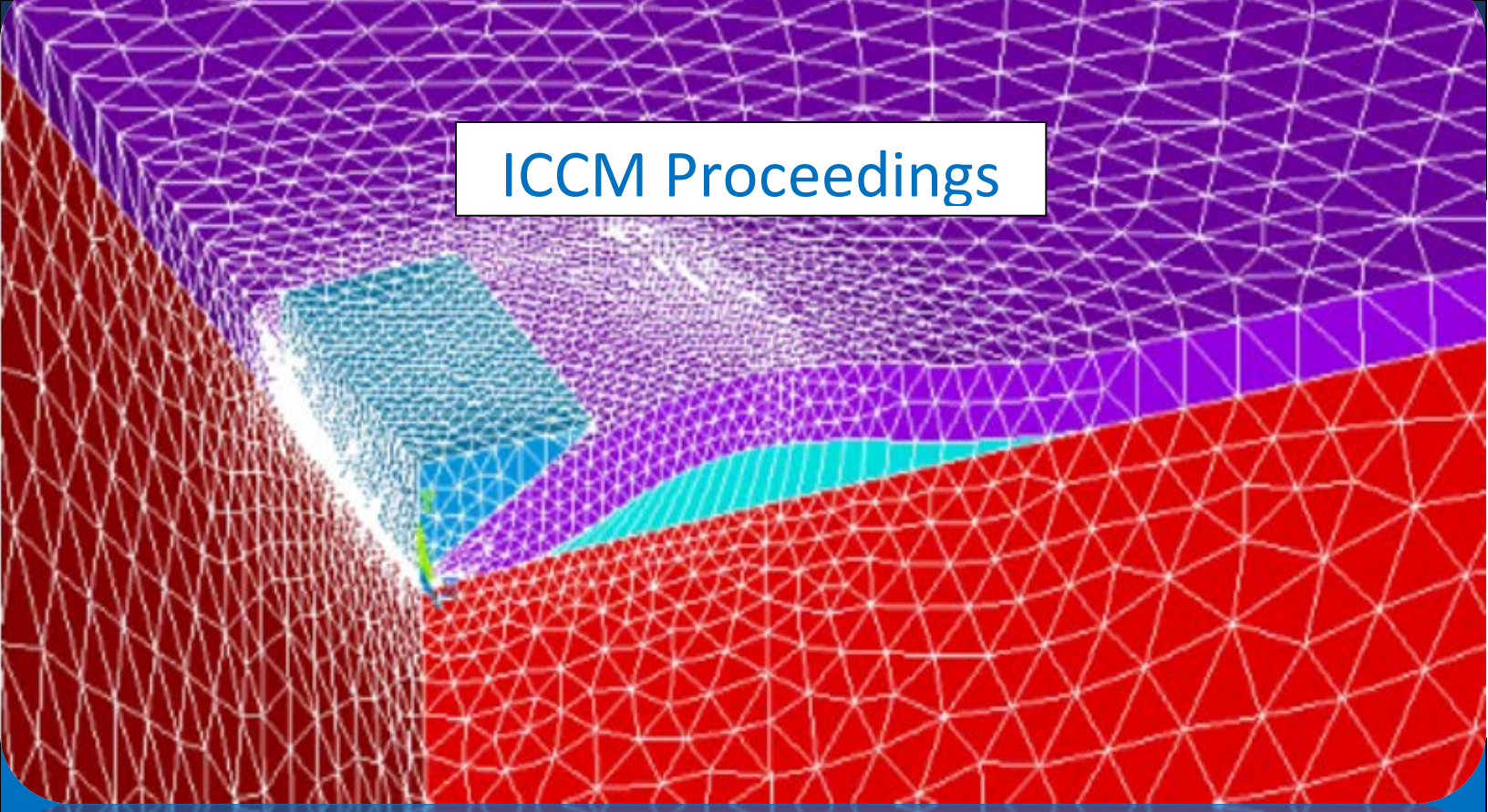


ICCM Proceedings

# Proceedings of the International Conference on Computational Methods

## (Vol. 6, 2019)

10th ICCM, 9th-13th July 2019, Singapore

Editors: **G. R. Liu, Fangsen Cui, George Xu**

ScienTech Publisher

ISSN 2374-3948 (online)

# ICCM2019

Proceedings of the International Conference on Computational Methods (Vol. 6, 2019)

10th ICCM, 9th-13th July 2019, Singapore

Edited by

**G. R. Liu**
University of Cincinnati, USA

**Fangsen Cui**
Institute of High Performance Computing, A*Star, Singapore

**George Xu Xiangguo**
Institute of High Performance Computing, A*Star, Singapore

Proceedings of the International Conference on Computational Methods, Vol.6, 2019

This volume contains full papers accepted by the 10th ICCM, 9th-13th July 2019, Singapore.

First Edition, July 2019

International Standard Serial Number: ISSN 2374-3948 (online)

Papers in this Proceedings may be identically cited in the following manner: Author names, *Paper title, Proceedings at the 10th ICCM2019, 9th-13th July 2019, Singapore*, Eds: G.R. Liu, Fangsen Cui, George Xu Xiangguo, ScienTech Publisher.



Published by
Scientech Publisher LLC, USA
http://www.sci-en-tech.com/

# WELCOME MESSAGE

Dear Colleagues and Friends,

On behalf of the organising committees, we are delighted to welcome you to the 10th Conference on Computational Methods (ICCM2019) at Singapore. The ICCM conference series is an international conference that provides an international forum for exchange of ideas on recent advances in areas related to computational methods, including computational mechanics, numerical modelling & simulation of manmade or natural systems, as well as their applications in engineering and sciences. It will accommodate presentations on a wide range of topics to facilitate inter-disciplinary exchange of ideas in science, engineering and related disciplines, and foster various types of academic collaborations in the internationally. All papers accepted for publication in the proceedings have been peer reviewed. Papers may also be selected and invited to be developed into a full journal paper for publication in special issues of some peer-reviewed journals.

The ICCM (International Conference on Computational Methods) conference series were originated in Singapore in 2004 by Professor GR Liu, followed by ICCM2007 at Hiroshima, Japan, ICCM2010 at Zhangjiajie, China, ICCM2012 at Gold Coast, Australia, ICCM2014 at Cambridge, England, ICCM2015 at Auckland, New Zealand, ICCM2016 at Berkeley, CA, USA, ICCM2017 at Guilin, Guangxi, China, ICCM2018 at Rome, Italy. Now ICCM2019 comes back to Singapore for celebrating its 10th event.

The ICCM2019 conference program includes over 330 presentations from more than 20 countries and regions scheduled in 47 technical sessions. There will be 3 Plenary Lectures, 8 Thematic Plenary Lectures, many Keynote Lectures and Invited Lectures at the conference. The conference sessions will cover a broad range of topics related to computational methods, including formulation theory, computational techniques, machine learning, modelling techniques and procedures, materials, deformation processing, materials removal processes, processing of new and advanced materials, welding and joining, surface engineering and other related processes.

We would like to express our gratitude to all the members of the Local Organizing Committee, International Scientific Committee, and the student helpers who have contributed significantly in this conference. Our sincere thanks and appreciation go to some international reviewers for their prompt review reports on the submitted abstracts and papers. Our appreciation goes also to all the Mini-Symposium Organizers for their effort and contribution in the organization. Special thanks go to Dr. George Xu, the secretary general, for handling a lot of tedious work, Prof Li Hua for the great support by organizing student helpers.

We hope that this conference will provide a great venue of presenting and exchanging information for your scientific work. We wish all of you have a great time in this beautiful garden city Singapore. Finally, we would like to thank you for your contribution to the ICCM2019 conference. We are looking forward to your participation and continued engagement for the future ICCM conferences.

Dr. Fangsen Cui
Conference Chairman
Institute of High Performance Computing, A*STAR
Singapore Association for Computational Mechanics, Singapore

Professor Gui-Rong Liu
Honorary Conference Chairman
University of Cincinnati
USA

Dr. George Xu Xiangguo
Conference Secretary
Institute of High Performance Computing, A*STAR
Singapore Association for Computational Mechanics, Singapore

# ORGANIZATION COMMITTEES

**Conference Chairman**
Fangsen Cui (Institute of High Performance Computing, A*STAR, Singapore)

**Honorary Chairman**
Guirong Liu (University of Cincinnati, United States)

**International Co-Chairs**
Moubin Liu (Peking University, China)
Yuantong Gu (Queensland University of Technology, Australia)
Zishun Liu (Xi'an Jiaotong University, China)
Patrizia Trovalusci (Sapienza University of Rome, Italy)
Sung-Kie Youn (Korea Advanced Institute of Science and Technology, South Korea)

**Local Co-Chairmen**
Hua Li (Nanyang Technological University, Singapore)
Vincent Tan Bing Chye (National University of Singapore, Singapore)

**Secretary General**
George Xu Xiangguo (Institute of High Performance Computing, A*STAR, Singapore)

**Local Organizing Committee**
Gideon Praveen Kumar Vijayakumar, Jingjie Yeo, Hongying Li, Menglong Liu, Jun Liu, Yuan Cheng, Ling Dai, Zhiqian Zhang, Zhoucheng Su, Yucheng Zhong, Dan Wang, Junyan Guo, Xiuqing Xing, Gioia Fusaro (Institute of High Performance Computing, A*STAR, Singapore)
Xingyu Zhang, Xiao Chen, Yifeng Wang, Kewei Chen, Clarence Tee, Brian Lee Zhan Rui, Elisa Ang Yun Mei, Simiao Fan, Junheng Ren (Nanyang Technological University, Singapore) Yao Sun (National University of Singapore, Singapore)

**International Scientific Advisory Committee**

Addessi Daniela (Italy)
Aoki Takayuki (Japan)
Beer Michael (Germany)
Birken Philipp (Sweden)
Blachut Jan (United Kingdom)
Bui Tinh Quoc (Japan)
Chen Chuin-Shan David (Taiwan)
Chen Gongfa (China)
Chen Jeng-Tzong (Taiwan)
Chen Lei (USA)
Chen Songying (China)
Chen Weiqiu (China)
Chen Zhen (USA)
Chiu Wing Kong (Australia)
Colaco Marcelo (Brazil)
Cui Fangsen (Singapore)
Dong Leiting (China)
Dulikravich George (USA)
Fan Chia-Ming (Taiwan)
Fantuzzi Nicholas (Italy)
Fu Zhuojia (China)
Gan Yixiang (Australia)
Gao Wei (Australia)
Gao Xiaowei (China)
Geers Marc (Netherlands)
Greco Fabrizio (Italy)
Gu Yuantong (Australia)
Han Xu (China)
Herrera Ismael (Mexico)
Hirose Sohichi (Japan)
Hori Muneo (Japan)
Huang Wei (China)
Huang Yu (China)
Huang Zhengming (China)
Hung Nguyen-Xuan (Viet Nam)
Jabareen Mahmood (Israel)
Jacobs Gustaaf (USA)
Jiang Chao (China)
Jin Feng (China)
Kanayama Hirpshi (Japan)
Kang Zhan (China)
Khoo Boo Cheong (Singapore)
Kim Woo Kyun (USA)
Kim Yoon Young (South Korea)
Kougioumtzoglou Ioannis (USA)
Le Canh Van (Vietnam)
Lee Chin-Long (New Zealand)
Lee Heow Pueh (Singapore)
Lee Ik-Jin (South Korea)
Lenci Stefano (Italy)
Leo Hwa Liang (Singapore)
Li Hua (Singapore)
Li Quanbing Eric (UK)
Li Wei (China)

Li Yueming (China)
Lim Keng Hui (Singapore)
Lim Kian Meng (Singapore)
Liu Feng (China)
Liu Jie (China)
Liu Moubin (China)
Liu Yan (China)
Liu Yijun (China)
Liu Zhanli (China)
Liu Zishun (China)
Ma Qingwei (UK)
Miller Karol (Australia)
Montenegro Armas Rafael(Spain)
Navarrina Fermin (Spain)
Ng Teng Yong (Singapore)
Nithiarasu Perumal (UK)
Niu Xiaodong (China)
Nogueira Xesús (Spain)
Ogino Masao (Japan)
Orlande Helci (Brazil)
Papadrakakis Manolis (Greece)
Peng Qing (USA)
Picu Catalin (USA)
Poh Leong Hien (Singapore)
Qin Qinghua (Australia)
Quek Sinsin Jerry (Singapore)
Reali Alessandro (Italy)
Rębielak Janusz (Poland)
Reddy Daya (South Africa)
Saavedra Flores Erick (Chile)
Sadowski Tomasz (Poland)
Saitoh Takahiro (Japan)
Shen Lian (USA)
Shen Luming (Australia)
Shen Yongxing (China)
Shu Chang (Singapore)
Song Chongmin (Australia)
Stefanou George (Greece)
Tadano Yuichi (Japan)
Tan B.C. Vincent (Singapore)
Tanaka Satoyuki (Japan)
Tian Zhaofeng (Australia)
Trovalusci Patrizia (Italy)
Trung Nguyen-Thoi (Vietnam)
Tsubota Kenichi (Japan)
Wan Decheng (China)
Wang Bo (China)
Wang Cheng (China)
Wang Dongdong (China)
Wang Hu (China)
Wang Jie (China)
Wang Lifeng (China)
Wang Wenquan (China)
Wu Yanling (China)
Xiang Zhihai (China)
Xiao Feng (Japan)
Xiao Jinyou (China)
Xie Mike (Australia)
Xu Chao (China)
Xu Minglong (China)
Xu Xiangguo George (Singapore)
Xu Xu (China)
Yang Chun Charles (Singapore)
Yang Judy (Taiwan)
Yang Yeong-Bin (Taiwan)
Yang Zhenjun (China)
Yao Jianyao (China)
Yao Yao (China)
Ye Hongling (China)
Ye Wenjing (Hong Kong)
Yeo Jingjie (Singapore)
Yoshimura Shinobu (Japan)
Youn Sung-Kie (South Korea)
Yu Chengxiang Rena (Spain)
Yuan Mingwu (China)
Zeidan Dia (Jordan)
Zhan Haifei (Australia)
Zhang Jian (China)
Zhang Lihai (Australia)
Zhang Lucy (USA)
Zhang Xiong (China)
Zhang Zhao (China)
Zhao Kejie (USA)
Zhao Liguo (UK)
Zhou Kun (Singapore)
Zhuang Zhuo (China)

# Table of Contents

# A hybrid reconstructed discontinuous Galerkin method for incompressible flows on arbitrary grids

***Fan Zhang[1], Jian Cheng[2], and †Tiegang Liu[1]**

[1]School of Mathematics and System Science, Beihang University, China
[2]Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, China

*Presenting author: zhangfan1990@buaa.edu.cn
†Corresponding author: liutg@buaa.edu.cn

## Abstract

The discontinuous Galerkin (DG) methods have attained increasing popularity for solving the incompressible Navier-Stokes (INS) equations in recent years. However, the DG methods have their own weakness due to the high computational costs and storage requirements. In order to tackle this problem, in this paper, a hybrid least-squares reconstruction DG (rDG) method, namely $P_1P_2$(HLSr), is presented to obtain a quadratic polynomial solution from the underlying linear DG solution by use of a hybrid recovery and reconstruction strategy. This hybrid rDG method combines the simplicity of the reconstruction-based DG method and the accuracy of the recovery-based DG method, and has the desired property of 2-exactness which is violated by the original least-squares rDG method. The inviscid term of the INS equations, which is split into the nonlinear convective term and the linear Stokes operator, is discretized by using a simplified artificial compressibility flux. More specially, the nonlinear convective term is discretized in divergency form by using the local Lax-Friedrichs flux, while the Stokes operator is discretized by using the artificial compressibility flux which is provided by the exact solution of a Riemann problem associated with a local artificial compressibility perturbation of the Stokes system. The discretization of the viscous term follows the simple direct DG (DDG) method. A number of incompressible flow problems, in both steady and unsteady forms, for a variety flow conditions are computed to numerically assess the spatial order of convergence of the $P_1P_2$(HLSr) method, which demonstrate its ability to achieve the designed optimal 3rd-order of accuracy at a significantly reduced computational costs.

Keywords: Incompressible Navier-Stokes equations, reconstructed methods, discontinuous Galerkin method, artificial compressibility, arbitrary grids

## Numerical results

*Kovasznay problem*

The analytic solution for the 2D INS equations was derived by Kovasznay. The analytical expression for the velocity and the pressure is

$$u(x,y) = 1 - e^{\lambda x}\cos(2\pi y),$$

$$v(x,y) = \frac{\lambda}{2\pi}e^{\lambda x}\sin(2\pi y),$$

$$p(x,y) = -\frac{1}{2}e^{\lambda x}.$$

Here, $\lambda = \frac{Re}{2} - \sqrt{\frac{Re^2}{4} + 4\pi^2}$. The computational domain is $\Omega = (-\frac{1}{2}, \frac{3}{2}) \times (0, 2)$ with prescribed Dirichlet boundary conditions on $\partial\Omega$. The Reynolds number is Re=10 and the artificial compressibility parameter is $c^2$=1.0.

The comparisons among the numerical results obtained by the DG($P_1$), DG($P_2$) and $P_1P_2$(HLSr) methods are presented in Tab.1. It can be seen that the $P_1P_2$(HLSr) method, as expected, adding one order of accuracy to the underlying DG($P_1$) method and even having higher order of accuracy for the pressure than the DG($P_2$) method. Although the DG($P_2$) method does yield a slightly more accurate solution than the $P_1P_2$(HLSr) method at the same grid resolution, however, it is obtained at the cost of more number of the degrees of freedom which leads to a dramatic increase of the computational cost. The detailed convergence history, which contains the number of iteration steps and the CPU time (s) to reduce the residual by 8 orders of magnitude are presented in Tab.2.

Tab1. Convergence results for the Kovasznay problem.

| Grid size | No. DOFs | $\|e_{\mathbf{u}}\|_2$ | | $\|e_p\|_2$ | | $\|e_{\nabla\cdot\mathbf{u}}\|_2$ | |
|---|---|---|---|---|---|---|---|
| | | Error | Order | Error | Order | Error | Order |
| DG($P_1$) | | | | | | | |
| $8\times8$ | 192 | 3.19e-1 | | 5.02e-1 | | 1.41e-0 | |
| $16\times16$ | 768 | 7.89e-2 | 2.02 | 1.25e-1 | 2.01 | 6.86e-1 | 1.04 |
| $32\times32$ | 3,072 | 1.92e-2 | 2.04 | 3.42e-2 | 1.87 | 2.40e-1 | 1.52 |
| $64\times64$ | 12,288 | 4.56e-3 | 2.07 | 9.92e-3 | 1.79 | 7.00e-2 | 1.78 |
| DG($P_2$) | | | | | | | |
| $8\times8$ | 384 | 6.00e-2 | | 6.29e-2 | | 4.48e-1 | |
| $16\times16$ | 1536 | 7.35e-3 | 3.03 | 1.18e-2 | 2.41 | 9.25e-2 | 2.25 |
| $32\times32$ | 6,144 | 9.21e-4 | 3.00 | 2.37e-3 | 2.32 | 1.95e-2 | 2.25 |
| $64\times64$ | 24,576 | 1.26e-4 | 2.87 | 5.15e-4 | 2.20 | 4.40e-3 | 2.15 |
| $P_1P_2$(HLSr) | | | | | | | |
| $8\times8$ | 192 | 1.39e-1 | | 2.84e-1 | | 8.65e-1 | |
| $16\times16$ | 768 | 1.71e-2 | 3.02 | 4.29e-2 | 2.73 | 2.00e-1 | 2.11 |
| $32\times32$ | 3,072 | 2.08e-3 | 3.04 | 6.63e-3 | 2.77 | 3.48e-2 | 2.57 |
| $64\times64$ | 12,288 | 2.24e-4 | 3.22 | 1.02e-3 | 2.70 | 5.49e-3 | 2.66 |

Tab.2 Convergence history for the Kovasznay problem.

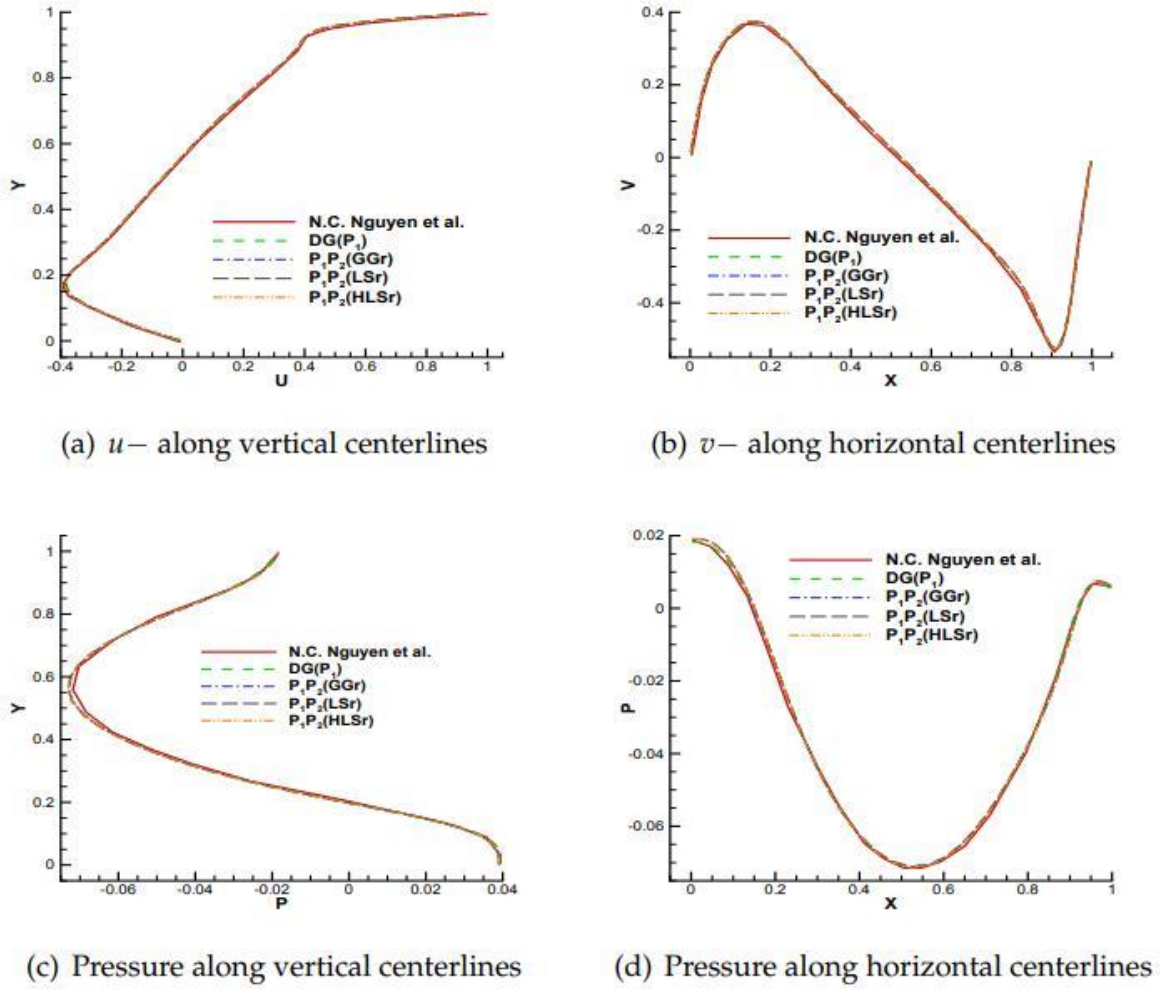
| Grid size | Iteration step | | | CPU time (s) | | |
|---|---|---|---|---|---|---|
| | DG($P_1$) | DG($P_2$) | $P_1P_2$(HLSr) | DG($P_1$) | DG($P_2$) | $P_1P_2$(HLSr) |
| $8\times8$ | 95 | 95 | 93 | 9.65 | 28.08 | 3.63 |
| $16\times16$ | 127 | 126 | 126 | 23.71 | 103.13 | 12.25 |
| $32\times32$ | 159 | 204 | 156 | 84.40 | 557.36 | 84.91 |
| $64\times64$ | 201 | 250 | 190 | 679.12 | 5289.36 | 785.05 |

*Lid-driven cavity flow*

The lid-driven cavity flow has been widely used as a validation case for numerical method of the INS equations. The problem has simple geometry and boundary conditions. The standard

case is fluid contained in a square domain $\Omega = (0,1)^2$ with homogeneous Dirichlet boundary conditions on all sides except on the upper side where the velocity is prescribed as $\mathbf{u} = (1,0)$. Here, we compare the performances of the DG($P_1$), DG($P_2$) and $P_1P_2$(HLSr)methods at high Reynolds number by the lid-driven cavity flow problem with Re=1,000, 5,000 and 10,000.

The u-velocity and pressure profiles along a vertical line and the v-velocity and pressure profiles along a horizontal line passing through the geometric center of the cavity respectively are presented in Fig.1-Fig.3. It can be seen that the present method is able to mimic the available results with great accuracy, the profiles are in good agreement of the reference results.



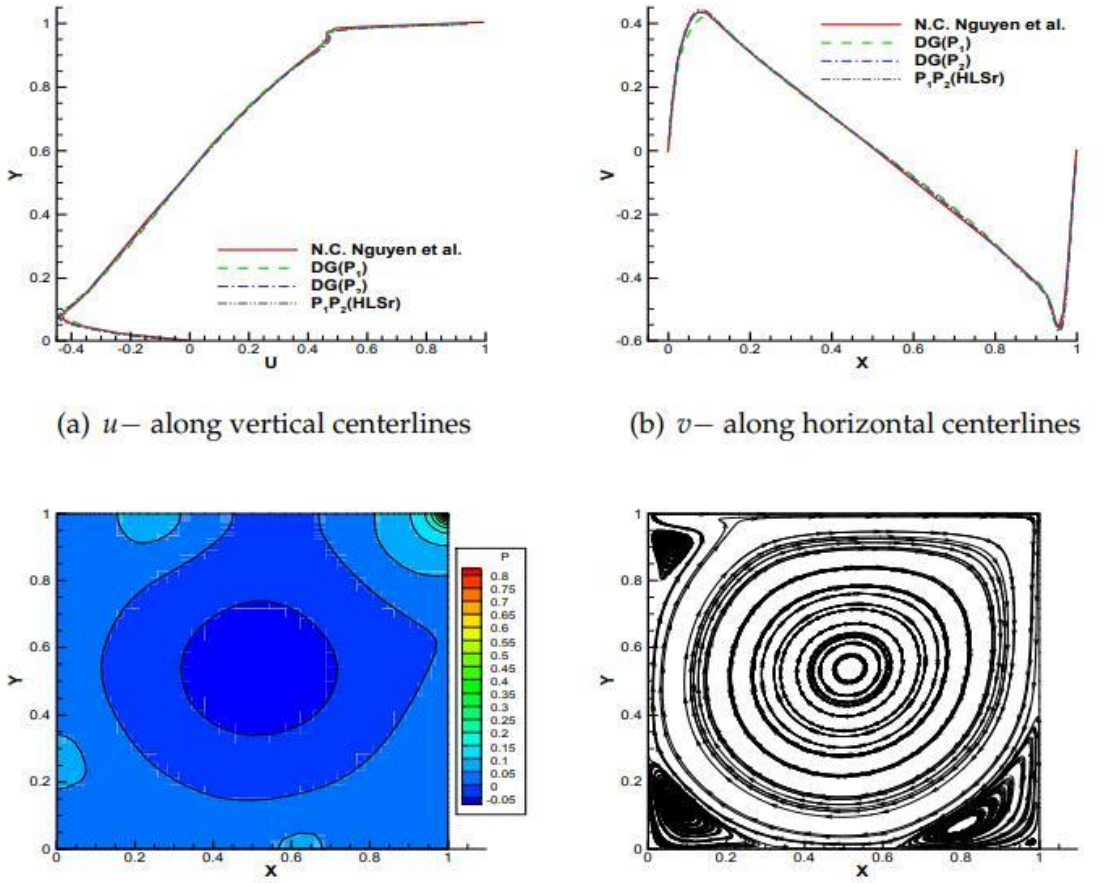

(a) $u-$ along vertical centerlines

(b) $v-$ along horizontal centerlines

(c) Pressure along vertical centerlines

(d) Pressure along horizontal centerlines

Fig.1 Results for the lid-driven cavity flow at Re=1,000.



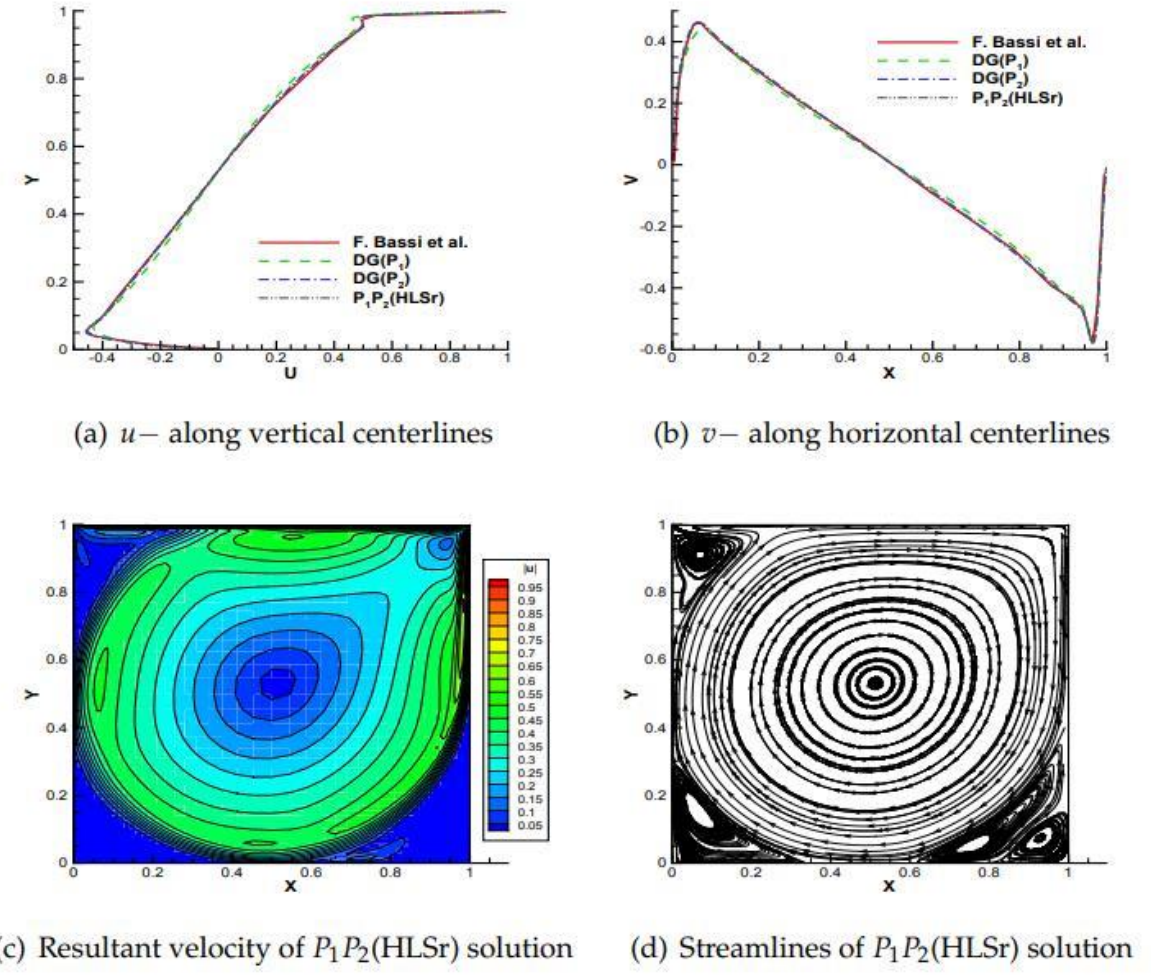

(a) $u-$ along vertical centerlines

(b) $v-$ along horizontal centerlines

(c) Pressure of $P_1P_2$(HLSr) solution

(d) Streamlines of $P_1P_2$(HLSr) solution

Fig.2 Results for the lid-driven cavity flow at Re=5,000.




(a) $u-$ along vertical centerlines

(b) $v-$ along horizontal centerlines

(c) Resultant velocity of $P_1P_2$(HLSr) solution

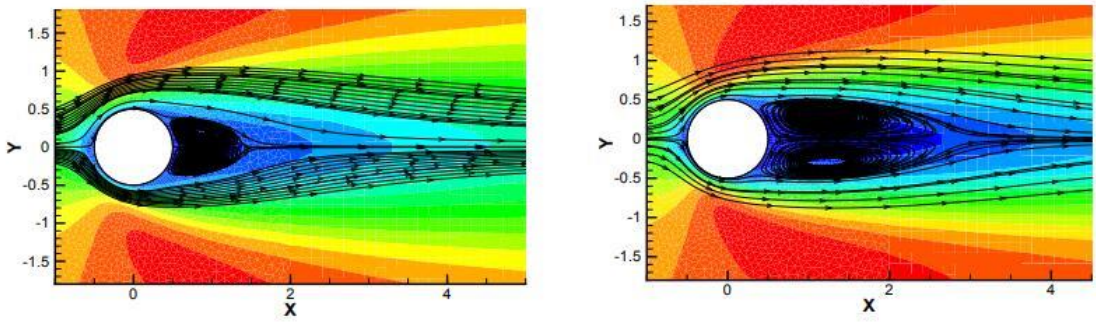
(d) Streamlines of $P_1P_2$(HLSr) solution

Fig.3 Results for the lid-driven cavity flow at Re=10,000.

*Steady flow over a circular cylinder*

A flow past a circular cylinder at a Reynolds number of 20 and 40 respectively based on a uniform free-stream velocity $\mathbf{u} = (1, 0)$ with no-slip boundary conditions on the cylinder surface is considered in this case. At both of these two Reynolds numbers, the flows are laminar and steady and were studied quite extensively in both measurements and numerical calculations.

Fig.4 shows the streamlines and the vortex behind the cylinder computed by the $P_1P_2$(HLSr) method at Re = 20 and Re = 40, respectively. It is clear to see that a pair of stationary recirculating regions appears in the wake of the cylinder for each condition and the length of the recirculating region increases with the Reynolds number.




(a) $Re=20$ (b) $Re=40$

Fig.4 Streamlines plot of steady flow past around a circular cylinder based on $P_1P_2$(HLSr).

Next, we calculate the friction and pressure drag coefficients, the total drag coefficients, the front and rear stagnation pressure coefficients, and recirculation lengths obtained for the steady flows at Re=20 and 40, respectively. The results are summarized in Tab.3.

Tab.3 Comparison of results for steady flow past a circular cylinder.

| Source | $C_{DF}$ | $C_{DP}$ | $C_D$ | $C_p(0)$ | $-C_p(\pi)$ | $L_w/D$ |
|---|---|---|---|---|---|---|
| $Re=20$ | | | | | | |
| R.P. Bharti et al. | 0.8211 | 1.2244 | 2.0455 | 1.2889 | 0.5457 | 0.9164 |
| DG($P_1$) | 0.7882 | 1.2241 | 2.0132 | 1.2918 | 0.5505 | 0.9301 |
| DG($P_2$) | 0.8078 | 1.2257 | 2.0336 | 1.2956 | 0.5492 | 0.9051 |
| $P_1P_2$(HLSr) | 0.8074 | 1.2256 | 2.0330 | 1.2943 | 0.5496 | 0.9051 |
| $Re=40$ | | | | | | |
| R.P. Bharti et al. | 0.5316 | 0.9976 | 1.5292 | 1.1636 | 0.4798 | 2.2252 |
| DG($P_1$) | 0.5072 | 0.9959 | 1.5031 | 1.1649 | 0.4821 | 2.0507 |
| DG($P_2$) | 0.5234 | 0.9961 | 1.5196 | 1.1674 | 0.4804 | 2.1233 |
| $P_1P_2$(HLSr) | 0.5232 | 0.9962 | 1.5195 | 1.1663 | 0.4810 | 2.2685 |

An excellent correspondence can be seen to exist between the present and literature results which demonstrate that our method can provide an attractive alternative for solving the INS equations on arbitrary grids.

# Blood flow simulation of left ventricle and aorta with translation motion

***Masashi Yamakawa[1], Yuto Yoshimi[1], Shinichi Asao[2] and Kyohei Tajiri[1]**

[1]Faculty of Mechanical Engineering, Kyoto Institute of Technology, Japan.
[2]Department of Mechanical Engineering, College of Industrial Technology, Japan

*Presenting and corresponding author: yamakawa@kit.ac.jp

## Abstract

To specify causes of heart diseases, it is very important to understand a blood flow state in an aorta. In this paper, the blood flow which pushed out to the aorta according to a contraction motion of a left ventricle was simulated. To express its complicated shape and the motion, the unstructured moving grid finite volume method was adopted. In this method, the control volume is defined for a space time unified domain. Not only a physical conservation law but also a geometric conservation law is satisfied in this approach. Then high accurate computation is conducted under the method. The left ventricle expands and contracts, at the same time, the ventricle and the aorta perform a translational motion. The model of its motion captured from the computed tomography images is also introduced to this computation. The result of flow calculation in left ventricle matches with the measurement result qualitatively. A flow in an aorta has a dramatic shift on its style in the contraction process of left ventricle. We also succeeded to capture its shift on our result of the flow in the aorta. Then, the tendency of the flow also matches with the computation and measurement result of others. Furthermore, the complicated vortex structure in the left ventricle was shown as the results of the simulation. Thus, the validity of the computational method and the possibility of calculation for capturing detail flows in left ventricle and aorta were shown in this paper.

**Keywords:** Computational fluid dynamics, Blood flow simulation, Left ventricle, Aorta

## Introduction

There are a lot of threat serious or life threatening disease in a heart and vascular diseases, for example arteriosclerosis or an aneurysm. It is known from clinical observation that the heart and vascular diseases often appear at bifurcation or flexure of thick blood vessel. Thus, the relation between the origin of the heart disease and blood flow has been pointed out. Then, various hypotheses regarding the fluid dynamics factor of the heart and vascular diseases have been made. To prove the validity of the hypotheses, flows in a heart or blood vessel have been studied through the method of experimentation or numerical simulation. Ku et al. [1] made a measurement the intimal thickening generated at the branching part of the human arteria carotis communis, as they were focused on the relations between the intimal thickening of an artery and the blood flow. Then, it was shown that the intimal thickening has a correlation with the time fluctuation of shear force measured on a glass tube flow made from specimens of blood vessel. Fukushima et al. [2] created a visualization of blood flow using the real blood vessel taken out from the body. The real blood vessel is made transparent by salicylic acid. Then, they determined whether vortex tube exist at the bifurcation of the blood vessel. While, by multi scale computing using the finite element method, Sugiura et al. [3] created the numerical heart in the supercomputer although it took a huge cost.

And now, the aorta connected to the left ventricle is comprised of three parts as the aorta ascendens expanding upward, the aortic arch taking a bend, and the aorta descendens expanding downward. Then, the three principal branched blood vessels expand from the

aortic arch. The blood current in the aorta behaves like an intermittent flow which alternates start and stop flow, because the aortic valve located between the aorta and the left ventricle has to alternate close and open to occur the flow from the left ventricle to the aorta. The complicated flow phenomenon in the aorta is created from the geometric feature and the pulsatility of the aorta. Then, it is very interesting from the hemodynamic standpoint regarding the relation between the origin of the heart disease and the factor of the fluid dynamics. It is at an increased risk for developing of the disease at the left ventricle and the aorta. Furthermore, it would become more serious when it develops. From the point of the view, a lot of researches [4] of the left ventricle and the aorta have been conducted. The objective of our paper is to develop an efficient computation method for flows of the left ventricle and the aorta with satisfying the expression of the complicated shapes and the function. To calculate more accurately, not only expansion and contraction of the left ventricle but also translational motion of the aorta is adopted as the motion for computation. In particular, to satisfy a physical conservation law and a geometric conservation law, the unstructured moving grid finite volume method [5][6] is adopted. In this method, a control volume is defined for a space time unified domain. The method made it possible to compute accurately for motion of the left ventricle and the aorta. Furthermore, the unstructured mesh approach was also able to express such the complicated shape. Then, the computation was carried out under the OpenMP parallel environment [7].

## Numerical Approach

*Governing Equations*

As governing equations, the continuity equation and the Navier-Stokes equations for incompressible flows are adopted and written as follows:

$$\nabla \cdot \mathbf{q} = 0, \tag{1}$$

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{E}_a}{\partial x} + \frac{\partial \mathbf{F}_a}{\partial y} + \frac{\partial \mathbf{G}_a}{\partial z} = -\left( \frac{\partial \mathbf{E}_p}{\partial x} + \frac{\partial \mathbf{F}_p}{\partial y} + \frac{\partial \mathbf{G}_p}{\partial z} \right) + \frac{1}{\mathrm{Re}} \left( \frac{\partial \mathbf{E}_v}{\partial x} + \frac{\partial \mathbf{F}_v}{\partial y} + \frac{\partial \mathbf{G}_v}{\partial z} \right), \tag{2}$$

where $\boldsymbol{q}$ is the velocity vector, $\boldsymbol{E}_a$, $\boldsymbol{F}_a$, and $\boldsymbol{G}_a$ are advection flux vectors in the $x$, $y$, and $z$ direction, respectively, $\boldsymbol{E}_v$, $\boldsymbol{F}_v$, and $\boldsymbol{G}_v$ are viscous-flux vectors, and $\boldsymbol{E}_p$, $\boldsymbol{F}_p$, and $\boldsymbol{G}_p$ are pressure terms. The elements of the velocity vector and flux vectors are

$$\begin{aligned} &\mathbf{q} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \ \mathbf{E}_a = \begin{bmatrix} u^2 \\ uv \\ uw \end{bmatrix}, \ \mathbf{F}_a = \begin{bmatrix} uv \\ v^2 \\ vw \end{bmatrix}, \ \mathbf{G}_a = \begin{bmatrix} uw \\ vw \\ w^2 \end{bmatrix}, \ \mathbf{E}_p = \begin{bmatrix} p \\ 0 \\ 0 \end{bmatrix}, \\ &\mathbf{F}_p = \begin{bmatrix} 0 \\ p \\ 0 \end{bmatrix}, \ \mathbf{G}_p = \begin{bmatrix} 0 \\ 0 \\ p \end{bmatrix}, \ \mathbf{E}_v = \begin{bmatrix} u_x \\ v_x \\ w_x \end{bmatrix}, \ \mathbf{F}_v = \begin{bmatrix} u_y \\ v_y \\ w_y \end{bmatrix}, \ \mathbf{G}_v = \begin{bmatrix} u_z \\ v_z \\ w_z \end{bmatrix}, \end{aligned} \tag{3}$$

where $u$, $v$, and $w$ are the velocity components of the $x$, $y$, and $z$ directions, respectively, and $p$ is pressure. The subscripts $x$, $y$, and $z$ indicate derivatives with respect to $x$, $y$, and $z$, respectively. Here, Re is the Reynolds number.

*The Unstructured Moving-Grid Finite-Volume Method*

In this simulation, expansion and contraction of the left ventricle and translation motion of the aorta are expressed using moving mesh approach. To assure a geometric conservation law in moving mesh, a control volume is defined in a space-time unified domain. For the discretization, Eq. (2) can be written in divergence form as

$$\widetilde{\nabla}\cdot\widetilde{\mathbf{F}}=\mathbf{0}, \tag{4}$$

where

$$\widetilde{\nabla}=\begin{bmatrix}\dfrac{\partial}{\partial x}\\ \dfrac{\partial}{\partial y}\\ \dfrac{\partial}{\partial z}\\ \dfrac{\partial}{\partial t}\end{bmatrix},\quad \widetilde{\mathbf{F}}=\begin{bmatrix}\mathbf{E}_a+\mathbf{E}_p-\dfrac{1}{\mathrm{Re}}\mathbf{E}_v\\ \mathbf{F}_a+\mathbf{F}_p-\dfrac{1}{\mathrm{Re}}\mathbf{F}_v\\ \mathbf{G}_a+\mathbf{G}_p-\dfrac{1}{\mathrm{Re}}\mathbf{G}_v\\ \mathbf{q}\end{bmatrix}. \tag{5}$$

The flow variables are defined at the center of the cell in the ($x$, $y$, $z$) space, as the approach is based on a cell-centered finite volume method. Thus, the control volume becomes a four-dimensional polyhedron in the ($x$, $y$, $z$, $t$)-domain. For the control volume, Eq. (4) is integrated using the Gauss theorem and written in surface integral form as:

$$\int_{\widetilde{\Omega}}\widetilde{\nabla}\cdot\widetilde{\mathbf{F}}\mathrm{d}\widetilde{V}=\oint_{\partial\widetilde{\Omega}}\widetilde{\mathbf{F}}\cdot\widetilde{\mathbf{n}}_u\mathrm{d}\widetilde{S}\approx\sum_{l=1}^{6}\left(\widetilde{\mathbf{F}}\cdot\widetilde{\mathbf{n}}\right)_l=\mathbf{0} \tag{6}$$

Here, $\widetilde{\mathbf{n}}_u$ is an outward unit vector normal to the surface, $\partial\widetilde{\Omega}$, of the polyhedron control volume $\widetilde{\Omega}$, and $\widetilde{\mathbf{n}}=(\widetilde{n}_x,\widetilde{n}_y,\widetilde{n}_z,\widetilde{n}_t)_l$, ($l$=1, 2,... 6) denotes the surface normal vector of control volume, and its length is equal to the boundary surface area in four-dimensional ($x$, $y$, $z$, $t$) space. The upper and bottom boundary of the control volume ($l$ = 5 and 6) are perpendicular to the $t$-axis, and therefore they have only the $\widetilde{n}_t$ component, and its length corresponds to the volume of the cell in the ($x$, $y$, $z$)-space at time $t^n$ and $t^{n+1}$, respectively.

## Computational Model and Conditions

*Geometric Model of Left Ventricle and Aorta*

The function of the left ventricle is draining blood to the aorta like a pomp. The mitral valve and the aortic valve are put on the inlet and the outlet of the ventricle, respectively. The shape of the left ventricle is structured, as shown Fig. 1. Bothe of the diameter of blood vessels at the mitral valve and the aortic valve are 3.0cm. The length from the base of heart to the cardiac apex is 7.8cm at lumen maximum volume. The cross-section shape of the left ventricle is ellipse. Then, the ratio of the major axis and minor axis on the ellipse is 5 to 4.
While, the aorta is comprised of three parts which are the ascending aorta expanding upward, the aortic arch taking a bend, and the descending aorta expanding downward. Furthermore, the three principal branched blood vessels which are called innominate artery, left common carotid artery and left subclavian artery expand from the aortic arch. Then the aortic arch itself curves three-dimensionally. In other words, the central axis of the aortic arch in not on a plane surface. Thus, the aorta is complicated shape with bending, bifurcation and three-

dimensional torsion. In this paper, the shape of the aorta model is created, as shown Fig. 2. Then, Fig. 3 shows aortic arch with three-dimensional curve around the bifurcation points.

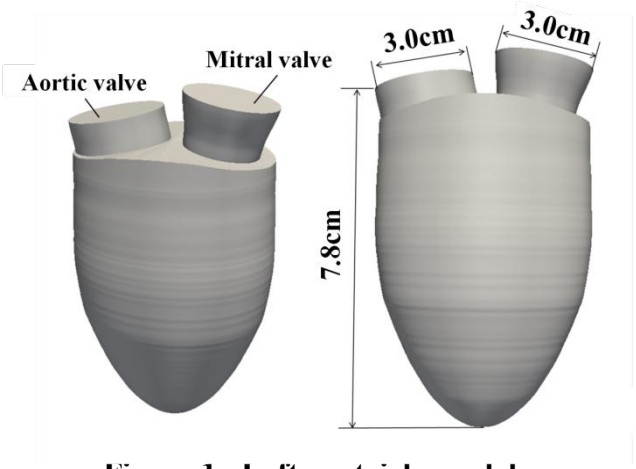



**Figure 1. Left ventricle model**

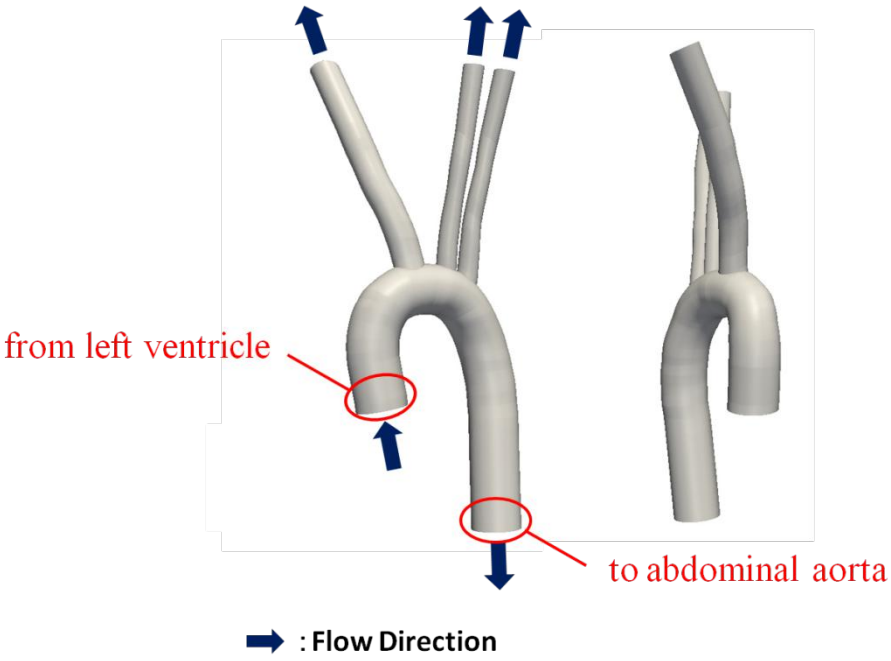



**Figure 2. Aorta model**

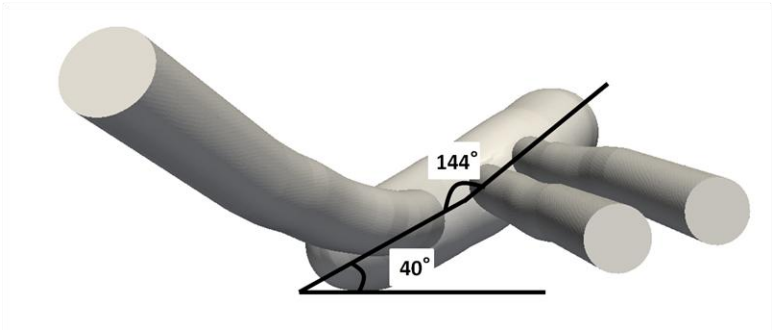



**Figure 3. Angle of torsion at aortic arch from top view**

*Motions of Left Ventricle and Aorta*

The left ventricle is draining blood to the aorta by expansion and contraction. Then the heart rate is determined the systole and diastole of the heart. Then, a period from starting point to the next starting point of heart rate is called the cardiac cycle. If a pulse rate is 60bpm, one cardiac cycle would be 1.0sec. Then, it is classified 0.49sec as the systole and 0.51sec as the diastole. The history of the left ventricle cavity volumetric change in one cardiac cycle is shown in Fig. 4. The expansion and contraction using moving mesh at the simulation are expressed under the history.

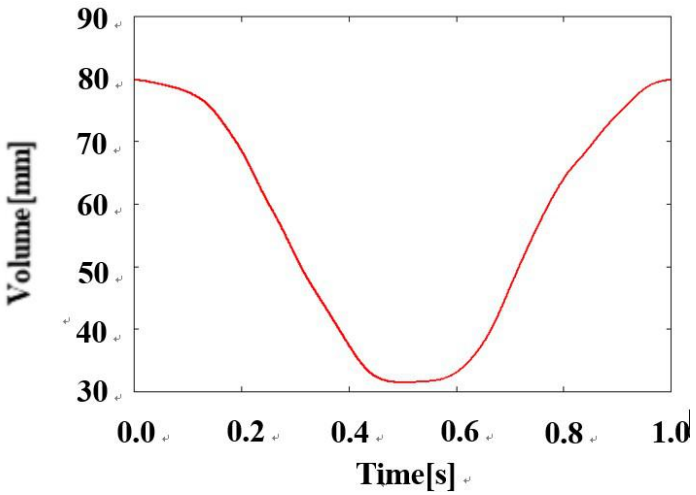



**Figure 4. History of left ventricle volumetric change**

Fig. 5 shows the computed tomography images of the left ventricle and the aorta [8]. The dark black line is a catheter injecting a contrast medium. The figure on the left is a front view of a human and the figure on the right is a side view.

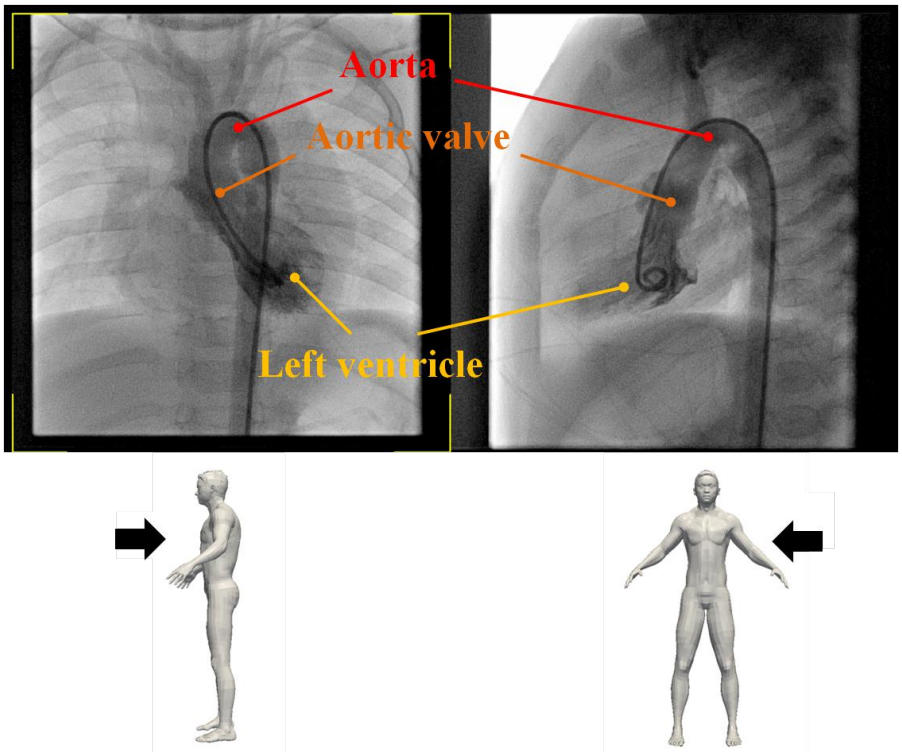



**Figure 5. CT images of left ventricle and aorta (left: front view, right; side view)**

Fig. 6 shows the CT images at the maximum volume of the left ventricle and Fig. 7 shows at the minimum one. On both figures, the red line indicates the point of the aortic valve at the maximum one. Then, the blue line indicates at the minimum one. From these figures, the translation motion of the position of the aortic valve is confirmed from the figures. Then, we can see the motion within a broader range from the left ventricle to the descending aorta. The motion can affect to the blood flow since the acceleration of the motion is relatively high. Thus, in this computation, the motion is introduced.

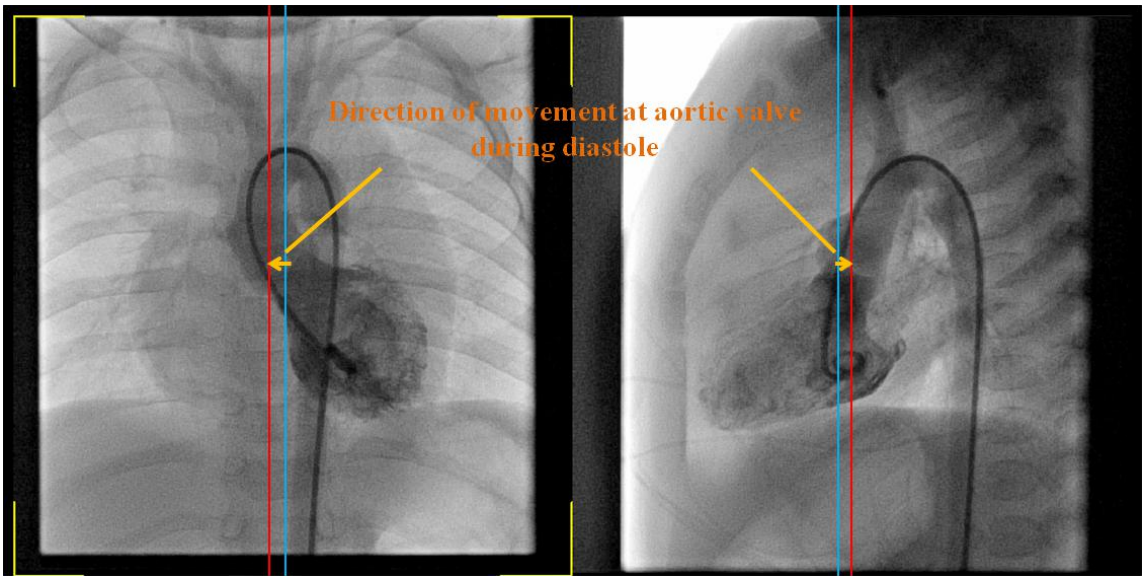



**Figure 6. CT image at maximum volume of ventricle (left: front view, right; side view)**

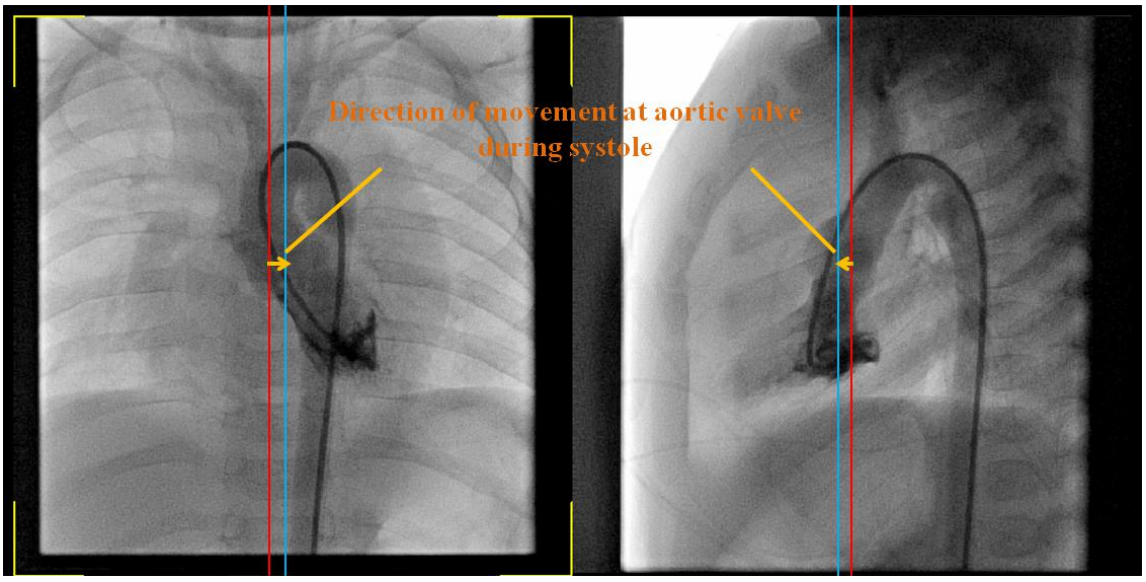



**Figure 7. CT image at minimum volume of ventricle (left: front view, right; side view)**

To know the length of the translation motion, the CT images are used. Several points are put on the images as shown in Fig. 8. These points correspond to the points at the computational model as shown in Fig. 9. The detail travel lengths are calculated using the height of the left ventricle as reference length L. Conducting the measurement by two aspects, the translation motion is estimated as three-dimensional movement.

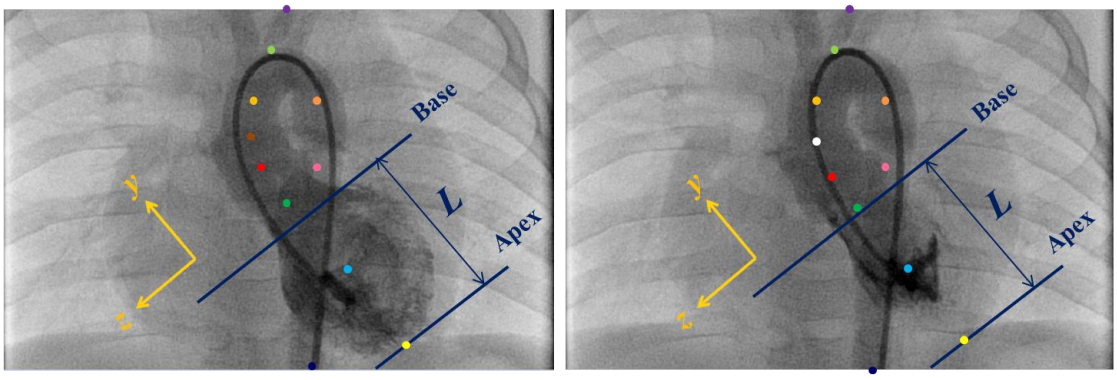



**Figure 8. Measurement length of translation motion from CT image (left: at maximum volume, right; at minimum volume)**

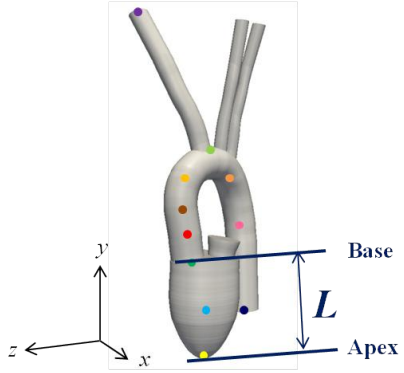



**Figure 9. Corresponding points on computational model**

*Computational Conditions*

The computational mesh is generated by MEGG3D [9] using tetrahedral and prism elements. The total number of the elements are 2,777,089. The heart rate is 60bpm, and the Reynolds number is 2,030. As an initial condition, pressure $p = 0$ and velocity for $x$, $y$, $z$ directions $u = v = w = 0$ are obtained for all elements. In the cardiac diastole, the mitral valve is open and the aortic valve is closed completely. Then, the velocity at the mitral valve is given as a linear extrapolation and pressure is fixed as $p = 0$. While, in the cardiac systole, the mitral valve is closed and the aortic valve is open completely. These open and closing motion are conducted instantly. On the four exit of blood vessels, velocity is determined as a linear extrapolation and pressure is zero. The velocity on all walls of the left ventricle and the aorta is given the moving velocity decided expansion, contraction and translation motion.

## Computational Results

*Verification of the validity for the Computation*

To verify the validity for the computational approach, results of the flows inside the left ventricle and the aortic arch are evaluated. Fig. 10 shows streamlines in the left ventricle at the third cardiac diastole from starting this calculation. By expansion of the left ventricle, the blood inflow through the mitral valve is seen. Then, we can confirm that the flow makes two vortexes. One is generated on the center of the left ventricle, which is larger than another one. Another is created between mainstream and the wall. The relatively large vortex is also seen in the measurement results of flow in the left ventricle by Kilner et al. as shown in Fig. 11.

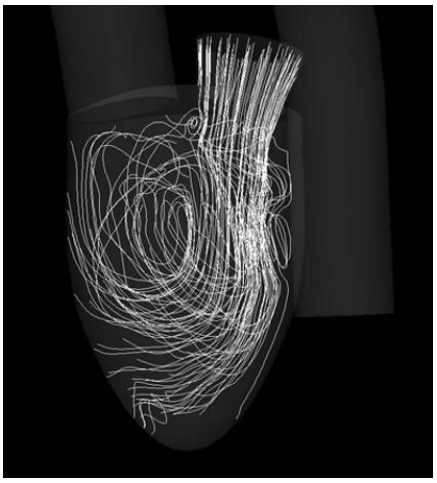


**Figure 10. Streamlines in left ventricle on this computation result**

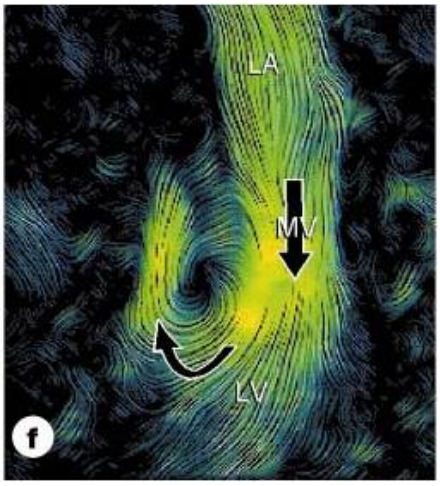



**Figure 11. Measurement result of flow in left ventricle**

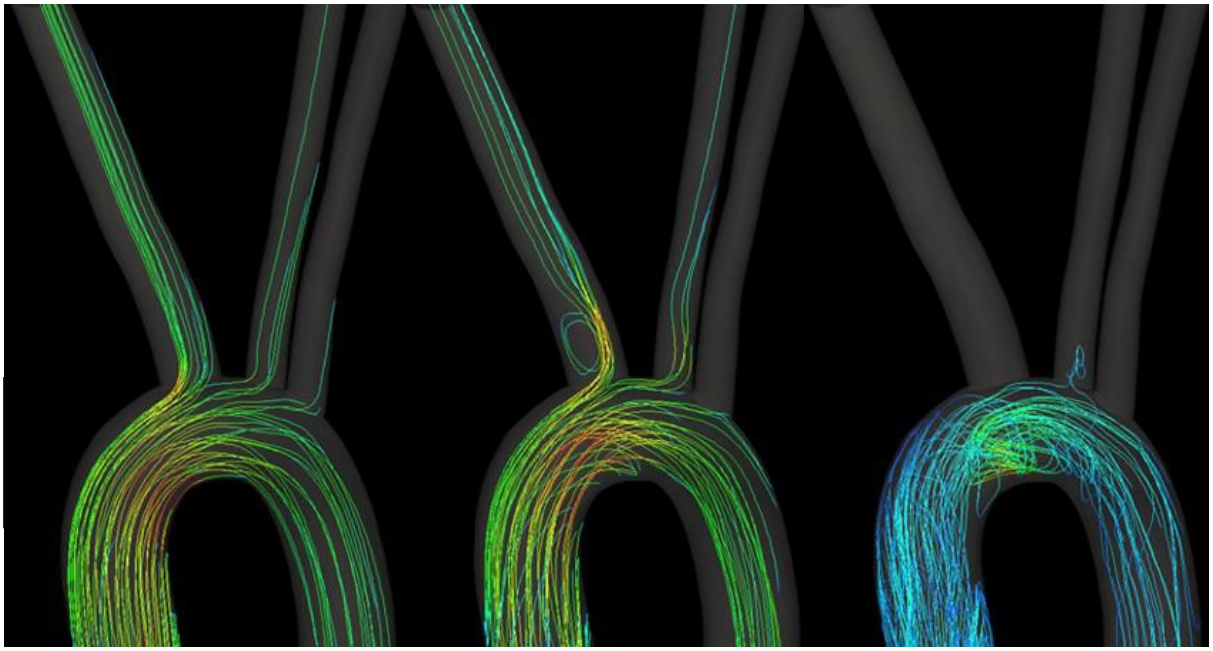


**Figure 12. Streamline in aorta of the computation**
**(left: in early systole, center: in mid to late systole, right: in end systole)**

Fig. 12 shows streamlines in the aorta at the third cardiac systole from starting this calculation. In this figure, left one is in early systole, centered one is in mid to late systole, and right one is in end systole. In the early systole, a strong flow along the aortic arc from the left ventricle is seen. In the mid to late systole, a spiral flow along the mainstream is confirmed. Then, small vortex is generated at the entrance of the innominate artery. In the end systole, mainstream itself becomes weak and circulating flow along the aortic wall. Then, we can hardly see flows in the tributary. These results are also compared with other computation and measurement results. The computational results by Wada are shown in Fig. 13, and the measurement results by Kilner et al. are shown in Fig. 14. The tendencies of our computation results on each systole are seen in the other computation and measurement results.
Comparing with other computation and measurement results of flows inside the left ventricle and the aorta, the qualitative correspondences are seen in both case. Thus, the validity of the computational approach is confirmed.

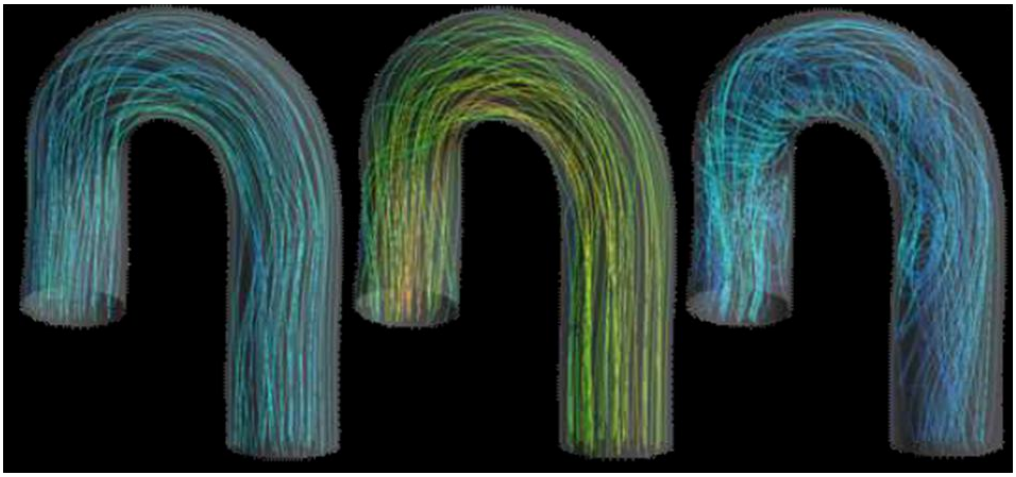


**Figure 13. Streamline in aortic arch computed by Wada**
**(left: in early systole, center: in mid to late systole, right: in end systole)**

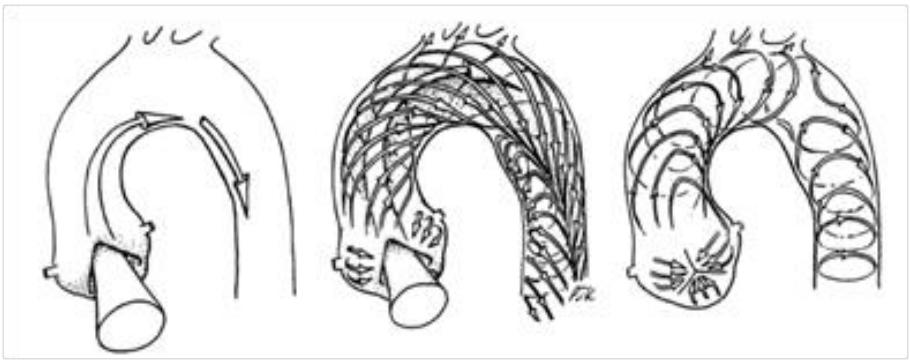
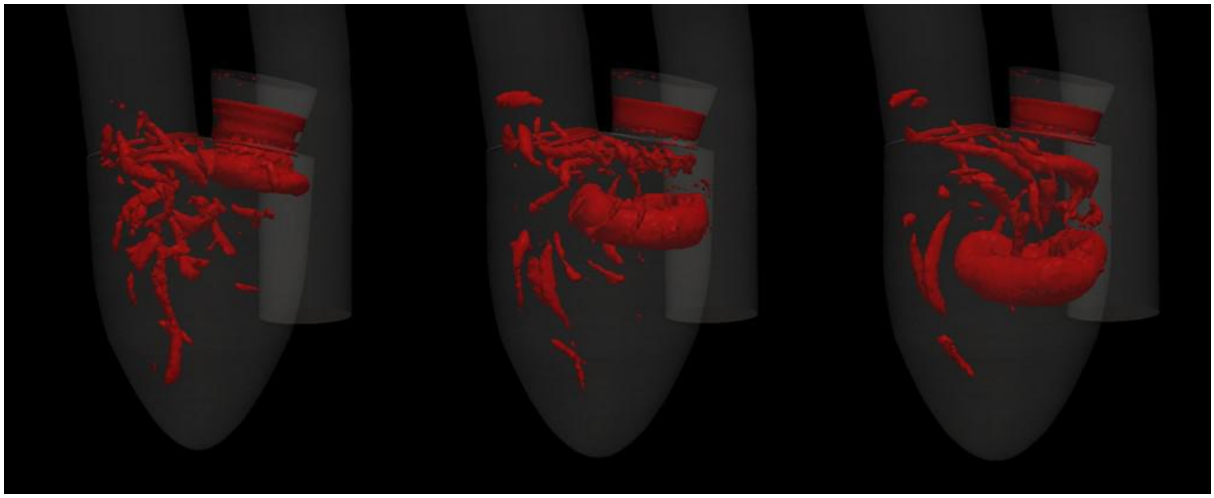


**Figure 14. Measurement results in aortic arch by Kilner**
**(left: in early systole, center: in mid to late systole, right: in end systole)**

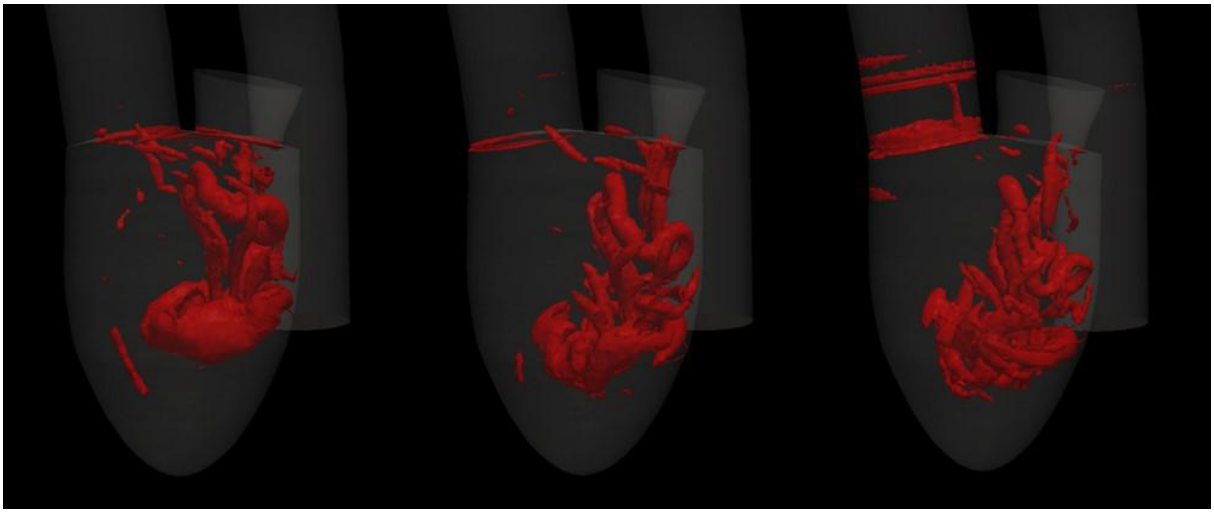
*Blood Flow in Left Ventricle*

The isosurface of Q criterion in the left ventricle at $t$ = 22.0, 23.0, 24.0, 25.0, 26.0 and 27.0 are shown in Fig. 15. These are from in the third early systole to the third in the mid to late systole. Vortex structures generated by the cardiac beat from the second period remain in the left ventricle. Into the domain, the inflow of blood with generating the ring-shaped vortex tube from the mitral valve is seen. The vortex tube is collapsing according to a decrease in the inflow of blood from the mitral valve. Then it becomes a complicated vortex structure and

spreads in the left ventricle. Thus, the possibility of calculation for capturing detail flows in left ventricle is shown.



$t = 22.0$ $t = 23.0$ $t = 24.0$



$t = 25.0$ $t = 26.0$ $t = 27.0$

**Figure 15. Isosurface of Q criterion in left ventricle ($t$ = 22.0 to 27.0)**

## Conclusions

To construct the computational approach for specifying cause of heart diseases, blood flows in the left ventricle and the aorta were calculated. For high order accurately, the unstructured moving-grid finite-volume method was adopted. Furthermore, not only the expansion of contraction of the left ventricle but also translation motion of the aorta which is captured from the computed tomography images was adopted in this paper. From the results of the flow in the left ventricle compared with measurement result, the large vortex is seen. Furthermore, the flows in the aorta were estimated using the other computation and measurement results. As the tendencies on each systole are seen in both results, the qualitative correspondences are confirmed. Thus, the validity of the computational approach is shown. In the computation for blood flows inside of the left ventricle at the diastole, complicated vortex structures are captured by the approach. Thus, the possibility to computed and specify the cause of the diseases was shown.

## Acknowledgments

This publication was subsidized by JKA through its promotion funds from KEIRIN RACE and by JSPS KAKENHI Grant Number 16K06079.

### References

[1] Ku, D., Giddens, et al, (1985) Pulsatile flow and atherosclerosis in the human carotid bifurcation. Positive correlation between plaque location and low oscillating shear stress., *Arteriosclerosis* Vol.5, 293-302
[2] Fukushima, T. et al. (1982) The horseshoe vortex: A secondary flow generated in arteries with stenosis, bifurcation, and branching, *Biorheology* Vol.19, 143-154.
[3] Sugiura, S., et al. (2012) Multi-scale simulations of cardiac electrophysiology and mechanics using the University of Tokyo heart simulator, *Progress in Biophysics and Molecular Biology*, 2012 Oct-Nov;110(2-3), 380-389
[4] Liang, F., et al. (2007) A multi-scale computational method applied to the quantitative evaluation of the left ventricular function, *Computers in Biology and Medicine*, Vol.37, 700-715
[5] Yamakawa, M., et al. (2012) Numerical Simulation for a Flow around Body Ejection using an Axisymmetric Unstructured Moving Grid Method, *Computational Thermal Sciences*, Vol.4, No.3, 217-223.
[6] Yamakawa, M., et al. (2017) Numerical Simulation of Rotation of Intermeshing Rotors using Add-ed and Eliminated Mesh Method, *Procedia Computer Science* 108C, 1883-1892.
[7] Yamakawa, M., et al. (2011) Domain decomposition method for unstructured meshes in an OpenMP computing environment, *Computers & Fluids*, Vol. 45, pp.168-171.
[8] Fukui, T., et al. (2017) Influence of geometric changes in the thoracic aorta due to arterial switch operations on the wall shear stress distribution, *Open Biomedical Engineering Journal*, 11, 9-16.
[9] Ito, Y. (2013) Challenges in Unstructured Mesh Generation for Practical and Efficient Computational Fluid Dynamics Simulations, *Computers and Fluids*, 85, 47-52.

# Comparison of transport methodologies on near-field pollutant dispersion in urban environments using CFD

**†*Clarence TEE[1], E.Y.K. NG[1] and George XU[2]**

[1]School of Mechanical & Aerospace Engineering, Nanyang Technological University, Singapore.
[2]Fluid Dynamics Department, Institute of High-Performance Computing, A*STAR, Singapore.
†Corresponding author: clarenceteee@gmail.com

## Abstract

Until recently, the use of Computational Fluid Dynamics (CFD) appears to be gaining traction over traditional Gaussian Dispersion Modeling to predict and understand pollutant dispersion processes in urban environments. Gaussian Dispersion Models, while computationally fast, lacks in physical representation and accuracy but still sufficed as evidenced by its use in numerous engineering designs and research applications. In CFD, two typical methodologies namely the passive scalar transport and the multi-species transport are used to track spatial dispersion of pollutants. The focus of this study is to understand and quantify the differences between the two models when applied to near-field dispersion of heavy gases. The two methodologies are validated by simulating the dispersion phenomena for two test cases which largely bears resemblance to urban settings: a three-dimensional street canyon setup and the Mock Urban Setting Test (MUST) field experiment. The pollutant used in the CODASC study is Sulphur Hexafluoride ($SF_6$) while Propylene ($C_3H_6$) is used in MUST, both heavier than air. It is found that numerical results are highly sensitive to the value of turbulent Schmidt number ($Sc_t$) in both test cases. Through parametric studies, the best accuracy is attained when the $Sc_t$ value of 0.5 is used in the street canyon case and $Sc_t$ value of 1.5 for MUST. Generally, better agreement between numerical and experimental results is reflected in the street canyon case study compared to MUST. For the CODASC case study, the passive scalar transport model yielded better results than the multi-species transport model while the opposite is true for MUST. With the preconceived notion that the multi-species transport model should outperform the passive scalar transport, a conclusion showing a lack of significant improvement the former has over the passive scalar transport model is indeed unexpected. With regards to computational efficiency, the passive scalar transport model requires much lesser resources such as CPU time and memory compared to the multi-species transport model, thus making it more efficient.

**Keywords:** CFD, pollutant dispersion, near-field, urban environment modeling, turbulence modeling

## Introduction

Pollutant dispersion modeling holds significant importance when determining the severity of disasters, be it natural or man-made. Incidents such as the Fukushima Daiichi accident in 2011 or the more recent 2018 gas leak in Zhangjiakou, China have shown the devastating consequences on air quality through the release of toxic materials into the atmosphere. Even eight years after the disaster, atmospheric air quality around the immediate vicinity of Fukushima still contains lethal levels of radiation [1] and the effects of the incident in Zhangjiakou led to tragic consequences, due to the dispersion of a flammable pollutant which was ignited, leading to fatalities [2]. It is therefore of paramount importance to understand and assess these effects on public health and safety, immediately following an accident or to conduct risk management planning for pre-emptive purposes. The importance of understanding atmospheric dispersion processes is further emphasized through the span of research in developing state-of-the-art models from regulatory bodies such as the United States Environmental Protection Agency (US EPA) and UK Met Office to research organizations such as the European Cooperation in Science & Technology (COST). Many of these regulatory

bodies have developed their own atmospheric dispersion models: for instance, the open-source AERMOD was developed by US EPA and NAME by the UK Met Office [3]. In general, dispersion models are classified into three different families of models: Gaussian, Lagrangian and Eulerian. These models have been used to good effect in disasters such as the Eyjafjallajökull eruption and the Fukushima nuclear accident to provide crucial information for timely and preventive measures. Private research organizations have also funded initiatives to understand and improve the quality of models used to predict transport phenomena such as the COST Action 732.

However, effectiveness and reliability of these dispersion models are confined to certain spatial scales namely the meteorological mesoscale which would not be suitable for analyzing dispersion in the urban environment, the scale of which is two orders of magnitude smaller than meteorological mesoscale. Hence Lagrangian and Eulerian models will not work well for modeling dispersion on a microscale level. Gaussian-based models do have its merits when used on that spatial scale but the interaction between complex flow fields and different building configurations limit its accuracy significantly. Besides Gaussian-based models, microscale Computational Fluid Dynamics (CFD) are commonly used. The cost involving CFD simulations is relatively low, detailed information about the flow field can be obtained depending on mesh resolution and scaling of simulations can be carried out with ease [4]. But since errors are introduced in each progressive stage of CFD modeling, accuracy and reliability of results are often questioned which require validation studies.

*Gaussian Dispersion Modeling*

The underlying equation of Gaussian-based models is given by:

$$c(x,y,z) = \frac{Q}{2\pi\sigma_y\sigma_z u}\exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left(\exp\left(-\frac{(z-H_e)^2}{2\sigma_z^2}\right)+\exp\left(-\frac{(z+H_e)^2}{2\sigma_z^2}\right)\right) \quad (1)$$

where $c$ is the concentration of a pollutant at a given location, $Q$ is the rate of pollutant emission, $H_e$ is the effective height of release which is the sum of actual stack height $H_p$ and plume rise $\Delta h$, $u$ is the speed of wind in the $x$-direction at height $H_e$. $\sigma_y$ and $\sigma_z$ are standard deviations of the pollutant concentration profile in the $y$ and $z$ direction respectively, both represented by a Gaussian distribution which depends heavily on atmospheric turbulence. Comprehensive experimental measurements have been carried out to relate various values of $\sigma$ to atmospheric stability. The resulting tabulated data is known as the Pasquill-Gifford-Turner Stability Classifications [5]. Stratification of the atmosphere into different stability classes is crucial to model the atmospheric boundary layer accurately. The boundary layers are differentiated into three main types: unstable, neutral and stable.

The popularity of Gaussian models is largely due to its low computational costs, with many environmental regulatory agencies using it as an initial assessment in determining the severity of industrial accidents or pollution levels. Complex dispersion related phenomena can be included which adds an edge of versatility to Gaussian models. However, its validity and accuracy are dampened by underlying assumptions used in the derivation of Eq. (1). Velocities in the $y$ and $z$ direction are assumed to be zero and diffusion in the $x$ direction is ignored. Neglecting diffusion in the $x$ direction can lead to inaccurate results especially in situations where there are low wind speeds which causes significant concern. Atmospheric turbulence is also assumed to be uniform and homogenous since $\sigma$ is directly proportional to sizes of eddies. Furthermore, $\sigma$ values are measured from rural terrains with flat and open surfaces, which reduce its applicability when the model is used in an urban environment [6, 7]. A sensitivity study on the parameters of the Gaussian model by Adel [6] showed that changes as small as

10% could result in 100% under or over prediction. Hence using Gaussian models for dispersion in urban environments raises questionable issues in accuracy and applicability.

Within urban environments, buildings and obstacles as well as their complex interaction with flow fields bring additional challenges in dispersion modeling. Near-field dispersion is characterized by interactions between atmospheric boundary layer flows and flow structures around buildings while in far-field dispersion, the horizontal motion proves to be more dominant over vertical motions and effects of buildings on flow fields are limited [8]. More specifically in near-field dispersion, there are features such as a fully three-dimensional flow structure around buildings which affects pollutant transport in ways that could not be fully described by Gaussian models. Flow separation, recirculation and various patterns of vortices (e.g. horseshoe vortices, vortex shedding etc.) generated by the presence of buildings adds further complication in determining pollutant distribution at the desired location.

*Computational Fluid Dynamics (CFD)*

Interest in CFD to model pollutant dispersion in urban environments is not newly found as seen by the review of Tominaga and Stathopoulos [8], which offered a comprehensive compilation of current modeling techniques. The use of CFD has allowed the approximation of flow field quantities such as velocities and concentration to be made known throughout the computational domain in all three-dimensions, which Gaussian-based models lack. However, CFD is not without uncertainties. The validation and verification of CFD models constitute a major role in forming a quantitative conclusion on the accuracy and the reliability of results. Availability of experimental data while reassuring, is to be used with caution as circumstances surrounding the experiment must be replicated in the CFD model through initial and boundary conditions, failing which will render the comparison between numerical and experimental data pointless [8]. Inherent in such a strict requirement is the assumption that every experiment must be carried out under identical conditions. This hints at some form of replicability which is mostly impossible when meteorological factors like wind and weather conditions are involved [9]. Even under more predictable conditions such as wind tunnel testing, drawbacks still exist, nonetheless. Therefore, a certain pre-existing error must be accepted into the overall model evaluation.

The motivation behind this study is largely due to limited efforts thus far to compare the differences between passive scalar transport and multi-species transport. With the passive scalar transport model, the concentration of pollutant does not have any impact on the flow field. As such, flow properties remain unchanged even if a different pollutant is used. The pollutant used can be interpreted to be weightless and its transport is mainly governed by advection and mass diffusion. Using the multi-species transport model, on the other hand, will result in the flow field changing depending on the chemical species being transported. Since the mixing law takes effect, any change in composition in the mixture of species will affect the density of the resulting mixture. Gravitational effects are also included where physical characteristics of the pollutant will affect the flow field. The two species considered in this study will be the pollutant and air. Chemical reactions between the two species are however, ignored in this study.

A review by Lateb et al. [4] on CFD to predict dispersion in urban environments stated that a common assumption used by various studies is that the pollutants are assumed to be passive and subsequently, the effects physical characteristics of pollutant particles have on the flow field are often neglected. A study previously done by Gromke and Ruck [11] [12] investigated pollutant dispersion in a three-dimensional tree free street canyon and found that concentration values predicted by FLUENT gave good agreement to wind tunnel measurements. Species transport was utilized but they did not study the effects of a passive pollutant. Bekka et al. [13]

carried out a study based on the MUST case and found that agreement with full-scale experimental data [14] varied with distance from the source. Better agreement was observed closer to the source while more deviation was observed further from the source but there was no indication on how transport of the pollutant was modeled, raising some doubt over the quality of their findings. Tominaga and Stathopoulos [15] carried out a comparison between neutrally buoyant (passive) scalars and heavy gases on flow and concentration fields and found that prediction performance of heavy gases was worse than that for neutral gases, but their scope was limited to just one building configuration.

*Objective*

The objective of this study is to present and evaluate CFD methods for the dispersion of heavy gases in urban environments using the commercial code FLUENT. Two different methodologies: 1) passive scalar transport and 2) multi-species are proposed to track dispersion of pollutant particles. The two methodologies are validated through two test cases representative of urban environments: a three-dimensional street canyon and the Mock Urban Setting Test (MUST). First, the respective mathematical models are analysed. Following that, the characteristics of both case studies used for validation purposes are identified. Besides visual observation and comparison of numerical results with experimental data, statistical performance measures are subsequently introduced to provide a succinct approach in quantifying the performance of each methodology. Results and discussion of the two methodologies are subsequently given, where the effects $Sc_t$ values have on concentration distributions are discussed. A comparison between both methodologies is rendered and their accuracy is evaluated.

**Fundamental Mathematical Models**

The Standard Gradient Diffusion Hypothesis (SGDH) is primarily used to model the turbulent scalar fluxes, which are postulated to be directly proportional to concentration gradients with a coefficient of proportionality known as turbulent diffusivity, $D_t$:

$$u'c' = D_t \frac{\partial \bar{c}}{\partial x} \tag{2}$$

$$v'c' = D_t \frac{\partial \bar{c}}{\partial y} \tag{3}$$

$$w'c' = D_t \frac{\partial \bar{c}}{\partial z} \tag{4}$$

Turbulent diffusivity can be further broken down into a ratio of two terms: turbulent kinematic viscosity ($\nu_t$) over the turbulent Schmidt number ($Sc_t$). It should be noted that $Sc_t$ is a free parameter, with multiple ranges of values that depend on flow fields and configurations of geometries with which the flow interacts with. This seems to suggest that $Sc_t$ can be altered to artificially increase or reduce turbulent diffusion which can drastically affect transport phenomena and subsequently, numerically predicted concentration.

Tominaga and Stathopoulos [17] showed that $Sc_t$ ranged from 0.2 to 1.3 which differed from the commonly used values of 0.7 to 0.9. Gualtieri et al. [18] discovered in his analysis of literature from both water and atmospheric systems that the range of best-fitting $Sc_t$ changes for different cases and that there were cases where best-fit $Sc_t$ value was found to be the same for extremely different flow conditions (i.e. cases of water and atmospheric systems having similar $Sc_t$ values). These two inconsistent observations contributed to a conclusion that it is impossible to determine a generic $Sc_t$ value. Gualtieri et al. subsequently questioned if $Sc_t$ had

different values in the same flow domain but was inconclusive as to which parameters controlled this variability.

The relevant governing equations begin with the continuity equation:

$$\nabla \cdot (\rho \vec{v}) = 0 \tag{5}$$

Reynolds Averaged Navier-Stokes, also known as the momentum equations:

$$\rho(\vec{v} \cdot \nabla u) = -\frac{\partial p}{\partial x} + (\mu + \mu_t)\nabla \cdot (\nabla u) \tag{6}$$

$$\rho(\vec{v} \cdot \nabla v) = -\frac{\partial p}{\partial y} + (\mu + \mu_t)\nabla \cdot (\nabla v) \tag{7}$$

$$\rho(\vec{v} \cdot \nabla w) = -\frac{\partial p}{\partial z} + (\mu + \mu_t)\nabla \cdot (\nabla w) \tag{8}$$

$k$-$\epsilon$ turbulence model:

$$\nabla \cdot (\rho \vec{v} k) = \nabla \cdot \left(\mu + \frac{\mu_t}{\sigma_k}\nabla k\right) + P_k - \rho\varepsilon \tag{9}$$

$$\nabla \cdot (\rho \vec{v} \varepsilon) = \nabla \cdot \left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\nabla \varepsilon\right) + C_{\varepsilon 1} P_k \frac{\varepsilon}{k} - C_{\varepsilon 2}\rho \frac{\varepsilon^2}{k} \tag{10}$$

$$\mu_t = \rho C_\mu \frac{k^2}{\varepsilon} \tag{11}$$

$$P_k = \mu_t (\nabla \vec{v})^2 \tag{12}$$

Passive scalar transport equation:

$$\nabla \cdot (\vec{v} c) = \nabla \cdot \left[\left(D_c + \frac{\nu_t}{Sc_t}\right)\nabla c\right] + S_c \tag{13}$$

where $c$ is the concentration of the scalar, $D_c$ is laminar diffusivity and $S_c$ is the source term. The multi-species transport model also utilizes Eqns. (5) to (12) but continues to Eqn. (14) and (15):

$$\nabla \cdot (\rho Y_i \vec{v}) = -\nabla \cdot J_i + S_i \tag{14}$$

$$J_i = -\left(\rho D_i + \frac{\mu_t}{Sc_t}\right)\nabla Y_i \tag{15}$$

where $S_i$ is the source term for the $i^{th}$ species, $J_i$ is the diffusion flux of species $i$ which is due to concentration gradients, $D_i$ is the mass diffusion coefficient for species $i$ in the mixture and $Y_i$ is the mass fraction of species $i$. Eq. (14) and Eq. (15) solve for the mass fraction, $Y_i$ as each constituent of the mixture is represented with $i = 1, 2$ …etc.

**Case Studies**

*CODASC*

As part of a larger effort in air quality studies to quantify the effects of tree planting on dispersion of traffic emissions by the Karlsruhe Institute of Technology (KIT), a database named Concentration Data of Street Canyons (CODASC) was established from the results of wind tunnel experiments with varying parameters and street canyon/tree avenue configurations

[20]. For the scope of this study, it suffices to consider only a tree-free street canyon configuration in three dimensions. The street canyon model is scaled down to 1:150, with $L = 180\ m$, $H = W = 18\ m$. Four lines of tracer pollutant release, with the intention of emulating traffic exhaust, are placed in between the two building blocks and it exceeds the street canyon by about 10% on each side to consider pollutant release from traffic junctions. $SF_6$ was used as a pollutant in this case and the release is distributed through equally spaced perforations. Dimensions of computational domain and the street canyon are shown in Fig. 1. Experimental data for various scenarios was made available through the online database [20].

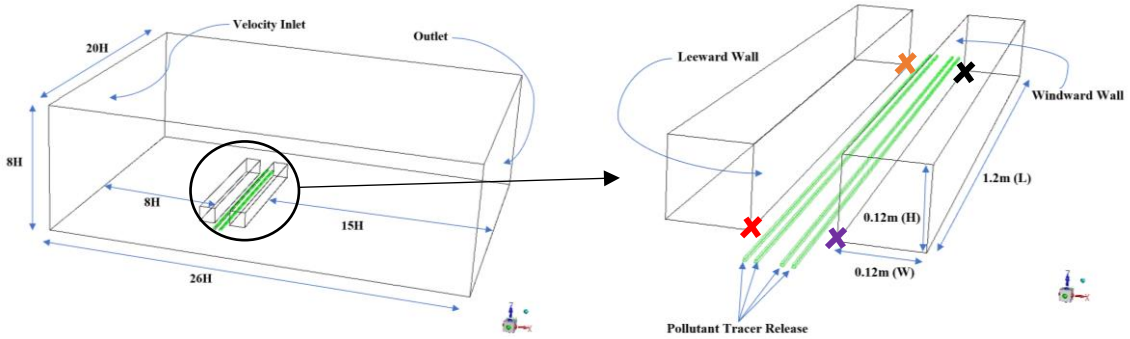



***Figure 1.*** *Dimensions of CODASC street canyon model.*

Inlet wind velocity is in the $x$-direction, perpendicular to the building blocks and it follows a power law profile in a neutrally stratified atmospheric boundary layer:

$$\frac{u(z)}{u(z_{ref})} = \left(\frac{z}{z_{ref}}\right)^{\alpha} \tag{16}$$

$z_{ref}$ is the reference height which is 0.12 $m$ and the flow velocity at the reference height, $u(z_{ref})$ is 4.70 $m/s$. Similarly, $u(z)$ is the velocity at a given height $z$. $\alpha$ is the wind shear exponent of 0.3. $k$ and $\epsilon$ profiles are specified as follows:

$$k = \frac{u_*^2}{\sqrt{C_\mu}}\left(1 - \frac{z}{\delta}\right) \tag{17}$$

$$\varepsilon = \frac{u_*^3}{\kappa z}\left(1 - \frac{z}{\delta}\right) \tag{18}$$

where $u_*$ is the friction velocity of 0.52 $m/s$ and $C_\mu$ is 0.09 which is the turbulence model constant. $\delta$ is the boundary layer thickness, $\kappa$ is von Kármán's constant and both have values of 0.96 and 0.4 respectively. The variable of interest is the non-dimensional concentration value $c^+$ which is normalized:

$$c^+ = \frac{cu_H H}{Q/l} \tag{19}$$

where $c$ is the measured concentration, $u_H$ is equivalent to $u(z_{ref})$ and $Q/l$ is the tracer source strength of $SF_6$ per unit length. The main point of interest is the normalized concentration distribution of the leeward and windward side of the street canyon.

*Mock Urban Setting Test (MUST)*

MUST is a full-scale, outdoor experiment free from laboratory limitations conducted at the U.S. Army Dugway Proving Ground (DPG) Horizontal Grid test site [14]. Motivation for MUST primarily grew out of a need for field data to be obtained outside laboratory conditions for the

verification and validation of models developed to understand dispersion mechanisms and the full effects of atmospheric boundary layers and surface roughness on transport phenomena in an urban setting. The setup configuration of MUST comprises of shipping containers each 12.2 $m$ long, 2.42 $m$ wide and 2.54 $m$ high arranged in a 12 by 10 array in a 200 $m$ squared area. With reference to the original report [14], trial name #2681829 is selected to be replicated in this present study by means of numerical methods. Propylene ($C_3H_6$) is used as a tracer gas and is released at a height of 1.8 $m$ above ground at a rate of 225 litres per minute and the location of release is marked by a cross in Fig. 2. 48 sensors were strategically placed at areas of interest to measure concentrations; 40 of which are distributed evenly across the array at a height of 1.6 $m$ above ground and are denoted by the orange points. The other 8 are placed 1, 2, 4, 6, 8, 10, 12 and 16 $m$ above ground level at a single location in the middle of the array, where the green point is. The dimensions of the computational domain are shown in Fig. 3.

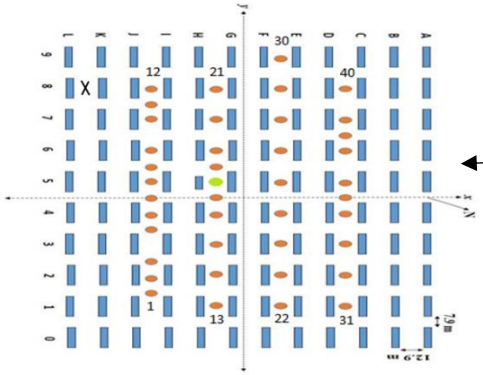



***Figure 2.*** *Layout of test case #2681829 and locations of sensors 1 to 48 (Image taken from [12]).*

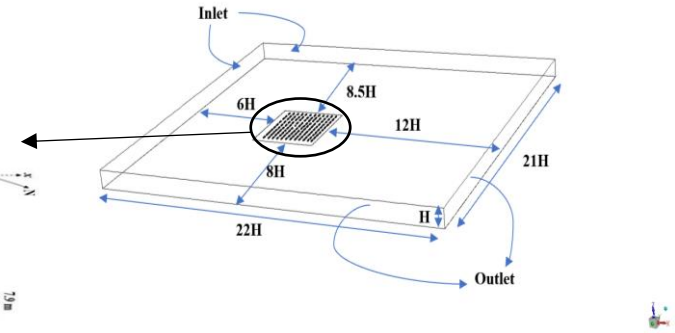



***Figure 3.*** *Dimensions of computational domain for MUST.*

Given the nature of MUST experiments (i.e. outdoor and full-scale), it is necessary to model the atmospheric boundary layer (ABL) to obtain accurate and reliable predictions of atmospheric-related phenomena [21]. More specifically, the modeled ABL must be horizontally homogenous which is achieved when inlet profiles of mean wind speed and turbulence quantities are in equilibrium with the wall functions used. With a horizontally homogenous ABL, streamwise gradients will be prevented from forming in these profiles as the flow progresses to simulate the phenomena of interest. This study focuses on a neutrally stratified ABL since thermal and buoyancy effects on turbulence are not considered. The most commonly used boundary conditions to simulate the ABL is first proposed by Richards & Hoxey [22]. For fully-developed, steady, incompressible and 2-dimensional ABL flows, the $k$-$\varepsilon$ turbulence model can be simplified to Eqs. (21) and (22) when Eq. (20) is assumed:

$$V = W = \frac{\partial U}{\partial x} = \frac{\partial k}{\partial x} = \frac{\partial \varepsilon}{\partial x} = 0 \tag{20}$$

$$\frac{\partial}{\partial z}\left(\frac{\mu_t}{\sigma_k}\frac{\partial k}{\partial z}\right) + G_k - \rho\varepsilon = 0 \tag{21}$$

$$\frac{\partial}{\partial z}\left(\frac{\mu_t}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial z}\right) + C_{\varepsilon 1}G_k\frac{\varepsilon}{k} - \rho C_{\varepsilon 2}\frac{\varepsilon^2}{k} = 0 \tag{22}$$

$$\text{where } G_k = \mu_t\left(\frac{\partial U}{\partial z}\right)^2$$

$$\kappa^2 = \sigma_\varepsilon(C_{\varepsilon 2} - C_{\varepsilon 1})\sqrt{C_\mu} \tag{23}$$

Inlet velocity, $k$ and $\varepsilon$ profiles proposed by Richards & Hoxey in 1993 follow the log-law profile as specified from Eqs. (24)-(26). It can subsequently be shown that Eqs. (24)-(26) are the analytical solutions to the $k-\varepsilon$ turbulence model of Eqs. (21) and (22) if Eq. (23) is satisfied. The constants are specified by Launder & Spalding [16], where $C_{\varepsilon 2} = 1.92$, $C_{\varepsilon 1} = 1.44$, $\kappa = 0.42$ and $C_\mu = 0.09$ which results in $\sigma_\varepsilon = 1.225$.

$$U(z) = \frac{u^*_{ABL}}{\kappa}\ln\left(\frac{z + z_0}{z_0}\right) \tag{24}$$

$$k(z) = \frac{(u^*_{ABL})^2}{\sqrt{C_\mu}} \tag{25}$$

$$\varepsilon(z) = \frac{(u^*_{ABL})^3}{\kappa(z + z_0)} \tag{26}$$

The freestream velocity of wind entering the computational domain has a magnitude and direction of 7.93 $m/s$ at a height of 4 $m$ above ground level and $-41°$ respectively. $u^*_{ABL}$ is the frictional velocity of 1.1$m/s$ and roughness length, $z_o$ has a value of 0.19368 $m$ when von Kármán's constant, $\kappa$ takes a value of 0.42. Concentration is measured in parts per million by volume (ppmv). Standard wall functions with sand-grain roughness modifications following the logarithmic law of the wall where the roughness function $\Delta B = \frac{1}{\kappa}\ln(1 + C_s k_s^+)$ for a fully rough regime ($k_s^+$>90) are used and $k_s^+$ is the non-dimensional roughness height. This results in Eq. (27) where $E = 9.793$.

$$u_p = \frac{u^*}{\kappa}\ln\left(\frac{E y^+}{1 + C_s k_s^+}\right) \tag{27}$$

Comparing Eq. (27) with the velocity profiles by Richards & Hoxey, Eq. (24) it is observed that both are similar and hence must be consistent. If profiles proposed by Richards & Hoxey are used and through first-order matching, $k_s$ is given by:

$$k_s = \frac{E z_0 z_p}{C_s(y_0 + y_p)} \tag{28}$$

According to FLUENT [19], the roughness constant $C_s$ is set to a default value of 0.5 which when used with $k$-$\varepsilon$ turbulence models, Nikuradse's original experimental data [23] for flow through a pipe tightly packed with a uniform sand-grain roughness can be replicated. FLUENT subsequently recommended that the range of $C_s$ should lie between 0 and 1 depending on how roughness varies from the uniform sand-grain. However, this range is limited to a specific case of pipe flow as evidenced from lack of a clear guideline on setting $C_s$ for arbitrary types of roughness. Using the default value of $C_s$ in FLUENT results in $k_s \approx 3.8\,m$, based on $z_0 = 0.19368\,m$. Given the multitude of studies [21] [24] [25] which strongly advised that the normal distance from the centre point *P* of the first cell adjacent to the wall $z_p$ to be greater than $k_s$, it follows that $z_p > 3.8\,m$ and subsequently the height of the first cell adjacent to the wall, $2z_p > 7.6\,m$. This requirement is not practical, given that the height of a container used in this case study is only a mere $2.54\,m$. The result would be a very coarse mesh which would not be able to resolve the flow field accurately. To overcome this, an alternative solution is considered and implemented in this study where the restriction on $C_s$ is relaxed and set equal to $E$ which results in $k_s = z_0$. This condition is enforced in the far-field region of the computational domain

surrounding the near-field region, where the arrays of containers are placed. In the near-field region, $k_s$ is set equal to 0 and $C_s$ equal to 0.5.

*Statistical Performance Measures*

To quantify the quality of numerical results and their agreement with experimental data, statistical performance measurements are introduced. Following recommendations by Chang et al. [10], the normalized mean square error (NMSE), fractional bias (FB) and the fraction of predictions within a factor of two of observations (FAC2) are used. The definitions of these statistical performance measures are defined below:

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\overline{C_o}\,\overline{C_p}} \tag{29}$$

$$FB = \frac{(\overline{C_o} - \overline{C_p})}{0.5(\overline{C_o} + \overline{C_p})} \tag{30}$$

$$FAC2 = \frac{N_{0.5 \le \frac{C_p}{C_o} \le 2.0}}{N_{total}} \tag{31}$$

where $C_p$ refers to model predictions while $C_o$ are experimental values. A perfect model would have a value of 1 for FAC2; a value of 0 for FB and NMSE. However, it is well-known that uncertainties will inevitably be present in simulating transport phenomena.

**Results & Discussion**

*CODASC*

Grid independence studies were carried out, ensuring the solution no longer changes with mesh density. A mesh size of 14 million cells was found to be adequate. As the flow approaches normal to the street canyon, flow separation results with elements of recirculation inside the canyon. Pollutants released from the ground will be carried by the downward movement of the recirculated flow and deposited on the leeward side. This recirculation is responsible for differences in magnitude of $c^+$ between both walls as the pollutants gradually accumulate on the leeward side whilst deposits of tracer gas are removed from the windward side. Flow recirculation also traverses in the $z$-direction, along the length of each building as seen in Fig. 4. Since the entering flow is symmetrical about the $z$-$x$ plane, the traversing of recirculated flow grows from each end of the street canyon into the $z$-direction and naturally meets in the middle, which explains why $c^+$ is highest in the middle of the street canyon. Combining all these elements together will result in concentration distribution in Fig. 5. The streamlines of the airflow in Fig. 4 are colored by velocity magnitude and the contour plot in Fig. 5 is distinguished by normalized concentration $c^+$.

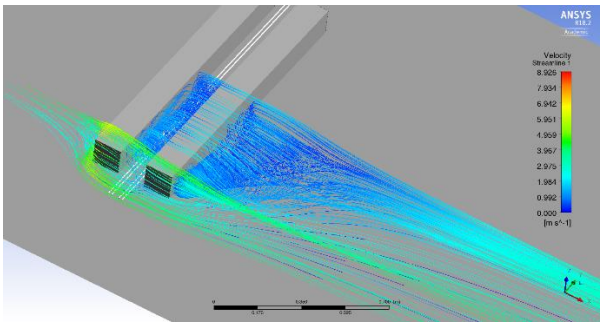


***Figure 4.*** *Velocity ($m/s$) streamlines obtained from a z-x cutting plane at edge of the model as flow approaches normal to building.*

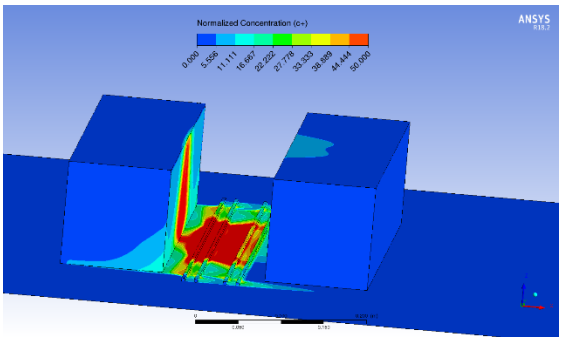


***Figure 5.*** *$c^+$ distribution in near-field region area of interest.*

*Dependence of $Sc_t$ Value on $c^+$*

One of the primary concerns in this study is determining the appropriate value of $Sc_t$, which is a free parameter, to be used. Given the nature of $Sc_t$ as described earlier, this free parameter can be tuned to artificially reduce or amplify turbulent diffusion or alternatively as it is more commonly done, the value to be used can be obtained from reviews of past studies. The practice of tuning the $Sc_t$ value is ill-advised as evidenced from multiple studies which have warned against it in order to fit the experimental data available [17] [18]. But the use of such tuning practices strengthens the premise that determining the appropriate $Sc_t$ value is problem dependent [17] as flow interacting with different configurations of obstacles have their own turbulent kinetic energy and turbulent diffusion rates. Referring to past studies, Tominaga et al. [17] concluded that dispersion in a street canyon configuration compared to that around a single building will be different, stating that turbulent diffusion is often underestimated in RANS models when applied to single building configuration. With more obstacles present, the increase in turbulent kinetic energy would compensate for that underestimation, hence it is expected that a value greater than 0.3 would provide better prediction results for a street canyon configuration. They went on to conclude that $Sc_t$ values should be considered depending on how dominant turbulent diffusion is in a given flow configuration. But Gromke et al. [12] used RANS simulations of a street canyon for multiple values of $Sc_t$ ranging from 0.2 to 1.0 and found that the value of 0.3 gave the best agreement with experimental results. However, they concluded by recommending a critical review of $Sc_t$ values for future studies on pollutant dispersion in urban environments.

In this case study, RANS simulations were conducted for selected $Sc_t$ values from a range of 0.1 to 0.9 to validate against past studies and to show the effects $Sc_t$ values have on turbulent diffusion. Normalized concentration, $c^+$ was monitored on both leeward and windward side in the street canyon. On each side, $y$ from 0 to 1.2 were divided into 100 equally spaced discrete points for every $z$ value from 0 to 0.12 in increments of 0.02. Referring to Fig. 6 below, points 1 to 700 represent distribution of $c^+$ on the leeward side of the street canyon. Similarly, points 701 to 1400 represent distribution of $c^+$ on the windward side of the street canyon. For example, with reference to Fig. 1, point 1 refers to the red cross marked on the leeward wall, where $z$ = 0 and $y$ = 0. Point 2 is subsequently on $z$ = 0 and $y$ = 0.012 and point 100, denoted by the orange cross is where $z$ = 0 and $y$ = 1.2. Point 101 is continued where $z = 0.02$ and $y = 0$ and so on. Point 701 is on the windward wall, where $z = 0$ and $y = 0$, is denoted by the purple cross. Point 800 is denoted by the black cross which is where $z = 0$ and $y = 1.2$. Point 801, by the same logic, is where $z = 0.02$ and $y = 0$. This accounts for the total of 1400 discrete points to determine the concentration distribution on both leeward and windward walls. By visual

observation, it can be seen that $c^+$ distribution has been consistently overpredicted on the leeward side and this overprediction increases as the $Sc_t$ value increases in Fig. 6. The same observation applies to the windward side but $c^+$ is underpredicted when $Sc_t$ has a value of 0.3. $c^+$ is larger in magnitude on the leeward side and this remains true for all cases simulated.

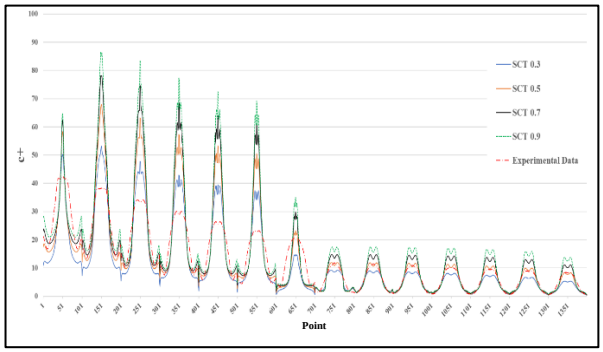



***Figure 6.*** *Plot of $c^+$ distribution on leeward side followed by windward side in the street canyon.*

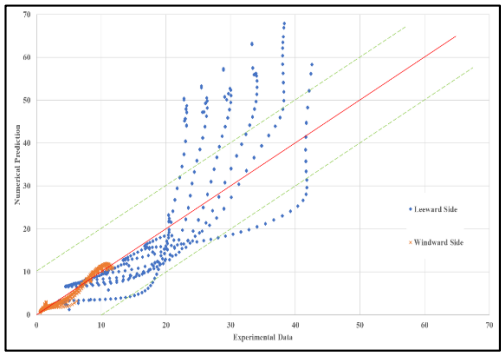



***Figure 7.*** *45° line plot of $c^+$ distribution when $Sc_t$ = 0.5.*

Fig. 7 shows the deviation between numerical prediction and experimental data for the case when $Sc_t$= 0.5. The leeward side and windward side are separated for ease of analysis. Data points should ideally lie along the red line and in between the two dotted green lines which provide a band of error on how far these points deviate from experimental measurements. $c^+$ distribution on the windward wall tend to lie along the 45° line compared to that of the leeward wall, where larger and more frequent deviation is encountered. Due to the extensive number of data points, statistical measures are necessary in order to give a quantitative view of the results which are reflected in Table. 1. Comparing the results, NMSE is lowest for $Sc_t$ 0.3 while FB for $Sc_t$ 0.5 is lowest. Both $Sc_t$ 0.3 and 0.5 share the same value for FAC2 which is close to the ideal value of 1 compared to other $Sc_t$ values. From this analysis, it can be concluded that $Sc_t$ 0.5 is the optimum value for this street canyon case study, simultaneously verifying Tominaga et al. [17] expectations for a $Sc_t$ value greater than 0.3.

***Table 1.*** *Statistical performance measurements for different $Sc_t$ values*

| | $Sc_t$ 0.3 | $Sc_t$ 0.5 | $Sc_t$ 0.7 | $Sc_t$ 0.9 | Model Perfects |
|---|---|---|---|---|---|
| **NMSE** | 0.20 | 0.31 | 0.45 | 0.59 | 0 |
| **FB** | 0.14 | -0.05 | -0.17 | -0.26 | 0 |
| **FAC2** | 0.95 | 0.95 | 0.92 | 0.90 | 1 |

Contour plots of the leeward wall of the street canyon are presented from Figs. 8 and 9, while contour plots of the windward wall are presented from Figs. 10 and 11. These are positioned on the $z - y$ plane. The contour plots are all symmetrical about the $z$-axis, which serves as a form of validation between model and physical reality. $c^+$ at the two ends of both leeward and windward walls is relatively lower due to more dominant convection forces from vortices generated from the interaction between the flow and the street canyon.

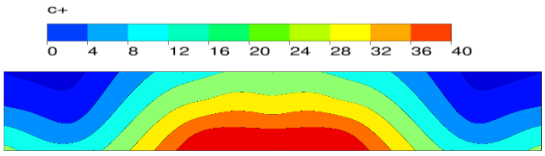



***Figure 8.*** *$c^+$ distribution of experimental result for leeward wall.*

***Figure 9.*** *$c^+$ distribution of numerical result on leeward wall when $Sc_t$ = 0.5.*

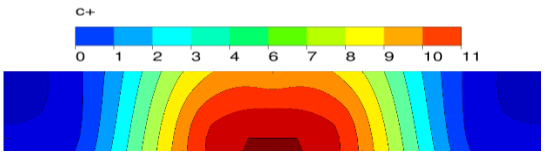



***Figure 10.*** *$c^+$ distribution of experimental result for windward wall.*

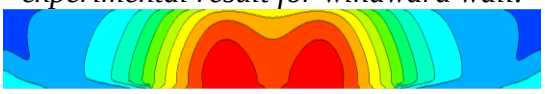


***Figure 11.*** *$c^+$ distribution of numerical result on windward wall when $Sc_t$ = 0.5.*

*Comparison Between Passive Scalar and Multi-Species Transport Models*

For the multi-species model, a $Sc_t$ value of 0.5 is used based on earlier findings from the passive scalar model. As seen in Table 2, the passive scalar transport model outperforms the multi-species model in every statistical benchmark, which is unexpected as it is initially believed that the multi-species model will instead outperform the passive scalar transport model since the pollutant considered is denser than air. Density of the flow field is expected to change due to changes in the mass fraction of pollutant in air because of the coupling of momentum equations with the continuity equation. Similarly, Fig. 12 to Fig. 15 show the respective contour plots.

***Table 2.*** *CODASC: Comparison of passive scalar and multi-species transport model through statistical performance measures.*

| | Passive Scalar Transport Model | Multi-Species Transport Model | Model Perfects |
|---|---|---|---|
| **NMSE** | 0.31 | 1.99 | 0 |
| **FB** | -0.05 | -0.48 | 0 |
| **FAC2** | 0.95 | 0.87 | 1 |

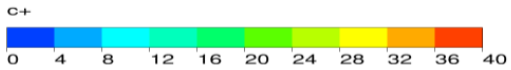
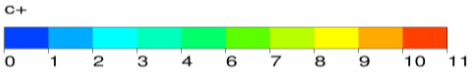



***Figure 12.*** *CODASC: $c^+$ distribution on the leeward wall (Passive Scalar Model).*

***Figure 13.*** *CODASC: $c^+$ distribution on the leeward wall (Multi-Species Model).*



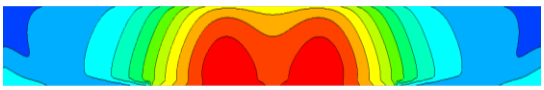



***Figure 14.*** *CODASC: $c^+$ distribution on the windward wall (Passive Scalar Model).*

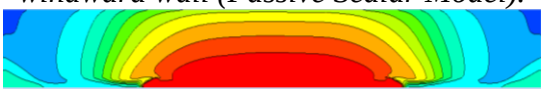


***Figure 15.*** *CODASC: $c^+$ distribution on the windward wall (Multi-Species Model).*

**Mock Urban Setting Test (MUST)**

In order to determine if some form of homogeneity has been achieved with the current parameters set, the profiles of velocity in the $y$-direction, $k$ and $\varepsilon$ are monitored. Referring to Fig. 16, profiles at $y\_1$ and $y\_2$ are tracked as flow progresses from the far-field region to the near-field region, especially since grid density changes are involved. Fig. 17 shows the $y$-velocity profiles with the analytical profile, the $k$ and $\varepsilon$ profiles with their respective analytical profiles as $z$ increases. The velocity profiles in Fig. 17 reflects a relatively good agreement between the analytical profile and $y\_1$ and $y\_2$ save for some deviations on the order of approximately 0.1. $k$ profiles, on the other hand, show an increasing deviation from the

analytical profile with increasing $z$. The profile of $y\_1$ deviate from $y\_2$ near the ground and generally show an underprediction of turbulent kinetic energy. According to Richards and Norris [26], the local maximum in turbulent kinetic energy $k$ at $y\_2$ is due to an inconsistency in the discretization of the production term $P_k$ instead of the turbulence model itself. The $\varepsilon$ profiles show no differences between the analytical profile, $y\_1$ and $y\_2$ except for the near-ground region.

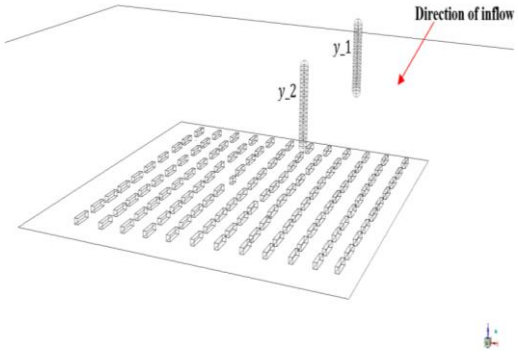



***Figure 16.*** *Locations in the MUST case where v, k and ε profiles are taken and monitored from.*

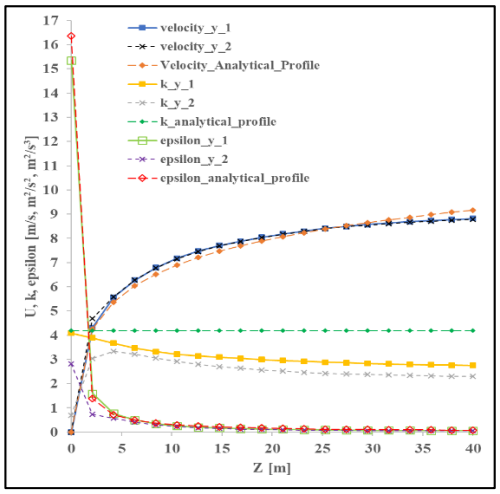



***Figure 17.*** *Plot of Z against mean velocity, k and ε profiles for MUST case.*

As the incoming flow approaches the near-field region, the presence of containers forces the flow to be diverted with recirculation around each individual container and is deflected upwards. Complexity surrounding this flow configuration is significantly greater than the street canyon case given that the flow enters at an inclined angle. The streamlines of the fluid flow are shown in Fig. 18, where it is colored by velocity magnitude. According to the coloration of the streamlines, it is observed that a slight acceleration of the flow persists despite enforcing the horizontal homogeneity condition when flow transits from the far-field to near-field region. It is also noted that behind each container, the velocity of streamlines drops to almost zero.

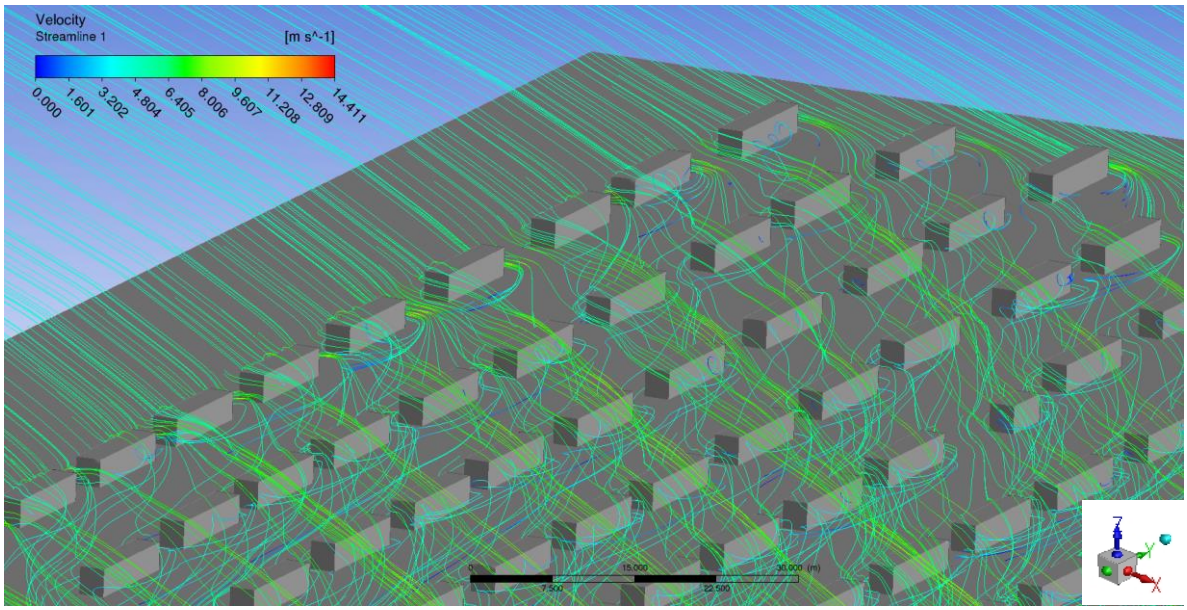



***Figure 18.*** *Velocity streamlines of flow entering the near-field domain at a $-41°$ angle of the MUST case.*

*Dependence of $Sc_t$ Value on $c^+$*

Given the flow configuration in MUST varies significantly from the street canyon model due to different geometries, it is expected that the optimal $Sc_t$ would be different. Experimental data is readily available from on-site measurements which adds ease to validation. Based on visual observations of Fig. 19, the most noticeable trend is that numerical results consistently underpredict $c^+$. Using statistical performance measures in Table 3., NMSE appears to be lowest for $Sc_t$ 1.5, while FB is lowest for $Sc_t$ 2. Both $Sc_t$ 1.5 and 2 share the same FAC2 of 0.47. Comparing NMSE and FB of $Sc_t$ 1.5 and $Sc_t$ 2 leads to conflicting results and can be concluded that there are limited differences between these two $Sc_t$ values. However, a conclusion can be drawn in that $Sc_t$ values of 1.5 and higher will result in better agreement with experimental data. Fig. 20 show the contour plot of $Sc_t$ value equal to 1.5 which ranges from 0 $ppmv$ to 6 $ppmv$. The inlet flow is at an angle of $-41°$, which is stated to be the mean angle based on data obtained by field experiments [14].

***Table 3.*** *Statistical performance measurements of various $Sc_t$ values for the MUST case.*

| | $Sc_t$ 0.5 | $Sc_t$ 0.7 | $Sc_t$ 0.9 | $Sc_t$ 1.5 | $Sc_t$ 2 | Model Perfects |
|---|---|---|---|---|---|---|
| **NMSE** | 1.68 | 1.08 | 0.78 | 0.55 | 0.59 | 0 |
| **FB** | 0.95 | 0.81 | 0.70 | 0.49 | 0.39 | 0 |
| **FAC2** | 0.03 | 0.21 | 0.35 | 0.47 | 0.47 | 1 |

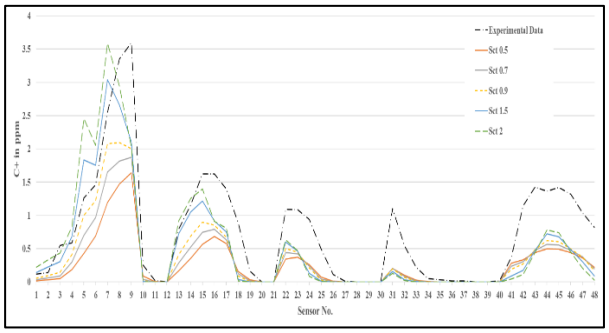



***Figure 19.*** *Plot of numerical $c^+$ data from 48 sensors placed at various locations with increasing $Sc_t$ values in the MUST case.*

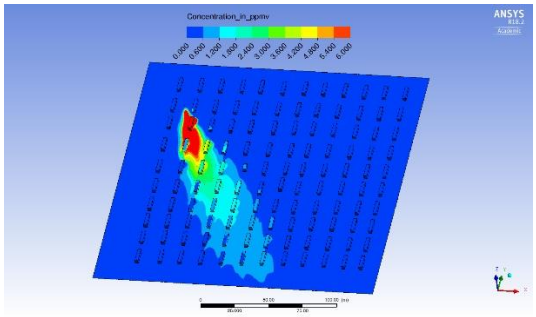



***Figure 20.*** *$c^+$ distribution with $Sc_t$ = 1.5 for MUST case.*

*Comparison Between Passive Scalar and Multi-Species Transport Models*

Next, a comparison of accuracy in predicting $c^+$ between the passive scalar transport model and multi-species model is made. $Sc_t$ value of 1.5 is taken for both cases. From Fig. 21, the multi-species model provides better agreement with experimental results compared to the passive scalar model. Significant underprediction of numerical results continue to persist, with increasing deviation away from the source release. A similar conclusion is also reached by examining the statistical performance measures in Table 4. NMSE and FB indicate that the multi-species model gave results closer to perfect model values than the passive scalar model did while both models hold the same values for FAC2 of 0.47. Fig. 22 and Fig. 23 show the degree of dispersion with different methodologies used.

***Table 4.*** *Comparison of statistical performance measures of passive scalar and multi-species transport model against the model perfects.*

| | Passive Scalar | Multi-Species | Model Perfects |
|---|---|---|---|
| NMSE | 0.55 | 0.45 | 0 |
| FB | 0.49 | 0.31 | 0 |
| FAC2 | 0.47 | 0.47 | 1 |

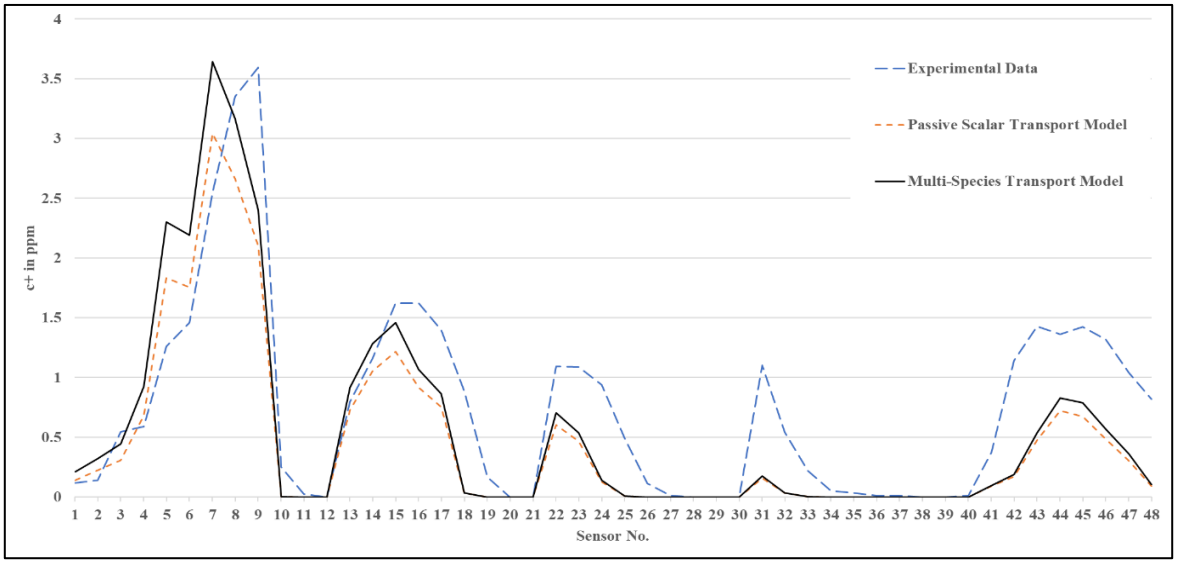



***Figure 21.*** *Plot of numerical $c^+$ data for passive scalar and multi-species transport model with $Sc_t$ = 1.5 of 48 sensors placed at various locations for MUST case.*

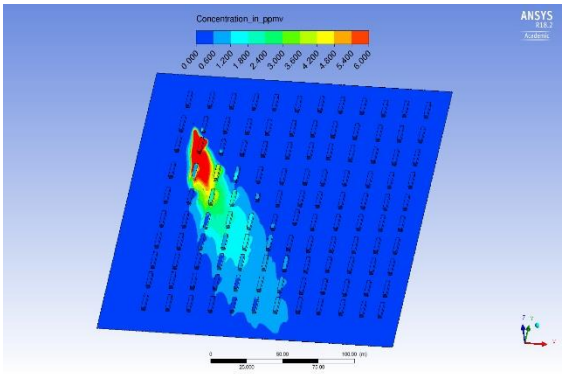



***Figure 22.*** *$c^+$ distribution of passive scalar transport model in MUST case with the mean inflow angle of $-41°$.*

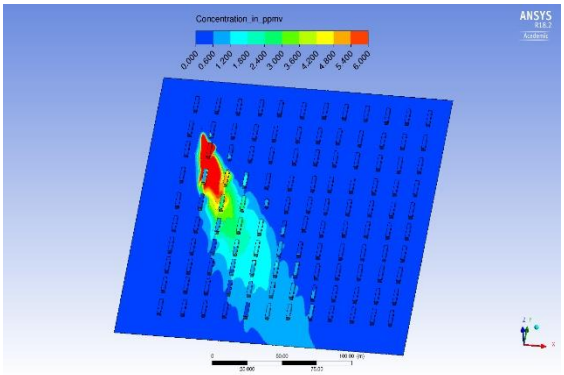



***Figure 23.*** *$c^+$ distribution of multi-species transport model in MUST case with the mean inflow angle of $-41°$.*

*Variation of Inflow Angles*

In an attempt to investigate further the discrepancy in concentration of pollutant far from the source release, the inlet boundary conditions are examined in further detail. A standard deviation of 9.5° in the instantaneous wind direction is provided as reflected by Yee and Biltoft [27]. Likewise, an additional point mentioned by Nadir et al. [13] was the standard deviation of the inlet flow angle of $-41°$ but this was overlooked as a potential source of error in his paper. This is considered in the present study and following the deviation of $-9.5°$, two separate additional cases with the inflow angle of $-31.5°$ and $-50.5°$ were simulated. In Fig. 24, the inflow at an angle of $-31.5°$ clockwise from the positive $x$-axis resulted in better agreement compared to the case with an angle $-41°$ clockwise from the positive $x$-aixs. Underprediction still occurs, albeit at a lower occurrence when inflow angle deviates to $-31.5°$ and it is mostly limited to the tower of sensors numbered 40 to 48. This is further evidenced in Table 5., where the case with inflow angle of $-31.5°$ performed significantly better on every statistical performance measure. When inflow is angled at $-50.5°$, the opposite occurs in that the model performed worse. These cases are simulated using a $Sc_t$ value of 1.5. This was repeated using the multi-species transport model and a similar conclusion can be drawn. The above attempts

in carrying out studies involving the change in inflow angles conclusively show a source of error originating from comparing field experiments with numerical studies. This error is expected and inevitable when carrying out full-scale outdoor experiments.

***Table 5.*** *Statistical performance measures against model perfects for varying inflow angles with the passive scalar transport model in MUST case.*

| | $-41°$ | $-31.5°$ | $-50.5°$ | **Model Perfects** |
|---|---|---|---|---|
| NMSE | 0.55 | 0.21 | 2.31 | 0 |
| FB | 0.49 | 0.28 | 0.76 | 0 |
| FAC2 | 0.47 | 0.65 | 0.18 | 1 |

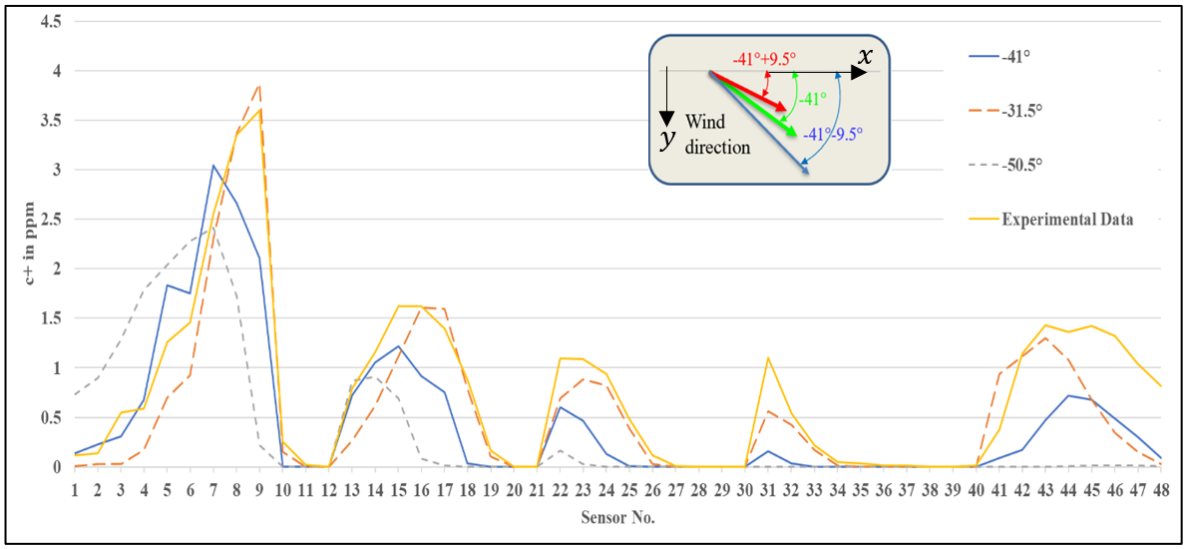



***Figure 24.*** *Plot illustrating the numerical $c^+$ data from 48 sensors placed at various locations in the MUST case, with different inflow angles using the passive scalar transport model compared against experimental data.*

## Conclusion

A comparison between two different transport methodologies was carried out through their respective validation against two test cases; the CODASC model and MUST. The CODASC experiment was carried out under controlled conditions in a wind tunnel whereas MUST is a full-scale outdoor experiment, taking into account effects of the atmospheric boundary layer. Results from CODASC study proved that the passive scalar transport model outperformed the multi-species transport model and the opposite for MUST. In reality, it is expected that the multi-species model will provide better accuracy since density changes caused by species compositions in the flow field are taken into account. Furthermore, pollutants heavier than air in terms of molecular weight were used in simulating transport phenomena for both cases, allowing the physical dispersion process to be more accurately represented.

Hence, it is concluded that since the multi-species transport model did not provide significant improvement over the passive scalar transport model in general, the multi-species model should be dismissed until further investigations can prove its worth. Furthermore, for engineering applications where computational efficiency is a major component in determining which methodology to use, it should be noted that the passive scalar transport model is much more efficient compared to the multi-species transport model. Time taken for a simulation using the passive scalar transport model can be up to three times shorter compared to using the multi-species model.

For CODASC case study, limited information on how the four lines of tracer source release were modeled in both wind tunnel experiments and numerical studies that it is believed the results from this present study could only be justified with the assumption of a dilute source. For MUST, given that it is a full-scale outdoor experiment, it would be unrealistic to expect the magnitude and direction of the inlet wind flow be maintained throughout the course of the experiment, though it is an assumption necessary for numerical studies lest the level of complexity be increased. Through investigations on the variations of inflow angles, it suggests that the inflow angle did not remain constant at $-41°$ during field experiments. Differences in configuration of obstacles in both cases could play a major role in prohibiting the full realisation of benefits in utilising the multi-species transport model. Efforts to remove the turbulent Schmidt number, $Sc_t$ should be a priority in reducing ambiguity in dispersion studies using CFD.

**References**

[1] https://www.independent.co.uk/news/world/asia/fukushima-nuclear-disaster-radiation-lethal-levels-leak-japan-tsunami-tokyo-electric-power-company-a8190981.html (Date of last access: 10/4/2019)

[2] https://www.nst.com.my/world/2018/11/436113/gas-leak-caused-deadly-blast-china-olympic-city (Date of last access: 10/4/2019)

[3] Á. Leelőssy, F. Molnár, F. Izsák, Á. Havasi, I. Lagzi, and R. Mészáros, "Dispersion modeling of air pollutants in the atmosphere: a review," *Central European Journal of Geosciences*, vol. 6, no. 3, pp. 257–278, Sep. 2014.

[4] Lateb, M., et al. "On the Use of Numerical Modelling for near-Field Pollutant Dispersion in Urban Environments −A Review." *Environmental Pollution*, vol. 208, 2016, pp. 271–283., doi: 10.1016/j.envpol.2015.07.039.

[5] M. Mohan, "Analysis of various schemes for the estimation of atmospheric stability classification," *Atmospheric Environment*, vol. 32, no. 21, pp. 3775–3781, 1998.

[6] Adel A. Abdel-Rahman, "On the Atmospheric Dispersion and Gaussian Plume Model," in *2nd International Conference on Waste Management, Water Pollution, Air Pollution, Indoor Climate (WWAI'08)*, Corfu, 2008.

[7] R. MacDonald, "Theory and Objectives of Air Dispersion Modelling", Modelling Air Emissions for Compliance MME 474A Wind Engineering, Dec. 2003.

[8] Y. Tominaga and T. Stathopoulos, "CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques," *Atmospheric Environment*, vol. 79, pp. 716–730, 2013.

[9] Y. Tominaga and T. Stathopoulos, "Ten questions concerning modeling of near-field pollutant dispersion in the built environment," *Building and Environment*, vol. 105, pp. 390–402, 2016.

[10] J. C. Chang and S. R. Hanna, "Air quality model performance evaluation," *Meteorology and Atmospheric Physics*, vol. 87, no. 1-3, pp. 167–196, Jun. 2004.

[11] C. B. Gromke, "CODASC: a database for the validation of street canyon dispersion models," *In Proceedings of the 15th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO)*, May 2013.

[12] C. Gromke and B. Ruck, "Dispersion study in a street canyon with tree planting by means of wind tunnel and numerical investigations – Evaluation of CFD data with experimental data," *Atmospheric Environment*, vol. 42, no. 37, pp. 8640–8650, Dec. 2008.

[13] N. Bekka, P. Kumar, A. A. Feiz, S. Singh, M. Sellam, E. Barbosa, P. Ngae, G. Turbelin, A. Chpoun, "A CFD MODELING APPROACH FOR A CONTAMINANT RELEASED IN A CITY", in *17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Budapest, Hungary, 9-12 May 2016.

[14] C. A. Biltoft, "Customer Report for Mock Urban Setting Test," *DPG Document No. WDTCFR- 01-121, West Desert Test Center, U. S. Army Dugway Proving Ground, Dugway, Utah*, 58 pp., 2001

[15] Y. Tominaga and T. Stathopoulos, "CFD simulations of near-field pollutant dispersion with different plume buoyancies," *Building and Environment*, vol. 131, pp. 128–139, 2018.

[16] B. E. Launder and D. B. Spalding, "The numerical computation of turbulent flows," *Computer Methods in Applied Mechanics and Engineering*, vol. 3, no. 2, pp. 269–289, 1974.

[17] Y. Tominaga and T. Stathopoulos, "Turbulent Schmidt numbers for CFD analysis with various types of flowfield," *Atmospheric Environment*, vol. 41, no. 37, pp. 8091–8099, 2007.

[18] C. Gualtieri, A. Angeloudis, F. Bombardelli, S. Jha, and T. Stoesser, "On the Values for the Turbulent Schmidt Number in Environmental Flows," *Fluids*, vol. 2, no. 2, p. 17, 2017.

[19] ANSYS® FLUENT Release 18.2 Theory Guide, August 2017.

[20] https://www.windforschung.de/CODASC.htm (Date of last access: 10/4/2019)

[21] B. Blocken, T. Stathopoulos, and J. Carmeliet, "CFD simulation of the atmospheric boundary layer: wall function problems," *Atmospheric Environment*, vol. 41, no. 2, pp. 238–252, 2007.

[22] P. Richards and R. Hoxey, "Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 46-47, pp. 145–153, 1993.

[23] J. Nikuradse, "Laws of Flow in Rough Pipes," *VDI Forschungsheft*, July/August 1933.

[24] Parente, Alessandro, "CFD boundary conditions, turbulence models and dispersion study for flows around obstacles," 2013.

[25] B. Blocken, J. Carmeliet, T. Stathopoulos, "CFD evaluation of wind speed conditions in passages between parallel buildings—effect of wall-function roughness modifications for the atmospheric boundary layer flow" *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 95, no. 9–11, pp. 941–962, 2007, https://doi.org/10.1016/j.jweia.2007.01.013.

[26] P. J. Richards and S. E. Norris, "Appropriate boundary conditions for computational wind engineering models revisited," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 99, no. 4, pp. 257–266, 2011.

[27] E. Yee and C. A. Biltoft, "Concentration Fluctuation Measurements in a Plume Dispersing Through a Regular Array of Obstacles," *Boundary-Layer Meteorology*, vol. 111, no. 3, pp. 363–415, 2004.

# Three-dimensional computer reconstruction of the airway- and the vascular systems of the lung of the domestic fowl, *Gallus gallus* variant *domesticus*

**Jeremy D. Woodward[1], Yolanda Ramonisi[2], Reatlegile Mashiteng[2], Lolo Mokae[2], John N. Maina[2]**

[1]Division of Medical Biochemistry and Structural Biology, Structural Biology Research Unit, University of Cape Town, Observatory 7925, South Africa.
[2]Department of Zoology, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa.

Corresponding and presenting author: jmaina@uj.ac.za

## Abstract

Invariably, any type of microscopic imaging modality can be used to reconstruct morphological data in three-dimensions (3-D). These techniques allow us to study the structural characteristics of cells, tissues and organs and gaining meaningful insights into their form and function. One of the oldest methods for studying biological entities in 3-D is serial histological-section reconstruction (SHSR) which dates back to the late 19$^{th}$ century. Despite the advent of competing modern techniques that are faster, cheaper and easier to apply and that work in an automated fashion, this method (SHSR) remains indispensable. It is because reconstructions resulting from sections generally display better contrast and signal-to-noise ratio. Additionally, dyes and labels can be used more effectively when applied to the surfaces of sections than to, e.g., resin sample blocks: functional units can therefore be identified by coloured- or fluorescent signals. SHSR is useful from ~0.2 µm resolution and can easily sample large volumes in the $1000^3$ voxel range. These characteristics make it particularly effective for preparing accurate, high-resolution 3-D representations of micro-anatomical data. Here, we report on computational use of SHSR in investigation of the structure of the airway- and vascular systems of the mature lung of the domestic fowl, *Gallus gallus* variant *domesticus*. The topographical relationships between the structures were thoroughly examined after preparation of movies that allowed us to rotate the reconstruction around different axes and extract and re-insert different parts back into the visualization. In the extant air-breathing vertebrates, the avian respiratory system (the lung-air sac system) is structurally the most complex and functionally the most efficient. Its unique design largely explains how, among only a few other extant animal taxa, birds attained volancy. Powered (active) flight is an exceptionally energetically costly mode of locomotion which requires large amounts of oxygen to sustain. While it has been continuously studied for over the last four centuries, some aspects of the structure and function of the avian lung remain uncertain and contentious. They include the airflow dynamics across the lung, the arrangement of the airway- and vascular systems and the shapes, sizes and spatial organization of the terminal respiratory units. We noted that in contrast to the mammalian lung, where arterial- and airway systems track each other and the venous system runs segmentally, i.e., between the other two systems, in the avian lung, the corresponding structures do not display similar close following. This may be explained by the unique development and evolution of the avian lung: two morphologically- and functionally distinct parts, namely the paleopulmo and the neopulmo, exist. The parts develop at different times which later combine into one organ. It explains how and why the airways- and the blood vessels do not track and even pattern each fundamentally because the conduits are dedicated to the distinctive parts of the lung with which they were associated.

**Keywords:** chicken, domestic fowl, bird, lung, serial sectioning, 3-D reconstruction

## Introduction

> *A number of techniques are available to generate volumetric ultrastructural models, and combination of a variety of strategies is now possible for tailoring to specific biological questions and applications.* [1]

Through the process of evolution by natural selection [2], living things develop in specific ways to enable them to best execute functions. Understanding cellular dynamics and processes is fundamental to characterizing the underpinnings of life. Most diseases and pathologies arise from cellular abnormalities which trigger intracellular biochemical changes. After the invention of the compound light microscope, some four-and-half centuries ago by Zacharias Jansen (1580-1638) in 1595 [3] [4] and for nearly three hundred years thereafter, microscopic study of biological structures was limited to two-dimensional (2-D) delineation, depiction and description. In the late 19$^{th}$ century, researchers desiring to understand the three-dimensional (3-D) attributes of biological entities employed different ingenious visualization aids to examine serial sections, such as projecting lines from the outlines of structures to produce an image of the object perpendicular to the direction of sectioning [5] or by tracing outlines of the object of interest onto stacks of glass sheets [6] [7]. Modifications on these types of techniques were still being used [8] when Sydney Brenner first applied computers to the problem of visualizing serial section reconstructions in the early 1970's [9]. More recently however, advances in computing power; software; imaging technology as well as visualization tools have spurred the development of a large number of 3-D reconstruction approaches. These techniques are revolutionizing the fields of molecular-, cellular- and tissue biology [1] [10]-[21]: even four-dimensional (4-D) preparations or visualizations are now possible [21]-[23]. The distinctive functions that are displayed by biological structures stem from a multitude of minute physical and biochemical events than occur between and among topologically connected structural components at the different levels of the organization of highly complex assemblages [24]. The spatial conformation (shape) of the structural components that comprise living matter is ubiquitously 3-D [25]-[28]. Recently, 3-D cell- and organoid culture have become particularly popular in studies of cell development, morphology, differentiation, host-pathogen interaction and effects of drug treatments compared to the traditional 2-D ones [23] [29]-[41]. It has been recognized that cells which are cultured in 3-D cell culture assays behave utterly differently compared to those 'grown' in 2-D ones [31] [32]. Comparative investigations have indicated that 2-D cell cultures lose tissue distinctive assemblage, generating changes in mechanical and biochemical signals and disrupting cell-cell or cell matrix connections [31]. In drug delivery studies, while many preparations may show success in 2-D cell cultures, this is not the case in *in vivo* studies because the 3-D environment of the large number of cells in the body may render it problematical for the drugs to equally affect all cells [30]. Furthermore, although in 2-D culture cells acquire a spheroid morphology, the clusters are formed mainly by aggregation of cell masses instead of dividing from a single cell and the groupings of cells are significantly larger than those obtained by 3-D culture [36]. 3-D cell culture optimally supports the 3-D shape of cells, thereby providing a physiologically appropriate environment similar to the one which exists in the living tissue [34]-[36]. It is by coiling and folding into proper 3-D shape that proteins can perform their complex biochemical function(s). Here, two examples are given to underscore the importance of 3-D configuration in proper function of biological components.

In microscopic biology, 3-D reconstruction is a highly instructive investigative technique [18] [42]-[45]. It involves constructing a spatial model of a biological entity from a sufficient

number of 2-D images [46]-[48]. Unfortunately, because of the high cost of time and materials as well as the special skills required to perform it, 3-D reconstruction has not been employed in morphological studies as much as it should have been. The 2-D images needed to prepare the 3-D shape can be acquired by manual preparation of serial sections or by tomographic methods using any measureable signal that is able to penetrate the specimen such as: light, ultrasound, gamma rays, X-rays, neutrons or electrons [18] [49]-[52]. Serial section 3-D reconstruction is the oldest method of obtaining 3-D spatial micro-anatomical data and dates back to the late 19$^{th}$ century [53]-[55]. Recent advances in optical microscopical sectioning techniques as well as automatic block-face image capturing methods like episcopic fluorescence image capturing (EFIC) or surface imaging microscopy (SIM) [56] [57] have considerably increased the efficiency and the accuracy of serial sectioning. Regarding possible sub-micron resolution, compared to the modern 3-D reconstruction techniques, conventional light microscopical (histological) sections have certain advantages of which the main ones are contrast and large sample size which is investigated [11] [18] [58] [59]. Only a few state-of-the-art 3-D reconstruction techniques, e.g., synchrotron-based microcomputer tomography (μCT) or focused ion beam scanning electron microscopy (FIB/SEM) tomography compare in resolution with imaging sections [60]. As more robust ways of utilizing 3-D reconstruction technology continue to be developed, 3-D imaging and animation will constitute an indispensable investigative arsenal in the morphologist's toolbox. Old questions will be revisited and investigated from different perspectives and new questions will be formulated and researched. 3-D reconstruction has lately found application in new research fields such as tissue engineering and regenerative medicine [60]-[63].

In this investigation, we have performed 3-D computer reconstruction on serial histological sections to study the spatial arrangement of the airway- and the vascular systems of a bird's (avian) lung. Among the extant air-breathing vertebrates, the avian respiratory system (the lung-air sac system) is structurally the most complex and functionally the most efficient [64]-[72] compared with the mammalian one. Although it has been continuously investigated for well over four centuries (i.e., since [73]), still, some important aspects of its biology remain unclear and/or controversial [74]. Like the invaginated gas exchangers of other vertebrates, the lungs of birds are ventilated tidally, i.e., in-and-out, and in addition the avian lung (specifically the paleopumonic part of it) is ventilated unidirectionally and continuously in a caudocranial direction, i.e., back-to-front. This is achieved by synchronized bellows-like actions of the air sacs. The path followed by the inspired air across the avian lung is controlled by aerodynamic valving [75] and not by anatomical valves or sphincters, as it was once believed. Morphometrically, the avian lung has thin blood-gas barrier, large respiratory surface area and large blood capillary volume, structural parameters which confer an exceptionally high pulmonary diffusing capacity for oxygen [71] [76]-[78]. Such specializations explain why except for bats, birds are the only other vertebrate animals that have attained powered (active) flight which is an energetically highly costly form of locomotion which requires particular specializations [79]. Some birds can fly nonstop over long distances and others fly under the extreme hypoxic conditions of the high altitude [78]-[82]. Here, we present a 3-D reconstruction method for multi-view image acquisition of microscopic samples combined with pre- and post-processing steps including correlation-based image registration, filtering and a combination of manual and automated segmentation.

## Materials and methods

### *Fixation and processing of the lung*

The Animal Ethics Committee of the University of Johannesburg approved all experimental procedures (Clearance Number: 2017-06-29/Maina). A mature domestic fowl (chicken), *Gallus gallus* variant *domesticus*, was killed by intravenous injection with pentobarbitone sodium (Euthanase®) into the brachial vein at a dosage of 2mg/kg. Thereafter, the lungs were fixed by intratracheal instillation of phosphate buffered 2.5% glutaraldehyde (350 mOsm $L^{-1}$, pH 7.4) at a pressure head of 3 kPa. The trachea was ligated and the fixative left *in situ* for six hours. Afterwards, the lungs were carefully dissected out from their costal attachments. Whole lungs were processed and embedded in paraffin wax according to routine histological procedures.

### *Serial sectioning and imaging*

Two-thousand six hundred and eighty-nine (2689) transverse serial sections were cut at 8 µm thickness, stained with haematoxylin and eosin and mounted onto glass slides. The whole series of sections constituted the entire lung volume. A total of thirty seven (37) sections were lost or damaged during sectioning. The rest were mounted onto glass slides. In most cases, the lost sections were non-consecutive, but in one case nine sections (72 µm) were lost in a row. An area measuring 12.88 × 9.655 mm, which included the entire transverse section through the lung was photographed using an Axioskop image analyser (Zeiss Instruments) at a magnification of x10 in uncompressed Tiff image format at a resolution of 2576 × 1931 pixels with a calibrated sampling of 5 µm/pixel.

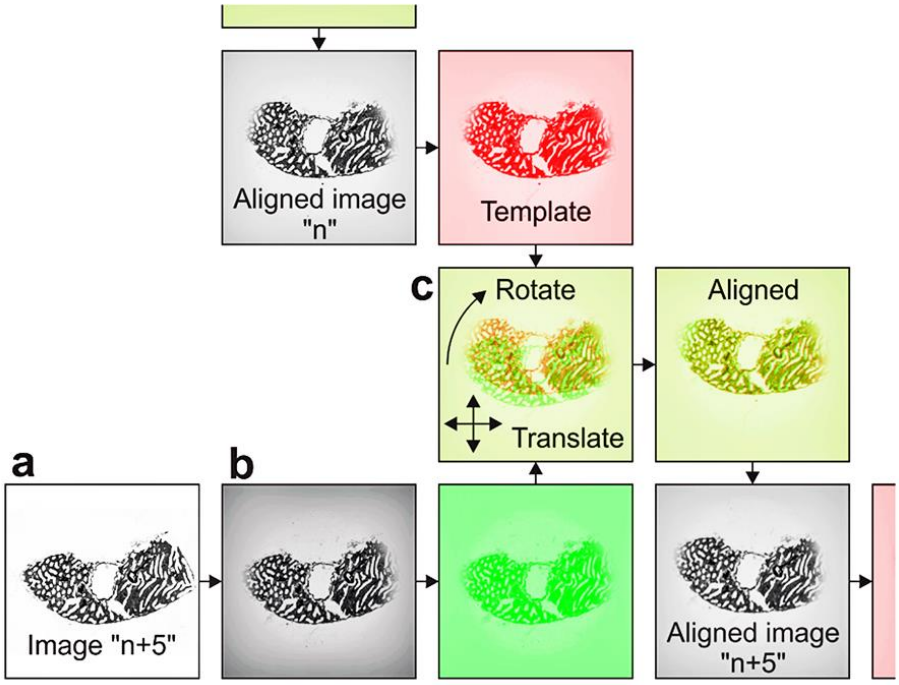



**Figure 1:** Image processing and alignment. (a). Every 5th image was selected, normalised and down-sampled. (b). The images were multiplied by a mask function to prevent the image border and particles of dust from contributing to the alignment score. (c). The previously aligned image ("n") is used as a template and the images are correlated with one another: this is illustrated using red/green. The image to be aligned (n+5) is rotated and translated relative to the template (n) until a maximum correlation score is obtained (aligned). This newly aligned image is then used as a template to align the next image (n+10) and the process is repeated.

***Image alignment***

Every fifth image was selected. In cases where this section was missing or showed obvious defects such as folds, tears, compression or inadequate staining, the previous- or subsequent section was chosen. This produced five hundred and thirty five (535) images (representing 40 µm in the Z-direction), which were manually corrected in brightness and contrast using ImageJ Version 1.4.0. [83] and imported into Spider V.15 [84]. Images were normalised to a mean of 0 and standard deviation of 1 and down-sampled by a factor of 8 using bicubic interpolation to yield a sampling of 40 µm in X and Y (Fig. 1a). The images were then multiplied by a mask function representing a Gaussian falloff (Fig. 1b) and aligned to one another by maximising the cross-correlation function in X, Y and in-plane rotation (Fig. 1c) [12]. The resulting alignment was sufficient for resolving the parabronchi and larger blood vessels (Fig. 2) and permitted satisfactory reconstruction.

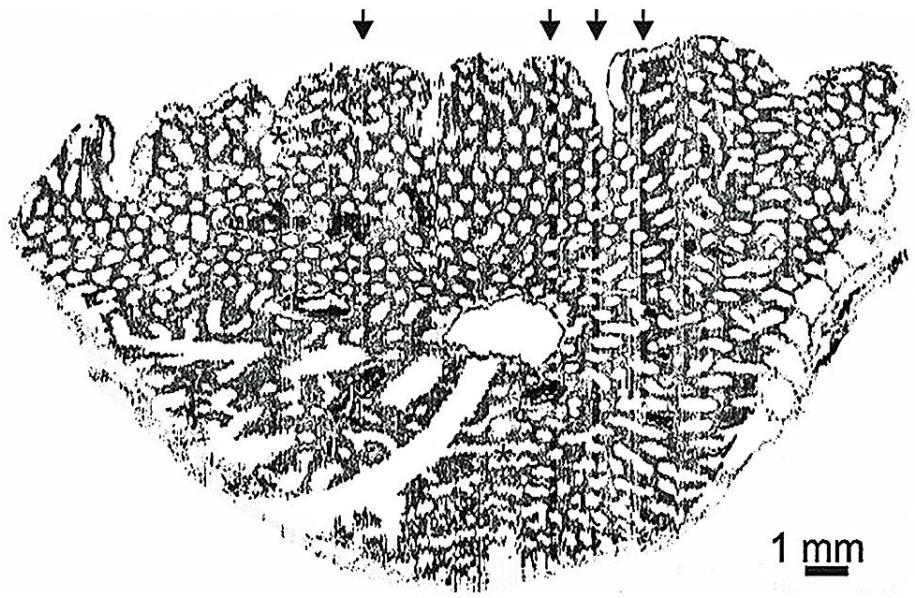



**Figure 2:** A cross-section (image 127) of the lung through the aligned image stack. All 535 individual sections can be seen as vertical lines. Errors in vertical alignment can be seen as "jitters" in individual structures. Areas where this is especially apparent are marked (*). Large differences in intensity between adjacent sections can be seen as vertical bands (arrows).

