat Rangsit Post Office, Pathum Thani 12121, Thailand*

<sup>C</sup> *E-mail: luckhana@siit.tu.ac.th*

## ABSTRACT

The interactions between *Mycobacterium tuberculosis* dihydrofolate reductase (mtbDHFR) and WR99210 (IUPAC Name: 6,6-dimethyl-1-[3-(2,4,5-trichlorophenoxy)propoxy]-1,3,5-triazine-2,4-diamine) including some of its analogues, were investigated by molecular modeling approaches. Their initial conformation were generated by the means of GaussView and their geometry optimizations were performed as included in the Gaussian09 program package. AutoDock 4.2.6 was used to dock 19 WR99210 analogues into the active site of mtbDHFR. The binding affinities of a number of analogues tested support the validity of the molecular docking prediction model. These approaches can determine the highest affinity compound which is compound 8 that lead for further development of effective anti-tuberculosis agents.

**Keywords:** Binding energy, Inhibition constant, Ligands-Protein Interaction

## 1. INTRODUCTION

Tuberculosis (TB) is a disease which mostly caused by *Mycobacterium tuberculosis*. It can be spread through the air from one person to another by sneezing and coughing. People who are infected by TB disease need a prescribed medical treatment. However, the estimated 9 million new TB cases had occurred globally, and more than 480,000 cases of multidrugresistant (MDR) occur every year which 9% of them are affected by extensively drugresistant (XDR) [1]. Therefore, the development of new drug is urgently needed.

Dihydrofolate reductase (DHFR) is an enzyme that reduces dihydrofolate (DHF) to tetrahydrofolate (THF) by using NADPH as a cofactor. The derivatives of THF are essential for the cell proliferation and cell growth because, folate is required for the synthesis of nucleotides and many amino acids. According to the function and property of DHFR, it is reasonably to presume that DHFR of pathogenic organism could be a good target for drug. Hence, the selected drug target to design tuberculosis drug candidate in this work is *Mycobacterium tuberculosis* dihydrofolate reductase (mtbDHFR).

In this study, WR99210 analogues are the interested inhibitors. Molecular modeling is used to investigate the interactions between mtbDHFR and WR99210 analogues. The optimization is performed to determine the most stable conformation of each analogue. Molecular docking is applied to dock all analogues into the active site of mtbDHFR. The results of this study will provide the characteristics of effective inhibitors for mtbDHFR. Furthermore, it could lead to development of effective anti-tuberculosis agents.

## 2. THEORY AND RELATED WORKS

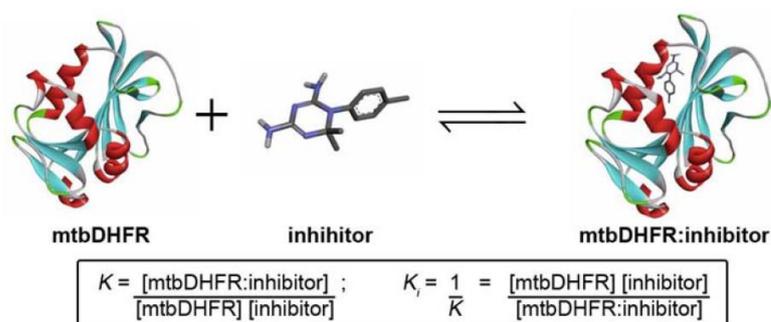
DHFR is an enzyme in reductase family that is normally expressed in all organisms where it sustains the intracellular level of reduced folates. It catalyzes the reduction of DHF to THF, using NADPH as a cofactor [2]. As THF, which is important cofactor for

cell growth and DNA synthesis, is produced by using DHFR. Thus, DHFR is an interesting pharmaceutical target that is inhibited by the inhibitor.

For tuberculosis treatment, MTX, TMP, pyrimethamine (PYR), cycloguanil (CYC), and WR99210 are discovered that they may be effective inhibitors for restraining mtbDHFR function. Following the table 1, binding energy in different inhibitors bound with mtbDHFR is negative sign which means the interaction between inhibitors and mtbDHFR can be possibly occurred. The results of this study are given below.

**Table 1.** RMSD of molecular superposition between docked inhibitors and their specific mtbDHFR binding energy (B.E.) and inhibition constant ( $K_i$ ) from the experimental data and molecular docking calculation [3].

Inhibitors	PDB code	RMSD (Å)	$K_i$ (nM)		B.E. (kcal/mol)	
			Exp.	Docking	Exp.	Docking
Br-WR99210	1DG7	1.67	187	223.46	-9.18	-9.07



**Figure 1.** The model of molecular docking study of antifolate drug into the binding site of mtbDHFR and the formula to calculate the inhibition constant ( $K_i$ ).

The binding energy of mtbDHFR inhibitor in molecular docking is related to both  $K$  and  $K_i$  values by the equation below:

$$B.E. = -RT \ln K = RT \ln K_i$$

The binding energy and  $K_i$  at  $T = 298.15$  K of molecular docking models and from experiments are presented in Table 1.

Actually, there are some research that already studied about the interactions between inhibitors and mtbDHFR. The experiment was performed by culture the bacteria in a specific area. Also, the computational method is applied in order to obtain the inhibition constant and binding energy to classify high and low affinity compound which may further lead to the development of new anti-TB drugs.

### 3. COMPUTATIONAL DETAILS

#### 3.1 Ligands structure preparation and optimization

The WR99210 series of compounds [4] against the *M. tuberculosis* structures are built using GaussView 5.0.8 [5] and the geometry optimizations are performed using calculations at HF/6-31G (d,p) level in the Gaussian 09 program [6]. The optimized conformations are saved as .mol2 file to put into the molecular docking step.

### 3.2 Protein structure preparation

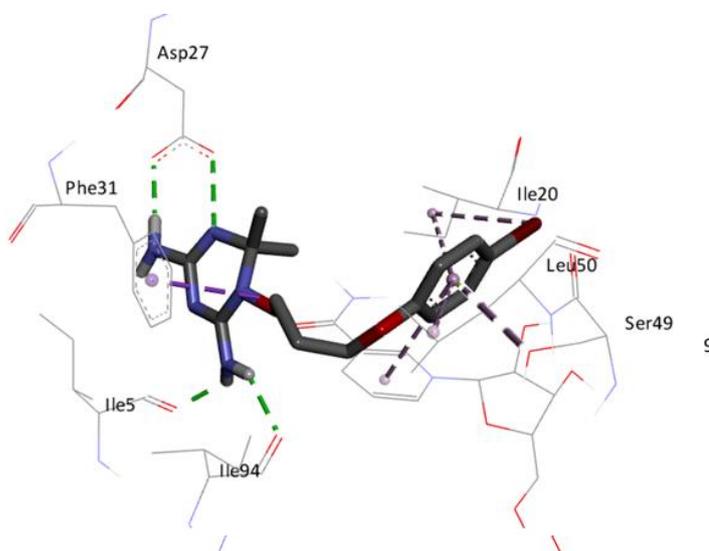
The structure of mtbDHFR complexes with NADPH and 4-bromo WR99210 (PDB code 1DG7) [4] is provided by X-RAY diffraction method in RCSB protein data bank. In order to remove inhibitors, ligands, and water molecules from the crystal structure, Discovery Studio Visualizer 4.0 is applied.

### 3.3 Molecular docking

Molecular docking of 1DG7 mtbDHFR with inhibitors is performed. The protein is fixed as a rigid molecule (host) while the rotatable bonds within the inhibitors molecules (guest) are allowed to be flexible during the docking simulation. At first, the structure of a ligand is opened to check for the bond order and charge. Then the file is saved as .pdbqt type ready to dock with 1DG7 host. In order to start docking ligands with the host, the grid specification is required. The results will be analyzed and the total binding energy will be calculated. The conformation with lowest binding energy and highest frequency will be chosen. Finally, all of the selected conformations will be brought to compare since they are differ in the identity of R group, the length of a carbon linker chain, and the substitution of sulfonyl group and oxygen atom. In addition, the compound that contains lowest binding energy and lowest value of the inhibition constant will provide highest affinity.

## 4. RESULTS AND DISCUSSION

In order to obtain the information of anti-mtbDHFR biological activity which WR99210 compounds bind at the active site of the NADH:mtbDHFR complex, the molecular docking is required. The molecular docking calculation reveals the interaction between two amino groups ( $\text{NH}_2$ ) of WR99210 analogues and protein residues Ile5, Aps27, and Ile94 to form three hydrogen bonds as indicated by green dash line in Figure 2. However, N1 of the pyrimidine ring of WR99210 is possible to interact via a contact of H-bond with Aps27. The R group of WR99210 analogues is in hydrophobic behavior, contacts with the side-chain of Leu50.

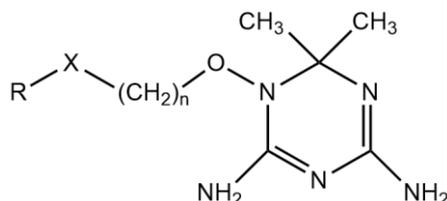


**Figure 2.** The interactions between mtbDHFR with NADPH and Br-WR99210

Molecular docking provides the binding energy to identify alteration of lead compound WR99210 that shows vary level of inhibition of mtbDHFR. The WR99210 analogues are distinguished by adjusting the structure of their compound into three different ways: the specification of the R group, the substitution of sulfonyl group or an oxygen atom at X position, and the length of carbon linker chain.

After docking 1DG7 with series of WR99210, the results including B.E. and  $K_i$  were calculated. The values of B.E. and  $K_i$  indicate the stability of interaction between WR99210 analogues (ligand) and 1DG7 (host). Therefore, the complex with lowest B.E. and lowest  $K_i$  is preferred.

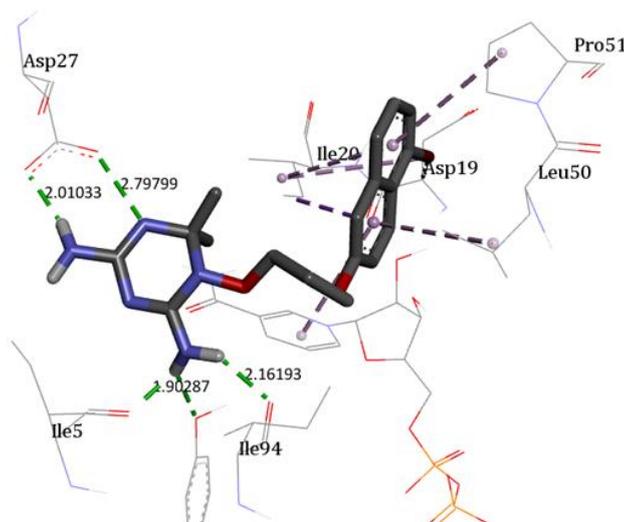
**Table 2.** Binding energy (B.E.) and inhibition constant ( $K_i$ ) of 1DG7 with WR99210 analogues from docking calculations.



No.	R	X	n	B.E. (kcal/mol)	$K_i$ (nM)
1	2,4,5-Trichlorophenyl (WR99210)	O	3	-9.20	180.16
2	2,4-Dichlorophenyl	O	3	-9.35	139.86
3	1-Naphthyl	O	3	-9.30	152.18
4	4-Bromophenyl (Br-WR99210)	O	3	-9.44	120.15
5	3,4-Dichlorophenyl	O	3	-9.40	128.54
6	1-(4-Chloronaphthyl)	O	3	-9.88	57.17
7	4- <i>t</i> -Butylphenyl	O	3	-9.66	82.88
8	2-(6-Bromonaphthyl)	O	3	-10.07	41.48
9	4-Methoxyphenyl	O	3	-8.85	325.27
10	4-Chlorophenyl	O	3	-9.27	160.08
11	2-Naphthyl	O	3	-9.79	66.55
12	4-Nitrophenyl	O	3	-8.73	398.30
13	4-Chlorophenyl	CH <sub>2</sub>	3	-9.43	122.19
14	4-Fluorophenyl	S	3	-8.71	411.97
15	2,4,5-Trichlorophenyl	O	2	-9.03	240.04
16	4-Chlorophenyl	SO <sub>2</sub>	3	-8.35	756.45
17	4-Fluorophenyl	O	3	-8.58	513.06
18	4-Chlorophenyl	O	4	-9.02	244.13
19	3,4-Dichlorophenyl	CH <sub>2</sub>	0	-8.73	398.30

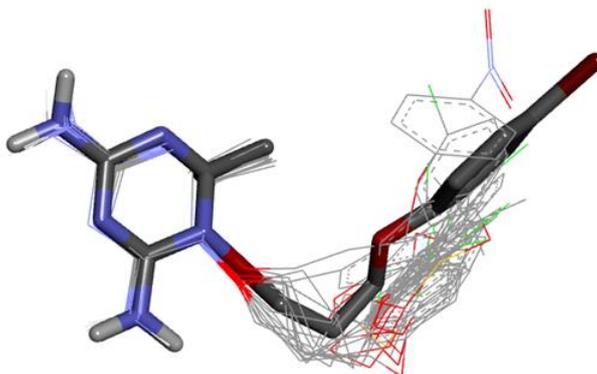
From table 2, our approach can classify the WR99210 series of compounds to be the effective compound (No.1-8, 10-11, and 13) with  $K_i$  lower than 181 nM, and non-effective compound (No.9, 12, and 14-19) with  $K_i$  greater than 181 nM. However, there is a small differ due to the experiment potency data obtained from an engineered strain of the budding yeast, but our docking calculations use the crystal structure of mtbDHFR. The lower in B.E., the fewer of  $K_i$  values which indicate the stronger binding affinity of compound on mtbDHFR.

The results indicate that compound No.8 is the most effective inhibitor with  $K_i$  value 41.48 nM. This molecule is composed of 2-(6-Bromonaphthyl) group at R position, three carbon atoms at the carbon linkage chain, and oxygen atom at X position.



**Figure 3.** Interactions of compound 8 with the amino acid residue within 3Å.

In order to determine the most effective inhibitor, the interaction of 19 potent inhibitors to bind with mtbDHFR and their binding energy are compared with the interaction of Br-WR99210 in mtbDHFR binding pocket (1DG7). Since the compound 8 provides lowest binding energy and it is superimposed in the active site of mtbDHFR with NADPH as shown in Figure 4 below, then this compound is selected to be the potent inhibitor which leads to be an effective anti-tuberculosis agent.



**Figure 4.** Superposition of WR99210 analogues which are bound in the same active site of mtbDHFR with NADPH.

The structure of compound 8 is rational to obtain the lowest B.E. and  $K_i$ . Therefore, compound 8 is expected to be the most potential inhibitor for further development of anti-tuberculosis drug.

## 5. CONCLUSION

According to the results from molecular docking of flexible ligand with fixed host, the compounds are classified to be an effective compound with  $K_i$  lower than 181 nM, and non-effective compound with  $K_i$  greater than 181 nM. The compounds which contain a naphthyl group show the stronger inhibition of the inhibitor to the mtbDHFR. In addition, for the carbon linker chain, the analogues with three carbons provide strongest inhibition. The compound that provide lowest B.E. and lowest  $K_i$  is the compound with 2-(6-bromonaphthyl). Therefore, choosing this compound as a ligand to bind with the enzyme of *M. tuberculosis* will be most effective.

## REFERENCES

1. D'Ambrosio, Lia, et al. "New anti-tuberculosis drugs and regimens: 2015 update." *ERJ Open Research* 1.1 (2015): 00010-2015.
2. Li, Rongbao, et al. "Three-dimensional structure of *M. tuberculosis* dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs." *Journal of molecular biology* 295.2 (2000): 307-323.
3. Pimonluck S., Pisanu T., Chawanee T., Ubolsree L., and Luckhana L. "Molecular Docking Study of *Mycobacterium tuberculosis* dihydrofolate reductase in complex with 2,4-Diaminopyrimidines Analogues." *Chiang Mai J. Sci.* 2015; 42(X): 1-15.
4. A'Lissa, B. Gerum, et al. "Novel *Saccharomyces cerevisiae* screen identifies WR99210 analogues that inhibit *Mycobacterium tuberculosis* dihydrofolate reductase." *Antimicrobial agents and chemotherapy* 46.11 (2002); 3362-3369.
5. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. *Gaussian 16*. Wallingford, CT2016.
6. *Gaussian 09, Revision B.01 Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT (2009)* by M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al.
7. Jnawali, Hum Nath, and Sungweon Ryoo. *First-and Second-Line Drugs and Drug Resistance*. INTECH Open Access Publisher, 2013.
8. Schnell, Jason R., H. Jane Dyson, and Peter E. Wright. "Structure, dynamics, and catalytic function of dihydrofolate reductase." *Annu. Rev. Biophys. Biomol. Struct.* 33 (2004): 119-140.
9. Schweitzer, BARRY I., Adam P. Dicker, and Joseph R. Bertino. "Dihydrofolate reductase as a therapeutic target." *The FASEB Journal* 4.8 (1990): 2441-2452.
10. GH, Shinde, S. S. Pekamwar, and S. J. Wadher. "An overview on dihydrofolate reductase inhibitors."
11. Kumar, Manoj, Rajakrishnan Vijayakrishnan, and Gita Subba Rao. "In silico structure-based design of a novel class of potent and selective small peptide inhibitor of *Mycobacterium tuberculosis* Dihydrofolate reductase, a potential target for anti-TB drug discovery." *Molecular diversity* 14.3 (2010): 595-604.
12. Bates, J. H. "Transmission and pathogenesis of tuberculosis." *Clinics in chest medicine* 1.2 (1980): 167.
13. Griffith, David E., and Clark M. Kerr. "Tuberculosis: disease of the past, disease of the present." *Journal of Perianesthesia Nursing* 11.4 (1996): 240-245.
14. Rao, A. S., and S. R. Tapale. "A STUDY ON DIHYDROFOLATE REDUCTASE AND ITS INHIBITORS: A REVIEW." *International Journal of Pharmaceutical Sciences and Research* 4.7 (2013): 2535.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by Thammasat University Research Fund under the TU Research Scholar, Contract No. TP 2/21/2559.

# Enantioselectivity and Enzyme-Ligand Docking Studies of pfDHFR and Cycloguanil compounds

S. Kulatee<sup>1</sup>, P. Toochinda<sup>1</sup>, L. Lawtrakul<sup>1, C</sup>

<sup>1</sup> School of Bio-Chemical Engineering and Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, 12121 Thailand

<sup>C</sup> E-mail: luckhana@siit.tu.ac.th; Fax: +66 2 986 9112~3; Tel. +66 2 986 9103~5 ext. 2301

## ABSTRACT

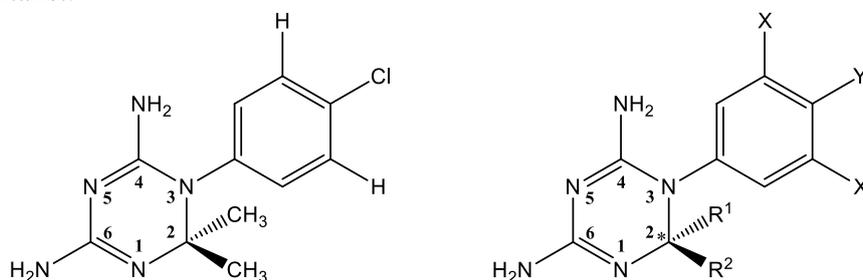
Cycloguanil (Cyc) derivatives were designed to tackle the fatal problem of *Plasmodium falciparum* Dihydrofolate Reductase (pfDHFR) resistance. However, there is no report on the drug's stereochemistry. This is very important because in the field of drug design, the consumption of pure enantiomeric drugs can lead to desirable as well as non-desirable biological responses. With the help of molecular modeling, Cyc derivatives were modeled into *R* and *S* enantiomers. pfDHFR enantioselectivity towards ligand binding was investigated. The binding interactions of wild-type and double mutant (A16V+S108T) pfDHFR complexed with Cyc and NADPH, with enantiomeric Cyc derivatives are discussed in details. Wild-type pfDHFR have preferential binding towards Cyc derivatives in *R* configuration with non-bulky substituents and vice-versa. While the mutant-type pfDHFR shows preferential binding only to Cyc derivatives in *R* configuration, irrespective to the substituent's bulkiness. The effect of *m*-Cl substitution at different *meta* positions lead to the changing of *m*-Cl positions after molecular docking simulation. This is the result of phenyl ring rotation to avoid steric clash with methyl group of residue 108 and flexible side chain of R<sup>1</sup> and R<sup>2</sup> substituents. As a result of steric clash, there is minor shifts in the phenyl ring and (1, 3, 5)-triazine ring.

**Keywords:** Chiral compounds, Antimalarial drugs, Cycloguanil enantiomers, Enzyme-ligand interaction, molecular docking, HF/6-31G (d,p)

## 1. INTRODUCTION

Malaria is caused by *Plasmodium falciparum* Dihydrofolate Reductase (pfDHFR). Cycloguanil (Cyc), one of the antimalarial, is used to treat and prevent the malaria infection by stopping the folic acid cycle in pfDHFR, which is responsible for the reproductive system of the *Plasmodium* parasites. Unfortunately, the treatment of malaria with antimalarial ceases due to the ability of pfDHFR to mutate itself to reduce the binding interactions of Cycloguanil and other antimalarials, rendering the inhibition of pfDHFR to be ineffective [1]. Thus, these antimalarials couldn't be used to treat complicated malarial cases. According to the literature review, the mutations observed in the mutant-type pfDHFR was associated with point mutations at the following amino acid positions: 16, 51, 59, 108 and 164 [2, 3]. Double point mutations at position 16 and 108 are conferred directly with Cyc resistance [1, 4]. The rapid emergence of antifolate resistant *P. falciparum* has unfortunately compromised the clinical utilities of drugs, and thus highlights the urgent need to search for new effective antifolate antimalaria [2]. To tackle this problem, Cyc derivatives were designed. The design can increase the inhibiting activity of Cyc derivatives with inhibition constant in the nanomolar level [2]. Unfortunately, there is no report on the drug's property towards inhibiting potential against pfDHFR (*R* and *S* enantiomer). This is important because stereochemistry plays an important role in the drug design and the responses of the drug to the body [2, 5, 6].

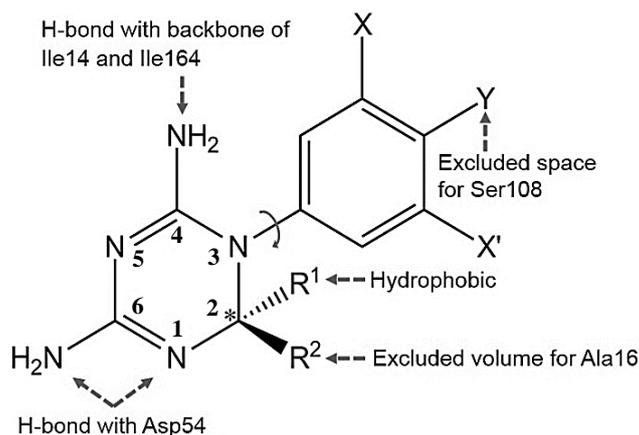
The chemical structure of Cyc and the general structure of Cyc derivatives are demonstrated in **Figure 1**. Cyc contains one 1, 3, 5-triazine ring (or s-triazine ring) bonded to the chlorophenyl ring at N3. In double mutant pfDHFR, Cyc compound experiences steric constraint with amino acid Valine16 (Val16) and Threonine108 (Thr108). The 2,2-dimethyl group of C2 experiences steric constraint with Val16 of mutant pfDHFR while the *para*-chlorine (or *p*-Cl) experiences steric constraint with Thr108. Cyc derivatives contain modification at C2 position and the position of chlorophenyl ring of the Cyc compound. At C2, the 2,2-dimethyl groups of Cyc is replaced with R<sup>1</sup> and R<sup>2</sup> substituents resulting in chiral carbon centre. Meanwhile, the chlorophenyl ring contains two *ortho* (non-labelled), two *meta* (X and X') and one *para* position (Y); refer to **Figure 1** for the abbreviations. Cyc derivatives were designed to have chlorine atom at either *meta* position (X in **Figure 1**) or *para* position (Y in **Figure 1**); refer to **Table 1** and **Table 2** for the substituents data. The new design of Cyc derivatives promised the clinical utilities with the inhibition constant as low as the nanomolar level [2] but there is no report on the stereochemistry of the designed drugs. As mentioned, taking of racemic drugs exert different biological responses to the body [2, 5, 6]. For some chiral drugs, only one enantiomer is effective, that would, in theory, only require half of the effective dose of a 50/50 racemic mixture [6]. Pure enantiomeric form of chiral drug that exert desirable effects and non-desirable effects is called eutomer and distomer respectively [6]. Taking mixture of eutomer and distomer (in a form of racemates) might lead to different biological responses like: (i) distomer is inactive, when compared to eutomer; (ii) distomer has the same biological activity as eutomer; (iii) distomer is less potent than eutomer; (iv) distomer acts as antagonist to eutomer; (v) distomer exert adverse effect to eutomer and (vi) distomer exert different therapeutic effect than eutomer [6]. The process to separate enantiomers (known as chiral resolution) is achievable by the following techniques: resolution by crystallization, chiral derivatizing agents, chiral column chromatography and enzyme-ligand complex formation. However, these processes are time-consuming and expensive and poses the limitations to experimental findings. Therefore, molecular modeling can be used as an alternative solution to predict and explain the experiment findings and help reduce time and cost for the experimentalist.



**Figure 1.** Chemical structure of Cyc (left) and Cyc derivatives (right).

## 2. THEORY AND RELATED WORKS

In order to verify the results obtained from AutoDock calculations, the essential characteristics of binding interactions are studied from the x-ray structure. According to [3], essential characteristics required for effective inhibiting activity are: (i) hydrogen bonding with both of the carboxyl oxygen of Asp54, as for the N1 and 6-N of dihydrotriazine ring of cycloguanil and equivalent groups of other inhibitors and (ii) hydrogen bonding with backbone oxygen of Ile164 and Ile14, as for the 4-N of dihydrotriazine ring of cycloguanil (iii) free space close to Ala16 in order to avoid steric clash and (iv) optimum rigid length from the 6-amino group to the distal end interacting with Ser108 (wild-type pfDHFR) and Thr108 (mutant-type pfDHFR). The essential binding characteristics of a good inhibitor of pfDHFR are illustrated in **Figure 2**. According to [7], Asp54 is the characteristic hydrogen bonding that is essential for pfDHFR catalytic activity. The analysis of binding interactions of x-ray structures of Cyc and pfDHFR verifies the hypothesis of essential binding characteristics [3, 7].



**Figure 2.** Essential characteristics of a good inhibitor of pfDHFR, as deduced from modeling of enzyme-inhibitor interaction [3].

### 3. COMPUTATIONAL DETAILS

The three-dimensional structures of Cyc derivatives were constructed and its geometry were optimized by HF/6-31G (d,p) calculations via GaussView05 and GUASSIAN09 software package respectively [8]. The crystal structure of wild-type and mutant-type pfDHFR were obtained from RSCB Protein Data Bank (PDB ID: 3UM8 and 3UM6 respectively) [4]. Water molecules were removed and hydrogen atoms were added to the pfDHFR molecules (enzymes) via Discovery Studio 4.0 [9]. Molecular docking of wild-type and mutant-type pfDHFR with enantiomeric Cyc compounds were carried out using AutoDock 4.2.6 software package [10]. All rotatable bonds within enantiomeric Cyc compounds were allowed to be freely flexible during the docking simulations, whereas enzyme structures were fixed as the rigid macromolecule. Gasteiger charges were assigned for the system. The Lamarckian Genetic Algorithm was used at 100 dockings for each Cyc compounds. The grid size of 0.375 Å spacing was set at specified grid points of 60 points in x-y-z directions. Other parameters were run at the program's default settings. Clusters obtained from AutoDock calculations are then compared to pfDHFR-Cyc interactions available in the crystal structures [4]. The best-fit configuration with the lowest binding energy was selected from each Cyc compounds for further analysis.

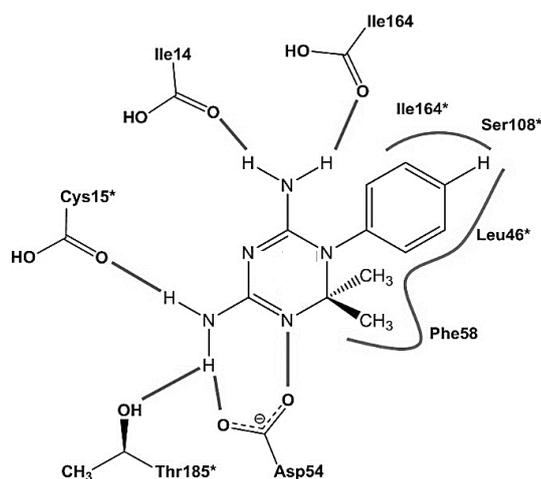
### 4. RESULTS AND DISCUSSION

The binding interaction of x-ray crystal structure (or reference structure) is shown in **Figure 3**. The hydrogen bondings present are the interactions between amine group (-NH<sub>2</sub>) of Cyc with Ile14, Cyc15, Asp54, Ile164 and Thr185 (shown as straight line) of pfDHFR binding pocket, whereas the hydrophobic interaction is observed to occur around the R<sup>1</sup> and R<sup>2</sup> flexible side chains and the chlorophenyl ring (shown as curved line). The comparison of binding interactions of Cyc derivatives with reference interaction shows that there are extra interactions with amino acid residues Cyc15, Leu46, Ser108 (wild-type pfDHFR), Thr108 (mutant-type pfDHFR), Ile164 and Thr185 are being observed (**Figure 3**). Amino acid residues with asterisk sign indicate the interactions that were observed apart from the reference structure.

The binding energy (B.E.) obtained from the AutoDock calculations is obtained via the formula  $B.E. = -RT \ln(K) = RT \ln(K_i)$ ;  $T = 298.15 \text{ K}$  and  $R = 8.314 \text{ JK}^{-1}\text{mol}^{-1}$ . Cluster with the lowest B.E. and also fulfill essential binding characteristics [3, 7] are selected for further analysis. The best clusters of Cyc derivatives in *R* and *S* configuration are then compared in terms of B.E. as shown in **Table 1** (wild-type) and **Table 2** (mutant-type). In **Table 1** and **Table 2**, Cyc derivatives are classified into *R* enantiomers and *S* enantiomers based on the priority of R<sup>1</sup> and R<sup>2</sup> substituents positions. Cyc derivatives with R<sup>1</sup> substituents projected into the paper and R<sup>2</sup> substituents projected out of the paper is classified as *R* enantiomer and vice-versa. Cyc derivatives with chlorine substitution at *para* position will have a total of two structures i.e., a structure in *R* configuration and a structure in *S* configuration. In contrary,

Cyc derivatives with chlorine substitution at two *meta* positions will have a total of four structures i.e., two structures in *R* configuration with *m*-Cl substitution at position X and X' respectively and two structures in *S* configuration with *m*-Cl substitution at position X and X' respectively.

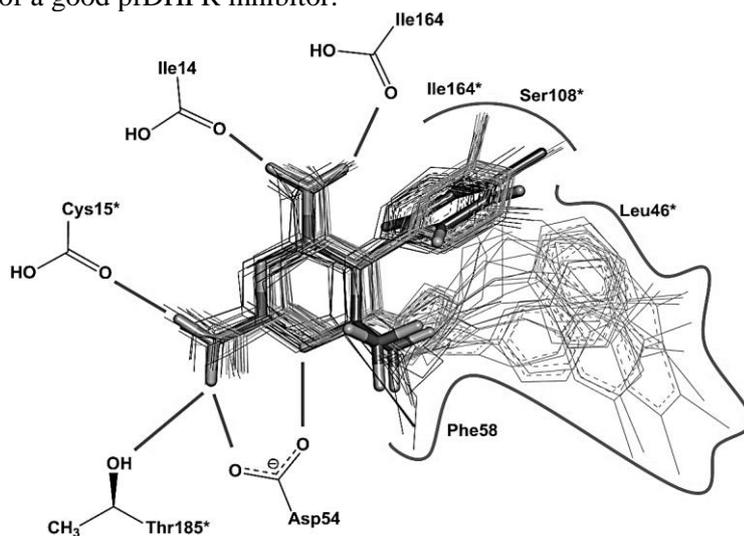
In **Table 1** (wild-type pfDHFR), Cyc derivatives with small substituents (non-bulky) have lower B.E.(s) when they are in *R* configuration and vice-versa. Cyc derivatives having lower B.E.(s) when in *R* configuration consist of Cyc24 up to 33 (except Cyc28 and 29) and the remaining Cyc derivatives which are Cyc34 up to 46 have lower B.E.(s) when in *S* configuration (except Cyc42). A closer examination reveals that R<sup>1</sup> and R<sup>2</sup> substituent's bulkiness have an effect on pfDHFR binding selectivity. This is probably due to the steric constraint that the substituents' side chain experience while binding to pfDHFR [2, 3]. In contrast to the **Table 1, Table 2** (mutant-type pfDHFR) data shows a completely different behavior. All Cyc derivatives have lower B.E.(s) when in *R* configuration except two candidates which are Cyc28 and 29 which have almost similar B.E. when in *R* and *S* configuration. In the binding interactions between Cyc derivatives and wild-type pfDHFR, Cyc derivatives with bulky substituents prefer to be in *S* configuration so that it can avoid steric clash with Ala16 side chain. The observed phenomena are not observed in mutant-type pfDHFR because all Cyc derivatives prefer to be in *R* configuration irrespective to the substituent's bulkiness. Thereby indicating the pfDHFR enantioselectivity to be more towards *R* enantiomers in this category.



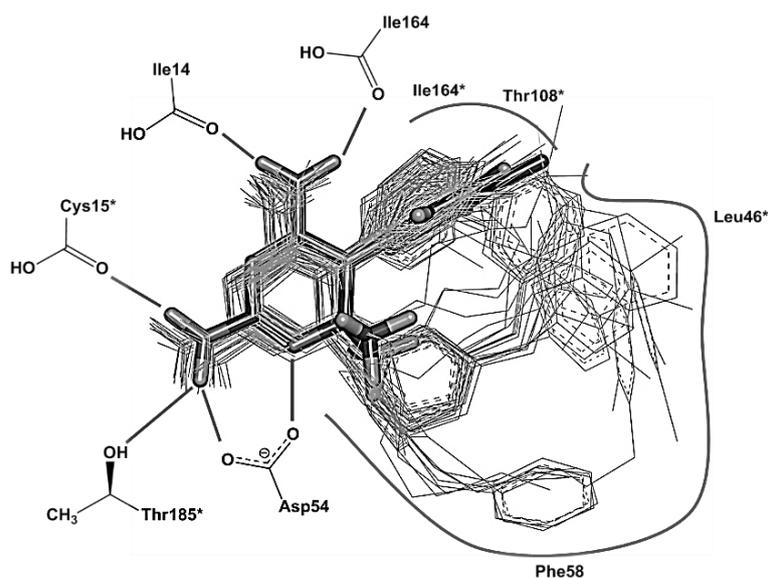
**Figure 3.** X-ray crystal structure (or reference structure) binding interactions. Hydrogen bondings and hydrophobic interactions are represented by straight lines and curve lines respectively. Amino acid residues with asterisk sign indicates extra interactions observed in Cyc derivatives but not in reference structure.

The effect of chlorine atom substitution at *para* and *meta* positions were investigated in the interactions of wild-type and mutant-type pfDHFR with enantiomeric Cyc derivatives. The results are drawn from the best-fit configuration with the highest frequencies. According to literature review, chlorine substitution at *meta* position can avoid the extension of the rigid length to the point of steric clash with the side chain of amino acid residue 108, and leading to good activity against resistant mutants (**Figure 2**) [2, 7, 11]. However, the B.E.(s) obtained do not show remarkable difference between *m*-Cl and *p*-Cl. The molecular docking simulation of Cyc derivatives with *p*-Cl and *m*-Cl reveals that both have the ability to form hydrophobic interaction with Leu46, Ser108 (wild-type pfDHFR), Thr108 (mutant-type pfDHFR) and Ile164. The binding interaction of *m*-Cl with Leu46 led to the decrease in overall binding energy while hydrophobic interactions with amino acid residue position 108 led to the increase in overall binding energy. The shift in the phenyl ring(s) of Cyc derivatives were observed after molecular docking simulation. This effect was due to a steric clash between

*m*-Cl and *p*-Cl with methyl group of residues 108 [4]. The effect propagates throughout the Cyc derivatives molecules resulting in a slight shift in *s*-triazine ring from the reference structure while maintaining essential binding characteristics. Likewise, the rotation of phenyl ring also contributes to minimize the effect of steric clash with methyl group of residue 108 and with the flexible side chains of substituents R<sup>1</sup> and R<sup>2</sup>. Such effect provides the interaction surface for *m*-Cl (*p*-Cl remains unaffected by the phenyl ring rotation). The avoidance of steric clash within pFDHFR binding pocket resulted in the change in the position of *m*-Cl from X to X' and vice-versa. The superposition images of Cyc derivatives with reference structure in the pFDHFR (both wild and mutant-type) binding pocket are shown in **Figure 4** and **Figure 5** respectively. There is a slight shift in the Cyc derivatives structure as described. However, most of them were able to fulfill the criteria of essential binding interactions of a good pFDHFR inhibitor.



**Figure 4.** Superposition of Cyc derivatives with inhibitor template (3UM8, wild-type); Cyc (stick model). Cyc derivatives in *R* and *S* configuration are shown as line model as light grey lines and dark grey lines respectively. Hydrogen bondings and hydrophobic interactions are viewed as straight lines and curve lines respectively.



**Figure 5.** Superposition of Cyc derivatives with inhibitor template (3UM6, mutant-type); Cyc (stick model). Cyc derivatives in *R* and *S* configuration are shown as line model as light grey lines and dark grey lines respectively. Hydrogen bondings and hydrophobic interactions are viewed as straight lines and curve lines respectively.

Another important factor that helps in predicting the superiority of enantiomeric form of Cyc derivatives is the absolute binding energy difference ( $|\Delta E|$  in kcal/mol). Because pure enantiomers can exert different biological responses [2, 5, 6], molecular modeling can be a helping hand in predicting which enantiomeric form of Cyc derivatives have better inhibiting potential. The values of  $|\Delta E|$  (in kcal/mol) are tabulated in **Table 1** and **Table 2** and the values range from negligible difference of 0.02 kcal/mol (Cyc28) up to very large difference of 4.19 kcal/mol (Cyc45). The  $|\Delta E|$  values increases as the substituents get bulkier. The authors believe that these data will help to predict whether that particular Cyc derivatives should be synthesized in either a mixture of enantiomers or a single pure enantiomeric form.