***Segmentation of the reconstruction***

To identify and segment the air-conducting elements of the reconstruction, an automated procedure was used. It involved applying a Gaussian filter and a threshold [12] [14]. In some cases, where the parabronchi were separated from adjacent ones by interparabronchial septa, a border was manually drawn around the periphery of the lung. Also, where two air-conducting elements lay adjacent to one another, to ascertain that they were resolved, lines were manually drawn between them. This process was iterated until a reasonable match between the unprocessed images and segmentation [85] was achieved (Fig. 3).

Blood vessel segmentation was achieved by manually defining the border of each blood vessel in the original (unprocessed) colour images (Fig. 3). These images were then down-sampled and aligned to one another by applying the alignment parameters obtained previously. To ensure that no blood vessels were mistaken for air-conducting elements

(especially those lacking erythrocytes in their lumen), the segmented blood vessels were subtracted from the air-conducting elements.

***Reconstruction processing and display***

The above procedure produced three image stacks: "air", "blood" and "original images". These were converted into volumes using Spider V.15 and low-pass Fourier filtered to a resolution of 160 µm with a Gaussian falloff [84] [85]. All three reconstructions were simultaneously displayed in UCSF Chimera 1.12 [86], the appropriate surface threshold value being determined by comparison with the "original images" volume. Larger blood vessels and airways were visualised by applying an additional low-pass Fourier filter to a resolution of 0.8 mm and adjusting the threshold value. The segmentation function in Chimera was used to further segment the blood vessels into arteries and veins.

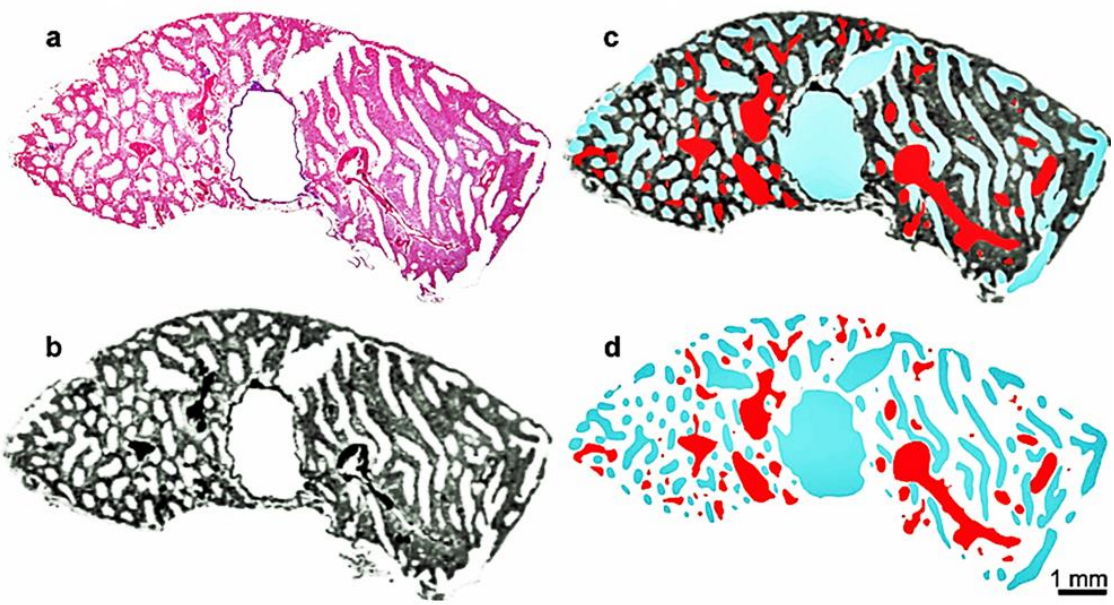



**Figure 3:** Segmentation of the reconstruction. (a). An unprocessed hematoxylin and eosin stained image (section 233) showing various air-conducting elements and blood vessels. (b). The same section shown in (a), but resampled, normalised and aligned. The larger vessels and lumina of the parabronchi are sufficiently well-resolved. (c). The manual blood vessel segmentation (red) and automated airways segmentation (cyan) superimposed on (b). An excellent match between the structures identifiable in the section was obtained. (d). The final image showing the airways (cyan) and blood vessels (red).

## Results

***Airways: bronchial system***

The trachea divided into right- and left extrapulmonary primary bronchi (EPPB) at the syrinx. The EPPB penetrated the lung at the hilus where they lie craniolateral to the pulmonary artery (PA) and caudomedial to the pulmonary vein (PV). On entering the lung, the EPPB becomes the intrapulmonary primary bronchus (IPPB) or the mesobronchus which changes in diameter and course as it passes through the lung to exit the lung at the abdominal air sac. On the various aspects of its lumen, as it passes through the lung, the IPPB gives off four sets of secondary bronchi. These are: the medioventral secondary bronchi (MVSB) that originate from the dorsomedial aspect of the lumen; the mediodorsal secondary bronchi (MDSB) that arise from the dorsal wall; the lateroventral secondary bronchi (LDSB) that arise from the caudoventral part and; the laterodorsal secondary bronchi (LDSB) that emanate from the

lateral aspect of the distal part of the IPPB. The parabronchi or the tertiary bronchi interconnect the secondary bronchi. The parabronchial system, which connects the MVSB to the MDSB, forms the paleopulmo or the 'old lung' while those which connect the MDSB to the LVSB and the LDSB form the neopulmo or the 'new lung'. The paleopulmonic parabronchi form a stack or pile of air conduits which largely occupy the dorsocranial part of the lung while the neopulmonic ones are mostly located on the caudoventral part of the lung. As they (paleopulmonic parabronchi) join the MVSB to the MDSB on the dorsal aspect of the lung, they form hoop-like shapes: the paleopulmonic parabronchi lie parallel to each other and sporadically anastomose with each other while the neopulmonic parabroonchi anastomose profusely, forming a dense network. Generally, the paleopulmonic parabronchi are larger in size compared to the neopulmonic ones.

***Vascular systems: pulmonary artery (PA)***

The PA enters the hilus ventral to the root of the first MVSB. On penetrating the lung, it divides into four main branches (= rami), namely the accessory-, the cranial-, the caudomedial- and the caudolateral branches. The branches supply blood to different parts of the lung: the accessory branch, which is the first blood vessel to arise from the PA, supplies blood to a small part of the lung ventral to hilus; the cranial branch supplies blood to the craniodorsal region of the lung cranial to the second costal sulcus; the caudolateral branch supplies blood to the ventrolateral part of the lung and; the caudomedial branch, which is the most direct extension of the PA, supplies blood to most of the lung caudal to the second costal sulcus. The four branches of the PA divide the lung roughly into a cranial- and a caudal arterial vascular region, with a vertical transverse line passing through the second costal sulcus forming the anatomical landmark or the dividing boundary: the cranial part of the lung is supplied with blood by the accessory- and the cranial branches while the caudal one is supplied by the caudomedial- and the caudolateral branches. Along the median longitudinal plane which divides the lung into a lateral and a medial half along the so-called *linea anastomotica*, i.e., the area marking connection between the parabronchi from the MVSB and those from the MDSB (the paleopulmonic parabronchi) meet, the former part of the lung is supplied by the caudolateral branch of the PA while the later one is supplied by the caudomedial branch. In this study, up to about the level of the interparabronchial arteries, no anastomoses were observed between the four branches of the PA.

***Vascular system: pulmonary vein (PV)***

At the hilus, the PV is separated from the PA by the intrapulmonary primary bronchus. The PV is formed by connection of three converging blood vessels (radices), namely the cranial-, the caudal- and the ventral radices. The radices join outside the lung to form the PV. In our study, the connection of the radices was not included in the 3-D reconstruction because that part of the lung was inadvertently cut off during trimming off of adhering connective tissue. Most of the craniodorsal part of the lung is drained by the cranial radix which is formed by confluence of three large veins; the caudal radix drains the part of the lung caudal to the third costal sulcus and is formed by as many as four radices which extend dorsally and ventrally and; the ventral radix, which drains the cranioventral part of the lung comprises two main branches which drain the part of the lung located between the second and the third costal sulci. Up to the level of the interparabronchial veins, no anastomoses were observed between the radices.

## Discussion

Employing different methods, the morphologies of the airways of the avian lung have been investigated by various investigators [64]-[66] [71] [87]-[90]. Consensus on the numbers of

airways, their shapes, connections, topographic locations and nomenclature has, however, not yet been reached [74. While interspecific differences have been reported [64]-[66] [68] [71] [91] [92], it cannot be completely ruled out that interspecimen differences occur largely from developmental abnormalities and irregularities and the environmental conditions under which the avian eggs are incubated. In this study, in addition to 3-D reconstruction of the airways- and the blood vessels, for more insightful visualization of the morphologies and the spatial associations between the different parts, movies which allowed the rotation of the constructions across different planes and extraction (removal) and re-introduction of the different parts to the constructions were prepared. Although certain weaknesses exist in all the techniques which have been used to study the morphology of the avian lung, this study corroborates most of the previous accounts. Indisputably, 3-D reconstruction is a very powerful means of studying the morphologies of biological structures, including those of structurally complex avian respiratory system.

The pulmonary vasculature of birds has been studied by [93]-[98]. Various techniques which included injection with markers and following the paths they follow, e.g., microspheres, in the blood vessels by light microscopic examination and microfilm, silicone, mercox or latex rubber injection followed by maceration and preparation of casts or replicas have been used. While our observations generally agree with those previously made by other investigators, certain differences exist. Unlike in the mammalian lung where arteries closely follow airways while veins run intersegmentally [99] [100], i.e., they are located between the airways and the arteries, in the avian lung, the airway- and the vascular systems do not display such arrangement. This may be explained by the complex development of the avian lung [101]-[103], where, in lungs of phylogenetically derived species (evolutionally advanced birds), the paleopulmonic part develops first to be followed by the neopulmonic one [102] [103]: the adult lungs of an evolutionally developed bird comprises two distinctive parts, namely the ‘paleopulmo’ and the ‘neopulmo’ which display structural and functional differences: the two regions are typically located in different regions of the lung [64] [68] [92] and the two parts are ventilated differently. The paleopulmo is continuously and unidirectionally ventilated in a caudocranial direction by concerted actions of the air sacs [67] [104] while the neopulmo is ventilated tidally, i.e., back-and-forward. Taking these properties into consideration, it is axiomatic that the airways and the blood vessels of the avian lung cannot follow each other in the same way as in the mammalian lung, where mesenchymal cells which contribute to the development of the airways and the blood vessels start at the same point (essentially the lung bud) and in close proximity grow outwards as the lungs develop [105]-[107], forming the various functional systems. Regarding the observations of [93] [94], in comparison with the observations noted in this study, certain structural discrepancies exist. The main ones are: a) while two main radices were reported to converge and form the PV [93] [94], here, three main blood vessels drained the lung and joined to form the PV and; b) the second costal sulcus and not the third one formed the boundary between the cranial- and the caudal blood supply- and the respective drainage regions of the lung by the PA and PV, respectively.

In conclusion, compared to the other techniques which have hitherto been employed to study the morphology of the airway- and the vascular systems of the avian lung, incontrovertibly, 3-D reconstruction is the more robust technique. When it is combined with preparation of movies which can be operated and closely viewed from different angles, the geometries of the structures can be thoroughly scrutinized and understood. It is important to underscore that while a powerful technique in its own right, scanning electron microscopy (SEM), which is conventionally applied for imaging biological structures, does not strictly generate 3-D images as generally wrongly interpreted: although they may appear so, the resulting images

do not have an aspect of depth. Presently, there are several SEM techniques that can be used to obtain 3-D information on a biological sample [108]-[110]. Some of them can be done on any microscope and some require specialist instrumentation, software, or microscopes.

## Acknowledgements

This work was funded by the National Research Foundation (NRF) of South Africa through a Research Incentive Grant.

## References

[1] Miranda, K., Girard-Dias, W., Attias, M., de Souza, W. and Ramos, I. (2015) Three dimensional reconstruction by electron microscopy in the life sciences: an introduction for cell and tissue biologists. *Molecular Reproduction and Development* **82**, 530-547.

[2] Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murry, London.

[3] Nachtigall, W. (1997) *Exploring with the Microscope: A Book of Discovery and Learning*. Sterling Publications, London.

[4] Savile, B. and Bracegirdle, B. (1998) *Introduction to Light Microscopy*. Springer-Verlag, New York.

[5] His, W. (1880) *Anatomie menschlicher Embryonen*. Vogel, Leipzig.

[6] Kastschenko, N. (1886) Methode zur genauen rekonstruktion kleinerer makroskopischer gegenstande. *Archive für Anatomie und Physiologie Abteilung* **1**, 388-394.

[7] Odhner, T. (1911) Zum naturlichen System der digenen Trematoden IV. *Zoologischer Anzeiger*. **38**, 513-531.

[8] Ohta, Y. and Millhouse, E.W., Glass Plate Reconstruction from Serial Sections Used in the Study of Neonatal Biliary Atresia, *Stereology*, Elias H. Ed.. Springer, Berlin, 1967, 302-303.

[9] Ward, S., Thomson, N., White, J.G. and Brenner, S. (1975) Electron microscopical reconstruction of the anterior sensory anatomy of the nematode, *Caenorhabditis elegans*. *Journal of Comparative Neurology* **160**, 313-337.

[10] Teutsch, H.F., Schuerfeld, D. and Groezinger, E. (199) Three-dimensional reconstruction of parenchymal units in the liver of the rat. *Hepatology* **29**, 494-505.

[11] Woodward, J.D. and Maina, J.N. (2005) A 3D digital reconstruction of the components of the gas exchange tissue of the lung of the muscovy duck, *Cairina moschata*. *Journal of Anatomy* **206**, 477-492.

[12] Woodward, J.D. and Maina, J.N. (2008) Study of the structure of the air- and blood capillaries of the gas exchange tissue of the avian lung by serial section three-dimensional reconstruction. *Journal of Microscopy* **230**, 84-93.

[13] Song, W.C., Hu, K.S., Kim, H.J. and Koh, K.S. (2007) A study of the secretion mechanism of the sebaceous gland using three-dimensional reconstruction to examine the morphological relationship between the sebaceous gland and the arrector pili muscle in the follicular unit. *British Journal of Dermatology* **157**, 325-330.

[14] Maina, J.N. and Woodward, J.D. (2009) Three-dimensional serial section computer reconstruction of the arrangement of the structural components of the parabronchus of the ostrich, *Struthio camelus* lung. *Anatomical Record* **292**, 1685-1698.

[15] Sun, K., Zhang, J., Chen, T., Chen, Z., Chen, Z., Li, Z., Li, H. and Hu, P. (2009) Three dimensional reconstruction and visualization of the median nerve from serial tissue sections. *Microsurgery* **29**, 573-577.

[16] Penczek, P.A. (2010) Fundamentals of three-dimensional reconstruction from projections. *Methods in Enzymology* **482**, 1-33.

[17] Wu, X., Yu, Z. and Liu, N. (2012) Comparison of approaches for microscopic imaging of skin lymphatic vessels. *Scanning* **34**, 174-180.

[18] Onozato, M.L., Klepeis, V.E, Yagi, Y. and Mino-Kenudson, M. (2012) A role of three-dimensional (3D)-reconstruction in the classification of lung adenocarcinoma. *Analytical Cellular Pathology* **35**, 79-84.

[19] Baghaie, A., Pahlavan, Tafti, A., Owen, H.A., D'Souza, R.M. and Yu, Z. (2017) Three-dimensional reconstruction of highly complex microscopic samples using scanning electron microscopy and optical flow estimation. *PLoS One* **12(4)**, e0175078.

[20] Chozinski, T.J., Mao, C., Halpern, A.R., Pippin, J.W., Shankland, S.J., Alpers, C.E., Najafian, R. and Vaughan, J.G. (2018) Volumetric, nanoscale optical imaging of mouse and human kidney via expansion microscopy. *Scientific Reports* **8,** 10396.

[21] Kartasalo, K., Latonen, L., Vihinen, J., Visakorpi1, T., Nykter, M. and Ruusuvuori1, P. (2018) Comparative analysis of tissue reconstruction algorithms for 3D histology. *Bioinformatics* **34**, 3013-3021.
[22] Lorenz, U.J. and Zewail, A. (2014) Observing liquid flow in nanotubes by 4D electron microscopy. *Science* **344**, 1496-1500.
[23] Bissell, M.J. (2017) Goodbye flat biology - time for the 3rd and the 4th dimensions. *Journal of Cell Science* **130**, 3-5.
[24] Turing, A.M. (1952) The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society (London) B* **237**, 37-72.
[25] French, R. (1988) *Invention and Evolution Design in Nature and Engineering*. Cambridge University Press, Cambridge.
[26] Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nature Reviews Genetics* **19**, 789-799.
[27] Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, S.M., Hermetz, K., Wang, P., Ruan, Y. and Corces, V.G. (2017) Evolutionarily conserved principles predict 3D chromatin organization. *Molecular Cell* **67**, 837-852.
[28] Duan, Z. and Blau, C.A. (2012) The genome in space and time: does form always follow function? How does the spatial and temporal organization of a eukaryotic genome reflect and influence its function? *Bioessays* 34, 800-810.
[29] Ravi, M., Paramesh, V., Kaviya, S.R., Anuradha, E. and Solomon, P.F.D. (2015) 3D cell Culture systems: advantages and applications. *Journal of Cell Physiology* **230**, 16-26.
[30] Cavo, M., Fato, M., Peñuela, L., Beltrame, F., Raiteri, R. and Scaglione, S. (2016) Microenvironment complexity and matrix stiffness regulate breast cancer cell activity in a 3D *in vitro* model. *Scientific Reports* **6:35367**. doi:10.1038/srep35367.
[31] Powell, K. (2017) Adding depth to sell culture. *Science Technology Feature* **361, 6402**.
[32] Zanoni, M., Piccinini, F., Arienti, C., Zamagni, A., Santi, S., Polico, R., Bevilacqua, A. and Tesei, A. (2016) 3D tumor spheroid models for *in vitro* therapeutic screening: a systematic approach to enhance the biological relevance of data obtained. *Scientific Report* **6:19103**. doi: 10.1038/srep19103
[33] Baker, B.M. and Chen, C.S. (2012) Deconstructing the third dimension – how 3D culture microenvironments alter cellular cues. *Journal of Cell Science* **125**, 3015-3024.
[34] Fang, Y. and Eglen, R. (2017) Three-dimensional cell cultures in drug discovery and development. *SLAS Discovery: Advancing Life Sciences* **22**, 456-472.
[35] Langhans, S. (2018) Three-dimensional *in vitro* cell culture models in drug discovery and drug repositioning. *Fronteers in Pharmacology* **22**, 456-472.
[36] Edmondson, R., Broglie, J., Adcock, A. and Yang, L. (2014) Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay Drug Development Technology* **12**, 207- 218.
[37] Lee, G.Y., Kenny, P.A., Lee, E.H. and Bissell, M.J. (2007) Three–dimensional culture models of normal and malignant breast epithelial cells. *Nature Methods* **4**, 359-365.
[38] Abbot, A. (2003) Biology's new dimension. *Nature* **424**, 870-872.
[39] Glauco, S. (2010) Three-dimensional tissue culture based on magnetic cell levitation. *Nature Nanotechnology* **5**, 291-296.
[40] Haycock, J.W. (2011) 3D cell culture: a review of current approaches and techniques. *Methods in Molecular Biology* **695**, 1-15.
[41] Derricott, H., Luu, L., Fong, W.Y., Hartley, C.S., Johnston, L.J., Armstrong, S.D., Randle, N., Duckworth, C.A., Campbell, B.J., Wastling, J.M., Coombes, J.L. (2019) Developing a 3D intestinal epithelium model for livestock species. *Cell and Tissue Research* 375, 409-424.
[42] Levinthal, C. and Ware, R. (19772) Three-dimensional reconstruction from serial sections. *Nature* **236**: 207-209.
[43] Perkins, W.J. and Green, R.J. (1982) Three-dimensional reconstruction of biological sections. *Journal of Biomedical Engineering* **4**, 37-43.
[44] Latamore, G.B. (1983) Creating 3-D models for medical research. *Computer Graphics World* **5**, 31-38.
[45] Mercer, R.R. and Crapo, J.D. Structure of the Gas Exchange Region of the Lungs Determined by Three Dimensional Reconstructions, *Toxicology of the Lung*, Gardner, D.E., Crapo, J.D., Massaro, E.J. Eds.. Raven Press, New York, 1988, 43-70.
[46] Anderson, J.R., Wilcox, M.J., Wade, P.R. and Barrett, S.F. (2003) Segmentation and 3D reconstruction of biological cells from serial slice images. *Biomedical Science Instrumentation* **39**,117-122.
[47] Rosenhain, S., Magnuska, Z.A., Yamoah, G.G., Rawashdeh, W.A., Kiessling, F. and Gremse, F. (2018) A preclinical micro-computed tomography database including 3D whole body organ segmentations. *Scientific Data* **5**. Article number 180294.

[48] Vints, K., Vandael, D., Baatsen, P., Pavie, B., Vernaillen, F., Corthout, N., Rybakin, V., Munck, S. and Gounko, N.V. (2019) Modernization of Golgi staining techniques for high-resolution, 3-dimensional imaging of individual neurons. *Scientific Reports* **9**. Article number: 130.
[49] Herman, G. T. (2009) *Fundamentals of Computerized Tomography: Image Reconstruction from Projection, 2nd Edition*. Springer-Verlag, Berlin.
[50] Handschuh, S., Schwaha, T. and Metscher, B. D. (2010) Showing their true colors: a practical approach to volume rendering from serial sections. *BMC Developmental Biology* **10**, 41.
[51] Wang, C.W., Gosno, E.B. and Li, YS (2015a) Fully automatic and robust 3D registration of serial-section microscopic images. *Scientific Reports* ***5*****, 15051**. Doi 10.1038/srep15051.
[52] Wang, Y., Xu, R., Luo, G. and Wu, J. (2015b) Three-dimensional reconstruction of light microscopy image sections: present and future. *Frontiers of Medicine* **9**, 30-45.
[53] Born, G. (1883) Die Plattenmodelliermethode. *Archives fur mikroskopie Anatomie* **22**, 584-599.
[54] Strasser, H. (1886) Ueber das Studium der Schnittserien und über die Hülfsmittel, welche die Reconstruction der zerlegten Form erleichtern. *Zeitschrifft Wissen Mikroskope* **3**, 179-195.
[55] Strasser, H. (1987) Ueber die Methoden der plastischen Rekonstruktion. *Zeittschrift Wissen Mikroskope* **4**,168-208.
[56] Weninger, W.J. and Mohun, T. (2002) Phenotyping transgenic embryos: a rapid 3-D screening method based on episcopic fluorescence image capturing. *Nature Genetics* **30**, 59-65.
[57] Ewald, A.J., McBride, H., Reddington, M., Fraser, S.E., Kerschmann, R. (2002) Surface imaging microscopy, an automated method for visualizing whole embryo samples in three dimensions at high resolution. *Developmental Dynamics* **225**, 369-375
[58] Liu, B., Gao, X.L., Yin, H.X., Luo, S.Q. and Lu, J. (2007) A detailed 3D model of the guinea pig cochlea. *Brain Structure and Function* **212**, 223-230.
[59] Rau, T.S., Hussong, A., Herzog, A., Majdani, O., Lenarz, T. and Leinung, M. (2011) Accuracy of Computer aided geometric 3D reconstruction based on histological serial microgrinding preparation. *Computationa*l *Methods Biomechanicaland Biomedical Engineering* **14**, 581-594.
[60] Carletti, E., Motta, A. and Migliaresi, C. (2011) Scaffolds for tissue engineering and 3D cell culture. *Methods in Molecular Biology* **695**, 17-39.
[61] Deluzio, TGB, Seifu, D.G.and Mequanint, K. (2011) 3D scaffolds in tissue engineering and regenerative medicine: beyond structural templates? *Pharmaceutical Bioprocessing* **1**, 267-281
[62] Yu, Y., Moncal, K.K., Li, J., Peng, W., Rivero, I., Martin, J.A. and Ozbolat, I.T. (2016) Three-dimensional bioprinting using self-assembling scalable scaffold-free tissue strands as a new biolink. *Scientific Reports* **6**:28714. doi: 10.1038/srep28714.
[63] Jensen, G., Morrill, C. and Huang, Y. (2018) 3D tissue engineering, an emerging technique for pharmaceutical research. *Acta Pharmaceutica Sinica B* **8**, 756-766.
[64] Duncker, H.R. (1971) The lung-air sac system of birds. A contribution to the functional anatomy of the respiratory apparatus. *Ergebinisse Anatomie Entwicklungsgeschichte* **45**, 1-171.
[65] Duncker, H-R. (1972) Structure of avian lungs. *Respiration Physiology* **14**, 44-63.
[66] Duncker, H-R. (1974) Structure of avian respiratory tract. *Respiration Physiology* **22**, 1-19.
[67] Scheid, P. (1979) Mechanisms of gas exchange in bird lungs. *Review of Physiology Biochemistry and Pharmacology* **86**, 137-186.
[68] McLelland, J., Anatomy of the Lungs and Air Sacs, *Form and Function in Birds, Vol. IV*, King, A.S., McLelland, J. Eds.. Academic Press, London, 1989, 221-279.
[69] Powell, F.L., Respiration, *Sturkie's Avian Physiology, 5th Edition*, Whittow, G.C. Ed.. San Diego, Academic Press, 2000, 233-264.
[70] Powell, F.L. and Hopkins, S.R. (2004) Comparative physiology of lung complexity: implications for gas exchange. *News in Physiological Science* **19**, 55-60.
[71] Maina, J.N. (2005) *The Lung Air Sac System of Birds: Development, Structure, and Function*. Springer-Verlag, Heidelberg.
[72] Harvey, E.P. and Ben-Tal, A. (2015) Robust unidirectional airflow through avian lungs: new insights from a piecewise linear mathematical mode. *PLoS Computational Biology* **12(2),** e1004637.
[73] Coiter, V., Anatomie Avium, *Externum et Internarum Praecipalium Humani Corporis Partium Tabulae atque Anatomicae Exercitationes Observationesque Varieae*. Norimbergae, Germany, 1573, 130-133.
[74] Maina, J.N. (2017) Pivotal debates and controversies on the structure and function of the avian respiratory system: setting the record straight. *Biological Reviews* **92,** 1475-1504.
[75] Maina, J.N., Singh, P. and Moss, E.A. (2009) Inspiratory aerodynamic valving occurs in the ostrich, *Struthio camelus* lung: computational fluid dynamics study under resting unsteady state inhalation. *Respiration Physiology and Neurobiology* **169**, 262-270.

[76] Maina, J.N., The Morphometry of the Avian Lung, *Form and Function in Birds, Vol. 4*, King, A.S., McLelland, J. Eds.. Academic Press, London, 1989, 307-368.
[77] Maina, J.N., King, A.S. and Settle, J.G. (1989) An allometric study of the pulmonary morphometric parameters in birds, with mammalian comparison. *Philosophical Transactions of the Royal Society (London) B* **326**, 1-57.
[78] Maina, J.N., McCracken, K.G., Chua, B., York, J.M. and Milsom, W.K. (2017) Morphological and morphometric specializations of the lung of the Andean goose, *Chloephaga melanoptera*: a lifelong high-altitude resident. *PLoS One* **12(3)**: e0174395.
[79] Maina, J.N. (2000) What it takes to fly: The novel respiratory structural and functional adaptations in birds and bats. *Journal of Experimental Biology* **203**, 3045-3064.
[80] Tucker, V.A., Energetics of natural avian flight. *Avian Energetics*, Paynter, R.A. Ed.. Nuttal Ornithological Club, Cambridge (MA), 1974, 298-333.
[81] Altshuler, D.L. and Dudley, R. (2006) The physiology and biomechanics of avian flight at high altitude. *Integrative Comparative Biology* **46**, 4-8.
[82] Scott, G.R. (2011) Elevated performance: the unique physiology of birds that fly at high altitudes. *Journal of Experimental Biology* **214**, 2455-2462.
[83] Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T. and Eliceiri, K.W. (2017) ImageJ2: ImageJ for the next generation of scientific image data. *BMCBioinfornatics* **18**, 529.
[84] Frank, J., Radermacher, M., Pencze, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *Journal of Structural Biology* **116**, 190-199.
[85] Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W. and Gossard, D.C. (2010) Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *Journal of Structural Biology* **170**, 427-38.
[86] Pattersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612.
[87] Makanya, A.N. and Djonov, V. (2008) Development and spatial organization of the air conduits in the lung of the domestic fowl, *Gallus* variant *domesticus*. *Microscopic Research Techniques* **71**, 689-702.
[88] Makanya, A.N. and Djonov, V. (2009) Parabronchial angioarchitecture in developing and adult chickens. *Journal of Applied Physiology* **106**, 1959-69.
[89] Makanya, A.N., Kavoi, B.M., Djonov, V. (2014) Three-dimensional structure and disposition of the air conducting and gas exchange conduits of the avian lung: the domestic duck (*Cairina moschata*). *Journal of Anatomy* **1**, 1-9.
[90] Pandey, A.K., Praveen, P.K., Ganguly, S., Para, P.A., Wakchaure, R., Saroj and Mahajan, T. (2015) Avian respiratory and physiology with its interspecies variations: A review. *World Journal of Pharmacology and Life Sciences* **1**, 137-148.
[91] West, N.H., Bamford, O.S. and Jones, D.R. (1977) A scanning electron microscope study of the microvasculature of the avian lung. *Cell Tissue Research* **176**, 553-564.
[92] King, A.S. (1966) Structural and functional aspects of the avian lung and its air sacs. *International Journal of Reviews in General Experimental Zoology* **2**, 171-267.
[93] Abdalla, M.A. and King, A.S. (1975) The functional anatomy of the pulmonary circulation of the domestic fowl. *Respiration Physiology* **23**, 267-290.
[94] Abdalla, M.A., The Blood Supply to the Lung, *Form and Function in Birds, Vol. 4*, King, A.S., McLelland, J., Eds.. Academic Press, San Diego, 1989, 281-306.
[95] Radu, C. and Radu, L. (1971) Le dispositive vasculaire du poumon chez les oiseaux domestiques. *Revue de Medecine Veterinaire* **122**, 1219-1226.
[96] Abdalla, M.A. and King, A.S. (1976a) Pulmonary arteriovenous anastomoses in the avian lung: do they exist. *Respiration Physi*ology **27**, 187-191.
[97] Abdalla, M.A and King, A.S. (1976b) The functional anatomy of the bronchial circulation of the domestic fowl. *Journal of Anatomy* **121**, 537-550.
[98] Abdalla, M.A. and King, A.S. (1977) The avian bronchial arteries: species variations. *Journal of Anatomy* **123**, 697-704.
[99] Weibel, E.R. (1984) *The Pathway for Oxygen: Structure and Function in the Mammalian Respiratory System*. Harvard University Press, Cambridge (MA).
[100] Hislop, A.A. (2002) Airway and blood vessel interaction during lung development. *Journal of Anatomy*
[101] Duncker, H-R., Development of the Avian Respiratory and Circulatory Systems, *Respiratory Function in Birds, Adult and Embryonic*, Ducker, H.R. Ed.. Springer-Verlag, Berlin, 1978, 260-273.

[102] Maina, J.N. (2003a) A systematic study of the development of the airway (bronchial) system of the avian lung from days 3 to 26 of embryogenesis: a transmission electron microscopic study of the domestic fowl, *Gallus gallus* variant *domesticus*. *Tissue and Cell* **35**, 375-391.

[103] Maina, J.N. (2003b) Developmental dynamics of the bronchial- (airway) and air sac systems of the avian respiratory system from days 3 to 26 of life: a scanning electron microscopic study of the domestic fowl, *Gallus gallus* variant *domesticus*. *Anatomy and Embryology* **207**, 119-134. **201**, 325-334.

[104] Fedde, R. (1980 The structure and gas flow pattern in the avian lung. *Poultry Science* **59**, 2642-2653.

[105] Schittny, J.N. and Burri, P.H. Development and Growth of the Lung, *Fishman's Pulmonary Disorders and Diseases, 4th edition*, Fishman, A.P. Ed., McGraw Hill, New York, 2008, 91-114.

[106] Burri, P.H. (1984) Fetal and postnatal development of the lung. *Annual Review of Physiology* **46**, 617-628.

[107] Burri, P.H., Development and Growth of the Human Lung, *Handbook of Physiology, Section 3: The Respiratory System*, Fishman, A.P., Fisher, A.B. Eds.. American Physiological Society, Bethesda (MD), 1985, 1-46.

[108] Hughes, L.C., Archer, C.W. and Gwynn, I. (2005) The ultrastructure of mouse articular cartilage: collagen orientation and implications for tissue functionality. A polarised light and scanning electron microscope study and review. *European Cells and Materials* **9**, 68-84.

[109] Peddie, C.J. and Collinson, L.M. (2014) Exploring the third dimension: volume electron microscopy comes of age. *Micron* **61**, 9-19.

[110] Woodward, J.D., Wepf, R. and Sewell, B.T. (2009) Three-dimensional reconstruction of biological macromolecular complexes from in-lens scanning electronmicro-graphs. *Journal of Microscopy* **234**, 287-292.

# 3D Nonlinear random vibrations of cable-moored offshore floating structures under wave excitations

**†Kun Wang[1], Guo-Kang Er[2], Zhihui Zhu[1] and Vai Pan Iu[2]**

[1]School of Civil Engineering, Central South University, Changsha, Hunan Province, China.
[2]Department of Civil and Environmental Engineering, University of Macau, Macau SAR, China.

† Presenting and corresponding author: vtkwang@csu.edu.cn

## Abstract

The nonlinear random vibrations of the cable-moored floating structures under wave excitations are studied in three dimensions. One ends of four mooring cables are connected to the floating structure and the other ends are fixed to the seabed. The nonlinear equations of motions of the mooring cables are derived using the 3D cable elements which are formulated based on the extended Hamilton principle. The floating structure is simplified as a rigid body with six degrees of freedom. Then the equations of motion of the floating structure and mooring cables are formulated as a whole system through their connection conditions. In the last, the equations of motion of the whole structure under random wave excitation are analyzed numerically. The influences of different sag-to-span ratios and inclination angles of the mooring cables on the responses of the floating structure and maximum cable tensile force are studied.

**Keywords:** Cable-moored floating structure, random wave excitation, 3D cable elements, connection conditions.

## Introduction

The cable-moored floating structures can find their applications in ocean engineering to exploit marine resources such as oil, gas and minerals. If the floating platform is subjected to horizontal excitations, the movements of floating platform can induce the geometry change of mooring cables. The geometric nonlinearity of the mooring cables plays an important role in the dynamical analysis due to their flexibility. Some researches simplified the mooring cables as linear springs [1, 2] or nonlinear springs [3, 4] to support the floating platform, which cannot reflect the real behavior and influence of the cables. A numerical approach was developed for analyzing the dynamic behavior of marine cables using lumped mass [5-8]. With this method, cables are discretized into linear segments connected by nodes and the equilibrium equations are established at each node. The mooring cables were fully modelled using the finite element method [9, 10], in which the equations of motions of the mooring cables and those of floating platform were solved separately and iteratively.

In this paper, the nonlinear random vibrations of three-dimensional floating structure and mooring system under wave excitations are studied. The nonlinear random equations of motions of the mooring cables are formulated using the 3D cable elements formulated based on the extended Hamilton principle [12]. The cable element is simplified as a flexible tension member without considering its bending and torsion stiffness because of the extremely large ratio of its length and cross-sectional dimension. The floating platform is considered as a rigid body with six degrees of freedom, i.e., three translational displacements and three rotational displacements. The equations of motions of both the floating platform and mooring system

are formulated as whole system through their connection conditions. Finally, the whole system under random wave excitation modelled using JONWSAP spectrum is solved numerically.

## Problem Statement

Consider a floating structure and mooring system as shown in Figure 1. It consists of the floating platform and four catenary mooring lines $C_1$, $C_2$, $C_3$ and $C_4$. The floating platform and mooring cables are connected through four nodes *A*, *B*, *C* and *D*. *O* is the mass center of the floating platform. The other ends of the mooring cables are fixed on the seabed. $w_a$, $w_b$ and $w_c$ are the length, height and width of the floating platform, respectively. The top view and side view of the three-dimensional floating system are shown in Figure 2. The mooring cables $C_1$, $C_2$ and $C_3$, $C_4$ are symmetric about the *y*-axis in the plane $x_1Oy$ and $x_2Oy$, respectively. $\theta$, $l$ and $d$ are the inclination angle, inclined length and sag of the mooring cable, respectively. $w_l$ is the length between the nodes *A* and *B*.

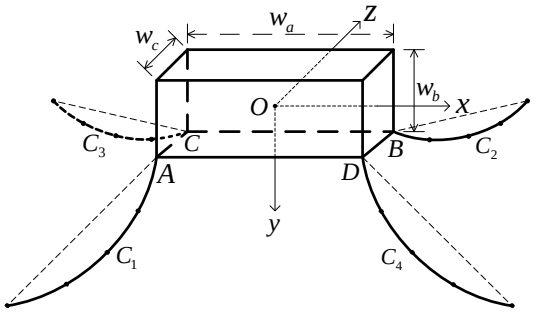
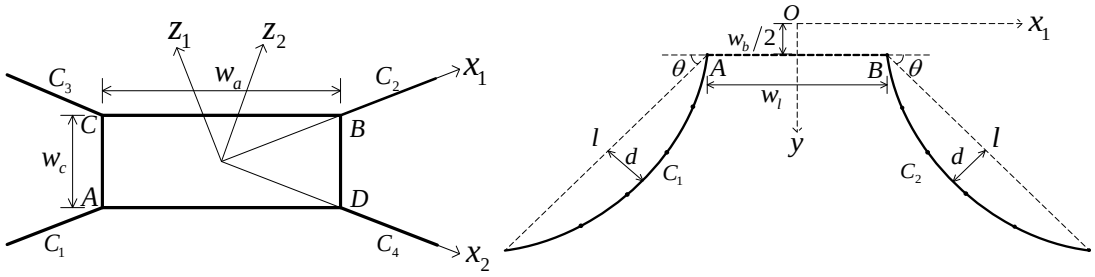



Figure 1. Configuration of the three-dimensional floating system




(a) (b)

Figure 2. (a) Top view (b) Side view of the three-dimensional floating system

## Nonlinear Random Vibrations of the Moored Floating System

*Finite Element Formulation for the Dynamics of Cable*

The equations of motion for the element *e* in the local coordinate systems $O$-$x_1yz_1$ and $O$-$x_2yz_2$ are derived based on the extended Hamilton principle and they are given as follows.

$$\left(\mathbf{M}_l^e + \mathbf{M}_a^e\right)\ddot{\mathbf{d}}_l^e + \mathbf{C}_l^e\dot{\mathbf{d}}_l^e + \mathbf{K}_l^e(\mathbf{d}_l^e)\mathbf{d}_l^e = \mathbf{f}_d^e \tag{1}$$

where $\mathbf{d}_l^e$ is the displacement vector of element *e* in the local coordinate systems $O$-$x_1yz_1$ and $O$-$x_2yz_2$; $\mathbf{f}_d^e$ is drag force vector of element *e*; $\mathbf{M}_a^e$ is the added mass matrix of element *e* which is expressed as

$$\mathbf{M}_a^e = \frac{\rho_s A l^e}{2} \mathbf{T}_1^T \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_c & 0 & 0 & 0 & 0 \\ 0 & 0 & C_c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_c & 0 \\ 0 & 0 & 0 & 0 & 0 & C_c \end{bmatrix} \mathbf{T}_1 \tag{2}$$

where $\mathbf{T}_1$ is the transformation matrix between the coordinate system $x_1yz_1$ $\left(x_2yz_2\right)$ and the coordinate system $x_3y_3z_3$, as shown in Figure 3; $C_c$ is the added-mass coefficient of the cable in the transverse direction.

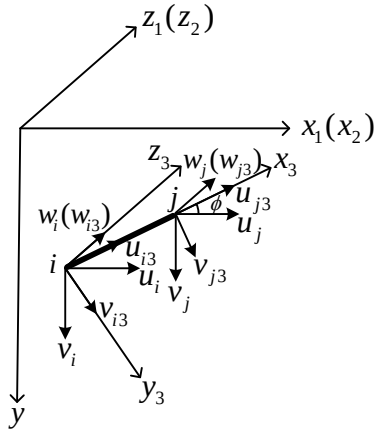



Figure 3. Differential element $e$ in the coordinate system $x_3y_3z_3$ of the cable element and the coordinate system $x_1yz_1$ $\left(x_2yz_2\right)$ of the cable

With Morison's equation, the drag forces that act along the $x_3$, $y_3$, $z_3$ directions of element $e$ are given as follows, respectively

$$\begin{aligned} F_{u3}^e &= -\frac{\pi}{2}\rho_s C_{dl} D_1 l^e \left(\bar{u}_3^e - V_{u3}\right)^2 \operatorname{sgn}\left(\bar{u}_3^e - V_{u3}\right) \\ F_{v3}^e &= -\frac{1}{2}\rho_s C_{dt} D_1 l^e \left(\bar{v}_3^e - V_{v3}\right)^2 \operatorname{sgn}\left(\bar{v}_3^e - V_{v3}\right) \\ F_{w3}^e &= -\frac{1}{2}\rho_s C_{dt} D_1 l^e \left(\bar{w}_3^e - V_{w3}\right)^2 \operatorname{sgn}\left(\bar{w}_3^e - V_{w3}\right) \end{aligned} \tag{3}$$

where $D_1$ is the diameter of the cable cross section; $C_{dl}$ and $C_{dt}$ are the drag coefficients in the longitudinal and transverse directions of the element $e$, respectively; $V_{u3}$, $V_{v3}$, $V_{w3}$ are the fluid velocities in the $x_3$, $y_3$, $z_3$ directions of element $e$, respectively; $\bar{u}_3^e$, $\bar{v}_3^e$, $\bar{w}_3^e$ are the average velocities of element $e$ in the $x_3$, $y_3$, $z_3$ directions of element $e$, respectively, which are expressed as

$$\bar{u}_3^e = \frac{\dot{u}_{i3} + \dot{u}_{j3}}{2}, \quad \bar{v}_3^e = \frac{\dot{v}_{i3} + \dot{v}_{j3}}{2}, \quad \bar{w}_3^e = \frac{\dot{w}_{i3} + \dot{w}_{j3}}{2} \tag{4}$$

and $\operatorname{sgn}(\bullet)$ denotes the sign function given by

$$\operatorname{sgn}(z) = \begin{cases} 1, & z > 0 \\ 0, & z = 0 \\ -1, & z < 0 \end{cases} \tag{5}$$

Therefore, the drag force vector $\mathbf{f}_d^e$ of element *e* in the coordinate system $x_1yz_1$ $\left(x_2yz_2\right)$ is expressed as

$$\mathbf{f}_d^e = \mathbf{T}_1^T \frac{1}{2}\left\{F_{u3}^e, F_{v3}^e, F_{w3}^e, F_{u3}^e, F_{v3}^e, F_{w3}^e\right\}^T \tag{6}$$

Using the transformation matrix $\mathbf{T}$ and the relationship $\mathbf{d}_l^e = \mathbf{T}\mathbf{d}_g^e$ in which $\mathbf{d}_g^e$ is the displacement vector of element e in the global coordinate system *O-xyz*, Eq. (1) becomes

$$\mathbf{M}_g^e \ddot{\mathbf{d}}_g^e + \mathbf{C}_g^e \dot{\mathbf{d}}_g^e + \mathbf{K}_g^e(\mathbf{d}_g^e)\mathbf{d}_g^e = \mathbf{F}_g^e \tag{7}$$

where $\mathbf{M}_g^e = \mathbf{T}^T\left(\mathbf{M}_l^e + \mathbf{M}_a^e\right)\mathbf{T}$, $\mathbf{C}_g^e = \mathbf{T}^T\mathbf{C}_l^e\mathbf{T}$, $\mathbf{K}_g^e = \mathbf{T}^T\mathbf{K}_l^e\mathbf{T}$, and $\mathbf{F}_g^e = \mathbf{T}^T\mathbf{F}_d^e$. The equations of motion of the mooring cables are

$$\mathbf{M}_m \ddot{\mathbf{U}}_m + \mathbf{C}_m \dot{\mathbf{U}}_m + \mathbf{K}_m(\mathbf{U}_m)\mathbf{U}_m = \mathbf{F}_m \tag{8}$$

where the subscript *m* denotes the number of mooring cables.

*Dynamics of the Floating Platform*

The floating platform has six degrees of freedom, which are displacements *u*, *v*, *w* along *x*, *y*, *z* axes and rotations *α*, *β*, *γ* in *xOy*, *xOz*, *yOz* plane, respectively. The equations of motion of the floating platform are given as follows based on Figure 4.

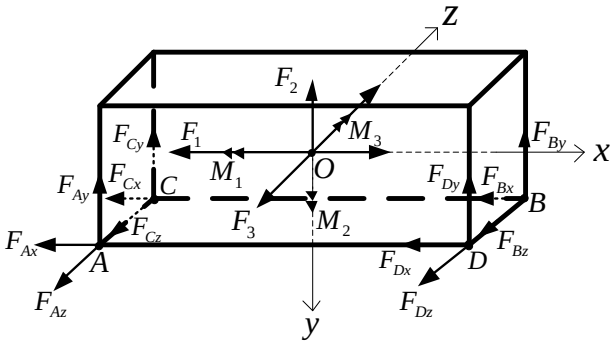



Figure 4. Forces applied on the floating platform ( $F_1 = M\ddot{u} + c_4\dot{u} - F_{dx}$ ,
$F_2 = M\ddot{v} + c_5\dot{v} - F_{dy} + F_b$ , $F_3 = M\ddot{w} + c_6\dot{w} - F_{dz}$ , $M_1 = J_x\ddot{\gamma} + c_9\dot{\gamma} - M_{dx} + F_{b2}\cdot 2w_c/3$ ,
$M_2 = J_y\ddot{\beta} + c_8\dot{\beta} - M_{dy}$ , $M_1 = J_z\ddot{\alpha} + c_7\alpha - M_{dz} + F_{b1}\cdot 2w_a/3$ )

$$\sum F_x = 0: \ \left(M + M_{ax}\right)\ddot{u} + c_4\dot{u} + F_{Ax} + F_{Bx} + F_{Cx} + F_{Dx} = F_{dx} \tag{9}$$

$$\sum F_y = 0: \ \left(M + M_{ay}\right)\ddot{v} + c_5\dot{v} + F_{Ay} + F_{By} + F_{Cy} + F_{Dy} + F_b = F_{dy} \tag{10}$$

$$\sum F_z = 0: \ \left(M + M_{az}\right)\ddot{w} + c_6\dot{w} + F_{Az} + F_{Bz} + F_{Cz} + F_{Dz} = F_{dz} \tag{11}$$

$$\begin{aligned}\sum M_z = 0: J_z\ddot{\alpha} + c_7\dot{\alpha} + F_{Ax}\frac{w_b}{2} + F_{Bx}\frac{w_b}{2} + F_{Ay}\frac{w_a}{2} - F_{By}\frac{w_a}{2} + F_{Cx}\frac{w_b}{2} \\ + F_{Dx}\frac{w_b}{2} + F_{Cy}\frac{w_a}{2} - F_{Dy}\frac{w_a}{2} + F_{b1}\frac{2w_a}{3} = M_{dz}\end{aligned} \tag{12}$$

$$\begin{aligned}\sum M_y = 0: J_y\ddot{\beta} + c_8\dot{\beta} + F_{Ax}\frac{w_c}{2} - F_{Bx}\frac{w_c}{2} - F_{Az}\frac{w_a}{2} + F_{By}\frac{w_a}{2} - F_{Cx}\frac{w_c}{2} \\ + F_{Dx}\frac{w_c}{2} - F_{Cz}\frac{w_a}{2} + F_{Dz}\frac{w_a}{2} = M_{dy}\end{aligned} \tag{13}$$

$$\begin{aligned}\sum M_x = 0 : J_x\ddot{\gamma} + c_g\dot{\gamma} + F_{Ay}\frac{w_c}{2} - F_{By}\frac{w_c}{2} + F_{Az}\frac{w_b}{2} + F_{Bz}\frac{w_b}{2} - F_{Cy}\frac{w_c}{2} \\ + F_{Dy}\frac{w_c}{2} + F_{Cz}\frac{w_b}{2} + F_{Dz}\frac{w_b}{2} + F_{b2}\frac{2w_c}{3} = M_{dx}\end{aligned} \tag{14}$$

where $M$ is the mass of the floating platform, $M_{ax}$, $M_{ay}$, $M_{az}$ are the added mass of the floating platform along the $x$, $y$ and $z$ axes, respectively, which are assumed as constant because the vertical displacement is small [12]. $J_z$, $J_y$ and $J_x$ are the moment of inertia of the floating platform in the $xOy$, $xOz$ and $yOz$ planes, respectively; $F_b$, $F_{b1}$ and $F_{b2}$ are the dynamical buoyancy of the floating body; $F_{Ax}$, $F_{Ay}$, $F_{Az}$, $F_{Bx}$, $F_{By}$, $F_{Bz}$, $F_{Cx}$, $F_{Cy}$, $F_{Cz}$, $F_{Dx}$, $F_{Dy}$, $F_{Dz}$ are the dynamical tensions from the cable at nodes $A$, $B$, $C$, and $D$ in the $x$, $y$, $z$ axes, respectively. $F_{dx}$, $F_{dy}$, $F_{dz}$, $M_{dz}$, $M_{dy}$, $M_{dx}$ are the hydrodynamic drag forces in the $x$, $y$, $z$ axes and $xOy$, $xOz$, $yOz$ planes, respectively. $F_b$, $F_{b1}$ and $F_{b2}$ are the dynamical buoyancy of the floating body due to the change of submerged volume of the floating body, which are expressed as

$$\begin{aligned}F_b &= \rho_s g w_a w_c v_f \\ F_{b1} &= \frac{1}{8}\rho_s g w_a^2 w_c \alpha \\ F_{b2} &= \frac{1}{8}\rho_s g w_a w_c^2 \gamma\end{aligned} \tag{15}$$

Referring to Figure 5 with $P_1(-w_a/2, y, z)$, $P_2(w_a/2, y, z)$, $P_3(x, w_b/2, z)$, $P_4(x, y, -w_c/2)$ and $P_5(x, y, w_c/2)$, the hydrodynamic drag forces or moments that act on the floating platform are given as follows.

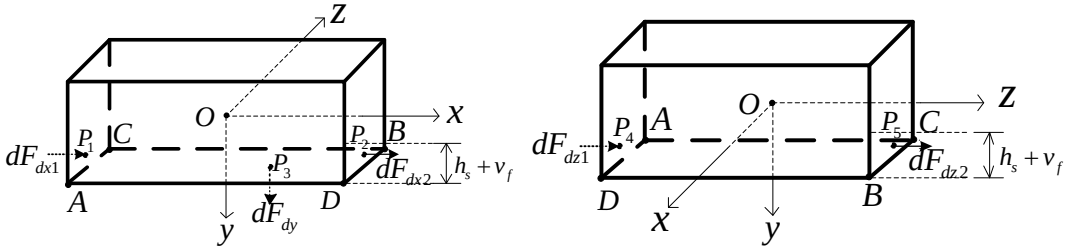



Figure 5. Drag forces act on the floating platform

$$\begin{aligned}F_{dx} &= \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2} dF_{x1} + \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2} dF_{x2} \\ F_{dy} &= \int_{-w_c/2}^{w_c/2}\int_{-w_a/2}^{w_a/2} dF_y \\ F_{dz} &= \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} dF_{z1} + \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} dF_{z2} \\ M_{dz} &= \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2} y dF_{x1} + \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2} y dF_{x2} + \int_{-w_c/2}^{w_c/2}\int_{-w_a/2}^{w_a/2}(-x) dF_y \\ M_{dy} &= \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2}(-z) dF_{x1} + \int_{-w_c/2}^{w_c/2}\int_{w_b/2-h_s-v_f}^{w_b/2}(-z) dF_{x2} \\ &\quad + \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} x dF_{z1} + \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} x dF_{z2} \\ M_{dx} &= \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} y dF_{z1} + \int_{-w_a/2}^{w_a/2}\int_{w_b/2-h_s-v_f}^{w_b/2} y dF_{z2} + \int_{-w_c/2}^{w_c/2}\int_{-w_a/2}^{w_a/2}(-z) dF_y\end{aligned} \tag{16}$$

where $dF_{dx1}$, $dF_{dx2}$, $dF_{dy}$, $dF_{dz1}$ and $dF_{dz2}$ are expressed by

$$
\begin{aligned}
dF_{dx1} &= -\frac{1}{2}\rho_s C_{dx} dydz\left(\dot{u}_f + y\dot{\alpha} + z\dot{\beta} - V_{fx1}\right)^2 \operatorname{sgn}\left(\dot{u}_f + y\dot{\alpha} + z\dot{\beta} - V_{fx1}\right) \\
dF_{dx2} &= -\frac{1}{2}\rho_s C_{dx} dydz\left(\dot{u}_f + y\dot{\alpha} + z\dot{\beta} - V_{fx2}\right)^2 \operatorname{sgn}\left(\dot{u}_f + y\dot{\alpha} + z\dot{\beta} - V_{fx2}\right) \\
dF_{dy} &= -\frac{1}{2}\rho_s C_{dy} dxdz\left(\dot{v}_f + x\dot{\alpha} + z\dot{\gamma} - V_{fy}\right)^2 \operatorname{sgn}\left(\dot{v}_f + x\dot{\alpha} + z\dot{\gamma} - V_{fy}\right) \\
dF_{dz1} &= -\frac{1}{2}\rho_s C_{dz} dydx\left(\dot{w}_f + y\dot{\gamma} + x\dot{\beta} - V_{fz1}\right)^2 \operatorname{sgn}\left(\dot{w}_f + y\dot{\gamma} + x\dot{\beta} - V_{fz1}\right) \\
dF_{dz2} &= -\frac{1}{2}\rho_s C_{dz} dydx\left(\dot{w}_f + y\dot{\gamma} + x\dot{\beta} - V_{fz2}\right)^2 \operatorname{sgn}\left(\dot{w}_f + y\dot{\gamma} + x\dot{\beta} - V_{fz2}\right)
\end{aligned}
\tag{17}
$$

where $C_{dx}$, $C_{dy}$ and $C_{dz}$ are the drag coefficients along the $x$, $y$, and $z$ directions, respectively; $V_{fx1}$, $V_{fx2}$, $V_{fy}$, $V_{fz1}$ and $V_{fz2}$ are the fluid velocities at specific locations along the $x$, $y$ and $z$ directions, respectively.

*Formulation of the Whole System*

In order to formulate the equations of motion of the mooring cables and the floating platform as a whole system, the connection conditions between the mooring lines and floating platform are required. Their relationships are

$$
\begin{aligned}
u_A &= u + \frac{w_b}{2}\alpha + \frac{w_c}{2}\beta,\ v_A = v + \frac{w_a}{2}\alpha + \frac{w_c}{2}\gamma,\ w_A = w - \frac{w_a}{2}\beta + \frac{w_b}{2}\gamma \\
u_B &= u + \frac{w_b}{2}\alpha - \frac{w_c}{2}\beta,\ v_B = v - \frac{w_a}{2}\alpha - \frac{w_c}{2}\gamma,\ w_B = w + \frac{w_a}{2}\beta + \frac{w_b}{2}\gamma \\
u_C &= u + \frac{w_b}{2}\alpha - \frac{w_c}{2}\beta,\ v_C = v + \frac{w_a}{2}\alpha - \frac{w_c}{2}\gamma,\ w_C = w - \frac{w_a}{2}\beta + \frac{w_b}{2}\gamma \\
u_D &= u + \frac{w_b}{2}\alpha + \frac{w_c}{2}\beta,\ v_D = v - \frac{w_a}{2}\alpha + \frac{w_c}{2}\gamma,\ w_D = w + \frac{w_a}{2}\beta + \frac{w_b}{2}\gamma
\end{aligned}
\tag{18}
$$

where $u_A$, $v_A$, $w_A$, $u_B$, $v_B$, $w_B$, $u_C$, $v_C$, $w_C$, $u_D$, $v_D$, $w_D$ are the displacements of the nodes *A*, *B*, *C*, and *D* in the *x*, *y*, *z* axes, respectively. Then the equations of motion about the nodes *A*, *B*, *C* and *D* in Eq. (8) are removed and replaced by Eqs. (9)-(14) using the connections conditions given by Eq. (18). The variables of displacements related to nodes *A*, *B*, *C* and *D* in other equations of motion in Eq. (8) are also expressed by Eq. (18). The final equations of motion of the whole system are obtained as

$$
\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}(\mathbf{U})\mathbf{U} = \mathbf{F}(t) \tag{19}
$$

where $\mathbf{U}$ is the global displacement vector; $\mathbf{K}(\mathbf{U})$ is the global stiffness matrix; $\mathbf{F}(t)$ is the wave force vector.

*Modeling of sea wave excitation*

The sea wave is assumed to propagate in the horizontal direction in the plane *yOz*. The kinematics of the water particles under wave excitation can be calculated based on the linear Airy wave theory [13]. The free surface elevation $\eta$ of the wave is introduced with a wave

spectrum $S_{\eta\eta}$. Then the surface elevation at location *z* and time *t* is expressed by using wave superposition as

$$\eta(z,t)=\sum_{i=1}^{N}\sqrt{2S_{\eta}(\omega_i)\Delta\omega}\cos(k_i z-\omega_i t+\theta_i) \tag{20}$$

where $\Delta\omega$ is the frequency interval; $\omega_i$ is the angular frequency of the *i*th wave component which equals $gk_i\tanh(k_i d)$ and *g* is the acceleration due to gravity; $k_i$ is the *i*th wave number which equals $2\pi/\lambda_i$ and $\lambda_i$ is the *i*th wavelength; *N* is the number of frequencies; $\theta_i$ is the statistically independent random phase angle which is uniformly distributed between 0 and $2\pi$.

The condition of deep water depth is considered in this paper because $h>\lambda/2$. Therefore, the fluid velocities $V_z$ and $V_y$ along the *z* and *y* directions at any point *P*(*x*,*y*,*z*) and time *t* are expressed as

$$\begin{aligned}V_z&=\sum_{i=1}^{N}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega}\omega_i e^{k_i(y+h_s-w_b/2)}\cos(k_i z-\omega_i t+\theta_i)\\ V_y&=\sum_{i=1}^{N}\sqrt{2S_{\eta\eta}(\omega_i)\Delta\omega}\omega_i e^{k_i(y+h_s-w_b/2)}\sin(k_i z-\omega_i t+\theta_i)\end{aligned} \tag{21}$$

During structural design, the significant height and average period of a random wave are specified. Therefore, the following approximate expression for the JONSWAP spectrum given by Goda [14] is adopted.

$$S_{\eta\eta}(\omega)=\alpha_1 H_s^2\frac{\omega^{-5}}{\omega_0^4}\exp\left[-1.25(\omega/\omega_0)^{-4}\right]\gamma^{\exp\left[-(\omega-\omega_0)^2/2\tau^2\omega_0^2\right]} \tag{22}$$

where

$$\alpha_1=\frac{0.0624}{0.23+0.0336\gamma-0.185(1.9+\gamma)^{-1}} \tag{23}$$

and $H_s$ is the significant height of the wave; $\omega_0=2\pi/T_0$ is the peak angular frequency of the wave and $T_0$ is the average period of the wave; $\gamma_1$ is the peakedness parameter which varies from 1 to 7; and $\tau$ is a shape parameter which is expressed as

$$\tau=\begin{cases}0.07, & \omega\le\omega_0\\ 0.09, & \omega>\omega_0\end{cases} \tag{24}$$

**Numerical Example**

Consider a 3D cable-moored floating platform with the parameters listed in Tables 1 and 2. The density of seawater is $\rho_s=1.025\times10^3$ kg/m$^3$. The power spectral density $S_{\eta\eta}$ is plotted in Figure 6 with $\gamma_1=3$, $H_s=0.8$ m and $T_0=8$ s. Each cable is discretized with 11 elements because further increasing the element number cannot make the precision of the results further increased obviously. The time step is 0.0008 s and the sample size used in Monte Carlo simulation (MCS) is $10^8$.

Table 1. Properties of mooring cables

| Parameter | Value |
|---|---|
| Young's modulus $E$ (N/m$^2$) | $1.9\times10^{11}$ |
| Diameter $D_1$ (m) | 0.1 |
| Mass density $\rho$ (kg/m$^3$) | $8.2\times10^{3}$ |
| Damping ratio $\xi$ | 0.03 |
| Sea depth $h$ (m) | 120 |
| Inclination angle $\theta$ (degree) | 45 |
| Sag-to-span ratio $d/l$ | 1/90 |
| Longitudinal drag coefficient $C_{dl}$ | 0.01 |
| Transverse drag coefficient $C_{dt}$ | 1 |
| Transverse added-mass coefficient $C_c$ | 1 |

Table 2. Properties of mooring cables

| Parameter | Value |
|---|---|
| Length $w_a$ (m) | 26 |
| Height $w_b$ (m) | 5 |
| Width $w_c$ (m) | 10 |
| Mass $M$ (kg) | $1.2\times10^{5}$ |
| Drag coefficient $C_{dx}$ along $x$ axis | 1 |
| Drag coefficient $C_{dy}$ along $y$ axis | 1 |
| Drag coefficient $C_{dz}$ along $z$ axis | 1 |

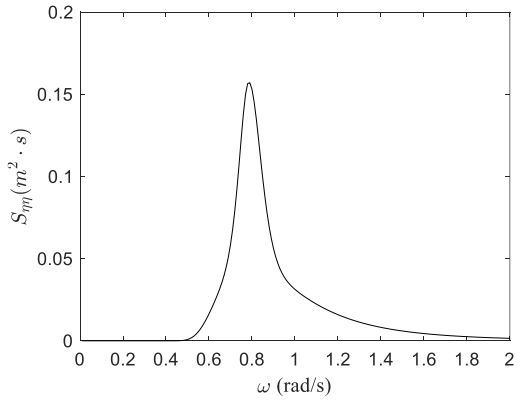



Figure 6. Power spectral density of wave surface with $\gamma_1 = 3$, $H_s = 0.8$ m and $T_0 = 8$ s

The PDFs of the responses of the floating platform and maximum cable tensile force at steady state are shown in Figure 7. The mean values of $v_f$, $w_f$, $\gamma$ and $T_c$ at steady state are 0.0059 m, 0 m, $0^o$ and $7.845\times10^5$ N, respectively and the corresponding standard deviations are 0.0112 m, 0.445 m, $1.95^o$ and $1.258\times10^5$ N.

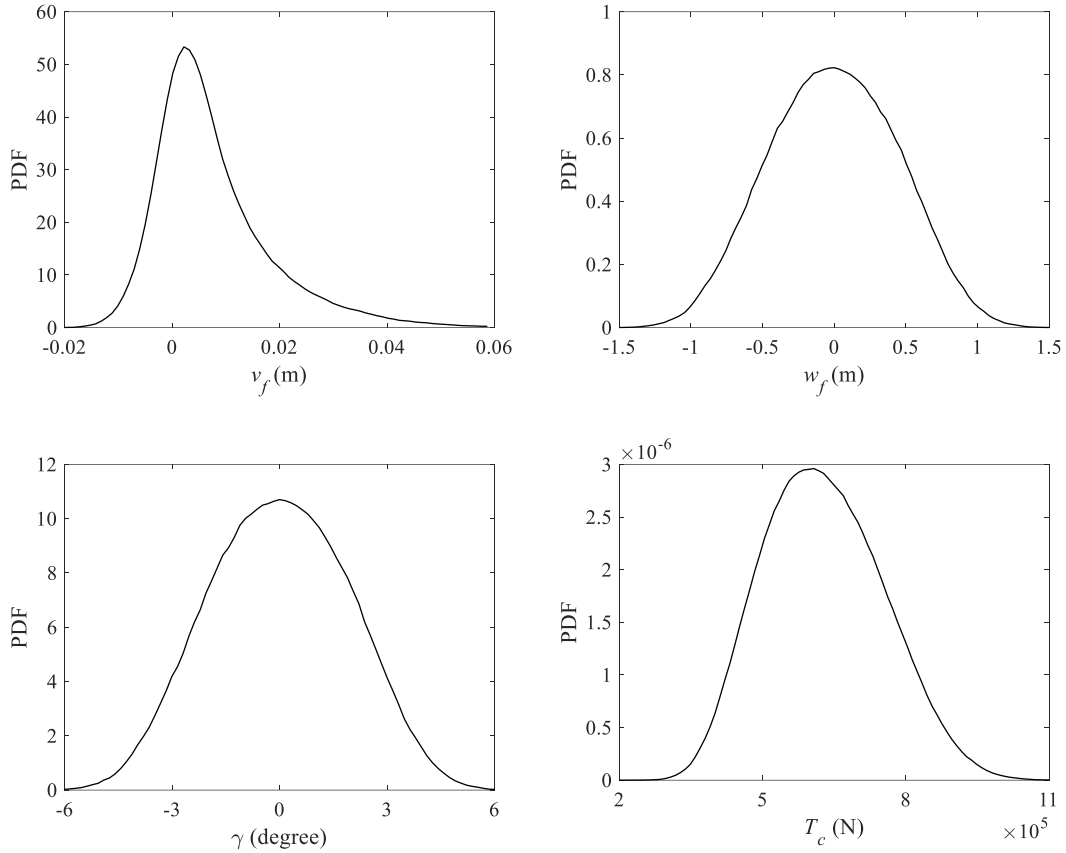
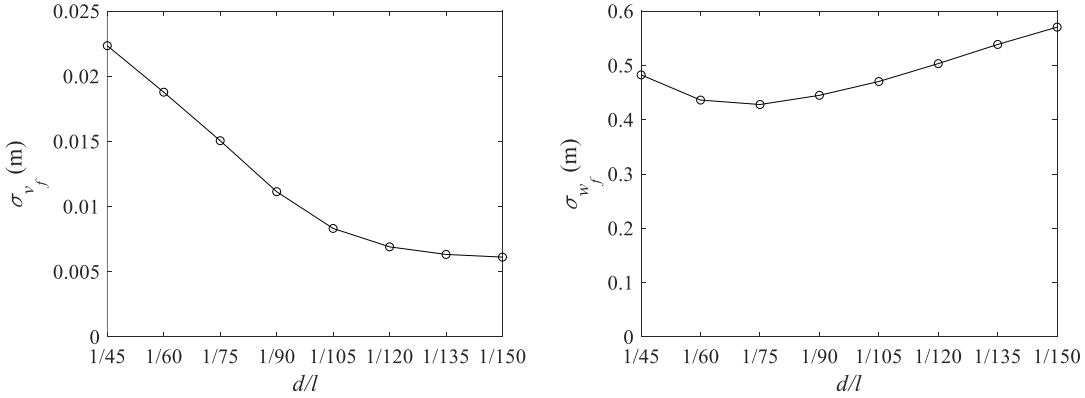



Figure 7. The PDFs of $v_f$, $w_f$, $\gamma$ and $T_c$ at steady state with $d/l = 1/90$, $\theta = 45°$

If the inclination angle of the cables keeps as $45^o$, the standard deviations of the responses of the floating platform and maximum cable tensile force at steady state are shown in Figure 8. It is observed from Figure 8 that the standard deviation of $w_f$ decreases as $d/l$ decreases from 1/45 to 1/75. Then it increases as $d/l$ further decreases from 1/75 to 1/150. This is due to the fact that as $d/l$ decreases from 1/45 to 1/75, the second natural frequency of the linear system increases from 0.94 to 1.103 rad/s, which is farther away from the dominant frequency 0.79 rad/s of $S_{\eta\eta}$. As $d/l$ decreases from 1/75 to 1/150, the fundamental natural frequency of the linear system increases from 0.586 to 0.83, which is closer to the dominant frequency of $S_{\eta\eta}$. It is also observed from that the standard deviation of $v_f$ always decreases and the standard deviations of $\gamma$ and $T_c$ always increase as $d/l$ decreases.



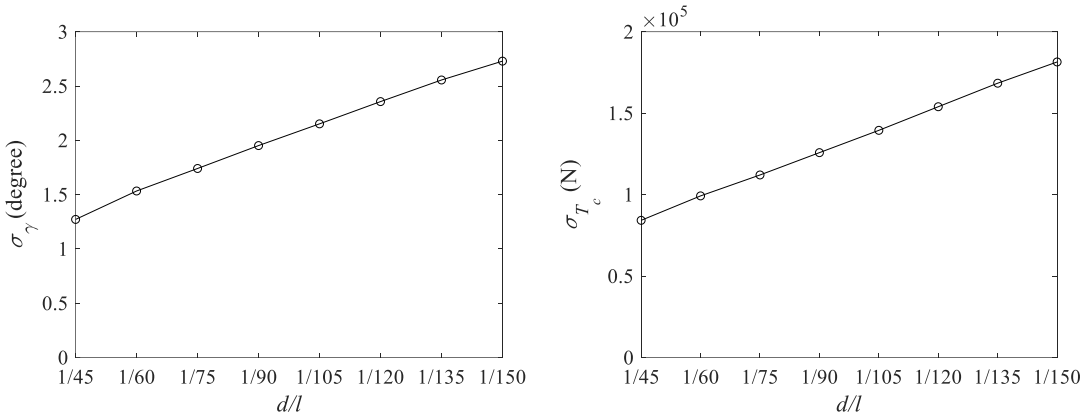



Figure 8. The standard deviations of $v_f$, $w_f$, $\gamma$ and $T_c$ at steady state for different $d/l$ with $\theta = 45°$

If the sag-to-span ratio of the cables keeps as 1/90, the standard deviations of the responses of the floating platform and maximum cable tensile force at steady state are shown in Figure 9 for different inclination angles of the cables. It is observed from Figure 9 that the standard deviations of $v_f$, $w_f$, $\gamma$, $T_c$ always increases as $\theta$ increases from 33° to 54° and they are much influenced by the inclination angles of the cables. This is due to the fact that as $\theta$ increases from 33° to 54°, the fundamental natural frequency of the linear system decreases from 0.679 to 0.593 rad/s and changes within a small interval, which is farther away from the dominant frequency of $S_{\eta\eta}$, but the second natural frequency of the linear system decreases from 1.576 to 0.832 rad/s, which is closer to the dominant frequency of $S_{\eta\eta}$.

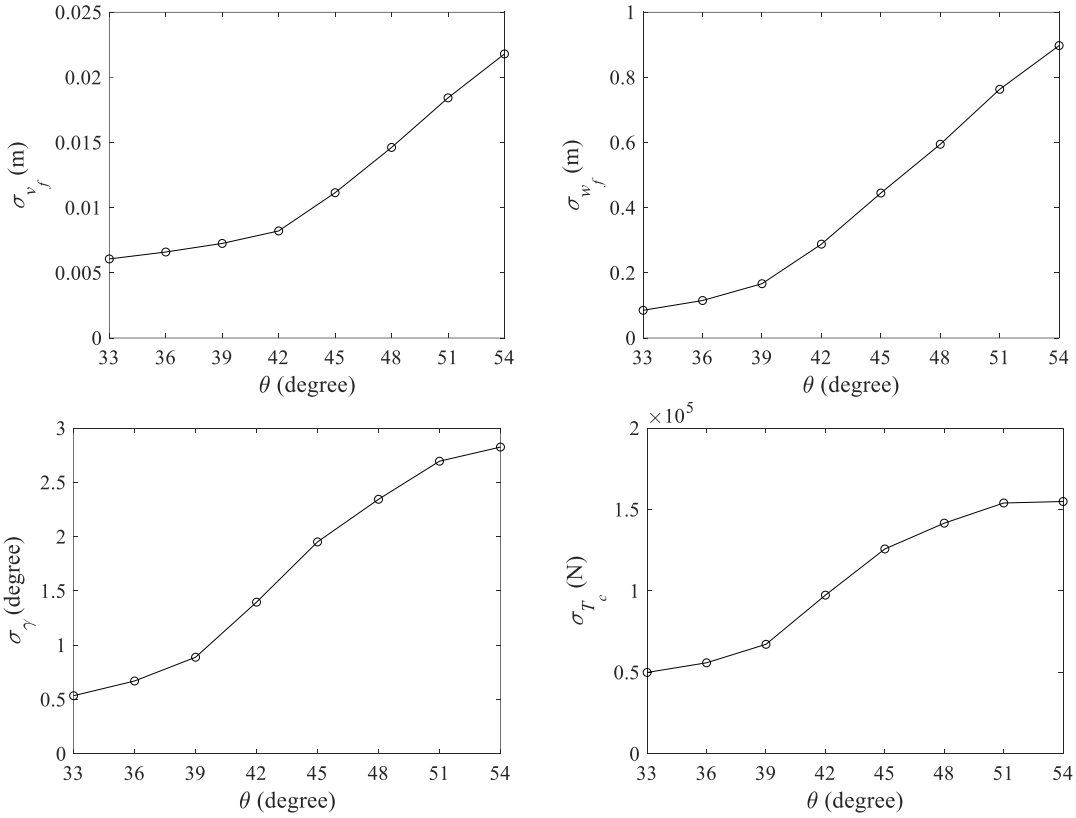



Figure 9. The standard deviations of $v_f$, $w_f$, $\gamma$ and $T_c$ at steady state for different $\theta$ with $d/l = 1/90$

## Conclusions

The nonlinear random vibrations of the cable-moored offshore floating structure are analyzed under wave excitation. The floating platform is modeled as a rigid body with six degrees of freedom. The mooring cables are modeled by using the nonlinear 3D cable elements which are formulated based on the extended Hamilton principle. The effects of added-mass and nonlinear hydrodynamic drag forces on both the floating platform and mooring cables are taken into consideration. Firstly, the equations of motion of the mooring cables and floating platform are formulated separately. After that, the connection conditions between the mooring cables and floating platform are introduced to make the nonlinear equations of motions of both the mooring cables and floating platform formulated as a whole system. The equations of motion of the whole system are solved numerically using MCS. The influences of the sag-to-span ratio and inclination angle of the mooring cables on the statistical properties of the moored floating structure and the maximum cable tensile force are studied. It is found from numerical results that the responses of the floating platform and the maximum cable tensile force are much influenced by both the initial sag-to-span ratio and inclination angle of the cables.

## Acknowledgement

The research work presented in this paper were obtained under the supports of the National Natural Science Foundation of China (China, Grant No. 51678576), the Research Committee of University of Macau (Grant No. MYRG2018-00116-FST) and the Science and Technology Development Fund of Macau (Grant No. 042/2017/A1).

### References

[1] Yamamoto, T., Yoshida, A. and Ijima, T. (1980) Dynamics of elastically moored floating objects, *Applied Ocean Research* **2**, 85–92.
[2] Tang, H. J., Chen, C. C. and Chen, W. M. (2011) Dynamics of dual pontoon floating structure for cage aquaculture in a two-dimensional numerical wave tank, *Journal of Fluids and Structures*, 27, 918-936.
[3] Esmailzadeh, E. and Goodarzi, A. (2001) Stability ananlysis of a CALM floating offshore structure, *International Journal of Non-Linear Mechanics* **36**, 917–926.
[4] Umar, A and Datta, T. K. (2003) Nonlinear response of a moored buoy, *Ocean Engineering* **30**, 1625-1646.
[5] Huang, S. (1994). Dynamic analysis of three-dimensional marine cables. *Ocean Engineering*, **21**, 587-605.
[6] Driscoll, F. R., Lueck, R. G. and Nahon, M. (2000). Development and validation of a lumped-mass dynamics model of a deep-sea rov system. *Applied Ocean Research*, **22**, 169-182.
[7] Buckham, B., Nahon, M., Seto, M., Zhao, X. and Lambert, C. (2003) Dynamics and control of a towed underwater vehicle system, part I: model development, *Ocean Engineering* **30**, 453-470.
[8] Zhu, X. Q. and Yoo, W. S. (2016) Dynamic analysis of a floating spherical buoy fastened by mooring cables, *Ocean Engineering* 121, 462–471.
[9] Garrett, D. L. (2005) Coupled analysis of floating production system, *Ocean Engineering* **32**, 802–816.
[10] Kim, B. W., Sung, H. G., Kim, J. H. and Hong, S. Y. (2013) Comparision of linear spring and nonlinear FEM methods in dynamic coupled analysis of floating structure and mooring system, *Journal of Fluid and Structures* **42**, 205–227.
[11] Pai, P. F. (2007) Highly Flexible Structure: Modeling, Computation, and Experimentation. American Institute of Aeronautics and Astronautics, Inc., Reston.
[12] Sarpkaya, T. and Isaacson M. (1981). Mechanics of wave forces on offshore structures. Van Nostrand Reinhold Co.
[13] Borgman, L. E. (1967) Ocean wave simulation for engineering design. No. HEL-9-13. University of California Berkeley Hydraulic Engineering Lab.
[14] Goda, Y. (1979) A review of statistical interpretation of wave data, Report of the Port and Harbour Research Institute, **18**, 5-32.

# Modeling penny-shaped crack using a single high order smooth element

**†*Hang Ma[1], Donghong He[2], and Zhao Guo[3]**

[1]College of Sciences, Shanghai University, China
[2]Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, China
[3]College of Civil Engineering and Urban Construction, Jiujiang University, China

†*Presenting and corresponding author: hangma@shu.edu.cn

## Abstract

A high order smooth element is constructed for modeling penny-shaped crack placed on but not limited to flat surface using a single element. The smoothness of the element is realized by repeated use of real nodes for interpolation in both the radial and circumferential directions of the element by taking advantage of geometrical features of the penny shape so that the end node/line effects existing in conventional low order elements have been removed. The technique of shape function manipulation is proposed to deal with the hyper-singular integrals in the boundary element method (BEM) for crack problems. The stress intensity factors under various loads are computed and compared with the analytical solutions in the numerical examples, showing the accuracy and effectiveness of the proposed high order smooth element.

**Keywords:** High order smooth element, Penny-shaped crack, Stress intensity factor, Hypersingular integral, Boundary element method

## Introduction

In the three-dimensional fracture analysis of structures, penny-shaped cracks have long been one of the most investigated cracks since their good representativeness to the defects in the three-dimensional solids. Conventionally, the low-order elements are employed in the numerical fracture analysis [1]-[2]. In the case of analyzing fine details near crack tips [3] or multiple-cracks [4], huge numbers of elements have to be used, resulting in large solution scale of the problem, especially for the FEM. In addition, the hypersingular integrals have to be treated carefully in the use of boundary element method (BEM). Based on the Chebyshev polynomials, Chen proposed a Gauss type quadrature to evaluate the hypersingular integrals encountered over the whole flat crack [5], much like using a single element. Recently, Gao et al. proposed a series of isoparametric closure elements [6] and improved by Ma et al. to high order smooth elements [7] for modeling closed surfaces using a single element. In the present work, the high order smooth elements are extended for modeling penny-shaped crack on flat and spherical surfaces, combined with the proposed technique of shape function manipulation to deal with the hypersingular integrals in the BEM. The stress intensity factors (SIF) under various loads are computed and compared with the analytical solutions [8], showing the accuracy and effectiveness of the high order smooth element.

## Construction of smooth element for crack

*Crack discretization*

Only the upper face of the penny-shaped crack needs to be discretized in the radial and circumferential directions. An example of the element with total nodes $N$=12 in the real and parametrical spaces are shown in Fig. 1a and 1b, respectively, where the symbols ○ and ◇ mean that the nodes are used repeatedly more than once in either circumferential or radial

directions for interpolations. The local intrinsic coordinates $\xi_1$ and $\xi_2$ correspond to the radial and circumferential lines. The double solid line Fig. 1b is the polar point of the element without placing node and the numbers with an apostrophe is for avoiding confusion of nodes which are used repeatedly. The digits in parentheses are the local counting numbers of nodes, which correlates with the global counting number, *m*, as follows

$$m = (k_1 - 1) \cdot N_2 + k_2 , \ (k_1 = 1,..., N_1 ; k_2 = 1,..., N_2) \tag{1}$$

where $N_1$ and $N_2$ represent the numbers of nodes on the radial and circumferential lines, and also the numbers of the circumferential and radial lines, respectively.

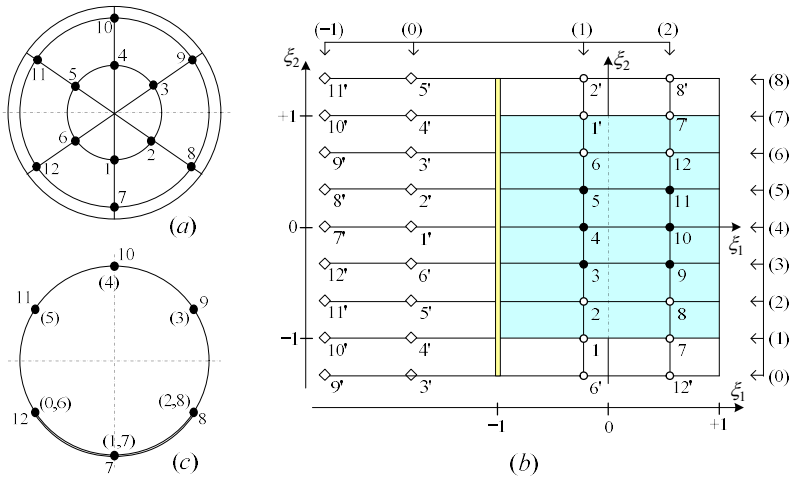



**Figure 1. Crack discretization in real space (*a*), parametrical space (*b*) and along one of the circumferential lines (*c*)**

*Shape functions*

The shape functions for the crack are formed by the product of the shape functions in both the circumferential and radial directions, just the same with the formation of closure elements [6]. However, the key idea for constructing smooth elements is the repeated use of nodes, with which the interpolation spans are expanded as shown in Fig. 1b. Along one of the circumferential lines as shown in Fig. 1c, three nodes with the global numbers 12, 7 and 8 are used twice. The digits in parentheses denote the local counting numbers from 0 to $N_2$+2, where the nodes used twice have two local counting numbers. The shape functions along circumferential lines are defined as follows

$$\phi_k^{(2)}(\xi) = l_k^{(N_2+2)}(\xi) + l_{N_2+k}^{(N_2+2)}(\xi), \ \ (k = 1,2) \tag{2a}$$

$$\phi_k^{(2)}(\xi) = l_k^{(N_2+2)}(\xi), \qquad (k = 3,..., N_2 - 1) \tag{2b}$$

$$\phi_k^{(2)}(\xi) = l_0^{(N_2+2)}(\xi) + l_k^{(N_2+2)}(\xi), \ \ (k = N_2) \tag{2c}$$

where $l_k^{(N_2+2)}$ represents the Lagrange interpolation polynomials of order $N_2$+2 as follows

$$l_k^{(N_2+2)}(\xi) = \prod_{j=0, j\neq k}^{N_2+2} \frac{(\xi - \xi_j)}{(\xi_k - \xi_j)}, \ (k = 0,1,..., N_2 + 2) \tag{3}$$

In radial directions, two auxiliary nodes denoted by ◇ as shown in Fig. 1b are supplemented across the pole so that the shape functions for crack surface are defined by

$$\phi_k^{(1)}(\xi) = l_k^{(N_1+1)}(\xi), \qquad (k = 1,..., N_1) \tag{4}$$

where $l_k^{(N_1+1)}$ represents also the Lagrange interpolation polynomials of order $N_1$+1 as follows

$$l_k^{(N_1+1)}(\xi)=\prod_{j=-1,j\neq k}^{N_1}\frac{(\xi-\xi_j)}{(\xi_k-\xi_j)},\quad (k=-1,0,...,N_1) \tag{5}$$

In this way, the shape functions for the crack surface can be obtained

$$\phi_m(\xi_1,\xi_2)=\phi_{k_1}^{(1)}(\xi_1)\cdot\phi_{k_2}^{(2)}(\xi_2),\quad (k_1=3,...,N_1) \tag{6a}$$

$$\phi_m(\xi_1,\xi_2)=\phi_{k_1}^{(1)}(\xi_1)\cdot\phi_{k_2}^{(2)}(\xi_2)+\phi_{1-k_1}^{(1)}(\xi_1)\cdot\phi_{M(k_2)}^{(2)}(\xi_2),\ (k_1=1,2) \tag{6b}$$

where $k_2=1,...,N_2$ and the counting numbers $m$, $k_1$ and $k_2$ in Eqs. (6a) and (6b) are correlated by Eq. (1). The subscript $M(k)$ is a mirror function about the pole defined as follows

$$M(k)=\begin{cases}k+N_2/2 & k\le N_2/2\\ k-N_2/2, & k>N_2/2\end{cases} \tag{7}$$

Thus in the construction of smooth element for cracks, the even number should be used for $N_2$. Although the variation spans of the two intrinsic variables $\xi_1$ and $\xi_2$ have been expanded, however, the integration spans for the smooth element remain still within [–1,+1], a shaded square region as shown in Fig. 1b. In consideration of the deformation feature of crack tip, the shape functions for the crack opening displacement (COD) take a different form as

$$\bar{\phi}_m(\xi_1,\xi_2)=\bar{\phi}_{k_1}^{(1)}(\xi_1)\cdot\phi_{k_2}^{(2)}(\xi_2),\quad (k_1=3,...,N_1) \tag{8a}$$

$$\bar{\phi}_m(\xi_1,\xi_2)=\bar{\phi}_{k_1}^{(1)}(\xi_1)\cdot\phi_{k_2}^{(2)}(\xi_2)+\bar{\phi}_{1-k_1}^{(1)}(\xi_1)\cdot\phi_{M(k_2)}^{(2)}(\xi_2),\ (k_1=1,2) \tag{8b}$$

where

$$\bar{\phi}_k^{(1)}(\xi)=\frac{\sqrt{1-\left[(1+\xi)/2\right]^2}}{\sqrt{1-\left[(1+\xi_k)/2\right]^2}}l_k^{(N_1+1)}(\xi),\qquad (k=1,...,N_1) \tag{9}$$

Therefore the smooth element for crack does not belong to the category of isoparametric elements. It need to be pointed out that the role of auxiliary nodes denoted by ◇ is nothing but to improve the fitting effect along radial lines, since the outward normal of the surface in the region indicated by ◇ turns upside-down from that of the shaded square region as shown in Fig. 1b. In addition, what needs to be emphasized is that the outward normal just at the pole is indeterminate since the circumferential line reduces to a single point at the pole. This is why no node be arranged at the pole for the crack element. It can be seen from the construction process that there is no end node and end line inside the smooth element. As a result, the interpolation accuracy of the smooth element will increase greatly with the removal of end node/line effects and the increase of the order of interpolation polynomials while the total number of nodes of the element is kept unchanged.

## Dealing with singularities of integrals

*Basic equations*

The boundary integral equation (BIE) for a crack in full space is given below [9]

$$\tau_i(y)=-\mathrm{HFP}\int_A\delta_j(x)\tau_{ij}^C(x,y)dA(x),\quad y\in A \tag{10}$$

where $A$ stands for the upper surface of the crack and $\tau_i$ the traction on $A$. $x$ and $y$ represent the field and source points, respectively. $\delta_j=u_j^+-u_j^-$ denotes the COD defined by the difference of displacements, $u_j^+$ and $u_j^-$, over the upper and lower surfaces of the crack. HFP means that the integral is evaluated in the Hadamard finite part sense. The kernel $\tau_{ij}^C$ is defined by

$$\tau_{ij}^C(x,y)=n_k(y)\tau_{ikj}^*(x,y) \tag{11}$$

In Eq. (11), $n_k$ stands for the component of the outward normal and

$$\tau^*_{ijk}(x,y)=\frac{\mu}{4\pi(1-\nu)r^3}\left\{3\frac{\partial r}{\partial n}\left[(1-2\nu)\delta_{ij}r_{,k}+\nu\left(\delta_{ki}r_{,j}+\delta_{jk}r_{,i}\right)-5r_{,i}r_{,j}r_{,k}\right]\right.$$
$$\left.+3\nu\left(r_{,j}r_{,k}n_i+r_{,k}r_{,i}n_j\right)+(1-2\nu)\left(3r_{,i}r_{,j}n_k+\delta_{ki}n_j+\delta_{jk}n_i\right)-(1-4\nu)\delta_{ij}n_k\right\} \tag{12}$$

where $\mu$ and $\nu$ are the shear modulus and Poisson ratio of the material, respectively. $r$ is the distance between the field and source points defined as

$$r=\sqrt{(x_k-y_k)(x_k-y_k)} \tag{13}$$

In a special case of a flat crack placed in the plane $x_1$-$x_2$ with only the normal load applied in $x_3$ direction, the expressions (10)-(12) can be written in much simplified forms [5][8]. However, in the present work, these forms are kept unchanged to cope with the general cases.

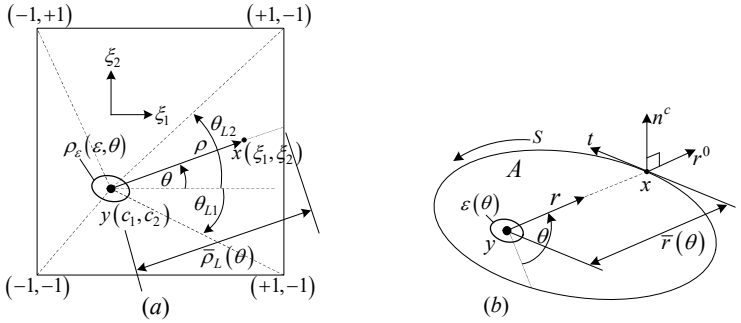



**Figure 2. Domains for evaluating strong-singular (*a*) and hypersingular (*b*) integrals**

*Shape function manipulation and evaluation of strong-singular integral*

Now rewrite Eq. (10) after discretizing the COD of the crack using the shape functions (8a) and (8b) as follows

$$\tau_i(y)=-\sum_{m=1}^{N}\delta_j^m I_{ij}^m \tag{14}$$

$$\delta_k(x)=\sum_{m=1}^{N}\delta_k^m\bar{\phi}^m\left[\xi_1(x),\xi_2(x)\right] \tag{15}$$

where

$$I_{ij}^m=\mathrm{HFP}\int_A\bar{\phi}^m\left[\xi_1(x),\xi_2(x)\right]\tau_{ij}^C(x,y)dA(x)$$
$$=\lim_{(\xi_1,\xi_2)\to(c_1,c_2)}\int_{-1}^{+1}\int_{-1}^{+1}\bar{\phi}^m(\xi_1,\xi_2)\tau_{ij}^C(\xi_1,\xi_2)J(\xi_1,\xi_2)d\xi_1 d\xi_2 \tag{16}$$

It needs to be pointed out that in Eq. (16) the upper crack surface $A$ is discretized using the shape functions (6a) and (6b), different from that for the COD. The integrals (16) are to be evaluated in polar systems as shown in Fig. 2a as follows

$$I_{ij}^m=\lim_{\rho_\varepsilon\to 0}\sum_L\int_{\theta_{L1}}^{\theta_{L2}}\int_{\rho_\varepsilon}^{\bar{\rho}_L(\theta)}\bar{\phi}^m(\rho,\theta)\tau_{ij}^C(\rho,\theta)J(\rho,\theta)\rho d\rho d\theta \tag{17}$$

where $(c_1,c_2)$ stand for the local coordinates of the source point $y$. Introduce the shape function manipulation as

$$\Delta\bar{\phi}^m=\bar{\phi}^m(\rho,\theta)-\bar{\phi}^m(c_1,c_2) \tag{18}$$

Rewrite integrals (17) as

$$I_{ij}^m=\lim_{\rho_\varepsilon\to 0}\sum_L\int_{\theta_{L1}}^{\theta_{L2}}\int_{\rho_\varepsilon}^{\bar{\rho}_L(\theta)}\Delta\bar{\phi}^m\tau_{ij}^C(\rho,\theta)J(\rho,\theta)\rho d\rho d\theta+\bar{\phi}^m(c)\lim_{\rho_\varepsilon\to 0}\sum_L\int_{\theta_{L1}}^{\theta_{L2}}\int_{\rho_\varepsilon}^{\bar{\rho}_L(\theta)}\tau_{ij}^C(\rho,\theta)J(\rho,\theta)\rho d\rho d\theta$$

$$= I_{ij\text{CPV}}^{m} + \bar{\phi}_m(c_1,c_2)\lim_{x\to y}\int_A \tau_{ij}^{C}(x,y)\,dA(x) = I_{ij}^{m\text{CPV}} + \bar{\phi}^m(c_1,c_2)\,I_{ij}^{\text{HFP}} \tag{19}$$

by subtracting then adding back a shape function at the singular source point. It is seen that the first integral $I_{ij}^{m\text{CPV}}$ at the right hand side of Eq. (19) is reduced to strong-singular integrals to be evaluated in the sense of Cauchy principal value (CPV). Expand the integrand of $I_{ij}^{m\text{CPV}}$ in (19) in truncated Laurent expansion at a fixed angle $\theta$ as follows

$$\tau_{ij}^{C}(\rho,\theta)\Delta\bar{\phi}^m J(\rho,\theta)\rho = \frac{1}{\rho}\tau_{ij}^{0}(\rho,\theta)\frac{\Delta\bar{\phi}^m}{\rho}J(\rho,\theta)\left(\frac{r}{\rho}\right)^{-3} = \frac{A_{ij}^{m}(\theta)}{\rho} + O(1) \tag{20}$$

where

$$\tau_{ij}^{0}(x,y) = r^3\tau_{ij}^{C}(x,y) \tag{21}$$

$$A_{ij}^{m}(\theta) = \lim_{\rho\to 0}\tau_{ij}^{0}(\rho,\theta)\frac{\Delta\bar{\phi}^m}{\rho}J(\rho,\theta)\left(\frac{r}{\rho}\right)^{-3} = \tau_{ij}^{0}(\theta)\bar{\phi}_{\rho}^{m}(\theta)J^{-2}(\theta) \tag{22}$$

$$\bar{\phi}_{\rho}^{m}(\theta) = \lim_{\rho\to 0}\frac{\Delta\bar{\phi}^m}{\rho} = \left[\frac{\partial\bar{\phi}^m}{\partial\xi_1}\cos\theta + \frac{\partial\bar{\phi}^m}{\partial\xi_2}\sin\theta\right]_{\rho=0} \tag{23}$$

In this way, the strong-singular integrals $I_{ij}^{m\text{CPV}}$ can be evaluated using the conventional Gauss quadrature in the following form

$$\begin{aligned} I_{ij}^{m\text{CPV}} &= \sum_L \int_{\theta_{L1}}^{\theta_{L2}} \left\{ \int_0^{\bar{\rho}_L(\theta)} \left[\Delta\bar{\phi}^m\tau_{ij}^{C}(\rho,\theta)J(\rho,\theta)\rho - \frac{A_{ij}^{m}(\theta)}{\rho}\right]d\rho + A_{ij}^{m}(\theta)\lim_{\rho_\varepsilon\to 0}\int_{\rho_\varepsilon}^{\bar{\rho}_L(\theta)}\frac{d\rho}{\rho}\right\}d\theta \\ &= \sum_L \int_{\theta_{L1}}^{\theta_{L2}}\int_0^{\bar{\rho}_L(\theta)}\left[\Delta\bar{\phi}^m\tau_{ij}^{C}(\rho,\theta)J(\rho,\theta)\rho - \frac{A_{ij}^{m}(\theta)}{\rho}\right]d\rho d\theta + \int_0^{2\pi} A_{ij}^{m}(\theta)\ln\left[\bar{\rho}_L(\theta)J_0(\theta)\right]d\theta \end{aligned} \tag{24}$$

*Evaluation of hypersingular integral*

The second integrals $I_{ij}^{\text{HFP}}$ at the right hand side of Eq. (19) have no shape function, resulted from the shape function manipulation stated previously. Noticed that the kernels appeared in these integrals describe a divergence-free field [10], having the properties of

$$\lim_{x\to y}\oint_\Gamma \tau_{ij}^{C}(x,y)\,d\Gamma(x) = 0 \tag{25}$$

over a closed surface, suggesting that the integrals $I_{ij}^{\text{HFP}}$ are surface independent, which can be and should be made use of. As shown in Fig. 2b, when the point *x* move along the boundary *S* of the crack, the straight line connecting *x* and *y*, or the generatrix, will form a new surface, over which the evaluation of integrals $I_{ij}^{\text{HFP}}$ can be carried out instead of the original crack surface *A*. The shape of the new surface would be planar or conical, depending on whether the original crack is a flat or curved surface. In either of the cases, however, the component of the kernel $\tau_{ij}^{C}$ on the generatrix is constant at a fixed angle $\theta$ because $n^c$ $\left(\vec{n}^c = \vec{r}^{\,0}\times\vec{t}\right)$ and $r_{,k}$ are all constants on the generatrix, where $n^c$ is the outward normal of the new surface, $r^0$ the unit vector in *r* direction. See Fig. 2b and Eqs. (11)-(12). Therefore

$$\begin{aligned} I_{ij}^{\text{HFP}} &= \lim_{x\to y}\int_A \tau_{ij}^{C}(x,y)\,dA(x) = \lim_{\varepsilon\to 0}\int_0^{\alpha}\int_{\varepsilon}^{\bar{r}(\theta)} r^{-3}\tau_{ij}^{0}(\theta)\,r\,dr\,d\theta \\ &= -\int_0^{\alpha}\bar{r}^{-1}(\theta)\tau_{ij}^{0}(\theta)\,d\theta + \lim_{\varepsilon\to 0}\int_0^{\alpha}\varepsilon^{-1}(\theta)\tau_{ij}^{0}(\theta)\,d\theta \end{aligned} \tag{26}$$

where $\alpha$ is the solid angle of the cone surface and in particular $\alpha = 2\pi$ for planar surface. For a physical problem, the integral above should exist, which means that the infinite term or the last term at the right hand side of Eq. (26) should be eliminated or should be cancelled out by free terms [11]. Finally, the integrals $I_{ij}^{\text{HFP}}$ can be evaluated using the conventional Gauss quadrature in the following form

$$I_{ij}^{\mathrm{HFP}} = -\oint_S \bar{r}^{-2}(x)\,\tau_{ij}^{0}(x,y)\,dS(x) \tag{27}$$

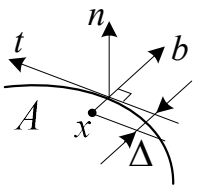
## Numerical examples

*Computation of SIF*

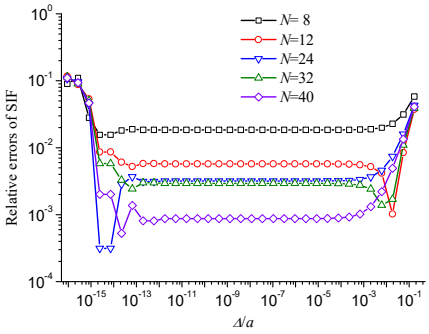
In the numerical examples, the SIFs, $K_1$, $K_2$ and $K_3$ are computed by the corresponding COD values, $\delta^n$, $\delta^b$ and $\delta^t$ in the open, shear and tear modes, respectively, at the point $x$ in the local coordinate system $\left(\vec{b}=\vec{t}\times\vec{n}\right)$ as shown in Fig. 3a using the following approximate equations

$$K_1 = \frac{E}{4\left(1-\nu^2\right)}\sqrt{\frac{\pi}{2\Delta}}\delta^n \;,\; K_2 = \frac{E}{4\left(1-\nu^2\right)}\sqrt{\frac{\pi}{2\Delta}}\delta^b \;,\; K_3 = \frac{E}{4\left(1-\nu^2\right)}\sqrt{\frac{\pi}{2\Delta}}\delta^t \tag{28}$$

where $\Delta$ represents a small distance from the point $x$ to the crack front and $E$ is Young's modulus of the material. In the situation that the crack surfaces are traction free in full space under far-field loads, the computation model needs to be decomposed into two parts in the numerical analysis of such cracks. In the first part of the model, the tractions equal and opposite to the applied loads are acting on the crack surfaces without the far-field loads. In the second part of the model, the full space is loaded by the far-field loads without cracks. The final response is the linear superposition of the two parts of the model. However, the second part has no direct contribution to the values of SIF.




(*a*)




(*b*)

**Figure 3. Local coordinate system (*a*) and errors of computed SIF as function of $\Delta$ (*b*)**

*Suitable distance check*

Firstly, the suitable distance $\Delta$ is checked by a penny-shaped crack of radius $a$ placed in the plane $x_1$-$x_2$ ($x_3$=0) in full space under a far-field unit normal load in $x_3$ direction, discretized using a single smooth element with a series of total node numbers from $N$=8 to $N$=40. The relative errors of computed SIF are depicted in Fig. 3b as a function of $\Delta$, showing that the accuracy is satisfactory with such few nodes using a single smooth element. The results are fairly stable in a wide range of $\Delta$ so that $\Delta/a$=$10^{-4}$ is adopted in the following examples.

*SIF under various normal loads*

Secondly, the SIFs of the penny-shaped crack in full space under various far-field normal loads are computed using a single smooth element with the total node number $N$=32 and compared with the analytical solutions [8]. The normal loads are expressed by

$$\sigma_0 = c_0 + c_1\frac{x_1}{a} + c_2\frac{x_2}{a} + c_3\frac{x_1x_2}{a^2} + c_4\left(\frac{x_1}{a}\right)^2 + c_5\left(\frac{x_2}{a}\right)^2 \tag{29}$$

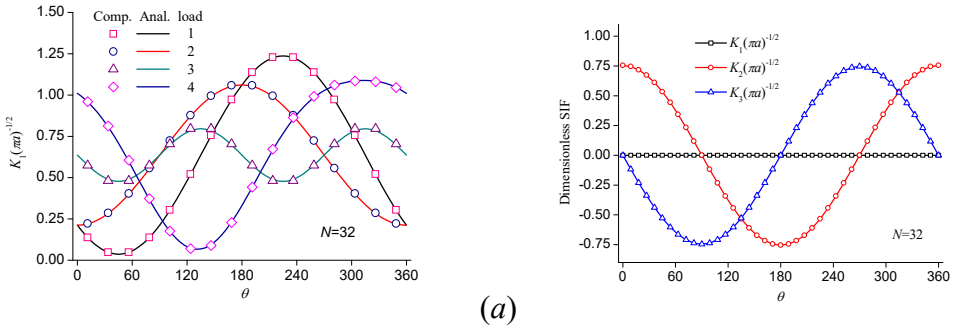
The coefficients for the 4 kinds of loads are listed in Table 1. The SIFs along the crack front are computed and shown in Fig. 4a while the angle $\theta$ is starting from the positive direction of $x_1$. It is seen from Fig. 4a that the computed results are in good agreement with those of analytical solutions, showing the accuracy and effectiveness of the proposed high order smooth element for the crack.

**Table 1. Coefficients in Eq. (29)**

| Load | $c_0$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ |
|---|---|---|---|---|---|---|
| 1 | 1 | -1 | -1 | 0 | 0 | 0 |
| 2 | 1 | -1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | -1 | 0 | 0 |
| 4 | 1 | 0.8 | -0.9 | 0.5 | 0.08 | 0.02 |

*SIF under unit shear load*

Next example considers the penny-shaped crack in full space under far-field unit shear load in $x_1$ direction using a single smooth element with the total node number $N$=32. It is seen from Fig. 4b that the computed SIF in shear and tear modes ($K_2$ and $K_3$) varies along the crack front while the SIF in open mode ($K_1$) keeps zero throughout as expected.

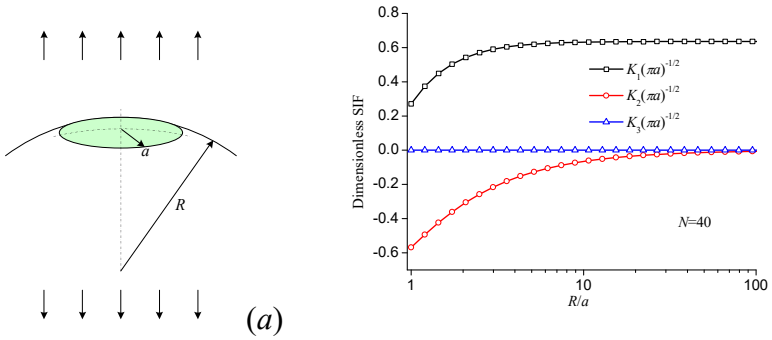



(*a*) (*b*)

**Figure 4. Comparison of computed SIF under various far-field normal loads (*a*) and computed SIF under far-field unit shear load (*b*)**

*SIF of cracks on spherical surface*

In the last example, the curved penny-shaped cracks on a spherical surface of radius $R$ are computed under far-field unit normal load as shown in Fig. 5a using a single smooth element with the total node number $N$=40. The computed SIFs are presented in Fig. 5b with $R/a$, where the shape of the crack becomes a half-sphere when $R/a$=1 but behave towards a flat crack when $R/a$ is very large. It is shown from Fig. 5b that that the SIF in open and shear modes ($K_1$ and $K_2$) increase gradually with $R/a$ while the SIF in tear mode ($K_3$) keeps zero throughout as expected.




(*a*) (*b*)

**Figure 5. Curved penny-shaped cracks on a spherical surface (*a*) with computed SIF (*b*) under far-field unit normal load**

## Conclusions

In the present work, a high order smooth element is constructed successfully for modeling penny-shaped crack placed on flat or curved surface using a single element. By making full use of geometrical features such as the symmetry and periodicity, the smoothness of the element is realized by repeated use of nodes in the radial and circumferential lines for interpolation. As a result, the accuracy of the crack modeling increases because of the raise of the order of interpolation polynomials as well as the removal of the end node/line effects existing in conventional low order elements. The technique of shape function manipulation is proposed to deal with the hyper-singular integrals in the BEM for crack problems. In the numerical examples, the accuracy and effectiveness of the proposed high order smooth element and the technique for hyper-singular integrals are verified by the computed SIFs, using a single element with such few nodes, compared with the analytical solutions.

**Acknowledgement:** The support of the present work by National Natural Science Foundation of China (Nos.11672173, 11272195, 11662005) is acknowledged.

### References

[1] Cruse, T.A. (1988) *Boundary Element Analysis in Computational Fracture Mechanics*, Kluwer Academic Publishers, Dordrecht, Netherlands.

[2] Aliabadi, M.H. and Rooke, D.P. (1991) *Numerical Fracture Mechanics*, Computational Mechanics Publications, Southampton, UK.

[3] Wang, X. (2004) Elastic T-stress solutions for penny-shaped cracks under tension and bending, *Engineering Fracture Mechanics* **71**, 2283–2298.

[4] Tsang, D.K.L., Oyadiji, S.O. and Leung, A.Y.T. (2004) Multiple penny-shaped cracks interaction in a finite body and their effect on stress intensity factor, *Engineering Fracture Mechanics* **70**, 2199–2214.

[5] Chen, Y.Z. and Lee, K.Y. (2001) Numerical solution of three-dimensional crack problem by using hypersingular equation, *Computer Methods in Applied Mechanics and Engineering* **190**, 4019–4026.

[6] Gao, X.W., Yuan, Z.C., Peng, H.F., Cui. M. and Yang, K. (2016) Isoparametric closure elements in boundary element method, *Computers and Structures* **168**, 1–15.

[7] Ma, H., Tian, Y. and He, D.H. (2019) High order isoparametric elements in boundary element method—smooth elliptical element, *Engineering Analysis with Boundary Elements* **101**, 34–47.

[8] Atroshchenko, E., Potapenko, S. and Glinka, G. (2009) Stress intensity factor for an embedded elliptical crack under arbitrary normal loading, *International Journal of Fatigue* **31**, 1907–1910.

[9] Telles, J.C.F., Castor, G.S. and Guimaraes, S. (1995) A numerical Green's function approach for boundary elements applied to fracture mechanics, *International Journal for Numerical Methods in Engineering* **38**, 3259–3274.

[10] Ma, H. and Kamiya, N. (2002) A general algorithm for the numerical evaluation of nearly singular boundary integrals of various orders for two- and three-dimensional elasticity, *Computational Mechanics* **29**, 277–288.

[11] Gao, X.W. (2010) An effective method for numerical evaluation of general 2D and 3D high order singular boundary integrals, *Computer Methods in Applied Mechanics and Engineering* **199**, 2856–2864.

# Nonlinear Interaction of Internal Waves Due to Two Point Vortices

**Zhen Wang**[1] **and Di Liu**[1]

[1]School of Mathematical Science, Dalian University of Technology, Dalian 116024 China

## Abstract

In this paper, two-dimensional, two-layer steady stratified flow about a equal-strength counter- or co-rotating vortex pair in the lower layer is concerned. Potential flow theory and boundary integral equation method are applied to establish boundary integral equations about the interfacial wave. These equations are solved numerically based on quasi-Newton method. The effects on interfacial wave profiles of distance of the vortex pair are analyzed. It is found that the wave height oscillates with $d$ and the extreme values are almost the sum of that for two vortices consists of the pair, especially for $d$ large enough. When $d$ is set that the wave height gets its maximum points the wave profiles is about the superposition of that for the two vortices, whereas when $d$ get its minimum points the difference between the wave profiles caused by the pair and the sum of profiles for these two vortices is large.

Keywords: **Stratified flow, Point vortex pair, Boundary integral equation, Nonlinear interfacial waves**

## Introduction

When there exists a disturbance source in the steady density-stratified fluids, an internal wave will be generated. Two-layer model of the vertical structure with different densities are often employed as a simplified model of internal waves of steady density-stratified fluids. Many researchers have studied the internal waves generated by various disturbance sources in two-layer fluids such as the moving point source in the upper[1] and lower layers[2], the moving dipole [3], the point vortex in the lower layer[4] and upper layer[5] and the hydrofoil in the lower layer[6].

Forbes applied this boundary integral equation method with arclength parameters to describe the surface while studying non-linear surface wave caused by a submerged point vortex [7] and a submerged hydrofoil[8] in two-dimensional ideal irrotational fluid of infinite depth. The obtained equations were solved numerically based on Newton method. Then this theory is used to study the interfacial waves in two-layer fluids by different disturbance by following researcher[9], [4].

The vortex pair is the basic element of fluid mechanics. Study of it to a large extent comes from the problem of trailing wakes. Many studies are concerned with dynamics and instability of vortex pairs. The literature [10] reviewed the characteristics and the behaviors of vortex pairs. Besides, some researchers focused on the interactions and the flow structures between vortex pairs and other objects like wall[11] and free surface[12],[13].

In this paper, the two-dimensional two-layer steady flow for a submerged vortex pair is con-

sidered. Both layers are inviscid and incompressible ideal fluids with consistent flow direction. The upper layer is of finite depth and bounded by a rigid lid, while the lower fluid is infinitely deep in which there exists a vortex pair set on a horizontal fixed position. The structure of this paper is as follows: at first integral-differential equations are established using the potential flow theory and boundary integral equation method. Secondly the problem is solved numerically based on the quasi-Newton method, which has been verified and gives a well performance in DoF. Then we compare the effects of different parameters on the wave profile, including Froude number, vortex strength and distance between two vortices.

## Model of the problem

Consider steady two-layer fluids of different densities. Both layers are ideal fluids and irrotational. Their upstream uniform speeds have consistent flow direction. Creating a Cartesian coordinate system such that the $x$ axis is placed at the undisturbed horizontal interface and point in the same direction of upstream uniform speed, as well as the $y$ axis points up vertically. The depth of upper fluid is $T$ and the upper surface satisfies the rigid-lid assumption. The lower fluid is infinitely deep with a point vortex pair placed where its center is at $(0, -H)$ . The distance between the two point vortices is $2D$ with circulation $K_1 < 0$ at $(-d, -1$ and $K_2 > 0$ at $(d, -1)$ respectively. In following context we use subscripts 1 and 2 to represent the physical variables associated with the upper fluid and the lower, respectively. Densities and upstream uniform speeds of two layers of fluid are $\rho_1, \rho_2$, and $\gamma_1, \gamma_2$.

For the convenience of discussion, use $\gamma_2$ as the speed scale, $H$ the length scale to get the dimensionless model, then introduce following dimensionless parameters:

$$F = \frac{\gamma_2}{\sqrt{gH}},\ \epsilon_1 = \frac{K_1}{\gamma_2 H},\ \epsilon_2 = \frac{K_2}{\gamma_2 H}\rho = \frac{\rho_1}{\rho_2},\ \gamma = \frac{\gamma_1}{\gamma_2},\ \lambda = \frac{T}{H}\ d = \frac{D}{H}$$

where $F$ is the Froude number, $\epsilon_1, \epsilon_2$ are the dimensionless vortex strengths of the two point vortices, $\rho$ is the ratio of density, $\gamma$ is the ratio of far upstream uniform speed, $\lambda$ is the nondimensional depth of the upper layer, and $2d$ is the nondimensional distance between two vortices. The elevation of fluid interface is described by a function $y = \eta(x)$.

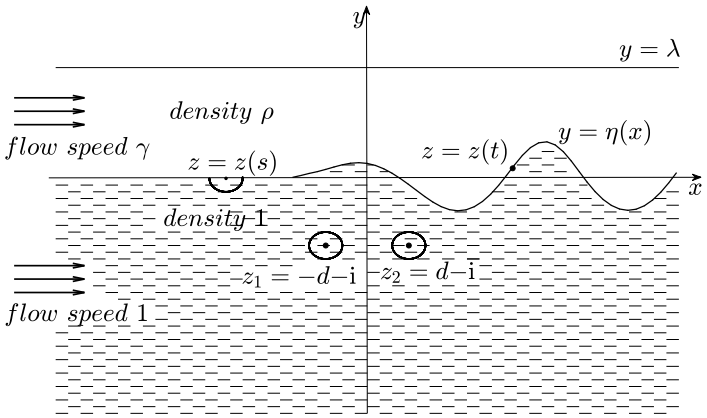



**Figure 1. The nondimensional problem of a two-layer flow about a submerged vortex pair located horizontally.**

Because two layer fluids are both ideal fluids and flows are irrotational, from potential flow theory two potential functions $\phi_1, \phi_2$, the stream functions $\psi_1, \psi_2$ satisfy the Cauchy-Riemann

equation: $(\partial/\partial x)\phi_j = (\partial/\partial y)\psi_j$, $(\partial/\partial y)\phi_j = -(\partial/\partial x)\psi_j$ $j = 1, 2$. Thus two analytic functions $f_j(z) = \phi_j(x, y) + \mathrm{i}\psi_j(x, y)$, $z = x + \mathrm{i}y$, complex velocity potential functions for upper and lower fluid separately are introduced. $z_1 = -d - \mathrm{i}$, $z_2 = d - \mathrm{i}$, the position of the two point vortices, are two singularities of $f_2$ where i is the imaginary unit, $\mathrm{i}^2 = -1$. $f_2$ satisfies

$$f_2 \to z + \frac{i\epsilon_1}{2\pi}\ln(z - z_1) + \frac{i\epsilon_2}{2\pi}\ln(z - z_2), \quad z \to z_1, z_2 \tag{1}$$

at $z_1$ and $z_2$. The upstream conditions are

$$f_1 \to \gamma z,\ f_2 \to z, \quad \mathrm{Re}[z] \to -\infty \tag{2}$$

here $\mathrm{Re}[z]$ means the real part of $z$.

The kinematic boundary condition for upper surface

$$\nabla\phi_1 \cdot \mathbf{n} = 0. \tag{3}$$

At the interface $y = \eta(x)$ it is

$$\nabla\phi_j \cdot \mathbf{n} = 0,\ j = 1, 2, \tag{4}$$

where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$.

Introduce the arclength parameter $s$ to parameterize the fluid interface $y = \eta(x)$, so the fluid interface is represented as $(x, y) = (x(s), y(s))$. The arclength condition is

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1. \tag{5}$$

The fluid interface satisfies the Bernoulli equation

$$\rho\left(\frac{d\phi_1}{ds}\right)^2 - \left(\frac{d\phi_2}{ds}\right)^2 + \frac{2(\rho - 1)y}{F^2} = \rho\gamma^2 - 1, \tag{6}$$

according to setting the pressure and velocity equally on the interface. For more details can be found in [4] and referring in it.

Choose a fixed point $z = z(s) = x(s) + \mathrm{i}y(s)$ randomly on the fluid interface, where $s$ corresponding arclength parameters. The Laplace equation for the upper and lower layer fluid could be replaced by $\oint_\Gamma f(z)dz = 0$ and $\sum \mathrm{Res} f, z_k$. Both two integral equations come form Cauchy integral and residue theorem with respect to integral contour-path on $z_1$ and $z_2$. Writing two functions $\phi_j' s$ as forms of integral equations, two integral-differential equations are established for the upper and lower layers by the boundary integral equation method The detailed derivation process is similar to [4].

Introduce the analytic function $G_1(z) = \frac{df_1}{dz} - \gamma$, then apply the Cauchy integral formula and take its imaginary part to establish the governing equation. Applying the equality

$$AB = A \cdot \bar{B} + \mathrm{i}\bar{A} \times B = (ac - bd) + \mathrm{i} \begin{vmatrix} a & -b \\ c & d \end{vmatrix}$$

where $A = a + \mathrm{i}b, B = c + \mathrm{i}d \in \mathbb{C}$, the governing equation can be written as follow

$$\begin{aligned} \pi(\gamma - x'(s)\phi_1'(s)) = \mathrm{Im}\left\{ \fint_{-\infty}^{+\infty} \overline{G_1(z(t))} \times d(\ln(z(t) - z(s))) \right\} \\ + \mathrm{Im}\left\{ \int_{-\infty}^{+\infty} \overline{G_1(\tilde{z}(t))} \times d(\ln(\tilde{z}(t) - z(s))) \right\} \end{aligned} \tag{7}$$

where $\tilde{z}(t) = x(t) + i(2\lambda - y(t))$ is the mirror point of $z(s)$ about $y = \lambda$. The first integral on the right side of the equation is the singular integral in the sense of the Cauchy principal value.

For the lower layer, introduce the analytic function $G_2(z) = \frac{df_2}{dz} - 1$ similarly, apply the residue theorem, and the equation $\oint_{\Gamma_2} \frac{G_2(\xi)d\xi}{\xi - z(s)} = 2\pi\mathrm{i} \sum_{k=1}^{2} \mathrm{Res}\left\{ \frac{G_2(\xi)}{\xi - z(s)}, z_k \right\}$ is obtained, where $z_1 = -d - \mathrm{i}$, $z_2 = d - \mathrm{i}$.

Calculating the residue and integrals in the above formula and taking the imaginary part. The residue contribution obeys the superposition law. The lower layer fluids following governing equation

$$\begin{aligned} \pi(x'(s)\phi_2'(s) - 1) =& \mathrm{Im}\left\{ \fint_{-\infty}^{+\infty} \overline{G_2(z(t))} \times d(\ln(z(t) - z(s))) \right\} \\ &+ \frac{\epsilon_1 \left(y(s) - \mathrm{Im}\, z_1\right)}{|z(s) - z_1|^2} + \frac{\epsilon_2 \left(y(s) - \mathrm{Im}\, z_2\right)}{|z(s) - z_1|^2} \end{aligned} \tag{8}$$

Whereas the vortex pair contains two isolated singular points $z_1$ and $z_2$, which requires two calculation of residual number and add one more term in the governing equation.

The governing equations (5), (6), (7) and (8) are derived. Based on them and the corresponding boundary condition (2), the unknowns $x(s)$, $y(s)$, $\phi_1(s)$ and $\phi_2(s)$ can be calculated.

**Numerical procedure**

The numerical calculation method is similar to [4], while the difference is that this paper applies the quasi-Newton iteration method [14, 15] to solve (5), (6), (7) and (8). If $y'$ is determined , then from (5) $x'$ is obtained, as well as $x(s) = \int_{-\infty}^{s} x'(t)dt$ and $y(s) = \int_{-\infty}^{s} y'(t)dt$ are also acquired. (7) can be written as integral equations with respect of $\phi_1^{'}$, then from (6) $\phi_2^{'}$ can be solved easily. As $y'$ is unknown, take the approximation of $y'$ as $\widetilde{y'}$ and (8) as the cost function to update the approximation with quasi-Newton method. The following is the detailed process.

The integral area $(-\infty, +\infty)$ is truncated to the finite interval $[s_1, s_N]$, then $N-1$ equally dividing it to get $N$ grid points $s_k = s_1 + (k-1)\Delta s, k = 1, \ldots N$. Here $\Delta s = (s_N - s_1)/(N-1)$ means the step size. $x_k$, $y_k$, $x_k'$, $y_k'$, $\phi_{1,k}^{'}$, $\phi_{2,k}^{'}$ are the approximation value of the responding unknown quantities. According to the infinity boundary condition (2), the equation (5) and (6) determine $y_1 = y_1' = 0,$, $x_1' = 1$, $x_1 = s_1$, $\phi_{1,1}' = \gamma$, $\phi_{2,1}' = 1$. To eliminate the effects of singularity in the integral, half grid points at at$s_{k-1/2} = (s_{k-1} + s_k)/2$: $x_{k-\frac{1}{2}}$, $y_{k-\frac{1}{2}}$, $x'_{k-\frac{1}{2}}$, $y'_{k-\frac{1}{2}}$, $\phi'_{1,k-\frac{1}{2}}$, $\phi'_{2,k-\frac{1}{2}}$,

$k = 2, \ldots, N$ are also calculated. Here $x_{k-\frac{1}{2}} = (x_{k-1} + x_k)/2$. So as other variables. The initial approximation of $y_2', \cdots, y_N' = 0$.

As $y_2', \cdots, y_N'$ is determined, calculate numerical integration on the finite interval $[s_1, s_N]$ with trapezoidal rule to get discrete equations $A[\phi_{1,1}', \ldots, \phi_{1,N}']^T = c$ while taking the approximation value of quantities at $s$ as its value in (7). Dealing with (8) in a similar way yields the equations of matrix form $E(y_2', \cdots, y_N') = B[\phi_{2,1}', \ldots, \phi_{2,N}']^T - d$, where $\phi_{2,k}'$ calculated from Bernoulli's equation (6). In these equations coefficient matrices $A, B$, constant terms $b, d$ are all concerned with $y_k'$, $x_k'$, $x_k$, $y_k$ , which can be calculated by (5) and trapezoidal rules: Finally, we get a system of equations for $y_2', \ldots, y_N'$. Solve it applying quasi-Newton method, and the iteration formula is[14]:

$$\begin{cases} u_{i+1} = u_i - A_i^{-1} E(u_i), \\ A_{i+1} = A_i + (b_i - A_i s_i) s_i^T / (s_i^T s_i) \end{cases} \qquad i = 0, 1, 2, \ldots \tag{9}$$

where $u = (y_2', \ldots, y_N')$, $u_i$ represents $i$th iteration approximation of $u$, $E(u) = (E_2[y_2', \ldots, y_N'], \ldots, E_N[y_2', \ldots, y_N'])$, $s_i = u_{i+1} - u_i$, and $b_i = E(u_{i+1}) - E(u_i)$. For $i$=0, $A_0$ could be chosen as $\left((E^T(y' + he_i) - E^T(y'))/h\right)$, the $(N-1 \times N-1)$ difference matrix of cost function $E$, $i = 2, \cdots, N$, where $y' = [y_2', \ldots, y_N']$ and $e_i$ is $N-1$ dimensional unit vector. This Calculating progress terminates when Calculate $\|\mathbf{E}\|_2$ $\|\mathbf{E}\|_2$ is less than the given number $\varepsilon$.

**Results analysis**

In the numerical calculation, the upper depth is set $\lambda = 20$, and far upstream uniform speed ratio $\gamma = 1$, namely two-layer fluids with equal spped. If we set $d = 0$, $\epsilon_1 < 0, \epsilon_2 = 0$, in fact it is the case for a single vortex $\epsilon < 0$. The calculation error precision is setting to be $\sigma = 10^{-9}$. The calculation domain is [-25, 30] and the grid number $N = 2201$, as well as $\Delta s = 0.025$.

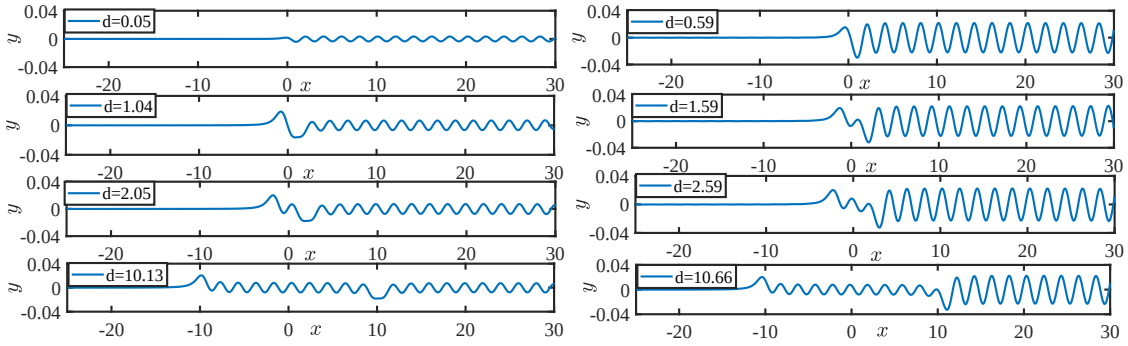



**Figure 2. Comparison of interfacial wave profiles when $d$ changes, where parameters $F = 0.13, \rho = 0.9, \epsilon_1 = -0.24, \epsilon_2 = 0.24$**

For counter-rotating vortex pair with $\epsilon_1 < 0$ and $\epsilon_2 > 0$, figure 2 represents how wave profiles change with $d$. As $d$ increasing the amplitude of upstream wave profiles increases and gradually stabilizes, as well as that of downstream steady wave profile oscillates. When $d$ is large enough there's steady wave profile between two vortices, which is close to that for the single vortex $\epsilon < 0$. For $d = 10.13$, the wave height and length of this steady wave profile are $h = 0.01587$, $L = 2.018$, close to that of the steady wave profile $h_- = 0.01580$, $L_- = 2.019$ for a single vortex $\epsilon = -0.24$ at $(0, -1)$. Figure 3 describes the periodical change of wave height $h$ of downstream wave profiles with $d$ clearly. When $d$ changes, the phase difference of two wave profiles caused

by two vortices changes periodically. If the phase difference is one/half a period, the amplitude of sum of these two profiles is the maximum of $h$. The period is close to $L_-/2$. That is to say that this period is almost the distance of two vortices(2d). As $d$ is large enough, the extreme values of downstream wave height $h$ are approximately $h_+ \pm h_-$, which are downstream steady wave height for a single vortex $\epsilon = -0.24$ and $\epsilon = 0.24$, respectively. Whereas for $d = 0.51$, $h/(h_+ + h_-) = 0.9775$ and $d = 1.01$,$h/|h_+ - h_-| = 0.8459$. These values of $d$ are the maximum and minimum of figure 3.

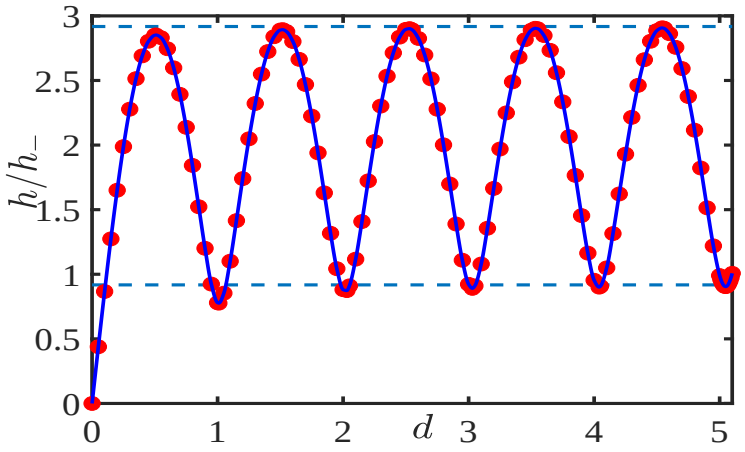



**Figure 3. Relationship of $h/h_-$ and $d$. Two horizontal lines of dashes represent $h_+ \pm h_-$. $h$ is the downstream wave height and $h_- = 0.01580$, $h_+ = 0.03038$ are wave height for a single vortex $\epsilon = -0.24$ and $\epsilon = 0.24$, respectively. Other parameters $F = 0.13$, $\rho = 0.9$, $\epsilon_1 = -0.24$, $\epsilon_2 = 0.24$.**

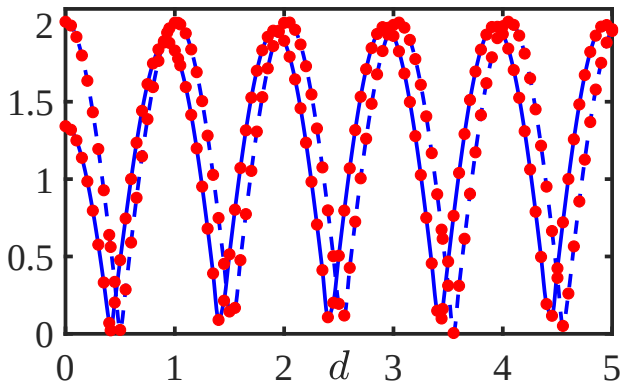



**Figure 4. Relationship of $h/h_-$(solid line) and $h_{12}/h_-$(dash line) with $d$, where $h$ is the downstream wave height and $h_- = 0.01580$ is the wave height for a single vortex $\epsilon = -0.23$. Other parameters $F = 0.13$, $\rho = 0.9$, $\epsilon_1 = \epsilon_2 = -0.23$.**

For co-rotating pair, figure 4 shows that the wave height $h$ also change with $d$ periodically and the extreme values are almost the sum and difference of wave height for the single vortex except $d = 0.413$ and 0.93. On the other hand, let $(x(s), y(s))$, $(x_1(s), y_1(s))$ and $(x_2(s), y_2(s))$ represent the interfaces for the vortex pair $\epsilon_1$ and $\epsilon_2$, the single vortex at $z_1$ and $z_2$, respectively. Figure 4 shows the variance of wave height of $y_1(s)) + y_1(s))$ (written as $h_{12}$) oscillates like $h$ with $d$ and moves backward except for about $d < 0.5$. To examine this behavior, try to construct an approximate analytic expression considering that for a single vortex $\epsilon < 0$ the wave profile consists of a large crest like a solitary wave and the downstream steady waves. Assume that this

crest can be written as $y = Asech^{\alpha}(x - x_0)$ and the downstream wave $y = Asin((2\pi)x/L + \phi)$, calculate curvature of highest point, the amplitude and wavelength and the positions of the maximum points to obtain the expression $y = Asech^{\alpha}(x-x_0)+0.0794sin(2\pi/2.025+0.562)H(x-x_1)$, where $x_0 = 0.275$, $A = 0.0205$, $\alpha = 2.957$ and $x_1 = (\pi - \phi)L/2\pi$. Figure 5 describes the

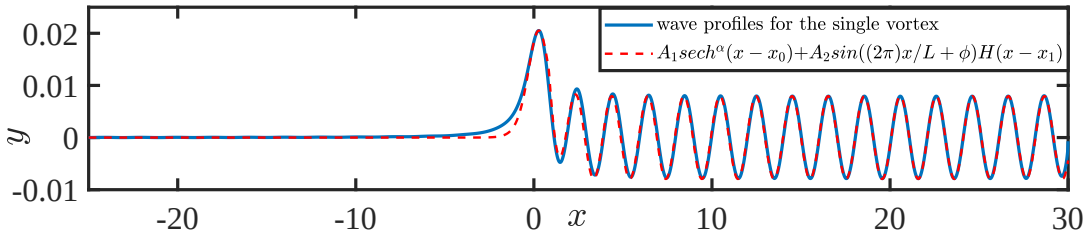
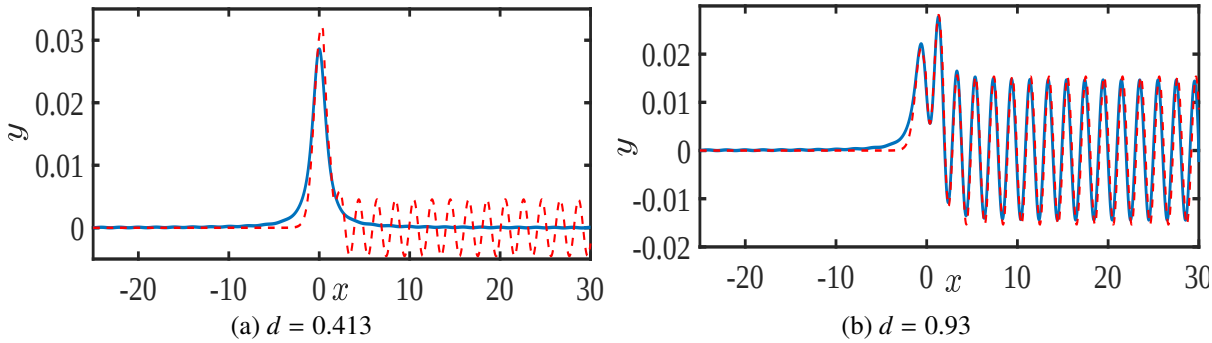



**Figure 5. Comparison of interfacial wave profiles calculating by boundary integral method(solid line) and the expression patched (dash line), where parameters $F = 0.13$, $\rho = 0.9$, $\epsilon = -0.23$.**

fitting effects. From figure 6, the wave profile is close to the superposition of two wave profiles




(a) $d = 0.413$ (b) $d = 0.93$

**Figure 6. Comparison of interfacial wave profiles calculating by boundary integral method(solid line) and the expression patched (dash line) for (a) $d = 0.413$, (b) $d = 0.93$**

.

for the single vortex as $d$ makes $h$ around its maximum (for instance $d = 0.93$) and significantly different as $h$ gets is minimum points(for example $d = 0.413$).

## Conclusion

In the steady two-dimensional two-layer flow with ideal irrotating fluids, a vortex pair submerged in lower layer generates interfacial waves . In this paper, two integral equations coupled with Bernoulli equations of nonlinear boundary waves are established by applying potential flow theory and boundary integral equation method, then a numerical method based on quasi-Newton method is carried out. The influences of $d$, which is the half of distance between two vortices on symmetric and asymmetric pair is discussed.

For symmetric/antisymmetric vortex pair, as $d$ increases the wave height of downstream wave $h$ oscillates and the extreme values are close to the sum/difference of wave heights for two vortices of the pair. When $d$ is taken near its maximum points, the wave profiles is close to the superposition of that for these two vortices, whereas if $d$ is taken other values the difference is large.

## References

[1] R. W Yeung and T. C Nguyen. Waves generated by a moving source in a two-layer ocean of finite depth. Journal of Engineering Mathematics, 35(1-2):85–107, 1999.

[2] Gang Wei, Jiachun Le, and Shiqiang Dai. Surface effects of internal wave generated by a moving source in a two-layer fluid of finite depth. Applied Mathematics and Mechanics (English Edition), 24(9):1025–1040, 2003.

[3] Gang Wei, Dongqiang Lu, and Shiqiang Dai. Waves induced by a submerged moving dipole in a two-layer fluid of finite depth. Acta Mechanica Sinica, 21(1):24–31, 2005.

[4] Zhen Wang, Li Zou, Hui Liang, and Zhi Zong. Nonlinear steady two-layer interfacial flow about a submerged point vortex. Journal of Engineering Mathematics, 103(1):1–15, 2016.

[5] Zhen Wang, Changhong Wu, and Li Zou. Study of nonlinear interfacial wave of stratified fluid due to a point vortex. Journal of Jiangsu University of Science and Technology (Natural Science Edition), 31(5):561–566, 2017.

[6] Zhen Wang, Changhong Wu, Li Zou, Qianxi Wang, and Qi Ding. Nonlinear internal wave at the interface of two-layer liquid due to a moving hydrofoil. Physics of Fluids, 29(7):65–69, 2017.

[7] Larry K. Forbes. On the effects of non-linearity in free-surface flow about a submerged point vortex. Journal of Engineering Mathematics, 19(2):139–155, 1985.

[8] Larry K. Forbes. A numerical method for non-linear flow about a submerged hydrofoil. Journal of Engineering Mathematics, 19(4):329–339, 1985.

[9] S. R Belward and Larry K. Forbes. Fully non-linear two-layer flow over arbitrary topography. Journal of Engineering Mathematics, 27(4):419–432, 1993.

[10] Thomas Leweke, Stéphane Le Dizès, and Charles H. K. Williamson. Dynamics and instabilities of vortex pairs. Fluid Dynamics Research, 46(1):507–541, 2016.

[11] Jason Rabinovitch, Vincent Brion, and Guillaume Blanquart. Effect of a splitter plate on the dynamics of a vortex pair. Physics of Fluids, 24(7):107, 2012.

[12] John G. Telste. Potential flow about two counter-rotating vortices approaching a free surface. Journal of Fluid Mechanics, 201(201):259–278, 2006.

[13] Hans J Lugt and Samuel Ohring. The oblique ascent of a viscous vortex pair toward a free surface. Journal of Fluid Mechanics, 236(236):461–476, 2006.

[14] C. G Broyden. A class of methods for solving nonlinear simultaneous equations. Mathematics of Computation, 19(92):577–593, 1965.

[15] John E. Dennis and J. More Jorge. Quasi-newton methods, motivation and theory. Siam Review, 19(1):46–89, 1977.

# A stochastic B-spline wavelet on the interval finite element method for elastic buckling of columns

***Shashank Vadlamani[1] and †Arun C.O.[1]**

[1]Department of Aerospace Engineering, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala-695547, India

*Presenting author: shashankvadlamani@gmail.com
†Corresponding author: shashankvadlamani@gmail.com

## Abstract

The current paper presents the solution of elastic buckling of columns using stochastic B-spline wavelet on the interval (BSWI) based wavelet finite element method (WFEM). In this work, the spatial variation of modulus of elasticity is modelled as a homogenous random field. BSWI scaling functions are used for the discretization of the random field. Columns under different boundary conditions are considered as numerical examples. The stochastic Eigen value problem is solved for the response statistics of buckling load with perturbation approach and the results are validated using Monte Carlo simulation (MCS). A parametric study is carried out by considering different coefficient of variation values by varying the standard deviation. A comparative study of computational time needed for the execution of perturbation approach and MCS is also done.

**Keywords:** B-spline wavelet on the interval; Multiresolution analysis; Random field; Auto-covariance function; Perturbation method; Monte Carlo simulation; Elastic buckling

## Introduction

Buckling is one of the predominant modes of failure which is observed when a structure is subjected to an axial compressive type external loading. It is a stability failure wherein, the entire structure collapses suddenly and the critical value of the applied external load causing this failure depends on the geometry of the structure and the stiffness of the material but not its strength [1]. Engineering structures have a high degree of uncertainty associated with its material properties, loads, geometry, operating environments, etc. [2]. The uncertainty in the design parameters will also result in uncertainty in buckling loads and its mode shapes. Therefore, a stochastic modelling approach leads to a robust design by providing additional statistical information on the stability of the structures. At the same time, a stochastic modelling also increases the complexity of the mathematical model and needs a higher computational effort to obtain the system response when compared with a deterministic approach. Nonetheless, widespread research has gone into the development of stochastic based numerical methods over the past few decades due to the availability of powerful computational resources.

Extensive research has gone into the development of stochastic finite element methods (SFEM) [3], wherein a stochastic mesh is generated to discretize the input random field and calculate the response statistics. Vanmarcke and Grigoriu [4] analysed simple beams with random elastic moduli using SFEM. Lin [5] developed a SFEM for the buckling analysis of frames with random initial imperfections, uncertain sectional and material properties. However, due to the high mesh dependency of finite element method (FEM), mapping the random field discretization onto response discretization becomes difficult. Hence, there is a

need for the development of stochastic based numerical methods, which can address the mesh dependency and re-meshing issues of FEM. Meshfree methods have been used in the stochastic analysis [6,7] to alleviate the mesh dependency of FEM. Gupta and Arun [8] proposed a stochastic meshfree method for elastic buckling of columns. In addition to meshfree methods wavelet finite element method (WFEM) is another alternate numerical tool which has shown to reduce the issues related to FEM considerably.

Wavelets are mathematical functions that are used in the approximation of other unknown functions at different levels of resolution. The multiresolution analysis (MRA) and two scale relation properties of wavelets lead to the development of a hierarchy of solutions during the approximation process. Wavelets have a scale varying local basis functions having a compact support that leads to a refinement of solution locally in the regions of high gradient. Therefore, issues related to slow convergence and re-meshing can be addressed using wavelet based numerical methods. B-spline wavelet on the interval (BSWI) has gained widespread popularity from among different wavelets that exist in the literature [9], due to its underlying properties [10,11] and hence, it is selected to be used in the current paper.

One-dimensional (1D) $C_0$ and $C_1$ BSWI elements for structural analysis using BSWI WFEM were constructed by Xiang et al. [12]. Deterministic buckling analysis of functionally graded beams and functionally graded plates was done by Zuo et al. in their papers [13] and [14] respectively. Yang et al. [15] carried out a deterministic study of free vibration and buckling analysis of plates.

Besides the discretization of random field, evaluation of response statics also needs to be computationally efficient. Monte Carlo simulation (MCS) has been used for the calculation of response statistics. Elishakoff [16] solved the problem of buckling of finite columns with initial imperfections, resting on a softening nonlinear elastic foundation by Monte Carlo method. But the usage of MCS makes the modelling process computationally expensive with increase in MCS sample size and number of random variables. Hence, a more viable procedure is needed that requires less computational effort. In this regard, perturbation methods have been extensively used for SFEM or stochastic meshless methods.

From the existing literature, it is noticed that a stochastic BSWI WFEM formulation for elastic buckling of columns using the perturbation method for calculating the response statistics, while material properties are modelled as random field does not exist. Hence, in the present study, the solution of elastic buckling of columns using stochastic BSWI WFEM is presented. The spatial variation of modulus of elasticity is modelled as a homogenous random field. BSWI scaling functions are used for the discretization of the random field and response. Columns under different boundary conditions are considered as numerical examples. The stochastic Eigen value problem is solved for the response statistics of buckling load with perturbation approach and the results are validated using MCS. A parametric study is carried out by considering different coefficient of variation (CV) values by varying the standard deviation. A comparative study of computational time needed for the execution of perturbation approach and MCS is also done.

In the next section, for the benefit of the reader, a brief description of BSWI and its properties is given.

**B-spline wavelet on the interval [0, 1]**

The theory of spline wavelets for whole square integrable real space $L^2(\mathbf{R})$ was developed by

Chui and Wang [17–19]. Wavelets defined on $L^2(\mathbf{R})$ cannot be directly used as interpolating functions as it results in numerical instability [20]. Hence, Chui and Quak [10] addressed this issue by constructing wavelet bases for the bounded interval [0, 1], which came to be known as BSWI. Spline wavelets are semi-orthogonal wherein, they retain inter-scale orthogonality and there is no necessity for the basis functions to be orthogonal to its translates within the same resolution level. By introducing multiple knots at the endpoints, splines can readily adapt to the case of the bounded interval [0, 1]. As a result, no truncation is needed when the function on $L^2(\mathbf{R})$ is restricted. By way of suitable adaptation at the endpoints, MRA of $L^2(\mathbf{R})$ can be implemented over to [0, 1]. Multiple knots exist at end points (0 and 1 in the case of BSWI) and they do not diminish the overall order of smoothness of the elements on [0, 1]. The continuity of B-splines depends on the selected order $m$ in such a way that B-splines with order $m$ are in $C_{m-2}$ continuity. The analytical expressions for the BSWI scaling functions $\phi$ and wavelet functions $\psi$ for given order $m$ and resolution $j=0$ can be found in the paper by Goswami et al. [21] and the expressions for order $m$ and any resolution $j$ were given by Xiang et al. [12] as,

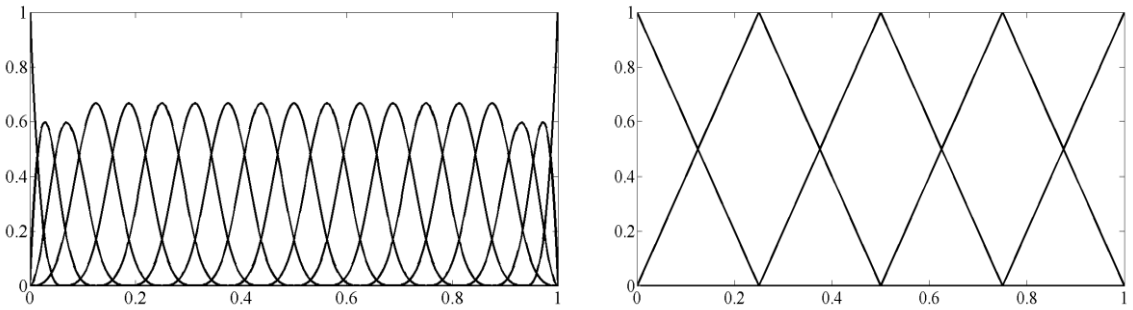
$$\phi_{m,k}^{j}(\xi) = \left\{ \begin{array}{l} \phi_{m,k}^{l}(2^{j-l}\xi), \quad k=-m+1,...,-1, \ \text{(0 boundary scaling functions)} \\ \phi_{m,2^j-m-k}^{l}(1-2^{j-l}\xi), \quad k=2^j-m+1,...,2^j-1, \\ \text{(1 boundary scaling functions)} \\ \phi_{m,0}^{l}(2^{j-l}\xi-2^{-l}k), \quad k=0,...,2^j-m, \ \text{(inner scaling functions)} \end{array} \right\} \tag{1}$$

$$\psi_{m,k}^{j}(\xi) = \left\{ \begin{array}{l} \psi_{m,k}^{l}(2^{j-l}\xi), \quad k=-m+1,...,-1, \ \text{(0 boundary wavelets)} \\ \psi_{m,2^j-2m-k+1}^{l}(1-2^{j-l}\xi), \quad k=2^j-2m+2,...,2^j-m \\ \text{(1 boundary wavelets)} \\ \psi_{m,0}^{l}(2^{j-l}\xi-2^{-l}k), \quad k=0,...,2^j-2m+1, \ \text{(inner wavelets)} \end{array} \right\} \tag{2}$$

The compactly supported intervals of wavelets are,

$$\text{supp}\ \psi_{m,k}^{j}(\xi) = \left\{ \begin{array}{l} [0,\ (2m-1+k)2^{-j}], \ \text{(0 boundary wavelets)} \\ [k2^{-j},\ 1], \ \text{(1 boundary wavelets)} \\ [k2^{-j},\ (2m-1+k)2^{-j}], \ \text{(inner wavelets)} \end{array} \right\} \tag{3}$$

BSWI scaling functions are categorized as the boundary scaling functions that exist at boundary points 0 and 1 on the domain and inner scaling functions that are dilations and translations of cardinal B-splines as shown in Eq. (1), (2) and (3). Eventually, the corresponding wavelets can be constructed from the scaling functions. BSWI scaling functions of different order and resolution which are used in the current study are shown in Figure 1.



a) $m = 4,\ j = 4$ b) $m = 2,\ j = 2$

Figure 1.BSWI scaling functions using different order and resolution

**Formulation of stochastic BSWI WFEM element for elastic buckling of columns**

In BSWI WFEM, the problem domain $\Omega$ is divided into sub-domains $\Omega_i$ ($i$=1, 2...) and each $\Omega_i$ is then mapped into the standard element solving domain $\Omega_e = \{\xi \mid \xi \in [0, 1]\}$, where instead of using the traditional polynomial interpolation, scaling or scaling and wavelet functions of BSWI can be used to form the shape functions over the elements $\Omega_e$. Here $\xi$ is the local co-ordinate used for solving 1D BSWI on [0, 1] along $y$ axis.

*Deterministic modelling*

The generalized functional of potential energy governing static buckling of columns is given as [22],

$$\prod = \frac{1}{2}\int_0^L EI\left(\frac{d^2 w_0}{dx^2}\right)^2 dx - \frac{P}{2}\int_0^L \left(\frac{dw_0}{dx}\right)^2 dx \tag{4}$$

Here, $w_0$ is the transverse deflection, $I$ is the second moment of area, $P$ is the axial compressive load, $E$ is the Young's modulus. One BSWI WFEM beam element based on Euler-Bernoulli theory (EBT) which was developed by Xiang et al. [12] is used in the present study. One BSWI EBT beam element with $C_1$ continuity is divided into $2^j + m - 3$ nodes with end nodes having both transverse and rotational degrees of freedom (DOF) and internal nodes having only transverse DOF as shown in Figure 2, where $m$, $j$ are the order and resolution of BSWI scaling functions respectively.

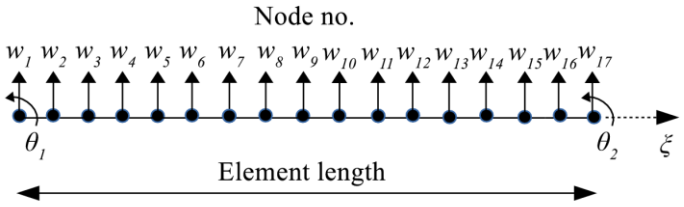



Figure 2.Distribution of nodes and degrees of freedom for one BSWI EBT based beam element with $C_1$ continuity using $m = 4,\ j = 4$

The unknown transverse deflection field function of Euler-Bernoulli beam element is approximated in the element solving domain $\xi$ in terms of wavelet scaling functions as,

$$w_0(\xi) = \sum_{k=-m+1}^{2^j-1} b^j{}_{m,k}\phi^j{}_{m,k}(\xi) = \boldsymbol{\varphi}\boldsymbol{b}^{\boldsymbol{e}} \tag{5}$$

where, $\boldsymbol{\varphi} = \{\phi^j{}_{m,-m+1}(\xi).....\phi^j{}_{m,2^j-1}(\xi)\}$ is the row vector of BSWI scaling functions and $\boldsymbol{b}^{\boldsymbol{e}} = \{b^j{}_{m,-m+1}\, b^j{}_{m,-m+2}....b^j{}_{m,2^j-1}\}^T$ is the column vector of wavelet coefficients that needs to be determined. The unknown transverse deflection field function can be expressed in terms of $C_1$ element type transformation matrix and physical DOF as,

$$w_0(\xi) = \boldsymbol{\varphi}(\boldsymbol{R}^{\boldsymbol{e}})^{-1}\boldsymbol{w}^{\boldsymbol{e}} = \boldsymbol{\varphi}\,\boldsymbol{T}^{\boldsymbol{e}}\boldsymbol{w}^{\boldsymbol{e}} = \boldsymbol{N}^{\boldsymbol{e}}\boldsymbol{w}^{\boldsymbol{e}} \tag{6}$$

where,

$$\left.\begin{aligned}
&\boldsymbol{w}^{\boldsymbol{e}} = \left\{ w_1\ \theta_1\ w_2\ w_3.....w_n\ w_{n+1}\ \theta_{n+1} \right\}^T, \\
&\theta_1 = \frac{1}{l_e}\left(\frac{dw_0(\xi_1)}{d\xi}\right), \theta_{n+1} = \frac{1}{l_e}\left(\frac{dw_0(\xi_{n+1})}{d\xi}\right), \\
&l_e = x_{n+1} - x_1,\ \boldsymbol{w}^{\boldsymbol{e}} = \boldsymbol{R}^{\boldsymbol{e}}\boldsymbol{b}^{\boldsymbol{e}},\ \boldsymbol{T}^{\boldsymbol{e}} = (\boldsymbol{R}^{\boldsymbol{e}})^{-1}, \\
&\boldsymbol{R}^{\boldsymbol{e}} = \left[ \boldsymbol{\varphi}^T(\xi_1)\ \ \frac{1}{l_e}\left(\frac{d\boldsymbol{\varphi}^T(\xi_1)}{d\xi}\right)\ \ \boldsymbol{\varphi}^T(\xi_2).....\boldsymbol{\varphi}^T(\xi_n)\ \ \boldsymbol{\varphi}^T(\xi_{n+1})\ \ \frac{1}{l_e}\left(\frac{d\boldsymbol{\varphi}^T(\xi_{n+1})}{d\xi}\right) \right]^T, \\
&\boldsymbol{N}^{\boldsymbol{e}} = \boldsymbol{\varphi}\,\boldsymbol{T}^{\boldsymbol{e}}
\end{aligned}\right\} \tag{7}$$

The elemental transformation matrix transforms the stiffness matrix from wavelet space into physical space. The transformation matrix also maintains the continuity and compatibility within the element and by using an assembly matrix, at the interface between the neighbouring elements. Upon substituting the deflection field of Eq. (6) into the weak form and invoking the stationary condition for variation of admissible deflections, the solution of static buckling of columns is obtained in the form of an Eigen value problem as,

$$\left(\boldsymbol{K}^{e} - \boldsymbol{P}\boldsymbol{G}^{e}\right)\boldsymbol{w}^{e} = 0 \tag{8}$$

where,

$$\left.\begin{aligned} \boldsymbol{K}^{e} &= \frac{EI}{\left(l_{ex}\right)^{3}}\int_{0}^{1}\left(\boldsymbol{T}^{e}\right)^{T}\left(\frac{d^{2}\boldsymbol{\varphi}}{d\xi^{2}}\right)^{T}\left(\frac{d^{2}\boldsymbol{\varphi}}{d\xi^{2}}\right)\left(\boldsymbol{T}^{e}\right)d\xi\,, \\ \boldsymbol{G}^{e} &= \frac{1}{l_{ex}}\int_{0}^{1}\left(\boldsymbol{T}^{e}\right)^{T}\left(\frac{d\boldsymbol{\varphi}}{d\xi}\right)^{T}\left(\frac{d\boldsymbol{\varphi}}{d\xi}\right)\left(\boldsymbol{T}^{e}\right)d\xi \end{aligned}\right\} \tag{9}$$

Here, $\boldsymbol{K}^{e}$ is the elemental stiffness matrix and $\boldsymbol{G}^{e}$ is the elemental geometric stiffness matrix. The Eigen values $\boldsymbol{P}$ from Eq. (8) correspond to the buckling loads and the Eigen vectors $\boldsymbol{w}^{e}$ correspond to the mode shapes.

*Stochastic modelling*

In the present work, the Young's modulus $E(\mathbf{x})$ is considered as a spatially varying homogeneous lognormal random field. As a result, the generalized functional of total potential as given in Eq. (4) along with response, will also become stochastic in nature. When $E(\mathbf{x})$ is a homogeneous lognormal field with mean $\mu_{E_l}$ and standard deviation $\sigma_{E_l}$ it can be expressed in terms of $\alpha(\mathbf{x})$ as,

$$E(\mathbf{x}) = C_l e^{\alpha(\mathbf{x})} \tag{10}$$

with

$$C_l = \frac{\mu_{E_l}^{2}}{\sqrt{\mu_{E_l}^{2} + \sigma_{E_l}^{2}}} \tag{11}$$

The auto-covariance kernel for $\alpha(\mathbf{x})$ can be written as [23],

$$\Gamma_{\alpha_l} = \ln\left(1 + \frac{\sigma_{E_l}^{2}}{\mu_{E_l}^{2}}\right)\exp\left[-\left(\Sigma_{i}^{n}\frac{|\Delta_i|}{c_i}\right)\right] \tag{12}$$

where, $n \in \mathbf{R}^{n}$, $\Delta_i$ is the distance between two points $x_a$, $x_b$ along $i$, $c_i$ is the correlation length parameter which determines the statistical correlation of field variable in the domain. Here, $\alpha(\mathbf{x})$ is a random field that does not possess an explicit expression and hence requires an approximation, which can be achieved by approximating a function over a set of random variables distributed in the domain obtained by discretization of the random field. In the current study, for modelling the random field a shape function method is proposed to be used.

Shape function method using Lagrange interpolation and moving least square shape functions has been employed in SFEM [24] and stochastic meshless methods [6] respectively. However, in the present study, BSWI scaling functions are used to model both the random field and response.

On similar lines, as the deflection field is approximated in Eq. (5), the unknown random field can be approximated in the element solving domain in terms of BSWI scaling functions as,

$$\alpha(\xi) = \sum_{k=-m_r+1}^{2^{j_r}-1} b^{j_r}{}_{m_r,k}\phi^{j_r}{}_{m_r,k}(\xi) = \boldsymbol{\varphi}_{\mathrm{R}}\boldsymbol{b}_{\mathrm{R}}^{e} \tag{13}$$

where, $\boldsymbol{\varphi}_{\mathrm{R}} = \{\phi^{j_r}{}_{m_r,-m_r+1}(\xi).....\phi^{j_r}{}_{m_r,2^{j_r}-1}(\xi)\}$ is the row vector of BSWI scaling functions, $\boldsymbol{b}_{\mathrm{R}}^{e} = \{b^{j_r}{}_{m_r,-m_r+1}\, b^{j_r}{}_{m_r,-m_r+2}....b^{j_r}{}_{m_r,2^{j_r}-1}\}^{T}$ is the column vector of wavelet coefficients that needs to be determined and $m_r$, $j_r$ are the order and resolution of BSWI scaling function chosen for the discretization of random field. The subscript $r$ is used here to denote the function or variable associated with the random field. Also, it can be noted that the order and resolution that is used for the discretization of the deflection field and random field can be different from each other. The unknown random field function is expressed in terms of $C_0$ element type transformation matrix as,

$$\alpha(\xi) = \boldsymbol{\varphi}_{\mathrm{R}}\left(\boldsymbol{R}_{\mathrm{R}}^{e}\right)^{-1}\boldsymbol{\alpha}_{\mathrm{R}}^{e} = \boldsymbol{\varphi}_{\mathrm{R}}\boldsymbol{T}_{\mathrm{R}}^{e}\boldsymbol{\alpha}_{\mathrm{R}}^{e} = \boldsymbol{N}_{\mathrm{R}}^{e}\boldsymbol{\alpha}_{\mathrm{R}}^{e} \tag{14}$$

where, $\boldsymbol{\alpha}_{\mathrm{R}}^{e} = \left\{\alpha_{1\mathrm{R}}\ \alpha_{2\mathrm{R}}.....\alpha_{(n+1)\mathrm{R}}\right\}^{T}$ is the set random variables distributed over the domain of the element. Thus, element stiffness coefficients and hence the element deflections will become functions of random variables $\boldsymbol{\alpha}_{\mathrm{R}}^{e}$ and Eq. (8) becomes a stochastic Eigen value problem as,

$$\left[\boldsymbol{K}^{e}\left(\boldsymbol{\alpha}_{\mathrm{R}}^{e}\right) - \boldsymbol{P}\left(\boldsymbol{\alpha}_{\mathrm{R}}^{e}\right)\boldsymbol{G}^{e}\right]\boldsymbol{w}^{e}\left(\boldsymbol{\alpha}_{\mathrm{R}}^{e}\right) = 0 \tag{15}$$

When $E(\mathbf{x})$ is modelled as a homogeneous lognormal field as given in Eq. (10), the $\boldsymbol{K}^{e}$ in Eq. (15) can be written as,

$$\boldsymbol{K}^{e} = \frac{\mu_{E_l} I}{\left(l_{ex}\right)^{3}\sqrt{1+\left(\dfrac{\sigma_{E_l}}{\mu_{E_l}}\right)^{2}}}\int_{0}^{1}\exp^{\left(\boldsymbol{\varphi}_{\mathrm{R}}\boldsymbol{T}_{\mathrm{R}}^{e}\boldsymbol{\alpha}_{\mathrm{R}}^{e}\right)}\left(\boldsymbol{T}^{e}\right)^{T}\left(\frac{d^{2}\boldsymbol{\varphi}}{d\xi^{2}}\right)^{T}\left(\frac{d^{2}\boldsymbol{\varphi}}{d\xi^{2}}\right)\left(\boldsymbol{T}^{e}\right)d\xi\Bigg\} \tag{16}$$

Here, $\boldsymbol{K}^{e}$ is the elemental stochastic stiffness matrix for beams based on EBT formulation. The element stiffness matrices $\boldsymbol{K}^{e}$ and $\boldsymbol{G}^{e}$ are obtained for all the sub-domains and assembled together to obtain the global stochastic Eigen value problem as,

$$\left[\boldsymbol{K}\left(\boldsymbol{\alpha}_{\mathrm{R}}\right)-\boldsymbol{P}\left(\boldsymbol{\alpha}_{\mathrm{R}}\right)\boldsymbol{G}\right]\boldsymbol{W}\left(\boldsymbol{\alpha}_{\mathrm{R}}\right)=0 \tag{17}$$

From Eq. (17), the second moment characteristics of buckling loads (Eigen values) $\boldsymbol{P}$ and mode shapes $\boldsymbol{w}$ (Eigen vectors) are obtained using the perturbation method which is discussed in the next section.

**Perturbation method**

Perturbation method uses the expansion of the global stiffness matrix $\boldsymbol{K}$, Eigen values $\boldsymbol{P}$ and Eigen vectors $\boldsymbol{w}$ via Taylor series [8]. It is based on the assumption that the variance of the random field should be small. Let $\Lambda=\{\alpha_i\}_{i=1}^{N}$ denote the vector of $N$ zero mean random variables representing the random field in the global domain $\Omega$. The Taylor series expansion of $\boldsymbol{K}$, $\boldsymbol{P}$ and $\boldsymbol{w}$ can be obtained as,

$$\boldsymbol{K}=\boldsymbol{K}_0+\sum_{i=1}^{N}\boldsymbol{K}_i^{I}\boldsymbol{\alpha}_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{K}_{ij}^{II}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j+....,\Bigg\} \tag{18}$$

$$\boldsymbol{P}=\boldsymbol{P}_0+\sum_{i=1}^{N}\boldsymbol{P}_i^{I}\boldsymbol{\alpha}_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{P}_{ij}^{II}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j+.....,\Bigg\} \tag{19}$$

$$\boldsymbol{W}=\boldsymbol{W}_0+\sum_{i=1}^{N}\boldsymbol{W}_i^{I}\boldsymbol{\alpha}_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{W}_{ij}^{II}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j+.....,\Bigg\} \tag{20}$$

where, $\boldsymbol{K}_0$, $\boldsymbol{P}_0$ , $\boldsymbol{W}_0$ are deterministic values evaluated at $\boldsymbol{K}(0),\boldsymbol{P}(0),\boldsymbol{W}(0)$; $(.)_i^{I}=\dfrac{\partial(.)}{\partial\boldsymbol{\alpha}_i}(0)$, and $(.)_{ij}^{II}=\dfrac{\partial^2(.)}{\partial\boldsymbol{\alpha}_i\partial\boldsymbol{\alpha}_j}(0)$. Upon substituting Eq. (18), (19) and Eq. (20) into Eq. (17) and rearranging the terms of the same order gives,

$$\left[\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right]\boldsymbol{W}_0=0 \tag{21}$$

$$\left[\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right]\boldsymbol{W}_i^{I}+\left[\boldsymbol{K}_i^{I}-\boldsymbol{P}_i^{I}\boldsymbol{G}\right]\boldsymbol{W}_0=0 \tag{22}$$

$$\left[\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right]\boldsymbol{W}_{ij}^{II}+\left[\boldsymbol{K}_i^{I}-\boldsymbol{P}_i^{I}\boldsymbol{G}\right]\boldsymbol{W}_j^{I}+\left[\boldsymbol{K}_j^{I}-\boldsymbol{P}_j^{I}\boldsymbol{G}\right]\boldsymbol{W}_i^{I}+\left[\boldsymbol{K}_{ij}^{II}-\boldsymbol{P}_{ij}^{II}\boldsymbol{G}\right]\boldsymbol{W}_0=0 \tag{23}$$

It is to be noted that $\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}$ is symmetric, which leads to,

$$\left[\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right]=\left[\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right]^{T} \tag{24}$$

Pre-multiplying Eq. (22) and (23) by $\boldsymbol{W}_0^{T}$ and using Eq. (24) leads to,

$$\left[\left(\boldsymbol{K}_0-\boldsymbol{P}_0\boldsymbol{G}\right)\boldsymbol{W}_0\right]^{T}\boldsymbol{W}_i^{I}+\boldsymbol{W}_0^{T}\left[\boldsymbol{K}_i^{I}-\boldsymbol{P}_i^{I}\boldsymbol{G}\right]\boldsymbol{W}_0=0 \tag{25}$$

$$\left[\left(\boldsymbol{K}_0 - \boldsymbol{P}_0\boldsymbol{G}\right)\boldsymbol{W}_0\right]^T \boldsymbol{W}_{ij}^{II} + \boldsymbol{W}_0{}^T\left[\boldsymbol{K}_i^I - \boldsymbol{P}_i^I\boldsymbol{G}\right]\boldsymbol{W}_j^I + \boldsymbol{W}_0{}^T\left[\boldsymbol{K}_j^I - \boldsymbol{P}_j^I\boldsymbol{G}\right]\boldsymbol{W}_i^I + \boldsymbol{W}_0{}^T\left[\boldsymbol{K}_{ij}^{II} - \boldsymbol{P}_{ij}^{II}\boldsymbol{G}\right]\boldsymbol{W}_0 = 0 \tag{26}$$

Upon simplification of Eq. (25) leads to,

$$\boldsymbol{P}_i^I = \left[\boldsymbol{W}_0{}^T\boldsymbol{G}\boldsymbol{W}_0\right]^{-1}\left[\boldsymbol{W}_0{}^T\boldsymbol{K}_i^I\boldsymbol{W}_0\right] \tag{27}$$

By substituting Eq. (27) into Eq. (25), $\boldsymbol{W}_i^I$ can be obtained, which can be further substituted into Eq. (26) to obtain,

$$\boldsymbol{P}_{ij}^{II} = \left[\boldsymbol{W}_0{}^T\boldsymbol{G}\boldsymbol{W}_0\right]^{-1}\left(\boldsymbol{W}_0{}^T\left[\boldsymbol{K}_i^I - \boldsymbol{P}_i^I\boldsymbol{G}\right]\boldsymbol{W}_j^I + \boldsymbol{W}_0{}^T\left[\boldsymbol{K}_j^I - \boldsymbol{P}_j^I\boldsymbol{G}\right]\boldsymbol{W}_i^I + \boldsymbol{W}_0{}^T\boldsymbol{K}_{ij}^{II}\boldsymbol{W}_0\right) \tag{28}$$

Upon the substitution of Eq. (28) into Eq. (26) $\boldsymbol{W}_{ij}^{II}$ can be obtained. By applying the expectation and variance operators on the first order or second order approximation of Eq. (19), the first and second order statistics of critical buckling load can be obtained as,

*First order approximation*

$$\left.\begin{aligned} \mu_{P_{cr}} &= \boldsymbol{P}_0 \quad , \\ \gamma_{P_{cr}} &= \sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{P}_i^I\left(\boldsymbol{P}_j^I\right)^T\Gamma_{\alpha}(\alpha_i,\alpha_j) \end{aligned}\right\} \tag{29}$$

*Second order approximation*

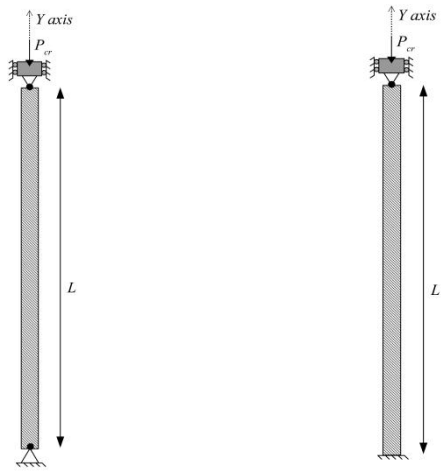
$$\left.\begin{aligned} \mu_{P_{cr}} &= \boldsymbol{P}_0 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{P}_{ij}^{II}\,\Gamma_{\alpha}(\alpha_i,\alpha_j) \ , \\ \gamma_{P_{cr}} &= \sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{P}_i^I\left(\boldsymbol{P}_j^I\right)^T\Gamma_{\alpha}(\alpha_i,\alpha_j) + \\ &\frac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\boldsymbol{P}_{ij}^{II}\left(\boldsymbol{P}_{kl}^{II}\right)^T\left\{\Gamma_{\alpha}(\alpha_i,\alpha_l)\Gamma_{\alpha}(\alpha_j,\alpha_k)+\Gamma_{\alpha}(\alpha_i,\alpha_k)\Gamma_{\alpha}(\alpha_j,\alpha_l)\right\} \end{aligned}\right\} \tag{30}$$

Similarly statistics of other response functions of interest, like $\boldsymbol{W}$ can also be found out. In the next section, a few 1D numerical examples are solved based on the preceding formulations and the results are analysed.

**Numerical examples**

Two numerical examples are solved with the proposed stochastic BSWI WFEM formulation for elastic buckling of columns. Columns with pinned-pinned (p-p) and fixed-pinned (f-p) boundary conditions under axial compressive loading as shown in Figure 3 are considered for the study. The response statistics for buckling loads and mode shapes are calculated via perturbation approach and the results are compared with the statistics obtained from MCS. From a convergence study, based on the calculation of relative percentage error

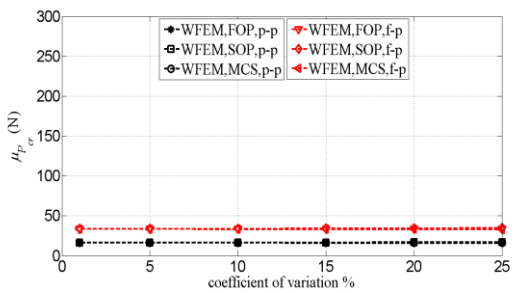
in $L_2$ norm of mean and standard deviation values of Young's modulus for various MCS sample size; it is noted that an error of less than 1% is obtained when the MCS sample size is 5000. Hence, MCS sample size of 5000 is considered in the current study. The mean value of Young's modulus is taken as $\mu_{E_l} = 2\times10^5$ MPa with $L$ = 100 mm, $b = 1$mm and $h = 1$mm. The entire domain of the column is modelled using one BSWI $C_1$ type continuity element. The deflection field is approximated using cubic ( $m = 4$ ) BSWI scaling functions with a resolution of $j = 4$ and the random field is approximated with linear ( $m = 2$ ) BSWI scaling functions with a coarse resolution of $j = 2$.

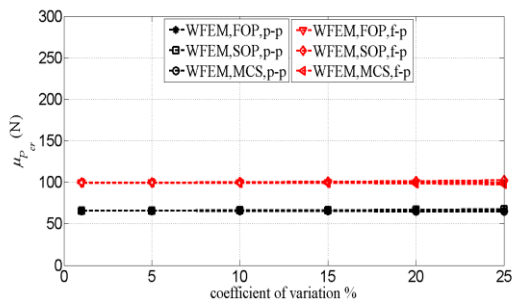



a) Pinned-Pinned b) Fixed-Pinned

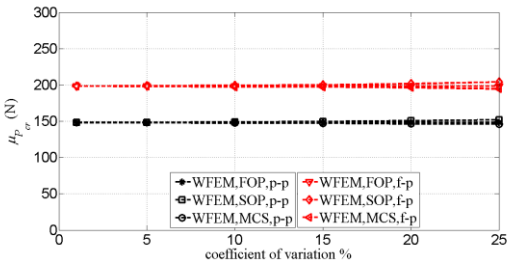
Figure 3.Columns with various boundary conditions under axial compressive loading

The mean values of the buckling loads (first, second and third) for a pinned-pinned (p-p) column obtained by using the perturbation approach are shown in Figure 4. These values are compared with the values obtained from MCS and the results are plotted for different values of CV, obtained by varying the standard deviation of Young's modulus $E(\mathbf{x})$. The correlation length parameter considered is 50. It can be observed from Figure 4 that at a CV of 20% the results obtained from perturbation approach are in good agreement with MCS for all the buckling loads. However, at a CV of 25%, a deviation of 3% is observed between the perturbation and MCS results in the case of third buckling load. The variation of standard deviation values of buckling loads against CV are shown in Figure 5. It can be observed that even at a CV of 25%, the values obtained from perturbation approach concur well with the MCS values for all the buckling loads.



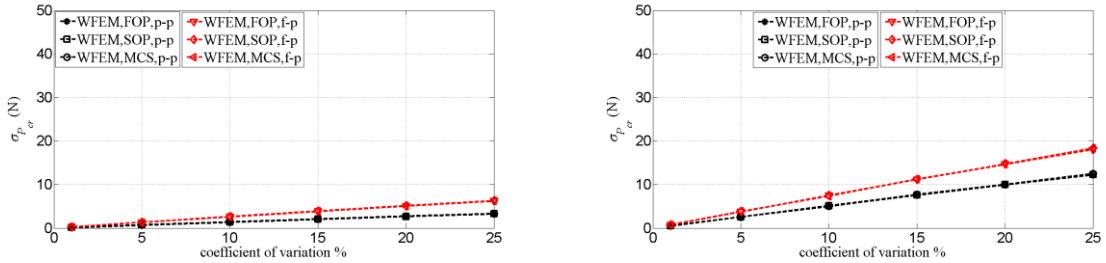



a) First buckling load b) Second buckling load

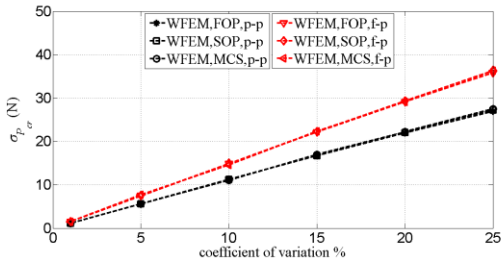



c) Third buckling load

Figure 4.Variation of mean values of buckling loads for columns with different boundary conditions against CV




a) First buckling load b) Second buckling load

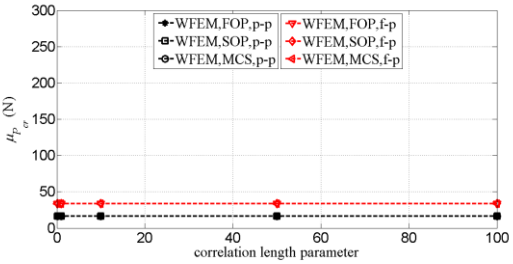



c) Third buckling load

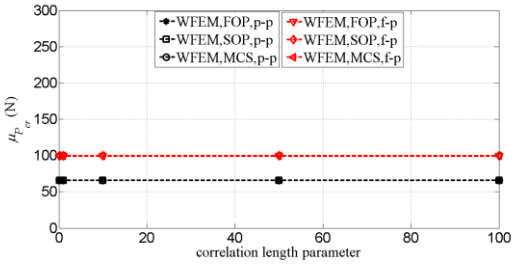
Figure 5.Variation of standard deviation values of buckling loads for columns with different boundary conditions against CV

The variation of mean and standard deviation values of buckling loads obtained by using the perturbation approach against varying correlation length parameter is shown in Figure 6 and Figure 7 respectively. A value of 5% is considered for CV during the process. Figure 6 and 7 shows that even at a small correlation length parameter the values obtained from WFEM based perturbation approach remain in good agreement with the MCS values. This shows that a coarse discretization of random field using BSWI WFEM is able to accurately capture the results even at extreme correlation length parameters unlike SFEM, wherein the dependency of correlation length parameter on random field mesh is well documented [3,24] and would require a higher number of random variables to be used for accurate results.

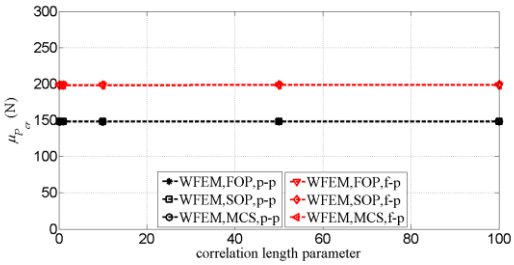
Besides the buckling loads, the first three mode shapes are also plotted for the pinned-pinned column in Figure 8. It can be seen that WFEM based perturbation approach accurately captures the first three mode shapes when compared with MCS results.




a) First buckling load

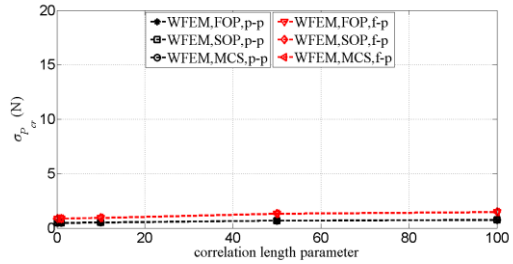



b) Second buckling load

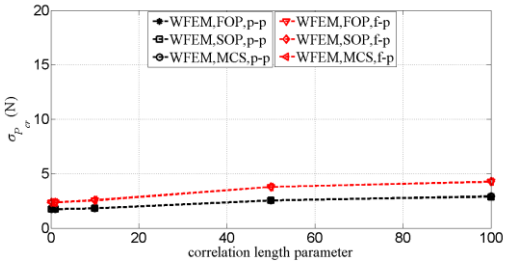



c) Third buckling load

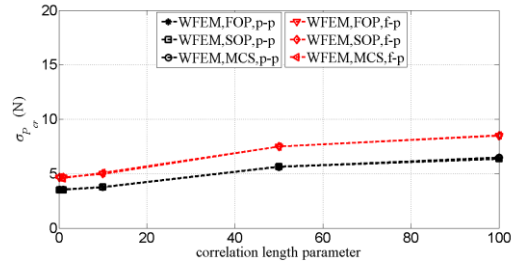
Figure 6.Variation of mean values of buckling loads for columns with different boundary conditions against correlation length parameter




a) First buckling load

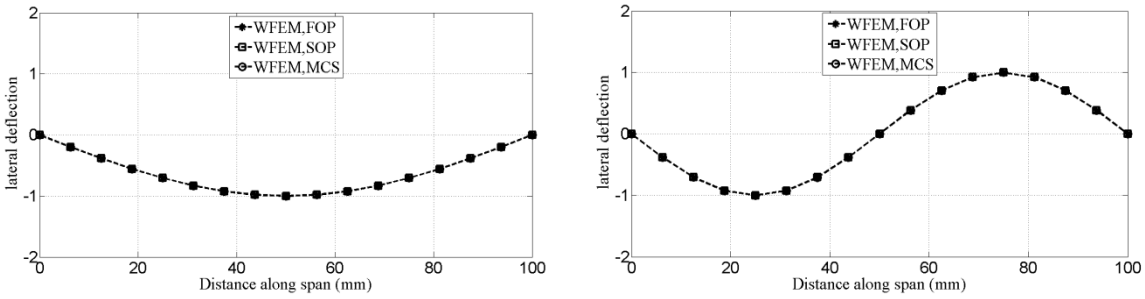



b) Second buckling load

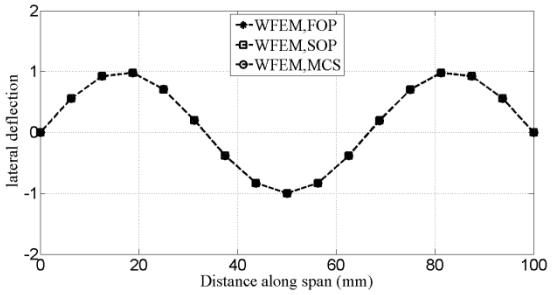



c) Third buckling load

Figure 7.Variation of standard deviation values of buckling loads for columns with different boundary conditions against correlation length parameter



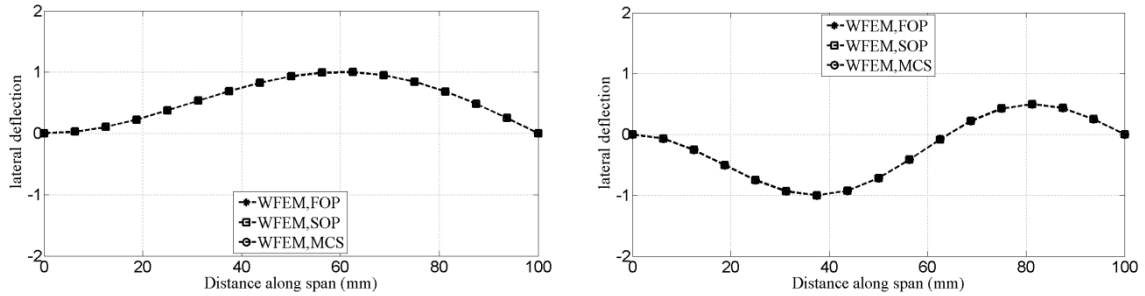

a) First mode b) Second mode



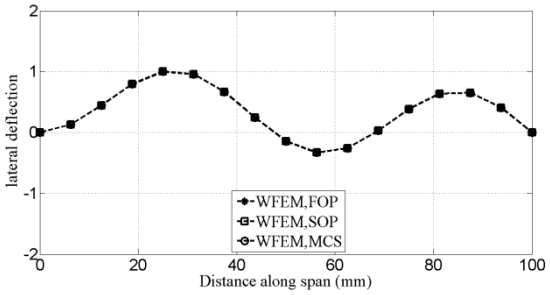

c) Third mode

Figure 8. Mode shapes for a pinned-pinned column

The mean values of the buckling loads for a fixed-pinned (f-p) column obtained by using the perturbation approach are shown in Figure 4. It can be observed from Figure 4 that at a CV of 25% there is a deviation of around 5% between the results obtained from perturbation approach and MCS for the third buckling load. However, no such deviation is observed in the standard deviation values obtained from perturbation approach and MCS as seen in Figure 5. Similar to the case of pinned-pinned columns, the mean and standard deviation values of buckling loads against varying correlation length parameters for a fixed-pinned column show a good agreement between the perturbation and MCS results as observed in Figure 6 and Figure 7. Furthermore, the first three mode shapes as shown in Figure 9 reinstates the accuracy of the WFEM based perturbation approach.



a) First mode b) Second mode




c) Third mode

Figure 9. Mode shapes for a fixed-pinned column

*Computational time*

Besides evaluating the mean and standard deviation values of the buckling loads, the normalized computational times required by the perturbation approach (FOP and SOP combined) and MCS (5000 simulations) is also calculated. It is noted that in the case of a pinned-pinned column, the execution time of MCS is 39.63 times more in comparison with WFEM based perturbation approach. Similarly, for a fixed-pinned column, the execution time of MCS is 38.28 times more than the perturbation method. Hence, the proposed BSWI WFEM based perturbation approach is not only accurate but also computationally more efficient in comparison with the MCS based approach.

**Conclusion**

The current paper proposes the formulation of stochastic BSWI WFEM formulation for elastic buckling of columns wherein, the spatial variation of modulus of elasticity is modelled as a homogeneous random field. In the present work, BSWI scaling functions are used for the approximation of deflection field as well as random field. The response statistics are calculated using the perturbation approach and validated by comparing with the results of MCS. The results obtained from the numerical examples show that WFEM based perturbation approach can be used to accurately capture the response statistics of the buckling load for values of CV up to 25%.

The domain of the column is discretized using only one BSWI WFEM element, due to which, there are no meshes and the programming effort needed in the pre-processing stage to form a global matrix from the assembly of multiple elements is reduced. The parametric study on correlation length parameters show that the values obtained from perturbation approach based on WFEM concur well with MCS values at extremely small or large correlation length parameters even when the random field is modelled using a coarse nodal discretization. Further, the normalized computational times are calculated for both the numerical examples and WFEM based perturbation approach takes less time in comparison with MCS in both the cases, thereby making it more efficient.

**References**


[1] Bauchau, O. A., and Craig, J. I., *Structural Analysis With Applications to Aerospace Structures*, Springer.

[2] Choi, S.-K., Grandhi, R. V, and Canfield, R. A., 2007, *Reliability-Based Structural Design*, Springer London.

[3] Stefanou, G., 2009, "The Stochastic Finite Element Method: Past, Present and Future," Comput. Methods Appl. Mech. Eng., **198**(9–12), pp. 1031–1051.

[4] Vanmarcke, E., and Grigoriu, M., 1983, "Stochastic Finite Element Analysis of Simple Beams," J. Eng. Mech., **109**(5), pp. 1203–1214.

[5] Lin, S. C., and Kam, T. Y., 1992, "Buckling Analysis of Imperfect Frames Using a Stochastic Finite Element Method," Comput. Struct., **42**(6), pp. 895–901.

[6] Rahman, S., and Rao, B. N., 2001, "A Perturbation Method for Stochastic Meshless Analysis in Elastostatics," Int. J. Numer. Methods Eng., **50**(8), pp. 1969–1991.

[7] Rahman, S., and Xu, H., 2005, "A Meshless Method for Computational Stochastic Mechanics," Int. J. Comput. Methods Eng. Sci. Mech., **6**(1), pp. 41–58.

[8] Gupta, A., and Arun, C. O., 2018, "Stochastic Meshfree Method for Elastic Buckling Analysis of Columns," Comput. Struct., **194**, pp. 32–47.

[9] Li, B., and Chen, X., 2014, "Wavelet-Based Numerical Analysis: A Review and Classification," Finite Elem. Anal. Des., **81**, pp. 14–31.

[10] Chui, C. K., and Quak, E., 1992, "Wavelets on a Bounded Interval," *Numerical Methods in Approximation Theory, Vol. 9*, Birkhäuser Basel, Basel, pp. 53–75.

[11] Chui, C. K., 1992, *An Introduction to Wavelets - Wavelet Analysis and Its Applications*, Academic press.

[12] Xiang, J. W., Chen, X. F., He, Z. J., and Dong, H. B., 2007, "The Construction of 1D Wavelet Finite Elements for Structural Analysis," Comput. Mech., **40**(2), pp. 325–339.

[13] Zuo, H., Yang, Z. B., Chen, X. F., Xie, Y., Zhang, X. W., and Liu, Y., 2014, "Static, Free Vibration and Buckling Analysis of Functionally Graded Beam via B-Spline Wavelet on the Interval and Timoshenko Beam Theory," C. - Comput. Model. Eng. Sci., **100**(6), pp. 477–506.

[14] Zuo, H., Yang, Z., Chen, X., Xie, Y., and Zhang, X., 2014, "Bending, Free Vibration and Buckling Analysis of Functionally Graded Plates via Wavelet Finite Element Method," C. Comput. Mater. Contin., **44**(3), pp. 167–204.

[15] Yang, Z., Chen, X., Zhang, X., and He, Z., 2013, *Free Vibration and Buckling Analysis of Plates Using B-Spline Wavelet on the Interval Mindlin Element*, Appl. Math. Model.

[16] Elishakoff, I., 1979, "Buckling of a Stochastically Imperfect Finite Column on a Nonlinear Elastic Foundation: A Reliability Study," J. Appl. Mech., **46**(2), p. 411.

[17] Chui, C. K., and Wang, J., 1991, "A Cardinal Spline Approach to Wavelets," Proc. Am. Math. Soc., **113**(3), p. 785.

[18] Chui, C. K., and Wang, J.-Z., 1992, "On Compactly Supported Spline Wavelets and a Duality Principle," Trans. Am. Math. Soc., **330**(2), pp. 903–915.

[19] Chui, C. K., and Wang, J. Z., 1993, "An Analysis of Cardinal Spline-Wavelets," J. Approx. Theory, **72**(1), pp. 54–68.

[20] Bertoluzza, S., Naldi, G., and Ravel, J. C., 1994, "Wavelet Methods for the Numerical Solution of Boundary Value Problems on the Interval," *Wavelets: Theory, Algorithms, and Applications*, C.K. Chui, and Others, eds., pp. 425–448.

[21] Goswami, J. C., Chan, A. K., and Chui, C. K., 1995, *On Solving First-Kind Integral Equations Using Wavelets on a Bounded Interval*, IEEE Trans.

[22] Bathe, K. J., 2010, *Finite Element Procedures*, PHI learning private limited, {New} {Delhi}.

[23] Xu, H., and Rahman, S., 2005, "Decomposition Methods for Structural Reliability Analysis," Probabilistic Eng. Mech., **20**(3), pp. 239–250.

[24] Sudret, B., and Kiureghian, A. Der, 2000, *Stochastic Finite Element Methods and Reliability: A State-of-the-Art Report*.

# A novel substructural damage detection approach for shear structures based on the combination of ARMAX model residual and Kullback-Leibler divergence

***Liu Mei[1,2], Huaguan Li[1,2], Yunlai Zhou[3], and †Wujian Long[1,2]**

[1]Guangdong Provincial Key Laboratory of Durability for Marine Civil Engineering, Shenzhen University, Shenzhen, 518060, P.R. China

[2]College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, 518060, P.R. China

[3]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

*Presenting author: meiliu@szu.edu.cn

†Corresponding author: longwj@szu.edu.cn

**Abstract**

In this paper, a novel substructural damage detection method combining autoregressive moving average with exogenous inputs (ARMAX) model residual and Kullback-Leibler divergence (KLD) is proposed to identify the damages of shear building structures. Firstly, based on the partition strategy of multi-input multi-output (MIMO) model, the overall structure is divided into series of substructures such that structural damage diagnosis process is able to be implemented on each substructure independently. For the sake of better resisting noise interference and enhancing damage detection robustness, every substructure is modeled by autoregressive-moving average with exogenous inputs (ARMAX) model, and its model residuals contain sensitive structural damage feature characterized by the changes in chi-square distribution function (CSDF) of the model residuals. Furthermore, KLD is utilized to measure the similarity between two probability distributions and used as structural damage indicator to quantify the damage in detail. Numerical simulation is conducted to evaluate the performance of proposed damage identification approach, and it shows the satisfactory results of structural damage localization and quantification.

**Keywords:** Damage detection; shear structure; substructure; ARMAX model residual; chi-square distribution function (CSDF); Kullback-Leibler divergence (KLD)

## 1. Introduction

In recent years, due to the aging of aerospace, civil, and mechanical infrastructures especially for the structures serving human society activities for a long time, structural health monitoring (SHM) has become an essential research field in maintaining the integrity of structures. Structural damage detection generally provides the fundamental information for SHM

practices and is one of the most challenging components in the construction of SHM system [1].

As the nondestructive evaluation (NDE) techniques, vibration based damage detection methods have become more effective and flexible than the traditional detection approaches in engineering applications in the latest decades. Modal properties are easily obtained from structural responses, and modal frequencies, mode shapes or mode shape curvatures are chosen as damage sensitive features for wide applications of damage detection. In addition, substructuring technique has been also developed for subtly designing and analyzing the complex large-scale structures in an efficient way that the whole structure is decomposed into a series of smaller substructures. By adopting a strategy of 'divide-and-conquer', the performance of dynamic system model fitting and the accuracy of structural parameters identification are not reduced especially for actual large-scale engineering structures, such as high-rise building structures and long span bridges. For damage detection, most damage indicators of previous substructuring method are based on modal parameters. A substructuring method combined with the difference between squared original frequency and squared damaged frequency was proposed in [2]-[3] for damage identification of shear structures. However, modal parameters usually represent the property of the whole structure such that the modal parameters based damage indicators indicate the global structural damage, which is not sensitive enough for local damage identification under complicated environmental conditions [4]. In contrast to modal parameters identification based detection methods, the time domain or frequency domain methods extract local damage sensitive features via signal processing and only concentrating on measured data of structural responses, meanwhile, these methods are in a data-driven way without a model of the structure and different from the model based method requiring an accurate finite element model. For linear time-invariant (LTI) systems, autoregressive process can well model the structural systems and distinguish various system dynamics through the 'black box' model structure containing system inputs and outputs; besides, the autoregressive model based methods are more intuitionistic than the frequency domain method by generating underlying observations process directly from the model parameters without spectral representation. The autoregressive model parameters, such as the model residuals and model coefficients, have been well utilized as the damage sensitive indexes for local damage detection. In the previous researches, most model residuals-based damage detection methods are based on pattern recognition approaches [5], and their damage indicators usually rely on pattern recognition tools, which often require a large amount of training data to extract damage sensitive features and inevitably produces huge computational complexity.

In order to improve the computational flexibility of existing methods and potentially locate and quantify damages for shear structures, this paper proposes a new substructural damage detection method based on autoregressive moving average with exogenous inputs (ARMAX) model and Kullback-Leibler divergence (KLD). At first, the substructural division strategy in [2]-[3] is employed so that the damage detection process can be carried out on each

substructure independently, which is suitable for a parallel and distributed SHM system. The ARMAX model combined with substructure of MIMO system is built to remove strong correlation of the responses and needed not to use pre-whitening filter, which is more convenient in signal processing than the autoregressive (AR) method of single output [6]-[7]. Besides, an ARMAX model is able to enhance the noise immunity of damage detection results by its moving average term of model residual. Furthermore, this study proposes an innovative damage indicator by incorporating ARMAX model residual and Kullback-Leibler divergence (KLD) for sensitive damage quantification in a data-driven way. KLD is an index widely used for measuring the similarity between two probability distributions in statistics; the value of KLD is close to zero when the two probability distributions are similar, otherwise, the value of KLD is close to one [8]. In this study, the distribution of model residual can be well described by the chi-square distribution function (CSDF), the KLD value between CSDFs in damage state and undamaged state can clearly indicate the damage, including the location and extent of substructural damage. Simulation of six degrees of freedom (DOFs) shear building structure subjected to mutually correlative white noises is conducted to verify the performance of proposed damage substructural damage detection approach, and specific conclusions are finally discussed.

## 2. Theoretical fundamentals

### *2.1. Dynamic system modeling with ARMAX model*

For the linear discrete-time system with multiple-input and multiple-output, the dynamic process can be described by an ARMAX model as following:

$$y(t) + \sum_{k=1}^{n_a} a_k y(t-k) = \sum_{k=1}^{n_b} b_k u(t-n_k-k+1) + \sum_{k=1}^{n_c} c_k e(t-k) \quad (1)$$

where $y(t)$ represents the system output at time $t$, $u(t)$ denotes the system input; $a_k$, $b_k$, and $c_k$ indicate the coefficients of autoregressive term, system input term, and moving-average term, respectively, $n_a$, $n_b$ and $n_c$ depict their corresponding model orders, respectively, $n_k$ means the time delay steps; $e(t)$ are the residuals of the estimation process at time $t$; the ARMAX model is efficient for its flexibility to availably handle the disturbance modeling through its moving-average coefficient $c_k$.

### *2.2. Substructure division*

Generally, the shear building structure can be simulated as a one-dimensional shear model with lumped masses through the below motion equation:

$$\boldsymbol{M\ddot{x} + C\dot{x} + K\dot{x} = -Mr\ddot{x}_g} \quad (2)$$

where $\boldsymbol{M}_{n\times n}$, $\boldsymbol{K}_{n\times n}$, and $\boldsymbol{C}_{n\times n}$ respectively depicted the mass, stiffness, and damping matrixes, *n* depicts the number of DOFs, $\boldsymbol{r}$ denotes the $n \times 1$ unit vector ($\boldsymbol{r} = [1 \cdots 1]^T$), $\boldsymbol{x}$ indicates the displacement vector of lateral vibration relative to the ground, $\ddot{x}_g$ means the ground acceleration.

The motion of each DOF is affected by the motion of adjacent DOFs; every mass and its adjacent masses are separated from the overall structure to construct series of substructures, as shown in Fig. 1. According to the principle of force balance at the lateral direction, the motion equation of substructure *i* ($1 \le i \le n$-1) can be expressed as

$$m_i\ddot{y}_i + (c_i + c_{i+1})\dot{y}_i + (k_i + k_{i+1})y_i = -m_i\ddot{z}_{i-1} + c_{i+1}\dot{y}_{i+1} + k_{i+1}y_{i+1} \tag{3}$$

where $m_i$ is the $i^{th}$ story mass, $k_i$ is the stiffness coefficient of the $i^{th}$ story, $c_i$ is the damping coefficient of the $i^{th}$ story; $y_i$ represents the displacement of the $i^{th}$ story relative to the $(i\text{-}1)^{th}$ story; $\ddot{z}_i$ means the absolute acceleration of the $i^{th}$ story, and especially $\ddot{z}_0$ denotes the ground acceleration $\ddot{x}_g$. Considering that the top mass $m_n$ is the free end and only one mass is adjacent to it, the motion equation of the top substructure is represented by following:

$$m_n\ddot{y}_n + c_n\dot{y}_n + k_n y_n = -m_n\ddot{z}_{n-1} \tag{4}$$

Introducing the difference expression

$$\dot{y}_i(t) = \frac{y_i(t+T) - y_i(t-T)}{2T} \tag{5}$$

$$\ddot{y}_i(t) = \frac{y_i(t+T) - 2y_i(t) + y_i(t-T)}{T^2} \tag{6}$$

where $\dot{y}_i$ and $\ddot{y}_i$ means the velocity and acceleration of the $i^{th}$ story relative to the $(i\text{-}1)^{th}$ story, respectively, $t$ represents the time index, $T$ depicts the sampling interval. By substituting Eq. (5) and Eq. (6) into Eq. (3), the motion equation of substructure *i* ($1 \le i \le n$-1) can be rewritten as

$$\begin{aligned}\ddot{y}_i(t) + a_1\ddot{y}_i(t-1) + a_2\ddot{y}_i(t-2) = {} & b_{11}\ddot{z}_{i-1}(t-1) + b_{12}\ddot{z}_{i-1}(t-2) \\ & + b_{21}\ddot{y}_{i+1}(t-1) + b_{22}\ddot{y}_{i+1}(t-2) \\ & + c_1 e(t-1) + c_2 e(t-2)\end{aligned} \tag{7}$$

In this, Eq. (7) can be regarded as an ARMAX model with two-input ($\ddot{z}_{i-1}$ and $\ddot{y}_{i+1}$) and single-output ($\ddot{y}_i$) [2]-[3], where *e*(*t*) represents the ARMAX model residuals. Likewise, the motion equation of top substructure *n* can be rewritten by substituting Eq. (5) and Eq. (6) into Eq. (4), and it can be identified as a single-input ($\ddot{z}_{n-1}$) and single-output ($\ddot{y}_n$) ARMAX model, that is

$$\ddot{y}_n(t) + a_1\ddot{y}_n(t-1) + a_2\ddot{y}_n(t-2) = b_1\ddot{z}_{n-1}(t-1) + b_2\ddot{z}_{n-1}(t-2) + c_1 e(t-1) + c_2 e(t-2) \quad (8)$$

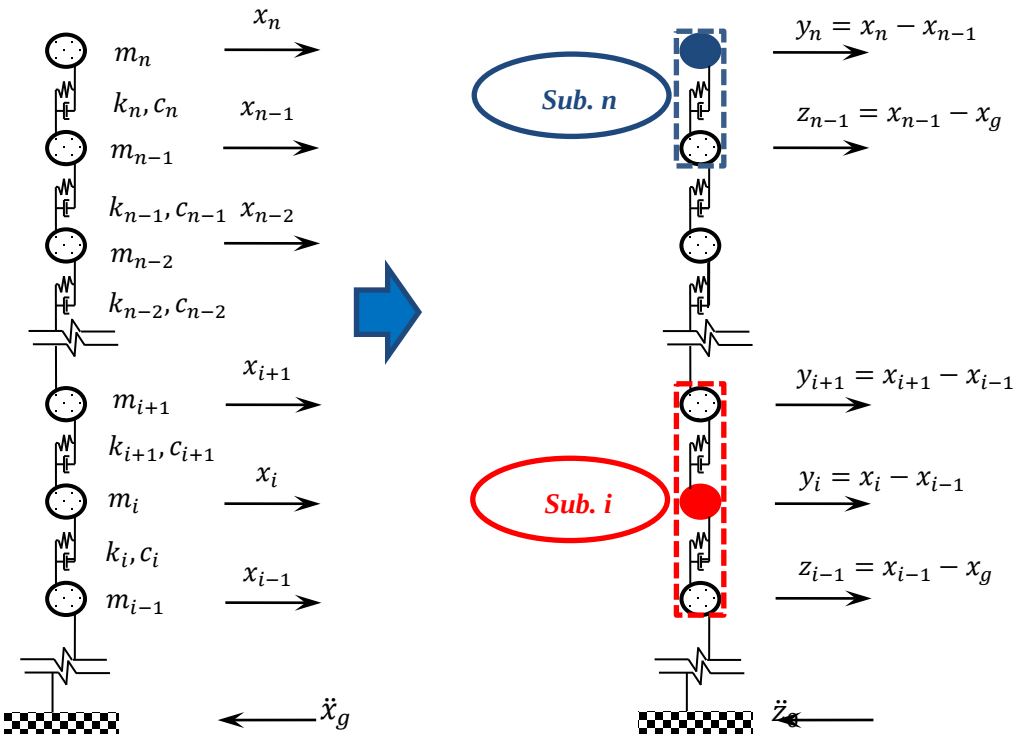



**Figure 1. Substructure division method**

Herein, in order to characterize each substructure with MIMO model and promptly acquired essential structural features, only three accelerometers are needed to establish the related ARMAX model while two is enough for the top substructure.

*2.3. Damage indicator*

In this paper, a novel structural damage indicator based on ARMAX model residual and Kullback-Leibler divergence is proposed to identify the damages of shear structures. Initially, the predicted system output $\hat{y}_u(t)$ modeling with ARMAX model in undamaged state can be calculated as

$$\hat{y}_u(t) = -\sum_{k=1}^{n_a} a_k y(t-k) + \sum_{k=1}^{n_b} b_k u(t-n_k-k+1) + \sum_{k=1}^{n_c} c_k\, e(t-k) \quad (9)$$

The model residuals of the undamaged state and damaged state can be generated between the measurement system output $y$ and the predicted system output $\hat{y}$ by comparing with the reference ARMAX model of undamaged state, that is

$$e_u(t) = y_u(t) - \hat{y}_u(t) \tag{10}$$

$$e_d(t) = y_d(t) - \hat{y}_u(t) \tag{11}$$

where $y_u(t)$ and $y_d(t)$ are the measurement output of undamaged and damaged state from the substructure being analyzed, respectively. For damage case caused by the degradation of story stiffness, structural responses generated from damaged system generally vary from responses of undamaged system, and it is hard to fit the structural responses in damaged state well by using the reference ARMAX model in undamaged state. In other words, model residuals from damaged system responses (Eq. (11)) are different from residuals of responses in undamaged state (Eq. (10)), which contains important structural information for damage examination. In addition, the model residual vector are normalized to a dimensionless vector so as to remove the effects of various response amplitudes, as shown as follows

$$\bar{\boldsymbol{e}}_{\boldsymbol{u}} = \frac{\boldsymbol{e_u}}{\|\boldsymbol{y_u}\|} \tag{12}$$

$$\bar{\boldsymbol{e}}_{\boldsymbol{d}} = \frac{\boldsymbol{e_d}}{\|\boldsymbol{y_d}\|} \tag{13}$$

where $\boldsymbol{e_u}$ and $\boldsymbol{e_d}$ represent the ARMAX model residual vector in the undamaged and damaged state, respectively; $\|\boldsymbol{y_u}\|$ and $\|\boldsymbol{y_d}\|$ mean the norm of output response vector in the undamaged and damaged state, respectively, $\bar{\boldsymbol{e}}_{\boldsymbol{u}}$ and $\bar{\boldsymbol{e}}_{\boldsymbol{d}}$ denote the corresponding normalized dimensionless residual vector, respectively. On the other hand, the discrepancy between the distributions of residual vectors in undamaged and damaged state is able to qualitatively reflect the existence of structural damage, and in this work we utilize the chi-square distribution function (CSDF) to characterize ARMAX residual vectors for structural damage identification:

$$f(x) = \begin{cases} \dfrac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{14}$$

where $x$ represent the random variable, $f(.)$ denotes the chi-square distribution function, $\Gamma(.)$ indicates the gamma distribution function, $k$ depicts a positive integer that specifies the number of degrees of freedom and affects the shape of the chi-square distribution function curves of residual vectors. Moreover, KLD is utilized to quantify the difference of distributions of ARMAX model residuals and used as the structural damage indicator in this study. At first, for the discrete random variable $X = \{x_1, x_2, \ldots, x_n\}\,(n \geq 2)$ and

$Y = \{y_1, y_2, \ldots, y_n\}$ from an uncertainty system, their corresponding probability distribution of each element are given as

$$P(X) = \{p_1(x), p_2(x), \ldots, p_n(x)\} \tag{15}$$

$$Q(Y) = \{q_1(y), q_2(y), \ldots, q_n(y)\} \tag{16}$$

where $p_i(x)$ and $q_i(y)$ represent the probability distribution function of the element $x_i$ and $y_i$, respectively; and $0 \leq p_i(x), q_i(y) \leq 1$, $\sum_{i=1}^{n} p_i(x)$ (or $q_i(y)) = 1$, $i = 1, \ldots, n$. The KLD [8] between the probability distributions of discrete random variables *X* and *Y* is defined as

$$D_{KL}(P(X) \| Q(Y)) = \sum_{i=1}^{n} p_i(x) \ln \frac{p_i(x)}{q_i(y)} \tag{17}$$

For the discrete random variables of ARMAX model residuals applied in linear time-invariant system,

$$\boldsymbol{e_u} = \{e_u(t), e_u(t-1), \ldots, e_u(t-n+1)\} \tag{18}$$

$$\boldsymbol{e_d} = \{e_d(t), e_d(t-1), \ldots, e_d(t-n+1)\} \tag{19}$$

where *t* indicates time index, $n \geq 2$ denotes the length of the residual vector, the corresponding probability distributions of each element of undamaged and damaged state are described as

$$P(\boldsymbol{e_u}) = \{p_1(e_u), p_2(e_u), \ldots, p_n(e_u)\} \tag{20}$$

$$P(\boldsymbol{e_d}) = \{p_1(e_d), p_2(e_d), \ldots, p_n(e_d)\} \tag{21}$$

Eventually, the structural damage indicator using KLD between distributions of ARMAX model residuals is defined as following:

$$D_{KL}(P(\tilde{\boldsymbol{e}}_{\boldsymbol{u}})\|P(\tilde{\boldsymbol{e}}_{\boldsymbol{d}})) = \sum_{i=1}^{n} p_i(\tilde{\boldsymbol{e}}_{\boldsymbol{u}}) \ln \frac{p_i(\tilde{\boldsymbol{e}}_{\boldsymbol{u}})}{q_i(\tilde{\boldsymbol{e}}_{\boldsymbol{d}})} \tag{22}$$

$$\tilde{\boldsymbol{e}}_{\boldsymbol{u}} = \text{sort}(\text{abs}(\bar{\boldsymbol{e}}_{\boldsymbol{u}})) \tag{23}$$

$$\tilde{\boldsymbol{e}}_{\boldsymbol{d}} = \text{sort}(\text{abs}(\bar{\boldsymbol{e}}_{\boldsymbol{d}})) \tag{24}$$

where $\tilde{\boldsymbol{e}}_{\boldsymbol{u}}$ and $\tilde{\boldsymbol{e}}_{\boldsymbol{d}}$ represent the rearrangement vectors of absolute value of $\bar{\boldsymbol{e}}_{\boldsymbol{u}}$ and $\bar{\boldsymbol{e}}_{\boldsymbol{d}}$ in ascending order, respectively; $P(.)$ means the corresponding chi-square distribution function.

## 3. Numerical simulation

### *3.1. Simulation setup*

In order to substantiate the performance of proposed method for damage detection, numerical simulation of damage detection on a six-story shear building structure has been conducted. As is depicted in Fig. 2, it is a six-story shear building model which can be simplified as a 6-DOF structure system, and the structure system is subjected to white noise excitation. The structural parameters are given as follows: the mass of every story is $1\times10^2$ kg, and the lateral stiffness is $1\times10^6$ N/m; damping ratio is assumed to be 3% for all modes; the first six natural frequencies of the shear model in undamaged state are given as 3.84 Hz, 11.29 Hz, 18.08 Hz, 23.83 Hz, 28.18 Hz and 30.91 Hz for the $1^{st}$ mode to the $6^{th}$ mode, respectively; the data sampling frequency is 200 Hz; taking into account the influence of environmental disturbance, measurement noises of 5% noise level are added into the acceleration data of all stories; there are totally 5×6=30 damage cases which consist of 10%, 20%, 30%, 40% and 50% reduction of lateral stiffness on every story. Fig. 3 shows the time series excitation of white noise.

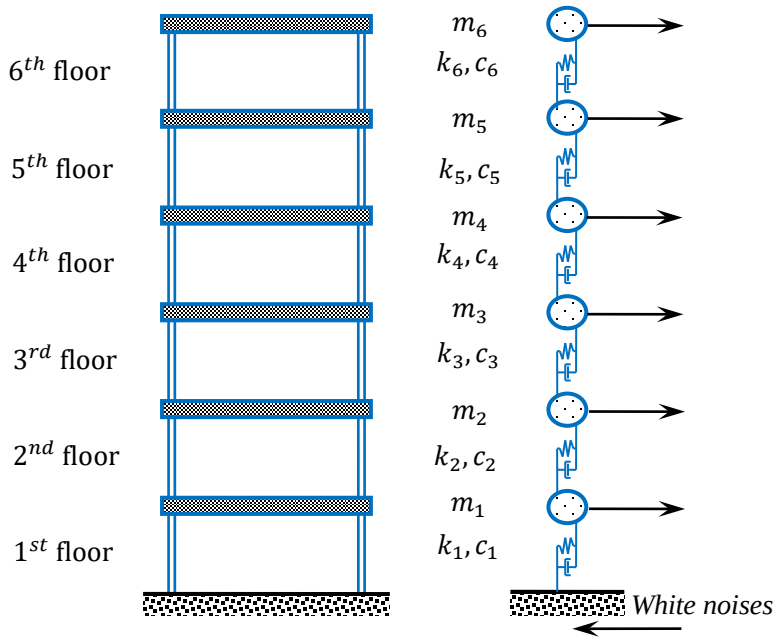



**Figure 2. 6-story shear building structure subjected to white noises excitaion**

*3.2. Procedure and results*

Primarily, the overall structure is divided into 6 substructures using the partition method mentioned in Section 2.2 (Eqs. (2) ~ (8)), as shown in Fig. 4. For each substructure from number 1 to 5, it can be modeled in good condition as a 2-input and 1-output ARMAX model, while the substructure 6 is modeled as 1-input and 1-output ARMAX model. For example, the ground acceleration ($\ddot{z}_g$) and the acceleration of the $2^{rd}$ DOF relative to the ground ($\ddot{y}_2$) are modeled as the input of substructure 1 while the $1^{st}$ DOF relative to the ground ($\ddot{y}_1$) is modeled as the output. Besides, the absolute acceleration ($\ddot{z}_{i-1}$) ($2 \leq i \leq 5$) of the $(i\text{-}1)^{th}$ DOF and the acceleration of the $(i+1)^{th}$ DOF relative to $(i\text{-}1)^{th}$ DOF ($\ddot{y}_{i+1}$) are modeled as the input of substructure $i$ while the acceleration of the $i^{th}$ DOF relative to the $(i\text{-}1)^{th}$ DOF ($\ddot{y}_i$) is modeled as the output. Especially for the top substructure, the absolute acceleration ($\ddot{z}_5$) of the $5^{th}$ DOF is modeled as the input and the acceleration of the $6^{th}$ DOF relative to $5^{th}$ DOF ($\ddot{y}_6$) is modeled as the output of substructure 6.

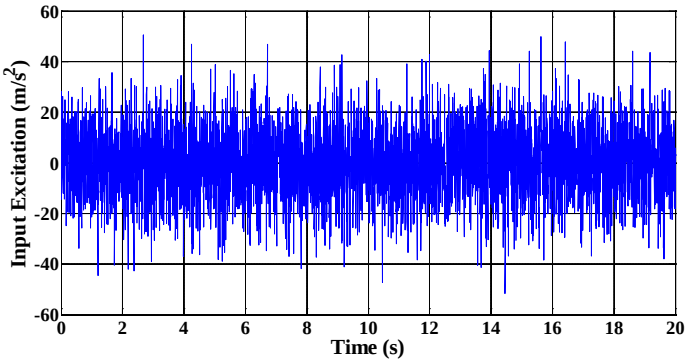



**Figure 3. Input excitation of white noises**

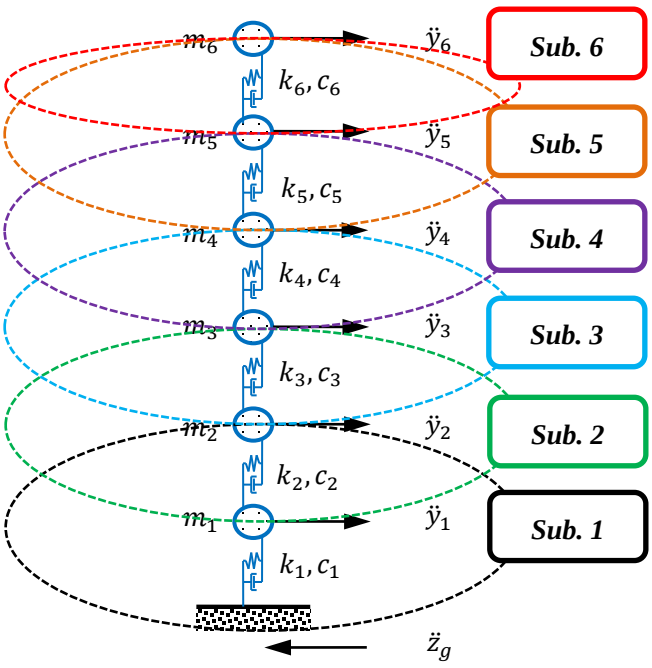



**Figure 4. Substructure division for 6-DOF simulated shear building structure**

The undamaged limit of each substructure is independently calculated through the mean value of KLD values between 10 data subsets in undamaged state, and the time duration of each subset is 20 secs. The changes in ARMAX model residuals from the undamaged system to the damaged systems can be reflected by its chi-square distribution function. All CSDF curves of 10% damage at the $1^{st}$ ~ $6^{th}$ floors are shown in Fig. 5 to indicate the damages existing in the structure. It can be observed from Fig. 5 that the CSDF curves shapes of damaged floors are conspicuous compared with others of undamaged floors. This is in good agreement with the theoretical expectation that ARMAX model residual of substructure in damaged state varies from the one of substructure in undamaged state since the substructural responses in damaged state cannot fit well the ARMAX model in undamaged state, as shown in Eqs. (9) ~ (14). Nevertheless, it is hard to exactly calculate the gap between the CSDF curves of different damage cases by only using the distinction of curves shapes. In view of this, KLD derived from Eqs. (15) ~ (24) is adopted to quantify the difference of CSDF curves between undamaged state and damaged state, and the complete identification of 10%, 20%, 30%, 40% and 50% damage cases is shown in the bar plots of Fig. 6. As a result, it reveals that there exists evident regularity in the damage location while the damage indicator can clearly quantify the damage with the damage degree increasing though it was interfered by the 5% noise. Therefore, it is explicitly reasonable that the proposed residual-based KLD is acceptably to reveal the linear relationship between the values of damage indicator and the structural stiffness reductions even in the case of a high severity of damage.

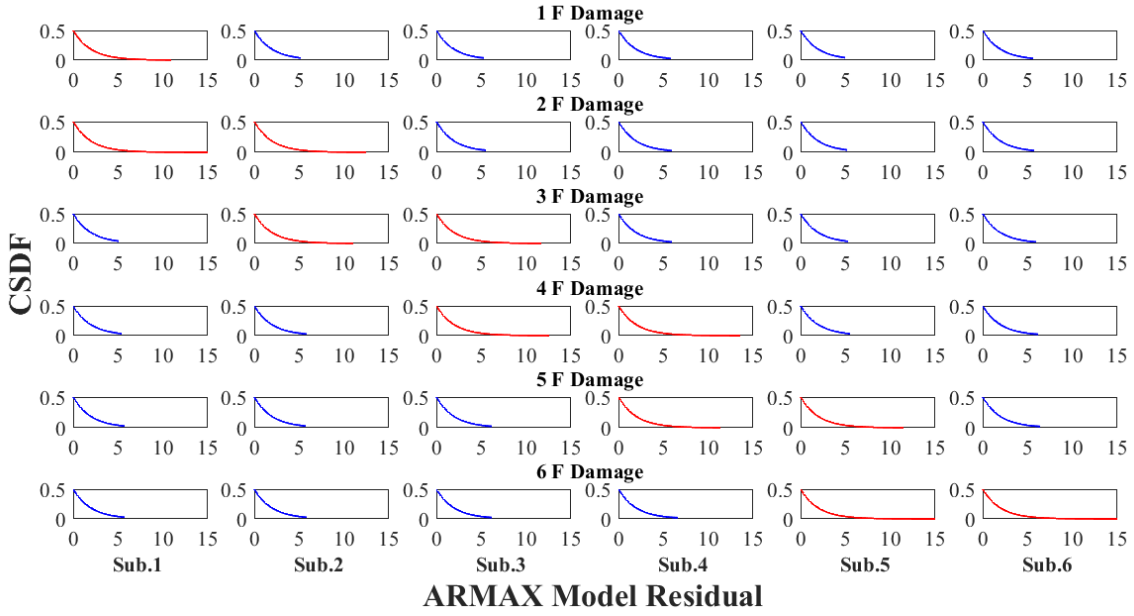



**Figure 5. Chi-square distribution function (CSDF) of ARMAX model residual** (white noise excitation, 5% noise, ARMAX model, data length = 4000, na = 2, nb = 3, nc = 3, and nk = 0; k = 2 (number of DOFs of CSDF))

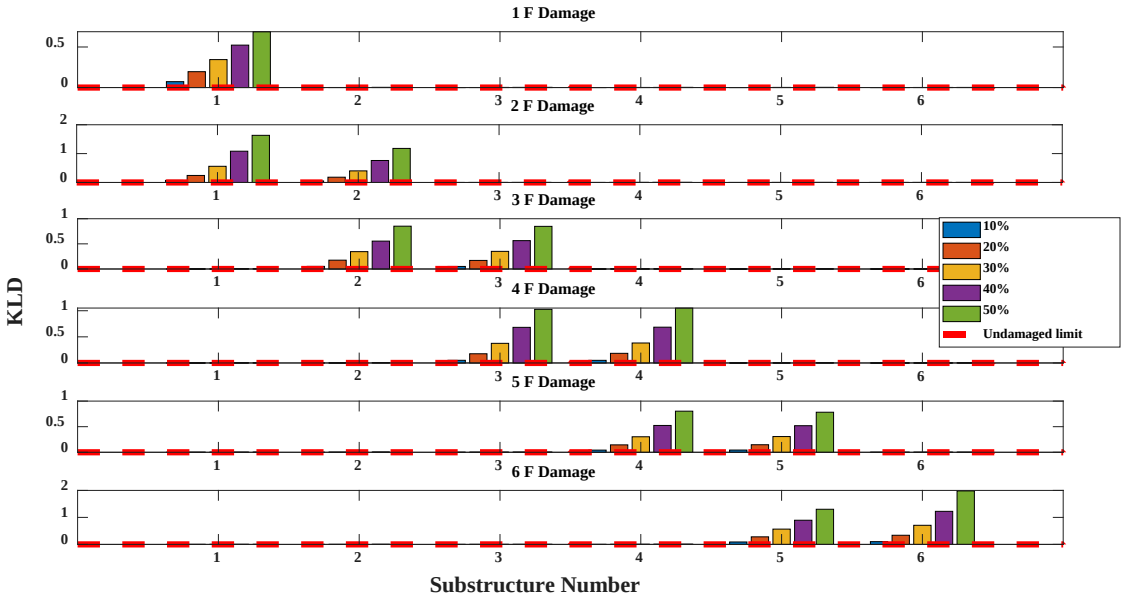



**Figure 6. Damage indicator of Kullback-Leibler divergence (KLD)** (white noise excitation, 5% noise, ARMAX model, data length = 4000, na = 2, nb = 3, nc = 3, and nk = 0; k = 2 (number of DOFs of CSDF))

## 4. Conclusions

This paper proposed an innovative substructural damage detection method based on damage indicator of ARMAX model residual-based KLD. Simulation of damage identification on a six-story shear building structure subjected to white noise is conducted to evaluate the performance of proposed damage detection strategy and damage indicator, and the results show that it can locate and quantify the damages of shear structures effectively by the proposed method. Due to the damage detection procedure can be implemented on each substructure independently, which suits for monitoring of key areas of actual engineering structure. On the other hand, the proposed CSDF curves of ARMAX model residual can clearly locate the structural damages in a visualized way with its distinguished tendencies; the proposed damage indicator of residual-based KLD can locate and quantify the damages in a data-driven way, which is suitable for local damage detection and does not rely on previous training data of various damage patterns. These mean that the proposed substructural damage detection approach is easy and efficient for local substructure damage detection of shear structures. In the following research, it is needed to further investigate about the identification of nonlinear damage in complex engineering structures with the proposed substructure damage identification method.

## Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (Grant No. 51708359), the Shenzhen Science, Technology and Innovation Commission (SZSTI) Basic Research Program (No. JCYJ20170817102125407), and the Scientific

Research Startup Fund of Shenzhen University (Grant No. 2016067) for which, the writers are grateful.

## References

[1] Mita, A. (2003) *Structural Dynamics for Health Monitoring*, Sankeisha: Nagoya, Japan.

[2] Xing, Z. and Mita, A. (2012) A substructure approach to local damage detection of shear structure, *Structural Control and Health Monitoring* **19**, 309–318.

[3] Mei, L., Mita, A. and Zhou, J. (2016) An improved substructural damage detection approach of shear structure based on ARMAX model residual, *Structural Control and Health Monitoring* **23**, 218–236.

[4] Farrar, C. R. and Iii, G. H. J. (1997) System identification from ambient vibration measurements on a bridge, *Journal of Sound & Vibration* **205**, 1–18.

[5] Figueiredo, E., Park, G., Farrar, C. R., Worden, K. and Figueiras, J. (2011) Machine learning algorithms for damage detection under operational and environmental variability, *Structural Health Monitoring* **10**, 559–572.

[6] Zheng, H. and Mita, A. (2008) Damage indicator defined as the distance between ARMA models for structural health monitoring, *Structural Control and Health Monitoring* **15**, 992–1005.

[7] Mei, L., Li, H., Zhou, Y., Wang, W. and Xing, F. (2019) Substructural damage detection in shear structures via ARMAX model and optimal subpattern assignment distance, *Engineering Structures* **191**, 625–639.

[8] Kullback, S. and Leibler, R.A. (1951) On information and sufficiency, *Annals of Mathematical Statistics* **22**, 79–86.

# A Mesh Refinement Algorithm for Mixed Boundary Value Elastic Problems

**†Md. Abdus Salam Akanda, and Aminul Islam Khan**

Department of Mechanical Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

†Corresponding and presenting author: masalamakanda@me.buet.ac.bd

## Abstract

The finite difference method (FDM) is a renounced numerical method for solving of complex problems of numerous fields. However, the efficacy of this method depends on the resolution of the mesh i.e. the size of the mesh used to obtain the solutions. In general, very small sized mesh, i.e., high mesh resolution is necessary to obtain an acceptable solution for various multi-scale physical problems. This high resolution of mesh consumes a significant amount of computational memory. Thus, huge wastage of computational resources occurs in refinement of sections of the domain where computation of the solution does not require high resolutions. This problem is effectively addressed by mesh refinement (MR) technique, a technique of local refinement of mesh only in sections where needed, thus allowing concentration of effort where it is required. The objective of this paper is to develop a mesh refinement algorithm for fourth order biharmonic equation which is widely used to solve boundary value elastic problems by using finite difference approach. Initial tests using the MR algorithm establish that the model adopted has considerable potential for mixed boundary value elastic problems. The results of initial test also show that consumption of computational resources is significantly less compare to uniform mesh (UM) while maintaining the quality of the solution.

**Keywords:** Finite Difference Method; Mesh Refinement; Fourth Order Biharmonic Equation; Mixed Boundary Value Problems

## Introduction

Finite difference method is extensively used to solve mixed boundary value elastic problems because of the simplicity of this method[1],[2],[3],[4],[5]. Usually mixed boundary value elastic problems are governed by fourth order biharmonic equation (FOBE) of potential function, $\Psi$. Analytical solution is not possible for FOBE which makes numerical method very popular for mixed boundary value problems. For instance, Ahmed et al. [1] analyzed stress-strain distributions of a both end fixed deep beams with mixed boundary conditions by applying FDM. Later, a generalized mathematical model for the solution of mixed-boundary-value elastic problems is depicted by Hossain et al. [4]. However, in these examples, the physical domains are discretized with high resolution uniform mesh, which consumes a significant amount of computational resources to store data. Moreover, high resolution of mesh involves the solution of a large matrix which ultimately accumulates huge amount of round of error during computation.

Previously, remedies of fine uniform mesh induced problems are sought by zooming the critical region, stress concentrated area [6]. However, the adaptability of this method is limited by the requirement of solutions for several times. Moreover, for this method, the

boundary conditions for current solution step depend on the previous step solution. So, if the solution of previous step is not acceptable due to less resolution, then current step solution would not be also acceptable. Recently, memory exhausting problems is significantly reduced with the development of powerful computers, however, an algorithm for the solution of mixed boundary value problems with less memory consumption is still demanding. To address this memory exhausting problem, the mesh refinement technique is extensively used in various field of study[7],[8],[9],[10].

Mesh refinement is a technique of local refinement of a mesh to allow computational resources and efforts where it is required. Sections of the physical domain needing high resolution are generally determined by means various criteria which includes comparing the solution to a threshold or the local rate of change to a solution. A mesh refinement algorithm based on the idea of multiple component grids for the solution of fourth order biharmonic equation using finite difference techniques is presented in this article. The solution of this equation is often smooth and easily approximated over large portions of their domains if there is no steep gradients, cracks or other discontinuity in the solution. However, most often the physical problem contains support boundary or locally isolated internal regions with steep gradients, cracks, or discontinuities, where the solution is difficult to approximate. We place locally finer grids in these regions over a coarse grid covering the domain. The solution on each line sub grid can then be approximated by standard finite difference techniques, as done on the coarse grid.

**Mathematical Formulation**

In terms of potential function, $\Psi$ under plain stress or plain strain condition, solution of boundary value elastic problems requires to solve the following fourth order biharmonic equations with appropriate boundary conditions [1].

$$\frac{\partial^4\Psi}{\partial x^4}+\frac{\partial^4\Psi}{\partial x^2\partial y^2}+\frac{\partial^4\Psi}{\partial y^4}=0 \tag{1}$$

The relation between potential function and displacement components are as follows

$$u_x=\frac{\partial^2\Psi}{\partial x\partial y} \tag{2}$$

$$u_y=-\frac{1}{1+\mu}\left[2\frac{\partial^2\Psi}{\partial x^2}+(1-\mu)\frac{\partial^2\Psi}{\partial y^2}\right] \tag{3}$$

where $u_x$ and $u_y$ are the displacement components in the *x*- and *y*- directions respectively. The relation between stress components, displacement components are as follows

$$\sigma_x(x,y)=\frac{E}{1-\mu^2}\left(\frac{\partial u_x}{\partial x}+\mu\frac{\partial u_y}{\partial y}\right)=\frac{E}{(1+\mu)^2}\left(\frac{\partial^3\Psi}{\partial x^2\partial y}-\mu\frac{\partial^3\Psi}{\partial y^3}\right) \tag{4}$$

$$\sigma_y(x,y)=\frac{E}{1-\mu^2}\left(\frac{\partial u_y}{\partial y}+\mu\frac{\partial u_x}{\partial x}\right)=-\frac{E}{(1+\mu)^2}\left(\frac{\partial^3\Psi}{\partial y^3}+(2+\mu)\frac{\partial^3\Psi}{\partial x^2\partial y}\right) \tag{5}$$

$$\sigma_{xy}(x,y)=\frac{E}{2(1+\mu)}\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right)=\frac{E}{(1+\mu)^2}\left(\mu\frac{\partial^3\Psi}{\partial x\partial y^2}-\frac{\partial^3\Psi}{\partial x^3}\right) \tag{6}$$

Since, the target of this paper is to develop a mesh refinement algorithm for fourth order biharmonic equation, we select a very simple mixed boundary value elastic problem as shown in Fig. 1. In the considered problem, a simple elastic member of length ‘2b’ and width ‘a’ has an embedded crack under the uniform axial loading. For simplicity, we considered there is no crack growth under this uniform loading condition. The material geometry of the problem is taken as a/b=1.0 and size of the crack is taken as one fourth (a/4) of the width of the member.

Referring to Fig. 1, for this problem, both the top and bottom edges are free and the both lateral edges are subjected to uniform tensile loading. Taking the advantages of symmetry, the right half section of the elastic member is solved under MR and finite element method (FEM) with necessary BCs as shown in Fig. 2.

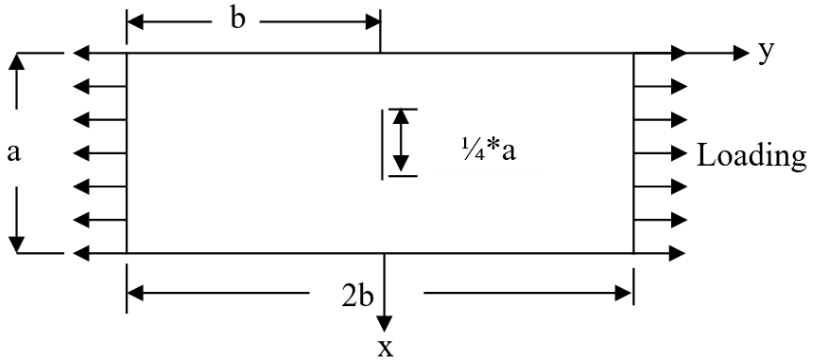



**Figure 1:** Simple bar with embedded crack under uniform tensile stress

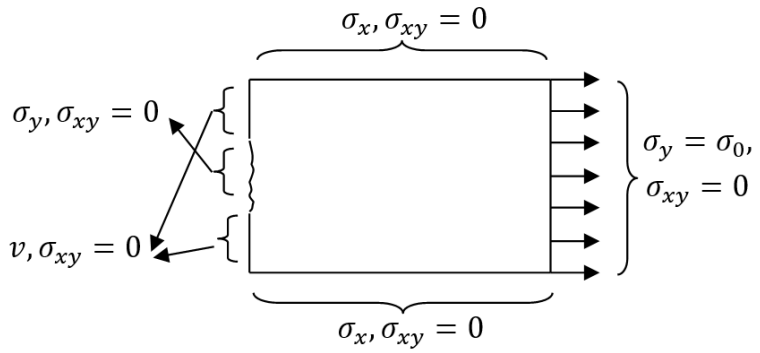



**Figure 2:** Half section of the problem with necessary boundary conditions

As far as numerical solution is concerned, it is evident from the expression of boundary conditions (please see Fig. 2) that all the boundary conditions of interest can easily be discretized in terms of displacement function, $\Psi$ by the finite-difference method.

**Numerical Method**

The replacement process of continuous problem by a discrete problem whose solution approximates the solution of continuous problem under numerical method is known discretization. Under mesh refinement technique, first the domain is discretized with a coarse grid. Then, finer grids are added in the region which requires more resolution. An example of discretization of the considered problem under mesh refinement is shown in Fig. 3. From theory of elasticity we know that the crack location is the critical region for this problem. So, under the mesh refinement technique, the finer mesh is taken in that region. Since, no time scale is associated with the problem, instead of adaptive mesh refinement, we introduce statistically refined mesh. Under any numerical method, the governing equation (Eq. 1) must be satisfied each interior nodal point. Since, the domain is discretized into variable sized mesh, uniform grid-based stencil (Stencil-1 of Fig. 4) cannot be used throughout the domain. Thus, to satisfy the governing equation throughout the domain, several stencils are formulated as shown Fig. 4 (Stencils 2-6). Details of the stencil formulation can be found in reference These stencils can fully satisfy the governing equation throughout the domain.

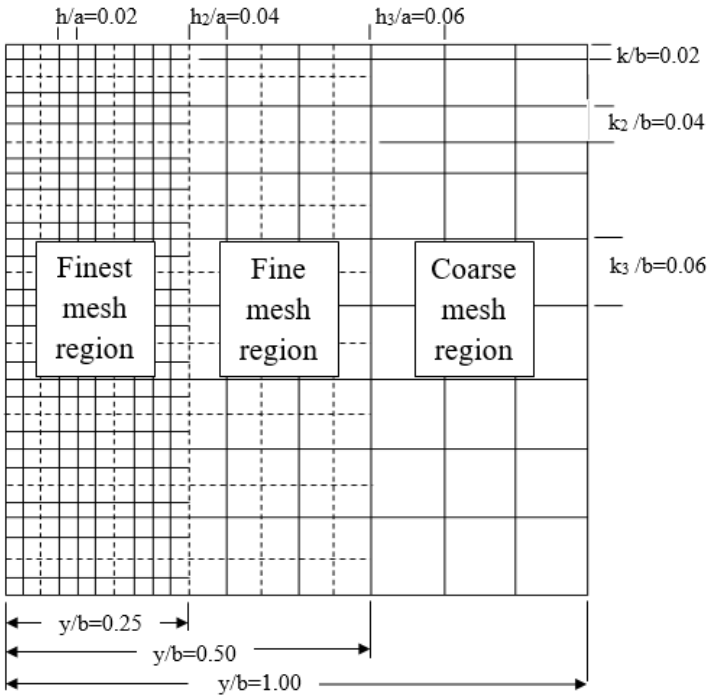



**Figure 3:** Discretization of domain under mesh refinement technique with three different size of mesh.

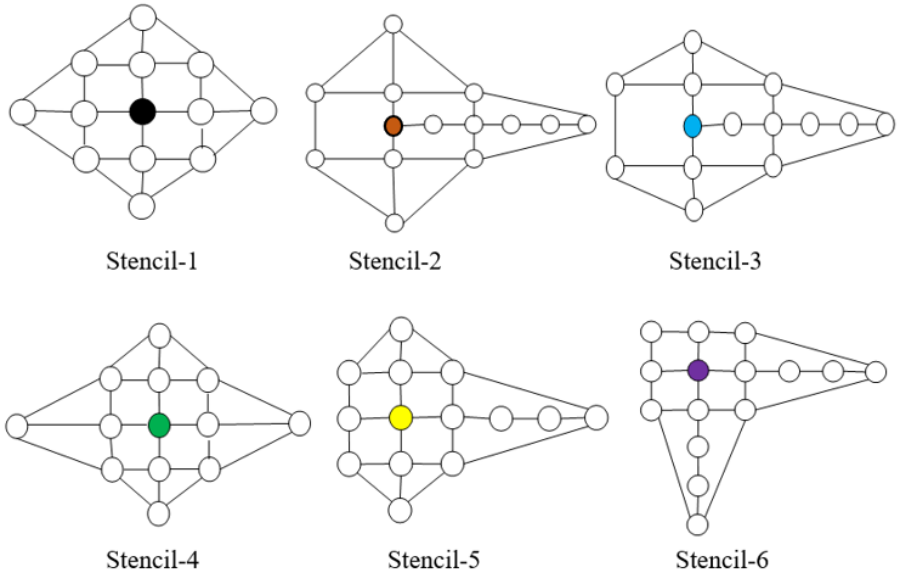



**Figure 4:** Types of stencils for governing equation

As seen from the problem definition, each physical boundary is defined by two conditions. This double conditions problem is satisfied by bringing an imaginary boundary [4]. The stencils of various boundary parameters over uniform mesh are shown in Fig. 5. However, these uniform mesh-based stencils are not applicable on the transition node, a node that connects two sizes of mesh. Special stencils are required for these transition nodes which is shown in Fig. 6.

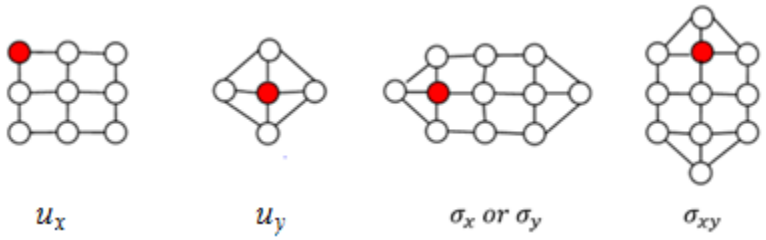



**Figure 5:** Stencils for displacement and stress components

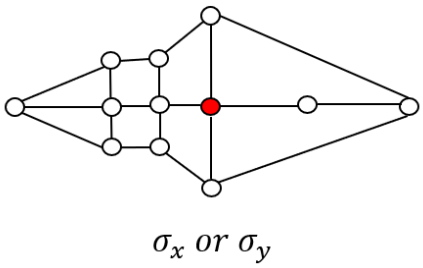



**Figure 6:** Special stencil for stress component

## Results and Discussions

In MR FDM, finest mesh size is taken as $0.01 \times 0.01$, fine mesh size is taken as $0.02 \times 0.02$ and coarse mesh size is taken as $0.04 \times 0.04$. A uniform mesh of size $0.01 \times 0.01$ is taken for FEM discretization. Mesh sensitivity analysis is performed for both methods (data not shown). The displacement components at y/b = 0.0 with MR FDM with refine mesh (RM) and FEM are shown in Fig. 7 as a comparative study. Fig. 7a shows displacement component, $u_x/a$ distribution at section y/b = 0.0. Except at the tip of the crack both methods show same amount of displacement in x-direction. At the tip of the crack FEM shows a little bit higher displacement, however, this disagreement is not significant. Fig. 7b shows displacement component $u_y/b$ distribution at section y/b=0.0 and the results of both are in good agreement. In other section of the member, both methods provide exactly same amount of displacement (data not shown). In every case, MR results are as good as FEM results although greater no of nodal points is considered under FEM discretization.

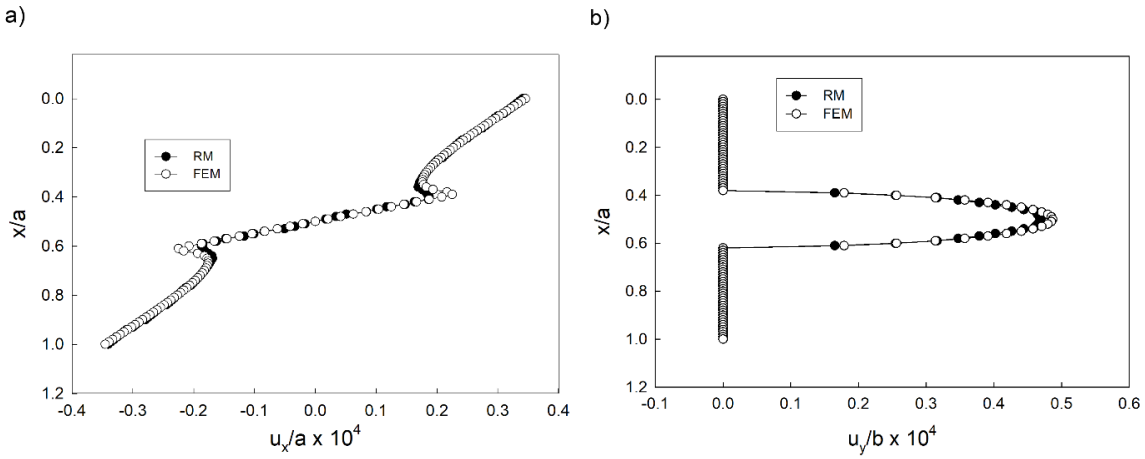



**Figure 7:** Comparison of MR FDM results with FEM results at section y/b=0.0: a) $u_x$ distribution and b) $u_y$ distribution.

The comparison of stress components at y/b = 0.0 is shown in Fig. 8. For this type of problems, the most desired parameter is the stress component in direction to applied stress and, in this case, it is $\sigma_y$ which is shown in Fig. 8a. From this figure it is seen that results of both methods are in good agreement. The maximum stress is observed as 3.75 times of applied stress for FEM and around 3.8 times of applied stress for MR FDM. The normal stress, $\sigma_x$ distribution is shown in Fig. 8b. From figure 8b, it is seen that the pattern of distribution is similar for both methods, but FEM give somewhat larger stress than that of mesh refinement technique. As stated in earlier example, this discrepancy arises due to the application of three BCs at the singularity points.

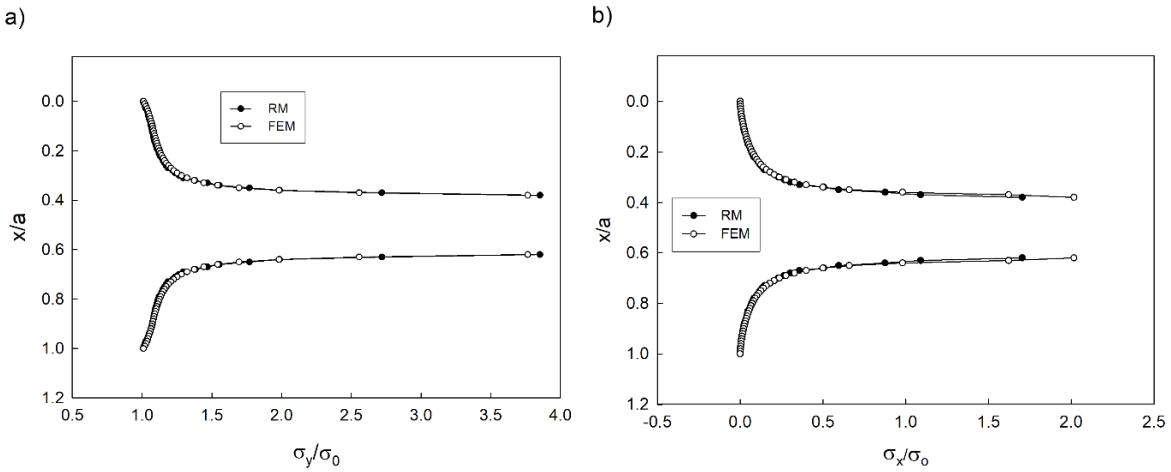



**Figure 8:** Comparison of MR FDM results with FEM results at y/b=0.0 in terms of stress components: a) $\sigma_y/\sigma_o$ distribution and b) $\sigma_x/\sigma_o$ distribution.

## Conclusions

A mesh refinement algorithm is developed for fourth order biharmonic equation which is widely used to investigate the displacements and stress analysis of mixed boundary value problems. The governing equation is discretized by finite difference method in various way to develop various stencils which are required to satisfy the governing equation throughout the domain. Due to mesh refinement, the boundaries are also discretized into irregular meshes. As a result, the boundary conditions also need to be discretized in different way than regular mesh. Our results show that the developed method can easily be used to obtain the solution of mixed boundary value elastic problems in terms of displacement and stress components. Our results also show that a reduced number of nodes can yields results as good as finite element method.

## Acknowledgements

This work has been carried out in the department of Mechanical Engineering, Bangladesh University of Engineering and Technology (BUET). The authors gratefully acknowledge the support and facilities provided by BUET.

### References

[1] Ahmed, S. R., Idris, A. B. M., and Uddin, M. W. (1996) Numerical solution of both ends fixed deep beams, C*omputers & Structures* **61**, 21-29.

[2] Akanda, M. A. S., Ahmed, S. R., Khan, M. R., and Uddin, M. W. (2000) A finite-difference scheme for mixed boundary value problems of arbitrary-shaped elastic bodies, *Advances in Engineering Software*, **31**, 173-184.

[3] Akanda, M. A. S., Ahmed, S. R., and Uddin, M. W. (2002) Stress analysis of gear teeth using displacement potential function and finite differences, I*nternational Journal for Numerical Methods in Engineering*, **53**, 1629-1640.
[4] Hossain, M. Z., Ahmed, S. R., and Uddin, M. W. (2005) Generalized mathematical model for the solution of mixed-boundary-value elastic problems, *Applied Mathematics and Computation*, **169**, 1247-1275.
[5] Afsar, A. M., Nath, S. K. D., Ahmed, S. R., and Song, J. L. (2008) Displacement potential based finite difference solution to elastic field in a cantilever beam of orthotropic composite, *Mechanics of Advanced Materials and Structures*, 15, 386-399.
[6] Akanda, M. A. S., Noman, K. N. A., and Bhuiyan, M. S., Zooming of Critical Zone to Refine Finite Difference Solution of Elastic Problem, *Proceedings of 4th International Mechanical Engineering Conference & 9th annual Paper Meet*, Dhaka, Bangladesh, 2004, 12-18.
[7] Berger, M. J., and Oliger, J. (1984) adaptive mesh refinement for hyperbolic partial-differential equations, *Journal of Computational Physics*, **53**, 484-512.
[8] Keppens, R., Nool, M., Toth, G., and Goedbloed, J. P. (2003) Adaptive Mesh Refinement for conservative systems: multi-dimensional efficiency evaluation, *Computer Physics Communications*, **153**, 317-339.
[9] Min, C. H., Gibou, F., and Ceniceros, H. D. (2006) A supra-convergent finite difference scheme for the variable coefficient Poisson equation on non-graded grids, *Journal of Computational Physics*, **218**, 123-140.
[10] Shen, C. P., Qiu, J. M., and Christlieb, A. (2011) Adaptive mesh refinement based on high order finite difference WENO scheme for multi-scale simulations, *Journal of Computational Physics*, **230**, 3780-3802.
[11] Khan, A. I., *Local mesh refinement scheme for finite difference solution of mixed boundary-value elastic problems*, M.Sc. Thesis, Bangladesh University of Engineering & Technology, Bangladesh, 2014.

# Mechanical analysis of widespread corrosion-damaged structure based on ES-PIM

**†Keliang Ren, *Jimin Li[1], Hua Ji[2] ,Ting Wang[3] ,Lihong Ding[4] and Jia Yang[5]**

† Department of physics and electrical and electronic engineering, Ningxia University, P. R. China

[1,3,5]Department of physics and electrical and electronic engineering, Ningxia University, P. R. China

[2]College of Mechanical Engineering, Ningxia University, P. R. China

[4]Department of mathematics and statistics, Ningxia University, P. R. China

*Presenting author: liangliang910201@126.com

†Corresponding author: kl_ren@nxu.edu.cn

## Abstract

In order to solve the problem of large limitation of simulation results caused by load factors and corrosion conditions respectively in the study of corrosion failure of metallic materials, the idea of combining cellular automata and edge-based smoothed point interpolation method (ES-PIM) is adopted. The corrosion process of aluminum under environmental action was simulated by cellular automata method and the corrosion topography was obtained which was used for the establishment of calculation model. To overcome overly-stiff property existing in the widely used finite element method (FEM) with linear triangular elements, the edge-based smoothed point interpolation method (ES-PIM) is used for calculation which has been found both spatially and temporally stable, and works well for both static and dynamic problems.

**Keywords:** Widespread corrosion damage Cellular Automaton Meshfree method

## 1.Introduction

As a kind of nonferrous metal structural materials, aluminum and its alloys are widely used in aviation, aerospace, automobile, machinery manufacturing and so on. In the process of its use, it will suffer different forms of damage due to the influence of the service environment, among which corrosion is a common form of damage. However, the failure of aluminum structure is not caused by the single factor of corrosion, usually accompanied by the action of load on it. Most scholars have simulated the corrosion behavior of aluminum based on the corrosion mechanism of aluminum, but have not carried out mechanical analysis on the corrosion defect structure generated subsequently. Therefore, it is of great significance to simulate the corrosion and failure behavior of aluminum and its alloys in the service environment.

In the field of aluminum corrosion behavior simulation, Engelhardt et al.[1-3] established a damage function analysis model to predict local corrosion such as pitting corrosion, crevice corrosion and stress corrosion. Urquidi-macdonald et al.[4] studied the relationship between the crack growth rate caused by pitting corrosion and various parameters (corrosion potential, pH, temperature and conductivity) using artificial neural network technology (ANN).In Cellular Automata (CA) model, space is divided into cells with finite states. These cells

evolve according to certain local rules. By the CA model based on the rules of local reactions in the system, can reflect the influence of different scales complex physical and chemical systems, by defining the molecular scale or the interaction of the atomic scale, within the scope of the macro qualitatively describe the nature of the complex system, so using the CA technology to rot corrosion in the process modeling more intuitive and convenient, like Wang Hui et al. [5-6] adopted cellular automata (CA) method to obtain the growth and evolution curve of corrosion pits and the change curve of geometric morphology with time.

After years of development, finite element method (FEM) has become an important tool for modeling and simulation of solid and complex geometric structures. However, the finite element method has some inherent defects. For example, the finite element model is "overly-stiff", which affects the accuracy of strain, especially for stress results; When the mesh is severely deformed, the precision of the solution will be affected.

During mesh generation, using triangle (for 2D) or tetrahedron (for 3D) elements, mesh generation becomes much easier and is usually done automatically without manual manipulation. But at the same time, the accuracy of finite element calculation results is often very poor.

Professor G.R. LIU[7-9]and his team used point-based polynomial interpolation (PIM)[10] or radial basis interpolation(RPIM)[11] to construct the shape function, and introduced smoothed Galerkin weakform and generalized gradient smooth operator [12,13]. Thus, the node-based smoothed point interpolation method [14,15](NS-PIM) can be obtained.

In practical application, this method shows that it can withstand mesh deformation better and still has good calculation results under triangular mesh. Moreover, this method provides the upper bound of energy norm. However, the stiffness matrix **K** obtained by this method is smaller than the actual stiffness matrix **K**, that is, the smoothed system is "overly-soft", resulting in a displacement larger than the actual displacement. In order to solve this problem, Edge-based smoothed Point Interpolation Method[16-18] (ES-PIM) is introduced.

In ES-PIM, every edge-related problem domain of the background grid is smoothed. Compared with the node-based smooth operator, the smoothness of edge-basis strain can weaken the degree of softening, so that the ES-PIM model is closer to the exact stiffness and has a good calculation effect in both static and dynamic problems.

For the above methods, professor G.R.LIU and his team established Galerkin weakened weak ($W^2$) formulation constructed by generalized gradient smoothing operator.

Weakened weak ($W^2$) formulation seeks solutions in G space[19]which is a function space containing both continuous and discontinuous functions. G space includes all the continuous and discontinuous displacement cases under the framework of FEM and Meshfree. Therefore, in the framework of finite element and meshless method, it is suitable for both compatible and incompatible displacements. By using the generalized strain smoothing technique, we can

obtain the generalized smooth Galerkin weak form for all the above methods.

In this paper, the widespread corrosion behavior of aluminum is simulated by using cellular automata method, and a computational model is established based on the simulated corrosion topography. The model is analyzed by introducing edge-based smoothed point interpolation method and the results are obtained. The combination of the two methods provides a new way to analyze the mechanical properties of structures with corrosion defects.

## 2. Establishment of widespread corrosion model

### *2.1 Definition of cellular automata*

Cellular Automata (CA) is defined as a dynamic system that evolves in discrete time dimensions in accordance with certain local rules in a cellular space composed of cells with discrete and finite states.

In the process of its evolution, each cell can change according to local rules state, namely based on cellular automata and its neighbor cell state with this to determine the next state of cellular automata, all belong to sync status updates, in accordance with local rules the entire cellular space show the change of state in discrete time.

### *2.2 Boundary conditions*

When simulating a given cellular automata rule, one cannot deal with an infinite lattice. The system must be finite and have boundaries. Clearly, a site belonging to the lattice boundary does not have the same neighborhood as other internal sites. In order to define the behavior of these sites, a different evolution rule can be considered, which sees the appropriate neighborhood. This means that the information of being, or not, at a boundary is coded at the site and, depending on his information, a different rule is selected.

Each variable of cellular automata has a finite number of states and is local in time and space. In order to keep each variable in the cellular space free from the influence of the external environment, the boundary conditions of the model are usually defined as follows: Periodic boundary conditions are used for the left and right boundary to keep the properties of the system and the element unchanged, and the theoretical infinity of cellular space is realized. The upper and lower boundaries adopt fixed boundary conditions to ensure the non-correlation of upper and lower boundary cells, as shown in Figure 1.

A cellular automata rule is local, by definition. The updating of a given cell requires one to know only the state of the cells in its vicinity. The spatial region in which a cell needs to search is called the neighborhood. In principle, there is no restriction on the size of the neighborhood, except that it is the same for all cells. However, in practice, it is often made up of adjacent cells only. If the neighborhood is too large, the complexity of the rule may be unacceptable (complexity usually grows exponentially fast with the number of cells in the neighborhood).

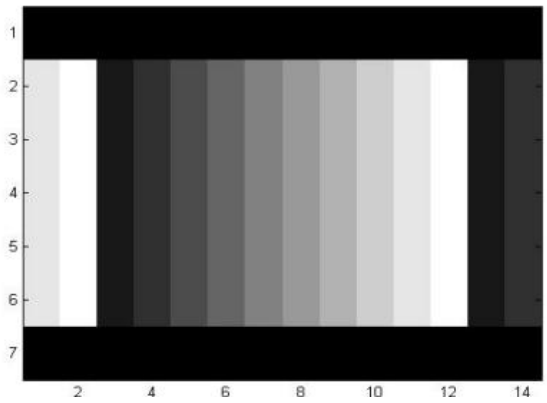
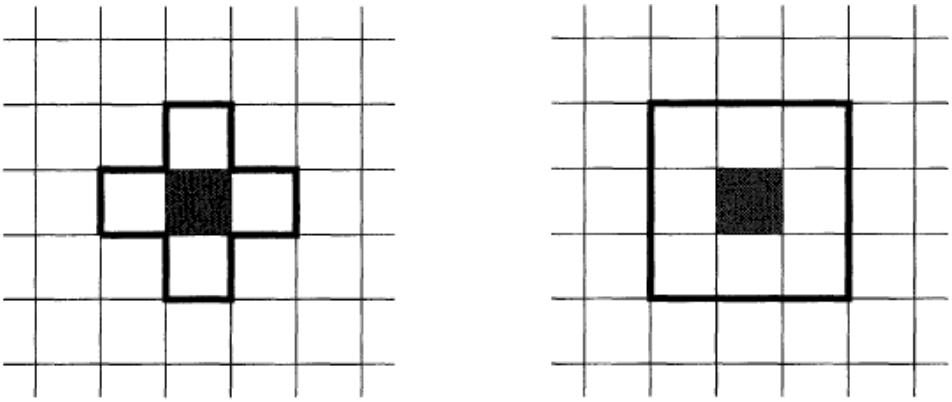



**Figure 1. Schematic Diagram of boundary conditions**

For two-dimensional cellular automata, two neighborhoods are often considered: the von Neumann neighborhood, which consists of a central cell (the one which is to be updated) and its four geographical neighbors north, west, south and east. The Moore neighborhood contains, in addition, second nearest neighbors northeast, northwest, southeast and southwest, that is a total of nine cells.[7]

Figure 2. illustrates these two standard neighborhoods.



(a) (b)

**Figure 2. (a) Von Neumann and (b) Moore neighborhoods. The shaded region indicates the central cell which is updated according to the state of the cells located within the domain marked with the bold line.**

### *2.3 transformation rule*

In the study on the formation process of metal corrosion products, the cellular transformation process is realized, which is expressed by the following conversion equation:

$$M + C \rightarrow P \tag{1}$$

In the top formula, M is the metal cell; C is corrosive cell; P is the cell of corrosion product. In particular, the concentration of oxygen in the atmosphere changes very little, and the key factor affecting the corrosion rate of metal oxygen absorption is the relative humidity of water molecules. Therefore, in order to simplify the CA model, only water molecules involved in the reaction are represented by C in the equation, and the effect of oxygen on corrosion is no

longer considered.

Cellular conversion rules: only corrosion cell C in cellular space has random mobility (top, bottom, left, right). If C moves in the direction of neighbor metal cell M, both M and C are converted to P. Conversely, if the neighbor is C, all cells remain unchanged. If the neighbor in the direction of C is a space bit, then C will jump to the space bit, and the original position will randomly become a space bit; If all neighbors of M are C, all cells remain unchanged. If at least one neighbor of M is C, and C does not move toward M, M also remains unchanged.

## 3. Numerical model

### *3.1 Brief on Basic Equations*

The stress-strain problem of corrosion-damaged structure belongs to the linear elastic problem of solid mechanics, so we first brief the basic equations for solid mechanics problem of linear elasticity.

Consider a two-dimensional solid mechanics problem with a physical domain of $\Omega \in R^2$ bounded by $\Gamma$.The static equilibrium equation for 2D solids in the domain $\Omega \in R^2$ can be written as:

$$\frac{\partial \sigma_{ij}}{\partial x_j} + b_i = 0, \qquad i, j = 1, 2 \tag{2}$$

where $b_i$ are given external body force and $\sigma_{ij}$ is the stress tensor which relates to the strains tensor $\varepsilon_{ij}$ via the constitutive equation or the Generalized Hook's law:

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{3}$$

where $C_{ijkl}$ is elasticity tensor of material property constants.

The strains tensor $\varepsilon_{ij}$ relates to the displacement by the following compatibility equation.

$$\varepsilon_{kl} = \frac{1}{2}\left( \frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right) \tag{4}$$

where $u_i$, $i = 1, 2$ is the displacement components in the xi-directions at a point in $\Omega$ .

In matrix form, the equilibrium Eq.(2) becomes:

$$\mathbf{L}_d \boldsymbol{\sigma} + \mathbf{b} = 0 \tag{5}$$

where $\mathbf{L}_d$ is a matrix of the differential operator defined as:

$$\mathbf{L}_d\left( \frac{\partial}{\partial x_1}, \quad \frac{\partial}{\partial x_2} \right) = \begin{bmatrix} \frac{\partial}{\partial x_1} & 0 \\ 0 & \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} \end{bmatrix} \tag{6}$$

The constitutive equation becomes:

$$\boldsymbol{\sigma} = \mathbf{C}\boldsymbol{\varepsilon} \tag{7}$$

Where $\mathbf{C}$ is matrix of material properties which entries of $C_{ijkl}$ , $\boldsymbol{\sigma} = \{\sigma_{11} \quad \sigma_{22} \quad \sigma_{12}\}^T$ and $\boldsymbol{\varepsilon} = \{\varepsilon_{11} \quad \varepsilon_{22} \quad \varepsilon_{\mathrm{r}2}\}^T$

The compatibility equation (4) can also be written in the matrix form of :

$$\boldsymbol{\varepsilon} = \mathbf{L}_d \mathbf{u} \tag{8}$$

where $\mathbf{u} = \{u_1 \quad u_2\}^T$ is the displacement vector. Substituting Eq.(8) into (7) and then into (5):

$$\mathbf{L}_d^T \mathbf{C} \mathbf{L}_d \mathbf{u} + \mathbf{b} = \mathbf{0} \tag{9}$$

There are two types of boundary conditions: Dirichlet (or essential / displacement ) boundary conditions and Neumann (or natural / stress) boundary conditions. Let $\Gamma_D$ denote a part of $\Gamma$ , on which homogenous Dirichlet boundary condition is specified, then we can obtain:

$$\Gamma_D \in \Gamma \quad \text{on} \quad u_i = 0, \tag{10}$$

Let $\Gamma_N$ denotes a part of $\Gamma$ , on which Neumann boundary condition is satisfied,

$$\sigma_{ij} n_j = t_i, \quad \text{on} \quad \Gamma_N \in \Gamma \tag{11}$$

where $n_j$ is unit outward normal vector, and $t$ is the specified boundary stress on $\Gamma_N$ , respectively. The matrix form of Eq.(11) is as follows:

$$\mathbf{L}_n^T \sigma = \mathbf{t}, \quad \text{on} \quad \Gamma_N \in \Gamma \tag{12}$$

Where

$$\mathbf{L}_n \begin{pmatrix} n_x & n_y \end{pmatrix} = \begin{bmatrix} n_x & 0 \\ 0 & n_y \\ n_y & n_x \end{bmatrix} \tag{13}$$

### *3.2 Displacement field approximation using the PIM*

The point interpolation method (PIM) obtains the approximation by letting the interpolation function pass through the function values at each scattered node within the local supporting domain.

Different types of point interpolation can be constructed by using different types of base functions. There are two common types: Polynomial point interpolation method (PIM) based on polynomial basis function and radial point interpolation method (RPIM) based on radial basis function.

For the polynomial PIM, the formulations start with the following assumption:

$$u(x) = \sum_{i=1}^{n} P_i(\mathbf{x}) a_i = \mathbf{P}^T(\mathbf{x})\mathbf{a} \tag{14}$$

Where $u(x)$ is a field variable function defined in the Cartesian coordinate space $\mathbf{x}^T = \{x \quad y\}$, $P_i(\mathbf{x})$ is the basis function of monomials which is usually built utilizing Pascal's triangles, $a_i$ is the corresponding coefficient, and $n$ is the number of nodes in the local support domain. The complete polynomial basis of orders 1 and p can be written as:

$$\mathbf{P}^T(\mathbf{x}) = \begin{pmatrix} 1 & x & x^2 & \cdots & x^p \end{pmatrix}, \quad 1D \tag{15}$$

$$\mathbf{P}^T(\mathbf{x}) = \begin{pmatrix} 1 & x & y & x^2 & xy & \cdots & x^p & y^p \end{pmatrix}, \quad pD \tag{16}$$

For the radial PIM, using radial basis functions augmented with polynomials, the field function can be approximated as follows:

$$u(x) = \sum_{i=1}^{n} R_i(\mathbf{x}) a_i + \sum_{j=1}^{m} P_i(\mathbf{x}) b_j = \mathbf{R}^T(\mathbf{x})\mathbf{a} + \mathbf{P}^T(\mathbf{x})\mathbf{b} \tag{17}$$

Where $R_i(\mathbf{x})$ and $P_j(\mathbf{x})$ are radial basis functions and polynomial basis functions respectively, $a_i$ and $b_j$ are corresponding coefficients, $n$ is the number of nodes in the local support domain and $m$ is the number of polynomial terms.

The coefficients in Eqs.(14) and (17) can be determined by enforcing the field function to be satisfied at the $n$ nodes within the local support domain. Finally, the PIM shape functions can be obtained and the field function can be expressed as:

$$u(\mathbf{x}) = \sum_{i=1}^{n} \varphi_i(\mathbf{x}) d_i = \mathbf{\Phi}^T(\mathbf{x})\mathbf{d} \tag{18}$$

where $d_i$ is a nodal function value and $\varphi_i(\mathbf{x})$ is the PIM shape function which possesses the Kronecker delta function property. In the above formulation, it is noticed that we need to properly select $n$ nodes for interpolation ensuring a nonsingular moment matrix.

### *3.3 T schemes for node selection*

In this paper, three-point triangular background element is adopted to discretize the problem domain. The element can be generated automatically without manual operation, and the mesh density of triangular background element can be adjusted according to the need of computational accuracy.

The T scheme used in ES-PIM method mainly includes T3 scheme, T6/T3 scheme and T6 scheme. The following mainly introduce the T3 and T3/T6 scheme which used for programming.

The T3 scheme mainly uses the three vertices of the triangular mesh where the calculation point is located to represent the displacement function of the calculation point, and its displacement field is a linear displacement field. This method has many similarities with the first-order finite element interpolation method. The T3 scheme is used only for creating linear PIM shape functions by using polynomial basis functions. As illustrated in Figure 3.

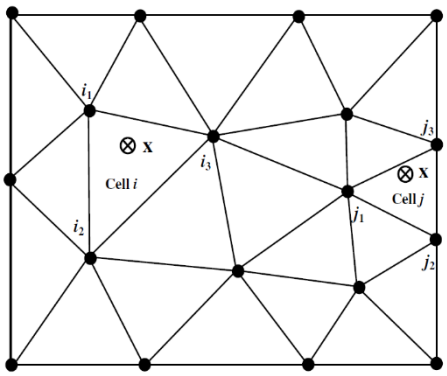



**Figure 3. Illustration of the T3 scheme of node selection**

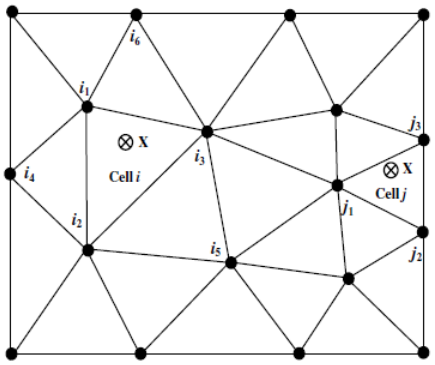



**Figure 4. Illustration of the T6/T3 scheme of node selection**

Whether the point of interest $(\mathbf{x})$ is located in an interior cell (element $i$ ) or a boundary cell (element $j$ ), only the three nodes of the home cell ( $i_1 - i_3$ or $j_1 - j_3$) are selected.

In the T6/T3 scheme, the point of interest located in the boundary cell only needs three nodes to interpolate and the linear displacement approximation function also constructed by these three nodes. For the point of interest located in an interior cell needs six nodes to interpolate and been used for construct a quadratic displacement approximation function. As illustrated in Figure 4.

It not only successfully overcomes the singular problem which exists in the process of PIM approximation by using the polynomial basis but also improves the efficiency of the method.

### *3.4 Edge-based smoothed strains*

In consideration of the displacement field is not continuous, the generalized smoothed Galerkin weak form or the weakened weak form which has exactly the same form as the standard Galerkin weak form need to be used.

$$\int_\Omega \delta\left(\hat{\varepsilon}\left(\mathbf{u}\right)\right)^T \mathbf{D}\left(\hat{\varepsilon}\left(\mathbf{u}\right)\right) d\Omega - \int_\Omega \delta\mathbf{u}^T \mathbf{b} d\Omega - \int_{\Gamma_t} \delta\mathbf{u}^T \mathbf{t} d\Gamma = 0 \tag{19}$$

Thus, the formulation procedure is exactly the same as that in the standard FEM and all we need to do is to use the edge-based smoothed strain $\hat{\varepsilon}$ in place of the compatible strain fields $\tilde{\varepsilon}$.

In the framework of W² formulation, the gradient of the field function (strains) will be obtained using the following generalized smoothing operation which considers both continuous and discontinuous displacement functions.[20]

$$\hat{\varepsilon}_k = \begin{cases} \dfrac{1}{A_k}\int_{\Omega_k} \tilde{\varepsilon}(x) d\Omega = \dfrac{1}{A_k}\int_{\Gamma_k} \mathbf{L}_n \mathbf{u}(x) d\Gamma & \textit{When } \mathbf{u}(x) \textit{ is continuous in } \Omega_k \\ \dfrac{1}{A_k}\int_{\Gamma_k} \mathbf{L}_n \mathbf{u}(x) d\Gamma & \textit{When } \mathbf{u}(x) \textit{ is discontinuous in } \Omega_k \end{cases} \tag{20}$$

where $\tilde{\varepsilon}$ is the compatible strain, $\hat{\varepsilon}_k$ is the smoothed strain over the smoothing domain($\Omega_k$),

$A_k$ is the area and $\Gamma_k$ is the boundary of the smoothing domain $\Omega_k$.

To perform the generalized strain smoothing, the problem domain is first discretized using three-node triangular background cells, and then the stationary and nonoverlapping smoothing domains are constructed based on these triangles such that $\Omega = \Omega_1 \cup \cdots \cup \Omega_{N_S}$ and $\Omega_i \cap \Omega_j = \varnothing, i \neq j$ in which $N_S$ is the number of smoothing domains.

Under the framework of ES-PIM theory the, smoothing domains are constructed for the edges of triangular cells by connecting two ends of the edge to the centroids of two adjacent cells. As illustrated in Figure 5. Thus, the number of smoothing domains ($N_S$) equals the number of edges of triangles ($N_{edge}$).

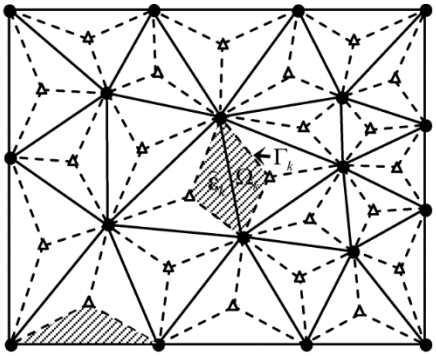



**Figure 5. Construction of edge-based strain smoothing domains, which are stationary, nonoverlapping and constructed based on the three-node triangular cells.**

Substituting Eq. (8) into Eq. (20), the edge-based smoothed strain $\hat{\varepsilon}_k$ can be written in the following matrix form of nodal displacements:

$$\hat{\varepsilon}_k = \frac{1}{A_k}\int_{\Gamma_k} \mathbf{L}_n \mathbf{\Phi} \mathbf{d}_i d\Gamma = \sum_{i \in N_{\text{infl}}} \hat{\mathbf{B}}_i(\mathbf{x}_k)\mathbf{d}_i \tag{21}$$

where $\mathbf{\Phi}$ is the matrix of PIM shape functions and $N_{\text{infl}}$ is the number of field nodes involved in constructing the smoothed strain fields within $\Omega_k$. $\hat{\mathbf{B}}_i(\mathbf{x}_k)$ is termed the smoothed strain matrix, which can be expressed as:

$$\hat{\mathbf{B}}_i(\mathbf{x}_k) = \begin{bmatrix} \hat{\mathbf{b}}_{ix}(\mathbf{x}_k) & 0 \\ 0 & \hat{\mathbf{b}}_{iy}(\mathbf{x}_k) \\ \hat{\mathbf{b}}_{iy}(\mathbf{x}_k) & \hat{\mathbf{b}}_{ix}(\mathbf{x}_k) \end{bmatrix} \tag{22}$$

In the above equation, elements of the smoothed strain matrix are obtained as:

$$\hat{\mathbf{b}}_{il}(\mathbf{x}_k) = \frac{1}{A_k}\int_{\Gamma_k} \varphi_i(\mathbf{x}_k) n_l(\mathbf{x}_k) d\Gamma \qquad (l = x, y) \tag{23}$$

Using the Gauss integration scheme, the above integration can be further expressed as follows:

$$\hat{\mathbf{b}}_{il} = \frac{1}{A_k}\sum_{m=1}^{N_{seg}} \left[ \sum_{n=1}^{N_{Gauss}} w_n \varphi_i(\mathbf{x}_{mn}) n_i(\mathbf{x}_m) \right] \qquad (l = x, y) \tag{24}$$

where $N_{seg}$ is the number of segments of the boundary $\Gamma_k$ , $N_{Gauss}$ is the number of Gauss points which located in each segment on $\Gamma_k$ , and $w_n$ is the corresponding weight number of the Gauss integration scheme.

By using the PIM with different T schemes we can construct the displacement field. Then, substituting the assumed displacements and the smoothed strains which given by Eq.(20) into Eq.(19), a set of discretized algebraic system equations can be obtained in the matrix form:

$$\hat{\mathbf{K}}\mathbf{d} = \hat{\mathbf{f}} \tag{25}$$

$\hat{\mathbf{f}}$ is the force vector, which can be obtain as:

$$\hat{\mathbf{f}} = -\int_{\Omega} \mathbf{\Phi}^T \mathbf{b} d\Omega + \int_{\Gamma_t} \mathbf{\Phi}^T \mathbf{t}_{\Gamma} d\Gamma \tag{26}$$

and the stiffness matrix $\hat{\mathbf{K}}$ is assembled from the substiffness matrix for all the integration cells which are exactly the edge-based smoothing domains for the present method:

$$\hat{\mathbf{K}}_{ij} = \sum_{k=1}^{N_a} \hat{\mathbf{K}}_{ij(k)} \tag{27}$$

where $\hat{\mathbf{K}}_{ij}$ is the substiffness matrix associated with the integration cell $k$ (i.e. smoothing domain $\Omega_k$), which is computed using the smoothed strain matrix, as follows:

$$\hat{\mathbf{K}}_{ij(k)} = \int_{\Omega_k} \hat{\mathbf{B}}_i^T D \hat{\mathbf{B}}_j d\Omega \tag{28}$$

## 4. The Solution of widespread corrosion-damaged structure

In the service environment, corrosion damage of engineering structures are usually not just single point corrosion damage, instead, widespread corrosion is the main form of structural corrosion. The occurrence of widespread corrosion is often random, which makes the cellular automata method more suitable for simulation

### *4.1 Establishment of widespread corrosion-damaged structure model*

In order to simulate the formation process of widespread corrosion, we use cellular automata method combined with MATLAB software for programming. The morphological model of widespread corrosion-damaged structure of aluminum can be obtained by simulation program. As illustrated in Figure 6.

**Figure 6. Topography of widespread corrosion damage structures**

Where the black part in Fig.6 is the appearance of the corroded aluminum metal, with Young's modulus $E = 7.1\times10^{10}\,\text{Pa}$ and Poisson's ratio $\nu = 0.33$, the white part is the liquid with a certain concentration of corrosion and causes widespread corrosion defects on the surface of aluminum. There are two main factors influencing corrosion morphology, one is relative humidity, the other is ambient temperature.Figure.6 shows the widespread corrosion-damaged structure morphology which forms in relative humidity 70% and the temperature of 50℃.

The solution domain was determined based on the simulated widespread corrosion morphology, and the computational model was established by combining Fortran programming, as illustrated in Figure 7.

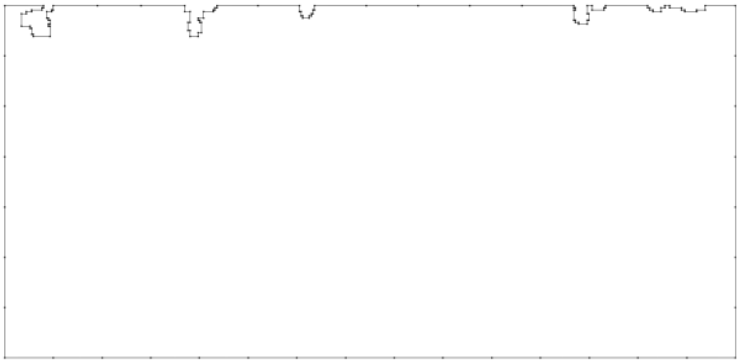


**Figure 7. Calculation model of widespread corrosion morphology**

Hinge constraints are applied to the left edge of the member containing widespread corrosion-damaged, and 15kN distributed load is applied to the right edge for stretching.

The problem domain and its boundary are modelled and represented by using 35920 nodes scattered in the problem domain and on its boundary. The density of the nodes depends on the accuracy required and resources available.

Near the upper edge of the corrosion morphological structure model, we use adaptive algorithms to improve the computational accuracy. As illustrated in Figure 8.

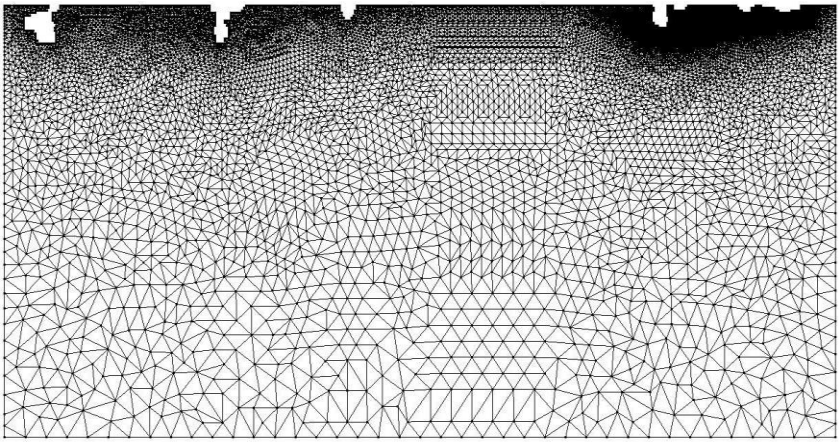


**Figure 8. Domain discretization of the widespread corrosion-damaged structure by using 35920 nodes and 70065 triangular meshes.**

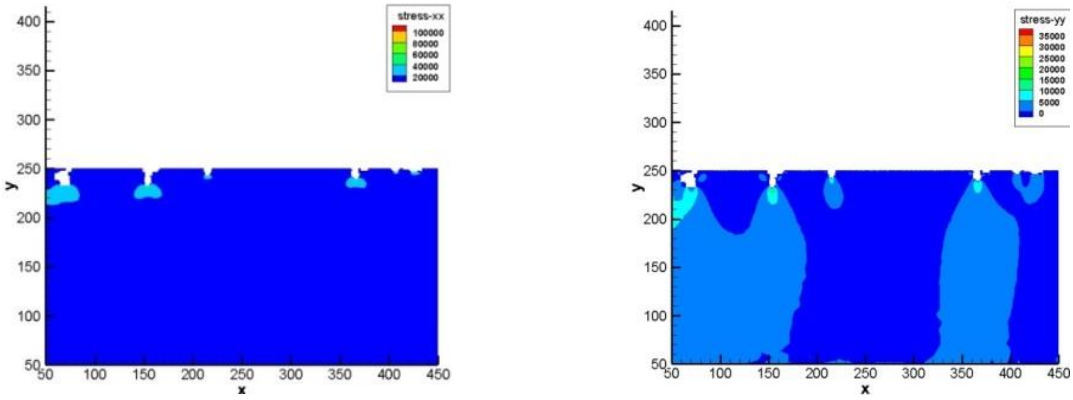
### *4.2 The solution of the widespread corrosion structure*

The results of stresses obtained using ES-PIM are listed in Tables 1.

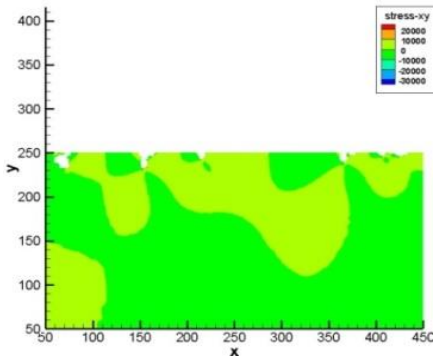
**Table 1. The output sample for stress obtained**

| No. of field nodes | $\sigma_{xx}$ | $\sigma_{yy}$ | $\tau_{xy}$ |
|---|---|---|---|
| 1 | 4.78E+03 | 1.90E+02 | -3.75E+02 |
| 2 | 4.80E+03 | -1.12E+03 | 2.02E+03 |
| 3 | 8.34E+03 | 1.15E+03 | -2.40E+03 |
| 4 | 5.24E+03 | -1.56E+02 | -2.18E+03 |
| 5 | 4.55E+03 | -1.13E+03 | -2.73E+03 |
| 6 | 4.37E+03 | -1.62E+03 | -2.90E+03 |
| 7 | 4.85E+03 | -1.54E+03 | -3.13E+03 |
| 8 | 5.93E+03 | -1.70E+03 | -3.72E+03 |
| 9 | 6.57E+03 | -1.87E+03 | -4.03E+03 |
| 10 | 7.78E+03 | -1.56E+03 | -4.45E+03 |
| ⋮ | | | |
| 35918 | 1.08E+04 | -1.20E+03 | 3.15E+03 |
| 35919 | 2.81E+04 | 4.76E+03 | -2.05E+03 |
| 35920 | 9.42E+03 | -7.85E+02 | 1.29E+03 |

The distribution of stresses in the domain obtained by ES-PIM are drawn by Tecplot shown in Figure 9. respectively.




(a) (b)




(c)

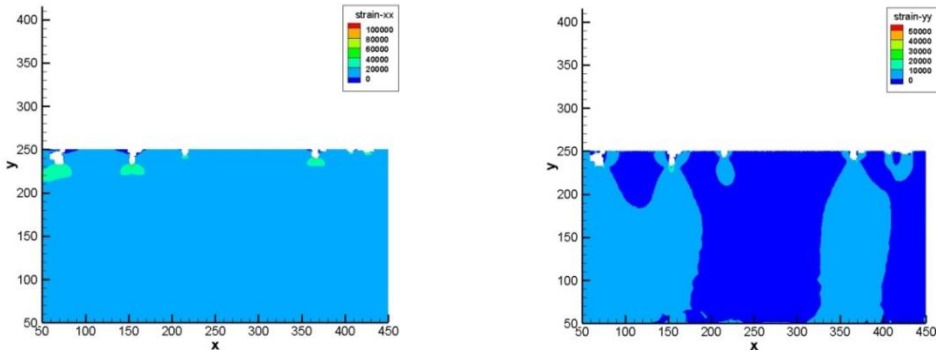
**Figure 9. (a), (b), (c) are stress nephogram of $\sigma_{xx}$,$\sigma_{yy}$ and $\tau_{xy}$ respectively.**

The results of displacements obtained using ES-PIM are listed in Tables 2.

**Table 2. The output sample for strain obtained**

| No. of field nodes | $u$ | $v$ |
|---|---|---|
| 1 | 4.75E+03 | 2.21E+02 |
| 2 | 4.00E+03 | -3.21E+02 |
| 3 | 7.42E+03 | 2.07E+03 |
| 4 | 4.13E+03 | 9.55E+02 |
| 5 | 2.51E+03 | 9.14E+02 |
| 6 | 2.11E+03 | 6.34E+02 |
| 7 | 2.29E+03 | 1.02E+03 |
| 8 | 2.94E+03 | 1.28E+03 |
| 9 | 3.61E+03 | 1.08E+03 |
| 10 | 4.53E+03 | 1.69E+03 |
| ⋮ | | |
| 35919 | 2.79E+04 | 4.94E+03 |
| 35920 | 9.25E+03 | -6.19E+02 |

The distribution of stresses in the domain obtained by ES-PIM are drawn by Tecplot shown in Figure 10. respectively.




(a) (b)

**Figure 10. (a), (b) are strain n ephogram of $u$ and $v$ respectively.**

## 5. Conclusions and discussions

In this work, an edge-based point interpolation method for widespread corrosion-damaged structure problems is formulated. And it provides a new way to solve the problem of widespread corrosion-damaged structure and lays a foundation for the research of integrated disruption of corrosion structure. The following conclusions can be drawn from the analysis:

The PIM shape functions used in the ES-PIM have the Kronecker delta function property. Thus, we can perform the straightforward imposition of point essential boundary conditions and no additional treatments are needed to apply continuity conditions along the interface to meet the interface condition. For the present ES-PIM models, the requirement for the nodes distribution and mesh generation along the interface is exactly the same as that for the FEM.

The influence range of stress concentration of corrosion pits with relatively close distance is wider than that of a single corrosion pit, therefore, the widespread corrosion damage structure is more likely to generate cracks in the stress concentration area and eventually lead to the failure of the structure.

**References**

[1] Engelhardt G, Macdonald D D, Urquidi-Macdonald M. Development of fast algorithms for estimating stress corrosion growth rate [J] Corrosion Science, 1999, 41(12):2267-2302.

[2] Engelhardt G, Macdonald D D. Deterministic prediction of pit depth distribution [J]. Corrosion, 1998,54(6):469-479

[3] Macdonald D D, Engelhardt G. Deterministic prediction of localized corrosion damage a reflective review of critical issues[J]. Journal of Corrosion Science and Engineering, 2003, 6:C066.

[4] LU P C, Urquidi-Macdonald M. Prediction of IGSCC in type 304 SS using artificial neural networks [C] Corrosion 1994. Houston, TX: NACE 1994:103.

[5] Wang Hui, Song Bifeng, Wang Le, Lv Guozhi Three-dimensional Computational Simulation of Corrosion Pit Growth Morphology. [J] Acta Aeronautica et Astronautica Sinica, 2009 30(11): 2186-2192.

[6] Wang Hui, Lv Guozhi, Zhang You Hong, Cellular automaton simulations of corrosion pit growth.Corrosion science and protection technology. 20(6) (2008):472-475.

[7] Bastien Chopard and Michel Droz, Cellular Automaton Modeling of Physical Systems, 1st edn. The Pitt Building Trumpington Street, Cambridge CB2 1RP, United Kingdom.

[8] G.R.Liu, M.B.Liu, Smoothed particle hydrodynamics: A meshfree particle method World Scientific Press.Singpore 2003.

[9] G.R.Liu, Meshfree method Moving Beyond the Finite Element Method, CRC Press, Boca Ratou, USA, 2000.

[10] G.R.Liu, Y.T.Gu, An introduction to meshfree methods and their programming, Springer, 2005.

[11] G.R.Liu, G.Y.Zhang, K.Y.Dai, Y.Y.Wang, Z.H.Zhong, G.Y.Li, X.Han. A linearly conforming point interpolation method (LC-PIM) for 2D solid mechanics problems, Int J Computat Methods, 2005, 2(4):645-655.

[12] G.R.Liu, Y.Li, K.Y.Dai, Z.H.Zhong, G.Y.Yi, X.Han, A linearly conforming RPIM for solid mechanics problems Int J Computat Methods, 2006, 3(4):401-428.

[13] G.R.Liu, A generalized gradient smoothing technique and the smoothing bilinear form for Galerkin

formulation of a wide class of computational methods, Int J Computat Meth 2008a, 5(2):199-236.

[14] G.R.Liu, Z.Wang, G.Y.Zhang, Z.Zong, S.Wang, An Edge-based smoothed point interpolation method for material discontinuity, Mech Adv Mater Struc, 2012, 19:3-17.

[15] G.Y.Zhang, G.R.Liu, T.T.Nguyen, C.X.Song, The upper bound property for solid mechanics of the linearly conforming radial point interpolation method (LC-RPIM), International Journal of Compu, 2007,4(3):521-541.

[16] G.Y.Zhang, G.R.Liu, Y.Y.Wang, H.T.Huang, Z.H.Zhong, G.Y.Li, X.Han, A linearly conforming point interpolation method(LC-PIM) for three-dimensional elasticity problems,International Journal of Co mpu,2005,72(13):1524-1543.

[17] G.R.Liu, T.T.Nguyen, K.Y.Lam, An edge-based smoothed finite element method(ES-FEM) for static and dynamic problems of solid mechanics, J.Sound.Vib.,DOI:10.1016/j.jsv.2008.08.027.

[18] G.R.Liu, A G space theory and a weakend weak($W^2$) form for a unified formulation of compatible incompatible methods: Part I Theory,Int.J.Numer.Meth.Eng.2010, 81:1093-1126.

[19] G.R.Liu, A G space theory and a weakend weak($W^2$) form for a unified formulation of compatibl and incompatible methods: Part II applications to solid mechanics problems.Int.J.Numer.Meth.Eng. 2010, 81:1127-1156.

[20] G.R.Liu A weakened weak ($W^2$) form for a unified formulation of compatible and incompatible displacement methods, Int. J. Numer. Meth. Eng. (revised).

# Fluid–rigid Structure Interaction of Moving Cuttlefish Robot Using Moving Mesh Method

**†*Shinichi ASAO[1], Masashi YAMAKAWA[2], and Yuto MORI[2]**

[1]Department of Mechanical Engineering, College of Industrial Technology, Japan
[2]Department of Mechanical Engineering, Kyoto Institute of Technology, Japan

*Presenting author: asao@cit.sangitan.ac.jp
†Corresponding author: asao@cit.sangitan.ac.jp

## Abstract

Cuttlefish are known to have superior characteristics for attitude control by using two fins. The objective of this study is to clarify the characteristics of cuttlefish-like motion for an optimal attitude control. To analyze both flows around a cuttlefish and its motion depending on fluid force, coupled simulation of the moving computational domain method and motion dynamics were applied. Furthermore, moving mesh techniques combined with a torsional spring approach and an angle-based-smoothing enabled the calculation of various fin motions. As a first step, it is confirmed whether fluid–rigid structure interaction of moving cuttlefish robot is correct. As a result of comparison with the movement of the experiment robot, the movement was almost in agreement. Therefore, this calculation method is effective. In addition, it was found that the influence of the movement of cuttlefish fins on the flow field was small.

**Keywords:** Cuttlefish, moving boundary problem, moving mesh method

## Introduction

Cuttlefish swim by undulating the fins on the sides of their bodies. This affects the surrounding flow and enables fine movement, resulting in excellent attitude control. By incorporating this movement into a fish-type robot, the robot can move underwater without raising mud on the sea floor or getting entangled in seaweed. The robot can also achieve fine attitude control when working with its arms.

The way cuttlefish swim is currently being studied. Kier et al. [1] suggested that in their fin movements, cuttlefish swim in almost the same way as do fish. Rahman et al. [2] are developing a robot that imitates the side fins of cuttlefish. They have indicated that propulsion and rotation can be performed by fin undulation. However, little has been reported on attitude control. Enabling a robot to achieve attitude control will make it possible to smoothly perform underwater work and ecological surveys even if disturbances such as tidal currents and waves largely affect the obtained results. It is therefore important to conduct research on the attitude control mechanism. Although experiments have been done using a robot, they are costly and difficult to implement. Therefore, studies based on numerical fluid dynamics are effective. On the other hand, as a simulation method, it is necessary to consider the interaction of moving fins and fluid flow around the fins and body. Therefore, we performed a coupled simulation of fluid dynamics and dynamics of structure.

In the work we report in this paper, we clarified fin movement as a means to achieve attitude control and performed coupled simulation of numerical fluid dynamics and structure

dynamics. As a first step, we determined whether the fluid–rigid structure interaction of a moving cuttlefish robot is correct.

## NUMERICAL APPROACH

*Governing equations*

The governing equations are the continuity equation and the incompressible Navier–Stokes equation. These are written as

$$\nabla \cdot \boldsymbol{q} = 0, \tag{1}$$

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{\partial \boldsymbol{E}_a}{\partial x} + \frac{\partial \boldsymbol{F}_a}{\partial y} + \frac{\partial \boldsymbol{G}_a}{\partial z} = -\left( \frac{\partial \boldsymbol{E}_p}{\partial x} + \frac{\partial \boldsymbol{F}_p}{\partial y} + \frac{\partial \boldsymbol{G}_p}{\partial z} \right) + \frac{1}{\mathrm{Re}} \left( \frac{\partial \boldsymbol{E}_v}{\partial x} + \frac{\partial \boldsymbol{F}_v}{\partial y} + \frac{\partial \boldsymbol{G}_v}{\partial z} \right), \tag{2}$$

where $\boldsymbol{q}$ is the velocity vector, $\boldsymbol{E}_a$, $\boldsymbol{F}_a$, and $\boldsymbol{G}_a$ are respectively advection flux vectors in the $x$, $y$, and $z$ directions, $\boldsymbol{E}_v$, $\boldsymbol{F}_v$, and $\boldsymbol{G}_v$ are viscous-flux vectors, and $\boldsymbol{E}_p$, $\boldsymbol{F}_p$, and $\boldsymbol{G}_p$ are pressure terms. The elements of the velocity vector and flux vectors are

$$\begin{gathered}
\boldsymbol{q} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \boldsymbol{E}_a = \begin{bmatrix} u^2 \\ uv \\ uw \end{bmatrix}, \boldsymbol{F}_a = \begin{bmatrix} uv \\ v^2 \\ vw \end{bmatrix}, \boldsymbol{G}_a = \begin{bmatrix} uw \\ vw \\ w^2 \end{bmatrix}, \\
\boldsymbol{E}_p = \begin{bmatrix} p \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{F}_p = \begin{bmatrix} 0 \\ p \\ 0 \end{bmatrix}, \boldsymbol{G}_p = \begin{bmatrix} 0 \\ 0 \\ p \end{bmatrix}, \\
\boldsymbol{E}_v = \begin{bmatrix} u_x \\ v_x \\ w_x \end{bmatrix}, \boldsymbol{F}_v = \begin{bmatrix} u_y \\ v_y \\ w_y \end{bmatrix}, \boldsymbol{G}_v = \begin{bmatrix} u_z \\ v_z \\ w_z \end{bmatrix}
\end{gathered} \tag{3}$$

where $u$, $v$, and $w$ are respectively the velocity components of the $x$, $y$, and $z$ directions and $p$ is pressure. The $x$, $y$, and $z$ subscripts respectively indicate derivatives derived from $x$, $y$, and $z$ and Re is the Reynolds number. We also took into account the combined translation and rotation motions of a cuttlefish. The rigid body equations of motion are

$$\frac{d\boldsymbol{p}_B}{dt} = \boldsymbol{f}_B \tag{4}$$

$$\frac{d\boldsymbol{L}_B}{dt} + \omega_B \times \boldsymbol{L}_B = \boldsymbol{N}_B \tag{5}$$

Here, $\boldsymbol{p}_\mathrm{B}$ is the momentum vector of the body, $\boldsymbol{f}_\mathrm{B}$ is the external force vector, $\boldsymbol{L}_\mathrm{B}$ is the angular momentum vector, $\boldsymbol{N}_\mathrm{B}$ is the torque vector, and $\omega_B$ is the angular velocity.

*Moving-grid finite-volume method*

To ensure the geometric conservation laws are followed, we used a control volume in the space-time unified domain ($x$, $y$, $z$, $t$), which is four-dimensional in the case of three-dimensional flows. This enables Eq. (2) to be written in divergence form as

$$\tilde{\nabla} \cdot \tilde{\boldsymbol{F}} = \boldsymbol{0}, \tag{6}$$

where

$$\tilde{\nabla} = \begin{bmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \\ \dfrac{\partial}{\partial t} \end{bmatrix}, \quad \tilde{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{E}_a + \boldsymbol{E}_p - \dfrac{1}{\mathrm{Re}} \boldsymbol{E}_v \\ \boldsymbol{F}_a + \boldsymbol{F}_p - \dfrac{1}{\mathrm{Re}} \boldsymbol{F}_v \\ \boldsymbol{G}_a + \boldsymbol{G}_p - \dfrac{1}{\mathrm{Re}} \boldsymbol{G}_v \\ \boldsymbol{q} \end{bmatrix}. \tag{7}$$

The present method is based on a cell-centered finite-volume method. Thus, the flow variables are defined at the center of the cell in the (*x*, *y*, *z*) space. The control volume becomes a four-dimensional polyhedron in the (*x*, *y*, *z*, *t*)-domain, as schematically illustrated in Figure 1.

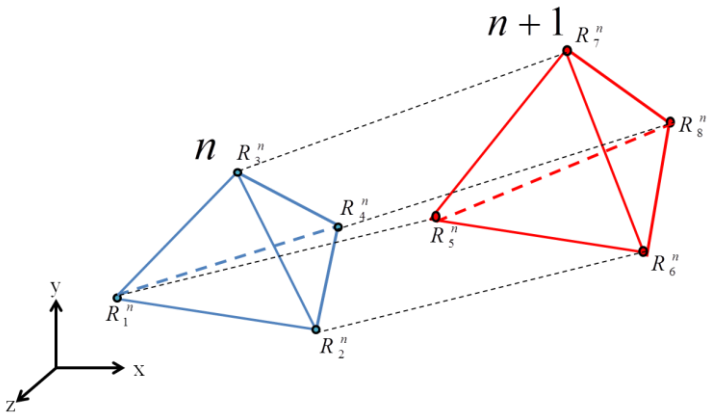



Figure 1 Schematic view of control volume $\tilde{\Omega}$ in (*x*, *y*, *z*, *t*) space-time unified domain.

We apply volume integration to Eq. (6) with respect to the control volume illustrated in Figure 1. Using the Gauss theorem, we can write Eq. (6) in surface integral form as

$$\int_{\tilde{\Omega}} \tilde{\nabla} \cdot \tilde{\boldsymbol{F}} \mathrm{d}\tilde{V} = \oint_{\partial\tilde{\Omega}} \tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{n}}_u \mathrm{d}\tilde{S} \approx \sum_{l=1}^{6} \left( \tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{n}} \right)_l = \boldsymbol{0} \tag{8}$$

Here, $\tilde{\boldsymbol{n}}_u$ is an outward unit vector normal to the surface of the polyhedron control volume $\tilde{\Omega}$ ($\partial\tilde{\Omega}$). The term $\tilde{\boldsymbol{n}} = (\tilde{n}_x, \tilde{n}_y, \tilde{n}_z, \tilde{n}_t)_l$, ($l$=1, 2,... 6) denotes the surface normal vector of control volume and its length is equal to the boundary surface area in four-dimensional (*x*, *y*, *z*, *t*) space. The upper and bottom boundaries of the control volume ($l$ = 5 and 6) are perpendicular to the *t*-axis. Therefore they have only the $\tilde{n}_t$ component, the length of which corresponds respectively to the cell volume in the (*x*, *y*, *z*)-space at time $t^n$ and $t^{n+1}$. Thus, Eq. (8) can be expressed as

$$\boldsymbol{q}^{n+1}(\tilde{n}_t)_6 + \boldsymbol{q}^{n}(\tilde{n}_t)_5 + \sum_{l=1}^{4}\left(\tilde{\boldsymbol{F}}\cdot\tilde{\boldsymbol{n}}\right)_l^{n+1/2} = \boldsymbol{0}. \tag{9}$$

*Moving computational domain method*

The basic coordinate system of the moving computational domain method is the general, fixed, stationary (*x*, *y*, *z*) Cartesian coordinate system. The computational domain itself, including the body inside, moves in the fixed (*x*, *y*, *z*)-space. The flow around the body is calculated as the moving boundary problem. Unknown flow variables, such as pressure *p* and *x*-directional velocity *u*, are defined at each grid cell center in the computational domain. The motion of the computational domain in accordance with the body motion y in the physical space is arbitrary. Accordingly, any kind of body motion can be simulated by the moving computational domain method. The flow field driven by the body is calculated in the computational domain in which the body fitted mesh system is used. The computational domain itself moves in the physical (*x*, *y*, *z*) space time-dependently. Thus, since the mesh system of the computational domain also moves in the (*x*, *y*, *z*) space, a flow solver has to be constructed for the moving grid system. In the present moving computational domain method, the moving-grid finite-volume method [3] is used. The only necessary and essential assumption is that the conditions in front of the moving computational domain have to be known because they are necessary as a boundary condition of the flow solver. The natural assumption may be the stationary fluid condition in front of the moving computational domain.

*Spring approach for moving grid*

Cuttlefish fins change their angle from 0 to 45 degrees. To express such fin movements, it is necessary to use a large movement defined by a computational grid. Thus, we used the spring approach [4], which adds a torsion spring effect to the conventional spring method. In this approach, to obtain robustness in the computation, we added a spring constant associated with the shape of the cell (cells sides' angles), as shown in Equation (10).

$$k_{ij} = k_{[spring]ij} + k_{[angle]ij} \tag{10}$$

In our work, we specifically used the following equations:

$$k_{[spring]ij} = \frac{1}{l_{ij}^2} \tag{11}$$

$$k_{[angle]ij} = \sum \frac{1}{\sin^2\theta_{in}} + \sum \frac{1}{\sin^4\theta_{dn}}, \tag{12}$$

where $l_{ij}$ shows the side length of the cell. Then, $\theta_{in}$ and $\theta_{di}$ are angles defined by two torsion sides of a tetrahedral cell, as shown in Figure 2.

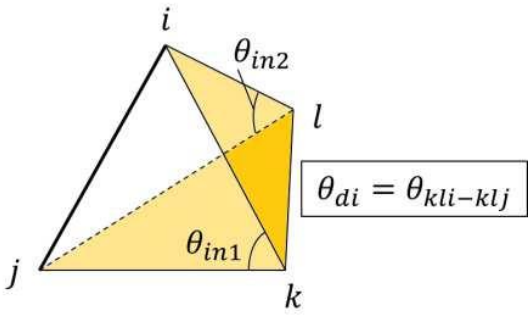



Figure 2 Spring constant for tetrahedron.

*Numerical procedure*

To solve Eq. (9), we apply the SMAC method [5]. Thus, Eq. (9) can be solved in the three following stages. The equation to be solved at the first stage contains unknown variables at $n$+1-time step in flux terms. Thus, the equation is iteratively solved using the LU-SGS method [6]. The equation to be solved at the second stage is the Poisson equation about the pressure correction. This equation is iteratively solved using the Bi-CGSTAB method [7]. The flux vectors are evaluated using the QUICK method, whereas the flux vectors of the pressure and viscous terms are evaluated in a central-difference-like manner. The incompressible fluid-body interaction is calculated in the first step of the SMAC method. Figure 3 shows the flowchart of the fluid-body interaction.

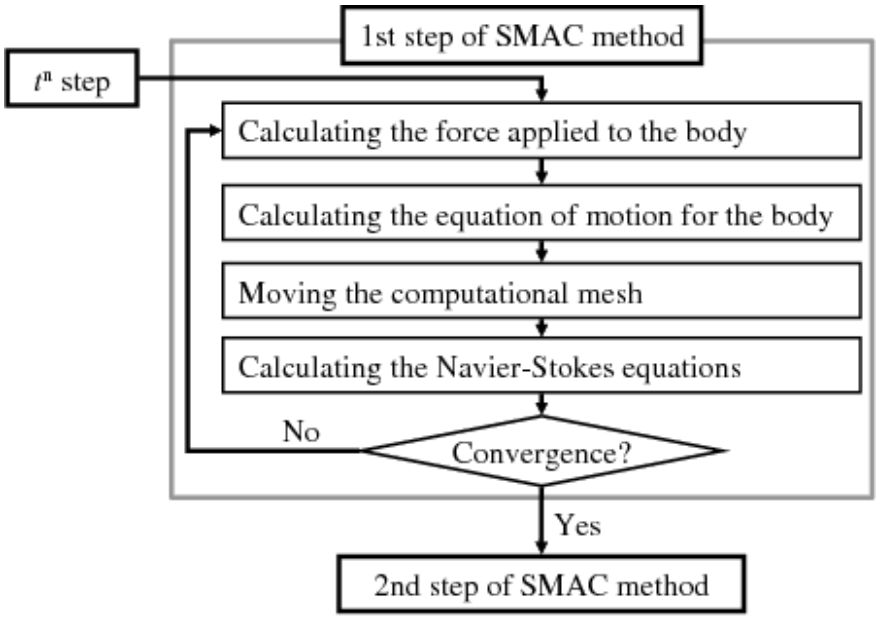



Figure 3 Flowchart for interaction calculation.

## NUMERICAL RESULTS

To verify the efficacy of the fluid–rigid structure interaction method, we used a cuttlefish robot to compare simulation results with the experiment results [2]. In the experiment, we

investigated circular movement due to the frequency differences between the left and right fins.

*Calculation model*

The experiment robot used in this calculation is shown in Figure 4. The robot has a total length of 1.3 m, a maximum width of 0.714 m, a thickness of 0.1 m, a fins length of 0.874 m, a fillet width of 0.075 m and a weight of 62.8 kg.

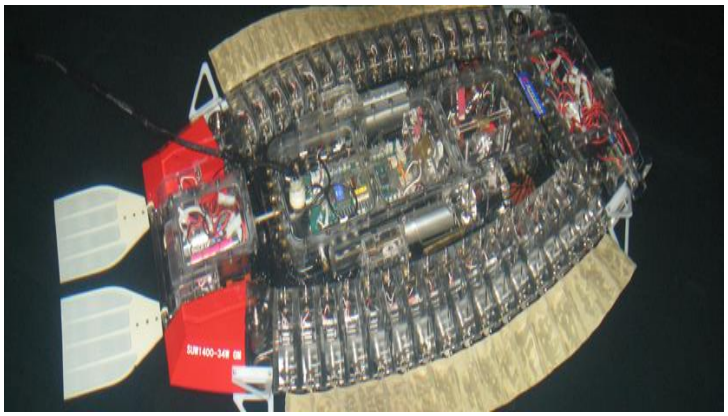


Figure 4 Cuttlefish robot [2].

The calculation model is shown in Figure 5. On a dimensionless basis, its total length $L$ amounts to the total length of the experiment robot. Figure 6 shows the model's surface mesh and Figure 7 shows the computational mesh around the model. We generated the computational mesh using MEGG3D supported by JAXA [8]. The calculation region is a sphere with diameter $30L$. The number of cells is 2,875,222. The coordinate system defines the $x$ axis in the length direction and the $y$ axis in the width direction, and the direction perpendicular to them is the $z$ axis.

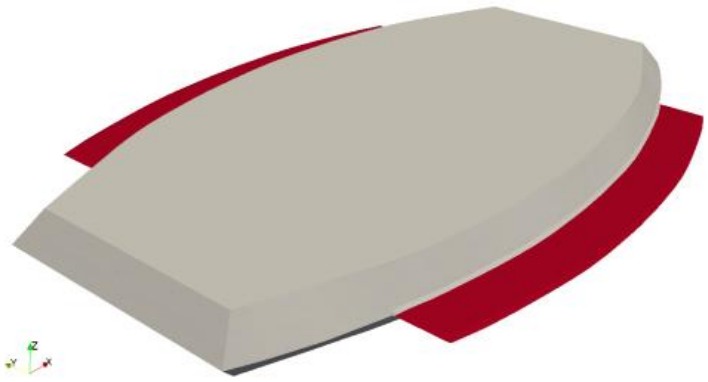


Figure 5 Cuttlefish robot model.

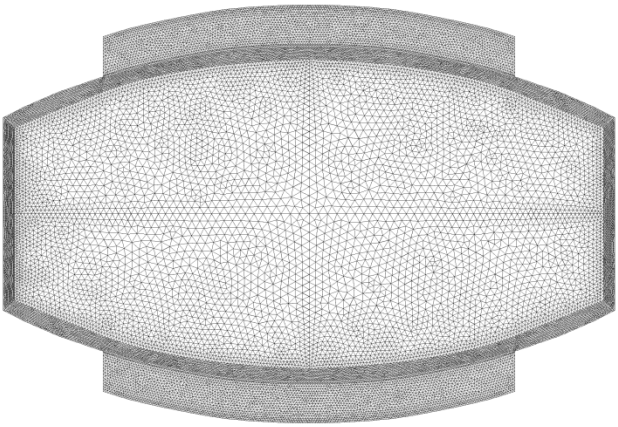
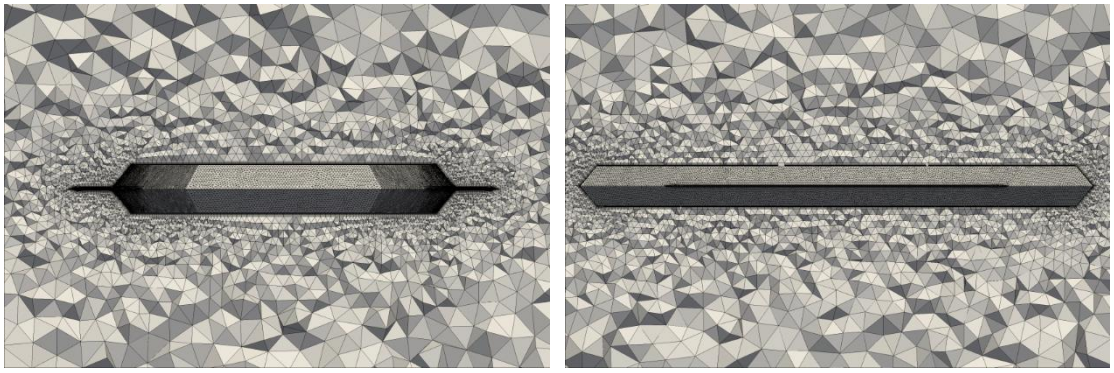


Figure 6 Cuttlefish robot surface mesh.



(a) *y*-*z* plane (b) *x*-*z* plane

Figure 7 Computational mesh around cuttlefish robot.

*Fin movements*

The cuttlefish fins move in a traveling wave motion and the fin deformation is represented by the following equations:

$$\theta = \Theta(s)\sin(2\pi Ks - 2\pi Nt), \tag{13}$$

$$\Theta(s) = \sin^{-1}\left[\left\{1 - 0.905\left(\frac{s}{fL} - 0.5\right)^2\right\}\sin\theta_{\max}\right], \tag{14}$$

$$y(t) = r\cos\theta, \tag{15}$$

$$z(t) = r\sin\theta. \tag{16}$$

In these equations, $\theta$ is the angle from the base of the fin to its tip, $K$ is wave number, $N$ is frequency, $fL$ is the fin length in the $x$ direction, $s$ is the length in the $x$ direction from the tip of the fillet, $\theta_{\max}$ is the maximum angle, and $r$ is the length from the base of the fins to the tip. In accordance with the experiment conditions, the calculation condition was set to 1 for the wave number and 45° for the maximum angle. The direction of traveling wave motion is the-$x$ direction. The left and right fins are defined as shown in Figure 8. The left fin frequency is 0.5 Hz and the right fin frequency is 1.0 Hz.

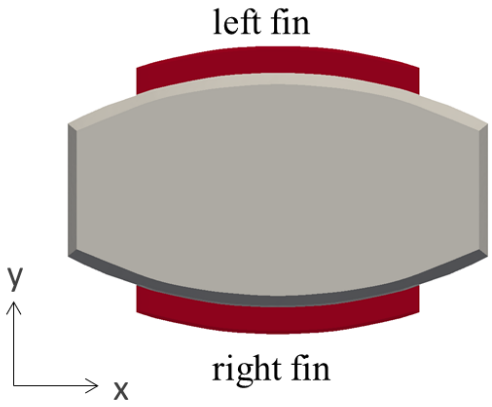
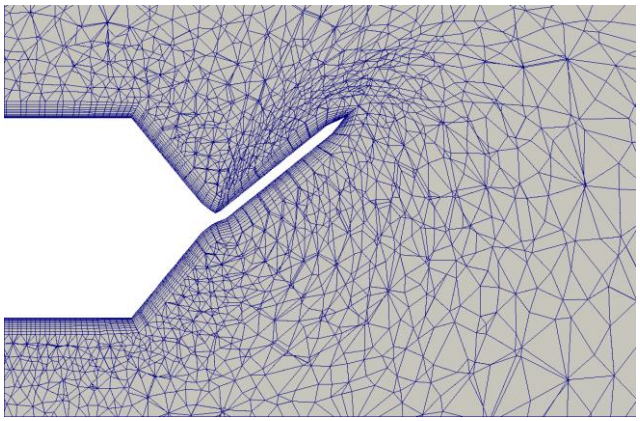


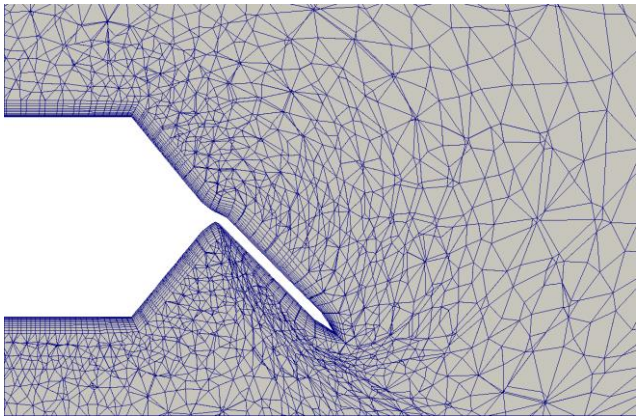

Figure 8 Definition of robot's left and right fins.

*Movement and deformation of computational mesh*

Figure 9 shows the movement and deformation of the space mesh. This figure shows that the fin movement is represented by the deformation of the computational mesh and the quality of the cell shape is maintained.



(a) Fin angle: 45 degrees



(b) Fin angle: -45 degrees

Figure 9 Mesh deformation for fin motion at *y*-*z* plane.

*Calculation conditions*

As computational conditions, the characteristic length is 1.3 m, the characteristic velocity is 0.5 m/s, the kinematic viscosity of water is $1.004\times10^{-6}$ $m^2/s$, and the Reynolds number is 647,000. The time step is 0.0001. It is assumed that the model moves in static fluid. The initial stationary condition of pressure and the velocity components in the *x*, *y*, and *z* directions are given by p = 0.0 and $u = v = w = 0.0$. The model movements are obtained by coupled analysis of fin movement and fluid dynamics.

*Trajectory of moving cuttlefish robot*

Figure 10 shows the movement trajectory of the side fin type robot up to $T = 51.4$ s. This figure shows the trajectory from the upper side, and the one that is not transparent is the position at $T = 51.4$ s. The robot gradually rotates in the acceleration from the stationary state and performs the circular movement. The obtained results confirmed that the simulation reproduces the movement in which the speed and the attitude angle change moment by moment.

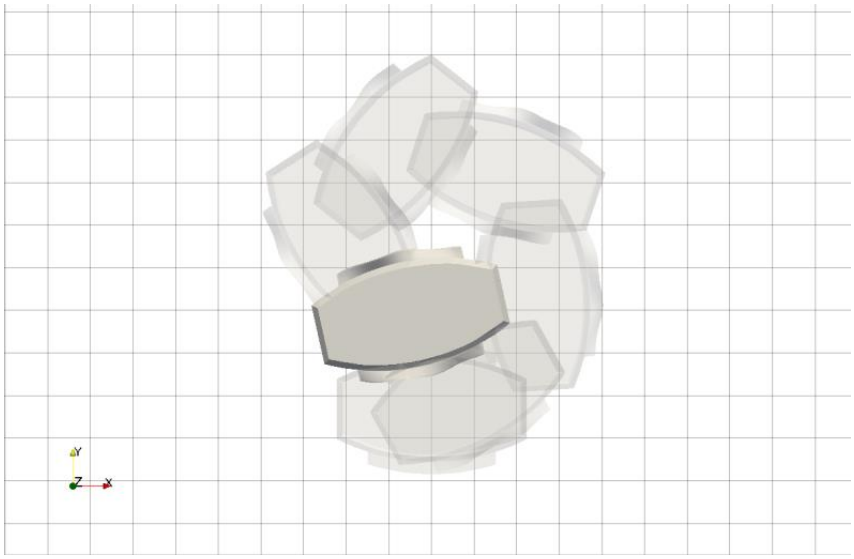


Figure 10 Trajectory of cuttlefish robot in circular motion.

We used the calculation results to compare the gravity position center coordinates of the robot obtained in the experiment Rahman et al. reported [2]. Figure 11 shows a comparison of the trajectory of the center of gravity position in the *x*-*y* plane. Red lines represent simulation results and blue lines represent experimental results. Figure 11 shows that both trajectories have the same circular motion. In comparing the circular motion diameters, we found the experimental result was 1.40 m, the simulation result was 1.31 m, and the error was 6.8%. The simulation result and the experimental result basically match. This confirms the effectiveness of the present calculation method of fluid-rigid structure interaction and the validity of the calculation result.

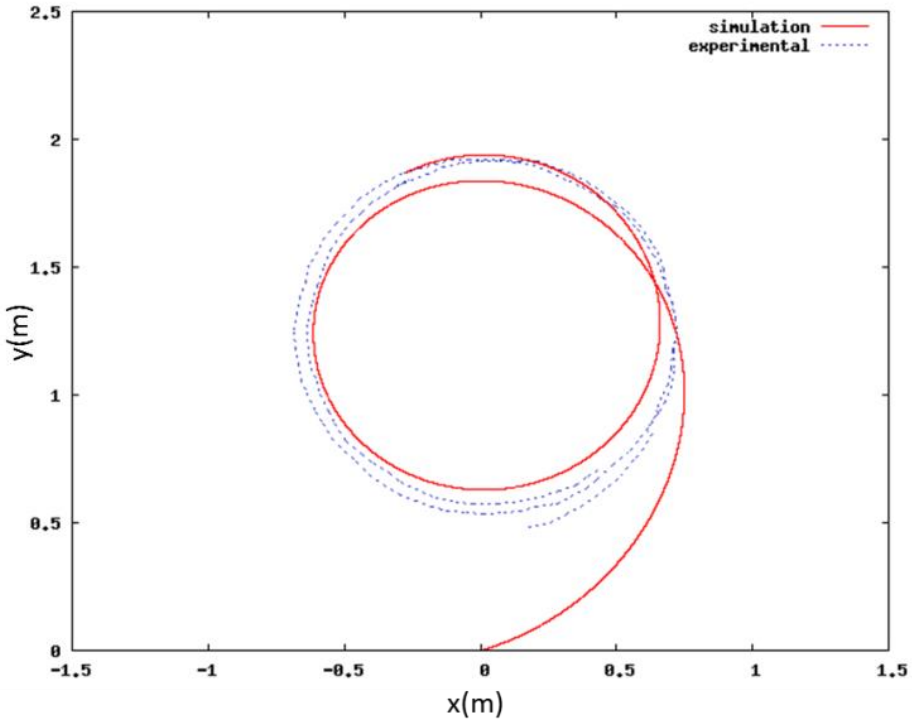
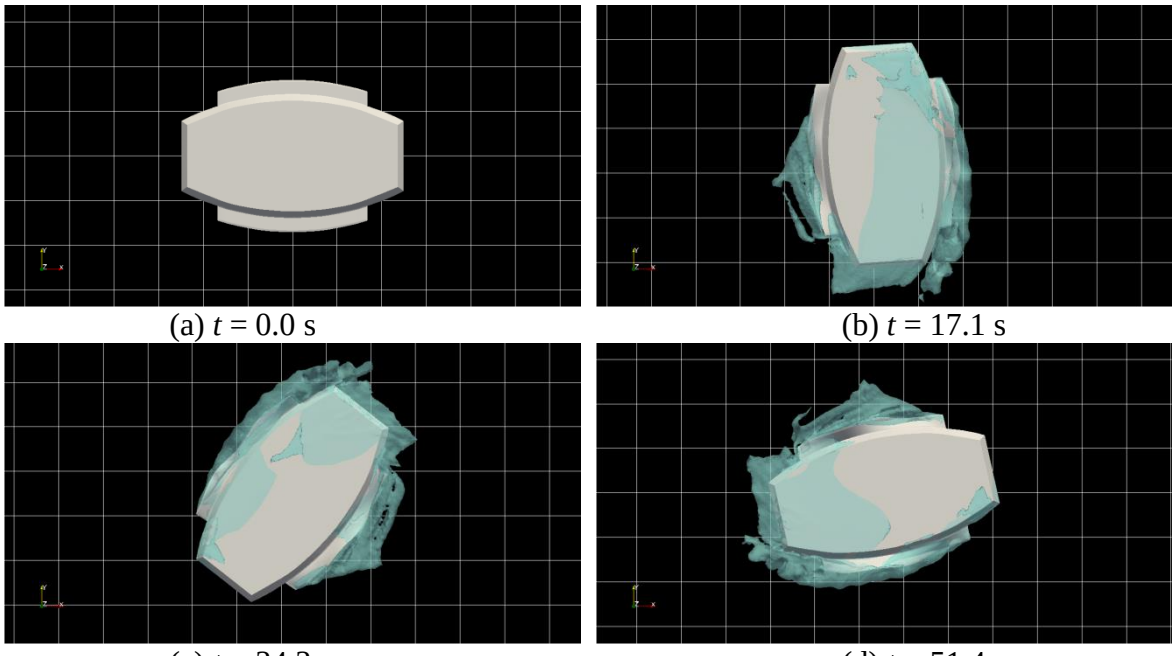



Figure 11 Center of gravity trajectory.