**Table 1.** AutoDock Binding Energy (B.E) comparison between Cyc derivatives in *R* and *S* configuration (wild-type pfDHFR)

Comp	3UM8 (wild-type pfDHFR)					<i>R</i>	<i>S</i>	$ \Delta E $
	X	X'	Y	R <sup>1</sup>	R <sup>2</sup>			
Cyc	H	H	Cl	Me	Me	-8.12	0.00	
23	Cl	H	H	Me	Me	-7.98	0.00	
23	H	Cl	H	Me	Me	-8.00 <sup>a</sup>	0.00	
24	H	H	Cl	Me	<i>n</i> Pr	-8.07	-6.85 <sup>c</sup>	1.22
25	Cl	H	H	Me	<i>i</i> Pr	-8.59	-7.30 <sup>c</sup>	1.29
25	H	Cl	H	Me	<i>i</i> Pr	-8.66 <sup>a</sup>	-7.32 <sup>d</sup>	1.34
26	H	H	Cl	Me	<i>i</i> Pr	-8.72	-7.12 <sup>c</sup>	1.60
27	Cl	H	H	Me	<i>n</i> Pr	-8.14	-8.01	0.13
27	H	Cl	H	Me	<i>n</i> Pr	-8.16 <sup>a</sup>	-8.08 <sup>a</sup>	0.08
28	H	H	Cl	Me	<i>n</i> Hex	-7.75 <sup>c</sup>	-8.26	0.51
29	Cl	H	H	Me	<i>n</i> Hex	-7.85	-8.17	0.32
29	H	Cl	H	Me	<i>n</i> Hex	-7.78 <sup>a</sup>	-8.29 <sup>a</sup>	0.51
30	H	H	Cl	H	Me	-8.34	-7.76	0.58
31	Cl	H	H	H	Me	-8.26 <sup>b</sup>	-7.83	0.43
31	H	Cl	H	H	Me	-8.24	-7.79 <sup>a</sup>	0.45
32	H	H	Cl	H	C <sub>6</sub> H <sub>5</sub>	-8.97	-8.39	0.58
33	Cl	H	H	H	C <sub>6</sub> H <sub>5</sub>	-8.80 <sup>b</sup>	-8.67	0.13
33	H	Cl	H	H	C <sub>6</sub> H <sub>5</sub>	-8.78	-8.61 <sup>a</sup>	0.17
34	H	H	Cl	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.57 <sup>c</sup>	-9.47	0.90
35	Cl	H	H	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.47 <sup>e</sup>	-9.97	1.50
35	H	Cl	H	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.40 <sup>c</sup>	-9.75 <sup>a</sup>	1.35

3UM8 (wild-type pfDHFR)								
Comp	X	X'	Y	R <sup>1</sup>	R <sup>2</sup>	R	S	\Delta E
36	H	H	Cl	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.60	-9.49	0.89
37	Cl	H	H	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.73 <sup>e</sup>	-9.85	1.12
37	H	Cl	H	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.62 <sup>d</sup>	-9.88 <sup>a</sup>	1.26
38	H	H	Cl	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-8.12	-8.82	0.67
39	Cl	H	H	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-9.31 <sup>e</sup>	-9.82	0.51
39	H	Cl	H	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-8.43 <sup>d</sup>	-9.82 <sup>a</sup>	1.39
40	H	H	Cl	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.82	-10.04	1.22
41	Cl	H	H	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-9.25 <sup>e</sup>	-10.40	1.15
41	H	Cl	H	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-9.19 <sup>d</sup>	-10.32 <sup>a</sup>	1.13
42	Cl	H	H	H	<i>n</i> C <sub>7</sub> H <sub>15</sub>	-8.39 <sup>b</sup>	-8.03	0.36
42	H	Cl	H	H	<i>n</i> C <sub>7</sub> H <sub>15</sub>	-8.25	-8.01 <sup>a</sup>	0.24
43	Cl	H	H	H	4-PrOC <sub>6</sub> H <sub>4</sub>	-7.43 <sup>e</sup>	-8.87	1.44
43	H	Cl	H	H	4-PrOC <sub>6</sub> H <sub>4</sub>	-7.55 <sup>d</sup>	-8.74 <sup>a</sup>	1.19
44	Cl	H	H	H	3-(3,5-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.90 <sup>e</sup>	-10.09	1.19
44	H	Cl	H	H	3-(3,5-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.69 <sup>c</sup>	-10.46 <sup>a</sup>	1.77
45	Cl	H	H	H	3-[2,4,5-Cl <sub>3</sub> C <sub>6</sub> H <sub>2</sub> O(CH <sub>2</sub> ) <sub>3</sub> O]C <sub>6</sub> H <sub>4</sub>	-9.18 <sup>e</sup>	-10.20	1.02
45	H	Cl	H	H	3-[2,4,5-Cl <sub>3</sub> C <sub>6</sub> H <sub>2</sub> O(CH <sub>2</sub> ) <sub>3</sub> O]C <sub>6</sub> H <sub>4</sub>	-8.26 <sup>c</sup>	-10.16 <sup>a</sup>	1.90
46	Cl	H	H	H	3-(3-CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.20 <sup>e</sup>	-9.98	1.78
46	H	Cl	H	H	3-(3-CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.28 <sup>c</sup>	-9.77 <sup>a</sup>	1.49

a = *m*-Cl at X' flip to X

b = *m*-Cl at X flip to X'

c = poor conformation

d = *m*-Cl at X' flip to X and poor conformation

e = *m*-Cl at X flip to X' and poor conformation

R = Cyc derivatives in R configuration

S = Cyc derivatives in S configuration

|\Delta E| = Absolute binding energy difference value between R and S configuration

**Table 2.** AutoDock Binding Energy (B.E) comparison between Cyc derivatives in *R* and *S* configuration (mutant-type pfDHFR)

3UM6 (Double mutant pfDHFR; A16V+S108T)								
Comp	X	X'	Y	R <sup>1</sup>	R <sup>2</sup>	<i>R</i>	<i>S</i>	\Delta E
Cyc	H	H	Cl	Me	Me	-7.70		0.00
23	Cl	H	H	Me	Me	-7.70		0.00
23	H	Cl	H	Me	Me	-7.77		0.00
24	H	H	Cl	Me	<i>n</i> Pr	-8.08	-7.09 <sup>c</sup>	0.99
25	Cl	H	H	Me	<i>i</i> Pr	-8.20 <sup>b</sup>	-7.41 <sup>c</sup>	0.79
25	H	Cl	H	Me	<i>i</i> Pr	-7.94	-7.47 <sup>d</sup>	0.47
26	H	H	Cl	Me	<i>i</i> Pr	-7.70	-7.37 <sup>c</sup>	0.33
27	Cl	H	H	Me	<i>n</i> Pr	-8.07	-7.57 <sup>b</sup>	0.50
27	H	Cl	H	Me	<i>n</i> Pr	-8.01 <sup>a</sup>	-7.59 <sup>a</sup>	0.42
28	H	H	Cl	Me	<i>n</i> Hex	-7.79	-7.81	0.02
29	Cl	H	H	Me	<i>n</i> Hex	-7.65 <sup>b</sup>	-7.62 <sup>b</sup>	0.03
29	H	Cl	H	Me	<i>n</i> Hex	-7.53 <sup>a</sup>	-7.79	0.26
30	H	H	Cl	H	Me	-7.98	-7.53	0.45
31	Cl	H	H	H	Me	-8.40 <sup>b</sup>	-7.48	0.92
31	H	Cl	H	H	Me	-8.37	-7.47 <sup>a</sup>	0.90
32	H	H	Cl	H	C <sub>6</sub> H <sub>5</sub>	-9.18	-7.21 <sup>c</sup>	1.97
33	Cl	H	H	H	C <sub>6</sub> H <sub>5</sub>	-9.34 <sup>b</sup>	-7.19	2.15
33	H	Cl	H	H	C <sub>6</sub> H <sub>5</sub>	-9.09	-6.97 <sup>d</sup>	2.12
34	H	H	Cl	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-9.19	-9.04 <sup>c</sup>	0.15
35	Cl	H	H	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-9.01 <sup>b</sup>	-7.22 <sup>c</sup>	1.79
35	H	Cl	H	H	4-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.86	-7.67 <sup>d</sup>	1.19
36	H	H	Cl	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.91	-6.81	2.10
37	Cl	H	H	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.50	-8.55 <sup>c</sup>	0.05
37	H	Cl	H	H	3-C <sub>6</sub> H <sub>5</sub> OC <sub>6</sub> H <sub>5</sub>	-8.41 <sup>a</sup>	-8.22 <sup>c</sup>	0.19
38	H	H	Cl	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-8.40	-6.74	1.66
39	Cl	H	H	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-8.14	-7.01	1.13

3UM6 (Double mutant pfDHFR; A16V+S108T)									
Comp	X	X'	Y	R <sup>1</sup>	R <sup>2</sup>	R	S	ΔE	
39	H	Cl	H	H	3-C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> OC <sub>6</sub> H <sub>4</sub>	-8.16	-8.10 <sup>c</sup>	0.06	
40	H	H	Cl	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.96	-7.27	1.69	
41	Cl	H	H	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.79	-7.39	1.40	
41	H	Cl	H	H	3-(4-ClC <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.57	-7.38 <sup>a</sup>	1.19	
42	Cl	H	H	H	<i>n</i> C <sub>7</sub> H <sub>15</sub>	-8.33 <sup>b</sup>	-7.59	0.74	
42	H	Cl	H	H	<i>n</i> C <sub>7</sub> H <sub>15</sub>	-8.19 <sup>a</sup>	-7.19 <sup>a</sup>	1.00	
43	Cl	H	H	H	4-PrOC <sub>6</sub> H <sub>4</sub>	-9.43 <sup>b</sup>	-8.05 <sup>e</sup>	1.38	
43	H	Cl	H	H	4-PrOC <sub>6</sub> H <sub>4</sub>	-9.43	-7.57 <sup>c</sup>	1.86	
44	Cl	H	H	H	3-(3,5-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.62	-6.93 <sup>c</sup>	1.69	
44	H	Cl	H	H	3-(3,5-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.30 <sup>a</sup>	-7.33 <sup>d</sup>	0.97	
45	Cl	H	H	H	3-[2,4,5-Cl <sub>3</sub> C <sub>6</sub> H <sub>2</sub> O(CH <sub>2</sub> ) <sub>3</sub> O]C <sub>6</sub> H <sub>4</sub>	-7.68	-3.49 <sup>b</sup>	4.19	
45	H	Cl	H	H	3-[2,4,5-Cl <sub>3</sub> C <sub>6</sub> H <sub>2</sub> O(CH <sub>2</sub> ) <sub>3</sub> O]C <sub>6</sub> H <sub>4</sub>	-7.88	-6.62	1.26	
46	Cl	H	H	H	3-(3-CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.62	-6.74	1.88	
46	H	Cl	H	H	3-(3-CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>4</sub>	-8.30 <sup>a</sup>	-8.61 <sup>d</sup>	0.31	

a = *m*-Cl at X' flip to X

b = *m*-Cl at X flip to X'

c = poor conformation

d = *m*-Cl at X' flip to X and poor conformation

e = *m*-Cl at X flip to X' and poor conformation

R = Cyc derivatives in R configuration

S = Cyc derivatives in S configuration

|ΔE| = Absolute binding energy difference value between R and S configuration

## 5. CONCLUSION

Molecular docking studies of wild-type and mutant-type pfDHFR with enantiomeric Cyc derivatives shows that the best-fit configurations have similar interactions with the reference structure with extra interactions like Cys15 and Thr185 (Hydrogen bonding) and Leu46, Ser108 (wild-type pfDHFR), Thr108 (mutant-type pfDHFR), and Ile164 (hydrophobic interactions) being present. A thorough analysis on binding interactions reveals Asp54 to be the characteristic hydrogen bonding. The comparison of B.E.(s) in **Table 1** predicted that Cyc derivatives with non-bulky substituents prefer to be in R configuration and vice-versa. While the comparison of B.E.(s) in **Table 2** predicted that Cyc derivatives with bulky and non-bulky substituents prefer to be in R configuration. The incidence of *m*-Cl changing its position after the course of molecular simulation is the result of phenyl ring rotation, in response to avoid steric clash with methyl group of Ser108 (wild-type pfDHFR) and Thr108 (mutant-type pfDHFR) and the flexible side chain of R<sup>1</sup> and R<sup>2</sup> substituents. As a result, there is a slight shift in phenyl rings being observed which propagated to the s-triazine ring of the compounds.

**REFERENCES**

1. Yuthavong, Y., et al., *J. Med. Chem.*, 2000, **43** (14), 2738-2744.
2. Kamchonwongpaisan, S., et al., *J. Med. Chem.*, 2004, **47** (3), 673-680.
3. Yuthavong, Y., *Microbes and Infect.*, 2002, **4** (2), 175-182.
4. Vanichtanankul, J., et al., *Antimicrob. Agents Chemother.*, 2012, **56** (7), 3928-3935.
5. Nguyen, L. A., He, H., and Pham-Huy, C., *Int. J. Biomed. Sci.: IJBS*, 2006, **2** (2), 85-100.
6. McConathy, J. and Owens, M. J., *Prim. Care Companion J. Clin. Psychiatry*, 2003, **5** (2), 70-73.
7. Sirawaraporn, W., et al., *Mol. Biochem. Parasitol.*, 2002, **121** (2), 185-193.
8. Gaussian 09, Revision A.02, Frisch, M. J., Gaussian, Inc., Wallingford CT, 2017.
9. Accelrys, D. s. m. e. r. A. I., San Diego, CA, 2017.
10. Morris, G. M., et al., *J. Comp. Chem.*, 2009, **16**, 2785-91.
11. McKie, J. H., et al., *J. Med. Chem.*, 1998, **41** (9), 1367-1370.

**ACKNOWLEDGMENTS**

S. Kulatee was financially supported by the Excellent Thai Students (ETS) scholarship program of Sirindhorn International Institute of Technology (SIIT). The authors gratefully acknowledge the Center of Nanotechnology, Kasetsart University for the GUASSIAN09 program package and the financial support provided by Thai Government Research Fund Contract No. 34/2560.

# Computational Docking Study of 4-oxo-*N*-(2-(piperidin-1-yl)ethyl)-4*H*-chromene-2-carboxamide as Acetylcholinesterase and Butyrylcholinesterase Inhibitors

P. Suwanhom<sup>1</sup>, V. S. Lee<sup>2</sup>, T. Nualnoi<sup>3</sup>, L. Lomlim<sup>1,C</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat Yai, Songkhla, 90110, Thailand

<sup>2</sup>Department of Chemistry, Faculty of Science, University of Malaya, Kuala Lumpur, 50603, Malaysia

<sup>3</sup>Department of Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat Yai, Songkhla, 90110, Thailand

<sup>C</sup>E-mail: luelak.l@psu.ac.th; Fax: +66 74 428 239; Tel. +66 74 288929

## ABSTRACT

Alzheimer's disease is disorder of the central nervous system that leads to loss of memory, impairment of language and changes of behavior. The majority of people with this disease are 65 years and older. Donepezil drug has been used for treatment of mild to moderate severe cases. This drug is an acetylcholinesterase inhibitor (AChEI) widely used in the clinic. Taking Donepezil and *N*-(4-((2-methoxybenzyl)(ethyl)amino)butyl)-5-hydroxy-4-oxo-4*H*-chromene-2-carboxamide (**B**) as lead molecules, we designed and synthesized 4-oxo-*N*-(2-(piperidin-1-yl)ethyl)-4*H*-chromene-2-carboxamide (**1**) as AChEI. This compound showed moderately potent cholinesterase inhibitory activities having IC<sub>50</sub> values 3.22±0.12 μM for *ee*AChE and 20.59±1.04 μM for *eq*BChE. In contrast, Donepezil showed potent AChEI activity (IC<sub>50</sub>=0.006±0.001 μM) and selective to AChE. In order to obtain insights of interactions between these ligands and the enzymes, molecular docking study was performed by using *Torpedo californica* acetylcholinesterase (*Tc*AChE, PDB ID: 1EVE) and *Human* butyrylcholinesterase (*Hu*BChE, PDB ID: 4BDS). In complex with *Tc*AChE, compound **1** showed higher interaction energy than **B** and Donepezil. The same trend was observed in complex with BChE. Compound **1** acted as dual binding site which had interaction with amino acid residues both in peripheral anionic site (PAS) and catalytic anionic site (CAS) of the enzymes. This information is valuable for further development of new AChEIs with higher potency.

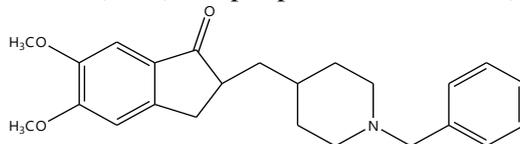
**Keywords:** Acetylcholinesterase inhibitor, Butyrylcholinesterase inhibitor, Molecular docking.

## 1. INTRODUCTION

Nowadays, more than six hundred thousand of Thai people are suffered from Alzheimer's disease (AD). This disease can affect different people in different ways, the most common symptom begins with gradual difficulty in remembering new information [1]. The cholinergic hypothesis states that deficits in function of the central nervous system result from a loss of acetylcholine (ACh). ACh can be demoted by two enzymes of cholinesterase, namely acetylcholinesterase (AChE) and butyrylcholinesterase (BChE).

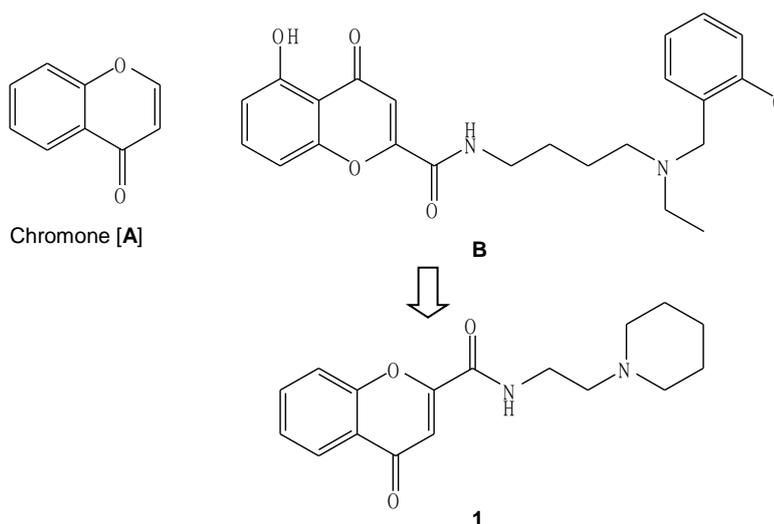
In healthy brains, AChE hydrolyzes the major of ACh, while BChE only plays a secondary role. However, in AD patients, 45% of AChE may be lost in certain brain regions during progression of this disease, while BChE activity inversely increases by up to 90%. Therefore, inhibition of both enzymes has been accepted as significant strategies for the effective management of this disease by increasing of synaptic ACh in the brain [2].

Donepezil is an AChE inhibitor (AChEI) that is currently approved for the palliative treatment of AD. The drug Donepezil has selectivity for AChE as compared to BChE (Figure 1). It also acts as a dual binding site inhibitor by interacting with various amino acid residues both in the catalytic anionic site (CAS) and peripheral anionic site (PAS) of the enzyme. [3].



**Figure 1.** Chemical structure of Donepezil.

Chromone derivatives have attracted much attention in recent years because of their diverse pharmacological properties such as anti-inflammatory effect, metal chelating ability and neuroprotective effect. Chromone [A] is a versatile scaffold for design of many bioactive molecules including acetylcholinesterase inhibitors (AChEIs) [4]. *N*-(4-((2-methoxybenzyl)(ethyl)amino)butyl)-5-hydroxy-4-oxo-4*H*-chromene-2-carboxamide (B) was reported as an AChEI ( $IC_{50} = 0.55 \pm 0.03 \mu\text{M}$ ) with pharmacological properties related to neurological disorders such as anti-inflammatory, anti-oxidant, neuroprotective effect, metal-chelating properties and the  $A\beta$  aggregation.



**Figure 2.** Designing strategy of compound 1.

Recently, we designed and synthesized 4-Oxo-*N*-(2-(piperidin-1-yl)ethyl)-4*H*-chromene-2-carboxamide (1) (Figure 2). This compound could bind to both CAS and PAS of AChE as suggested by molecular docking study [5]. AChE and BChE inhibitory activity of compound 1 was determined by Ellman's method using AChE from electric eel (*ee*AChE) and BChE (*eq*BChE) from horse serum with Donepezil used as a positive control [6]. Compound 1 displayed moderately potent *ee*AChE inhibitory activity at the micromolar range ( $IC_{50} = 3.22 \pm 0.12 \mu\text{M}$ ) but demonstrated weak *eq*BChE inhibitory activity ( $IC_{50} = 20.59 \pm 1.04 \mu\text{M}$ ). In contrast, as compared to Donepezil, compound 1 is more potent and selective to AChE ( $IC_{50} = 0.006 \pm 0.001 \mu\text{M}$  for *ee*AChE and  $IC_{50} = 1.06 \pm 0.01 \mu\text{M}$  for *eq*BChE, Table 1).

In this study, molecular docking was carried out by using AutoDock Vina 1.1.2 and Discovery Studio 2.5 to explore binding affinities and interactions of the ligands including compounds 1, B and donepezil with AChE and BChE.

**Table 1.** Acetylcholinesterase and butyrylcholinesterase inhibitory activities.

Compound	IC <sub>50</sub> ± SEM ( <i>ee</i> AChE)	IC <sub>50</sub> ± SEM ( <i>eq</i> BChE)	Selectivity* index
<b>1</b>	3.22 ± 0.12	20.59 ± 1.04	6.39
<b>Donepezil</b>	0.006 ± 0.001	1.06 ± 0.01	176.67
<b>B</b>	0.55 ± 0.03	ND	ND

\* Selectivity index = IC<sub>50</sub> of *eq*BChE/IC<sub>50</sub> of *ee*AChE

## 2. THEORY AND RELATED WORKS

Molecular docking is the common method nowadays to predict the protein-ligand binding structure with a protein of known three-dimensional (3D) structure. In this study, AutoDock Vina 1.1.2 [7] and Discovery Studio 2.5 programs was used for molecular docking and visualization, respectively. The improving of the speed and accuracy of docking with a new empirical scoring function, efficient optimization and multithreading on multi-core machines with AutoDock Vina and the automatic calculation of the grid maps and clusters make the program easier to use. The calculation of the Gibbs free energy ( $\Delta G$ ) of binding between the compound of interest and the enzyme was based on evaluation of the total energy of intermolecular forces including hydrogen bond, hydrophobic, electrostatic and van der Waal's forces. The conformation of ligands were arranged by utilizing the calculated  $\Delta G$  values; lower  $\Delta G$  values correspond to more desirable ligand binding, while higher  $\Delta G$  values are less desirable [8]. The docking score is a predicted binding affinity parameter expressed in kcal/mol.

## 3. COMPUTATIONAL DETAILS

**Molecular Docking Study.** To predict the binding mode of each compound in the active site of each enzyme, 3D structures of *Torpedo californica* AChE (*Tc*AChE, PDB ID: 1EVE) and human BChE (*Hu*BChE, PDB ID: 4BDS) obtained from protein data bank (<http://www.pdb.org>) were used. The 3D structure of *Tc*AChE used was AChE in complex with Aricept, a member of N-benzylpiperidine-based AChEIs whereas that of *Hu*BChE used was in form of BChE-Tacrine complex. To facilitate the docking studies, the 3D structures of enzymes were prepared under the protein preparation protocol implemented in Discovery Studio 2.5. PROPKA was used to assign protonation states at pH=7 while the ligands (compound **1**, **B**, Donepezil) were optimized using density functional theory at B3LYP/6-311G (d, p) level. Molecular docking studies with AutoDock Vina 1.1.2 [7] were performed towards the catalytic anionic site (CAS) and peripheral anionic site (PAS) of enzymes. For *Tc*AChE, the center of the grid box was placed at the bottom of the active site gorge ( $x=2.023$ ,  $y=63.295$ ,  $z=67.062$ ) and the dimensions of the active site box were set at 50 Å x 50 Å x 50 Å. For *Hu*BChE, the center of the grid box determined at  $x=135.117$ ,  $y=119.222$ ,  $z=40.667$  and the dimensions of the active site box were set at 34 Å x 46 Å x 34 Å. The exhaustiveness was set to the default value of eight [8] for all the docking runs. Different conformations of the ligands, known as binding modes, were obtained with their respective binding energy/affinity. The structure with the lowest binding energy/affinity was picked as the pose and was employed in the post-docking analysis for the residue interaction energy of the docked structure using the Discovery Studio 2.5 with CHARMM force field [9].

## 4. RESULTS AND DISCUSSION

The docking studies of compound **1**, **B**, Donepezil resulted in the low negative binding energy (below -10 kcal/mol) which indicated strong favorable binding for both enzymes (*Tc*AChE and *Hu*BChE). The lowest binding energy conformer(s) of each ligand were picked for further investigation of the residue interaction. Total interaction energy values and key interactions between the ligands and the enzymes were summarized in Table 2.

Compound **1**-*TcAChE* showed total interaction energy of -98.79 kcal/mol. This compound occupied the PAS, mid-gorge and CAS of the enzyme (Figure 3a). The key interactions include: chromone moiety interacted with PAS of the active site of enzyme and exhibited  $\pi$ - $\pi$  interaction with Trp278 (distance of 3.84 Å); methylene chain of the compound showed hydrophobic interaction with Phe330 in the mid-gorge; and piperidine ring of the compound formed cation- $\pi$  interaction with Phe329 (distance of 3.54 Å), similarly to the interactions found in compound **B**-*TcAChE* complex shown in Figure 4a. For compound **1**-*HuBChE*, it had a total interaction energy of -72.40 kcal/mol. The key interactions include: hydrogen bondings between chromone? and Phe115 and Gln116 of the enzyme (distance 2.90 Å and 2.80 Å, respectively); and cation- $\pi$  interaction of piperidine ring with Trp79 (distance 3.97 Å) (Figure 3b).

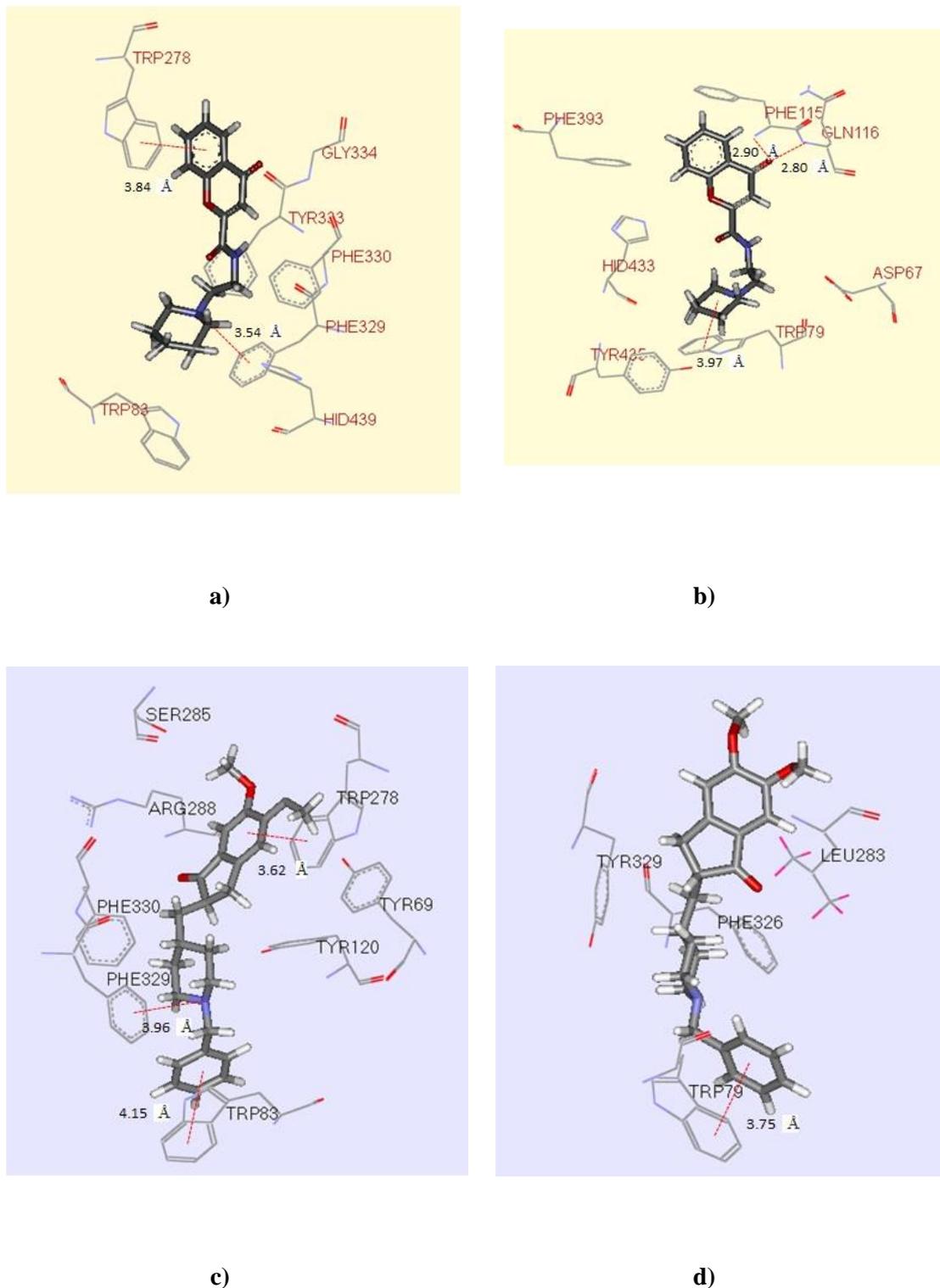
**Table 2.** Total interaction energy and key enzyme-ligand interactions.

Compound	<i>Torpedo californica</i> Acetylcholinesterase ( <i>TcAChE</i> ) PDB ID : 1EVE		<i>Human</i> Butyrylcholinesterase ( <i>HuBChE</i> ) PDB ID : 4BDS	
	Type of interaction (amino acid residue)	Total interaction energy (kcal/mol)	Type of interaction (amino acid residue)	Total interaction energy (kcal/mol)
<b>1</b>	$\pi$ - $\pi$ (Trp278), cation- $\pi$ (Phe329), Hydrophobic (Phe330)	-98.79	cation- $\pi$ (Trp79), Hydrogen bond (Phe115, Gln116)	-72.40
<b>Donepezil</b>	$\pi$ - $\pi$ (Trp83), $\pi$ - $\pi$ (Trp278), Cation- $\pi$ (Phe329), Hydrophobic (Phe330)	-119.63	$\pi$ - $\pi$ (Trp79), Hydrophobic (Tyr329)	-102.80
<b>B</b>	$\pi$ - $\pi$ (Trp83), $\pi$ - $\pi$ (Trp278), Hydrophobic (Phe329, Phe330)	-109.45	$\pi$ - $\pi$ (Trp79), Hydrogen bond (Ser195)	-94.29

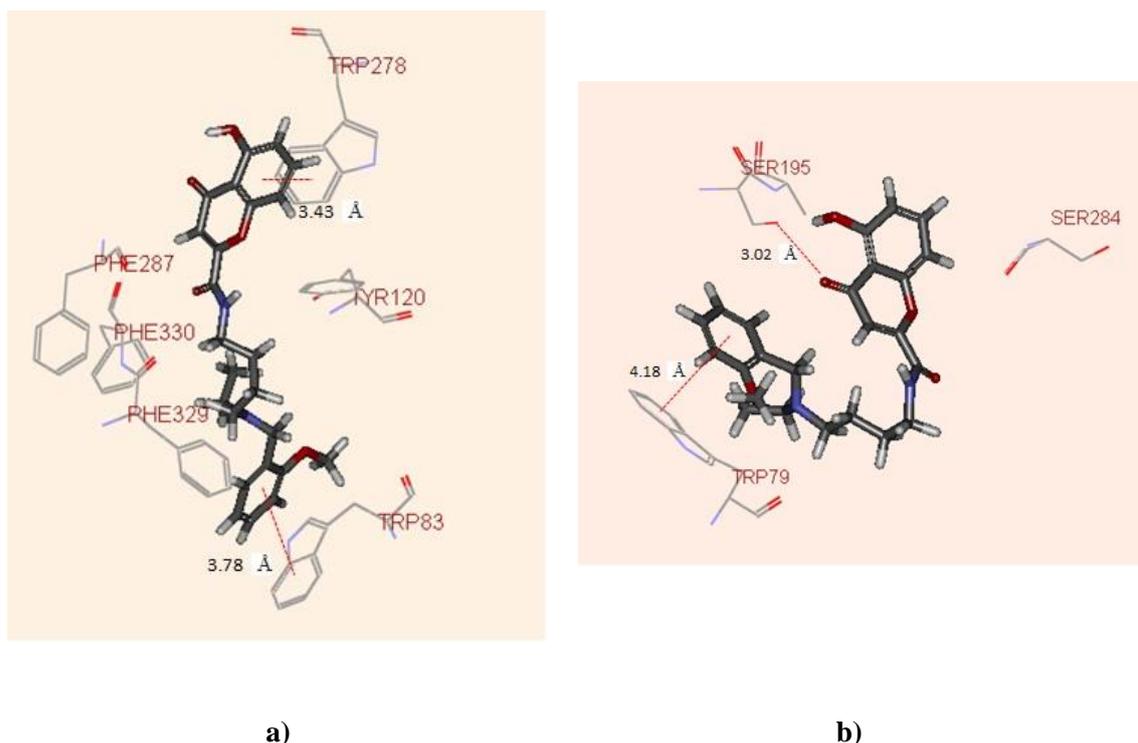
Total interaction energy of Donepezil-*TcAChE* was -119.63 kcal/mol. As shown in Figure 3c, chromone showed  $\pi$ - $\pi$  interaction with Trp278 (distance 3.62 Å) at the PAS of the enzyme. At the mid-gorge, the long chain of methylene interacted with Phe330 via hydrophobic interaction. At the CAS, piperidine ring exhibited cation- $\pi$  interaction with Phe329 (distance 3.96 Å) and benzyl group of the compound formed  $\pi$ - $\pi$  interaction with Trp83 (distance 4.15 Å). For Donepezil-*HuBChE*, the total interaction energy calculated was -102.80 kcal/mol. In the complex structure, the hydrophobic interaction of the compound with Tyr329 at mid-gorge of the enzyme and  $\pi$ - $\pi$  interaction between the benzyl group of the compound and Trp79 (distance 3.75 Å) were observed (Figure 3d).

The docking study presented that Compound **B** could bind to PAS, mid-gorge and CAS of *TcAChE*. This complex structure showed a total interaction energy of -109.45 kcal/mol. As illustrated in Figure 4a, the chromone moiety interacted with PAS of the enzyme and showed  $\pi$ - $\pi$  interaction with Trp278 (distance 3.43 Å). At the mid-gorge, the methylene chain of Compound **B** interacted with Phe329 and Phe330 via the hydrophobic interaction. At the CAS,  $\pi$ - $\pi$  interaction between benzyl group of the compound and Trp83 (distance 3.78 Å) was observed as well. This might be responsible for higher affinity of compound **B** as compared to that of compound **1**. For compound **B**-*HuBChE*, a total interaction energy was -94.29 kcal/mol. As shown in Figure 4b, 4-carbonyl group at the chromone formed hydrogen bond with Ser195 (distance 3.02 Å) and benzyl group of the compound exhibited  $\pi$ - $\pi$  interaction with Trp79 (distance 4.18 Å).

Therefore, compound **B** showed higher total interaction energy than compound **1** but both compounds had less total interaction energy than Donepezil (Table 2). These results were consistent with the IC<sub>50</sub> values (Table 1).



**Figure 3.** Molecular docking study of compound **1** and Donepezil. a) **1**-*TcAChE*, b) **1**-*HuBChE*, c) Donepezil-*TcAChE*, and d) Donepezil-*HuAChE*.



**Figure 4.** Molecular docking study of compound **B**. a) **B-TcAChE** and b) **B-HuBChE**.

## 5. CONCLUSION

Computational docking revealed that **1**, **B** and Donepezil acted as dual binding site inhibitors which had interactions with various amino acid residues both in the CAS and PAS of the enzymes. Compound **1** yielded lower negative value of total interaction energy than **B** and Donepezil which indicated that Donepezil has the strongest interaction with AChE and BChE. These findings correlated well with the biological assay results. This information is valuable for further development of new AChEIs with higher potency.

## REFERENCES

1. Alzheimer's Association, *Alzheimer's & Dementia*, 2009, **5**, 270-234.
2. Greig, N.H., Lahiri, D.K., and Sambamurti, K., *Int Psychogeriatr*, 2002, **14**, 91-77.
3. Mayeux, R., M.D., and Sano, M., *N Engl J Med*, 1999, **341**, 1679-1670.
4. Grazul, M., and Budzisz, E., *Coord Chem Rev*, 2009, **253**, 2598-2588.
5. Liu, Q., Qiang, X., and Li, Y., *Bioorg Med Chem*, 2015, **23**, 923-911.
6. Ellman, G., Courtney, K., and Andres, V., *Biochem Pharm*, 1961, **7**, 95-88.
7. Trott, O., and Olson, A.J., *J. Comput. Chem*, 2010, **31**(2), 461-455.
8. Jacob, R.B., Andersen, T., and McDougal, O.M., *PLoS Comput Biol*, 2012, **8**(5).
9. Brooks, B.R., Brooks III, C.L., MacKerell Jr, A.D., Nilsson, L., and et al., *J Comput Chem*, 2009, **30**(10), 1614-1545.

## ACKNOWLEDGMENTS

This work was supported by Prince of Songkla University (Contract Number PHA570404S).

# Molecular Modelling of 5-Lipoxygenase Enzyme with Antiasthmatic Substances from Plai

K. Jitapunkul<sup>1</sup>, O. Poachanukoon<sup>2</sup>, S. Hannongbua<sup>3</sup>, P. Toochinda<sup>1</sup>, and L. Lawtrakul<sup>1,C</sup>

<sup>1</sup> School of Bio-Chemical Engineering and Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, 12121, Thailand

<sup>2</sup> Center of Excellence for Allergy, Asthma and Pulmonary diseases, Faculty of Medicine, Thammasat University, Pathum Thani, 12121, Thailand

<sup>3</sup> Department of Chemistry, Kasetsart University, Chatuchak, Bangkok, 10900, Thailand

<sup>C</sup> E-mail: luckhana@siit.tu.ac.th; Tel. +66 2 986 9009 ext. 2301

## ABSTRACT

Bioactive compounds from Plai or *Zingiber cassumunar Roxb.* have antiasthmatic effect such as Compound D and DMPBD. Unfortunately, the inhibition mechanisms of these compounds are not clearly understood. Consequently, molecular modellings are used to study about complex systems between these compounds and a possible protein target, 5-Lipoxygenase enzyme (5-LO), which involve with stimulation of asthma symptoms. Our simulations showed that Compound D and DMPBD have opportunities to be antiasthmatic drugs with competitive inhibition effect because they bind at the same binding site of 5-LO natural substrate (arachidonic acid, AA) and commercial asthma drug (Zileuton).

**Keywords:** Molecular Dynamics Simulation, Asthma, Ligand-Protein Interaction.

## 1. INTRODUCTION

In 2014, The Global Asthma Network reported that approximately 334 million people suffered from asthma<sup>1</sup>. In general, asthma drug (inhaled corticosteroids) is used to control symptoms until it reaches resting-state. By using inhaled corticosteroids, the long-term side effects may occur. However, the corticosteroids are still used in asthma treatment due to their low market price. Therefore, the development for alternative asthma drugs with lower price and less side effects are necessary<sup>2-3</sup>. In the past decade, many research groups aimed to study about anti-leukotriene substances in order to use them as alternative asthma drugs. Some are used in clinical practice for curing asthma such as Zileuton (ZYFLO CR®)<sup>4</sup>. Nevertheless, those medications have lower efficacy than corticosteroids and contain many side effects<sup>5</sup>. Plai or *Zingiber cassumunar Roxb.* is a herb found in the South East Asia region. The major bioactive components of this plant are Compound D {(E)-4-(3',4'-dimethoxyphenyl)but-3-en-2-ol} and DMPBD {(E)-1-(3',4'-dimethoxyphenyl)-butadiene}, which have antiasthmatic activities<sup>6</sup>. Unfortunately, their mechanisms are not fully understood. The possible target for these substances is 5-Lipoxygenase enzyme (5-LO) because the constriction and inflammation of guinea pigs airways can be reduce by 5-LO inhibition<sup>7</sup>. Thus, molecular modelling approaches are used to investigate ligand-protein complexes in order to suggest opportunities for using Compound D and DMPBD as the novel antiasthmatic drugs in comparison with 5-LO natural substrate (arachidonic acid, AA) and asthma drug (Zileuton).

## 2. THEORY AND RELATED WORKS

Human 5-Lipoxygenase (5-LO) is an enzyme with 673 amino acids residues. It consists of two major domains: N-terminal  $\beta$ -sandwich domain (residues 1-114), and C-terminal catalytic domain (residues 121-673)<sup>8</sup>. The substrate binding site is located on the catalytic domain, composed of Phe177, Tyr181, His367, Leu368, His372, Leu414, Leu420, Phe421, Ala603, Leu607, and Ile673<sup>9</sup>. It catalyzes the oxygenation reaction of arachidonic acid (AA)

to form 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and then converts 5-HPETE to leukotriene A<sub>4</sub> (LTA<sub>4</sub>) by dehydration<sup>10,11</sup>. After those enzymatic processes, LTA<sub>4</sub> is immediately converted to leukotriene B<sub>4</sub> (LTB<sub>4</sub>) and cysteinyl leukotrienes (CysLTs)<sup>12,13</sup>. CysLTs mediate a biological response through G-protein couple receptors called cysteinyl leukotriene receptors (CysLTRs)<sup>14</sup>, which lead to airway injuries and smooth muscle contraction<sup>15</sup>.

Plai or *Zingiber cassumunar Roxb.* is a traditional herb that can be used in medical practices. Thai herbal pharmacopoeia (Vol. 1) indicated that 50% of Thai prescriptions for asthma treatment contain Plai<sup>16</sup>. Moreover, Plai has antiasthmatic activity in children with no acute toxicology effects. Pharmacological properties of isolated compounds from Plai revealed that there are two compounds in the phenylbutanoid group which have antiasthmatic activities. These compounds are Compound D ((E)-4-(3, 4'-dimethoxyphenyl) but-3-en-1-ol)) and DMPBD ((E)-1-(3',4'-dimethoxyphenyl)-butadiene)<sup>6</sup>. Jeenapongsa et al. (2003) and Pongprayoon et al. (1997) reported that DMPBD and Compound D can be used to reduce ear edema in rats (ear edema induced by AA or tetradecanoylphorbol acetate, TPA). Therefore, they suggested that DMPBD and Compound D can inhibit 5-LO pathways<sup>7,17</sup>.

### 3. COMPUTATIONAL DETAILS

#### 1. Molecular preparation and optimization

AA structure was extracted from x-ray structure downloaded from RCSB protein data bank (PDB ID: 3V99). Zileuton, Compound D, and DMPBD molecules were constructed by GaussView05 program. Afterward, perform structural optimization by density functional theory (DFT) at B3LYP/6-31G (d,p) level by using Gaussian09 program. Unfortunately, the x-ray of substrate-bound 5-LO structure (PDB ID: 3V99) is not complete, some amino acids are missing. Thus, the complete x-ray structure of substrate-free 5-LO was downloaded (PDB ID: 3O8Y) in order to use it as host for molecular modelling simulations.

#### 2. Molecular docking calculations

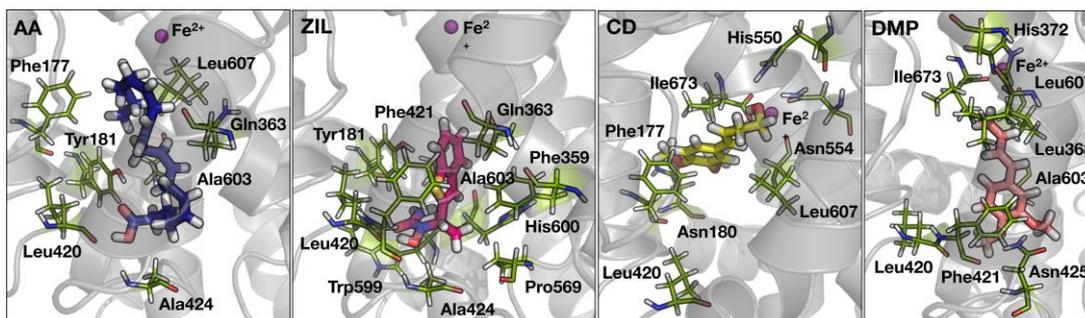
Molecular dockings were performed using the AutoDock 4.2 software package. All torsional angles within ligands were set free to perform flexible ligand docking. Gasteiger charges were assigned for all molecules. The Lamarckian Genetic Algorithm was used at 100 dockings for each ligand. All other parameters were run at program's default settings.