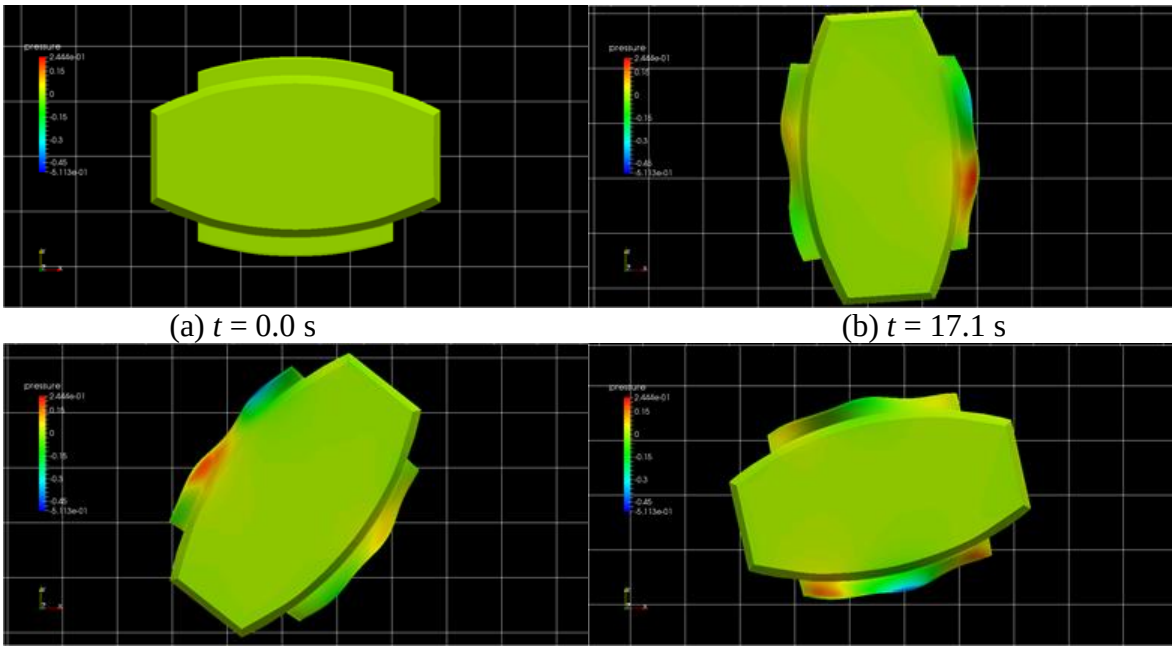
The flow field calculation results we obtained were a 0.1 m/s velocity isosurface around the robot at $t$ = 0 s, 17.1 s, 34.3 s, and 51.4 s (Figure 12). We consider that the velocity isosurface exist only around the fins, which little affects the surrounding flow. Figure 13 shows the pressure distribution on the cuttlefish robot surface. The obtained results confirm that in frequency terms the pressure changes more in the right fins than in the left fins, and that thus a difference in thrust force exists between the left and right fins.



(a) $t$ = 0.0 s (b) $t$ = 17.1 s

(c) $t$ = 34.3 s (d) $t$ = 51.4 s

Figure 12 Velocity isosurface at $V$ = 0.1 m/s



(a) $t$ = 0.0 s (b) $t$ = 17.1 s

(c) $t$ = 34.3 s (d) $t$ = 51.4 s

Figure 13 Pressure distribution on the cuttlefish robot surface.

Figure 14 shows the time history of the forces applied to the whole object, i.e. the right and left fins in the $x$ direction. These $x$ directions refer to the object fixed coordinate system. Figure 14 shows that a propulsion force applied to the right fin is larger than that applied to the left fin. The difference in the thrust force applied to the left and right fins makes a circular motion.

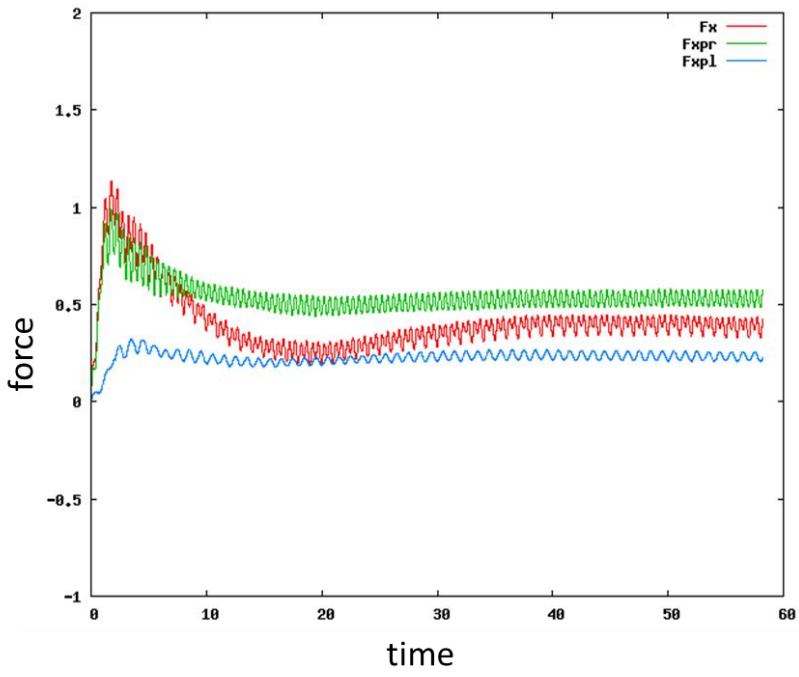



Figure 14 Time history of force applied to the robot.

## CONCLUDING REMARKS

We confirmed whether fluid–rigid structure interaction of a moving cuttlefish robot was correct for attitude control. In comparing experiment results with calculation results, we found that they agreed well. We ascertained that the calculation method results obtained with our method agreed well with simulation results and showed our method's efficacy. We also confirmed that the fin movement did not affect the flow field very significantly. We also found that the difference in thrust force on the left and right fins makes a circular motion.

**Acknowledgements**

This publication was subsidized by JKA through its promotion funds from KEIRIN RACE, and by JFE 21st Century Foundation.

**References**

[1] S. Johnsen and W. M. Kier, Intramuscular crossed connective tissue fibres: skeletal support in the lateral fins of squids and cuttlefish, Department of Biology, Land, 231, pp. 311–338, DOI: 10.1111/j.1469-7998.1993.tb01921.x (1993).

[2] M. M. Rahman, H. Miki, S. Sugimori, Y. Sanada, and Y. Toda, Development of a Real Time Simulator Based on the Analysis of 6-Degrees of Freedom Motion of a Biomimetic Robot with Two Undulating Side Fins. Journal of Aero Aqua Bio-mechanisms. 3 (1), pp. 71–78, DOI: 10.5226/jabmech.3.71 (2013).

[3] M. Yamakawa and K. Matsuno, An Iterative Finite-Volume Method on an Unstructured Moving Grid: 1st Report, The Fundamental Formulation and Validation for Unsteady Compressible Flows, JSME, Series B, Vol. 69-683, pp. 1577–1582, DOI: 10.1299/kikaib.69.1577 (2003).

[4] G. A. Markou, Z. S. Mouroutis, D. C. Charmpis, and M. Papadrakakis, The ortho-semi-torsional (OST) spring analogy method for 3D mesh moving boundary problems, ScienceDirect, 196, pp. 747–765, DOI: 10.1016/j.cma.2006.04.009 (2007).

[5] A. Amsden and F. Harlow, "A Simplified MAC Technique for Incompressible Fluid Flow Calculations," Journal of Computational Physics, Vol. 6, pp. 322–325 (1970).

[6] S. Yoon and A. Jameson, Lower-Upper Symmetric-Gauss-Seidel Method for the Euler and Navier-Stokes Equations, AIAA, Vol. 26, pp. 1025–1026 (1988).

[7] H. A. van der Vorst, "Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems," SIAM Journal on Scientific Computing, Vol. 13-2, pp. 631–644 (1992).

[8] Y. Ito and K. Nakahashi, Surface Triangulation for Polygonal Models Based on CAD Data, Internal Journal for Numerical Methods in Fluids, Vol. 39, Issue 1, pp. 75–96 (2002).

# Wrinkling analysis of pre-stressed membranes using element free Galerkin method

K R Unnikrishnan[1], C O Arun[2], and I R Praveen Krishna[3]

[1]Tata Consultancy Services Pune
[2]Indian Institute of Space Science and Technology

## ABSTRACT

In the current study, element free Galerkin method, a meshless method, is proposed for wrinkling analysis of pre-stressed membranes. The mathematical model for studying wrinkling of pre-stressed membranes is derived by considering the bending stiffness, though it is negligible. Moving least approximation for deflection is constructed by considering three degrees of freedom per node. Essential boundary conditions are imposed using scaled transformation matrix method. Initially, compression induced wrinkling of a homogeneous thin plate without pre-stress is solved to validate the method and then a pre-stressed homogeneous membrane is analyzed for both compression induced and shear induced wrinkling. Capabilities of proposed method for membrane analysis is compared with that of finite element method(FEM). Comparative study on wrinkling analysis using EFGM and different FEM element types in a FEM package shows that, in lower modes both methods shows satisfying consistency in eigenvalues with respect to total of number of nodes, while at higher modes EFGM shows better consistency than FEM. Further the study is extended to wrinkling of non-homogeneous membranes subjected to linearly varying in-plane load. The results obtained from EFGM analysis is compared and found to be matching well with those available in literature.

## INTRODUCTION

Thin membranes are largely used in the field of aerospace engineering applications such as solar collectors, parachutes, sun-shield, antennae, balloons, space radars, solar sails [1, 2] etc, because of its light weight and low space requirements. In most of these applications, the membrane will be kept folded and deployed to the required shape at the time of application. This process may produce wrinkles in the membrane as it has very small bending rigidity. Also, due to its lower bending stiffness, it cannot support compressive and bending loads. Most of the application needs highly smooth surface without wrinkles, in order to have maximum efficiency. This is usually achieved by giving a pre-stressing in the plane of the sheet. However, though these are pre-stressed, chances of wrinkling due to various loads acting on the membranes still exists.

A large number of studies are available in the area of membrane analysis [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Tension field theory [3] and bifurcation analysis [15] are two early developed approaches used in wrinkling analysis. Wagner[3] proposed the tension field theory for the wrinkling problem and estimated the maximum shear load that can be carried by a membrane. Wong and Pellegrino[10, 11, 12], conducted experimental and analytical studies on shear induced wrinkling of rectangular membrane and verified it numerically using FEM. Also they have extended their research into wrinkling due to corner loads on a square plate.

Miyamura[16] studied wrinkling of pre-stressed circular membrane due to in-plane torsional load. Stress in the membrane are found from experiments and are compared with the results obtained from bifurcation analysis. Xiao et al.[13] analyzed wrinkles on a square planar thin film under pure shear and validated using experimental results. The influence of shear force, pre-stress and boundary conditions were also investigated. Kumar et al.[14] studied wrinkling of a membrane due to tensile and shear loading using commercial finite element package ABAQUS. The work was concentrated on the variation in eigenvalues and number of wrinkles with thickness and aspect ratios for different materials. Leissa et al.[17] found exact solution for buckling and vibration of thin plate, having two opposite edges simply supported and other two edges clamped, subjected to linearly varying in-plane load, using power series method. They extended the studies to other boundary conditions as well [18]. Wang et al.[19] used differential quadrature (DQ) technique to solve the plate buckling problems studied by Leissa et al.[17]. Lal and Saini[20] studied buckling and vibration analysis of non-homogeneous thin plates subjected to linearly varying in-plane loads using DQ method. The variation of critical load with various parameters like $a/b$ ratio, non homogeneity parameter, loading parameter etc were studied. Though there exist a number of analytical, empirical and numerical studies, currently the most common practice in structural analysis is the use of FEM with the help of any commercially available packages. FEM, due to its simplicity and robustness, found its way into both industrial and academic fields. Also there are works on wrinkling analysis of membranes using FEM [12, 21]. However, use of non-conforming shell/plate elements available in commercial FEM softwares can lead to erroneous eigenvalues [22, 23]. Very small thickness of membranes also play a role in restricting the use of regular FEM plate/shell elements [22] and make the analysis cumbersome. Moreover, wrinkling is a highly mesh dependent problem and hence usage of FEM invites high computational time and cost. One way of handling these complexities is to re-mesh the problem domain at every stage or to refine mesh on that particular area. This helps in preventing severe distortion of element and also deals with the discontinuities developed at each stage. However, these are computationally expensive and less accurate. The use of very fine elements also leads to generation of localized modes in problems like wrinkling, vibration etc. Meshless methods are another way of avoiding the difficulties associated with FEM such as high cost in meshing low accuracy in stress difficulty in adaptive analysis etc [24], by alleviating the discretization of problem domain into elements or meshes. Only a set of nodes scattered on the problem domain are needed to represent the problem domain. The main advantage of these methods is its simplicity in adaptive analysis and problems with moving boundary discontinuities.

Among these meshless methods, the EFGM is particularly simple and has faster rate of convergence [25, 26, 27]. EFGM uses moving least square (MLS) method [28] for approximating the function. Krysl and Belytschko[29] introduced EFGM to the bending analysis of thin plates using Kirchhoff thin plate theory and extended the same to the analysis of thin shells [30]. EFGM had been used for vibration analysis of beams, plates and shells [31, 32, 33]. Enforcing essential boundary conditions in EFGM is little complicated when compared to methods like FEM. This is owing to lack of Kronecker delta property of MLS shape functions. Several techniques like, Lagrange multiplier [25], penalty function [34], scaled transformation method [27] etc are used to address this issue. For the current paper scaled transformation method is used for enforcing essential boundary conditions. Overall EFGM is found to be computationally efficient tool in many engineering applications. However, wrinkling analysis of pre-stressed membranes and non-homogeneous membranes using EFGM or any other meshless methods is missing in the available literature.

The current study proposes to use EFGM for wrinkling analysis of pre-stressed membranes.

In the proposed approach, governing differential equation is derived by coupling bending stiffness along with membrane forces using the principles of Love- Kirchoff thin plate theory. Though the present work does not do a study on post wrinkling behavior or study on effect permanent folding lines which necessitate the incorporation of bending stiffness [22], thin plate bending theory is used to obtain more general and simple mathematical model. The proposed method uses a deflection approximation with three degree of freedom (DOF) per node for MLS. The capabilities of EFGM in wrinkling of membrane is examined through different examples. In the first case, wrinkling analysis of a homogeneous membrane, without pre-stress is solved using EFGM and results are compared with those obtained analytically and from FEM. Wrinkling due to compressive loading and due to shear loading of a pre-stressed membrane is studied in the second problem. The results obtained from both these problems are compared with FEM, using different element types available. The dependency of eigenvalues on the number of nodes are studied and a convergence study is also conducted with the different element types and EFGM. In the third case, wrinkling of a non-homogeneous membrane subjected to linearly varying uni-axial load is studied. The results obtained are compared with the solutions available in the literature.

## VARIATIONAL FORMULATION FOR MEMBRANE WRINKLING

The governing differential equation of an initially flat, isotropic, non-homogeneous thin membrane defined on a $x-y$ Cartesian plane; assuming classical plate theory for incorporating the flexural rigidity and considering in-plane forces is given by,

$$\frac{\partial^2 M_x}{\partial x^2} - 2\frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} = F_x\left(\frac{\partial^2 w}{\partial x^2}\right) + F_y\left(\frac{\partial^2 w}{\partial y^2}\right) + 2F_{xy}\left(\frac{\partial^2 w}{\partial x \partial y}\right), \tag{1}$$

where $M_x$, $M_y$ are bending moments with respect to $y, x$ axis and $M_{xy}$ is the twisting moment, given by,

$$\begin{aligned} M_x &= -D(x,y)\left[\frac{\partial^2 w}{\partial x^2} + \nu\frac{\partial^2 w}{\partial y^2}\right] \\ M_y &= -D(x,y)\left[\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2}\right] \\ M_{xy} &= D(x,y)(1-\nu)\frac{\partial^2 w}{\partial x \partial y}, \end{aligned} \tag{2}$$

where $w$ is the deflection in $z$ direction, $F_x, F_y$ are in plane loads per unit length in $x$, $y$ direction and $F_{xy}$ is shear load per unit length acting on the middle plane of the membrane. $D(x,y)$ is membrane flexural rigidity defined as,

$$D(x,y) = \frac{E(x,y)h^3}{12(1-\nu^2)}, \tag{3}$$

where $h$ is thickness of membrane and $E(x,y)$ is modulus of elasticity, and $\nu$ is Poisson's ratio, which is assumed to be constant. The increment in the total potential energy of the thin plate upon buckling is given by [35],

$$\begin{aligned} \Delta\Pi = \frac{1}{2}\iint_A D\Big[\Big(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\Big)^2 + 2(1-\nu)\Big[\Big(\frac{\partial^2 w}{\partial x \partial y}\Big)^2 - \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2}\Big)\Big]dxdy \\ + \frac{1}{2}\int_0^a \int_0^b \Big[F_x\Big(\frac{\partial w}{\partial x}\Big)^2 + F_y\Big(\frac{\partial w}{\partial y}\Big)^2 + 2F_{xy}\Big(\frac{\partial w}{\partial x}\frac{\partial w}{\partial y}\Big)\Big]dxdy. \end{aligned} \tag{4}$$

Equation 4 is solved using EFGM as explained in the following sections.

## EFGM FORMULATION

In EFGM, the unknown field variable $\mathbf{w}^h(\mathbf{x})$ can be written as[25].,

$$\mathbf{w}^h(\mathbf{x}) = \sum_{i=1}^{m} p_i(x)a_i(x) = \mathbf{p}^T(\mathbf{x})\mathbf{a}(\mathbf{x}), \tag{5}$$

where $\mathbf{p}^T(\mathbf{x})$ is the basis function of order $m$ and $\mathbf{a(x)}$ are the unknown coefficients which depends on the position $\mathbf{x}$. To model thin membrane using Love-Kirchhoff's plate assumption, a quadratic basis function is used, which is given by,$\mathbf{p}^T(\mathbf{x}) = \begin{Bmatrix} 1 & x & y & x^2 & xy & y^2 \end{Bmatrix}$

The unknown coefficients at any point $\mathbf{x}$, are determined by performing a weighted least square fit of the local approximation, which in turn determined by minimizing the difference between local approximation at that point and nodal parameters [26]. A support domain or domain of influence is considered such that the weight function chosen has finite value inside this domain and has a value zero outside. A rectangular domain of influence is considered here[26]. Weight function of cubic type [24] is used for the present study Here in case of thin membrane wrinkling, considering three DOF per node, $\mathcal{L}2$ norm can be defined as,

$$J = \sum_{i}^{n} W(\mathbf{x} - \mathbf{x}_i)\Big\{[\mathbf{p}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - w_i]^2 + [\mathbf{p}_{,y}{}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - \theta_{xi}]^2 + [\mathbf{p}_{,x}{}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - \theta_{yi}]^2\Big\}, \tag{6}$$

where $W(\mathbf{x} - \mathbf{x}_i)$ is the weight function, $n$ is the number of nodes inside domain of influence and $\theta_x$,$\theta_y$ are rotation in $x, y$ direction respectively. The unknown coefficients can be found out by minimizing the $\mathcal{L}2$ norm. By differentiating $J$ with respect to $a_j$,

$$\begin{aligned} \frac{\partial J}{\partial a_j} = \sum_{i}^{n} W(x - x_i)\Big\{ & p_j(\mathbf{x}_i)[\mathbf{p}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - w_i] + p_{j,y}(\mathbf{x_i})[\mathbf{p}_{,y}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - \theta_{xi}] \\ & + p_{j,x}(\mathbf{x_i})[\mathbf{p}_{,x}{}^T(\mathbf{x}_i)\mathbf{a}(\mathbf{x}) - \theta_{yi}]\Big\} = 0, \\ & j = 1, 2, 3....m \end{aligned} \tag{7}$$

Assembling all $j$ and representing in matrix form,

$$\mathbf{A(x)a(x)=C(x)}. \tag{8}$$

where, $\mathbf{A}(\mathbf{x})$ is called the weighted moment matrix [24] defined by,

$$\mathbf{A}(\mathbf{x}) = \sum_{i=1}^{n} W_i(x - x_i)[\mathbf{p}(\mathbf{x}_i)\mathbf{p}^T(\mathbf{x}_i) + \mathbf{p}_{,y}(\mathbf{x}_i)\mathbf{p}_{,y}{}^T(\mathbf{x}_i) + \mathbf{p}_{,x}(\mathbf{x}_i)\mathbf{p}_{,x}{}^T(\mathbf{x}_i)], \tag{9}$$

$\mathbf{C(x)}$ is given by,

$$\mathbf{C(x)} = \mathbf{C_w w} + \mathbf{C}_{\theta_\mathbf{x}}\theta_\mathbf{x} + \mathbf{C}_{\theta_\mathbf{y}}\theta_\mathbf{y}, \tag{10}$$

with,

$$\begin{aligned}
\mathbf{C}_{\mathbf{w}} &= \begin{bmatrix} W(\mathbf{x}-\mathbf{x}_1)\mathbf{p}(\mathbf{x}_1) & W(\mathbf{x}-\mathbf{x}_2)\mathbf{p}(\mathbf{x}_2) & . & . & . & W(\mathbf{x}-\mathbf{x}_n)\mathbf{p}(\mathbf{x}_n) \end{bmatrix} \\
\mathbf{C}_{\theta_{\mathbf{x}}} &= \begin{bmatrix} W(\mathbf{x}-\mathbf{x}_1)\mathbf{p}_{,y}(\mathbf{x}_1) & W(\mathbf{x}-\mathbf{x}_2)\mathbf{p}_{,y}(\mathbf{x}_2) & . & . & . & W(\mathbf{x}-\mathbf{x}_n)\mathbf{p}_{,y}(\mathbf{x}_n) \end{bmatrix} \\
\mathbf{C}_{\theta_{\mathbf{y}}} &= \begin{bmatrix} W(\mathbf{x}-\mathbf{x}_1)\mathbf{p}_{,x}(\mathbf{x}_1) & W(\mathbf{x}-\mathbf{x}_2)\mathbf{p}_{,x}(\mathbf{x}_2) & . & . & . & W(\mathbf{x}-\mathbf{x}_n)\mathbf{p}_{,x}(\mathbf{x}_n) \end{bmatrix}
\end{aligned} \tag{11}$$

.

From equation 8 and equation 10, $\mathbf{a}(\mathbf{x})$ can be obtained as,

$$\mathbf{a}(\mathbf{x}) = \mathbf{A}(\mathbf{x})^{-1}[\mathbf{C}_{\mathbf{w}}\mathbf{w} + \mathbf{C}_{\theta_{\mathbf{x}}}\theta_{\mathbf{x}} + \mathbf{C}_{\theta_{\mathbf{y}}}\theta_{\mathbf{y}}]. \tag{12}$$

Finally, substituting equation 12 in equation 5, the approximation function $\mathbf{w}^h(\mathbf{x})$ can be written as,

$$\mathbf{w}^h(\mathbf{x}) = \mathbf{\Phi}_{\mathbf{w}}^T(\mathbf{x})\mathbf{w} + \mathbf{\Phi}_{\theta_{\mathbf{x}}}^T\theta_{\mathbf{x}} + \mathbf{\Phi}_{\theta_{\mathbf{y}}}^T\theta_{\mathbf{y}}, \tag{13}$$

where,

$$\begin{aligned}
\mathbf{\Phi}_{\mathbf{w}}^T &= \mathbf{p}^T(\mathbf{x})\mathbf{A}^{-1}(\mathbf{x})\mathbf{C}_{\mathbf{w}} \\
\mathbf{\Phi}_{\theta_x}^T &= \mathbf{p}^T(\mathbf{x})\mathbf{A}^{-1}(\mathbf{x})\mathbf{C}_{\theta_{\mathbf{x}}}. \\
\mathbf{\Phi}_{\theta_y}^T &= \mathbf{p}^T(\mathbf{x})\mathbf{A}^{-1}(\mathbf{x})\mathbf{C}_{\theta_{\mathbf{y}}}
\end{aligned} \tag{14}$$

Thus the equation 13 can be rewritten in matrix form as,

$$\mathbf{w}^h(\mathbf{x}) = \begin{bmatrix} \mathbf{\Phi}_{\mathbf{w}}^T & \mathbf{\Phi}_{\theta_x}^T & \mathbf{\Phi}_{\theta_y}^T \end{bmatrix} \begin{Bmatrix} \mathbf{w} \\ \theta_x \\ \theta_y \end{Bmatrix} = \mathbf{N}^T\mathbf{d}. \tag{15}$$

However, due to the lack of Kronecker delta properties in EFGM, imposing essential boundary conditions involves some additional complication. Here scaled transformation method [27] is used, in which the displacement approximations given in equation 13 are scaled using a scaled transformation matrix in such a way that, nodal values and nodal parameters will coincide along the nodes on essential boundary.

In scaled transformation method, an identity matrix say $\mathbf{\Lambda}$ having number of rows and columns equal to the total DOF is constructed. The rows corresponding to the degrees of freedom, in which essential boundary conditions has to be applied, is then populated with shape function associated with it. A representation of $\mathbf{\Lambda}$ can be shown as,

$$\mathbf{\Lambda} = \begin{Bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & . & . & \phi_{1m} & . & . \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & . & . & \phi_{2m} & . & . \\ 0 & 0 & 1 & 0 & . & . & 0 & . & . \\ 0 & 0 & 0 & 1 & . & . & 0 & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ \phi_{n1} & \phi_{n2} & \phi_{n3} & . & . & . & \phi_{nm} & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \end{Bmatrix}, \tag{16}$$

where, $\phi_{ij}$ represents shape function of node $i$ evaluated at node $j$. Here $1^{st}$,$2^{nd}$ and $n^{th}$ are rows corresponding to degrees of freedom in which boundary condition has to be applied. Detailed

discussions on this are not included here and reader may advised to go through [27]. Once, $\mathbf{\Lambda}$ is formed, the discretized system of equations can be written from 4 and 15 as,

$$\frac{1}{2}\overline{\mathbf{d}}^T\overline{\mathbf{K_f}\mathbf{d}} + \frac{F_x}{2}\overline{\mathbf{d}}^T\overline{\mathbf{Kg}_x}\overline{\mathbf{d}} + \frac{F_y}{2}\overline{\mathbf{d}}^T\overline{\mathbf{Kg}_y}\overline{\mathbf{d}} + F_{xy}\overline{\mathbf{d}}^T\overline{\mathbf{Kg}_{xy}}\overline{\mathbf{d}} = 0, \tag{17}$$

where, $\overline{\mathbf{K}}$ is the modified stiffness matrix, $\overline{\mathbf{Kg_x}}, \overline{\mathbf{Kg_y}}$ and $\overline{\mathbf{Kg_{xy}}}$ are modified geometric stiffness matrices, given by,

$$\begin{aligned}
\overline{\mathbf{K}} &= \mathbf{\Lambda^{-T} K \Lambda^T} \\
\overline{\mathbf{Kg_x}} &= \mathbf{\Lambda^{-T} Kg_x \Lambda^T} \\
\overline{\mathbf{Kg_y}} &= \mathbf{\Lambda^{-T} Kg_y \Lambda^T} \\
\overline{\mathbf{Kg_{xy}}} &= \mathbf{\Lambda^{-T} Kg_{xy} \Lambda^T},
\end{aligned} \tag{18}$$

where $\mathbf{K_f}$ is stiffness matrix and $\mathbf{Kg}_x, \mathbf{Kg}_y$ and $\mathbf{Kg}_{xy}$ are geometric stiffness matrices. The sign of geometric stiffness terms in the equation depends upon the direction of in-plane loads. The terms will be positive for stretching and negative for compression. Stiffness matrix and geometric stiffness matrices are given by,

$$\begin{aligned}
\mathbf{K_f} &= \iint \mathbf{N''}^T \mathbf{D} \mathbf{N''} dxdy \\
\mathbf{Kg}_x &= \iint \mathbf{N}_{,x}^T \mathbf{N}_{,x} dxdy \\
\mathbf{Kg}_y &= \iint \mathbf{N}_{,y}^T \mathbf{N}_{,y} dxdy \\
\mathbf{Kg}_{xy} &= \iint \mathbf{N}_{,x}^T \mathbf{N}_{,y} dxdy,
\end{aligned} \tag{19}$$

where $\mathbf{N''}$ represents matrix containing the double derivatives of shape function, $\mathbf{N_{,x}}, \mathbf{N_{,y}}$ represents partial derivative of $\mathbf{N}$ with respect to $x, y$ respectively. Consequently, the essential boundary conditions are imposed simply by following FEM methodology.

## MODEL DESCRIPTION

For the present study, four cases are taken in to consideration. Case(a):Wrinkling analysis of homogeneous thin plate subjected to uni-axial compressive loading Case(b): Wrinkling analysis of a pre-stressed, homogeneous membrane, case(c): Wrinkling analysis of non-homogeneous membranes, subjected to linearly varying, uni-axial load. Though inhomogeneity in modulus of elasticity is considered for case(c) and case(d), Poisson's ratio is assumed to be constant. Moreover, for all the cases material is assumed to be isotropic.

### Case(a): Wrinkling analysis of a homogeneous thin plate with uni-axial compressive loading

A homogeneous thin plate of dimension $0.3m \times 0.3m$ with thickness $0.2mm$, which is free of any initial pre-stress and loaded with a uni-axial compressive load is considered for the study. The material is assumed to be Kapton which has an Young's modulus of $3500 \times 10^6 N/m^2$ and Poisson's ratio of 0.31. All sides of the membrane are given simply supported (SSSS) boundary condition. The results obtained using EFGM, are compared with analytical solution and the results obtained using different element types in FEM. Analytical solution obtained using Navier's solution method is used here [36]. Five different quadrilateral element types, S4 (four noded element with six DOF per node and with full integration), S4R(four noded element,

with six DOF per node and with reduced integration), S4R5 (four noded element with 5 DOF per node with reduced integration), S8R(8 noded element with six DOF per node and with reduced integration) and S8R5 (8 noded element with 5 DOF per node and with reduced integration) are taken for modeling. For comparison, discretization in EFGM is done with regularly distributed nodes, similar to that of linear element types (S4, S4R and S4R5) in FEM.

Three different modes, first ($m = 1, n = 1$), $20^{th}(m = 5, n = 2)$ and $30^{th}(m = 9, n = 2)$ modes are considered for study, where $m, n$ are number of half sine waves in $x, y$ direction respectively. This is selected with an objective to study the capability of the methods in an initial, an intermediate and a higher mode. Mode shapes of corresponding modes, obtained from EFGM are shown in Fig. 1. A convergence study on eigen values obtained using EFGM and all element types in FEM for the the three different modes under consideration have been carried out.

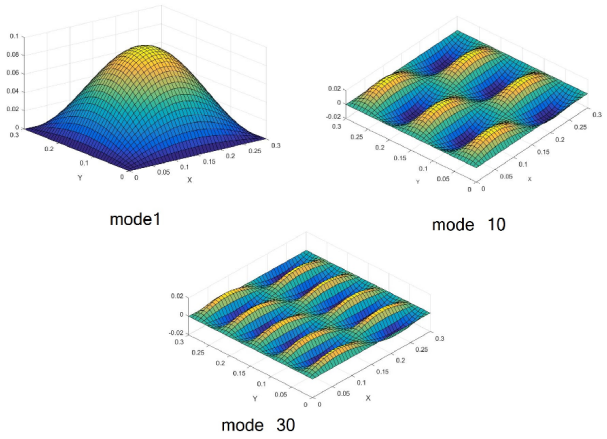
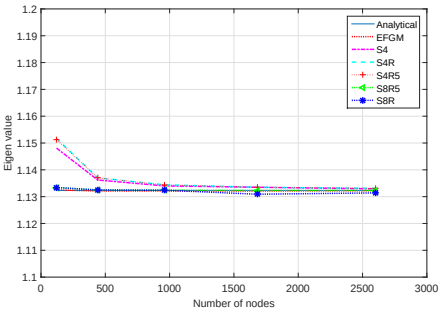


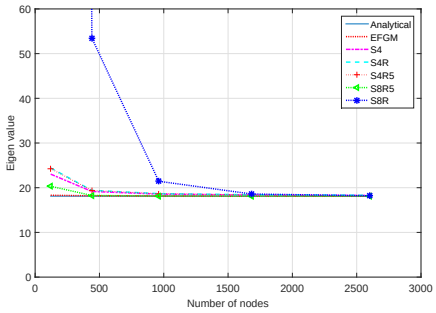

**Fig. 1.** Mode shapes obtained from EFGM, for wrinkling of membrane without pre-stress and simply supported on all sides

Fig. 2a shows the variation of eigenvalue with number nodes for the the first mode. It can be observed from Fig. 2a that, FEM results obtained from models which uses 4 node quadrilateral elements are showing larger variations from analytical solution when compared to other element types and EFGM, for lower number nodes. However, it is to be noted that the difference from analytical solution for S4 element are is within 2%. Moreover, a convergence is observed while number of nodes are increased. Accurate and consistent results are obtained while 8 node quadrilateral elements and EFGM are used for the analysis for all the set of nodes. Variation of $20^{th}$ mode eigenvalues with respect to number of nodes is shown in Fig. 2b. From the figure it can be observed that the variation of the results corresponding to element types S8R, S4R, S4R5 and S4 are very large for lesser number of nodes. For higher number of nodes all the element types are providing satisfactory results. However element type S8R5 and EFGM are showing very small error even with the use of lower number of nodes and hence shows a comparatively consistent results. Eigenvalues obtained for higher mode under consideration is given in Fig. 2c.

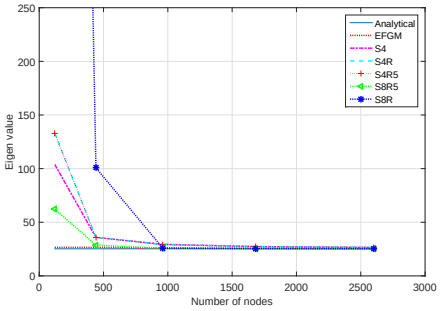
A large error is observed for all FEM elements at the lesser number of nodes, while EFGM gives fairly accurate and consistent results. Table 1 shows the percentage variation of eigenvalues obtained using EFGM and FEM from analytical solution for the set 121 nodes. It can be observed that S8R is showing a huge variation of 8239.83% and other FEM element types also show unacceptable variation from analytical solution at lesser number of nodes. However EFGM results are within 6%variation from analytical solutions.




**(a)** $1^{st}$ mode$(m = 1, n = 1)$




**(b)** $20^{th}$ mode$(m = 4, n = 4)$




**(c)** $30^{th}$ mode$(m = 9, n = 2)$

**Fig. 2.** Variation of $1^{st}$ mode$(m = 1, n = 1)$ eigenvalues with number of nodes for the case (a)

**Table 1.** Percentage variation of eigenvalues from analytical value with the use of 121 nodes for $30th$ mode$(m = 9, n = 2)$

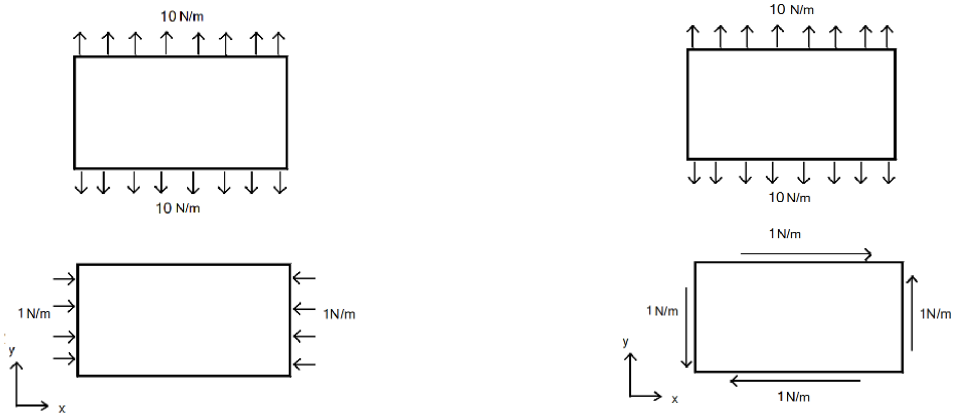
| Method | Element type | Percentage variation(%) |
|---|---|---|
| FEM | S4 | 311.36 |
| | S4R | 427.50 |
| | S4R5 | 426.36 |
| | S8R | 8239.83 |
| | S8R5 | 146.65 |
| EFGM | | 5.38 |

Thus EFGM with only 3 DOF per nodes is observed to provide accurate results with lesser number of nodes compared to FEM with 5 or 6 DOF per node.

**Case(b): Wrinkling analysis of a pre-stressed, homogeneous membrane**

A pre-stressed homogeneous membrane with same size and material as that of case(a) is considered for analysis. Two different numerical examples are examined in this case. In both the examples, the membrane is being pre-stressed in the $y$ direction and given simply supported boundary condition on all sides. In example(i), Compression induced wrinkling of pre-stressed membrane due to compressive load in $x$ direction Fig. 3a was studied and in example(ii), shear induced wrinkling of pre-stressed membrane as shown in Fig. 3b is studied. Both the problems are analyzed using EFGM and compared with results from FEM analysis using different element types available. Three modes (first mode, and an intermediate mode and a higher mode) are

taken into consideration. A convergence study for eigenvalues corresponds to $1^{st}$ mode, $20^{th}$ mode and $30^{th}$ mode have been carried out for both the examples. Percentage variation of eigenvalues obtained with a reference value is also analyzed for each element type and EFGM for studying the rate of convergence. The reference value is taken as the eigenvalue obtained when maximum number of nodes, that is 3721 nodes, are used. That is, reference value is different for EFGM and every element type.




**(a)** Example 1 **(b)** Example 2

**Fig. 3.** Problem considered for wrinkling analysis of pre-stressed membrane due to compressive and shear loading

*Example 1: Compression induced wrinkling*

Wrinkling of a pre-stressed homogeneous membrane due to compressive loading in $x$ axis is analyzed using EFGM and different element types available. A pre stress of $10N/m$ is applied to the membrane initially in $y$ direction. The mode shapes of the modes taken into consideration for study is shown in Fig. 4

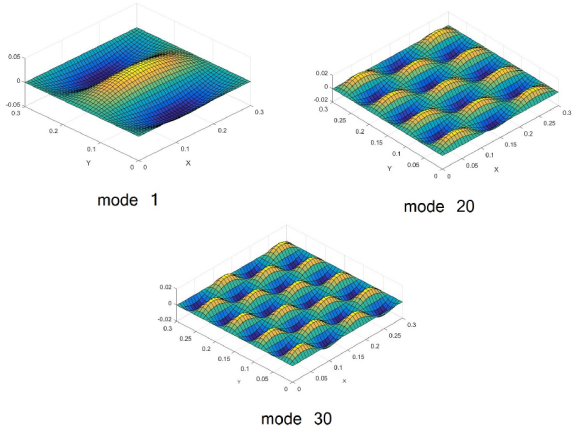


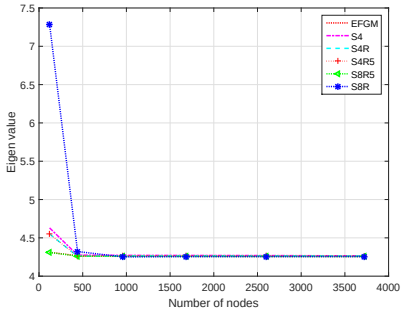

**Fig. 4.** Mode shapes obtained from EFGM, for wrinkling of pre-stressed membrane due to compressive loading

The results obtained from first mode is shown in Fig. 5a. From Fig. 5a it is clear that for initial mode, all FEM element types and EFGM provides fairly consistent eigenvalues with

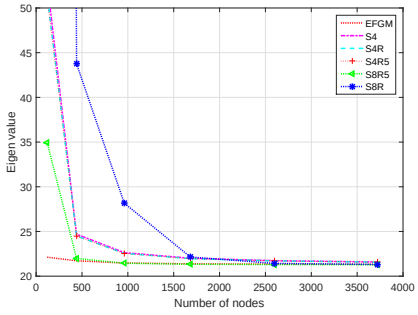
respect to number of nodes, except for lowest number of nodes. At lowest number of nodes considered, eigenvalue corresponding to S8R shows high variation. Percentage variation of eigenvalues from reference value for all element types and EFGM are shown in table 2. From Table 2, it can be observed that, except for S8R5 and EFGM all other element types shows error values more than 5% at 121 nodes. S8R is showing very high percentage variation of 71% at 121 number of nodes. But at 441 nodes percentage variation is reduced to 1.54%. For higher number of nodes, all element types and EFGM show consistent result with percentage variation within 0.3%.

For an intermediate mode, as shown in Fig. 5b, element types S8R, S4, S4R and S4R5 of FEM are showing lower convergence rate, while element types S8R5 and EFGM results shows higher convergence rate. The percentage variation of eigenvalues from the reference values are shown in Table 3. From the table it can be observed that element type S8R shows huge variation of 3584.40% with the use of 121 nodes. Element types S4, S4R, S4R5 and S8R5 also shows large variation at lower number of nodes. All element types except S8R shows percentage variation within 5% for number of nodes higher than 961. However, EFGM shows consistent results with percentage variation of 3.76% even at 121 nodes.

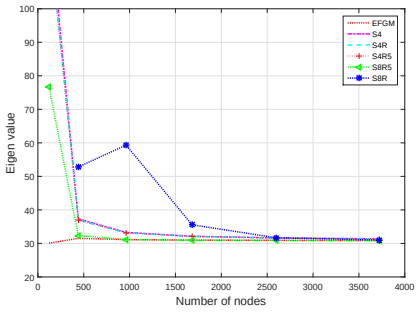
For a higher mode, from Fig. 5c, all element types are showing lower rate of convergence compared to EFGM results. EFGM gives consistent results at higher modes. Percentage variation in eigenvalues are shown in Table 4. As element type S8R fails to capture the $30^th$ mode with the use of 121 nodes, the percentage variation correspond to which is not included in Table 4. From the Table 4, it is clear that element type S8R is showing huge variation at lower number of nodes, element types S4, S4R, S4R5 and S8R5 also show considerable variation with the use of 121 nodes. However EFGM shows more consistent results compared to all element types in FEM.




**(a)** $1^{st}$ mode




**(b)** $20^{th}$ mode




**(c)** $30^{th}$mode

**Fig. 5.** Variation of eigenvalues with number of nodes for compression induced wrinkling

**Table 2.** Percentage variation of eigenvalue for $1^{st}$ mode compressive loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 8.50 | 0.22 | 0.15 | 0.12 | 0.10 | 0 |
| S4R | 6.74 | 0.087 | 0.04 | 0.09 | 0.07 | 0 |
| S4R5 | 6.70 | 0.09 | 0.04 | 0.09 | 0.07 | 0 |
| S8R5 | 1.24 | 0.12 | 0.04 | 0.01 | 0.01 | 0 |
| S8R | 71.2 | 1.54 | 0.09 | 0.02 | 0.01 | 0 |
| EFGM | 1.26 | 0.19 | 0.1 | 0.04 | 0.01 | 0 |

**Table 3.** Percentage variation of eigenvalue for $20^{th}$ mode compressive loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 144.53 | 14.38 | 4.83 | 1.94 | 0.67 | 0 |
| S4R | 138.24 | 13.57 | 4.55 | 1.83 | 0.63 | 0 |
| S4R5 | 136.96 | 13.46 | 4.51 | 1.81 | 0.62 | 0 |
| S8R5 | 64.02 | 3.28 | 0.72 | 0.22 | 0.06 | 0 |
| S8R | 3584.40 | 105.37 | 32.20 | 3.94 | 0.57 | 0 |
| EFGM | 3.76 | 1.88 | 0.72 | 0.30 | 0.10 | 0 |

**Table 4.** Percentage variation of eigenvalue for $30^{th}$ mode compressive loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 307.72 | 18.85 | 6.11 | 2.43 | 0.84 | 0 |
| S4R | 293.83 | 17.93 | 5.81 | 2.31 | 0.79 | 0 |
| S4R5 | 289.53 | 17.73 | 5.75 | 2.28 | 0.79 | 0 |
| S8R5 | 148.48 | 4.85 | 1.04 | 0.32 | 0.09 | 0 |
| S8R | * | 70.27 | 91.31 | 14.65 | 2.16 | 0 |
| EFGM | 2.56 | 2.04 | 0.80 | 0.33 | 0.16 | 0 |

*S8R fails to capture $30^{th}$ mode at 121 nodes

*Example 2: Shear induced wrinkling*

Wrinkling of a pre-stressed homogeneous membrane due to shear loading is analyzed using EFGM and FEM. An in-plane pre-tension of $10N/m$ is applied. The mode shapes of the corresponding modes taken into consideration for study is shown in Fig. 6.

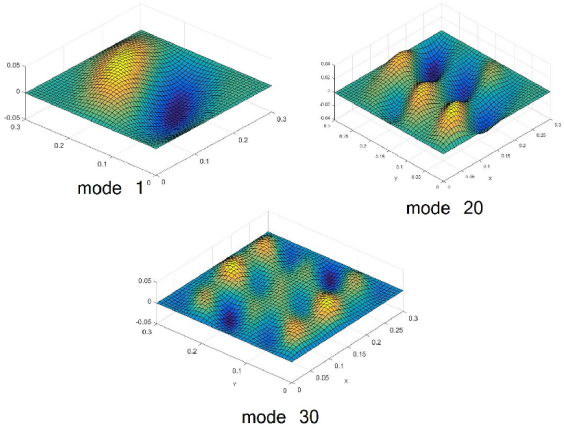
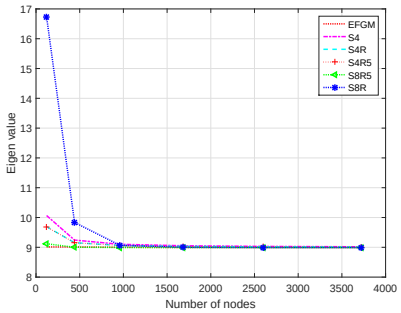



**Fig. 6.** Mode shapes obtained from EFGM, for wrinkling of pre-stressed membrane due to shear loading

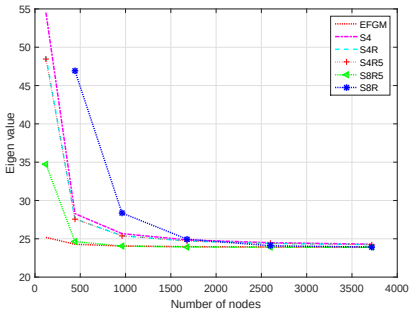
The variation of eigenvalues with the number of nodes for the first mode is shown in Fig. 7a. From Fig. 7a it is observed that, element type S4R and S8R shows less consistency in results at lower number of nodes, while other element types show consistent results. Particularly, element type S8R5 and EFGM shows consistency and high rate of convergence in the results. Table 5 shows percentage variation in eigenvalues of each element types and EFGM at lower number of nodes. It is observed that at 121 nodes all element types except S8R5 and EFGM are show percentage variation higher than 5%. Particularly element type S8R shows a huge variation of 86.13% at 121 nodes. With the use of 441 number of nodes, except for element type S8R, all other element types and EFGM has percentage variation value within 3%. Particularly, element type S8R5 and EFGM shows very low percentage variation of 0.12% and 0.21 % respectively.

Fig. 7b shows the variation in $20^{th}$ mode eigenvalues for shear induced wrinkling. From Fig. 7b it is observed that, all element types except S8R5 shows inconsistent results. Also EFGM provides faster rate of convergence compared to all element types. Table 6 shows the percentage variation of $20^{th}$ mode eigenvalues with number of nodes. As element type S8R fails to capture $20^{th}$ mode, the results corresponding to which is not included in Table 6 and Fig. 7b. From the Table 6, it is clear that EFGM shows very low percentage variation and higher rate of convergence compared to all element types in FEM.

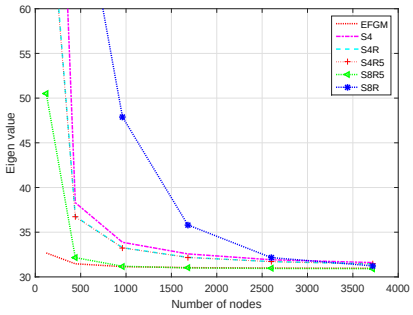
Fig. 7c shows the variation of eigenvalues corresponding to the higher mode under consideration with respect to total number of nodes used for discretization. It can be clearly observed that only EFGM is able to produce consistent results irrespective of the number of nodes, in comparison with FEM. Table 7 shows the percentage variation of eigenvalues for all the elements types in FEM and EFGM for different set of nodes with respect to its own reference value. It is very much clear from the table that EFGM shows faster rate of convergence with consistent result at higher modes too. FEM results are highly dependent on the number of nodes and element types used. Similar to $20^{th}$ mode results, element type S8R fails to capture $30^{th}$ mode also at 121 number of nodes




**(a)** $1^{st}$ mode




**(b)** $20^{th}$ mode




**(c)** $30^{th}$mode

**Fig. 7.** Variation of eigenvalues with number of nodes for for shear induced wrinkling

**Table 5.** Percentage variation of eigenvalue for $1^{st}$ mode shear loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 11.63 | 2.49 | 0.92 | 0.38 | 0.13 | 0 |
| S4R | 7.61 | 1.62 | 0.60 | 0.25 | 0.09 | 0 |
| S4R5 | 7.52 | 1.60 | 0.59 | 0.24 | 0.08 | 0 |
| S8R5 | 1.42 | 0.12 | 0.03 | 0.01 | 0.01 | 0 |
| S8R | 86.13 | 9.39 | 0.85 | 0.10 | 0.01 | 0 |
| EFGM | 0.22 | 0.21 | 0.08 | 0.03 | 0.01 | 0 |

**Table 6.** Percentage variation of eigenvalue for $20^{th}$ mode shear loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 124.43 | 13.92 | 5.66 | 2.30 | 0.80 | 0 |
| S4R | 101.34 | 13.92 | 4.83 | 1.96 | 0.68 | 0 |
| S4R5 | 100.02 | 13.82 | 4.80 | 1.94 | 068 | 0 |
| S8R5 | 45.40 | 3.13 | 0.68 | 0.2 | 0.06 | 0 |
| S8R | * | 96.23 | 18.56 | 4.24 | 0.69 | 0 |
| EFGM | 5.31 | 1.54 | 0.60 | 0.25 | 0.09 | 0 |

*S8R fails to capture $20^{th}$ mode at 121 nodes

**Table 7.** Percentage variation of eigenvalue for $30^{th}$ mode shear loading

| Method/Element type | Number of nodes | | | | | |
|---|---|---|---|---|---|---|
| | 121 | 441 | 961 | 1681 | 2601 | 3721 |
| S4 | 283.14 | 21.22 | 7.20 | 3.09 | 1.01 | 0 |
| S4R | 146.12 | 16.91 | 5.76 | 2.32 | 0.81 | 0 |
| S4R5 | 143.88 | 16.72 | 5.67 | 2.32 | 0.81 | 0 |
| S8R5 | 63.36 | 3.98 | 0.86 | 0.26 | 0.01 | 0 |
| S8R | * | 146.58 | 53.33 | 14.67 | 2.96 | 0 |
| EFGM | 5.55 | 1.64 | 0.61 | 0.25 | 0.09 | 0 |

*S8R fails to capture $30^{th}$ mode at 121 nodes

Thus it can be concluded that in wrinkling analysis of pre-stressed membrane, EFGM shows rate of convergence and consistent results, while the values obtained using FEM are highly dependent on the element type and size used.

**Case(c):Wrinkling analysis of non-homogeneous membrane, subjected to linearly varying, uni-axial load**

Wrinkling analysis of non-homogeneous membrane, subjected to linearly varying, uni-axial compressive load as shown in Fig. 8 is carried out using EFGM. The load $F_x$ is taken as zero and $F_y$ is taken as compressive and linearly varying along edges $y = 0, y = b$.

The variation of $F_y$ is given by [20],

$$F_y = -F_0(1 - \gamma\overline{x}), \tag{20}$$

where $\gamma$ is the loading parameter. $F_0$ is the load at $\overline{x} = 0$, where $\overline{x}$ is given by,

$$\overline{x} = \frac{x}{a} \tag{21}$$

similarly for consistency other non-dimensional parameters are defined as,

$$\begin{aligned} \overline{h} &= \frac{h}{a} \\ F_0^* &= \frac{12F_0(1-\nu^2)}{aE_0\overline{h}^3}, \end{aligned} \tag{22}$$

where, $a, b$ and $h$ are length, width and thickness of the membrane respectively. $F_0^*$ is the non-dimensional force parameter. The modulus of elasticity $E_0$, is assumed to be exponentially varying, which is given by,

$$E = E_0 e^{\mu\overline{x}} \tag{23}$$

where $\mu$ is the non-homogeneity parameter.

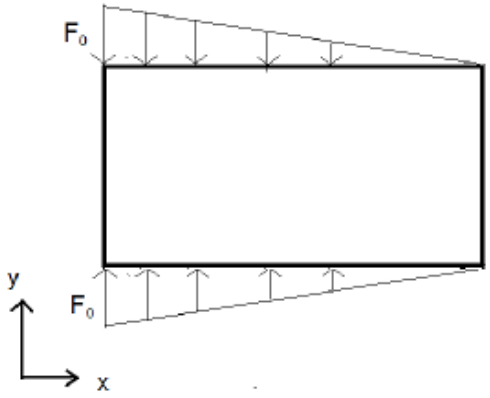



**Fig. 8.** Problem taken for non-homogeneous membrane wrinkling analysis

A non-dimensional critical force parameter is defined as [20],

$$F^*_{0cr} = \frac{12F_{0cr}(1-\nu^2)}{aE_0\overline{h}^3}, \tag{24}$$

where $F_{0cr}$ is the critical load obtained from wrinkling analysis.

Two type of boundary conditions are taken for analysis. In both of the cases, loading side is given simply supported boundary condition ($y = 0$ and $y = b$). In the first case, the other two supports are considered clamped (C-C). In the second problem, one of the non-loading edge is given clamped support and the other edge is given simply supported (C-SS) boundary condition. Analysis are done for two different loading parameters, $\gamma = 0$ ie uniform loading and $\gamma = 1$ ie, linearly varying loading from maximum value $F_0$ at $\overline{x} = 0$ to zero at $\overline{x} = 1$. Table 8 and Table 9 shows a comparison of critical force parameter obtained using EFGM for different values of non-homogeneous parameter, $\mu$. The same problem has been solved by [20] by DQ method. Here results available in [20] is taken for comparison. Table 8 shows the results for $b/a = 1$ and $\gamma = 0$ and Table 9 shows the results of $b/a = 1$ and $\gamma = 1$

**Table 8.** A comparison of critical force parameter for $b/a = 1$ and $\gamma = 0$, obtained using EFGM and using DQ given in[20] for different values of non-homogenity parameter, $\mu$

| $\mu$ | C-SS | | C-C | |
|---|---|---|---|---|
| | EFGM | [20] | EFGM | [20] |
| -0.5 | 43.9491 | 43.9425 | 58.8788 | 66.0304 |
| -0.3 | 48.6693 | 48.6598 | 65.2750 | 73.0507 |
| 0 | 56.6673 | 56.6536 | 75.9725 | 84.9225 |
| 0.3 | 65.9110 | 65.8938 | 88.1120 | 98.6082 |
| 0.5 | 72.8551 | 72.8365 | 97.0748 | 108.8660 |

**Table 9.** A comparison of critical force parameter for $b/a = 1$ and $\gamma = 1$, obtained using EFGM and using DQ given in [20] for different values of non-homogenity parameter, $\mu$

| $\mu$ | C-SS | | C-C | |
|---|---|---|---|---|
| | EFGM | [20] | EFGM | [20] |
| -0.5 | 61.5015 | 61.4897 | 117.7998 | 133.780 |
| -0.3 | 67.8166 | 67.8000 | 128.2408 | 146.139 |
| 0 | 78.4570 | 78.4334 | 145.34 | 166.7380 |
| 0.3 | 90.6731 | 90.6433 | 164.2914 | 190.086 |
| 0.5 | 99.8002 | 99.7673 | 178.0288 | 207.3480 |

From the Table 8 and 9 it can be observed that for membrane with C-SS boundary conditions, EFGM provides matching $F^*_{0cr}$ values with the reference value. However, for membrane with C-C boundary conditions EFGM results do not match with the reference values available in literature. Hence for comparison, a homogeneous membrane whose results are available in literature [17, 19] is taken into consideration. $F^*_{0cr}$ values obtained for homogeneous membrane ($\mu = 0$) of $b/a$ ratio 1 and $\gamma = 0$, with C-C boundary conditions using different methods are shown in table 10. It can be observed that, EFGM results match well with that obtained using all methods except that available in *et al*[20]. Hence, it is evident from table 10 that results of C-C boundary condition shown in [20] is erroneous and that is the reason for mismatch of results in table 8 and table 9 for C-C boundary condition.

**Table 10.** A comparison of critical force parameter for $b/a = 1$ and $\gamma = 0$ for a homogeneous($\mu = 0$) membrane

| Method | $F^*_{0cr}$ |
|---|---|
| EFGM (1296 nodes) | 75.9725 |
| Power series ([17]) | 75.9100 |
| DQM ([19]) | 75.9100 |
| DQM([20]) | 84.9225 |

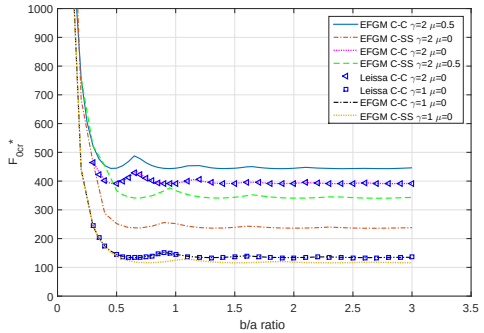


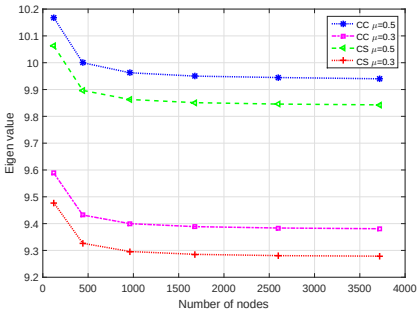

**Fig. 9.** Wrinkling curves for different cases of study using EFGM.

Wrinkling curves for non-homogeneous membrane corresponds to four combinations of $\mu$ and boundary conditions (C-C membrane with $\mu = 0$, C-SS membrane with $\mu = 0$, C-C membrane with $\mu = 0.5$ and C-SS membrane with $\mu = 0.5$) are shown in Fig. 9. Wrinkling

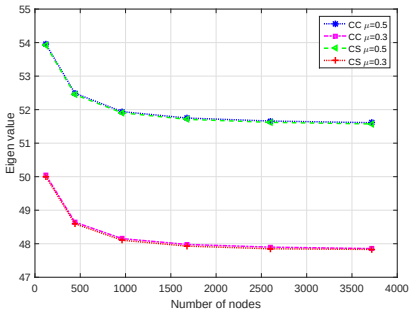
curve for a homogeneous membrane with $\mu = 0$ and loading parameter $\gamma = 2$ and $\gamma = 1$, are available in [17]. From the Fig. 9, it can be observed that, the curves plotted using EFGM are found to match with that available in [17].

**Case(d): Wrinkling analysis of pre-stressed non-homogeneous membrane due to compressive load**

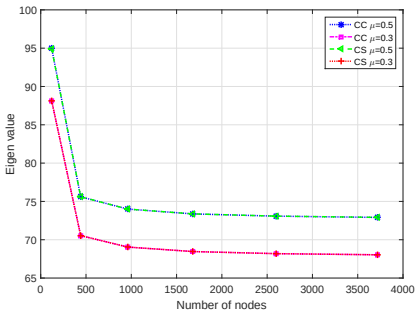
Compression induced wrinkling of non-homogeneous membrane pre-stressed by $10N/m$ in $y$ direction is considered for study. Membranes with non-homogeneity parameter $\mu$ equal to 0.3 and 0.5 and boundary conditions C-C and C-SS are considered. Initial, intermediate and higher modes of wrinkling is examined for all these four cases and the resulting plots showing the variation of eigenvalues with respect to number of nodes are shown in Fig. 10. From the figure it is clear that EFGM shows a faster rate of convergence in results for all the three modes taken into consideration. It is observed that, though at 121 nodes eigenvalues corresponds to $1^{st}$ and $20^{th}$ modes shows variation within 5%, eigenvalue corresponds to $30^{th}$ mode shows a variation of 30% with respect to converged value. However, as number of nodes are increased to 441, variation in all the modes are reduced to 3% or less. It is also found that, in pre-stressed case, the eigenvalues obtained for C-C boundary condition and C-SS boundary condition do not show much difference at higher modes. Also as expected, it is observed that, with the increase in $\mu$ value, the eigenvalue also increases.




**(a)** $1^{st}$ mode




**(b)** $20^{th}$ mode




**(c)** $30^{th}$ mode

**Fig. 10.** Variation of eigenvalue with number of nodes for pre-stressed non-homogeneous membrane

## CONCLUSION

The current work discusses the use of EFGM, a meshless method, for wrinkling analysis of pre-stressed membranes and proves its capabilities in capturing the eigenvalues even at higher modes. Classical plate formulation with three DOF per node is considered for study.

A general EFGM formulation for non-homogeneous membrane with in-plane loads derived by considering the bending stiffness, though it is negligible. Initially, a simple wrinkling analysis of homogeneous thin plate due to uni-axial compressive load is carried out and compared with analytical and FEM solutions to validate the proposed method. The results shows that, EFGM provides accurate results for all the modes under consideration, even with the use of less number of nodes. Further, wrinkling analysis of pre-stressed membrane due to shear and compressive loading is studied. The results obtained from both these problems are compared with result obtained from that of FEM using different element types. A study on consistency of the numerical tools are also conducted by taking respective converged eigenvalues as reference for every element types and EFGM. It is found that, consistent results adn faster rate of convergence are obtained when EFGM is employed regardless of modes under consideration. However many FEM element types shows larger variation at lower number of nodes at higher modes. The study is further extended to wrinkling analysis of non-homogeneous membrane subjected to linearly varying in-plane load, to demonstrate the capability of the proposed method. The results obtained matches well with the results available in the literature. Moreover wrinkling curves for non-homogeneous membranes are proposed. Finally, EFGM is used for wrinkling analysis of pre-stressed non-homogeneous membrane for two different non-homogeneity parameter and boundary conditions.It is observed that, EFGM produces consistent results and shows a faster rate of convergence. Thus, EFGM is proved to be an efficient and accurate numerical tool which can be used for wrinkling analysis of membranes.

**REFERENCES**

[1] David Sleight, Yuki Michii, David Lichodziejewski, Billy Derbes, Troy Mann, Kara Slade, and John Wang. Finite element analysis and test correlation of a 10-meter inflation-deployed solar sail. In *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, page 2121, 2005.

[2] Christopher Talley, William Clayton, Paul Gierow, Greg Laue, Jennie McGee, and James Moore. Advanced membrane materials for improved solar sail capabilities. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1561, 2002.

[3] Herbert Wagner. Flat sheet metal girders with very thin metal web. *Z. Flugtechn. Motorluftschiffahrt*, 20:200–314, 1929.

[4] Eric Reissner. On tension field theory. *Proc. of the 5th Int. Congr. for Applied Mechanics Harvard Univ. & MIT*, pages 88–92, 1938.

[5] Manuel Stein and John M Hedgepeth. *Analysis of partly wrinkled membranes*. National Aeronautics and Space Administration Washington, 1961.

[6] E H Mansfield. Tension field theory, a new approach which shows its duality with inextensional theory [c]. In *Proceeding XII International Congress of Applied Mechanics*, pages 305–320, 1968.

[7] Chien H Wu and Thomas R Canfield. Wrinkling in finite plane-stress theory. *Quarterly of Applied Mathematics*, 39(2):179–199, 1981.

[8] Allen C Pipkin. The relaxed energy density for isotropic elastic membranes. *IMA journal of applied mathematics*, 36(1):85–99, 1986.

[9] D J Steigmann. Tension-field theory. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 429, pages 141–173. The Royal Society, 1990.

[10] Wesley Wong and Sergio Pellegrino. Wrinkled membranes i: experiments. *Journal of Mechanics of Materials and Structures*, 1(1):3–25, 2006.

[11] Wesley Wong and Sergio Pellegrino. Wrinkled membranes ii: analytical models. *Journal of Mechanics of Materials and Structures*, 1(1):27–61, 2006.

[12] Wesley Wong and Sergio Pellegrino. Wrinkled membranes iii: numerical simulations. *Journal of Mechanics of Materials and Structures*, 1(1):63–95, 2006.

[13] Wei-wei Xiao, Wu-jun Chen, and Gong-yi Fu. Wrinkle analysis of a space planar film reflect-array. *Journal of Zhejiang University-SCIENCE A*, 12(1):24–32, 2011.

[14] Satish Kumar, S H Upadhyay, and Anil C Mathur. Wrinkling simulation of membrane structures under tensile and shear loading. *Journal of Vibration Analysis*, 3(1):17–33, 2015.

[15] Y Tomita and A Shindo. Onset and growth of wrinkles in thin square plates subjected to diagonal tension. *International Journal of Mechanical Sciences*, 30(12):921–931, 1988.

[16] Tomoshi Miyamura. Wrinkling on stretched circular membrane under in-plane torsion:: bifurcation analyses and experiments. *Engineering Structures*, 22(11):1407–1425, 2000.

[17] Arthur W Leissa and Jae-Hoon Kang. Exact solutions for vibration and buckling of an ss-c-ss-c rectangular plate loaded by linearly varying in-plane stresses. *International Journal of mechanical sciences*, 44(9):1925–1945, 2002.

[18] Jae-Hoon Kang and Arthur W Leissa. Exact solutions for the buckling of rectangular plates having linearly varying in-plane loading on two opposite simply supported edges. *International Journal of Solids and Structures*, 42(14):4220–4238, 2005.

[19] Xinwei Wang, Lifei Gan, and Yongliang Wang. A differential quadrature analysis of vibration and buckling of an ss-c-ss-c rectangular plate loaded by linearly varying in-plane stresses. *Journal of Sound and Vibration*, 298(1):420–431, 2006.

[20] Roshan Lal and Renu Saini. Buckling and vibration of non-homogeneous rectangular plates subjected to linearly varying in-plane force. *Shock and Vibration*, 20(5):879–894, 2013.

[21] Vishal Nayyar, K Ravi-Chandar, and Rui Huang. Stretch-induced stress patterns and wrinkles in hyperelastic thin sheets. *International Journal of Solids and Structures*, 48(25):3471–3483, 2011.

[22] P Frank Pai. *Highly flexible structures: modeling, computation, and experimentation*. AIAA (American Institute of Aeronautics & Ast, 2007.

[23] OC Zienkiewicz and RL Taylor. *The Finite Element Method. Solid and Fluid Mechanics. Dynamics and Non-Linearity, Vol. II*. McGraw-Hill, New York, 1991.

[24] Gui Rong Liu and Yuan Tong Gu. *An introduction to meshfree methods and their programming*. Springer Science & Business Media, 2005.

[25] Ted Belytschko, Yun Yun Lu, and Lei Gu. Element-free galerkin methods. *International journal for numerical methods in engineering*, 37(2):229–256, 1994.

[26] John Dolbow and Ted Belytschko. An introduction to programming the meshless element free galerkin method. *Archives of computational methods in engineering*, 5(3):207–241, 1998.

[27] C O Arun, B N Rao, and S M Srinivasan. Continuum damage growth analysis using element free galerkin method. *Sadhana*, 35(3):279–301, 2010.

[28] Peter Lancaster and Kes Salkauskas. Surfaces generated by moving least squares methods. *Mathematics of computation*, 37(155):141–158, 1981.

[29] Petr Krysl and Ted Belytschko. Analysis of thin plates by the element-free galerkin method. *Computational Mechanics*, 17(1):26–35, 1995.

[30] Petr Krysl and Ted Belytschko. Analysis of thin shells by the element-free galerkin method. *International Journal of Solids and Structures*, 33(20-22):3057–3080, 1996.

[31] C M Tiago and Vitor M A Leitao. Analysis of free vibration problems with the element-free galerkin method. In *Proceedings of the 9th International Conference on Numerical Methods in Continuum Mechanics*, 2003.

[32] L Liu, L P Chua, and D N Ghista. Element-free galerkin method for static and dynamic analysis of spatial shell structures. *Journal of sound and vibration*, 295(1):388–406, 2006.

[33] Ehsan Bahmyari, Mohammad Mahdi Banatehrani, Mohammad Ahmadi, and Marzieh Bahmyari. Vibration analysis of thin plates resting on pasternak foundations by element free galerkin method. *Shock and Vibration*, 20(2):309–326, 2013.

[34] G R Liu and K Y Yang. A penalty method for enforce essential boundary conditions in element free galerkin method. In *Proc. 3rd HPC Asia*, volume 98, pages 715–721, 1998.

[35] Eduard Ventsel and Theodor Krauthammer. *Thin plates and shells: theory: analysis, and applications*. CRC press new york, 2001.

[36] Stephen P Timoshenko and Sergius Woinowsky-Krieger. *Theory of plates and shells*. McGraw-hill, 1959.

# An Investigation of Matching 3D Crushed Sand Particles

**†Mengmeng Wu* and Jianfeng Wang**

Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China.

*Presenting author: mengmwu2-c@ad.cityu.edu.hk

†Corresponding author: mengmwu2-c@ad.cityu.edu.hk

## Abstract

Characterizing the fracture surfaces within a single sand particle precisely and describing it quantitatively plays an essential role in understanding the breakage behavior of sands. This paper presents two novel methods to obtain the realistic fracture surface from 3D particle fragment reassembly and the application of the point cloud registration technique. In the first phase of this research, a fracture region matching algorithm was developed to reassemble fractured sand particles by using a variety of image processing and matching techniques including the *minimum spanning tree* (*MST*), *simple chordless cycles* (*CSC*) and modified *4-points congruent set* algorithms (*4PCS*), etc. [1]. In that study, image matching was only performed between fragments and the original particle was not involved in the matching exercise. In the second phase, the target of research was set to match the individual fragments to their mother particle directly, which entailed matching the original face (if any) on the fragment surface to the mother particle and then identifying the fracture face on a fragment surface that is generated from the mechanical crushing event. That target was achieved by employing the *Standard Iterative Closest Point* (*Standard ICP*) algorithm in the 3D point cloud registration [2]. The effectiveness and efficiency of the tracking methods were demonstrated using the tomography data of 9 crushed Leighton Buzzard sand particles.

**Keywords:** Particle fracture; Tracking; Particle crushing; Registration

## 1 Introduction

Particle crushing plays an important role in the macroscopic mechanical behaviour of sands. A significant amount of research work (e.g., physical experiment and numerical simulation) has been done to investigate the crushing of sand particles. In spite of these research progress, difficulties still exist in accurately depicting the 3D fracture surfaces resulting from particle crushing in a mathematical way, and no method has been developed so far for identifying and tracking individual fragments generated during a particle crushing event.

In recent years, some efforts have been dedicated to the development of discrete particle tracking methods [3] [4]. These methods mainly rely on the use of particle volume for particle tracking. Very recently, Cheng and Wang [5] extended the method proposed by Andò et al. [4] by adopting both particle surface area and particle volume as the matching criterion for particle tracking. However, this approach suffers from the use of a small search window which may only be valid in the case of small sample deformation, and the searching efficiency could decrease dramatically when the sample is subject to a large deformation. Furthermore, it can only handle problems in which sand particles experience no or very little crushing so that the particle volume and surface area almost remain unchanged. As an alternative, Zhou et al. [6] incorporated a set of spherical harmonics invariants into the development of a novel particle tracking method, but again it will be highly complicated and challenging for the method to be employed in the matching of crushed particles. Most recently, the authors have achieved a

success in reassembling fractured LBS particles by using a fracture region matching algorithm, but this method cannot be directly used to match a fragment to the mother particle from which the fragment is generated.

The objective of this paper is to propose two methods for reassembling sand particles and registering 3D point clouds that were crushed and scanned by μCT. In the first phase of this research: based on the work of Zhao et al. [7], particle fragments were separated from each other, and every fragment was converted to a point cloud. According to the magnitude of the surface curvature calculated from the second fundamental form, the points with a higher curvature was extracted. Then, we used the *minimum spanning tree* (*MST*) to connect all these points into a curve network. The *bottom-up graph pruning* algorithm was essential for removing the short branches from the tree. To improve the work efficiency, all simple chordless cycles (*CSC*s) were identified from the *MST*. After that, the *Hausdorff distance* and the modified *4PCS* algorithms were adopted to identify potential *CSC*s for matching and to match them. The registered result which was evaluated by the *ray-triangle intersection* algorithm to avoid the substantial penetration effect provides further information about the potential broken region within the sand particle.

In the second phase, the point cloud registration technique which has been widely used in many areas, including computer vision, medical diagnosis and archaeology, etc., is adopted for quantitatively characterizing the fracture surface and matching a fragment to the mother particle. The *Standard Iterative Closest Point* (*Standard ICP*) technique, proposed by Besl and McKay [8], is the most well-known algorithm among the numerous registration methods for efficiently registering 3D point clouds. The *Standard ICP* is used in this study to mathematically characterize the surfaces of fragments of 9 LBS particles that were subjected to single particle crushing tests with in-situ μCT scanning and then match these child particles to their mother particle.

## 2 The first phase: reassembling sand particles

A number of μCT images of a fractured particle with a resolution of about 3 μm were obtained by using the GE Phoenix v|tome|x m. The CT data visualised as a stack of μCT images, could not be utilised directly for image reconstruction because of the existing noise. Performing image processing on digital images is an essential way to gain a precise description of fracture patterns. Scripting language was compiled to separate the fragments according to the magnitude of volume for the convenience of performing image registration. It shall be noted that not all the fragments were used in the image registration due to the difficulty of finding the right feature curve networks of the fragments as they get smaller and the goal of obtaining the principal fracture surfaces.

In this study, an OBJ file which consists of point and face information was read by the MATLAB software to obtain curvature at a point on a curve. Aiming at extracting the feature curve network the points with high mean curvatures were selected based on the distribution of curvature. Then, we used the *Prim's* algorithm to build a *MST* by connecting the extracted points in the first step. *MST* is a subset of the edges of (un)directed graph that forms a tree that includes every vertex, where the total weight of all the edges of the tree is minimized. the *bottom-up graph pruning* algorithm was used to remove short branches. The purpose of data segmentation is to obtain all *CSC*s. The *CSC*, which is a set of points in which a subset cycle of those points does not exist, can be extracted. Aiming at improving the work efficiency, the *Hausdorff distance* method was adopted to identify the similarity between two *CSC*s. Then the combination with the largest degree of similarity is prioritised for matching by sorting.

### *2.1 Modified 4PCS algorithm*

A modified *4PCS* algorithm was used to express the invariant and to obtain four-points wide bases *S* from chordless cycle *C*. It should be noted that the set of four-points wide bases *S* (*S* for chordless cycle *C*) is a group of four points which must satisfy the following two conditions: 1) the four points are coplanar; and 2) the quadrilateral composed by four points is the convex quadrilateral.

The main steps of the $S_1$ ($S_1$ for simple chordless cycles $C_1$ in fragment 1) search algorithm are: First, three points *A*, *B*, *C* were arbitrarily chosen from $C_1$. Note the fact that any three points in the space can form a plane. However, we need to eliminate the possibility that the three points selected are collinear. Then we selected the fourth point *D* beyond these three points in the set of *CSC* and determined whether the fourth point and the first three points were coplanar. If the four points are coplanar, we could then build a four-points wide base $X = \{A, B, C, D\}$. Next, the crossover point *O* of these two diagonals (*AC* and *BD*) is determined. If the following conditions are true, we can conclude it is a convex quadrilateral: A 5-dimensional descriptor vector $v_1 = \{l_1, l_2, \theta, \eta_1, \eta_2\}$ was constructed to depict the four-points wide base $X = \{A, B, C, D\}$, where $l_1=\|AC\|$, $l_2=\|BD\|$, $\theta$ is the angle between them, $\eta_1=\|AO\|/\|AC\|$, $\eta_2=\|BO\|/\|BD\|$.

The main steps of the $S_2$ ($S_2$ for simple chordless cycles $C_2$ in fragment 2) search algorithm are: We first determined the set of all point pairs $P_a = \{P_{a,1}, P_{a,2}, P_{a,3}, \ldots, P_{a,m}\}$ in $C_2$ (*m* is the number of point pairs), in which the distance between each pair of points was equal to $l_1$. Likewise, another set of point pairs $P_b= \{P_{b,1}, P_{b,2}, P_{b,3}, \ldots, P_{b,n}\}$ in $C_2$ (*n* is the number of point pairs) in which the distance between each pair of points was equal to $l_2$ was also determined. All the point pairs in set $P_a$ and set $P_b$ are alternately combined to form a set of four-points wide bases, $S_c= \{S_{c,1}, S_{c,2}, S_{c,3}, \ldots, S_{c,h}\}$, in which the number of all combinations was $h = m \times n$. Then the same algorithm used for $S_1$ search described above was used for $S_c$ search. Lastly, we could construct another 5-dimensional descriptor vector $v_2$.

### *2.2 Results and discussion*

In this section, we show the results of the fragment reassembly of 4 LBS particles tested by Zhao et al. [7] in which the results of fracture pattern and morphology evolution were demonstrated. The 4 particles are denoted as LBS-1, LBS-2, LBS-3 and LBS-4, where the numbering of particles follows Zhao et al. [7]. We only show the result of reassembly of LBS-1 particle, which has an initial volume of 2.18 $mm^3$ (Fig. 1). The crushing of LBS-1 resulted in a few hundred fragments, from which 16 fragments were successfully reassembled, making it the most successful case out of the 4 LBS particles. The volume of the reassembled LBS-1 is about 97.7% of the original volume. The smallest fragment volume matched successfully is 0.007 $mm^3$ and equal to 0.32% of the original particle volume. The shapes of all 16 fragments are shown in Fig. 1(e). Figs. 1(a)-1(d) indicate that LBS-1 was reassembled accurately, demonstrating again the high capacity of the proposed matching algorithm. Apparently, this fracture mode is different from those of other three particles and is resulted from the combined influence of overall less spherical shape (sphericity 0.80), unsmooth surface, little initial void and loading direction.

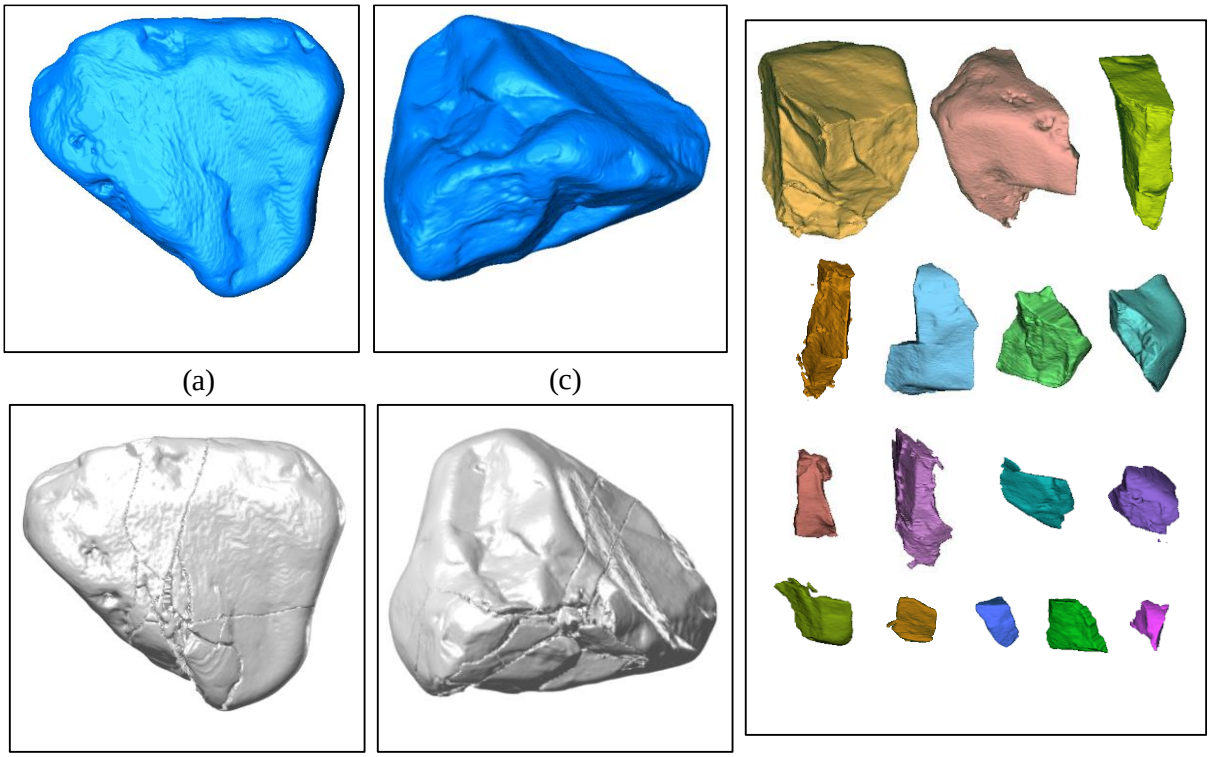



**Fig. 1. LBS-1 model: (a, c) the original particle shape; (b, d) the reassembly result; (e) 3D fragment model**

## 3 The second phase: point cloud registration

In this study, another 5 LBS particles were tested and scanned, in the same manner as Zhao et al. [7] using the μCT system (v|tome|x m, Phoenix|X-ray, General Electric Company (GE)) of Shanghai Yinghua NDT Equipment Trade Co., Ltd. The 5 LBS particles were randomly selected and had a size between 1.2 and 1.6 mm (due to the load capacity of the apparatus), which is not considered to have a significant effect on the crushing behavior. The limited CT resolution will create some difficulties for the matching algorithm, particularly for those very tiny fragments which tend to have similar morphologies. In fact, this is the main reason for choosing LBS particles for this study, which have relatively smooth surfaces and textures, reducing the difficulty level for the matching exercise. Other kinds of rough sand particles, for example, highly decomposed granite (HDG), are purposely avoided, due to its high surface roughness, which becomes intractable for the fragment identification and matching. More than 500 raw images (i.e., 2D slices) were obtained for each scan, containing a considerable amount of noises and cannot be directly used for image analysis. The voxel size of these images is 5.69 μm. The resolution of the CT scan was not changed during the scan. More details of the image processing can also be found in Zhao et al. [7].

### *3.1 Child particle surface segmentation*

In recent years, the point cloud registration is a popular topic and has been widely used in the driverless vehicle, medical diagnose and archaeology, etc. Although the technique has received a strong development and assisted us considerably in solving graphics-related problems, shortcomings like its being sensitive to the noise and hard to recognize the useless points still exist in its working process. For this study, points consisting of the fracture face of a child

particle, regarded as useless points, will affect the matching efficiency and accuracy, because there are no corresponding points to them on the mother particle surface. Therefore, a data pre-processing step, called the *contour-based mesh segmentation*, whose function is to partition the child particle surface into several faces, is necessary in the point cloud registration process. However, it shall be noted that distinguishing the fracture face from the original face in advance is highly challenging. As a result, it is necessary to try matching each face of a child particle to the mother particle surface and then use an index called the distance error to evaluate the matching degree and eliminate the wrong matching. More details of the procedure of *contour-based mesh segmentation* can also be found in Rodrigues et al. [9].

*3.2 Iterative closest point algorithm*

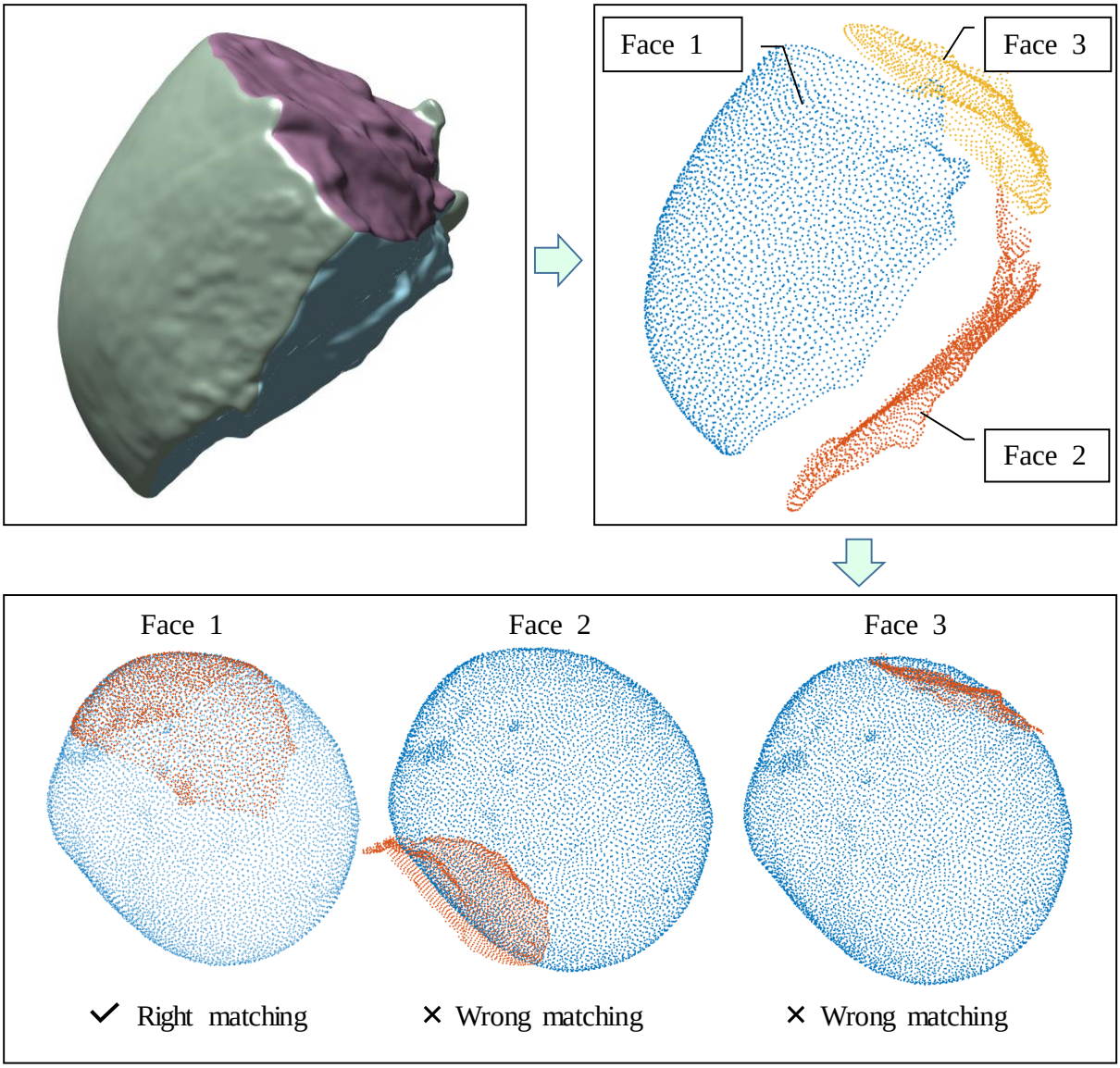



**Fig. 2. The relationship between *contour-based mesh segmentation* and the *Standard ICP***

In the point cloud registration, the order of child-mother particle image matching is determined by the volume of the child particle. The face with the largest area on a given child particle is prioritized for the point cloud registration. The *Standard ICP* algorithm, which is an optimal registration method based on *Least Squares Method*, was adopted in this paper. The core idea of this algorithm is to make trial selections of corresponding points repeatedly and calculate an optimal rigid body transformation until the convergence of matching is satisfied. Accordingly,

the first step in the *Standard ICP* algorithm is to find the initial corresponding points from the source cloud and the target cloud, respectively. Then in the second step, the rotation and translation matrices are calculated and applied to the source cloud, which results in the transformation of the corresponding points. In the last step, the calculation process returns to the part of adjusting and reselecting the correspondence points. These steps are iteratively performed until the termination condition is satisfied. The performance of the *Standard ICP* was tested by registering the point cloud of each face on a given child particle to the point clouds of the mother particle, as shown in Fig. 2.

*3.3 Results and discussion*

9 LBS particles are examined and divided into two groups, namely Group A and Group B in this section. Group A involves 4 LBS particles tested by Zhao et al. [7] in which the results of fracture pattern and morphology evolution were demonstrated. The 4 particles are denoted as LBS-1, LBS-2, LBS-3 and LBS-4, where the numbering of particles follows Zhao et al. [7] Group B includes 5 LBS particles which were tested in this study and labelled as LBS-5, LBS-6, LBS-7, LBS-8 and LBS-9. We only show the results of reassembly of LBS-1 and 5 particles (Figs. 3 and 4).

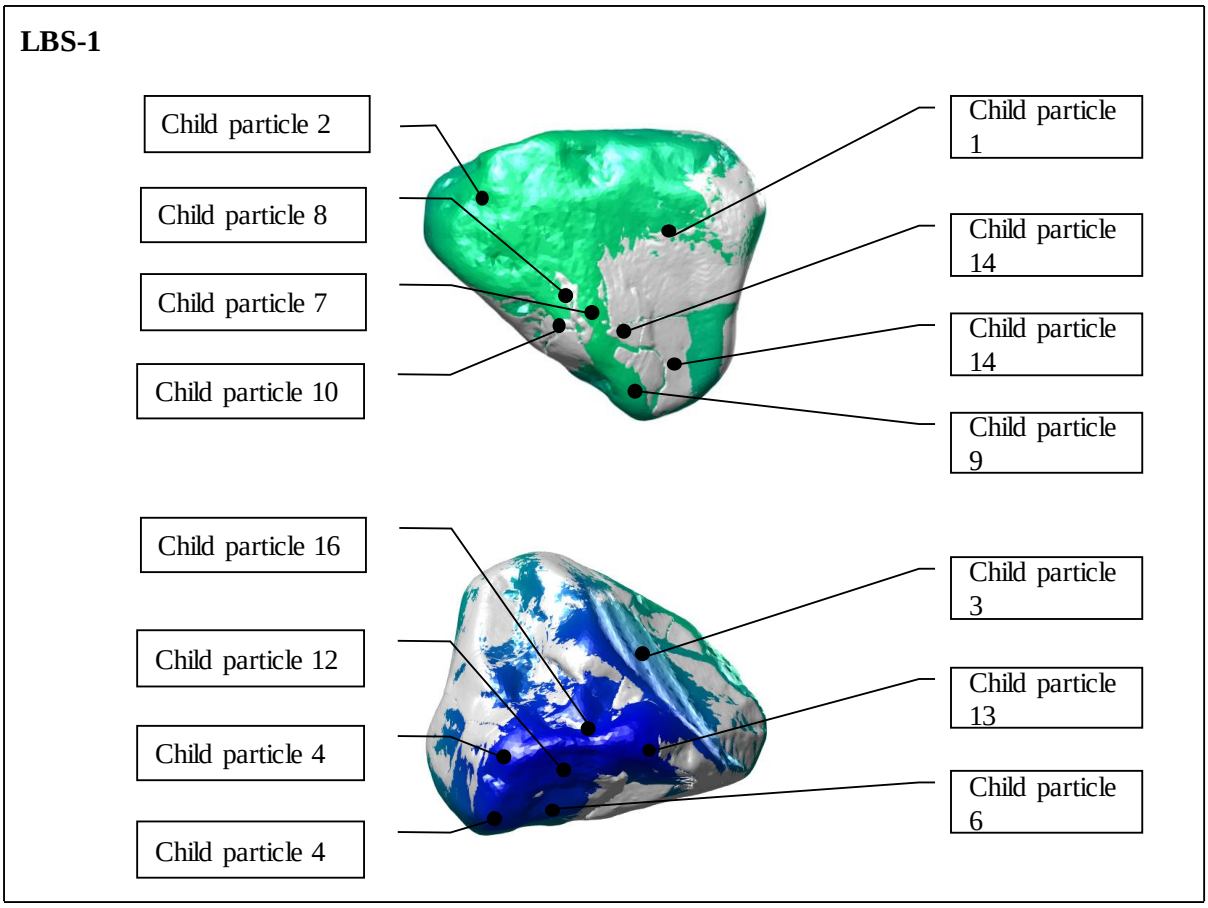



**Fig. 3. Group A: the original and fracture face information for a given child particle (The grey color expresses the child particle and the mother particle is presented by the rest color.)**

The point cloud registration can identify the locations of child particles in the mother particle, and upon the successful matching of more and more child particles, can restore the morphology of the mother particle. However, some limitations still exist in this algorithm. It will not be successful if there is no original face on the child particle. It will also be a challenge if the

original faces for two given child particles have similar morphologies. In addition to the above backwards, calculation errors are unavoidable in this algorithm, making it difficult to match the child particle with the mother particle completely. Fig.4 shows further the matching results with colour-labeled particle surfaces.

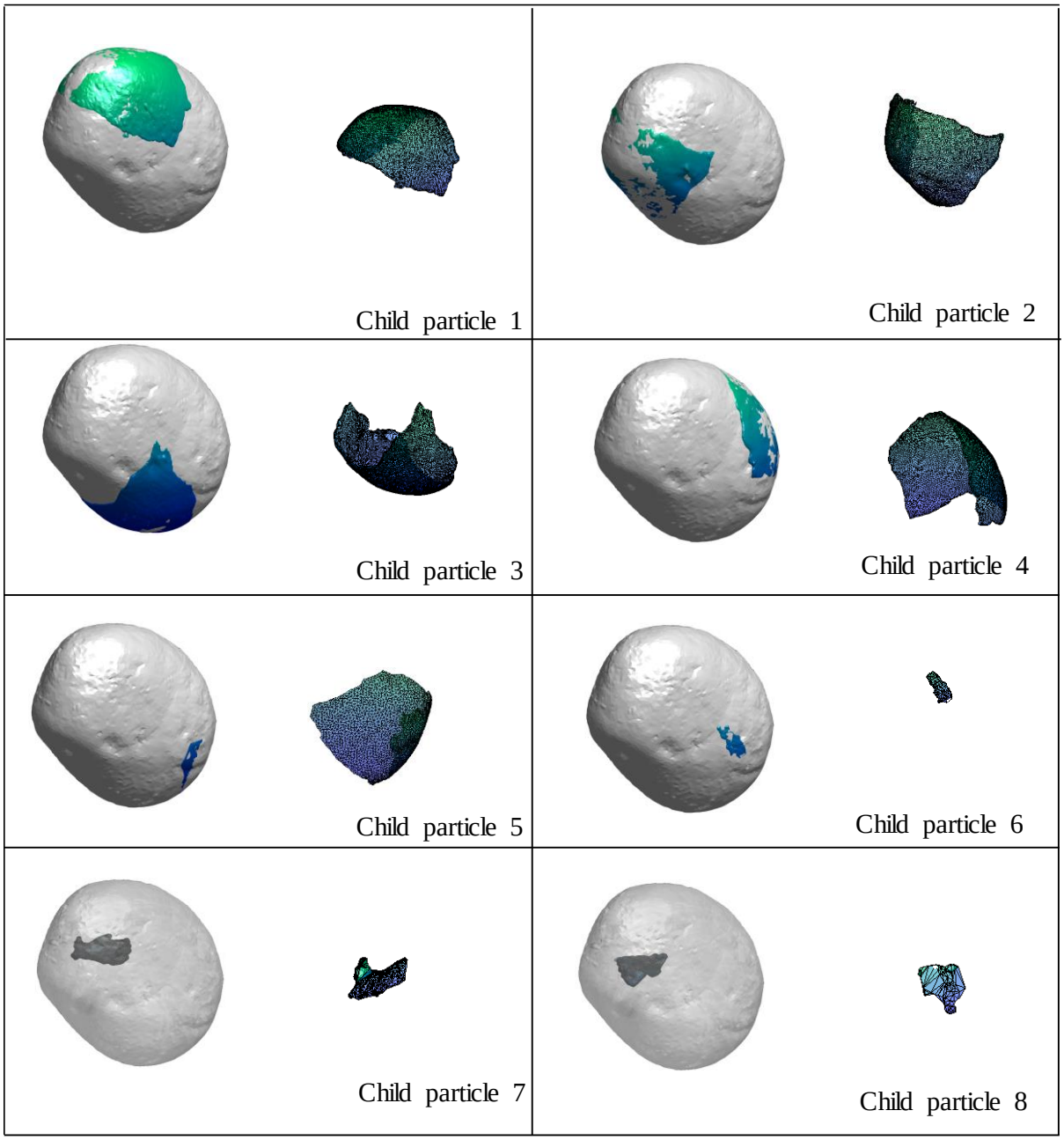



**Fig. 4. The results of image matching between child particle and mother particle (The grey color expresses the mother particle and the child particle is presented with green color. The mesh expresses the region with higher point density. Translucent state exists in the child particle 7 and 8 for better observing.)**

## 4 Concluding remarks

The main contribution of this paper is the proposal of two innovative algorithms for reassembling fractured sand particles based on the results of μCT scanning of single sand particle crushing tests.

For the first phase, the matching algorithm, which has been widely used in medical diagnosis, computer vision, archaeology, forensic investigations and other related fields, was applied to the micromechanical study of sands for the first time, making this study a pioneering one in the investigation of the 3D particle morphology restoration upon particle breakage. Although only four LBS particles were reassembled in this study, the results of particle reassembly demonstrated the high capability and robustness of the proposed algorithm, and contributed to our further understanding of the fracture pattern and morphology evolution during the sand particle crushing process. It serves as a starting point for the further investigation of the fracture mechanics of sands.

For the second phase, the point cloud registration method, which has been widely used in many areas such as image processing, computer vision, machine vision and medical diagnosis, etc., was applied for the first time to characterize fractured sand particles and match them to the original particles. This was achieved by implementing the *Standard ICP* algorithm to segment and identify the original faces and fracture faces of the child particles resulting from the crushing of LBS particles, which were subjected to the single particle crushing test and scanned by the μCT. 9 LBS particles were successfully reassembled using the above technique, demonstrating the high competence and robustness of the technique in quantifying the fragment morphologies and matching them to the original particles. Our next goal is to enhance the algorithm to enable the automatic ID-tracking of fractured particles within a loaded sand sample.

**References**

[1] Wu, M. and Wang, J. (2018), "Reassembling fractured sand particles using fracture-region matching algorithm", *Powder Technology*, Vol. 338, pp. 55–66.

[2] Wu, M. and Wang, J. (2019), "Registration of point cloud data for matching crushed sand particles", under review.

[3] Kingston, T.A., Geick, T.A., Robinson, T.R. and Heindel, T.J. (2015), "Characterizing 3D granular flow structures in a double screw mixer using X-ray particle tracking velocimetry", *Powder Technology*, Vol. 278, pp. 211–222.

[4] Andò, E., Hall, S.A., Viggiani, G., Desrues, J. and Bésuelle, P. (2012), "Grain-scale experimental investigation of localised deformation in sand: A discrete particle tracking approach", *Acta Geotechnica*, Vol. 7 No. 1, pp. 1–13.

[5] Cheng, Z. and Wang, J. (2018), "A particle-tracking method for experimental investigation of kinematics of sand particles under triaxial compression", *Powder Technology*, Vol. 328, pp. 436–451.

[6] Zhou, B., Wang, J. and Wang, H. (2018), "A novel particle tracking method for granular sands based on spherical harmonic rotational invariants", *Géotechnique*, pp. 1–8.

[7] Zhao, B., Wang, J., Coop, M.R., Viggiani, G. and Jiang, M. (2015), "An investigation of single sand particle fracture using X-ray micro-tomography", *Géotechnique*, Vol. 65 No. 8, pp. 625–641.

[8] Besl, P.J. and McKay, N.D. (1992), "A method for registration of 3-D shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14 No. 2, pp. 239–256.

[9] Rodrigues, R.S.V., Morgado, J.F.M. and Gomes, A.J.P. (2015), "A contour-based segmentation algorithm for triangle meshes in 3D space", *Computers and Graphics (Pergamon)*, Vol. 49, pp. 24–35.

# Parallel Computing for Trajectory Prediction of Aircraft

**†Momoha Nishimura[1], Masashi Yamakawa[1], Ryuta Sakashita[1], Shinichi Asao[2] and Mitsuru Tanaka[1]**

[1]Department of Mechanical and System Engineering, Kyoto Institute of Technology, JAPAN.
[2]Department of Mechanical Engineering, College of Industrial Technology, JAPAN.

†Corresponding author: d7821007@edu.kit.ac.jp

## Abstract

GPU parallel computing was applied to a trajectory prediction of an aircraft. An aerobatic maneuver was simulated by a coupled method of 6-DOF motion and MCD method. Because the simulation contained a complex interaction of grid movement and flow dynamics and thus it was quite expensive, the acceleration by GPU was attempted to demonstrate its ability compared to CPU parallelization. The GPGPU code was constructed by OpenACC because of the directive-based programming. The GPU computing accomplished a remarkable speedup, which confirmed that GPGPU is useful for the acceleration of this system.

**Keywords:** CFD, Parallel Computing, GPGPU, OpenACC, Flight Simulation.

## Introduction

The numerical flight simulation of an aircraft is an important problem in Computational Fluid Dynamics (CFD) for aerospace. With this system called the Digital Flight [1], aerodynamic coefficient of an airplane can be safely calculated even in a risky flight without real aircrafts. Moreover, a risk of stall can be predicted when the flight is simulated within a real time.

To achieve the flight simulation based on physics, the authors have proposed the Moving Computational Domain (MCD) method [2], which is one of the moving grid methods based on the Moving-grid Finite-Volume method [3]-[5]. In this method, the computational domain itself with an aircraft inside moves by following a movement of the airplane, therefore any restrictions of the computational domain for three-dimensional space can be removed [6][7]. Recently, we have integrated the six-degrees-of-freedom (6-DOF) motion to this CFD method [8][9], and simulated unconstrained motions of objects influenced by fluid and the motion. As applications of this coupled method to a flight simulation, various aerobatics of an airplane were computed. The rotation of the propeller and the moving control surface were installed in this system as well to simulate flight as if pilot operated the aircraft.

In this paper, parallel computing on graphics processing unit (GPU) is introduced towards the calculation within an actual time. Although OpenMP or MPI is generally adopted to shorten the calculation time [10], general-purpose computing on GPU (GPGPU) has been recently in the spotlight. GPU has a number of cores, thus GPUs could calculate much faster than CPU. Although major examples for GPGPU are NVIDIA's CUDA or OpenCL, much time and labor is required for code rewriting. Therefore, we employed OpenACC which can accelerate using compiler directives for our first coding on GPU to see whether GPGPU can help for faster computing compared to OpenMP.

## Numerical Scheme

*Flow Solver*

The governing equations are the three-dimensional Euler equations for compressible flow written in the conservation form (1),(2) as follows:

$$\frac{\partial \mathbf{q}}{\partial t}+\frac{\partial \mathbf{E}}{\partial x}+\frac{\partial \mathbf{F}}{\partial y}+\frac{\partial \mathbf{G}}{\partial z}=0, \tag{1}$$

$$\mathbf{q}=\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ \rho w\\ e\end{bmatrix},\mathbf{E}=\begin{bmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ \rho uw\\ u(e+p)\end{bmatrix},\mathbf{F}=\begin{bmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ \rho vw\\ v(e+p)\end{bmatrix},\mathbf{G}=\begin{bmatrix}\rho w\\ \rho uw\\ \rho vw\\ \rho w^2+p\\ w(e+p)\end{bmatrix}. \tag{2}$$

The equations are discretized by the Moving-grid Finite-Volume method with four-dimensional control volume combined with time and space. The inviscid flux vectors are estimated by Roe's flux difference splitting [11] at the interfaces. MUSCL approach and Venkatakrishnan limiter [12] are employed to provide second order accuracy in space. The 2-stage rational Runge-Kutta method is applied as a time stepping scheme.

*6-DOF Motion*

Because the aircraft is treated as a rigid body in this paper, 6-DOF equations of motion govern the flight of the aircraft. The Newton's equation of motion (3) is applied to the translation of the mass center, and the Euler's rotation equation (4) is applied to the rotation in body axes. The rotational motion is calculated in the body-fixed axis subscripted with B. Here quaternion [13] is used to avoid gimbal-lock. 6-DOF equations are integrated into the inner iteration of flow solver as a strong coupling method [14].

$$\frac{d\mathbf{p}}{dt}=\mathbf{f} \tag{3}$$

$$\frac{d\mathbf{L}_B}{dt}+\boldsymbol{\omega}_B\times\mathbf{L}_B=\mathbf{N}_B \tag{4}$$

$$\mathbf{p}=\begin{bmatrix}m\dot{r}_x\\ m\dot{r}_y\\ m\dot{r}_z\end{bmatrix},\mathbf{f}=\begin{bmatrix}f_x\\ f_y\\ f_z-mg\end{bmatrix},\mathbf{L}_B=\begin{bmatrix}I_{x_B}\omega_{x_B}\\ I_{y_B}\omega_{y_B}\\ I_{z_B}\omega_{z_B}\end{bmatrix},\boldsymbol{\omega}_B=\begin{bmatrix}\omega_{x_B}\\ \omega_{y_B}\\ \omega_{z_B}\end{bmatrix},\mathbf{N}_B=\begin{bmatrix}N_{x_B}\\ N_{y_B}\\ N_{z_B}\end{bmatrix} \tag{5}$$

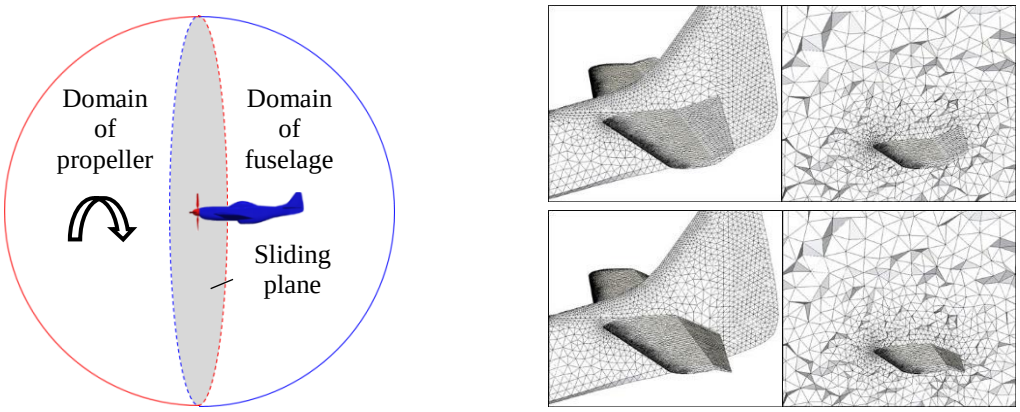
## Grid Movement and Deformation

*MCD Method*

When it comes to the movement of the whole grid in a large area, Moving Computational Domain (MCD) method is applied. The computational domain itself with an object inside moves by following the moving aircraft.

*Grid Deformation*

The rotation of the propeller and the motion of the control surface are installed by sliding-mesh approach and tension-torsion spring analogy [15] (Fig. 1(a)). To apply the sliding mesh technique, computational domain is divided into two domains, one is the fuselage domain and the other is the propeller domain. The conservative quantities are interpolated at the interface where tetrahedral grids overlap each other. When the control surface is moved, grids are deformed by tension-torsion spring analogy (Fig. 1(b)).




**(a) Sliding-mesh method** **(b) Tension-torsion spring analogy**

**Figure 1. Grid movement and deformation**

## Application

*Calculating Condition*

P-51 propeller aircraft model with 1,612,350 unstructured grid points are generated by MEGG3D [16][17] are shown in Fig. 2. The center of gravity is assumed to be located in 25%MAC (Mean Aerodynamic Chord), and the moment of inertia is generated by the engine, the fuel tank and the skin of the airplane.

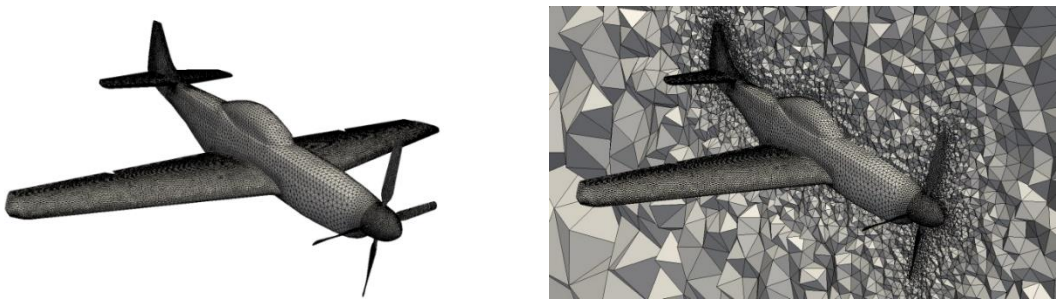
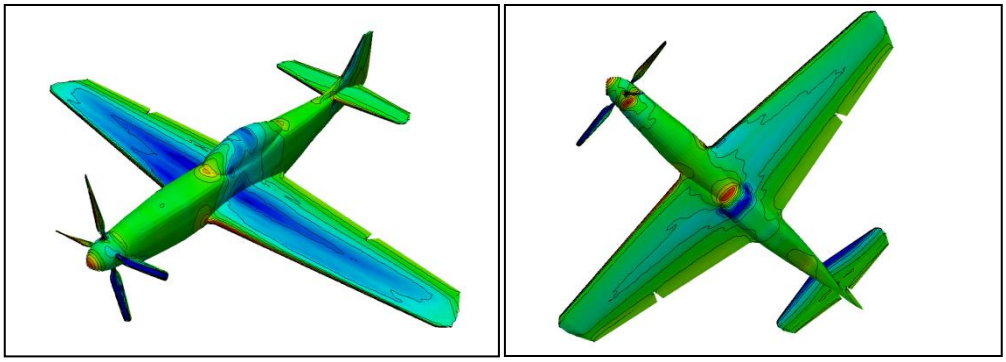


**Figure 2. Computational grid**

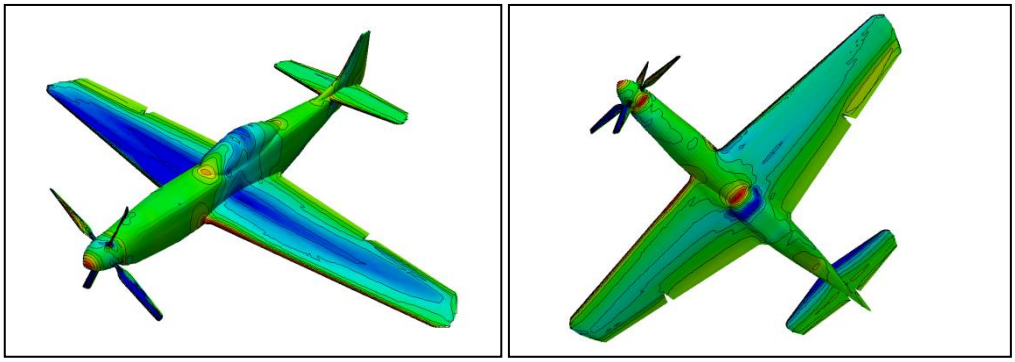
The straight flight, two clockwise and two counterclockwise aileron roll are simulated with this system. At first, the aircraft performs straight flight to avoid the initial turbulence at $V = 0.45$, where the speed of sound is 1.0. After that, to complete the clockwise rolling, the right aileron is manipulated up to 10 degrees, and left aileron is manipulated down to 10 degrees. The rudder is also controlled to avoid adverse yaw. In counter-clockwise rolling, the ailerons and rudder are operated oppositely.

## Results

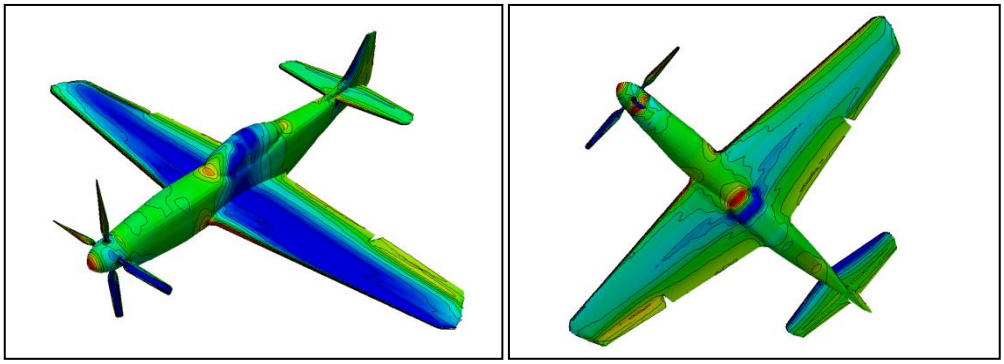
The surface pressure distributions at the forward straight flight, clockwise rolling and counter-clockwise rolling are shown in Fig. 3. It can be seen in Fig. 3(a) that the rotating propeller generated the vortex and pressure distribution on main wing became asymmetry. After ailerons are operated, the pressure distribution at the main wing changed significantly, as seen in Fig. 3(b) and (c). The pressure at the bottom side of the main wing is lower than the upper side, which yields the rolling moment to complete the aileron roll.



**(a) Forward flight**



**(b) Clockwise flight**




**(c) Counterclockwise flight**

**Figure 3. Pressure contours in the aileron roll**

## GPU Parallel Computation

For high-speed calculations, we constructed the GPU calculation code of this system by OpenACC. The computer for GPU computing (PC1) has CPU of Intel® Core™ i7-3930K Processor and GPU of GIGABYTE6.0 V-NTITANBLKGHZ-6GD-B GEFORCE GTX TITAN BLACK. The operating system is Windows7 64bit and the compiler is PGI Accelerator Fortran Workstation. For comparison of the calculating speed, this system is calculated only with CPU (PC2). The computer, which has the highest performance for CPU we could use, has Intel® Core™ i7-3930K Processor. The operating system is Cent OS 6.3, and the compiler is Intel Composer XE 2013.

We constructed the FORTRAN code accelerated by OpenACC and measured the calculating speed of this system on 100 steps 10 times with PC1 and PC2. The speedup is defined as follows:

$$\text{Speedup} = \frac{\mathrm{T_{PC1}}}{\mathrm{T_{PC2}}} \tag{6}$$

where T represents the calculation process time. With the GPU parallelization, we achieved 13.6x speedup over the serial CPU solver. In contrast to the fact that approximately 2x speedup can be achieved with OpenMP, GPU computing accomplished the significant speedup. The flight simulation so far was carried out for some maneuvers only because of its long calculation time. However, with GPU computing, it is expected that flight simulation from take-off to landing can be calculated in future.

## Conclusions

In this study, the coupled computational method of the unstructured MCD method and six-degrees-of-freedom flight dynamics was constructed, and GPU parallelization accelerated this system. In the simulation of aerobatic maneuver, the ailerons and rudder are operated, and then pressure distribution at main wing changed, which completes the aileron roll. Then, GPGPU by OpenACC was carried out to achieve the high-speed computation. GPU acceleration attained 13.6x speedup over the serial CPU computing. It is confirmed that GPU computing is effective for the acceleration of this system.

## Acknowledgments

This publication was subsidized by JKA through its promotion funds from KEIRIN RACE and by JSPS KAKENHI Grant Number 16K06079.

### References

[1] M. D. Salas, “Digital Flight: The Last CFD Aeronautical Grand Challenge”. Journal of Scientific Computing, Vol.28, Nos. 2/3, pp. 479-505, 2017.

[2] K. Watanabe and K. Matsuno, “Moving Computational Domain Method and Its Application to Flow Around a High-Speed Car Passing Through a Hairpin Curve”, Journal of Computational Science and Technology, Vol.3, No.2, pp. 449-459, 2009.

[3] K. Mihara, K. Matsuno and N. Satofuka, “An Iterative Finite−Volume Scheme on a Moving Grid (1st Report, The Fundamental Formulation and Validation)”, Transactions of the JSME (in Japanese), 65-637, No. 99-0060, pp. 2945-2953, 1999.

[4] M. Yamakawa and K. Matsuno: “An Iterative Finite−Volume Method on an Unstructured Moving Grid (1st Report, The Fundamental Formulation and Validation for Unsteady Compressible Flows)”, Transactions of the JSME (in Japanese), 69-683, No. 02-1277 pp. 1577-1582 ,2003.

[5] M. Yamakawa, D. Takekawa, K. Matsuno and S. Asao, "Numerical Simulation for a Flow around Body Ejection using an Axisymmetric Unstructured Moving Grid Method", Computational Thermal Sciences, Vol.4, No.3, pp.217–223, 2012.

[6] S. Asao, K. Matsuno and M. Yamakawa, "Simulations of a Falling Sphere with Concentration in an Infinite Long Pipe Using a New Moving Mesh System", Applied Thermal Engineering, Vol. 72, pp.29–33, 2014.

[7] S. Asao, K. Matsuno and M. Yamakawa, "Simulations of a Falling Sphere in a Long Bending Pipe with Trans-Mesh Method and Moving Computational Domain Method", Journal of Computational Science and Technology, Vol.7, No.2, pp.297–305, 2013.

[8] S. Asao, K. Matsuno and M. Yamakawa, "Parallel Computations of Incompressible Fluid-Rigid Bodies Interaction Using Transmission Mesh Method", Computers & Fluids, Vol.80, pp.178–183, 2013.

[9] S. Asao, K. Matsuno and M. Yamakawa, "Parallel Computations of Incompressible Flow Around Falling Spheres in a Long Pipe Using Moving Computational Domain Method", Computers & Fluids, Vol.88, pp.850–856, 2013.

[10] M. Yamakawa, Y. Kita and K. Matsuno, "Domain decomposition method for unstructured meshes in an OpenMP computing environment" Computers & Fluids, Vol. 45, pp.168–171, 2011.

[11] Roe, P. L: "Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes", Journal of Computational Physics, 43 pp. 357–372, 1981.

[12] V. Venkatakrishnan: "On the Accuracy of Limiters and Convergence to Steady State Solutions", 31st Aerospace Sciences Meeting and Exhibit, AIAA 93-0880, 1993.

[13] M. Yatabe: "Handy Note for Quaternions", MSS Technical Report (in Japanese), Vol.18, pp. 29–34, 2007.

[14] K. Kawamura, Y. Ueno and Y. Nakamura: "Flight Simulation of Taketombo Based on Computational Fluid Dynamics and Computational Flight Dynamics", Journal of the JSASS (in Japanese), Vol. 56, No. 654, pp. 324–330. 2008.

[15] M. Murayama, K. Nakahashi and K. Matsushima: "Unstructured Dynamic Mesh for Large Movement and Deformation", 40th AIAA Aerospace Sciences Meeting & Exhibit, AIAA 2002-0122, 2002.

[16] Y. Ito and K. Nakahashi: "Surface Triangulation for Polygonal Models Based on CAD Data", International Journal for Numerical Methods in Fluids, Vol. 39, Issue 1, pp. 75–96, 2002.

[17] Y. Ito: "Challenges in Unstructured Mesh Generation for Practical and Efficient Computational Fluid Dynamics Simulations", Computers & Fluids, Vol. 85 pp. 47–52, 2013.

# Shape Optimization of 3D Diaphragm Pumps using the Continuous Adjoint Approach to the Cut-Cell Method

***Konstantinos Samouchos[1], Dimitrios Kapsoulis[1], Xenofon Trompoukis[1], and Kyriakos Giannakoglou[1]**

[1]Parallel CFD & Optimization Unit, National Technical University of Athens, Greece.
[*]Presenting and Corresponding author: ksamouchos@gmail.com

## Abstract

This paper is concerned with tools that can be used for the design/optimization of 3D valveless and valved diaphragm pumps. These pumps comprise a main chamber, an inlet and outlet duct and a periodically moving diaphragm that causes the unsteady/periodic fluid flow within the device; the inlet and outlet ducts might either be valveless diffusers or tubes of constant cross-section with valves (valved pumps). Obtaining the desired net mass-flow rate per period with the minimum (hopefully zero) back-flow at the exit (especially, in the absence of valves) are the usual objectives. The minimization of viscous losses within some pump's components such as inlet or outlet valved ducts could be an extra target. Regarding CFD analysis, apart from the cut-cell method mentioned in the title, an in-house flow solver using body-fitted unstructured grids is also used. Both codes are based on the finite volume approach by sharing the same numerical features and run on GPU-clusters (the former) and on many-processor platforms (the latter). To support gradient-based optimization, with either CFD tool, the continuous adjoint method computing the gradient of the objectives with respect to the design variables has been developed and programmed. According to the proposed parameterization scheme, all design variables are related to the diaphragm geometry and its periodic motion. An alternative optimization method that was also made available for the needs of this paper is a (gradient-free) evolutionary algorithm, assisted by metamodels (on-line trained artificial neural networks) and the principal component analysis for efficiently searching the design space. Using this tool, many-objective optimizations can be performed to compute the front of non-dominated solutions. This paper focuses on particular features of the analysis and optimization tools for this type of diaphragm pumps, including micropumps for medical applications.

**Keywords:** Cut-cell method; Adjoint method; Diaphragm pump; Valved pump; Evolutionary Algorithm.

## Introduction

Valveless or valved diaphragm pumps are formed by an inlet and outlet duct as well as a chamber with a periodically moving diaphragm controlling the fluid's flow. Depending on the application, they are often preferred over bladed pumps since they can pump various types of fluids in a noiseless manner. They are manufactured in large or small scales, with the large (usually valved) ones used for cleaning tank bottoms or pumping sewage, while the small (usually valveless) ones (micropumps) used as medical analysis devices, in biochemical-processing applications or to deliver drug to patients [1]. The inlet and outlet ducts might be tubes of constant cross-section equipped with valves of different types (valved pumps) preventing back-flow. In medical applications and drug injections using valveless (micro)pumps, the exit flow must be

uninterruptible and, in some cases, the valves are replaced by diffusers [2, 3]. Unfortunately, during their operation, back-flow might instantaneously occur at the outlet. Therefore, obtaining the desired net volume flux per period with minimum (hopefully zero) back-flow at the exit are standard objectives/constraints during design-optimization. In some cases, keeping viscous losses low in pump's components is an additional objective.

Regarding the CFD analysis of the diaphragm pumps this paper is dealing with, two in-house codes are used to model the periodic viscous flow within the pump. The first one, referred to as PUMA [4], uses body-fitted unstructured grids, continuously adapted to the moving diaphragm, using the spring analogy method. The software is GPU-enabled and uses the MPI protocol to perform on a many-GPU platform. However, in cases where large boundary movements occur or two bodies approach and finally touch each other, such as in valved pumps, grid deformation/adaptation to the changing boundaries becomes costly, delicate or even impossible. For these reasons, a flow solver based on the cut-cell variant [5] of the general class of Immersed Boundary Methods (IMB) [6] has been developed, too. Starting point is a Cartesian grid enclosing the flow domain. The background (coarse) grid remains stationary while the immersed solid bodies are allowed to move covering and uncovering grid cells. Grid is continuously refined close to the moving geometry, to increase the accuracy of the flow simulation. The two CFD codes are based on the finite volume approach and share the same numerical features.

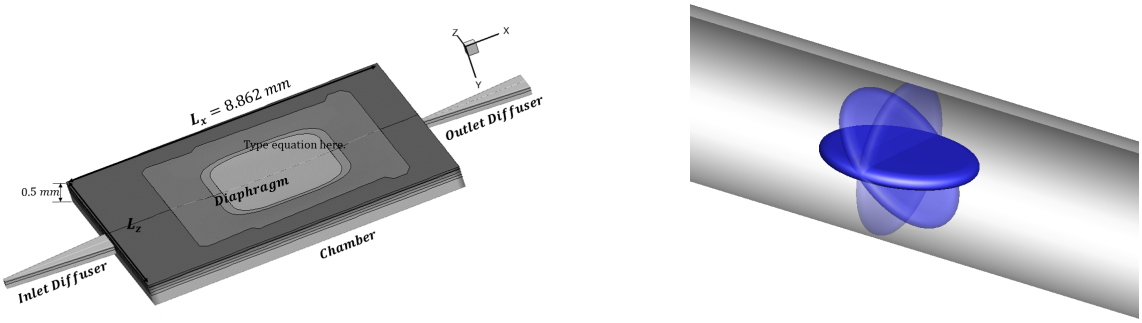
For the optimization, the pump's diaphragm motion must be parameterized. The values of the design variables minimizing the selected objective functions are the unknowns in the optimization problem. The manufacturability of the diaphragm and the mechanism controlling its motion is beyond the scope of this paper.

The GB method is supported by the continuous adjoint method that computes the gradient of the objective function w.r.t. the design variables. The main advantage of the adjoint method is that its cost is independent of the number of the design variables. In case of more than one objectives, these are concatenated to a single function to be minimized. However, since the computation of the Pareto front of non-dominated solutions is of great value for the designer, an Evolutionary Algorithm (EA) is additionaly used. The EA is assisted by locally trained metamodels and the Kernel Principal Component method reducing the optimization's computational cost, which is the main drawback of EAs [7].

In this paper, CFD analyses and optimizations with either code and either optimization method, for diaphragm pumps or their components (such as inlet/outlet tubes equipped with a rotating disc-shaped shutter) are demonstrated.

## Description/Parametrization of the Diaphragm Pumps

Fig. 1a shows the valveless micropump studied in this paper. Its geometry is based on an existing micropump found in the literature [2]. Regarding the valved pump it is assumed that they use cylindrical inlet and outlet ducts with a constant cross-section equipped with valves, fig. 1b, in place of the diffusers used in the valveless devices. The micropump's scale is of millimeters, with $1cm$ length and chamber's volume $\sim 40mm^3$. Valved pump's dimensions can be much larger, since this kind of devices are used for different purposes. Due to manufacturing reasons, the inlet and outlet diffusers or ducts and valves are identical. The pressure distribution inside the chamber and its volume is periodically changing by moving the diaphragm, this motion being essential for the pump's functionality. Technically, a piezoelectric device moves the diaphragm with a predetermined frequency, by pushing it from the inlet to the outlet causing the fluid to flow.




**(a)** **(b)**

**Figure 1: Left: Valveless diaphragm micropump with diffusers at the inlet and outlet. Right: Valved (butterfly valve) duct replacing the diffusers. The rotating disc, at three different positions, is shown in blue. The disc is always present in the flow, so it induces a total pressure drop, even when open.**

The diaphragm's motion is parameterized, defining thus the design variables for the ensuing optimizations. The resting position of the diaphragm lays on the $y=0$ plane, where the $x$ and $z$ axes point in the longitudinal and span-wise directions, respectively. The first two design variables ($b_1$, $b_2$) define a rectangular area limiting the moving part of the diaphragm, fig. 1. The maximum displacement $y_{max}(t)$ over all time instants is given by $y_{max}(t) = b_3 exp\big(-b_4(t-T)^2\big)\big(1-|1-\frac{2t}{T}|\big)$ where $T=0.02\ sec$ is the period, $b_3$ controls the maximum overall displacement, which is much smaller than the chamber's height, achieved at the half period and $b_4$ controls the abruptness of the exponential function. The longitudinal deformation over time is defined by $y(x,t) = y^{max}(t)(6\tau_x^2 - 8\tau_x^3 + 3\tau_x^4),\ \tau_x = \frac{x+\delta x}{Dx}$, where $\delta x = b_5 min(x + L_x/2, L_x/2 - x)$ and $Dx = (1-b_6)\delta x$, with $L_x$ and $L_z$ being the total length and width of the chamber. Similarly, the span-wise diaphragm's deformation follows a similar parameterization. The design variables $b_7$ and $b_8$ define $\delta z = b_7 min(z + L_z/2, L_z/2 - z)$ and $Dz = (1-b_8)\delta z$, which are required to compute $y(z,t) = y_{max}(t)(6\tau_z^2 - 8\tau_z^3 + 3\tau_z^4),\ \ \tau_z = \frac{z+\delta z}{Dz}$. Different time instants of the deformation of the arbitrarily selected motion along the symmetry plane can be seen, later on, in fig. 11.

### Flow Equations and Discretization

The 3D grid generator developed and used by the cut-cell method is likely the most challenging and important part of any cut-cell software. The implementation of an octree data structure makes grid generation fast with low memory requirements. Starting point is a uniform Cartesian grid covering the whole computational space including also the solid bodies (i.e. the pump's boundaries). Then, each cell intersected by the pump's boundaries is subdivided into eight sub-cells. The process is repeated until a minimum cell volume is reached. During the grid refinement process, each cell should have at most four neighbouring cells per face, preventing this way the creation of big cells next to much smaller ones. The computation of the exact intersection of cells with the pump's boundaries is based on the Sutherland-Hodgman algorithm [8]. The part of the cell inside the solid body is discarded and the remaining part creates a polyhedral finite volume (fig. 2a). Special treatment is needed whenever a cell splits into two or more polyhedra, which are treated as different finite volumes (fig. 2b).

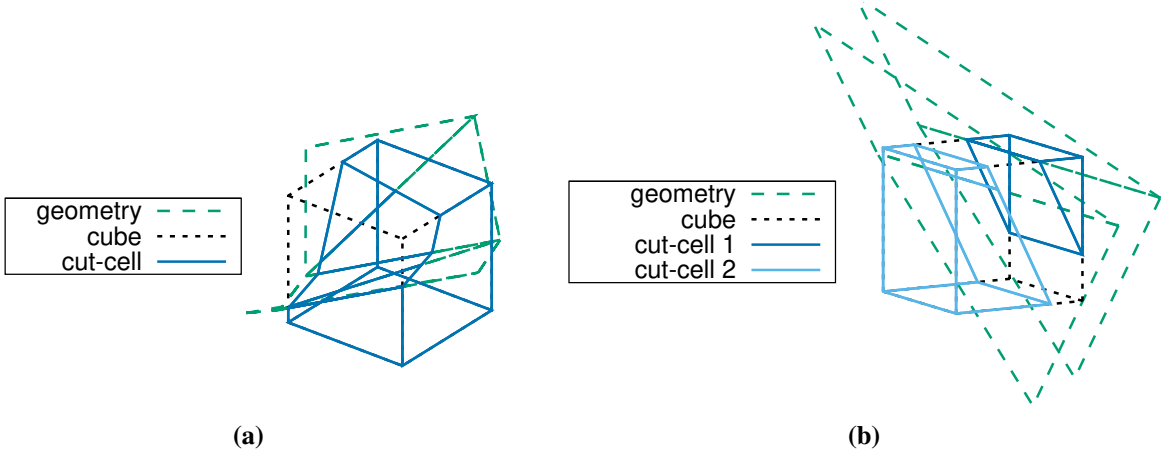



**Figure 2: (a) A cube (black dashed line) is cut by the triangulated solid surface (green dashed line) creating a cut-cell (blue line). (b) A cube is separated into two different cut-cells. The flow and adjoint equations are solved only in the part of the cube lying inside the fluid domain.**

A known downside of the cut-cell method is the formation of very small cut-cells leading to instabilities during the flow and adjoint numerical solution. A cell-merging technique is applied to surpass this difficulty. According to that, small cells are merged with bigger neighbours, named master-cells, creating hyper-cells. Geometrical criteria are used for the selection of the master-cell among the neighbours of a small-cell. In specific, for all neighbouring cells with volume greater than a threshold value, the one with the largest common face is chosen. If the first criterion fails, the neighbour with the biggest volume is chosen as the master-cell. Generally, more than two cells can be merged to form a hyper-cell.

PUMA [4] uses an unstructured grid with hexahedral elements, the generation of which will not be elaborated here, in the interest of space.

Flow prediction in a diaphragm pump requires the numerical solution of the unsteady incompressible Navier-Stokes equations. These are written as,

$$\Gamma_{ij}^{-1}\frac{\partial V_j}{\partial \tau}+\frac{\partial U_i}{\partial t}+\frac{\partial f_{ik}^{inv}}{\partial x_k}-\frac{\partial f_{ik}^{vis}}{\partial x_k}=0, \quad i=1,4, \quad \Gamma=\begin{bmatrix}\beta^2 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{1}$$

where the Einstein notation is used. Index $k=1,2,3$ refers to the Cartesian directions; $(x_1,x_2,x_3)$ stand for $(x,y,z)$ and $(u_1,u_2,u_3)$ to the corresponding Cartesian velocity components, $\vec{V}=\begin{bmatrix}p & u_1 & u_2 & u_3\end{bmatrix}^T$, $\vec{U}=\begin{bmatrix}0 & u_1 & u_2 & u_3\end{bmatrix}^T$, $\vec{f}_k^{inv}=\begin{bmatrix}u_k & u_1u_k+p\delta_k^1 & u_2u_k+p\delta_k^2 & u_3u_k+p\delta_k^3\end{bmatrix}^T$ and $\vec{f}_k^{vis}=\begin{bmatrix}0 & \tau_{1k} & \tau_{2k} & \tau_{3k}\end{bmatrix}^T$, where $\delta_i^j$ is the Kronecker symbol, $p$ is the pressure divided by the density, $t$ is the real time, $\tau$ is the pseudo-time and $\tau_{ik}=\nu(\frac{\partial u_i}{\partial x_k}+\frac{\partial u_k}{\partial x_i})$ are the viscous stresses. Finally, $\Gamma$ is the preconditioning matrix used to stabilize the numerical solution, depending on $\beta$ which stands for the pseudo-compressibility coefficient. Flows considered in this paper are laminar. In PUMA, the discretization of eqs. (1) is based on the vertex-centered finite volume approach. On the other hand, the cell-centered approach is used in the cut-cell software. In both solvers, inviscid flux $f_{ik}^{inv}$ is computed based on the Roe's approximate Riemann solver, [9].

The diaphragm motion is taken into account in the flow equations' discretization. Both solvers use the Reynolds theorem to compute the "physical" time derivatives, namely, the integration of the time derivative of $U_i$ in a deformable finite volume $\Omega_{FV}$ is given as

$$\begin{aligned} \int\limits_{\Omega_{FV}^{k+1}} \frac{\partial U_i}{\partial t}\, d\Omega &= \frac{d}{dt} \int\limits_{\Omega_{FV}^{k+1}} U_i\, d\Omega - \int\limits_{S_{FV}^{k+1}} u_m^{grid} n_m U_i\, dS \\ &\simeq \frac{1}{2\Delta t}\left(3U_i^{k+1}\Omega_{FV}^{k+1} - 4U_i^{k}\Omega_{FV}^{k} + U_i^{k-1}\Omega_{FV}^{k-1}\right) - \sum_{faces} u_m^{grid} n_m U_i^{k+1} \Delta s \end{aligned} \tag{2}$$

where $\Omega_{FV}^{k+1}$, $\Omega_{FV}^{k}$, $\Omega_{FV}^{k-1}$ is the volume of the same finite volume at time instants $k+1$, $k$ and $k-1$, respectively. $u_m^{grid}$ stands for the grid velocity (i.e. the grid deformation rate) and is computed in a way ensuring the satisfaction of the geometric conservation law. $n_m$ is the unit normal pointing outwards the finite volume and $\Delta S$ the area of the finite volume faces. PUMA morphs the grid at each time instant employing a spring analogy technique, [10]. Fig. 3 shows two grids (used by PUMA) where, on the left, the diaphragm is in its resting position whereas, on the right, the diaphragm is pushed down. Regarding the cut-cell method, the Cartesian grid is

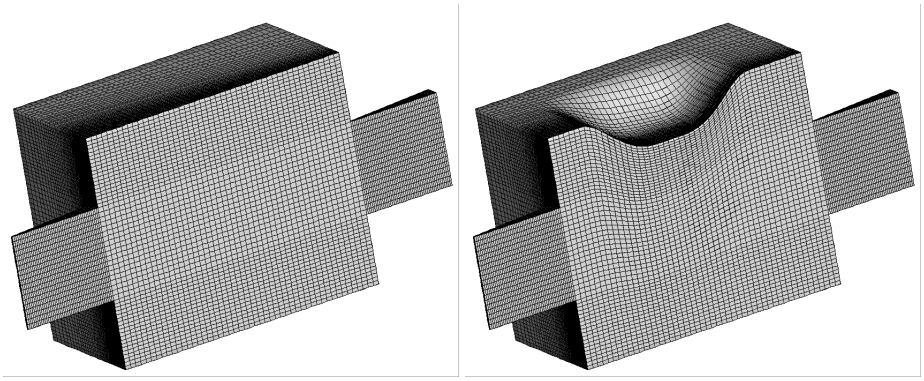
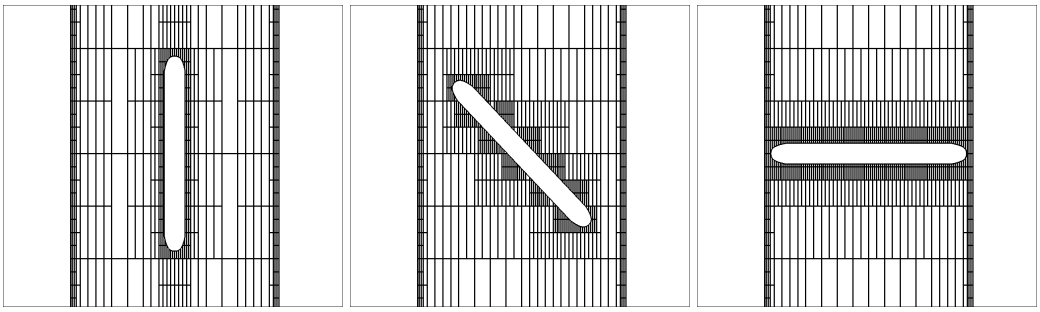
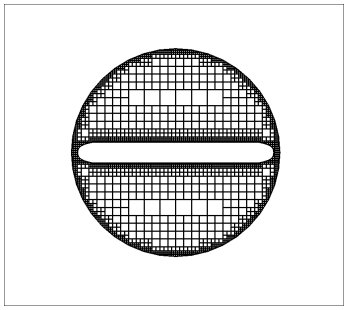


**Figure 3: Perspective views of the unstructured CFD grids used by PUMA code at time instants $0$ and $0.45T$. Axes not in scale, y-axis is 14 times greater.**

continuously adapted to the diaphragm's motion. First, a coarsening procedure applies and cells in the vicinity of the diaphragm are coarsened. Then, cells cut by the new diaphragm's geometry are refined and the neighbouring cells are adapted accordingly. The flow field computed at the previous time instant(s) is transferred to the new grid for computing time derivatives of the flow variables. Transfer is conservative by taking into account that cut-cells change in shape or migrate from the fluid to the solid region. Among other, solidified cells merge their values with those of neighbouring fluid cells. Fig. 4 shows three time-steps while the valve moves from open to closed position. A Cartesian grid, adapted to the valve's motion, is generated anew at each time-step.



**(a)**



**(b)**

**Figure 4: (a) Adapted Cartesian grid for three positions of the butterfly valve. (b) View of the valve in the fully open position.**

### Adjoint Equations and Gradient Computation

Three metrics are used to measure the quality of the resulting flow in a diaphragm pump or its components (valved ducts), namely

- The volume of the fluid entering pump's exit (back-flow, $Q_{bf}$; a non-negative quantity, by definition) within a period $T$. It is defined as $Q_{bf} = -\int_T \int_{S_O} min(0, u_m n_m^O) dS dt$, where $n^O$ is the unit normal to the pump's outlet ($S^O$).
- The net volume of fluid pumped within $T$, i.e. $Q_{net} = \int_T \int_{S_O} u_m n_m^O dS dt$.
- The total pressure losses ($Q_{pt} = p_t^I - p_t^O$) between (valved) duct's inlet and outlet.

Depending on the application, the objective function could be any of these metrics ($F \equiv Q$) or a combination of them; in regard to $Q_{net}$ since the usual target is to achieve a desired volume flow rate, this might give an objective function of the form $F = |Q_{net} - Q_{net,ref}|$. Among other, below, a single-objective optimization of the pump is performed to minimize $F = \frac{Q_{bf}}{Q_{net}}$.

The gradient of $F$ w.r.t. $b_i$ (defined in section ) is computed using the continuous adjoint method. The augmented objective function (to be differentiated instead of $F$) is

$$F_{aug} = F + \int_0^T \int_\Omega \Psi_i R_i d\Omega dt \tag{3}$$

where $R_i$ are the residuals of the flow equations, $\Psi_i$ are the adjoint variable fields, $\Omega$ is the fluid domain at $t$. By differentiating $F_{aug}$ w.r.t. $b_i$ and setting the multipliers of variations in the flow variables to zero, the unsteady adjoint equations along with their boundary conditions arise.

The unsteady adjoint equations are

$$\Gamma_{ji}^{-1}\frac{\partial \Psi_j}{\partial \tau} - \frac{\partial \bar{\Psi}_i}{\partial t} - A_{jik}\frac{\partial \Psi_j}{\partial x_k} - \frac{\partial f_{ik}^{A,vis}}{\partial x_k} = 0 \tag{4}$$

where $\vec{\bar{\Psi}} = \left[0\, \Psi_2\, \Psi_3\, \Psi_4\right]^T$, $\vec{f}_k^{A,vis} = \left[0\, \tau_{1k}^A\, \tau_{2k}^A\, \tau_{3k}^A\right]^T$ stands for the adjoint viscous flux and $\tau_{ik}^A = \nu\left(\frac{\partial \Psi_{i+1}}{\partial x_k} + \frac{\partial \Psi_{k+1}}{\partial x_i}\right)$ are the adjoint stresses. Similarly to the flow equations, a pseudo-time derivative multiplied with the inverse of the preconditioning matrix $\Gamma$ has been added in eq. 4. The integration of the temporal term on deformable finite volumes is similar to the flow equations solver as in eq. 2.

For the pump's wall boundaries (including the diaphragm), the adjoint velocity is set to zero. Given the fact that the inlet/outlet duct shapes do not change in time, the adjoint inlet/outlet condition reads

$$\Psi_i A_{ijk} n_k^{I/O}\frac{\partial V_j}{\partial q_l^{I/O}} + \frac{\partial \hat{F}}{\partial q_l^{I/O}} = 0 \tag{5}$$

where $\hat{F}$ is the integrand of $F$. Working with $Q_{bf}$ in particular, since function "min" cannot be differentiated, this is replaced with a sigmoid function in eq. 5. $q_l^{I/O}$ stands for flow quantities extrapolated from the CFD domain (like the velocity components at the outlet or the velocity magnitude at the inlet).

Considering that period $T$ is constant, the gradient of $F$ w.r.t. $b_i$ becomes an expression of the computed primal and adjoint fields, as follows

$$\begin{aligned}\frac{\delta F}{\delta b_i} = &-\int_0^T\int_\Omega \left[\Psi_n\frac{\partial V_n}{\partial x_l}\frac{\partial}{\partial t}\left(\frac{\delta x_l}{\delta b_i}\right) + \left(\Psi_n\frac{\partial f_{nk}^{inv}}{\partial x_l} - \Psi_n\frac{\partial f_{nk}^{visc}}{\partial x_l}\right)\frac{\partial}{\partial x_k}\left(\frac{\delta x_l}{\delta b_i}\right)\right] d\Omega dt \\ &-\int_0^T\int_\Omega \tau_{km}^{adj}\frac{\partial u_k}{\partial x_l}\frac{\partial}{\partial x_m}\left(\frac{\delta x_l}{\delta b_i}\right) d\Omega dt + \int_0^T\int_{S^W}\left(\Psi_1 n_k - \tau_{km}^{adj} n_m\right)\frac{\delta u_k^{grid}}{\delta b_i} dS dt \\ &+\int_0^T\int_{S^{I/O}}\frac{\partial \hat{F}}{\partial x_l}\frac{\delta x_l}{\delta b_i} dS dt\end{aligned} \tag{6}$$

where $S^I$, $S^W$ stand for the inlet and diaphragm surface, respectively. Note that in the cut-cell method, $\frac{\delta x_l}{\delta b_i}$ takes non-zero values only along the diaphragm.

### PCA-Driven Metamodel-Assisted Evolutionary Algorithms

A $(\mu, \lambda)$ EA, with $\mu$ parents and $\lambda$ offspring in each generation, is the background stochastic optimization tool. Metamodels assist the EA (Metamodel Assisted EA-MAEA), so as to reduce its computational cost. Metamodels are Radial Basis Function (RBF) network trained on-line during the evolution on individuals evaluated on the CFD tool and used, when needed, to predict the objective function value(s) of new individuals at practically negligible cost compared to a CFD run. In each generation (excluding the very first ones, in which there is no enough data to train the metamodels), each population member is pre-evaluated with personalized locally trained metamodels and, then, only the few most promising ones are re-evaluated with the CFD tool. Training patterns are selected, using several criteria, from the pool (database) of

individuals evaluated during the preceding generations. Aiming at maximum performance, the Kernel Principal Component Analysis (PCA) method drives the EA as mentioned in [7]. Briefly, prior to applying the evolution operators, the design space is transformed into the feature space in which the evolution operators perform better and the metamodels are trained only with the significant variables as computed by the method. This EA assisted by metamodels and the PCA, referred to as the PCA-driven MAEA [11] [12] [7], has been developed within the EASY platform [13] of NTUA, is also used for the optimization.

### Method Demonstration in a Valved Inlet/Outlet Duct

The valve is modeled as a disc-shaped shutter rotating around an axis/rod (fig. 1b). Applications like the flow inside a duct with a rotating valve, perfectly suit to the cut-cell CFD software, as this avoids morphing a body-fitted grid (without changing grid topology). Thus, only the cut-cell software is used here to compute the flow field inside the duct while the valve rotates from the open to the closed position. Velocity magnitude and pressure fields, for different valve's positions, are shown in fig. 5, respectively. Flow trajectories around the rotating disc are presented in fig. 6. The unsteady adjoint equations have also been solved giving the necessary information to compute the sensitivity derivatives of the $\Delta p_t$ function, defined between the inlet and the outlet of the duct. A time history of the adjoint velocity magnitude and adjoint pressure fields is shown in fig. 7. 3D views of the adjoint field around the disc-shaped shutter, at different positions is shown in 8.

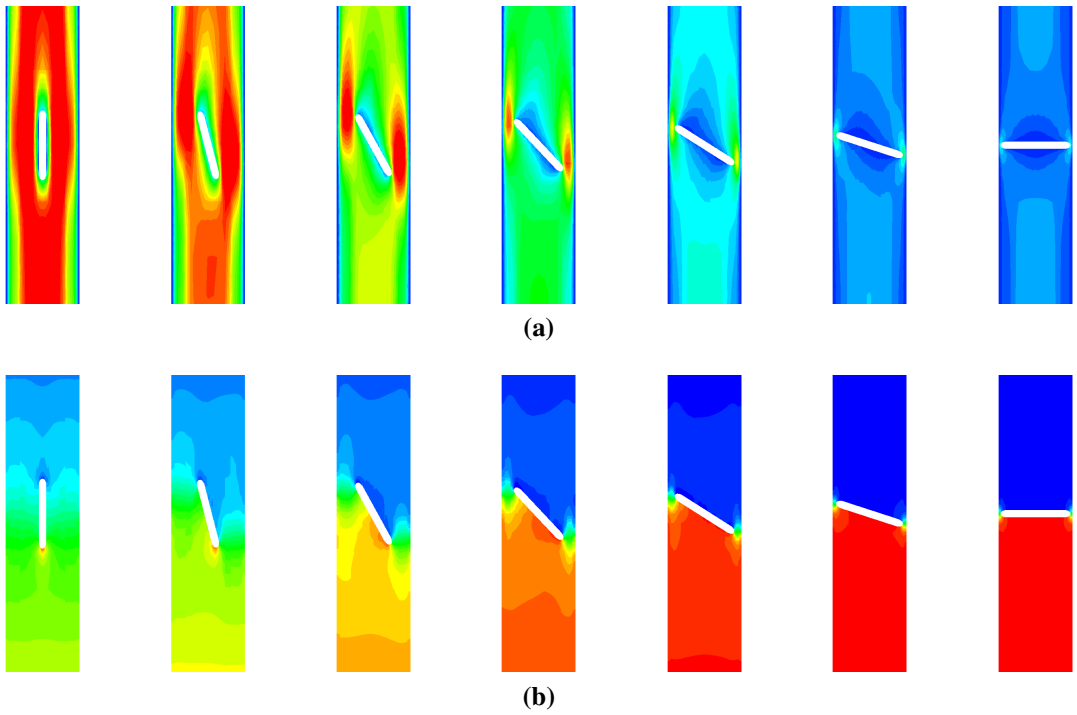



**Figure 5: Top: Instantaneous velocity magnitude fields within the valved duct, at equally distributed time instants. Bottom: Instantaneous pressure fields at the same time instants.**

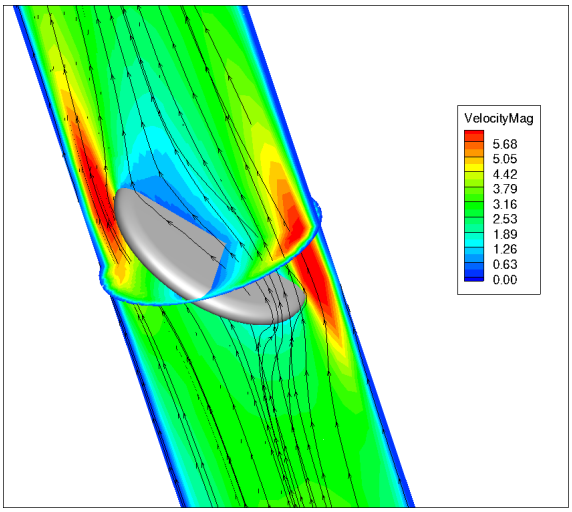
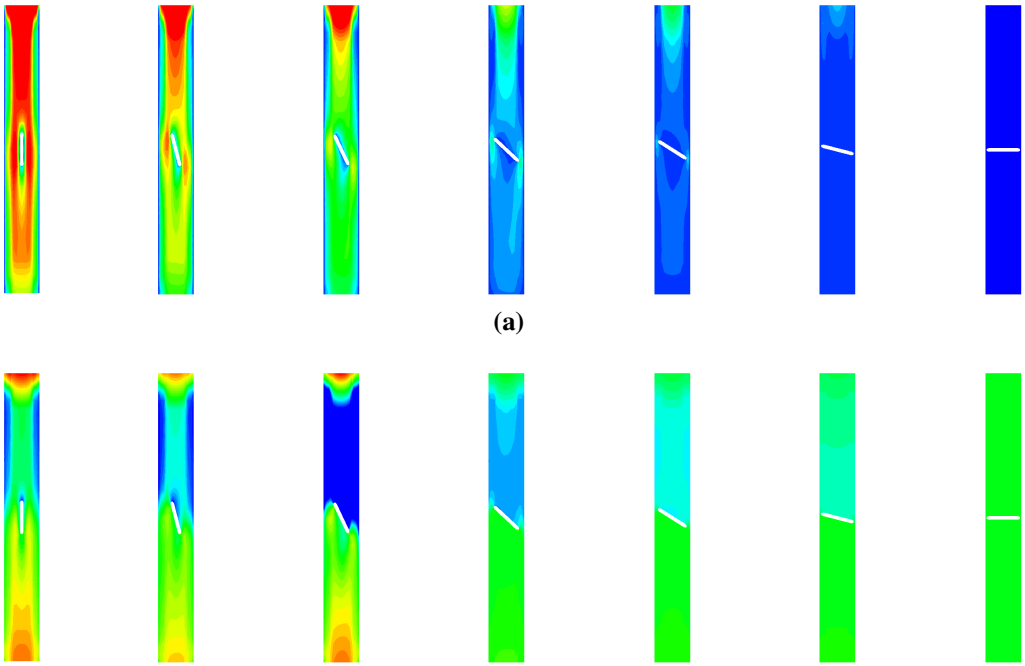



**Figure 6: Flow trajectories within the valved duct with the disc half-open.**



**(a)**

**(b)**

**Figure 7: Top: Instantaneous adjoint velocity magnitude fields within the valved duct, at equally distributed time instants. Bottom:Instantaneous adjoint pressure fields at the same time instants.**

### Method Demonstration in a Valveless Micropump

A first valveless micropump (the way this has been constructed is beyond the scope of this paper) is optimized at first; this pump (to be referred to as the "reference" one; index "ref") delivers the desired net volume of fluid per period ($T = 0.02s.$), with a non-negligible backflow rate though. Therefore, it was decided to run a two-objective optimization aiming at minimum $|Q_{net} - Q_{net,ref}|$ and minimum $Q_{bf}$. Next to this, two single-objective optimizations are also performed, both with the same target; one of them is based on the PCA-driven MAEA and the other on the GB method supported by the adjoint solver. It should be noted that, in the single-objective optimizations, the (common) target was to minimize $\frac{Q_{bf}}{Q_{net}}$; thus, any comparison with

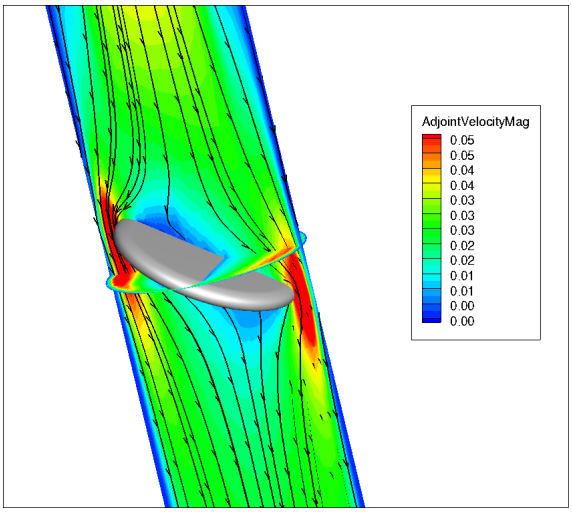



**Figure 8: Adjoint flow trajectories (for $\Delta p_t$ as the objective function) within the valved duct with the shutter half-open.**

the two-objective optimization is made only for the purpose of comparing computational cost and not the quality of the optimal solutions obtained by the various methods.

The PCA-driven MAEA is configured with $\mu = 6$ and $\lambda = 12$ and both metamodels and PCA are activated after the first generation. The computational budget is limited to 200 CFD runs. The convergence history of the single-objective optimization is shown in fig. 10. For the single-objective problem, the resulted optimal motion has reduced the objective function $\left(\frac{Q_{bf}}{Q_{net}}\right)$ by two orders of magnitude from the initial/reference one. Regarding the two-objective optimization, the final front of non-dominated solutions is presented in fig. 9. Notice that the optimal solution of the first optimization is the bottom right end of the front.

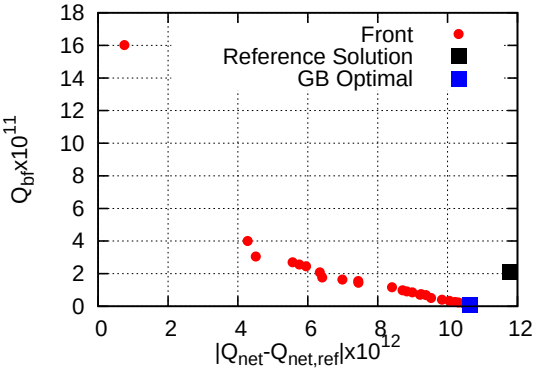



**Figure 9: Front of non-dominated solutions (red points) computed by the PCA-driven MAEA in the two objective optimization. Reference (black point) and the optimal (found by the GB optimization) solution (blue point) are included, too.**

In regard to the GB optimization, in which the cost of the solution of the adjoint equations is, on the average, equal to the cost of solving the flow equations, each optimization cycle approximately costs as much as two CFD calls. Steepest-descent is used with an adaptive step size per optimization step. The convergence of the GB optimization is shown in fig. 10. After 36 CFD calls, this practically leads to the same optimal solution with the one found by the

PCA-driven MAEA for the same objective. Note that the PCA-driven MAEA could have been stopped earlier since, during the last generations, the optimal solution does not change.

Regarding the flow quality metrics for the GB optimization, the net volume flow is increased from $Q_{net,ref} = 5.18mm^3$ for the reference motion to $Q_{net} = 6.23mm^3$ for the optimal one and the back-flow is reduced from $Q_{bf,ref} = 2.12mm^3$ to $Q_{bf} = 0.255mm^3$. The optimal and reference diaphragm's motion are shown at various time instants in figs. 11, 12, respectively. Fig. 13 shows that the optimal micropump has reduced the internal vortical flow in the last time instants and the maximum deformation of the diaphragm has been limited which assists at minimizing the back-flow at the exit.

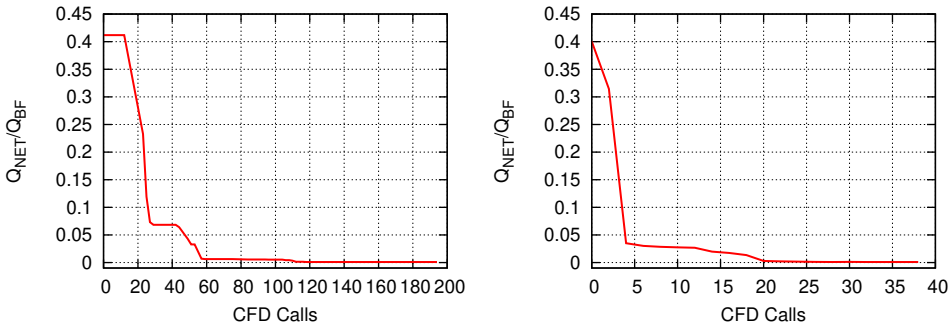



**Figure 10: Convergence history of the micropump's optimization. (Left) PCA-driven MAEA, (right) Gradient-Based method.**

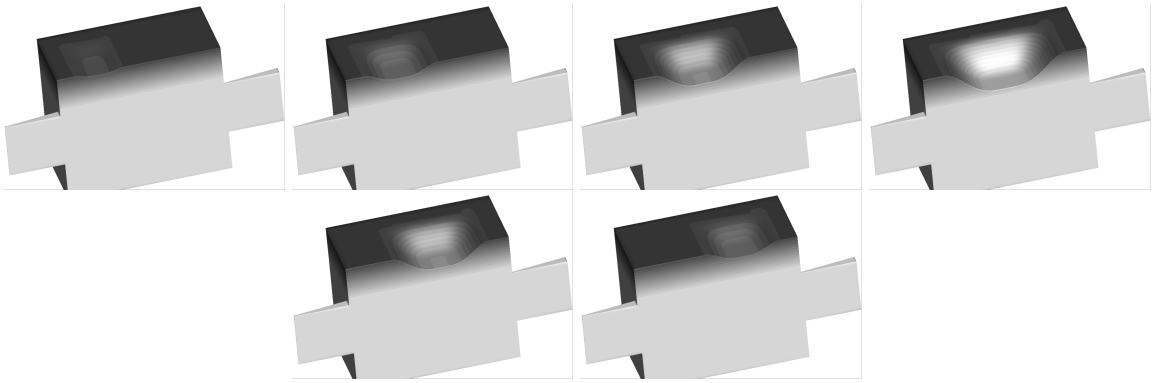


**Figure 11: Reference diaphragm motion at the symmetry plane. Time instants:** $0.20T$**,** $0.30T$**,** $0.40T$**,** $0.50T$**,** $0.60T$ **and** $0.70T$**, from top-left to bottom-right respectively.**

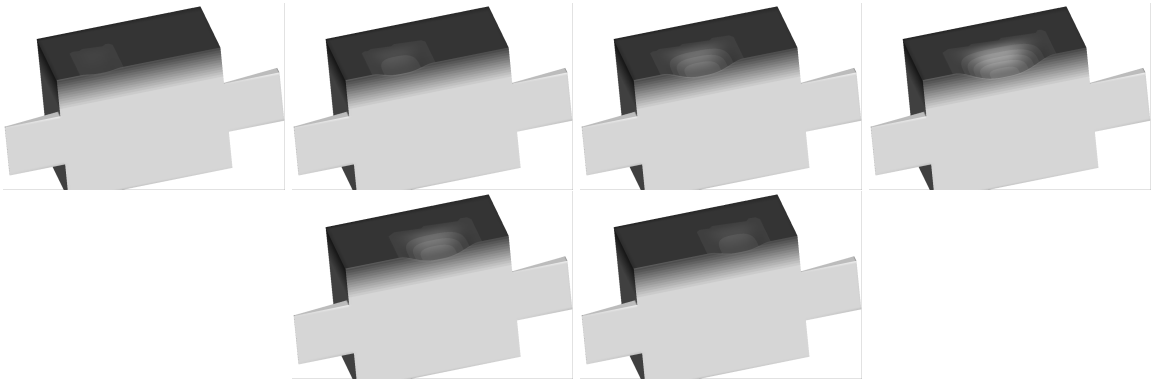


**Figure 12: Optimal diaphragm motion at the symmetry plane. Time instants:** $0.20T$**,** $0.30T$**,** $0.40T$**,** $0.50T$**,** $0.60T$ **and** $0.70T$**, from top-left to bottom-right respectively.**

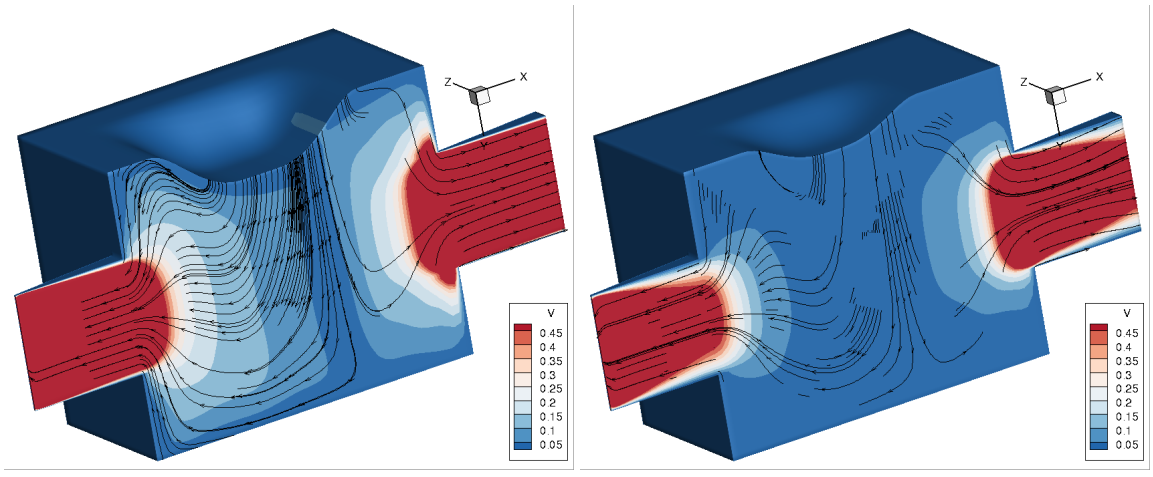



**Figure 13: Velocity magnitude iso-areas on the symmetry plane at time instant** $0.45T$**. Reference (left) and optimized (right) motion. Axes not in scale,** $y$**-axis is 14 times greater.**

## Conclusion

This paper focuses on particular features of the analysis and optimization tools for valveless and valved diaphragm pumps. For both pump types, the diaphragm motion is parameterized and metrics quantifying the quality of the resulting flow inside the pump or its components are defined. Performing the optimization through gradient-free methods, their computation is all we need. In this paper, the gradient-free optimization is performed using a metamodel-assisted evolutionary algorithm, additionally supported by the principal component analysis of the population, for the purpose of cost reduction. For the use of gradient-based optimization, though, the objective functions in use should also be differentiated; for this purpose, the continuous adjoint method is employed. From the CFD viewpoint, it is very convenient to possess more than one tools. Our first CFD tool is based on a cut-cell method, being more appropriate in case of valves that periodically open and close. Next to this, a standard CFD solver with body-fitted unstructured grids is used. For both codes, the same adjoint methods for the same objective functions have been developed. Some selected analyses and optimizations have been performed and demonstrated the capabilities of the available tools. Their integration in an automatic workflow, which will involve both optimization methods and both CFD tools, is in progress.

## Acknowledgment

Research presented in this paper has been funded by the Business Plan "Development of Human Resources, Education and Lifelong Learning" entitled "Support Researchers with Emphasis on Young Researchers" with the co-financing of Greece and the European Union. The project title is "*Design-Optimization of Diaphragm Pumps under Operational/Manufacturing Uncertainties using the Cut-Cell Method and Polynomial Chaos Expansion*".

## References

[1] H. Chun-Wei, H. Song-Bin, and L. Gwo-Bin, "A microfluidic device for precise pipetting," *Journal of Micromechanics and Microengineering*, vol. 18, no. 3, pp. 35–39, 2008.

[2] L. Songjing, J. Liu, and D. Jiang, "Dynamic characterization of a valveless micropump considering entrapped gas bubbles," *Journal of Heat Transfer*, vol. 135, 2013.

[3] K. Samouchos, D. K. Kapsoulis, X. Trompoukis, and K. Giannakoglou, "Design of a diaphragm pump under uncertainties using the continuous adjoint to the cut-cell method.," in *6th European Conference on Computational Mechanics (ECCM 6) - 7th European Conference on Computational Fluid Dynamics (ECFD 7)*, June 11-15 2018.

[4] I. Kampolis, X. Trompoukis, V. Asouti, and K. Giannakoglou, "CFD-based analysis and two-level aerodynamic optimization on graphics processing units," *Computer Methods in Applied Mechanics and Engineering*, vol. 199, no. 9-12, pp. 712–722, 2010.

[5] K. Samouchos, S. Katsanoulis, and K. Giannakoglou, "Unsteady adjoint to the cut-cell method using mesh adaptation on GPUs," in *ECCOMAS Congress, Crete Island, Greece*, July.

[6] C. Peskin, "Flow patterns around heart valves: a numerical method," *Journal of Computational Physics*, vol. 10, no. 2, pp. 252–271, 1972.

[7] D. Kapsoulis, K. Tsiakas, X. Trompoukis, V. Asouti, and K. Giannakoglou, "Evolutionary multi-objective optimization assisted by metamodels, kernel pca and multi-criteria decision making techniques with applications in aerodynamics," *Applied Soft Computing*, vol. 1, no. 2, pp. 182–197, 2017.

[8] M. Aftosmis, M. Berger, and J. Melton, "Robust and efficient cartesian mesh generation for component-based geometry," *AIAA Journal*, vol. 36, 11 2000.

[9] P. Roe, "Approximate Riemann solvers, parameter vectors, and difference schemes," *Journal of Computational Physics*, vol. 43, no. 2, pp. 357–372, 1981.

[10] Y. Yang and S. Özgen, "Comparison of various spring analogy related mesh deformation techniques in two-dimensional airfoil design optimization," *Progress in Flight Physics*, vol. 9, pp. 189 – 204, 2017.

[11] A. Giotis and K. Giannakoglou, "Low-cost genetic optimization based on inexact pre-evaluations and the sensitivity analysis of design parameters," *Advances in Engineering Software*, vol. 29, no. 2, pp. 129–138, 1998.

[12] M. Karakasis, A. Giotis, and K. Giannakoglou, "Inexact information aided, low-cost, distributed genetic algorithms for aerodynamic shape optimization," *International Journal for Numerical Methods in Fluids*, vol. 43, no. 10-11, pp. 1149–1166, 2003.

[13] K. Giannakoglou. The EASY (Evolutionary Algorithms SYstem) software, http://velos0.ltt.mech.ntua.gr/EASY., 2008.

# High-precision uncertainty propagation involving multimodal probability distributions

**Z. Zhang, C. Jiang***

State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, College of Mechanical and Vehicle Engineering, Hunan University, Changsha, P. R. China, 410082

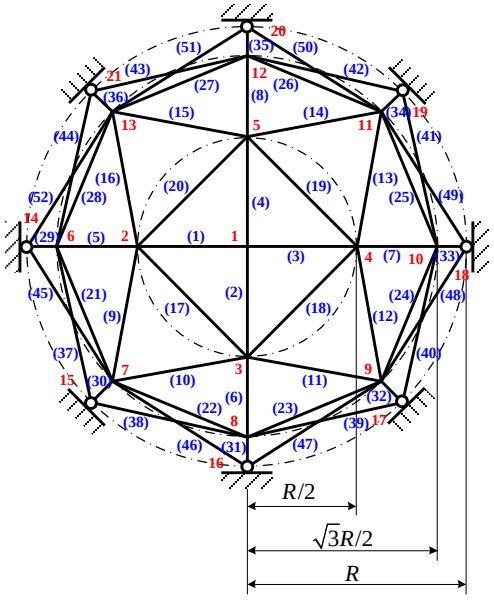
Corresponding author: jiangc@hnu.edu.cn

**Abstract**

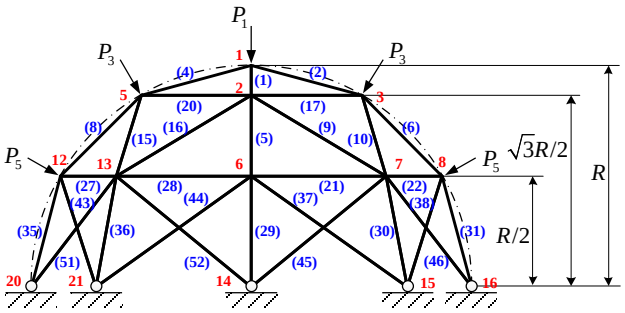
In practical engineering applications, random variables may follow multimodal distributions with multiple modes in the probability density functions, such as the structural fatigue stress of a steel bridge carrying both highway and railway traffic and the vibratory load of a blade subject to stochastic dynamic excitations, etc. Traditional probabilistic uncertainty propagation methods are mainly used to treat random variables with only unimodal distributions, which, therefore, tend to result in large computational errors when multimodal distributions are involved. In this paper, a high-precision probabilistic uncertainty propagation method is proposed for problems involving multimodal distributions. Firstly, the multimodal probability density functions of input random variables are constructed based on the Gaussian mixture model. Secondly, the high-order moments of the response function are calculated using the univariate dimension reduction method, based on which the input uncertainty is effectively propagated. Thirdly, the probability density function of the response is estimated using the maximum entropy method. Finally, a convergence mechanism is formulated to help ensure the uncertainty propagation accuracy. Two mathematical problems and two truss structures are investigated to demonstrate the effectiveness of the proposed method.

**Keywords:** Probabilistic uncertainty propagation; Multimodal distribution; High-order moment;

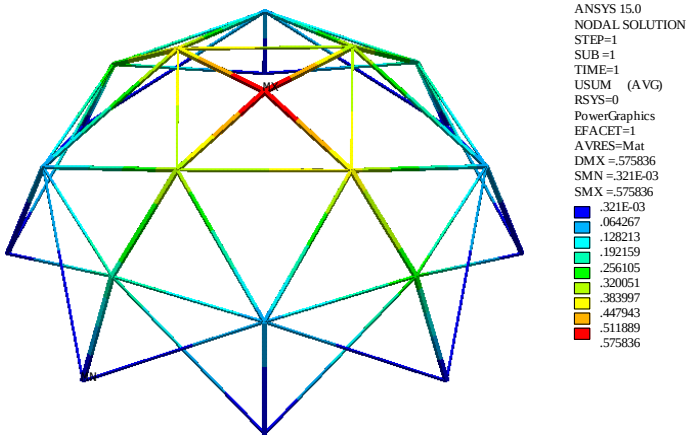
**Example: A 52-bar space truss**




(a) Top view




(b) Side view




(c) FEM model

**Fig. 1.** The 52-bar space truss and its FEM model

As shown in Fig. 1, a hemispherical space truss (like a dome) is considered, which contains 52 bars. The cross-sectional areas of bars 1-8 and 29-36 are $A_1 = 2\text{in}^2$. The cross-sectional areas of bars 9-16 are $A_2 = 1.2\text{in}^2$ and that of the other bars are $A_3 = 0.6\text{in}^2$. The radius of the hemispherical space truss is $R = 240\text{in}$. Six external loads are applied on the space truss, which are $p_1$ in the inner normal direction of point 1, $p_2$ in the inner normal direction of points 2 and 4, $p_3$ in the inner normal direction of points 3 and 5, $p_4$ in the inner normal

direction of points 6 and 10, $p_5$ in the inner normal direction of points 8 and 12, $p_6$ in the inner normal direction of points 7, 9, 11,13. The response function is defined as follows:

$$\delta_1 = g(E, P_1, P_2, P_3, P_4, P_5, P_6) \tag{1}$$

The detailed information of the random variables $P_1, P_2, P_3, P_4, P_5, P_6$ and $E$ are presented in Table 1.

**Table 1** The distribution parameters of random variables $P_1, P_2, P_3, P_4, P_5, P_6$ and $E$

| Random variables | Distribution types | Distribution parameters | | |
|---|---|---|---|---|
| | | Coefficients | Mean values | Standard deviations |
| $P_1$(kip) | Multimodal | $\boldsymbol{\alpha}$ =(0.2,0.2,0.3,0.3) | $\boldsymbol{\mu}$ =(45,55,75,85) | $\boldsymbol{\sigma}$ =(4,4.5,4,4.5) |
| $P_2$(kip) | Multimodal | $\boldsymbol{\alpha}$ = (0.2,0.2,0.3,0.3) | $\boldsymbol{\mu}$ = (40,50,70,80) | $\boldsymbol{\sigma}$ =(4,4.5,4,4.5) |
| $P_3$(kip) | Multimodal | $\boldsymbol{\alpha}$ = (0.2,0.2,0.3,0.3) | $\boldsymbol{\mu}$ = (35,45,65,75) | $\boldsymbol{\sigma}$ =(4,4.5,4,4.5) |
| $P_4$(kip) | Multimodal | $\boldsymbol{\alpha}$ = (0.18,0.18,0.32,0.32) | $\boldsymbol{\mu}$ =(30,35,55,60) | $\boldsymbol{\sigma}$ =(3,6,3,6) |
| $P_5$(kip) | Multimodal | $\boldsymbol{\alpha}$ = (0.18,0.18,0.32,0.32) | $\boldsymbol{\mu}$ = (25,30,50,55) | $\boldsymbol{\sigma}$ =(3,6,3,6) |
| $P_6$(kip) | Multimodal | $\boldsymbol{\alpha}$ = (0.25,0.25,0.25,0.25) | $\boldsymbol{\mu}$ = (20,25,45,50) | $\boldsymbol{\sigma}$ =(3,6,3,6) |
| $E$(ksi) | Normal | -- | $\mu$ =2.5e+04 | $\sigma$ =1.0e+03 |

The PDF results of $\delta_1$ obtained by the proposed method and MCS method are plotted in Fig. 2. It can be observed that the PDF results obtained by the proposed method coincides well with that obtained by MCS, which indicates the high uncertainty propagation accuracy of the proposed method. Especially, the bimodal characteristic of the PDF obtained by MCS is well captured by that obtained by the proposed method. Besides, the CDF results for a series of response functions $\delta_{10}$=$g(P_1, P_2, P_3, P_4) - \bar{\delta}_{10}$ and their relative errors are presented in Table 2. It can be observed that small relative errors are achieved by the proposed method at all cases. For example, the largest relative error of the proposed method is only 3.5088 percents when $\bar{\delta}_{10} = -1.5$ in.

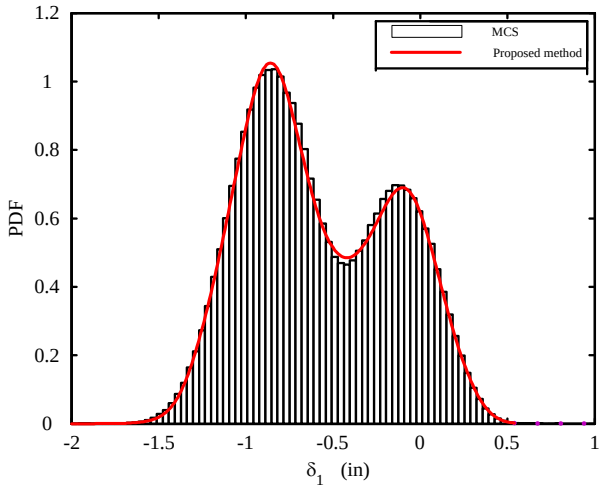



**Fig. 2.** Comparison of the PDF results between the proposed method and MCS method

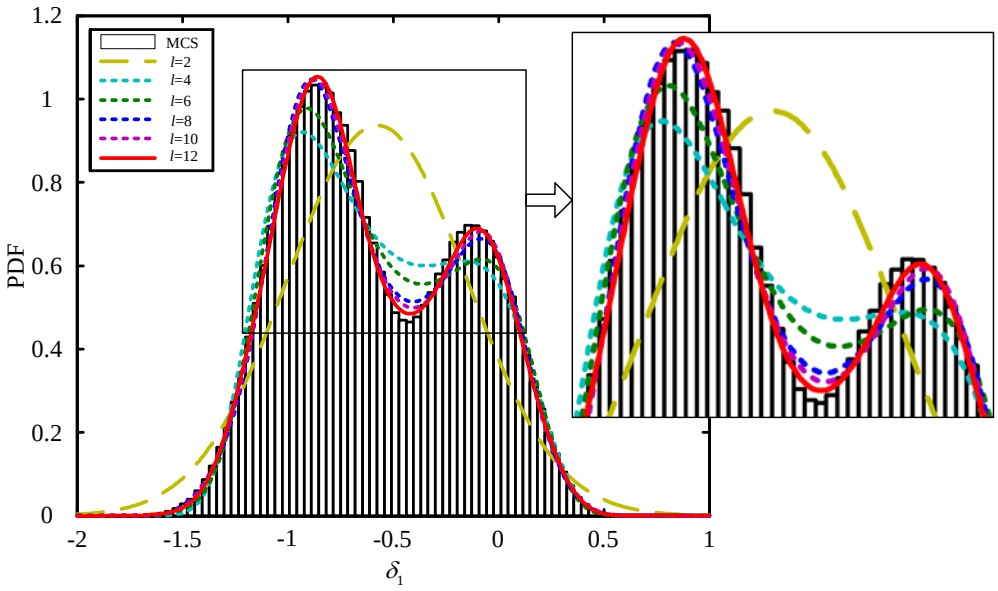
**Table 2** Comparison of the CDF results between the proposed method and MCS method

| $\delta_1$ (in) | MCS | | The proposed method | |
|---|---|---|---|---|
| | CDF | $\varepsilon_0$(%) | CDF | $\varepsilon_1$(%) |
| -1.5 | 1.7100e-03 | -- | 1.6500e-03 | 3.5088 |
| -1 | 0.1714 | -- | 0.1694 | 1.1350 |
| -0.5 | 0.5947 | -- | 0.5933 | 0.2337 |
| 0 | 0.8857 | -- | 0.8856 | 2.9700e-03 |
| 0.5 | 0.9997 | -- | 0.9998 | 3.5500e-03 |

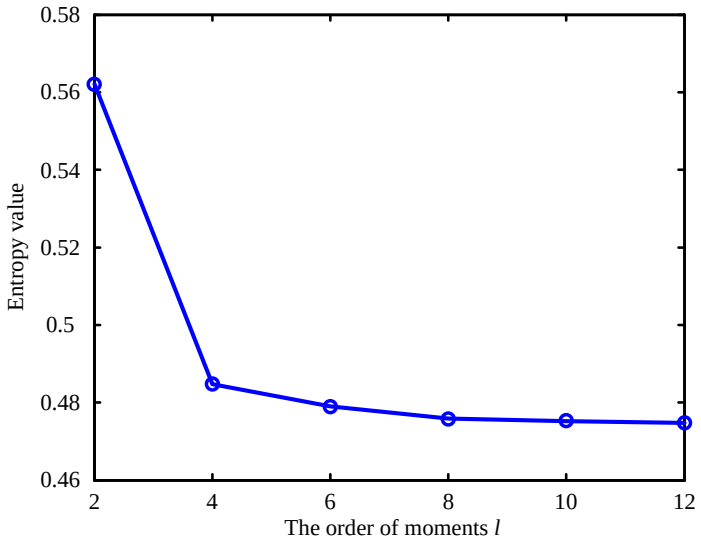
The order of moments that are required for uncertainty propagation is determined as $l$=12 by the convergence mechanism. The evolution process of the estimated PDF and its Shannon entropy under different order of moments are shown in Fig. 3. It can be found that when $l$ increases from 2 to 12, the response PDF calculated using the proposed method gradually approaches to the reference PDF obtained by MCS. When $l$=12, the estimated response PDF is of the highest precision. Furthermore, the Shannon entropy of the response PDF gradually converges to a steady value when $l$ increases from 2 to 12. Table 3 presents the first 12 raw moments of the response function calculated by UDRM and their relative errors compared with the results of MCS. It can be observed that the raw moments of the response function are calculated with satisfied accuracy using the UDRM. The largest relative error of the raw moments is just 9.2600 percents, which occurs at the calculation of $m_{12}$.

Table 4 presents the number of function evaluations of the proposed method and MCS method.

The MCS method is conducted with $1\times10^6$ function evaluations, while the proposed method operates with only 12×7+1=85 function evaluations. Therefore, the proposed method is of satisfied computational efficiency.




(a)




(b)

**Fig. 3.** Evolution of the estimated PDF and its Shannon entropy with the variation of *l*

**Table 3** The calculated raw moments of the response function

| Raw moments | The proposed method | MCS | Relative error（%） |
|---|---|---|---|
| $m_1$ | -0.5758 | -0.5764 | 9.7620e-02 |
| $m_2$ | 0.5120 | 0.5133 | 0.2672 |
| $m_3$ | 0.4327 | 0.4327 | 0.5930 |
| $m_4$ | 0.4930 | 0.4981 | 1.0214 |
| $m_5$ | -0.5243 | -0.5327 | 1.5811 |
| $m_6$ | 0.5800 | 0.5934 | 2.2272 |
| $m_7$ | -0.6631 | -0.6842 | 3.0931 |
| $m_8$ | 0.7807 | 0.8138 | 4.0523 |
| $m_9$ | -0.943 | -0.9952 | 5.1494 |
| $m_{10}$ | 1.1687 | 1.2484 | 6.3843 |
| $m_{11}$ | -1.4789 | -1.6033 | 7.7761 |
| $m_{12}$ | 1.9094 | 2.1043 | 9.2600 |

**Table 4** Comparison of computational efficiency between different methods

| | MCS | The proposed method |
|---|---|---|
| Number of function evaluations | $1\times10^6$ | 85 |

# Combination of Independent Component Analysis and Singular Value Decomposition for emotional EEG source localization

***Wenhui Sun[1], Wenchao Li[1], and †Zhao Lv[1,2]**

[1] School of Computer Science and Technology, Anhui University, China.
[2]Institute of Physical Science and Information Technology, Anhui University, China.

*Presenting author: 20150415swh@gmail.com
†Corresponding author: kjlz@ahu.edu.cn

## Abstract

Focusing on improving the performance of EEG-based emotion recognition and exploring emotional scalp region, this paper presents a novel emotion-related independent components selection method based on Independent Component Analysis (ICA). Specifically, we first establish an optimal spatial-domain filter based on whole channel ICA to extract Independent Components (ICs). On this basis, the Emotion Related Independent Components (ERICs) are determined by evaluating the performance of these ICs through the "leave one IC out" method. Besides, the Singular Value Decomposition (SVD) is help to extract the spatial features. Average recognition accuracy using support vector machine as the classifier achieves 86.49%, which reveals the superiority of the proposed algorithm for emotion recognition.

**Keywords:** Independent Component Analysis, Singular Value Decomposition, emotional EEG source, emotion recognition

## 1. Introduction

Electroencephalogram (EEG) signal, generated from Autonomous Nervous System (ANS), can describe the relationship between psychological changes and emotions [1][2]. Recently, some remarkable EEG-based emotion recognition works have been carrying out explorations on the locations of emotion-related scalp region. Among them, Heller, W. found that alpha-power (8-12Hz) and gamma spectral (30-50Hz) changing at right parietal lobe are related to emotional responses [3][4]. Li, M et al. and Coan, J. A et al. investigated the relationship between the region of temporal/frontal lobe in gamma band and emotion tasks [5][6]. Whereas, emotion-related areas of the cerebral cortex are still uncertain so far, this would lead to a limitation on performance improvement in emotion recognition. Nowadays, researches are mainly concentrating on time/frequency characteristics to analyse the sources of emotion-related scalp region while the independence and spatial information of the source may be ignored. To explore the location of emotion-related sources and improve the recognition accuracy, we develop a novel selection method basing on the Independent Component Analysis (ICA) to obtain Emotion-Related Independent Components (ERICs).

## 2. Method

The fundamental flowchart of the proposed algorithm based on ICA and SVD is elaborated in Fig.1 and the key steps are as follows:

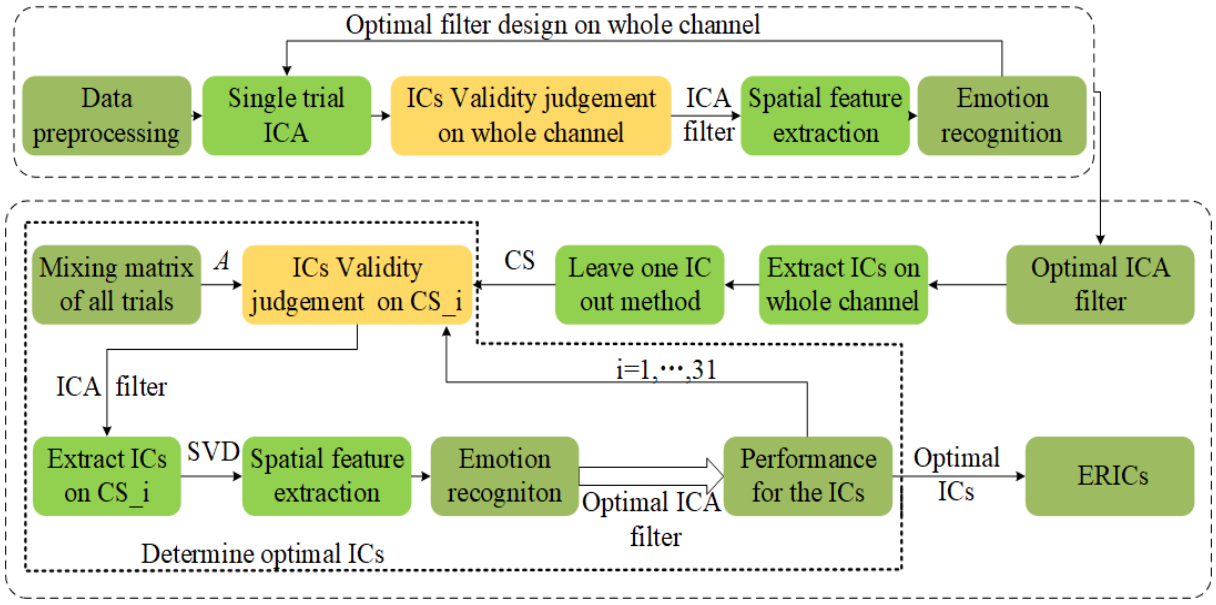



**Figure 1. Extraction of ERICs for emotional EEG source localization.**

### *2.1. Data preprocessing*

The 32-channel EEG datasets used in our work are obtained from the MAHNOB-HCI-TAGGING DATABASE [7].The classification task is performed among three classes of emotion (pleasant, neutral valence and unpleasant) according to the valence dimension. Considering the fact that the gamma band plays a significant role than other band in emotion recognition [5][8], we first use a bandpass filter of 30-50Hz to extract gamma rhythm. To ensure the stability of emotion elicitation and avoid the multi-emotion in an observation period, we further split the pre-processed EEG signals into segments of 8s with 50% overlap using the rectangle window [9], and regard one data segment as a trial.

### *2.2. Single trial ICA analysis*

We consider a single trial $\mathbf{x}=[x_1,\ldots x_n]^T$ mentioned above as instantaneous mixture, which can be separated by information maximization approach [11] combined with natural gradient [12]. Instantaneous mixture is the simplest form of ICA algorithm which can be modeled as $\mathbf{x}=\mathbf{As}$, where $\mathbf{A}$ represents mixing matrix and $\mathbf{s}=[s_1,\ldots,s_n]^T$ denotes the source signals. The goals of employing ICA algorithm are to learn the unmixing matrix $\mathbf{W}$ and obtain the estimate of source signals $\mathbf{y}=[y_1,\ldots,y_n]^T$, which means $\mathbf{W}=\mathbf{A}^{-1}$ and $\mathbf{y}=\hat{\mathbf{s}}=\mathbf{Wx}$, where $\hat{\bullet}$ denotes the estimate of $\bullet$.

Each row of $\mathbf{y}$ can be regarded as an Independent Component(IC), and a random column $a_i,(i=1,\ldots,n)$ in $\mathbf{A}$ includes projection coefficients from $i^{th}$ IC to each electrodes. To acquire the maximum projected position of each IC, we calculate the maximum value of each column for $|\mathbf{A}|$, and then save the index of the maximum values into a matrix $\mathbf{D}_{1\times n}$, in which the elements of $\mathbf{D}_{1\times n}$ represent the indexes of EEG channels.

*2.3. ICs validity judgement and spatial feature extraction*

In order to obtain the ICs related to the specific scalp region at the same time, we continue to perform the validity judgement of selected ICs on a channel set. We give a k-channels set **CS** to explain the process, specifically, if the matrix $\mathbf{D}_{1\times n}$ includes $No.Chan_{CS_1},\ldots,No.Chan_{CS_i},\ldots,No.Chan_{CS_k}$ simultaneously (here, $No.Chan_{CS_i}$ is the index of EEG channel, and $Chan_{CS_i}$ is the electrode label of the $i^{th}$ channel in $CS$ ), we infer that these ICs on the **CS** is valid.

For a random trial **x** that have conformed validity judgement of ICs, we choose the corresponding column vector to establish the ICA filter bank $\{w_1,\ldots,w_k\}$ and employ the filter bank to linearly project each trial to extract ICs ($u_1^n,\ldots,u_k^n$) for different emotional states. Let $\hat{\mathbf{S}}=\left[u_1^n,\ldots,u_k^n\right]^T$ and perform Singular Value Decomposition (SVD) on $\hat{\mathbf{S}}$ :

$$\hat{\mathbf{S}}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{1}$$

**U** , **V** are orthogonal matrices, $\boldsymbol{\Sigma}$ is diagonal matrix. Finally the **SF** = $[\lambda_1 v_1,\ldots,\lambda_j v_j,\ldots,\lambda_{k-1}v_{k-1}]$ is regarded as the feature. **SF** is the abbreviation of spatial feature. $\lambda_j$ is the $j^{th}$ non-zero element on the diagonal of $\boldsymbol{\Sigma}$, $v_j$ is the $j^{th}$ column element vector of **V**.

*2.4. ERICs selection*

First, the optimal ICA filter that designed on whole channels is applied to extract the corresponding ICs. Thus, 32 ICs for 32 observation channels, which according to the ICs-electrode mapping mode, can be acquired. In order to assess the relevance between emotion and ICs, we then apply the "leave one IC out" method for ERICs selection, that is one IC ($u_j$, $j=1,\ldots,32$) is taken out and the **SF** for emotion recognition is extracted on the rest of the ICs. In this way, we can acquire a recognition accuracy $ac_j$ in the absence of $u_j$, and attain recognition accuracy vector $\mathbf{ChanAc}=\{ac_1,\ldots,ac_j,\ldots,ac_{32}\}$ through 32-rounds tests in same way. To select the ERICs by evaluating the decreased recognition accuracy that induced by the absence of special IC, we define $\mathbf{ACC}=\{da_1,\ldots,da_j,\ldots,da_{32}\}$ as follows:

$$da_j=\max(\mathbf{ChanAc})-ChanAc_j \tag{2}$$

where, $\max(\bullet)$ denotes the selection of the maximum value in observation vector. It is worthy to note that the greater the degradation of the performance, the higher emotion correlation this IC has, then we get 31 test channel sets in Fig.2 for selecting ICs. Finally, the step named "Determine optimal channel set" in Fig. 1 is performed repeatedly on the test channels sets, the ICs that on the optimal channel set are regarded as the ERICs.

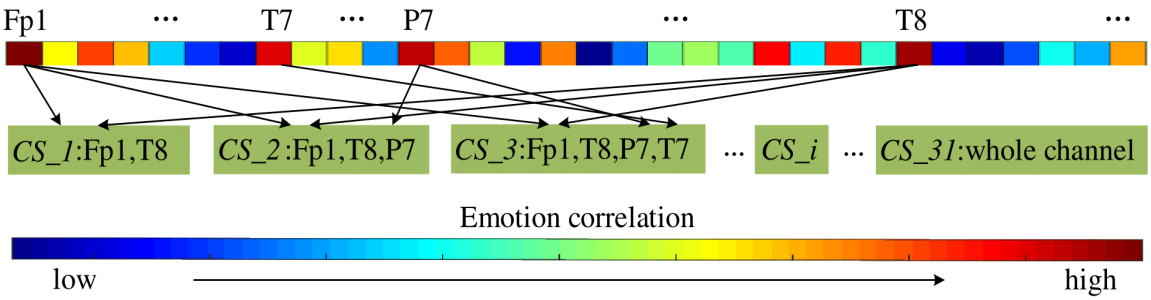



**Figure 2. The generation process of 31 test channel sets for a subject. *CS _ i* represents a test channel set for selecting ICs. The colored rectangular boxes show the emotional correlation of the channels, which obtained by ACC from a single subject.**

## 3. Experiments and results

To validate the feasibility of the proposed method, ten subjects' data from the database mentioned above are involved in our experiment. We divide single subject's EEG data to average two parts, one part is to generate the channel sets *CS* and another part is used to select the optimal channel set. The recognition ratios based on these test channel sets are illustrated in Fig. 3. It shows that the recognition accuracy is rising steadily and stabilized at a high level. For each subject, we choose ICs on the channel set with the highest ratio (marked with a white triangle) as the ERICs.

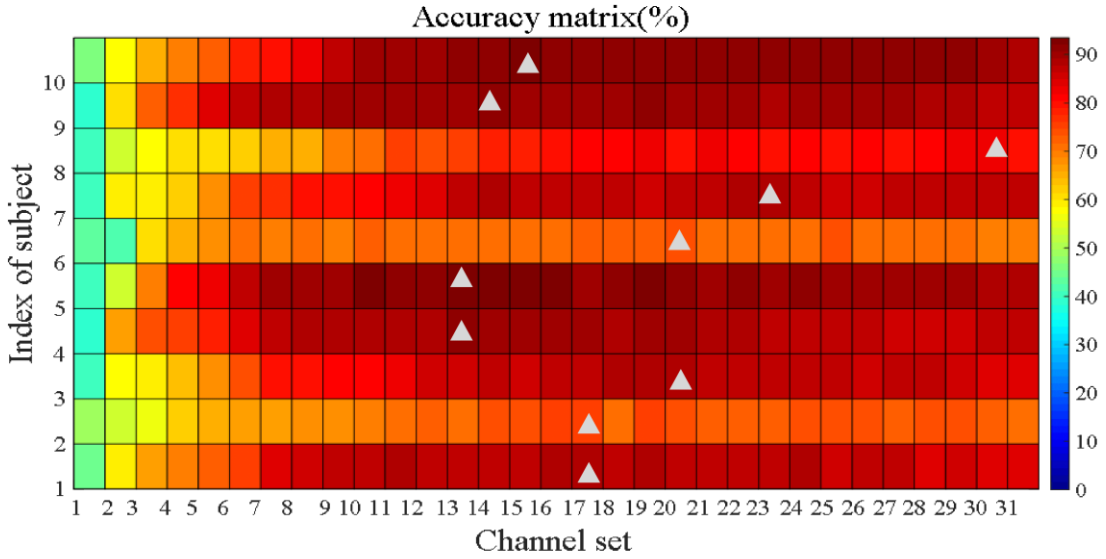
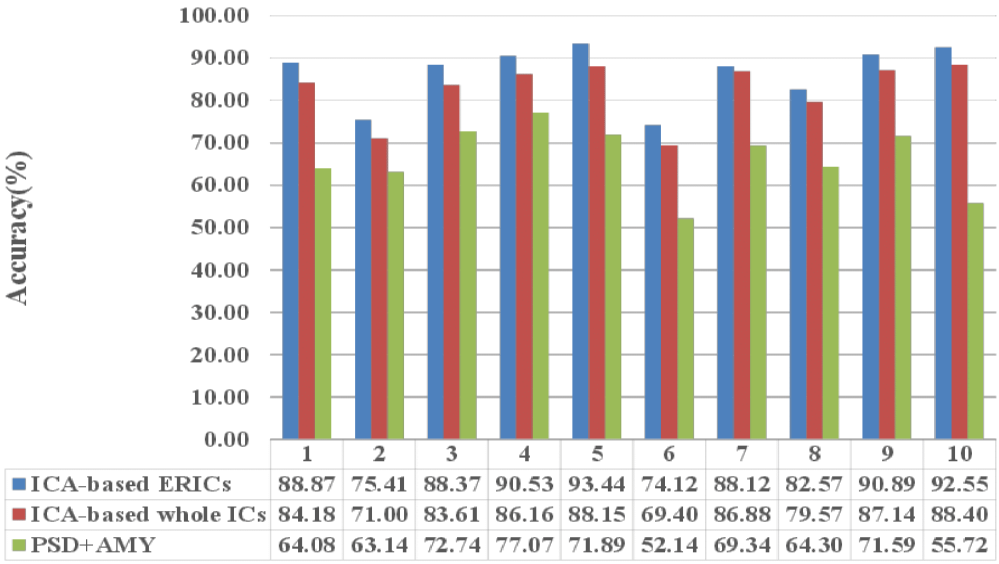



**Figure 3. Recognition accuracy based on test channel sets for each subject. The box with a white triangle indicates the best recognition performance among 31 channel sets.**

Furthermore, the comparison experiment results can be seen from Fig.4, which correspond to whole ICs, ERICs and traditional method based on power spectral density and asymmetry features [7], respectively. It is obviously that methods based on ICA algorithm achieve better performance than traditional time/frequency domain method, and the experiment results can also prove that spatial-domain feature provides richer distinguishable information to accurately identify different emotional states. Compared with result obtained by using whole

ICs, result of only using ERICs reaches higher accuracy since it removes irrelevant ICs which may have influence on the performance of emotion recognition. In a word, the experiment validates the feasibility of the proposed ERICs selection strategy as well as the ability of improving the recognition performance.




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ICA-based ERICs | 88.87 | 75.41 | 88.37 | 90.53 | 93.44 | 74.12 | 88.12 | 82.57 | 90.89 | 92.55 |
| ICA-based whole ICs | 84.18 | 71.00 | 83.61 | 86.16 | 88.15 | 69.40 | 86.88 | 79.57 | 87.14 | 88.40 |
| PSD+AMY | 64.08 | 63.14 | 72.74 | 77.07 | 71.89 | 52.14 | 69.34 | 64.30 | 71.59 | 55.72 |

**Figure 4. Recognition accuracy for different methods.**

To evaluate the performance of the ERICs, the F1 scores and recognition rates for the classification in different modalities are given in the Table 1. Among three classes of emotion recognition, the "Pleasant" state achieves the highest accuracy ratio (88.23%), while the "Neutral-valence" state shows the lowest one (81.5%). It indicates that the emotion independent component in the "pleasant" state is more effective than that in the "unpleasant" and "Neutral-valence" states.

**Table 1. The recognition accuracy and F1 score under three emotion tasks in case of the "ERICs".**

| Emotion | Modality | ERICs |
|---|---|---|
| Pleasant | Accuracy | 88.23% |
| | F1-score | 0.8697 |
| Neutral-valence | Accuracy | 81.50% |
| | F1-score | 0.8319 |
| Unpleasant | Accuracy | 82.92% |
| | F1-score | 0.8431 |

Moreover, we draw the topography map to analyse the emotion-related scalp region according to the mean of **ACC***s* over all subjects. From Fig. 5, we can observe that the ICs located on the lateral temporal, prefrontal and occipital scalp regions are crucial for emotion recognition. This result is consistent with the reports of literature [5] and [10].

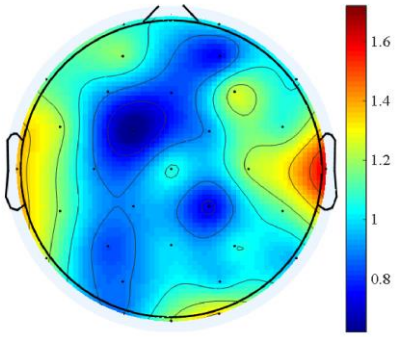



**Figure 5. The topography map based on the mean of ACC*s* over all subjects.**

## 4. Conclusion

In this work, we present an EEG-based emotion recognition method using ICA to improve the performance of emotion recognition. The main properties of this method are: (i) the spatial features obtained by ICA and SVD are first applied to EEG for emotion recognition, (ii) the independence of the emotion related sources is first considered, (iii) the emotion-related independent components can describe the emotion-related scalp region and (iv) experiment results confirm both the validity of ERICs in the cerebral cortex and the ability to recognize three-class emotion tasks with a high accuracy (86.49%).

## Acknowledgment

This work was supported by the National Natural Science Fund of China (grant 61401002), Anhui provincial natural science research project of colleges and Universities (grant KJ2018A0008) and Open fund for Discipline Construction， Institute of Physical Science and Information Technology, Anhui University.

### References

[1] Wioleta S. Using physiological signals for emotion recognition[C]//2013 6th International Conference on Human System Interactions (HSI). IEEE, 2013: 556-561, doi:10.1109/HSI.2013.6577880
[2] Ghare P S, Paithane A N. Human emotion recognition using non linear and non stationary EEG signal[C]//2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT). IEEE, 2016: 1013-1016, doi: 10.1109/ICACDOT.2016.7877739
[3] Heller W. Neuropsychological mechanisms of individual differences in emotion, personality, and arousal[J]. Neuropsychology, 1993, 7(4): 476, doi:10.1037/0894-4105.7.4.476
[4] Sarlo, M., Buodo, G., Poli, S., & Palomba, D.:'Changes in EEG alpha power to different disgust elicitors: the specificity of mutilations', Neuroscience letters, 2005, 382, (3), pp. 291-296, doi:10.1016/j.neulet.2005.03.037
[5] Li, M., & Lu, B. L.: 'Emotion classification based on gamma-band EEG', Annual International Conference of the IEEE, Medicine and Biology Society, September, 2009, pp. 1223-1226, doi:10.1109/IEMBS.2009.5334139
[6] Coan, J. A., & Allen, J. J.: 'Frontal EEG asymmetry as a moderator and mediator of emotion', Biological psychology, 2004, 67, (1-2), pp. 7-50, doi:10.1016/j.biopsycho.2004.03.002
[7] Soleymani, M., & Pantic, M.: 'Multimedia implicit tagging using EEG signals', IEEE International Conference, Multimedia and Expo, July 2013, pp. 1-6,doi:10.1109/ICME.2013.6607623
[8] Candra H, Yuwono M, Chai R, et al. Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine[C]//2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015: 7250-7253.
[9] Wang, X. W., Nie, D., & Lu, B. L.: 'Emotional state classification from EEG data using machine learning approach', Neurocomputing, 2014, pp. 94-106, doi:10.1016/j.neucom.2013.06.046

[10] Nie, D., Wang, X. W., Shi, L. C., & Lu, B. L.: 'EEG-based emotion recognition during watching movies', International IEEE/EMBS Conference, Neural Engineering, April 2011, pp. 667-670, doi:10.1109/NER.2011.5910636
[11] Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution[J]. Neural computation, 1995, 7(6): 1129-1159.
[12] Amari S I. Natural gradient works efficiently in learning[J]. Neural computation, 1998, 10(2): 251-276.

# Stress/Displacement Field Calculation for Discontinuous Mechanical Structure Based on Layered Elastic Theory

**†Qingchao Sun, *Liangguo Meng, Yunlong Wang and Lin Xue**

School of Mechanical Engineering, Dalian University of Technology, China

*Presenting author: menglg@mail.dlut.edu.cn
†Corresponding author: qingchao@dlut.edu.cn

## Abstract

Stress/displacement field analyzing of mechanical assembly is the basis of predicting the mechanical property of the assembly and optimizing the structure and the assembly process. However, the discontinuity structures in the mechanical make it difficult to calculate. In this paper, a new kind of stress/displacement layered mapping and calculation method based on layered elastic theory is proposed. With considering the mechanical continuous/discontinuous characteristics, a layered model of assembly structure is established and the layered constraint conditions is determined by its position in assembly. Stress/displacement field could be obtained by using traditional solution of layered elastic system which is modified to adjust to the model of mechanical assembly. Finally, a discontinuous mechanical structure is solved by using the layered model and the comparison between the analytically calculation, FEA and experiment data proves the effectiveness of the model.

**Keywords:** Stress/displacement field; Layered elastic theory; Structural discontinuity; Elastic mechanics

## 1 Introduction

In general, machines like lathe are not a continuous whole. They're assembly by various parts according to requirements. The parts that are connected to each other are called mechanical joints, such as bolted joint, sliding guide joint, et al. The contact region between different parts of the ministry is often referred to as a joint interface.

According to the research, the characteristics of mechanical joint interface have an important influence on the performance of mechanical parts, such as contact fatigue strength, friction power consumption, wear life [1,2,3]. In addition, the dynamic performance, vibration and noise of mechanical equipment depend on interface's stiffness and the damping. Burdekin et al [4] pointed out that the deformation of machine tool's joints accounted for most of the total deformation and the contact stiffness of joints was about 60%~80% of the total rigidity of the machine tool. The research of Yagi[5] found that the mechanical joint is important to the dynamic stiffness of reconfigurable machine tool. There is a view [6] joint's damping is larger than the damping of the structure itself and Beards [7,8,9] studied structure damping with the interface's sliding. He puts the integration that 90% of the total damping is coming from joint. Now that the joint is significant on the static and dynamic performance of mechanical system, a lot of researches have been carried out. For instance, Zhang and Mr. Huang [10, 11] had studied joint's normal and tangential stiffness, damping through a lot of experiments and summarized the influence of normal stress, the media, materials, processing methods and roughness. Mi [12] studied the influence of the pretension on the dynamic stiffness of the machine

tool. In addition, a lot of achievements have been made in the identification of the parameters and the mechanics modeling of the joint [1,13-15].

In fact, joint's behavior is important for mechanical design/operation/maintenance of mechanical equipment. People has accumulated experiences to analyze it in long-term engineering practice. With the development of science and technology, the joint interface's characticts of mechanical equipment will become an important part of scientific and technical.

Joint is a space with a certain thickness. In order to understand joint comprehensively and accurately, it is necessary to figure out how the load transforming in joint interface. Traditional mechanical analysis methods, either material mechanics or elastic mechanics, are based on the theory of continuum mechanics and the assumption that the object is composed of a continuous medium filled in the space [16]. The joint in mechanical equipment destroys the continuity of the whole machine structure, so that the continuum mechanics is not completely suitable for the analysis of the assembly mechanical. Moreover, the discontinuity of the structure makes the transfer of load between the components more complex, including non-linearity, which making it difficult to analyze the stress/displacement field in machine.

Layered elastic theory (LET) is the theoretical foundation in pavement design and calculation. LET belong to the elasticity. It developed on the basis of elastic semi-space theory. In 1885, Bossiness proposed a theoretical solution to the stress/displacement of elastic semi-space under vertical concentrated force. In 1943 and 1945, the Burmister used the Love functions to obtain the theoretical solution of the stress/displacement of the double-layer and multi-layer elastic system under the symmetrical vertical load [17-19]. In 1951, Sneddon firstly introduced the Hankel integral transforms into axisymmetric problems [20]. Since then, the LET has been developed rapidly and extended to non-axisymmetric and multi-layer elastic systems [21-24]. With the improvement of computer and computing method, the system of layered elastic mechanics is applied to the engineering. There are lots of algorithms for computing, such as BISAR, JULEA, DIPLOMAT, Kenlayer, LEAF, et al [24]. At present, the layered elastic system mechanics and its algorithm have been widely used in the engineering practice of multi-level road and multi-layer foundation all over the world.

LET is widespread in road construction, but it is seldom used in mechanical. The theory comprehensively considers both overall stress transfer and the discontinuous effect of interface to analyze the stress/ displacement field, which provides the possibility for the analysis of mechanical characteristics of discontinuous mechanical structures.

Therefore, this paper introduces a new method which is based on layered elastic mechanics to analysis stress/ displacement field in discontinuous mechanics. An example of double-layer discontinuous mechanical under vertical load is given and in order to verify the effectiveness, we also have experimental and finite element analysis. Comparing the results of three methods, the new method is effective in structural discontinuity.

## 2 The Basic Concept of Layered Elastic Theory

The LET is the theoretical basis of multi-layer pavement and foundation design. The pavement system is layered structure on the soil foundation and composed of different materials. In general, the external load on the road surfaces is vertical or horizontal. LET assumes object as an elastic system, including a series of elastic layers and a semi-infinite layer. It is used to analyze stress and displacement of the elastic system when load is acting on the surface of pavement. LET is based on the following assumptions:

- Each layer is ideal linear elasticity, completely uniform, continuous, and isotropic.
- Initial stress is 0 and body forces are ignored in system without external load.
- Strains and displacement are assumed to be small.
- Stress, deformation and displacement vanish in infinity point.

Fig. 1 shows a n-layer elastic system. We denote the number of layers as n. $E_i$、$\mu_i$、$h_i$ are Young's modulus, Poisson's ratio and thickness of layers respectively. Each layer is infinite along radial direction $r$. Except the $n_{th}$-layer is a semi-infinite space ($h_n = \infty$), the other layers' thickness is limited. The cylindrical coordinate system is established in Fig. 1 and there is only vertical load acting on the surface.

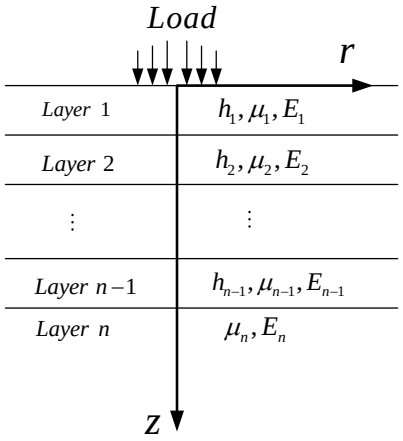



**Fig. 1. A n-layer elastic system**

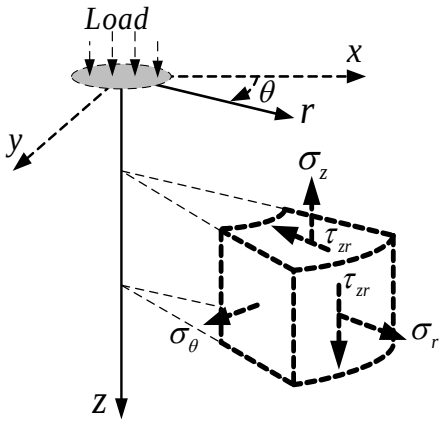



**Fig. 2. axisymmetric problem.**

There are two main methods based on LET to obtain stress/displacement filed in layered elastic system[23]. The traditional method gives the relation between mechanical component and correlation function firstly. Then the undetermined function is obtained through the Hankel transformation. For example, Burmister presents a unique stress function, which solves the stress and displacement of the two-layer and three-layer elastic system under the axisymmetric vertical load[17]−[19]. Maina and Matsui used the Michell equation and Boussinesq equation to calculate the elastic response of the elastic layered structure under horizontal and vertical loads[22]. It should be pointed out the method is simple and practical, but we must know the relation of displacement/stress component and the displacement/stress function. It is suitable for simple problems such as spatial axisymmetric.

The other method[23] calculates basic equation with the Laplace and Hankel integral transformation to obtain stress/displacement field. Although it does not need to obtain the relation between the displacement/stress component and the displacement/stress function, the process of solution is quite complicated.

The first method (displacement method) is used to solve spatial axisymmetric problem in the next. It is typical method to establish mechanical model in layered elastic system.

The whole solution begins with acquiring stress/displacement's general result in axisymmetric spatial. In cylindrical coordinates system, assuming Love displacement function $\phi(r,z)$ and according to Lame equation of elasticity, the relation between displacement components and displacement function is expressed as follows:

$$\left.\begin{aligned} u &= -\frac{1+\mu}{E}\frac{\partial^2\phi}{\partial r\partial z} \\ w &= \frac{1+\mu}{E}[2(1-\mu)\nabla^2\phi - \frac{\partial^2\phi}{\partial z^2}] \end{aligned}\right\} \tag{1}$$

The stress component is shown by the displacement function

$$\left.\begin{aligned} \sigma_r &= \frac{\partial}{\partial z}(\mu\nabla^2\phi - \frac{\partial^2\phi}{\partial r^2}) \\ \sigma_\theta &= \frac{\partial}{\partial z}(\mu\nabla^2\phi - \frac{1}{r}\frac{\partial\phi}{\partial r}) \\ \sigma_z &= \frac{\partial}{\partial z}[(2-\mu)\nabla^2\phi - \frac{\partial^2\phi}{\partial z^2}] \\ \tau_{zr} &= \frac{\partial}{\partial r}[(1-\mu)\nabla^2\phi - \frac{\partial^2\phi}{\partial z^2}] \end{aligned}\right\} \tag{2}$$

Where $\nabla^2$ denote the Laplace operator and is given by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} \tag{3}$$

In fact, the Love displacement function should satisfy the following re-harmonic equation

$$\nabla^4\phi = 0 \tag{4}$$

Using the Hankel integral transformation on both ends of the re-harmonic Eq. (3) it can be shown that

$$\int_0^\infty r\nabla^4\phi(r,z)J_0(\xi r)dr = (\frac{d^2}{dz^2} - \xi^2)^2\bar{\phi}(\xi,z) = 0 \tag{5}$$

Where $\bar{\phi}(\xi,z)$ is the zero-order Hankel integral transformation function

$$\bar{\phi}(\xi,z) = \int_0^\infty r\phi(r,z)J_0(\xi r)dr \tag{6}$$

Eq. (4) is the ordinary differential equation, and the expression of its solution is

$$\bar{\phi}(\xi,z) = (C_\xi + A_\xi z)e^{\xi z} + (D_\xi + B_\xi z)e^{-\xi z} \tag{7}$$

The solution of displacement function $\phi(r,z)$ is obtained by the means of Hankel inverse transform

$$\phi(r,z) = \int_0^\infty \xi[(C_\xi + A_\xi z)e^{\xi z} + (D_\xi + B_\xi z)e^{-\xi z}]J_0(\xi r)d\xi \tag{8}$$

Where $\xi$ is integral variable and $A_\xi$、$B_\xi$、$C_\xi$、$D_\xi$ are integral constant associated with $\xi$.

Assuming $A = \xi^3 A$, $B = \xi^2 B$, $C = \xi^3 C$, $D = \xi^2 D$ and substituting Eq. (6) into Eq. (1) and (2), component of stress and displacement are obtained

$$\left.\begin{aligned}
\sigma_r &= \int_0^\infty \xi J_0(\xi r)\left\{\left[C+(1+2\mu+\xi z)A\right]e^{\xi z}-\left[D-(1+2\mu-\xi z)B\right]e^{-\xi z}\right\}d\xi-\frac{1}{r}U \\
\sigma_\theta &= 2\mu\int_0^\infty \xi J_0(\xi r)(Ae^{\xi z}+Be^{-\xi z})d\xi+\frac{1}{r}U \\
\sigma_z &= -\int_0^\infty \xi J_0(\xi r)\left\{\left[C-(1-2\mu-\xi z)A\right]e^{\xi z}-\left[D+(1-2\mu_i+\xi z)B\right]e^{-\xi z}\right\}d\xi \\
\tau_{zr} &= -\int_0^\infty \xi J_1(\xi r)\left\{\left[C+(2\mu+\xi z)A\right]e^{\xi z}+\left[D-(2\mu-\xi z)B\right]e^{-\xi z}\right\}d\xi \\
u &= \frac{1+\mu}{E}\int_0^\infty J_1(\xi r)\left\{\left[C+(1+\xi z)A\right]e^{\xi z}-\left[D-(1-\xi z)B\right]e^{-\xi z}\right\}d\xi \\
w &= -\frac{1+\mu}{E}\int_0^\infty J_0(\xi r)\left\{\left[C-(2-4\mu-\xi z)A\right]e^{\xi z}+\left[D+(2-4\mu+\xi z)B\right]e^{-\xi z}\right\}d\xi
\end{aligned}\right\} \tag{9}$$

Where

$$U=\int_0^\infty J_1(\xi r)\left\{\left[C+(1+\xi z)A\right]e^{\xi z}-\left[D-(1-\xi z)B\right]e^{-\xi z}\right\}d\xi \tag{10}$$

Where $J_0$, $J_1$ are order 0 and order 1 of the first kind Bessel functions; denoting $\xi$ as integral constant. $A_\xi$、 $B_\xi$、 $C_\xi$、 $D_\xi$ are integral constants associated with $\xi$ and solved through definite condition of question.

If the load is axisymmetric, the mechanical problem in the layered elastic system also belongs to spatial axisymmetric problems. Since Eq. (9) is a general solution of stress/displacement in spatial axisymmetric problems, the stress displacement at any point in the layered elastic system can be solved. The stress and displacement expressions in Eq. (9) contain four integral constants $A$、 $B$、 $C$、 $D$. As long as the integral constant is determined, the stress/displacement of the entire layered elastic system can be obtained. Because the boundary conditions of each layer are different, the integral constants in the stress/displacement expression of each layer are also different. In this paper, the subscript i is used to number the layers. In a N-layer elastic system, there are 4N unknown integral constants, $A_i$, $B_i$, $C_i$, $D_i$ $(i=1,2,\ldots,N)$.

It is known when the surface of layers contacts or bond, the normal stress (displacement) on both sides of interface is continuous and the tangential stress (displacement) may have some correlation. Furthermore, it is easy to obtain the stress boundary conditions in the surface of the layered elastic system. By using these stress conditions of surface, bonding conditions between layers and other definite conditions, equation set can be established to solve integral constants. The above is a common method for solving layered elastic mechanics system.

## 3 Calculation of Discontinuous Mechanical Structure.

### *3.1 Model Assumption*

Different from roadbed structure, the structure of mechanical equipment is more complicated in general. It is necessary to ignore some unimportant factors and simplify the practical problems. In this paper, the discontinuous mechanical structure with vertical load is simplified as the following model, a double-layer elastic system. The system bases on the assumption of LET.

In Fig. 3, we denote the thickness of layer 1 as $h$ and the layer 2 is a semi-infinite substrate. $E_i$、$\mu_i$ are layers' Young's modulus and Poisson's ratio respectively. The uniform load $q(r)$ is distributed vertically on a circular area (radius $\delta$) on the 1st layer's surface. Because the distribution of load is axisymmetric, the stress, strain and displacement components are also axisymmetry. Stress and displacement of any point in the system could be solved through the traditional method, Love displacement function method. The stress of the axisymmetric problem is shown in.

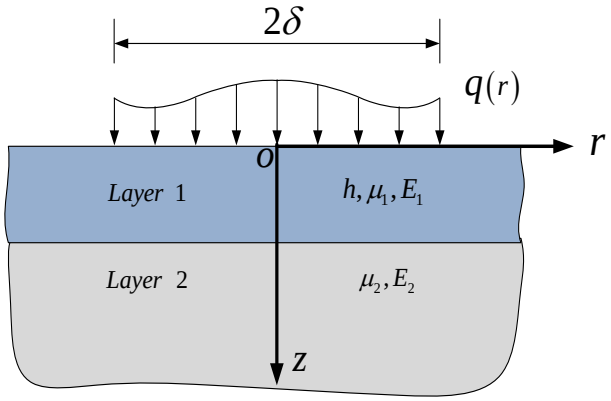



**Fig. 3. A double-layer elastic system**

### *3.2 Boundary Conditions and Governing Equations*

Since the system is subjected to vertical downward normal force, its surface stress boundary condition is as follows

$$\left.\begin{aligned} \sigma_{z1} &= -q(r) \quad (z=0) \\ \tau_{zr1} &= 0 \quad (z=0) \end{aligned}\right\} \tag{11}$$

For the continuous mechanical structure, the normal stress and normal displacement of interface are continuous

$$\left.\begin{aligned} \sigma_{z1} &= \sigma_{z2} \quad (z=h) \\ w_1 &= w_2 \quad (z=h) \end{aligned}\right\} \tag{12}$$

The shear stress of interface can be calculated by its displacement, as following

$$\tau_{zr1} = \tau_{zr2} = K(u_2 - u_1) \qquad (z=h) \tag{13}$$

Where $K$ is the coefficient of layers combination.

In fact, when there is only vertical load, friction in layers has little effect on mechanical components distribution. we assume interface is smooth and frictionless to simplify the model. So $K=0$ and the shear stress of interface is as follows

$$\tau_{zr1} = \tau_{zr2} = 0 \qquad (z=h) \tag{14}$$

If $r$ and $z$ approach infinity( $r \to \infty$ and $z \to \infty$), respectively, stress and displacement components in Eq. (9) will be zero

$$A_2 = C_2 = 0 \tag{15}$$

### *3.3 Solving the Stress and Displacement Components*

By substituting stress/displacement of components in Eq. (9) into Eq. (11), (12), (14) and using Hankel integral transformation ,the following linear algebraic equations are obtained

$$\left.\begin{aligned}
&C_1 - (1-2\mu_1)A_1 - D_1 - (1-2\mu_1)B_1 = \bar{q}(\xi) \\
&C_1 + 2\mu_1 A_1 + D_1 - 2\mu_1 B_1 = 0 \\
&\left[C_1 - (1-2\mu_1-\xi h)A_1\right]e^{\xi h} - \left[D_1 + (1-2\mu_1+\xi h)B_1\right]e^{-\xi h} \\
&= \left[C_2 - (1-2\mu_2-\xi h)A_2\right]e^{\xi h} - \left[D_2 + (1-2\mu_2+\xi h)B_2\right]e^{-\xi h} \\
&\left[C_1 + (2\mu_1+\xi h)A_1\right]e^{\xi h} + \left[D_1 - (2\mu_1-\xi h)B_1\right]e^{-\xi h} = 0 \\
&\left[C_2 + (2\mu_2+\xi h)A_2\right]e^{\xi h} + \left[D_2 - (2\mu_2-\xi h)B_2\right]e^{-\xi h} = 0 \\
&m\left[C_1 - (2-4\mu_1-\xi h)A_1\right]e^{\xi h} + \left[D_1 + (2-4\mu_1+\xi h)B_1\right]e^{-\xi h} \\
&= \left[C_2 - (2-4\mu_2-\xi h)A_2\right]e^{\xi h} + \left[D_2 + (2-4\mu_2+\xi h)B_2\right]e^{-\xi h}
\end{aligned}\right\} \tag{16}$$

Where

$$\begin{aligned}
&\bar{q}(\xi) = \int_0^\infty rq(r)J_0(\xi r)dr \\
&m = \frac{(1+\mu_1)E_2}{(1+\mu_2)E_1}
\end{aligned} \tag{17}$$

Substituting $A_2 = C_2 = 0$ into Eq. (16), $A_i$ , $B_i$ , $C_i$ , $D_i$ $(i = 1,2)$ could be represented as follows:

$$\left.\begin{aligned}
A_1 &= -\frac{\bar{q}(\xi)e^{-2\xi h}}{\Delta}[(N-1-\xi h)-(N-1)e^{-2\xi h}] \\
B_1 &= -\frac{\bar{q}(\xi)}{\Delta}[N-(N-\xi h)e^{-2\xi h}] \\
C_1 &= \frac{\bar{q}(\xi)e^{-2\xi h}}{\Delta}[(2\mu_1+\xi h)(N-1-\xi h)+N\xi h-2\mu_1(N-1)e^{-2\xi h}] \\
D_1 &= -\frac{\bar{q}(\xi)}{\Delta}\{2\mu_1 N-[2N(\mu_1-\xi h)+(1-2\mu_1+\xi h)\xi h]e^{-2\xi h}\} \\
B_2 &= -\frac{\bar{q}(\xi)}{\Delta}(2N-1)[(1+\xi h)-(1-\xi h)e^{-2\xi h}] \\
D_2 &= -\frac{\bar{q}(\xi)}{\Delta}(2N-1)(2\mu_2-\xi h)[(1+\xi h)-(1-\xi h)e^{-2\xi h}] \\
A_2 &= C_2 = 0
\end{aligned}\right\} \tag{18}$$

Where

$$\Delta = N + [2\xi h(2N-1) - (1+2\xi^2 h^2)]e^{-2\xi h} - (N-1)e^{-4\xi h}] \tag{19}$$

$$N = \frac{(1-\mu_1)m}{2(1-\mu_2)} + \frac{1}{2} \tag{20}$$

Integral constant $A_i$ , $B_i$ , $C_i$ , $D_i$ is related to $\bar{q}(\xi)$ . If the distribution equation of load $q(r)$ is known, components of stress and displacement is obtained by substituting $\bar{q}(\xi)$ into Eq. (18).

If the load is axial symmetrical and circular distribution of vertical loads, the distribution function of load circular range with assuming radius $\delta$ is expressed as:

$$q(r) = \begin{cases} q_\delta(r) & (0 \le r \le \delta) \\ 0 & (\delta < r < \infty) \end{cases} \tag{21}$$

Accordingly,

$$\bar{q}(\xi) = \int_0^\delta r q_\delta(r) J_0(\xi r) dr \tag{22}$$

In particular, if load circular is uniform $q_\delta(r) = q$ ,it is shown that

$$\bar{q}(\xi) = q\frac{\delta}{\xi} J_1(\xi\delta) \tag{23}$$

In order to separate $\bar{q}(\xi)$ from the integral constant expression, assuming $A_i = \tilde{A}_i \bar{q}(\xi)$ , $B_i = \tilde{B}_i \bar{q}(\xi)$ , $C_i = \tilde{C}_i \bar{q}(\xi)$ , $D_i = \tilde{D}_i \bar{q}(\xi)$ in Eq. (18). If adding a new integral variable $x = \xi\delta$ and considering $r$ , $z$ into dimensionless form $\frac{r}{\delta}$ , $\frac{z}{\delta}$ , respectively. the final expression of the stress strain and displacement component is subjected to the vertical load is obtained by considering Eq. (23).

$$\left.\begin{aligned}
\sigma_{ri} &= q\int_0^\infty J_0\left(\frac{r}{\delta}x\right) J_1(x)\left\{\left[\tilde{C}_i + \left(1+2\mu_i + \frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x} - \left[\tilde{D}_i - \left(1+2\mu_i - \frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx - U_i \\
\sigma_{\theta i} &= 2\mu_i q\int_0^\infty J_0\left(\frac{r}{\delta}x\right) J_1(x)(\tilde{A}_i e^{\frac{z}{\delta}x} + \tilde{B}_i e^{-\frac{z}{\delta}x})dx + U_i \\
\sigma_{zi} &= -q\int_0^\infty J_0\left(\frac{r}{\delta}x\right) J_1(x)\left\{\left[\tilde{C}_i - \left(1-2\mu_i - \frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x} - \left[\tilde{D}_i + \left(1-2\mu_i + \frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx \\
\tau_{zri} &= -q\int_0^\infty J_1\left(\frac{r}{\delta}x\right) J_1(x)\left\{\left[\tilde{C}_i + \left(2\mu_i + \frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x} + \left[\tilde{D}_i - \left(2\mu_i - \frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx \\
u_i &= \frac{1+\mu_i}{E_i} q\delta\int_0^\infty \frac{J_1\left(\frac{r}{\delta}x\right) J_1(x)}{x}\left\{\left[\tilde{C}_i + \left(1+\frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x} - \left[\tilde{D}_i - \left(1-\frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx \\
w_i &= -\frac{1+\mu_i}{E_i} q\delta\int_0^\infty \frac{J_0\left(\frac{r}{\delta}x\right) J_1(x)}{x}\left\{\left[\tilde{C}_i - \left(2-4\mu_i - \frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x} + \left[\tilde{D}_i + \left(2-4\mu_i + \frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx
\end{aligned}\right\} \tag{24}$$

Where

$$U_i = q\int_0^\infty \frac{J_1(\frac{r}{\delta}x)J_1(x)}{\frac{r}{\delta}x}\left\{\left[\tilde{C}_i+\left(1+\frac{z}{\delta}x\right)\tilde{A}_i\right]e^{\frac{z}{\delta}x}-\left[\tilde{D}_i-\left(1-\frac{z}{\delta}x\right)\tilde{B}_i\right]e^{-\frac{z}{\delta}x}\right\}dx$$

$$\tilde{A}_1 = -\frac{e^{-2\frac{h}{\delta}x}}{\Delta}[(N-1-\frac{h}{\delta}x)-(N-1)e^{-2\frac{h}{\delta}x}]$$

$$\tilde{B}_1 = -\frac{1}{\Delta}[N-(N-\frac{h}{\delta}x)e^{-2\frac{h}{\delta}x}]$$

$$\tilde{C}_1 = \frac{e^{-2\frac{h}{\delta}x}}{\Delta}[(2\mu_1+\frac{h}{\delta}x)(N-1-\frac{h}{\delta}x)+N\frac{h}{\delta}x-2\mu_1(N-1)e^{-2\frac{h}{\delta}x}] \tag{25}$$

$$\tilde{D}_1 = -\frac{1}{\Delta}\{2\mu_1 N-[2N(\mu_1-\frac{h}{\delta}x)+(1-2\mu_1+\frac{h}{\delta}x)\frac{h_1}{\delta}x]e^{-2\frac{h}{\delta}x}\}$$

$$\tilde{B}_2 = -\frac{1}{\Delta}(2N-1)[(1+\frac{h}{\delta}x)-(1-\frac{h}{\delta}x)e^{-2\frac{h}{\delta}x}]$$

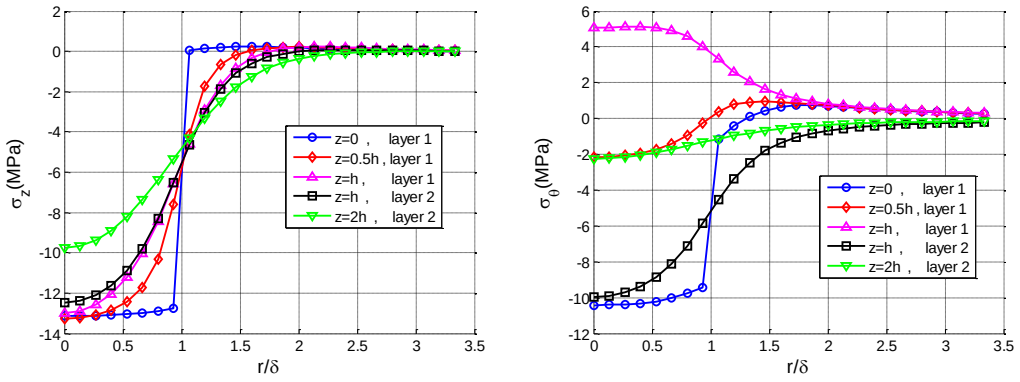
$$\tilde{D}_2 = -\frac{1}{\Delta}(2N-1)(2\mu_2-\frac{h}{\delta}x)[(1+\frac{h}{\delta}x)-(1-\frac{h}{\delta}x)e^{-2\frac{h}{\delta}x}]$$

$$\tilde{A}_2 = \tilde{C}_2 = 0$$

$$\Delta = N+[2\frac{h}{\delta}x(2N-1)-(1+2\frac{h^2}{\delta^2}x^2)]e^{-2\frac{h}{\delta}x}-(N-1)e^{-4\frac{h}{\delta}x}] \tag{26}$$

### *3.4 Calculation Example*

Assuming first layer's parameters $h=9mm$, $\delta=15mm$, Specimen's material $Q235$, Young's modulus $E_1=E_2=2\times10^5 MPa$, Poisson's ratio $\mu_1=\mu_2=0.3$. The vertical load on the surface is $Q=9000N$, $q=Q/\pi\delta^2$. The stress and displacement field are calculated by Eq. (24), as shown in Fig. 5.




(a) (b)

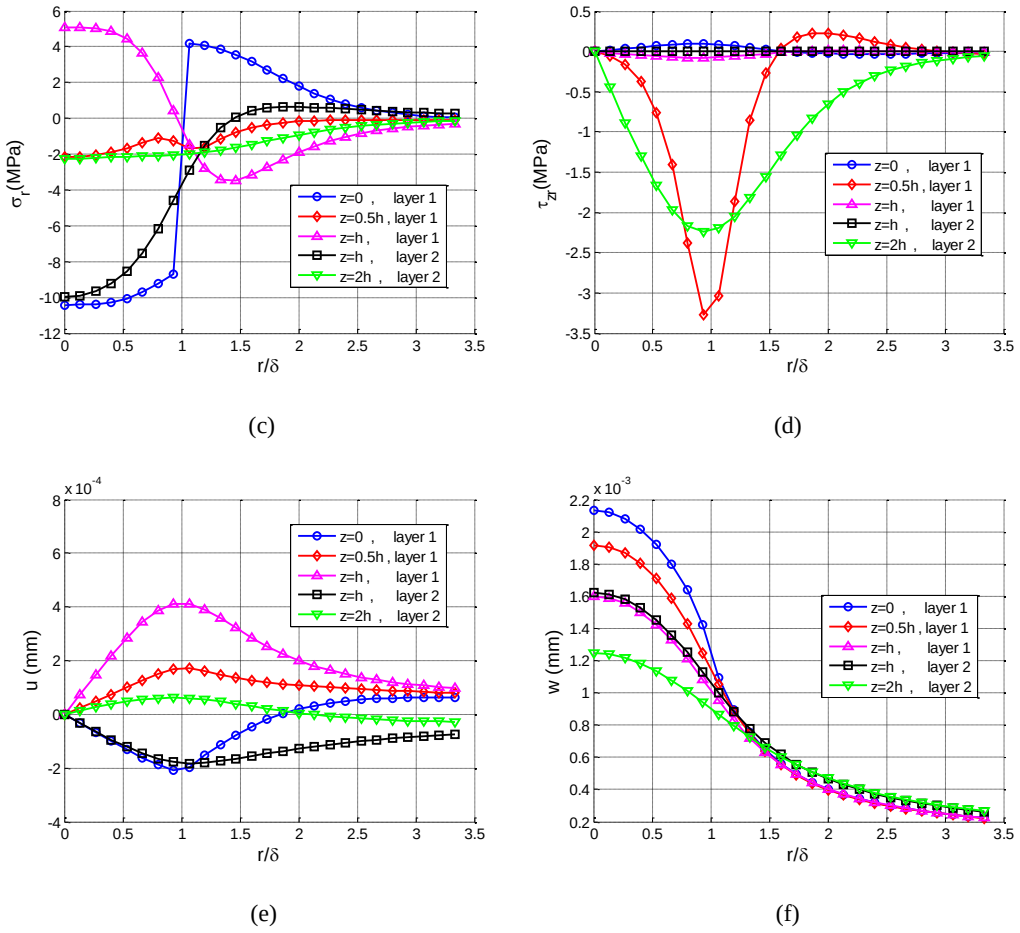
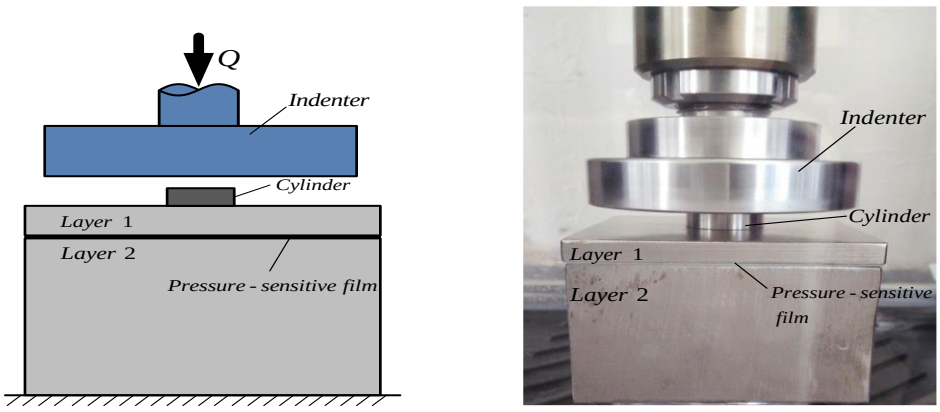



**Fig. 4. Stress and displacement distribution curves along radial direction**

In Fig. 4(a), the curve $z=0$ mutates at r= $\delta$ and when r> $\delta$ , $\sigma_z=0$ ; when r< $\delta$ , $\sigma_z=-q$ , satisfying surface stress boundary condition. The curve $z=h,\ layer\ 1$ and curve $z=h_1,\ layer\ 2$ are overlapped, corresponding with interlayer stress boundary condition $\sigma_{z1}=\sigma_{z2}$ . In Fig. 4(f), curve $z=h,layer\ 1$ and curve $z=h,\ layer\ 2$ are overlapped, satisfying interlayer displacement boundary condition $w_1=w_2$ . In Fig. 4 (d), the value of curve $z=0,layer\ 1$ and curve $z=h,layer\ 1$ are close to 0, satisfying shear stress equals to 0 in surface and interlayer. When $r$ gradually increasing, all the components in Fig. 4 will tend to zero, corresponding to the condition when r approaches infinite, all of components will be zero. These results prove the reasonability of LET.

## 4 Experiment and finite element analysis

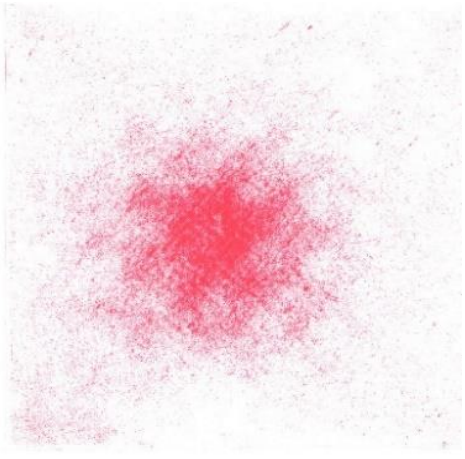
In order to verify the LET's effectiveness in the calculation of stress /displacement field of structural discontinuity, we extract the normal stress on interface and compare it with the results of experimental and finite element analysis.




(a) (b)

**Fig. 5. Experimental schematic diagram and field photos.**

Taking two cuboid metals as experimental specimen and both the material of them are $Q235$,Young's modulus $E_1 = E_2 = 2\times10^5 MPa$, Poisson's ratio $\mu_1 = \mu_2 = 0.3$. We denote the size of metal block as $100mm\times100mm\times9mm$ in the first layer and the size of another metal block as $100mm\times100mm\times60mm$. The contact region between two metals were grounded to $R_a = 0.8\mu m$. The contact stress was measured by the pressure-sensitive film, which is placed on the interface. The cupping machine was used to provide load and the load was transformed into circular load by a metal cylinder of radius $\delta = 15mm$, high $10mm$, placed on the top of the first layer. At first, the cupping machine loaded downwards slowly until the pressure reaches $9000N$. Then the pressure was kept for some time so that the color of pressure-sensitive film was fully displayed. Photos of experimental schematics and site are shown in Fig. 5.



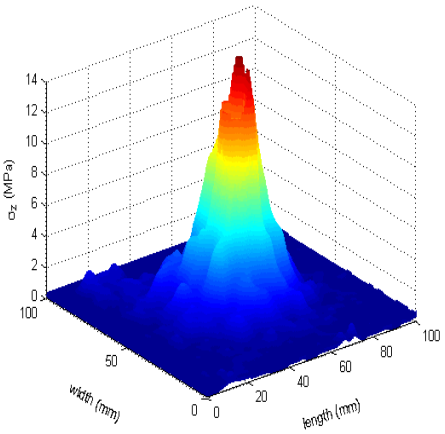



**Fig. 6. Pressure-sensitive film** **Fig.7. Pressure distribution of interface.**

Because of the pressure, the white pressure-sensitive film turns red, and the red concentration increases with the increase of pressure intensity. The contact stress can be measured by evaluating the color concentration of the film. Fig. 6 shows the scanning image of the pressure-sensitive film after experiment. Obviously, the color distributed as an axisymmetric pattern. The high color concentration in the center indicates that the contact pressures are large. The color concentration rapidly decreases in the region far from center, indicating that the pressure drops quickly. After denoising and fliting, the color density of film was converted into pressure value

and a three-dimensional image which displaying pressure distribution of interface in Fig. 6 and Fig. 7. The “steep peak” distribution of the contact pressure is shown clearly in the figure.

Due to machining error and measurement error, the result in Fig. 6 is not absolutely axisymmetric. In order to eliminate the impact of these errors, we used the average value of axial pressure to represent the pressure in the radial.

Fig. 8 shows the contact stress distribution of LET, experimental and finite element method (FEM). Because the existing of pressure-sensitive film, the state of contact pressure on interface has been changed in a way. There were two comparative analysis of FEM with different contact conditions. FEM (with film) adds a new layer which has same mechanical parameter with film between the original layers and FEM (without film) contacts directly. The value of pressure in the figure is negative, indicating the compressive stress.

In Fig. 8, four pressure distribution curves basically coincide. The curve of LET and FEM (without film) almost overlapped since both of them regardless the influence of film layer. The curve of experimental and FEM (with film) is lower in center and higher in edges than LET and FEM (without film). It indicates that the presence of films has a certain effect on the state of mechanical interface. In general, the pressure distribution of experimental and finite element analysis is consistent with the result of layered elastic system.

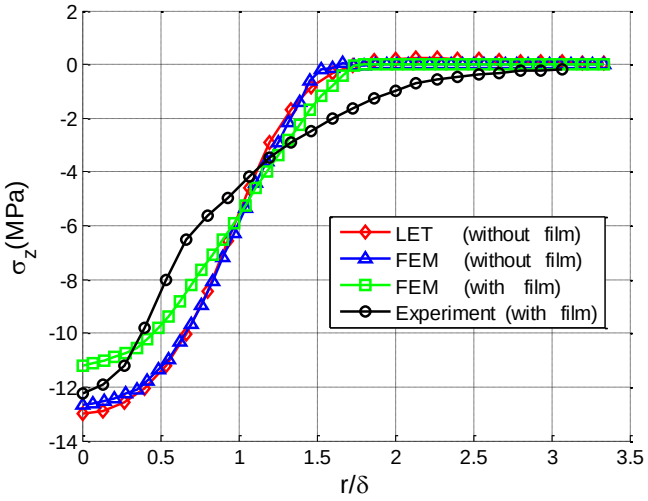



**Fig. 8. Data comparison of contact stress**

## 5 Conclusions

A stress/displacement field calculation model based on layered elastic mechanics is established to solve the mechanical calculation problem in the discontinuous structure in joints.

The calculation model based on layered elastic mechanics is a new way to calculate the stress / displacement field in joints. This method can be extended to more mechanical structure, considering different types of load, friction and other complex conditions and it is effective to analyze problems such as pressure distribution and small sliding of interface.

The model still has some shortcomings: For instance, because of ignoring the and roughness, flatness and waviness of the contact surfaces, there will be a deviation between the calculation

results and the actual situation to some extent. The analytical model of the mechanics characteristics of the complex geometry parts subjected to non-axisymmetric loads or horizontal load (unidirectional load, rotational load) needs further study.

**References**


[1] Ibrahim R A, Pettit C L. Uncertainties and dynamic problems of bolted joints and other fasteners. Journal of sound and Vibration, 279(3), 857-936.

[2] Chen, W., & Deng, X. Structural damping caused by micro-slip along frictional interfaces. International Journal of Mechanical Sciences 47(8), 1191-1211 (2016).

[3] Bosman, R., Hol, J., & Schipper, D. J. Running-in of metallic surfaces in the boundary lubrication regime. Wear, 271(7), 1134-1146 (2011).

[4] Burdekin, M., Back, N., & Cowley, A. Analysis of the local deformations in machine joints. ARCHIVE Journal of Mechanical Engineering Science 1959-1982 (vols 1-23), 21(1), 25-32 (1979).

[5] Yigit, A. S., & Ulsoy, A. G. Dynamic stiffness evaluation for reconfigurable machine tools including weakly non-linear joint characteristics. Proceedings of the Institution of Mechanical Engineers Part B Journal of Engineering Manufacture, 216(1), 87-101 (2002).

[6] Back N, Burdekin M, Cowley A. Review of the research on fixed and sliding joints. In 13th International Machine Tool Design and Research Conference, 87-97. MacMillan, London, (1973).

[7] Beards C F. The damping of structural vibration by controlled interfacial slip in joints. Journal of Vibration and Acoustics, 105(3), 369-373 (1983).

[8] Beards C F, Woowat A. The control of frame vibration by friction damping in joints. Journal of Vibration and Acoustics, 107(1), 26-32 (1985).

[9] Beards C F. Damping in structural joints. Vibration Inst., The Shock and Vibration Digest, 21(4), 3-5 (1989).

[10] Fu W P, Huang Y M, Zhang X L. Experimental investigation of dynamic normal characteristics of machined joint surfaces. ASME Journal of Vibration and Acoustics, 122(4), 393 (2000).

[11] Zhang X L, Wen S H, Lan G S, et al. Experiment research on tangential dynamic characteristics of machined plane joint interfaces. In Advanced Materials Research., 145: 584-589 (2011).

[12] Mi L, Yin G, Sun M, et al. Effects of preloads on joints on dynamic stiffness of a whole machine tool structure. Journal of mechanical science and technology, 26(2), 495-508 (2012).

[13] Shimizu S, Kabaya Y, Sakamoto H, et al. Identification Method of Dynamic Characteristics of Joints in Jointed Structures. Journal ref: International Journal of Automation Technology, 7(2), 221-227 (2013).

[14] Guo T, Li L, Cai L, et al. Alternative method for identification of the dynamic properties of bolted joints. Journal of mechanical science and technology, 26(10), 3017-3027 (2012).

[15] Tian H, Li B, Liu H, et al. A new method of virtual material hypothesis-based dynamic modeling on fixed joint interface in machine tools. International Journal of Machine Tools and Manufacture, 51(3), 239-249 (2011).

[16] Mase G T, Smelser R E, Mase G E. Continuum mechanics for engineers. CRC press, (2011).

[17] Burmister D M. The general theory of stresses and displacements in layered systems. I. Journal of applied physics, 16, 89-94 (1945).

[18] Burmister D M. The general theory of stresses and displacements in layered soil systems. II. Journal of applied Physics, 16, 126-127 (1945).

[19] Burmister D M. The general theory of stresses and displacements in layered soil systems. III. Journal of applied Physics, 16, 296-302 (1945).

[20] Muki R. A three-dimensional problem of a semi-infinite elastic solid under the compressive action of a rigid body. Proceedings of the Fujihara Memorial Faculty of Engineering Keio University, 8(30), 68 (8)-80 (20) (1955).

[21] Kai W. Analysis and calculation of stresses and displacements in layered elastic systems. Acta Mechanica Sinica, 3(3), 251-260 (1987).

[22] Maina J W, Matsui K. Developing software for elastic analysis of pavement structure responses to vertical and horizontal surface loadings. Transportation Research Record Journal of the Transportation Research Board, 1896(1), 107-118 (2004).

[23] Wang D, Roesler J R, Guo D Z. Innovative algorithm to solve axisymmetric displacement and stress fields in multilayered pavement systems. Journal of Transportation Engineering, 137(4), 287-295 (2010).

# A Bézier Extraction based XIGA Approach for Vibration Analysis of Cracked FGM Plate using Simple First-Order Shear Deformation Theory

**†S. K Singh[1], *I. V. Singh[1], and B. K. Mishra[1]**

[1]Department of Mechanical and Industrial Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India.

†Presenting author: ssingh3@me.iitr.ac.in

*Corresponding author: indrafme@iitr.ac.in

## Abstract

An extended isogeometric analysis (XIGA) approach based on Bézier extraction and simple first-order shear deformation theory (S-FSDT) is proposed for the free vibration analysis of cracked functionally graded material (FGM) plate. The XIGA relies on the concept of partition of unity to model a crack. By decomposing the NURBS basis functions into Bernstein basis functions and Bézier extraction operator, the implementation of XIGA becomes simple. The S-FSDT uses four parameters for displacement field approximation which overcomes the shear-locking and captures the shear deformation effect. The S-FSDT requires $C^1$ continuity which is easily achieved through non-uniform rational B-spline (NURBS) basis functions. The material properties of the FGM vary by power law along the thickness of plate. Several numerical examples are solved to validate the accuracy of the proposed approach. The effects of various parameters such as length to thickness ratio, crack length and boundary conditions are investigated on the natural frequencies and mode shapes.

**Keywords**: XIGA; Bézier extraction; NURBS; FGM; Vibration

## 1. Introduction

Functionally graded material (FGM) is a class of composite material made by mixing the two different material phases such as ceramic and metal. Unlike composite material, the material properties of the FGM vary smoothly and continuously in a certain direction and able to avoid the inter-laminar stresses and debonding phenomenon. These advantageous features of FGM are extensively used in variety of engineering applications [1]. In order to ensure the reliability of components made from functionally graded materials (FGMs), it is essential to analyze their behavior in the presence of crack, which can be done by evaluating the static and dynamic behavior, of few standard crack problems. Over the years, several researchers have performed the vibration analysis of cracked plates using different numerical techniques and plate theories. Guan-Liang *et al*. [2] employed the finite element method (FEM) to perform the free vibration analysis of cracked square plate based on the classical plate theory (CPT). Bachene *et al*. [3]

uses extended finite element method (XFEM) in context of first-order shear deformation theory (FSDT) to investigate the free vibration behavior of cracked homogenous rectangular and square plates. Further, Natarajan *et al*. [4] explored the XFEM based on FSDT to study the free vibration analysis of cracked FGM plate. Huang *et al*. [5] used the Ritz method and 3D elasticity theory to perform the free vibration analysis of cracked rectangular FGM plates.

In the present study, a simple first-order shear deformation theory (S-FSDT) is utilized for the free vibration analysis of cracked FGM plates. The S-FSDT model requires four parameters for displacement field approximation and completely overcomes the shear locking effect associated with the original FSDT model [6]. Moreover, S-FSDT model requires $C^1$ continuity of generalized displacement field which cannot be easily attainable using lower order Lagrangian shape functions. However, this necessity is easily attainable by the NURBS basis functions utilized by isogeometric analysis (IGA) [7]. Moreover, in order to capture the discontinuities in the domain, partition of unity (PU) enrichment functions are incorporated with IGA approximation and called as extended isogeometric analysis (XIGA) [8]. Over the years, XIGA is widely used for solving the stationary and propagating cracks in 2D [9], 3D [10], cracked plates [11, 12] and shell structures [13]. Furthermore, the implementation of XIGA can be further simplified by incorporating the Bézier extraction approach [10]. Recently, Tan *et al*. [14] employed XIGA based on Bézier extraction using refined plate theory for the free vibration analysis of cracked FGM plates. Hence, the present work aims to extend the XIGA based on Bézier extraction and S-FSDT for the free vibration analysis of cracked FGM plates. Numerous examples are solved to validate the accuracy of the proposed approach and the obtained results are compared to other published results.

### *2.1 Functionally Graded Plates*

Let us consider a ceramic-metal functionally graded plate of uniform thickness *h*. The upper surface of the plate is assumed to be ceramic rich whereas the bottom surface is fully composed of the metal. As shown in the Fig. 1, the *x*-*y* plane is assumed as the mid-plane of the plate, and the positive *z*-axis is directed above from the mid-plane. Moreover, along the thickness direction (*z*) of the plate, Young's modulus and density are varied using power law as [6],

$$E(z) = E_m + (E_c - E_m)\left(\frac{1}{2} + \frac{z}{h}\right)^n \tag{1}$$

$$\rho(z) = \rho_m + (\rho_c - \rho_m)\left(\frac{1}{2} + \frac{z}{h}\right)^n \tag{2}$$

where, *n* refers as gradient index and subscripts *m* and *c* denote the metal and ceramic constituents, respectively.

**Table 1**: Material properties of FGM plate [5]

| | $E$ (GPa) | $v$ | $\rho\left(kg/m^3\right)$ |
|---|---|---|---|
| Aluminum (Al) | 70 | 0.30 | 2702 |
| Alumina ($Al_2O_3$) | 300 | 0.30 | 3800 |

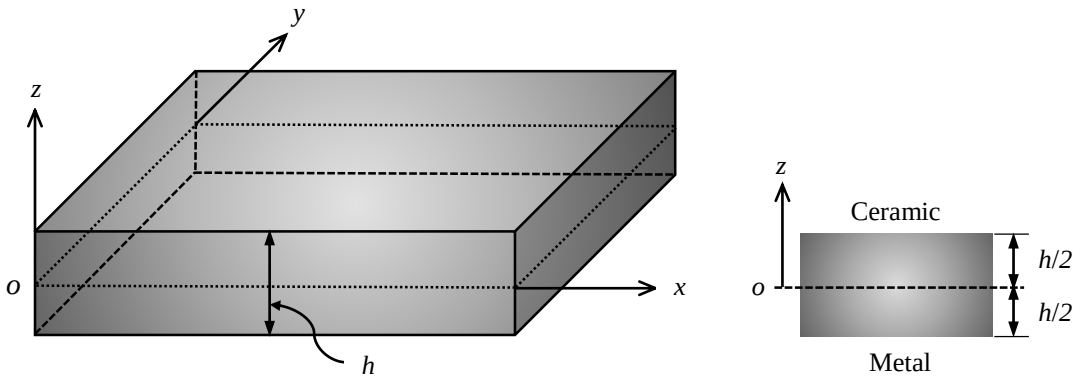



**Figure 1**: A schematic of cracked FGM plate

### *2.2 Simple First-Order Shear Deformation Plate Theory*

The displacement field at any point ($x$, $y$, $z$) in the plate based on S-FSDT is given as [6],

$$\begin{aligned} u(x,y,z) &= u_o(x,y) - z\frac{\partial w_b(x,y)}{\partial x} \\ v(x,y,z) &= v_o(x,y) - z\frac{\partial w_b(x,y)}{\partial y} \\ w(x,y,z) &= w_b(x,y) + w_s(x,y) \end{aligned} \tag{3}$$

where, $u_o$ and $v_o$ represent the mid-plane displacements in $x$ and $y$ directions respectively; $w_b$ and $w_s$ represent the bending and shear components of transverse displacement ($w$), respectively.

Assuming the small strain condition, the non-zero strains are related with the displacement field given in Eq. (3) as,

$$\begin{Bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\gamma} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{\varepsilon}_0 - z\boldsymbol{\kappa} \\ \boldsymbol{\gamma}_s \end{Bmatrix} \tag{4}$$

where,

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \gamma_{xy} \end{Bmatrix},\ \boldsymbol{\gamma} = \begin{Bmatrix} \gamma_{xz} \\ \gamma_{yz} \end{Bmatrix},\ \boldsymbol{\varepsilon}_0 = \begin{Bmatrix} \dfrac{\partial u_0}{\partial x} \\ \dfrac{\partial v_0}{\partial y} \\ \dfrac{\partial v_0}{\partial y} + \dfrac{\partial u_0}{\partial x} \end{Bmatrix},\ \boldsymbol{\kappa} = \begin{Bmatrix} \dfrac{\partial^2 w_b}{\partial x^2} \\ \dfrac{\partial^2 w_b}{\partial y^2} \\ 2\dfrac{\partial^2 w_b}{\partial x \partial y} \end{Bmatrix} \text{ and } \boldsymbol{\gamma}_s = \begin{Bmatrix} \dfrac{\partial w_s}{\partial x} \\ \dfrac{\partial w_s}{\partial y} \end{Bmatrix}$$

The relationship between the stress and strain are related by the following equation as,

$$\boldsymbol{\sigma} = \mathbf{Q}(z)(\boldsymbol{\varepsilon}_0 - z\boldsymbol{\kappa}),\ \boldsymbol{\tau} = \mathbf{G}(z)\,\boldsymbol{\gamma} \tag{5}$$

$$\boldsymbol{\sigma} = \begin{Bmatrix}\sigma_{xx} & \sigma_{yy} & \sigma_{xy}\end{Bmatrix}^T,\ \boldsymbol{\tau} = \begin{Bmatrix}\tau_{xz} & \tau_{yz}\end{Bmatrix}^T,\ \mathbf{Q}(z) = \frac{E(z)}{1-\nu^2}\begin{bmatrix}1 & \nu & 0\\ \nu & 1 & 0\\ 0 & 0 & (1-\nu)/2\end{bmatrix} \text{ and } \mathbf{G}(z) = \frac{k\,E(z)}{2(1+\nu)}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

where, $k$ is the shear correction factor (SCF). In the present work, SCF is taken as $k = 5/6$. Next, using the Hamilton principle, the weak form for free vibration analysis of a FGM plate can be expressed as,

$$\int_\Omega \delta\boldsymbol{\varepsilon}_b^T \mathbf{D}^b \boldsymbol{\varepsilon}_b\, d\Omega + \int_\Omega \delta\boldsymbol{\gamma}_s^T \mathbf{D}^s \boldsymbol{\gamma}_s\, d\Omega = \int_\Omega \delta\tilde{\mathbf{u}}\,\mathbf{m}\,\ddot{\tilde{\mathbf{u}}}\, d\Omega \tag{6}$$

where, $\boldsymbol{\varepsilon}_b = \begin{bmatrix}\boldsymbol{\varepsilon}_0\\ \boldsymbol{\kappa}\end{bmatrix}$, $\mathbf{D}^b = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{D}\end{bmatrix}$, $\mathbf{D}^s = \int_{-h/2}^{h/2} \mathbf{D}_s(z)\,dz$, $\mathbf{m} = \begin{bmatrix}I_0 & I_1\\ I_1 & I_2\end{bmatrix}$ and $\tilde{\mathbf{u}} = \begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix}$

with $\mathbf{A},\mathbf{B},\mathbf{D} = \int_{-h/2}^{h/2}(1, z, z^2)\mathbf{Q}(z)\,dz$, $I_0, I_1, I_2 = \int_{-h/2}^{h/2}(1, z, z^2)\boldsymbol{\rho}(z)\,dz$ and $\boldsymbol{\rho} = \begin{bmatrix}\rho & 0 & 0\\ 0 & \rho & 0\\ 0 & 0 & \rho\end{bmatrix}$

## 3. Bézier Extraction of NURBS

Bézier extraction represents the NURBS basis function over each element in the form of Bernstein polynomial defined over $C^0$ continuous isogeometric Bézier element. Bézier element representation is given by Borden *et al*. [15] for the NURBS and further explored by Scott *et al*. [16] for T-spline. In order to decompose the NURBS basis functions in to Bernstein polynomial basis, Bézier decomposition is used. For more detail interested readers are encouraged to follow these papers [15-17].

## 4. Extended Isogeometric Analysis (XIGA)

The XIGA uses the merits of IGA and partition of enrichment (PU) concept for the fracture analysis of stationary and quasi-static crack growth [8, 9, 14]. In XIGA, the crack is modeled through enrichment functions added in the standard IGA approximation. At a particular point $\mathbf{x} = (x, y)$, the displacement approximation for the crack based on Bézier extraction of NURBS is written as,

$$\mathbf{u}^{\mathbf{h}}(\mathbf{x}) = \sum_{i=1}^{n_{en}} R_i(\mathbf{x})\mathbf{u}_i + \underbrace{\sum_{j=1}^{n_{cf}} R_j(\mathbf{x})\left[H(\mathbf{x}) - H(\mathbf{x}_j)\right]\mathbf{a}_j + \sum_{k=1}^{n_{ct}} R_k(\mathbf{x})\left\{\sum_{\alpha=1}^{4}\left[\beta_\alpha(\mathbf{x}) - \beta_\alpha(\mathbf{x}_k)\right]\mathbf{b}_k^\alpha\right\}}_{\text{Crack}} \tag{7}$$

where, $R_i(\mathbf{x})$ is the NURBS basis functions which is written in the terms of Bernstein polynomial basis functions and Bézier extraction, $n_{en}$ indicates the total number of control points per element and $\mathbf{u}_i = \{u_o, v_o, w_b, w_s\}^T$ indicates the degrees of freedom (DOFs) per control point $i$ in any NURBS element. Moreover, $n_{cf}$ and $n_{ct}$ represent the set of control points associated with all those elements which possess crack face and crack-tip respectively. Additionally, $n_{cf}$ enriched with Heaviside function, $H(\mathbf{x})$ whereas $n_{ct}$ enriched with asymptotic crack tip enrichment functions, $\beta_\alpha(\mathbf{x})$. The $\beta_\alpha(\mathbf{x})$ are taken from the Ref. [14]

Substituting Eq. (7) into Eq. (4), the strains are given as,

$$\begin{bmatrix}\boldsymbol{\varepsilon}_o & \boldsymbol{\kappa} & \boldsymbol{\gamma}_s\end{bmatrix} = \sum_{i=1}^{n_{en}} \begin{bmatrix}\mathbf{B}_i^m & \mathbf{B}_i^b & \mathbf{B}_i^s\end{bmatrix} \{\mathbf{d}_i\} \tag{8}$$

where, $\mathbf{B} = \left[\, \mathbf{B}^{std} \middle| \mathbf{B}^{enr} \,\right]$ and $\mathbf{d} = \{\mathbf{u}, \mathbf{a}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$

Substituting Eq. (8) into Eq. (6), the following form is obtained,

$$\left(\mathbf{K} - \omega^2 \mathbf{M}\right)\mathbf{d} = 0 \tag{9}$$

where, the **K** and **M** are global stiffness and mass matrix, respectively. The expression of **K** and **M** are obtained as provided in Ref. [6].

**5. Results and Discussions**

In this section, the free vibration analysis of cracked FGM plates using S-FSDT in the context of XIGA based on Bézier extraction approach is performed. Several rectangular and square FGM plates having center crack configuration are considered. Unless stated otherwise, ceramic-metal FGM plates whose material properties given in Table 1 are considered. Cubic NURBS basis functions are used in either direction throughout this study, as it provides faster convergence [11]. In all examples, a full integration using $(p+1)\times(q+1)$ Gauss points are used for standard (non-enriched elements) and sub-triangulation scheme for the enriched elements [12]. Moreover, three different boundary conditions are considered on the edges of plate such as; SSSS, FCFF and CFCF, where, S, F and C represent simply supported, free and clamped respectively. The simply supported boundary condition (S) used in this paper is represented as,

$$\begin{aligned} v_o = w_b = w_s = 0, \;\; x = 0, a \\ u_o = w_b = w_s = 0, \;\; y = 0, b \end{aligned} \tag{10}$$

whereas the clamped boundary condition is given as [6],

The percentage difference of normalized natural frequencies is obtained as,

$$\text{Percentage diff.} = \left( \frac{2 \times \left| \bar{\omega}^{\text{Present}} - \bar{\omega}^{\text{Reference}} \right|}{\bar{\omega}^{\text{Present}} + \bar{\omega}^{\text{Reference}}} \right) \times 100 \tag{11}$$

As shown in Fig. 2, the rectangular FGM plate with planar dimension $(a \times b)$ and uniform thickness $h$ containing a through-thickness center crack of length $d$ is considered. Before proceeding to the free vibration analysis of cracked FGM plates, initially a convergence study of the normalized natural frequency of cracked homogeneous plate is performed. A fully simply supported (SSSS) homogeneous rectangular plate with $a/b = 1$, $b/h = 10$, $d/a = 0.3$ and material properties of aluminum alloy as given in Table 1 is considered. The normalized natural frequency $\left( \text{i.e. } \bar{\omega} = \omega \times \frac{b^2}{h} \times \sqrt{\frac{\rho}{E}} \right)$ obtained using S-FSDT and XIGA based on Bézier approach is presented in Table 2. It is observed that normalized frequencies obtained using S-FSDT based XIGA match well with the 3-D elasticity results [5]. Moreover, as the number of control points increases from $32 \times 32$ to $42 \times 40$ the results converge to two significant figures. Hence, for the subsequent examples $32 \times 32$ or more number of control points will be used.

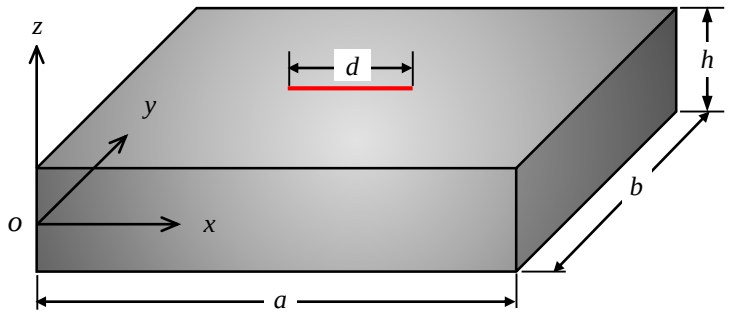



**Figure 2**: A rectangular FGM plate with center crack

To further illustrate the accuracy of proposed method, the normalized natural frequencies is obtained for different $b/h$ and $d/a$ ratios for the SSSS square homogeneous plate. The material properties are taken same as the previous example. Table 3 presents the normalized natural frequencies evaluated using S-FSDT based XIGA are compared with 3D elasticity approach [5]. It is found that the for both thick and thin plates the normalized natural frequencies obtained using present approach are in good agreement with 3D elasticity results. The maximum percentage difference between their results are within 4.21% for Mode 2 with $b/h = 20$ and $d/a = 0.5$.

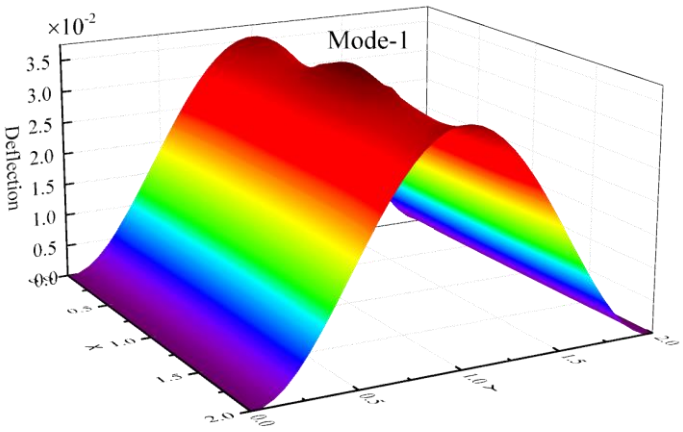
**Table 2**: Normalized natural frequency of SSSS square homogeneous plate ($b/h$ = 10) with center crack ($d/a$ = 0.3)

| Method | Number of control points | Mode | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| S-FSDT based XIGA | $20\times20$ | 5.4532 | 13.4416 | 13.7265 |
| | $24\times24$ | 5.4710 | 13.4095 | 13.7282 |
| | $28\times28$ | 5.5244 | 13.4065 | 13.7373 |
| | $32\times32$ | 5.5234 | 13.3988 | 13.7369 |
| | $36\times36$ | 5.5225 | 13.3962 | 13.7368 |
| | $40\times40$ | 5.5226 | 13.3962 | 13.7369 |
| 3D elasticity [5] | | 5.421 | 13.22 | 13.76 |

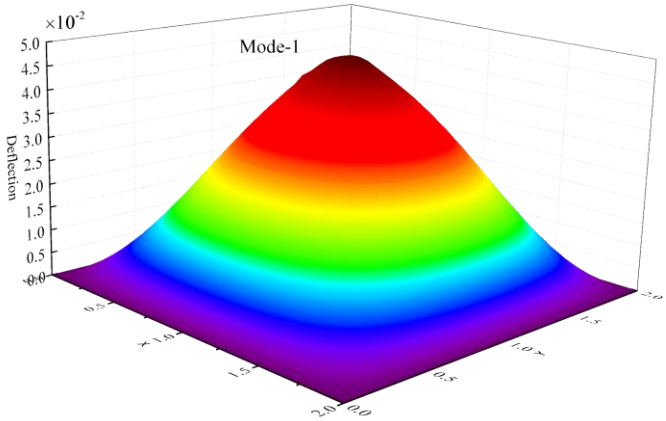
**Table 3:** Normalized natural frequency of SSSS square homogeneous plate with center crack

| *d/a* | *b/h* | Method | Mode | | | | |
|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 |
| 0.3 | 5 | S-FSDT based XIGA | 5.0799 | 11.0129 | 11.5279 | 16.6322 | 18.6180 |
| | | 3D elasticity [5] | 4.960 | 10.84 | 11.61 | 16.64 | 18.06 |
| | | % Difference | 2.39 | 1.58 | 0.71 | 0.05 | 3.04 |
| | 10 | S-FSDT based XIGA | 5.5224 | 13.3887 | 13.7369 | 21.0819 | 23.7780 |
| | | 3D elasticity [5] | 5.421 | 13.22 | 13.76 | 20.97 | 23.13 |
| | | % Difference | 1.85 | 1.27 | 0.17 | 0.53 | 2.76 |
| | 20 | S-FSDT based XIGA | 5.6573 | 14.3151 | 14.5760 | 23.0514 | 26.1400 |
| | | 3D elasticity [5] | 5.590 | 14.21 | 14.57 | 22.94 | 25.62 |
| | | % Difference | 1.2 | 0.74 | 0.04 | 0.48 | 2.01 |
| | 100 | S-FSDT based XIGA | 5.7031 | 14.6623 | 14.8871 | 23.8391 | 27.1033 |
| | | 3D elasticity [5] | 5.701 | 14.65 | 14.89 | 23.82 | 27.11 |
| | | % Difference | 0.04 | 0.08 | 0.02 | 0.08 | 0.02 |
| 0.5 | 5 | S-FSDT based XIGA | 4.8180 | 8.7417 | 11.4390 | 15.4412 | 16.7323 |
| | | 3D elasticity [5] | 4.633 | 8.764 | 11.43 | 15.97 | 16.89 |
| | | % Difference | 3.91 | 0.25 | 0.08 | 3.37 | 0.94 |
| | 10 | S-FSDT based XIGA | 5.2063 | 11.0595 | 13.6018 | 20.7129 | 22.0429 |
| | | 3D elasticity [5] | 5.069 | 11.10 | 13.55 | 20.35 | 21.44 |
| | | % Difference | 2.67 | 0.37 | 0.38 | 1.77 | 2.77 |
| | 20 | S-FSDT based XIGA | 5.3232 | 12.8075 | 14.4199 | 22.7085 | 24.0426 |
| | | 3D elasticity [5] | 5.238 | 12.28 | 14.37 | 22.44 | 23.60 |
| | | % Difference | 1.61 | 4.21 | 0.35 | 1.19 | 1.86 |
| | 100 | S-FSDT based XIGA | 5.3628 | 13.1284 | 14.7227 | 23.5087 | 24.8437 |
| | | 3D elasticity [5] | 5.353 | 12.98 | 14.72 | 23.46 | 24.79 |
| | | % Difference | 0.18 | 1.14 | 0.02 | 0.21 | 0.22 |

Next, an Al/$Al_2O_3$ center cracked square FGM plate is considered and the effect of various parameters such as; different boundary conditions, length to thickness ratio (*b*/*h*) and gradient index (*n*) on normalized natural frequencies is analyzed as presented in Table 4. In this case, the normalized natural frequency is obtained as $\bar{\omega} = \omega \times \frac{b^2}{h} \times \sqrt{\frac{\rho_c}{E_c}}$, where *c* represents the material properties corresponding to ceramic ($Al_2O_3$) in Al/$Al_2O_3$ FGM plate. Table 4 reveals that the normalized natural frequencies obtained using proposed method are well matched with the 3D elasticity results. However, for FCFF boundary condition the maximum percentages in normalized Mode 2 frequency is seen. Besides, the normalized frequencies increase with increasing the *b*/*h* ratios and decreases as gradient index (*n*) increases. It is also observed that the normalized frequencies for FCFF boundary condition is less as compared to SSSS and CFCF boundary conditions. Finally, the contour of first mode shape of square Al/$Al_2O_3$ FGM plate (*b*/*h* = 50) having center crack (*d*/*a* = 0.3) with CFCF and SSSS boundary conditions is shown in Fig. 3.




(a) CFCF




(b) SSSS

**Figure 3**: First mode shape of square Al/$Al_2O_3$ FGM plate (*b*/*h* = 50) having center crack (*d*/*a* = 0.3)

**Table 4:** Normalized natural frequency of $Al/Al_2O_3$ square FGM plate with center crack ($d/a$ = 0.3)

| BCs | *n* | *b/h* | Method | Mode | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 |
| SSSS | 0 | 5 | S-FSDT based XIGA | 5.0799 | 9.7269 | 9.7417 | 11.0129 | 11.5279 |
| | | | 3D elasticity [5] | 4.959 | 9.728 | 9.742 | 10.84 | 11.60 |
| | | | % Difference | 2.41 | 0.01 | 0.00 | 1.58 | 0.62 |
| | | 50 | S-FSDT based XIGA | 5.6973 | 14.6173 | 14.8468 | 23.7352 | 26.9756 |
| | | | 3D elasticity [5] | 5.665 | 14.58 | 14.84 | 23.68 | 26.71 |
| | | | % Difference | 0.57 | 0.26 | 0.05 | 0.23 | 0.99 |
| | 0.2 | 5 | S-FSDT based XIGA | 4.7338 | 9.2672 | 9.2813 | 10.3189 | 10.7870 |
| | | | 3D elasticity [5] | 4.627 | 9.266 | 9.280 | 10.18 | 10.88 |
| | | | % Difference | 2.28 | 0.01 | 0.01 | 1.36 | 0.86 |
| | | 50 | S-FSDT based XIGA | 5.2876 | 13.5678 | 13.7804 | 22.0323 | 25.0406 |
| | | | 3D elasticity [5] | 5.259 | 13.53 | 13.78 | 21.99 | 24.80 |
| | | | % Difference | 0.54 | 0.28 | 0.00 | 0.19 | 0.97 |
| | 5 | 5 | S-FSDT based XIGA | 3.3182 | 6.3156 | 6.3259 | 7.1625 | 7.4900 |
| | | | 3D elasticity [5] | 3.185 | 6.274 | 6.296 | 6.823 | 7.322 |
| | | | % Difference | 4.10 | 0.66 | 0.47 | 4.86 | 2.27 |
| | | 50 | S-FSDT based XIGA | 3.7498 | 9.6192 | 9.7695 | 15.6156 | 17.7484 |
| | | | 3D elasticity [5] | 3.725 | 9.581 | 9.760 | 15.56 | 17.53 |
| | | | % Difference | 0.66 | 0.40 | 0.10 | 0.36 | 1.24 |
| FCFF | 0 | 5 | S-FSDT based XIGA | 1.0164 | 2.4343 | 3.2024 | 5.3565 | 6.7674 |
| | | | 3D elasticity [5] | 1.016 | 2.195 | 3.221 | 5.359 | 6.285 |
| | | | % Difference | 0.04 | 10.3 | 0.58 | 0.05 | 7.39 |
| | | 50 | S-FSDT based XIGA | 1.0494 | 2.5723 | 6.4260 | 7.8198 | 9.3068 |
| | 0.2 | 5 | S-FSDT based XIGA | 0.9445 | 2.2627 | 3.0515 | 5.0015 | 6.3027 |
| | | | 3D elasticity [5] | 0.9441 | 2.049 | 3.069 | 5.010 | 5.869 |
| | | | % Difference | 0.04 | 9.91 | 0.57 | 0.17 | 7.13 |
| | | 50 | S-FSDT based XIGA | 0.9739 | 2.3872 | 5.9641 | 7.2575 | 8.6378 |
| | 5 | 5 | S-FSDT based XIGA | 0.6674 | 1.5880 | 2.0887 | 3.4704 | 4.3767 |
| | | | 3D elasticity [5] | 0.6633 | 1.406 | 2.098 | 3.394 | 3.992 |
| | | | % Difference | 0.62 | 12.2 | 0.44 | 2.23 | 9.19 |
| | | 50 | S-FSDT based XIGA | 0.6907 | 1.6929 | 4.2288 | 5.1463 | 6.1245 |
| CFCF | 0 | 5 | S-FSDT based XIGA | 5.2039 | 6.6281 | 8.8744 | 10.3010 | 12.1163 |
| | | 50 | S-FSDT based XIGA | 6.4333 | 7.9599 | 12.7911 | 18.2150 | 20.1967 |
| | 0.2 | 5 | S-FSDT based XIGA | 4.8705 | 6.1928 | 8.4553 | 9.6195 | 11.3993 |
| | | 50 | S-FSDT based XIGA | 5.9712 | 7.3879 | 11.8719 | 16.9093 | 18.7481 |
| | 5 | 5 | S-FSDT based XIGA | 3.4024 | 4.3168 | 5.7698 | 6.6574 | 7.8857 |
| | | 50 | S-FSDT based XIGA | 4.2342 | 5.2385 | 8.4171 | 11.9857 | 13.2888 |

## 6. Conclusions

In this work, the free vibration analysis of cracked FGM plates using S-FSDT in the context of XIGA based on Bézier extraction approach is successfully performed. The gradation of material properties is taken along the thickness of the plate. The bottom end of the plate possesses 100% alloy while top end possesses 100% ceramic. The material properties (i.e. Young's modulus & density) vary using power law from bottom to top end of the plate. NURBS basis functions obtained from Bézier extraction technique are used for defining the geometric description and solution approximation. The values of normalized frequencies are obtained using present method are found in good agreement with 3D elasticity solutions. Moreover, the normalized natural frequencies are significantly affected by the $b/h$ ratios, crack aspect ratios ($d/a$), gradient index ($n$) and boundary condtions.

## Acknowledgments

The first authors would like to thank Ministry of Human Resource Development (MHRD), Government of India, New Delhi for providing the financial support in the form of fellowship to this work. The authors would like to thank Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, New Delhi for providing financial support under the scheme of International travel support to this work through grant no. ITS/2019/002691. The authors would also like to thank Alumni Association of Indian Institute of Technology, Roorkee for providing partial financial support to this work.

## References

1. Gupta, A. and Talha, M. (2015): Recent development in modeling and analysis of functionally graded materials and structures, *Progress in Aerospace Sciences*, Vol. 79, pp.1-14.
2. Guan-Liang, Q., Song-Nian, G. and Jie-Sheng, J. (1991): A finite element model of cracked plates and application to vibration problems, *Computers & Structures*, Vol 39, pp. 483-487.
3. Bachene, M., Tiberkak, R. and Rechak, S. (2009): Vibration analysis of cracked plates using the extended finite element method, *Archive of Applied Mechanics*, Vol. 79, pp. 249-262.
4. Natarajan, S., Baiz, P.M., Bordas, S., Rabczuk, T. and Kerfriden, P. (2011): Natural frequencies of cracked functionally graded material plates by the extended finite element method, *Composite Structures*, Vol. 93, pp. 3082-3092.
5. Huang, C.S., Yang, P.J. and Chang, M.J. (2012): Three-dimensional vibration analyses of functionally graded material rectangular plates with through internal cracks, *Composite Structures*, Vol. 94, pp. 2764-2776.

6. Yin, S., Hale, J.S., Yu, T., Bui, T.Q. and Bordas, S.P. (2014): Isogeometric locking-free plate element: a simple first order shear deformation theory for functionally graded plates, *Composite Structures*, Vol. 118, pp. 121-138.
7. Hughes, T.J.R., Cottrell, J.A. and Bazilevs, Y. (2005): Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer Methods in Applied Mechanics and Engineering*, Vol. 194, pp. 4135-4195.
8. Bhardwaj, G., Singh, S.K., Singh, I.V., Mishra, B.K. and Rabczuk, T. (2016a): Fatigue crack growth analysis of an interfacial crack in heterogeneous materials using homogenized XIGA, *Theoretical and Applied Fracture Mechanics*, Vol. 85, pp. 294-319.
9. Bhardwaj, G., Singh, I.V. and Mishra, B.K. (2015): Fatigue crack growth in functionally graded material using homogenized XIGA, *Composite Structures*, Vol. 134, pp. 269-284.
10. Singh, S.K., Singh, I.V., Bhardwaj, G. and Mishra, B.K. (2018a): A Bézier extraction based XIGA approach for three-dimensional crack simulations, *Advances in Engineering Software*, Vol. 125, pp. 55-93
11. Singh, S.K., Singh, I.V., Mishra, B.K. Bhardwaj, G. and Singh, S.K. (2018b): Analysis of cracked plate using higher-order shear deformation theory: Asymptotic crack tip fields and XIGA implementation. *Computer Methods in Applied Mechanics and Engineering*, Vol. 336, pp. 594-639.
12. Singh, S.K., Singh, I.V., Mishra, B.K. and Bhardwaj, G. (2019): Analysis of Cracked Functionally Graded Material Plates using XIGA based on Generalized Higher-Order Shear Deformation Theory, *Composite Structures*, https://doi.org/10.1016/j.compstruct.2019.111038
13. Nguyen-Thanh, N., Valizadeh, N., Nguyen, M.N., Nguyen-Xuan, H., Zhuang, X., Areias, P., Zi, G., Bazilevs, Y., De Lorenzis, L. and Rabczuk, T. (2015): An extended isogeometric thin shell analysis based on Kirchhoff–Love theory, *Computer Methods in Applied Mechanics and Engineering*, Vol. 284, pp. 265-291.
14. Tan, P., Nguyen-Thanh, N. and Zhou, K., 2017. Extended isogeometric analysis based on Bézier extraction for an FGM plate by using the two-variable refined plate theory, *Theoretical and Applied Fracture Mechanics*, Vol. 89, pp. 127-138.
15. Borden, M.J., Scott, M.A., Evans, J.A. and Hughes, T.J.R. (2011): Isogeometric finite element data structures based on Bézier extraction of NURBS, *International Journal for Numerical Methods in Engineering*, Vol. 87, pp. 15-47.
16. Scott, M.A., Borden, M.J., Verhoosel, C.V., Sederberg, T.W. and Hughes, T.J.R. (2011): Isogeometric finite element data structures based on Bézier extraction of T-splines, *International Journal for Numerical Methods in Engineering*, Vol. 88, pp. 126-156.
17. Singh, S.K., Singh, I.V., Mishra, B.K., Bhardwaj, G. and Bui, T.Q. (2017): A simple, efficient and accurate Bézier extraction based T-spline XIGA for crack simulations, *Theoretical and Applied Fracture Mechanics*, Vol. 88, pp. 74-96.

# Improved moving particle semi-implicit method with adaptive variable-size particles

***Kai Zhang, †Zhongguo Sun, Xiao Chen, Qixin Liu, and Guang Xi**

School of Energy and Power Engineering, Xi'an Jiaotong University, China

*Presenting author: zk2017@stu.xjtu.edu.cn

†Corresponding author: sun.zg@xjtu.edu.cn

## Abstract

In order to conduct simulations with high accuracy using particle methods, numerous particles with small size are required to increase the resolution of the calculation domain, and furthermore an improved MPS method with variable-size VSP-MPS has been proposed to achieve the object in acceptable time. In this paper, the scheme is improved with adaptive variable-size particle as AVSP-MPS to increase computational efficiency. First, we made linear programming for the selection and changed the decision method of particle entering the different resolution to improve the robustness at coarse/fine interface, as a result, the accuracy of calculation was improved. As the high resolution of the previous multi-resolution MPS method is static, we get a moving high resolution region and proved that the region movement has no adverse effect on the flow field. At last, the scheme with adaptive variable-size particle as AVSP-MPS is introduced, the shape and area of the domains with high resolution can be dynamically adjusted during the calculation. The proposed method was verified by simulating dam-break case with a moving obstacle. The computing time for the cases with and without AVSP-MPS was analyzed to prove its capability on reducing the computational cost.

**Keywords:** adaptive variable-size particles, coalescence, split, moving multi-resolution region, computational efficiency

## 1 Introduction

In classical grid-based computational methods, variable resolution can be easily achieved though refined structured/unstructured grids to improve accuracy in specific computational domains. Particle-based methods such as moving particle semi-implicit (MPS)[1] method is widely used for analyzing unsteady flow with large deformation, however, with the increase of calculation accuracy and scale, it takes large amount of particles to adopt a single resolution, and the calculation is time-consuming. To reduce the computational cost, several methods have been developed in smoothed particle hydrodynamics (SPH)[2] and MPS simulations.

Different with incompressible flow simulation in MPS, SPH is generally used to solve compressible flow, the incompressible flow is usually solved by introducing a weakly compressible scheme (WCSPH). In order to reduce computer time, Omidvar[3] produced a variable particle mass distribution with fine resolution near the body and coarse resolution further away. Though two well-defined test cases of waves generated by a heaving semi-immersed cylinder and progressive waves interacting with a fixed cylinder, the variable mass distribution leads to a computer run speedup of nearly 200%. Feldman[4] proposed a dynamic particle refinement method where candidate particles are split into several 'daughter' particles according to a given refinement pattern. In such a method, the daughter particle properties

such as mass, volume, density, velocity and pressure are chosen so that both energy and mass are conserved. Vacondio[5] modified dynamically the particle sizes by means of splitting and coalescing (merging) individual particles, their simulations have shown that the particle refinement procedure is able to increase the efficiency while maintaining the same level of accuracy as a uniform distribution with the most refined resolution. Chiron[6] presented the basics of an Adaptive Particle Refinement (APR) technique, inspired by Adaptive Mesh Refinement (AMR)in mesh-based methods. This approach ensures robustness at coarse/fine interfaces with alleviated constraints. Sun[7-8] implemented a particle shifting technique (PST) in the framework of $\delta^+$-SPH combining with APR which is a numerical technique adopted to refine the particle resolution in the local region and de-refine particles outside that region. The problems of high computational costs and tensile numerical instability are avoided in $\delta^+$-SPH scheme since APR and Tensile Instability Control (TIC) have been implemented.

Most of the above attempts based on SPH were implemented with the explicit algorithm to produce the pressure field. However, a semi-implicit algorithm is often adopted to obtain the pressure field in MPS method, which need to solve the pressure Poisson equation and makes it much more difficult developing the local refine technique in the MPS than that in the SPH. Shibata[9-10] developed a multi-resolution technique, the overlapping particle technique (OPT), the OPT expresses a whole simulation domain with partially overlapping sub-domains with their own spatial resolutions and particle shape. Despite the improvements, because the mass or volume conservation of the particle size conversion procedure is not thoroughly discussed in OPT, the total mass conservation of the algorithm needs to be studied and guaranteed. Tanaka[11] developed further a multi-resolution technique for the MPS method in two dimensions, however, the formulation was derived for the classical MPS method and thus it suffers from inaccuracy and stability issues. Tang[12] extended this method for three dimensions, however, no splitting or merging algorithms were adopted and therefore the spatial resolution cannot be changed dynamically. Chen[13] developed a multi-resolution MPS method with variable-size particles based on an algorithm for dynamic particle coalescing and splitting. For the existing surface detection technique could not avoid misdetection, the dynamic particle refinement is also incorporated based on an improved MPS method with no surface detection (NSD-MPS)[14]. Besides, a new gradient model associated with different particles sizes is used and all effective radii of particles remain the same ensures the conservation of mass and momentum in VSP-MPS method. Tanaka[15] developed novel boundary conditions for the treatment of wall and pressure boundaries for the multi-resolution least square MPS method, the new boundary condition makes the method easier to be used in flow simulations of channel flows.

However, for all of the above MPS methods, the domains with different sized particles are set before calculation and their location/area are fixed, which makes it cumbersome when simulating the fluid flow cases with a moving object. In this paper, we improved the moving particle semi-implicit method with adaptive variable-size particles as AVSP-MPS. The spatial resolution varies dynamically with the location of the interface between liquid and the moving objects, the different resolution areas don't need to be known beforehand. In order to optimize the splitting and coalescing algorithm, we made linear programming for the selection of resolution and got the optimal resolution interval. In addition, we changed the decision method of particle entering the different resolution to improve the robustness at coarse/fine interfaces. Based on the above improvements, the dam-break case with a moving object was simulated with different methods. Compared with the VSP-MPS method which can only delimit the resolution region in advance, with an adaptive algorithm, AVSP-MPS method can

further reduce the number of particles needed in the simulation and improve the calculation efficiency.

## 2 VSP-MPS Methods

This section recalls the MPS method with variable-size particles, the governing equations for an incompressible flow are:

$$\frac{D\rho}{Dt} + \rho\nabla \cdot \mathbf{u} = 0 \tag{1}$$

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + \mu\nabla^2\mathbf{u} + \mathbf{f} \tag{2}$$

where **u** represents velocity vector, $t$ is time, $\rho$ is the constant density, $p$ is pressure, $\mu$ is dynamic viscosity and **f** is the volumetric force, such as gravity.

### *2.1 Kernel Function*

The cubic spline kernel function, which is usually used in SPH, was used in this paper:

$$w_{ij}(R_{ij}) = \alpha \times \begin{cases} 2/3 - R^2 + 0.5R^3 & (R \le 1) \\ 1/6 * (2-R)^3 & (1 < R \le 2) \\ 0 & (R > 2) \end{cases} \tag{3}$$

$$R = r / h \tag{4}$$

$$\alpha = 15/(7\pi h^2) \tag{5}$$

where $r$ is the distance between neighboring particles, $h = 2.1l_0$ is used for all particles in this paper, $l_0$ is the initial particle size, and $\alpha$ is the normalization coefficient. With the kernel function, the particle number density can be calculated as:

$$n = \sum w_{ij} V_j \tag{6}$$

where $w_{ij}$ is the kernel function between particle $i$ and $j$, and $V_j$ is the volume of particle $j$. The constant particle number density is $n^0 = 1.0$ for standard particle distribution.

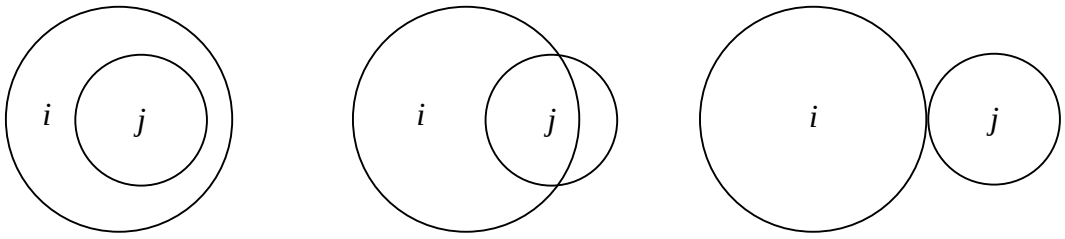
Figure 1 shows three possible cases between two adjacent particles. For example, particle $i$ may overlap with particle $j$ as shown in Figure 2(a), in this case, the density attribution of $j$ to $i$ is $w_{ij}V_j$ using Eq. (3), while the density of $i$ is $\rho(V_i + V_j)/\max(V_i, V_j)$. In order to avoid particle clustering, an additional weight function[13] is introduced:

$$w_{a,ij} = \begin{cases} 0 & (R_{a,ij} > 1) \\ (1 - 1.5R_{a,ij}^2 + 0.5R_{a,ij}^3)^2 / V_{\max} & (1 \ge R_{a,ij} \ge 0) \\ 1/V_{\max} & (R_{a,ij} < 0) \end{cases} \tag{7}$$

$$R_{a,ij} = \frac{r_{ij} - 0.5\left|l_j - l_i\right|}{0.5(l_j + l_i - \left|l_j - l_i\right|)} \tag{8}$$

where the subscript $a$ represents the additional function, $V_{\max} = \max(V_i, V_j)$ is the maximum volume between the two particles (where $V_i = l_i^2$ and $V_j = l_j^2$), $l_i$ and $l_j$ are the diameters of particle $i$ and $j$ respectively.

When $r_{ij} \geq 0.5(l_i + l_j)$ (Figure 2(c)), the additional weight function $w_{a,ij}$ is zero. When $r_{ij} < 0.5\left|l_j - l_i\right|$ (Figure 2(a)), the density attribution of $j$ to $i$ is $V_j / V_{\max}$ , $w_{a,ij} = 1/V_{\max}$ . When $0.5\left|l_j - l_i\right| \leq r_{ij} \leq 0.5(l_i + l_j)$ (Figure 2(b)), a monotone decreasing function $(1-1.5R_{a,ij}^2 + 0.5R_{a,ij}^3)^2$ is used to represent the value of $w_{a,ij}$ from $1/V_{\max}$ to zero.




(a) overlapping particles (b) clustering particles (c) adjacent particles

**Figure 1. Position relationships between two particles**

*2.2 Particle Interaction Models*

A new gradient model is used in this paper to ensure the conservation of momentum[16].

$$\left\langle \nabla \phi \right\rangle_i = \frac{d}{\lambda_1 n^0} \sum_{j \neq i} \left[ \frac{\phi_j + \phi_i}{\left| r_{ij} \right|} \vec{r}_{ij} w_{ij} \right] \tag{9}$$

The parameters $\lambda_1$ is defined as

$$\lambda_1 = \frac{\int_v w(r) r dv}{\int_v w(r) dv} \tag{10}$$

The gradient (9) is totally irrelevant from particles' distance, and the influence from the particles with different sizes could be ignored.

The original Laplace model introduced by Koshizuka[1] is used in this paper

$$\left\langle \nabla^2 \phi \right\rangle_i = \frac{2d}{n^0 \lambda_2} \sum_{j \neq i} \left[ \left( \phi_j - \phi_i \right) w_{ij} \right] \tag{11}$$

The parameters $\lambda_2$ is defined as

$$\lambda_2 = \frac{\int_v w(r) r^2 dv}{\int_v w(r) dv} \tag{12}$$

The pressure Poisson equation introduced by Tanaka[16] is used in this paper:

$$\nabla^2 P_i = (1-\gamma) \frac{\rho}{\Delta t} \nabla \cdot \vec{u}_i^* + \gamma \frac{\rho}{\Delta t^2} \frac{n^0 - n_i^n}{n^0} \tag{13}$$

where $n_i^n$ is the particle number density in the $n^{th}$ steps and $\gamma = 0.008$ is a coefficient.

*2.3 Free Surface boundary Conditions*

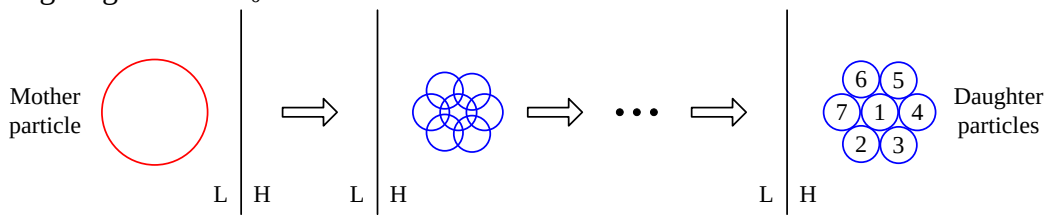
The original MPS method may misjudge surface particles when the solving the pressure Poisson equation with multi-resolution, which has an adverse effect on the accuracy of calculation. Thus, a new algorithm of MPS method with no surface detection (NSD-MPS)[14] can improve computational stability by avoiding surface particle detection. The NSD-MPS method ensures that all real particles have the particle number density more than $n^0$ in the whole region by introducing conceptual particles to compensate for the loss of particle number density. The Dirichlet boundary condition of the pressure Poisson equation is enforced by the conceptual particles, which have the free surface pressure $P_{\text{free}}$, that is to say, the conventional surface particles are replaced by conceptual particles to take the zero-pressure condition. You can get more details in reference [14].

*2.4 Particle Splitting and Coalescing Methodology*

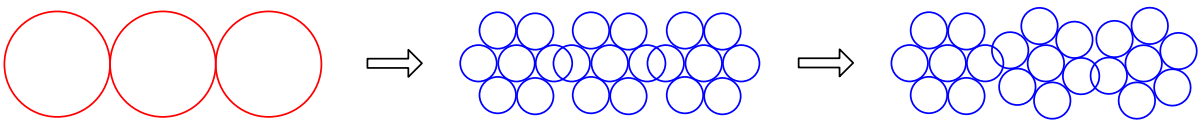
In order to increase the resolution in certain areas of the computational domain, particle splitting and coalescing were involved in the algorithms. In previous methods, the momentum tends to be non-conservation, especially in the process of fine resolution to coarse resolution, particles need to wait or be deleted in the simulation. In VSP-MPS method, we cancel the restriction, different particle sizes are allowed in the region, which can achieve momentum conservation. At the same time, different resolutions are determined by a maximum volume $V_{\max}$ and a minimum volume $V_{\min}$. A particle that is larger than $V_{\max}$ would be split into seven daughter particles, while a particle that is smaller than $V_{\min}$ would be coalesced with a neighboring particle. With several processes of splitting and coalescing, the size of particles in the region is limited to a certain range, which is conducive to controlling the number of particles and improving the computational efficiency.

*2.4.1 Particle splitting*

In particle splitting processes[13], a mother MPS particle splits into 7 daughter particles, and the volumes of the daughter particles are equal to 1/7 of mother particle. In order to meet the Newton's third law, all particles including the new daughter particles share the same smoothing length $h = 2.1l_0$.




(*a*) splitting process



(*b*) random splitting angles to avoid particle clustering

**Figure 2. Large particles splitting into daughter particles**

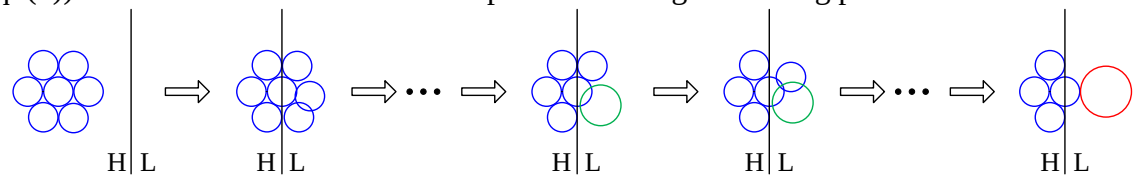
where the solid straight line represents the demarcation line of different resolutions. Here, the left side of the demarcation line is the low resolution area, as expressed with L at the left side, and the right side is the high resolution area, as expressed with H at the right side.

Figure 2(*a*) is the sketch of a process of a particle splitting into seven daughter particles. The daughter particle 1 locates at the same position as the mother particle, and the other six

daughter particles are distributed to form a regular hexagon around particle 1. Daughter particles are distributed with a small overlap in the first splitting process to decrease particle clustering, then the daughter particles move away from other particles with a small distance in the rest steps so that the CFL condition of MPS is met, the splitting process will be finished in five steps. Furthermore, an algorithm will be used to lead the daughter particles to form a random angle as shown in Figure 2(*b*), and to guide the particles moving to proper positions. The algorithm would further reduce the possibility of particle clustering. The velocity vectors of the daughter particles are set equal to that of their mother particle for linear momentum conservation in the five steps during the splitting process. The daughter particles would have no angular velocity at the first step of splitting because the mother particle has no angular momentum.

*2.4.2 Particle coalescing*

Particle coalescing process[13] will conduct between two neighboring particles as shown in Figure 3, fine particles will coalesce with other particles if their volume is smaller than given minimum volume $V_{\min}$. Similar to the splitting process described previously, a five-step coalescing process will be used to avoid particle clustering, and the additional weight function (Eq. (6)) won't be used between the two particles during coalescing process.




Two particle coalescing processes with ten steps

**Figure 3. Sketch of particle coalescing processes**

Considering the incompressibility and mass conservation, the volume of the new particle (new particle is represented by M) $V_M$ is calculated as:

$$V_M = V_i + V_j \tag{14}$$

For momentum conservation, during coalescing, the position and velocity vector of mass center are:

$$\vec{r}_M = \frac{V_i \vec{r}_i + V_j \vec{r}_j}{V_M} \tag{15}$$

$$\vec{v}_M = \frac{V_i \vec{v}_i + V_j \vec{v}_j}{V_M} \tag{16}$$

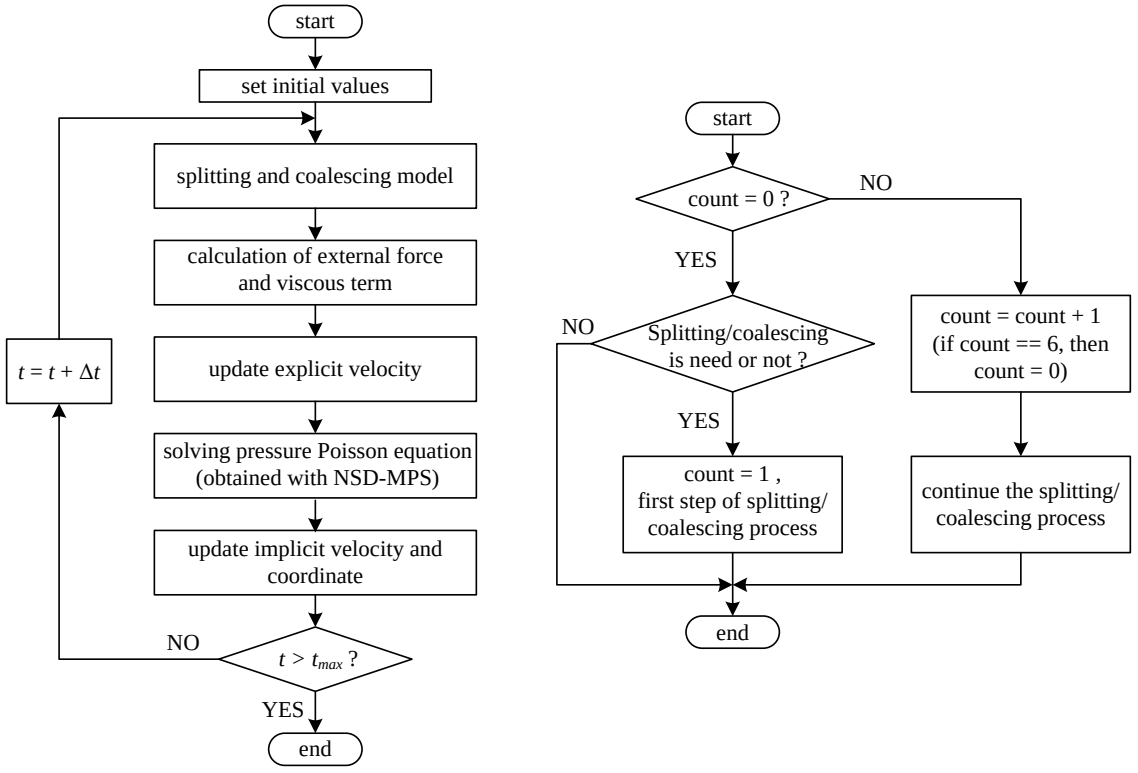
And for angular momentum, the coalescing two particles would have velocities

$$\vec{\theta}_M = \frac{\left(\vec{r}_i - \vec{r}_M\right) \times \left(\vec{v}_i - \vec{v}_M\right) \cdot V_i + \left(\vec{r}_j - \vec{r}_M\right) \times \left(\vec{v}_j - \vec{v}_M\right) \cdot V_j}{\left|\vec{r}_i - \vec{r}_M\right|^2 \cdot V_i + \left|\vec{r}_j - \vec{r}_M\right|^2 \cdot V_j} \tag{17}$$

$$\vec{v}_i = \vec{v}_M + \left(\vec{r}_i - \vec{r}_M\right) \times \vec{\theta}_M \tag{18}$$

$$\vec{v}_j = \vec{v}_M + \left(\vec{r}_j - \vec{r}_M\right) \times \vec{\theta}_M \tag{19}$$

where $\vec{\theta}_M$ is the angular velocity. After the coalescing, the two particles merged into one large particle without angular velocity.




(*a*) VSP-MPS (*b*) the splitting/coalescing model

**Figure 4. Flowcharts of the VSP-MPS method and the splitting/coalescing model**

Figure 4(a) shows the flowchart of the VSP-MPS method, and Figure 4(b) shows the flowchart of the splitting/coalescing model. The number count represents the status of the splitting/coalescing model, count = 0 means no the splitting/coalescing model, and count=1-5 represents the step of the splitting/coalescing process from 1 to 5. To save computational cost, splitting and coalescing processes share a same five-step process, in other words, the splitting and coalescing processes are initiated every five steps by checking their criteria respectively, there is no new splitting and coalescing process during the five steps.

## 3 Numerical Verification

A Dam-break case is simulated to verify the VSP-MPS method along with the particle splitting and coalescing schemes in this section. The case is simulated with single resolution and variable resolution respectively to verify the accuracy of the VSP-MPS method and its effectiveness on improving the computational efficiency.

The initial setup of the dam-break case is shown in Figure 5. $V_{max}$ and $V_{min}$ are the maximum and minimum volumes to control the particle splitting and coalescing processes in our algorithm. The physical parameters of fluid particle used in this case are $\rho$ = 988kg/m$^3$ and $\nu$ = 1.0 × 10$^6$m$^2$/s. In the single resolution case, the initial diameter of water particles is $l_0$ = 0.0019m, the particle is 5672, in which 3200 are fluid particles. In the variable resolution case, the initial diameter of water particles is $l_0$ = 0.005m, the fluid particle will split or coalesce during the simulation, and the diameter of the particles will be 0.0019m after once splitting process.

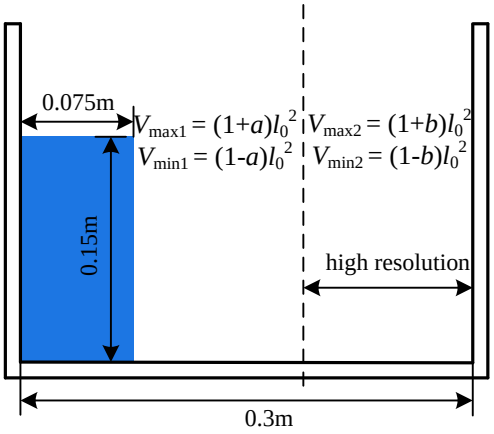



**Figure 5. Geometry of the dam-break case**

*3.1 Repartition of the Resolution*

The resolution is realized by setting the maximum and minimum volume of particles in different regions. In order to improve efficiency, we made linear programming for the selection of resolution and got the optimal resolution interval.

The upper and lower bounds of the intervals in different resolution regions are shown in Figure 5. Although the errors caused by the inconsistency of particle sizes can be mitigated by using the kernel function and the interaction model given in section 2. If the difference between $V_{min}$ and $V_{max}$ is too large in the same resolution region, all the particles meet this region would stay and coalesced would not happen, the number of the particles was large and the calculation efficient was low. On the other hand,, if the difference between $V_{min}$ and $V_{max}$ is too small, the split and coalesce process may happen and repeat until the size of the particle meet the narrow interval. The calculation efficient was also low by carrying out more split/coalesce process.
In order to reduce the number of splitting and coalescing process, after the splitting of the smallest particle and the largest particle, the split particles will not split and coalesce again in the low resolution region.

$$1-a \le \frac{1/7+b+1+a}{7} \le 1+a \tag{20}$$

$$1-a \le \frac{1+a}{7} \le 1+a \tag{21}$$

In high resolution regions, the largest and smallest particles that need to be split do not split and coalesce after once splitting.

$$\frac{1}{7}-b \le \frac{1+a}{7} \le \frac{1}{7}+b \tag{22}$$

$$\frac{1}{7}-b \le \frac{(1+b)/7}{7} \le \frac{1}{7}+b \tag{23}$$

At the same time, in order to maintain the continuity of particle diameter, the $V_{min}$ of low resolution region should be smaller than the $V_{max}$ of high resolution region.

$$1-a \le \frac{1}{7}+b \tag{24}$$

It is certain that $V_{min1}$ and $V_{min2}$ should be greater than zero to ensure the non-negativity of particle volume.

$$1-a\geq 0 \tag{25}$$

$$\frac{1}{7}-b\geq 0 \tag{26}$$

Simultaneous equation (20) - (26) gives the feasibility interval as shown in Figure 6. In order to reduce the difference of particle diameter in the same resolution region, in low resolution region, the particle diameter should be close to $V_0=l_0^2$, and the particle diameter should be close to $V_0/7=l_0^2/7$ in high resolution region, On the basis of satisfying the preceding conditions(equation (20) - (26)), the values of *a* and *b* should be as small as possible. When the weights of *a* and *b* are the same, the optimal solution is a=3/4 and b=6/50(the point marked P as shown in Figure (6)), when the weights of *a* and *b* are different, the optimal solution will not change. So after the repartition of the resolution, we get $V_{max1}=1.75l_0^2$, $V_{min1}=0.25l_0^2$, $V_{max2}=0.263l_0^2$, $V_{min2}=0.0229l_0^2$.

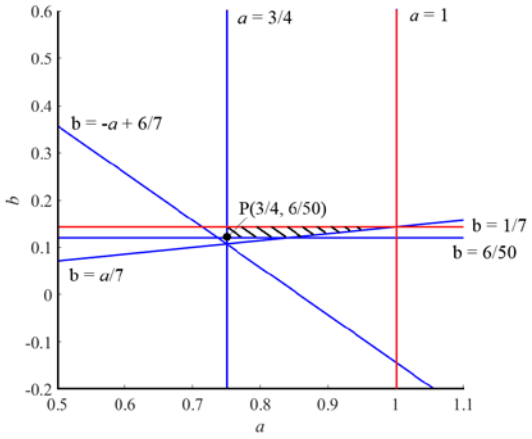



**Figure 6. Optimal solution of resolution repartition**

*3.2 Dam-break test*

Figure 7 shows the experimental and simulation results of the position of water front in the dam-break case. The horizontal axis is the non-dimensional time, $t(2g/L)^{0.5}$, and the vertical axis is the non-dimensional position of the water column's leading edge (Z/L). The VSP-MPS result with variable resolution is compared with the single resolution MPS result and the experimental results by Martin[17]. The simulation results agree with the experimental results, and compared with the single resolution MPS result, the VSP-MPS result has even higher precision in high resolution area(as shown in Fig. 7).

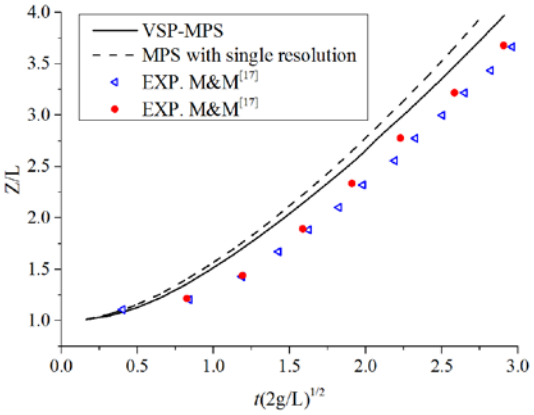


**Figure 7. Experimental and simulation results of the position of water front in the dam-break problem**

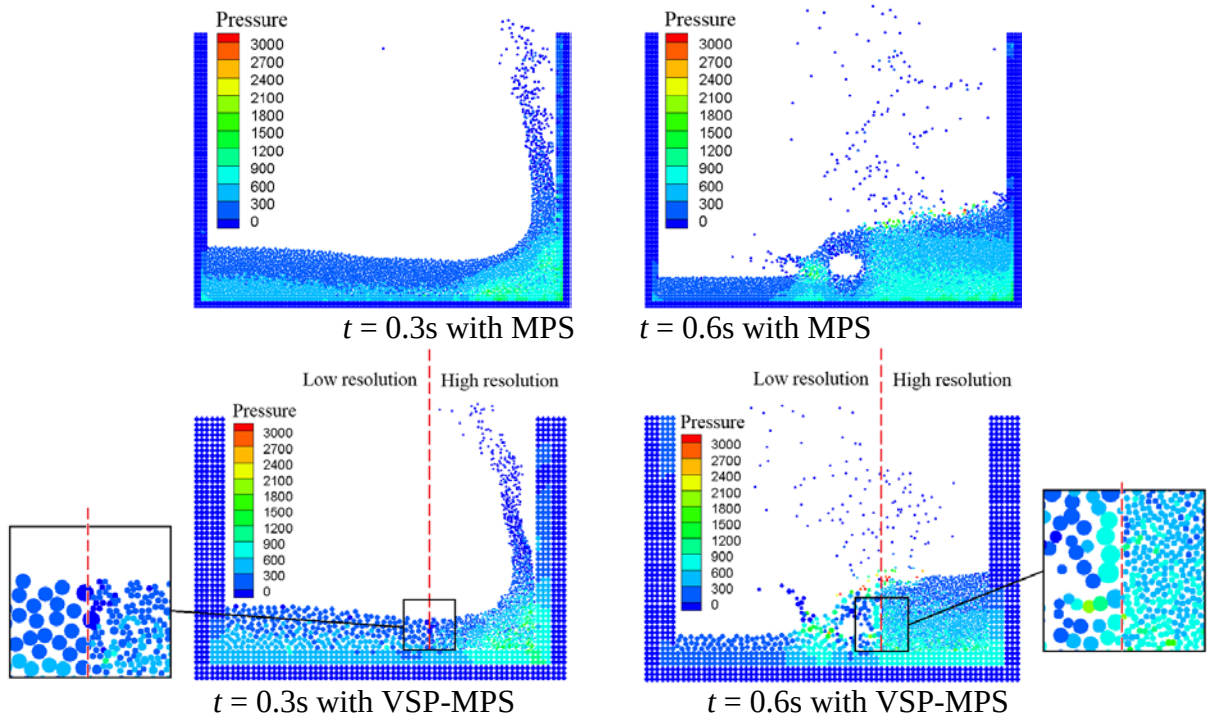
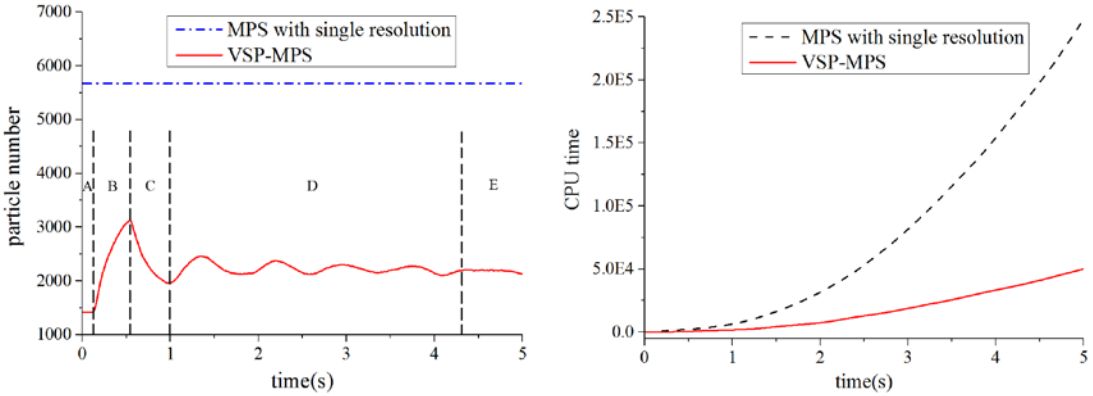



**Figure 8. Snapshots from the dam-break simulation**

Figure 8 shows the pressure distribution of fluid at two typical times with different methods, which shows that the VSP-MPS method can accurately simulate the pressure distribution of fluid in the flow process. In addition, it can be observed that coarse particles can split into fine particles after entering the high resolution region, and fine particles will coalesce with neighboring particles after entering the low resolution region. The splitting and coalescing processes are implemented until the particle volume remains within the given range. It is notable that the splitting and coalescing processes occur in the whole computational domain, instead of just at coarse/fine interface.

Figure 9 shows the comparison of particle number and CPU time with the two different methods. Compared with the single resolution MPS method, the VSP-MPS method requires fewer particles and greatly reduces the calculation time. For example, there are 5672 particles in the dam-break case with single resolution MPS, and the particle number is constant during the simulation. In VSP-MPS method, the particle number is various during the simulation, it can be divide into 5 phases during the simulation (see Figure 9(a)). There are 1422 initial setting particles at the begin of the simulation (region A), when the particle pass through the high resolution, the particle split into finer particles, so the particle number increases in this phase (region B). The particle number decrease in region C due to the fluid flow back into low resolution region and fine particle coalesces with neighboring particle. Then the fluid sloshes between the two solid walls (region D) flowing through the resolution boundary for several times before totally dissipated, At last, the fluid tends to be stable gradually(region E), and the particle number converge to be a constant.




(a) Particle number (b) CPU time

**Figure 9. Comparison of particle number and CPU time with different methods**

As shown in Fig 9(b), the calculation time of VSP-MPS is much short than that of MPS with single resolution. Compared with single resolution MPS method, when the simulation runs to five seconds, the CPU time is reduced by 79.8% using VSP-MPS method. The above simulation results prove that the VSP-MPS method could improve the simulation accuracy and significantly raise the calculation efficiency.

## 4 Numerical Simulation

In VSP-MPS method, the resolution regions are mandatory to be set before the simulation, and the shapes and locations of which are fixed during the whole calculation. It is difficult for a fixed high resolution region to cover a complex flow with moving object.

In this paper, a new scheme AVSP-MPS was proposed based on VSP-MPS, the high resolution region could be adjusted during the calculation and even could tracking the location and shape of the target flow. It is much more flexible for complex flows, and the accuracy and efficiency would both be improved.

### *4.1 Improvement of the Algorithm Accuracy*

Though the VSP-MPS method allows finer particles in specific areas and coarser particles elsewhere, the robustness on coarse/fine interface needs to be improved. In the previous algorithm, the particle will split into finer particles when the center of mass of the particle enters the high-resolution region. As a result, there will be a large number of particles of different diameters at coarse/fine interface, the numerous different diameters at coarse/fine interface will reduce the accuracy of the simulation. In order to improve the robustness at coarse/fine interface, we improved the algorithm by changing the decision method of particle entering the different resolution. The improved approach is that only when the whole particle (the boundary of the particle) enters the high resolution region the particle splitting into finer particles.