#### 3. Molecular dynamics (MD) simulations

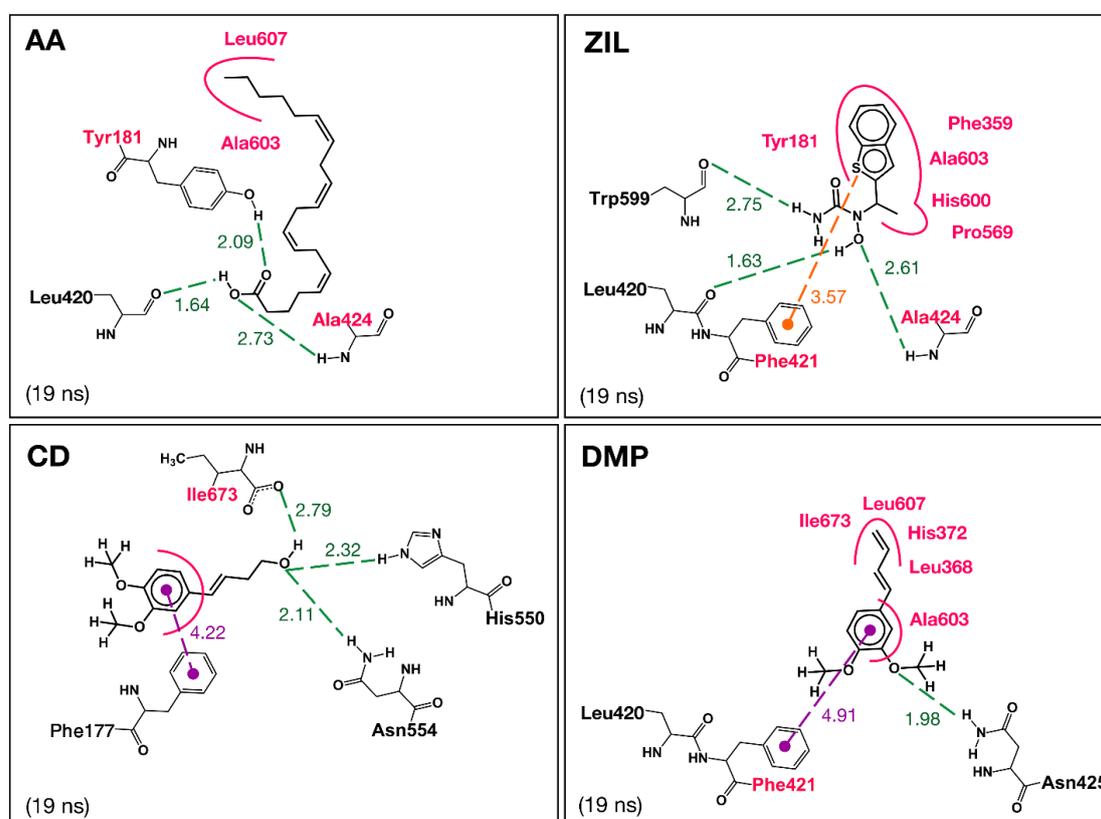
MD simulations were performed with AMBER force field in AMBER12 program. The stable prediction of 5-LO complex systems from molecular docking are solvated with a periodic truncated octahedral water box. Each of 5-LO system was started from a temperature of 0 K and gradually heated up to 300 K over 20 ps with constant volume. The whole system was equilibrated for 20 ns at 300 K and 1 atm (isothermal–isobaric ensemble, NPT) with 2 fs time steps. Several thermodynamic properties can be extracted from the output files. In addition, the free binding energy estimation by MM/GBSA (Molecular Mechanics Generalized Born Surface Area) method is used to calculate the binding energy between each ligand and specific amino acid residues of 5-LO.

### 4. RESULTS AND DISCUSSION

MD simulations showed that Zileuton, Compound D, and DMPBD are possible to bind at substrate (AA) binding site of 5-LO as shown in Figure 1. AA and Zileuton have similar binding mode with different binding affinity. The hydroxyl group of AA and Zileuton occur two hydrogen bonds with Leu420 and Ala424. Even though Compound D and DMPBD binding at the same catalytic site but molecular interactions with each amino acid residues are different as shown in Figure 2.



**Figure 1.** 5-LO binding site and interacting amino acid residues from MD models at 19 ns. 5-LO structure shown as ribbon. The amino acid residues shown as stick models. Arachidonic acid (AA), Zileuton (ZIL), Compound D (CD) and DMPBD (DMP) shown as blue, magenta, yellow, and pink stick models, respectively.



**Figure 2.** Ligand-enzyme interactions at 19 ns of MD simulations. Hydrogen bonds are shown as green dashed lines. Hydrophobic interactions are shown as pink curved lines. The pi-pi interactions are shown as purple dashed lines, and the pi-sulfur interaction is shown as an orange dashed line. The distances are in Å units.

The binding energy of 5-LO with AA (-48.21 kcal/mol) is much lower than binding with Zileuton (-29.40 kcal/mol), Compound D (-26.83 kcal/mol) and DMPBD (-29.15 kcal/mol) due to the van der Waals interaction between its hydrocarbon chain with many amino acid residues. The overall structures of 5-LO complexed with each ligands are very similar (RMSD 1.66 Å - 1.79 Å) which indicated that the 5-LO enzyme has the same dynamics motion even though it was binding with different ligands. This also support our assumption that Compound D and DMPBD are competitive inhibitors, same as Zileuton. The binding affinities of Compound D and DMPBD on 5-LO enzyme also reveal the opportunities for using them as the novel antiasthmatic drugs.

## 5. CONCLUSION

Our MD simulations showed that Compound D and DMPBD molecules bind at the same binding site of 5-LO natural substrate (AA) which is similar to the binding of a commercial asthma drug (Zileuton). The binding energy ( $\Delta G$ ) of the 5-LO complex with Compound D and DMPBD from Molecular Mechanics Generalized Born Surface Area (MM/GBSA) calculations are -26.83 and -29.15 kcal/mol, respectively. The dynamics and the thermodynamics properties of 5-LO, complexed with Compound D and DMPBD, from our simulation models reveal opportunities for using them as novel antiasthmatic drugs. However, further experiment on inhibition activity of these natural compounds with 5-LO could be a good validation of our calculations.

## REFERENCES

1. Asher, I. & Pearce, N., *J. Tuberc. Lung Dis.*, 2014, **18**, 1269-1278.
2. Pandya, D., Puttanna, A. & Balagopal, V., *Open Respir. Med. J.*, 2014, **8**, 59-65.
3. Luhadia, S. K., *J. Assoc. Physicians India*, 2014, **62**, 38-40.
4. O'Byrne, P. M., *Can. Respir. J.*, 1998, **5** (Suppl A), 64a-70a.
5. Werz, O. & Steinhilber, D., *Pharmacol. Ther.*, 2006, **112**, 701-718.
6. Panichyupakaranunt, P. & Latae, L., *J. Thai Tradit. Altern. Med.*, 2013, **11**, 123-137.
7. Jeenapongsa, R., Yoovathaworn, K., Sriwatanakul, K. M., Pongprayoon, U. & Sriwatanakul, K., *J. Ethnopharmacol.*, 2003, **87**, 143-148.
8. Rådmark, O. & Samuelsson, B. *Biophys. Res. Commun.*, 2010, **396**, 105-110.
9. Gilbert, N. C. et al., *Science*, 2011, **331**, 217-219.
10. Rådmark, O. & Samuelsson, B., *Prostaglandins Other Lipid Mediat.*, 2007, **83**, 162-174.
11. Rådmark, O., *Prostaglandins Other Lipid Mediat.*, 2002, **68-69**, 211-234.
12. Mullane, K., *Biochem. Pharmacol.*, 2011, **82**, 586-599.
13. O'Byrne, P. M., *Chest*, 1997, **111**, 27s-34s.
14. Hedi, H. & Norbert, G., *J. Biomed. Biotechnol.*, 2004, **2004**, 99-105.
15. Dahlen, S. E., *Eur. J. Pharmacol.*, 2006, **533**, 40-56.
16. Department of Medical Sciences, Ministry of Public Health, *Thai Herbal Pharmacopoeia*, Vol. 1, 2009.
17. Pongprayoon, U. et al., *Phytomedicine*, 1997, **3**, 323-326.

## ACKNOWLEDGMENTS

K. Jitapunkul was financially supported by the Excellent Thai Students (ETS) scholarship program of Sirindhorn International Institute of Technology (SIIT). The authors gratefully acknowledge the Center of Nanotechnology, Kasetsart University for the GAUSSIAN09 program package and the financial support provided by Thai Government Research Fund Contract No. 34/2560.

# A theoretical study of small schiff base complexes with heavy metal

**Kunanon Chattrairat<sup>1</sup>, Piyawan Leepheng<sup>2</sup>, Songwut Suramitr<sup>3</sup> and Darinee Phromyothin<sup>1,2,\*</sup>**

<sup>1</sup> College of Nanotechnology, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Rd., Ladkrabang, Bangkok, Thailand 10520

<sup>2</sup>Nanotec-KMITL Center of Excellence on Nanoelectronic Devices, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Rd., Ladkrabang, Bangkok, Thailand 10520

<sup>3</sup>Department of Chemical, Faculty of Science, Kasetsart University, Ladyao, Chatuchak, Bangkok, Thailand 10900

<sup>c</sup>**E-mail:** darinee.ph@kmitl.ac.th ; **Fax:** +66 2 329 8265 ext 3034 ; **Tel.** +66 2 329 8000 ext 3133

## ABSTRACT

Hugo Schiff base ligand is an imine compound that can applied for detection of metal ions via ligand to ligand charge transfer (LLCT) or ligand to metal charge transfer (LMCT) mechanisms. It is easy to functionalize and design to desired structure. This study, Schiff base was designed and investigated to detect the heavy metal ions ( $\text{Hg}^{2+}$ ,  $\text{Cu}^{2+}$  and  $\text{Pb}^{2+}$ ) by using density functional theory (DFT) calculations. The binding energies, optical properties and electron distribution of complexes were compared and analyzed. The mechanism of detection also was examined. The results indicate that small Schiff base interacted with  $\text{Hg}^{2+}$  presents lowest binding energy and change in optical properties.

**Keywords:** Schiff base, DFT, heavy metal.

## 1. INTRODUCTION

Treatment of heavy metal has been concerned as  $\text{Hg}^{2+}$ ,  $\text{Cu}^{2+}$  and  $\text{Pb}^{2+}$  owing to their high toxicity with both human and environment. Mercury ( $\text{Hg}^{2+}$ ) is one of the tremendous toxic at very low concentrations to human due to residue in the primary to top predator [1-3]. The toxicity of  $\text{Hg}^{2+}$  can highly affect to human in the immune system, genetic and nervous system [1, 4].  $\text{Hg}^{2+}$  absorption is extremely hazardous, especially fetuses, but the toxic cognition performance is slow leading to developed the  $\text{Hg}^{2+}$  detection performance [5]. Lead ( $\text{Pb}^{2+}$ ) has high toxicity. The excessive of it leads to headaches, irritability and memory problems [5]. Copper ( $\text{Cu}^{2+}$ ) is toxic as well. It is a neurotoxic that linked to physical and psychiatric symptoms[6]. The detection has desired to measure them. The one way that is low cost and less time consuming is computational method. For example, density functional theory (DFT) calculation was used to determine the derivative schiff base complex with heavy metal ions[7].

In this work, the molecular structure of schiff base was designed and studied the interaction with heavy metal ions ( $\text{Hg}^{2+}$ ,  $\text{Pb}^{2+}$  and  $\text{Cu}^{2+}$ ). Density functional theory calculation was used to calculate the structures and their complexes. Then electron distribution was analyzed by time dependent density functional theory (TD-DFT). Furthermore, binding energies, optical properties and mechanism of electron transition were also investigated.

## 2. THEORY AND RELATED WORKS

The molecular structure of schiff base was created by following the hard soft acid base (HSAB) theory. According to the HSAB theory, hard acids prefer to coordinate to hard bases and soft acids prefer to coordinate soft bases [8-10]. For example, the chemical hardness in complexes formation with  $\text{Ca}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Cu}^{2+}$  ions of pyridone ligand and its derivatives or the formation of the  $\text{MN}_2\text{O}_2$  or  $\text{NM}_2\text{S}_2$  coordination node composition (M=Be,

Hg) by using DFT with the prediction based on the HSAB principle was studied based on the HSAB theory[11]. Thus, the interaction side of metal ion to the schiff base molecule was proposed by this theory. The optimizations were obtained by B3LYP/6-311G(d,p) method for C, N, H, S and O and LANL2DZ basis set for heavy metal ions [12-15]. Then the LMCT or LLCT mechanism was investigated by TD-DFT calculation [16-19]. Moreover, TD-DFT was used to calculate the absorbance and electron distribution.

### 3. COMPUTATIONAL DETAILS

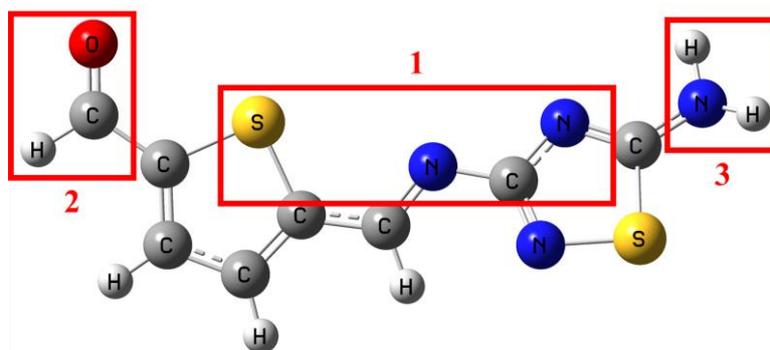
Quantum chemical calculation was brought to investigate structure geometry and electronic properties of schiff base structure and complexes. All schiff base molecule and complexes were carried out based on DFT calculation by using Gaussian09 [14]. The ground state geometries of schiff base molecule was fully optimized using hybrid functional B3LYP with 6-311G(d,p) basis set to determine the most stable orientation. Schiff base used in this study which contain 21 atoms (8 carbon atoms, 4 nitrogen atoms, 2 sulfide atoms, 1 oxygen atom and 6 hydrogen atoms). Then the metal ions ( $\text{Cu}^{2+}$ ,  $\text{Hg}^{2+}$  and  $\text{Pb}^{2+}$ ) were added to schiff base molecule following by hard soft acid base theory. The complex optimization was performed to achieve the stable structure using B3LYP/6-311G(d,p) method for C, N, H, S and O atom and LANL2DZ basis set for heavy metal ions. Then the binding energy was calculated following equation 1.

$$\text{Binding Energy (BE)} = E_{\text{complex}}^{\text{opt}} - E_{\text{pristine carbon}}^{\text{opt}} - E_{\text{metal ion}}^{\text{opt}} \quad (1)$$

Here,  $E_{\text{complex}}^{\text{opt}}$  represents the energy of the optimized structure of schiff complex with metal ion. The  $E_{\text{schiff base}}^{\text{opt}}$  and  $E_{\text{metal ion}}^{\text{opt}}$  represent the energy of the optimized structure of schiff base molecule and metal ion, respectively. The excited state of complexes, absorbance and electron distribution was determined by TD-DFT calculation.

### 4. RESULTS AND DISCUSSION

First, in this study is looking for the most stable structure for schiff base molecule interact with metal ions, corresponding to different metal ions and orientations. The schiff base molecule was used in this study as shown in Figure 1. The orientation was also varied and considered in three orientations as show in Figure 1. The first orientations have the sulfur and



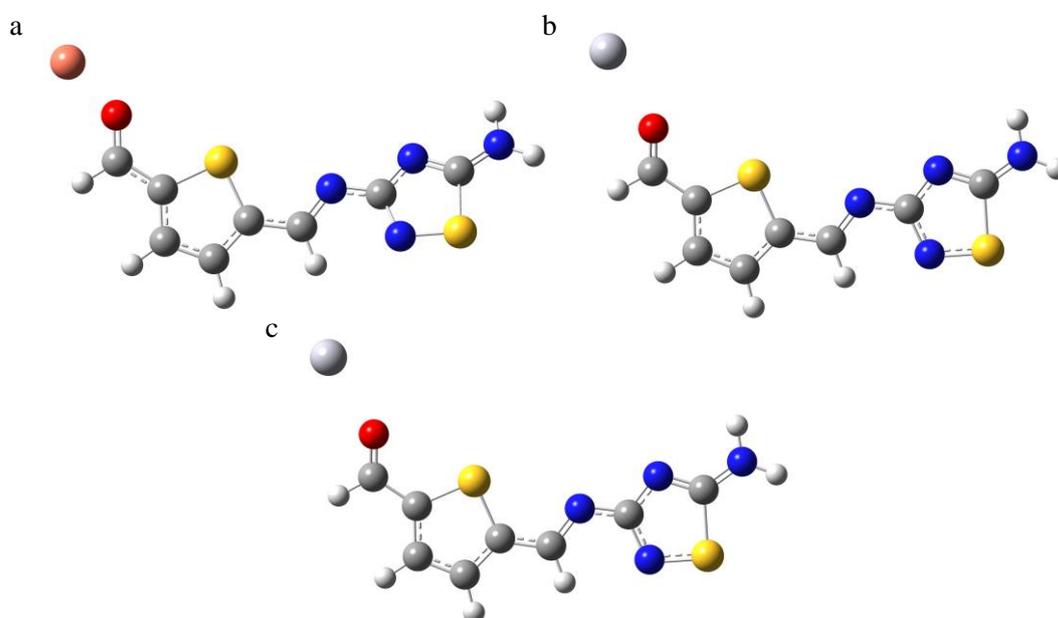
**Figure 1** The schiff base molecule and the side interact to metal ion

nitrogen atoms which interacted to the metal ion. The second orientation focuses at an aldehyde group end side and the last one is the amine group at another end side. Then, the optimization of molecular structure and binding energy were calculated as presented in Table 1. The complexes of  $\text{Cu}^{2+}$  in second orientation present the lowest binding energy of -910.76 kcal/mol among the complex of  $\text{Cu}^{2+}$  orientations. Therefore, the second orientation of complex of  $\text{Cu}^{2+}$  is the most stable as shown in Figure 2(a). The second orientation of complex of  $\text{Hg}^{2+}$  also has the lowest binding energy of -1016.91 kcal/mol as presented in Figure 2(b). While the first and third orientation has the same energy because after full relaxation, the molecular structures have the same orientation. However, the second orientation of complex of  $\text{Hg}^{2+}$  is the most stable molecular structure. Finally, after the

optimization of the complex of  $\text{Pb}^{2+}$ , the complex has equal binding energy in the first and second orientation as the same orientation as well. The first and second orientation show the most stable molecular structure that have the lowest binding energy at  $-355.76$  kcal/mol as shown in Figure 2(c). Then the bond length of oxygen atom-metal ions (O-M) was measured. Owing to the bond radius of metal ions, the radii are in the order:  $\text{Cu}^{2+} < \text{Hg}^{2+} < \text{Pb}^{2+}$  [20-23]. The bond length of O-M is in the order:  $\text{O-Cu}^{2+} < \text{O-Hg}^{2+} \approx \text{O-Pb}^{2+}$  due to the strength and favorable of interaction as shown in Table 2.

**Table 1.** The relative binding energy of metal complexes compared with the lowest energy of each complexes

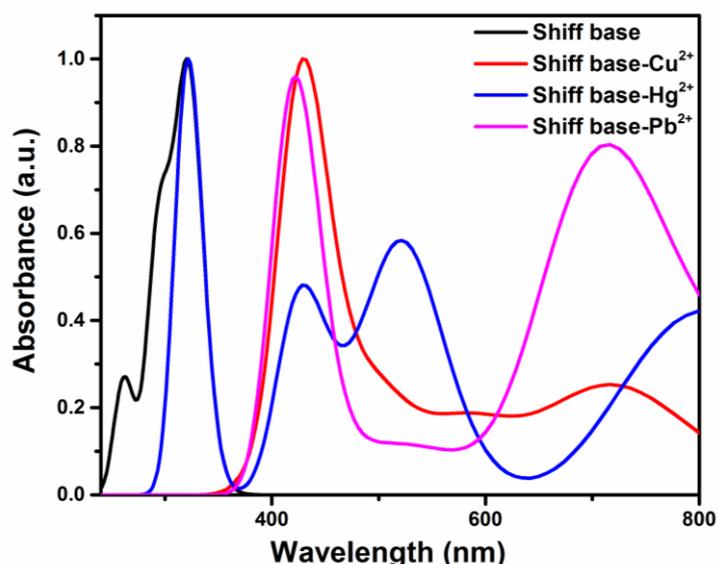
Orientation	Relative binding energy (kcal/mol)		
	1	2	3
$\text{Cu}^{(\text{II})}$ Complex	-9.89	0	-9.74
$\text{Hg}^{(\text{II})}$ Complex	-10.7	0	-10.7
$\text{Pb}^{(\text{II})}$ Complex	0	0	-26.56



**Figure 2** (a)  $\text{Cu}^{(\text{II})}$ Complex (b)  $\text{Hg}^{(\text{II})}$ Complex (c)  $\text{Pb}^{(\text{II})}$ Complex

**Table 2.** The bond length between oxygen atom in schiff base molecule and metal ions.

Complexes	Bond length (Å)
$\text{Cu}^{(\text{II})}$ Complex	1.93
$\text{Hg}^{(\text{II})}$ Complex	2.55
$\text{Pb}^{(\text{II})}$ Complex	2.13

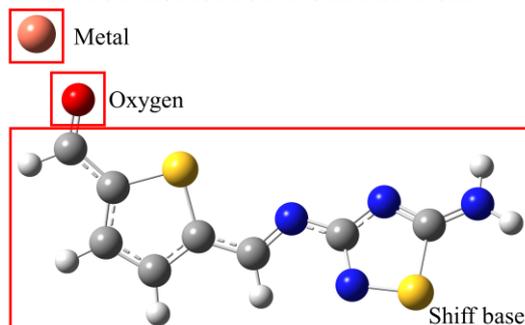


**Figure 3** The absorbance of the schiff base and complexes.

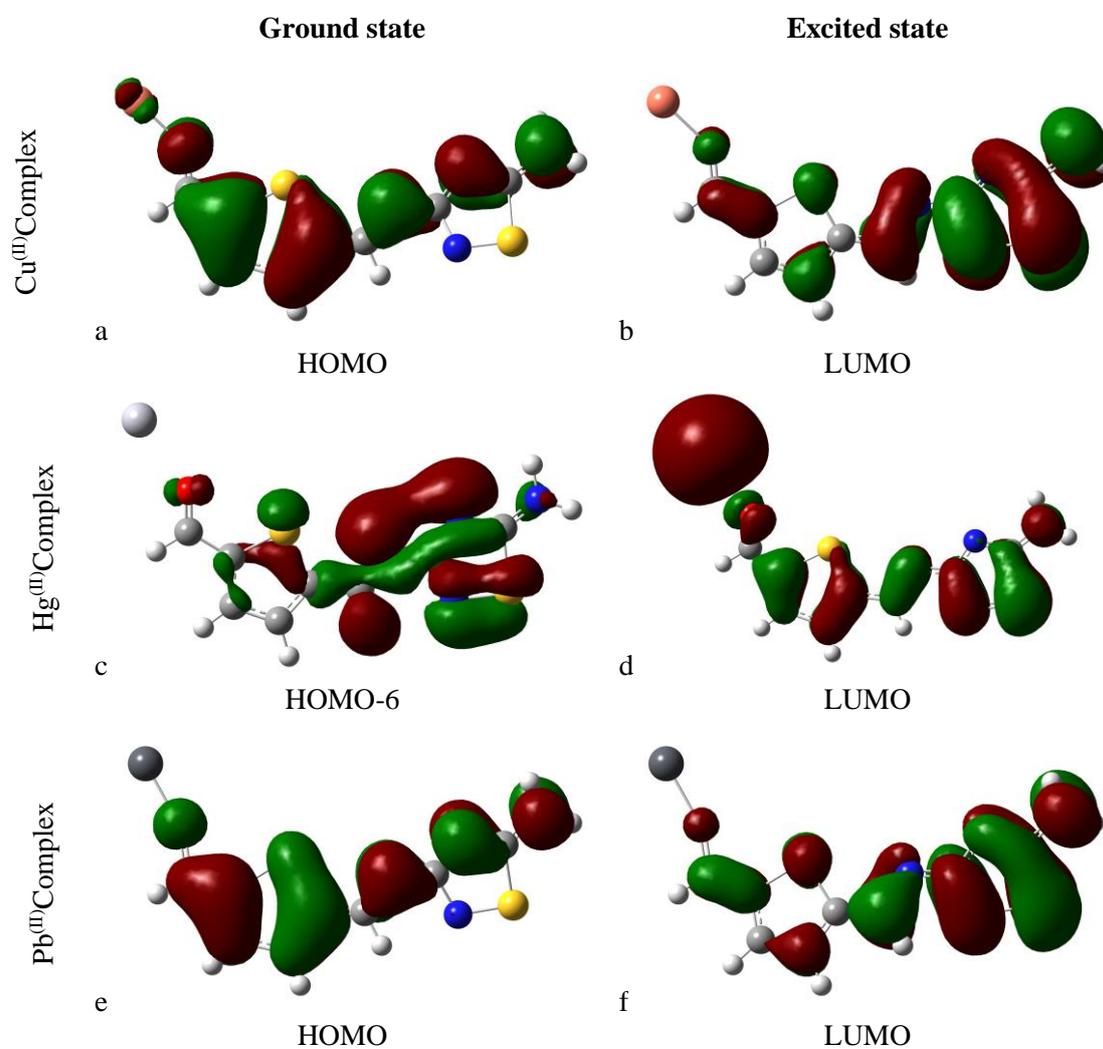
The absorbance of the schiff base and complexes was calculated by TD-DFT calculation as presented in Figure 3. The results indicate that the complexes of  $\text{Cu}^{2+}$  and complexes of  $\text{Pb}^{2+}$  present the red shift from 320 nm to 421 nm and 430 nm, respectively. Moreover, the new absorption peak at 723 nm and 713 nm was noticed, respectively. The schiff base- $\text{Hg}^{2+}$  also has the red shift with lower intensity and clearly new adsorption peak at 525 nm. The transition of electron distribution at the new absorption peak was calculated as shown in Table 3. The electron distribution of schiff base- $\text{Cu}^{2+}$  and schiff base- $\text{Pb}^{2+}$  has the transition from HOMO( $\beta$ ) to LUMO( $\beta$ ) (57%) and HOMO( $\beta$ ) to LUMO( $\beta$ ) (66%), respectively, corresponding to the percentage of electron distribution at selected atom was calculated as shown in Table 4. The results show the electron distribution of schiff base- $\text{Cu}^{2+}$  and schiff base- $\text{Pb}^{2+}$  at HOMO contain electron at the metal ion only 3 percent. After electrons were excited to excited state at LUMO, the electron distribution has transition in the schiff base molecule as shown in Figure 4. Thus, the schiff base- $\text{Cu}^{2+}$  and schiff base- $\text{Pb}^{2+}$  have the electron transition by LLCT mechanism. That indicate schiff base can interact to  $\text{Cu}^{2+}$  and  $\text{Pb}^{2+}$  with weak interaction but also has the new peak absorption detection. The transition of schiff base- $\text{Hg}^{2+}$  at 525 nm was calculated as well. The results demonstrate the transition from HOMO-6 to LUMO (65%) with the highest oscillator strength (Table 3). In the Table 4 presents that the percentage of electron distribution in schiff base- $\text{Hg}^{2+}$  at HOMO contains most of electron at the schiff base molecule. Then the excited state of schiff base- $\text{Hg}^{2+}$  has the transition of electron distribution from schiff base molecule to  $\text{Hg}^{2+}$  through the oxygen atom in schiff base molecule by LMCT mechanism. Which has the electron transition from HOMO-6 to LUMO as seen in Figure 4(c, d). Therefore, schiff base can interact to  $\text{Hg}^{2+}$  with the new absorption peak detection and electron transition by LMCT mechanism.

**Table 3.** Wavelength at absorption peak, oscillator strength and molecular orbital transition of schiff base and metal ion

	Wavelength (nm)	Oscillator Strength	Transition	Assignment
$\text{Cu}^{(II)}$ Complex	723	0.0724	HOMO( $\beta$ ) $\rightarrow$ LUMO( $\beta$ ) (57%)	LLCT
$\text{Hg}^{(II)}$ Complex	525	0.1903	HOMO-6 $\rightarrow$ LUMO (65%)	LMCT
$\text{Pb}^{(II)}$ Complex	713	0.1831	HOMO( $\beta$ ) $\rightarrow$ LUMO( $\beta$ ) (66%)	LLCT

**Table 4.** Percent of electron distribution at selected atom.

Complexes	State	Metal	Schiff base	Oxygen
Cu <sup>(II)</sup> Complex	LUMO	0	98	2
	HOMO	3	90	7
Hg <sup>(II)</sup> Complex	LUMO	38	58	4
	HOMO-6	0	96	3
Pb <sup>(II)</sup> Complex	LUMO	0	98	2
	HOMO	3	90	7

**Figure 4** Electron distribution of complexes by TD-DFT calculation

## 5. CONCLUSION

This research studied the interaction of schiff base molecule and metal ions ( $\text{Cu}^{2+}$ ,  $\text{Hg}^{2+}$  and  $\text{Pb}^{2+}$ ). Three possible orientation of complexes was investigated and presented that the lowest energy of complex was the interaction of aldehyde moiety at the end of the schiff base molecule and metal ions. The binding energy indicated that the  $\text{Hg}^{\text{II}}$  complexes showed the lowest binding energy compared other metals. It represents the stable and strong interaction between schiff base with  $\text{Hg}^{2+}$ . The TD-DFT was used to calculate the absorbance and electron distribution of these complexes. The results demonstrate that the schiff base molecule can interact to  $\text{Cu}^{2+}$  and  $\text{Pb}^{2+}$  with weak interaction by LLCT mechanism which the complexes have adsorption peak detection of metal ion with transition from HOMO( $\beta$ ) to LUMO( $\beta$ ) (57%) and HOMO( $\beta$ ) to LUMO( $\beta$ ) (66%), respectively. The  $\text{Hg}^{\text{II}}$  complex has the new peak adsorption at 525 nm with electron transition from HOMO-6 to LUMO (65%). Furthermore, the electron distribution of  $\text{Hg}^{\text{II}}$  complex presented electron transfer from the schiff base molecule to  $\text{Hg}^{2+}$  metal ion by LMCT mechanism.

## REFERENCES

1. Squadrone, S., et al., *Mercury and selenium in European catfish (Silurus glanis) from Northern Italian Rivers: can molar ratio be a predictive factor for mercury toxicity in a top predator?* Chemosphere, 2015. **119**: p. 24-30.
2. Suresh, M., et al., *Fluorescent and Magnetic Mesoporous Hybrid Material: A Chemical and Biological Nanosensor for Hg(2+) Ions*. Sci Rep, 2016. **6**: p. 21820.
3. Bernhoft, R.A., *Mercury toxicity and treatment: a review of the literature*. J Environ Public Health, 2012. **2012**: p. 460508.
4. Nordberg, G.F., B.A. Fowler, and M. Nordberg, *Handbook on the Toxicology of Metals*. 2014: Academic Press.
5. Avuthu, S.G.R., et al., *A Screen Printed Phenanthroline-Based Flexible Electrochemical Sensor for Selective Detection of Toxic Heavy Metal Ions*. IEEE Sensors Journal, 2016. **16**(24): p. 8678-8684.
6. Pal, A., *Copper toxicity induced hepatocerebral and neurodegenerative diseases: an urgent need for prognostic biomarkers*. Neurotoxicology, 2014. **40**: p. 97-101.
7. Jafari, M., et al., *DFT studies and antioxidant activity of Schiff base metal complexes of 2-aminopyridine. Crystal structures of cobalt(II) and zinc(II) complexes*. Inorganica Chimica Acta, 2017. **462**: p. 329-335.
8. Kaya, S., et al., *Effect of Some Electron Donor and Electron Acceptor Groups on Stability of Complexes According to the Principle of HSAB*. Journal of New Results in Science, 2014. **3**.
9. Pearson, R.G., *Hard and soft acids and bases, HSAB, part 1: Fundamental principles*. Journal of Chemical Education, 1968. **45**(9): p. 581.
10. Pearson, R.G., *Recent advances in the concept of hard and soft acids and bases*. Journal of Chemical Education, 1987. **64**(7): p. 561.
11. Kharabayev, N.N., A.G. Starikov, and V.I. Minkin, *The Pearson's HSAB principle in the quantum-chemical model of formation of the MN<sub>2</sub>O<sub>2</sub> or MN<sub>2</sub>S<sub>2</sub> coordination node in the bischelates of Be(II) and Hg(II) with ambidentate ligands based on azomethines and their cyclic analogs*. Russian Journal of General Chemistry, 2015. **85**(11): p. 2629-2633.
12. Li, P. and S. Chen, *A Review on Gaussian Process Latent Variable Models*. CAAI Transactions on Intelligence Technology.
13. Tshipis, A.C., *DFT flavor of coordination chemistry*. Coordination Chemistry Reviews, 2014. **272**: p. 1-29.
14. Frisch, M.J., et al., *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.
15. Thanakit, P., et al., *A Density Functional Theory Study on the Metal Binding Properties and Electronic Transitions of Dithienopyrrole Derivatives*. Integrated Ferroelectrics, 2014. **155**(1): p. 119-125.

16. Zhong, K., et al., *A water-soluble highly sensitive and selective fluorescent sensor for Hg<sup>2+</sup> based on 2-(2-(8-hydroxyquinolin-yl)benzimidazole via ligand-to-metal charge transfer (LMCT)*. RSC Advances, 2014. **4**(32): p. 16612.
17. Citek, C., et al., *Chemical Plausibility of Cu(III) with Biological Ligation in pMMO*. J Am Chem Soc, 2015. **137**(22): p. 6991-4.
18. Thanakit, P., et al., *A Study of Fluorescent Chemosensor for Fe(II) Based On Cyanoacrylic Derivatives*. Advanced Materials Research, 2013. **802**: p. 94-98.
19. Tongkate, P., D. Phromyothin, and J. Sumranjit, *Copper(II) ion-sensing mechanism of oligo-phenylene vinylene derivatives: syntheses and theoretical calculations*. Tetrahedron, 2012. **68**(16): p. 3329-3335.
20. Breviglieri, S.T., É.T.G. Cavalheiro, and G.O. Chierice, *Correlation between ionic radius and thermal decomposition of Fe(II), Co(II), Ni(II), Cu(II) and Zn(II) diethanoldithiocarbamates*. Thermochimica Acta, 2000. **356**(1): p. 79-84.
21. Sabounchei, S.J., et al., *X-ray crystal structure and spectral characterization of pseudo five-coordinate Hg(II) polymeric and four-coordinate binuclear complexes of an ambidentate sulfonium ylide*. Comptes Rendus Chimie, 2014. **17**(12): p. 1257-1263.
22. Canty, A.J. and G.B. Deacon, *The van der Waals radius of mercury*. Inorganica Chimica Acta, 1980. **45**: p. L225-L227.
23. Coughlin, B.R. and A.T. Stone, *Nonreversible Adsorption of Divalent Metal Ions (MnII, CoII, NiII, CuII, and PbII) onto Goethite: Effects of Acidification, FeII Addition, and Picolinic Acid Addition*. Environmental Science & Technology, 1995. **29**(9): p. 2445-2455.

## ACKNOWLEDGMENTS

The authors would like to thank the computational resources at the Laboratory for Computational & Applied Chemistry (LCAC), Kasetsart University, Thailand and the large scale simulation research laboratory of NECTEC, Thailand.

# Influence of Metal Species in Porphyrin-based Metal-organic Frameworks on Carbon Dioxide Capture

Panchanit Piyakeeratikul<sup>1</sup>, Bundet Boekfa<sup>2</sup>, Sippakorn Wannakao<sup>1</sup>, Taweesak Pila<sup>3</sup>  
and Kanokwan Kongpatpanich<sup>1,c</sup>

<sup>1</sup>*Department of Materials Science and Engineering, School of Molecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand*

<sup>2</sup>*Computational Chemistry Laboratory, Department of Chemistry, Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand*

<sup>3</sup>*Department of Chemical and Biomolecular Engineering, School of Energy Science and Engineering, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand*

<sup>c</sup> *E-mail:* kanokwan.k@vistec.ac.th; *Tel.* +66 3 301 4154

## ABSTRACT

Tuning of pore surface in metal-organic frameworks (MOFs) by functionalization with open metal sites is one of the key strategies to improve CO<sub>2</sub> adsorption selectivity over N<sub>2</sub> at ambient conditions. Herein, we prepared the porphyrin-based MOFs with and without the metal center in the porphyrin rings to study the influence from the additional metal species in MOF structures. Furthermore, density functional theory (DFT) calculation has been used to investigate the interaction between CO<sub>2</sub> and metal centers to explain the roles of metal species on CO<sub>2</sub> adsorption.

**Keywords:** Metal-organic framework, Porphyrin, CO<sub>2</sub> capture, DFT calculation.

## INTRODUCTION

Accumulation of carbon dioxide (CO<sub>2</sub>) from human activities has posed challenges to researchers for many decades. Several porous materials have been developed for CO<sub>2</sub> capture at cryogenic temperature or at high pressure. However, it remains challenging to develop materials for CO<sub>2</sub> capture at ambient conditions (room temperature and atmospheric pressure) with high CO<sub>2</sub>/N<sub>2</sub> selectivity. Metal-organic frameworks (MOFs), which are crystalline porous materials constructed from metal ions and ligands, are particularly attractive for CO<sub>2</sub> separation and storage due to their large surface area, versatile framework topologies, and tunable chemical functionality.

Tuning of pore surface in MOF by functionalization with polar organic groups or open metal sites is the key strategies to improve CO<sub>2</sub> adsorption selectivity over N<sub>2</sub>. Open metal sites in MOF are coordinatively unsaturated sites generated upon removal of weakly-coordinated solvent molecules from the metal centers. A recent report shows different metals influence on the CO<sub>2</sub> adsorption process. [1] However, these sites are generally sensitive to air and moisture because the open metal sites are typically in low valence states and easy to interact with moisture resulting in several handling problems in practical applications. In recent years, incorporating Zr<sub>6</sub>O<sub>8</sub> nodes in MOFs is a route to improve the stability of MOF materials by strongly coordinated Zr-O bonds and a high coordination number of Zr. However, use of Zr ions as metal nodes makes the MOF structure lack of open metal sites and limits the choices of metal species in MOFs. Thanks to the diversification of organic ligands, porphyrin is interesting because its structure has a free center site for metal ions to coordinate to form open metal sites in ligand. Therefore, incorporation of coordinatively unsaturated metal centers into porphyrin-based ligands could be an effective strategy to increase active units for CO<sub>2</sub> binding in MOFs.

ANSCSE21 Thailand Science Park, Pathum Thani, Thailand  
August 3-4, 2017

Herein, we study the effect of metal centers of a porphyrin ligand using adsorption experiments and density functional theory (DFT) calculations. Porphyrin-based MOFs were synthesized the pristine porphyrin MOF (PCN-222) and the isostructural MOFs synthesized from iron-porphyrin (Fe-PCN-222). Then, we evaluated the effect of metals on CO<sub>2</sub> capture by adsorption measurement under ambient conditions. Moreover, DFT calculation has been used to investigate the interaction between CO<sub>2</sub> and metal centers of porphyrins, which is useful to explain the influence of metal species on CO<sub>2</sub> adsorption. The study provides a guideline to design the suitable metal centers or metal complexes in the center of porphyrin ligands to enhance the adsorption and selectivity of CO<sub>2</sub>.

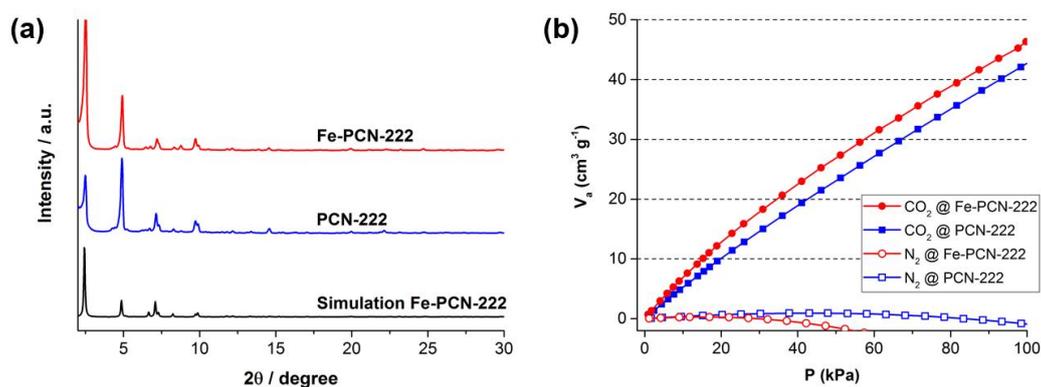
## EXPERIMENTAL AND COMPUTATIONAL DETAILS

PCN-222 and Fe-PCN-222 were synthesized by solvothermal reactions of ZrCl<sub>4</sub> (75 mg), either tetrakis(4-carboxyphenyl)porphyrin (H<sub>2</sub>-TCPP) or Fe-TCPP (50 mg) and benzoic acid (2700 mg) in *N,N'*-diethylformamide (DEF) as modified from literature. [2] The structures and CO<sub>2</sub> uptakes of PCN-222 and Fe-PCN-222 were characterized by power X-ray diffraction (XRD) and CO<sub>2</sub> adsorption experiments measured at 283 K. The heats of CO<sub>2</sub> adsorption were calculated from the CO<sub>2</sub> adsorption isotherms.

For DFT calculations, C<sub>44</sub>H<sub>30</sub>N<sub>4</sub> and FeC<sub>44</sub>H<sub>28</sub>N<sub>4</sub> clusters were used to represent PCN-222 and Fe-PCN-222, respectively. The cluster model of PCN-222 topology was obtained from the crystallographic data of Fe-PCN-222. [2] For PCN-222, the structure was created by substituting Fe for two H atoms. To obtain the most stable structure, the Fe-PCN-222 were optimized with several spin states. Various adsorption modes of CO<sub>2</sub> on the clusters were studied. All structures were optimized with M06-L functional, which has been successfully used for studying the adsorption and reaction mechanism over several metal catalysts. [3] The 6-31G(d,p) basis set was applied for C, O, N, Cl and H atoms while Stuttgart basis functional was used for Fe atom. Terminated hydrogen atoms (4H) were fixed with the crystallography coordinate while the rest atoms were allowed to relax during a geometry optimization. Partial charges and population analysis were determined by natural atomic orbital (NAO) method. All calculations were performed using the Gaussian 09 program.

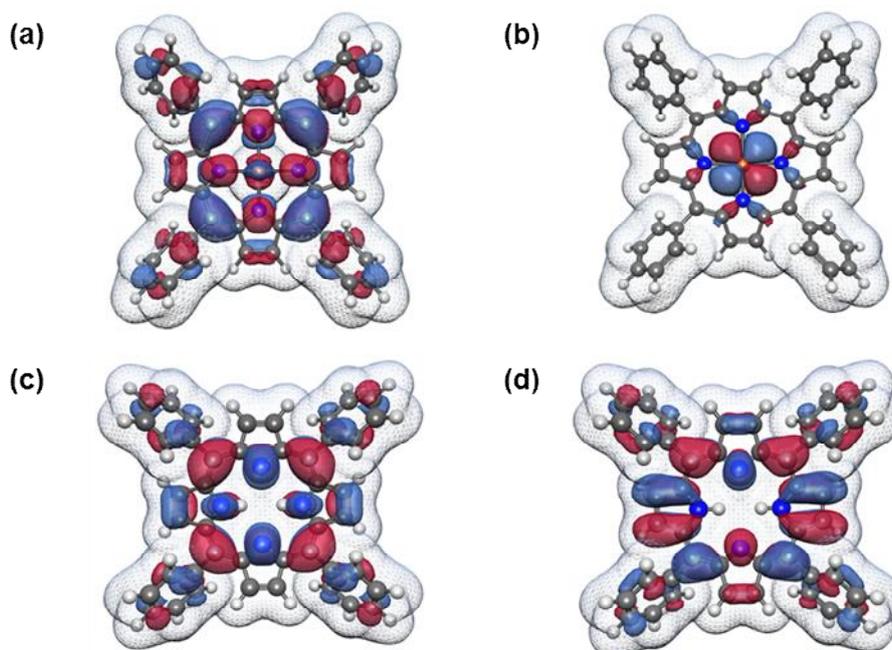
## RESULTS AND DISCUSSION

PCN-222 is Zr-based MOFs synthesized from the porphyrin ligands (H<sub>2</sub>-TCPP), and has been demonstrated on the high thermal and chemical stability of the framework topology. [2] Isostructures with other metalloporphyrin ligands such as Fe, Mn, Co, Ni and Cu have been reported. The framework is constructed from the Zr<sub>6</sub> clusters linked by carboxylate groups of either the uncoordinated or the metallated TCPP ligand to create the 3D frameworks of M-PCN-222. We synthesized M-PCN-222 by solvothermal reactions of ZrCl<sub>4</sub> and the corresponding TCPP ligands in DEF, and obtained M-PCN-222 as crystalline solids. The XRD patterns of all M-PCN-222 samples matched well with the simulated pattern of the reported Fe-PCN-222 indicating the same topology (**Fig 1a**). We evaluated CO<sub>2</sub> and N<sub>2</sub> adsorption capacities up to an atmospheric pressure at 283 K (**Fig 1b**). Both pristine PCN-222 and Fe-PCN-222 adsorbed only CO<sub>2</sub> indicating the CO<sub>2</sub> selectivity over N<sub>2</sub>. The CO<sub>2</sub> adsorption isotherms of the pristine and Fe-PCN-222 show the linear adsorption isotherms indicating the chemisorption of CO<sub>2</sub> to the porous structures at ambient conditions. Fe-PCN-222 provides the higher CO<sub>2</sub> uptake compared to the pristine PCN-222.



**Fig 1.** (a) XRD patterns and (b) CO<sub>2</sub> and N<sub>2</sub> adsorption isotherms of the pristine PCN-222 and Fe-PCN-222 measured at 283K

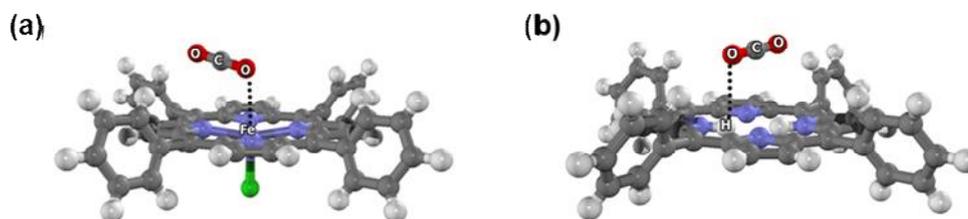
To clarify the effect of metal centers in porphyrin ligand on the CO<sub>2</sub> binding affinity, DFT calculation with M06-L functional, which was previously demonstrated to be cost-effective and accurate for systems involving transition metals, was conducted to obtain a heat of adsorption of CO<sub>2</sub> on Fe-PCN-222 and the pristine PCN-222. [3] We first considered the structure of Fe-PCN-222 optimized with M06-L functional. The Fe-PCN-222 with the Fe<sup>3+</sup> atom located in the middle of four pyrrole subunits, has been chosen to represent the additional Lewis acid sites from the Fe centers. For the Fe-PCN-222, energies relative to an energy of the most stable spin state were 20.6, 5.7, 0.0 and 37.2 kcal mol<sup>-1</sup> for the doublet, quintet, sextet and octet spin states, respectively. The most stable states were selected for further structural analysis. The average Fe-N, Fe-Cl and C-N bond distances were 2.09, 2.23 and 1.37 Å compared well with the experimental data of 2.03, 2.36 and 1.40 Å. The singlet state of pristine PCN-222 was optimized with same functional. The H-N and N-C bond distances were 1.01 and 1.37 Å, respectively.



**Fig 2.** The molecular orbitals with electron density (a) HOMO of the Fe-PCN-222, (b) LUMO of the Fe-PCN-222, (c) HOMO of the pristine PCN-222 and (d) LUMO of the pristine PCN-222.

The Mulliken charges suggest the electron transfers from N atoms of pyrrole to the metal to form bonds. In the Fe-PCN-222, charges of Fe and Cl atoms were  $0.943|e|$  and  $-0.243|e|$ , respectively. The charges of Fe and Cl atoms were  $0.941|e|$  and  $-0.308|e|$  with the Natural atomic charges. The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) with electron density for Fe-PCN-222 were showed in **Fig 2 (a,b)**. The HOMO of the Fe-PCN-222 locates at the pyrrole groups and Cl atom which are surrounding the metal atom. The LUMO, on the other hand, mostly exhibits at the Fe site of the porphyrin. For the pristine PCN-222, the HOMO and LUMO were showed in **Fig 2 (c,d)**. Both HOMO and LUMO locate around the nitrogen of pyrrole ring. The energy gaps were 1.45 and 1.82 eV for Fe-PCN-222 and the pristine PCN-222, respectively.

The geometries of  $\text{CO}_2$  adsorbed on Fe-PCN-222 ( $\text{CO}_2 @ \text{Fe-PCN-222}$ ) and the pristine PCN-222 ( $\text{CO}_2 @ \text{PCN-222}$ ) were shown in **Fig 3**. One can be considered that both structures demonstrate the contribution on the adsorption from Lewis and Brønsted acidity of similar materials. The  $\text{CO}_2$  adsorbs on the Fe atom in the end-on mode with the intermolecular distance between Fe and O of 3.06 Å. For the Brønsted acid site of pristine PCN-222, the intermolecular distance between O atom of  $\text{CO}_2$  and proton (H atom) became shorter (2.84 Å). Despite a longer intermolecular distance, the Fe site can induce the larger difference in the charge density on  $\text{CO}_2$  molecule during the adsorption. Fe sites in Fe-PCN-222 induced the change in the charge density of  $\text{CO}_2$  molecule to be  $\sim 20$  times greater than the pristine PCN-222. The  $\text{CO}_2$  became positively charged by  $+0.02 |e|$  on the Fe-PCN-222, while the charge of  $\text{CO}_2$  was less than  $+0.001 |e|$  on the pristine PCN-222. Charges of the Fe and Cl atoms were  $+0.942|e|$  and  $-0.238|e|$  indicating a direction of the electron density transfer. The adsorption energies were calculated to be  $-6.8$  and  $-5.9 \text{ kcal mol}^{-1}$  for Fe-PCN-222 and the pristine PCN-222, respectively. Hence, the  $\text{CO}_2$  adsorption energy on Fe-PCN-222 increased by 15% when compared to the pristine PCN-222. The trend of  $\text{CO}_2$  adsorption energies calculated from DFT agrees well with the adsorption experiments, which provide the adsorption energy increasing from  $-5.3 \text{ kcal mol}^{-1}$  for the pristine PCN-222 to  $-6.1 \text{ kcal mol}^{-1}$  for Fe-PCN-222 (15% increasing of adsorption energy). It can be concluded that the metal site influences in the improvement of  $\text{CO}_2$  adsorption by inducing electron density transfer of the  $\text{CO}_2$  molecule.



**Fig 3.** Optimized structures for (a)  $\text{CO}_2 @ \text{Fe-PCN-222}$  and (b)  $\text{CO}_2 @ \text{PCN-222}$

## CONCLUSION

In summary, we combined adsorption experiments and density functional investigations to study the effect from metal site in porphyrin ligand of MOFs on  $\text{CO}_2$  capture. The Fe site was used to represent the Lewis acid site in porphyrin ligand. The incorporating of Fe sites in Fe-PCN-222 increases the  $\text{CO}_2$  uptake and improves the  $\text{CO}_2$  adsorption energy by 15% compared to the pristine PCN-222. Adsorption energies obtain from DFT calculations are consistent with the experimental finding. Structural and electronic structure analyses suggest the roles of the metal in  $\text{CO}_2$  charge inducing, which leads to the improvement of  $\text{CO}_2$  capture capability by the Lewis acid site of Fe-PCN-222. Using the suitable metal centers or metal complexes in the center of porphyrin ligands is promising to enhance the adsorption and selectivity of  $\text{CO}_2$  and other polar gases.

**REFERENCES**

1. Queen, W. L.; Hudson, M. R.; Bloch, E. D.; Mason, J. A.; Gonzalez, M. I.; Lee, J. S.; Gygi, D.; Howe, J. D.; Lee, K.; Darwish, T. A.; James, M.; Peterson, V. K.; Teat, S. J.; Smit, B.; Neaton, J. B.; Long, J. R. and Brown, C. M., *Chem. Sci.* 2014, **5** (12), 4569-4581.
2. Feng, D.; Gu, Z.-Y.; Li, J.-R.; Jiang, H.-L.; Wei, Z. and Zhou, H.-C., *Angew. Chem. Int. Ed.* 2012, **51** (41), 10307-10310.
3. Wannakao, S.; Boekfa, B.; Khongpracha, P.; Probst, M. and Limtrakul, J., *ChemPhysChem* 2010, **11** (16), 3432-3438.

**ACKNOWLEDGMENTS**

This work was supported by Thailand Research Fund to K.K. (MRG 6080278) and postdoctoral research fellowship from Vidyasirimedhi Institute of Science and Technology (to S.W.). Support on research equipment by the Frontier Research Center, Vidyasirimedhi Institute of Science and Technology, is gratefully acknowledged.

# The Numerical Solution of Fractional Angiogenesis Problem by Meshless Local Petrov-Galerkin Method

K. Phramrung<sup>1,C</sup>, A. Luadsong<sup>1,2</sup>, and N. Ascharyaphotha<sup>2</sup>

<sup>1</sup> *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Bang Mod, Thung Khru, Bangkok 10140, Thailand*

<sup>2</sup> *Ratchaburi Learning Park, King Mongkut's University of Technology Thonburi (KMUTT), Rang Bua, Chom Bueng, Ratchaburi 70150, Thailand*

<sup>C</sup> **E-mail:** kunwithree.kp@mail.kmutt.ac.th; Fax: +66 2 428 4025; Tel. +66 9 9097 5232

## ABSTRACT

The angiogenesis model was extended to the concept of the new fractional differential equation proposed by Caputo and Fabrizio. The approximations of the angiogenesis model are used by the meshless local Petrov-Galerkin and the implicit finite difference method for the spatial and temporal discretization, respectively. In the meshless local Petrov-Galerkin method, the moving Kriging interpolation is employed to construct the shape function which processes the Kronecker delta property. And the Dirac delta function is applied in the local weak forms as the test functions. The results of the approximation are compared with the integer order to confirm the accuracy and efficiency of the proposed scheme.

**Keywords:** Angiogenesis, Fractional Differential Equation, Caputo and Fabrizio Fractional, Meshless Local Petrov-Galerkin Method, Keller-Segel Model.

## 1. INTRODUCTION

In recent years, chemotaxis received much attention due to its important role in the phenomenon of biological diversity. Chemotaxis is the movement of bacteria in response to gradients of chemical signals in their environment. In 1953, Patlak proposed the first mathematical model of chemotaxis. In 1970, Keller and Segel model proposed aggregation process of the cellular slime mold [11]. An example for Keller-Segel model, angiogenesis, is used to describe the evolution of creatures responding to chemicals in their environment [19]. Angiogenesis is the process of developing the physical properties of new blood vessel growth.

The fractional calculus is a field of study mathematics that grows out of the traditional definitions of calculus integral and derivative operators. For example, these equations are increasingly used to model problems in fluid flow, diffusion, oscillation, turbulence, electric networks, polymer physics, chemical physics, electrochemistry of corrosion, and many other physical processes. However, it is usually difficult or even impossible to obtain the exact solution. As a consequence, it is necessary to search for new numerical methods for the fractional differential equations. In 1695, the first idea of fractional calculus starts with Leibniz's letter to L'Hospital [8, 9, 12 and 15]. Caputo and Fabrizio have proposed a new fractional derivative in 2015 also designed with the concept of convolution. The convolute filter is an exponential function, which helps to reduce the risk of singularity [3, 17 and 21].

The meshless method is a widely used numerical method. It has advantages over conventional numerical methods. It is to reduce the problem to create the mesh. The meshless methods can save computational time because no mesh is needed [10]. In 1998, Atluri and Zhu proposed a different approach in constructing a meshfree method, the meshless local Petrov-Galerkin (MLPG), it is based on the local weak form which eliminates the need of the background element and performs the numerical integration. And later in 2002, Atluri expanded the meshless local Petrov-Galerkin formulation to six meshless approximation

techniques [1, 6, 14, 16 and 20]. Table 1 shows the characteristics that apply to the MLPG1-MLPG6 method.

**Table 1.** Meshless local Petrov-Galerkin methods as characteristics by Atluri and Shen [5].

Type of the method	Test function utilized	Type of the integral in the weak form
MLPG1	MLS weight function	Domain integral
MLPG2	Dirac delta function	None
MLPG3	Discrete least-squares	Domain integral
MLPG4	Fundamental solution	Singular boundary integral
MLPG5	Heaviside unit step function	Regular boundary integral
MLPG6	Same as the trial function	Domain integral

The purpose of this research is to develop the meshless local Petrov-Galerkin methods for solving the fractional angiogenesis model using Caputo and Fabrizio definition.

## 2. THEORY AND RELATED WORKS

In 2001, Levine et al. developed methods of predicting the effects of various interventions by the mathematical modeling in the beginning of capillary formation initiating angiogenesis. They proposed a model of the first two stages of angiogenesis that captures the bimodal migration of endothelial cells. They define their model on the interactions between one cellular species, endothelial cells, and the receptors on the endothelial cells. The model can be written as partial differential equations as follows:

$$\frac{\partial \eta(\mathbf{x}, t)}{\partial t} = D \frac{\partial}{\partial \mathbf{x}} \left( \eta \frac{\partial}{\partial \mathbf{x}} \ln \left( \frac{\eta}{\tau} \right) \right) \quad (2.1)$$

$$\frac{\partial c(\mathbf{x}, t)}{\partial t} = \frac{\lambda_1 \eta v}{1 + v_1 v} \quad (2.2)$$

$$\frac{\partial v(\mathbf{x}, t)}{\partial t} = \frac{-\lambda_1 \eta v}{1 + v_1 v} \quad (2.3)$$

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = \frac{-\lambda_2 c f}{1 + v_2 v} + \beta f (f_M - f) \eta \quad (2.4)$$

where  $v$  is the concentration of tumor angiogenesis factor,  $c$  is the concentration of proteolytic enzymes,  $r$  is the density of receptors to tumor angiogenesis factors on the endothelial cells,  $l$  is the concentration of an intermediate receptor complex,  $\eta$  is the density of endothelial cells,  $f$  is the density of fibronectin,  $D$  is the diffusion limit of the density of endothelial cells,  $\tau$  is the transition probability rate function depends on the concentration of the proteolytic enzyme and the fibronectin at that the lattice point. And  $\mathbf{x} \in \Omega$ ,  $t > 0$ ,  $\Omega \in R^d$ ;  $d = 1, 2, 3, \dots$

The values  $k$  with a subscript are the rates of chemical reaction,  $v_1 = \frac{k_1}{k_{-1} + k_2}$ ,  $\hat{k} = k_2 v_1 = k_1 - k_{-1} v_1$ ,  $\delta \approx \frac{r(\mathbf{x}, t)}{\eta(\mathbf{x}, t)} \approx \frac{r(\mathbf{x}, 0)}{\eta(\mathbf{x}, t)}$ ,  $\lambda_1 = \hat{k} \delta$ ,  $\lambda_2 = k_3$ ,  $v_2 = \frac{k_4}{k_3}$ . And the given initial distributions are  $\eta(\mathbf{x}, 0) = \eta_0(\mathbf{x}) \geq 0$ ,  $v(\mathbf{x}, 0) = v_0(\mathbf{x}) \geq 0$ ,  $c(\mathbf{x}, 0) = c_0(\mathbf{x}) \geq 0$  and  $f(\mathbf{x}, 0) = f_0(\mathbf{x}) \geq 0$ .

In 2015, Caputo and Fabrizio proposed a fractional order derivative without singularity kernel [7].

**Definition:** Let  $f \in H^1(a, b)$ ;  $b > a$ ,  $\alpha \in [0, 1]$  then the Caputo and Fabrizio fractional of order  $\alpha$  is given by

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{M(\alpha)}{1-\alpha} \int_{\alpha}^t f'(s) \exp\left[-\alpha \frac{t-s}{1-\alpha}\right] ds \quad (2.5)$$

where  $M(\alpha)$  is a normalization function such that  $M(0) = M(1) = 1$ . However, if the function does not belong to  $H^1(a, b)$  then, the derivative can be redefined as

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{\alpha M(\alpha)}{1-\alpha} \int_{\alpha}^t (f(t) - f(s)) \exp\left[-\alpha \frac{t-s}{1-\alpha}\right] ds \quad (2.6)$$

If  $\sigma = \frac{1-\alpha}{\alpha} \in [0, \infty]$ ,  $\alpha = \frac{1}{1+\sigma} \in [0, 1]$  then the above equation assumes the form

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{N(\sigma)}{\sigma} \int_{\alpha}^t f'(s) \exp\left[-\frac{t-s}{\sigma}\right] ds \quad (2.7)$$

where  $N(0) = N(\infty) = 1$ .

Thus, the equations (2.1) to (2.4) were transformed into a fractional differential equation by the definition of Caputo and Fabrizio as follows:

$$\frac{\partial^\alpha \eta(\mathbf{x}, t)}{\partial t^\alpha} = D \frac{\partial}{\partial \mathbf{x}} \left( \eta \frac{\partial}{\partial \mathbf{x}} \ln \left( \frac{\eta}{\tau} \right) \right) \quad (2.8)$$

$$\frac{\partial^\alpha c(\mathbf{x}, t)}{\partial t^\alpha} = \frac{\lambda_1 \eta v}{1 + v_1 v} \quad (2.9)$$

$$\frac{\partial^\alpha v(\mathbf{x}, t)}{\partial t^\alpha} = \frac{-\lambda_1 \eta v}{1 + v_1 v} \quad (2.10)$$

$$\frac{\partial^\alpha f(\mathbf{x}, t)}{\partial t^\alpha} = \frac{-\lambda_2 c f}{1 + v_2 v} + \beta f (f_M - f) \eta \quad (2.11)$$

### 3. SPATIAL AND TEMPORAL DISCRETIZATION

In this topic, present the approximation solution of the angiogenesis model for the spatial discretization used by the meshless local Petrov-Galerkin method [5, 18]. In the meshless local Petrov-Galerkin method, the moving Kriging interpolation is employed to construct the shape function which has the Kronecker delta property. And the Dirac delta function is applied in the local weak form as the test functions. For the temporal discretization used implicit finite difference method [2, 22].

The meshless local Petrov-Galerkin method creates the local weak form over local sub-domain,  $\Omega_s$ , which is a small region taken for each node in the global domain,  $\Omega$ . To multiplying test function  $w_i$  into equations (2.8) to (2.11) and then integrate over a local sub-domains associated with the point  $\mathbf{x}_i$ ;  $i = 1, 2, 3, \dots, N$  where  $N$  is the number of node in the sub-domain.

$$\int_{\Omega_s^i} \left( \frac{\partial^\alpha \eta(\mathbf{x}, t)}{\partial t^\alpha} - D \frac{\partial}{\partial \mathbf{x}} \left( \eta \frac{\partial}{\partial \mathbf{x}} \ln \left( \frac{\eta}{\tau} \right) \right) \right) w_i(\mathbf{x}) d\Omega = 0 \quad (3.1)$$

$$\int_{\Omega_S^i} \left( \frac{\partial^\alpha c(\mathbf{x}, t)}{\partial t^\alpha} - \frac{\lambda_1 \eta v}{1 + v_1 v} \right) w_i(\mathbf{x}) d\Omega = 0 \quad (3.2)$$

$$\int_{\Omega_S^i} \left( \frac{\partial^\alpha v(\mathbf{x}, t)}{\partial t^\alpha} + \frac{\lambda_1 \eta v}{1 + v_1 v} \right) w_i(\mathbf{x}) d\Omega = 0 \quad (3.3)$$

$$\int_{\Omega_S^i} \left( \frac{\partial^\alpha f(\mathbf{x}, t)}{\partial t^\alpha} + \frac{\lambda_2 c f}{1 + v_2 v} - \beta f (f_M - f) \eta \right) w_i(\mathbf{x}) d\Omega = 0 \quad (3.4)$$

Then substituting trial functions  $\eta^h(\mathbf{x}, t)$ ,  $c^h(\mathbf{x}, t)$ ,  $v^h(\mathbf{x}, t)$  and  $f^h(\mathbf{x}, t)$  for  $\eta(\mathbf{x}, t)$ ,  $c(\mathbf{x}, t)$ ,  $v(\mathbf{x}, t)$  and  $f(\mathbf{x}, t)$ , respectively, into the local weak forms in equations (3.5) to (3.8) as follows:

$$\begin{aligned} \sum_{j=1}^N \left( \int_{\Omega_S^i} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \frac{d^\alpha \hat{\eta}_j(t)}{dt^\alpha} - \sum_{j=1}^N \left( \int_{\Omega_S^i} D \frac{\partial^2}{\partial \mathbf{x}^2} \ln \left( \frac{\eta}{\tau} \right) \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{\eta}_j(t) \\ - \sum_{j=1}^N \left( \int_{\Omega_S^i} D \frac{\partial}{\partial \mathbf{x}} \ln \left( \frac{\eta}{\tau} \right) \phi_{j,\mathbf{x}}(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{\eta}_j(t) = 0 \end{aligned} \quad (3.5)$$

$$\sum_{j=1}^N \left( \int_{\Omega_S^i} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \frac{d^\alpha \hat{c}_j(t)}{dt^\alpha} - \sum_{j=1}^N \left( \int_{\Omega_S^i} \frac{\lambda_1 \eta}{1 + v_1 v} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{v}_j(t) = 0 \quad (3.6)$$

$$\begin{aligned} \sum_{j=1}^N \left( \int_{\Omega_S^i} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \frac{d^\alpha \hat{v}_j(t)}{dt^\alpha} + \sum_{j=1}^N \left( \int_{\Omega_S^i} \frac{\lambda_1 \eta}{1 + v_1 v} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{v}_j(t) \\ = 0 \end{aligned} \quad (3.7)$$

$$\begin{aligned} \sum_{j=1}^N \left( \int_{\Omega_S^i} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \frac{d^\alpha \hat{f}_j(t)}{dt^\alpha} + \sum_{j=1}^N \left( \int_{\Omega_S^i} \frac{\lambda_2 c}{1 + v_2 v} \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{f}_j(t) \\ - \sum_{j=1}^N \left( \int_{\Omega_S^i} \beta (f_M - f) \eta \phi_j(\mathbf{x}) w_i(\mathbf{x}) d\Omega \right) \hat{f}_j(t) = 0 \end{aligned} \quad (3.8)$$

where

$$\eta^h(\mathbf{x}, t) = \sum_{j=1}^N \phi_j(\mathbf{x}) \hat{\eta}_j(t), \quad c^h(\mathbf{x}, t) = \sum_{j=1}^N \phi_j(\mathbf{x}) \hat{c}_j(t),$$

$$v^h(\mathbf{x}, t) = \sum_{j=1}^N \phi_j(\mathbf{x}) \hat{v}_j(t), \quad f^h(\mathbf{x}, t) = \sum_{j=1}^N \phi_j(\mathbf{x}) \hat{f}_j(t).$$

The shape function,  $\phi_j$ , is constructed by moving Kriging interpolation which has the Kronecker delta property. This research use MLPG2 then the test function is chosen by Dirac delta function. The test function will define significance for each node in a sub-domain is a unit value and gives zero value at all other nodes then integrate over the sub-domain,  $\Omega_S^i$ , as follows:

$$\begin{aligned} \sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{d^\alpha \hat{\eta}_j(t)}{dt^\alpha} - \sum_{j=1}^N \phi_j(\mathbf{x}_i) D \frac{\partial^2}{\partial \mathbf{x}^2} \ln\left(\frac{\eta}{\tau}\right) \Big|_{\mathbf{x}_i} \hat{\eta}_j(t) \\ - \sum_{j=1}^N \phi_{j,x}(\mathbf{x}_i) D \frac{\partial}{\partial \mathbf{x}} \ln\left(\frac{\eta}{\tau}\right) \Big|_{\mathbf{x}_i} \hat{\eta}_j(t) = 0 \end{aligned} \quad (3.9)$$

$$\sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{d^\alpha \hat{c}_j(t)}{dt^\alpha} - \sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{\lambda_1 \eta(\mathbf{x}_i, t)}{1 + v_1 v(\mathbf{x}_i, t)} \hat{v}_j(t) = 0 \quad (3.10)$$

$$\sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{d^\alpha \hat{v}_j(t)}{dt^\alpha} + \sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{\lambda_1 \eta(\mathbf{x}_i, t)}{1 + v_1 v(\mathbf{x}_i, t)} \hat{v}_j(t) = 0 \quad (3.11)$$

$$\begin{aligned} \sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{d^\alpha \hat{f}_j(t)}{dt^\alpha} + \sum_{j=1}^N \phi_j(\mathbf{x}_i) \frac{\lambda_2 c(\mathbf{x}_i, t)}{1 + v_2 v(\mathbf{x}_i, t)} \hat{f}_j(t) \\ - \sum_{j=1}^N \phi_j(\mathbf{x}_i) \beta (f_M - f) \eta(\mathbf{x}_i, t) \hat{f}_j(t) = 0 \end{aligned} \quad (3.12)$$

The equations (3.9) to (3.12) can be written in a matrix form as follows:

$$A \frac{d^\alpha U}{dt^\alpha} + B(U)U = 0 \quad (3.13)$$

where

$$A = [A_{ij}]_{N \times N}; \quad A_{ij} = \phi_j(\mathbf{x}_i) = \begin{cases} 0; & i \neq j \\ 1; & i = j \end{cases}, \quad B = \begin{bmatrix} B_{11} & 0 & 0 & 0 \\ 0 & 0 & B_{23} & 0 \\ 0 & 0 & B_{33} & 0 \\ 0 & 0 & 0 & B_{44} \end{bmatrix},$$

$$B_{11} = \left[ -\phi_j(\mathbf{x}_i) D \frac{\partial^2}{\partial \mathbf{x}^2} \ln\left(\frac{\eta}{\tau}\right) \Big|_{\mathbf{x}_i} - \phi_{j,x}(\mathbf{x}_i) D \frac{\partial}{\partial \mathbf{x}} \ln\left(\frac{\eta}{\tau}\right) \Big|_{\mathbf{x}_i} \right]_{N \times N},$$

$$B_{23} = \left[ -\phi_j(\mathbf{x}_i) \frac{\lambda_1 \eta(\mathbf{x}_i, t)}{1 + v_1 v(\mathbf{x}_i, t)} \right]_{N \times N}, \quad B_{33} = \left[ \phi_j(\mathbf{x}_i) \frac{\lambda_1 \eta(\mathbf{x}_i, t)}{1 + v_1 v(\mathbf{x}_i, t)} \right]_{N \times N},$$

$$B_{44} = \left[ \phi_j(\mathbf{x}_i) \frac{\lambda_2 c(\mathbf{x}_i, t)}{1 + v_2 v(\mathbf{x}_i, t)} - \phi_j(\mathbf{x}_i) \beta (f_M - f) \eta(\mathbf{x}_i, t) \right]_{N \times N},$$

$$U = [\hat{H} \quad \hat{C} \quad \hat{V} \quad \hat{F}]^T, \quad \hat{H} = [\hat{\eta}_1 \quad \hat{\eta}_2 \quad \dots \quad \hat{\eta}_N]^T, \quad \hat{C} = [\hat{c}_1 \quad \hat{c}_2 \quad \dots \quad \hat{c}_N]^T,$$

$$\hat{V} = [\hat{v}_1 \quad \hat{v}_2 \quad \dots \quad \hat{v}_N]^T, \quad \hat{F} = [\hat{f}_1 \quad \hat{f}_2 \quad \dots \quad \hat{f}_N]^T.$$

Since the shape function that is constructed by the moving Kriging interpolation satisfies the Kronecker delta property,  $A$  is the identity matrix.

Then used implicit finite difference method based on the definition of Caputo and Fabrizio fractional for temporal discretization. Approximation of the time fractional derivative,  $\left. \frac{d^\alpha U}{dt^\alpha} \right|^{t_n}$ , that approximates the exact solution at time level  $n$ . The grid side in time for forward difference scheme is defined as  $\Delta t = \frac{T}{n}$  for some positive integer. The grid points in the time interval  $[0, T]$  are labeled  $t_i = i\Delta t + t_0$  where  $i = 1, 2, 3, \dots, n$  and can be modified using the forward difference method and then integrate over  $\tau$  for approximation [4] as follows:

$$\left. \frac{d^\alpha U}{dt^\alpha} \right|^{t_n} = \frac{M(\alpha)}{\alpha} \sum_{j=0}^{n-1} \left( \frac{U^{j+1} - U^j}{\Delta t} \right) d_j + \frac{M(\alpha)}{\alpha} \sum_{j=0}^{n-1} d_j O(\Delta t) \quad (3.14)$$

where  $d_j = \left( \exp \left[ -\alpha \frac{t_n - t_{j+1}}{1-\alpha} \right] - \exp \left[ -\alpha \frac{t_n - t_j}{1-\alpha} \right] \right)$ .

Let  $\Delta t = t_{j+1} - t_j$ ;  $j = 1, 2, 3, \dots, n-1$  and consider the summation of  $d_j$  for the accuracy of the equation (3.14). Then  $\sum_{j=0}^{n-1} d_j \approx \alpha \frac{n\Delta t}{1-\alpha}$ ;  $\exp \left[ -\alpha \frac{n\Delta t}{1-\alpha} \right] \approx 1 - \alpha \frac{n\Delta t}{1-\alpha}$ . The discrete approximation to the Caputo and Fabrizio fractional became

$$\left. \frac{d^\alpha U}{dt^\alpha} \right|^{t_n} = \frac{M(\alpha)}{\alpha} \sum_{j=0}^{n-1} \left( \frac{U^{j+1} - U^j}{\Delta t} \right) d_j + \frac{M(\alpha)n}{1-\alpha} O(\Delta t)^2 \quad (3.15)$$

Let  $\left. \frac{d^\alpha U}{dt^\alpha} \right|^{t_n} = D_t^\alpha U^n + O(\Delta t)^2$ . Concluded that the computation of the Caputo and Fabrizio fractional is given by

$$D_t^\alpha U^n = \frac{M(\alpha)}{\alpha} \sum_{j=0}^{n-1} \left( \frac{U^{j+1} - U^j}{\Delta t} \right) d_j \quad (3.16)$$

The equation (3.16) is substituted time fractional derivative as follows:

$$\frac{M(\alpha)}{\alpha} \sum_{j=0}^{n-1} \left( \frac{U^{j+1} - U^j}{\Delta t} \right) d_j + B(U^n)U^n = 0 \quad (3.17)$$

Let  $B^n = B(U^n)$ . At the time level  $n$ , it can be seen that

$$\left[ \sigma I + \frac{\alpha \Delta t}{M(\alpha)} B^n \right] U^n = d_0 U^0 + \sum_{j=1}^{n-1} d_j U^j - \sum_{j=0}^{n-2} d_j U^{j+1} \quad (3.18)$$

$$U^n = G^n \left( d_0 U^0 + \sum_{j=1}^{n-1} (d_j - d_{j-1}) U^j \right) \quad (3.19)$$

where

$$\sigma = 1 - \exp \left[ -\alpha \frac{\Delta t}{1-\alpha} \right], \quad G^n = \left( \sigma I + \frac{\alpha \Delta t}{M(\alpha)} B^n \right)^{-1}.$$

The matrix  $G^n$  must be non-singular and invertible. Now to get solution values at the time level  $n$ , on the right-hand side of equation (3.18). This research requires the previous solutions at all time levels  $0, 1, 2, \dots, n - 1$ .

#### 4. NUMERICAL EXAMPLE

The numerical results were compared with integer of order differential equations to confirm the accuracy and effectiveness of the proposed method. In conclude the results of a numerical experiment conducted by the figures 1 to 4 show the approximation of the system obtained numerically. The format of the transitional probability function used is

$$\tau = \tau_1(c)\tau_2(f) = \left(\frac{\alpha_1 + c}{\alpha_2 + c}\right)^{\gamma_1} \left(\frac{\beta_1 + f}{\beta_2 + c}\right)^{\gamma_2},$$

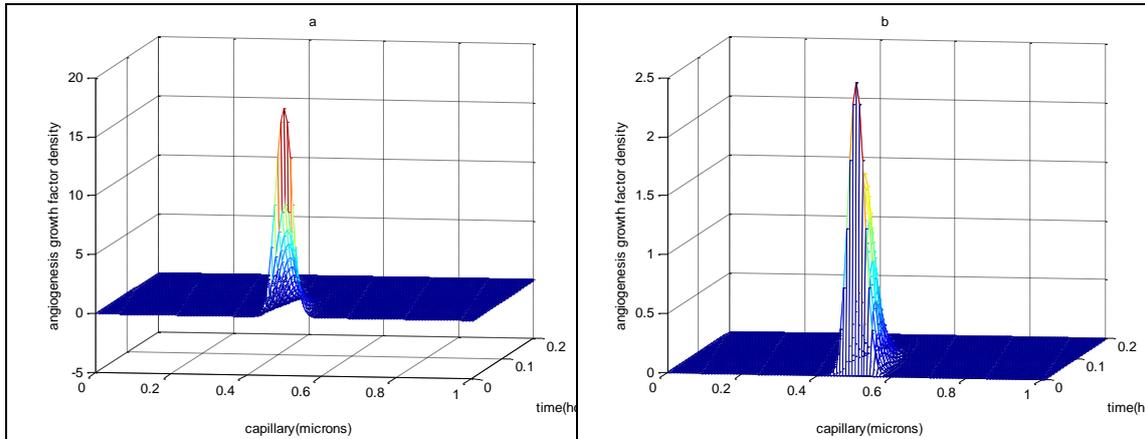
which is intended to describe both chemotaxis in response to protease  $c$  and haptotaxis in response to fibronectin  $f$ . So the initial conditions on the tumor angiogenesis factor of the numerical experiment [13] as follow:

$$\eta(\mathbf{x}, 0) = 1, \quad c(\mathbf{x}, 0) = 1, \quad f(\mathbf{x}, 0) = 1, \quad \kappa_{100} = 0.933 \times 10^{-30},$$

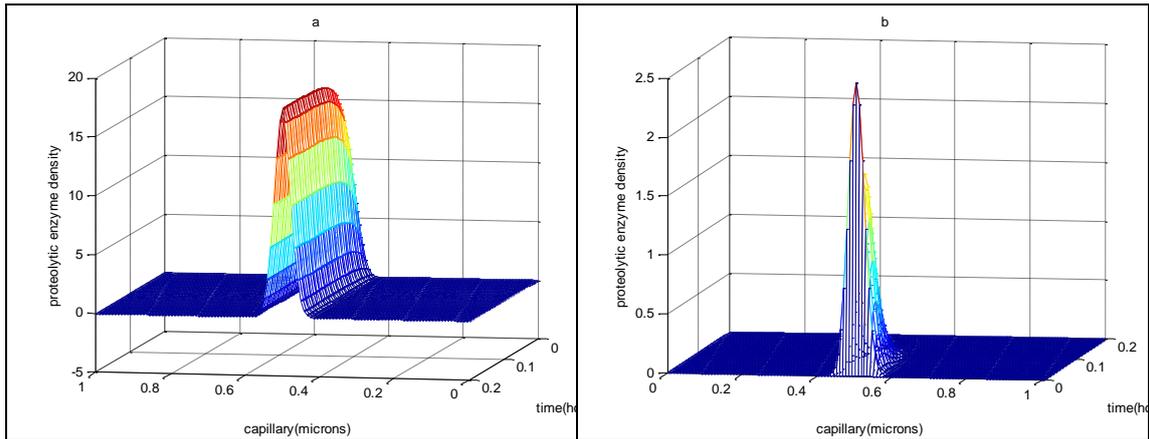
$$v(\mathbf{x}, 0) = 15\kappa_{100}(1 - \cos(2\pi x))^{100}, \quad D = 3.5e^{-5}, \quad \alpha_1 = 0.001, \quad \alpha_2 = 1.0,$$

$$\beta = 0.222, \quad \beta_1 = 1.0, \quad \beta_2 = 0.001, \quad \gamma_1 = 1.2, \quad \gamma_2 = 1.2, \quad \lambda_1 = 73.0,$$

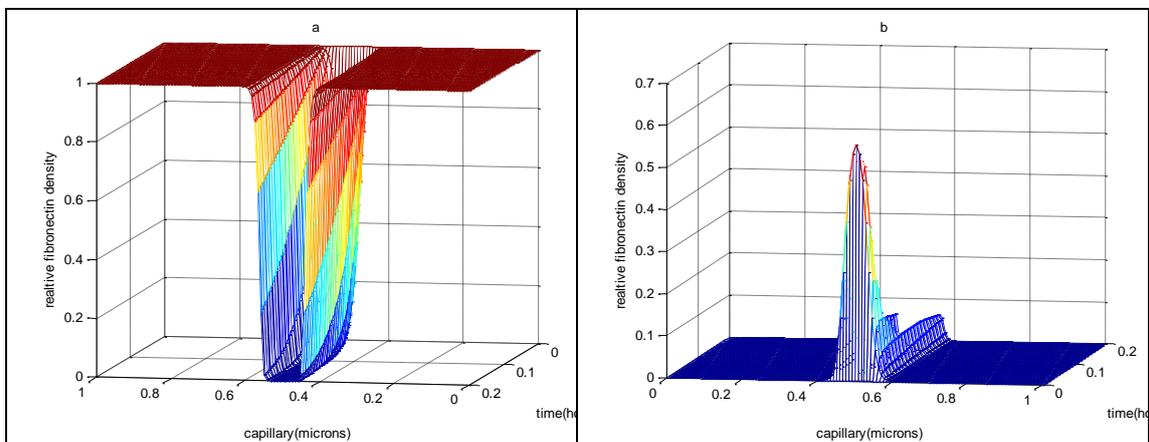
$$\lambda_2 = 19.0, \quad v_1 = 0.007, \quad v_2 = 1.28.$$



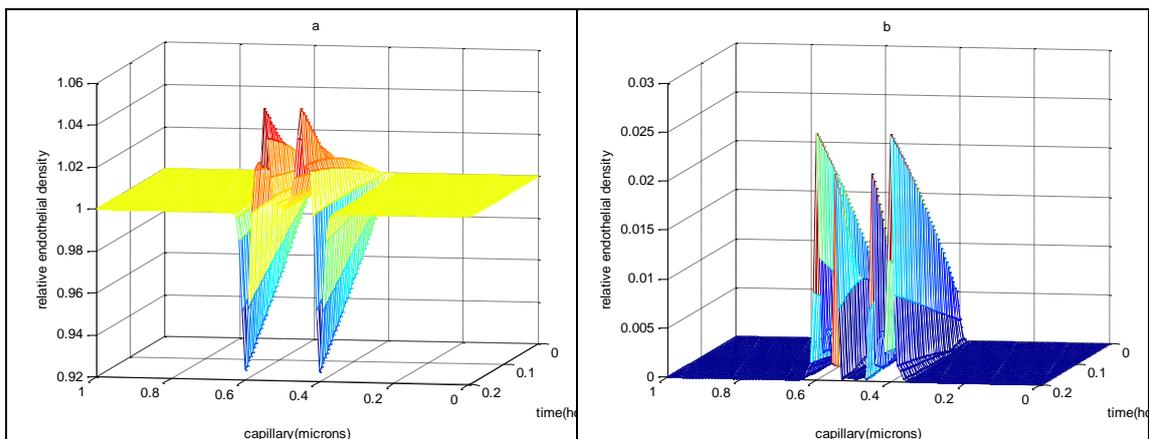
**Figure 1.** (a) The approximation solution of the fractional order differential equation of the evolution of the growth factor decay for  $\alpha = 0.99$  and (b) The absolute error of the approximation solution of integer order and fractional order differential equation.