In order to validate the improved algorithm, we simulated the dynamic behavior of a single particle passing through the high resolution boundary with the unimproved algorithm and the improved algorithm. The center of the particle is initially located at coarse/fine boundary, and the particle is given a uniform rightward initial velocity $u$=0.1m/s. Gravity is not considered in the whole process. The comparison of particles dynamic behavior before and after improvement is shown in Figure 10.

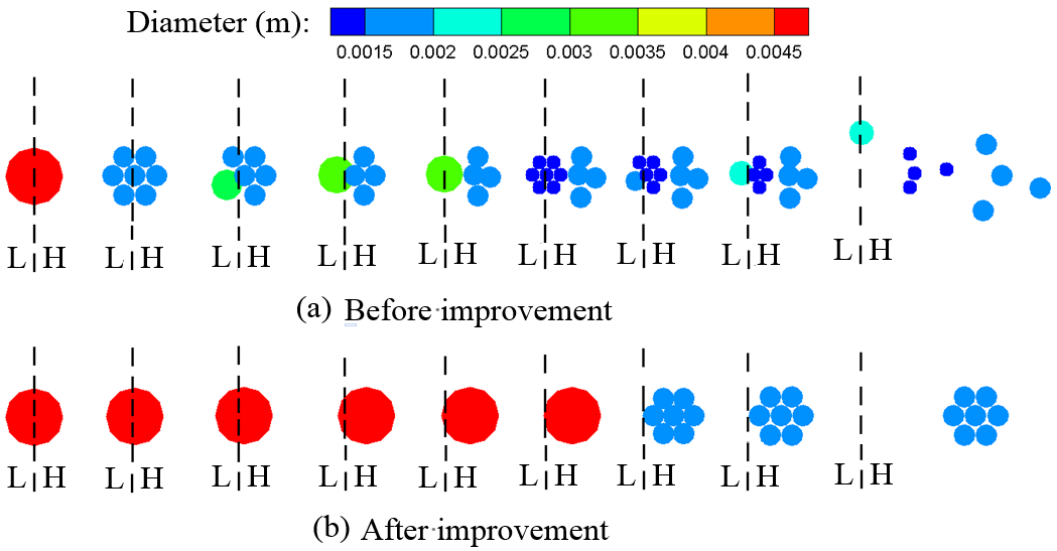


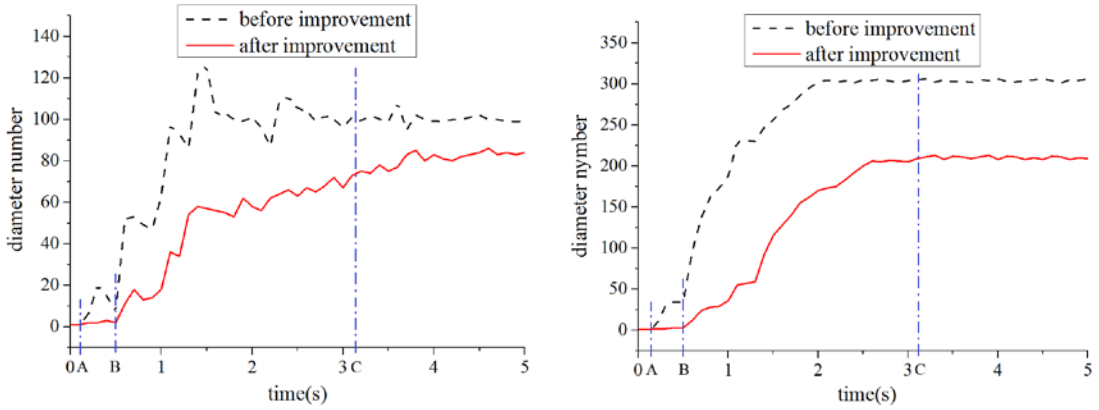

**Figure 10. Comparison of particles dynamic behavior before and after improvement**

where the black solid line represents the coarse/fine boundary, the left region of the boundary is low resolution, and the right region of the boundary is high resolution.

Before the improvement, the particle is divided into seven fine particles when the center of mass of the particle passes through the coarse/fine boundary. However, at this time, three fine particles stay in the low-resolution region. After two coalescing processes, one coarse particle and four fine particles are formed. Then the coarse particles will continue to split and form seven finer particles when it passes through the coarse/fine boundary. The finer particles, reside in the low resolution region, are coalesced twice, and finally nine particles with three different diameters are obtained.

After the improvement, the particle splits into 7 finer particles when the left boundary of the particle enters the high resolution region, all finer particles have a velocity of $u$ = 0.1m/s due to momentum conservation. The particles keep moving rightward at a constant speed. On the contrary, before the improvement, the particle interaction with different velocities makes the particle motion more complex, and the velocity of particles changes after they passing through the coarse/fine boundary, it has an adverse effect on the algorithm accuracy.

After the improvement, there is no finer particle staying in low resolution region, the dynamic behavior of the particle is simpler, the number of splitting and coalescing processes decreases, and the robustness of the coarse/fine interface is improved. We further simulate the dam-break case as shown in Figure 5 with the unimproved algorithm and the improved algorithm respectively. The physical parameters of fluid particle used in this paper are $\rho$ = 988kg/m$^3$ and $v = 1.0 \times 10^6$m$^2$/s, the initial diameter of the fluid particles is $l_0$ = 0.005m. Figure 11 shows the comparison of particle diameter numbers before and after the improvement near coarse/fine interface and in the whole calculation area.




(a) Diameter number near coarse/fine interface (b) Diameter number in the whole area

**Figure 11. Comparison of particle diameter number before and after improvement**

Before the improvement, the diameter number of the particles start to rise when particles pass through the coarse/fine interface (time = A), so when the fluid hits the right wall and falls back (time = B), the diameter number of particles increases sharply. As the fluid continues to flow across the coarse/fine interface, the difference in diameter number is accumulated. When the simulation runs to C (time = C), the fluid flow tends to stabilize gradually, and the diameter number of particles has basically reached saturation. Up to the five seconds of the simulation, after the improvement, the number of particle diameter is reduced by 15% near coarse/fine interface, and in the whole calculation area, the number of particle diameter decreased by 32%. Although the particle interaction model we used in 2.2 can reduce the influence of the particle diameter difference on the calculation accuracy, after improvement, the robustness at coarse/fine interface is improved.

In order to verify the accuracy of the algorithm has been improved, we compared the experimental with simulation results of the position of water front in the dam-break problem as shown in Figure 12, after improvement, the simulation results are more consistent with the experiment compared with the previous algorithm. The algorithm accuracy of the scheme is improved by changing the decision method of particle entering the different resolution.

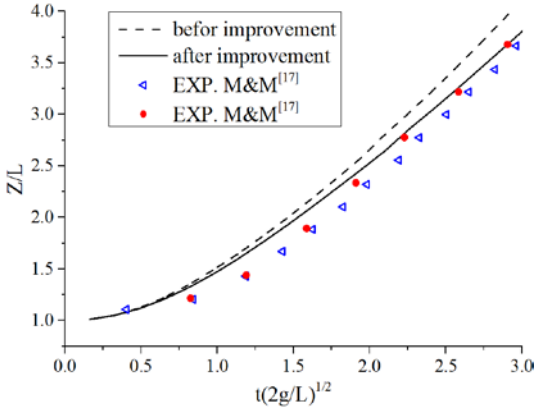



**Figure 12. Experimental and simulation results of the position of water front in the dam-break problem**

*4.2 Moving High Resolution Region*

In previous multi-resolution MPS methods, the high resolution region is fixed. In order to make the high resolution region move around the solid adaptively, we studied the influence of high resolution region movement on the flow field. We simulated the dam-break case as shown in Figure 5 with static high resolution region and moving high resolution region respectively. Figure 13 shows experimental and simulation results of the position of water

front in the dam-break problem, the simulation results obtained by the two methods almost coincide with each other, which proves that with moving high resolution region, the dynamic behavior of the fluid is not affected.

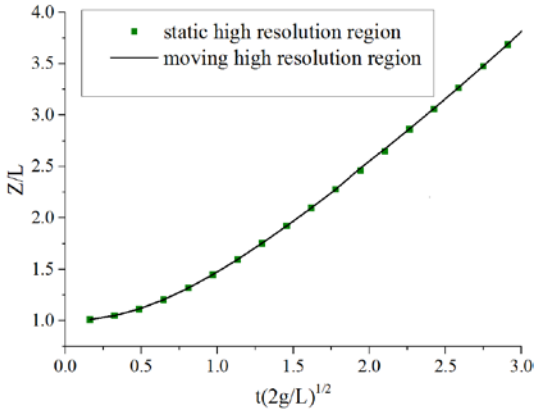
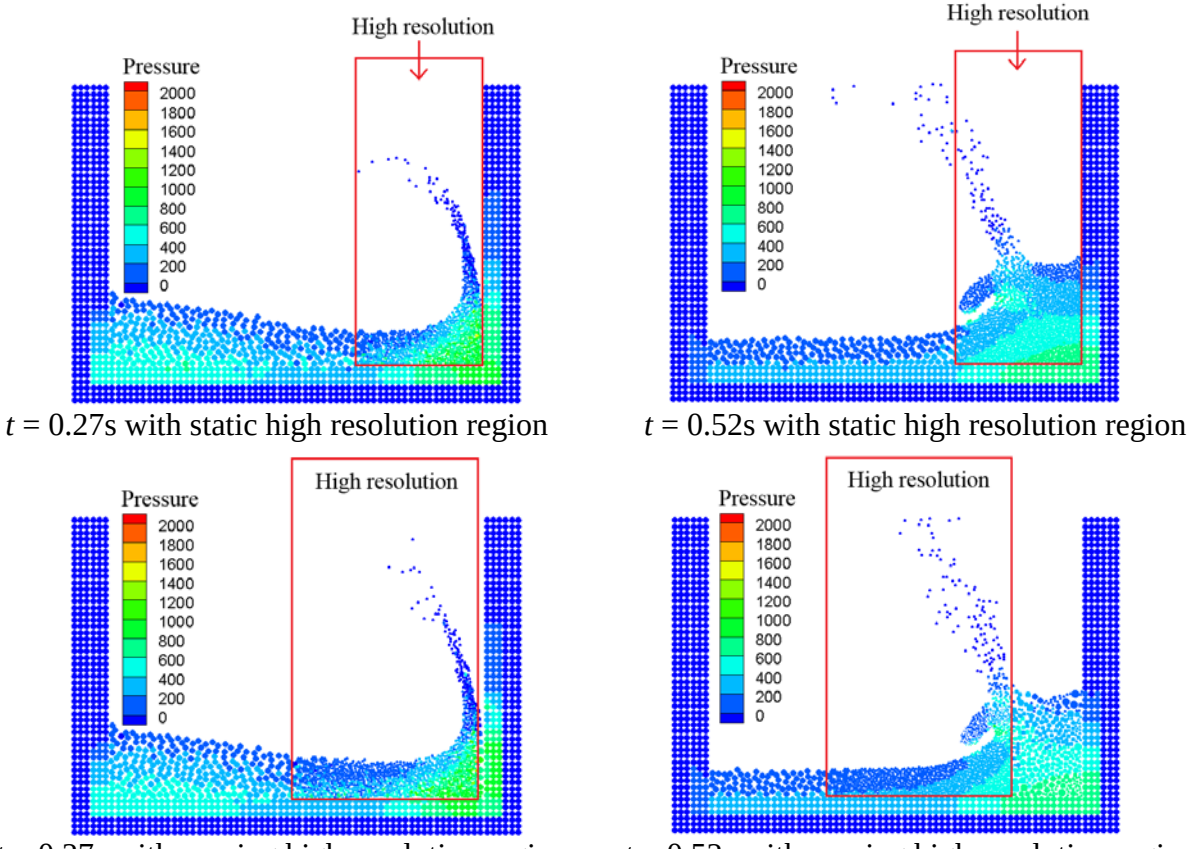



**Figure 13. Experimental and simulation results of the position of water front in the dam-break problem**




*t* = 0.27s with static high resolution region *t* = 0.52s with static high resolution region

*t* = 0.27s with moving high resolution region *t* = 0.52s with moving high resolution region

**Figure 14. Snapshots from the dam-break simulation**

Figure 14 shows the pressure distribution of fluid at two typical times with different methods, with moving high resolution region, the method can also accurately simulate the pressure distribution of fluid in the flow process.

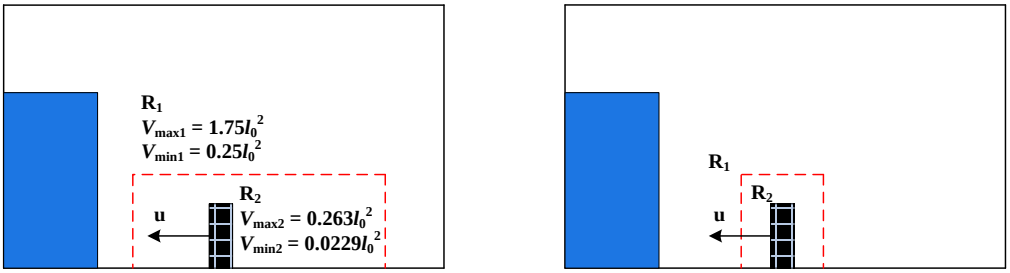
In fact, as described earlier, the algorithm has realized that the splitting and coalescing processes can be conducted in the whole area, and the system can always keep the conservation of mass and momentum. The only difference between static resolution and moving resolution is the location of finer particles, so movable scheme can be very helpful if the location of the flow details we want to describe may change in the simulation. All the

above proves that the region movement has no adverse effect on the flow field, the scheme has been extended and the multi-resolution regions is fixed any more.

*4.3 Adaptive Variable Size Particle MPS Model*

In previous algorithm, the high resolution regions are static in space and known beforehand; all the particles that enter the high resolution regions are split into smaller particles. Since the high resolution region is static, it is not adaptive, in this paper, the scheme is improved with adaptive variable-size particle as AVSP-MPS, which means the shape and area of the domains with high resolution can be dynamically adjusted during the calculation according to the distance from a fluid particle to a movable object.

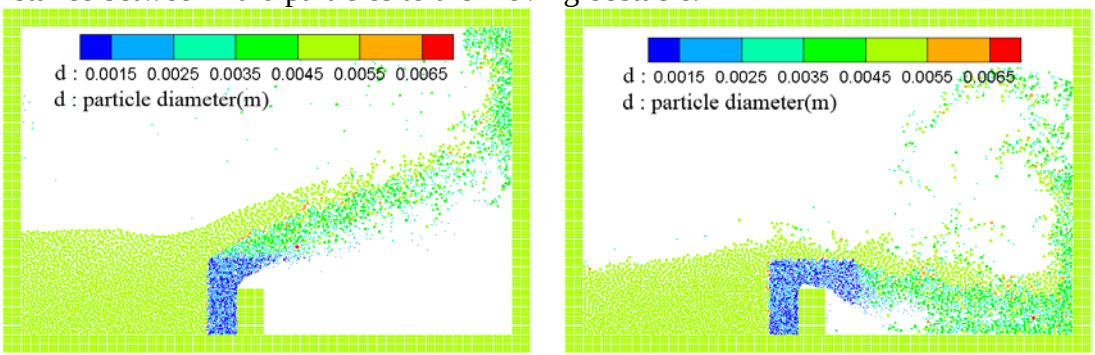
In this section, we use three methods, single resolution MPS, VSP-MPS and AVSP-MPS, to simulate the dam-break case with a moving obstacle, respectively. The schematic diagram is shown in Fig. 15. The density of fluid particles is $\rho$ = 988kg/m$^3$, the viscosity is $\nu$ = 1.0 × 10$^6$m$^2$/s, the gravity is G = 9.8m/s$^2$, and the initial diameter of water particles is $l_0$ = 0.005m. The object moves in horizontal direction and the equation of motion can be expressed as **u** = -0.5×sin(2$\pi t$). The high resolution area in VSP-MPS simulation is static, so it is necessary to divide a large area into high resolution region (as shown in $R_2$ (red box) in Figure 15(a)). As a contrast, the high resolution region does not need to be given beforehand with AVSP-MPS method, the high resolution region can dynamically change in the simulation to ensure fine particles are always used around the moving obstacle (as shown in $R_2$ (red box) in Figure 15(b)).



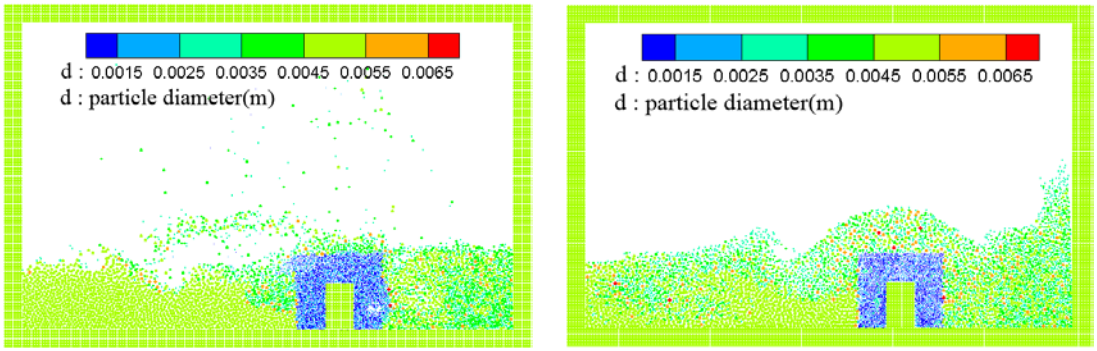

(a) VSP-MPS (b) AVSP-MPS

**Figure 15. Geometry of the dam-break with obstacle**

Fig. 16 shows the diameter distribution of fluid particles at several typical times using AVSP-MPS method, in the simulation, fine particles are only used around the moving obstacle, coarse particles are used in other area. It proves that the adaptive model has been successfully added to AVSP-MPS method, the high resolution area can dynamically change according to the distance between fluid particles to the moving obstacle.



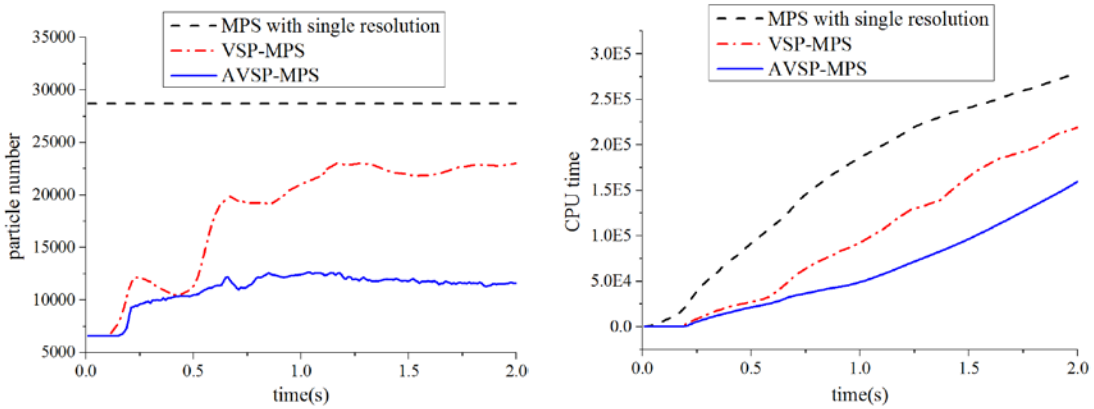

$t$ = 0.4s $t$ = 06s




$t$ = 1.0s $t$ = 2.0s

**Figure 16. Diameter distribution of the dam-break with obstacle**

Fig. 17 shows the time-varying curves of particle number and CPU obtained by three methods. Compared with the single resolution MPS method, AVSP-MPS can greatly reduce the particle number and CPU time in the simulation process by 59.5% and 43.0%. Compared with the VSP-MPS method which can only delimit the resolution region in advance, AVSP-MPS method can reduce particle number by 49.5%, CPU time by 27.1%. With an adaptive algorithm, AVSP-MPS method can reduce the number of particles needed in the simulation and improve the calculation efficiency.




(a) Particle number (b) CPU time

**Figure 17. Comparison of particle number and CPU time in different cases**

## 5 Conclusions

In this paper, the scheme of MPS method is improved with adaptive variable-size particle as AVSP-MPS to increase computational efficiency. First, we made linear programming for the selection of resolution and got the optimal resolution interval, compared with MPS with single resolution, VSP-MPS method can reduce the number of particles needed for simulation and greatly shorten the calculation time. Then the algorithm is improved by changing the decision method of particle entering the different resolution to improve the robustness at coarse/fine interface. The improved approach is that only when the whole particle (the boundary of the particle) enters the high resolution region the particle splitting into finer particles. The algorithm accuracy of the scheme is improved by changing the decision method of particle entering the different resolution. As the high resolution of the previous multi-resolution MPS method is static, we get a moving high resolution region and studied the influence of high resolution region movement on the flow field, the region movement will not affect the flow field. At last, the scheme with adaptive variable-size particle as AVSP-MPS is introduced, the shape and area of the domains with high resolution can be dynamically adjusted during the calculation according to the distance from a fluid particle to a movable object. Compared with the VSP-MPS method which can only delimit the resolution region in

advance, with an adaptive algorithm, AVSP-MPS method can further reduce the number of particles needed in the simulation and improve the calculation efficiency.

**References**


[1] Koshizuka S, Oka Y. (1996) Moving-particle semi-implicit method for fragmentation of incompressible fluid. *Nuclear Science and Engineering* **123**:421–434.

[2] Liu, G. R. Liu, M. B. (2003) Smoothed particle hydrodynamics, a meshfree particle method, *World Scientific*, Singapore, 50–59.

[3] Omidvar, P., Stansby, P. K., Rogers, B. D. (2012) Wave body interaction in 2D using smoothed particle hydrodynamics (SPH) with variable particle mass, *International Journal for Numerical Methods in Fluids* **68**:686-705.

[4] Feldman, J., Bonet, J. (2007) Dynamic refinement and boundary contact forces in SPH with applications in fluid flow problems, *International Journal for Numerical Methods in Engineering* **72**:295-324.

[5] Vacondio, R., Rogers, B. D., Stansby, P. K. (2013) Variable resolution for SPH: A dynamic particle coalescing and splitting scheme, *Computer Methods in Applied Mechanics and Engineering* **256**:132-148.

[6] Chiron, L., Oger, G., Leffe, M. D., Touzé, D. L. (2018) Analysis and improvements of Adaptive Particle Refinement (APR) through CPU time, accuracy and robustness considerations, *Journal of Computational Physics* **354**:552-575.

[7] Sun, P. N., Colagrossi, A., Zhang A. M. (2018) Numerical simulation of the self-propulsive motion of a fishlike swimming foil using the $\delta^+$-SPH model, *Theoretical & Applied Mechanics Letters* **8**, 115-125.

[8] Sun, P. N., Colagrossi, A., Marrone, S., Zhang A. M. (2017) The δplus-SPH model: Simple procedures for a further improvement of the SPH scheme, *Computer Methods in Applied Mechanics and Engineering* **315**. 25–49.

[9] Shibata, K., Koshizuka, S., Tamai, T., Murozono, K. (2012) Overlapping particle technique and application to green water on deck. *International Conference on Violent Flows* pp:160-111.

[10] Shibata, K., Koshizuka, S., Matsunaga, T., Masaie, I. (2017) The overlapping particle technique for multi-resolution simulation of particle methods. *Computer Methods in Applied Mechanics and Engineering* **325**, 434-462.

[11] Tanaka, M., Masunaga, T., Nakagawa, Y. (2009) Multi-resolution MPS method. *Trans. Jpn. Soc. Comput. Eng. Sci.* (in Japanese).

[12] Tang, Z.Y., Wan, D.C., Chen, G., Xiao, Q. (2016) Numerical simulation of 3D violent free-surface flows by multi-resolution MPS method. *J. Ocean Eng. Mar. Energy* **2**, 355-364.

[13] Chen, X., Sun, Z. G., Liu, L. Xi, G. (2016) Improved MPS method with variable-size particles. *International Journal for Numerical Methods in Fluids* **80**, 358-374.

[14] Chen, X., Xi, G., Sun, Z. G. (2014) Improving stability of MPS method by a computational scheme based on conceptual particles. *Computer Methods in Applied Mechanics and Engineering* **278**, 254-271.

[15] Tanaka, M., Cardoso, R. Bahai, H. (2018) Multi-resolution MPS method. *Journal of Computational Physics* **359**, 106-136.

[16] Tanaka, M., Masunaga, T. (2010) Stabilization and smoothing of pressure in MPS method by quasi-compressibility. *Journal of Computational Physics* **229**, 4279-4290.

[17] Martin J. C., Moyce W. J. (1952) An experimental study of the collapse of fluid columns on a rigid horizontal plane. *Philosophical Transactions of the Royal Society of London Series A* **244**, 312-324.

# *J*-integral Decomposition Approach for 3-D Elasto-Plastic Fatigue Crack Growth Simulations

***M. Kumar, I.V. Singh, and B.K. Mishra**

Department of Mechanical and Industrial Engineering,
Indian Institute of Technology Roorkee, Roorkee 247667 India

*Presenting author: mkumar2@me.iitr.ac.in

## Abstract

In the present work, elasto-plastic fatigue crack growth (FCG) in three-dimensional (3D) domains is numerically performed using the extended finite element method (XFEM). The XFEM does not require conformal mesh and remeshing for crack growth modeling as required in the standard finite element method. The crack front in 3D is modeled by joining the small size line segments. The fatigue crack growth rate (FCGR) is computed by the stress intensity factor (SIF) dependent Paris Law. The main challenge in the elasto-plastic FCG modeling is the evaluation of three modes of SIFs, which is properly handled by the *J*-decomposition approach. The field variables are decomposed into their symmetric and anti-symmetric portions across the crack surface in the *J*-decomposition approach. These decomposed portions of fields are used to compute the symmetric and anti-symmetric *J*-integrals. The numerical issues such as the derivative of stress and strain energy density; evaluation of stress in the virtual domain during the *J*-integral calculation are properly addressed. The numerically predicted FCG behavior of Ni-based superalloy is validated experimentally at elevated temperature.

**Keywords:** Stress intensity factor; Fatigue crack growth (FCG); *J*-decomposition; XFEM.

## Introduction

Advancement in the industries and technologies demands the highly efficient and reliable design of the structures/components. To fulfill this objective, all the complex loading effect, environmental factors, flaws in materials like heterogeneity, micro-defects, cracks are necessary to involve in the designing phase. In general, finite element method (FEM) is employed to assist the designing process of the structures/components but FEM is not suitable for designing when material flaws like cracks are considered in the structures/components. In FEM, there is the need of conformal mesh about the crack surface to produce the jump effect in displacement and a very fine mesh is required to capture the stress singularity at the crack front. The modeling to crack propagation requires the remeshing with every crack growth and transfer of data from old mesh to the new mesh. The remeshing procedure is a time-consuming process and data transfer introduces the inaccuracies in the solution. All these complications inspired the researchers to develop new methods to overcome these issues.

In the past two decades, many numerical methodologies are developed by the researchers to overcome the problem of conformal mesh and remeshing for crack modeling such as boundary element method [1], meshfree methods [2]-[3], extended finite element method [4]-[5], extended isogeometric analysis [6]-[7], coupled meshfree and finite element method [8], gradient damage

models [9], phase field method [10] and many more. Among these methods, XFEM is one of the most robust and successfully implemented method to model the stationary cracks, elasto-plastic fatigue crack growth [11], creep crack growth [12], crack growth in heterogeneous materials [13], dynamic crack growth [14], etc. In XFEM, two types of enrichment functions are added to the standard FEM displacement approximation via the partition of unity. The jump enrichment function is utilized to model the jump in displacement field about the crack surface whereas the crack front singularity is captured by the front enrichment functions.

In this paper, the methodology to model the elasto-plastic FCG in the 3D domain is presented. The SIF depended Paris Law is used to calculate the FCGR at the ends of the line segments of the crack front. The individual modes of SIFs are evaluated using the *J*-decomposition approach [15]. All the fields i.e. strain, stress and displacement derivatives are decomposed into symmetric and anti-symmetric portions across the crack surface in this approach. A virtual cylindrical domain is created at the ends of the line segments of crack front to calculate the *J*-integral. The interpolation functions are used to calculate all the required fields at the virtual domain from the nodal data. However, the stress field cannot be obtained by directly interpolation due to plasticity thus; a data transfer scheme is employed to calculate the stress field at the virtual domain. Due to the presence of plasticity, direct derivatives of stress and strain energy density are not possible hence function approximation is utilized to compute the derivative of stress and strain energy density. The numerically computed FCG for Ni-based superalloy is compared with the experimental results and found in a good match.

## Mathematical Formulation

In this section, XFEM based methodology to simulate the elasto-plastic FCG is explained. The FCGR is computed by SIF range based Paris Law. The individual modes of SIF are calculated by the *J*-decomposition approach, which depends on the decomposed fields. During the evaluation of *J*-integral, several numerical issues are faced that are discussed in detail in this section. The maximum principal stress criterion is used to obtain the crack growth direction.

### *Governing Equations*

A residual stress-free domain of isotropic homogeneous material is considered for the formulation. The domain is assumed of volume $\Omega$ and bounded by the surface $\Gamma$ as shown in Fig. 1. Prescribed traction and displacement is applied to the surface $\Gamma_t$ and $\Gamma_u$ of the domain respectively. A sharp traction free crack in the domain is also considered and denoted by the $\Gamma_c$ in Fig. 1. The equilibrium equation and the associated boundary conditions for the domain are defined as

$$\sigma_{ij,j} = 0 \qquad \text{in} \quad \Omega \qquad \forall\ i, j \in \{1,2,3\} \tag{1}$$

$$\sigma_{ij}\, n_j = 0 \qquad \text{on} \quad \Gamma_c \tag{2}$$

$$\sigma_{ij}\, n_j = \bar{t}_i \qquad \text{on} \quad \Gamma_t \tag{3}$$

$$u_i = \bar{u}_i \qquad \text{on} \quad \Gamma_u \tag{4}$$

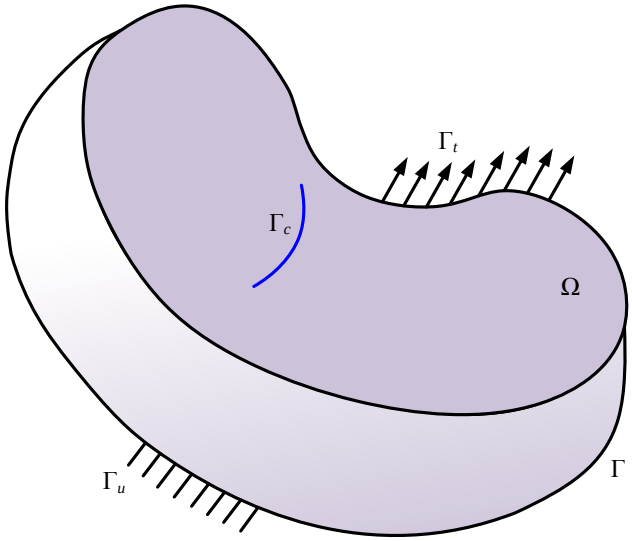



**Figure 1. An illustration on 3D cracked domain along with boundary conditions**

where $\sigma_{ij}$ is the Cauchy stress, $u_i$ is the displacement, $n_j$ is the unit normal vector, $\bar{u}_i$ and $\bar{t}_i$ are the applied displacement and traction on the surface $\Gamma_u$ and $\Gamma_t$ respectively. The strong form of Eq. (1) is converted into weak form by employing the principle of virtual work as

$$\int_{\Omega} \sigma_{ij} \left(\delta u_i\right)_{,j} d\Omega - \int_{\Gamma_t} \bar{t}_i \delta u_i d\Gamma = 0 \tag{5}$$

This weak form of equilibrium equation is written into discrete equations using discretization of the domain as follows

$$\mathop{\mathrm{A}}_{e=1}^{n_e} \int_{\Omega_e} \mathbf{B}^T \mathbf{C} \mathbf{B} u_i d\Omega - \int_{\Gamma_t} \mathbf{N}^T \bar{t}_i d\Gamma = 0 \tag{6}$$

where $\mathbf{B}$ is the gradient matrix of shape functions, $\mathbf{C}$ is the elasto-plastic constitutive matrix, $n_e$ is the number of elements and $\mathbf{N}$ is the shape function vector. The simultaneous solution of Eq. (6) gives the displacement field that is further used to compute the strain and trial stress field using displacement derivatives and constitutive relation respectively. The trial stress is checked for yielding at each integration point via *J*2 plasticity yielding criterion. For the yielded integration point, generalized Ramberg-Osgood material model along with associated flow rule is used to calculate the plastic strain and stress field [16] whereas for non-yielded integration point trial stress is taken as final stress field. The equilibrium of the system is ensured by the global convergence parameter that is defined as follows

$$\Upsilon = \frac{\left\| \mathbf{F}_{ext} - \mathbf{F}_{int} \right\|}{\left\| \mathbf{F}_{ext} \right\|} \quad \textit{where} \quad \mathbf{F}_{ext} = \int_{\Gamma_t} \mathbf{N}^T \bar{t}_i d\Gamma \; ; \; \mathbf{F}_{int} = \mathop{\mathrm{A}}_{e=1}^{n_e} \int_{\Omega_e} \mathbf{B}^T \sigma_{ij} d\Omega \tag{7}$$

If the convergence parameter is less than the tolerance that means the equilibrium has been attained and the next load step is initiated, otherwise solution of the discrete equations is performed again with the updated residual force and updated elasto-plastic constitutive matrix $(\mathbf{D})$ as

$$\mathop{\mathrm{A}}_{e=1}^{n_e}\int_{\Omega_e}\mathbf{B}^T\mathbf{D}\mathbf{B}u_i d\Omega=\int_{\Gamma_t}\mathbf{N}^T\overline{t}_i d\Gamma-\mathop{\mathrm{A}}_{e=1}^{n_e}\int_{\Omega_e}\mathbf{B}^T\sigma_{ij}d\Omega \tag{8}$$

This process is continued until the convergence is achieved. If the solution diverges then load step is taken as half and the solution of discrete equations is computed from the previously converged load step.

*Extended Finite Element Method*

XFEM has been successfully used to model the propagating cracks without the need of conformal mesh and remeshing of the domain during crack propagation. In this method, two types of enrichment are added to the standard FEM displacement approximation via the partition of unity. The jump enrichment function is used to model the crack surface whereas the crack front singularity is mimicked by the front enrichment functions. Due to the introduction of these enrichment functions, the total number of degree of freedom (DOF) of the system increased slightly. The enriched displacement approximation for a domain [17] can be written as

$$\mathbf{u}^h(\mathbf{x})=\sum_{i\in n}N_i(\mathbf{x})\mathbf{u}_i+\sum_{j\in n_c}N_j\left[H(\mathbf{x})-H(\mathbf{x}_j)\right]\alpha_j+\sum_{k\in n_t}N_k\sum_{l=1}^{4}\left[\varsigma_l(\mathbf{x})-\varsigma_l(\mathbf{x}_k)\right]\beta_k^l \tag{9}$$

where $H(\mathbf{x})$ and $\varsigma_l(\mathbf{x})$ are the jump and front enrichment functions respectively; $n$, $n_c$ and $n_t$ presents all the nodes in the domain, nodes associated with completely cut elements and nodes associated with partially cut elements respectively; $\alpha$ and $\beta$ are the DOFs related with jump enrichment function and front enrichment function respectively. The jump enrichment function and front enrichment function [18] are given below

$$H(\mathbf{x})=\begin{cases}-1 & \text{for } \Psi(\mathbf{x})<0\\ 1 & \text{for } \Psi(\mathbf{x})<0\end{cases} \tag{10}$$

$$\varsigma(\mathbf{x})=\left[r^\eta cos\frac{\theta}{2},r^\eta sin\frac{\theta}{2},r^\eta cos\frac{\theta}{2}sin\theta,r^\eta sin\frac{\theta}{2}sin\theta\right]\qquad\forall\ \eta=\frac{1}{\overline{n}+1} \tag{11}$$

where $\Psi(\mathbf{x})$ is the normal distance from the crack surface; $\overline{n}$ is the hardening constant of material; $r$ and $\theta$ are the polar coordinates with respect to crack front. To trace the crack surface in the domain level set is used.

*Fatigue Crack Growth*

The Paris law is used to estimate the FCGR in the domain as

$$\frac{da}{dN} = C\left(\Delta K_{I\,eq}\right)^{m} \tag{12}$$

where $da$ is the crack growth, $dN$ is the number of cycles required for the crack growth $da$, $\Delta K_{I\,eq}$ is the equivalent SIF range, $C$ and $m$ are the Paris law constants. The equivalent stress intensity factor range for the constant amplitude fatigue loading is defined as

$$\Delta K_{I\,eq} = K_{I\,eq}^{max} - K_{I\,eq}^{min} \tag{13}$$

where $K_{I\,eq}^{max}$ and $K_{I\,eq}^{min}$ are the equivalent stress intensity factor corresponding to $\sigma^{max}$ and $\sigma^{min}$ of applied fatigue load. The individual modes of SIFs (mode-I and mode-II) are used to compute the equivalent stress intensity factor as

$$K_{I\,eq} = K_I\, cos^3\left(\frac{\theta_c}{2}\right) - 3K_{II}\, cos^2\left(\frac{\theta_c}{2}\right) sin\left(\frac{\theta_c}{2}\right) \tag{14}$$

where $K_I$ and $K_{II}$ are the SIFs of mode-I and mode-II respectively that are evaluated by *J*-decomposition approach, $\theta_c$ is the critical angle for crack growth. The maximum principal stress criterion [19] is used to compute the critical angle for crack growth as given below

$$\theta_c = 2\,tan^{-1}\left(\frac{1}{4}\frac{K_I}{K_{II}} - \frac{1}{4}\sqrt{\left(\frac{K_I}{K_{II}}\right)^2 + 8}\right) \tag{15}$$

*J-decomposition Approach*

The individual modes of SIFs are evaluated from *J*-integral, which is calculated at the ends of the crack front line segments using the *J*-decomposition approach [20]. A virtual cylindrical domain as shown in Fig. 2 is created at these ends of crack front to perform the *J*-integral computation. The decomposed form of *J*-integral at these ends is defined as

$$\begin{aligned} J^N = \frac{1}{2L_e}\Bigg(&\int_\Lambda \left(\sigma_{ij}^N \left(u_{i,1}\right)^N - W^N \delta_{1j}\right) q_{,j}\, d\Lambda + \int_\Lambda \left(\sigma_{ij}^N \left(u_{i,1j}\right)^N - W_{,1}^N\right) q\, d\Lambda \\ &+ \int_\Lambda \left(\sigma_{i1,1}^N + \sigma_{i2,2}^N\right)\left(u_{i,1}\right)^N d\Lambda - \int_\Lambda \sigma_{i3}^N \left(u_{i,13}\right)^N d\Lambda\Bigg) \qquad \forall\, N \in \{I, II, III\} \end{aligned} \tag{16}$$

where $I, II, III$ represent the mode-I, mode-II and mode-III respectively, $W$ is the strain energy density, $L_e$ is the length of the virtual domain along the crack front, $q$ is function having value one at the crack front and zeros at the boundary of the virtual domain. The symmetric portion of fields provides the mode-I SIF whereas anti-symmetric portion is further divided to obtain the mode-II and mode-III SIFs.

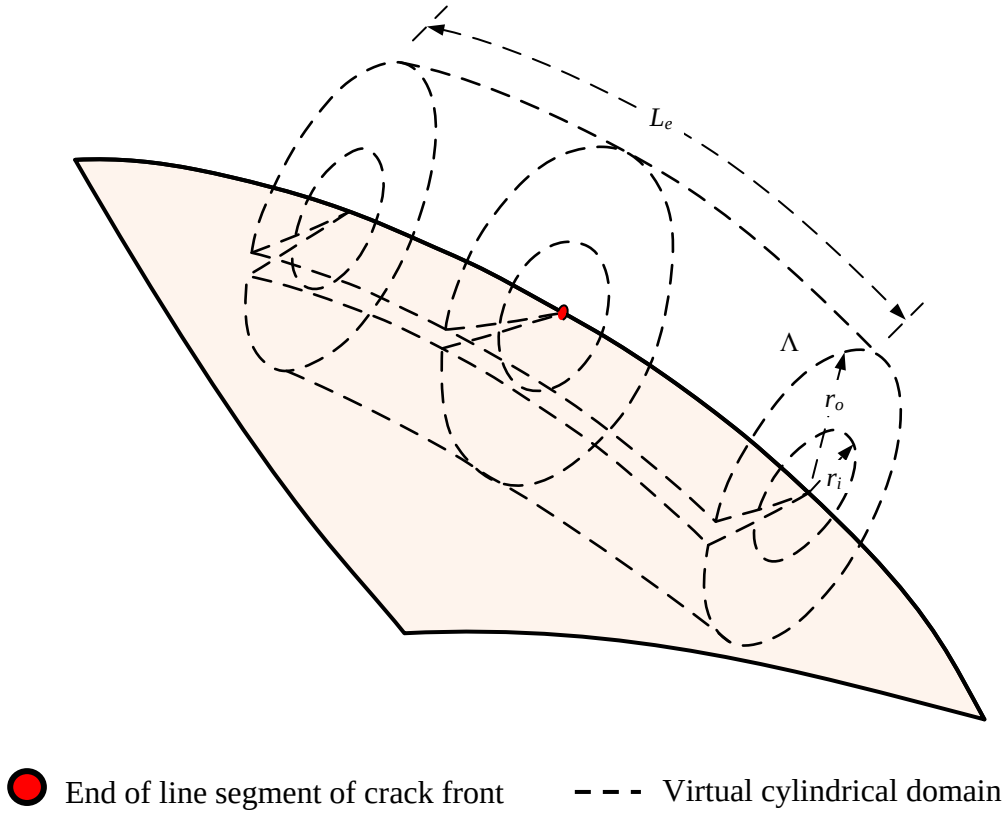



**Figure 2. A virtual domain at the end of crack front line segment for the calculation of *J*-integral**

In order to compute the *J*-integral by this approach, the required fields are interpolated from the original mesh to the virtual cylindrical domain via shape function interpolation and decomposed into symmetric and anti-symmetric portions across the crack surface. For the decomposition of fields, all the fields are required at the mirror point of the integration point that can be calculated from the nodal data by interpolation functions. Due to the presence of plasticity, stress field cannot be directly obtained at the virtual domain and mirrored point from the displacement field. Therefore, to calculate the stress field at the required point, a data transfer scheme is utilized. In this scheme, the stress field is transferred from the integration points to the nodes [21] by

$$\chi_n = \left(\bar{\mathbf{N}}^T \bar{\mathbf{N}}\right)^{-1} \bar{\mathbf{N}}^T \chi_{ip} \tag{17}$$

where $\bar{\mathbf{N}}^T$ is the matrix contains the value of shape functions at the integration points, $\chi_n$ and $\chi_{ip}$ are the fields at node and integration point respectively. After that, the nodal stress is interpolated at the required point using shape functions of the element. Due to the use of sub-tetrahedralization for enriched elements, the stress field is extrapolated at the nodes of sub-tetrahedron using Eq. (17) and stored for each tetrahedron separately. The interpolation of the stress field in the enriched element is performed in two steps. In the first step, sub-tetrahedron is identified that contains the mirrored point while in the second step; the stress field is interpolated using the extrapolated nodal

stress and shape functions of the identified sub-tetrahedron. The decomposed stress field for all the modes at the spatial point *P* across the crack surface 1-3 (as given in Fig. 3) is expressed as

$$\sigma_{ij\,P}^{I} = \begin{Bmatrix} \sigma_{11\,P}^{S} \\ \sigma_{22\,P}^{S} \\ \sigma_{33\,P}^{S} \\ \sigma_{12\,P}^{S} \\ \sigma_{23\,P}^{S} \\ \sigma_{13\,P}^{S} \end{Bmatrix} = \frac{1}{2} \begin{Bmatrix} \sigma_{11\,P} + \sigma_{11\,P'} \\ \sigma_{22\,P} + \sigma_{22\,P'} \\ \sigma_{33\,P} + \sigma_{33\,P'} \\ \sigma_{12\,P} - \sigma_{12\,P'} \\ \sigma_{23\,P} - \sigma_{23\,P'} \\ \sigma_{13\,P} + \sigma_{13\,P'} \end{Bmatrix} \tag{18}$$

$$\sigma_{ij\,P}^{II} + \sigma_{ij\,P}^{III} = \begin{Bmatrix} \sigma_{11\,P}^{AS} \\ \sigma_{22\,P}^{AS} \\ \sigma_{33\,P}^{AS} \\ \sigma_{12\,P}^{AS} \\ 0 \\ 0 \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sigma_{23\,P}^{AS} \\ \sigma_{13\,P}^{AS} \end{Bmatrix} = \frac{1}{2} \begin{Bmatrix} \sigma_{11\,P} - \sigma_{11\,P'} \\ \sigma_{22\,P} - \sigma_{22\,P'} \\ \sigma_{33\,P} - \sigma_{33\,P'} \\ \sigma_{12\,P} + \sigma_{12\,P'} \\ 0 \\ 0 \end{Bmatrix} + \frac{1}{2} \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sigma_{23\,P} + \sigma_{23\,P'} \\ \sigma_{13\,P} - \sigma_{13\,P'} \end{Bmatrix} \tag{19}$$

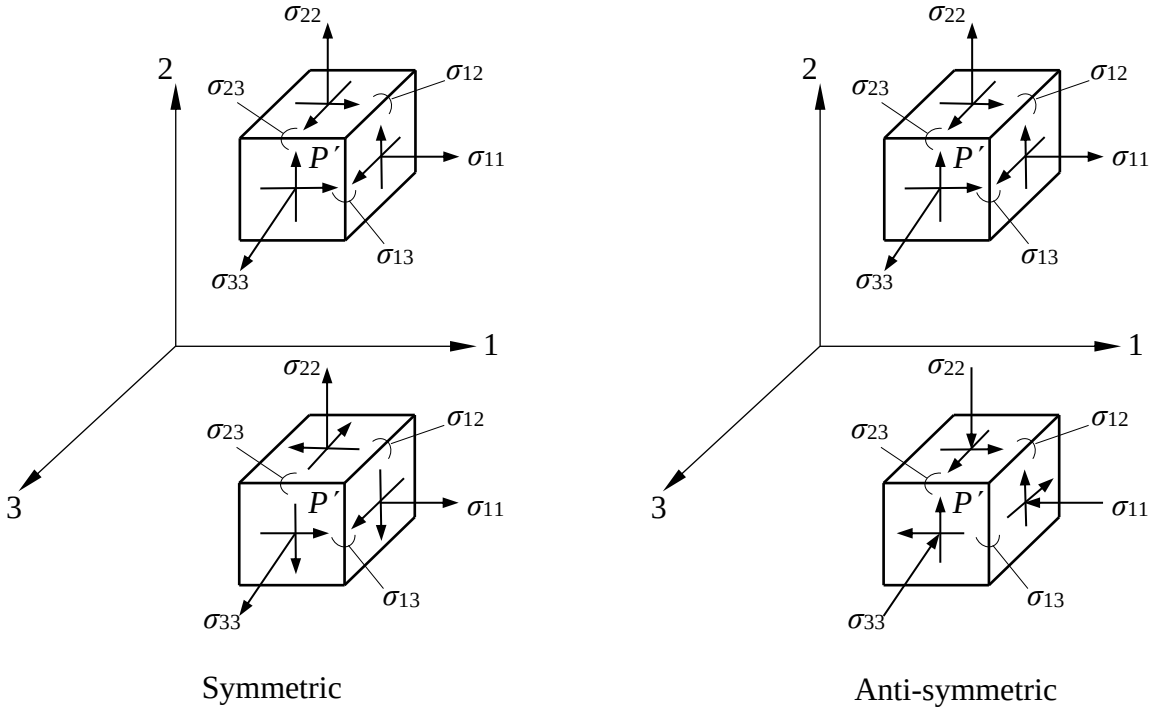



**Figure 3. Symmetric and anti-symmetric portion of stress field across the crack surface (1-3) at point *P***

In a similar way, other fields can also be decomposed [22]. The analytical derivatives of strain energy density and stress are not possible due to the plasticity hence; it is evaluated by the function approximation. The required field of all the integration points of an element is fitted into a quadratic function by nonlinear least squares method as

$$\varpi^N = f_1 + f_2\xi + f_3\eta + f_4\zeta + f_5\xi^2 + f_6\eta^2 + f_7\zeta^2 + f_8\xi\eta + f_9\eta\zeta + f_{10}\xi\zeta \tag{20}$$

where $\xi,\eta,\zeta$ are the local coordinates of the integration points and $f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}$ are the fitting constants. The derivative of Eq. (20) is used to compute the derivative of stress and strain energy density.

**Numerical Implementation**

The flowchart for the implementation of the FCG methodology is presented in Fig. 4. The discretized domain along with boundary conditions is given as input to the computational model. The crack front is divided into small line segments to perform the simulation. The elasto-plastic solution is performed on the computational model using a typical load step as discussed in the previous section. After obtaining the converged solution on complete loading, a virtual cylindrical domain is created at the ends of the line segments of crack front to evaluate the *J*-integral using the *J*-decomposition approach. The data is transferred from the original mesh to the virtual cylindrical domain to compute the *J*-integral. The stress evaluation at the virtual domain and mirror point is performed via the data transfer scheme. The derivatives of stress and strain energy density are computed by function approximation as given in Eq. (20). The decomposed fields are used to calculate the *J*-integral for all the modes, which are further converted into SIFs of individual modes as

$$K_i = \sqrt{\bar{E}\, J^i} \qquad \forall\ i \in \{I, II\} \tag{21}$$

$$K_{III} = \sqrt{2\bar{G}\, J^{III}} \tag{22}$$

where $\bar{E}$ and $\bar{G}$ is the Young's modulus (for plane stress condition) and shear modulus respectively. The equivalent SIF and critical angle are calculated from the SIFs of individual modes. The crack growth rate at the ends of line segments of the crack front is evaluated using equivalent SIF range and the Paris Law. The crack increment is computed for a particular number of cycles and the current crack front is updated with this crack increment. For the next step of the analysis, the updated crack front is considered as the crack front and the whole process of loading and calculation of SIF is repeated. This process is continued until the equivalent SIF is less than the fracture toughness.

**Numerical Results and Discussion**

A compact tensile specimen consists of Ni-based superalloy having 32 mm width and 6 mm thickness is considered for the validation of the presented FCG methodology. A through crack of 7.2 mm is considered in the specimen as shown in Fig. 5.

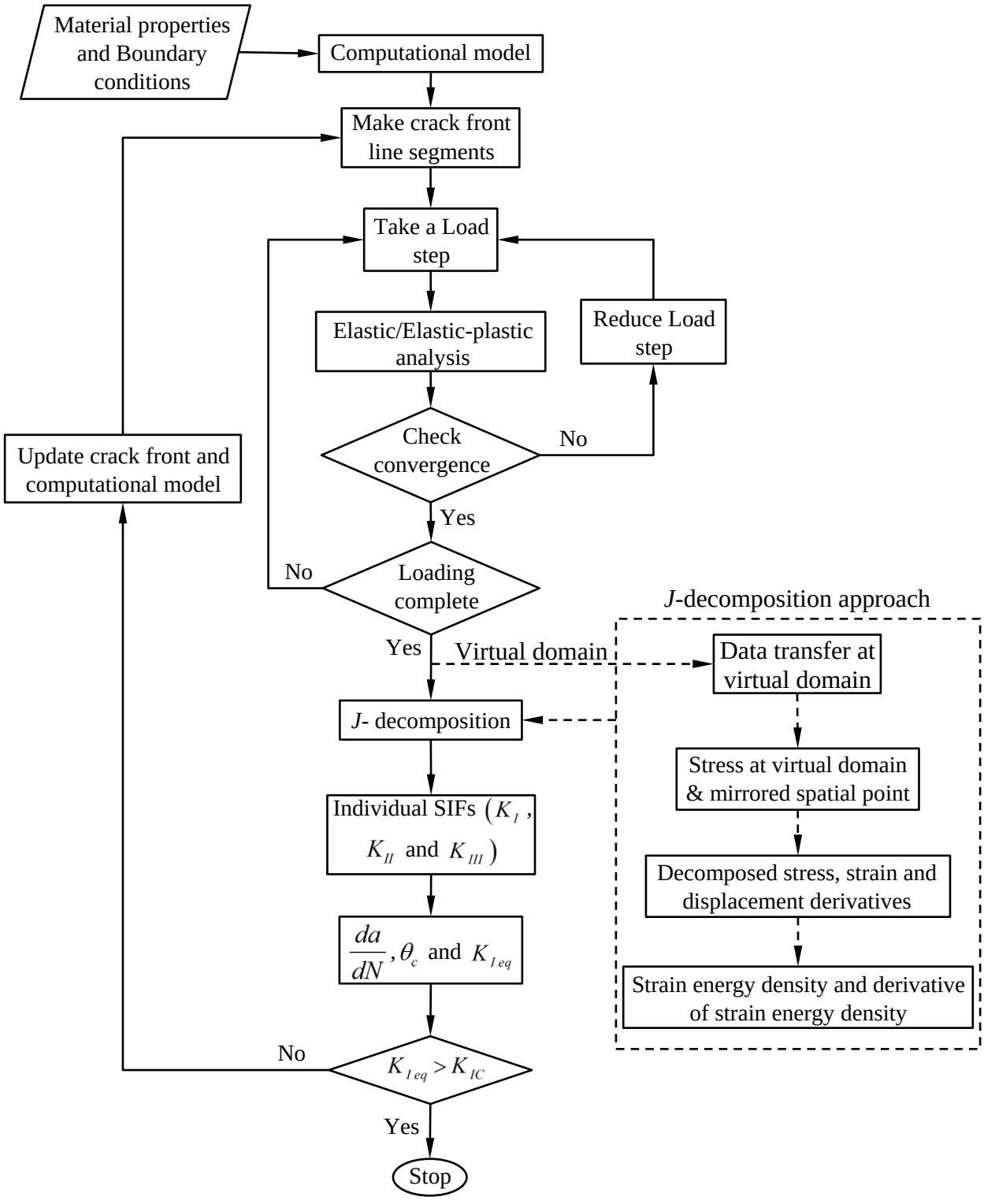



**Figure 4: The flowchart for the implementation of proposed FCG methodology**

The fatigue load of $F_{max} = 3500\,\text{N}\ (R = 0.1)$ is applied at the elevated temperature of 650 °C. The mechanical and fatigue properties of Ni-based superalloy at the elevated temperature are provided in Table 1. The specimen is discretized into 15 × 15 × 3 elements while the crack front is divided into 8 line segments. A virtual cylindrical domain is created at the endpoints of the line segments of crack front except for the corner points of the crack front. The radius and length of the virtual cylindrical domain are taken as 1 mm and 0.5 mm respectively. The decomposed fields are used to calculate the individual modes of SIFs as described in the previous sections.

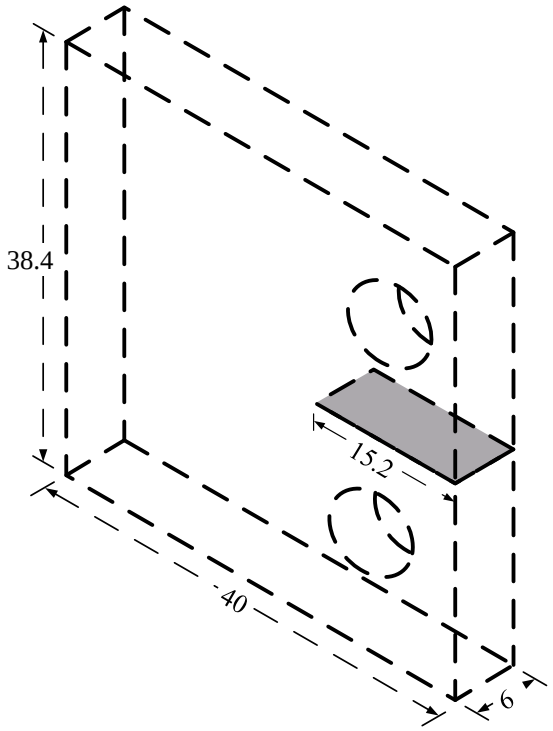



**Figure 5. A schematic of compact tensile specimen considered for simulation**

**Table 1. Material properties of Ni-based superalloy at 650 °C**

| **Mechanical Properties** | **650 ºC Value** |
|---|---|
| Young's modulus, $E$ (GPa) | 180 |
| Poisson ratio, $v$ | 0.33 |
| Yield strength, $\sigma_{yts}$ (MPa) | 653 |
| Ultimate tensile strength, $\sigma_{uts}$ (MPa) | 987 |
| Paris Law constant, $C$ | $1.78\times10^{-8}$ |
| Paris Law constant, $m$ | 2.89 |

The crack growth is evaluated from the crack growth rate at each endpoint of line segments of crack front for a particular number of cycles. Initially, the number of cycles is kept high but when the crack growth is in the range of element size then the number of cycles is reduced to capture the very high rate of crack growth. The numerically computed FCG is shown in Fig. 6 and compared with the experimental results [23]. The numerical results are found in a good agreement with the experimental results. The numerically obtained crack front at different stages of the simulation is also presented in Fig. 7. The predicted growth of the crack front in the middle of the specimen is high as compared to the surface of the specimen, which is consistent with the theoretical expectations.

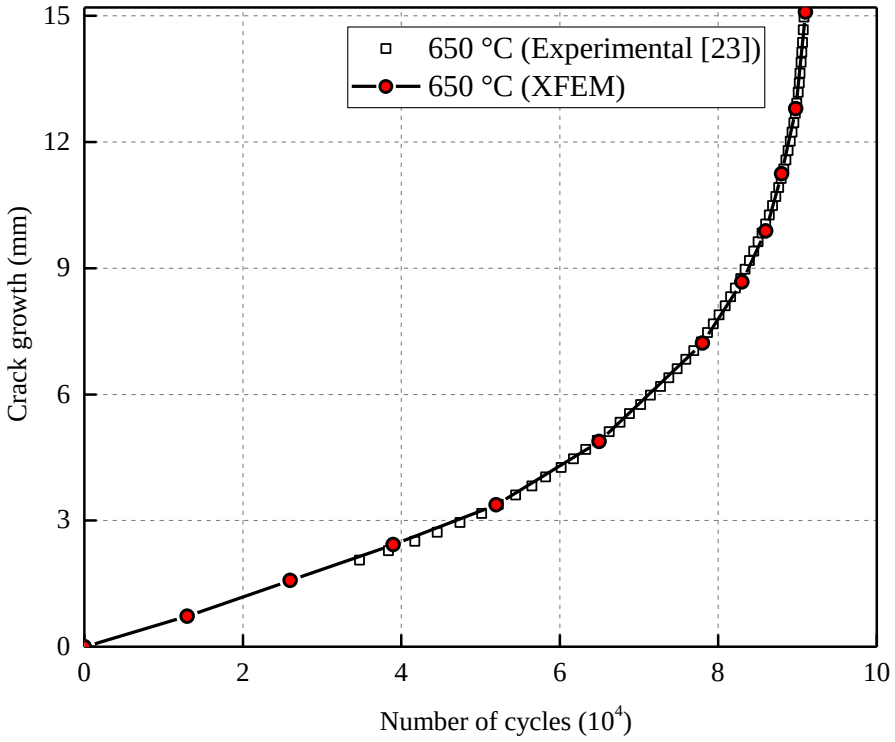



**Figure 6. A comparison of numerically predicted fatigue crack growth and experimental results for Ni-based superalloy at elevated temperature**

## Summary

In this paper, elasto-plastic FCG in the 3D domain is simulated using the XFEM. In the current study, FCGR is computed by the SIF based Paris law. The SIFs of individual modes at the ends of the line segments of the crack front are calculated via *J*-integral through *J*-decomposition approach. The variable fields are decomposed into symmetric and anti-symmetric portions across the crack surface in this approach. A virtual cylindrical domain has created at the ends of the line segments of the crack front to compute the *J*-integral. The nodal data is used to calculate all the required fields at the virtual domain by interpolation functions. In the presence of plastic deformation, the stress field at the virtual domain is obtained by a data transfer scheme. The direct

derivatives of stress and strain energy density are not possible hence function approximation is employed to compute the derivative of stress and strain energy density. The obtained fatigue crack growth for the compact tensile specimen is compared with the experimental results and found in a good match.

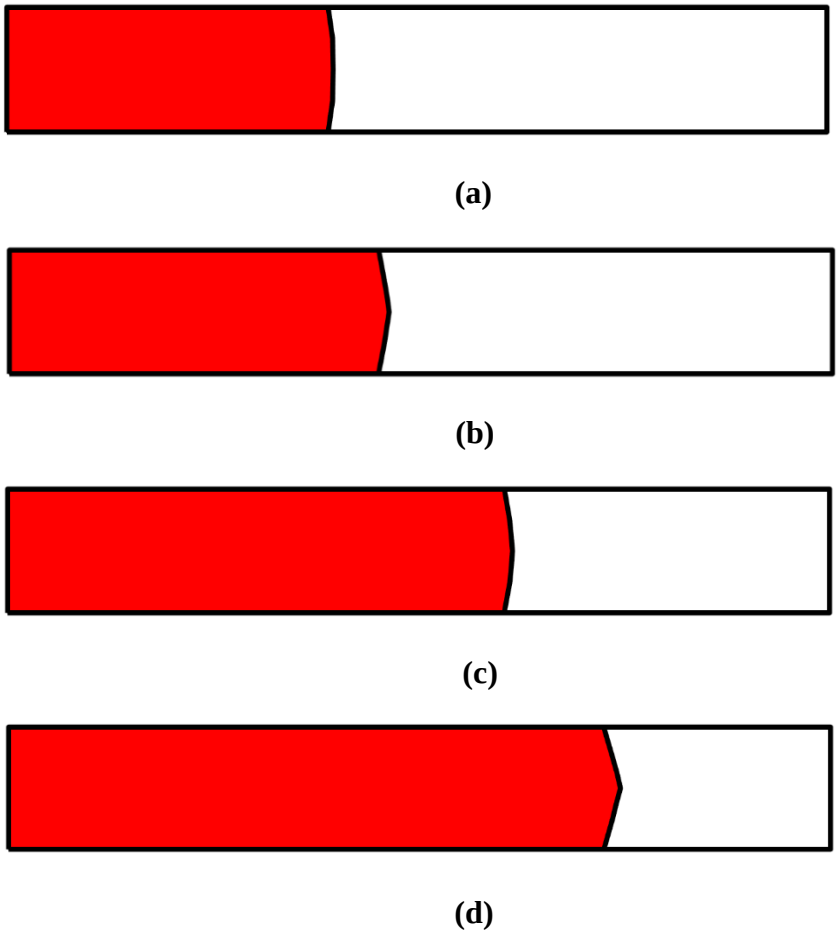



**Figure 7. Numerically predicted crack path at different stages of the simulation after (a) 13000 cycles (b) 52000 cycles (c) 86000 cycles (d) 91000 cycles**

## Acknowledgment

This research work is performed as the part of doctoral work of Mr. Manish Kumar under the fellowship of Ministry of Human Resource and Development, Government of India. This research work is also partially funded by the Alumni Association of Indian Institute of Technology Roorkee, India.

### References

[1] Yan, X.Q. (2006) A boundary element modeling of fatigue crack growth in a plane elastic plate, *Mechanics Research Communications* **33**, 470-481.

[2] Belytschko, T., Lu, Y.Y., and Gu, L. (1994) Element-Free Galerkin Methods, *International Journal for Numerical Methods in Engineering* **37**, 229-256.
[3] Pant, M., Singh, I.V., and Mishra, B.K. (2011) Evaluation of mixed mode stress intensity factors for interface cracks using EFGM, *Applied Mathematical Modelling* **35**, 3443-3459.
[4] Belytschko, T. and Black, T. (1999) Elastic crack growth in finite elements with minimal remeshing, *International Journal for Numerical Methods in Engineering* **45**, 601-620.
[5] Singh, I.V., Mishra, B.K., Bhattacharya, S., and Patil, R.U. (2012) The numerical simulation of fatigue crack growth using extended finite element method, *International Journal of Fatigue* **36**, 109-119.
[6] Luycker, E.D., Benson, D.J., Belytschko, T., Bazilevs, Y., and Hsu, M.C. (2011) X-FEM in isogeometric analysis for linear fracture mechanics, *International Journal for Numerical Methods in Engineering* **87**, 541-565.
[7] Singh, S.K., Singh, I.V., Mishra, B.K., Bhardwaj, G., and Singh, S.K. (2018) Analysis of cracked plate using higher-order shear deformation theory: Asymptotic crack-tip fields and XIGA implementation, *Computer Methods in Applied Mechanics and Engineering* **336**, 594-639.
[8] Shedbale, A.S., Singh, I.V., and Mishra, B.K. (2016) A coupled FE–EFG approach for modelling crack growth in ductile materials, *Fatigue & Fracture of Engineering Materials & Structures* **39**, 1204–1225.
[9] Sarkar, S., Singh, I.V., Mishra, B.K., Shedbale, A.S., and Poh, L.H. (2019) A comparative study and ABAQUS implementation of conventional and localizing gradient enhanced damage models, *Finite Elements in Analysis and Design* **160**, 1-31.
[10] Patil, R.U., Mishra, B.K., and Singh, I.V. (2018) A local moving extended phase field method (LMXPFM) for failure analysis of brittle materials, *Computer Methods in Applied Mechanics and Engineering* **342**, 674-709.
[11] Kumar, S., Shedbale, A.S., Singh, I.V., and Mishra, B.K. (2015) Elasto-plastic fatigue crack growth analysis of plane problems in the presence of flaws using XFEM, *Frontiers of Structural and Civil Engineering* **9**, 420-440.
[12] Pandey, V.B., Singh, I.V., Mishra, B.K., Ahmad, S., Rao, A.V., and Kumar, V. (2019) Creep crack simulations using continuum damage mechanics and extended finite element method, *International Journal of Damage Mechanics* **28**, 3-34.
[13] Patil, R.U., Mishra, B.K., and Singh, I.V. (2017) A new multiscale XFEM for the elastic properties evaluation of heterogeneous materials, *International Journal of Mechanical Sciences* **122**, 277-287.
[14] Kumar, S., Singh, I.V., Mishra, B.K., and Singh, A. (2016) New enrichments in XFEM to model dynamic crack response of 2-D elastic solids, *International Journal of Impact Engineering* **87**, 198-211.
[15] Rigby, R.H. and Aliabadi, M.H. (1998) Decomposition of the mixed-mode J-integral - Revisited, *International Journal of Solids and Structures* **35**, 2073-2099.
[16] Kumar, M., Singh, I.V., Mishra, B.K., Ahmad, S., Rao, A.V., and Kumar, V. (2018) Mixed mode crack growth in elasto-plastic-creeping solids using XFEM, *Engineering Fracture Mechanics* **199**, 489-517.
[17] Moës, N., Dolbow, J., and Belytschko, T. (1999) A finite element method for crack growth without remeshing, *International Journal for Numerical Methods in Engineering* **46**, 131-150.
[18] Shedbale, A.S., Singh, I.V., Mishra, B.K., and Sharma, K. (2017) Ductile failure modeling and simulations using coupled FE-EFG approach, *International Journal of Fracture* **203**, 183-209.
[19] Bhardwaj, G., Singh, I.V., and Mishra, B.K. (2015) Fatigue crack growth in functionally graded material using homogenized XIGA, *Composite Structures* **134**, 269-284.
[20] Kumar, M., Bhuwal, A.S., Singh, I.V., Mishra, B.K., Ahmad, S., Rao, A.V., and Kumar, V. (2017) Nonlinear fatigue crack growth simulations using J-integral decomposition and XFEM, *Plasticity and Impact Mechanics* **173**, 1209-1214.
[21] Durand, R. and Farias, M.M. (2014) A local extrapolation method for finite elements, *Advances in Engineering Software* **67**, 1-9.
[22] Kumar, M., Singh, I.V., and Mishra, B.K. (2019) Fatigue Crack Growth Simulations of Plastically Graded Materials using XFEM and J-Integral Decomposition Approach, *Engineering Fracture Mechanics* **216**.
[23] Kumar, M., Ahmad, S., Singh, I.V., Rao, A.V., Kumar, J., and Kumar, V. (2018) Experimental and numerical studies to estimate fatigue crack growth behavior of Ni-based super alloy, *Theoretical and Applied Fracture Mechanics* **96**, 604-616.

# Experimental analysis of T-pipe joints forming

**†*Jarosław Bartnicki[1], Jarosław Świtacz[1] and Zbigniew Pater[1]**

[1] Lublin University of Technology, Poland

*Presenting author: j.bartnicki@pollub.pl

†Corresponding author: j.bartnicki@pollub.pl

## Abstract

The solutions for different liquid, sand, gases transportations are very different. Many of them concern applications of different advanced polymer technologies but for high pressure pipe ducts steel materials are commonly applied. This paper presents results of numerical calculations and experimental verification of T-pipe joints forming process. FEM analyses were conducted for the chosen process technological parameters and tools designs. The kinematics of metal flow in the area of the formed flanges was analyzed. Distributions of stresses, strains and damage criterion during forming were determined. Calculated values of forces and moments acting on tools and workpieces allow for designing of tools geometry for experimental verification of the proposed forming process. Worked out numerical calculations of T-pipe joints forming show practical possibility of this process application. Regarding the existing solutions for this type of parts manufacturing it should be interesting to develop special device for this activity dedicated to the hydraulic press or for handy operated devices for smaller pipe diameters.

**Keywords**: FEM, design optimization, cold forming

## Introduction

There are various solutions for transportation of different liquid, sand and gases. A lot of them concern applications of different advanced polymer technologies but for high pressure pipe ducts steel materials are commonly applied. For this reason, it is of great importance to preserve good sealing of the designed pipelines. Different welding technologies are normally used for connections making but especially difficult case among them is T-pipe joints (tee) forming and welding with another pipe structures. In many cases connections made with T-pipe joints are performed by drilling the wholes inside one pipe and its welding into another. A flanging technology is applied for increasing mechanical characteristics and durability of these connections. This solution is favorable for pressure distribution and makes the welding processes easier for making and verifying – Fig. 1. In comparison with a traditional solution, this one is also less material consuming and it can be also applied in existing installations using portable devices for drilling and tee making. Typical applications areas include food processing, pharmaceutical, pulp and paper industry, stainless steel piping systems, water treatment, shipbuilding and conventional and nuclear submarines [1]. The proposed solution for bigger pipe installations must be equipped with special machine for tee making. This aggregate consists of drilling section for elliptic hole making and a special flanging head. Combining rotations and axial movement of this element results in flanges (collars) creation. It is possible to reduce smaller pipes tee making by means of portable tools as this process does not require application of bigger values of forces and torques. For the planned new devices for T-pipe joints development, the observations of this problematic and numerical verification of material flow and technological parameters are fundamental. Tools geometry

and load parameters knowledge combined with final parts dimensions accuracy [2-4].are the most significant for new solution designing.

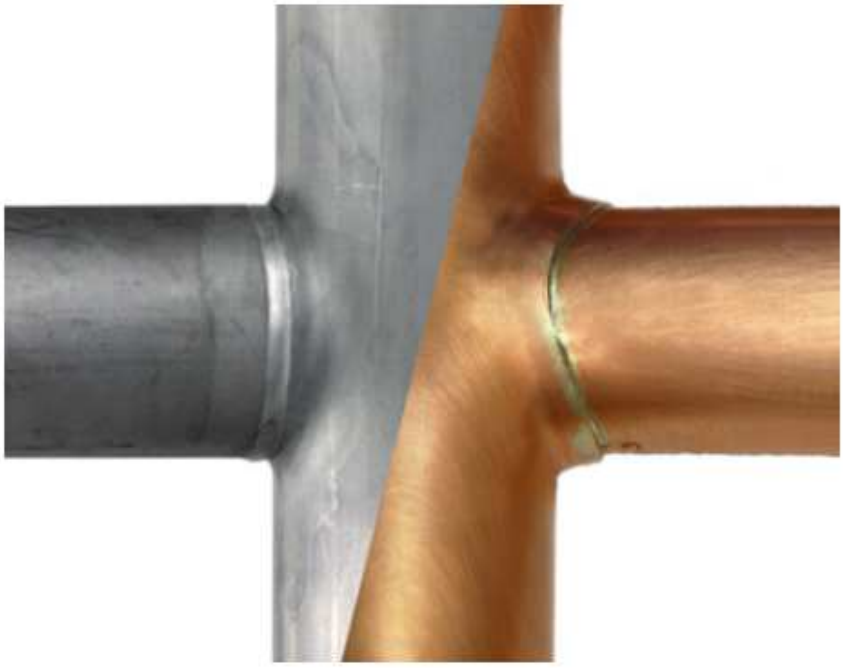


Fig. 1.The example of traditional (left) and formed by flanging T-pipe joint (right) [1]

**Numerical modeling**

Deform 3D software was used for numerical calculations of the proposed scope of flanging processes. Pipes with external diameter Ø150 and created collars with diameters Ø50 and were applied in numerical models. Weldable fine-grain structural steel for pressure vessels P355NH was used [2] as the material model. For the calculations needs in the cold metal forming conditions all tools and workpieces have the same initial temperature T = 20°C.The constant friction model was assumed for all cases with friction factor m = 0.25, as the processes are completed with good lubrication. Some calculations were realized with workpiece mounted inside the jaws with a diameter corresponding to the pipe dimensions. After verification of state of stress and strain, the workpiece were limited to the halfpipes fixed by boundary conditions at the external edges. For the proper circular collar shape forming the initial holes inside the pipes was designed as elliptic calculating geometrically created collar. For example, the collar Ø50 mm formed into the pipe Ø150 mm required initial elliptic whole 30 x 37 mm. Fig. 2 presents the examples of worked out models of the analyzed cases. The first model consists of formed pipe with initial elliptic whole, outer pipe as mounting for workpiece and two bars – tools which flange the collar. In the second, developed model physical fixation is replaced by boundary conditions with fixed edges (red nodes in Fig 2). The box with fine mesh was used for time calculation optimization in later realized simulation These modifications enable faster simulations.. All these changes were made after verifications of state of stress with the first model without any changes.

**Feed rates and tools configurations**

Feed rates f written in mm of working tools movement per rotation were crucial for the analyzed cases technological values. This parameter is strictly combined with dimensional accuracy and effective time of the realized collar tee connections. It was assumed, analyzing accessible solution [1], to apply solution with two rotating and axially moving tools. The forming tools – bars have the same diameter Ø10 mm and specially rounded ends. According to the chosen option, the feed rates applied in calculations were 0.5; 1.0 and 2.0 mm per one rotation. Tools – bars forming angle α were regarded as another very important technological parameter in this forming process The values for this parameter were taken initially within the range of α = (35÷60)°. Finally, this parameter was changeable during whole process starting from the biggest values α = 60° during necking and reaching value α = 0° at the final sizing stage of collar flanging. Final strain distribution is also changed by the forming angle choice. Bigger values of strain with significant part of redundant strains were obtained due to smaller inclination angle and feed rates.

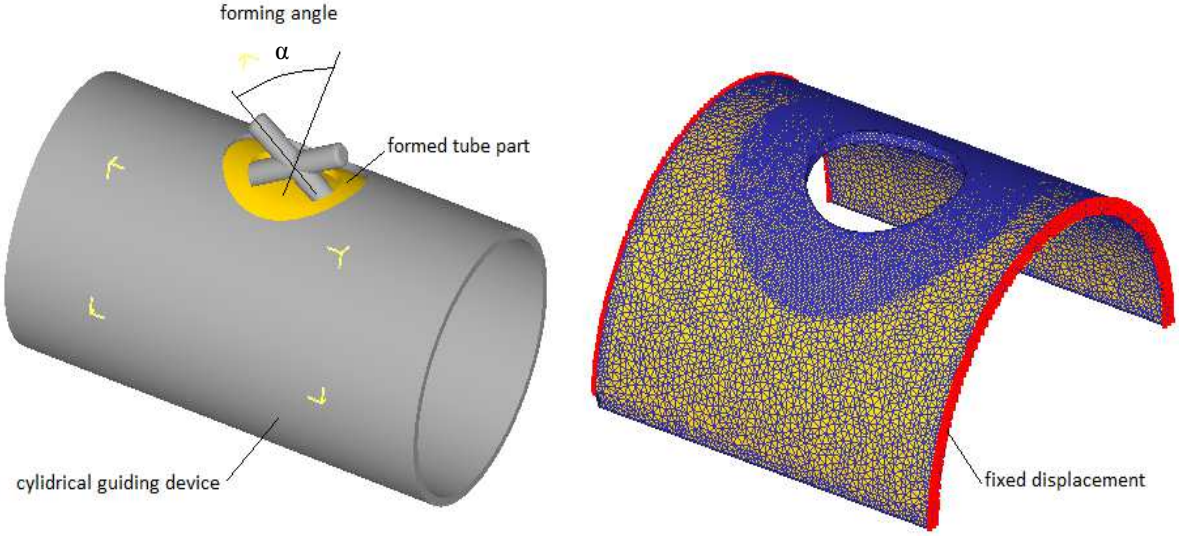



Fig.2. Worked out numerical models for calculations with physical pipe stabilization (left) and with boundary conditions with fixed displacement (right)

During numerical simulations realized with the presented assumptions and technological parameters it was possible to verify distributions of stress and strain during collars flanging. In Fig. 3 and Fig.4 the results of calculations for Ø50 collar forming into Ø150 mm pipe are presented. As it is shown in Fig. 3 the distribution of effective stress is really local and maximum values of this parameter exceed 1000MPa [5]. In zones of contact between tools and workpiece material is moved towards the direction of axial movement of tools. Presented in Fig. 4 effective strain distributions show real range of material deformation during flanging process. Values of this parameter are slightly different according to number of moving tools but the distributions are depended on different feed rates applications Bigger values of feed rate f = 1 and 2 mm per rotation provide to obtaining effective strain values reaching 12÷14. Normally this information is important considering fine grain material structures into plastically formed zones. However this observation must be verified in experimental tests in real flanging collars device [6-8]. After short verification of numerical calculations results, it was decided to design a handy operated device for smaller T-pipe forming up to 150 mm tube diameters with 50 mm flanges. Due to choosing two working bars head (caused by symmetry of loads), it was also decided to apply changeable forming angle values during forming.

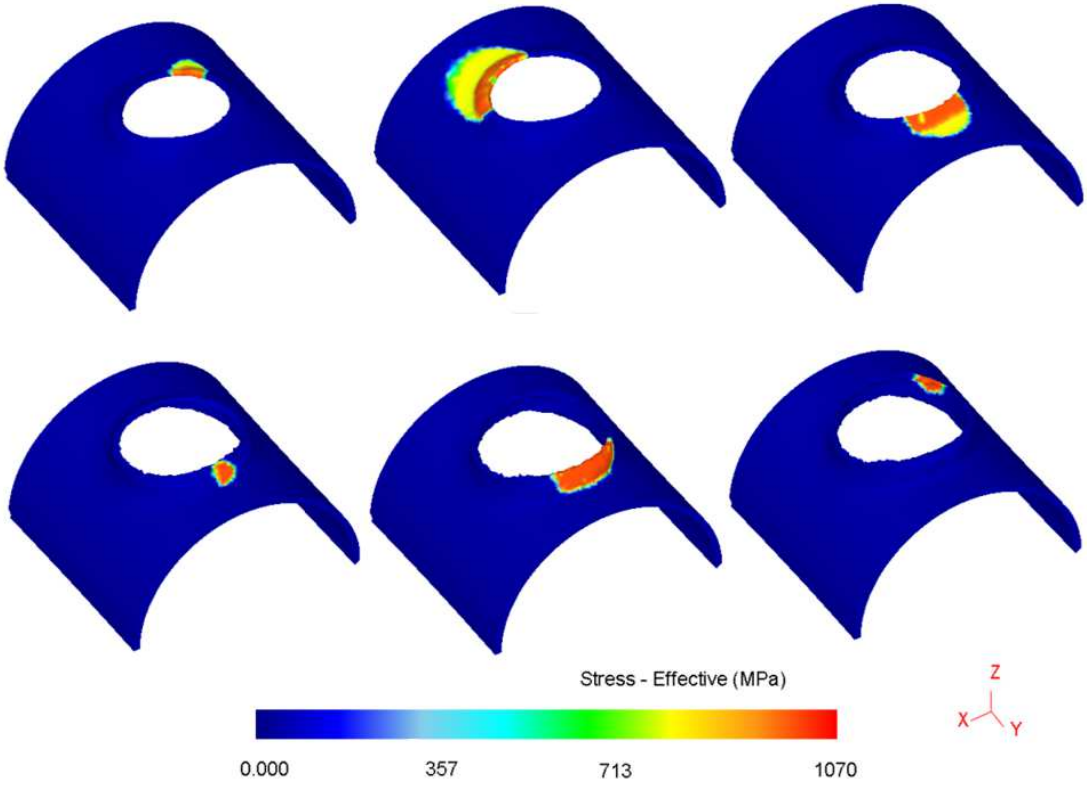



Fig.3. Progression of shape and distribution of effective stress during Ø50 collar forming into Ø150 mm pipe

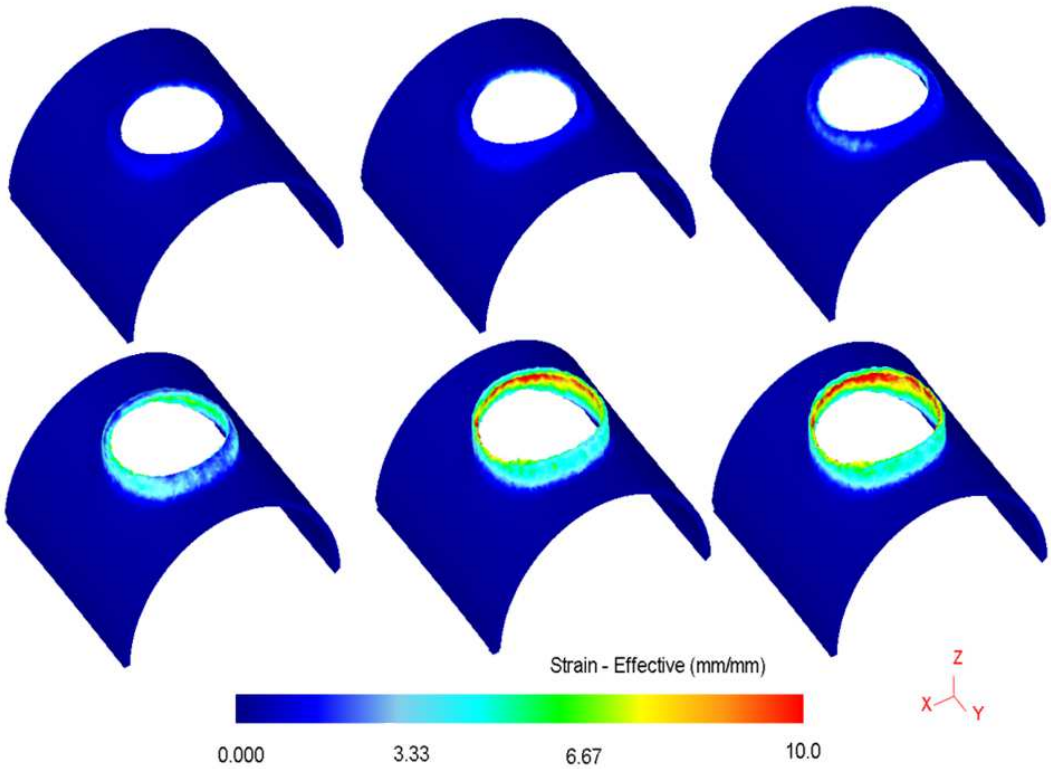



Fig.4. Progression of shape and distribution of effective strain during Ø50 collar forming into Ø150 mm pipe

Virtual model and real device for T-pipes flanges forming are presented in Fig. 5. Additionally, for verification of different forming angle α influence on the process course and final flange accuracy, three different working head with two working rounded edges were prepared – Fig. 6.

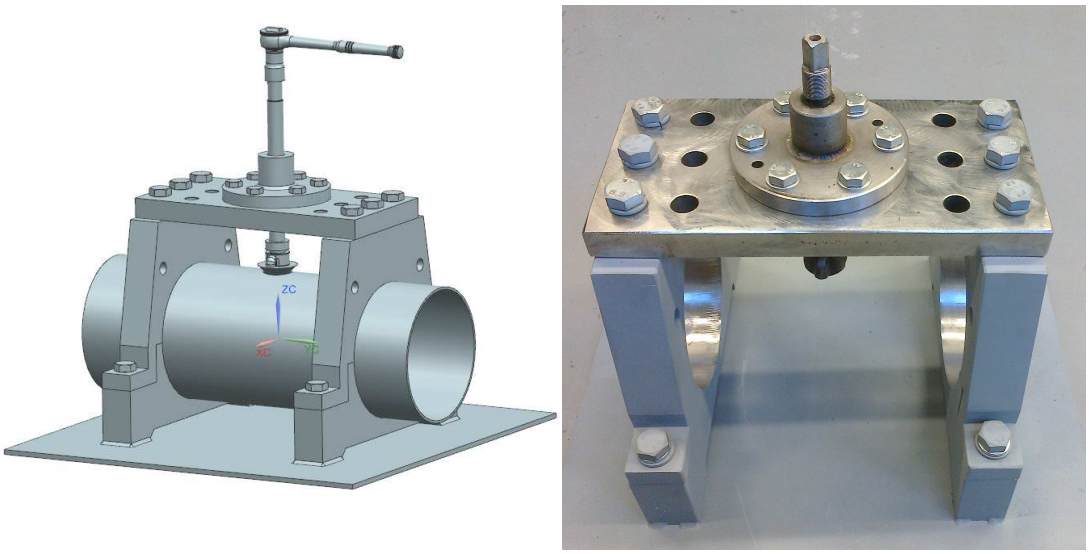


Fig. 5. Designed virtual model and worked out device for T-pipe forming

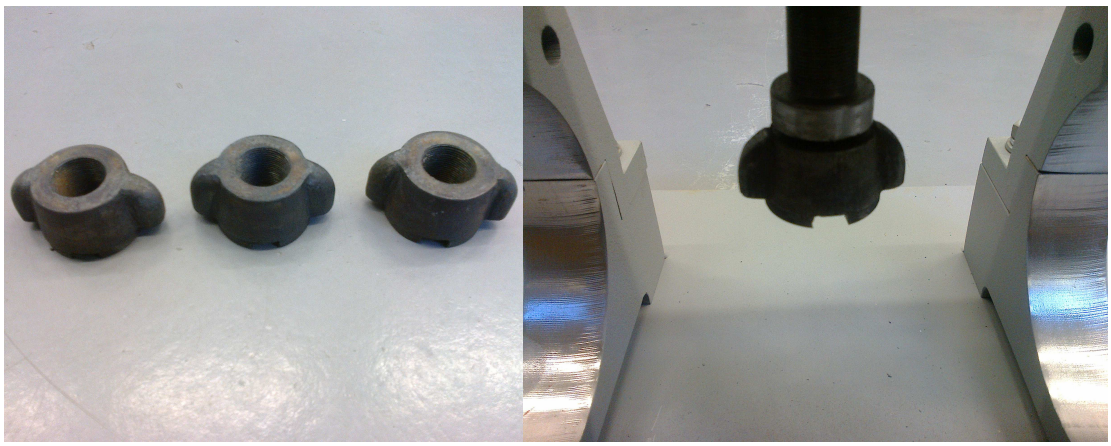


Fig. 6. Three different working heads with two working rounded edges and tools mounted on towing screw into the workspace

Designed for experimental tests three working heads have initial forming angle value α equal 65°; 50° and 35°. Finally in all cases this angle decreases up to 0° for effective flange internal diameter sizing. Results of 50 mm flange forming on 150 mm diameter tubes are presented in Fig. 7. Presented flanges were formed with feed rate f = 1 mm per rotation. Obtained after this experiment the assumed shape of collars shows the influence of working tools geometry on final wall thickness distribution at formed area and torque values. In all analysed zones of collar forming there are no significant decreases of wall thickness. This information is very important for the future scope of usage of T-pipe joint in welding of high pressure pipe ducts with very elevated mechanical characteristics. The differences between calculated and obtained in experiment distributions of collar wall thickness are not significant. The application of the pipe with initial wall thickness 4.00 mm results in minimal calculated value 3.48 mm and measured in experiment 3.56 mm. During the process observation it was stated that the friction conditions are very important for this values distribution. The application of

old pipes with rusted inner surface in experiments results in bigger wall thinning during flange forming. The application of parts from carbon steel with analogical dimensions and process conditions results in final collar wall thickness equal 3.16 mm. Presented result were obtained using working head with initial value of forming angle α = 50°.

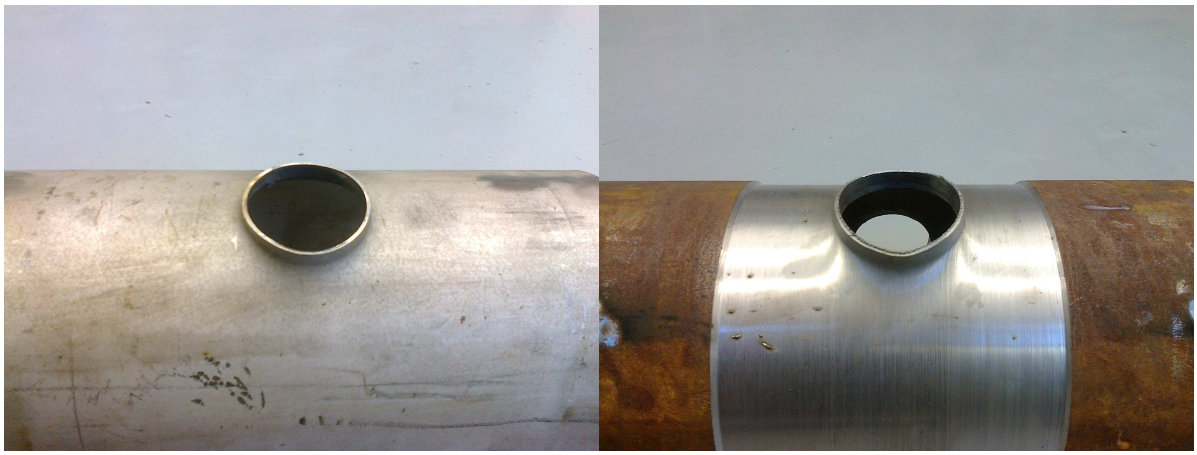


Fig.7. Flanges of 50 mm formed on 150 mm diameter tubes (new inox steel tube – left and used carbon steel tube - right)

Obtained during experimental tests maximum values of torques during flange forming depending on working head forming angles and types of tubes are presented in Tab 1. Assumed maximum values for handy – operated device was 300 Nm. This value was not exceed in whole experiment range for new tubes. During tests with presented in Fig. 7 (right side) used carbon steel tube measured maximum value of torque equal 380 Nm. It results mentioned above the biggest reduction of wall thickness. However for bigger tube diameters or wallthickness new device with servomotor is designed and it will be realized in practise.

Tab.1 Measured torques values during flange forming

| Initial α angle, ° | 65° | 50° | 35° |
|---|---|---|---|
| Inox steel tube torque, Nm | 240 | 250 | 270 |
| Carbon steel tube torque, Nm | 200 | 210 | 240 |

**Conclusions**

Worked out numerical calculations of T-pipe joints forming show practical possibility of this process application. Regarding the existing solutions for this type of parts manufacturing it should be interesting to develop special device for this activity dedicated to the hydraulic press. Some results concerning especially feed rate f and inclination angle α and initial elliptic hole dimensions will be very useful for this process recognizing. The observed significant torque values reduction depending on applied feed rates is very useful for handy operated worked out portable device. In this case, the crucial limitation is pipe fixation during collar forming in existing installation. The device body stiffness is also very important because it will decide about whole dimensions accuracy and tools positions stability. Another very important factor are friction conditions determining parts wall thickness distributions important for certification procedures of these T-joints. Finally, all these issues will decide on the scope of the proposed solution applicability and safety rules for users. Presently realized works at this field concern designing and working out stationary device for flanging of tubes up to 400 mm diameters.

## References

1. https://t-drill.com/technologies/tee-forming/
2. Montanari L, Cristino VA, Silva MB, Martins PAF Int J Adv Manuf Technol 69(5) (2013)
3. Thipprakmas S, Phanitwong W Int J Adv Manuf Technol 61(5) (2012)
4. Cao TT, Lu B, Ou H, Long H, Chen J Int J Mach Tools Manuf 110(2016)
5. Bartnicki J, MATEC Web of Conferences, Vol 190, p 04005 (2018)
6. Yang C, Wen T, Liu LT, Wang H J Mat Process Technol 214(2014)
7. Ben, Ning-Yu; Zhang, Da-Wei; Liu, Nan; Zhao, Xiang-Ping; Guo, Zhi-Jun; Zhang, Qi; Zhao, Sheng-Dun. Int J of Adv Manuf Technol 2017, Vol. 93, 9-12, p3189-3201. 13p.
8. Yantao Yin; Shicong Li; Liangbi Wang; Mei Lin; Qiuwang Wang. Energies, Vol 11, Iss 4, p 864 (2018)

# Fatigue life estimation of riveted joints using crack growth concept

**†*Jakub Šedek[1], Tomáš Mrňa[1], Ivan Mlch[1], Pavel Kucharský[1]**

[1]Department of Strength of Structures, Czech Aerospace Research Centre, Beranových 130, 19905 Prague – Letňany, Czech Republic.

*Presenting author: sedek@vzlu.cz
†Corresponding author: sedek@vzlu.cz

## Abstract

Fatigue life estimation of riveted joints with countersunk head was performed by the crack growth analysis and results were compared with the experimental observation. The multi-site damage (MSD) in the skin made from the aluminium alloy D16č ATV was assumed so that the simultaneous development of fatigue cracks emanating from holes in the same structural element was studied. The equivalent initial flaw size approach (EIFS) was employed. The series of ten simultaneously growing cracks was analysed by methods of linear elastic fracture mechanic and crack growth concept utilizing FASTRAN model. Stress intensity factors used in the crack growth law were determined from the finite element model created in ABAQUS FE package. Cracks development under loading at several load levels was simulated. Assuming initial corner cracks of 0.125 mm in size, the analysis results show the fatigue lives close to the experimentally determined S-N curve of tested specimens. The considerable part of the fatigue life of the joint can be attributed to the crack growth. The analysis results exhibit that the smaller load level yields the longer crack initiation. The simulated fatigue life of tested specimens determined by the crack growth concept can be used as the upper bound for the design purposes in involved joint configuration very well.

**Keywords:** Crack, Fatigue, FEM, MSD, Rivets

## Introduction

The connection of skins in an airplane structure is commonly made by rivets. Several types including solid rivets and blind rivets with round, flat, countersunk or pan head are being used. The countersunk head is special for its smooth surface of final outer side and for that it is used in aerospace widely. The disadvantage is the sharp shape in the hole caused by the countersunk and due to this fact, the fatigue life can be reduced considerably in comparison with rivets that require blind hole [1]. Expected stress peak on the hole edge does not need to be crucial because it can be reduced by residual stresses created during riveting process [2][3][4]. The residual stress can beneficial to the fatigue performance of riveted joints [5].

The durability of the joint is influenced by many factors and the fatigue life prediction is far from easy. The experimental evaluation is used for S-N curve determination which depicts cyclic stress amplitude versus number of cycles to failure. The fatigue life under constant amplitude loading at different load level can be predicted based on the S-N curve. Alternatively, more complex prediction models can be employed [6][7]. The means of fracture mechanics are also utilized [8][9]. The fatigue life of the joint is analysed according to a crack advance, but some initial crack should be assumed. The analysis can be conducted for different joint configurations without experimental S-N curve data, but the crack growth

rate data together with crack growth model are necessary. The introduced analysis is performed by this approach and the comparison with the experimental S-N curve is finally presented.

**Experimental evaluation of fatigue life**

In test laboratories of the Czech Aerospace Research Centre, the experimental fatigue life verification of several rivet types was carried out [1]. The result of countersunked rivets were different from other types due to the failure occurring in the centre part representing a skin, while the failure of straps was common in other rivet types (see Figure 1). The quadriplicated configuration of the test specimen was used. The methodology of fatigue tests was based on fatigue loading until failure. Whereas, after the failure of one joint, the specimen was split up to enable further testing of remaining joints of the specimen. The joint consisted of the central part 1.5 mm thick and two 0.8 mm thick strips symmetrically placed on the top and the bottom. The rivets in each joint were arranged in four rows with five rivets. The rivets were specified as 5DuZz 3x7P according to the manufacturer standards. The main head of the rivet is countersunked, the shank is 3 mm in diameter and the secondary head is being created by squeezing the rivet shank with final cylindrical shape. The type of rivets was the same in all joints in the test specimen. A set for the fatigue life evaluation consisted of six pieces of test specimens.

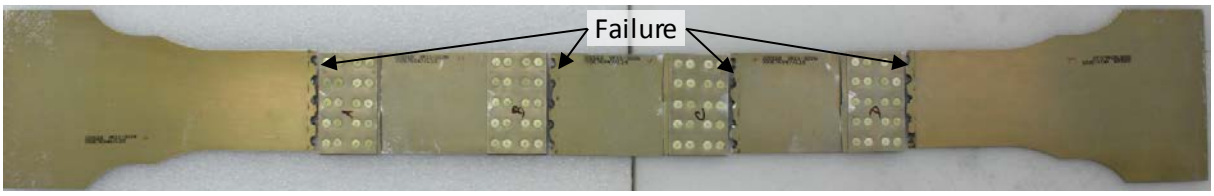



**Figure 1: Tested and analysed fatigue lives of the riveted joints with countersunk head.**

Specimens were loaded by monotone loading with constant amplitude of the force, the stress ratio $R$ = 0.05 and the frequency from 3 Hz up to 8 Hz. The tests were performed at room temperature and ambient laboratory conditions. The uniaxial hydraulic test machines INOVA ZUZ 100 with load cell capacity of 100 kN was used for fatigue tests. The maximum stress values $\sigma_{max}$ of the central part gross section were defined on five levels from 83 MPa up to 152 MPa to cover S-N curve uniformly.

During the fatigue experiment the failure of the central part only has been occurring. No failure of the rivets happened. Unfortunately, due to the cover by straps, the crack growth was not able to monitor. The cracks developed through rivet holes and only the final failure was noticed. The life of the joints was in the range of 1e4 up to 2e6 cycles. The fatigue data were evaluated using the linear regression by means of a linear model represented by the equation

$$\log N_f = A_1 + A_2 \log \sigma_{\max} \tag{1}$$

with determined coefficients for tested configuration $A_1$=12.87 and $A_2$=3.98 and coefficient of determination $r^2$ = 0.98.

After the failure the crack surface was photographically documented. The example tested under load conditions with $\sigma_{max}$ = 48.1 MPa is shown in Figure 2. Based on the fractographic examination carried out on a macro scale utilising optical microscopy the crack initiation

points were identified. In the Figure 2 they are marked by red arrows. Fatigue cracks initiated in the rivet holes. The crack initiation was clear at site holes shown in Figure 3. In the detail the corner crack shape is obvious and the surrounding crack surface structure shows changes in the macromorphology of the surface. The finding indicates the damage of the surrounding area in the cross section due to the final failure.

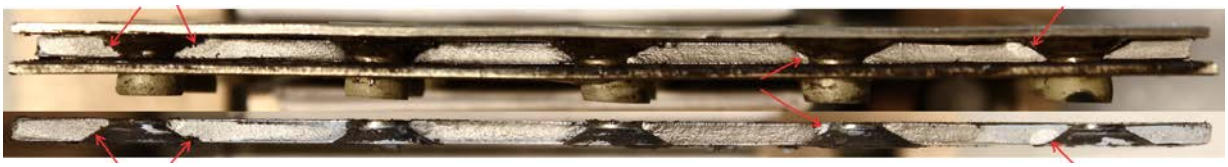


**Figure 2: Several crack initiations from the rivet holes; $\sigma_{max}$ = 48.1 MPa.**

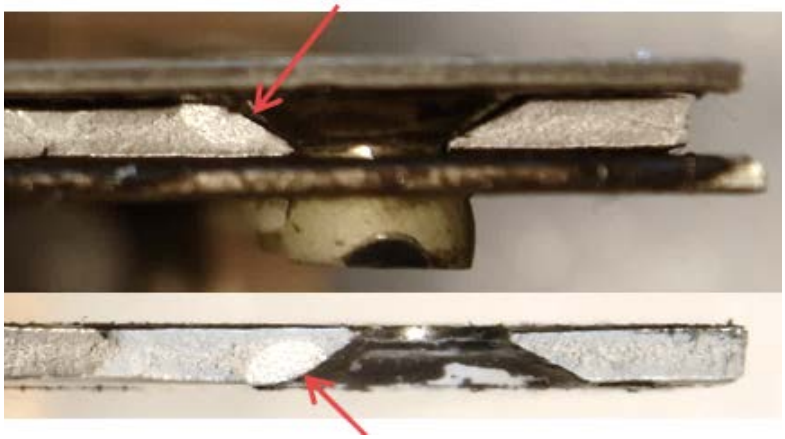


**Figure 3: Detail of crack initiation point at the side hole; $\sigma_{max}$ = 48.1 MPa.**

The interesting note is that the crack initiation spots were located mainly at the corner of intersection of the countersunk and the outer surface and not at the opposite sharp edge as can be expected due to stress concentration effect.

**Simulation of the fatigue life**

The joints life can be split into the period of a crack initiation and a crack growth. The boundary between these periods is not strict and is very difficult to distinguish it quantitatively. In microscale the transition depends on the microcrack size and the microstructural barriers. The size of the crack at the transition point can be different for different types of materials. The point of the transition may be specified rather qualitatively. The initiation period is the surface phenomenon and the crack growth period starts if the crack growth resistance of the material is controlled by the crack growth rate [8]. The initiation period can be significant portion of the life. However in the airplane structures the crack growth period is being essential and the considerable attention is paid to it.

The macromorphology of inspected failure surfaces of tested specimens indicated several cracks emanating from the rivet holes. The cracks growth was not documented due to the straps that covered the cracks for the hole life of the joints. The simulation of crack growth could point out the severity of the period of crack growth in the fatigue life.

A simultaneous crack growth scenario was chosen to analyse employing equivalent initial flaw size (EIFS) approach. According to Joint Service Specification Guide of USAF for

aircraft structures (JSSG-2006) [9] small imperfections equivalent to an 0.127 mm radius corner flaw resulting from material and structure manufacturing and processing operations are assumed to exist in each hole of each element in the structure. In slow crack growth structure at holes and cut-outs, the assumed initial flaw is a 1.27 mm radius corner flaw at one side of the hole. When the primary damage and crack growth originates in a fastener hole and terminates prior to member or element failure, continuing damage should be an 0.127 mm radius corner flaw plus the amount of growth which occurs prior to primary element failure emanating from the diametrically opposite side of the fastener hole at which the initial flaw was assumed to exist. The presumption of initial 1.27 mm flaw can be sometimes conservative but it is prescribed to declare the crack resistant feature of the structure.

In the presented cracks scenario the cracks were developing simultaneously from the same size in configuration of corner cracks of 0.13 mm in size located at the edge of countersunked holes. This arrangement was chosen in order not to give priority to a one larger crack. The analysis of crack growth was performed by step-by-step routine so that the cracks were extended from the previous configuration based on the computed crack increment. The characteristics of linear elastic fracture mechanics for each modeled step were determined from finite element model and the crack extension between modeled steps was determined by the crack growth model.

**Numerical model**

The numerical model of cracks configuration in each step was created in the FE package ABAQUS 2017. A symmetry was applied on the model so that only one quarter of the model was analysed. The example of crack configuration is shown in Figure 4. Model was discretized by tetrahedral (C3D4) and hexahedral elements (C3D8I). The mesh was refined around a crack front using only hexahedral elements stacked in ten layers (see Figure 5) with the smallest element dimension of about 0.002 mm. Regular mesh around a crack front is used to have a good results of contour integral for stress intensity factor calculation. Isotropic linear elastic material model with E = 72 000 MPa and ν = 0.33 was employed. The model was loaded by stress applied on tension site and boundary conditions respected the used symmetry.

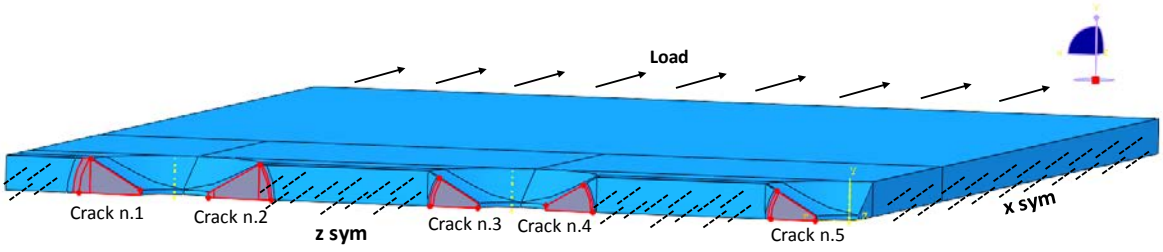



**Figure 4: 1/4 model of the skin part including cracks; top-front view on the geometry.**

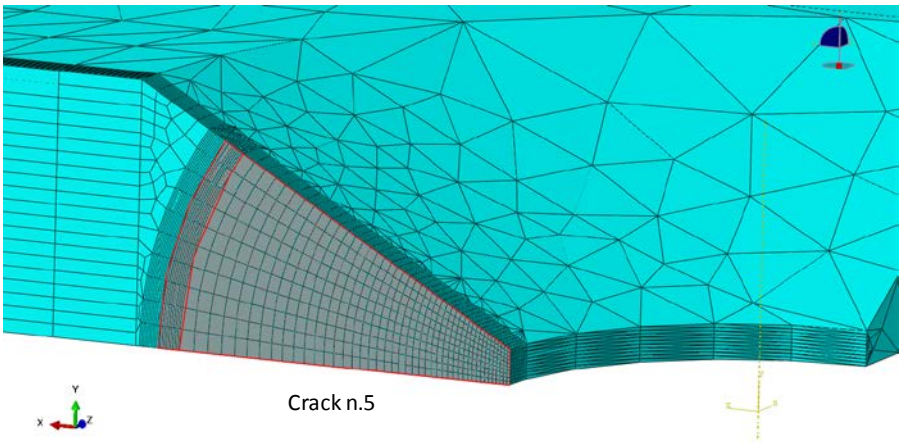



**Figure 4: Detail of the mesh around the crack n. 5.**

The stress intensity factor $K_I$ was determined by internal ABAQUS routine using contour integral method. Only the opening mode I was assumed in fracture mechanics parameters determination thus the $K$ always corresponds to mode I if stated without index in this work. There were from ten up to twenty contours evaluated in ten layers around a crack front typically. The mean value of the stress intensity factor from each contour was determined and one mean value $K_{char}$ from all layers was used as characteristic for a crack. Typical normalized stress intensity factors $K/K_{char}$ according to contours are plotted in Figure 5. The value of $K$ stabilizes after some contours away from the crack front except the one on the side of the crack (the blue dots in Figure. 5). The mean value eliminates this discrepancy.

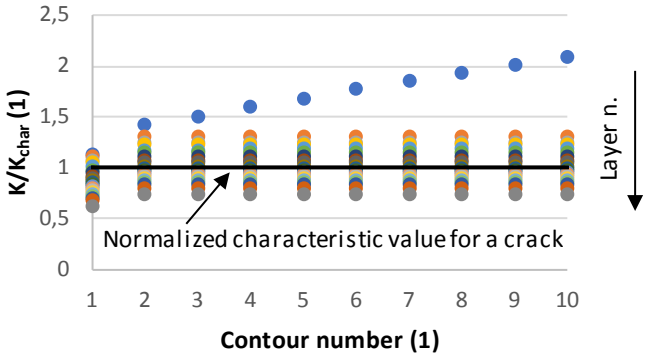



**Figure 5: Stress intensity factor around the crack front of the crack n.5 with the length of 2.1 mm.**

## Material data

In the numerical model cracks were extended at the same moment of elapsed cycles. The crack increments were determined by the crack growth model FASTRAN. The specimens were loaded with constant amplitude and despite the basic Paris law could be sufficient, more sophisticated model was used. At high crack growth rates, the effect of constrained loss can be expected, but only some models can capture this effect.

The crack growth rate data were used according to [11], where the data of Russian alloy D16CzATWH are documented. The crack growth properties are considered to be similar to the 2024-T3 alloy [12]. These data were used in presented work based on the results from

preliminary data testing. The results of simulated crack growth under constant amplitude loading in M(T) specimens using D16czATWH data for FASTRAN model were close to experimentally observed growth in specimens made from D16cATV. The crack growth rate data were interpolated by linear relationship in the form

$$\frac{\Delta a}{\Delta N} = C\left(\Delta K_{eff}\right)^{m} \tag{2}$$

with coefficient $C$ = 1.2e-10 and $m$ = 3.5. The effective value of stress intensity factor $K_{eff}$ depends on the opening stress intensity factor $K_{open}$ according to

$$\Delta K_{eff} = K_{\max} - K_{open} \tag{3}$$

$K_{open}$ is determined by the FASTRAN model considering load history by semi-analytical method. It requires the value of constraint factor α. The implementation in AFGROW [13] enables using one constraint value through the whole crack growth or two values according to crack growth rates at transitions points from plane strain to plane stress. These rates are attributed to points of flat-to-slant transition visible on fracture surfaces [8, 14]. The rate of 1e-7 m/cycle together with α = 2 were used for plane strain transition point and the rate of 2.5e-6 together with α = 1.2 were used for plane stress transition point.

The cracks were characterised by the shape function $\beta$. Its values were determined in step points normalizing the values of stress intensity factor $K_{FEM}$ determined from the numerical model by the term $\sigma\sqrt{\pi a}$, where $\sigma$ denotes the remote stress and $a$ is the crack length.

$$\beta = \frac{K_{FEM}}{\sigma\sqrt{\pi a}} \tag{4}$$

The crack increments to find the crack fronts in the step $i+1$ were determined from the actual step $i$ extrapolating the $\beta$ values linearly from previous points $i-1$ and actual point $i$. The resulting values of $\beta$ therefore occur in zig-zag pattern around middle theoretical curve. In order to avoid this behaviour, the iteration process should be employed as presented in [15], but it was not used in this work accepting some loss of accuracy.

## Cracks growth

The analysis of crack growth was carried out step-by step. The multi-side damage (MSD) scenario was assumed and all cracks were influencing each other through the simulation during the whole fatigue life. This type of analysis is not common due to its cost. More convenient is to analyse crack extension of the most loaded crack and after reaching the opposite side in the load carrying cross-section, the crack increment of the other crack is determined. Note that the crack increment during the growth of primary crack must be included. The initial crack length for all cracks was 0.13 mm assuming the crack shape as a circle section. A crack was therefore corner shaped up to reaching the whole thickness.

## Results

The analysis of fatigue life was performed with the $\sigma_{\max}$ = 48, 83 and 112 MPa independently. At each step of crack increments the numerical model was adapted in order to obtain the values of shape function. The shape function plots are in Figure 6.

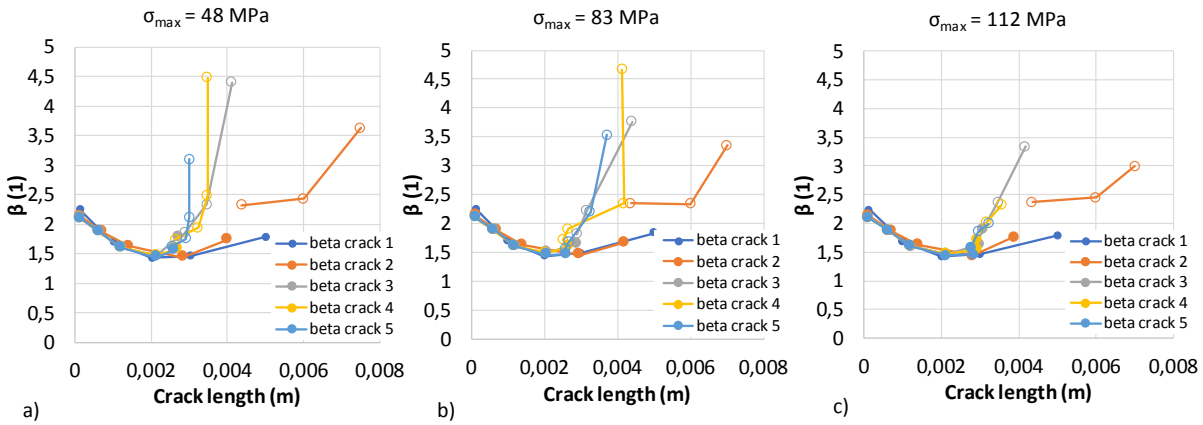



**Figure 6: Shape functions of cracks in analyses with a) σmax = 48 MPa b) σmax = 83 MPa c) σmax = 112 MPa.**

The shape function for small crack lengths is identical up to the size of 2 mm that corresponds to the change of the shape from the corner crack to the through crack. Up to this point cracks do not influence each other. The other change of the shape function occurs when the crack terminates. The full dots mark the state before terminating of the crack n. 1 and the circle dots correspond to the crack after that. The rapid rise of $\beta$ value for the crack n. 2 immediately after terminating the crack n. 1 is obvious.

From the plot of crack growth curves shown in Figure 7, it is clear that the fatigue life is determined by the crack n. 1. After terminating the crack n. 1 the other cracks growth is very fast.

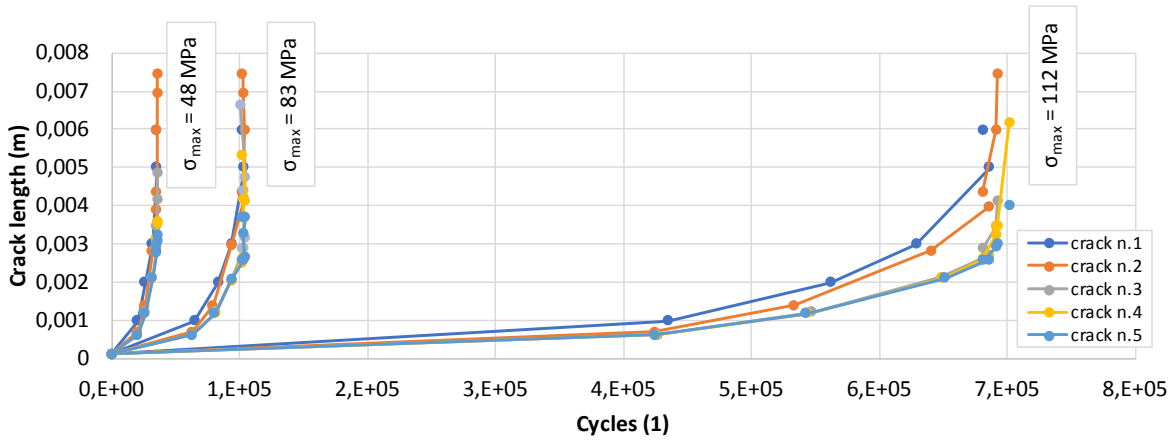



**Figure 7: Shape functions of cracks in analyses with σmax = 48, 83 and 112 MPa.**

Based on the result that the fatigue life is mainly determined by the crack n.1 the fatigue life with the levels $\sigma_{max}$ = 64 MPa and 152 MPa was determined using the shape function of the crack n. 1. The results are shown in logarithmical plot in Figure 8 together with the experimental results of specimen lives. The test data can be interpolated by power function drawn in log-log plot by linear curve.

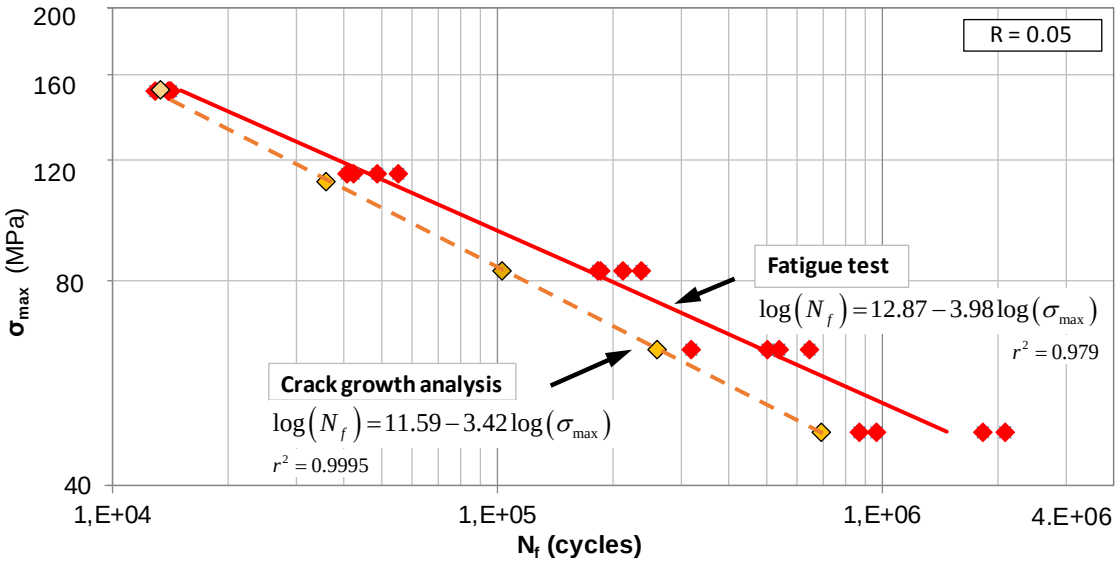



**Figure 8: Fatigue life of joint specimens; crack growth analysis and test data.**

## Discussion and conclusion

The fatigue life of the specimens representing riveted joint of airplane skin made from aluminium alloy was analysed. The experimental evaluation of the life was performed on several load levels under load cycle asymmetry $R = 0.05$. In logarithmical axes plot the test data can be interpolated by linear regression curve.

The analysis of fatigue live was also performed by means of linear elastic fracture mechanics and crack growth model. The multi-site damage (MSD) was assumed so that the numerical model of the part representing the skin was created and the shape function for each crack was determined. The analysis of fatigue life was carried out using equivalent initial flaw size approach (EIFS). The simultaneous crack growth was assumed. Provided initial corner cracks of 0.125 mm in size, the analysis results show the fatigue lives close to the experimentally determined S-N curve of tested specimens. The considerable part of the fatigue life of the joint can be therefore attributed to the crack growth. The analysis results exhibit the smaller load level yields the longer crack initiation and also the higher scatter of the fatigue life.

The crack initiation process is the surface phenomenon and can be covered computationally very hardly without relevant material data [8]. The presented analysis shows that in involved joint configuration the simulated fatigue life determined by the crack growth concept can be used as the upper bound for the design purposes successfully.

## Acknowledgement

This work was funded by the Ministry of Industry and Trade of the Czech Republic in the framework of the Institutional Support of Research Organizations.

### References

[1] T. Mrňa et al. (2018) The fatigue strength of various mechanical double shear rivet joints of D16čA TV sheets, Lufinka A. et al., (eds.). *Experimental Stress Analysis 2018*, Book of extended abstracts, June 5th – 7th, 2018, Harrachov, Czech Republic: Czech Society for Mechanics, 286-291, ISBN 978-80-270-4061-2.

[2] Wronicz, W. and Kaniowski, J. (2014) The analysis of the influence of riveting parameters specified in selected riveting instructions on residual stresses, *Fatigue of Aircraft Structures* **1** 63- 71, doi: 10.1515/fas-2014-0005.
[3] Zheng, B., Yu, H., Lai, X. and Lin, Z. (2016) Analysis of Residual Stresses Induced by Riveting Process and Fatigue Life Prediction *Journal of Aircraft* **53**, No. 5, DOI: 10.2514/1.C033715
[4] Szymczyk, L. and Godzimirski, J. (2012) The influence of riveting process on sheets fatigue life - The stress state analysis, *Acta Mechanica et Automatica* **6** (1), 74-81.
[5] Ransa, C.D., Alderliestena, R.C. and Straznicky, P.V. (2009) Assessing the effects of riveting induced residual stresses on fatigue crack behaviour in lap joints by means of fractography, *International Journal of Fatigue* **31** (2), 300-308.
[6] Skorupa, M. et al. (2017) Fatigue life predictions for riveted lap joints, *International Journal of Fatigue* **94,** 41–57. doi: 10.1016/j.ijfatigue.2016.09.007.
[7] Skorupa, M. et al. (2015) Fatigue life prediction model for riveted lap joints, *Engineering Failure Analysis* **53,** 11-123.
[8] Schijve, J. (2010) Fatigue of structure and materials, 2nd edition, *Springer*, 621 pages, ISBN-13: 987-1-4020-6807-2.
[9] JSSG-2006 (1998) Joint service Specification Guide: Aircraft Structures, *Department of Defense of USA*, 148 pages.
[10] Zamani P. et al. (2015) Numerical Investigation on Optimizing Fatigue Life in a Lap Joint Structure, World Academy of Science, *Engineering and Technology International Journal of Civil and Environmental Engineering* **9** (5), 647-653.
[11] Ziegler, B. et al. (2011) Application of a strip-yield model to predict crack growth under variable-amplitude and spectrum loading – Part 2: Middle-crack-tension specimens, *Engineering Fracture Mechanics* **78**, 2609–2619.
[12] Schijve J. et al. (2004) Fatigue crack growth in the aluminum alloy D16 under constant and variable amplitude loading. *International Journal of Fatigue* **26**, 1–15.
[13] Harter, J. A. (2008) AFGROW users guide and technical manual, *Air Force Research Laboratory*, AFRL-VA-WP-TR-2008-XXXX.
[14] Běhal J and Nováková L. (2013) Stress state factor evaluation based on a fractographic analysis for use in the crack growth FASTRAN retardation model of the AFGROW computing code, *Engineering Failure Analysis* **35**, 645–651.
[15] Šedek, J., Běhal, J. and Siegl, J. (2015) Structure overloading evaluation based on the identification of subcritical crack increments, *Engineering Failure Analysis* **56**, 265-274.

# Weak Shock Reflection from Blunt Bodies with the High-Order Numerical Simulation Using an Immersed Boundary Method

***Hyun Jin Kwon[1], †Se-Myong Chang[2],**

[1]Department of Mechanics Engineering, Kunsan National University, Republic of Korea
[2]School of Mechanical Convergence Engineering , Kunsan National University, Republic of Korea.

*Presenting author: kint12@naver.com
†Corresponding author: smchang@kunsan.ac.kr

## Abstract

In this paper, we used an immersed boundary method to simulate the compressible flow field around a curvilinear body. The reflection of a moving shock wave is studied from a cylinder and other blunt bodies. WENO(weighted essentially non-oscillatory) scheme is a widely known high-order method for the simulation of compressible flow, but the implementation for an arbitrarily given boundary condition is hard to make using coordinate transformation in FDM(finite difference method) or polygonal control volume in FVM(finite volume method). Therefore, we have set a numerical code without skewness of grids by use of simple Cartesian coordinates system. At the boundary, the conservative condition with the slip velocity is guaranteed with a simple mathematical treatment called the immersed boundary condition. The reflected shock waves computed from blunt bodies are compared with experimental results.

**Keywords:** Euler equation, WENO scheme, Immersed boundary method

## Introduction

Immersed Boundary Method(IBM) can be used for the numerical computation using the structured meshes in Cartesian coordinate for a complex geometry. Peskin[1], proposing modifications of original IBM, analyzed the blood flow around the heart valves, and Chudhuri et al.[2], using WENO scheme in conjunction with IBM, simulated the interaction of high Mach number shock waves with wedge and blunt obstacles.

In this paper, we simulate the weak shock waves reflected from blunt bodies such as a cylinder with WENO scheme[3] linked with IBM.

## Governing equation

The Euler equation governing two-dimensional compressible inviscid flows can be written in conservative form as,

$$\frac{\partial Q}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}=0 \tag{1}$$

$$\begin{aligned} Q &= [\rho, \rho u, \rho v, E] \\ F &= [\rho u, \rho u^2 + p, \rho uv, u(E+p)] \\ G &= [\rho v, \rho uv, \rho v^2 + p, v(E+p)] \end{aligned} \tag{2}$$

where $Q$ is convection variables vector; $F, G$ are inviscid flux variables vectors; $\rho, u, v, p, E$ are density, $x$-direction velocity, $y$-direction velocity, pressure, and total energy, etc.

**Immersed Boundary Method(IBM)**

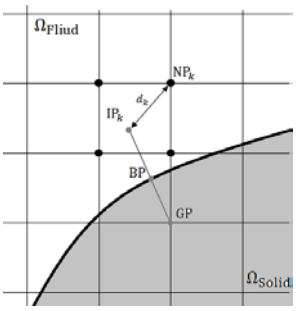



**Figure 1. Immersed boundary treatment**

Flow variables $\phi_{IP}$ in imaginary point can be evaluate as,

$$\phi_{IP} = \sum_{k=1}^{4} \delta_k \phi_{NP} \tag{3}$$

where $\delta_k$ is Inversed Distance Weight(IDW) of interpolation, and the flow variable $\phi_{BP}$ in the boundary point can be evaluate from Ref. [4]. In Eq. (3), the IDW is defined as

$$\delta_k = \eta_k \left( \sum_{k=1}^{4} \eta_k \right)^{-1}, \quad \eta_k = 1/d_k \tag{4}$$

where $d_k$ is the distance between the imaginary point($IP_k$) and a neighboring grid point($NP_k$), which are marked in Fig. 1.

To calculate the flow variable $\phi_{GP}$ in ghost point using a Dirichlet boundary condition.

$$\phi_{GP} = 2\phi_{BP} + \phi_{IP} \tag{5}$$

where the $\phi_{BP}$ means the variable at the boundary.

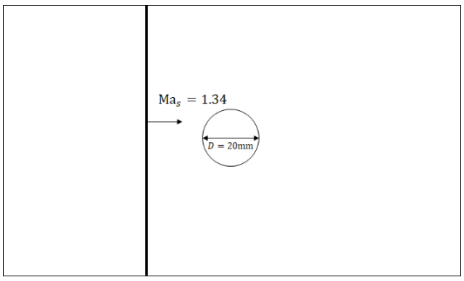



**Figure 2. Schematic diagram for the weak shock wave reflected blunt body.**

**Numerical simulation**

Fig. 2 shows the computational domain and the initial condition for the planar moving shock impinging into a cylinder, for example. The shock Mach number is set $Ma_s$ = 1.34, and the pressure before and after shock can be simply computed with Rankine-Hugoniot condition. The specific heat ratio of air is fixed to 1.4.

The density contours at each time step can be visualized in Fig. 3, and the fifth-order WENO scheme is used with third-order TVD(total variation diminishing) Runge-Katta time integration for the numerical result in the present figure. In every time step, holographic interferograms[5] are compared with isopycnics, or density contours obtained from computation. The number of nodes is 258,496 and number of elements is 257,336.

Mach reflection consisting of Mach triple points and Mach stems is reflected in front of the circular cylinder. A slip line is visible as the disturbed isolines, and the Mach stems intersect each other to form a secondary Mach reflection at the aft of the cylindrical body, which is called a 'shock-shock reflection'.

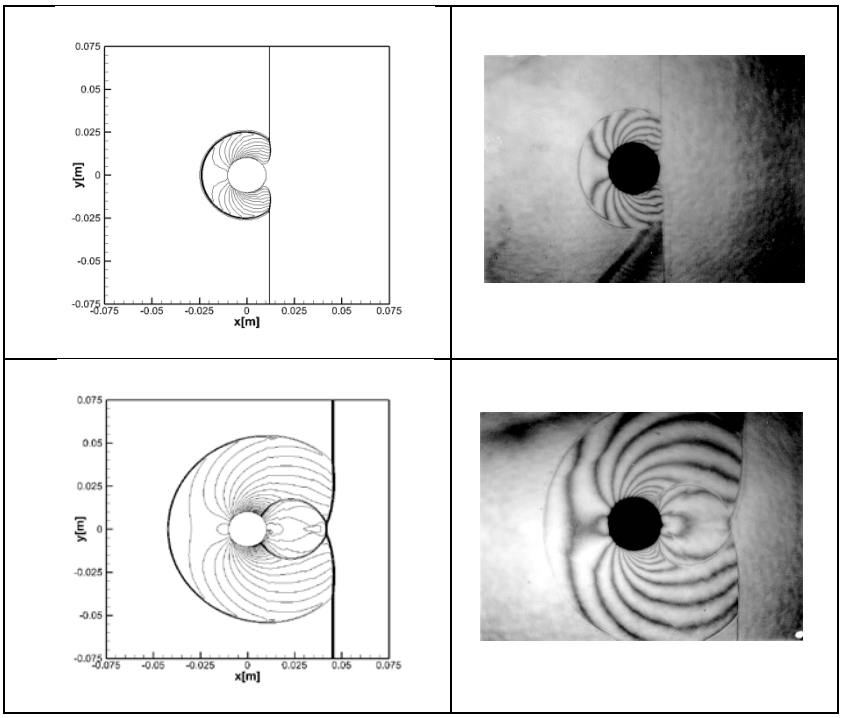



**Figure 3. Numerical isopycnics using WENO with IBM(left) and experimental interferogram(right): 48, and 120$\mu s$[5]**

**Summary**

In this paper, we have simulated two-dimensional compressible flow induced from a moving shock reflection from a circular cylinder. The incident Mach number is $Ma_s$ = 1.34, and the boundary treatment from the implementation of an immersed boundary method is successfully applied for the high-order WENO method. The numerical results coincide with those of experiment, and describe delicate wave physics in detail for every time step. This method is also valid for other geometries if we just change the shape of blunt bodies.

**Acknowledgement**

This work is financially supported by NRF(National Research Foundation 2016R1D1A3-B01015543) and KETEP(Korea Institute of Energy Technology Evaluation and Planning, No. 20174010201350).

**References**

[1] Peskin, C.S., (1972) Flow Patterns Around Heart Valves: A Numerical Method, *Journal of Computational Physics*, **10**, 252-271.
[2] Chaudhuri, A., Hadjadj, A., and Chinnayya, (2011) On The Use of Immersed Boundary Methods for Shock/Obstacle Interactions, *Journal of Computational Physics*, **230**, 1731-1748.
[3] Jiang, G.S. and Shu, C.W. (1996) Efficient Implementation of Weighted ENO Schemes, *Journal of Computational Physics*, **126**, 202-228.
[4] Dadone, A., (1998) Symmetry Techniques for the Numerical Solution of the 2D Euler Equations at Impermeable Boundaries, *International Journal for Numerical Methods in Fluids*, **28**, 1093-1108.
[5] Chang, S.M. and Chang, K.S. (1999) Weak Shock Waves Reflected from a Blunt Body, *Transactions of KSME*, **23**, 901-910.

# Evaluation of a Semi-automated Photogrammetry-based Method for Geometry Creation for Urban Simulations

***†Chinchun Ooi[1], Raymond Quek[1], Zhengwei Ge[1], Hee Joo Poh[1], and George Xu[1]**

[1]Department of Fluid Dynamics, Institute of High Performance Computing, Singapore.

*Presenting author: ooicc@ihpc.a-star.edu.sg

†Corresponding author: ooicc@ihpc.a-star.edu.sg

## Abstract

With the increasing accessibility of computing power in recent years, computational fluid dynamics (CFD) has become a vital and routine part of the building design process. However, these simulations require explicit modeling of major structures in the vicinity of the building of interest, as the orientation and placement of these structures can have significant impact on the local wind flow patterns around and within the building of interest. However, the manual generation of these geometries can be extremely tedious. Thus, we propose a semi-automated photogrammetry-based approach to regenerate simplified building geometries for urban simulations. We also examine the actual discrepancy in heights between the photogrammetry-generated buildings and the actual reported building heights from an online database for a sample location in Singapore, and report a mean percentage error of less than 10% under this approach, thus suggesting the applicability of this method to a wide range of urban simulations.

**Keywords:** Computational Fluid Dynamics, Urban Simulations, Photogrammetry, Automated Geometry Regeneration

## Introduction

With the recent advances in availability of computing power, computational fluid dynamics (CFD) has become a vital and routine part of the urban planning and building design process. Singapore's Building and Construction Authority, for example, now has established guidelines for conducting CFD simulations in their Green Mark methodology to aid in performance-based urban planning and building design [1].

Even while these urban simulations become increasingly common, scientific literature now recommends explicit modeling of major structures in the vicinity of the building of interest in the urban built environment, as these local structures can have significant impact on the local wind flow patterns around the building of interest [2, 3]. Assuming a typical building length and spacing of approximately 10 x 10 $m^2$ to 100 x 100 $m^2$, there can be on the order of 100 buildings around the building of interest that needs to be explicitly modeled. While one option is to purchase the explicit building footprints and geometries directly from relevant governmental agencies, this may not always be an option due to a variety of issues such as cost, especially for academics, or lack of documentation, especially for older structures and buildings with no appropriately digitized records. A common solution is thus to capitalize on open-source resources such as Google Maps or Open Street Maps to manually obtain the building footprint, and re-generate the buildings in a computer-assisted design (CAD) program by hand. This is

laborious, and often also requires the CFD practitioner to resort to other methods to obtain the building heights, such as the enumeration of building floors through on-site visits.

Conversely, photogrammetry is a long developed technique in other applications that has been previously described in literature as an effective means of obtaining 3D-model information for various urban simulations [4, 5]. While image acquisition in the past might have been difficult, the increasingly widespread availability of drone photography now makes photogrammetry for geometry regeneration a feasible solution. In addition, while photogrammetry can produce extremely high fidelity geometries, the sheer scale of CFD urban simulations often necessitate simplifications in the geometry, hence necessitating an additional processing step post-photogrammetry.

In this work, we thus propose a semi-automated photogrammetry-based approach to regenerate building geometries for urban simulations. While this method cannot regenerate the actual building of interest in adequate detail for modeling, it can help expedite the generation of surrounding buildings for explicit modeling which often do not need to be represented in as fine detail. We describe the workflow in the following section, along with a comparison of the actual discrepancy in heights between the photogrammetry-generated buildings and the actual reported building heights in an online database. This is anticipated to provide a quantitative measure for interested urban simulation practitioners to decide if photogrammetry is indeed suitable for further routine use.

## Methods and Results

### *Semi-Automated Photogrammetry Workflow*

There are 3 components to the current workflow, with occasional human input required. Two sources of images are required, with the first being 3-dimensional aerial images of the area of interest, such as can be acquired by aerial drone photography, and the second being 2-dimensional images or schematics of the building footprints, similar to the maps available from open source resources such as Open Street Maps. The combination of these images to create an actual water-tight CAD geometry for subsequent urban simulations is described in greater detail in the following subsections. Additionally, the scripts used for the automatic acquisition and conversion of acquired images to the actual CAD geometries are available for referencing at https://github.com/ooichinchun/Maps2Geometry.

### *Aerial Image Acquisition*

Images acquired in this work are 3-dimensional screen grabs from Google Maps instead of actual drone photography images. However, it is anticipated that aerial drone photography would be able to obtain images of a similar quality and type. The set of images acquired should span a complete rotation around the area of interest, and should ideally comprise a set of images at different azimuths as well. A representative example of a potential image that can be used is presented in Figure 1.


**Figure 1. Example aerial photograph of the area of interest to be modeled.**

*Building Footprint Modeling*

Images of the simplified building footprints can be acquired from sources such as Open Street Maps as .svg vector files. From these files, it is possible to obtain the vertices and lines that make up individual building polygons. In addition, the convex hull algorithm can be used to simplify the geometry, as illustrated in Figure 2 [6]. The convex hull algorithm essentially seals off small gaps and undulations within the buildings, which are often not meshed in actual CFD urban simulations due to their comparatively small length scales. For example, the two buildings in the top right corner of Figure 2 with the undulating edges are simplified by the convex hull algorithm into relatively simpler polygons. The resulting set of vertices and edges can then be written out into a CAD file for import into any CAD program.

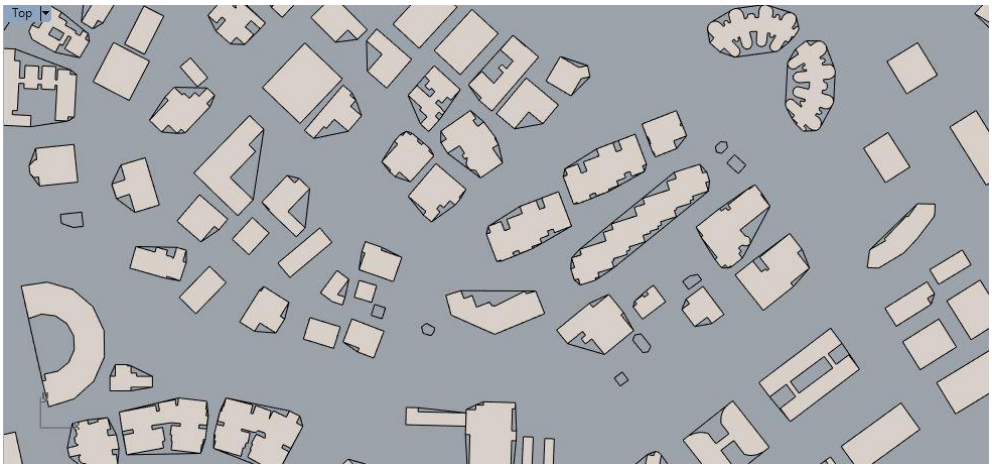



**Figure 2. Example of the geometries regenerated from Open Street Maps (objects with the solid brown fill) and the convex hull geometries generated (black lines).**

*Geometry Regeneration*

The aerial images acquired in the prior step are then used to re-generate the geometry via photogrammetry. Agisoft Photoscan is used in this work to convert the set of images into an actual 3-dimensional point cloud. Rescaling and re-alignment to a North-South orientation are required at this stage to ensure consistency with the geometry file generated from Open Street Maps. The individual mesh points generated by photogrammetry are then filtered by their Cartesian coordinates for the building locations as defined by the geometry file such that the respective building heights can be obtained and the individual building footprints can be extruded accordingly. Representative output images for this process are displayed in Figure 3.

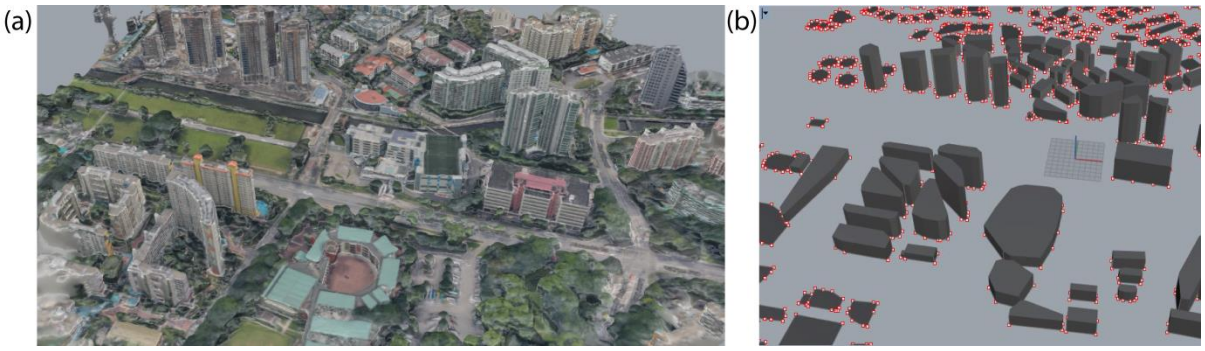



**Figure 3. (a) Sample output of the mesh from photogrammetry for the aerial images obtained (b) Corresponding output of the scripts for the simplified extruded buildings around the region of interest.**

*Quantification of Error in Geometry Regeneration*

The generated geometries are then analyzed for discrepancies to assess the accuracy of this particular method. Heights as obtained from the photogrammetry point cloud are compared to heights obtained from an online reference (www.emporis.com), and the results are plotted in Figure 4.

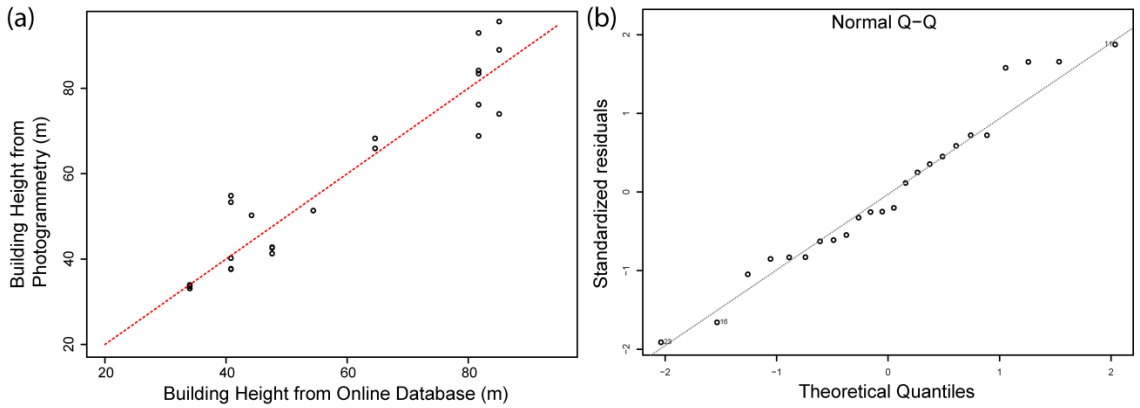



**Figure 4. (a) Plot of the building heights as obtained from an online database against the building heights obtained by the photogrammetry-based method described in this work. (b) A normal Q-Q plot for the standardized residuals for the regression line from (a).**

The results indicate that there is a very good match between the building heights as reported by the online reference and as produced from photogrammetry. The gradient of the line in Figure 4a is 0.97, while the $R^2$ of the regression is 0.88. The Q-Q normal plot of the scaled residuals as per Figure 4b also indicates that there is no systemic error in the photogrammetry, with the regression exhibiting normally-distributed errors. More critically, we determine the mean and median absolute percentage error in heights to be 9.7% and 7.7% respectively, which can be a helpful measure for the CFD practitioner to determine if this method is appropriate for their purposes.

## Conclusions

In this work, we demonstrate the application of photogrammetry to rapidly regenerate simplified building geometries for urban CFD simulations. More critically, we show that the discrepancy in heights obtained via this method are less than 10%, which is probably much less than the typical uncertainty in other parameters such as appropriate inlet wind velocity for urban simulations. We anticipate that this method would be of increasing interest to CFD practitioners, as the necessity for accurate representation of surrounding buildings in CFD simulations gradually becomes more evident to industry.

## References

1. Authority BaC. BCA GREEN MARK : Technical Guide and Requirements. 2015.
2. Tominaga Y, Mochida A, Yoshie R, Kataoka H, Nozu T, Yoshikawa M, et al. AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings. Journal of wind engineering and industrial aerodynamics. 2008;96(10-11):1749-61.
3. Blocken B. Computational Fluid Dynamics for urban physics: Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations. Building and Environment. 2015;91:219-45.
4. Hu J, You S, Neumann U. Approaches to large-scale urban modeling. IEEE Computer Graphics and Applications. 2003;23(6):62-9.
5. Shiode N. 3D urban models: Recent developments in the digital modelling of urban environments in three-dimensions. GeoJournal. 2000;52(3):263-9.
6. Barber CB, Dobkin DP, Dobkin DP, Huhdanpaa H. The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS). 1996;22(4):469-83.

# ONE DIMENSIONAL FINITE ELEMENT MODELLING ON AN FRP-TO-CONCRETE BONDED JOINT ANCHORED WITH INCLINED U-JACKETS

***Chuying Cui[1], Li Lijuan[1], †Fu Bing[1], Liu Feng[1]**

[1]School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou 510006, China

*Presenting author: cuichuying@hotmail.com

†Corresponding author: cefubing@gdut.edu.cn

## Abstract

Debonding failure in the form of either IC debonding or concrete cover separation commonly controls the load carrying capacity of RC beams flexurally strengthened with an externally bonded FRP plate, leading to a very low utilization rate (e.g., 20%-40%) of the FRP strength. A number of experimental studies carried by the authors' group and some other researchers indicate inclined FRP U-jackets as the end anchorage show to be highly effective in delaying or suppressing the debonding failures, resulting in significant enhancements in the structural performance of FRP-plated RC beams and utilization rate of the FRP strength. This paper presents a one-dimensional finite element model for FRP-to-concrete bonded joint anchored with inclined FRP U-jackets, in which interfacial cohesive elements are used to represent the force-slip behavior of the anchored joints. Comparisons between the FE predictions and test results have been made to demonstrate the accuracy of the proposed FE model. On the basis of the FE predictions, expressions of the force-slip model are proposed for the inclined FRP U-jackets, which considers the effect of vital parameters (e.g., inclination angle, width and thickness of U-jackets), and can be directly used in an FE model for the FRP-plated RC beams with inclined U-jackets.

**Keywords**: Fiber reinforced polymer (FRP), concrete, debonding, fiber anchor, finite element modelling

## 1. Introduction

Externally bonding of fiber reinforced polymer (FRP) plates has become a widely-used technique for the strengthening of reinforced concrete (RC) beams due to superior properties of the FRP materials, such as high strength-to-weight ratio and excellent corrosion resistance[1].Debonding failure in the form of either IC debonding (intermediate crack debonding) or concrete cover separation commonly controls the load carrying capacity of RC beams flexurally strengthened with an externally bonded FRP plate, leading to a very low utilization rate (e.g., 20%-40%) of the FRP strength[2].A number of experimental studies carried by the authors' group and some other researchers indicate that inclined FRP U-jackets

as the end anchorage show to be highly effective in delaying or suppressing the debonding failures, resulting in significant enhancements in the structural performance of FRP-plated RC beams and utilization rate of the FRP strength. FRP U-jacketing at 45° to the beam axis at end anchorage is much more effective in delaying or suppressing the above two failure modes than vertical U-jacketing[3][4]. FRP U-jackets of different parameters have significantly various effects on anchorage, however, the force-slip behavior between the FRP U-jackets and concrete interface remains unclear. Bond testing on of FRP-to-concrete bonded joints anchored with inclined U-jackets is an attractive test approach to investigate the anchoring mechanism of FRP U-jackets in FRP-plated RC beams. Experimental studies including single shear tests, double shear tests and modified beam tests and theoretical studies by means of fracture mechanics analysis and finite element analysis are two vital measures for investigating interfacial debonding behavior of FRP-to-concrete bonded joints [5]-[9].

The authors' group has conducted several series of NES single-shear pull tests for a CFRP plate-to-concrete bonded joint anchored with inclined U-jackets, which considers the effect of vital parameters (e.g., inclination angle, width and thickness of U-jackets). In this paper, a one-dimensional finite element model for FRP-to-concrete bonded joint anchored with inclined FRP U-jackets is proposed, in which interfacial cohesive elements are used to represent the force-slip behavior of the anchored joints. Comparisons between the FE predictions and test results have been made to demonstrate the accuracy of the proposed FE model. On the basis of the FE predictions, expressions of the force-slip model are proposed for the inclined FRP U-jackets, and can be directly used in an FE model for the FRP-plated RC beams with inclined U-jackets.

## 2. Near-end supported single-shear pull test

The author has conducted a series of NES single-shear pull tests, taking the inclination angle, width and thickness of U-jackets as experimental variables, to study debonding behavior between FRP and concrete anchored with FRP U-jackets of different forms. The test program consisted of 10 concrete specimens. All used 220×300mm×700mm concrete prisms. In this test program, the cylinder compressive strength of concrete $f_c^{`}$ was 34MPa. All CFRP plates had the same dimensions: 50mm in width and 1.2mm in thickness. The bond length of CFRP plates was 400mm, which was longer than the effective bond length. And the nominal thickness of single ply of CFRP plate used for U-jackets is 0.167mm. Details of specimens dimensions of U-jackets in different forms can be found in Fig. 1 and Table 1.

Figure 1. Test rig.

Table 1. Details of specimens and test results

| Specimen | U-jackets | | | | Rounder radius (mm) |
|---|---|---|---|---|---|
| | Width $b_U$ (mm) | Plies | Bond length $L_U$ (mm) | angle (°) | |
| CS1 | — | — | — | — | — |
| I30P2W40L250C25 | 40 | 2 | 250 | 30 | 25 |
| I45P2W40L250C25 | 40 | 2 | 250 | 45 | 25 |
| I60P2W40L250C25 | 40 | 2 | 250 | 60 | 25 |
| I90P2W40L250C25 | 40 | 2 | 250 | 90 | 25 |
| I135P2W40L250C25 | 40 | 2 | 250 | 135 | 25 |
| I45P2W50L250C25 | 50 | 2 | 250 | 45 | 25 |
| I45P2W60L250C25 | 60 | 2 | 250 | 45 | 25 |
| I45P1W40L250C25 | 40 | 1 | 250 | 45 | 25 |
| I45P3W40L250C25 | 40 | 3 | 250 | 45 | 25 |

## 3. One-dimensional finite model

### *3.1. General*

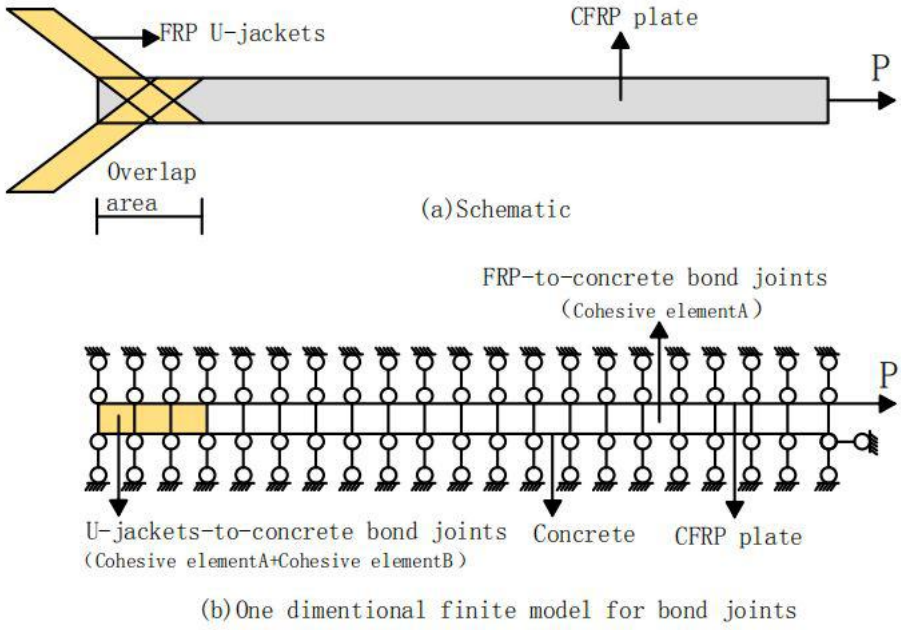



Figure 2. Pull test: (a) Schematic; (b) FE model

The one-dimensional interface finite element model in Fig. 2 contains two kinds of interfacial elements COH2D4, referred to hereafter as CA and CB, adopting two different constitutive laws to represent the debonding behavior of two different interfaces. In the CFRP-to-concrete bonded joint, the properties of the interfacial elements CA are defined by using the bi-linear bond-slip model of Lu et al. [9], while the proposed elastic-brittle bond-slip model accounts for the bond behavior of U-jacket-to-concrete using the interfacial elements CB. By employing the interfacial elements CB along the overlap area, as shown in Fig. 2, the bond bond strength among the overlap area can be significantly enhanced to reflect the anchoring effect of FRP U-jackets. The debonding failure of bonded joints depends only on the

bond-slip behavior parallel to the interface, and the vertical displacement of top and bottom surfaces of interfacial elements are constrained as a result. So that top and bottom surfaces can be assumed to be the CFRP plate and the concrete prism, respectively.

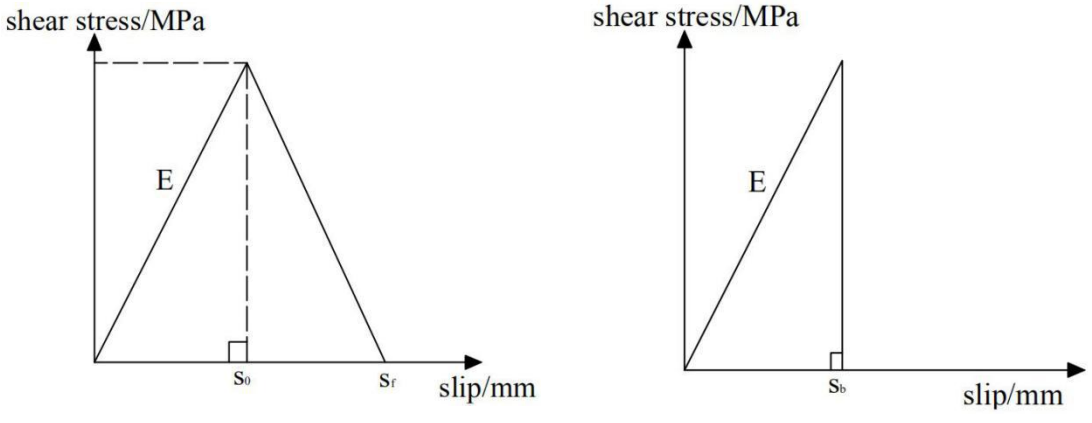
The bonded joint model used in this paper has the following dimensions: the length of the bonded joint is 400mm, which is equal to the bond length of CFRP plates, and the out-of-plane thickness is 50mm, which is equal to the width of CFRP plates. The element size of 1mm is used for the interfacial element, and the shear strain of the element equals to the shear slip hence. After constraining the horizontal displacement of the bottom surface of interfacial elements at the loaded end, the interface slip can be accessed directly by the horizontal displacement of the top surface of interfacial elements at the loaded end.

*3.2. Proposed bond-slip model*

The proposed bi-linear bond-slip model is shown in Fig. 3(b), which features a linear ascending branch followed by the rapid failure of the interface element CB. According to the above model, the bond shear stress increases linearly with the interface slip until it reaches the peak value at which the interfacial slip corresponding to the applied load is defined as $s_b$ hereafter. When the interfacial slip is greater than $s_b$, the bond stress reduces to zero immediately, indicating the failure of a local interface element. And the description of the local bond-slip relationship is proposed by following equations:

$$\tau = \begin{cases} Es & if \quad s \le s_b \\ 0 & if \quad s > s_b \end{cases} \tag{1}$$

where $E$ is the slope of the ascending branch. Based on the above discussions, the local bond-slip model can be precisely denoted, as long as the value of the key parameters, including the slope of the ascending branch $E$ and the ultimate slip $s_0$, are determined.




(a) (b)

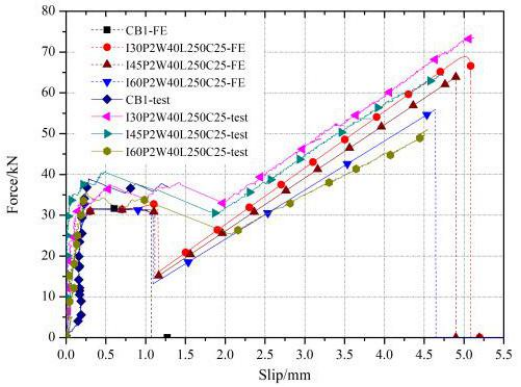
Fig 3. Bond slip model: (a) Lu's bi-linear model; (b) Proposed elastic-brittle model

## 4. Analysis results and discussions

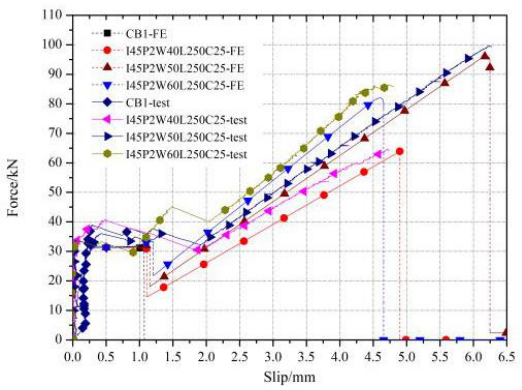
Table 2. Key parameters of proposed bond slip model

| Specimen | Slope E(MPa) | Ultimate slip $s_0$(mm) |
|---|---|---|
| I30P2W40L250C25 | 3.0 | 3.2 |
| I45P2W40L250C25 | 4.3 | 3.0 |
| I60P2W40L250C25 | 5.6 | 2.8 |
| I45P2W50L250C25 | 5.2 | 3.4 |
| I45P2W60L250C25 | 6.1 | 2.3 |
| I45P1W40L250C25 | —— | —— |
| I45P3W40L250C25 | 6.4 | 2.2 |

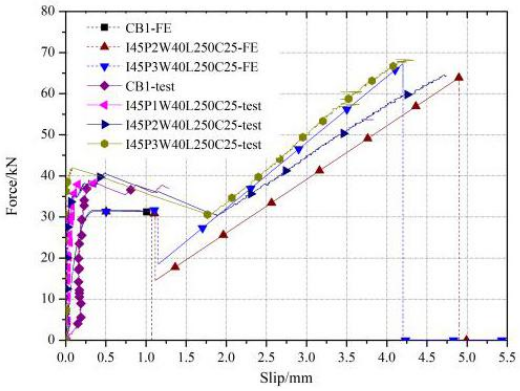
Through multiple calculations, the values of slope $E$ and the ultimate slip $s_b$ in the proposed bond-slip model adopted by interfacial elements CB, which are employed among the overlap area, are shown in Table 2. Comparisons between the FE predictions and test results have been made to demonstrate the accuracy of the proposed FE model in Fig. 4. It can be seen that the proposed bond-slip model gives results in close agreement with the test results, showing the same development trend of load-slip curves, excluding the stage 4 of load-slip curves. It can be found that a sudden drop of the load-carrying capacity occurs in the stage 4 of the load-slip curve when it reaches a interface slip $s_1$ at a value around 1.1~1.2mm, followed by the load's linear increasing with the interfacial slip until the failure of the interface model. Nevertheless, the experimental curve only showed a slight downward trend in stage 4, revealing that there is only a mild decrease in the load-carrying capacity in stage 4. When the interface slip reaches a value around 2.0 mm, where the complete debonding of the CFRP-to-concrete bonded joint emerges, the load-carrying capacity increases linearly at a constant rate in stage 5 and converges with the FE prediction until the failure of the specimen. This inevitable difference appears in stage 4 of the force-slip response, on account of different loading schemes between the FE analysis and the experiment program. The force-control loading scheme is adopted by the test program, while the displacement-control loading scheme is used for the FE model, which makes it possible to capture rapid drop of the load-carrying capacity.




(a)




(b)




(c)

Fig 4.FE versus test load-slip curves: (a) Inclination angle of U-jackets; (b) Width of U-jackets; (c) Thickness of U-jackets

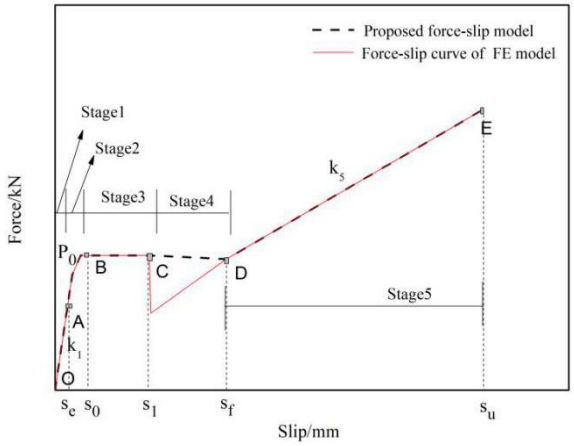



Figure 5. Proposed force-slip model and force-slip curve of FE model

$P_0$ is the peak load of the debonding duration of CFRP plates. $s_0$ is the slip corresponding to the test load reaches $P_0$; $s_f$ is the slip of complete debonding of CFRP plates; $P_u$ is the ultimate test load; $s_u$ is the ultimate slip.

For the stage 4 of load-slip curves from FE predictions, the following simplification can be made to get a better fitting with experimental results. Make the extension line for the segment BC, intersecting with stage 5 at point D, where the corresponding interface slip can be defined as $s_f$. Accordingly, a load-slip model of the FRP-to-concrete bonded joint anchored with the inclined U-jackets is proposed through the finite element analysis by replacing the fold line in the stage 4 of the FE model with the linear segment CD, as the dotted line shown in Fig. 5. Based on the observations, the load-slip response can be closely predicted by the proposed bond-slip model.

## 5. Conclusions

This paper presents a one-dimensional finite element model for FRP-to-concrete bonded joint anchored with inclined FRP U-jackets, in which interfacial cohesive elements adopting the proposed bond-slip model are used to represent the force-slip behavior of anchored joints. Comparisons between the FE predictions and test results have shown that the FE model can

provide close predictions of the load-slip response of the bonded joint, demonstrating the accuracy of the proposed FE model with the bi-linear bond-slip model. The shape and several key parameters of the force-slip model of the FRP-to-concrete with the end anchorage can be determined through the proposed FE model.

## 6. References

[1] Smith, S. T., & Teng, J. G. (2002). FRP-strengthened RC beams. Ⅰ: review of debonding strength models. Engineering Structures, 24(4), 385-395.
[2] Kalfat, R. , Almahaidi, R. , & Smith, S. T. . (2013). Anchorage devices used to improve the performance of concrete structures retrofitted with FRP composites: a state-of-the-art review. Journal of Composites for Construction, 17(1), 14-33.
[3] Fu, B. , Chen, G. M. , & Teng, J. G. . (2017). Mitigation of intermediate crack debonding in FRP-plated RC beams using FRP U-jackets. Composite Structures, (176), 883-897.
[4] Fu, B., Tang, X. T., Li, L. J., Liu, F., & Lin, G. (2018). Inclined FRP U-jackets for enhancing structural performance of FRP-plated RC beams suffering from IC debonding. Composite Structures, 200, 36-46.
[5] Chen, J. F. , & Teng, J. G. . (2001). Anchorage strength models for FRP and steel plates bonded to concrete. Journal of Structural Engineering, 127(7), 784-791.
[6] Yao, J. , Teng, J. G. , & Chen, J. F. (2005). Experimental study on FRP-to-concrete bonded joints. Composites Part B Engineering, 36(2), 99-113.
[7] Yuan, H. , Teng, J. G. , Seracino, R. , Wu, Z. S. , & Yao, J. (2004). Full-range behavior of FRP-to-concrete bonded joints. Engineering Structures, 26(5), 553-565.
[8] Lu, X. Z. , Ye, L. P. , Teng, J. G. , & Jiang, J. J. (2005). Meso-scale finite element model for FRP sheets/plates bonded to concrete. Engineering Structures, 27(4), 564-575.
[9] Lu, X. Z. , Teng, J. G. , Ye, L. P. , & Jiang, J. J. (2005). Bond-slip models for FRP sheets/plates bonded to concrete. Engineering Structures, 27(6), 920-937.

# 3D numerical study of wet foam behavior and channels by MPS method

**†Zhongguo Sun, * Yong Zhang, Qixin Liu, Yijie Sun and Guang Xi**

School of Energy and Power Engineering, Xi'an Jiaotong University, China

*Presenting author: 2738700649@qq.com

†Corresponding author: sun.zg@xjtu.edu.cn

## Abstract

Wet foam consists of many bubbles. Its geometry and mechanical property are complex. The contact between bubbles generates structure like porous media with Plateau channels, and the surface tension force acts on both sides of the liquid film. Seldom model or method can simulate the phenomenon of liquid foam in detail. In this study, the moving particle semi-implicit (MPS) method is employed to study wet foam behavior and channels. The Lagrangrian description and particles with flow information are applied, which has inherent advantages in dealing with large deformation of interface. A double gas-liquid interface surface tension model is studied based on the surface free energy model, and the interface tension was introduced to study the interaction between bubbles and solid wall. Further, the behavior of three-dimensional multiple bubbles is investigated in detail. The typical behavior of bubbles are analyzed. The effects of liquid viscosity, surface tension coefficient and bubble size on the behavior of bubbles are discussed. Subsequently, the forming process of different kinds of Plateau channel structure were displayed and discussed in detail.

**Keywords:** MPS method; double gas-liquid interface surface tension; interface tension; Plateau channels

## 1 Introduction

Foams includes two phases of gas and liquid, along with the surfactant molecules in the liquid which reduce the surface tension of the liquid at the liquid-gas interface to facilitate foam formation. As shown in Figure 1, the liquid volume fraction,Ø may vary from less than 1% (dry foam) to around 36% (wet foam) under gravity effect [1]**.** Typically, foam channels are divided into internal channels and external channels. The internal channels can be considered as flow through a collection of long channels that are called Plateau borders (PB) and junctions of four PBs that is called node. The external channels' fundamental difference with internal channels is the existence of a no-slip wall[2]. However, foams are often inside of containers and has one side in contact with air, so it is necessary to classify the foam channels more comprehensively. In this study, the foam channels that come into contact with air are called the open channels, and the wet foam channels are discussed and studied emphatically.

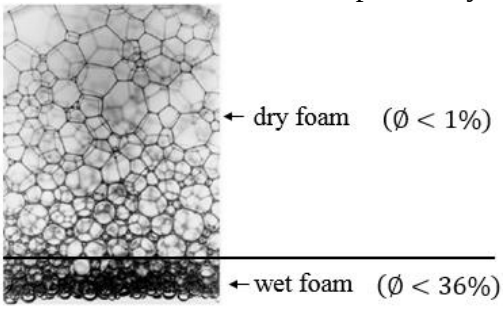



**Figure 1. Foam gas-liquid two-phase distribution**

Foams have numerous applications in the industries, agriculture and food. Industrially, their ability to preferentially gathering desired materials has made them greatly useful in the mineral industry for collecting desirable elements, and they can also be used for crop spraying in agriculture. These successful applications of foam rest on the capability of water-based foam to distribute a small volume of liquid over a wide area[2][3][4]. Therefore, one of the major question about the dynamical behavior of foams is the forming process of liquid channel structure in foams. Understanding the factors that influence bubbles' behavior such as surfactant materials and the forming process of different kinds of channels structure in foam is crucial in increasing its efficiency for current applications as well as creating potential new applications for foams. To better understand the factors that influence bubbles' behavior and the formation mechanism of various foam channel structures, requires a more complete description and study of the flow behavior and different kinds of foam channel structure respectively.

Along with the extensive application of foams in many aspects, great progress has been made in making the science of foams a sound including theories models, numerical simulation methods and experiments respectively. Hot topics have included static structure (the shapes and arrangements of bubbles), and dynamic evolution such as coarsening due to diffusion of gas, capillary and gravity driven drainage, rheology and coalescence due to rupture of the films [1][5][6]**.** Koehler et al. expanded the basic microscopic model by a theoretical study and experimental investigations at the scale of a single Plateau border for the flow rates, obtaining experimental results of velocity profiles inside the both interior and exterior PBs[2][7]. D. G. T. Barrett et al. employed experiments to help us to understand the dynamics of foam instability, and static (or quasi-static) simulations using the Surface Evolver to establish equilibrium film shapes in different frame shapes and sizes[8]. Besides, in gravity environment, the hydrodynamic process will result in uneven distribution of liquid film thickness from top to bottom. At the same time, the rapid liquid loss also makes it difficult for the bubbles to evenly and stably distribute in the foam. However, the buoyancy disappears and gravity drainage is suppressed under low microgravity, capillary drainage is slowed down due to the quasi-spherical shape of the bubbles, so the overall behavior of foam will show completely different characteristics from that under normal gravity. The microgravity environment of 20~25 seconds (0.01G) was realized by parabolic aircraft flight method, and then the liquid evolution process of two-dimensional foam was observed and analyzed, but these experiments are relatively isolated and their results are not well analyzed[9].

There have been many numerical method of the foam in recent years. Among the approaches, VOF[10] and level set method[11] are applied extensively. In fact, the VOF method tracks the volume fraction of each phase or component instead of tracking the interface itself. The interface is reconstructed from the values of volume fraction which imply that it would be complicated for VOF to be extended to the three-dimensional case. For the level set method, it uses the level set function to store the information of the interface. The advantage is that the level set function varies smoothly across the interface while the volume fraction is discontinuous. Besides, the curvature can be easily evaluated from the level set function. However, the level set method requires a re-initialization procedure to keep the distance property when large topological changes occur around the interface. This may violate the mass conservation for each phase or component. The LBM method [12] was also applied to study the deformation of a single bubble or several bubbles in the liquid, but the surface tension force acts only on the single-gas-liquid-interface in the liquid.

Most studies focused on the bubble's shape and deformations in the liquid or the tube where the surface tension force acts only on the single-gas-liquid-interface in the liquid. And most of

the mentioned studies' numerical models are two-dimensional simulation which are suitable for the specific individual parts of foam, whereas a three-dimensional model can overcome the shortcomings of two-dimensional simulation by analyzing the overall structure of foam. Besides the study of wet foams is essential to understand and control foaming processes. However, foams are created in a transient wet state and evolve rapidly afterwards under normal gravity. Indeed, a micro- or zero-gravity study of wet foam hydrodynamics removes the various instabilities experienced under normal gravity, but these experimental methods are difficult to achieve, or the weightlessness time is too short, or the preparation cycle is too long, resulting in scattered experimental research results in this field. All these questions have motivated the use of numerical methods to carry out wet foam investigations without considering the gravity effect.

To model the behavior of a single film bubbles with double-gas-liquid-interfaces in the absence of gravity, two main issues including surface tension force modeling and interface recording have to be considered. The moving particle semi-implicit (MPS)[13] is employed in this study because of the advantage in dealing with the large deformation. So far, the MPS method was successfully used in engineering and science, such as the deforming process of bubbles in the air[14], and the multiphase flows with deformable interfaces movement[15]. Figure 2 is the schematic diagram of the framework of this study. The aim of this study is to present a three-dimensional model for the foam system including interior node-PB and exterior node-PB as well as open node-PB without considering the effect of gravity. Firstly, the factors of the effect on the dynamic behavior of bubbles including surface tension coefficient and fluid viscosity were studied respectively to present a comprehensive comparison on foam behavior in the presence of various physical conditions and bubble sizes. Subsequently, the forming process of different kinds of foam channel structure were displayed and discussed. Once the forming process of different kinds of foam channels structure including exterior and interior as well as open channels were studied, a complete geometrical model for the foam can be used in foam stability study in different environments. This study is anticipated to broaden the recognition of the foam behavior and the formation mechanism of various foam channel structures without considering gravity.

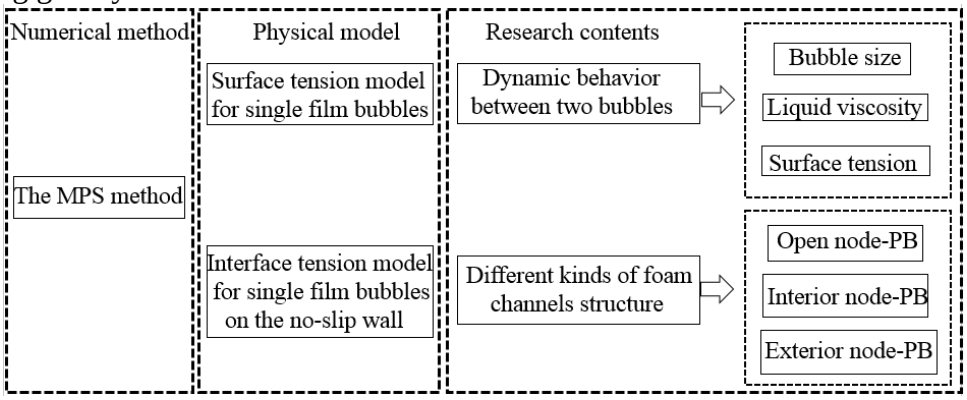



**Figure 2. Schematic diagram of the framework**

## 2 Numerical method

### 2.1 The MPS method

In the MPS method, the governing equations of the flow mechanics include the conservation equations of the mass and the momentum. For an incompressible flow, they can be written in a Lagrangian form as:

$$\frac{D\rho}{Dt}+\rho\nabla\cdot\boldsymbol{u}=0 \tag{1}$$

$$\rho \frac{\mathrm{D}\boldsymbol{u}}{\mathrm{Dt}} = -\nabla P + \mu \nabla^2 \boldsymbol{u} + \rho \boldsymbol{f} \tag{2}$$

where $\rho$ is density, t is time, $\boldsymbol{u}$ is velocity, P is pressure, $\mu$ is dynamic viscosity coefficient, $\boldsymbol{g}$ is acceleration of gravity, and $\boldsymbol{f}$ is the volumetric force, such as gravity and surface tension force, $\nabla$ is gradient, $\nabla^2$ is Laplacian. The calculation area and its boundary are discretized by a set of particles, which are divided into three kinds including fluid particles, gas particles and solid particle with different physical parameters in this study. The interaction between adjacent particles is defined by a kernel (weight) function:

$$w(r) = \begin{cases} \frac{r_e}{r} - 1 & (r \le r_e) \\ 0 & (r > r_e) \end{cases} \tag{3}$$

Where $r$ is the distance between the two particles. This effective radius $r_e$ is $2.1l_0$ , and $l_0$ is the initial distance between two neighboring particles. It is obvious that the kernel function is only valid within the effective radius. Similar to the physical density of the liquid, a particle density is defined as:

$$n_i = \sum_{j \ne i} w(r) \tag{4}$$

It is a constant value $n_0$ for an incompressible fluid but a little smaller on the free surface. Any movement of the particles would change the particle density to a temporal $n_i^*$. In order to modify it back to $n_0$, the pressure is calculated by solving the Poisson equation with Incomplete Cholesky Conjugate Gradient (ICCG) method:

$$\langle \nabla^2 P^{n+1} \rangle_i = \frac{\rho}{\Delta t^2} \frac{\langle n_i^* \rangle - n_0}{n_0} \tag{5}$$

By treating the foam system as a multi-density multi-viscosity fluid, the mesh-free particle method for incompressible multiphase flow has been introduced in this study based on the MPS. At the interface, a particle with viscosity $\mu_1$ (or density $\rho_1$ ) may interact with particles with viscosity $\mu_2$ (or density $\rho_2$ ). The multi-viscosity and multi-density models are derived from the interaction between particles with different properties.In this study, the harmonic mean viscosity model is used to calculate the viscous momentum transfer process.

$$\frac{1}{\rho} \nabla \cdot (\mu \nabla \varphi) = \frac{2d}{\rho_i \lambda n_0} \sum_{j \ne i} \left[ \frac{2\mu_i \mu_j}{\mu_i + \mu_j} (\varphi_j - \varphi_i) w(|\boldsymbol{r}_j - \boldsymbol{r}_i|) \right] \tag{6}$$

where $\varphi$ is an arbitrary scalar, d=3 indicates a three-dimensional problem. Similar to the harmonic mean viscosity model, the harmonic average method is applied to the average density model so that the change of pressure field in the transition region was relatively gentle. In this way, the pressure field distribution in the transition region is conducive to ensuring the stability of the simulation. Poisson equation of pressure can be written as:

$$\nabla \cdot \left( \frac{1}{\rho} \nabla P \right) = \frac{2d}{\lambda n_0} \sum_{j \ne i} \left[ \frac{P_j - P_i}{2\rho_i \rho_j / (\rho_i + \rho_j)} w(|\boldsymbol{r}_j - \boldsymbol{r}_i|) \right] \tag{7}$$

After obtaining the distribution of the multiphase pressure field, the following pressure gradient model was used to calculate the pressure gradient term:

$$\nabla P = \frac{2d}{n_0} \sum_{j \ne i} \left[ \frac{P_j + P_i}{|\boldsymbol{r}_j - \boldsymbol{r}_i|^2} (\boldsymbol{r}_j - \boldsymbol{r}_i) w(|\boldsymbol{r}_j - \boldsymbol{r}_i|) \right] \tag{8}$$

Only brief outline of the MPS method is given in this section, more detailed models and algorithms could be found in relative references[13].

### 2.2 Surface tension model

According to the single film bubble's geometric feature (Figure 3), it could be considered as a liquid film shell with an amount of gas on both sides in the air. Since the flow of the liquid film and the deformation of the bubble often happen in a very low Reynolds Number with a very small velocity, the gas outside of the bubble is assumed that it has a very limited influence on the bubble. In the algorithm of the MPS, the influence of the outside gas on the bubble will be totally ignored except as the atmosphere pressure in this study. In other words, there will be no gas particles in the gas phase outside the bubble system, and we have zero pressure constantly on the outside free surface open in the air, but both sides of the film formed between bubbles are affected by gas.

Though the single liquid film is very thin, it has two gas-liquid interfaces including outside interface and inside interface. In order to accurately calculate the surface tension force on both sides, an interlayer of viscous fluid particles is employed to represent the liquid between the two interfaces. On the other hand, the gas inside the bubble is taken as incompressible and has a uniform density and pressure so that the surface tension force which acts on both sides of the film formed between bubbles in foams can be calculated.

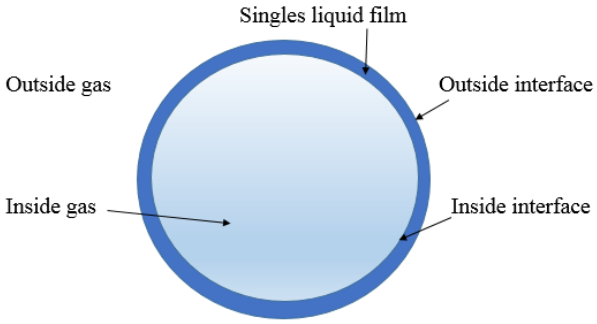



**Figure 3. The geometric feature of single film bubble**

Within the framework of the MPS method, the surface tension forces on the two interfaces will be calculated integrally using the surface free energy surface tension model[16]. The potential energy between two particles is denoted by P(r), and then the force $\boldsymbol{f}$ between the particles is:

$$\boldsymbol{f} = \frac{\partial \mathrm{P}}{\partial \mathrm{r}}\boldsymbol{n} \tag{9}$$

According to the requirement of repulsion when the distance between particles is less than a critical value and attraction when it is larger than a critical value. The force between the particles can be calculated by different formulas[16]**.** For the particles with particles of uniform and symmetrical distribution around, the resultant force on them is 0, when the particles are on or adjacent to the free surface, the resultant force is not 0, that is the surface tension. Therefore, the formula for calculating the potential energy between particles is[17]:

$$P(r) = \begin{cases} \frac{1}{3}C\left(r - \frac{3}{2}l_{min} + \frac{1}{2}r_e\right)(r - r_e)^2 & (\mathrm{r} < r_e) \\ 0 & (r \geq r_e) \end{cases} \tag{10}$$

where C is a modified parameter, which can be obtained according to the physical properties of fluids. $l_{min}$ is the boundary of repulsion and attraction, which is also the extreme point of surface free energy. $l_{min}$ is $1.5l_0$, and $r_e$ is $3.1l_0$ in the subsequent numerical calculation and research[18].

### 2.3 Interface tension

When the foam contacts the container, the external channels are created. For small foam containers, the relative number of exterior and interior channels is significant. Therefore, it is necessary to study the formation mechanism and the factors influencing the formation of external channels.

Interface tension exists between different substances that are in contact but not mutually soluble such as liquid-gas, liquid- liquid and liquid-solid. The wetting effect between multiple interfaces can be described by the following equations[18]:

$$\begin{cases} \sigma_{\alpha\theta}cos\alpha + \sigma_{\alpha\beta} + \sigma_{\beta\theta}cos\beta = 0 \\ \sigma_{\alpha\theta} + \sigma_{\alpha\beta}cos\theta + \sigma_{\beta\theta}cos\beta = 0 \\ \sigma_{\alpha\theta}cos\alpha + \sigma_{\alpha\beta}cos\theta + \sigma_{\beta\theta} = 0 \\ \alpha + \beta + \theta = 360° \end{cases} \tag{11}$$

where $\alpha$, $\beta$ and $\theta$ are contact angles, $\sigma_{\alpha\theta}$, $\sigma_{\alpha\beta}$ and are the surface tension coefficients of non-gaseous contact material.

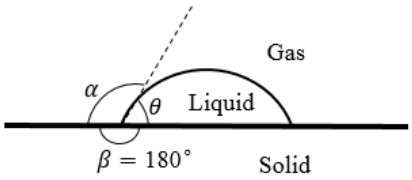



**Figure 4. The relationship between contact angles and interface tensions**

For the liquid-solid interface tension case, since the solid is considered as rigid body, β in Figure 4 is 180 °. Contact angle $\theta$ of liquid-solid system demonstrates the wettability of solid surface. If $\theta$ is smaller than 90 °, the solid surface is hydrophobic, or else the solid surface is hydrophilic. In this study, we focus on the effect of hydrophobic/hydrophilic solid surface on the forming process of the external channels.

## 3 Numerical validation

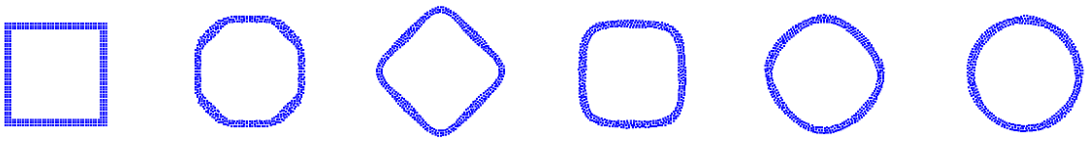
### 3.1 Oscillation of square single film bubble

The square single film bubble oscillation process is calculated with the surface tension model as a classic validation case. The results are shown in Figure 5. Initially, a square single film bubble is arranged with 60×60 particles including liquid particles and gas particles. The side length of the square bubble is 18 mm and the spacing between the particles is $l_0$=0.0003m. The time step is t = 0.00001s, and the physical parameters are shown in the captions of Table 1.

**Table 1. Particle Physical property parameters**

| Material | Density（kg/m$^3$） | Dynamic Viscosity Coefficient（N·S/m$^2$） | Surface Tension Coefficient（N/m） |
|---|---|---|---|
| Water | 1000 | $1.01\times10^{-3}$ | 0.0725 |
| Gas | 1.5 | $1.79\times10^{-5}$ | — |

The single film is composed of three layers of particles including the outside layer, the inside layer and the inter layer between them. The gas outside the bubble is ignored, and the gas inside the bubble is incompressible and uniform which is not shown in these figures (Figure 5).



(a) t=0 s (b) t=0.02 s (c) t=0.04 s (d) t=0.08 s (e) t=0.2 s (f) t=0.86s

**Figure 5. Oscillation process of square single film bubble under surface tension forces**

The square bubble firstly shrinks from the four corners with large curvatures until it is deformed into a diamond shape. Then the similar deforming process repeats, the square shape and the diamond shape appear alternately. However, since the viscous dissipation of the liquid and the gas, their interaction reduce the energy gradually in every time step, and the oscillation amplitude shows damped oscillation attenuation with time. Finally, the deformation will end when the amplitude approaches zero in Figure 6. As a result, a perfect round shape (t = 0.86 s in Figure 5) is generated.

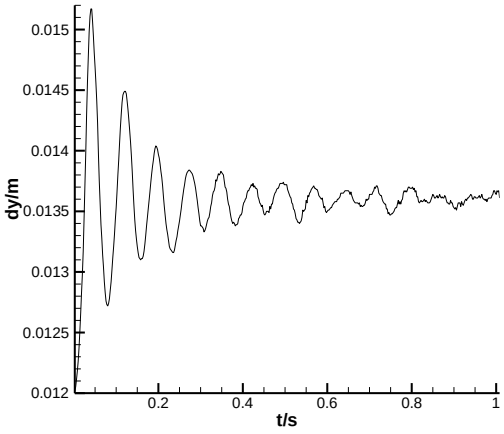



**Figure 6. Amplitude variation curve of y-axis oscillation during bubble deformation**

The square single film bubble oscillation process is a benchmark test to verify surface tension model. We know that Laplace law for the two-dimensional case is[19]:

$$\Delta P = \frac{\sigma}{R} \tag{12}$$

Where ΔP is the pressure jump across the interface. R is the radius of the bubble. When the amplitude approaches zero, ΔP can be calculated by counting the average pressure o the liquid film with the corresponding bubble radius R. The radius is set to be 3, 4, 5, 6, 7 and 8mm respectively. The surface tension coefficient $\sigma$ is taken as 0.072N/m. Then, the numerical results and the analytical solution (Eq. (12)) are drawn as a function of the radius of the bubble in Figure 7. As can be seen from the Figure 7, the numerical results agrees well with the analytical solution. These show that the Laplace law is accurately satisfied.

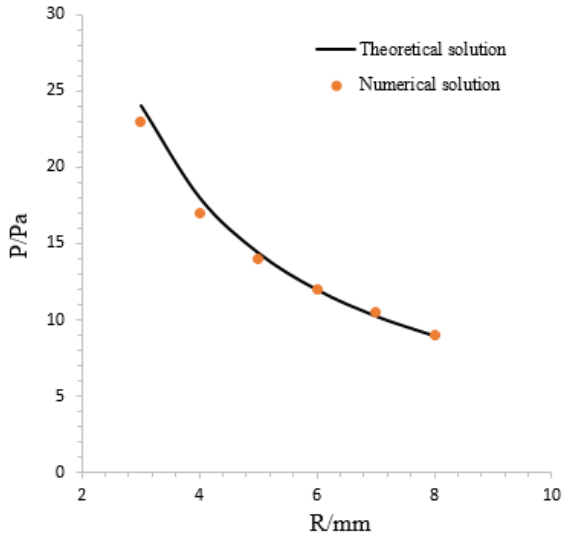



**Figure 7. The verification of Laplace law**

### 3.2 The solid-liquid wetting effect

A two-dimensional water droplet wetting on a solid wall without gravity was simulated in this section to verify the interface tension mentioned above. Theoretical solutions exist for a droplet wetting on solid under zero gravity as shown in Figure 8. The stable contact angle θ is formed at the contact point between the three phases[18]. Due to the surface tension, the free surface of droplet presents a regular sphere, and it's a regular arc in two-dimensional space. The relation of the parameters on Figure 8 is:

$$V=\pi R_0^2 = \frac{1}{2} \times 2\pi R \frac{2\theta}{2\pi} \times R - R\sin\theta \times R\cos\theta \quad (13)$$

$$S=2R\sin\theta \quad (14)$$

Where V is the volume of the droplet, S is the area of the contact part, R is the radius of the droplet surface,$R_0$ is the initial radius of the droplet.

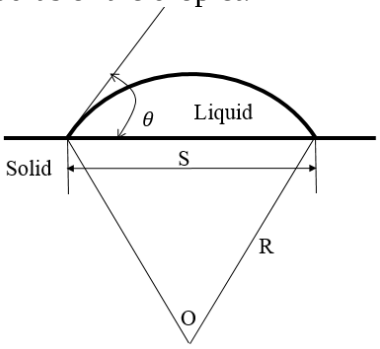



**Figure 8. The relationship between contact angle and contact area in solid-liquid wetting**

Figure 9 is the time serial of a water droplet wetting process on a solid wall under zero gravity with a static contact angle of 60°. The initial droplet radius is 0.02m with particle size $l_0$=0.001m and the time step is 0.001s. As shown in Figure 9, the wetting length increases rapidly with time, and the dynamic contact angle and droplet height decrease gradually.

(a) t=0 s (b) t=0.5 s (c) t=1.5 s (d) t=2 s (e) t=3 s (f) t=5 s

**Figure 9. Droplet wetting process on a solid wall**

As shown in Figure 9(f), the droplet reaches finally stable at about t=5s with steady contact angle and area like Figure 10.

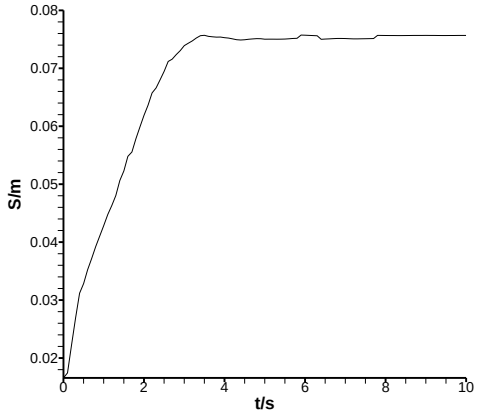



**Figure 10. Curve of wetting length over time**

Figure 11 shows the contact area of numerical and theoretical results at different contact angle. The simulation results indicate that the surface tension model and the interface tension mentioned could simulate solid-liquid wetting effect correctly.

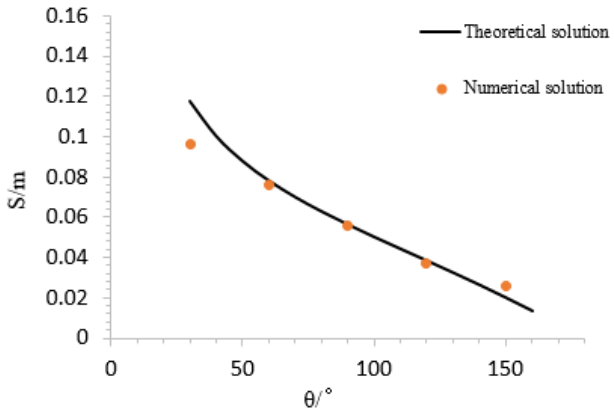



**Figure 11. Contact area of numerical and theoretical results at different contact angle**

## 4 Numerical simulation and discussions

### 4.1 Dynamic Behavior Between Two Bubbles

In order to study the dynamic behavior of bubbles in detail, the flow field velocity of the liquid film in the calculation area is counted, and the average velocity of the liquid film is obtained from the following formula.

$$velo_L = \frac{\sum_i^{N_L} velo_i}{N_L} \tag{15}$$

Where $velo_L$ is the average velocity of liquid film particles, $N_L$ is the total number of liquid particles, $velo_i$ is the velocity of liquid particle $i$.

Without considering the effect of gravity, the process of bubble coalescence at the same scale is simulated and the effects of different physical parameters on its dynamic behavior are analyzed. Then the more general phenomenon of bubbles' dynamic behavior at different scales is simulated and analyzed.

#### 4.1.1 Simulation of interaction between the same sizes of single film bubbles

Figure 12 is the time serial of the coalescence process of three-dimensional bubbles with the same size. The initial position of the bubbles is shown in Figure 12(a), and Figure 12(b-f) is obtained by cutting along the plane of the center of two bubbles where blue particles and red particles represent liquid film and gas respectively. The radius of the bubble is 4.5 mm with particle size $l_0$=0.3mm, the distance between the center of mass of the bubble is 9.3 mm, the thickness of the liquid film is 0.9 mm, and the time step is 0.00001 s. The physical parameters are shown in the table 2.

**Table 2. Particle Physical property parameters**

| Material | Density（$kg/m^3$） | Dynamic Viscosity Coefficient（N S/$m^2$） | Surface Tension Coefficient（N/m） |
|---|---|---|---|
| Water | 1000 | $2\times10^{-2}$ | 0.032 |
| Gas | 1.5 | $1.79\times10^{-5}$ | — |

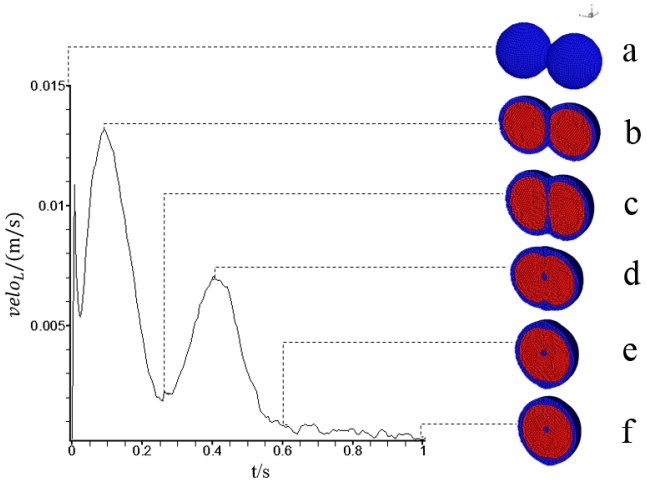



**Figure 12. The coalescence process of three-dimensional bubbles with the same size**

It can be seen from the Figure 12 that the initial velocity of two equal-scale bubbles is zero at the initial time. Since the bubble is not a sphere in the strict sense when it is arranged in Cartesian coordinates, the bubble shrinks into a sphere under the action of surface tension .Therefore, the average velocity of liquid film increases sharply in a very short time .In the process of forming spheres, two bubbles contact each other (there is a small degree of extrusion), and liquid film is formed between bubbles, then the average velocity of liquid film decreases sharply. The above process takes place in a very short time and can be regarded as an adaptive process of the initial physical environment of bubbles, so the influence of fluid properties on this stage can be neglected.

In the initial stage (a-b) of the contact process between two bubbles, a large curvature is formed at the angle between the outside of the liquid film. The surface tension drives the surface of the liquid film to deform and fuse rapidly, which makes the average velocity of liquid film increase.

With the continuous approaching of bubbles (b-c), the average velocity of liquid film decreases because the squeezing degree between bubbles increases. Besides, the intermediate liquid film gradually expands to both sides so that the liquid film becomes thinner through continuous drainage.

Subsequently, under the action of inertia, bubbles continue to move toward the center (c-d) and the liquid film breaks up when the liquid film thickness reaches a critical value. Because the liquid film formed between bubbles is not uniform, there is a curvature difference, the liquid film gradually stretches longitudinally, and the average velocity of liquid film shows an upward trend.

When the liquid film is stretched to a certain extent, it gradually shrinks to the center (d-e) under the action of surface tension with ellipsoidal bubbles transiting to spherical bubbles. Finally, under the action of surface tension and viscous dissipation, the average velocity of liquid film also shows a trend of attenuation through continuous contraction oscillation.

Finally, the oscillation amplitude becomes smaller and smaller, and the average velocity of liquid film basically remains unchanged, and approaches zero. The two bubbles merge into a large bubble with a stable shape (e-f).

Besides, the broken liquid film disperse in the bubble without the gravity, and under the action of surface tension, the small spherical droplet are formed. Because of the symmetry, the small spherical droplet suspends in the center of the bubble.

### 4.1.2 The influence of different physical parameters on the interaction between bubbles

From the above analysis, it can be found that when bubbles merge, the velocity curve shows damping oscillation attenuation with the shape of bubbles changed. Finally, the average velocity of liquid film becomes zero, and the two bubbles merge into a static spherical bubble. From the point of view of energy, the two bubbles are driven by chemical potential energy in a static state with a tendency to form a bubble. The shape of bubble deforms during coalescence, which will cause viscous dissipation and consume a part of energy. The chemical potential energy is related to the surface tension of bubbles, so the effects of surface tension and fluid viscosity on the dynamic behavior of bubbles will be studied below.

For the coalescence process of three-dimensional bubbles with the same size above, Figure 13 is the average velocity of liquid film under different viscous coefficients with the same surface tension coefficient, and other parameter settings are the same as table 2 except for the dynamic viscosity coefficient.

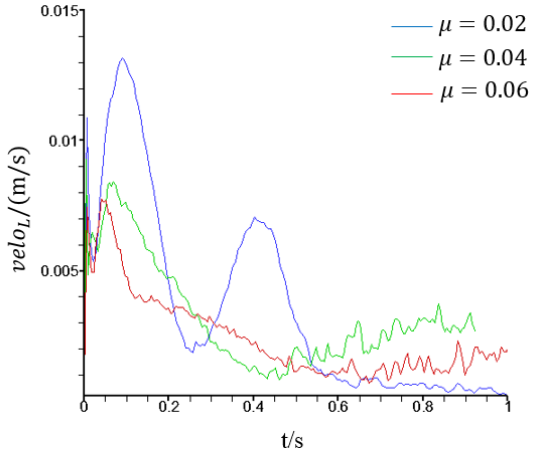



**Figure 13. Average velocity of liquid film under different viscous coefficients**

The smaller the viscous coefficient is, the larger the first peak value of the bubble coalescence speed and the larger the oscillation amplitude of the velocity curve are, which shows that the

chemical potential energy in the process of bubble coalescence is dissipated through the viscous term. The larger the viscous coefficient is, the more the energy dissipation per unit time is. With the increase of viscosity, the peak velocity of liquid film decreases, and the time required to reach the first peak velocity decreases.

With the increase of viscous coefficient, the deformation of bubbles is smaller in unit time. This is precisely because the viscous coefficient reflects the strong degree of fluid impediment to bubble motion. If the viscous coefficient is large, the resistance of the bubbles' coalescence will be large, and the viscous dissipation will be large in the course of motion. Therefore, the larger the viscous coefficient is, the less likely the liquid film held by the bubbles in the interaction will crack.

Similarly, Figure 14 is the average velocity of liquid film under different surface tension coefficients with the same dynamic viscosity coefficient, and other parameter settings are the same as table 2 except for surface tension coefficient.

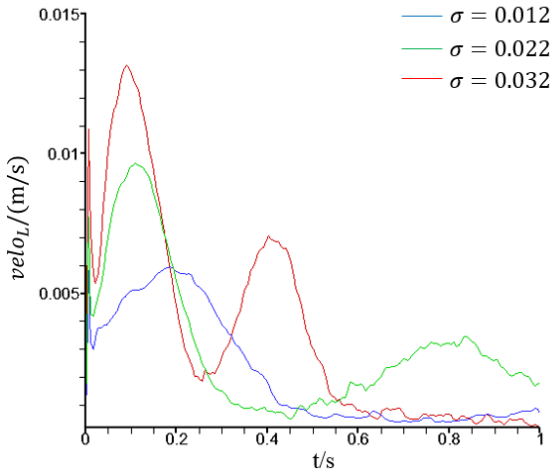



**Figure 14. Average velocity of liquid film under different surface tension coefficients**

The relationship between the first peak value of velocity and surface tension is opposite to that of viscosity. The higher the surface tension is, the earlier the bubbles begin to polymerize, and the higher the average peak velocity of liquid film is.

With the increase of surface tension, the deformation of bubbles increases in unit time, which shows that the chemical potential energy is related to the surface tension of bubbles. Under the effect of surface tension, the surface free energy of bubbles changes into kinetic energy, and the average velocity of particles increases.

### 4.1.3 Simulation of interaction between the different sizes of single film bubbles

It can be seen from the above analysis that when the viscosity coefficient is increased or the surface tension is reduced, the stable shape of the connection between bubbles is more likely to appear. The new physical parameters are shown in the table 3. Figure 15 is the time serial of the connection process of three-dimensional bubbles with the different sizes. The initial position of the bubbles is shown in Figure 15(a). The radius of the bigger bubble is 6 mm and the radius of the smaller bubble is 3 mm with particle size $l_0$=0.3mm, the distance between the center of mass of the bubble is 9.3 mm, the thickness of the liquid film is 0.9 mm, and the time step is 0.00001 s.

**Table 3. Particle Physical property parameters**

| Material | Density（$kg/m^3$） | Dynamic Viscosity Coefficient（$N·S/m^2$） | Surface Tension Coefficient（N/m） |
|---|---|---|---|
| Water | 1000 | $2\times10^{-2}$ | 0.012 |
| Air | 1.5 | $1.79\times10^{-5}$ | — |

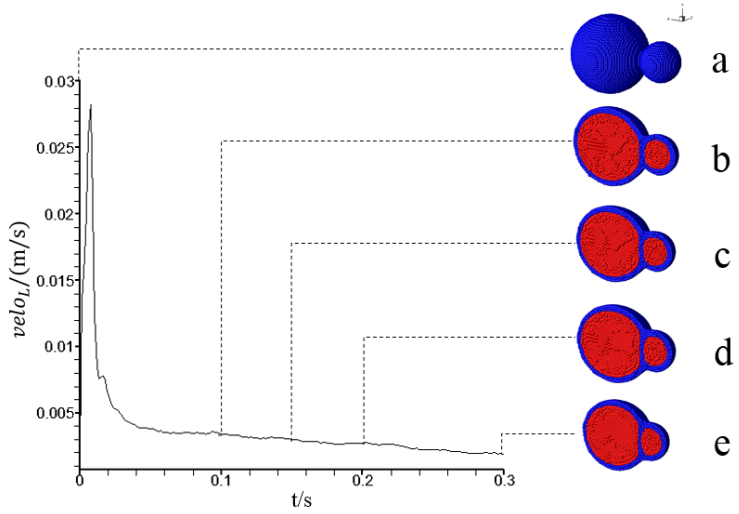



**Figure 15. The connection process of three-dimensional bubbles with the different size**

It can be seen from the Figure 15 that when the two bubbles start to contact each other at the beginning, the curvature of the place where the bubble contacts is large, and the liquid film is rapidly deformed and fused under the surface tension (a-c). Then, a liquid film is formed between the bubbles (c-e).

Due to the symmetry, the liquid film held by the two bubbles with the same size is a circular plane when they are in contact with each other. However, when the non-equal two bubbles are in contact with each other, the liquid film is a spherical surface having a certain curvature and is convex toward the larger bubble direction (d-e), because the small bubble curvature is larger than the large bubble, and the additional pressure of the small bubble is greater than the large one.

### 4.2 The forming process of different kinds of foam channels structure

By simulating the dynamic behavior of three-dimensional single-film bubbles of equal and non-equal scales, typical flow phenomena and deformation characteristics and rules of liquid film are obtained, and the effects of basic physical parameters including surface tension and dynamic viscosity on flow and deformation were studied. On the above basis, further research on the formation mechanism of various foam channel structures has been carried out through the following work.

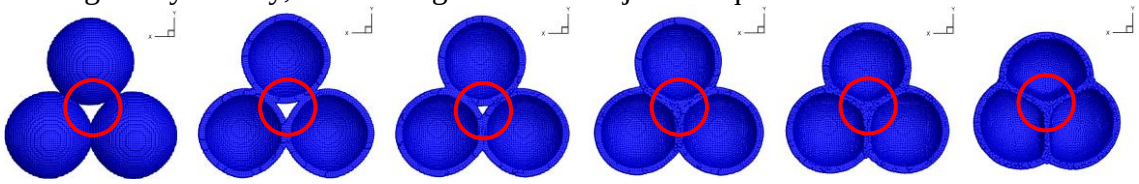
To better understand the formation mechanism of various foam channel structures, requires a more complete description the different kinds of foam channel structure firstly. In liquid foam, bubbles stack with each other, and the gaps between bubbles form a network channel for the flow of trace liquid, which is composed of Plateau channel and the junction point. In this study,

the foam channels are divided into three types including interior node-PB and exterior node-PB as well as open node-PB. Subsequently, the forming process of different kinds of foam channel structure were displayed and discussed as follows.

### 4.2.1 Open node-PB

In the study of the open Plateau channel where the bubbles in the foam are in contact with the air, in order to save computing resources and ensure the rationality of the calculation results, three bubbles with the same size are selected as the basic unit of calculation.

Figure 16 is the time serial of a cross-section of three bubbles with the same size during connection and the gas inside the bubbles is not shown in the figures. The initial arrangement of three bubbles with the same size is shown at t = 0s, the radius of a single bubble is 6 mm, and the thickness of the liquid film is 0.9 mm. Under the action of surface tension, the three bubbles contact with each other, then the liquid film deform rapidly and stick together with each other. During this process like Figure 16(a-d), the liquid film held between bubbles expands continuously, and the space occupied by air between bubbles is filled by liquid film continuously. Finally, the liquid film held by the three bubbles does not change significantly, and form a more stable system like Figure 16(f). The liquid film in the stabilized bubble system shows good symmetry, and the angle between adjacent liquid films is about 60 °.



(a) t=0 s (b) t=0.02 s (c) t=0.04 s (d) t=0.06 s (e) t=0.1 s (f) t=0.3 s

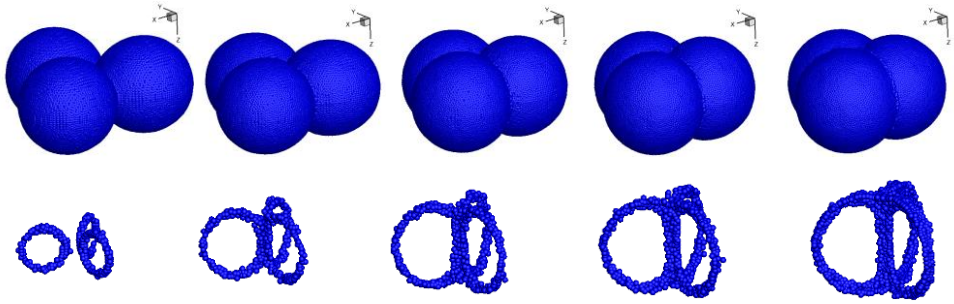
**Figure 16. Three equal-scale bubbles connection processes**

At the same time, in order to further study the details of bubble evolution from sphere to polyhedron, triangle in polygon configuration is selected as the basic research unit, and the process of gas disappearance and polygon formation is analyzed in Figure 16(a-f) where the part marked with red circles. At the initial moment, the bubble is completely circular, and the gas in the triangle region is the largest. With the structure evolution of adjacent bubbles under surface tension, the triangular region begins to sew up and the internal gas decreases gradually. At t=0.3 s, all the gas disappear with the curved surface held by bubbles gradually becoming flat. At this time, it can be predicted that when the space is filled with multiple bubbles, the polygon structure can be formed completely.

In order to better analyze the formation mechanism of liquid film channel for micro-liquid flow in the process of bubble interaction, the channel structure at a specific time is specially extracted in this study.

Figure 17 is the time serial of three bubbles forming an Open node-PB during the connection process. Under the action of surface tension, bubbles keep approaching each other, and the space occupied by air between bubbles is continuously filled by liquid film. Each bubble liquid film contacts with each other to form a new liquid film and a liquid film channel encapsulating the liquid film. Later, the distance between bubbles decreases further, and the area of liquid film held by bubbles enlarges and the channel of liquid film encapsulating liquid film increases. When the channel of liquid film increases to a certain extent, the channel of liquid film contacts with each other to generate a new liquid film channel and two nodes. As shown in Figure 17(a-

e), there is no independent channels which are connected by nodes. Each node has four channels, and the channel centered on the nodes is umbrella-shaped. If the formation mechanism of a single node is analyzed, the basic process is similar to the above process. It is very similar to the formation of the liquid film channel and the nodes between the channels in the foam system, which is directly contacted with the air. Therefore, this type of channel is an Open node-PB in this study.



(a) t=0.05 s (b) t=0.1 s (c) t=0.15 s (d) t=0.2 s (e) t=0.3 s

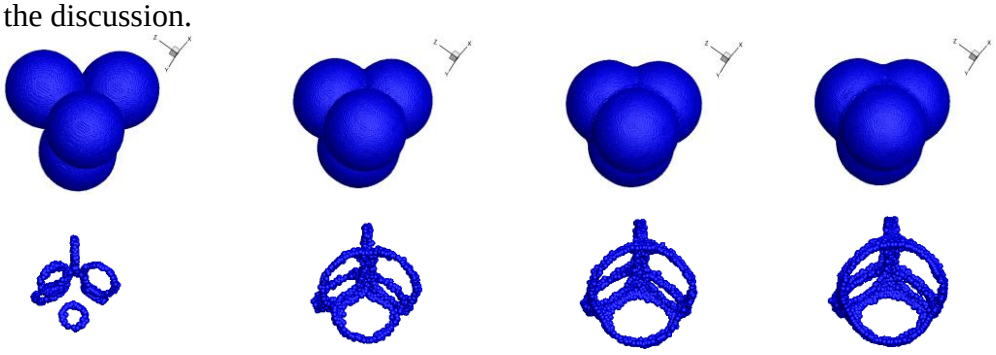
**Figure 17. Three equal-scale bubbles connection processes with film channels**

### 4.2.2 Interior node-PB

Similarly, in the study of the internal Plateau channel in the foam system, in order to save computing resources and facilitate the analysis and discussion of the formation of the internal Plateau channel, four bubbles with the same size are selected to form a basic unit of interaction in this study.

Figure 18 is the time serial of formation of Plateau Channels in four bubbles during the connection process. Initially, the four bubbles are tetrahedral distribution. The radius of a single bubble is 6 mm, and the thickness of the liquid film is 0.9 mm. Under the action of surface tension, the four bubbles contact each other, and make the liquid film rapidly deform and bond with each other. In this process, the liquid film held by the bubbles expands continuously, and the space occupied by the air between bubbles is filled and occupied by the liquid film continuously. Finally, the liquid film held by the four bubbles does not change significantly, and form a relatively stable system, which is similar to the behavior of the three bubbles mentioned above.

As bubbles keep approaching each other, six new liquid films and the liquid film channels encapsulating the liquid membranes are formed. When the liquid film channel increases to a certain extent, the liquid film channel contacts each other, and four new liquid film channels are generated and five nodes are generated. This paper considers that the external liquid film channel (outline) is a part of the interaction between the basic unit and other basic units, so it is neglected in the discussion.



(a) t=0.05 s (b) t=0.15 s (c) t=0.3 s (d) t=0.45

**Figure 18. Four equal-scale bubbles connection processes with film channels**

In order to better analyze the channel structure, the internal Plateau channel at t=0.45 s is specially extracted and rotated every 60 degrees along the x-axis in this study. As shown Figure 19 below, the four bubbles share a central interior node, where each of the three bubbles forms a concave triangular inner Plateau border, and the four bubbles form four inner Plateau channels.

**Figure 19. Interior node-PB structure at different angles**

Figure 20 is the internal structure of four bubble stabilization units, where the different colors represent the different pressure values. As can be seen from the figure 20, the pressure values at channels and nodes are all higher than the pressure values of the nearby liquid film. The area with a smaller pressure value means the thickness of the liquid film there is smaller, in other words, it contains less liquid.

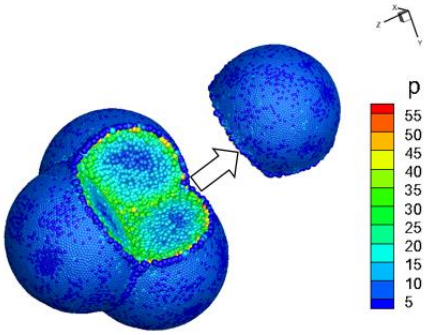



**Figure 20. Internal structure of four bubble stabilization units**

Figure 21 is the Plateau structural unit. For the whole foam system, the liquid is mainly distributed in the liquid films and liquid film channels, and the liquid content in the liquid film channels is more. If gravity is taken into consideration, the liquid stored in the liquid film channels will naturally drain, making it difficult for the entire foam system to remain stable for a long time. This is also one of the difficult problems in the experimental study of the formation mechanism and structural characteristics of the foam channel. In this section, the influence of gravity is neglected and the drainage process is suppressed, making the bubble system easier to achieve and maintain stable state. After considering the size of bubbles and the number of bubbles, the formation mechanism and structural characteristics of the whole foam system can be further analyzed.

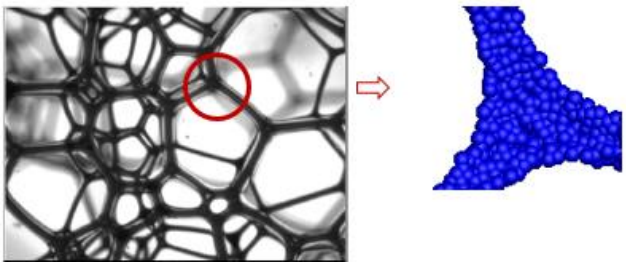


**Figure 21. Plateau structural unit**

### 4.2.3 Exterior node-PB

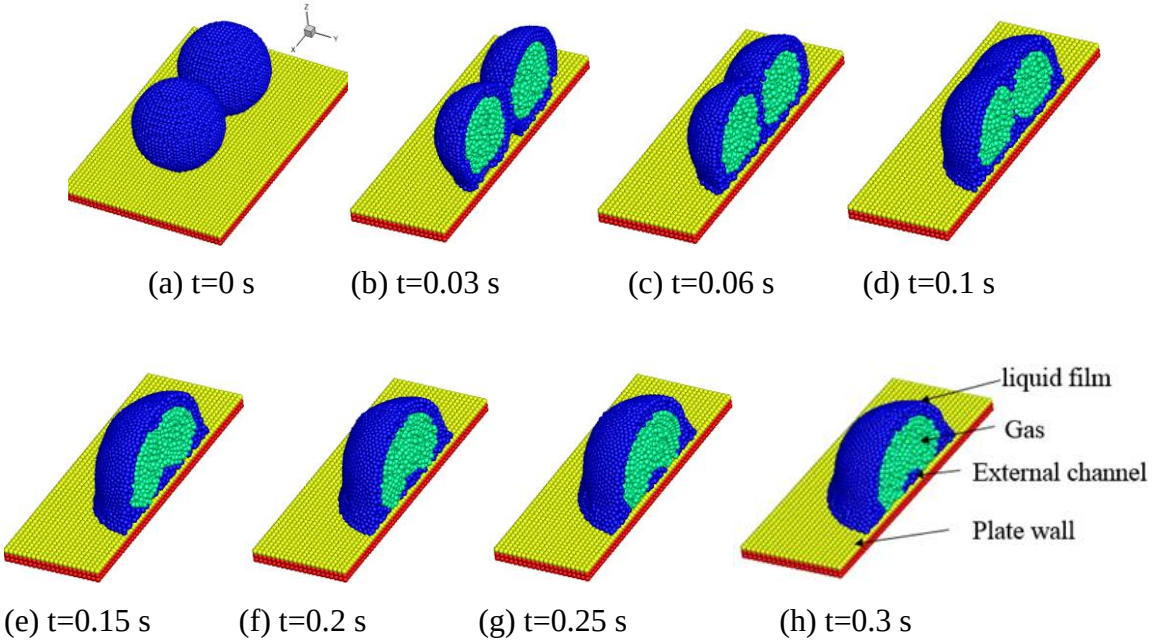
Foam systems usually exist in containers. Therefore, the relative number of exterior and interior channels is significant for small foam containers. The container wall is treated as flat wall in

this paper. The exterior foam's fundamental difference with interior foam is the existence of a no-slip wall.

The influence of vessel wall on channel formation is mainly in two aspects. One is that the existence of the wall restricts the space expansion of the channel to the normal direction of the wall during the formation process. The other is that the different wettability of the wall material to the liquid will affect the formation of the channel and the shape of the stable channel. In this study, the interaction of two bubbles on a flat plate with different wettability is simulated, and the effect of wall surfaces with different wettability on the formation of the channel and the shape of the stable channel is studied.

In order to facilitate the observation and analysis of the formation of external channels, the bubbles are split vertically along the center of the bubbles. Figure 22 is the time serial of the two bubbles interacting and infiltrating on a plate wall under zero gravity with a static contact angle of 30 °. The initial arrangement of two bubbles with the same size is shown at t = 0s, the radius of a single bubble is 3 mm, and the thickness of the liquid film is 0.9 mm. and the time step is 0.00001s.

As shown in Figure 22(a-d), the coalescence of two bubbles is accompanied by the infiltration of bubble liquid film into a flat plate. During the coalescence of the two bubbles, as the liquid film continuously infiltrates the wall, the bubble liquid film begins to contact the wall surface and connect with each other, and the liquid film held between the two bubbles undergoes the process of forming, expanding and breaking up like Figure 22(a-d). Then, the two bubbles coalesce into a larger bubble. Under the action of gas pressure and interfacial tension in the larger bubble, the liquid film in contact with the plate in the bigger bubble changes constantly and cracks finally like Figure 22(c-e). Subsequently, the fractured liquid film gathers on the plate to form an external channel in Figure 22(e-g). As shown in Figure 22(g-h), the external channel reaches finally stable at about t=0.35s with steady contact angle and area finally.



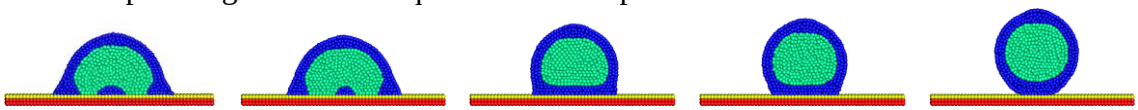

(a) t=0 s (b) t=0.03 s (c) t=0.06 s (d) t=0.1 s

(e) t=0.15 s (f) t=0.2 s (g) t=0.25 s (h) t=0.3 s

**Figure 22. Bubbles wetting process on a solid wall**

Figure 23 shows the axial cross section shape of the coalesced bubbles at different contact angles (30 °, 60 °, 90 °, 120 ° and 150 °), when the coalesced bubbles are in stable state on the plate. From the simulation results, it can be concluded that the different wall wettability will

affect the dynamic behavior of bubbles on the flat plate when bubbles are in contact with a flat plate. When the wall wettability is hydrophilic, it is easy to produce an exterior node-PB during the coalescence of multiple bubbles on a flat plate, and the shape of the axial cross section of the channel is closely related to the hydrophilicity. On the contrary, when the wall wettability is hydrophobic, it is hard for bubbles to form an exterior node-PB in the process of coalescence on the plate, but to spread a layer of liquid film on the plate, and the spreading area of the liquid film on the plate is related to the hydrophobicity. The stronger the wall hydrophobicity, the smaller the spreading area of the liquid film on the plate.



$\theta$=30 ° $\theta$=60 ° $\theta$=90 ° $\theta$=120 ° $\theta$=150 °

**Figure 23. Bubbles wetting result on a solid wall with different contact angles**

## 5 Conclusions

Based on the MPS method, the bubbles behavior and the forming process of different kinds of foam channel structure were simulated in three-dimensional space by introducing the surface tension and interfacial tension model. The MPS method is a Lagrange method which avoids the occurrence of convection terms in the flow control equation, and eliminates the numerical dissipation phenomenon that may occur in the Euler method. The modeling process is more reasonable and the physical meaning of the parameters is clear, which provides a new idea for the study of foam dynamics .The main conclusions of this study are as follows.

By simulating the dynamic behavior of three-dimensional single-film bubbles of equal and non-equal scales, typical flow behaviors including coalescence and connection, and deformation characteristics and rules of liquid film are obtained. The effects of basic physical parameters on flow and deformation were studied. It was found that decreasing the surface tension coefficient or increasing the viscous coefficient would weaken the dominant role of surface tension in bubble deformation. As the simulation progresses, the bubble motion tends to be stable due to viscous dissipation.

Further, the Plateau channel structure in the foam system is further classified. The liquid channels in the foam system are divided into interior node-PB and exterior node-PB as well as open node-PB. The formation of different kinds of the Plateau channel structure is simulated and the results of each simulation are analyzed and discussed.

With the continuous approaching of bubbles, the length and cross-sectional area of channels formed between bubbles are increasing. Finally, when the multi-bubble system is stable, the main characteristics of liquid film channels are as follows:

For the open node-PB, there is no independent channels which connected by nodes. There are two nodes. Each node has four channels, and the channel centered on the nodes is umbrella-shaped. For the interior node-PB, the four bubbles share a node, where each of the three bubbles forms a concave triangular inner Plateau border, and the four bubbles form four inner Plateau channels. For the exterior node-PB, the different wall wettability will affect the dynamic behavior of bubbles on the flat plate when bubbles are in contact with a flat plate. When the wall wettability is hydrophilic, it is easy to produce an exterior node-PB during the coalescence of multiple bubbles on a flat plate, and the shape of the axial cross section of the channel is closely related to the hydrophilicity. On the contrary, when the wall wettability is hydrophobic,

it is hard for bubbles to form an exterior node-PB in the process of coalescence on the plate, but to spread a layer of liquid film on the plate, and the spreading area of the liquid film on the plate is related to of hydrophobicity. The stronger the wall hydrophobicity, the smaller the spreading area of the liquid film on the plate.

**References**

[1] D. Weaire, S. Hutzler, The Physics of Foams, Oxford University Press, Oxford, 2000

[2] S.A. Koehler, S. Hilgenfeldt, H.A. Stone. (2004) Foam drainage on the microscale I. Modeling flow through single Plateau borders, Journal of Colloid and Interface Science 276, 420–438.

[3] R.K. Prud'homme, S.A. Khan (Eds.), Foams, Theory, Measurements and Applications, Dekker, New York, 1996.

[4] L.J. Gibson, M.F. Ashby, Cellular Solids, Cambridge University Press, Cambridge, UK, 1997

[5] P. Stevenson (Editor), Foam Engineering: Fundamentals and Applications (Wiley, 2012).

[6] I. Cantat, S. Cohen-Addad et al, Foams – Structure and Dynamics (Oxford University Press, 2013).

[7] S.A. Koehler, S. Hilgenfeldt, H.A. Stone, (2004)Foam drainage on the microscale II. Imaging flow through single Plateau borders, Journal of Colloid and Interface Science 276, 439–449.

[8] D. G. T. Barrett , S. Kelly E. J. Daly et al, (2008) Taking Plateau into Microgravity: The Formation of an Eightfold Vertex in a System of Soap Films ,Microgravity Sci. Technol, 20:17–22

[9] Noever DA. Foam fractionation of particles in low gravity[J]. Journal of Spacecraft and Rockets, 1994, 31(2): 319-322.

[10] C.W. Hirt, B.D. Nichols, Volume of fluid (VOF) methods for the dynamics of free boundaries, J. Comput. Phys. 39 (1981) 201–225.

[11] M. Sussman, E. Fatemi, P. Smereka, S. Osher, An improved level set method for incompressible two-phase flows, Comput. Fluids, 27(1998) 663–680.

[12] H.W. Zheng, C. Shu , Y.T. Chew, A lattice Boltzmann model for multiphase flows with large density ratio, Journal of Computational Physics 218 (2006) 353–371.

[13] Koshizuka S. Oka Y. Moving-particle semi-implicit method for fragmentation of incompressible fluid [J]. Nuclear Science and Engineering, 1996, 123(3): 421-434.

[14] Zhangguo Sun,Ni Ni,Yijie Sun,Gung Xi,Modeling of single film bubble and numerical study of the plateau structure in foam system ,China Ship Scientific Research Center.https://www.jhydrod.com/2018,30(1):79-86

[15] Guangtao Duana, Bin Chena,Seiichi Koshizukab, Hao Xiang.Stable multiphase moving particle semi-implicit method for incompressible interfacial flow,Comput. Methods Appl. Mech. Engrg. 318 (2017) 636–666

[16] Tartakovsky A M, Meakin P. Modeling of surface tension and contact angles with smoothed particle hydrodynamics[ J]. Physical Review E, 2005, 72:026301.

[17] Kondo M, et al. A development of surface tension model using inter-particle force[A].Proceeding of Computational Engineering Conference[C],2006,19:337-338.

[18] Xiao Chen, Zhongguo Sun, Guang Xi. Numerical Investigation on Droplet Wetting Effect With the MPS Method ,Proceedings of the ASME 2014 4th Joint US-European Fluids Engineering Division Summer Meeting FEDSM2014 August 3-7, 2014, Chicago, Illinois, USA.

[19] H.W. Zheng, C. Shu , Y.T. Chew. A lattice Boltzmann model for multiphase flows with large density ratio, Journal of Computational Physics 218 (2006) 353–371.

# Calculating the Forced Response of a Mistuned Blisk via Surrogate Models

***Shiyuan DENG[1], †Jianyao YAO ¦ Jianqiang XIN,[2] Chenxu JIANG[1], Teng LI[1], Dongyang SUN[1], Zejun WU[1]**

[1]College of Aerospace Engineering, Chongqing University, China
[2]Research and Development Center, China Academy of Launch Vehicle Technology, China

*Presenting author: dsycircle@163.com
†Corresponding author: yaojianyao@cqu.edu.cn

## Abstract

The forced responses of blisks are highly sensitive to inevitable random mistuning, which can cause severe damage. Considerable computational efforts are required for the sampling process to acquire the statistical vibration properties of mistuned blisks via finite element models. Therefore, efficient surrogate models are preferred. In this paper, four methods are utilized to construct the relation between random mistuning and forced response amplitudes. These four methods include polynomial chaos expansion (PCE), response surface method (RSM), Kriging interpolation and artificial neural networks (ANN). A lumped parameter model of a 24-sector blisk is used to investigate the vibration patterns. Each sector has 2 degrees-of-freedom (DOF). Mistuning is simulated by treating the stiffness of blades as independently identically distributed (i.i.d) random variables and Sobol sequence is applied for designing the sample sites. On assessing the result, mean squared error (MSE) and Kolmogorov-Smirnov test are applied to numerically determine the accuracy. Results show that PCE can yield the most accurate and stable predictions of the statistical characteristics of the forced responses; Kriging interpolation and ANN are also remarkable while RSM does not show any priority on this issue.
**Keywords**: mistuned blisks, forced response, polynomial chaos expansions, response surface method, Kriging interpolation, artificial neural networks

## Introduction

Mistuning refers to inevitable variations of blisks and arises during the manufacturing and assembly process due to wear, maintenance, machining error, material dispersion, etc. These small deviations can lead to much larger response level than the ideal, tuned design[1]. The severe amplification of vibration altitude may cause high cycle fatigue (HCF) and premature failure of the blades. As was estimated in 1998, about 30% of all engine maintenance costs were caused by HCF[2]. Therefore, it is of great interest to accurately predict the adverse effect of mistuning on the vibration of blisks.

The mistuning across blades is randomly distributed[3,4], thus we usually have to use the statistical properties of the forced responses of blisks to assess the effects of mistuning. The most commonly used method is Monte Carlo Simulation (MCS), which requires an amount of repetitive analysis and tiny deviations of parameters are needed in each process[5]. Typically, finite element (FE) models are applied to carry out the calculation process[3,3,6].

There have been numerous studies on the mistuned blisks using FE models. Slater pointed out that a complete FE model instead of a single sector was necessary because mistuning could break the cyclic symmetry[7]. Petrov considered it as an optimization problem to search for the best and worst vibration pattern using an FE model[8]. Castanier gave a complete FE model of a mistuned blisk[1]. He discussed about some of the fundamental issues of mistuned disks, including coupling and mode localization, and proposed an algorithm to accelerate MCS. Laxalde built a multistage FE model for mistuned blisks and also confirmed the validity[9]. More recently, accurate modeling of small manufacturing errors or geometric mistuning of blades is realized using the coordinate measurement methods[10]. One of the most significant advantages that make FE models irreplaceable is the high accuracy. However, although computational capacity of modern computers has greatly improved, it is still challenging to carry out a large number of MCS based on full FE models. Therefore, much attention has been paid to reduced order models (ROM) for mistuned blisks, mainly including the component-mode-based methods and the system-mode-based methods[11–14].

The lumped parameter models are one kind of ROM. They treat each sector as a spring-mass oscillator connected to the ground and coupled to neighboring sectors by linear springs. Although not capable of precisely representing the actual engineering structures, these models can capture some basic features of the mistuned blisks, such as the modal localization and vibration amplification[15,16]. Also, the required computational efforts are far less than FE models. Therefore, they provide a good insight of the vibrational mechanism of mistuned blisks together with MCS.

Surrogate models are another popular approach for uncertainty analysis in engineering. Instead of direct operations on each individual sample, the surrogate models pursue to establish an analytical relationship between the random input and output based on the results obtained at designed sampling points, which is much easier to use, and once the model is obtained, the computational efforts for new points are negligible. Since the forced responses of mistuned blisks are highly sensitive to the random mistuning, the relationship between the response amplitude and mistuning is highly nonlinear and difficult to define, thus the application of surrogate models for mistuned analysis has not drawn much attention. In a previous study, we tried to use RSM and Kriging metamodel to obtain the response amplification factor over a frequency range, but only succeeded when 4 sectors out of 24 were mistuned. When the number of mistuned sectors increases, the accuracy and the required number of samples of the metamodels are not acceptable. Sinha applied the method of PCE to analytically compute the statistics of the forced response of a mistuned blisk assembly[17]. But the proposed method is an intrusive one, which means that we have to modify the solver correspondingly. In practice, we prefer non-intrusive methods, which only need the input and output to construct the surrogate models.

In this paper, we will try to build the relation between the forced response amplitude and the mistuning via four surrogate models, namely the polynomial chaos expansions (PCE), the response surface method (RSM), the Kriging interpolation and the artificial neural network (ANN), and compare their validities. A lumped parameter model is used to generate the training data. In later parts, we will introduce: 1) The basic theories and implementations in mistuned blisks of the aforementioned four methods; 2) A numerical example of lumped parameter blisk model; 3) The validation of the four methods, followed by the results and discussions; 4) Some conclusions.

**Briefs of Response Surface Method, Artificial Neural Networks, Kriging Interpolation and Polynomial Chaos Expansion**

*Surrogate Models for Forced Vibration Analysis of Mistuned Blisks*

For a blisk, the amplitude of sector $i$ can be denoted as $A_i$. If only the stiffness mistuning is considered, $A_i$ is determined by system stiffness matrix $\boldsymbol{K}$, which consists of stiffness of the sectors:

$$A_i = A_i(\boldsymbol{K}) \tag{1}$$

If we have $n$ random variables $\boldsymbol{v} = [v_1, v_2, \cdots, v_n]^{\mathrm{T}}$ representing the input random mistuning for $n$ sectors, and the scalar output $u(f_j)$, which is the amplitude of forced response with excitation frequency $f_j$, the target surrogate model can be written as

$$u(f_j) \approx f(\boldsymbol{v}) \tag{2}$$

In this paper, $f$ refers to PCE, RSM, Kriging interpolation and ANN and can be obtained using the designed training points, which will be described in the following subsections.

We use the cross validation method to verify the surrogate models. Typically, the sample set is partitioned into two parts. One part is called the training set, noted as $D$, used to build a model, and the other is called the test set, noted as $T$, used to verify the model. Generally, $D$ and $T$ are guaranteed to be i.i.d. The sample sizes are respectively written as $|D|$ and $|T|$. Error on $T$ will be a significant indicator of the model validity. If the accuracy of the constructed surrogate model is acceptable, we can use it for the response analysis with other random mistuning to obtain the statistical characteristics. If not, we need to update or add some new training points to create a new model and validate it again. The implementation procedure is illustrated in figure 1.

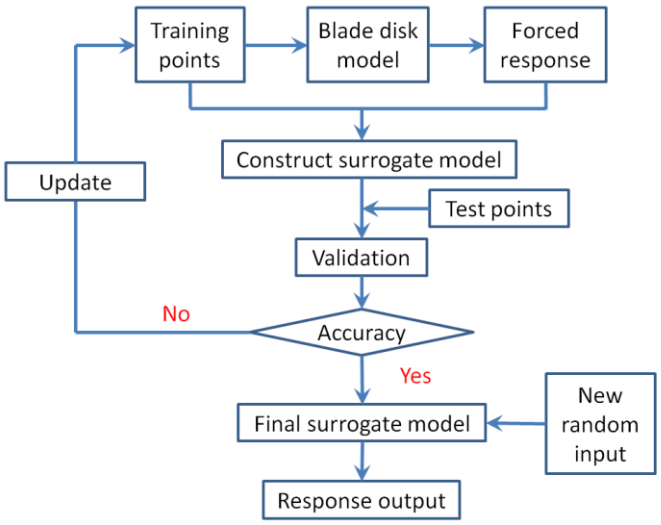



Figure 1. The flowchart for establishing the surrogate models for forced response analysis of mistuned blisks

*Polynomial Chaos Expansion*

The method of PCE expands the stochastic process $u = u(\boldsymbol{v})$ into an infinite series[18]:

$$u = u(\boldsymbol{v}) = \sum_{k=0}^{\infty} \alpha_k \, \phi_k(\boldsymbol{v}) \tag{3}$$

where $\phi_k(\cdot)$ is a polynomial basis and $\alpha_k$ is the corresponding coefficient. Often this expression is truncated to $m$ terms as an approximation of the infinite orthogonal series:

$$\sum_{k=0}^{\infty} \alpha_k \, \phi_k(\boldsymbol{v}) \approx \sum_{k=1}^{m} \alpha_k \, \phi_k(\boldsymbol{v}) \tag{4}$$

Apparently, the key to validate Eq. (4) is determine the coefficients $\alpha_k$'s.
Generally, the number of terms, $m$, is governed by $n^r$, where $r$ refers to the order of polynomial basis that are included. In other words, there is an exponential growth in $m$ as $r$ gets large, making high order PCE inappropriate. Second-order PCE is used later in this paper.

In addition, orthogonality of the polynomials has to be guaranteed, namely the inner product of $\phi_k$ and $\phi_l$ is always equal to 0 when $k \neq l$:

$$\langle \phi_k, \phi_l \rangle = \int_{-\infty}^{\infty} \phi_k \, (\boldsymbol{v}) \phi_l(\boldsymbol{v}) p(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} = 0 \tag{5}$$

where $p(\boldsymbol{v})$ is the weight function, generally substituted by the probability density function (PDF) of $\boldsymbol{v}$. Make an inner product of $u$ and $\phi_l$ and one can obtain

$$\langle u, \phi_k \rangle = \alpha_k \langle \phi_k, \phi_k \rangle \tag{6}$$

due to the orthogonality.

Therefore, to compute the coefficients $\alpha_k$, we only need to divide $\langle u, \phi_k \rangle$ by $\langle \phi_k, \phi_k \rangle$:

$$\alpha_k = \frac{\langle u, \phi_k \rangle}{\langle \phi_k, \phi_k \rangle} = \frac{\int_{-\infty}^{\infty} u \, (\boldsymbol{v}) \phi_k(\boldsymbol{v}) p(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}}{\int_{-\infty}^{\infty} \phi_k^2 \, (\boldsymbol{v}) p(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}} \tag{7}$$

Now that we have the coefficients, numerical integral schemes are often applied to obtain the upper part of Eq. (7). On designing the integral sites of $u(\boldsymbol{v})$, we use Sobol sequence to accomplish the above integral.

*Response Surface Method*

RSM is a traditional but quite widely-used method. The main idea is to use a sequence of designed experiments to obtain an optimal response. Box and Wilson acknowledge that this model is only an approximation, but such a model is easy to estimate and apply, even when little is known about the process[19,20]. RSM is actually an extension of least square regression. It treats $u(\boldsymbol{v})$ as a linear combination of first and second order terms of $\boldsymbol{v}$:

$$u(\boldsymbol{v}) \approx f(\boldsymbol{v}) = \boldsymbol{\beta}^{\mathrm{T}} \, \widetilde{\boldsymbol{v}} = \beta_0 + \sum_{i=1}^{n} \beta_i \, v_i + \sum_{j=1}^{n} \sum_{i=1}^{n} \beta_{ij} \, v_i v_j \tag{8}$$

where $\widetilde{\boldsymbol{v}} = [1, v_1, \cdots, v_n, v_1^2, v_1 v_2, \cdots, v_n^2]^{\mathrm{T}}$ is the augmentation of $\boldsymbol{v}$, and $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_{nn}]^{\mathrm{T}}$ is the coefficient vector. If we note matrices $\boldsymbol{V}$ and $\boldsymbol{F}$ as the samples from $D$:

$$\boldsymbol{V} = [\widetilde{\boldsymbol{v}_1}, \widetilde{\boldsymbol{v}_2}, \cdots, \widetilde{\boldsymbol{v}_d}] = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ v_{11} & v_{12} & \cdots & v_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1}^2 & v_{n2}^2 & \cdots & v_{nd}^2 \end{bmatrix}, \boldsymbol{F} = [f(\boldsymbol{v}_1), f(\boldsymbol{v}_2), \cdots, f(\boldsymbol{v}_d)]$$

Then we have

$$\boldsymbol{F} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{V} \tag{9}$$

and the least square estimation of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})^{-1} \boldsymbol{V}\boldsymbol{F}^{\mathrm{T}} \tag{10}$$

*Kriging Model*

Kriging interpolation can be expressed as a two-step process: first, the covariance structure of the samples in $D$ is determined by fitting a variogram; second, weights derived from this covariance structure are used to interpolate values for samples in $T$. The Kriging model can be considered as a parametric model combined with a random process which simulates the prediction error, denoted as Eq. (11):

$$u(\boldsymbol{v}) = f(\boldsymbol{\beta}, \boldsymbol{v}) + \epsilon(\boldsymbol{v}) \tag{11}$$

where $f(\boldsymbol{\beta}, \boldsymbol{v})$ is a combination of several basis functions; $\boldsymbol{\beta}$ is a vector containing the parameters in need; $\epsilon(\boldsymbol{v})$ is used to model the error. $\epsilon(\boldsymbol{v})$ satisfies the following properties:

$$\begin{aligned} \mathrm{E}[\epsilon(\boldsymbol{v})] &= 0 \\ \mathrm{Var}[\epsilon(\boldsymbol{v})] &= \sigma^2 \\ \mathrm{Cov}\left[\epsilon\left(\boldsymbol{v}^{(i)}\right), \epsilon\left(\boldsymbol{v}^{(j)}\right)\right] &= \sigma^2 R\left(\boldsymbol{v}^{(i)}, \boldsymbol{v}^{(j)}\right) \end{aligned} \tag{12}$$

The matrix $R(\boldsymbol{v}^{(i)}, \boldsymbol{v}^{(j)})$ is a correlation function which evaluates how close $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$ are to each other. And this function is:

1) always positive but no larger than 1;
2) negative correlated to the distance between $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$;
3) equal to 1 if and only if $\boldsymbol{v}^{(i)} = \boldsymbol{v}^{(j)}$.

In a nutshell, Kriging interpolation predicts a desired point by summing all the acquired samples based on different weights. One can infer to [21] for more detailed derivation.

*Artificial Neural Networks*

The method of artificial neural networks (ANN) was first invented in the 1940s, as an attempt to simulate the network of neurons that made up a human brain, and has been one of the main tools used in machine learning in recent years. A typical feed-forward ANN uses multiple layers of mathematical processing to make sense of the information it is fed, as is shown in Fig. 2. Here we omit the derivation and just come up with the essence that we only have to care about the input and the output and leave the calculation.

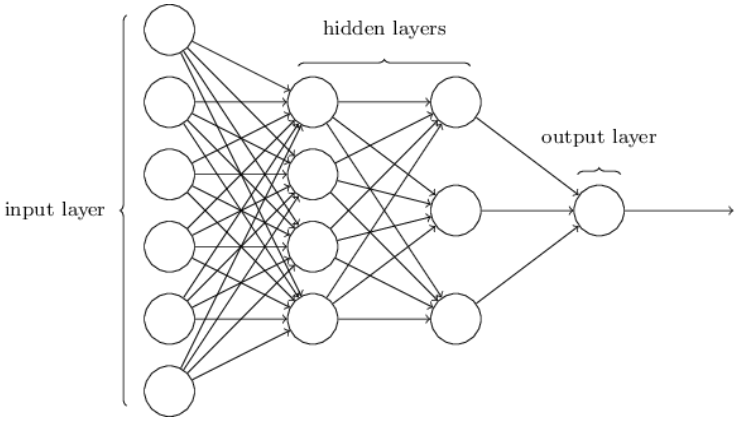



Figure 2. An example of a 4-layer feed-forward ANN

The first layer on the left in figure 2 is the input layer, which accepts the input that is fed. The first layer on the right gives the output. And the layers between, which are called hidden layers, accomplish the calculating process. In this manner, a model $u \approx f(\boldsymbol{v})$ is established, where $u$, obviously, implies the output layer and $\boldsymbol{v}$ the input layer.

Typically, it plays an important role to design the structure of an ANN and to adjust parameters. In this paper, we make use of 'Sci-Kit Learn [1]', a scientific computation package for Python to accomplish the construction, and we only focus on the performance of ANN when applied to the analysis of mistuned blisks.

## Numerical Example of a Mistuned Blisk

*Lumped Parameter Model*

The lumped parameter model of the blisk is illustrated in figure 3 with a cyclic chain of spring-mass oscillators with several degrees of freedom for each sector[22]. For each sector of the blisk, the springs are massless, $x_1$ and $x_2$ describe the vibration of blade and disk, $m_1$, $m_2$, $k_1$ and $k_2$ are the equivalent mass and stiffness of blade and disk, respectively. $k_c$ denotes the coupling stiffness between every two sectors. $c$ is the damping of the blade. Here, we set $n = 24$, $c = 0.005, f = 1, m_1 = 1, m_2 = 426, k_1 = 1, k_2 = 1.1, k_c = 493$.

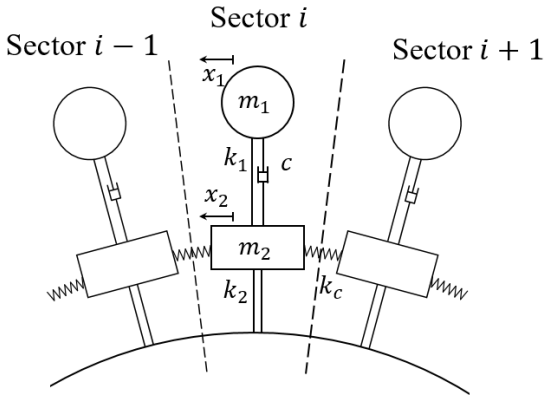



Figure 3. A lumped parameter model of a blisk

The sector stiffness matrix can be written as

$$\boldsymbol{K}_i = \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1 + k_2 + 2k_c \end{bmatrix}, \boldsymbol{K}_c = \begin{bmatrix} 0 & 0 \\ 0 & -k_c \end{bmatrix} \tag{13}$$

Therefore, for an $n$-sector blisk, its system stiffness matrix $\boldsymbol{K}$, mass matrix $\boldsymbol{M}$ and damping matrix $\boldsymbol{C}$ can be denoted as:

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_c & 0 & \cdots & \boldsymbol{K}_c \\ \boldsymbol{K}_c & \boldsymbol{K}_2 & \boldsymbol{K}_c & \cdots & 0 \\ 0 & \boldsymbol{K}_c & \boldsymbol{K}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{K}_c & 0 & 0 & \cdots & \boldsymbol{K}_n \end{bmatrix}, \boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{M}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{M}_n \end{bmatrix}, \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{C}_n \end{bmatrix}$$

where

$$\boldsymbol{M}_i = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \boldsymbol{C}_i = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix}, i = 1,2,\cdots,n$$

The equations of motion of the blisk can be written as

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{C}\dot{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{x} = \boldsymbol{F} \tag{14}$$

[1] https://scikit-learn.org/stable/

in which, $\mathbf{F} = [F_1, F_2, \cdots, F_n]^{\mathrm{T}}$ refers to the excitation vector whose $i$-th component $F_i$ excites the $i$-th sector. Usually, the excitation on each sector has the same amplitude with a fixed phase lag with the following form

$$F_i = f e^{j\psi_i} e^{j\omega t} \tag{15}$$

where $j = \sqrt{-1}$ is the imaginary unit, $f$ is the excitation force acting on the $i$-th sector, $\omega$ is the excitation frequency and

$$\psi_i = \frac{2i\pi E}{n}$$

describes the blade's relative angle where $E = 4$ is the engine order.
The forced response of the blisk can be written as

$$\boldsymbol{A} = \boldsymbol{H}^{-1}\boldsymbol{f} \tag{16}$$

where

$$\begin{aligned} \boldsymbol{H} &= -\omega^2\boldsymbol{M} + j\omega\boldsymbol{C} + \boldsymbol{K} \\ \boldsymbol{A} &= [A_1, A_2, \cdots, A_{2n}]^{\mathrm{T}} \\ \boldsymbol{f} &= f\left[e^{j\psi_1}, e^{j\psi_2}, \cdots, e^{j\psi_{2n}}\right]^{\mathrm{T}} \end{aligned} \tag{17}$$

We carried out a simple analysis of the vibration amplitudes. A tuned blisk with 24 sectors was investigated to capture the basic rules of excitation-response relations. Then we substituted the frequencies back in Eq. (21) to calculate the resonance amplitudes, shown in figure 4.

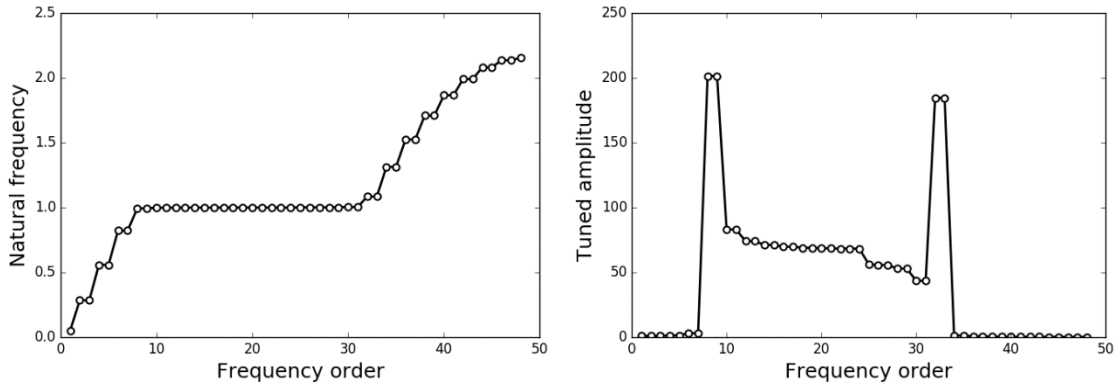



Figure 4. Natural frequencies and resonance amplitudes of the blisk

As is shown in figure 4, the largest amplitude is 201.3505, and appears when excitation frequency is 0.993. Therefore, we will set a frequency sweep near 0.993 in later sections in order to concentrate on the resonance performance of the blisk.

*Stiffness Mistuning*

One can simulate mistuning via a couple of methods, of which the simplest and most common one is to choose a different equivalent stiffness for every blade[16]. This method is widely applied in relative works because deviations among blades don't change the vibration modes.

Let $\delta_i$ be the relative variation on equivalent stiffness $k_1$ of the $i$-th blade and we can use

$$k_1^i = k_1(1 + \delta_i) \tag{18}$$

to represent the stiffness. For convenience, we write Eq. (18) as

$$k_i = k(1 + \delta_i) \tag{19}$$

Generally, we assume that $\delta_i$'s are normally i.i.d, namely

$$\delta_i \sim N(\mu, \sigma^2), i = 1,2, \cdots ,24$$

In this paper, $\mu = 0$ and $\sigma = 0.03$.

We've known that the response amplitudes vector $\boldsymbol{A}$ is influenced by the stiffness matrix $\boldsymbol{K}$, controlled by the equivalent stiffnesses $k_i$'s. So $\boldsymbol{A}$ can be seen as a stochastic process governed by $k_1 \sim k_{24}$:

$$\boldsymbol{A} = A(k_1, k_2, \cdots, k_{24}) \tag{20}$$

In the following section, we will illustrate the results of the four surrogate models, and make a comparison.

## Results and Discussions

In this section, we will build the four surrogate models based on sample set $D$ and testify the results on $T$ which contains 1000 samples.

Two issues will be considered:
1) to investigate the relation between $|D|$ and the prediction accuracy;
2) to investigate the forced responses under 5 excitation frequencies near 0.993.

In terms of evaluating the error, we care about the mean squared error (MSE) and K-S test result, which is used to assess the similarity of two probability distributions.

*Relation Between Sample Size and Accuracy*

It is commonly acknowledged that sample size $|D|$ has a profound impact on the prediction performance of models. So firstly we want to explore the relationship between prediction accuracy and $|D|$.

Setting the excitation frequency to be $\omega = 0.993$, we examine the performance of the four methods when $|D| = 50,100,200$ and 500 respectively.

Figure 5 shows the relative errors comparing the four methods with MCS results under different sample sizes. Basically, PCE generates the best result, followed by Kriging interpolation and ANN. RSM is barely of use when $|D| < 500$. The relative error of PCE and RSM prediction, as one can see, shrinks dramatically when $|D|$ increases, meaning that simply adding training samples can lead to remarkable performance promotion. The result of RSM is hardly credible when $D$ contains less samples. On the other hand, Kriging interpolation and ANN have similar results. And these two methods perform an independence on the sample size, namely, increase of $|D|$ hardly improves the performance. Further more, the prediction histograms of the four methods based on different sample sizes are plotted in figures 6 to 9.

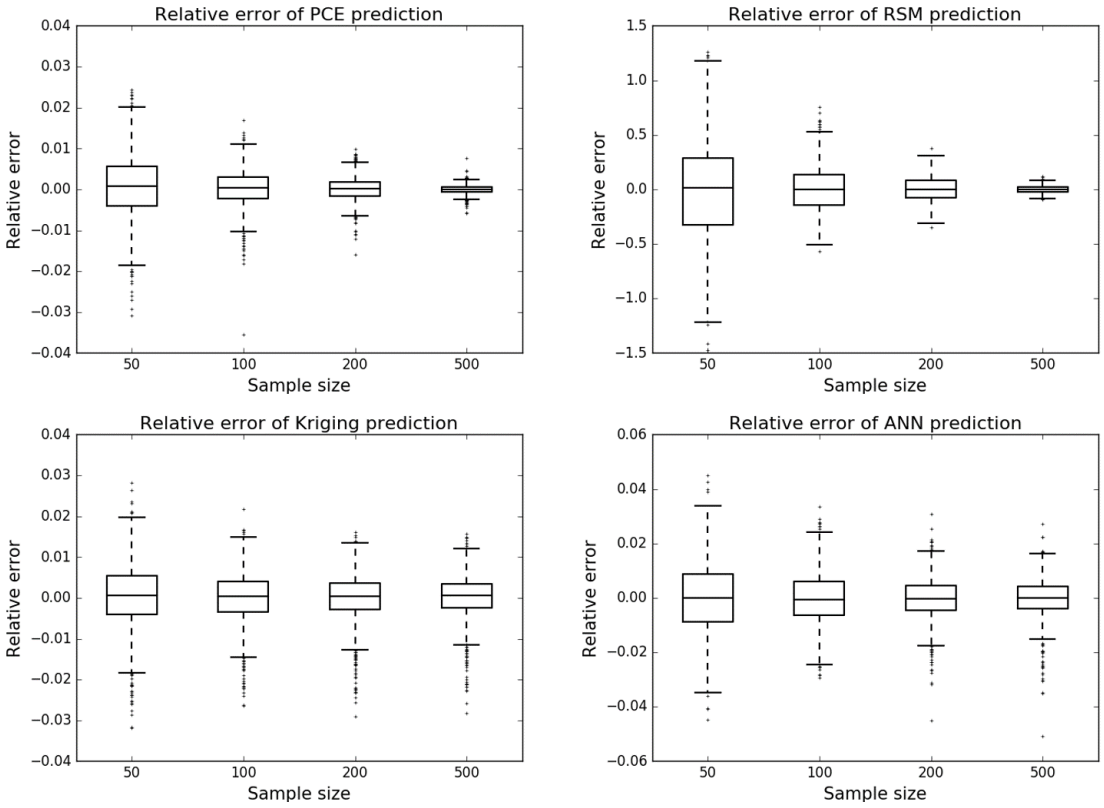



Figure 5. Boxplots of relative errors of the four methods under different sample sizes

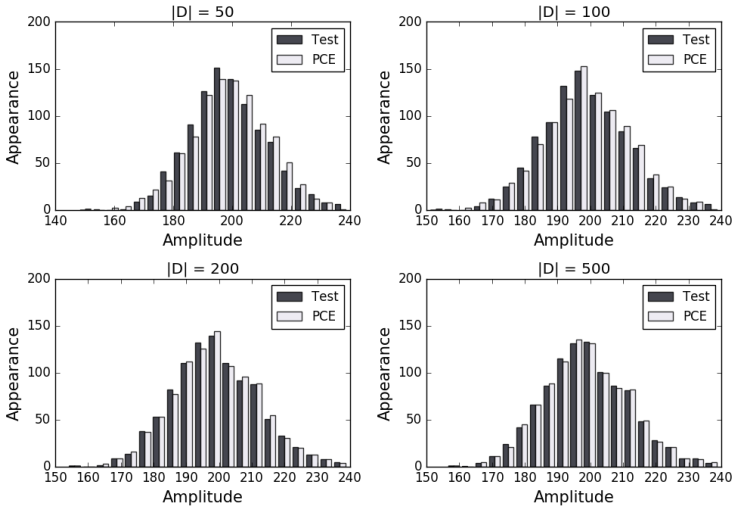



Figure 6. Histograms of PCE prediction based on four sample sizes

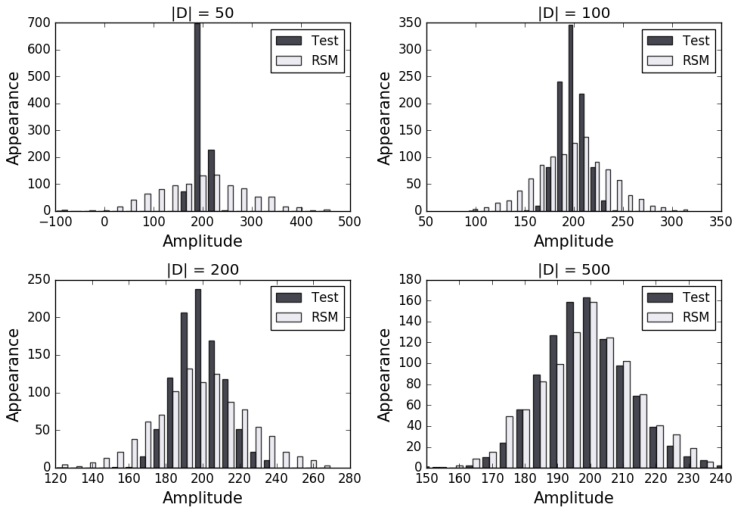



Figure 7. Histograms of RSM prediction based on four sample sizes

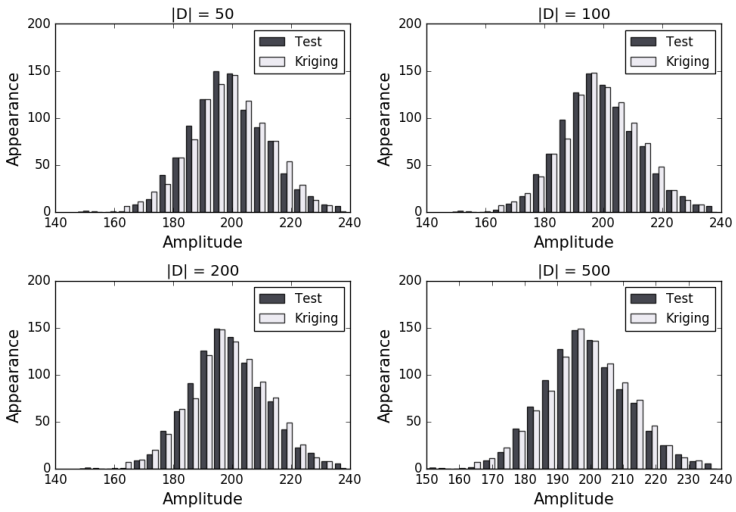



Figure 8. Histograms of Kriging prediction based on four sample sizes

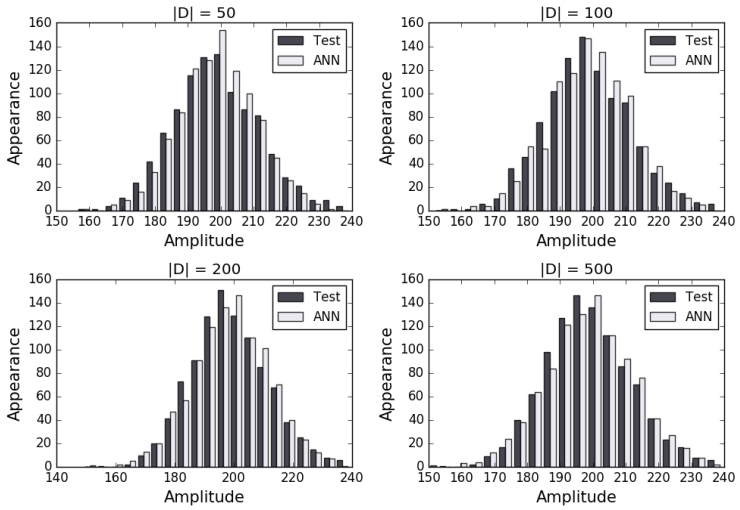



Figure 9. Histograms of ANN prediction based on four sample sizes

One can clearly see that the distribution of PCE prediction matches best with that of MCS result. When $|D| \geq 200$, the data sets look almost identical. So we will use a training set with $|D| = 200$ in later parts. Although there seems to be apparent differences in the middle parts, PCE predictions always perform well in the tail parts, namely extreme values. And extreme values should draw more attention because they are more likely to cause failure. On the other hand, Kriging interpolation and ANN also perform quite good utility. RSM hardly captures the essence of the distribution when $|D| \leq 500$.

Moreover, MSE and K-S test result of the four methods are listed in table 1, where KST represents the result of K-S test.

Table 1. Similarity of predictions and samples with different sample sizes

| Sample size | PCE | | RSM | | Kriging | | ANN | |
|---|---|---|---|---|---|---|---|---|
| | MSE | KST | MSE | KST | MSE | KST | MSE | KST |
| 50 | 6.30e-4 | 0.95 | 0.21 | 0.00 | 8.12e-4 | 0.68 | 9.62e-4 | 0.00 |
| 100 | 3.55e-4 | 0.91 | 4.36e-2 | 0.00 | 5.96e-4 | 0.91 | 1.19e-3 | 0.46 |
| 200 | 1.04e-4 | 0.99 | 1.38e-2 | 0.00 | 5.09e-4 | 0.98 | 1.54e-3 | 0.08 |
| 500 | 2.87e-5 | 0.99 | 1.43e-3 | 0.34 | 4.53e-4 | 0.88 | 7.89e-4 | 0.31 |

*Prediction under Different Frequencies*

Next, we set a frequency sweep: $\omega$ = 0.991, 0.992, 0.993, 0.994 and 0.995, trying to figure out to which frequency the blades resonate the most and whether the four methods work when $|D| = 200$. The error boxplots and the histograms are shown in figures 10 to 13, and specific statistical results are listed in table 2.

The result of PCE is shown in figure 10, containing the relative error and the histograms paired with that of MCS. Apparently, the accuracy of PCE is sufficiently guaranteed.

Figure 11 shows the prediction of RS. One can clearly see that the result is much worse than that of PCE predictions. And RSM gives some predictions that differ from test samples by over 30%, which are hard to be treated as valid results. Moreover, we can see from the histograms that RSM captures the mean values well, but it fails to predict the flank parts.

One can see from figure 12 that Kriging prediction histograms resemble those of test samples well, a little worse than PCE but far better than RSM and ANN.

In this paper, a one-hidden-layer ANN is used. The number of neurons in hidden layer is 20; activation function is radial basis function (RBF); solver algorithm is gradient descent; step size is 0.0001 and max iteration number is 1000. Above work is done in Python, "scikit-learn" package specifically. The result is shown in figure 13.
As is shown in the histograms of figure 13, ANN can yield sufficiently accurate result. In a word, this method has a promising convergence comparing to RSM when its model parameters are appropriately chosen, which often needs manual intervention, making it not as good as PCE and Kriging interpolation.

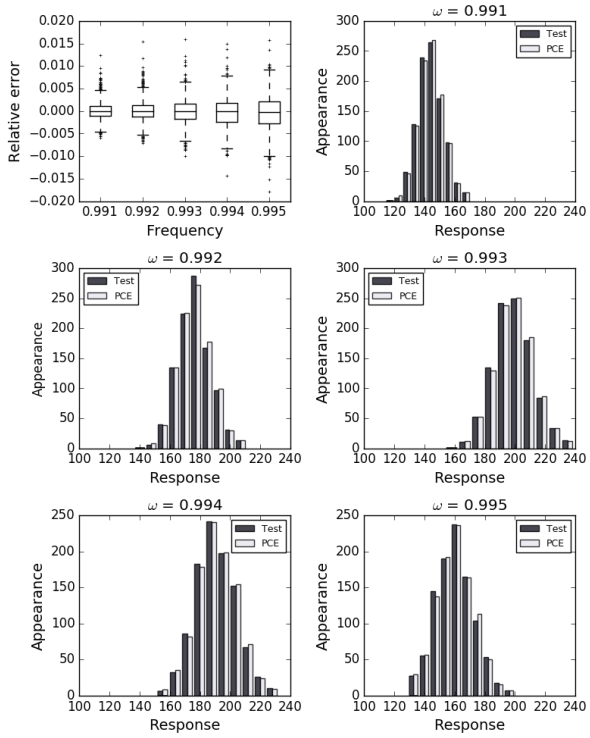



Figure 10. PCE predictions under different excitation frequencies based on 200 samples

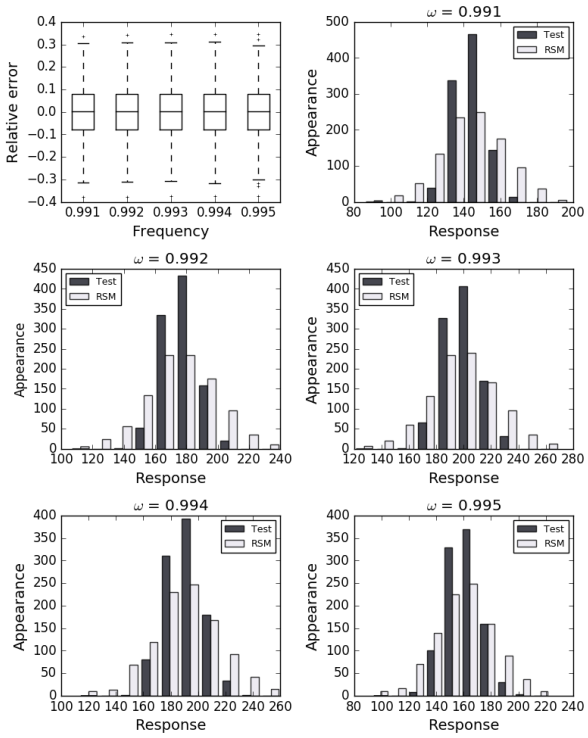



Figure 11. RSM predictions under different excitation frequencies

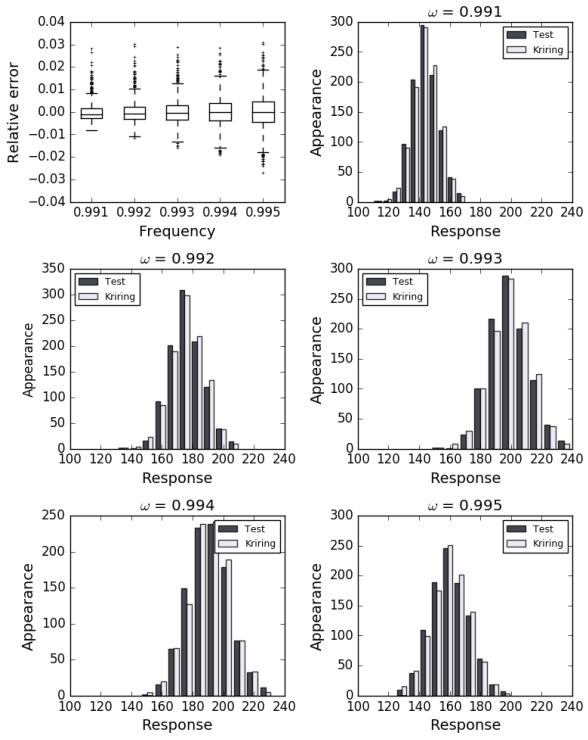



Figure 12. Kriging predictions under different excitation frequencies

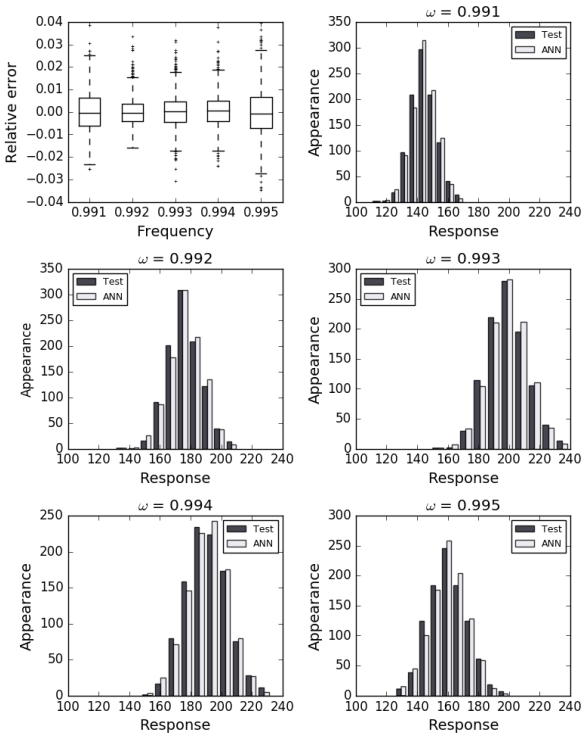



Figure 13. ANN predictions under different excitation frequencies

Table 2. Similarity of predictions and samples under different excitation frequencies

| Freq. | PCE | | RSM | | Kriging | | ANN | |
|---|---|---|---|---|---|---|---|---|
| | MSE | KST | MSE | KST | MSE | KST | MSE | KST |
| 0.998 | 6.25e-5 | 0.99 | 1.31e-2 | 0.00 | 1.13e-4 | 0.98 | 9.62e-4 | 0.08 |
| 0.999 | 1.07e-4 | 0.99 | 1.33e-2 | 0.00 | 3.39e-4 | 0.88 | 2.77e-3 | 0.60 |
| 1.000 | 1.04e-4 | 0.99 | 1.38e-2 | 0.00 | 5.09e-4 | 0.98 | 1.54e-3 | 0.61 |
| 1.001 | 9.62e-5 | 0.98 | 1.38e-2 | 0.00 | 3.36e-4 | 0.72 | 4.21e-3 | 0.00 |
| 1.002 | 5.59e-5 | 0.98 | 1.36e-2 | 0.88 | 1.31e-4 | 0.99 | 5.73e-3 | 0.00 |

Obviously, PCE provides the most valid and stable results, because MSE and KST of PCE are both the best among the four methods. And the distribution best resembles that of test set. Kriging interpolation and ANN are also remarkable on this issue. RSM fails to model the vibration properly, making it hard to be of actual use.

## Conclusion

The statistics of forced response of a mistuned blisk have been calculated via second-order PCE and three other methods. mean square error in addition with the K-S test result of predictions of the four methods have been computed and discussed. With 200 training samples, PCE can yield the most accurate and convergent result, followed by Kriging interpolation and ANN. RSM seems to be inappropriate on this issue. The accuracy of PCE and RSM increases when the sample size gets large. In addition, PCE provides a functional relationship between mistuning variables and the response amplitudes, which can be used

for further analysis on the sensitivity of amplitudes to the variables for numerical optimization.

The analysis and results presented in this paper can be easily applied to lumped-parameter models even with more degrees of freedom in future works. Only the $k_1^i$'s will have to be redefined.

**References**


[1] CASTANIER M P, PIERRE C. Modeling and analysis of mistuned bladed disk vibration: status and emerging directions. *Journal of Propulsion and Power*, 2006, 22(2): 384–396.

[2] THOMSON D E, GRIFFIN J T. The national turbine engine high cycle fatigue program. *Global Gas Turbine News*, 1999, 39(1): 14–17.

[3] DYE R, HENRY T. Vibration amplitudes of compressor blades resulting from scatter in blade natural frequencies. *Journal of Engineering for Power*, 1969, 91(3): 182–187.

[4] KAZA K R V, KIELB R E. Vibration and flutter of mistuned bladed-disk assemblies. *Journal of Propulsion and Power*, 1985, 1(5): 336–344.

[5] GRIFFIN J, HOOSAC T. Model development and statistical investigation of turbine blade mistuning. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, 1984, 106(2): 204–210.

[6] EWINS D. Vibration characteristics of bladed disc assemblies. *Journal of Mechanical Engineering Science*, 1973, 15(3): 165–186.

[7] SLATER J, MINKIEWICZ G, BLAIR A. Forced response of bladed disk assemblies-a survey. *34th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*. 1999: 3743.

[8] PETROV E P. Explicit finite element models of friction dampers in forced response analysis of bladed disks. *Journal of engineering for gas turbines and power*, 2008, 130(2): 022502.

[9] LAXALDE D, LOMBARD J-P, THOUVEREZ F. Dynamics of multistage bladed disks systems. *Journal of Engineering for Gas Turbines and Power*, 2007, 129(4): 1058–1064.

[10] KASZYNSKI A A, BECK J A, BROWN J M. Uncertainties of an automated optical 3d geometry measurement, modeling, and analysis process for mistuned integrally bladed rotor reverse engineering. *Journal of Engineering for Gas Turbines and Power*, 2013, 135(10): 102504.

[11] CASTANIER M P, OTTARSSON G, PIERRE C. A reduced order modeling technique for mistuned bladed disks. *Journal of Vibration and Acoustics*, 1997, 119(3): 439–447.

[12] LIM S-H, BLADH R, CASTANIER M P et al. Compact, generalized component mode mistuning representation for modeling bladed disk vibration. *AIAA journal*, 2007, 45(9): 2285–2298.

[13] VARGIU P, FIRRONE C M, ZUCCA S et al. A reduced order model based on sector mistuning for the dynamic analysis of mistuned bladed disks. *International Journal of Mechanical Sciences*, 2011, 53(8): 639–646.

[14] ELHAMI A, LALLEMENT G, MINOTTI P et al. Methods that combine finite group theory with component mode synthesis in the analysis of repetitive structures. *Computers & structures*, 1993, 48(6): 975–982.

[15] CHEN S, SINHA A. Calculating the statistics of the maximum amplitude of a mistuned bladed disk assembly. *33rd Structures, Structural Dynamics and Materials Conference*. 1992: 2217.

[16] SINHA A. Computation of the Statistics of Forced Response of a Mistuned Bladed Disk Assembly via Polynomial Chaos. *Journal of Vibration and Acoustics*, 2006, 128(4): 449.

[17] SINHA A. Calculating the statistics of forced response of a mistuned bladed disk assembly. *AIAA journal*, 1986, 24(11): 1797–1801.

[18] THAPA M, MULANI S B, WALTERS R W. A new non-intrusive polynomial chaos using higher order sensitivities. *Computer Methods in Applied Mechanics and Engineering*, 2018, 328: 594–611.

[19] BOX G, WILSON K. On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1951, 13(1): 1–45.

[20] JONES D R, SCHONLAU M, WELCH W J. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 1998, 13(4): 455–492.

[21] CRESSIE N. The origins of kriging. *Mathematical geology*, 1990, 22(3): 239–252.

[22] YAO J, WANG J, LI Q. Improved modal localization and excitation factors for understanding mistuned bladed disk response. *Journal of Propulsion and Power*, 2011, 27(1): 50–60.

# Numerical Study of FIO of Circular Cylinder with Different Heights of Passive Turbulence Control Strips

**Chunhui Ma and Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

## Abstract

Vortex-induced vibrations (VIV) and galloping are commonly referred to as fluid induced oscillation(FIO). For a spring-mounted rigid cylinder inflow, an isolated smooth circular cylinder could only undergo VIV but not gallop. However, non-circular sections or circular sections with attachments could be subjected to large amplitude galloping oscillations. In this paper, two-dimensional RANS equations with SST k – ω turbulence model are used to simulate the flow induced oscillation of seven single circular cylinders (P0, P1, P1.4, P2, P3, P5, P10) with different height PTC (Passive Turbulence Control) strips and supported by spring stiffness in steady uniform flow. The simulation is carried out by the in-house CFD solver naoe-FOAM-SJTU, based on the open source toolbox OpenFOAM. The oscillation response of P0 to P10 at $Re = 3.71 \times 10^4$ and $Re = 1.03 \times 10^5$ are being presented and analyzed. The amplitude ratio and frequency ratio has different trends following the height of the PTC changes. However, vibration responses of P10 are very similar regardless of the Reynolds number. Moreover, KC number as an important parameter of vortex regime in FIO, vortex structures are analyzed using vorticity images from the CFD result. Single vortex shedding structures are found in the wake as $KC < 1.8$. 3P+S vortex shedding appears when $KC > 4$.

**Keywords:** VIV; galloping; naoe-FOAM-SJTU solver; RANS; passive turbulence control

## 1 Introduction

Vortex-induced vibrations (VIV) and galloping are the two common fluid-structure interaction phenomenon which can be observed in crossflow. They are commonly referred to as fluid induced oscillation(FIO). The excitation of the VIV is caused by the alternating shedding of eddy current from each side of the cylinder. Also, time-varying forces at the frequency of vortex shedding give rise to periodic changes in the pressure distribution on the body surface. For a spring-mounted rigid cylinder, an isolated smooth circular cylinder could only undergo VIV does not gallop. However, such as triangle, square non-circular sections could be subjected to large amplitude galloping oscillations (Parkinson and Sullivan, 1979; Bokaian and Geoola, 1984b) [1][2]. And circular sections with like splitter plate attachments also could experience galloping (Nakamura et al., 1994) [3]. Furthermore, the proximity of another cylinder would also induce galloping excitation in circular cylinders (Bokaian and Geoola, 1984a) [4]. Therefore, the phenomenon of galloping is not due to the normal vortex shedding, which is an instability phenomenon caused by the motion of the shear layers on both sides of the cylinder.

Some scholars applied straight rough strips (PTC, passive turbulence control) to the surface of the cylinder to change the geometry section (Chang et al., 2010; Chang et al., 2011; Park et al., 2011) [5-7]. Width, roughness, and circumferential location on the cylinder are the parameters tested experimentally in the Water Channel located in the Marine Renewable Energy Laboratory (MRELab) of the University of Michigan. Also, many scholars are devoted to

studying the motion mechanism of passive turbulence control system and applying them to new energy fields. (Bernitsas et al., 2009; Lee et al., 2011, Sun et al.,2017; Sun et al., 2019) [8-11]

However, the replacement of different PTC strip in most experiments is very time consuming and requires recalibration, while the CFD method makes this efficient. Meanwhile, CFD make the local vortex shedding visualization, so this is easier to analyze the mechanism of FIO. In this study, two-dimensional RANS equations with SST k – ω turbulence model are used to simulate the flow induced oscillation of seven single circular cylinders (P0, P1, P1.4, P2, P3, P5, P10) with different height PTC (Passive Turbulence Control) strips and supported by spring stiffness in steady uniform flow at $Re = 3.71 \times 10^4$ and $Re = 1.03 \times 10^5$ 103000 (primarily the TrSL3 flow regime (Zdravkovich, 1997) [12]). The simulation is carried out through the CFD solver naoe-FOAM-SJTU developed by the open source toolbox OpenFOAM. The objective is to understand the impact of PTC on flow induced oscillation by changing the height of strips attached on the cylinder, and analyze vortex structures using high-resolution imaging from the CFD result. This paper first introduces the numerical methods used in naoe-Foam-SJTU solver, including turbulence simulation method, dynamic overset grid technique and spring system in section 2. Section 3 illustrates computational physical model, geometric parameters, computational domain, grid and boundary conditions. Then in section 4, the free decay of the spring system is verified, the amplitude and frequency response of seven cylinders are conduct under two Reynolds numbers to evaluate the effect of the height of PTC on the FIO. Also, vortex structures are analyzed from the CFD result. Finally, section 5 concludes this paper.

## 2 Mathematical model and numerical method

All numerical simulations in this study are based on the in-house 6DoF CFD solver naoe-FOAM-SJTU (Shen and Wan, 2015) [13], which was developed on the open-source CFD software OpenFOAM. And the dynamic overset grid technology was implemented into the solver in coupled with Suggar ++(Noack, 2005, Noack et al., 2009) [14,15] to facilitate large amplitude hull motions. The naoe-FOAM-SJTU has been validated via a majority of marine and ocean engineering problems. (Cao and Wan, 2010; Zhou et al., 2013; Wang et al., 2017) [16-18].

### *2.1 Governing equations and turbulence modeling*

The turbulent flow is a highly complex three-dimensional unsteady, irregular flow with rotation. The simulations for turbulence are currently divided into direct numerical simulation (DNS), large eddy simulation (LES), and the use of Reynolds average NS equation simulation (RANS). Because DNS and LES have high requirements on a computing resource, so its applicability to engineering problems needs to be improved. RANS equation is adopted in this study. The Reynolds average NS equation is as follows:

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}\left(\rho u_i u_j\right) = \frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_j}(2\mu S_{ij} - \rho\overline{u_j' u_i'}) \tag{2}$$

In order to solve the equations, a turbulence model is introduced to determine the Reynolds stress term $-\rho\overline{u_j' u_i'}$. The currently used turbulence models are k − ε and k − ω models. This paper uses the SSTk − ω model, which combines the advantages of the above two types of

models to ensure the accuracy and reliability of the solution near the wall, and can better solve the flow problem with negative pressure gradient. Specifically, the problems discussed in this paper require the capture of flow separation and wake near-field detailing. The corresponding mesh should satisfy the near-walled body-fitted grid placed within the viscous bottom layer and the mesh near the wall is dense. The SST k − ω model can better meet the above requirements. The equation for the SST k − ω model in the OpenFOAM solver is referred to Menter,1994. [19]

### *2.2 Overset Grid*

In this paper, the overset grid program Suggar++ is used to calculate Domain Connectivity Information (DCI). The information mainly includes unit information (hole unit, interpolation unit, contribution unit, orphan unit) and interpolation weight coefficient. The solver implements fully parallelized flow field solving and overset mesh burrow interpolation calculations by running OpenFOAM and Suggar++ in different processes. Grid motion and DCI information exchange between different processes are implemented through the Message Passing Interface (MPI).

### *2.3 Spring system*

The spring system is an important factor in studying the FIO response of a rigid cylinder with elastic support. The spring system in the experiment called Vck (Virtual spring-damping device) can realize the oscillator of the real spring by changing the parameters such as system stiffness and damping (Sun, 2015) [20]. Also, the mooring system of naoe-FOAM-SJTU is adopted to simulate the spring and more convenient to adjust the parameters (Zhao et al., 2018) [21]. Currently the mooring system supports four types of mooring line, which include linear spring, catenary, PEM (piecewise extrapolation method) and LMM (lumped mass method). Moreover, it is convenient to extend more complex mooring system and update new mooring line types based on the current framework. In the present study, all mooring lines are treated as linear springs.

## 3 Physical model and simulation design

### *3.1 Physical model*

The physical model is referred to MRELab of University of Michigan (Chang, 2011; Bernitsas, 2009) [6][8]. A simple schematic of the circular cylinder oscillation system in the present work is shown in Figure 1. The elements of this oscillator include a rigid circular cylinder of diameter *D*, length *L*, stiffness *K*, and system damping *C*. The cylinder is limited to oscillating in one direction, that is, perpendicular to the direction of incoming flow.

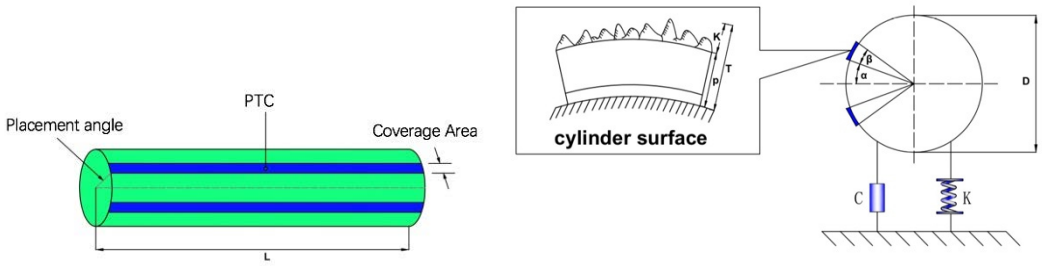



**Figure 1. Physical model**

PTC strips acts like ‘step’ are attached to the smooth cylinder, the total thickness (*T*) of PTC strips is defined as *T*=*k*+*P*, where *k* is the roughness height and p is the paper thickness.

Meanwhile $\alpha$ is the strip placement angle and value of $\beta$ indicates coverage area. The parameters for this system are listed in table 1. The roughness act as a 'catalyst' to promote galloping fully-developed in the experiment (Chang, 2010)[5], but it does not significantly influence the oscillation response, so the wall function with roughness temporarily not be considered in this paper, which means $k$ in all case should be set as zero. Table 2 shows 7 cases (P0, P1, P1.4, P2, P3, P5, and P10), which are the circular cylinder of the same diameter with various height $T$, by changing the parameter $p$. It should be noted that P0 is a smooth cylinder without PTC strip, 0.847mm is setting as P1's strip height references the experiment (Sun,2016) [20], and P1.4 to P10 respectively set the PTC height to 1.4 times to 10 times this value.

**Table 1. Parameters of the oscillatory system**

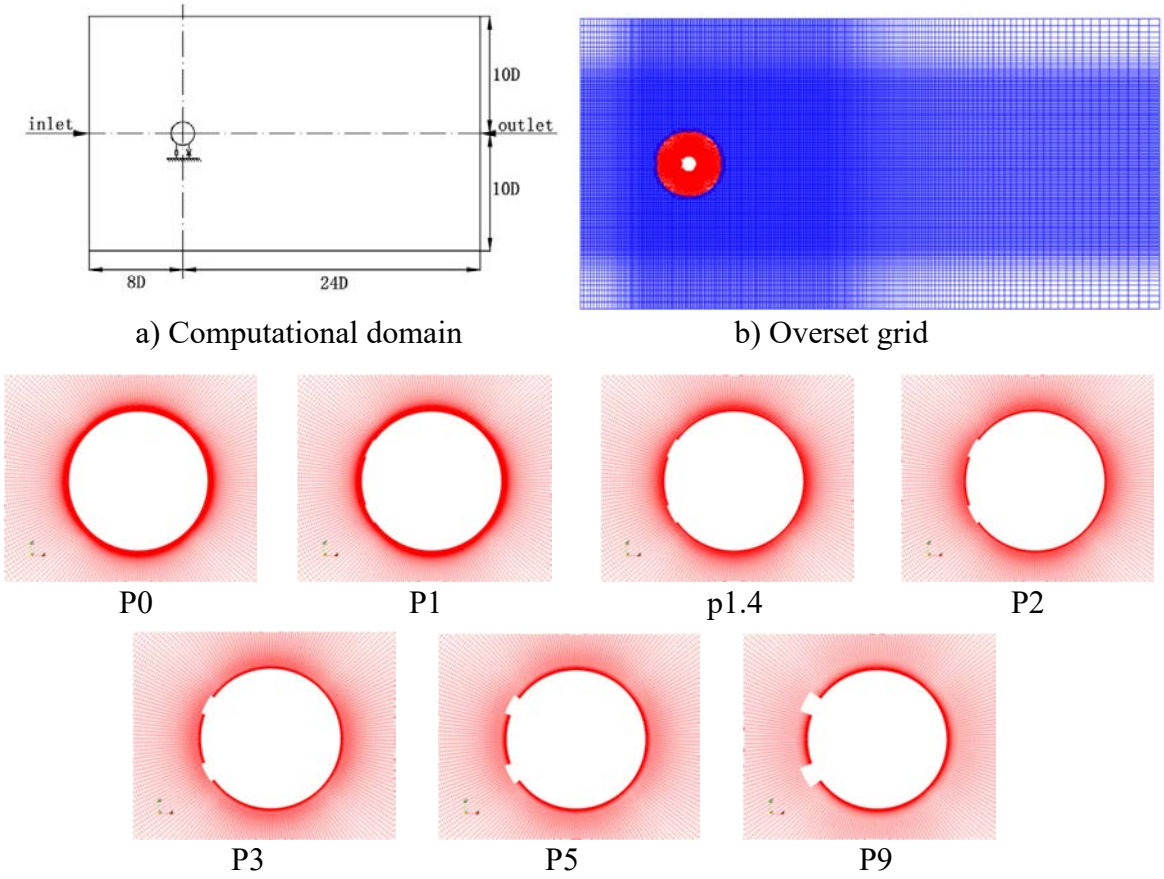
| Parameters | Symbol(unit) | Dimension |
|---|---|---|
| Diameter | $D$ [m] | 0.0889 |
| Length | $L$ [m] | 0.894 |
| strip placement angle | $\alpha$[°] | 20 |
| strip coverage angle | $\beta$[°] | 16 |
| Spring stiffness | $k$ [N/m] | 600 |
| Damping ratio | $\zeta$ | 0.02 |
| Mass | $m$ [kg] | 7.286 |
| Mass ratio | $m^*$ | 1.343 |
| Natural frequency in air | $f_{n,air}$ [Hz] | 1.44 |
| Natural frequency in water | $f_{n,water}$ [Hz] | 1.09 |

**Table 2. Different height of PTC**

| | P0 | P1 | P1.4 | P2 | P3 | P5 | P10 |
|---|---|---|---|---|---|---|---|
| T(mm) | 0 | 0.847 | 1.198 | 1.694 | 2.541 | 4.235 | 8.47 |

### *3.2 Computational Domain and Grid Generation*

Figure 3(a) shows the computational domain for all simulations in this study, which size is 20D*32D, D is the diameter of the cylinder. The center of gravity of the oscillating system is located 8D downstream of the inlet boundary condition along the longitudinal centerline of the domain. The boundary condition for velocity is set as U at inlet and zero gradient at outlet. A zero gradient boundary condition is specified for both inlet and outlet for pressure and the value is zero. The boundary condition of front and back lateral are set as empty for this case is a two-dimension simulation. The top and bottom conditions are considered as far-field boundary. For the surface of PTC as moving wall, nutUSpaldingWallFunction is a wall function that can solve any of the conditions of the near wall surface including the buffer layer, so it is selected as a wall function of moving wall.




a) Computational domain

b) Overset grid

c) close-up grid of P0 to p10

**Figure 2. Computational domain and grid**

The overset grid technique has been verified as a reliable grid application in a fixed circular flow numerical simulation within a vast Reynold number ($6.31 \times 10^4 \sim 7.57 \times 10^5$) (Ye et al, 2018) [22], which can achieve better numerical results than dynamic grid techniques. So an overset grid system is used throughout the present study to simulate the FIO response of cylinders. The overset grid system Contains two sets of independent grids, as shown in Figure 2(b), the blue grid is the background and the red part is hull grid, which are generated separately using the Pointwise grid-generating software. Then background and hull grids are merged into one set of grid in naoe-Foam-SJTU solver. The two mesh-blocks do not share any points, edges or faces. Flow information are exchanged by interpolation using domain connectivity information (DCI) generated by Suggar++. Figure 2(c) shows a close-up of a cylindrical grid of seven different PTC heights. For all P0 to P10 cases, the dimensionless wall spacing of the first layer close to the wall mesh satisfies y + <5, which makes sure the first layer cells are located in the viscous sublayer although nutUSpaldingWallFunction is applied. The total cell number is around 80,000 to 86,000. Also, the PIMPLE (merged PISO-SIMPLE) algorithm is used to solve the coupled pressure and velocity. PIMPLE treats every single time step as steady-state and performs SIMPLE correctors outside the PISO loop. It can run robustly at larger time step where Courant number is larger than one.

## 4 Results and discussion

### *4.1 Free decay*

For validate spring stiffness, free decay tests are carried out. The free decay test allows the object to oscillate according to a specified initial offset or speed without inflow. The spring system in noae-FOAM-SJTU only works effectively in the tension state, so symmetrical spring with pre-tension force is arranged along the direction of the vibration of the cylinder. The stiffness and pre-tension are adjusted to ensure that the system is always effective and accurate. The effective stiffness must be consistent with the experiment. Linear spring system test parameters refer to experimental data (sun et al, 2015) [23]. That is, mass of cylinder is 5.4878kg, mass ratio $m^* = \frac{m_{osc}}{m_{dis}} = 1.012$, $m_{dis}$ is the displacement mass of water when cylinder is merged in the channel. Spring stiffness is set to *600N/m*, damping coefficient is 0.0453, pre-tension is given 210N to ensure that the linear spring system is always tensioned as long as the amplitude of cylinder is less than 4. The parameters and results of free decay test is shown in Table 3, while Figure 5 presents the time history curve and spectrum analysis of the free decay test. The spring system and experimental error are less than 1.373%, indicating that the current spring system provides the correct effective stiffness.

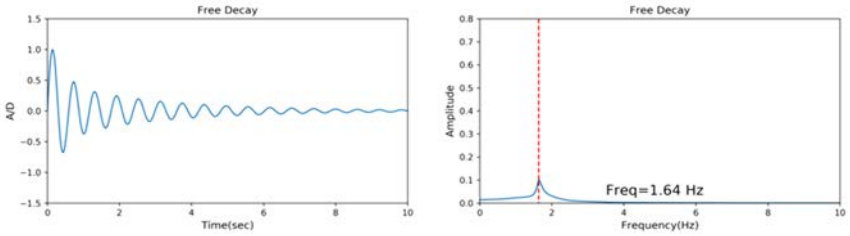



**Figure 3. Time history and spectrum of free decay**

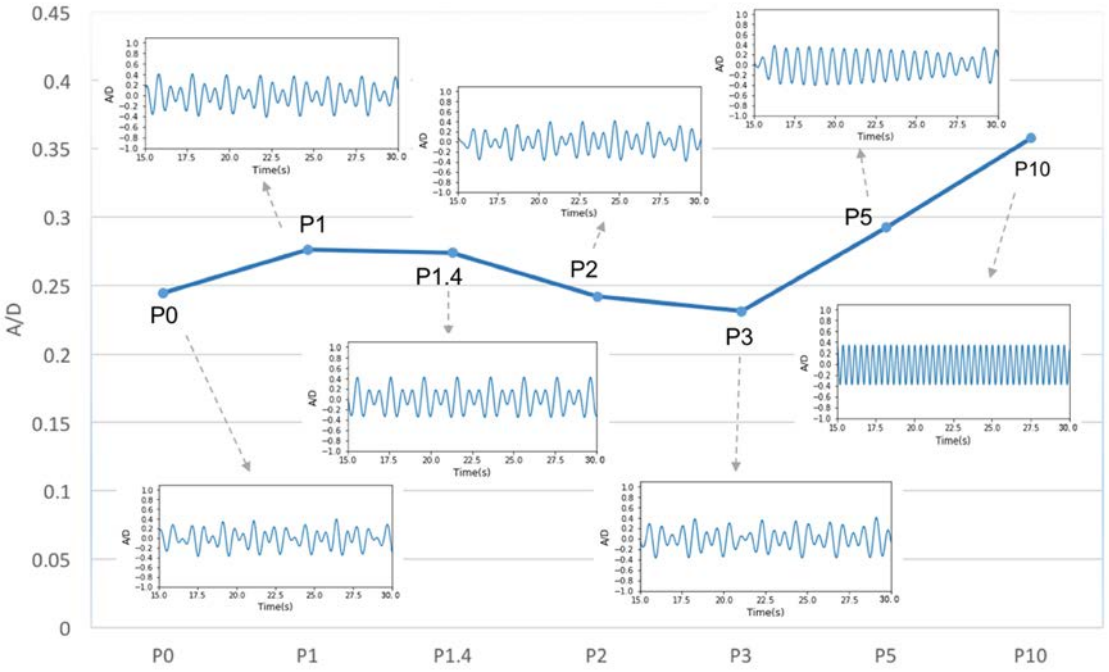
**Table 3. Free decay parameters**

| k[N/m] | Damping ratio | c(Ns/m) | m(kg) | m* | $T_{exp}$ | $T_{cfd}$ | Dev. |
|---|---|---|---|---|---|---|---|
| 600 | 0.0453 | 5.204 | 5.4878 | 1.012 | 0.6015 | 0.6098 | 0.01373 |

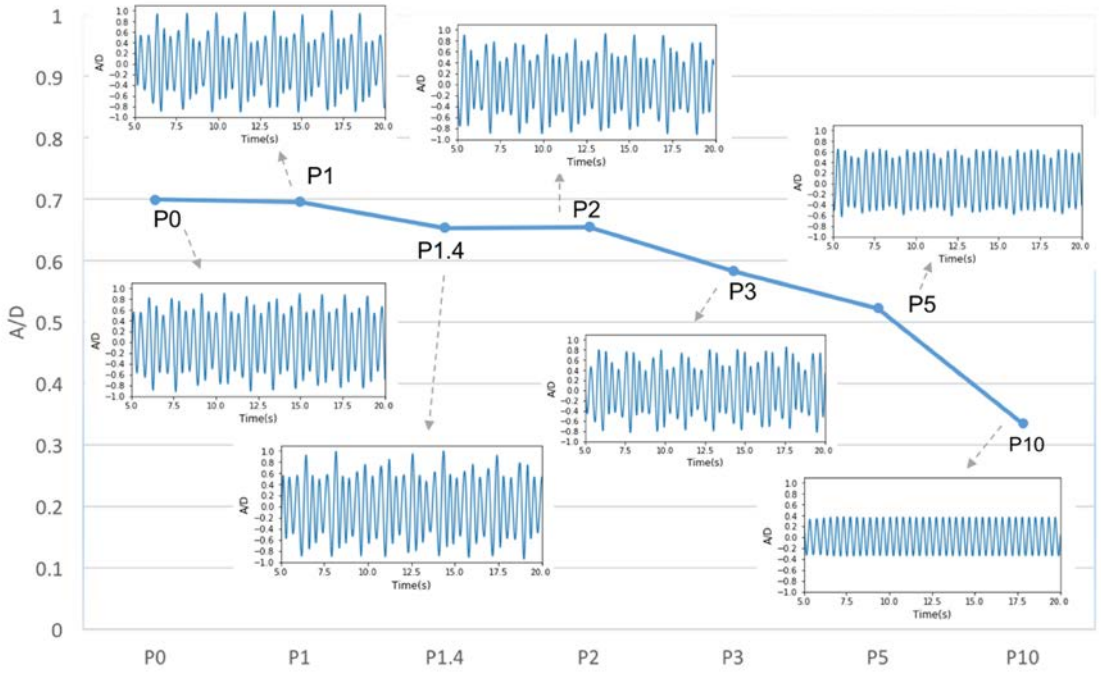
### *4.2 Amplitude response*

The amplitude of the cylinder is one of the critical properties which can adequately describe and quantify FIO. The amplitude ratios A/D of seven cases (P0 to P10) are plotted in Figure 4 at $Re = 3.71 \times 10^4$ and $Re = 1.03 \times 10^5$. The corresponding time history is also marked in the figure. The experimental values of P1 and the data of seven present numerical case are listed in Table 4. At $Re = 3.71 \times 10^4$, the amplitude ratio of P1 in the experiment is 0.352, and the value of P1 of the present study is 0.276, which is 20% smaller than experimental data. Since this is a complex flow separation of high Reynolds number, it can still be considered that the current numerical calculations and experiments consistently achieve the initial branch of VIV, even with a 20% error. Then, the amplitude ratio of the cylinders with P0 to P10 can be compared and analyzed. As can be seen from the figure4 (a), while the height of PTC increases from zero, A/D increase slightly at first, then slowly bend down at P1.4 heights. Till the height of PTC up to P3, the amplitude reaches the lowest point, after that, it tends to a steady rise again. When the PTC height is up to the highest of this group study, amplitude ratio is also up to the maximum. At $Re = 1.03 \times 10^5$, that should be a large amplitude galloping region as depicted in experimental data. However, galloping phenomenon has not appeared in any present

numerical simulation cases. As can be seen from Figure 4(b), the range A/D keep decreasing constantly from P0 to P10, except a slight increase at P2. So the higher the PTC, the more obvious the amplitude decline trend. It is revealed here that the increase in the height of the PTC possibly causes the vibration to be suppressed at $Re = 1.03 \times 10^5$.




a) $Re = 3.71 \times 10^4$




b) $Re = 1.03 \times 10^5$

**Figure 4. Amplitude ratio and displacement history**

**Table 4. Amplitude ratio**

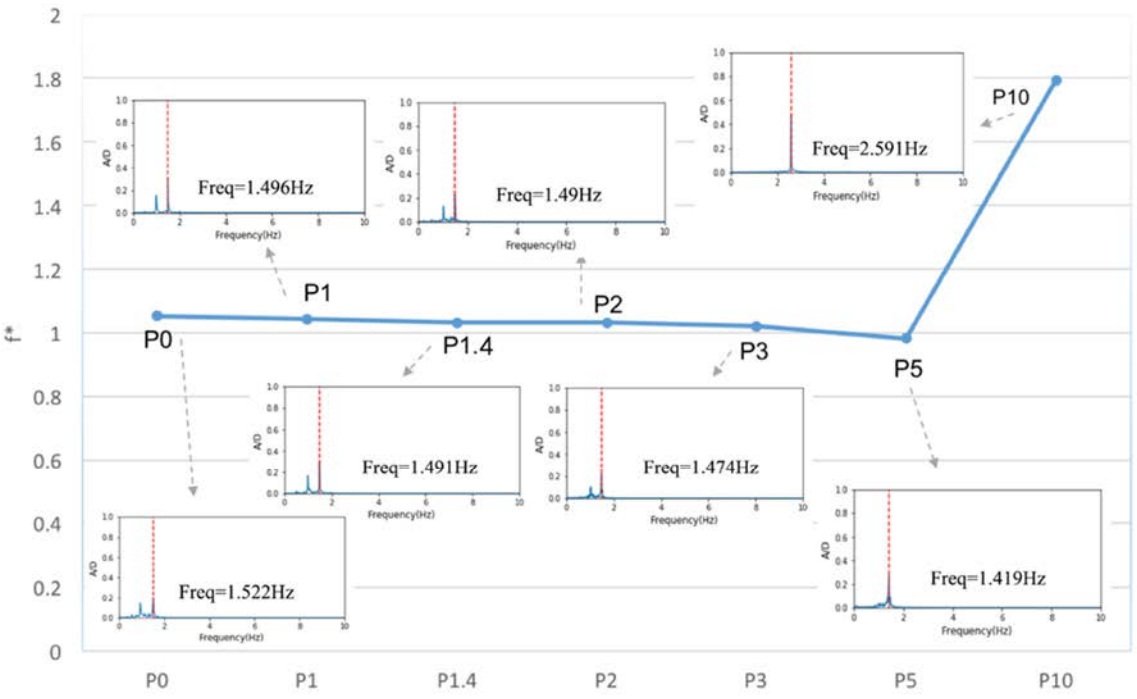
| | $Re$ | $P1_{exp}$ | P0 | P1 | P1.4 | P2 | P3 | P5 | P10 |
|---|---|---|---|---|---|---|---|---|---|
| A/D | $3.71 \times 10^4$ | 0.352 | 0.244 | 0.276 | 0.274 | 0.242 | 0.231 | 0.292 | 0.358 |
| | $1.03 \times 10^5$ | 1.662 | 0.699 | 0.696 | 0.653 | 0.655 | 0.583 | 0.522 | 0.335 |

*4.3 Frequency response*

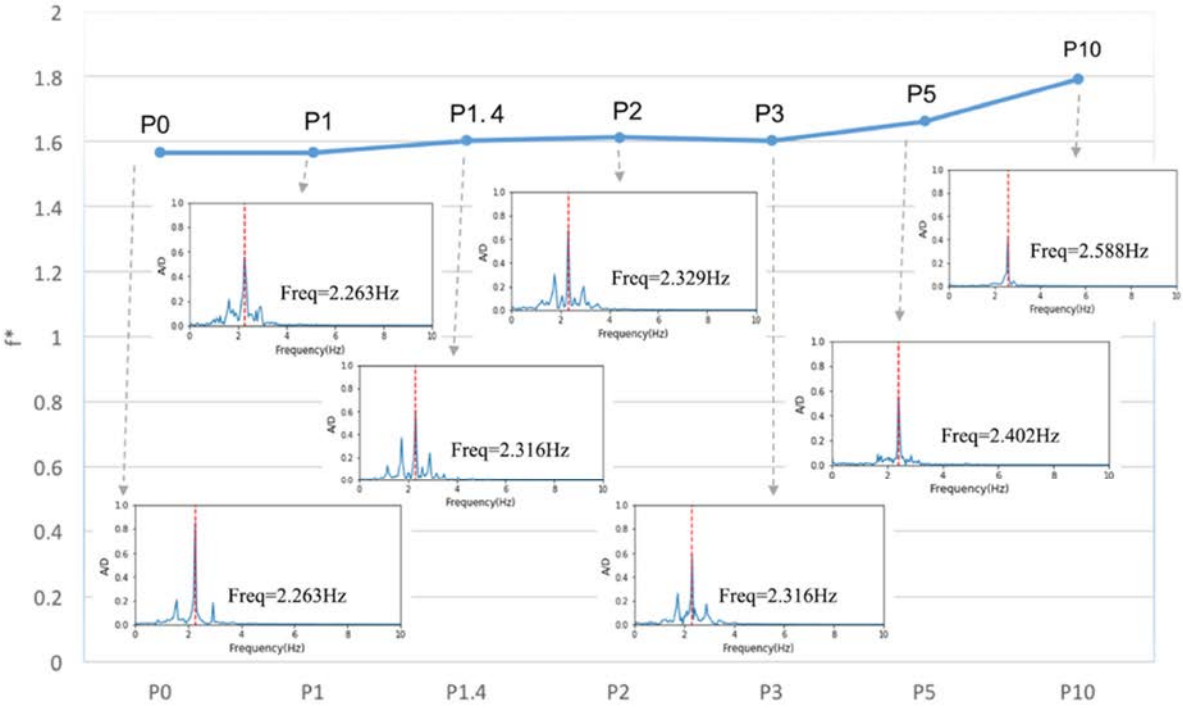
The oscillation frequency results are shown in Figure 5. Where the frequency ratio $f^* = f_{osc}/f_{n,air}$ of the cylinder is plotted vs. Reynolds. The oscillation frequency $f_{osc}$ is calculated by FFT (Fast Fourier Transform) of the time history of the cylinder over the recorded period. At $Re = 3.71 \times 10^4$, the frequency ratio is around 1 and from P0 to P5, which assumes a practically constant value, which looks like not be affected significantly by the height of the PTC except P10. From P5 to P10 height, there is a sharp increase, the change at this time is the same as the amplitude ratio. At $Re = 1.03 \times 10^5$, all cases have a very high frequency (Table 5, and Fig.5b). The data of frequency ratio appears a slow upward trend. According to the experimental result (Sun et al. 2016)[20], at the Reynolds number is higher than $10^5$, there should be low frequency and high amplitude galloping by changing the height of passive turbulence control. However, the numerical results show that the oscillation still maintains high-frequency oscillation and does not reach the instability of galloping.

**Table 5. Frequency ratio**

| | $Re$ | $P1_{exp}$ | P0 | P1 | P1.4 | P2 | P3 | P5 | P10 |
|---|---|---|---|---|---|---|---|---|---|
| f* | $3.71 \times 10^4$ | 0.649 | 1.054 | 1.043 | 1.032 | 1.032 | 1.021 | 0.982 | 1.794 |
| | $1.03 \times 10^5$ | 0.846 | 1.567 | 1.567 | 1.604 | 1.613 | 1.604 | 1.663 | 1.792 |




a) $Re = 3.71 \times 10^4$




b) $Re = 1.03 \times 10^5$

**Figure 5. Frequency ratio and spectral analysis**

*4.4 Vortex shedding*

The vortex shedding works have been considerable light on the understanding of the complex behavior of vortex motion in various regime. Williamson (1985) has described the vortex trajectory patterns in quite a systematic manner [24]. And KC number is an important parameter of vortex regime in FIO, the physical meaning can probably be best explained by reference to Equation 3, where A is the amplitude of the motion.

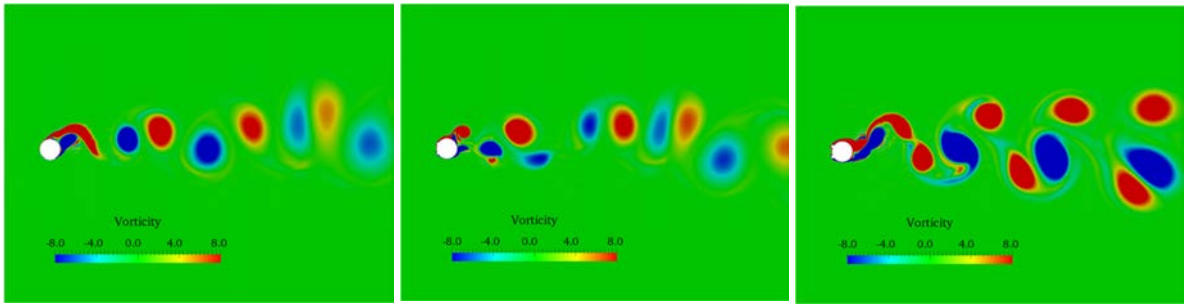
$$KC = \frac{2\pi A}{D} \tag{3}$$

**Table 6. KC number**

| | ***Re*** | P0 | P1 | P1.4 | P2 | P3 | P5 | P10 |
|---|---|---|---|---|---|---|---|---|
| *KC* | $3.71 \times 10^4$ | 1.535 | 1.735 | 1.720 | 1.519 | 1.452 | 1.837 | 2.247 |
| | $1.03 \times 10^5$ | 4.394 | 4.372 | 4.105 | 4.114 | 3.665 | 3.284 | 2.103 |

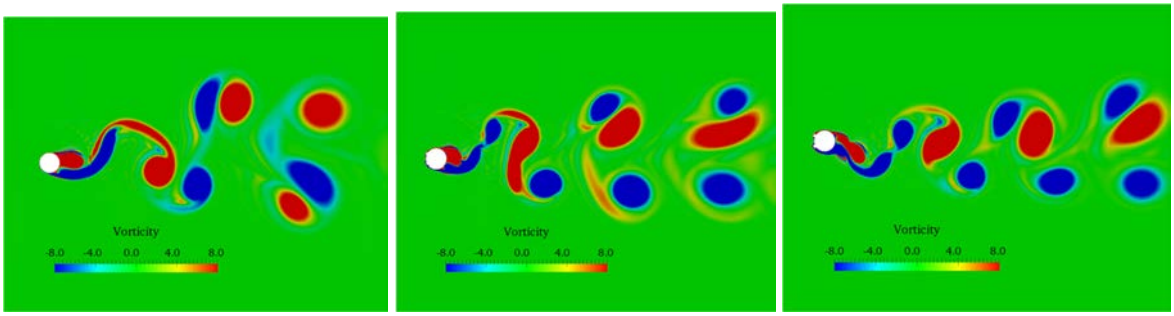
Low KC number therefore means that the orbital motion of the water particles is small relative to the total width of the cylinder D. When $KC$ number is tiny, separation behind the cylinder may not even occur. On the other hand, large $KC$ numbers mean that the water particle travels quite vast distances relative to the diameter of cylinder D, resulting in separation and probably vortex shedding (Sarpkaya (1986) and Honji (1981)) [25,26], most of their study concentrated their attention on the $KC$ number dependence or low Re number less than $10^4$. As the Reynolds number is less than $10^3$, $KC < 1.1$ is laminar flow around the cylinder. Vortex shedding appears when $KC > 7$. Moreover, vortex regime can be divided $7 < KC < 15$ (single pair), $15 < KC < 24$ (double pair), $24 < KC < 32$ (three pair), $32 < KC < 40$ (four pair), etc.

However, for higher Reynolds numbers, the relationship between kc and regime is not clearly defined. In this study, the Reynolds number exceeded $10^5$, and the $KC$ values for several cases are given in Table 5, according to the range of kc and the analysis of amplitude and frequency, since the vibration is similar, the vortex diagrams of P1, P5 and P10 under two Reynolds are given here. At $Re = 3.71 \times 10^4$, when $KC < 1.8$, only S (single vortex) shedding appears. When $KC > 1.8$, a slight pair vortex shedding can be found in Figure (b). In the case of P10, an asymmetrical vortex appears on both sides of the cylinder. From the figure, it can be seen that both P (pair vortex) and S appear simultaneously. At $Re = 1.03 \times 10^5$, when $KC > 4$ (P0, P1, P1.4, P2), vortex shedding is 3P+S, when $KC = 3.284$ and $KC = 2.103$, pair vortex and single vortex are alternately appearing regularly. Vortex shedding is suppressed and tend to single pair vortex shedding at P10.




a) P1 b) P3 c)P10

**Figure 6. Vortex pattern at $Re = 3.71 \times 10^4$**



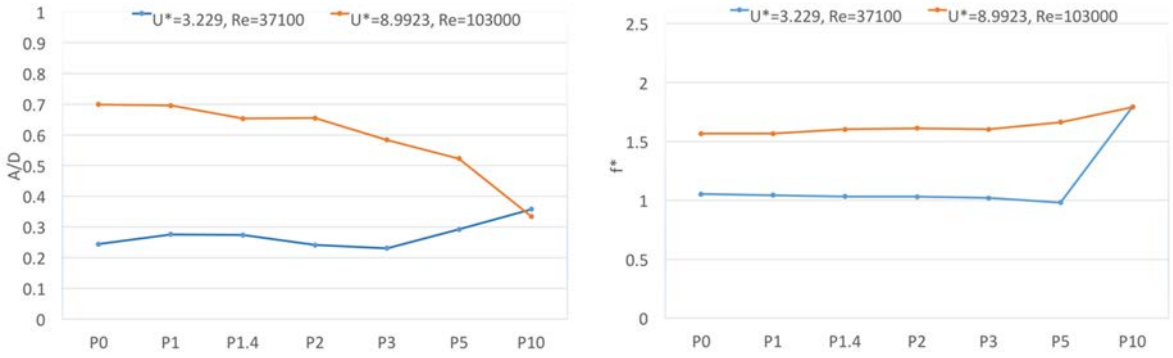

a) P1 b) P3 c)P10

**Figure 7. Vortex pattern at $Re = 1.03 \times 10^5$**

*4.5 Comparison in different flow region*

The amplitude and frequency response of $Re = 3.71 \times 10^4$ and $Re = 3.71 \times 10^4$ are put together for comparative analysis. As can be seen from the figure 8(a), under different Reynolds numbers, the amplitude ratio of P0, P1, P1.4, P2, P3 under higher Reynolds number is about twice times compared with lower Reynolds number. As the height increases, the amplitude difference between the same cylinders becomes smaller and smaller in the case of two Reynolds numbers. There is almost no difference in the amplitude of P10, which can be seen from the amplitude displacement history of Figure 4(a)(b). The frequency ratio at $Re = 3.71 \times 10^4$ *and* $Re = 1.03 \times 10^5$, as shown in Figure 8(b). Similarly, the frequency is the same at P10 case, regardless of the Reynolds number. It can be seen from the time history curve that the vibration of p10 is very stable, and the consistency also is conducted from their analysis of the vortex pattern. The vortex structure of Figures 6 and Figures 7 also illustrates the

consistency of the P10 oscillation.




a) Amplitude ratio b) Frequency ratio

**Figure 8 Amplitude ratio and frequency ratio of P10**

## 5 Conclusion

The Flow Induced Oscillation of a single, rigid, circular cylinder on linear spring with different height passive turbulence control (PTC) strips were investigated using Reynolds-Averaged Naiver-Stokes equations with $SSTk-\omega$ model at $Re = 3.71 \times 10^4$ and $Re = 1.03 \times 10^5$. All numerical simulations in this study are based on the in-house 6DoF CFD solver naoe-FOAM-SJTU, which developed by open source toolbox OpenFOAM. The following conclusions can be drawn.

(1) At $\mathrm{Re} = 3.71 \times 10^4$, the amplitude ratio and frequency ratio of P1 can realize the initial branch of VIV compared with experimental data. With the height increase of the PTC strips, oscillation amplitude first increases slowly and then show a downward trend, but rise again at p5 and p10. Moreover, oscillation frequency responds almost keep stable from P0 to P5, only with suddenly high-frequency oscillation at P10 case.

(2) At $\mathrm{Re} = 1.03 \times 10^5$, the amplitude ratio decreases as the PTC height increases. The value of amplitude at P10 is almost half of the value of P0, which illustrates PTC can suppress the oscillation in this Reynold number. At the same time, the frequency response at this Reynolds number is gradually increasing.

(3) KC numbers are introduced to analyze wake vortex shedding. Single vortex appears as $\mathrm{KC} < 1.8$. Vortex shedding is 3P+S pattern as $\mathrm{KC} > 4$. When $\mathrm{KC} = 3.284$ and $\mathrm{KC} = 2.103$, pair vortex and single vortex are alternately appearing regularly at $\mathrm{Re} = 1.03 \times 10^5$.

(4) The existence of PTC is similar to a “step”, which determines the flow separation point of the incoming stream. However, the galloping with large amplitude and lower frequency phenomenon does not found in the CFD simulation. The reason may be the FIO at high Reynolds number involves strong flow separations and RANS is not appropriate to employ for it employs statistical averaging procedure to model the mean flow quantities the turbulent fluctuations are eliminated during averaging.

The above study is the beginning of research on passive turbulence control in the flow around a cylinder. The purpose is to better analyze the oscillation response of VIV and galloping by parameter adjustment, to study how to suppress or stimulate FIO. Next, delayed DES (DDES) based on the two-equation shear stress transport (SST) model would be applied in the FIO turbulence modeling and the roughness wall function will be studied in depth.

## Acknowledgement

This work is supported by the National Natural Science Foundation of China (51879159,

51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

**References**

[1] Parkinson, G.V., Sullivan, P.P., 1979. Galloping response of towers. J. Ind. Aerodyn. 4, 253–260.
[2] Bokaian, A.R., Geoola, F., 1984b. Hydroelastic instabilities of square cylinders. J. Sound Vib. 92, 117–141.
[3] Nakamura, Y., Hirata, K., Kashima, K., 1994. Galloping of a circular cylinder in the presence of a splitter plate. J. Fluids. Struct. 8, 355–365.
[4] Bokaian, A., Geoola, F., 1984a. Wake-induced galloping of two interfering circular cylinders. J. Fluid. Mech. 146, 383–415.
[5] Chang, C.C., 2010. Hydrokinetic Energy Harnessing by Enhancement of Flow Induced Motion using Passive Turbulence Control, Naval Architecture and Marine Engineering. University of Michigan, Ann Arbor, MI, USA (Ph.D. Dissertation).
[6] Chang, C.C., Kumar, R.A., Bernitsas, M.M., 2011. VIV and galloping of single circular cylinder with surface roughness at $3 \times 10^4 \leq \mathrm{Re} \leq 1.2 \times 10^5$Ocean Engineering 2011(38), 1713–1732.
[7] Park, H., Kumar, R.A., Bernitsas, M.M., 2013. Enhancement of flow-induced motion of rigid circular cylinder on springs by localized surface roughness at $3 \times 10^4 \leq \mathrm{Re} \leq 1.2 \times 10^5$.Ocean Engineering 2011(72),403–415.
[8] M.M. Bernitsas, K. Raghavan, Converter of Current, Tide, or Wave Energy, United States Patent and Trademark Office, Patent# 7,493,759 B2 issued on February 24, 2009.
[9] J.H. Lee, M.M. Bernitsas, High-damping, high-Reynolds VIV tests for energy harnessing using the VIVACE converter, J. Ocean Eng. 38 (16) (2011) 1697e1712.
[10] Sun, H., Ma, C., Kim, E. S., Nowakowski, G., Mauer, E., & Bernitsas, M. M. (2017). Hydrokinetic energy conversion by two rough tandem-cylinders in flow induced motions: Effect of spacing and stiffness. Renewable Energy, 107, 61-80.
[11] Sun, H., Bernitsas M. M. (2019) Bio-Inspired Adaptive Damping in Hydrokinetic Energy Harnessing using Flow-Induced Oscillations. Energy,
[12] Zdravkovich, M.M., 1997. Flow Around Circular Cylinders, Volume 1: Fundamentals. Oxford University Press, England.
[13] Shen, Z., Wan, D., Carrica, P.M., 2015. Dynamic overset grids in OpenFOAM with application to KCS self-propulsion and maneuvering. Ocean Eng. 108, 287–306.
[14] Noack, R.W., 2005. SUGGAR: a general capability for moving body overset grid assembly. In: Proceedings of the 17th AIAA Computational Fluid Dynamics Conference. Toronto, Ontario, Canada, AIAA 2005–5117.
[15] Noack, R.W., Boger, D.A., Kunz, R.F., Carrica, P.M., 2009. Suggar++: an improved general overset grid assembly capability. In: Proceedings of the 19th AIAA Computational Fluid Dynamics Conference. San Antonio, Texas, USA, AIAA 2009–3992.
[16] Cao, H., Wan, D.C., 2010. Application of OpenFOAM to simulate three-dimensional flows past a single and two tandem circular cylinders. In: Proceedings of the 20th International Offshore and Polar Engineering Conference. Beijing, China, ISOPE, vol. 3. pp. 702–709.
[17] Zhou, H., Cao, H., Wan, D.C., 2013. Numerical predictions of wave impacts on the supporting structures of Shanghai Donghai-Bridge offshore wind turbines. In: Proceedings of the 23rd International Offshore and Polar Engineering Conference. Anchorage, Alaska, USA, ISOPE, vol. 1. pp. 216–224.
[18] Wang, J., Zou, L., Wan, D.C., 2017. CFD simulations of free running ship under course keeping control. Ocean Eng. 141, 450–464.
[19] F.R. Menter, Two-equation eddy-viscosity turbulence models for engineering applications, AIAA J. 32 (1994) 1598–1605.
[20] Sun, H., Kim, E. S., Nowakowski, G., Mauer, E., & Bernitsas, M. M. (2016). Effect of mass-ratio, damping, and stiffness on optimal hydrokinetic energy conversion of a single, rough cylinder in flow induced motions. Renewable Energy, 99, 936-959.
[21] Zhao, W.W., Zou, L., WAN, D.C., Hu, Z.Q, Numerical investigation of vortex-induced motions of a paired-column semi-submersible in currents. Ocean Engineering 2018(164), 272–283.
[22] Ye, H.X, Wan, D.C., Benchmark computations for flow around a stationary cylinder with high Reynolds numbers by RANS-overset grid approach. Applied Ocean Research 65(2017), 315-326
[23] H. Sun, E.S. Kim, P.M. Bernitsas, M.M. Bernitsas, Virtual spring damping system for flow-induced motion experiments, J. Offsh. Mech. Arct. Eng. 137 (6) (2015) 061801.