**Figure 2.** (a) The approximation solution of the fractional order differential equation of the evolution of the protease for  $\alpha = 0.99$  and (b) The absolute error of the approximation solution of integer order and fractional order differential equation.



**Figure 3.** (a) The approximation solution of the fractional order differential equation of the evolution of fibronectin decay for  $\alpha = 0.99$  and (b) The absolute error of the approximation solution of integer order and fractional order differential equation.



**Figure 4.** (a) The approximation solution of the fractional order differential equation of the evolution of the endothelial cells for  $\alpha = 0.99$  and (b) The absolute error of the approximation solution of integer order and fractional order differential equation.

Figure 1(a) shows the phenomenon of rapid uptake of the tumor angiogenic factors by the endothelial cells. Figure 2(a) shows the density of the protease attains a near constant in the short period. Figure 3(a) shows the decay of fibronectin which the tumor angiogenic factors are originally concentrated. Figure 4(a) shows the corresponding bimodal concentration of endothelial cells at the ends of the interval which is indicative of the growth of a new capillary tube structure. And figures 1(b) to 4(b) show the comparison of the absolute error of the approximation solution for the integer order and the fractional order differential equation.

## 5. CONCLUSION

The angiogenesis model was extended to the concept of the derivative fractional using the derivative with the fractional order without the kernel singular proposed by Caputo and Fabrizio. In this research solve a system of fractional nonlinear differential equation by the meshless local Petrov-Galerkin and implicit finite difference method are used for discretizing the spatial and time variable, respectively. In the meshless local Petrov-Galerkin method, the moving Kriging interpolation is employed to construct a shape function which processes the Kronecker delta property, and the Dirac delta function is applied in the local weak form as the test functions. The results of the approximate solution are showed about evolution of the growth factor decay, the protease, fibronectin decay and the endothelial cells. Compared to the approximation solution for integer order differential equation, it can be found that the results obtained by the proposed scheme to have the accurate and the effective similar.

## REFERENCES

1. Arefmanesh, A., et al., IUST International Journal of Engineering Science, 2007, **18**(34), 39-45.
2. Atangana, A. and Alkahtani B. S. T., *Entropy*, 2015, **17**, 4439-4453.
3. Atangana, A. and Alkahtani, B. S. T., *Advances in Mechanical Engineering*, 2015, **7**(6), 1-6.
4. Atangana, A. and Alqahtani R. T., *Advances in Difference Equations*, 2016, **1**(156).
5. Atluri, S. N. and Shen, S., *Computer Modeling in Engineering & Sciences*, 2002, **3**(1), 11-51.
6. Augarde, C. and Heaney, C., Proceedings of the 1st International Symposium on Computational Geomechanics (COMGEO I), 2009.
7. Caputo, M. and Fabrizio, M., *Progress in Fractional Differentiation and Applications*, 2015, **1**(2), 73-85.
8. Dalir, M., *Applied Mathematical Sciences*, 2010, **4**(21), 1021-1032.
9. Ding, H., *Journal of Computational and Applied Mathematics*, 2016, **299**, 221-228.
10. Ghandehari, M. A. M. and Ranjbar, M., *Journal of Nonlinear Science*, 2014, **17**(2), 105-110.
11. Keller, E. F. and Segel, L. A., *Journal of Theory Biology*, 1970, **26**, 399-415.
12. Kumar, D., Singh, J., and Kumer, S., *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 2014, **1**, 177-183.
13. Levine, H. A., Sleeman, B. D., and Nilsen-Hamilton, M., *Journal of Mathematical Biology*, 2001, **42**, 195-238.
14. Li, S. and Liu, W. K., *Applied Mechanics Reviews*, 2002, **55**(1), 1-34.
15. Loverro, A., *Department of Aerospace and Mechanical*, Engineering University of Notre Dame Notre Dame, U.S.A., 2004, 1-28.
16. Nguyen, V. P., et al., *Mathematics and Computers in Simulation*, 2008, **79**(3), 763-813.
17. Oliveira, E. C. D. and Machado, J. A. T., *Mathematical Problems in Engineering*, 2014, **1**.
18. Phaochoo, P., Luadsong, A., and Ascharyaphotha, N., *Springer Plus*, 2016, **5**(305).
19. Ritter, L. R., *A Short Course in the Modeling of Chemotaxis*, 1st ed., Texas A & M University, Texas, 2004, 36-52.
20. Sladek, J. et al., *Computer Modeling in Engineering & Sciences*, 2013, **92**(5), 423-475.
21. Thamareerat, N., Luadsong, A., and Ascharyaphotha, N., *Advances in Difference Equations*, 2017, **1**(74).

22. Yimnak, K., and Luadsong, A., Hindawi Publishing Corporation Advances in Mathematical Physics, 2014.

#### **ACKNOWLEDGMENTS**

This research was supported by King Mongkut's University of Technology Thonburi (KMUTT). The researcher would like to thank the advisor for providing this advice and taking care of this research and Kanchanaburi Rajabhat University for providing a scholarship.

# Automatic Control for Water Droplet on Metal Surface

Prasarnphun Saisin<sup>1</sup>, Santi Wannippanto<sup>2</sup>, Yongyut Pattanapong<sup>3</sup>,

<sup>1,2,3</sup>*Graduate of Electrical Engineering, Faculty of Science and Technology*

*Pathumwan Institute of Technology, Bangkok 10330, Thailand*

**E-mail** : saisin\_p@hotmail.com, <sup>e2</sup>santi@pit.ac.th, <sup>e3</sup>yongyut.pattanapong@gmail.com ;

**Fax** : +66 4 241 2458; **Tel.** +66840286923

## ABSTRACT

The purpose of this research is to design an automatic control unit to control the droplet on metal surface. Using the principle of temperature and relative humidity in the air to calculate the dew point and target temperature to control the temperature at the surface of the metal. Experiment in a room with a temperature of 28 °C with an aluminum plate. Surface dimensions 7.62 cm in width, 11.43 cm in length and 0.5 cm in thickness. Experiment with 2 aluminum plate at temperatures above and below the dew point temperature of 1 °C. The results of the experiment when the dew point temperature was 19 °C. It took 60 seconds to reduce the temperature of the aluminum surface from 28 °C to 20 °C. Which is higher than the dew point temperature, no droplets occur. It takes 80 seconds to reduce the temperature from 28 °C to 18 °C, which is lower than the dew point temperature, causing the droplets to drip and the droplets drift over the plate for 5 minutes. which shows that the automatic control unit to control the droplet on the metal surface design can control the droplet on the metal surface.

**Keywords** : an automatic control, water droplet, dew point.

## 1. INTRODUCTION

Factories in Thailand mainly import machinery used in manufacturing to the use of force and to produce large quantities [1]. Machinery used mainly made of metal, which is the component of steel and shaped characteristics vary according to use. And require regular maintenance. The problems encountered in the maintenance of the rust caused a lot of water vapor in the air condenses on the surface of the metal and the condensation reaction between iron and oxygen in the air. Rust stains are red or reddish brown on a surface of the metal. And the Condensation on the surface of the metal, machinery and equipment industry is another reason that causes rust. As a result, cause a short circuit of the electronics and cause damage [2]. The dew point refers to the point at which condensation of water vapor occurs due to the amount of water vapor in the air is fully saturated and the air temperature drop causes the condensation water naturally. If we can control the condensation on the surface, especially in the machining of large industrial plants to prevent damage different. The aim of researchers are interested to learn how to control condensation on metal surfaces and cause condensation to occur by increasing the temperature to a metal surface [3]. This method seems to control condensation and may be another way to reduce the damage of a mechanical device [4].

## 2. THEORY AND RELATED WORKS

The relative humidity ( $RH$ ) and the dew point temperature ( $T_d$ ) are two widely used indicators of the amount of moisture in air. The exact conversion from  $RH$  to  $T_d$ , as well as highly accurate approximations, are too complex to be done easily without the help of a calculator or computer. However, there is a very simple rule of thumb that found to be quite useful for approximating the conversion for moist air.

Relative humidity is commonly defined in one of two ways, either as the ratio of the actual water vapor pressure  $p_w$  to the equilibrium vapor pressure over a plane of water  $p_{ws}$  or often called the “saturation” vapor pressure [5].

#### Change Equation

Select to solve a different unknown relative humidity

- Solve for relative humidity

$$RH = 100 \times \frac{p_w}{p_{ws}} \quad \dots\dots\dots(1)$$

- Solve for actual vapor pressure.

$$p_w = \frac{RH \times p_{ws}}{100} \quad \dots\dots\dots(2)$$

- Solve for saturated vapor pressure.

$$p_{ws} = 100 \times \frac{p_w}{RH} \quad \dots\dots\dots(3)$$

#### Dew Point Equations Formulas Calculator Meteorology Weather Water Vapor.[6-10]

- Solving for dew point temperature

$$T_d = \left( \frac{RH}{100} \right)^{\frac{1}{8}} (112 + 0.9T) + 0.1T - 112 \quad \dots\dots\dots(4)$$

#### Relative humidity

- relative humidity

$$RH = 100 \left( \frac{112 - 0.1T + T_d}{112 + 0.9T} \right)^8 \quad \dots\dots\dots(5)$$

- temperature

$$T = \left( \frac{T_d - 112 \left( \frac{RH}{100} \right)^{\frac{1}{8}} + 112}{0.9 \left( \frac{RH}{100} \right)^{\frac{1}{8}} + 0.1} \right) \quad \dots\dots\dots(6)$$

- And Set point temperature

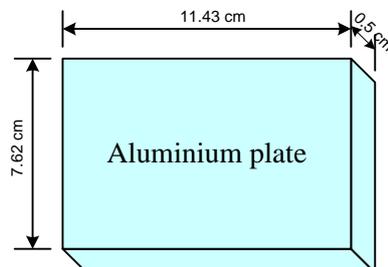
$$T_{sp} = (T_d \pm 1^\circ\text{C}) \quad \dots\dots\dots(7)$$

### 3. EXPERIMENTAL

#### 3.1 Preparation Materials

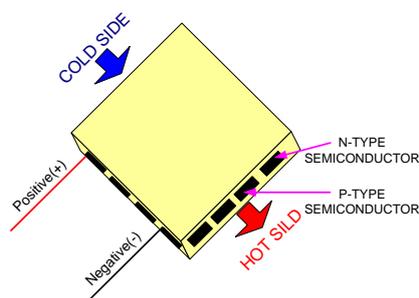
The materials for this study consist of as follows:

3.1.1 Aluminium plate with size width 7.62 cm, length 11.43 cm, thick 0.5 cm and the constructed surface of the Aluminium as the lightweight substance that effectively conducts heat to the surface heat are shown in Figure 1.



**Figure 1.** Aluminium plate.

3.1.2 Thermoelectric device has the size 92 watts 12 volts. The special properties of this material can transform the heat energy into electricity and then electricity transformed to the cool by the vibration principle of the internal structure of the solid material. Thermoelectric in terms of quantum physics called phonon, Electron and hole. When the thermoelectric material has high different temperatures, it is transferred to a lower temperature. That is movement of electrons and hole, it will generate the electric which leads to the electric power generator and refrigerator as shown in Fig.2



**Figure 2.** Component of the thermoplastic dielectric.

3.1.3 Fin heat sink and the cooling fan with size 12 volts are the equipment of conductor and convector the heat, respectively, in order to lead the high heat source to the low heat source. The fin heat sink is made from copper, Aluminium, and mixture of copper and Aluminium. It is designed with fin depending on the usage and the efficient characteristics that specify the direction of cooling and the exposure area of cooling plate. Most structure of fin heat sink usually composes of fan as the blower to cool from the cooling plate.

3.1.4 Solid State Relays (SSR) are normally-open semiconductor equivalents of the electromechanical relay that can be used to control electrical loads without the use of moving parts. This SSR can switch current loads of up to 40 A with 3-32V DC input that has both single phase and three phase, which solid state relay is compatible with temperature control to control the heater as shown in Figure 3.



**Figure 3.** The solid state relays (SSR).

3.1.5 Heater or resistance wire is a wire made from the alloy made the thin wire and it has a high resistance. For example, Nichrome wire, an alloy of Nickel and Chromium, is a resistance wire. The electric current flows through the wire. This produces the more heat. Finally, the power supply of 12 volts supplies the current with maximum, 30 amps. This power supply is employed for feeding the thermoelectric plate.

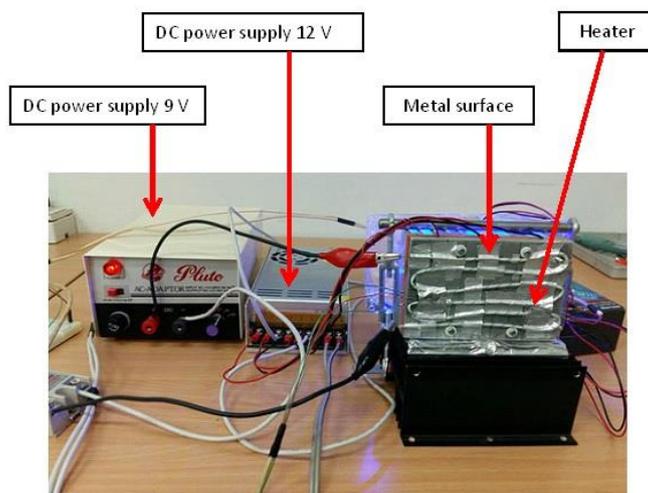
3.1.6 The DHT-22 (also named as AM2302) is a digital-output relative humidity and temperature sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin as shown in Figure 4.



**Figure 4.** The DHT-22.

### 3.2 Preparation of Laboratory

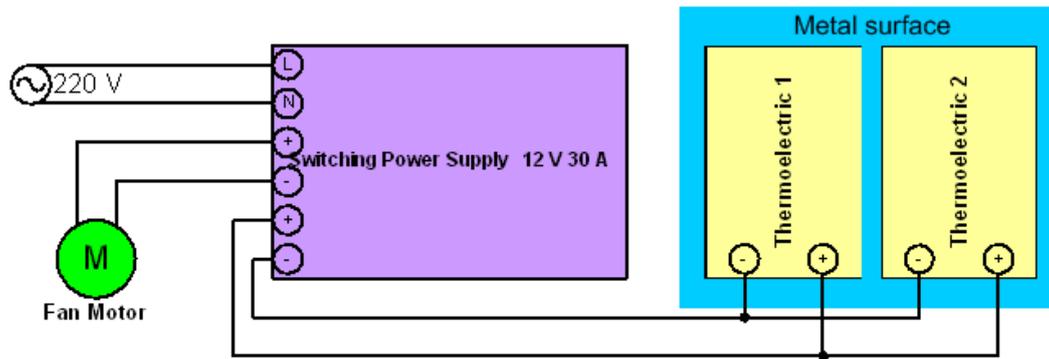
For preparation of laboratory, all materials are composed and designed circuit are made for studying the behavior of the water droplets on the specimen and controlling the occurrence of water droplets on Aluminium surface as shown in Figure 5.



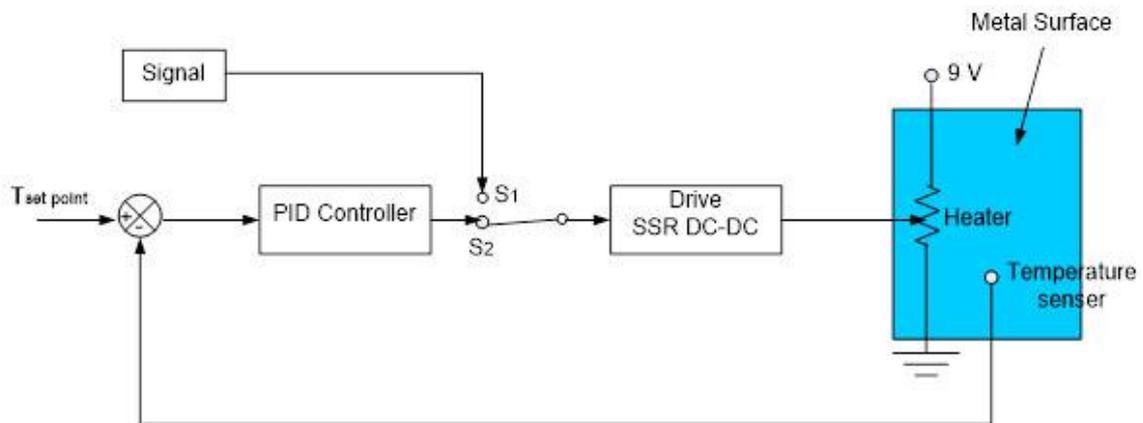
**Figure 5.** Experimental setup.

### 3.3 Method of Experiment

In this the circuit of design shown in Figure 6, 12 volts DC power was fed through the semiconductor type: n-type and p-type, which is inside the thermoelectric plate. One of the thermoelectric plate sides would be cool and another side of the plate would be hot. Cool thermoelectric plate side would attach to metal surface, its made by Aluminium, can cause droplets on metal surface. However, the controller designed could get rid of the water droplets on the surface as shown in Figure 7.



**Figure 6.** Circuit of design.



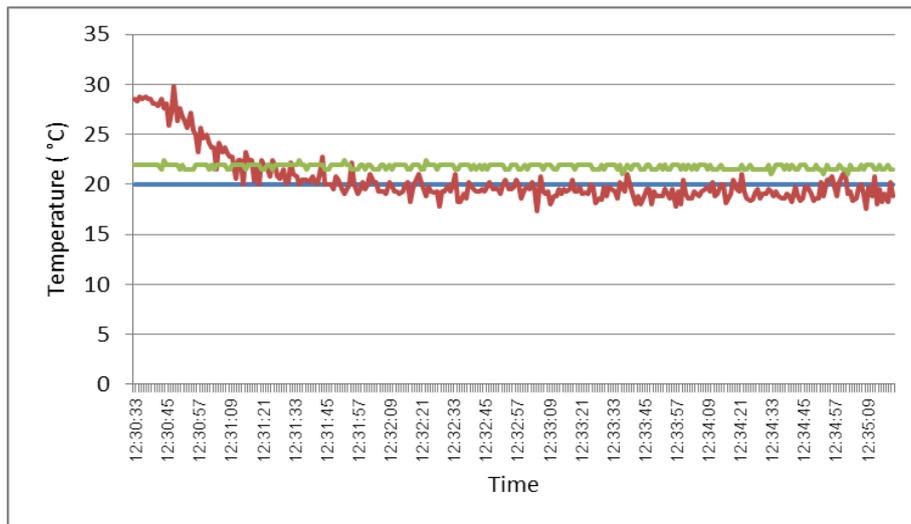
**Figure 7.** Control system of design.

From Figure 7, when the switch to S1 is pressed, the heat conductor connected to the 9V supply will cause the sheet to rise to a higher temperature, called open loop control. By switching to S2, the set point temperature can be set to greater or less than the dew point temperature through the PID controller and then passed to the solid state to control the heat conductor, causing the metal to rise to a higher temperature. And using a temperature sensor attached to the metal surface sends the feedback signal back to the set point temperature, which control the temperature at the metal surface, called closed loop control.

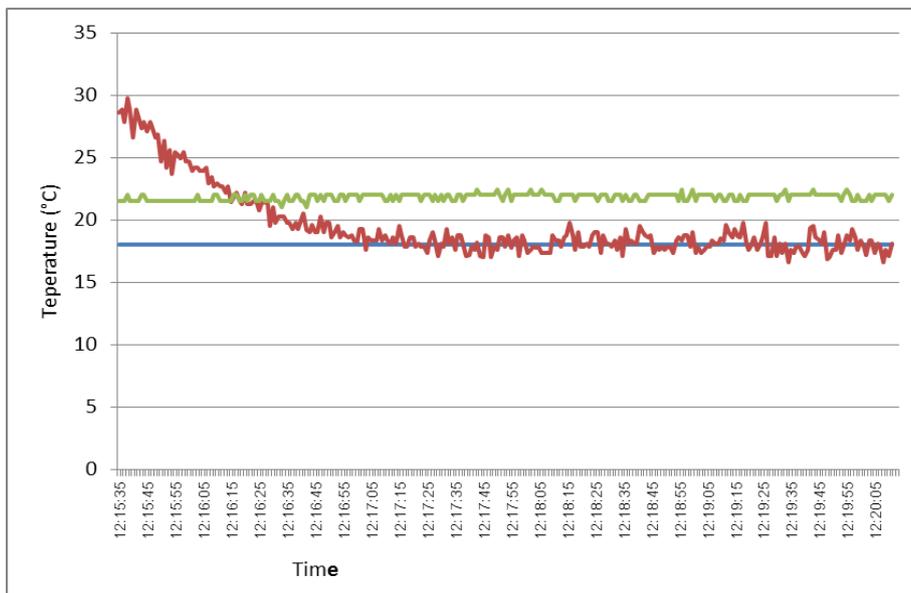
#### 4. RESULTS AND DISCUSSION

The experimental results in Figure 8. Dew point temperature is 19 °C. We can control the temperature more than the dew point temperature to 20 °C (the blue line). This brings to the set point temperature. The temperature (the red line) of the metal surface runs to the set point temperature in 60 sec. The water droplet was not occur on the metal surface as shown in Figure 10.

From the graph, the experiment results in Figure 9. Dew point temperature is 19 °C. We can control the temperature lower the dew point temperature to 18 °C (the blue line). This brings to the set point temperature. The temperature (the red line) of the metal surface runs to the set point temperature in 80 sec. And water droplet occur all over the metal surface in 5 min as shown in Figure 11.



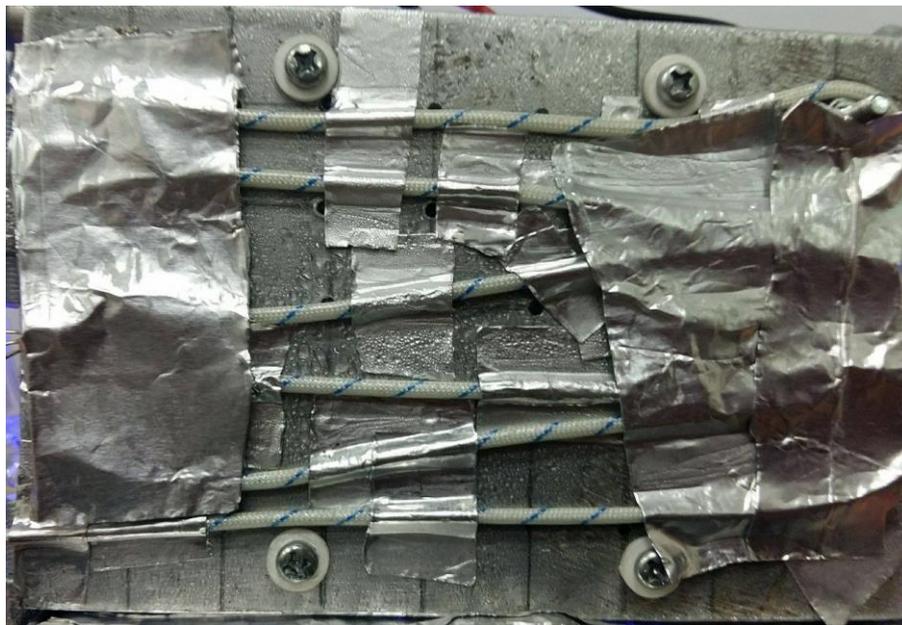
**Figure 8.** Temperature set point at 20 degree



**Figure 9.** Temperature set point at 18 degree



**Figure 10.** The dry metal surface.



**Figure 11.** The water droplet on metal surface.

## **5. CONCLUSION**

The results showed that the automatic controller for controlling the droplet on the designed surface could control the droplet at a temperature below the dew point temperature of 1 °C. And if the temperature is above the dew point of 1 °C, there will be no droplets. This shows that the automatic control unit is designed to control the droplet on the metal surface for its intended purpose. And in the future will design the control circuit in the smart controller anyway.

## REFERENCES

1. Ivan J. S. Lopes: Shesha H. Jayaram and Edward A. Cherney ,“A Study of Partial Discharges from Water Droplets on a Silicone Rubber Insulating Surface”, IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 8 No. 2, pp. 262-268, April 2001.
2. Laurent Latorre, Joonwon Kim, Junghoon Lee, Associate Member, IEEE, Peter-Patrick de Guzman, Hyesog J. Lee, “Electrostatic Actuation of Microscale Liquid-Metal Droplets”, JOURNAL OF MICROELECTRO MECHANICAL SYSTEMS, VOL. 11, NO. 4, pp. 302- 308, August 2002.
3. Luca Ferraris and Paolo Ferraris, “Approach to the problem of water condensation by means of Photo Voltaic energy” , IEEE Transactions on Dielectrics and Electrical Insulation, 2006.
4. Luca FERRARIS, Paolo FERRARIS, “Improvement of a static system for water condensation supplied with Photo Voltaic energy”, IEEE TRANSACTIONS ON PLASMA SCIENCE, pp. 1-5, 2016.
5. Mark G. Lawrence. “The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air.” AMERICAN METEOROLOGICAL SOCIETY, February 2005, pp. 225-233.
6. Titipong Kudeepun, Pichnaree Lalitapom, and Monthon Thanuttamavong. “Dew Formation from Thermoelectric Cooler.” The 54<sup>th</sup> Kasetsart University Annual Conference, 2015, pp.415-421.
7. Dae Heon Park, Chul Young Park, Sung Eon Cho, Jang Woo Park, “Greenhouse environment monitoring and Automatic control system based on dew condensation prevention”, 2010.
8. Dr. D. Senthil Kumar, R. Yuvaraj, “Simulation of Dropwise Condensation on a Superhydrophobic Inclined Substrate”, 2013, pp. 217-222.
9. Xiao Tan, Yunhao Qiu, Yong Yang, Dawei Liu, *Member,IEEE*, XinPei Lu, *Senior Member, IEEE*, and Yuan Pan, “Enhanced Growth of Single Droplet by Control of Equivalent Charge on Droplet”, 2016, pp. 1-5.
10. Martin Wanielista, Robert Kersten and Ron Eaglin. 1997. Hydrology Water Quantity and Quality Control. John Wiley & Sons. 2<sup>nd</sup> ed.

## ACKNOWLEDGMENTS

The authors would like to Acknowledge support received from Energy Control Research Unit and Pathumwan Institute of Technology.

# The Meshless Local Petrov-Galerkin Method for Solving the New Black-Scholes-Schrodinger Model

N. Nualsaard<sup>1,C</sup>, A. Luadsong<sup>1,2</sup>, and N. Ascharyaphotha<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-utid Road, Bangkok 10140, Thailand

<sup>2</sup> Ratchaburi Learning Park, King Mongkut's University of Technology Thonburi (KMUTT), Rang Bua, Chom Bueng, Ratchaburi 70150, Thailand

<sup>C</sup> E-mail: naravadee401@gmail.com; Fax: +66 24284025; Tel. +66 0817345895

## ABSTRACT

The meshless local Petrov-Galerkin (MLPG) method is applied for solving the new Black-Scholes-Schrodinger model. The MLPG method is used for spatial discretization and the  $\theta$ -weighted method is chosen for temporal discretization. In MLPG method, there are creating the local weak form over a local sub-domain. The shape function is constructed by the moving kriging interpolation. Finally, the Kronecker delta function is chosen as the test function in each sub-domain for simplifying the equation. The numerical experiment is compared with the semi-classical solution to verify the results. The results show that the option price from MLPG method are agreed to semi-classical solution.

**Keywords:** Arbitrage, New Black-Scholes-Schrodinger Model, Meshless Local Petrov-Galerkin Method, Moving Kriging Interpolation, Option Pricing, Quantum Mechanics,  $\theta$ -Weighted Method.

## 1. INTRODUCTION

In 1973, Black and Scholes [5] proposed the Black-Scholes equation, which is the famous financial model for option pricing. This model has more important for the calculation of the price and options trading due to the simplicity of the model and it can approach to the solution in a short time. The Black-Scholes model is used for evaluating European (the option can only be exercised on the expiration date) or American (the option can be exercised at any time before the expiration date) call and put options. The basic assumptions of the Black-Scholes model as following: there are no arbitrage opportunities, trading in all securities is continuous, no transaction costs or taxes, no dividends, and it is possible to lend and borrow cash at a constant risk-free interest rate. During the last decades, there are many researchers studied the numerical method for solving the Black-Scholes model using several methods in [1-3, 6-9, 12-13, 17].

The Black-Scholes equation can be interpreted from the point of view of quantum mechanics, as the imaginary time Schrodinger equation of a free particle [11]. Due to some conditions, the Black-Scholes model cannot explain such as the effects of arbitrage on the option pricing dynamics. Haven E. (2002) [14] presented that an arbitrage is a necessary condition of the Black-Scholes model in a more general of quantum physics. Contreras, M., et al. (2010) [11] propose the new Black-Scholes-Schrodinger model which can be described the Black-Scholes model with the arbitrage possibility in the sense of the Schrodinger equation. The semi-classical approximation method is applied to this new Black-Scholes-Schrodinger model.

In recent decades, the meshless or meshfree method is more interested and popular for solving an engineering and practical science because it have several advantage. The main advantage of meshless method is it can save computational time because this method do not use meshes, only nodes. This method can easily for increasing the support domain, changing basis and adopting nodal densities. They can easily construct high-order shape functions [15].

In addition, process of meshless method is not complexity for solving the results. There are many of meshless method, such as the element free Galerkin (EFG) method, the point interpolation method (PIM), the radial point interpolation method (RPIM), the point assembly method (PAM), the meshless local Petrov-Galerkin (MLPG) method, etc. The meshless local Petrov-Galerkin (MLPG) method is one of most popular meshless method which was first discovered by Atluri and Zhu (1998) [4]. MLPG method is different from other meshless method because it not requires background cells for numerical integration or interpolation, so the MLPG is a truly meshless method. Several examples for applications of MLPG method can be found in wide area such as in electrochemistry, diffusion-reaction processes, viscoelasticity, financial modeling etc. Phaochoo P., et al. (2016) [16] present a numerical study of the European option by the meshless local Petrov-Galerkin (MLPG) method with moving kriging interpolation.

The purpose of this paper, the meshless local Petrov-Galerkin (MLPG) method is applied for solving the new Black-Scholes-Schrodinger model which is proposed by [11]. The MLPG method is used for spatial discretization and the  $\theta$ -weighted method is chosen for temporal discretization. In MLPG method, the shape function is constructed by the moving kriging interpolation. The Kronecker delta function is chosen as the test function. The numerical experiment is compared with the semi-classical solution to verify the results.

## 2. THE NEW BLACK-SCHOLES-SCHRODINGER MODEL

In 2010, the new Black-Scholes-Schrodinger model is proposed by [11]. This new Black-Scholes-Schrodinger model can be described the Black-Scholes model with the arbitrage possibility in the sense of the Schrodinger equation. Transforming the Black-Scholes model with the arbitrage possibility to the new Black-Scholes-Schrodinger model as following:

$$\frac{\partial \pi}{\partial t} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 \pi}{\partial s^2} + r \frac{\left(\sigma - \frac{\alpha f}{r}\right)}{(\sigma - f)} \left(s \frac{\partial \pi}{\partial s} - \pi\right) = 0, \quad (s, t) \in \mathbb{R}^+ \times [0, T] \quad (2.1)$$

where  $t$  is time variable,  $s$  is underlying asset price,  $\pi(s, t)$  represents the option price,  $\sigma$  is the volatility of underlying asset price,  $r$  is the risk-free interest rate,  $f(t)$  is arbitrage bubble and  $T$  is the expiration date. Transforming Eq. (2.1) by taking the variable change  $\xi = \ln s$ , to obtain

$$\frac{\partial \pi}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 \pi}{\partial \xi^2} + \left(r - \frac{1}{2} \sigma^2\right) \frac{\partial \pi}{\partial \xi} + \frac{(r - \alpha)f}{\sigma - f} \left(\frac{\partial \pi}{\partial \xi} - \pi\right) - r\pi = 0. \quad (2.2)$$

Making a second (time dependent) change of variable  $x = \xi - \left(r - \frac{1}{2} \sigma^2\right) t$ , and become to

$$\begin{aligned} \frac{\partial \pi}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 \pi}{\partial x^2} + \frac{(r - \alpha) \check{f}}{\sigma - \check{f}} \left( \frac{\partial \pi}{\partial x} - \pi \right) - r\pi \\ = 0 \end{aligned} \quad (2.3)$$

where  $\check{f}(x, t) = f\left(e^{x + (r - \frac{1}{2}\sigma^2)t}, t\right)$ .

Given the Black-Scholes model with arbitrage and defining by [11]

$$\begin{aligned} \pi(x, t) \\ = e^{-r(T-t)} \psi(x, t). \end{aligned} \quad (2.4)$$

The new Black-Scholes-Schrodinger model is following.

$$\begin{aligned} \frac{\partial \psi(x, t)}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 \psi(x, t)}{\partial x^2} + v(x, t) \left( \frac{\partial \psi(x, t)}{\partial x} - \psi(x, t) \right) = 0, \quad (x, t) \\ \in \mathbb{R}^+ \times [0, T] \end{aligned} \quad (2.5)$$

where  $\psi(x, t)$  is wave function,  $v(x, t)$  is the potential energy and

$$\begin{aligned} v(x, t) \\ = \frac{(r - \alpha) \check{f}(x, t)}{\sigma - \check{f}(x, t)}. \end{aligned} \quad (2.6)$$

The semi-classical solution of the Black-Scholes model with arbitrage, in presence of a time dependent arbitrage bubble  $f = f(t)$  can be computed [11] as

$$\begin{aligned} \pi_{sc}(s, t) \\ = \frac{1}{e^{\rho(t, T)}} \pi_{BS}(e^{\rho(t, T)} s, t) \end{aligned} \quad (2.7)$$

where  $\pi_{BS}(s, t)$  is the arbitrage-free Black-Scholes solution for the specific option with contract  $\Phi(s)$ . In this way, the function  $\rho(t, T)$  renormalizes the bare arbitrage-free BS solution [11].

### 3. SPATIAL AND TEMPORAL DISCRETIZATION

In this section, the meshless local Petrov-Galerkin (MLPG) method is introduced for spatial discretization and the  $\theta$ -weighted method is chosen for temporal discretization. Both techniques for discretization are described in detail.

#### 3.1 Spatial Discretization

The meshless local Petrov-Galerkin (MLPG) method is used for spatial discretization. There are creating the local weak form over local sub-domain,  $\Omega_s$  which is a small region taken for each node in the global domain,  $\Omega \in \mathbb{R}$ . First of all, the new Black-Scholes-Schrodinger equation, Eq. (2.5), can be weighted by test functions,  $w_i(x)$

$$\begin{aligned}
& \frac{\partial \psi(x, t)}{\partial t} w_i(x) \\
&= \left[ -\frac{1}{2} \sigma^2 \frac{\partial^2 \psi(x, t)}{\partial x^2} \right. \\
&\quad \left. - v(x, t) \left( \frac{\partial \psi(x, t)}{\partial x} - \psi(x, t) \right) \right] w_i(x) \quad (3.1)
\end{aligned}$$

Next, integrating over a local sub-domain which associate with the point  $x_i$ ;  $i = 1, 2, \dots, N$ , where  $N$  is the number of nodes surrounding point  $x$ .

$$\begin{aligned}
& \int_{\Omega_s^i} \left( \frac{\partial \psi(x, t)}{\partial t} \right) w_i(x) d\Omega \\
&= \int_{\Omega_s^i} \left[ -\frac{1}{2} \sigma^2 \frac{\partial^2 \psi(x, t)}{\partial x^2} \right. \\
&\quad \left. - v(x, t) \left( \frac{\partial \psi(x, t)}{\partial x} - \psi(x, t) \right) \right] w_i(x) d\Omega \quad (3.2)
\end{aligned}$$

where  $\Omega_s^i$  is a local sub-domain which associates with the point  $x_i$ , and  $w_i(x)$  is a test function.

Substituting the trial function,  $\psi^h(x, t) = \sum_{j=1}^N \phi_j(x) \hat{\psi}_j(t)$  for  $\psi(x, t)$  into the local weak form Eq. (3.2)

$$\begin{aligned}
& \int_{\Omega_s^i} \left( \sum_{j=1}^N \phi_j(x) \hat{\psi}_j'(t) w_i(x) \right) d\Omega \\
&= \int_{\Omega_s^i} \left[ -\frac{1}{2} \sigma^2 \sum_{j=1}^N \phi_{j,xx}(x) \hat{\psi}_j(t) \right. \\
&\quad \left. - v(x, t) \left( \sum_{j=1}^N \phi_{j,x}(x) \hat{\psi}_j(t) - \sum_{j=1}^N \phi_j(x) \hat{\psi}_j(t) \right) \right] w_i(x) d\Omega
\end{aligned}$$

(3.3)

where  $\phi_j(x)$  is the shape function which is constructed by moving kriging interpolation [18],  $\hat{\psi}_j(t)$  is value of  $\psi$  for position  $x_j$  at time,  $t$ ,  $\phi_{j,xx}(x) = \frac{\partial^2 \phi_j(x, t)}{\partial x^2}$  and  $\phi_{j,x}(x) = \frac{\partial \phi_j(x, t)}{\partial x}$ .

Rearranging Eq. (3.3) as following

$$\begin{aligned}
\sum_{j=1}^N \left( \int_{\Omega_s^i} \phi_j(x) w_i(x) d\Omega \right) \hat{\psi}_j'(t) &= -\frac{1}{2} \sigma^2 \sum_{j=1}^N \left( \int_{\Omega_s^i} \phi_{j,xx}(x) w_i(x) d\Omega \right) \hat{\psi}_j(t) \\
&\quad - \sum_{j=1}^N \left( \int_{\Omega_s^i} v(x,t) \phi_{j,x}(x) w_i(x) d\Omega \right) \hat{\psi}_j(t) \\
&\quad + \sum_{j=1}^N \left( \int_{\Omega_s^i} v(x,t) \phi_j(x) w_i(x) d\Omega \right) \hat{\psi}_j(t), \\
i &= 1, 2, \dots, N. \tag{3.4}
\end{aligned}$$

The Kronecker delta function is chosen as the test function in each sub-domain:

$$w_i(x) = \begin{cases} 0, & x \neq x_i \\ 1, & x = x_i \end{cases}; \quad i = 1, 2, \dots, N.$$

Substituting the test function into Eq. (3.4) and integrate over sub-domain as following

$$\begin{aligned}
\sum_{j=1}^N \phi_j(x_i) \hat{\psi}_j'(t) &= \sum_{j=1}^N \left[ -\frac{1}{2} \sigma^2 \phi_{j,xx}(x_i) \right. \\
&\quad \left. - v(x_i, t) (\phi_{j,x}(x_i) - \phi_j(x_i)) \right] \hat{\psi}_j(t) \tag{3.5}
\end{aligned}$$

for  $i=1, 2, \dots, N$ . Equation (3.5) can be written in the matrix form as

$$A \frac{d\Psi}{dt} = B\Psi, \tag{3.6}$$

where

$$A = [A_{ij}]_{N \times N}, \quad A_{ij} = \phi_j(x_i)$$

$$B = [B_{ij}]_{N \times N}; \quad B_{ij} = -\frac{1}{2} \sigma^2 \phi_{j,xx}(x_i) - v(x_i, t) (\phi_{j,x}(x_i) - \phi_j(x_i)) \quad ; i = 1, 2, \dots, N$$

$$j = 1, 2, \dots, N$$

$$\Psi = [\hat{\psi}_1 \hat{\psi}_2 \dots \hat{\psi}_N]^T.$$

Since the moving kriging interpolation which has the delta function property is used for constructing nodal shape function. For this reason,  $A$  is the identity matrix. Therefore, Eq. (3.6) can be written as

$$\frac{d\Psi}{dt} = B\Psi. \tag{3.7}$$

### 3.2 Temporal Discretization

The implicit method for approximate solution is used for solving the new Black-Scholes-Schrodinger equation. There are modifying PDE operator through a finite difference approximation of time derivative, using the  $\theta$ -weighted method. The nodal points in the time interval  $[0, T]$  are given by  $t_n = n\Delta t, n = 0, 1, 2, \dots$ . The finite difference approximation of the time derivative of equation (3.7) for the  $\theta$ -weighted method as

$$\frac{\hat{\psi}^{n+1} - \hat{\psi}^n}{\Delta t} = \theta B \hat{\psi}^{n+1} + (1 - \theta) B \hat{\psi}^n \quad (3.8)$$

Rearranging Eq. (3.8) lead to

$$(I - \theta \Delta t B) \Psi^{n+1} = [I + (1 - \theta) \Delta t B] \Psi^n \quad (3.9)$$

where  $\Psi = [\hat{\psi}_1 \hat{\psi}_2 \dots \hat{\psi}_N]^T$ ,  $B$  is the space discretization matrix and  $I$  is the identity matrix.

## 4. RESULTS AND DISCUSSION

In this section, some numerical results are presented to illustrate the implementation of the present MLPG method for solving the new Black-Scholes-Schrodinger model. The numerical results are compared with a semi-classical solution to verify the result.

From the Black-Scholes model with the arbitrage possibility, Eq. (2.1)

$$\frac{\partial \pi}{\partial t} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 \pi}{\partial s^2} + r \frac{\left(\sigma - \frac{\alpha f}{r}\right)}{(\sigma - f)} \left(s \frac{\partial \pi}{\partial s} - \pi\right) = 0, (s, t) \in \mathbb{R}^+ \times [0, T] \quad (4.1)$$

Contract function,  $\Phi(s)$  is used for a binary put

$$\Phi(s) = \begin{cases} 1, & 0 < K \\ 0, & K < s \end{cases} \quad (4.2)$$

where  $K$  is the exercise price.

The semi-classical solution for the Black-Scholes model with arbitrage, in presence of a time dependent arbitrage bubble  $f = f(t)$ , Eq. (2.7) can be computed [11] as

$$\pi_{sc}(s, t) = \frac{1}{e^{\rho(t, T)}} \pi_{BS}(e^{\rho(t, T)} s, t) \quad (4.3)$$

where  $\pi_{BS}(s, t)$  is the arbitrage-free Black–Scholes solution for the specific option with contract  $\Phi(s)$ . The pure Black–Scholes solution  $\pi_{BS}(s, t)$  as following

$$\pi_{BS}(s, t) = e^{-r(T-t)} [1 - N(d_2(s, t))] \quad (4.4)$$

where  $N(x)$  is the normal distribution function and

$$d_2(s, t) = \frac{\ln \frac{s}{K} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{(T-t)}} \quad (4.5)$$

Transforming the Black-Scholes model with arbitrage to the new Black-Scholes-Schrodinger model from Eqs. (2.1) – (2.5) as

$$\frac{\partial \psi(x, t)}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 \psi(x, t)}{\partial x^2} + v(x, t) \left( \frac{\partial \psi(x, t)}{\partial x} - \psi(x, t) \right) = 0, (x, t) \in \mathbb{R}^+ \times [0, T] \quad (4.6)$$

with initial condition

$$\psi(x, T) = \begin{cases} 1, & 0 < x < \ln K - \left(r - \frac{1}{2}\sigma^2\right)T \\ 0, & \ln K - \left(r - \frac{1}{2}\sigma^2\right)T < x \end{cases} \quad (4.7)$$

The boundary condition as following

$$\begin{aligned} \psi(x_1, t) &= e^{r(T-t)} \pi_{sc} \left( e^{x_1 + \left(r - \frac{1}{2}\sigma^2\right)t}, t \right) \\ \psi(x_N, t) &= e^{r(T-t)} \pi_{sc} \left( e^{x_N + \left(r - \frac{1}{2}\sigma^2\right)t}, t \right). \end{aligned} \quad (4.8)$$

#### 4.1 Time Step Arbitrage bubble

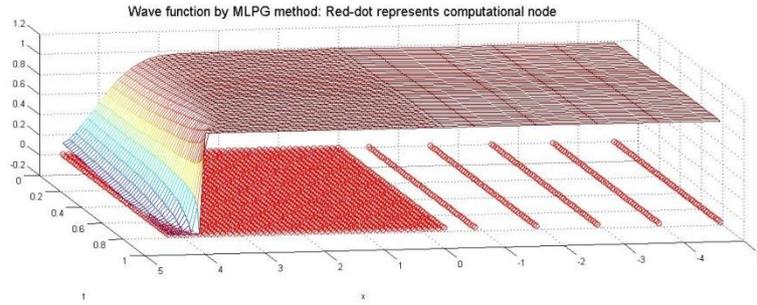
Time step bubble is given by

$$f(t) = \begin{cases} 0, & 0 < t < T_1 \\ H, & T_1 < t < T_2 \\ 0, & T_2 < t < T \end{cases} \quad (4.9)$$

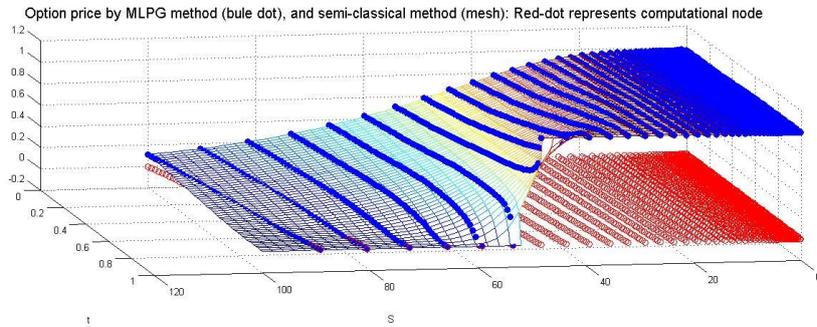
We obtain for the  $\rho$  factor

$$\rho(t, T) = \begin{cases} (T_2 - T_1) \frac{(r - \alpha)H}{\sigma - H}, & 0 < t < T_1 \\ (T_2 - t) \frac{(r - \alpha)H}{\sigma - H}, & T_1 < t < T_2 \\ 0, & T_2 < t < T \end{cases} \quad (4.10)$$

In this paper, we analyze the option price for  $\alpha = -0.6$ ,  $\sigma = 0.5$ . We use  $T_1 = 0.3$ ,  $T_2 = 0.6$  and  $H = 0.1\sigma$  for numerical experiment. The results are presented in figures 4.1 and 4.2. Figures 4.1 shows wave function which is solved by MLPG method. In figures 4.2, option price is solved by MLPG method and semi-classical method. From the numerical experiment, we can conclude that the option prices from MLPG method are agreed to semi-classical solution.



**Figure 4.1.** The wave function is solved by MLPG method.



**Figure 4.2.** The option price is solved by MLPG method and semi-classical method.

## 5. CONCLUSION

In this paper, the meshless local Petrov-Galerkin (MLPG) method is introduced as a new numerical method for solving the new Black-Scholes-Schrodinger model which is proposed by [11]. The MLPG method is used for spatial discretization process. The moving kriging interpolation which has the delta function property is used for constructing nodal shape function. The Kronecker delta function is chosen as the test function in each sub-domain for simplifying the equation. The  $\theta$ -weighted method is chosen for temporal discretization. The numerical experiment is compared with the semi-classical solution to verify the results. The results show that the option price from MLPG method are agreed to semi-classical solution.

## REFERENCES

1. Amster, P., Averbuj, CG., and Mariani, MC., *J Math Anal Appl*, 2002, **276**, 231-238.
2. Amster, P., Averbuj, CG., and Mariani, MC., *Appl Numer Math*, 2003, **47**, 275-280.
3. Ankudinova, J., and Ehrhardt, M., *Comput Math Appl*, 2008, **56**, 799-812.
4. Atluri, SN., and Shu, T., *Computational Mechanics*, 1998, **22**, 117-127.
5. Black, F., and Scholes, M., *Journal of Political Economy*, 1973, **81**(3), 637-654.
6. Bohner, M., and Zheng Y., *Appl Math Lett*, 2009, **22**, 309-313.
7. Cen, Z., and Le, A., *J Comput Appl Math*, 2011, **235**, 3728-3733.
8. Company, R., Navarro, E., Pintos, JR., and Ponsoda, E., *Comput Math Appl*, 2008, **56**, 813-821.
9. Company, R., Jodar, L., and Pintos, JR., *Math Comput Model*, 2009, **50**, 910-920.
10. Contreras, M., Montalva, R., Pellicer, R., and Villena, M., *PhysicaA*, 2010, **389**, 3552-3564.
11. Contreras, M., Pellicer, R., Villena, M., and Ruiz, A., *PhysicaA*, 2010, **389**, 5447-5459.
12. Fabiao, F., Grossinho, MR., and Simoes, OA., *Nonlinear Anal*, 2009, **71**, 4624-4631.
13. Gulkac, V., *J Stat Comput Simul*, 2010, **80**, 1349-1354.
14. Haven, E., *PhysicaA*, 2002, **304**, 507-524.
15. Khankham, S., Luadsong, A., and Ascharyaphotha, N., *Journal of King Saud University- Science*, 2015, **27**, 292-301.
16. Phaochoo, P., Luadsong, A., and Ascharyaphotha, N., *SpringerPlus*, 2016, **5**(305), 1-14.
17. Sekhara Rao and Manisha, S., *Procedia Computer Science*, 2016, **80**, 1765 – 1776.
18. Yimnak, K., and Luadsong, A., *Advance in Mathematical Physics*, 2014, **2014**, 1-7.

## ACKNOWLEDGMENTS

This research was supported by Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT). The authors would like to thank their advisor for providing advice and taking care of this research. Finally, the authors would like to thank Uttaradit Rajabhat University for providing a scholarship.

# The Development of a VM Auto-Scaling Software for OpenStack Cloud

Krailerk Manopattanagorn<sup>1,c</sup> and Putchong Uthayopas<sup>1</sup>

<sup>1</sup> Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand.

<sup>c</sup> E-mail: krailerkm@gmail.com, putchong@ku.th; Tel. +66814944962

## ABSTRACT

In this paper, the implementation of an Auto-Scaling support for OpenStack called Media Auto-scale Streaming Service (MASS) system is presented. The MASS system is built as a part of OpenStack system can employ various subsystem in OpenStack to function using OpenStack API. The MASS system enables users to automatically scale the number of VM according to increases and decreases workload. The experiments have been conducted and show that the use of MASS help increases the performance of data services of OpenStack system to match the increases benchmark workload. The work can be employed for heavy loaded services such as media streaming from the cloud, massive web and e-commerce service, large scale social media site.

**Keywords:** Auto-Scaling, Cloud Computing, OpenStack, Virtualization.

## 1. INTRODUCTION

In the past decade, cloud computing system (Mell, P. and Grance, T., 2011) becomes an important platform for the next generation IT services. One of the substantially important cloud services is a media streaming service that can scale to support many users. As the number of users increases, an ability to automatically scale to match the users' workload becomes a critical issue. In this paper, an implementation and design of a software called *MASS (Media Auto-scale Streaming Service)* is presented. The goal of this software is to add an auto-scaling capability to the widely used OpenStack (V. K. Cody and Bumgardner, 2016) open source cloud computing software. This software enables the developers to easily add a smart auto-scaling policy to improve the performance an OpenStack Cloud system.

## 2. THEORY AND RELATED WORKS

There are many solutions that can be employed to build the private and public clouds. One of them is OpenStack, a widely used open source cloud distribution platform. The OpenStack Cloud can offer services such as VM hosting for application services that enable more efficient resources sharing among applications. The auto-scaling is an ability to automatically scale resources to fit the demand of users. This feature requires the coordination between multiple modules inside OpenStack cloud. An intermediate module called *Heat* is developed that provides basic auto-scaling capability (F. F. Moghaddam, A. Gherbi and Y. Lemieux, 2016). Another module names, Ceilometer is used as a telemetry that gives the system status of resources under OpenStack (B. I. Santoso, M. R. S. Idrus and I. P. Gunawan, 2016). Moreover, Neutron is a module that provides the network configuration together with the load balancing as a service (B. Dongmyoung and L. Bumchul, 2015). The feature of auto-scaling is one of the most popular services on the public cloud since this can enable a more flexible use of resources for users. The auto-scaling on a public cloud usually starts by users setting up the minimum and a maximum number of machines in a group and deploy the applications on that group of resources. After that, the cloud will measure the workload and adjust a number of machines accordingly. The speed of measurement of resources usage will have some impact on the performance of the system (F. L. Ferraris et al., 2012). One of the approaches that are used to control the scaling is to set the threshold of workload scale up

when a threshold is reached (F. Al-Haidari, M. Sqalli and K. Salah, 2013). This method is simple and works well for a normal workload. Anyway, when more precise control is needed, the historical data can be employed to predict the proper time to create a new machine (N. Roy, A. Dubey and A. Gokhale, 2011). The use of auto-scaling not only help maintain the performance of application user increases workload but also help reduce the energy consumption of the application as well (K. Kanagala and K. C. Sekaran, 2013) (M. Tighe and M. Bauer, 2014). Most of the system increase new machine having the same capacity but the use of the small and large machine to match the increased workload can help improve the system performance as well (K. Hwang, Y. Shi and X. Bai, 2014).

The system architecture of the OpenStack platform, which is used in this work, is as depicted in Figure 1. OpenStack consists of many interrelated subsystems creating from many cloud projects. OpenStack coordinates these modules to function together to provide the cloud services. For instance, users can create a new VM using the dashboard which is called Horizon subsystem. Anyway, the user request must be authenticated using a subsystem called Keystone. New VM is the created using the Nova subsystem. The new VM will rely on image template obtained from the glance service. The new VM is then linked to the network using neutron subsystem. To create MASS software, the new MASS subsystem has to be built and linked to another subsystem to perform the auto-scale function as well.

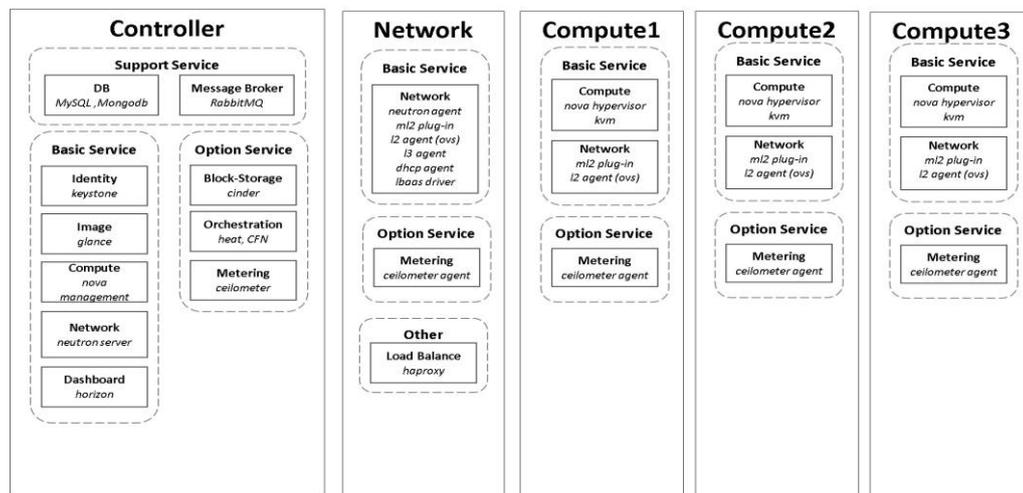


Figure 1. OpenStack configuration environment

The MASS software is designed bases on an OpenStack environment. The MASS system works by coordinating with OpenStack modules to increase the number and size of the resource via OpenStack controller API. The mass system is developed in Python language. The main components of MASS and related subsystems is as illustrated in Figure 3.

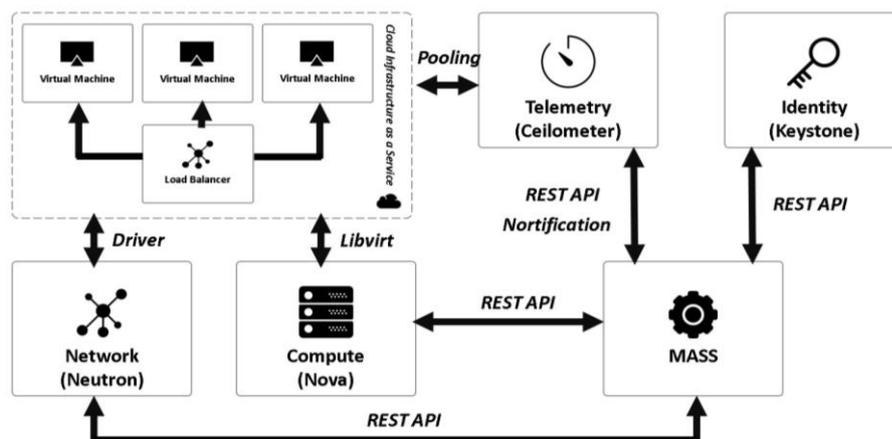


Figure 2. Media Auto-scale Streaming Service (MASS)

MASS relies on OpenStack identity service for authentication. Then, the Ceilometer telemetry service is employed to measure the utilization, generate the notification when there is a need to increase or decrease resources. As the load measured by Ceilometer increases, MASS will wait for the threshold to be reached. Once the user defined threshold has been reached, the notification will be received by MASS. MASS will increase the VM according to the demand of users. The control of VM has been done using OpenStack API that connects to Nova subsystem. In MASS, templates of three VMs have been created. The small size (S) VM is composed of 1 vCPU, RAM 512MB, Disk 5GB. The medium size VM (M) has 2 vCPU, RAM 512MB, Disk 5GB. Finally, the large size (L) has 4 vCPU, RAM 512MB, Disk 5GB. The different size VMs enable us to select the proper set of VMs based on application characteristics. As the workload increases, the system will automatically deploy more VMs in the VM group. However, when the workload decreases, MASS will shut down and delete some VMs automatically. The execution of the MASS system is as explained in Figure 3.

MASS auto-scale algorithm	
Input: Minimum and maximum number of VM	
Output: Resource scheduling scale = (Number of VM)	
1:	Initialize token id for define auto-scale group
2:	Create pool at load balance service
3:	Create virtual IP at pool load balance
4:	Associate floating IP to VIP pool load balance
5:	Create health service monitoring
5:	Create the notification high for scale-out from monitoring if threshold CPU over 70%
6:	Create the notification low for scale-in from monitoring if threshold CPU lower 15%
7:	While(true)
8:	If(VM=0 or Action=high or Count VM < maximum bound of VM)
9:	If (Count VM % 3 = 0)
10:	VM Size = Small
11:	LB Weight = 1
12:	Else if (Count VM % 3 = 1)
13:	VM Size = Medium
14:	LB Weight = 2
15:	Else if (Count VM % 3 = 2)
16:	VM Size = Large
17:	LB Weight = 4
18:	Create VM = VM Size
19:	Add VM to pool load balance by Weight = LB Weight
20:	Else if(action=low or Count VM > minimum bound of VM)
21:	Remove IP new VM out of pool load balance
22:	Delete new VM
23:	Action = receive notification

**Figure 3.** Algorithm used by MASS

### 3. EXPERIMENTAL DETAILS

In this work, the test environment is composed of a set of 10 computers. One computer is used to run the MASS subsystem. Two computers are used as the OpenStack controller and network services. These computers are based on CPU Intel E750 2.93Ghz, Memory 4GB DDR3, HDD160GB SATA installed with Linux CentOS Version 7.2. In the test-bed, 3 computers are used as OpenStack compute nodes. These computers use CPU Intel Q6600 2.4Ghz, Memory 6GB DDR2, HDD80GB SATA. The test has been conducted using JMeter version 3.1 (B. Erinle, 2014) as the workload generator. The master load test computer is based on CPU Intel I7-4700MQ 2.4Ghz, Memory 16GB DDR3, SSD240GB SATA installed with Windows 7. Three more load test slaves are based on CPU Intel E750 2.93Ghz, Memory 2GB DDR3, HDD160GB SATA. For the master JMeter, the operating system used is Linux

CentOS version 7. The rest is installed with Linux CentOS Version 7.2. The test configuration is as shown in Figure 4.

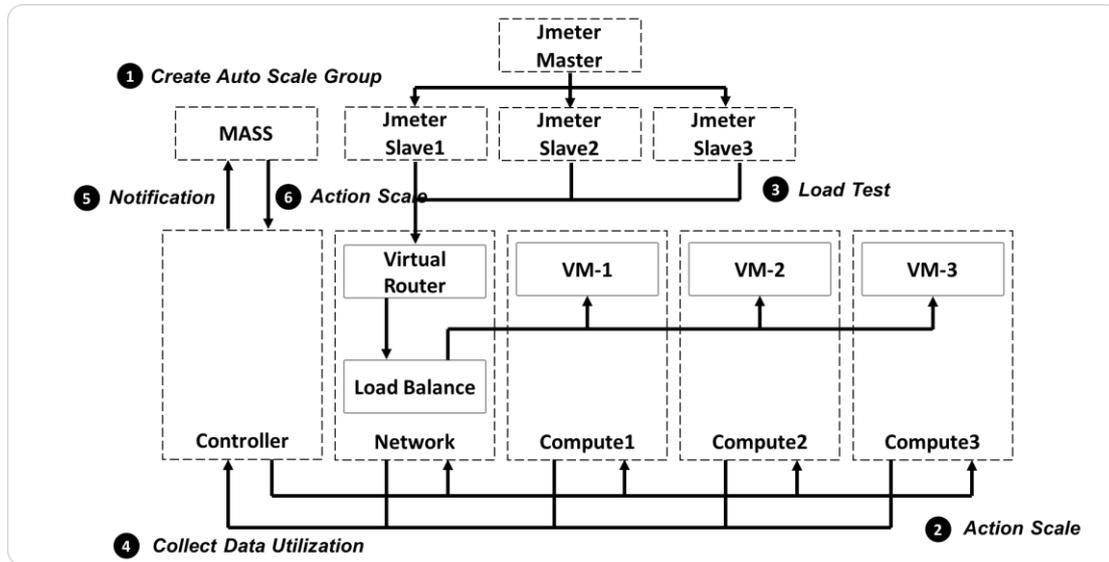


Figure 4. Test system configuration environment.

#### 4. RESULTS AND DISCUSSION

The experiments have been conducted using JMeter cluster sending workload to the testbed system. The workload is increased on three machines one thread at a time and reaching 150 threads in 24 minutes. First, the test has been conducted using a single VM system and 3 VM system. Each VM is a small size VM. The results are as shown in Figure 5.

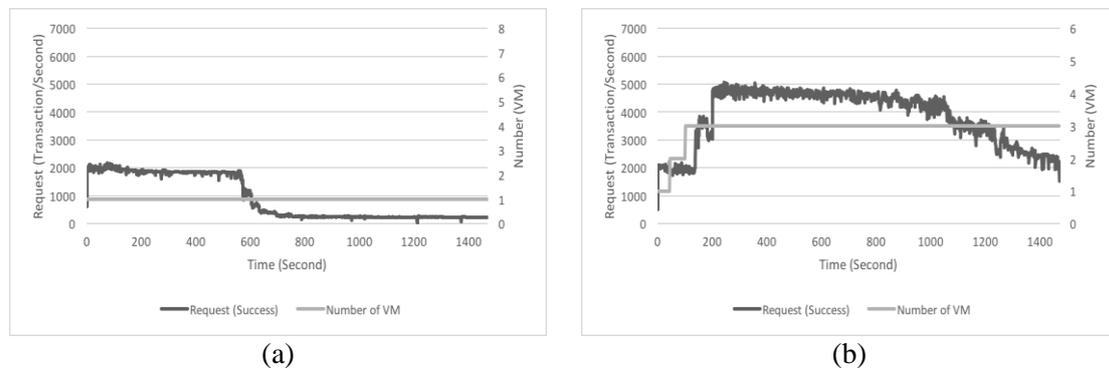
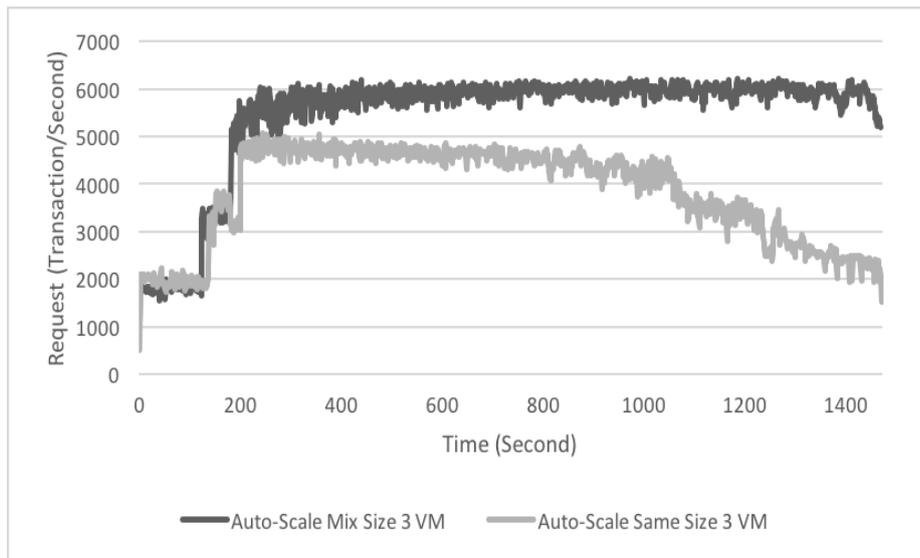


Figure 5. (a) Performance of a Standalone 1 VM System (Small size)  
(b) Performance of an Auto-Scale 3 VM (Small size)

From Figure 5, it can be seen that a single VM can deliver a throughput about 2000 transaction per second. After a while, the system will have a problem and cannot deliver the workload anymore. When the MASS system is used the system can scale and deliver a much better throughput at almost 5000 transactions per second. This show that the auto-scale implementation of MASS works and help improve the capacity of the system.

One of an interesting concept is to also increase the size of VM and the number of VM at the same time. So, the experiment is conducted using 3 VMs. The comparison of using 3 VM with the same VM size and 3 VMs with the scaled size has been conducted. The results are as shown in Figure 6.



**Figure 6.** The comparison of the auto-scaling with the same and scaled VM size.

The results clearly show the performance advantage of using scaled VM size over the fixed VM size. Again, the MASS system can help deliver a better performance for the OpenStack cloud due to the auto-scaling capability.

## 5. CONCLUSION

In this work, the design of an Auto-scaling system for OpenStack cloud called MASS has been presented. MASS integrate itself as a subsystem to OpenStack and use OpenStack API to implement an auto-scaling feature. The experiment has been conducted and it is found that MASS system can increase the performance of VM service in OpenStack. This MASS system can be applied to build a flexible and scalable cloud system for web application and media streaming as well. In the future, more complex and better scaling algorithm can be implemented to enable MASS to perform better and help users create a cost effective and power efficient system.

## REFERENCES

1. Mell, P. and T. Grance, *The NIST Definition of Cloud Computing*, 2011, NIST Special Publication, 800-145.
2. V. K. Cody Bumgardner, *OpenStack in Action*, Manning Publication Co., Shelter Island, 2016, 113-216.
3. B. Erinle, *JMeter Cookbook*, Packt Publication, Birmingham, 2014, 265-270.
4. F. F. Moghaddam, A. Gherbi and Y. Lemieux, *Self-Healing Redundancy for OpenStack Applications through Fault-Tolerant Multi-Agent Task Scheduling*, 2016, IEEE International Conference on Cloud Computing Technology and Science, 572-577.
5. B. Dongmyoung and L. Bumchul, *Analysis of telemetering service in OpenStack*, 2015, International Conference on Information and Communication Technology Convergence, 272-274.
6. F. L. Ferraris et al., *Evaluating the Auto Scaling Performance of Flexiscale and Amazon EC2 Clouds*, 2012, International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 423-429.
7. B. I. Santoso, M. R. S. Idrus and I. P. Gunawan, *Designing Network Intrusion and Detection System using signature-based method for protecting OpenStack private cloud*, 2016, International Annual Engineering Seminar, 61-66.
8. K. Kanagala and K. C. Sekaran, *An Approach for Dynamic Scaling of Resources in Enterprise Cloud*, 2013, IEEE International Conference on Cloud Computing Technology and Science, 345-348.

9. F. Al-Haidari, M. Sqalli and K. Salah, *Impact of CPU Utilization Thresholds and Scaling Size on Autoscaling Cloud Resources*, 2013, IEEE International Conference on Cloud Computing Technology and Science, 256-261.
10. N. Roy, A. Dubey and A. Gokhale, *Efficient Autoscaling in the Cloud Using Predictive Models for Workload Forecasting*, 2011, IEEE International Conference on Cloud Computing, 500-507.
11. M. Tighe and M. Bauer, *Integrating cloud application autoscaling with dynamic VM allocation*, 2014, IEEE Network Operations and Management Symposium, 1-9.
12. K. Hwang, Y. Shi and X. Bai, *Scale-Out vs. Scale-Up Techniques for Cloud Performance and Productivity*, 2014, IEEE International Conference on Cloud Computing Technology and Science, 763-768.

# Application of Intensified Current Search to Design Optimal PIDD<sup>2</sup> Controller for BLDC Motor Speed Control with Back EMF Detection

Danupon Kumpanya<sup>1C</sup>, and Chookiat Kiree<sup>2</sup>

<sup>1</sup>*Department of Electrical Engineering, Faculty of Engineering and Architecture, Rajamangala University of Technology Suvarnabhumi (RUS), Suphanburi 72130, Thailand*

<sup>2</sup>*Department of Electrical, Saraburi Technical College, Saraburi 18000, Thailand*

<sup>C</sup> **E-mail:** *kdanupon@yahoo.com*; **Fax:** +66 3 5434 014; **Tel.** +66 8 1850 3236

## ABSTRACT

This paper proposes an optimal PIDD<sup>2</sup> controller design for brushless DC (BLDC) motor speed control by the intensified current search (ICS), one of the newest metaheuristic optimization search techniques for solving the continuous optimization problems. The proposed control system is implemented on the TMS320F28335 DSP board interfacing to MATLAB/SIMULINK. With Back EMF detection, the proposed system is considered as a class of sensorless control. This scheme leads to the speed adjustment of the BLDC motor by PWM. In this work, the BLDC motor of 100 watt is conducted to investigate the control performance. As results, it was found that the speed response of BLDC motor can be regulated at the operating speed of 800 and 1200 rpm in both no load and full load conditions. Very satisfactory responses of the BLDC motor system can be successfully achieved by the proposed control structure and ICS-based design approach.

**Keywords:** BLDC Motor, PIDD<sup>2</sup> Controller, Intensified Current Search, Back EMF Detection.

## 1. INTRODUCTION

Since 1970's, the brushless dc (BLDC) motor has been used in different applications, for example, industrial automation, automotive, aerospace, instrumentation and appliances [1]. The BLDC motor uses permanent magnets for excitation. Rotor position sensors of such the motor are needed to perform electrical commutation. Usually, three Hall Effect sensors are used as rotor position sensors for a BLDC motor. However, the rotor position sensors cause several disadvantages from the standpoint of total system such as cost, size and reliability. For these reasons, it is desired to eliminate these sensors from the motor. This leads the sensorless control. In recent years, sensorless BLDC motor has been studied [2, 3]. The rotor position signal of the motor can be detected by motor voltage and current, etc. Among the methods of rotor position estimation, the Back EMF estimation is one of the major approaches [4]. However, the Back EMF is very small in the process of motor starting, and the commutation point is not easily achieved. For these reasons, the conventional sensorless drive needs a complicated starting procedure [2-4]. Design and implementation of the BLDC motor drive by PIDD<sup>2</sup> (proportional plus integral plus derivative plus double-derivative) controller [5] based on TMS320F28335 DSP board [6] interfacing to MATLAB/SIMULINK [7] is proposed in this paper. In order to obtain the optimal PIDD<sup>2</sup> controller, the intensified current search (ICS) is one of the newest and most efficient methods. The ICS is firstly proposed in 2014 [8] for solving engineering optimization problems. Algorithms of the ICS are inspired and conceptualized from the electric current flowing through electric network. With both diversification and intensification properties, the ICS has been successfully applied to many engineering problems [8, 9]. The ICS is applied to achieve the optimal tracking and regulating responses. The trapezoidal Back EMF waveforms are modeled as a function of rotor position, while the switching function concept is adopted to model the voltage source

inverter (VSI). In this work, the tabu search (TS) [10, 11] is conducted to design the PID controller for comparison in simulation results. By the proposed experimentation, speed and current waveforms of the controlled system can be easily obtained.

## 2. BLDC MOTOR MODEL AND DSP BOARD

### A. Mathematical model of BLDC motor

In practice, the commutation of the brushless DC motor is done electronically. The position of rotor is very important for electrical commutation. Usually, the Hall sensors are placed in 120-degree intervals and the common operation of BLDC motor is achieved by six-sector. The Back EMF and Hall sensor signals are depicted in Figure 1.

Once the magnetic poles of rotor come to Hall sensor, the sensory signals are generated. According to the six-step in Figure 1, the commutation sequence is performed. The motor phases are supposed to conduct for 120 electrical degrees two times per cycle. The two phases are only conducted at one time. The Hall sensor signal has the rising and falling edges for each phase. This means that the six-trigger signals are generated per one cycle. Using these trigger signals, motor control is carried out. The switching sequence for commutation phase is given in Table 1. For forward and reverse directions of rotor, switching sequence is different.

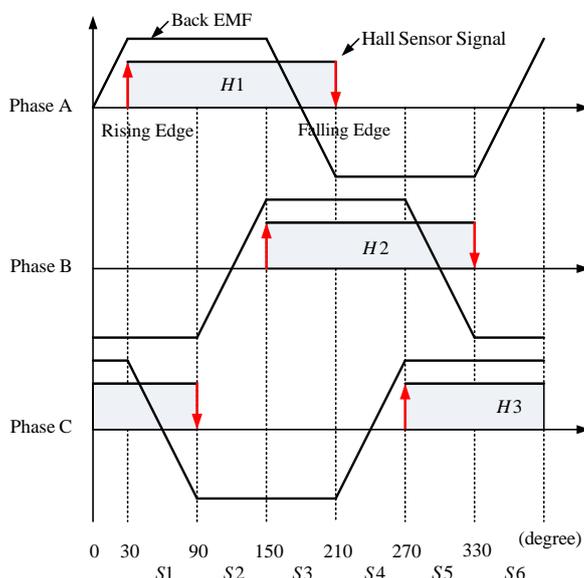


Figure 1. Back EMF and Hall-Effect sensor signals.

Table 1. Switching sequence.

Directions	Switching sectors					
	S1	S2	S3	S4	S5	S6
Forward	$a^+b^-$	$a^+c^-$	$b^+c^-$	$b^+a^-$	$c^+a^-$	$c^+b^-$
Reverse	$b^+a^-$	$c^+a^-$	$c^+b^-$	$a^+b^-$	$a^+c^-$	$b^+c^-$

Modeling of the BLDC motor is similar to three-phase synchronous machine. Since there is permanent magnet mounted on the rotor, some dynamic characteristics are different. Model of the BLDC motor can be formulated through the electrical equivalent circuit represented in Figure 2 [12].

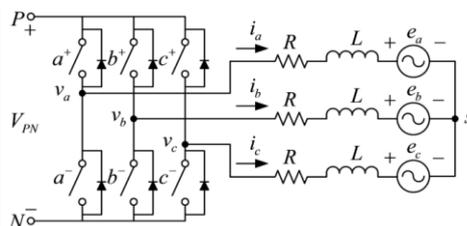


Figure 2. Equivalent circuit of the BLDC motor.

The voltage equation in (1) consists of resistance, inductance and Back EMF voltages for  $a$ ,  $b$  and  $c$  phases, where  $v_x$  is the input voltage,  $i_x$  is the current,  $R$  is the resistance,  $L$  is the inductance and  $e_x$  is the Back EMF, respectively. The subscript  $x$  stands for  $a$ ,  $b$  and  $c$  phases. In upper voltage equation,  $L$  is the difference between self-inductance  $L_s$  and mutual-inductance  $L_m$ . That is,  $L = L_s - L_m$ .

$$v_x = Ri_x + L \frac{di_x}{dt} + e_x \quad (1)$$

$$e_x = K_w f \left( \theta_e - \frac{2n\pi}{3} \right) \omega_m \quad (2)$$

In the 3-phase BLDC motor, the Back EMF  $e_x$  in Eq.(2) is related to a function of rotor position, where  $K_w$  is Back EMF constant of one phase,  $\theta_e$  is the electrical rotor angle and  $\omega_m$  is the rotor angular velocity. The subscript  $x$  presents  $a$ ,  $b$  and  $c$  phases. If  $x$  is  $a$ ,  $b$  and  $c$  phase,  $n$  is 0, 1 and 2, respectively.

Unlike the induction motor or synchronous motor control, phase variables are used directly without any transformation equation in BLDC motor. The physical properties as current, flux and Back EMF perform rectangular forms. This implied that the coordinate transformation is not needed. The output power  $P_e$  and the torque  $T_e$  of the BLDC motor are expressed in Eq.(3) and Eq.(4), respectively.

$$P_e = e_a i_a + e_b i_b + e_c i_c \quad (3)$$

$$T_e = \frac{P_e}{\omega_m} = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega_m} \quad (4)$$

If a motor is operated with two-phase conduction type, the torque equation is modified. In S1 mode as shown in Table 1, the current ( $i_a = I$ ,  $i_b = -I$  and  $i_c = 0$ ) and Back EMF ( $e_a = E$  and  $e_b = -E$ ). The torque  $T_e$  of the BLDC motor can be expressed in Eq.(5).

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega_m} = \frac{2EI}{\omega_m} \quad (5)$$

The total output torque is a simple equation of Back EMF  $E$  and current  $I$ . Regarding to Eq.(5) for producing an electromagnetic torque, the sum of  $e_a i_a$ ,  $e_b i_b$  and  $e_c i_c$  is constant as far as a certain speed is concerned. As shown in Figure 2, this means that the rectangular phase currents with the corresponding Back EMF are required [4].

### B. DSP controller board

The DSP controller board used in this work is the Texas Instrument TMS320F28335 [6] consisting of a 32-bit CPU and a single-precision 32-bit floating-point. The 150 MHz system clock is provided by an on-chip oscillator including the MATLAB/SIMULINK software platform [7]. The module board includes 3-phase PWM inverter 90V, 5A upto 25 kHz PWM frequency and measurement of motor currents in all 3 phases with DC voltage supply of 5Vdc.

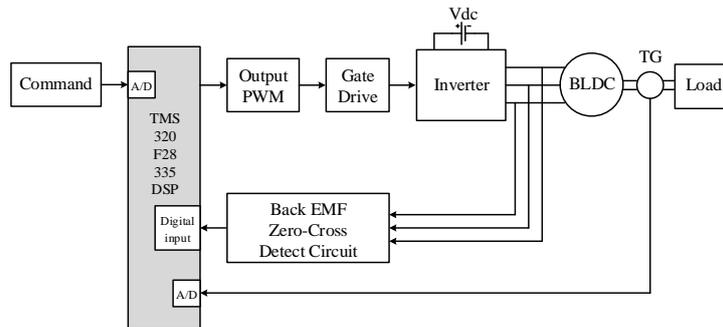


Figure 3. DSP-based BLDC motor speed control system.

The BLDC motor speed control system based on TMS320F28335 DSP board can be represented in Figure 3. The MATLAB/SIMULINK platform is integrated to monitor communication program and DSP applications. Using the PC communication module, it can be accessible to the TMS320F28335 DSP controller board and execute any application program in order to analyze and evaluate the overall system performance.

### 3. ICS ALGORITHM AND PIDD<sup>2</sup> DESIGN PROBLEM

#### A. ICS Algorithms

The intensified current search (ICS) is one of the trajectory-based metaheuristic optimization techniques based on the stochastic approach and multiple local search [8]. It is the improved version of the conventional current search (CS). The ICS possesses the memory list (ML) to escape from entrapment, the adaptive radius (AR) and the adaptive neighborhood (AN) to speed up the search process. Algorithms of the ICS can be represented by the pseudo code as shown in Figure 4. Some movements of the ICS over 3-D search space can be represented as visualized in Figure 5.

---

```

Initialized:
- Objective function  $f(x)$ ,  $x=(x_1, \dots, x_d)^T$ ,
- Initialize the search space  $\Omega$ , maximum allowance of solution cycling  $j_{max}$ ,
  number of initial solutions  $N$ , number of neighborhood members  $n$ ,
  search radius  $R$ ,  $k=j=1$ .
- Random initial solution  $X_j$  within  $\Omega$ 
- Evaluate  $f(X_j)$  then rank and  $X_j$  into  $\Psi$ . Let  $x_0=X_k$  as selected initial solution.
-  $X_{global}=X_{local}=x_0$ .
while ( $k \leq N$  or termination criteria: TC);
  while ( $j \leq j_{max}$ );
    Random neighborhood  $x_j$  around  $x_0$  within  $R$ .
    Evaluate  $f(x_j)$  and set the best one as  $x^*$ .
    if  $f(x^*) < f(x_0)$ ; Update  $x_0=x^*$  and set  $j=1$ .
    else Update  $j=j+1$ . end
    Activate AR by  $R=\rho R$ ,  $0 < \rho < 1$ .
    Invoke AN by  $n=\alpha n$ ,  $\alpha > 1$ .
  end Update  $X_{local}=x_0$ .
  if  $f(X_{local}) < f(X_{global})$ ; Update  $X_{global}=X_{local}$ . end
  Update  $k=k+1$  and set  $j=1$ . Let  $x_0=X_k$  as selected initial solution.
end
  
```

---

Figure 4. Pseudo code of ICS.

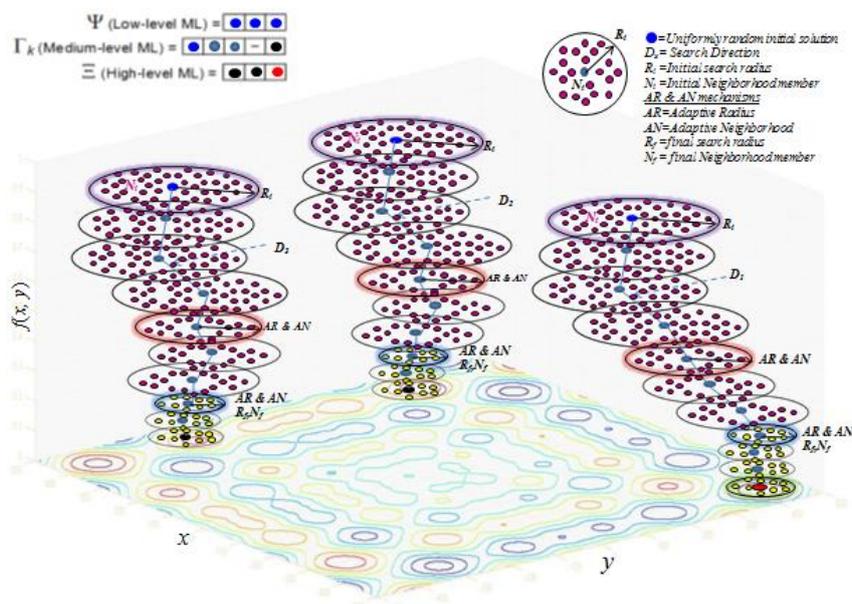
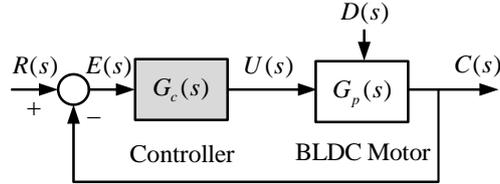


Figure 5. Some movements of the ICS.

### B. ICS-based PIDD<sup>2</sup> design problem

According to the classical control context, a conventional control loop is represented by the block diagram in Figure 6. The PIDD<sup>2</sup> controller receives the error signal,  $E(s)$ , and produces the control signal,  $U(s)$ , to control and regulate the output response,  $C(s)$ , referred to the reference input,  $R(s)$ , while  $D(s)$  is an external disturbance signal,  $G_p(s)$  and  $G_c(s)$  are the plant and the controller transfer functions. The transfer function of the PIDD<sup>2</sup> controller is expressed in Eq.(6), where  $K_P$ ,  $K_I$ ,  $K_D$ , and  $K_{DD}$  are the proportional, integral, derivative, and double-derivative gains, respectively, while, the closed loop transfer function with PIDD<sup>2</sup> controller is given in Eq.(7) with the following parameters:  $K_P$ ,  $K_I$ ,  $K_D$ , and  $K_{DD}$  respectively.

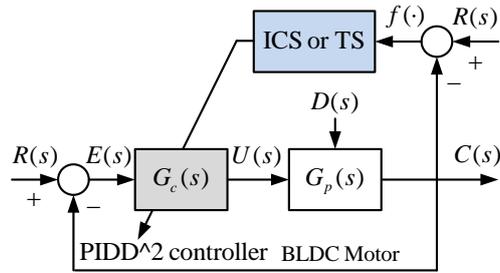


**Figure 6.** Conventional control loop.

$$G_c(s) = K_P + \frac{K_I}{s} + K_D s + K_{DD} s^2 \quad (6)$$

$$\left. \frac{C(s)}{R(s)} \right|_{\text{PIDD}^2} = \frac{G_c(s)G_p(s)}{1 + G_c(s)G_p(s)} \quad (7)$$

In this work, the ICS is applied to design an optimal PIDD<sup>2</sup> controller in order to gain the optimal response of the BLDC system. The ICS-based PIDD<sup>2</sup> controller design for BLDC system can be represented by the block diagram in Figure 7, where  $C(s)$  and  $R(s)$  stand for actual and desired responses. The sum-squared error (SSE) between  $R(s)$  and  $C(s)$  set as the objective function if  $f(\cdot)$  stated in Eq.(8) will be fed back to the ICS block to be minimized to obtain an appropriate parameters, i.e.  $K_P$ ,  $K_I$ ,  $K_D$ , and  $K_{DD}$  of PIDD<sup>2</sup> controller giving a satisfactory response.



**Figure 7.** ICS-based PIDD<sup>2</sup> controller design.

$$\text{Minimize} \quad f(\cdot) = \sum_{i=1}^N [r(i) - c(i)]^2 \quad (8)$$

## 4. RESULTS AND DISCUSSIONS

### A. BLDC model identification

The plant  $G_p(s)$  of the BLDC model can be identified via MATLAB/SIMULINK and ICS technique [13]. The third-order transfer function model of BLDC obtained by this approach can be expressed in Eq.(9).

$$G_p(s) = \frac{0.04421}{5.423 \times 10^{-9} s^3 + 5.936 \times 10^{-6} s^2 + 0.001618s + 0.03683} \quad (9)$$

## B. Simulation results

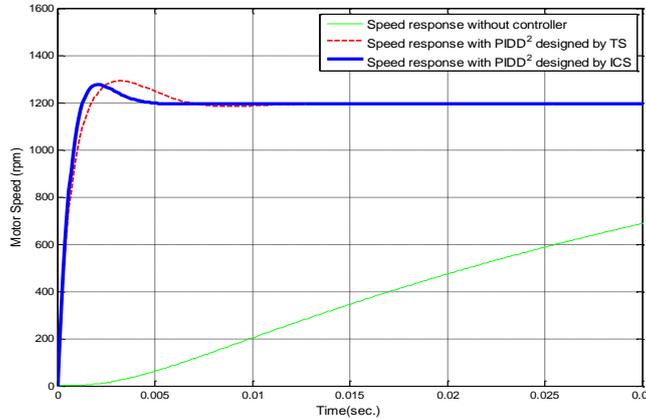
Regarding to Figure 7, the ICS is applied to design the PIDD<sup>2</sup> controller for the BLDC system. The algorithms of TS and ICS were coded by MATLAB run on Intel(R) Core(TM) i5-2430M, 2.4 GHz, 8 GB-RAM computer. The search parameters of two candidate metaheuristic techniques are priori set for a fair comparison. For the TS, the number of neighborhood member  $n = 20$ , the search radius  $R = 20$  and maximum allowance of solution cycling  $j_{max} = 10$ . For AR mechanism; (i) if  $f(\cdot) < 100$ , then  $R = 10$ , (ii) if  $f(\cdot) < 75$ , then  $R = 5$ , (iii) if  $f(\cdot) < 50$ , then  $R = 2$ . For the ICS, the number of initial solutions (search directions)  $N = 3$ , the number of neighborhood member  $n = 10$ , the search radius  $R = 1.0$ , maximum allowance of solution cycling  $j_{max} = 10$  and the maximum iteration of each search direction = 200. For AR and AN mechanisms; (i) if  $f(\cdot) < 0.5$ , then  $R = 0.25$  and  $n = 15$ , (ii) if  $f(\cdot) < 0.1$ , then  $R = 0.1$  and  $n = 20$ , (iii) if  $f(\cdot) < 0.01$ , then  $R = 0.01$  and  $n = 25$ . For both TS and ICS, the search will be terminated once the  $f(\cdot) \leq 0.0010$  set as the termination criteria (TC). The maximum boundaries of the model parameters are preset to perform their corresponding search spaces as stated in Eq.(10), respectively.

$$K_P \in [0, 150], K_I \in [0, 1000], K_D \in [0, 0.5], K_{DD} \in [0, 0.0005] \quad (10)$$

The proposed design performs the search of 20 trials with different random initial solutions in order to obtain the best solution. To find the optimal  $K_P$ ,  $K_I$ ,  $K_D$ , and  $K_{DD}$ , TS and ICS are applied under the boundary of equation in (10). After searching with TS and ICS, the optimal parameters of PIDD<sup>2</sup> controller for the BLDC are demonstrated in Table 2.

**Table 2.** Optimal parameters PIDD<sup>2</sup> controller for the BLDC motor.

Methods	Obtained PIDD <sup>2</sup> parameters				$f(\cdot)$ (Obj.Func.)
	$K_P$	$K_I$	$K_D$	$K_{DD}$	
TS	111.4147	800.8196	0.2148	0.00018	1.0751
ICS	108.4159	800.8245	0.3176	0.00020	1.0696

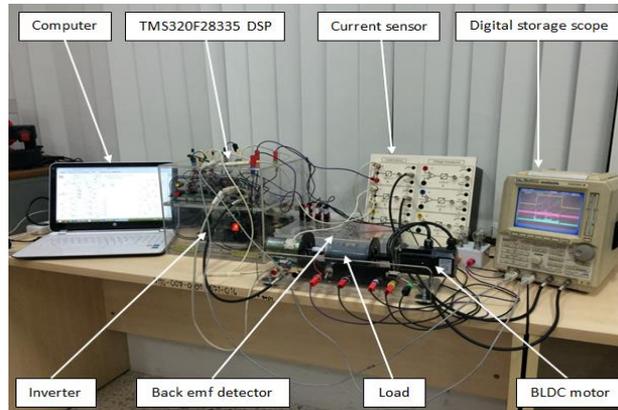


**Figure 8.** System responses of BLDC with PIDD<sup>2</sup> controller.

According to Figure 8, there are two search methods of BLDC with the PIDD<sup>2</sup> controller, namely, TS and ICS. In TS method, the system provides rise time ( $t_r$ ) of 0.00168 sec., settling time ( $t_s$ ) of 0.0056 sec., maximum percentage of overshoot ( $M_p$ ), 9.33%, steady-state error ( $e_{ss}$ ) of 0.00% and search time of 71.70 sec.. For ICS method, the system yields are as  $t_r = 0.00125$  sec.,  $t_s = 0.0036$  sec.,  $M_p = 6.85\%$ ,  $e_{ss} = 0.00\%$  and search time = 48.84 sec., respectively. ICS can provide an optimal PIDD<sup>2</sup> controller for the BLDC motor speed control system superior to the TS does. The speed response without controller is an unstable response. Consequently, the speed response with PIDD<sup>2</sup> controller by TS and ICS have stable response.

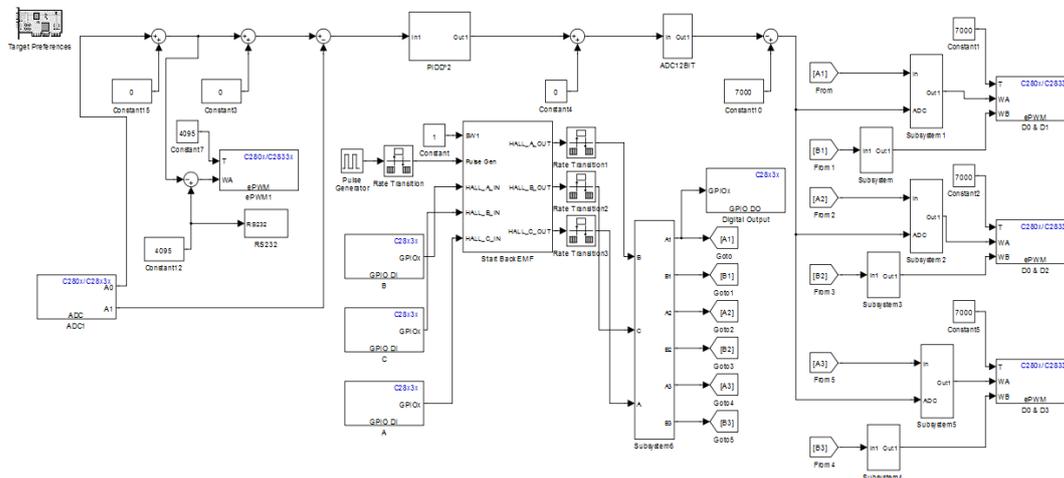
### C. Experimental Results

The BLDC motor speed control system is implemented by using the 100 watt, 1600 rpm BLDC motor as shown in Figure 9. The speed of motor ranging from 0 to 1600 rpm can be measured by tacho-generator of 0.825V/500 rpm. A speed transformed to be a voltage ranging from 0 to 3.3 V will be sent to A/D convertor. This scheme enables the user can adjust the speed of the motor by the power amplifier. The performance of implemented BLDC motor based on TMS320F28335 DSP board with the MATLAB/SIMULINK can be tested by sampling rate of 0.0001 sec.



**Figure 9.** Experimental setup.

The measurements can be divided into two groups, the first is a step change of the speed reference at constant load torque and the second is a step change of the load torque at constant speed reference. Speed responses of the system via the step change from 800 to 1200 rpm of speed reference without and with rated load torque are investigated. The Simulink block diagram of control system is represented in Figure 10. The closed-loop responses of the system at 800 rpm and 1200 rpm can be simulated by such the simulation block diagram.



**Figure 10.** Simulink block diagram of control system.

As results shown in Figure 11 – Figure 12, it was found that the speed responses of the BLDC system controlled by the PIDD<sup>2</sup> controller designed by ICS from 800 to 1200 rpm both no load and full load conditions are very satisfactory with fast response and  $e_{ss} = 0.00\%$ . Supply phase voltage, phase current and DC current waveforms of BLDC at 800 rpm and 1200 rpm are measured by digital storage scope YOKOGAWA DL1540C in which the ratio of a current sensor is 2.424V/A. Referring to Figure 11 – Figure 12, the entire system performance evaluated in this PIDD<sup>2</sup> controller designed by ICS is summarized in Table 3.

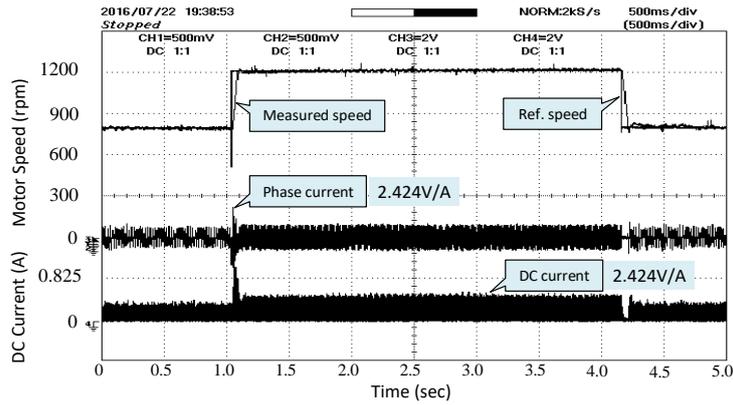


Figure 11. Step change from 800 to 1200 rpm (no load).

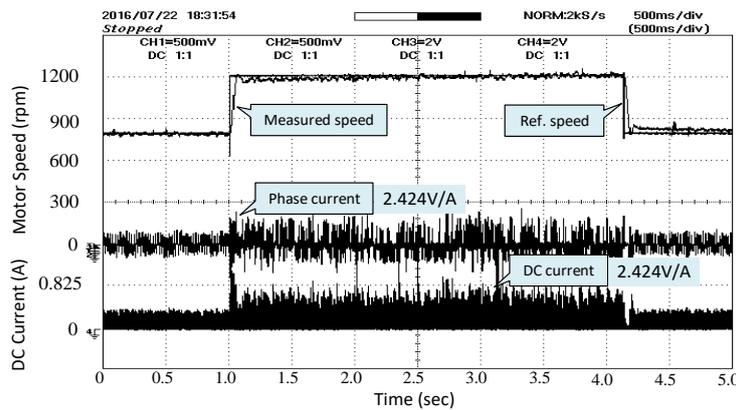


Figure 12. Step change from 800 to 1200 rpm (full load).

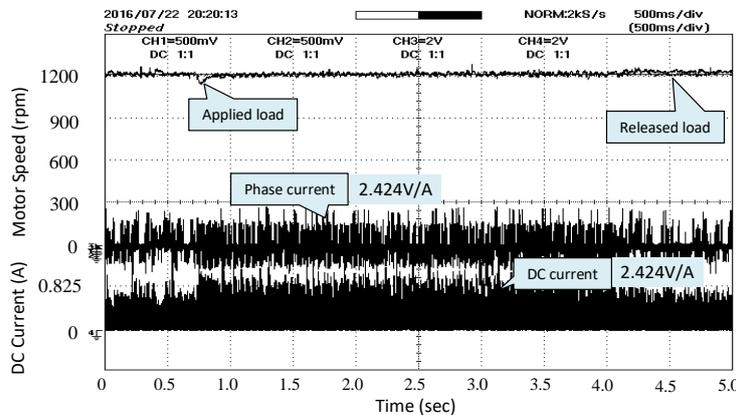


Figure 13. Load regulation at constant speed of 1200 rpm.

Table 3. Entire System Performance.

Step Changes	Step Responses			
	$t_r$ (sec.)	$M_p$ (%)	$t_s$ (sec.)	$e_{ss}$ (%)
No load	0.0491	1.37	0.0592	0.00
Full load	0.0595	0.82	0.0881	0.00

Referring to Table 3, it was found that at the speed from 800 rpm to 1200 rpm the system response reaches to desired steady state level. In case of no load, the step response declares values of  $t_r$ ,  $M_p$ ,  $t_s$  and  $e_{ss}$  almost as same as the those values obtained from simulation as can be observed in Figure 8. However,  $t_r$  and  $t_s$  of the full load is a little bit slower than that of no load condition. Also,  $M_p$  of the full load is lower than that of no load condition. At both

conditions, the BLDC system responses have no steady state error. Responses of load rejection by PIDD<sup>2</sup> controller at 1200 rpm can be observed in Figure 13, respectively. It was found that the fast regulation responses once load torque is occurred can be successfully achieved by the proposed scheme. This can be concluded that the satisfactory responses of the BLDC system can be obtained by the PIDD<sup>2</sup> controller designed by the ICS and a good performance can be successfully achieved by the proposed approach.

## 5. CONCLUSION

Designing an optimal PIDD<sup>2</sup> controller for brushless DC (BLDC) motor speed control based on the intensified current search (ICS) has been proposed in this paper. Considered as sensorless control by using back EMF detection based on TMS320F28335 DSP board, the proposed control structure has been interfaced to MATLAB/SIMULINK for adjusting the BLDC motor speed via PWM. The ICS has been applied to design the PIDD<sup>2</sup> controller to obtain the optimal tracking and regulating responses. Performance of the proposed system has been evaluated against the BLDC motor of 100 watt. As simulation results comparing with the TS, it was found that with the shortest search time consumed the ICS could provide the optimal PIDD<sup>2</sup> controller giving the fastest response with shortest rise and settling times as well as smallest overshoot. As experimental results, it was found that the satisfactory tracking and regulating speed response of BLDC motor can be achieved from 800 and 1200 rpm with no load and full load conditions. Experimental results strongly agree to simulation ones. This can be concluded that the proposed design approach and control structure are most efficient and alternative to control the BLDC motor effectively.

## REFERENCES

1. Yedamale, P., *Brushless DC (BLDC) Motor Fundamentals*: AN885 Microchip Technology Inc., 2003.
2. Patel, V. K. S. and Pandey, A. K., Modeling and simulation of brushless DC motor using PWM control Technique, *IJERA*, vol.3(3), 2013, pp.612–620.
3. Alaeinovin, P. and Jatskevich, J., Hall-sensor signals filtering for improved operation of brushless DC motors, *IEEE International Symposium on Industrial Electronics (ISIE)*, 2011, 613–618.
4. Yun, S. Y., Lee, H. J., Han, J. H. and Lee, J., Position control of low cost brushless DC motor using hall sensor, *The 6<sup>th</sup> International Conference on Electromagnetic Field Problems and Applications (ICEF)*, 2012, 1–4.
5. Lee, K.C., Jeong, Y.H., Koo, D. H. and Ahn, H. J., Development of a radial active magnetic bearing for high speed turbo-machinery motors, *SICE-ICASE, International Joint Conference 2006*, 18–21 Oct. 2006, 1543–1548.
6. *Texas Instruments*, TMS320F28335, Digital Signal Controller, 2007.
7. *MATLAB/SIMULINK User's Guide*, The Math Works Inc., Natick, MA, 1998.
8. Nawikavatan, A., Tunyasrirut, S. and Puangdownreong, D., Application of intensified current search to optimum PID controller design in AVR system, *AsiaSim-2014*, 2014, 255–66.
9. Nawikavatan, A., Tunyasrirut, S., and Puangdownreong, D., Optimal PID controller design for three-phase induction motor speed control by intensified current search, *The 19<sup>th</sup> International Annual Symposium on Computational Science and Engineering (ANSCSE19)*, 2015, 104–109.
10. Glover, F., Tabu search - part I, *ORSA Journal on Computing*, 1989, 1(3), 190–206.
11. Glover, F., Tabu search - part II, *ORSA Journal on Computing*, 1990, 2(1), 4–32.
12. Krishnan, R., *Electric Motor Drives Modeling, Analysis, and Control*, Prentice-Hall International Inc., New Jersey, 2001.
13. Kumpanya, D., Thaiparnat, S., Puangdownreong, D., Parameter Identification of BLDC Motor Model via Metaheuristic Optimization Techniques, *Proceeding of Industrial Engineering and Service Science, (IESS2015)*, Indonesia, 2015, 326–331.

# A Simple Numerical Model of an Active Magnetic Regenerator in Operation

**Pairat Tulyaprawat<sup>1,2</sup>, Ratchatee Techapiesancharoenkij<sup>1,2,3</sup>, Yuranan Hanlumyung<sup>1,2,3,C</sup>**

<sup>1</sup>*Department of Materials Engineering, Faculty of Engineering, Kasetsart University, Bangkok, 10900, Thailand*

<sup>2</sup>*Thailand Center of Excellence in Physics, Chiang Mai University, Chiang Mai, 50202, Thailand*

<sup>3</sup>*Materials Innovation Center, Faculty of Engineering, Kasetsart University, Bangkok, 10900, Thailand*

<sup>C</sup>*E-mail: yuranan.h@ku.th Fax: +66 2 955 1811; Tel. +66 2 7970999 ext. 2119*

## ABSTRACT

The rising energy consumption has become one of the most important problems in modern world. It is known that more than 50% the energy is spent in cooling systems, both in industry and household. Therefore, there is a need, now more than ever, to develop a better energy-saving refrigerant alternative. Recently, cooling-system engineers and scientists have spurred into investigating magnetic refrigerators because of their efficiency; 40-60% improvement over those of current vapor-compression refrigerators. However, there is still a gap of knowledge on the interplay among magnetic work, thermal work, and fluid flow inside a magnetic refrigerator. Our work is aimed to develop a simple mathematical model representing the magnetic regenerator in operation. The mathematical model does not require an experimental magnetization curve as an input, but rely on a principle derived from the mean-field theory. The developed numerical model allows experimentalists and engineers to optimize the cooling-system design, and predict the operating temperature span.

**Keywords:** Numerical Model, Magnetocaloric Effect, Active Magnetic Regenerator

## 1. INTRODUCTION

Conventional refrigerative system, based on vapor compression/expansion systems, has efficiency around 40% - 60% of Carnot cycle [1]. Furthermore, it is susceptible to leakage of chlorofluorocarbon (CFC) which is an ozone-depleting gas. On the contrary, magnetic refrigeration is more efficient and environmentally benign. Presently there have been a number of reports stating a realistic potential of magnetic refrigeration entering larger consumer markets, where this novel refrigeration system should find applications both in households and mainstream industries [2][3].

The working of a magnetic refrigerator (MR) is based on the thermodynamics principles of a magnetocaloric (MC) material whose temperature can be altered by applying or removal of external magnetic fields [4]. An MC material can replace a typical cooling mechanism which is based on liquid-to-gas condensation. The MC material in the MR is mostly configured in the form of a magnetic regenerator (AMR) whose name derived from its role as a heat regenerator. The AMR is therefore the heart of the magnetic refrigerator. The AMR may consist of a packed MC particles where the porosity between particles serves as the main heat-exchanging surface between the MC material and the flowing fluid [5]. The fluid is the major heat convection medium within a cycle of the MR. The heat regeneration at the AMR leads to a temperature span sufficient for utilizing as a cooling system. Too low temperature renders the magnetic refrigerators impractical [6].

A major contribution to the change in temperature of the MR is the magnetization/demagnetization behaviors of the MC material making up the AMR. It is therefore essential to understand magnetization behaviors of the AMR. Presently a magnetization curve of a magnetic material is usually obtained using a vibrating sample magneto meter (VSM). However, due to limiting resources obtaining a magnetization curve could be a daunting task. This work is aimed at creating a simplified mathematical model which does not require an exact magnetization curve, however is capable to capture major operations of a magnetic refrigerator. Though our model is

oversimplified, it will be shown that it is able to find maximum temperature of gadolinium plate in AMR. Our model gives results which are comparable to data obtained from an experimental magnetization curves. By incorporating the equations of fluid-heat convection to our model of magnetization/demagnetization, the evolution of the AMR into a steady state in a thermodynamic cycle can be predicted. The model also contains a number of necessary physical parameters all of which affect the temperature span of the system.

## 2.THEORY AND RELATED WORK

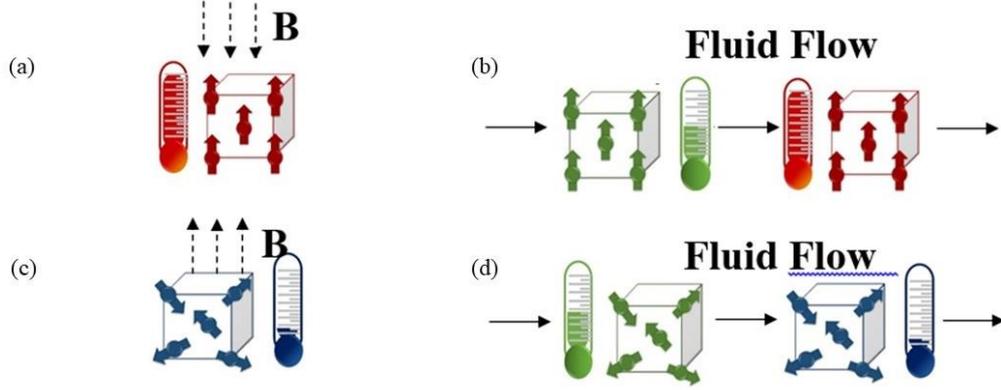


Figure1: Four process of magnetic refrigeration cycle; (a) adiabatic magnetization, (b) heat transfer from solid to liquid, (c) adiabatic demagnetization, and (d) heat transfer from liquid to solid

In order to understand magnetocaloric effect, one needs to consider entropy change within four important processes as shown in Figure 1. The first process involves an adiabatic magnetization (Figure 1(a)). The magnetic field is applied to the magnetocaloric material without any heat transfer from/to the outside. The applied magnetic field causes a decrease in the magnetic entropy of the MC material which in turn increases its thermal entropy, and hence temperature. Here, the total entropy is a sum of two contributions: (a) the magnetic entropy  $S_m$  and (b) the thermal entropy  $S_t$ , i.e. [7]

$$S_{\text{total}} = S_m + S_t. \quad (1)$$

Since there is no heat input/output in this process the total entropy remains constant. As a result, an adiabatic magnetization causes  $S_t$  to increase resulting in higher temperature. In the second process shown in Figure 1(b), a fluid is passed through the MC material leading to a heat flow from the solid MC materials to the fluid, and consequently the reduction of the temperature of the solid. In the third process shown in Figure 1(c), an adiabatic demagnetization enhances the magnetic entropy of the MC material, and consequently reducing its thermal entropy and overall temperature. Lastly (Figure 1(d)), the fluid is allowed to pass through the MC material again, but this time it causes a heat transfer from the fluid to MC material, leading to an increase in temperature of the MC material. The adiabatic magnetization/demagnetization is key to the thermal cycles leading to a temperature span while the AMR is in operation. According to principles of thermodynamics and Maxwell relations, the rate of change in temperature  $dT/dt$  of the MC material under an applied field  $H$  is proportional to the product of  $(\partial M/\partial T)_H$  and  $dH/dt$  as [8]

$$\frac{dT}{dt} = -\frac{T}{C_H} \mu_0 V \left( \frac{\partial M}{\partial T} \right)_H \frac{dH}{dt}, \quad (2)$$

where  $T$  is the temperature,  $H$  is the applied magnetic field,  $C_H$  is the working AMR's specific heat capacity at constant field,  $M$  is the internal magnetization of the AMR,  $t$  is the time,  $V$  is specific volume, and  $\mu_0$  is permeability of free space. During a thermodynamic cycle, the sinusoidal profile of the applied magnetic field with time determines the rate of change of the applied field  $dH/dt$ . The change in internal magnetization as a function of temperature at a constant applied field  $(\partial M/\partial T)_H$  can be derived from a magnetization curve obtained from a constant-field environment. An example magnetization curve of gadolinium at 1 Tesla is shown in Figure 2. The curve is derived from fitting experimental data in [9].

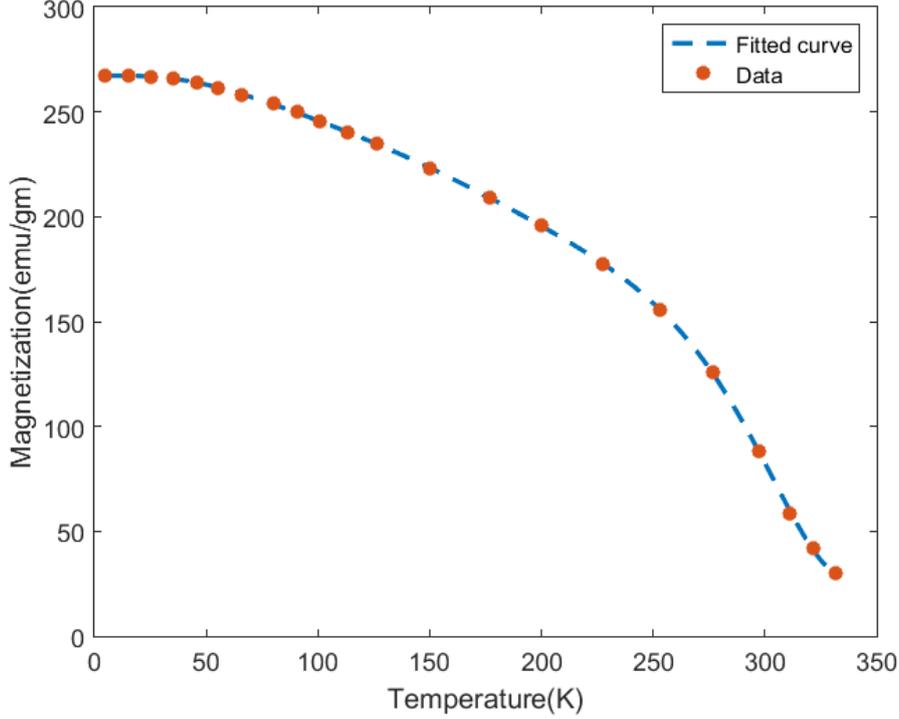


Figure2: Experiment data of magnetization of gadolinium at filed constant [9].

Setting up an experiment to obtain an accurate magnetization curve may pose some challenges such as the cost of equipment, the cost of the samples, and sample preparations, among others. Here we wish to estimate the temperature changes without having to resource to experimental data, but instead using theoretical results from the mean field theory (MFT) which states that [8]

$$\Delta T \approx -\frac{T\Delta S_m}{c_p}, \quad (3)$$

where  $\Delta S_m$  is the change in magnetic entropy while the MC material is driven by an external applied field, and the temperature is risen/lowered from room temperature to a targeted level. To the best of our knowledge, there has been no report on using this MFT result in predicting the temperature span of the AMR in operation. Our simplified physical approximation is that the MC effects only result in the adiabatic temperature  $T_{ad}$  during the adiabatic magnetization and demagnetization processes (as shown in Figure 1(a) and 1(c)) as

$$T_{ad} = T \pm \Delta T, \quad (4)$$

where  $\Delta T$  is the results of MFT in Eq. (3). The values of  $\Delta T$  for gadolinium are usually tabulated in a table such as those in [8]. Magnitude of adiabatic temperature change of gadolinium about 2.3 K. It should be noted that the MC material exhibits good MC properties while it is in the state of

paramagnetism [10]. Curie temperature is a transition temperature between ferromagnetic (lower than Curie temperature) and paramagnetic (higher than Curie temperature) phases. To achieve maximum  $\Delta T$ , it is required that MC material is in the paramagnetic phase at room temperature. In other words, the MC materials should possess a low Curie temperature. Due to its low Curie temperature, gadolinium is therefore widely accepted as a major material in the MR research.

In an AMR cycle, the flow of fluid is key to heat convection mechanism. The fluid carries heat emanating from MC material and promotes heat regenerator in the AMR. In this work, the fluid flow and interaction with the MC material are modeled using a set of partial differential equations representing the heat transfer between the fluid and solid. The governing equations for the fluid and solid are [11]:

$$\frac{\partial T_f}{\partial t} + v_f \frac{\partial T_f}{\partial x} + \frac{\alpha A}{\rho_f c_{pf} V_f} (T_s - T_f) = 0 \quad (5)$$

$$\frac{\partial T_s}{\partial t} + \frac{\alpha A}{\rho_s c_{ps} V_s} (T_f - T_s) = 0, \quad (6)$$

where  $v$ ,  $x$ ,  $\rho$ ,  $A$ ,  $V$  and  $c_p$  are velocity, axial position, density, cross-sectional area between the fluid and the solid, volume, and specific heat at constant pressure respectively. The subscripts  $f$  and  $s$  represent fluid and solid refrigerants respectively. The heat transfer and fluid properties depend on the geometric shape of the AMR bed. The geometric effect is reflected by the parameter  $\alpha$  which is given by [12]

$$\alpha = \frac{Nu k_f}{D_h}, \quad (7)$$

where  $Nu$ ,  $D_h$ , and  $k_f$  are the convection heat transfer coefficient, hydraulic diameter, and thermal diffusivity respectively. The boundary conditions of the system investigated are

$$T_f(0, t) = T_{\text{hot}}^{\text{in}} \quad (8)$$

$$T_f(x = L, t) = T_{\text{cold}}^{\text{in}}, \quad (9)$$

where  $T_{\text{hot}}^{\text{in}}$  and  $T_{\text{cold}}^{\text{in}}$  represent temperature of hot inlet fluid and temperature of cold inlet fluid respectively. The initial conditions are temperatures of the inlet fluid and the solid at any value of  $x$ 's and  $t = 0$  are room temperature. That is

$$T_f(x, t = 0) = T_{\text{room}} \quad (10)$$

$$T_s(x, t = 0) = T_{\text{room}}, \quad (11)$$

Solving equation (5) and (6) subjected to the initial conditions (10) and (11), and boundary conditions (8) and (9) will give a temperature profile of the working AMR.

### 3. COMPUTATIONAL DETAILS

To calculate the temperature profile of the AMR in operation, the computational steps start form using the MFT to calculate the temperature due to adiabatic magnetization of the MC material. Then, the set of equations from (5)-(11) are then solved to account for fluid flow through the AMR bed. In the next step, the temperature change due to adiabatic demagnetization is calculated using results from the MFT. Lastly, the set of equations involving fluid flow and heat exchanging between the fluid and AMR is solved. All calculations are cycled so that the temperature profile of the AMR bed and temperature spans are quantitatively achieved. The logical workflow of all calculation steps is summarized in Figure 3.

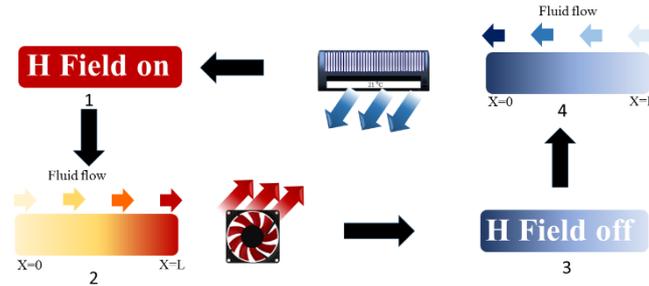


Figure 3: Heat transfer model of a fluid flow through a solid at various positions.

Finite difference method is utilized to solve the set of governing equations described. The developed numerical scheme contains inputs of all four processes in an AMR cycle. After adjusting the temperature due to the adiabatic magnetization, the governing equations of fluid flow in Eq. (5) and (6) are solved using the forward difference technique. On the other hand, following adiabatic demagnetization the backward difference technique is employed. In every thermodynamic cycle, the computer will calculate the temperature level as a function of time and spatial positions. Figure 4 shows the flow chart detailing the numerical workflow, particularly the numerical loops and inputs to each step. In this work, the time step is set to 0.001 second, while the spatial interval is set to 0.01 m.

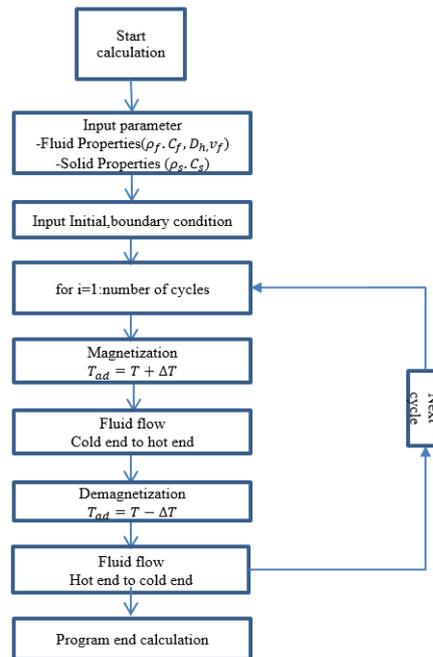


Figure4: Flow chart of the numerical model.

#### 4. RESULTS

Solving the above set of equations while incorporating the temperature change due to the MFT yields temperature as a function of the time evolution. Four initial and transient time-temperature cycles of the AMR test bed at the cold end are shown in Figure 5. The calculated profile agrees with intuitions based on the aforementioned principles of the thermodynamics processes of the AMR. On the cold end, the temperature of the AMR reduces from room temperature. It is then excited to a higher level by an adiabatic magnetization process. The inlet of cold fluid in turn reduces the temperature of the hot AMR, and finally the AMR is adiabatically demagnetized. The processes are then repeated. In overall operation, the temperature of the cold end reduces to lower than room temperature.

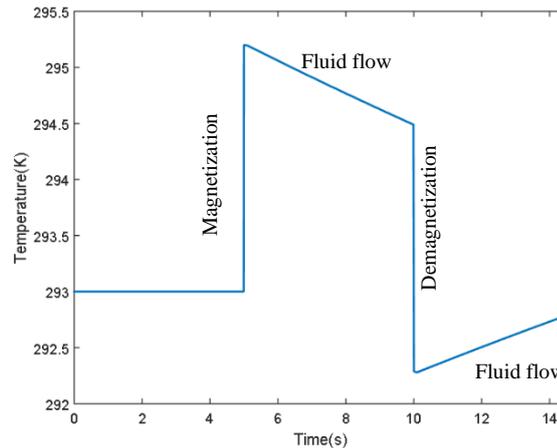


Figure 5: Calculated temperature versus time at the cold end of the AMR test bed. The curve contains initial four sequential cycles.

Similar to Figure 5, Figure 6 shows temperature-time dependence of the AMR but with the number of cycles increased to 25 cycles. Temperature profiles of both cold and hot ends are illustrated. The curves represent the temperature span of approximately 10 K. Our calculated results are also plotted in comparison to the mathematical model using a full magnetization curve [9]. It is evident that our model gives almost identical results to the other model. The differences in the maximum and minimum temperatures of the two models are only within the range of 0.2 K. Therefore, our simplified model is capable of capturing key features of the temperature profile and should be advantageous in aiding the design of the MR. The size, density of fluid, density of solid, heat capacity of solid, heat capacity of fluid, specific volume are  $1.5 \times 500 \times 50 \text{ mm}^3$ ,  $1000 \text{ kg/m}^3$ ,  $7900 \text{ kg/m}^3$ ,  $240 \text{ J/(kg} \times \text{K)}$ ,  $4185.5 \text{ J/(kg} \times \text{K)}$ , and  $0.126 \text{ g/cm}^3$  respectively.

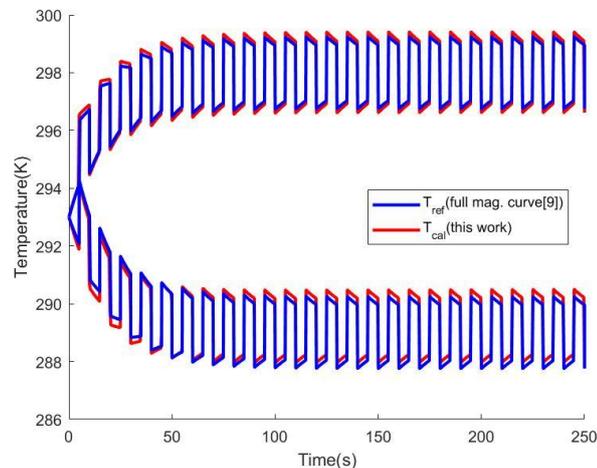


Figure 6: Comparison of time-temperature dependences of two AMR models

Regarding the effects of size of the AMR bed to the over temperature profile, Figure 7(a) is the time-temperature dependence of the AMR of size 0.15(thickness) x 50(length) x 5 mm<sup>3</sup>(width), while Figure 7(b) represents the AMR of size 1.5(thickness) x 500(length) x 50 mm<sup>3</sup>(width). Figure 7 shows the influences of the geometrical parameters on the time-temperature dependence. In the case of the smaller AMR, it takes less time to reach a steady state due the size effects on efficiency of heat exchanging between the solid and fluid.

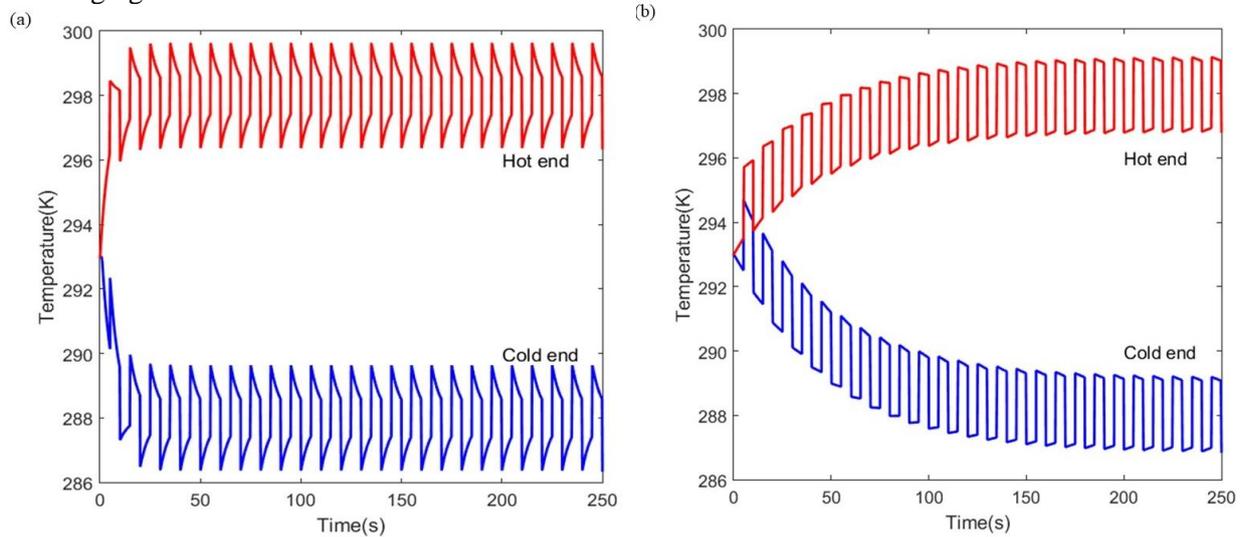


Figure 7: Computed results of the temperature profiles at both ends of magnetic regenerative refrigeration test bed given that the sizes of the AMR of both panels are different.

By examining Eq. (3)-(11), it can be seen that our numerical model is quite flexible in term of the choice of parameters such as those influenced by shapes, flow properties, and cycle frequency. The model can be applied to an optimization problem involving the operation of the MR. Given that the shape parameter is fixed, the time-temperature profile can be easily obtained using this model. Furthermore, the effects of adjusted parameters suitable for design purposes can be straightforwardly estimated. It is also interesting to note that our numerical model can be developed into the two- or three-dimensional versions because of the simple set of equations involved.

## 5. CONCLUSION

Active magnetic refrigeration is a new technology to the further improvement of energy efficiency. This article presents a one-dimensional model of a gadolinium plate in magnetic refrigerator. Thermal losses and heat exchange between the magnetocaloric material and the fluid are represented by simple partial differential equations. A simple approximation based on a mean-field theory is made so that the developed numerical routine can capture the adiabatic magnetization/demagnetization while bypassing the complex magnetization curves. The calculated results agree well with results utilizing full magnetization curve at a constant applied field. Furthermore, due to simple equation involved, our model has potential to be developed into the two- and three-dimensional versions. It is also capable of aiding the optimization of a realistic magnetocaloric material.

## REFERENCES

1. Franco, V., Blázquez, J. S., Ingale, B., & Conde, A., *Annual Review of Materials Research*, 2012, 42.
2. Kitanovski, Andrej, and Peter W. Egolf., *Journal of Magnetism and Magnetic Materials*, 2009, 321.7, 777-781.
3. Zheng, Z. G., Yu, H. Y., Zhong, X. C., Zeng, D. C., & Liu, Z. W., *International Journal of Refrigeration*, 2009, 32.1, 78-86.
4. Engelbrecht, K. L., G. F. Nellis, and S. A. Klein., *Cryocoolers*, 2005, 13 471-480.
5. Silva, D. J., Ventura, J., Araújo, J. P., & Pereira, A. M., *Applied Energy*, 2014, 113, 1149-1154.
6. Nanto, D., Akbar, H., Soegijono, B., Kurniawan, B., Ghosh, N., Hwang, J. S., & Yu, S. C. (2017)., *Physica B: Condensed Matter*, 2017.
7. Hashimoto, T., Numasawa, T., Shino, M., & Okada, T. *Cryogenics*, 1981, 21.11, 647-653.
8. Shir, F., Mavriplis, C., Bennett, L. H., & Della Torre, E., *International Journal of Refrigeration*, 2005, 28.4, 616-627.
9. E. P. Wohlfarth, Amsterdam, The Netherlands: North-Holland, 1980, 215–216.
10. Lei, T., Engelbrecht, K., Nielsen, K. K., Bez, H. N., & Bahl, C. R. H., *Journal of Physics D: Applied Physics*, 2016, 49.34 , 345001.
11. Allab, F., Kedous-Lebouc, A., Fournier, J. M., & Yonnet, J. P., *IEEE transactions on magnetics*, 2005, 41.10, 3757-3759.
12. Nikkola, P., Mahmed, C., Balli, M., & Sari, O, *International Journal of Refrigeration*, 2014, 37, 43-50.

## ACKNOWLEDGMENTS

This work was supported by Thailand Center of Excellence in Physics under funding number THEP-60PIP-MU1. The authors also acknowledge e-Science Infrastructure Consortium for providing necessary computing resources.

# System Tuning for Energy Efficient Big Data Infrastructure

Soratoch Pornmaneerattanatri<sup>1,C</sup> and Putchong Uthayopas<sup>1</sup>

<sup>1</sup> *Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand*

<sup>C</sup> *E-mail: soratoch.p@ku.th, putchong@ku.th; Tel. +66852201444*

## ABSTRACT

Big Data is an emerging area of applying complex data analytic to a massive number of data. Big Data has a broad application in science, engineering, business, financial, and industrial applications. For a massive multi-terabyte dataset, a Big Data platform consists of hundreds of computing servers and middleware such as Spark or Hadoop are employed for the processing. One of the major operating cost for such a large infrastructure is an energy cost. In this paper, various approaches of system tuning of Big Data infrastructure is proposed. This enables the Big data platform to achieve the results with much less time and energy. Hence, the total operating cost is substantially reduced for an organization that depends on large-scale Big Data analytic for their strategic advantages.

**Keywords:** Big Data, Energy Efficient, Green IT, Performance Tuning.

## 1. INTRODUCTION

Currently, Big data is one of the most widely used applications in the world. Everyone want to extract useful information from the massive amount of data that are generated globally from many sources such as the social media, web board, sensors, and commercial transactions. This information is very useful for every modern digital organization. For example, a business can analyze the product purchase behavior using user's comments on a social network. This kind of analysis enables companies to discover a new way to sell the products and services. To perform the Big data analytic, one requires the massively amount of data to be processed on a substantially large cluster infrastructure. This cluster infrastructure usually consumes a lot of energy to operate. Hence, this is one of the major operating cost for this kind of Big Data infrastructure.

In this paper, techniques used to enable an energy efficient Apache Hadoop execution is investigated. Since the default Hadoop configuration and system configuration may not lead to an energy efficient execution, the smart tuning of the configuration will reduce the execution time and overall energy consumption. This paper is organized as follows. Section 2 presents some related works followed by the presentation of the proposed approaches in Section 3. Section 4 explains the experimental results. Finally, some finding and conclusion are summarized in Section 5.

## 2. RELATED WORKS

This application of Big data [1] is the analysis of a massive amount of data to gain an insight into the data such relationship, grouping, trends. Big data has 3 main properties; Volume, Velocity, and Variety. There is many software being used to perform the Big data analysis. One of the most used software is *Apache Hadoop* [2]. Apache Hadoop is the open source Java project that develops by Apache foundation. Hadoop is designed to deploy in a cluster of computers using distributed computing model. This approach enables Hadoop to scale to a very large system. The performance of Hadoop comes from the distribution of data and processing to multiple computers and execute in parallel. Hadoop developer ensures that the reliability of the system can be achieved using the failure detection in application level. If

any machine in Hadoop cluster fails, Hadoop will repair itself and continue working until the work is finished. Hadoop consists of 4 main components[3]: Hadoop Common, Hadoop Distributed File System or HDFS, YARN and Hadoop MapReduce. Hadoop common is the part that manages other modules in Hadoop system. HDFS function is to manage the distribution and storage of data. YARN is a resources management and work schedule. Finally, Hadoop Map Reduce is the module that enables the programmability for Hadoop system. Many researchers are working on reducing energy consumption. Nidhi Tiwari[4] measure consumption of energy energy for large Hadoop system and propose the concept of controlling CPU performance to reduce energy consumption. Shadi Ibrahim[5] works on the algorithm to control the frequency of CPU according to the processing demands. But the improvement of energy consumption is not much different that the standard module being used. Shadi work shows that CPU is not the only source of energy consumption challenge. Many factors involved in minimizing the energy consumption of the system.

### 3. PROPOSED APPROACH

For a typical computer server system, there are 2 parts of energy usage. First, the hardware consumes a constant amount of energy in the idle state when no processing has been done. This “*base-line*” energy consumption is the minimum energy consumption required by each computer. Second, the energy is consumed for the processing of application code. As the processing demands increases, the CPU frequency will be increased as well. Thus, more energy will be consumed accordingly. This “*working*” energy consumption depends heavily on the application logics and hardware used in the execution. Therefore, these two parts of energy consumption together comprise the total energy consumption of the server computer. The total energy consumption depends as well on the total execution time of the application. Thus, the minimization of the execution time is also an important factor in reducing energy consumption since the base-line energy consumed by each computer participating in the execution is reduced.

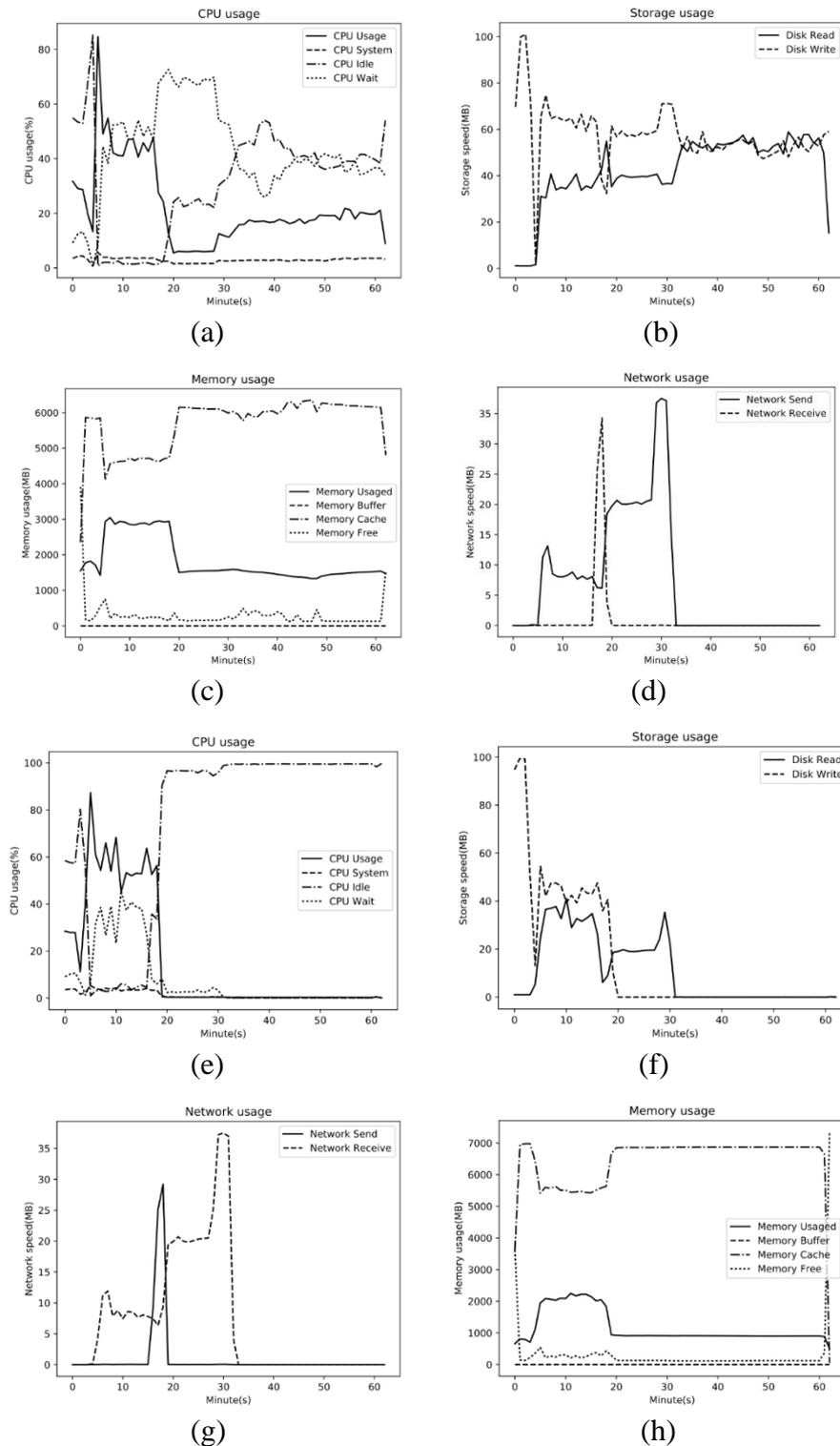
According to the concept mention earlier, the energy consumption can be given by  $E = (E_b + E_w)t$  where  $E$  is the total energy consumption,  $E_b$  is the base-line energy,  $E_w$  is the working energy, and  $t$  is the total execution time. In the real scenario, there are many machines involved. Let  $n$  depicts the number of machines in the system. The total energy for the system with  $n$  machines is given by  $E = ntE_b + t \sum E_w$ . To improve the equation further, two elements are needed. First, the energy from network switch must be taken into account. Second, the runtime overhead of the system on each machine must be included. This overhead usually comes from runtime system configuration and hardware. Thus, the total energy consumption of the Hadoop system can be given  $E = ntE_b + ntE_s + t \sum (E_w + E_o)$  where  $E_s$  energy from network switch and,  $E_o$  is an energy consumption from the overhead.

When closely consider the energy consumption  $E = ntE_b + ntE_s + \sum (E_w + E_o)$ , it can be seen that some parameters cannot be easily altered. For example,  $E_w$  depends on algorithm need to be processed and directly depends on the problem size. Thus, the optimization must be done on algorithm not the Big data runtime system. To reduce base energy the hardware needs to be carefully selected when building the infrastructure. In this work, the approach is to reduce the overhead of the computing process and try to shorten the execution time. Hadoop has overhead in many parts, from configuration to concept of MapReduce can cause of many areas of both hardware and software to have overhead. This can show by monitoring on Hadoop.

### 4. EXPERIMENTS AND DISCUSSION

In this work, the experiment has been conducted on 2 machines. The configuration of each machine is Intel i5 CPU 3330 3.00GHz 4 cores, 8GB of memory, Storage Seagate 1TB. Some experiments have been conducted using SSD OCZ ARC100 240GB. The operating system used is Linux Centos 7 with Java OpenJDK version 1.7

and Apache Hadoop version 2.6.5. The Terasort benchmark has been used for this study. The experiments have been conducted and the results have been shown in Figure 1,2,3 and 4.



**Figure 1.** Hadoop runtime monitoring results (a),(b),(c),(d) is measured from master node, (e),(f),(g),(h) is measured from compute node.

In Figure 1, the results from running Teragen and Terasort benchmark using input data of 40 gigabytes created by Teragen is presented. The first step is to consider tuning up the CPU

usage. From the CPU usage monitoring graph, one can observe show the massive overhead from CPU being in waiting and idle. Thus, the processing power is wasted. Therefore, the base-line energy becomes a dominant factor due to the unnecessary long execution time. From the closer investigation, it has been discovered that this massive CPU wait and idle comes from the writing and loading data between memory and storage due to the unmatched transfer speed. Thus, this problem must be fixed to increase the CPU processing period. However, the configuration does not accommodate the tuning to reduce this problem. Therefore, the solid-state drive which is much faster should be added to reduce this problem. After considering the tuning mentioned earlier, the memory can be tuned such as the disable of huge page and swap. This prevents the unnecessary swapping of memory to reduce the storage usage and ensure that all memory has been utilized efficiently. If the data size computed is larger than the memory available. One should consider adding more machine to practically add more memory rather than wasting processing time in memory swapping. After the tuning, the system has been tested using data size of 8,16,24,32,40GB accordingly. Then, the execution time and energy consumption is measure. The energy consumption of the whole system is measured using a watt meter. The results are as presented in Table 1 and Table 2. In Table 1 and Table 2, C represents the CPU tuning only. CM represents the tuning of both CPU and Memory. Finally, CMS represents the tuning for all CPU, memory, and Storage.

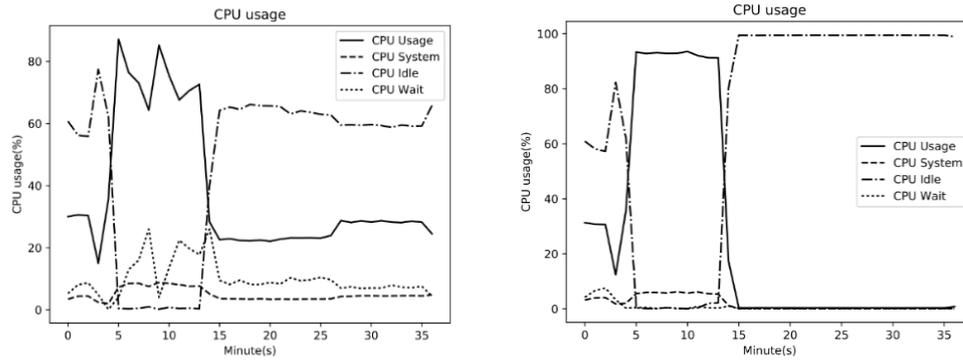
**Table 1.** The execution time for the test data size

Data Size (GB)	Runtime (second)			
	Default	C	CM	CMS
8	534	534	492	362
16	1237	1304	1194	742
24	2005	1990	2076	1203
32	2970	3043	2887	1633
40	3780	4724	3920	2097

**Table 2.** The energy consumption for the test data size.

Data Size (GB)	Energy (Watt)			
	Default	C	CM	CMS
8	8.67	8.67	8.00	6.00
16	20.67	21.00	18.67	13.00
24	32.33	31.33	32.00	19.33
32	46.00	46.67	44.67	27.00
40	61.67	68.67	62.00	34.33

From Table 1 and Table 2, the tuning shows that the reduction of energy and improved speed has been achieved. Anyway, the tuning of memory can only improve the system performance and energy consumption slightly. The substantial improvement happened when the solid state drive has been used. The tuning of storage to a faster one can reduce the runtime by 44.52% and reduce the energy consumption by 44.32% for 40GB data size. Figure 4, shows that CPU utilization resulted from the storage tuning is higher than the default setting.



**Figure 4.** CPU usage graph after the storage tuning.

## 5. CONCLUSION

In this paper, the characteristics of Hadoop system regarding the execution speed and energy consumption has been investigated. Several tuning up of the system has been proposed. From the benchmark, it has been seen that the source of in-efficiency that lead to more energy consumption is the idle time caused by the runtime system waiting for the storage access. Once, the storage has been changing to a much faster one, the speed and efficiency have been improved substantially. In the future, the careful study can be conducted to find more ways of improving the system energy efficiency. The results of this research can lead to a large cost reduction for the operation of a large scale Big data system in the future.

## REFERENCES

1. Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins, Nina Kruschwitz, *Big Data, Analytics and the Path from Insights to Value*, MIT Sloan Management Review, 2011, **52** (2), 21-32.
2. Tom White, *Hadoop: The definitive guide*, O'Rilly Media Inc., 2012.
3. Jeffrey Dean, Sanjay Ghemawat, *MapReduce: Simplified Data Processing on Large Cluster*, Communications of the ACM, 2008, 51 (1), 107-113.
4. Nidhi Tiwari, Umesh Bellur, Maria Indrawan, Santonu Sarkar, *Improving the Energy Efficiency of MapReduce Framework*, Ph.D. Forum, in Proceedings of International Conference on Distributed Computing and Networking ICDCN, 2014.
5. Shadi Ibrahim, Tien-Dat Phan, Alexandra Carpen-Amarie, Housseem-Eddine Chihoub, Diana Moise, Gabriel Antoniu, *Governing Energy Consumption in Hadoop through CPU Frequency Scaling: an Analysis*, Future Generation Computer System, 2016, 54 (1), 219-232

# Express Lane Services on Software-Defined Networks

Peeranon Wattanapong, Kittipat Jutawongcharoen, and Vara Varavithya

*Department of Electrical and Computer Engineering, Faculty of Engineering  
King Mongkut's University of Technology North Bangkok  
Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800  
Email: varavithya@gmail.com*

## ABSTRACT

Modern science cultivates enormous amounts of data. Demand for bandwidth is upsurging, especially for applications related to high energy physics, bioinformatics, and big data. Due to best-effort resource sharing nature of the Internet, the network providers cannot offer services that match user's bandwidth requirements. Resource reservations, dedicated networks, QoS, overlay networks, and virtual networks are examples of techniques in which high speed networks can be offered. In this work, we describe express lane services over the research DMZ framework on the UniNet. The NetFGPA based SDN nodes were deployed on 7 nodes. The express lane application was developed to command the RYU-SDN controller which, in turn, markdwns flows in SDN switches. The primitive express lane functions were tested in the real environment. A set of applications, which includes iPerf, ftp, and GridFTP, was used to suggest the available bandwidth using the proposed system. The performance levels in the vicinity of 550, 300, and 900 Mbps from iPerf, sftp, and GridFTP, respectively, were observed.

**Keywords:** Computer networks, software define networks, express lane services on demand services.

## 1. INTRODUCTION

Research and Education Networks (RENs) are considered as crucial infrastructure for conducting high quality education and research. In many cases, bandwidth level, offered by RENs, is an order of magnitude higher than that of commercially available obtained from commercial ISPs. Examples of RENs include GÉANT, ESnet, and Internet2. In Thailand, the office of higher education commission under the ministry of education has operated the UniNet to support research and education. The Internet serves education mainly for information browsing. Network usage characteristics from researchers might have more stringent requirements in terms of low latency, low jitter in realtime communication, and high bandwidth. Current bandwidth offered by ISPs cannot effectively serve the demands from researchers. Providing guarantee services between two end points involves complex system configuration. Based on these requirements, a research and education network has to support on-demand high speed networks [1].

To met special needs, several techniques, such as adopting L2-L3 virtual private networks (VPN) to create an overlay network or leasing dedicated links, have been adopted. The cost of dedicated link is prohibitively expensive, therefore virtual network concept is more attractive in managing network resources. An overlay network isolates network accesses among different classes of user to shared resources according to service level agreement. Due to the fact that the new generation of network devices are increasingly programable, software defined network (SDN) concept extends flexibility in providing personalized services and

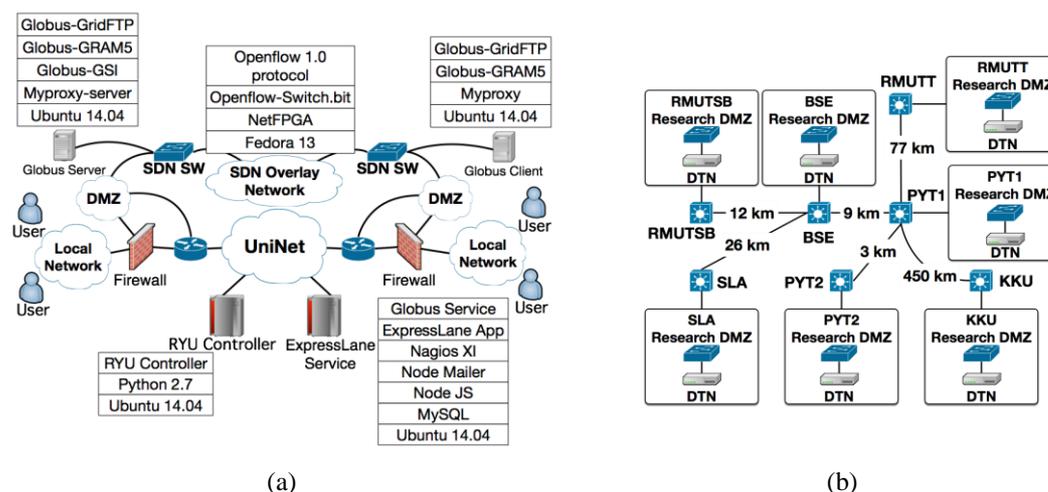
enhances richer set of network functions. Thus, functionalities of SDN have been used in RENs to benefit different needs.

In this work, we present the integrations of software defined network (SDN) and L2-VPN networks to create a new high performance services. The NetFPGA [2] cards are deployed as the SDN switches and the RYU is implemented as a controller [3, 4, 5]. The system can allocate resources by sending commands to set the designated flows at the SDN switches via OpenFlow protocol [6, 7]. A new service, called Express Lane, is introduced. The main objective of express lane services is to provide high bandwidth connections for research and education institutes. We demonstrated the use of sftp and GridFTP [8] to transfer large data at the high speed. Section 2 describes architecture of the express lane. The service deployments are explained in Section 3. Experimental results are given in Section 4 and the conclusions are drawn in the Section 5.

## 2. EXPRESS LANE ARCHITECTURE

At the lowest level, a user has access to the network resources via wireline connection(s) at his/her desk. A connection between two end-points is accomplished by forwarding data from user storage to the campus network. The campus network transports data to the edge router(s) where the REN service point is located. The bandwidth of local area network in campus environment usually deployed at 1 Gbps where 10 Gbps arrangement is possible in the near future. Internet security and traffic management devices, such as firewalls and load balancers, hinder performance perceived by the user because of fair usage and security policies. The DMZ concept, borrowed from security field, has been used to bypass these bottleneck devices. Several RENs offers high speed services via the Science DMZ [9]. As shown in Figure 1 (a), DMZs are located at the edge of the campus networks. A high speed VLAN path can be arranged from the DMZ to a user station. This establishment alleviates performance degradation from sharing of resources.

The UniNet extends their MPLS service to support multipoint L2-VPN services (VPLS) for 7 educational institutes in which the NetFPGA based SDN nodes were commissioned at each site. Figure 1 (b) shows a collection of participated organizations and the network topology as well as distance between the nodes.



**Figure 1.** System architecture of the UniNet's Express Lane Services (a) hardware and software architecture (b) Sites installation that geographically distributed in the kingdom.

Software components used in this work are also shown in Figure 1 (a). The system consists of five components, the express lane system, an SDN overlay network, a research

DMZ, the SDN controller, and SDN switches. The SDN overlay network is considered as the high bandwidth network infrastructure which is an overlay network over research and education network, the UniNet in our case. The overlay network is supported by underlining MPLS network where traffic engineering policy is imposed. The RYU controller was selected in our works.

The RYU controller reads the request and generates respective entries in the participated SDN switches. Each PC based SDN switch is a linux box with a NetFPGA card. The NetFPGA was configured to support the OpenFlow 1.0. Due to the older version of NetFPGA cards we possess, the version of OpenFlow is limited to 1.0. We developed an express lane application using Node.js to process service requests from users and coordinate works with the RYU controller. A set of auxiliary applications, Nagios XI for server monitoring and Globus Service for authentication is also resided in the express lane service server.

To demonstrate high speed data transfer, the GridFTP software was installed at the Globus end points to test Disk-Disk Bandwidth. The GridFTP is a highly optimized software that help transfer data at the high bandwidth where the path between two end points is created by the express lane system.

### 3. EXPRESS LANE SERVICE DESCRIPTIONS

Users in the express lane system are classified into three groups, namely: audience, member, and administrator. As shown in Figure 2 (a), the use cases are assigned to each group. The system administrator is responsible for managing users and approving service requests as well as monitoring system operations. Audiences are general users who can apply for services and get through the approval process. In creating a service request, a member user has to provide

[Src\_MAC, Dest\_MAC, Start\_Time, End\_Time]

to the system.

The service duration is specified by *Start\_Time* and *End\_Time*. Accordingly, the user can configure both communication end points of high speed connection within the requested period. The mapping between a user request and SDN flows for each switch is generated by the express lane system and communicates with the SDN controller. After approval, the administrator sends *Src\_IP* and *Des\_IP* addresses to the user. The express lane system maintains three tables, *requested\_service*, *approved\_service*, and *active\_service*. A request is first inserted to the *requested\_service* table and move to *approved\_service* when it is authorized by the administrator. When the time to perform

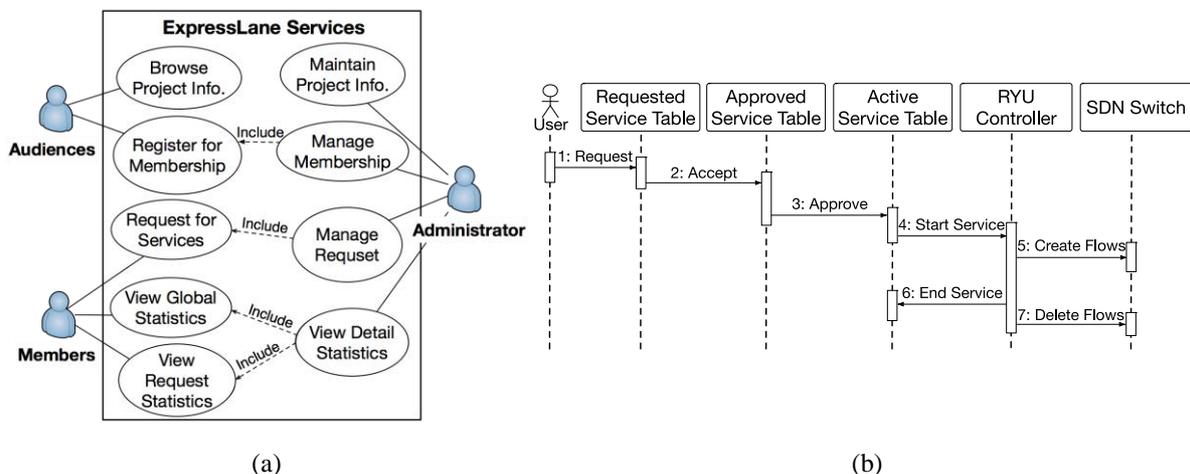


Figure 2. Express lane applications: (a) use cases (b) service flows.

arrives, the request is moved to the `active_service` table. The `active_service` is used to activate the RYU controller where respective flows are created and set at the group of participated switches. After the service time expire, the RYU will remove all related flows for that request.

Figure 2 (b) shows the sequence diagram specifying a request process. It is possible to merge these tables using flags to specified states of a service. However, there are more flexibility in implementing multiple types of services if the tables are kept separate from each other. It is easier to extend this application to support other services or include different controllers.

The express lane front end was developed on the Node.js framework. A set of REST APIs were designed and implemented to decouple the state of express lane services from front end interface. This design has advantages when we want to change presentation frameworks or build more services.

#### 4. RESULTS AND DISCUSSION

In this work, the SDN nodes were deployed on 7 nodes in the UniNet backbone, as shown in Figure 1 (b). The objectives of the performance evaluations are to test the functionalities of the express lane in real environment and measure performance level obtained from the research DMZ. We use ping, iPerf, perfSONAR, ftp, and GridFTP as tools to measure performance. Entry level PCs were used as the data transfer nodes in the tests. Software suites, described in Section 2, were configured. The process of express lane operation is as follows:

1. The request [00:1c:f0:cf:eb:ae, 00:18:f8:0b:59:17, 08:00 26-06-2017, 12:00 28-06-2017] was created and submitted to the express lane system. The first tuple is the MAC addresses of two end points and the second tuple is the start and end time.
2. The system administrator approved the request and specified the IP address. In our experiment a pair of [10.10.1.4, 10.10.1.7] was used for communication end points. This IP information was given to the user.
3. The user configured their machines using respective IP addresses.
4. When the time to perform arrived, the Node.js moved the request from `approved_service` to `active_service` in the database.
5. At SDN switches, a new packet injected to the system induces an event. All new packet-in events were forwarded to the RYU controller. The RYU controller checked the `active_service` table to match IP addresses with outstanding requests. If the matching request found, the flows were created and sent to the SDN switches. The SDN switches were then update the flow table. Example of the flows in this request, dumped from the SDN switch at the SLA node, is as follow:

```
[cookie=0, duration_sec=19s, table_id=0, priority=32768,
n_packets=0, n_bytes=0, idle_timeout=0, hard_timeout=14400,
in_port=3, dl_dst=00:1c:f0:cf:eb:ae, actions=output:4]
```

```
[cookie=0, duration_sec=24s, table_id=0, priority=32768,
n_packets=22, n_bytes=1320, idle_timeout=0, hard_timeout=14400,
in_port=4, dl_dst=00:18:f8:0b:59:17, actions=output:3]
```

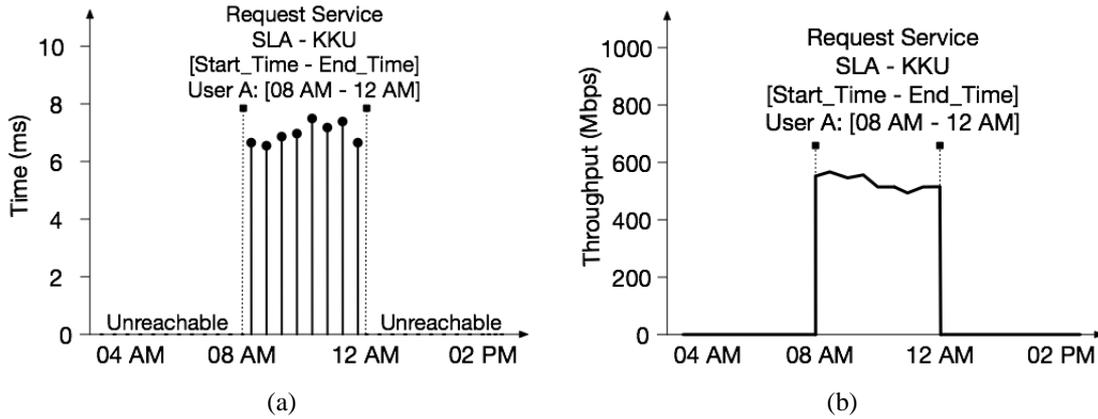
The `n_packets` is the number of packets passed through a switch and `n_bytes` is the total number of bytes forwarded. The `hard_timeout` specifies time-to-live of the flow. The `in_port` is the input port that connect to the user machine with specified MAC address and the `out_port` is the forwarding port to the backbone.

6. During the express lane system provided service to the request, the RYU controller maintain the following flow:

```
1:[{action:[OUTPUT:3], cookie:0, priority:32768, table_id:0,
idle_timeout: 0, hard_timeout:14,400, match:{dl_dst:
00:18:f8:0b:59:17, in_port:4}}]
```

The above flow indicates that, the flow no. 1, a packet from input port 4 with matched MAC address is sent to the output port 3 with time to live of 14,400 seconds.

7. After service time expired, the express lane application (Node.js) removed the entry from the active\_service table and the hard time out expired at the SDN switch. Any packet forwarded to the switch was then treat as a new packet-in and forwarded to the RYU controller. If there is no service entry in the active\_service table, no connection is allowed. The service ended for this request.



**Figure 3.** Operations of the Express Lane services between SLA and KKU nodes. The express lane requests for connection between 8-12 AM (a) the ping result between connection period (b) bandwidth result using iPerf.

From the experiment, the express lane system can control the traffic flow and gives satisfactory bandwidth results as shown in Figure 3 (a) and (b). To test the benefit of express lane system over research DMZ network, we performed the all-to-all bandwidth tests covered every nodes. Figure 4 (a) shows the bandwidth performance results using the iPerf tool. The bandwidth results are range between 496 to 685 Mbps which is quite high considering that the measurement was performed between user desks from different institutions. The sftp results represent disk-to-disk transfer where the bandwidth of 300 Mbps is attained, shown in Figure 4 (b). Tuning of data transfer node is very crucial to achieve high speed connections. There are several aspects that needed to consider such as TCP buffer size, NIC library, maximum MTU, etc.

iPerf (Mbps)	PYT 1	PYT 2	BSE	SLA	RMU TT	RMU TSB	KKU	10GB sftp (Mbps)	PYT 1	PYT 2	BSE	SLA	RMU TT	RMU TSB	KKU
PYT 1	-	685	660	455	587	604	396	PYT 1	-	317.6	311.2	325.6	149.8	388.8	150.4
PYT 2	603	-	607	603	610	608	570	PYT 2	362.4	-	316.6	317.6	150.4	318.4	318.4
BSE	589	592	-	587	602	596	593	BSE	392	318.6	-	298.4	224.2	308.8	304.4
SLA	588	589	588	-	589	587	585	SLA	364.8	322.4	323.2	-	158.4	344	189.6
RMU TT	597	598	602	593	-	576	609	RMU TT	247.6	218.4	310.4	233.6	-	315.6	257.6
RMU TSB	585	589	593	566	564	-	584	RMU TSB	367.6	321.6	319.2	314.4	309.4	-	181.1
KKU	466	556	396	541	559	597	-	KKU	129.6	248	237.2	237.2	151.2	310	-

BSE - RMUTSB Private IP, 1 Gbps Port			
File Size	FTP	SFTP	GSIFTP
1 GB	900.03	373.12	902.45
10 GB	888.18	382.4	900.14
100 GB	861.51	376.32	899.98

in Mbps

(a) (b) (c)  
**Figure 4.** Performance result of express lane services over the Research DMZ the UniNet (a) iPerf bandwidth result (b) ftp performance for 10 GB file (c) GridFTP bandwidth performance

A well-tune data transfer software as well as optimizing operating system parameters can help maintain high data rate. The GridFTP is well serve this purpose. Figure 4 (c) gives comparison results of ftp, sftp and GridFTP. Bandwidth of 900 Mbps is reached. The express lane service provides on-demand network services for shared network resources which are shown in the above results for both functionalities and bandwidth performance.

## 5. CONCLUSIONS

The Internet has intertwined in the way we conducting research. In many cases, researchers demand premium network services for a certain period of time. The research DMZ concept gives virtual dedicated network links among research institutes at the high data rate. In this paper, we describe the deployment and experimental results of the high speed on-demand network service, called express lane. A user can request for high speed network service and use facility for a specified period. The express lane application is explained and deployed over 7 nodes in the UniNet. We can achieve up to 900 Mbps of data transfer from desk-to-desk with resource reservation system. The UniNet can lunch this service to the research community in Thailand. New services can be further developed and experimented in the research DMZ testbed with SDN nodes.

## REFERENCES

1. D. Adami, L. Donatini, S. Giordano, and M. Pagano, "A network control application enabling software-defined quality of service," *in the proceeding of the IEEE International Conference on Communications (ICC)*, 2015, 6074–6079.
2. J. Naous, D. Erickson, G. Adam Covington, G. Appenzeller, and N. McKeown, "Implementing an OpenFlow switch on the NetFPGA platform", *in the proceedings of the Symposium on Architectures for Networking and Communications Systems*, 2008.
3. N. Guge, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, "NOX: Towards an operating system for networks," *SIGCOMM Computing Communication Review*, July 2008, 38(3):105-110.
4. S. Kaur, J. Singh, and N. S. Ghumman, "Network programmability using POX controller," *in the proceeding of the International Conference on Communication, Computing and Systems (ICCS 2014)*, 2014.
5. Using OpenFlow 1.3 RYU SDN framework: RYU project team. Web site <https://osrg.github.io/ryu-book/en/Ryubook.pdf>
6. N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling innovation in campus networks," *SIGCOMM Comput. Commun. Rev.*, 2008, 38(2):69–74.
7. "OpenFlow Switch Specification Version 1.0.0," <http://www.openflow.org/documents/openflow-spec-v1.0.0.pdf>, 2009.
8. W. Allcock, J. Bresnahan, R. Kettimuthu, M. Link, C. Dumitrescu, and I. Raicu, I. Foster, "The Globus striped GridFTP framework and server," *SC'05*, ACM Press, 2005.
9. K. Jutawongcharoen, V. Varavithya, K. Lekdee, A. Chaichit, and T. Sribuddee, "The implementation of the UniNet's research DMZ," *in the proceeding of the International Computer Science and Engineering Conference (ICSEC)*, 2016, 1-5.

## ACKNOWLEDGMENTS

This research was supported in part by Office of Information Technology Administration for Educational Development, Office of Higher Education Commission, MOE and the Department of Electrical and Computer Engineering, King Mongkut's University of Technology North Bangkok.