[24] Williamson, C.H.K. Sinusoidal flow relative to circle cylinders. J. Fluid Mech. Vol. 155, p. 141-174.
[25] Sarpkaya, T. Force on a circular cylinder in viscous oscillatory flow at low Keulegan-Carpenter number. J. Fluid Mech., 1986(a)165: 61-71
[26] Honji, H. Streaked flow around an oscillating circular cylinder. J. Fluid Mech., 1981(107): 509-520

# Thick Strip Model for Numerical Investigations on VIV of a Flexible Cylinder

**Di Deng, Lei Wu, Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

## Abstract

With the increase of oil exploitation, the aspect ratio of marine risers increases so apparently that leads to the prediction on VIV response becoming more difficult and the computational domain becoming larger. In this paper, numerical simulations on VIV of a flexible cylinder experiencing uniform and stepped flow adopting the thick strip method are conducted through the in-house viv-FOAM-SJTU solver. Hydrodynamic forces are calculated in each thick fluid strip distributed equidistantly along the cylinder. The connection among all thick strips is realized through the calculation of the structural vibration in both in-line and cross-flow directions, using the finite element method (FEM) with the Euler-Bernoulli beam model. Comparisons between experimental results and present studies show good agreement, which represents the feasibility of the thick strip method in solving VIV response of flexible cylinder. Meanwhile, the apparent three-dimensional effect is captured. It can also be found that vortex shedding features vary at each thick strip along different locations of the cylinder.

**Keywords**: vortex-induced vibration; thick strip method; flexible cylinder

## 1. Introduction

Alternant vortex shedding phenomenon will happen when viscous flow goes through the circular cylinder, which contributes to the vortex-induced vibration problem especially for the offshore structures and risers. VIV of flexible cylinders has been extensively studied during the past decades. Overviews on VIV researches can be referred to Sarpkaya (2004), Huang et al (2009), Chen et al (2016) and Wan et al (2017). In order to solve the problems of high costs of computational resources and long computing time, the traditional strip method, which considered that the fluid flow was locally two-dimensional without spanwise correlation and simplified the three-dimensional fluid field into several two-dimensional strips, was proposed and used to predict the VIV response of risers by Willden and Graham (2001, 2004). Adopting the strip method, Duan et al (2016, 2018) developed the viv-FOAM-SJTU solver and carried out numerical simulations on VIV of a vertical riser exposed to the stepped current basing on the experiments of Chaplin et al (2005). Fu et al (2016) conducted parameter researches on VIV of a flexible cylinder using the viv-FOAM-SJTU solver. Fu et al (2018) then further expanded the capacity to simulations of a flexible cylinder experiencing oscillatory flow and validate its reliability with standard model experiments of Fu et al (2013).

In order to solve the drawback in simulating the axial three-dimensional correlation of vortex shedding, the three-dimensional simulation method was adopted and used by researchers in predicting VIV response. Holmes et al (2006) combined three-dimensional CFD solutions

with structural models in simulating VIV of a straked riser. Huang et al (2007a, 2007b) presented a CFD approach for riser VIV prediction using the URANS method on an three-dimensional overset grid system. Numerical results on VIV of a flexible riser experiencing uniform flow and sheared flow were in good agreement with experimental results and previous publications. Wang and Xiao (2016) adopted the large eddy simulation (LES) method with the Arbitrary Lagrangian-Eulerian (ALE) scheme in simulating VIV response of a riser in uniform flow and sheared flow respectively using the ANSYS MFX multi-field solver.

Although the three-dimensional numerical simulation method can cover disadvantages in simulating flow fields comparing with the traditional strip method, the cost of computational resources are extremely larger especially for detailed flow fields simulation. Combining advantages of the axial correlation for the three-dimensional method and the low computational resource cost for the traditional strip method, Bao et al (2016) proposed the thick strip model for VIV prediction of long flexible cylinder using the direct numerical simulation (DNS) method. Comparisons between experimental results and simulation results validate the validity of the thick strip method.

In this paper, the modified viv-FOAM-SJTU solver is used to predict VIV response of a flexible cylinder exposed to uniform flow and stepped flow respectively basing on the thick strip method. This paper is organized as follows: The first section gives a brief introduction to the referenced experiments and the numerical methodology. The second section presents the results and the final section concludes the paper.

## 2. Method

### 2.1 Thick Strip Method

The schematic of the strip theory is shown in Fig.1. The flow field is discreted into several two-dimensional fluid strips equal-distantly located along the span of the cylinder. The strip method owns high computational efficiency and accuracy that has been verified through previous researches, such as Meneghini et al (2004) and Yamamoto et al. (2004). The original viv-FOAM-SJTU solver developed basing on the OpenFOAM and the traditional strip method and detailedly validated by Duan et al (2016). However, the simplification of ignoring the spanwise correlation will lead to the lower prediction of hydrodynamic forces in both inline and crossflow directions. Then, the predicted vibration amplitude will be smaller than that of the experimental results, especially in high Reynolds Number conditions with non-negligible three-dimensional characteristic of vortex shedding along the cylinder span. Bao et al(2016) proposed the generalized thick strip modelling method considering the spanwise correlation in the flow field locally. And this method has been validated in simulating VIV responses of a flexible cylinder experiencing the uniform flow by Bao et al(2019). In this paper, the thick strip model has been used to modify the original viv-FOAM-SJTU solver through transforming the two-dimensional fluid strips into three-dimensional thick fluid strips. While the correlation of fluid strips are realized through the vibration of the cylinder using the Finite Element Method (FEM).

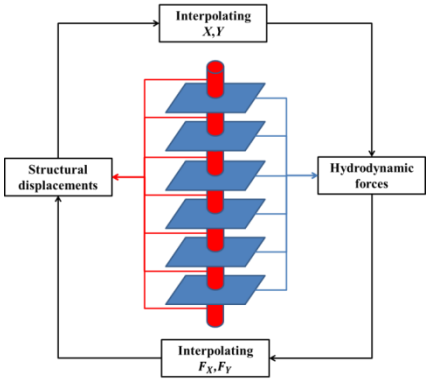



Fig.1 Schematic diagram of the strip theory

## 2.2 Hydrodynamics Governing Equations

The Reynolds-averaged Navier-Stokes equations (RANS) are used as the hydrodynamics governing equations in all thick flow strips as shown in Eq(1) and Eq(2). The flow field is supposed to be incompressible, with constant dynamic viscosity $\mu$ and constant density $\rho$.

$$\frac{\partial \overline{u}_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial}{\partial t}\left(\rho\overline{u}_i\right)+\frac{\partial}{\partial x_j}\left(\rho\overline{u}_i\overline{u}_j\right)=-\frac{\partial \overline{p}}{\partial x_i}+\frac{\partial}{\partial x_j}\left(2\mu\overline{S}_{ij}-\rho\overline{u_j'u_i'}\right) \tag{2}$$

where $\overline{S}_{ij}=\frac{1}{2}\left(\frac{\partial \overline{u}_i}{\partial x_j}+\frac{\partial \overline{u}_j}{\partial x_i}\right)$ is the mean rate of strain tensor, $-\rho\overline{u_j'u_i'}$ is referred as Reynolds stress $\tau_{ij}$ computed through $\tau_{ij}=-\rho\overline{u_j'u_i'}=2\mu_t\overline{S_{ij}}-\frac{2}{3}\rho k\delta_{ij}$, where $\mu_t$ is the turbulent viscosity and $k=(1/2)\overline{u_i'u_i'}$ is the turbulent energy, computing from the fluctuating velocity field.

## 2.3 Structural Dynamic Governing Equations

Each two-dimensional fluid strip is independent with the connection between all strips realized through the in-line and cross-flow vibration of the cylinder. The flexible cylinder is simplified to be an Euler-Bernoulli bending beam model. The vibration of the model is solved through the FEM method and the structural governing equations in each element are shown in Eq(3) and Eq(4).

$$m\ddot{x}+c\dot{x}+kx=f_x \tag{3}$$

$$m\ddot{y}+c\dot{y}+ky=f_y \tag{4}$$

where $m$, $c$, $k$ are the mass, the damping and the stiffness of the structural element; $f_x, f_y$ are the in-line and the cross-flow hydrodynamic forces respectively. Hence, the mass-spring-damping (MCK) equations of the system can be expressed as Eq(5) and Eq(6):

$$\mathbf{M}\{\ddot{\mathbf{X}}\}+\mathbf{C}\{\dot{\mathbf{X}}\}+\mathbf{K}\{\mathbf{X}\}=\{\mathbf{F_{HX}}\} \tag{5}$$

$$\mathbf{M}\{\ddot{\mathbf{Y}}\}+\mathbf{C}\{\dot{\mathbf{Y}}\}+\mathbf{K}\{\mathbf{Y}\}=\{\mathbf{F_{HY}}\} \tag{6}$$

where **M**, **C**, **K** are the mass, damping and stiffness matrixes of the system; $\{\mathbf{F_{HX}}\},\{\mathbf{F_{HY}}\}$ are hydrodynamic force vectors in the in-line and cross-flow directions. While, the Rayleigh damping is adopted to generate the damping matrix replacing the practical damping as shown in Eq(7) and Eq(8).

$$\mathbf{C}=a_0\mathbf{M}+a_1\mathbf{K} \tag{7}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{2\varsigma}{f_{n1} + f_{n2}} \begin{bmatrix} 2\pi f_{n1} f_{n2} \\ 1/(2\pi) \end{bmatrix} \tag{8}$$

where $a_0$ and $a_1$ are proportionality coefficient; $\varsigma$ is the damping ratio; $f_{n1}$ and $f_{n2}$ are the first two order of natural frequencies of the cylinder.

### 2.4 Problem Description

In the present study, two types of flow conditions are mainly considered for VIV of a vertical flexible cylinder, such as uniform flow and stepped flow. For the uniform flow condition, the model experiments were conducted at the MARINTEK by ExxonMobil (Lehn, 2003). Main parameters of the model cylinder and the flow condition are listed in Table 1. Totally, 10 thick strips are equal-distantly located along the cylinder.

For the stepped flow condition, model cylinder experiments are carried by Chaplin et al (2005a, 2005b). Main parameters and the selected flow condition are listed in Table 2. In the stepped flow condition, only the lower 45% part of the cylinder is immersed in the uniform flow with the other part being in the still water. 20 thick strips are adopted to generate the computational model, so that the lower 9 strips set the uniform flow condition and other strips set the still water condition.

Table 1 Main parameters of model cylinder in uniform flow condition

| Properties | Values | Unit |
|---|---|---|
| *L* | 9.63 | m |
| *D* | 20 | mm |
| *EI* | 135.4 | $Nm^2$ |
| *T* | 817 | N |
| *m** | 2.23 | - |
| *L/D* | 481.5 | - |
| *U* | 0.2 | m/s |

Table 2 Main parameters of model cylinder in stepped flow condition

| Properties | Values | Unit |
|---|---|---|
| *D* | 0.028 | m |
| *L* | 13.12 | m |
| *L/D* | 469 | - |
| *EI* | 29.88 | $Nm^2$ |
| *m** | 3.0 | - |
| *T* | 405 | N |
| *U* | 0.16 | m/s |

## 3. Results

### 3.1 Vibration response

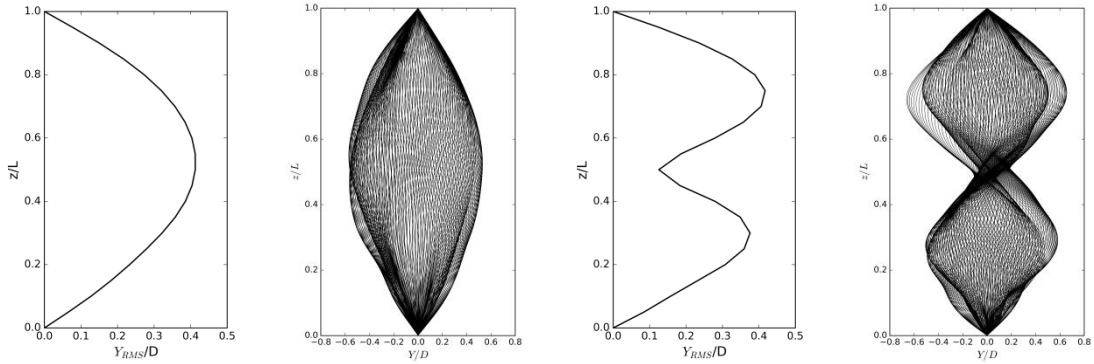
For the uniform flow condition, CFD simulation is conducted using a computational mesh with 7.9 million cells. The maximum crossflow root-mean-square (RMS) response amplitude comparison among experiment (Lehn, 2003), previous simulation of Wang and Xiao (2016) and present simulation is shown in table 3. It can be concluded that both the maximum crossflow RMS amplitude and the corresponding axial location are in good agreement, which validate the validity of the modified viv-FOAM-SJTU solver in predicting VIV response of flexible cylinder. For the stepped flow condition, the maximum crossflow vibration amplitude

is around 0.7D in present simulation, while the experimental result is around 0.75D and the corresponding computational error is acceptable.

Crossflow spatial RMS amplitude and the corresponding response envelops in both flow conditions are shown in Fig.2. The crossflow vibration is dominated by the first mode in the uniform flow and the maximum vibration amplitude locates at around the mid-span of the cylinder as shown in Fig.2(a) and Fig.2(b). While the dominant vibration mode presents the second mode for the stepped flow condition in Fig.2(c), which shows good agreement with the experimental results by Chaplin et al (2005). The peak points of the spatial vibration along the cylinder span locate at 0.25$L$ and 0.75$L$ respectively as shown in Fig.2(d), where $L$ is the total length of the cylinder.

Table 3 Maximum RMS amplitude comparison for the uniform flow condition

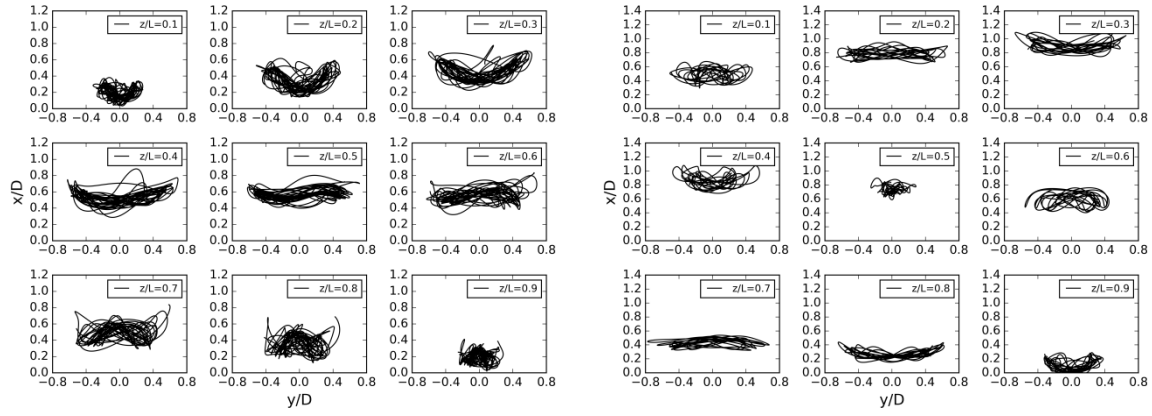
| | CF RMS | z/L |
|---|---|---|
| Lehn (2003) | 0.408 | 0.549 |
| Wang and Xiao (2016) | 0.4 | 0.523 |
| Present simulation | 0.417 | 0.552 |




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

Fig.2 RMS amplitude and response envelopes of the cylinder: (a) uniform flow (b) uniform flow (c) stepped flow (d) stepped flow

Vibration trajectories of specific nodes along the cylinder span are presented in Fig.3. It can be known that the vibration trajectory presents an approximate 'V' type at $z/L$=0.1. With the increase of the axial location from the bottom end to the top end, the vibration trajectory turns to be thinner in the inline direction ($x$ direction). The vibration trajectory shape changes to the converse 'V' type at $z/L$=0.8.




(a) (b)

Fig.3 Vibration trajectories along the cylinder span: (a) uniform flow (b) stepped flow

This variation tendency can owe to the interaction between the crossflow vibration and the inline vibration. When the crossflow vibration of lower nodes on the cylinder reaches its maximum value, the corresponding inline vibration amplitude reaches its maximum value at the same time. While, upper nodes on the cylinder present opposite phase between the crossflow vibration and the inline vibration contributing to the converse 'V' type of vibration trajectory. As for the middle node on the cylinder, the standing wave point appears that leads to the minimum vibration value and the thinnest trajectory thickness in the inline direction.
For the stepped flow condition, the vibration trajectory presents the reverse 'U' type at $z/L$=0.1, 0.2, 0.6, 0.7 and the 'U' type at $z/L$=0.3, 0.4, 0.8, 0.9. This variation tendency is similar to that of the uniform flow condition owing to the second mode vibration in the crossflow direction and the fourth vibration mode in the inline direction. The maximum inline vibration locates at around $z/L$=0.3. Both the crossflow thickness and the inline thickness of the vibration trajectory turn to the minimum value at the mid span cylinder owing to the existence of standing wave points in both directions.

### 3.2 Modal response

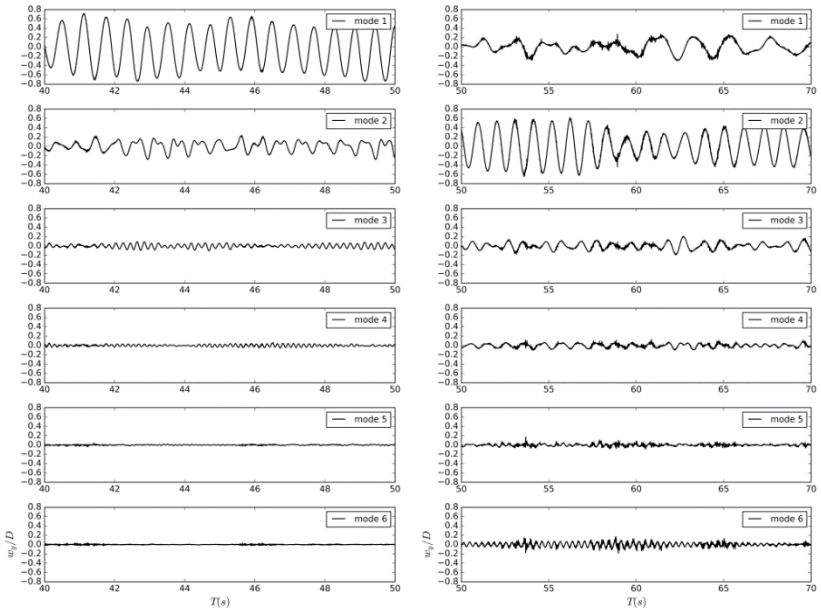
In order to get the vibration features of the flexible cylinder, the modal decomposition method is adopted, which has been verified to be available by Chaplin et al (2005). The cross-flow and in-line time varying shape of the cylinder can be expressed as the sum of a series of mode shapes as followed:

$$\varphi_n(z) = \sin(\frac{n\pi}{L} z) \tag{9}$$

$$x(z,t) = \sum_{n=1}^{N} u_n(t) \cdot \varphi_n(z) \tag{10}$$

$$y(z,t) = \sum_{n=1}^{N} v_n(t) \cdot \varphi_n(z) \tag{11}$$

where $z$ is the node location along the flexible cylinder span; $L$ is the length of the cylinder; n=1, 2, 3, etc; $u_n(t)$ is the time-dependent modal weight in the in-line direction; $v_n(t)$ is the time-dependent modal weight in the cross-flow direction; $N$ is the mode number.



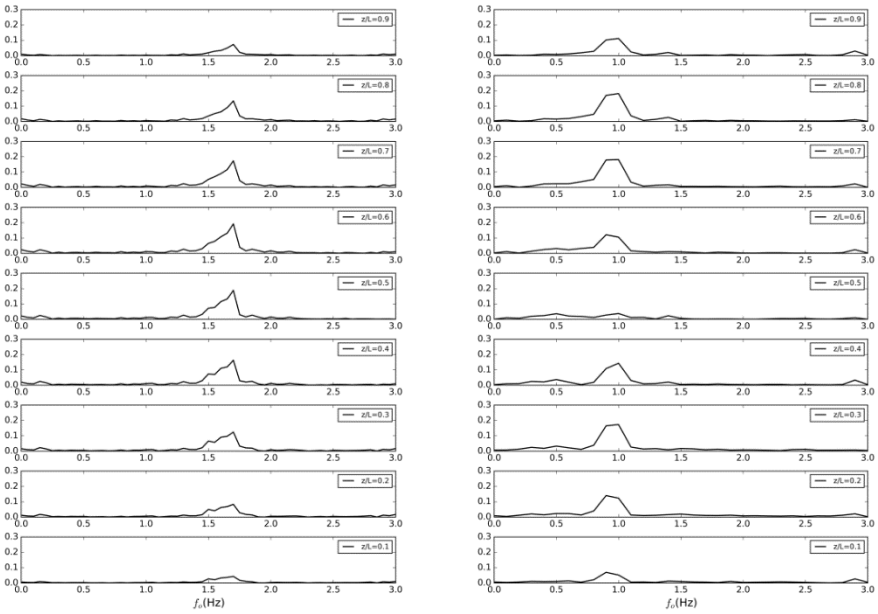

(a) (b)

Fig.4 Modal weight of the cylinder vibration: (a) uniform flow (b) stepped flow

Modal weights of the first six order of vibration mode in both flow conditions are shown in Fig.4. It can be known that the dominant vibration mode presents the first mode for the uniform flow condition. Although the second vibration mode is comparatively apparent comparing with other higher modes, it is still unable to transit the vibration of the cylinder to the second mode by the controlling effect of the first mode as shown in Fig.4 (a). As for the stepped flow condition, the dominant vibration mode of the cylinder presents the second mode with effects of other modes on the vibration can be neglected. The mode decomposition results agree well with the experimental results of Lehn(2003) and Chaplin et al(2005) respectively, which also correspond with the response envelopes of the cylinder as shown in Fig.2.

### 3.3 Frequency response

In order to analyze the frequency feature of the crossflow vibration, the Fast Fourier transform (FFT) method is adopted to calculate the frequency components at specific nodes along the cylinder as shown in Fig.5. From Fig.5 (a), it can be seen that the dominant vibration frequency of nodes along the cylinder is around 1.7Hz in the uniform flow condition. And the vibration response is more conspicuous when close to the mid-span of the cylinder, corresponding to the place where drastic crossflow VIV phenomenon happens. In the stepped flow condition as shown in Fig.5 (b), the dominant vibration frequency is around 0.9Hz. The apparent crossflow VIV phenomenon happens at around $z/L$=0.3 and 0.7 where peak and points appears as shown in Fig.2 (d). The generation of the standing wave point at $z/L$=0.5 leads to the reduction of VIV feature comparing with that of other nodes. The variation of the vibration frequency response along the cylinder is quite similar to that of the amplitude response.



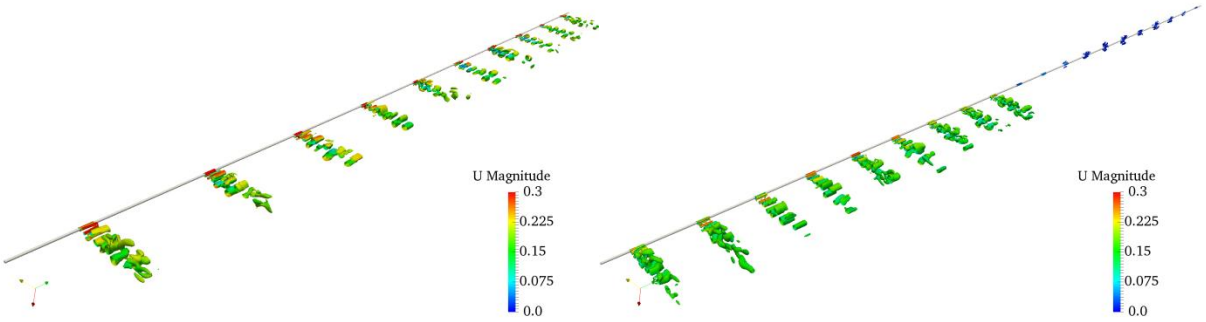

(a) (b)

Fig.5 Vibration frequency along the cylinder span: (a) uniform flow (b) stepped flow

### 3.3 Wake fields

Adopting the flow visualization method, wake flow contours along the cylinder in different flow conditions are shown in Fig.6. On account of the crossflow and the inline vibration, the

wake field turns to be chaotic and shows apparent three-dimensional features. The axial three-dimensional feature becomes more drastic at location close to the mid-span of the cylinder in the uniform flow condition where maximum crossflow vibration amplitude appears. In the stepped flow condition, apparent three-dimensional wake fields appear in the lower 9 strips where vortex shedding direction is the same as the flow speed direction. As for the upper strips that locate in the still water, vortices generate due to the vibration of the cylinder. Apparent vortex shedding phenomenon happens at around $z/L$=0.7 where maximum crossflow vibration amplitude occurs. And the vortex shedding direction is the same as the crossflow vibration direction.




(a) (b)

Fig.6 Wake field contour: (a) uniform flow (b) stepped flow

## 4. Conclusions

In this paper, the thick strip method is adopted to modify the original viv-FOAM-SJTU solver based on the two-dimensional strip method. Two simulations based on model experiments of VIV of a flexible cylinder in uniform flow (Lehn, 2003) and stepped flow (Chaplin, 2005) are conducted using the modified solver. Simulation results are in good agreement with experiment results, which verifies the validity of the thick strip method in predicting VIV response.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

## Reference

[1] Bao Y., Palacios R., Graham M., et al. Generalized thick strip modelling for vortex-induced vibration of long flexible cylinders[J]. Journal of Computational Physics, 2016, 321(800), 1079–1097.
[2] Bao Y., Zhu H.B., Huan P., et al. Numerical prediction of vortex-induced vibration of flexible riser with thick strip method[J]. Journal of Fluids and Structures, 2019, (xxxx), 1–8.
[3] Chaplin J.R., Bearman P.W., Cheng Y., et al. Blind predictions of laboratory measurements of vortex-induced vibrations of a tension riser. Journal of Fluids and Structures, 2005a, 21(1 SPEC. ISS.), 25–40.
[4] Chaplin J.R., Bearman P.W., Huera Huarte F.J., et al. Laboratory measurements of vortex-induced vibrations of a vertical tension riser in a stepped current[J]. Journal of Fluids and Structures, 2005b, 21(1 SPEC. ISS.), 3–24.
[5] Chen W.M., Fu Y.Q., Guo S.X., et al. Fluid-solid coupling and hydrodynamic response of vortex-induced vibration of slender ocean cylinders[J]. Advances in Mechanics, 2016, 38(05):604.

[6] Duan M.Y., Wan D.C., Xue H.X. Prediction of response for vortex-induced vibrations of a flexible riser pipe by using multi-strip method[C] Proceedings of the Twenty-sixth (2016) International Ocean and Polar Engineering Conference Rhodes, Greece, June 26-July 1, 2016, pp. 1065-1073.
[7] Duan M.Y., Zou L., Wan D.C. Numerical analysis of multi-modal vibrations of a vertical riser in step currents[J]. Ocean Engineering, 2018, 152: 428-442.
[8] Fu B.W., Duan M.Y., Wan D.C. Effect of mass ratio on the vortex-induced vibrations of a top tensioned riser[C]. Proceedings of the Second Conference of Global Chinese Scholars on Hydrodynamics, November 11-14, 2016, Wuxi, China, pp. 431-435.
[9] Fu B.W., Zou L., Wan D.C. Numerical study of vortex-induced vibrations of a flexible cylinder in an oscillatory flow[J]. Journal of Fluids and Structures, 2018, 77:170-181.
[10] Fu S.X., Wang J.G., Baarholm R., et al. VIV of Flexible Cylinder in Oscillatory Flow[C]. International Conference on Offshore Mechanics and Arctic Engineering, OMAE, June 9–14, 2013, Nantes, France, Volume 7: CFD and VIV ():V007T08A021.
[11] Holmes S., Oakley O.H., Constantinides Y. Simulation of Riser VIV Using Fully Three Dimensional CFD Simulations[J]. International Conference on Offshore Mechanics and Arctic Engineering, OMAE, June 4–9, 2006, Hamburg, Germany, OMAE2006-92124.
[12] Huang K., Chen H., Chen C.R. Riser VIV Analysis by a CFD Approach[C]. Proceedings of the Seventeenth (2007) International Ocean and Polar Engineering Conference, Lisbon, Portugal, July 1-6, 2007, pp2722–2729.
[13] Huang K. and Chen C.R. Time-Domain Simulation of Riser VIV in Sheared Current[C]. International Conference on Offshore Mechanics and Arctic Engineering, OMAE, June 10–15, 2007, San Diego, California, USA, OMAE2007-29363.
[14] Huang X.D., Zhang H., Wang X.S. An overview on the study of vortex-induced vibration of marine riser[J]. Journal of Marine Sciences, 2009, 27(04):95-101.
[15] Lehn E. VIV Suppression Tests on High L/D Flexible Cylinders. 2003, Norwegian Marine Technology Research Institute, Trondheim, Norway.
[16] Sarpkaya T. A critical review of the intrinsic nature of VIV[J]. Fluid Mechanics and Its Applications, 2004, 75:159–161.
[17] Meneghini J.R., Saltara F., Fregonesi R.D.A., et al. Numerical simulations of VIV on long flexible cylinders immersed in complex flow fields[J]. European Journal of Mechanics / B Fluids, 2004, 23(1):51-63.
[18] Wan D.C., Duan M.Y. A Recent Review of Numerical Studies on Vortex-Induced Vibrations of Long Slender Flexible Risers in Deep Sea[J]. Chinese Quarterly of Mechanics, 2017, 38(02):179-196.
[19] Wang E. and Xiao, Q. Numerical simulation of vortex-induced vibration of a vertical riser in uniform and linearly sheared currents[J]. Ocean Engineering, 2016, 121, 492–515.
[20] Willden R.H.J., Graham J.M.R. Numerical prediction of VIV on long flexible circular cylinders[J]. Journal of Fluids & Structures, 2001, 15(3-4):659-669.
[21] Willden R.H.J., Graham J.M.R. Multi-modal Vortex-Induced Vibrations of a vertical riser pipe subject to a uniform current profile[J]. European Journal of Mechanics, B/Fluids, 2004, 23(1):209–218.
[22] Yamamoto C.T., Meneghini J.R., Saltara F., et al. Numerical simulations of vortex-induced vibration on flexible cylinders[J]. Journal of Fluids & Structures, 2004, 19(4):467-489.

# Comparison of liquid sloshing in different tanks based on MPS method

**Fengze Xie, Xiang Chen, Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

## Abstract

The selection of tanks' shape may affect its life-span and safety because different shape tanks show different sensitivity to liquid sloshing. In this paper, responses of different shape tanks under external excitation are compared. The in-house solver MLParticle-SJTU solver based on moving particle semi-implicit method is employed for the simulation. The convergent validation is conducted to verify the reliability of present solver to simulate the sloshing in cylindrical tank. Sloshing in the rectangular tank and cylindrical tank with different filling ratios is simulated. The characteristics of flow field and pressure time history are presented. In addition, Three-dimensional effect can be observed in simulation, which have great influence on the sloshing and the load applied on the wall.

**Keywords:** Moving particle semi-implicit method; Liquid sloshing; MLParticle-SJTU solver; Three-dimensional effect

## Introduction

Due to the uneven distribution of energy in different region, a large amount of energy needs to be transported from one area to another every year. Liquid sloshing is a significant issue in the transportation of liquefied natural gas, oil and liquefied petroleum gas. The liquid inside a partially filled tank will be induced to violent oscillations and large impact pressure on the tank under external excitations which are large amplitude or resonance frequency of sloshing. Therefore, many researchers have investigated the characters and mechanisms of sloshing.

There are many methods to investigated sloshing problem. Compared with other traditional methods, Computational Fluid Dynamic (CFD) technology has some extra advantages. It can provide more detailed flow field information, which facilitate people to analyze the evolution process of sloshing fluid flow field, discover the physical mechanism behind sloshing phenomenon, and provide guidance for experiment. This is the reason why more researchers focus their attention on CFD technology. Faltisnen (1978) firstly applied Boundary-Element-Method (BEM) to investigate the sloshing problem [1]. Milkelis et al. (1984) used the Marker-And-Cell (MAC) method to capture the free surface and simulate the 2-D liquid sloshing in a rectangular tank and a membrane tank [2]. Liu et al. (1994) adopted Arbitrary-Lagrangian-Euler (ALE) to capture the free surface and simulate the large -amplitude sloshing [4]. Sussman (1998) studied the 2-D liquid sloshing based on the modified Level-set method, which can simulate complex free surface deformation successfully [3]. Kim et al. (2001) applied the Finite-Difference-Method (FDM) and Impulse-Response-Function (IRF) method to study the coupling effects of ship motion and sloshing [5]. Belakroum et al. (2010) used Finite-Element-Method (FEM) to simulate the sloshing of rectangular tank and proposed a new method to reduce sloshing [6]. Zhuang et al. (2016) used the naoe-FOAM-SJTU solver based

on Finite-Volume-Method (FVM) to conduct the numerical simulation of FPSO motion coupled with LNG sloshing [7].

Particle method has the superiority to handle the problem of large deformation of free surface, which is proven to be valid and efficient in simulating liquid sloshing in previous work. Iglesias et al. (2004, 2006) used Smoothed Particle Hydrodynamics (SPH) method to simulate the sloshing in the anti-rolling tank of a fishing vessel [8] [9]. Delorme et al. (2009) used SPH method to investigate the impact pressure in the case of shallow water sloshing. Zhang et al. (2012) used Moving Particle Semi-implicit method to simulate liquid sloshing in LNG tank [11]. Koh et al. (2013) adopted Consistent Particle Method (CPM) to investigate the sloshing problem of a rectangular tank with constrained floating baffles under sway excitation [12]. Zhang et al. (2017) applied MPS method and FEM method to study the liquid sloshing in rectangular tank with elastic bulkhead [13]. Chen et al. (2017) introduced the GPU technology to MPS method to accelerate the simulation in 3-D sloshing [14]. Wen et al. (2018) simulate the three-layer-liquid sloshing in the rigid tank with multiphase MPS method [15].

In this study, an in-house solver MLParticle-SJTU based on modified MPS is employed to simulate the three-dimensional sloshing. In the first section, the description of modified MPS method is presented briefly. In the second section, the convergent validation is carried out to verify the accuracy of present solver in the simulation of cylindrical tank and the simulation result shows good agreements with experimental data. Then, the sloshing in three-dimensional cylindrical tank and rectangular tank with different filling ratios is simulated at their respective natural frequencies. The slamming pressure in different location is measured and the comparison of those two type tanks is conducted.

## Numerical Method

MPS method is proposed by Koshizuka et al. (1996) for viscous incompressible fluid [16]. In this section, basic theories and discretization process of MPS will be presented in detail.

### *Governing Equations*

The governing equations contain continuity equation and Navier-Stokes equation.

$$\nabla \cdot \vec{V} = 0 \tag{1}$$

$$\frac{D\vec{V}}{Dt} = -\frac{1}{\rho}\nabla P + \nu\nabla^2\vec{V} + \vec{g} \tag{2}$$

Where the $\vec{V}$ is the velocity vector, the $t$ is the time, the $\rho$ is the fluid density, $P$ is the pressure, $\nu$ is the kinematic viscosity, $\vec{g}$ is the gravity acceleration vector.

### *Kernel Function*

In MPS method, the interaction between particles is controlled by kernel function, which plays a role of weight function in the discretization process. In order to avoid non-physical pressure oscillation, the kernel function presented by Zhang et al. (2014) is employed here [17].

$$W(r)=\begin{cases}\dfrac{r_e}{0.85r+0.15r_e}-1 & 0\le r<r_e\\ 0 & r_e\le r\end{cases} \tag{3}$$

Where $r$ is the distance between two particles and $r_e$ is the radius of the particle interaction.

*Gradient Model*

The gradient vector of particle $i$ is the weighted average of the gradient vectors between particle $i$ and all its neighboring particles $j$. The model adopted in this paper is proposed by Tanaka et al. (2010), which meets the law of conservation of momentum.

$$<\nabla\phi>_i=\frac{d}{n^0}\sum_{j\neq i}\frac{\phi_j+\phi_i}{|\vec{r}_j-\vec{r}_i|^2}(\vec{r}_j-\vec{r}_i)\cdot W(|\vec{r}_j-\vec{r}_i|) \tag{4}$$

Where $\phi$ is a physical quantity, $d$ is the number of space dimension, $n^0$ is the initial particle density, $\vec{r}$ is the position vector relative to origin.

*Laplacian Model*

The Laplacian model is the weighted average of the distribution of a quantity $\phi$ from particle $i$ to neighboring particle $j$, which is needed in the solution of viscosity term of N-S equation and the space discretization of Pressure Poisson Equation.

$$<\nabla^2\phi>_i=\frac{2d}{n^0\lambda}\sum_{j\neq i}(\phi_j-\phi_i)\cdot W(|\vec{r}_j-\vec{r}_i|) \tag{5}$$

$$\lambda=\frac{\sum_{j\neq i}W(|\vec{r}_j-\vec{r}_i|)\cdot|\vec{r}_j-\vec{r}_i|^2}{\sum_{j\neq i}W(|\vec{r}_j-\vec{r}_i|)} \tag{6}$$

*Pressure Poisson Equation* （*PPE*)

In MPS method, the acquirement of pressure is through solving PPE. In this paper, mixed source term method is used to solve PPE, which is developed by Tanaka et al. (2010) [18].

$$<\nabla^2P^{k+1}>_i=(1-\gamma)\frac{\rho}{\Delta t}\nabla\cdot\vec{V}_i^*-\gamma\frac{\rho}{\Delta t^2}\frac{<n^*>_i-n^0}{n^0} \tag{7}$$

Where $P^{k+1}$ is the pressure of the step $k+1$, $\gamma$ is a blending parameter, $\Delta t$ is the time step, $\vec{V}_i^*$ is the temporal velocity, $n^*$ is the temporal particle density. In this paper, the value of $\gamma$ set to be 0.01.

*Divergence Model*

Divergence model is similar to the gradient model and it is used to discrete velocity divergence in the PPE.

$$\nabla \cdot \vec{V} = \frac{d}{n^0} \sum_{j \neq i} \frac{\left(\vec{r}_j - \vec{r}_i\right) \cdot \left(\vec{V}_j - \vec{V}_i\right)}{\left|\vec{r}_j - \vec{r}_i\right|^2} W\left(\left|\vec{r}_j - \vec{r}_i\right|\right) \tag{8}$$

*Free Surface Detection*

An improved surface particle detection method developed by Zhang et al. (2010) is adopted in this paper [19], which can distinguish the particles of free surface from the others efficiently. This approach is based on the asymmetry distribution of neighboring particles.

$$< n >_i^* \ < \beta n^0 \tag{9}$$

$$< \vec{F} >_i = \frac{d}{n^0} \sum_{j \neq i} \frac{1}{|\vec{r}_i - \vec{r}_j|} (\vec{r}_i - \vec{r}_j) W\left(\left|\vec{r}_j - \vec{r}_i\right|\right) \tag{10}$$

$$< |\vec{F}| >_i \ > \alpha |\vec{F}_0| \tag{11}$$

Where $\beta$ and $\alpha$ are parameters, $\vec{F}$ is a vector which represents the asymmetry distribution of neighboring particles. When $\beta \leq 0.8$ or $\beta \geq 0.97$, the type of particles can be set to free surface. When the $0.8 < \beta < 0.97$, formulas (9) and (10) are used to judge free surface particles. The value of $\alpha$ is set to be 0.9.

*Boundary Condition*

There are multilayer particles arranged at the solid boundary. One layer of wall particles is arranged near the fluid particles and their pressure is solved by PPE. Two layers of ghost particles are configured because fluid particles lack neighbor particles on the side of the solid wall. The pressure of ghost particles is obtained by extrapolation. Both the wall particles and the ghost particles don't update their velocity and displacement after they gain the pressure.

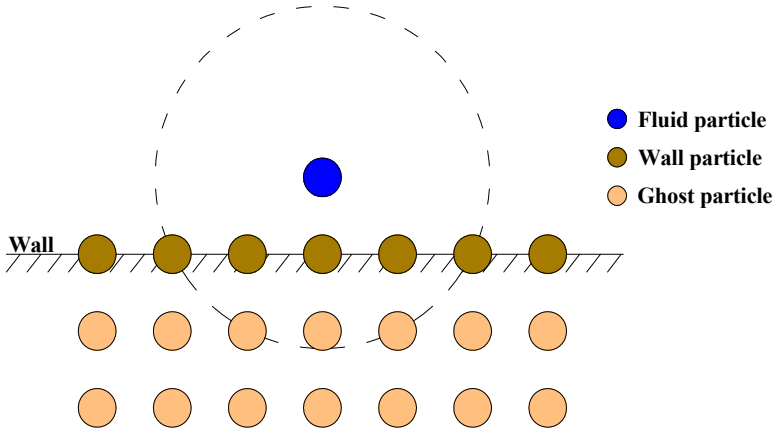



**Figure 1. Diagram of boundary particles**

## Numerical Simulations

In this section, the responses of a cylindrical tank and a rectangular tank under external

excitation are compared. The cylindrical tank comes from the experiment conducted by Kobayashi et al. (1986) [20]. The parameters of the rectangular tank are selected according to the cylindrical tank and its cross-sectional area is a square. Figure 2 shows the geometry of those two liquid tanks. The dimensions of tanks are 0.94m ($L$), 0.47m ($D$), 0.47m ($B$) and 0.47m ($H$). Those two liquid tanks sway harmonically under the external excitation.

$$x = A\sin(\omega t) \tag{12}$$

Where $A$ is the amplitude of motion, $\omega$ is the excitation frequency.

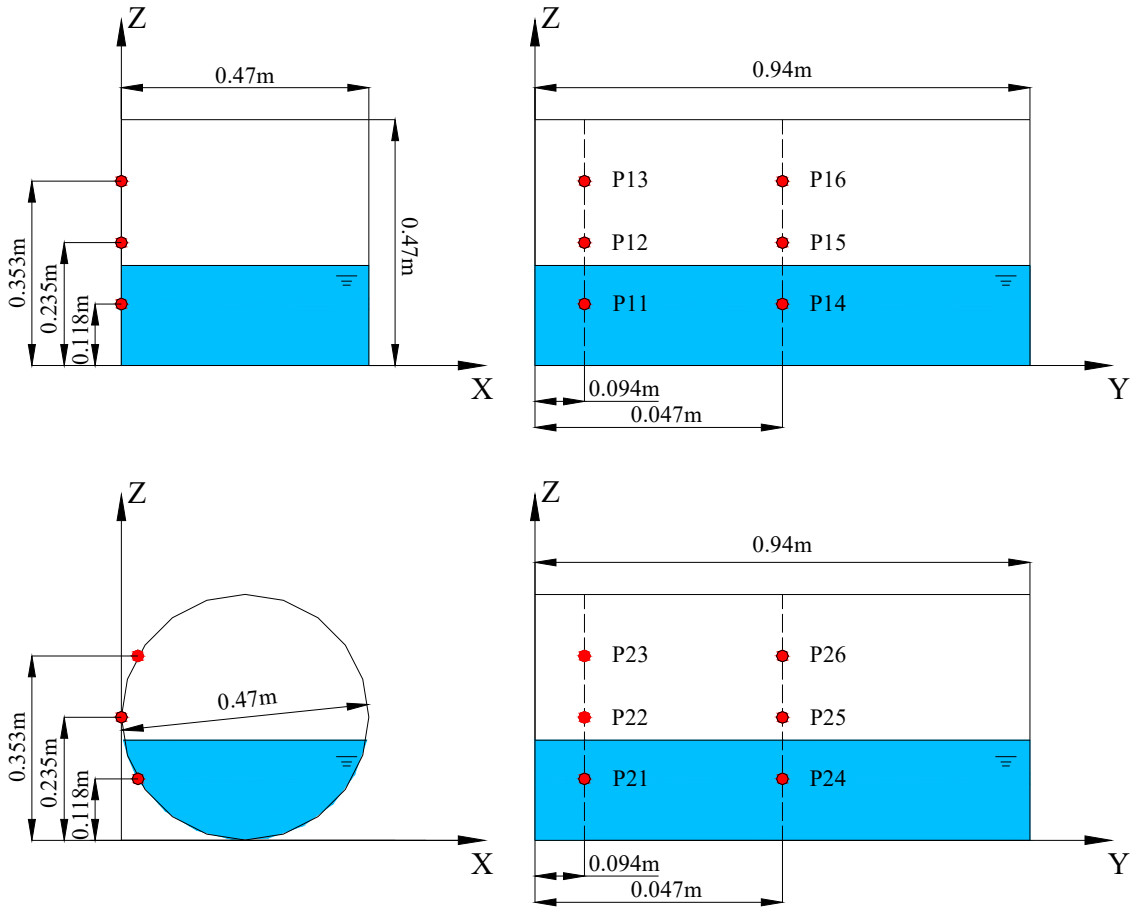



**Figure 2. The sketch of numerical models**

Six pressure probes are arranged on walls of both tanks to measure the time history of pressure and their specific locations are listed in Table 1.

**Table 1 Arrangements of pressure probe**

| rectangular | X/m | Y/m | Z/m | cylinder | X/m | Y/m | Z/m |
|---|---|---|---|---|---|---|---|
| P11 | 0 | 0.094 | 0.118 | P21 | 0.0315 | 0.094 | 0.118 |
| P12 | 0 | 0.094 | 0.235 | P22 | 0 | 0.094 | 0.235 |
| P13 | 0 | 0.094 | 0.353 | P23 | 0.0781 | 0.094 | 0.353 |
| P14 | 0 | 0.470 | 0.118 | P24 | 0.0315 | 0.470 | 0.118 |
| P15 | 0 | 0.470 | 0.235 | P25 | 0 | 0.470 | 0.235 |
| P16 | 0 | 0.470 | 0.353 | P26 | 0.0781 | 0.470 | 0.353 |

*Verification*

The accuracy of MLParticle-SJTU solver to simulate sloshing in rectangular tank has been validated in previous work. In this sub-section, the convergence verification of particle distance is conducted and the time history of resultant force in the Z direction is compared with the experimental data (Kobayashi et al., 1986) and the reliability of the solver to simulate sloshing in cylindrical tank is confirmed. The initial depth of water ($h$) is 0.235 m, corresponding filling ratio is 50%. The tank is forced to sway with the frequency ($\omega = 7.536\,\mathrm{rad/s}$) and the amplitude ($A = 0.015\,\mathrm{m}$). The model with initial spacing sizes of 0.0045 m, 0.005 m and 0.006 m is simulated to check the convergence of numerical results. Fig.3 compares the present numerical results with the experimental results. It can be noticed that the results of models with different spatial resolutions all agree well with the experimental results， which shows the accuracy and stability of the solver. Considering the computational efficiency and the refinement of flow field, initial distance between fluid particles is set to 0.005 m for the following simulation in this paper.

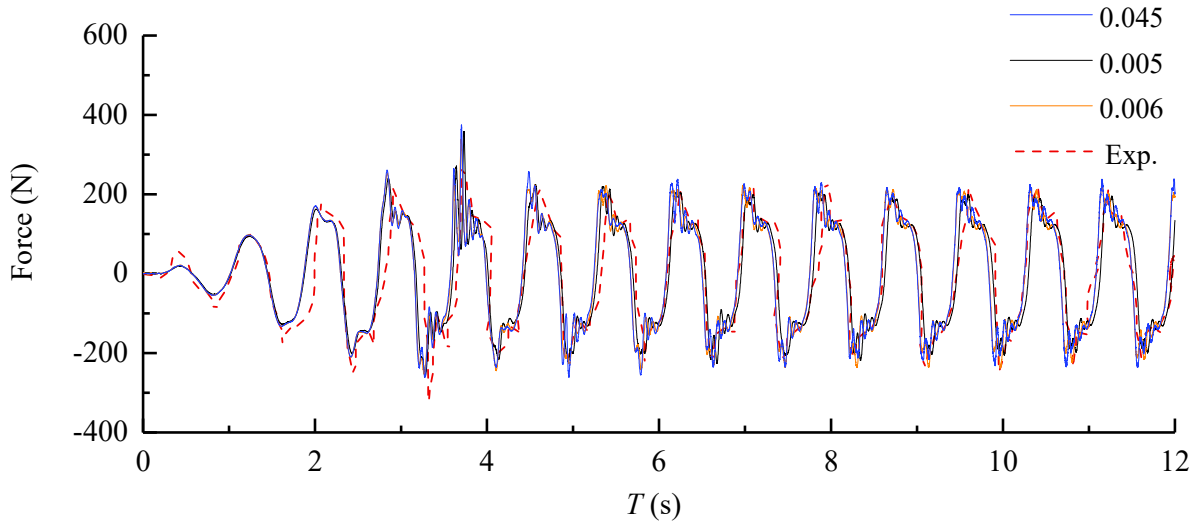



**Figure 3. Time history of force in Z direction**

*Numerical conditions*

In present paper, the sloshing in the cylindrical tank and the rectangular tank is compared at different filling ratios. The amplitude of motion is set to 0.015m. Considering the most extreme condition, tanks are excited at their respective natural frequencies. For rectangular tanks, the natural frequencies are calculated according to the formula (13). For cylindrical tanks, the natural frequencies are acquired according to the curve (Wiesche et al., 2008 [21]), which are presented in the Figure 4. Detailed parameters for numerical conditions are presented in Table 2.

$$\omega=\sqrt{\frac{g\pi}{L}\tanh\frac{\pi h}{L}} \tag{13}$$

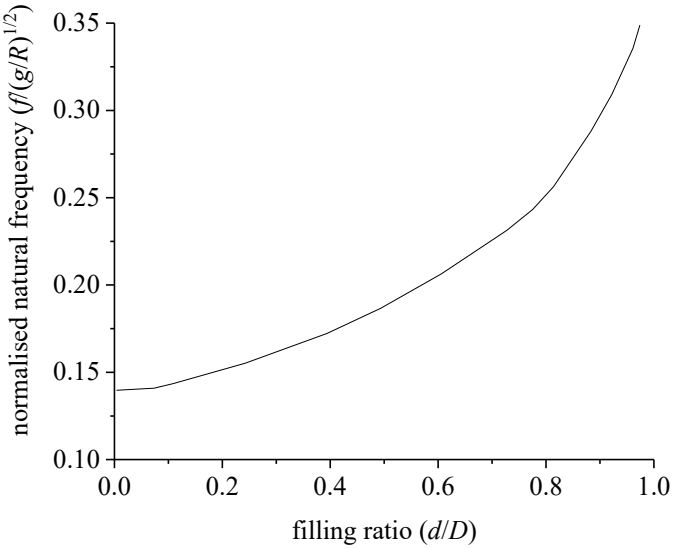



**Figure 4. First transverse natural slosh frequency for horizontal cylindrical tanks(Wiesche et al. ,2008)**

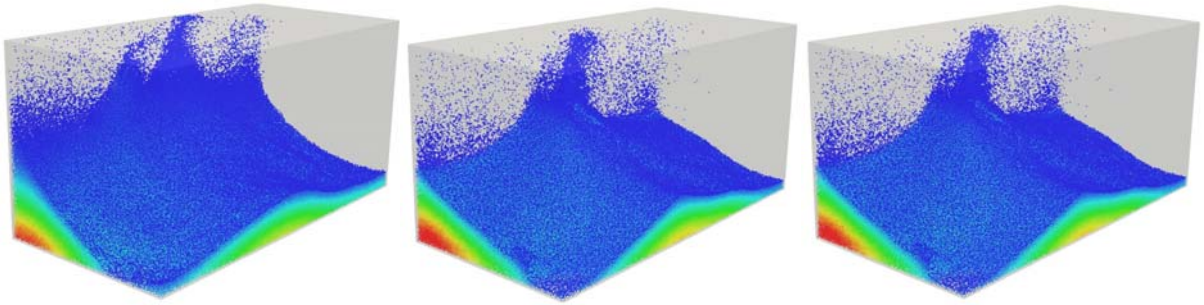
**Table 2 Numerical conditions**

| rectangular tank | | | cylindrical tank | | |
|---|---|---|---|---|---|
| filling ratio ($h/H$) | nature frequency $\omega_0$ (rad/s) | particle number | filling ratio ($h/H$) | nature frequency $\omega_0$ (rad/s) | particle number |
| 0.25 | 6.558 | 709005 | 0.25 | 6.535 | 520060 |
| 0.5 | 7.755 | 1126389 | 0.5 | 7.805 | 939800 |
| 0.75 | 8.025 | 1491600 | 0.75 | 9.566 | 1321792 |

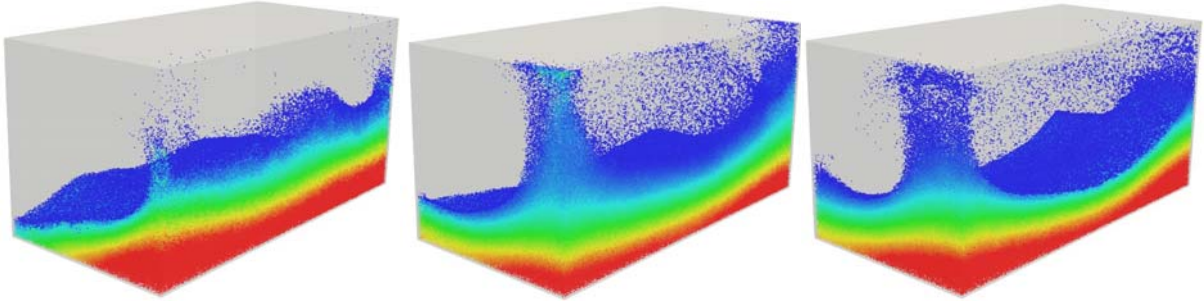
*Numerical results*

In this sub-section, there are some comparisons between cylindrical tank and rectangular tank presented.

Firstly, the tanks with low filling ratio (0.25) are compared. Figure 5 shows some snapshots of numerical flow field. The 3D effect in rectangular tank is more obvious than cylindrical tank's and a jumping phenomenon can be observed in rectangular tank. Besides, it can be noticed that slamming in rectangular tank is asymmetric. When the fluid flows to the left wall, the fluid in the middle of the wall is faster than that on both sides, impacting the roof of the tank. When the fluid flows to the right wall, the fluid on both sides of the wall is faster than that in the middle, reaching the higher position of the wall. The pressure time histories measured at different probes in the same tank are shown as Figure 6. For rectangular tank, the peak pressures and phases measured at P11 and P14 are very different, which indicates that traveling waves are generated along the longitudinal direction. For cylindrical tank, the peak pressures measured at P21 and P24 are slightly different and the pressure time histories are similar in general, which indicates that the waves traveling along the longitudinal direction are small and their influence to transverse sloshing is limited. In addition, the pressure time histories measured at different probes in different tanks are compared, as shown in Figure 7. The peak values of pressure measured at P14 and P24 are very close and the double pressure peaks can be observed at both probes when the waves in longitudinal direction haven't formed. Due to interaction of transverse sloshing mode and longitude sloshing mode, one of pressure peak disappears. There is a radian on the cylindrical wall and the momentum will be changed as fluid climb along the wall, so the peak values of pressure measured at P15 are generally higher than those measured at P25. However, the last few peak values of pressure at P15 are higher than those at P25 because of 3D effect. The probe P26 arranged at the highest position hardly detect the pressure, which indicates that the fluid have detached the wall before it arrived at the probe.



(a1) 16.35s (a2) 16.47s (a3) 16.59s



(a4) 16.71s (a5) 16.83s (a6) 16.95s

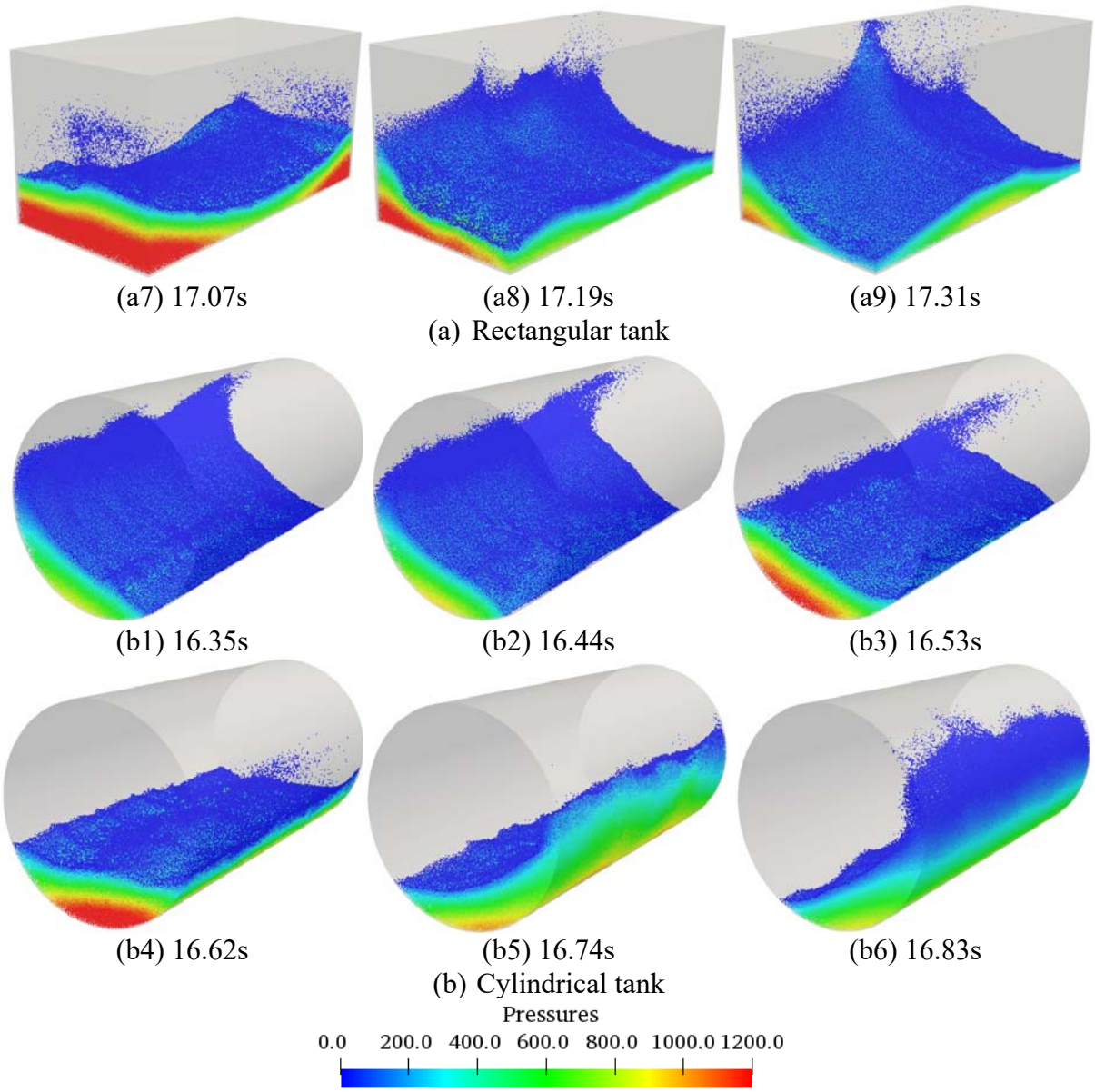
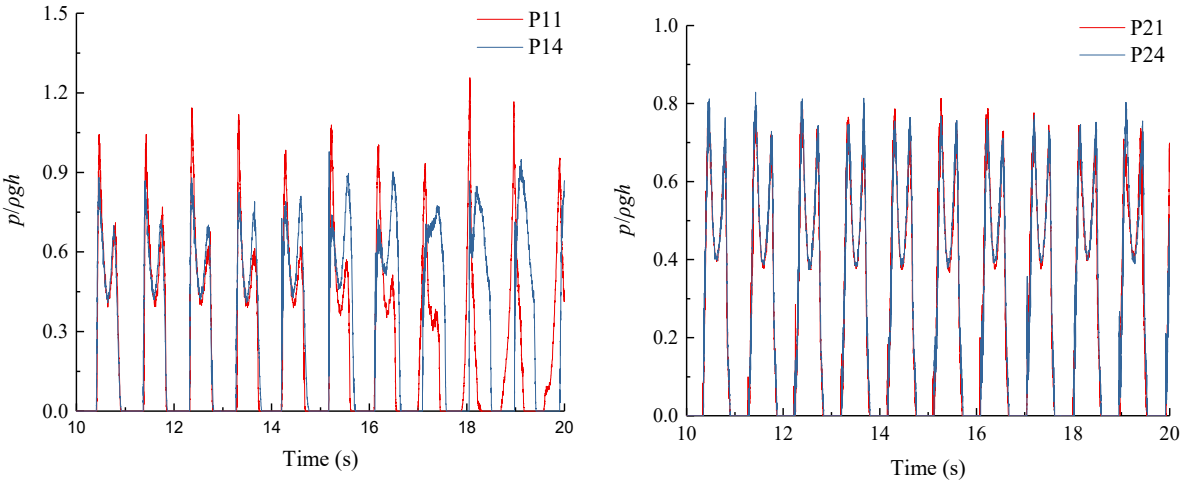
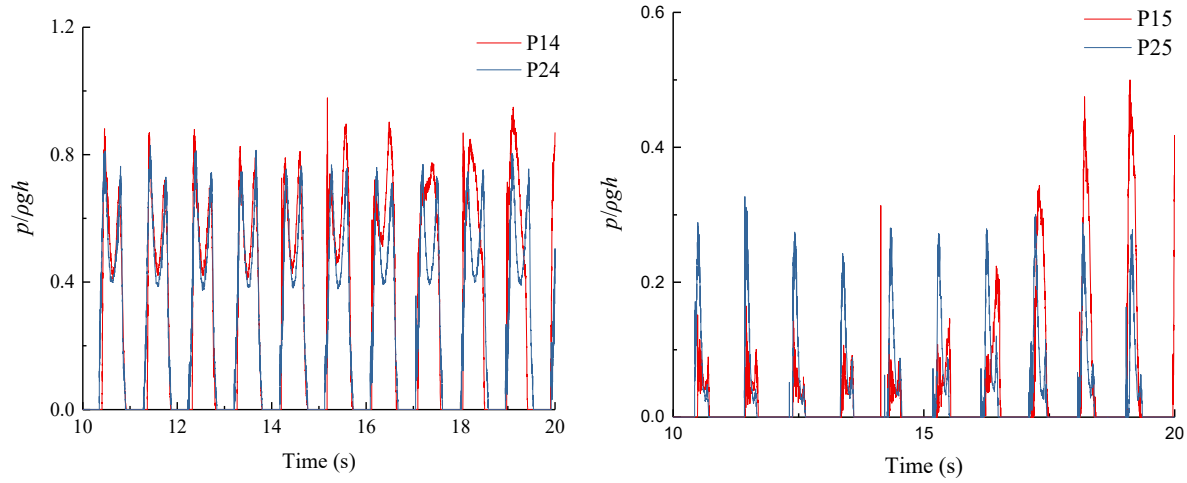



**Figure 5. The flow fields in rectangular tank and cylindrical tank($h/H$=0.25)**

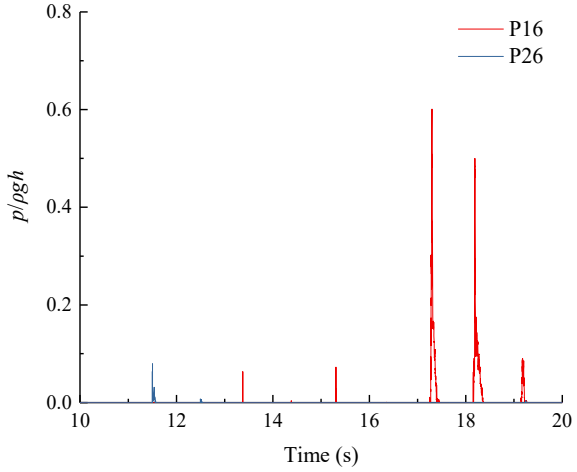



(a) P11 and P14 (b) P21 and P24

**Figure 6. Time history of pressures measured at different probes in the same tank(*h*/*H*=0.25, Left: rectangular tank, Right: cylindrical tank)**



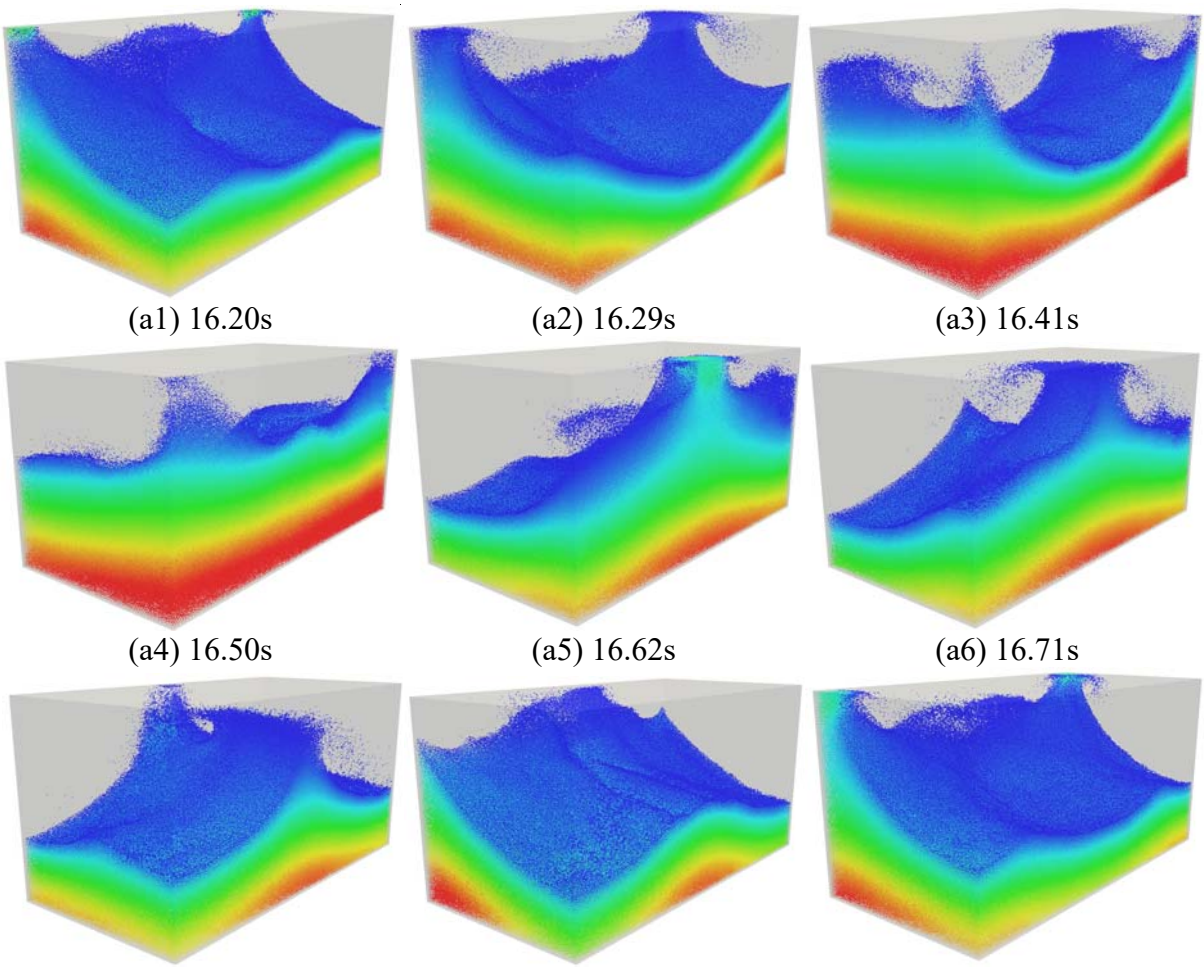

(a) P14 and P24 (b) P15 and P25




(c) P16 and P26

**Figure 7. Time history of pressures measured at different probes in different tanks(*h*/*H*=0.25, red line: rectangular tank, blue line: cylindrical tank)**

Secondly, the tanks with the filling ratio of 0.5 are compared. Figure 8 shows some snapshots of numerical flow field and Figure 9 shows the comparisons of pressure time histories measured at different probes in the same tank. The conclusions drawn from this simulation are similar to those obtained from simulation with low filling ratio. Figure 10 shows the comparisons of pressure time histories measured by different probes in different tanks. Before 15s，the 3D effect is at initial stage and peak values of pressure measured at probes of cylindrical tank are all slightly higher than those measured at corresponding probes of rectangular tank. After 15s，the 3D effect is at mature stage and the peak values of pressure measured by cylindrical tank probes all far less than those measured at corresponding probes in the rectangular tank. Besides, it can be noticed that time histories of pressure measured in rectangular tank is more randomness than those measured in cylindrical tank.



(a1) 16.20s (a2) 16.29s (a3) 16.41s

(a4) 16.50s (a5) 16.62s (a6) 16.71s

(a7) 16.80s (a8) 16.92s (a9) 17.01s

(a) Rectangular tank

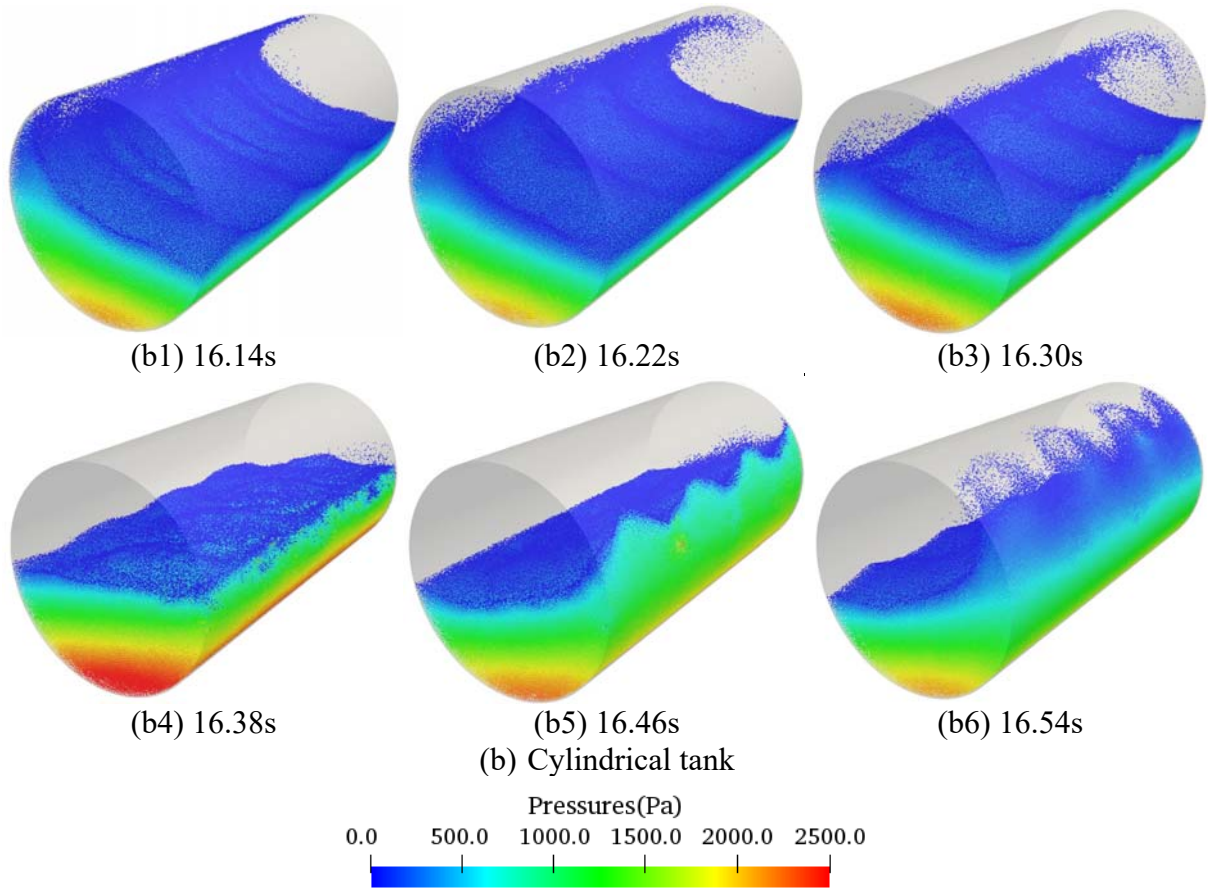



**Figure 8. The flow fields in rectangular tank and cylindrical tank (*h*/*H*=0. 5)**

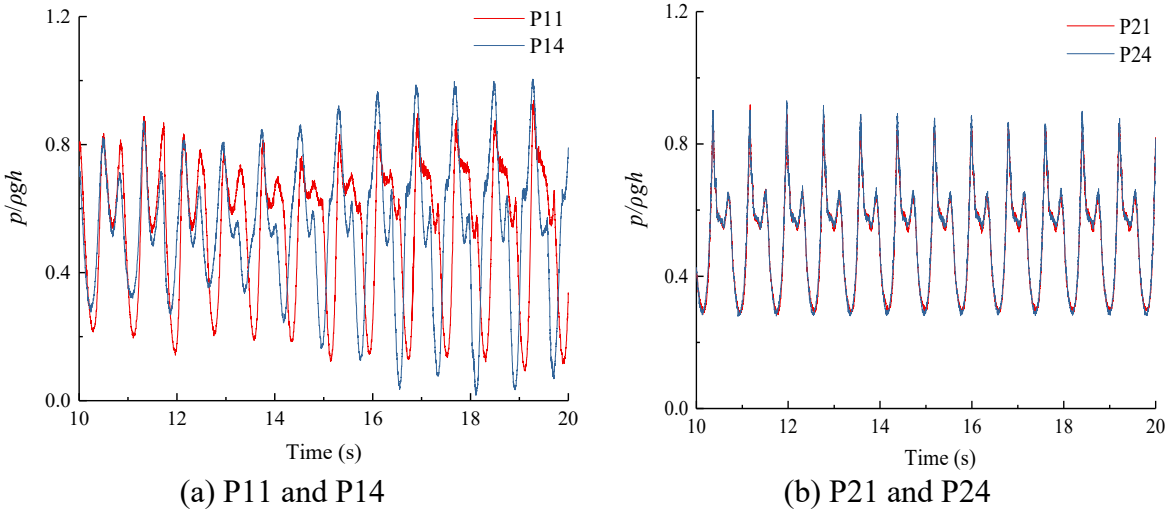
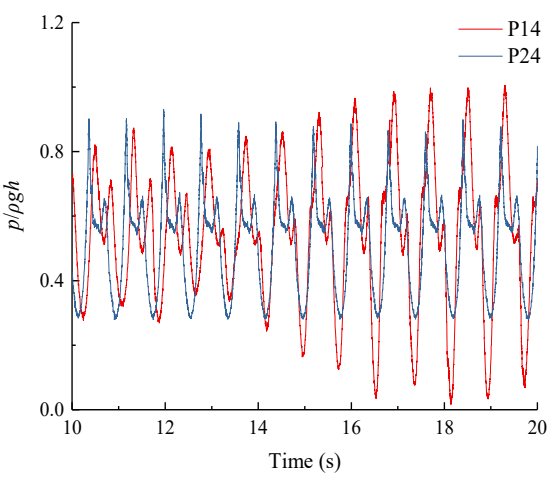
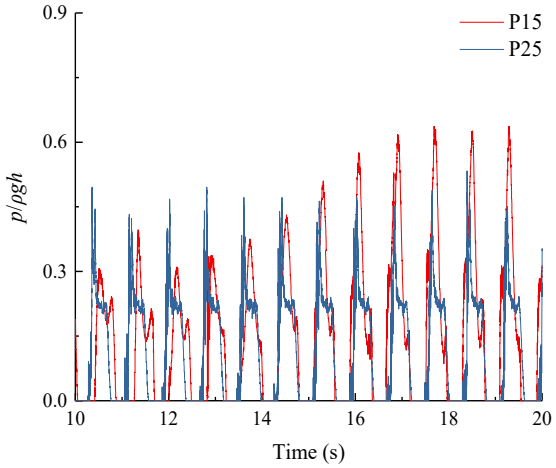


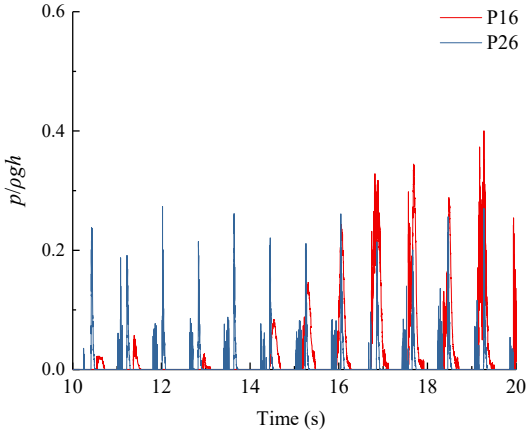

**Figure 9. Time history of pressures measured at different probes in the same tank(*h*/*H*=0.5, Left: rectangular tank, Right: cylindrical tank)**




(a) P14 and P24




(b) P15 and P25




(c) P16 and P26

**Figure 10. Time history of pressures measured at different probes in different tanks(*h*/*H*=0.5, red line: rectangular tank, blue line: cylindrical tank)**

Finally, the tanks with high filling ratio (0.75) are compared. There is no obvious 3D effect observed in either type of tank, as shown in Figure 11 and Figure 12. This is because the roof plays a role as a horizontal buffer and prevents the further development of 3D effect. Figure 13 shows the comparisons of pressure time histories measured at different probes in different tanks. It is noticed that the pressure peak of cylindrical tank is higher than rectangular tank's. Besides, two successive peaks of pressure can be observed at each probe and the difference between those two peaks of cylindrical tank is much greater than rectangular tank's. Most of water will flow through the roof and fall to the fee surface in cylindrical tank. However, much water will go back along the wall in rectangular bank.

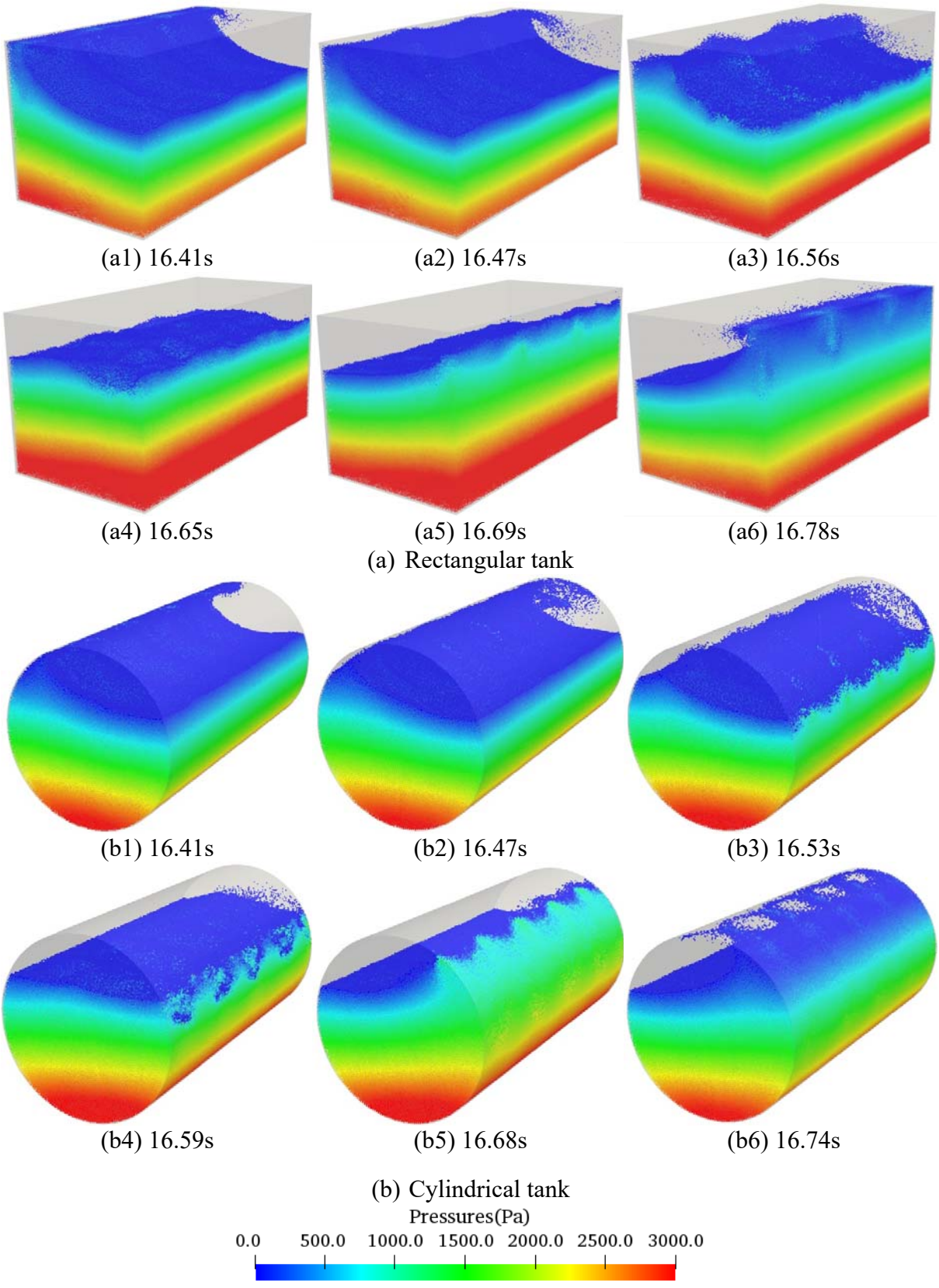
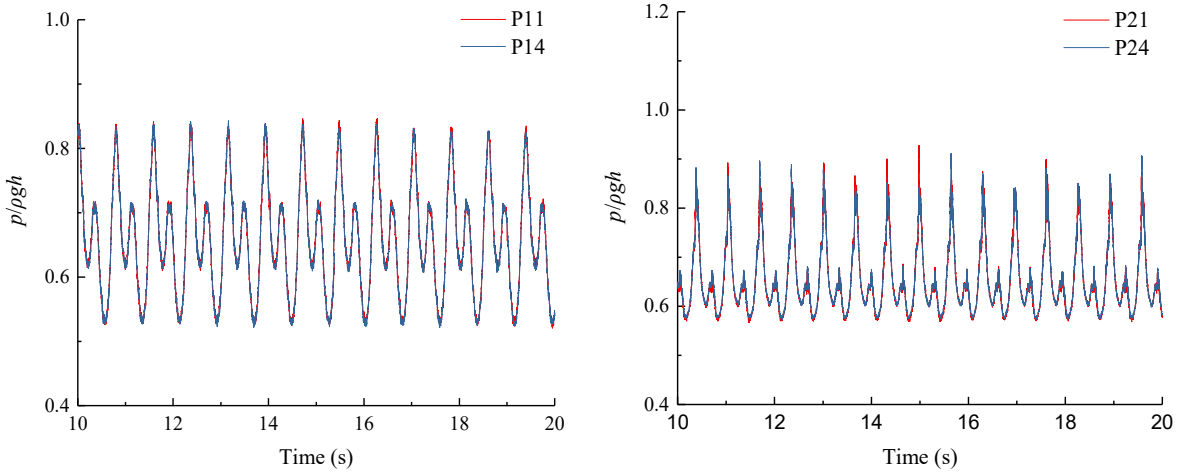
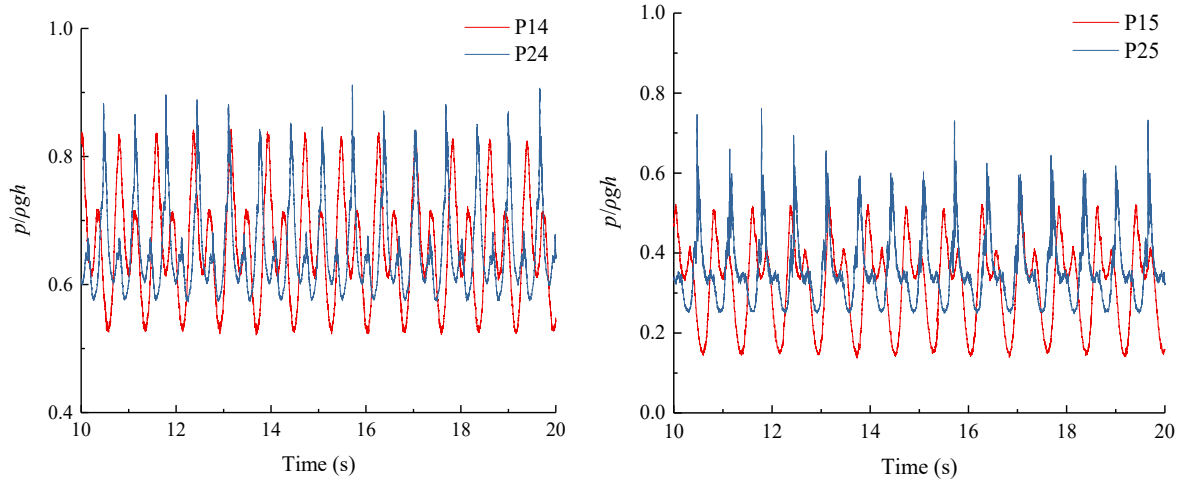



**Figure 11. The flow fields in rectangular tank and cylindrical tank (*h*/*H*=0. 75)**

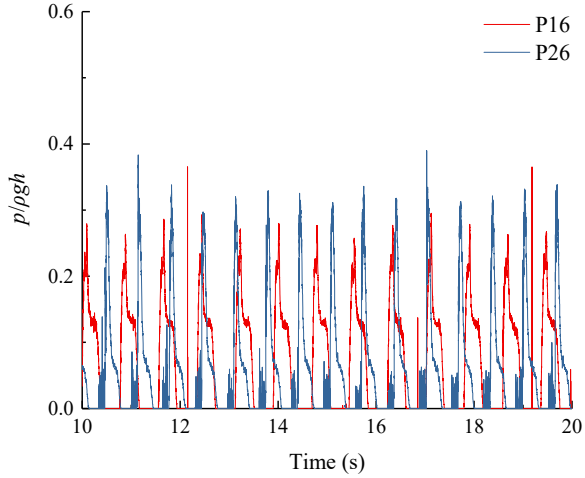



(a) P11 and P14 (b) P21 and P24

**Figure 12. Time history of pressures measured at different probes in the same tank(*h*/*H*=0.75, Left: rectangular tank, Right: cylindrical tank)**




(a) P14 and P24 (b) P15 and P25




(d) P16 and P26

**Figure 13. Time history of pressures measured at different probes in different tanks(*h*/*H*=0.75, red line: rectangular tank, blue line: cylindrical tank)**

## Conclusions

In this paper, the MLParticle-SJTU solver based on modified MPS is employed to investigate difference of sloshing in rectangular tank and cylindrical tank. The convergence verification is conducted to confirm the reliability of the solver. The comparison of pressure measured by different probes and the numerical flow field are presented. Based on the results of simulations, the following conclusions can be summarized:

- The 3D effect observed in rectangular tank with filling ratio of 0.25 and 0.5 is more obvious than that in cylindrical tank and a jumping phenomenon can be observed. The same conclusion can be drawn throw the comparison of pressure time history measured at different probes in the same tank at the same height. The peaks and phases are different in rectangular tank while those show a good agreement in cylindrical tank.
- When the 3D effect has not formed, the pressure peaks measured in rectangular tank is less than or close to those measured by corresponding probes in cylindrical tank. As the 3D effect become more severe, the pressure peaks measured in rectangular tank is far higher than those in cylindrical tank.
- With filling ratio of 0.75，The difference of two successive pressure peak in cylindrical tank is much greater than those in rectangular tank. Because, after slamming the roof, most of water will flow through the roof and fall to the fee surface in cylindrical tank while much water will go back along the wall in rectangular bank.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

### References

[1] Faltinsen, O. M. (1978) A numerical nonlinear method of sloshing in tanks with two-dimensional flow, *Journal of Ship Research* **22**(3), 193-202.

[2] Milkelis, N. E. and Robinson, D. W. (1985) Sloshing in arbitrary shaped tanks, *Journal of the Society of the Naval Architects of Japan* **158**, 246-255.

[3] Sussman, M., Fatemi, E., Smereka, P., et al. (1998) An improved level set method for incompressible two-phase flows, *Computers & Fluids* **27**(5), 663-680.

[4] Liu, Z., and Huang, Y. (1994) A new method for large amplitude sloshing problems, *Journal of Sound and Vibration* **175**(2), 185-195.

[5] Kim, Y., Nam, B. W., Kim, D. W. and Kim, Y. S. (2007) Study on coupling effects of ship motion and sloshing, *Ocean Engineering* **34**(16), 2176-2187.

[6] Belakroum, R., Kadja, M., Mai, T.H., Maalouf, C. (2010) An efficient passive technique for reducing sloshing in rectangular tanks partially filled with liquid, *Mechanics Research Communications* **37**(3), 341-346.

[7] Zhuang, Y. and Wan, D. C. (2017) Numerical study on ship motion fully coupled with LNG tank sloshing in CFD method, *International Journal of Computational Methods*, 1840022.

[8] Iglesias, A. S., Rojas, L. P. and Rodriguez, R. Z. (2004) Simulation of anti-roll tanks and sloshing type problems with smoothed particle hydrodynamics, *Ocean Engineering* **31**(8), 1169-1192.

[9] Iglesias, A. S., Delorme, L., Rojas L. P., et al. (2006) Liquid moment amplitude assessment in sloshing type problems with smooth particle hydrodynamics, *Ocean Engineering* **33**(11). 1462-1484.

[10] Delorme, L., et al. (2009) A set of canonical problems in sloshing, Part I: Pressure field in forced roll—comparison between experimental results and SPH, *Ocean Engineering* **36**(2), 168-178.

[11] Zhang, Y. X. and Wan, D. C. (2012) Apply MPS method to simulate liquid sloshing in LNG tank, *Phytotherapy Research* **29**(12), 1843–1857.

[12] Koh, C. G., Luo, M., Gao, M., et al. (2013) Modelling of liquid sloshing with constrained floating baffle, *Computers and Structures* **122**, 270-279.

[13] Zhang, Y. L. and Wan, D. C. (2017) MPS-FEM coupled method for sloshing flows in an elastic tank, *Ocean Engineering* **152**, 416-427.

[14] Chen, X., Zhang Y. L. and Wan, D. C. (2017) *GPU acceleration of MPS for three-dimensional sloshing*, Proceedings of the 8th International Conference on Computational Methods (ICCM2017), Guilin, Guangxi, China.

[15] Wen, X. and Wan, D. C. (2018) *Numerical Simulation of Three-Layer-Liquid Sloshing by Multiphase MPS Method*, the ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering, Madrid, Spain.

[16] Koshizuka, S. and Oka, Y. (1996) Moving-Particle Semi-Implicit Method for Fragmentation of Incompressible Fluid, *Nuclear Science and Engineering* **123**(3), 421-434.

[17] Zhang, Y. X., Wan, D. C. and Hino, T. (2014) Comparative study of MPS method and level-set method for sloshing flows, *Journal of hydrodynamics* **26**(4), 577–585.

[18] Tanaka, M. and Masunaga, T. (2010) Stabilization and smoothing of pressure in MPS method by quasi-compressibility, *Journal of Computational Physics* **229**(11), 4279-4290.

[19] Zhang, Y. X. and Wan, D. C. (2012) *Apply MPS method to simulate liquid sloshing in LNG tank*, Proceedings of the 22nd International Offshore and Polar Engineering Conference, Rhodes, Greece.

[20] Kabayashi, N., Mieda, T. and Shibata, H. (1989) A Study of the Liquid Slosh Response in Horizontal Cylindrical Tanks, *Journal of Pressure Vessel Technology* **111**, 32-38.

[21] Wiesche and Stefan, A. D. (2008) Sloshing dynamics of a viscous liquid in a spinning horizontal cylindrical tank, *Aerospace Science and Technology* **12**(6), 448-456.

# Numerical Analysis of Yawed Turbine Wake under Atmospheric Boundary Layer Flows

**Xu Ning, Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

## Abstract

Yaw is the most common working condition of a wind turbine and the key of reducing the fatigue loads and improving the performance of a wind farm is to understand the wake characteristics of a wind turbine in yaw condition. A neutral boundary layer flow in the atmosphere is simulated by the LES technique using the solver developed based on OpenFOAM and the wake flow of a yawed wind turbine modeled by the actuator line is studied. The time-average velocity field proves the practicability of the yaw control operation in optimizing the total power output of a wind farm, but from the cross section contours, the velocity distribution in the wake of a yawed turbine is not completely symmetric and the vertical wake deflection cannot be neglected, which is the main source of errors of the analytical wake models based on the gaussian distribution assumption. The time history curves and frequency spectrum of the wake meandering gained from the filtered flow data indicate that the yaw conditions have limited impacts on the wake meandering and the downstream position $x$ is the main factor that affects the lateral movement of the wake. As the wake moves downstream, the meandering intensity increases and two distinct dominant frequencies of the wiggling wake are found in the far wake, of which the Strouhal numbers are 0.09 and between 0.1 ~ 0.3 respectively.

**Keywords**: Yawed wind turbine; Atmospheric boundary layer; Large eddy simulation; Wake meandering

## Introduction

A wind turbine works in yaw condition when the turbine rotor is not perpendicular to the incoming wind, which is in fact the most frequent working condition in a real wind farm. The inevitable misalignment between the horizontal axis of the wind turbine and the wind speed vector could, to a certain degree, reduce the aerodynamic performance and influence the development of the wake flow[1]. However, instead of aligning all the wind turbines perfectly with the wind direction, appropriate active yaw operation of the wind turbines located upstream will significantly decrease the production loss of downstream wind turbines caused by the turbine wake and thus achieve the overall optimal performance of the wind farm[2][3]. To build up a control algorithm that computes the optimal yaw angle of each turbine given the wind regime and the layout of the wind farm, a systematic study of the wake characteristics of a yawed wind turbine under various inflow conditions is necessary.

Grant et al.[4] conducted the very early wind-tunnel experiments to observe the vortex structure and the deflection phenomenon in the near wake region of a yawed model turbine. Similar experimental studies focusing on the near wake flow of wind turbine in yaw condition were also performed by Haans et al.[5] and Krogstad & Adaramola[6], with the later work revealing that the power output of a yawed wind turbine is proportional to $\cos^3\gamma$ ($\gamma$ is the yaw angle).

Bartl et al.[7] used Laser Doppler anemometry in the experiment to investigate the influence of the inflow shear and turbulence intensity on the evolution of the deflected wake. Bastankhah and Porté-Agel[8]-[10] performed a series of experimental and theoretical researches on both the near and far wake flow structure and dynamics of yawed wind turbines and integrated the yaw operation into the optimization algorithm of the wind farm. Moreover, the CFD technique also plays significant role in related studies. Jiménez et al.[11] introduced the body force computed by the actuator disk model to the LES equations for incompressible flow to simulate the wake deflection and trajectories and the results agreed well with the experimental measurements as well as the prediction of an analytical model. Also from LES case studies, which consider the different atmospheric stabilities and the uncertainties of wake defection, Vollmer et al.[12] emphasized the importance of the accurate measurement of the shear and turbulence of incoming flow in the prediction of the downstream wake position and indicated that the active yaw operation could effectively improve the performance of the wind farm in condition of an atmosphere with relatively high stability. Based on the work of Bastankhah and Porté-Agel[9], Guo-Wei Qian and Takeshi Ishihara[13] proposed an improved analytical wake model for yawed wind turbines able to predict both velocity deficit and turbulence intensity, of which the results showed good agreements with those of the RANS simulation cases.

The work mentioned above mainly concentrate on the time averaged characteristics of the turbine wake, aiming to precisely predict the velocity and turbulence of the wake downstream. However, under atmospheric ambient flow, the extension and expansion of the wake is not subject to a mathematically expressible rule at a certain instant. The stochastic behavior of turbine wake, meandering for example, has noticeable influence on the performance of downstream wind turbines. Therefore, in the present work the wake of yawed wind turbine under neutral boundary layer flow is simulated and the features of wake meandering as well as the time averaged flow field are studied.

## Numerical Method

### *Governing Equations*

Considering the superiority in simulating the unsteady large scale turbulent structures in the atmospheric flow, large eddy simulation technique is introduced in the present study. In order to simulate the ubiquitous effects of the earth's rotation and the buoyance caused by the spatial difference of temperature, the incompressible Navier-Stokes equation consists additionally of a Coriolis term and a buoyancy term, which, together with the continuity equation, constitutes the governing equation set as follows:

$$\frac{\partial \overline{u_i}}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \overline{u_i}}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u_j}\,\overline{u_i}) = \underbrace{-2\varepsilon_{i3k}\Omega_3\overline{u_k}}_{\text{I}} \underbrace{-\frac{1}{\rho}\frac{\partial}{\partial x_i}p_0(x,y)}_{\text{II}} \underbrace{-\frac{\partial \tilde{p}}{\partial x_i}}_{\text{III}} \underbrace{-\frac{\partial \tau_{ij}^D}{\partial x_j}}_{\text{IV}} \underbrace{+g\left(\frac{\overline{\theta}-\theta_0}{\theta_0}\right)\delta_{i3}}_{\text{V}} \underbrace{+\frac{1}{\rho}f_i}_{\text{VI}} \tag{2}$$

The Coriolis force is computed by term I, where $\varepsilon_{ijk}$ is the alternating tensor and $\Omega = \omega[0, \cos(\phi), \sin(\phi)]$ (the planetary rotation rate $\omega = 7.27\times10^{-5} rad/s$ ) is the rotation rate vector. The gradient of resolved-scale pressure $\overline{p}$ is divided into term II, the background pressure gradient, and term III, which consists of the gradient of one third of the stress tensor trace $\tau_{kk}/3$ and $\hat{p} = \overline{p} - p_0(x,y) + \rho_0 gz$ . $\tau_{ij}$ is the fluid stress tensor and in term

$\tau_{ij}^D = \tau_{ij} - \delta_{ij}\tau_{kk}/3$, where $\delta_{ij}$ is the Kronecker delta. The Smargorinsky eddy viscosity model is used to deal with the $\tau_{ij}^D$ in term :

$$\tau_{ij}^D = -2\upsilon^{SFS}\overline{S_{ij}} \tag{3}$$

$$\overline{S_{ij}} = \frac{1}{2}\left(\frac{\partial\overline{u_i}}{\partial x_j} + \frac{\partial\overline{u_j}}{\partial x_i}\right) \tag{4}$$

Where the sub-filter scale viscosity $\upsilon^{SFS} = (C_s\Delta)^2(\overline{S_{ij}}\,\overline{S_{ij}})^{1/2}$, $C_s$ is a constant, set to 0.14 in this work. The density $\rho$ is considered uniform in the whole domain, so the buoyancy effect has to be modeled by the Boussinesq approximation as shown in term V, where $\overline{\theta}$ and $\theta_0 = 300K$ represent the resolved scale potential temperature and reference potential temperature respectively. A temperature transport equation decoupled from the N-S equation needs to solved to obtain the potential temperature field:

$$\frac{\partial\overline{\theta}}{\partial t} + \frac{\partial}{\partial x_j}(u_j\overline{\theta}) = -\frac{\partial q_j}{\partial x_j} \tag{5}$$

$$q_j = -\frac{\upsilon^{SFS}}{\mathrm{Pr}_t}\frac{\partial\overline{\theta}}{\partial x_j} \tag{6}$$

Equation (6) models the molecular and sub-filter effects in the temperature diffusion, where the turbulent Prandtl number $\mathrm{Pr}_t = 1/3$. Term VI represents the aerodynamic force of the wind turbine blades exerted to the flow field, which is modeled by the actuator line model presented in the next part.

*Actuator Line Model*

First proposed by Sørensen and Shen[14], the actuator line model is widely used in the researches of wind turbine aerodynamics. The main idea is to directly exert the equivalent aerodynamic force to the flow field instead of building up the physical model of the turbine blades. Concretely, the blade is firstly divided into tens of airfoil segments and of each the lift force and drag force generated is expressed as:

$$L = \frac{1}{2}C_l\rho U_{rel}^2 c\,dr \tag{7}$$

$$D = \frac{1}{2}C_d\rho U_{rel}^2 c\,dr \tag{8}$$

Where $C_l$ and $C_d$ are lift and drag coefficients as function of the attack angle $\alpha$, $c$ the chord length and $dr$ the width of a airfoil segment. The relative velocity of a certain blade segment is computed as:

$$U_{rel} = \sqrt{U_x^2 + (\Omega r - U_\theta)^2} \tag{9}$$

$U_x$, $U_\theta$ and $\Omega$ are axial velocity, tangential velocity and rotation speed of the rotor respectively, among which the relationship is depicted in Fig. 1. The vector sum of the lift force and drag force $f$ is considered equal to the aerodynamic force generated by the corresponding airfoil segment. Before applying $f$ to the flow field, it is necessary to redistribute and smooth the concentrated force from one point to a ball-shaped region to avoid the possible numerical divergence. As following, the projection is implemented by taking the convolution of $f$ and the regularization kernel $\eta_\varepsilon$ :

$$f_\varepsilon = f \otimes \eta_\varepsilon \quad , \quad \eta_\varepsilon(d) = \frac{1}{\varepsilon^2 \pi^{3/2}} \exp\left[ -\left( \frac{d}{\varepsilon} \right)^2 \right] \tag{10}$$

Where $d$ represents the distance between the grid points and a certain actuator point and $\varepsilon$ is a constant that controls the concentration of the aerodynamic force.

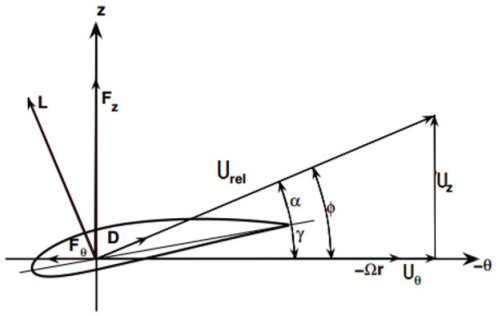



**Figure 1. Velocity vector in the cross section of an airfoil segment**

## Simulation Setup

*Computation Domain*

Basically, there are two methodologies to realize the atmospheric boundary layer inflow condition: (i) directly producing the pseudo atmospheric boundary layer inflow in the inlet plane through mathematical model based on the spectral tensor model[15] or synthetic harmonic waves[16]; (ii) simulating a fully developed boundary layer flow in a precursor case and collecting the flow data in the cross sections as the inflow condition in the successor case. In the present work, the latter is adopted because the synthetic turbulent energy cannot avoid decay as flowing downstream since no energy is generated to compensate the dissipation[17]. In order to simulate the divers scale turbulent structures in the atmosphere, which could range from a few millimeters to more than a kilometer, the width and height of the precursor domain should be large enough. Therefore, both the width and height are set to 1008m and the length is set to 2016m, so as to investigate the flow in far wake region. The layout and main dimensions of the domain are indicated in Fig. 2.

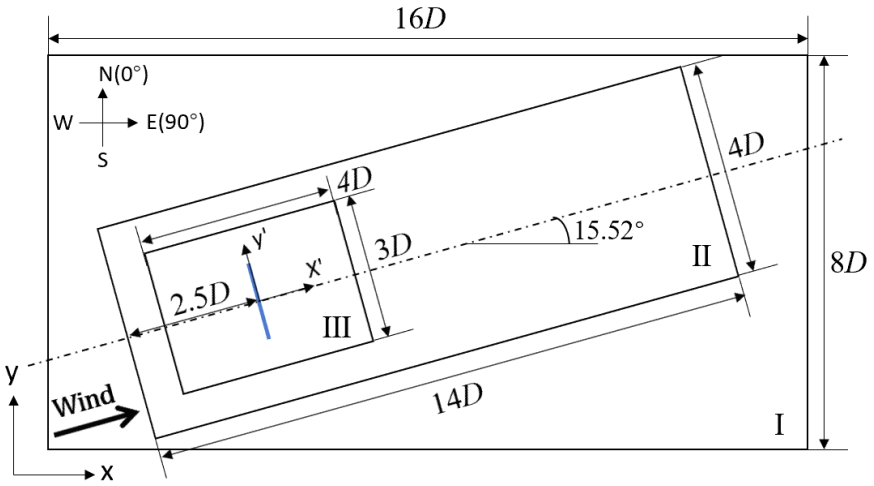



**Figure 2. Layout and main dimensions of the computation domain**

The unit grid scale in the background mesh $\Delta_I = 8m$ and in the region II that covers the vicinity

of the wind turbine and the whole wake region, the side length of every single mesh cell $\Delta_{II}$ is set to 4m. Region III with the finest mesh resolution $\Delta_{III} = 2m$ includes the turbine rotor and its vicinity so that the wind turbine aerodynamics and the vortex structure induced behind the blade tip and root can be simulated correctly. For analyzing the wake data more conveniently, a new coordinate system is created and used in the following work, with the original point located at the position of the wind turbine tower and the x-y plane rotating $15.52^\circ$ around the z axis in anticlockwise direction to align the x axis with wind direction. This coordinate system is also shown in Fig. 2.

*Boundary Conditions and Case Setup*

For the precursor case, all four vertical planes of the domain is set to cycle boundary condition, allowing the boundary flow to fully develop within a finit computation domain. The free-slip condition is applied to the top, meaning no friction and flux at this plane. Considering the real ground is not smooth but with a rough surface which cannot be directly simulated, a wall model proposed by Moeng[18] with the roughness length set to 0.001, is introduced to approximate the flow near the ground. The mean wind direction is at an angle of $15.52^\circ$ to the original x-axis and the wind speed at the hub height is prescribed to the rated wind speed of the NREL-5MW wind turbine $11.4m/s$. The computation runs with a time step of $0.5s$ until the boundary layer flow reached quasi-equilibrium state under the horizontal driving pressure. Then the simulation continues but with the time step reset to $0.02s$ and the flow data is collected as the inflow database of the successor cases.

**Table 1: Main parameters of the NREL-5MW wind turbine**

| | **Values** | **Units** |
|---|---|---|
| Rating | 5.00 | MW |
| Rotor orientation | Upwind | – |
| Blade number | 3 | – |
| Rotor diameter | 126 | m |
| Hub height | 90 | m |
| Rated wind speed | 11.4 | m/s |
| Rated rotation speed | 12.1 | rpm |

Three successor cases were set up, among which the first case contains a wind turbine rotating at $12.1rpm$ with its rotor directly facing the wind direction and in the other two cases, the wind turbine is $30^\circ$ yawed in clockwise and counterclockwise direction respectively while keeping the other conditions unchanged. The main object is to investigate the wake deflection and wake meandering phenomenon in the far wake region of the yawed wind turbine under neutral boundary layer flow through a comparative study with the non-yaw condition case. The main properties of the wind turbine is listed in Table 1.

**Results**

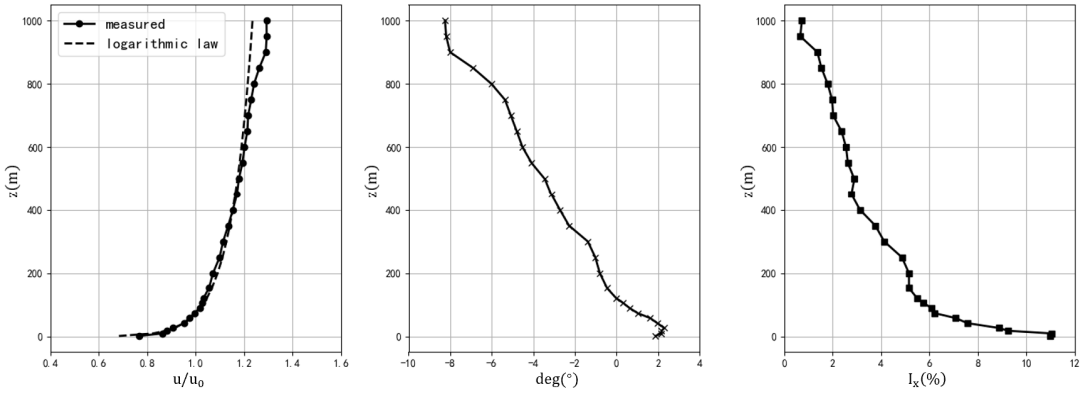
*NBL Flow Validation*

The boundary layer flow is considered fully developed after the simulation of $18000s$, with the wind speed and direction at the hub height stable at the prescribed values. Fig. 4 shows the time averaged vertical profiles of streamwise wind speed, wind direction and the streamwise

turbulent intensity which is defined as the following formula:

$$I_x(z) = \frac{\sqrt{\overline{u'u'}(z)}}{\overline{u}(z)} = \frac{\sigma_u}{u_0(z)} \tag{10}$$

Where the overbar represents time averaging, $u' = u(t) - \overline{u}$, and $u_0 = \overline{u}$ is considered as the ambient flow velocity in streamwise direction. In order to validate the velocity profile, the logarithmic law is also plotted in Fig. 3(a) and the measured velocity in the Prandtl layer accords well with the theoretical value. The three red dash lines in the picture mark the top, bottom and the hub height of the turbine rotor. Strong wind shear occurs around the hub height and the speed difference between the wind speeds at top and bottom of the rotor attains $1.68m/s$, which is a typical characteristic of the neutral boundary layer flow. As shown in Fig. 3(b), under the Coriolis effect, the wind turbine also experiences a wind direction shear of two degree. The influence of wind direction shear on the development of the wake cannot be neglected, because a deviation of $2^\circ$ across the rotor will stretch the wake transversely by $0.18D \sim 0.28D$ after the advection of $5D \sim 8D$, which is the normal streamwise spacing between two wind turbines in a wind farm. From the Fig.3(c), the streamwise turbulent intensity near the ground reaches 10% and decreases rapidly as the height rises to $200m$, indicating that properly raising the hub height of the wind turbine could significantly reduce the fatigue loads on the blades. The turbulent intensity at the hub height in the precursor case is 6.4%, a typical value of the ABL flow over a calm sea or an empty ground, which corresponds to the roughness length of 0.001 prescribed in the wall model.




(a) (b) (c)

**Figure 3. Time averaged characteristics of the atmospheric boundary layer flow in precursor case: (a)streamwise wind speed profile; (b)wind direction profile; (c)streamwise turbulent intensity profile**

*Velocity Deficit and Wake Deflection*

The velocity deficit represents the loss of the streamwise wind speed in the wake region normalized by the ambient inflow velocity, as defined by the following expression:

$$u_d = 1 - u/u_0 \tag{10}$$

The velocity deficit distributions in different downstream positions in the central vertical plane and the hub height horizontal plane are plotted in Fig. 4, with the origin of height set to hub height and the scale normalized by rotor radius $R$. As seen in Fig. 4(a), the velocity deficit profiles of the non-yawed turbine show symmetry until $6D$ downstream, after which the wind speed in the upper part of the wake region recovers faster than that of the lower part, because

the expansion of the wake is blocked by the ground and the compensation of the kinetic energy is mainly acquired through the shear layer turbulent flow in the top part of the wake. The velocity deficit in the wake of a yawed turbine is relatively small compared with that of a non-yawed turbine, because only the velocity component perpendicular to the rotor plane can be utilized. Moreover, the crosswise thrust induced by the misalignment between the turbine shaft and wind direction push the whole wake aside. For this reason, the flow data probed in the central longitudinal plane of the domain shown in Fig. 4(a) show very low level velocity deficit in the far downstream positions, which indicates a low power loss of downstream wind turbines and demonstrates the practicability of yaw operation in the wind farm optimization algorithm. From Fig. 4(b), the wake center in yaw condition deflects aside as moving downstream, with the lateral displacement reaching $1R$ in $10D$ position, and it is clearly seen in all cases that the velocity deficit curves transform from M shape in the near wake to gaussian curves after $x = 4D$. This self-similarity feature in the far wake is also one of the basic assumptions in the analytical models of wind turbine wake.

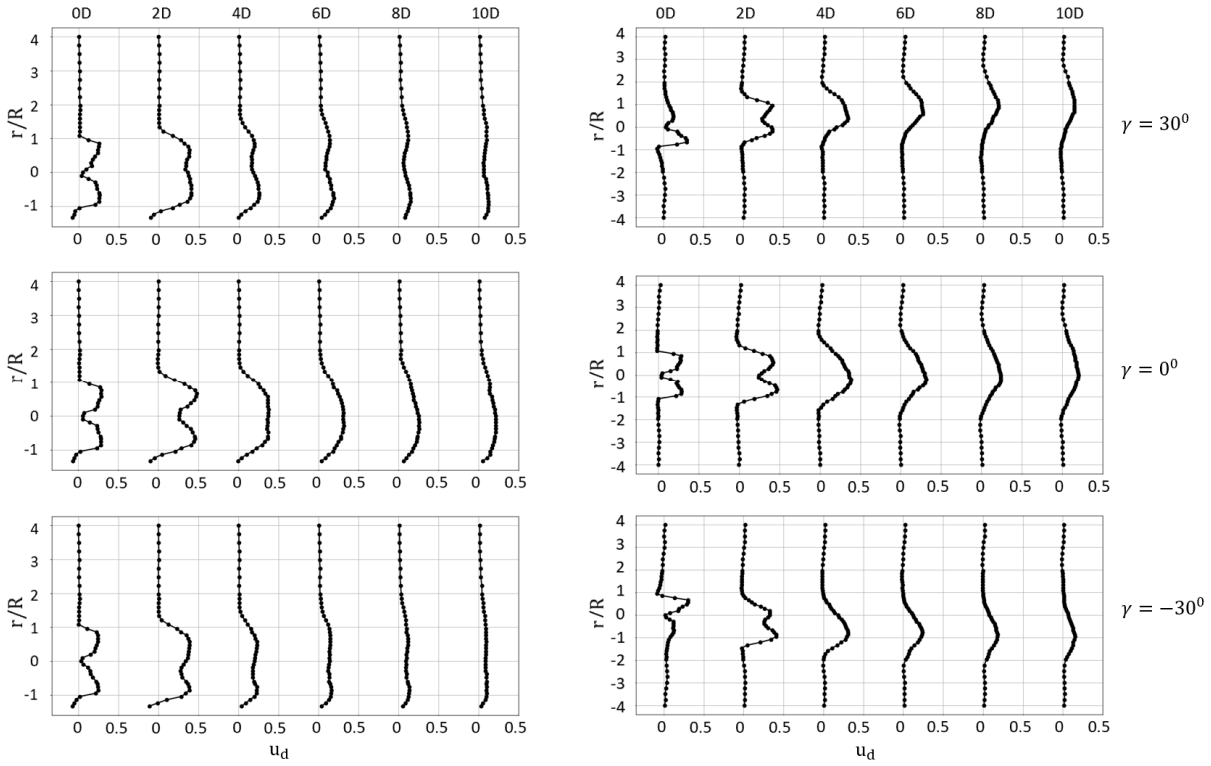



**Figure 4. Time averaged velocity deficit in different downstream positions: (a)central longitudinal plane; (b) hub height horizontal plane**

Fig. 5(a) shows the time averaged velocity deficit contours in the hub height horizontal plane of case 1~3. The profile in every $x$ position is fitted by a gaussian curve, of which the parameter $\mu$ indicates the coordinate of the wake center and plotted as black dotted line in this figure. The velocity deficit in the near wake in case 0 is much more serious than those in the other two cases and due to the larger wake width, it takes longer distance for the shear layer to fully develop and recover the wind speed around the wake center. By contrast, the effective area against the wind reduces when a turbine yaws, so the wake region narrows and the shear layers in both sides meet earlier with each other in the wake core region. Moreover, the wake deflection of the yawed turbine is also clearly seen and the time averaged wake traces of the two turbines yawed in opposite direction show symmetry in lateral displacement. The wake

skew angle ( $\theta \approx \tan\theta = u/v$ ) reveals the tendency of the lateral movement of the wake. From the wake skew angle contour in Fig. 5(b), in case 0, the outwards radial flow induced by the turbine blades weakens immediately behind the rotor, while in cases with yawed turbines, the windward part of the rotor induces the air to flow inward and this tendency becomes even more strong downstream in the near wake, dominating the whole wake region and gradually decreasing in the far wake region.

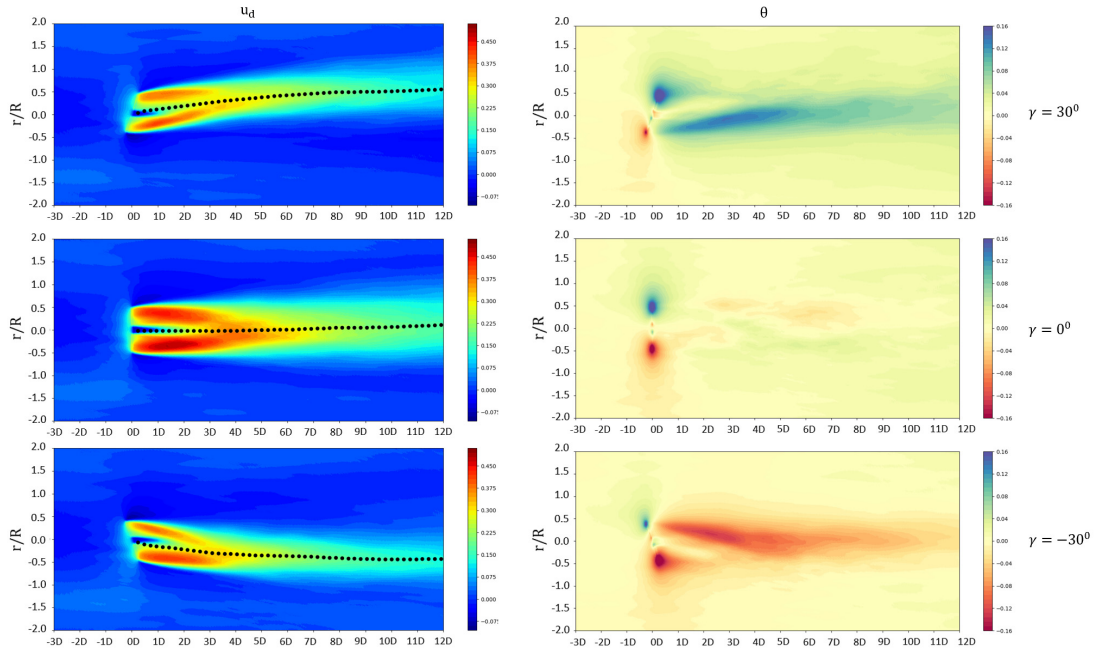



**Figure 5. Contours of time averaged wake field in the hub height horizontal plane: (a) velocity deficit; (b) wake skew angle**

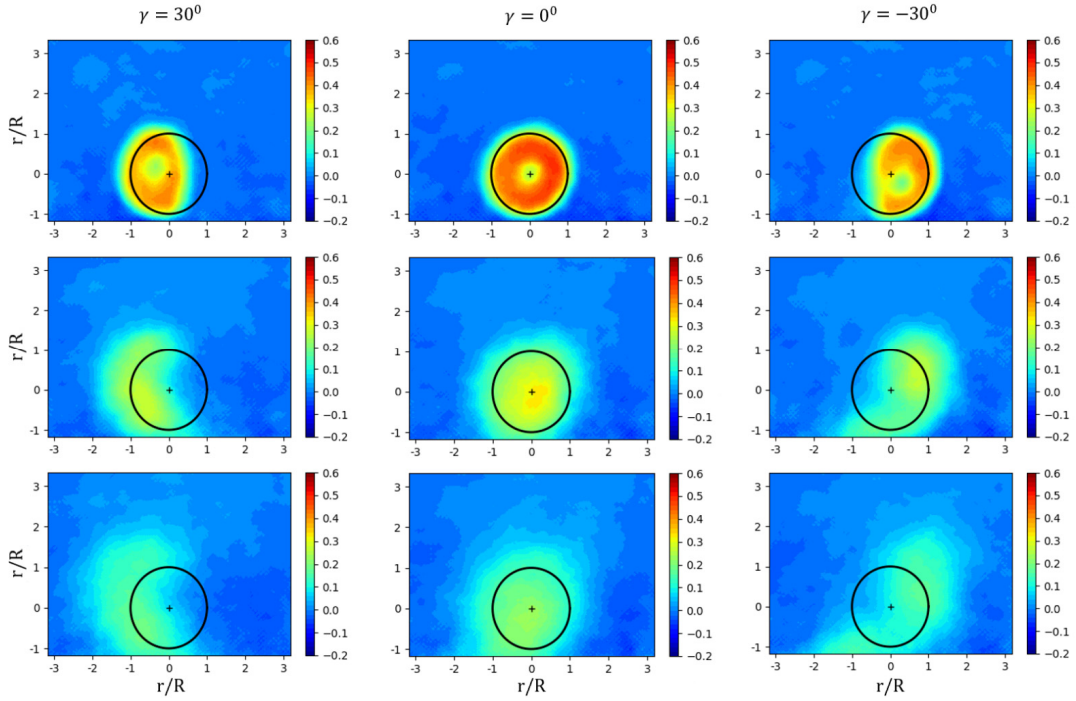



**Figure 6. Contours of time averaged velocity deficit in different cross sections**

To illustrate the wake deflection in both $y$ and $z$ direction, the contours of time averaged velocity deficit of $x = 2D, 6D, 10D$ cross sections are shown in Fig. 6, with the hub center marked by a plus sign and the rotor edge marked by a black circle. For $x = 2D$ position, the wakes in case 1 and 2, compared with the round wake region in case 0, show the shape of a ellipse and the wake center of the turbine with $\gamma = 30°$ moves upward and the other moves downward. In $6D$ downwind location, the wake changes into kidney-shape, caused by the forming of a counter-rotating vortex pair in the far wake region of a turbine with high yaw angle[8] and the wake remain this cross section shape until $x = 10D$. Furthermore, the two wakes of turbines yawed in opposite directions are not completely symmetric and the vertical deflection of the wake center seems to have the same order of the magnitude as the lateral. Thus the yaw direction and the vertical wake deflection cannot be neglected when predicting and describing the wake flow of a turbine in yaw condition.

*Wake Meandering*

As seen in the Fig. 7, which shows the instantaneous velocity deficit field in the hub height plane, The trace of the wake center of an instant marked by the solid line is not a smooth line, but wiggles around the average center line. This phenomenon called wake meandering is caused by the large-scale turbulent structure in the atmosphere and the meandering is not completely random but demonstrates certain statistical properties. It should be noted that the velocity data used in this part are all filtered temporally with a $\tau$-window:

$$\hat{u} = \frac{1}{\tau}\int_{t-\tau/2}^{t+\tau/2} u(t)dt, \quad \tau=3\text{s} \tag{11}$$

The purpose is to conserve the large scale flow structure and eliminate the high frequent turbulence, which could lead to the bias in the estimation of the wake meandering. The value $\tau = 0.6T \approx 3s$ is chosen as same as the window width used by Foti[19].

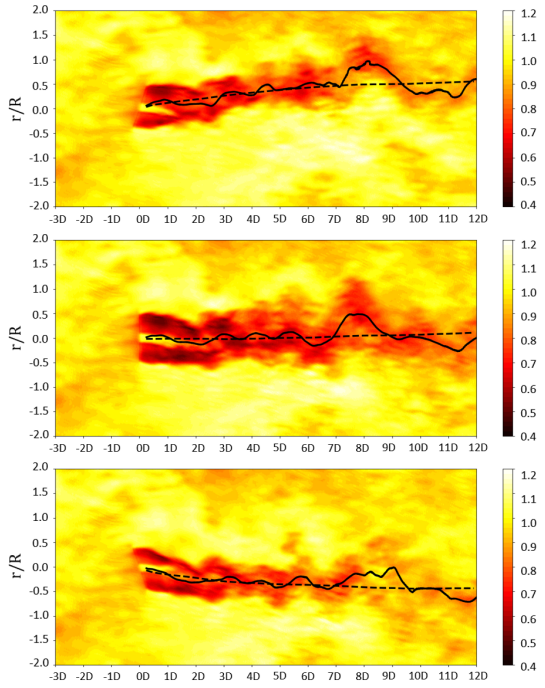



**Figure 7. Filtered velocity deficit field in hub height plane of $t = 380s$**

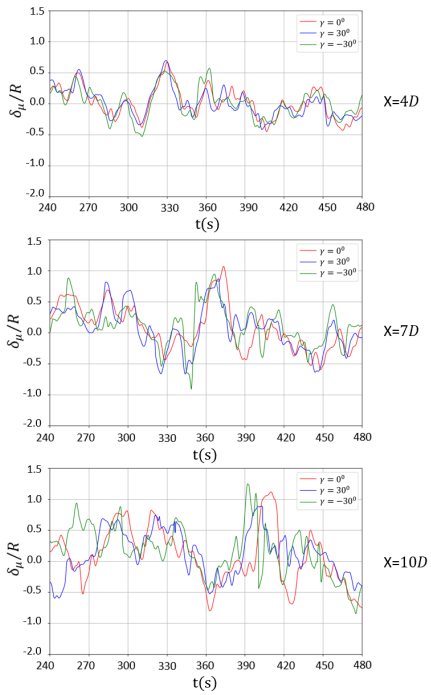



**Figure 8. Time history curves of $\delta_\mu$ in different downstream positions**

The Fig. 7 shows that the levels of the wake meandering in all three cases are nearly the same, although the velocity deficit in case 0 is larger than those in yaw condition, and the amplitudes of the wiggle increases as it moves downstream. The lateral displacement of an instantaneous wake center relative to its average value in a certain downstream position, defined as $\delta_\mu(t) = \mu(t) - \bar{\mu}(t)$ is used to describe the scale of the wake meandering. Because $4D \sim 10D$ is the common streamwise spacing between turbines in the wind farm, the study of the wake meandering in this range has more reference value for the engineering practice. The $\delta_\mu(t)$ in the $x = 4D,\ 7D,\ 10D$ locations are acquired by fitting the gaussian curve to the temporally filtered velocity deficit profile of the three positions and the time history curves of $\delta_\mu(t)$ normalized by rotor radii $R$ for case 1~3 are plotted in Fig. 8. The results show that there is strong correlation between the curves for the same position of three cases. Considering the inflow data of case 1~3 are exactly the same (collected from the same precursor case), the similarity of the time history meandering curves of different cases provides the strong evidence for the assumption made by Larsen et al. in his work[20], which indicates that the wake acts as a passive tracer driven by the large-scale turbulence structures in the atmospheric boundary layer.

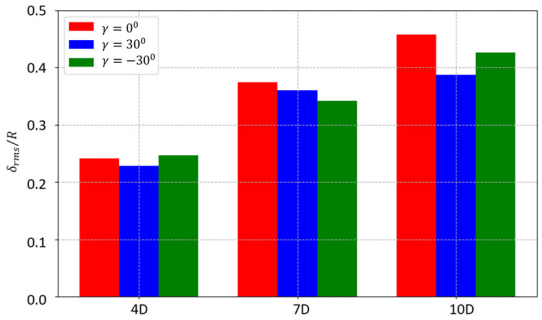



**Figure 9. Root mean square of $\delta_\mu$**

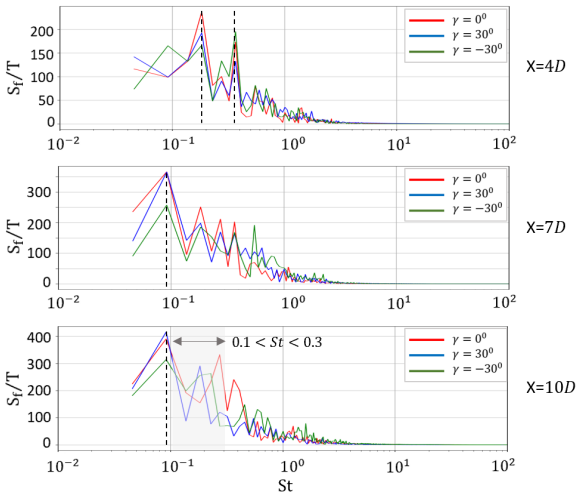



**Figure 10. Frequency spectrum of $\delta_\mu$**

To investigate the influence of yaw condition and downstream distance on the intensity of the wake meandering, the root mean square of $\delta_\mu(t)$ is computed with the results shown in Fig. 9. Apparently the wake meandering intensity raises as the wake moves downstream, which is also demonstrated in Fig. 7 and 8. Moreover, by comparison, the influence of yaw condition on the wake meandering intensity is inappreciable, because the difference of rms among three cases is less than $0.05R$, which can be neglected in the performance prediction of a wind farm. The normalized frequency spectrums gained by the FFT of $\delta_\mu(t)$ series are depicted in Fig. 10. The spectrum for $x = 4D$ shows two distinct peaks corresponding to $St = 0.18$ and $St = 0.36$ respectively in all cases, while in the spectrum for $x = 7D$, one dominant frequency $St = 0.09$ and several subpeaks with $0.1 < St < 0.4$ and almost the same amplitude are detected. The main frequency with the highest amplitude of the meandering for $4D$ downstream position corresponds to a period $T = D/(u_0 St) \approx 120s$, and this dominant frequency keeps the same value for $10D$ downstream position. But the subpeaks with relatively high frequency weaken as $x$ increases and only the secondary peaks, of which the Strouhal number is between 0.1 ~ 0.3, remains remarkable until $x = 10D$, as illustrated in the last graph in Fig 10. The dominant low frequent meandering from $x = 4D$ to $x = 10D$ with a period up to 120s

is supposed to be affected by the kilometer scale flow structure in the atmospheric boundary layer, and the secondary frequency component of the wake meandering, with $0.1 < St < 0.3$, is also found in the experimental work by Okulov[21].

## Conclusion

Performance optimization of a wind farm through yaw control is a realizable and promising technique but still needs a deep understanding of the wake characteristics of a yawed wind turbine operating in the atmospheric boundary layer. In the present work, the wake flow of a single wind turbine with and without a yaw angle is studied through actuator line model and LES technique. The velocity and turbulence intensity profiles of the flow generated in the precursor case show a typical stratification of a neutral boundary layer flow in the atmosphere and accord well with the prescribed wind velocity and direction.

For a turbine operating in yaw condition, the time-average velocity deficit is generally smaller than that of a non-yawed turbine, because the misalignment leads to the decrease of the effective wind speed, which is a function of the yaw angle $\gamma$. In the central longitudinal plane, the velocity deficit in far wake region behind a turbine with a yaw angle of $\pm 30°$ is much weaker than that behind a non-yawed turbine, indicating the power output of a turbine downwind could be raised significantly by yawing the turbines located upstream. Moreover, the profiles of the velocity deficit in all cases show self-similarity and obey well with the gaussian distribution in the horizontal plane. Nevertheless, through the contours of $u_d$ in different cross sections downstream, the velocity distribution in the far wake of a yawed wind turbine is not symmetric but shows a kidney-shape and the direction of the vertical deflection of the wake center depends on the yaw direction, which reduces the accuracy of the analytical wake model based on the gaussian distribution assumption.

A filtering process proposed by Howard[22] is introduced to smear out the high frequent turbulence and highlight the wake meandering movement. The time history curves of the meandering for case 1~3 show high correlation, demonstrating that the large-scale movement of the wake follows the flow structure with the same scale in the incoming flow. The rms values of $\delta_\mu$ for all cases increase as the wake flows downstream and show nearly the same magnitude until $x = 10D$. Furthermore, regardless of the yaw condition, the frequency spectrum analysis shows a constant dominant frequency in the far wake region, caused by the low frequent turbulence in the atmosphere and other subpeaks frequency components with $0.1 < St < 0.3$ are found from $4D$ to $10D$ downstream positions, which is a main factor that influences the fatigue loads of the wind turbines in a wind farm. Future work should focus on the wake meandering characteristics of wind turbines under different atmospheric stratifications and its influence on the performance and aerodynamic loads of downstream turbines.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

## Reference


[1] Fleming, P. A., Scholbrock, A. K., Jehu, A., Davoust, S., Osler, E., Wright, A. D., & Clifton, A. (2014). Field-test results using a nacelle-mounted lidar for improving wind turbine power capture by reducing yaw misalignment. In *Journal of Physics: Conference Series* (Vol. 524, No. 1, p. 012002). IOP Publishing.

[2] Fleming, P., Gebraad, P. M., Lee, S., van Wingerden, J. W., Johnson, K., Churchfield, M., ... & Moriarty, P. (2015). Simulation comparison of wake mitigation control strategies for a two-turbine case. *Wind Energy*, *18*(12), 2135-2143.

[3] Gebraad, P. M. O., Teeuwisse, F. W., Van Wingerden, J. W., Fleming, P. A., Ruben, S. D., Marden, J. R., & Pao, L. Y. (2016). Wind plant power optimization through yaw control using a parametric model for wake effects—a CFD simulation study. *Wind Energy*, *19*(1), 95-114.

[4] Grant, I., Parkin, P., & Wang, X. (1997). Optical vortex tracking studies of a horizontal axis wind turbine in yaw using laser-sheet, flow visualisation. *Experiments in fluids*, *23*(6), 513-519.

[5] Haans, W., van Kuik, G. A. M., & Van Bussel, G. J. W. (2007). Experimentally observed effects of yaw misalignment on the inflow in the rotor plane. In *Journal of Physics: Conference Series* (Vol. 75, No. 1, p. 012012). IOP Publishing.

[6] Krogstad, P. Å., & Adaramola, M. S. (2012). Performance and near wake measurements of a model horizontal axis wind turbine. *Wind Energy*, *15*(5), 743-756.

[7] Bartl, J. M. S., Mühle, F. V., Schottler, J., Sætran, L. R., Peinke, J., Adaramola, M. S., & Holling, M. (2018). Wind tunnel experiments on wind turbine wakes in yaw: effects of inflow turbulence and shear.

[8] Bastankhah, M., & Porté-Agel, F. (2015). A wind-tunnel investigation of wind-turbine wakes in yawed conditions. In *Journal of Physics: Conference Series* (Vol. 625, No. 1, p. 012014). IOP Publishing.

[9] Bastankhah, M., & Porté-Agel, F. (2016). Experimental and theoretical study of wind turbine wakes in yawed conditions. *Journal of Fluid Mechanics*, *806*, 506-541.

[10] Bastankhah, M., & Porté-Agel, F. (2019). Wind farm power optimization via yaw angle control: A wind tunnel study. *Journal of Renewable and Sustainable Energy*, *11*(2), 023301.

[11] Jiménez, Á., Crespo, A., & Migoya, E. (2010). Application of a LES technique to characterize the wake deflection of a wind turbine in yaw. *Wind energy*, *13*(6), 559-572.

[12] Vollmer, L., Steinfeld, G., Heinemann, D., & Kühn, M. (2016). Estimating the wake deflection downstream of a wind turbine in different atmospheric stabilities: an LES study. *Wind Energy Science*, *1*(2), 129-141.

[13] Qian, G. W., & Ishihara, T. (2018). A New Analytical Wake Model for Yawed Wind Turbines. *Energies*, *11*(3), 665.

[14] Sørensen, J. N., & Shen, W. Z. (2002). Numerical modeling of wind turbine wakes. Journal of fluids engineering, 124(2), 393-399.

[15] Mann, J. (1998). Wind field simulation. *Probabilistic engineering mechanics*, *13*(4), 269-282.

[16] Kraichnan, R. H. (1970). Diffusion by a random velocity field. *The physics of fluids*, *13*(1), 22-31.

[17] Troldborg, N., Sørensen, J. N., & Mikkelsen, R. F. (2009). Actuator line modeling of wind turbine wakes.

[18] Moeng, C. H. (1984). A large-eddy-simulation model for the study of planetary boundary-layer turbulence. Journal of the Atmospheric Sciences, 41(13), 2052-2062.

[19] Foti, D., Yang, X., Guala, M., & Sotiropoulos, F. (2016). Wake meandering statistics of a model wind turbine: Insights gained by large eddy simulations. Physical Review Fluids, 1(4), 044407.

[20] Larsen, G. C., Aagaard, H. M., Bingöl, F., Mann, J., Ott, S., Sørensen, J. N., ... & Larsen, T. J. (2007). Dynamic wake meandering modeling.

[21] Okulov, V. L., Naumov, I. V., Mikkelsen, R. F., Kabardin, I. K., & Sørensen, J. N. (2014). A regular Strouhal number for large-scale instability in the far wake of a rotor. *Journal of Fluid Mechanics*, *747*, 369-380.

[22] Howard, K. B., Singh, A., Sotiropoulos, F., & Guala, M. (2015). On the statistics of wind turbine wake meandering: An experimental investigation. Physics of Fluids, 27(7), 075103.

# Unsteady Aerodynamics of a Spar-type Floating Offshore Wind Turbine Induced by Platform Pitch Motion

**Yang Huang  Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

## Abstract

There are strong interactions between the aerodynamics and hydrodynamics in the floating offshore wind turbine (FOWT) system. The aerodynamic loads of wind turbine acting on the floating support platform via turbine tower enlarge the motion responses amplitudes, which in return alters the aerodynamics of wind turbine. How the platform pitch motion interact with the wind turbine aerodynamics under operating wind-wave conditions is an attractive research. In the present work, the unsteady aerodynamics of the FOWT are investigated numerically with an unsteady actuator line model (UALM) that take the additional relative wind speed induced by the platform motions into consideration. To investigate the influence of platform pitch motion on unsteady aerodynamic characteristic, coupled aero-hydrodynamic simulations of a spar-type FOWT with and without pitch motion under shear wind and regular wave are both performed. The aerodynamic characteristics including the rotor power, thrust, fatigue loads and detailed wake field information are analyzed. Furthermore, the relative wind velocity, attack angle and the bending moments at blade root for different simulation conditions are compared and discussed to explore the intrinsic relationship between platform pitch motion and unsteady aerodynamics. It can be found that the average and oscillating amplitude of pitch responses under operating wind-wave loads increase remarkably due to aerodynamic forces. The dramatic change of aerodynamic loads significantly alters the forces acting on the rotating blades with a result of rapidly increased fatigue loads and instability problem. Thus, complicated control strategies are supposed to apply in the FOWT system to suppress the motion responses of floating platform.

**Keywords**: Floating wind turbines; Pitch motion; Unsteady aerodynamics; Fatigue loads; Unsteady actuator line model

## Introduction

Wind energy is thought to be one of the most promising renewable energy due to enormous reserves, and the wind power technology have become more mature in the past decade [1]. With the depletion of land resources for onshore wind farms, the offshore regions with much stronger and smoother wind become a better choice. In shallow waters, bottom-fixed offshore wind turbines have achieved great success. However, the noise restriction and visual pollution limit its further development and application. The offshore wind turbines are advancing into deep water areas [2][3]. Considering that the cost of offshore wind turbines mounted on bottom-fixed structures increases sharply with water depth, the floating offshore wind turbine (FOWT) is generally believed to an alternative [4][5]. Several countries have planned to build floating wind farms. From the perspective of practical deployment, there are still some challenges for the

FOWT, especially the stability problem. Different from the onshore wind turbine and bottom-fixed wind turbine, the FOWT consisting of a wind turbine, a floating platform and a mooring system is a rather complex system and suffers various environment excitations from winds, waves and currents. There are strong interactions between the wind turbine and floating support platform. The aerodynamic loads of wind turbines acting on the floating support platform via turbine tower will enlarge the motion responses amplitudes, which in return alters the aerodynamics of wind turbine [6]. The study of Sebastian [7] indicated that the unsteady aerodynamics of wind turbine were significantly influenced by the surge, pitch and yaw motions of the floating support platform. The platform pitch motion that results in a non-uniform flow on the turbine rotor and more complex blade-wake interaction significantly affects the aerodynamic performance.

In order to investigate how the platform pitch motion influences the unsteady aerodynamics of wind turbine and explore the interaction mechanism, a number of experimental and numerical researches have been conducted. Leble and Barakos [8] studied the aerodynamic performance of DTU 10MW wind turbine with prescribed sinusoidal pitching and yawing motions. They found that the mean power for pithing amplitudes of 5 deg is 32.8% larger than that without pitching motion. It indicated the pitch motion was advantageous to improve the mean power output of the wind turbine. However, the platform pitch motion also had adverse effects on the stability of aerodynamic power. The numerical research conducted by Tran and Kim [9] showed that the instantaneous aerodynamic power of the NREL 5MW wind turbine varied from 0MW to 15MW when the wind turbine experienced a platform pitch motion with an amplitude of 4 deg and a period of 5s. Besides, they have conducted a series of CFD simulations focusing on the influence of platform pitch motion on the unsteady aerodynamics and wake characteristics. The rotating blades and generated wake vortices were found to have strong interactions with each other. And the aerodynamic loads presented highly unsteady characteristics due to the platform pitch motion [10][11]. Consequently, the aerodynamic performance of the FOWT greatly affected by the platform pitch motion. Several studies have tried to reveal the interaction mechanism between the platform pitch motion and the unsteady aerodynamics of the wind turbine. Wen et al. [12] investigated the influence of the platform pitch amplitude and frequency on the power performance using Free Vortex Method (FVM). It was concluded that the impacts of platform pitch motion on mean power output had great discrepancy when the tip speed ratio changed. And they proposed a platform-pitch-induced (PPI) wind shear model in their later work to explain the influence of platform pitch motion on unsteady aerodynamics and investigated the influencing factors of the PPI wind shear [13]. Wind tunnel experiments were also carried out to model the wake characteristics of wind turbine. Rockel et al. [14] used Particle Image Velocimetry (PIV) technique to observe the development of a model wind turbine wake and discussed the influence of platform pitch motion on aerodynamic power and wake characteristics. Khosravi et al. [15] performed an experimental study with 1:300 scaled model wind turbine to analyze the influence of platform pitch motion on aerodynamic loading and turbine wake characteristics. The fatigue loads were detected to be increased remarkably due to the platform pitch motion, which will lead to the decrease of the lifetime of wind turbine blades.

It can be found that the existing researches about the effects of platform motions on the aerodynamic performance of wind turbine mainly focus on the prescribed platform pitch motion while not the pitch response of floating support platform in realistic environment. In our previous work, the coupling effects between the aerodynamics of the wind turbine and the six-degree-of-freedom platform motions under wind-wave conditions have been investigated by comparing the coupled aero-hydrodynamics including aerodynamic forces, platform motion responses and mooring tensions [16]. Simplified force model that assumes the time-varying

aerodynamic forces acting on the floating support platform as constant loads and coupled analysis model FOWT-UALM-SJTU were both applied to simulate the aero-hydrodynamic performance of the FWOT to detect the coupling relationship between aerodynamics and hydrodynamics in FOWT system. However, how the different degree-of-freedom platform motions under realistic wind-wave loads, especially for the pitch responses, influence the unsteady aerodynamics of wind turbine is still to be further discussed. And it is exactly what we concern in the present work. In order to explore the impacts of platform pitch motion on the unsteady aerodynamics and wake characteristics under operating conditions, coupled aero-hydrodynamic simulations for a spar-type FOWT under combined wind and waves are conducted. And the coupled CFD analysis tool FOWT-UALM-SJTU [17] is chosen again to model the aero-hydrodynamic performance. The aerodynamic characteristics including the rotor power, thrust, fatigue loads and detailed wake field information are analyzed to reveal the detailed relationship between the platform pitch motion and the unsteady aerodynamics of wind turbine.

## Numerical Method

*Unsteady Actuator Line Model*

Full scale CFD simulations for the wind turbine are quite time consuming and requires a significant amount of computing resources while detailed flow field information can be obtained. The Blade Element Theory (BEM) and Free Vortex Method (FVM) have the characteristics of fast accurately calculating, but the detailed wake characteristics cannot be acquired. The actuator line model (ALM) proposed by Sørensen and Shen [18], which combined the advantages of these methods, is chosen in the present work to model the aerodynamics performance of the FOWT. The real blade surfaces of the wind turbine are replaced with virtual actuator lines in the ALM. Each actuator line is further discretized into a serial of actuator points and each actuator point represents a section of the blade. The aerodynamic force acting on the blades are calculated from the local attack angle and a look-up table of airfoil data. Then the calculated body forces smeared by regularization kernel function are introduced into the moment equations to reproduce the turbulent wake flow.

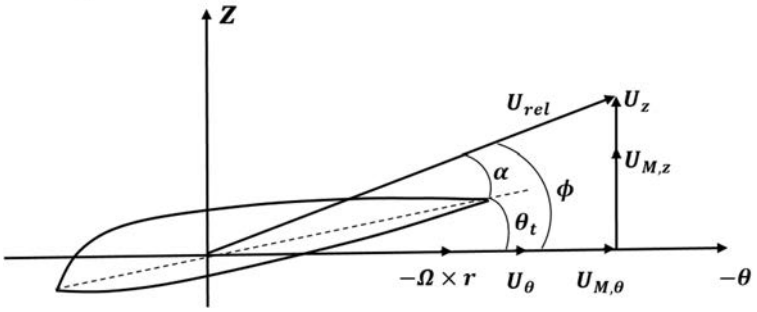



**Fig. 1 Velocity at cross-sectional airfoil element**

In order to model the unsteady aerodynamic characteristics of the FOWT, modifications are made to the initial ALM. As Fig. 1 shows, the additional velocity $\boldsymbol{U_M}$ induced by the platform motion, which intensify the interaction phenomenon between rotating blades and wake field, is taken into consideration in the calculation of local attack angle. To determine the body forces acting on the blades, a blade element approach combined with two-dimensional airfoil characteristics is used. To illustrate the relationship between different velocities at cross-sectional airfoil element, a cross-sectional element at radius r defines the airfoil at the ($\theta$, z) plane. Denoting the tangential and axial velocity in the inertial frame of reference as $\boldsymbol{U_\theta}$ and $\boldsymbol{U_z}$, respectively.

The local velocity $U_{rel}$ relative to the rotating blade is defined as:

$$\boldsymbol{U}_{rel} = \boldsymbol{U}_{\theta} - \boldsymbol{\Omega} \times \boldsymbol{r} + \boldsymbol{U}_{z} + \boldsymbol{U}_{M} \tag{1}$$

Where $\Omega$ is the rotating speed of the wind turbine. Then the attack angle can be calculated by the following equation:

$$\alpha = \phi - \theta_t \tag{2}$$

Where $\phi$ is the inflow angle. $\theta_t$ is the local twist angle. The aerodynamic forces can be obtained by the following equation:

$$\boldsymbol{f} = (\boldsymbol{L}, \boldsymbol{D}) = \frac{\rho |U_{rel}|^2 c N_b}{2 r d\theta dz} (C_L \boldsymbol{e}_L + C_D \boldsymbol{e}_D) \tag{3}$$

Where $c$ is the chord length; $N_b$ is the number of blades; $C_L$ and $C_D$ are the lift and drag coefficient, respectively; $\boldsymbol{e}_L$ and $\boldsymbol{e}_D$ denote the unit vectors in the directions of the lift and the drag, respectively.

To reproduce the turbulent wake flow, the calculated aerodynamic forces need to be smeared before they are introduced into the moment equations to avoid singular behavior in numerical simulations.

$$\boldsymbol{f}_{\varepsilon} = \boldsymbol{f} \otimes \eta_{\varepsilon}(d) = \frac{1}{\varepsilon^3 \pi^{3/2}} exp\left[ -\left(\frac{d}{\varepsilon}\right)^2 \right] \tag{4}$$

Here $\boldsymbol{f}_{\varepsilon}$ is the source term added into the right hand of momentum equation. $d$ is the distance between the measured point and the initial force points on the rotor. $\varepsilon$ is a constant which serves to adjust the strength of regularization function.

*FOWT-UALM-SJTU solver*

Using the UALM to model the unsteady aerodynamics of wind turbine, a coupled CFD analysis tool FOWT-UALM-SJTU for FOWT is established by interpolating the UALM into in-house code naoe-FOAM-SJTU. This tool is utilized in the present work to achieve the coupled aero-hydrodynamic simulations for the FOWT. The in-house code naoe-FOAM-SJTU solver based on the open source CFD toolbox OpenFOAM is developed to investigate typical hydrodynamic problems of ship and marine engineering. It is composed of a 3D numerical wave tank module, a 6DOF motion module, a mooring system module and the interface module with OpenFOAM. It is applied to study the hydrodynamics of the spar-type floating support platform with a mooring system. The volume of fluid (VOF) method with bounded compression technique is utilized to capture the free surface, and a dynamic deformation mesh approach is employed to handle structure motions. The piecewise extrapolating method (PEM) is chosen to solve the constraint of the mooring line system.

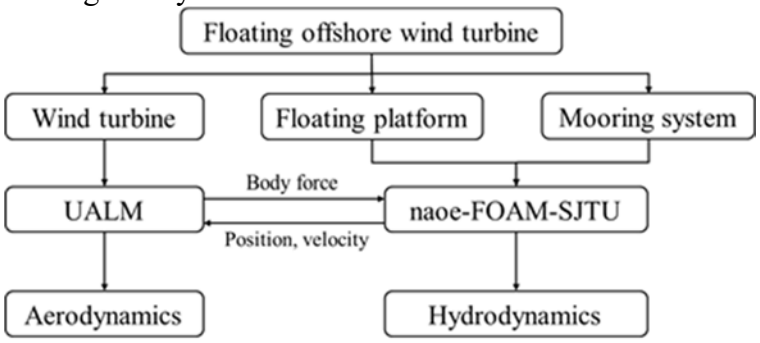



**Fig. 2 Schematic diagram of the coupled analysis tool**

*Governing equations*

Considering the wind speed is low, the air phase is regarded as incompressible liquid as the water phase is. The same governing equations are adopted to solve these two-phase flow conditions. The three-dimensional Reynolds-Averaged Navier-Stokes (RANS) equations are selected as governing equations for the FOWT-UALM-SJTU model, and the k-ω SST turbulence model is employed for closure of RANS equations.

$$\nabla \cdot \boldsymbol{U} = 0 \tag{5}$$

$$\frac{\partial \rho \boldsymbol{U}}{\partial t} + \nabla \cdot \left(\rho\left(\boldsymbol{U} - \boldsymbol{U}_g\right)\right)\boldsymbol{U} = -\nabla p_d - \boldsymbol{g} \cdot x \nabla \rho + \nabla \cdot \left(\mu_{eff} \nabla \boldsymbol{U}\right) + (\nabla \boldsymbol{U}) \cdot \nabla \mu_{eff} + \boldsymbol{f}_\sigma + \boldsymbol{f}_s + \boldsymbol{f}_\varepsilon \tag{6}$$

Where $\boldsymbol{U}$ is flow velocity in computation domain; $\boldsymbol{U}_g$ is the velocities of flow field on the grid nodes; $p_d = p - \rho \boldsymbol{g} \cdot x$ is the dynamic pressure; $\boldsymbol{g}$ is the gravitational acceleration vector; $\rho$ is the mixture density with two phases; $\mu_{eff} = \rho(\nu + \nu_t)$ is effective dynamic viscosity, in which $\nu$ and $\nu_t$ are kinematic viscosity and eddy viscosity respectively; $\boldsymbol{f}_\sigma$ is the surface tension term in two phases model and takes effect only on the liquid free surface; $\boldsymbol{f}_s$ is the source term for sponge layer, which is set to avoid the wave reflection at the end of the computation domain and takes effect only in sponge layer.

## Simulation descriptions

*Analysis object*

A spar-type FOWT consisting of the NREL 5-MW baseline wind turbine and the OC3 Hywind spar-type floating platform with catenary mooring lines is selected in the present work to investigate the influence of platform pitch motion on the unsteady aerodynamics of wind turbine. The NERL 5-MW wind turbine is a conventional three-bladed, upwind, variable-speed and variable blade-pitch-to-feather controlled turbine [19]. The floating support platform a spar-type concept platform called Hywindspar [20]. Three catenary mooring lines are arranged around the platform to limit the platform motions and keep the stability of the FOWT. The sketch of the spar-type FOWT is shown in Fig. 3. To simplify the aerodynamics modelling, the tower, hub and nacelle are not taken into account. And there is no control strategy for the wind turbine. Different from our previous work [16] that simply the aerodynamics of wind turbine with simplified force model to focus on the hydrodynamic responses of floating platform and the coupling effects between aerodynamic and hydrodynamics in the FOWT system, the major objective of the present work is to investigate the unsteady aerodynamics of the wind turbine with realistic platform pitch motion under operating wind-wave loads. In order to detect how the platform pitch motion under operating wind-wave conditions affects the aerodynamics of the wind turbine, two simulation cases with different platform state are performed using FWOT-UALM-SJTU model in the present work. In case 1, the floating support platform remains stationary. While the platform pitch motion is taken into consideration in case 2. Wind and wave conditions in these two cases keep the same, which are referenced to Jonkman's work [20]. The exponential model is adopted to describe the characteristic of height-dependent wind speed. The wind speed at the height of z is defined by following equation:

$$u_Z = u_0 \times \left(\frac{z}{90}\right)^{0.143} \tag{7}$$

Where $u_0$ is the wind speed at the height of the turbine hub. In this study, the rated wind speed $u_0 = 11.4$m/s is selected for the analysis of unsteady aerodynamics. The corresponding turbine rotor speed is 12.1 rpm. The incident wave is first order Stocks wave. The wave height and wave length are 6m and 10s, respectively. There is no control algorithm for the wind turbine in the present simulations.

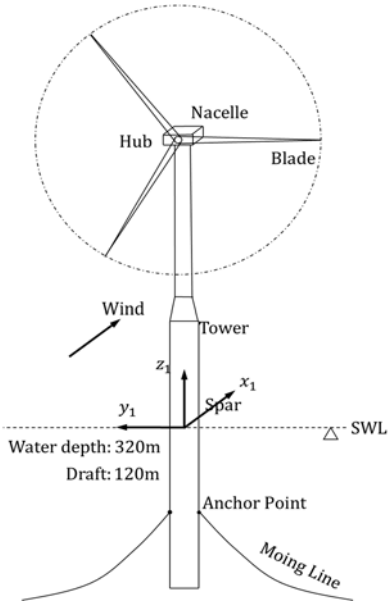



**Fig.3 Diagram of the FOWT system**

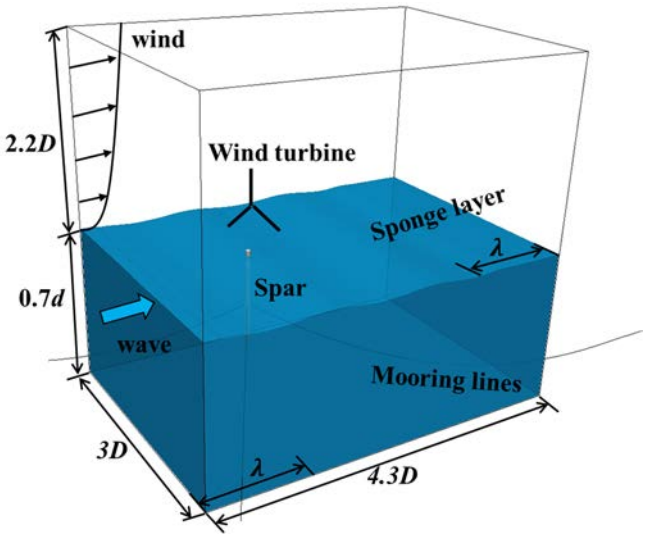



**Fig. 4 Computation domain**

*Computation Domain*

A hexahedral computational domain is applied for the present numerical simulations. The length and width are set to 540m and 400m respectively, about a dimension of 4.3$D$×3$D$ ($D$=126m is the diameter of the NREL 5-MW wind turbine). Considering the expansion of turbine wake, the height of the air phase is set to 280m (about 2.2$D$). To decrease the grid number and limit the computation resources, the depth of water phase is set to 224m that is the 70% of real water depth ($d$=320m), for this water depth is deep enough to neglect the influence of water depth on platform motion response. A rectangle sponge layer with length is selected to avoid the effects of wave reflection. The FOWT system is located in the middle of computation domain along the x direction, 1$\lambda$ ($\lambda$=156m is the wave length of incident wave) behind the inlet boundary. The main parameters of the computation domain are shown in Fig. 4.

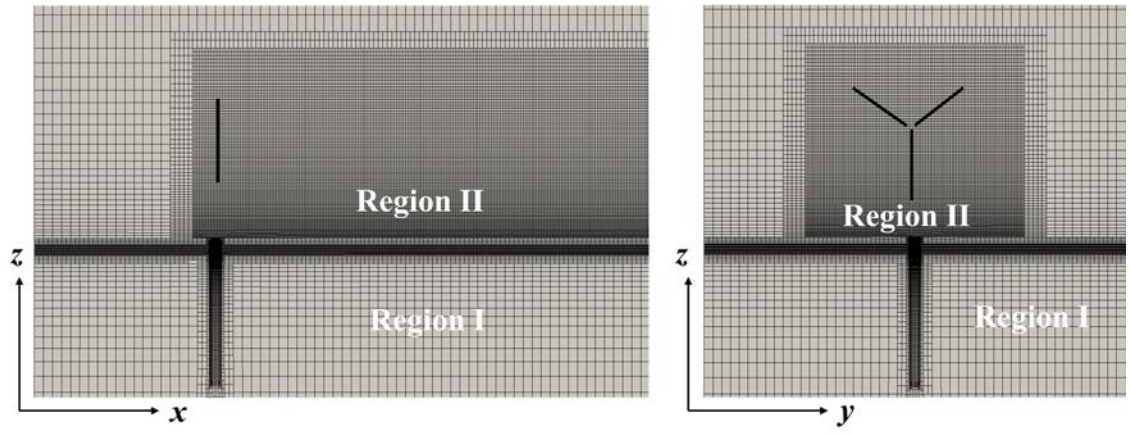



**Fig. 5 Grid distribution in longitudinal section and cross section**

To capture detailed wake behavior and water surface, the grids behind the wind turbine and near the free surface are refined. As Fig. 5 shows, region I is the background mesh, where the grid size is 8m × 8m × 8m. Region II represents the refined mesh with the grid size of

2m×2m×2m. And the grids near free surface are generated with the size of 2m×2m×0.5m. The total grid number is 3.5 million, which is affordable to achieve the coupled aero-hydrodynamic simulations for FOWT.

## Results and Discussions

*Aerodynamic load characteristics*

In this study, aeroydnamic performance of the FOWT with platform pitch motion is compared with that of fixed wind turbine to explore the influence of pitch responses the floating support platform in operating wind and wave conditions on the unsteady aerodynamic characteristics. Simulation time for all coupled cases is 180s while only the aero-hydrodynamics in the last 20s are analyzed, as the coupled performance of the FOWT have stabilized during this period. Duo to the contribution of wind loads, the pitch response of floating support platform becomes great and oscillates periodically, which furhtermore makes the aerodynamic characteristics change in the same tendency. It can be seen from the Fig. 6 that the oscillating amplitude of pitch motion for the flaoting support platform under wind and wave loads is nearly 2 degrees, almost half of the average value of the pitch response. The similar phenomenon can also be found in our previous work [16]. This significantly variation of the platform pitch moiton transfers to the turbine rotor via tower and potentially influences the relative wind velotiсy at the cross sections of rotating blades, amplifying the cyclical change of the local attack angle. To clearly show the effects of pitch motion, the relative wind velocity including the axial and tangential wind velocities during the rotation of blades and attack angle with respect to azimuth angle for the blade #1 at a typical blade section $r/R$ = 0.8 ($R$=63m is the radius of the turbine blade) are presented in Fig. 7 and Fig. 8, respectively. It should be noted here that the cyclical change of relative wind velocity consisting of axial wind velocity and tangential wind velocity in fixed case is because of the height-dependent characteristics of wind speed.

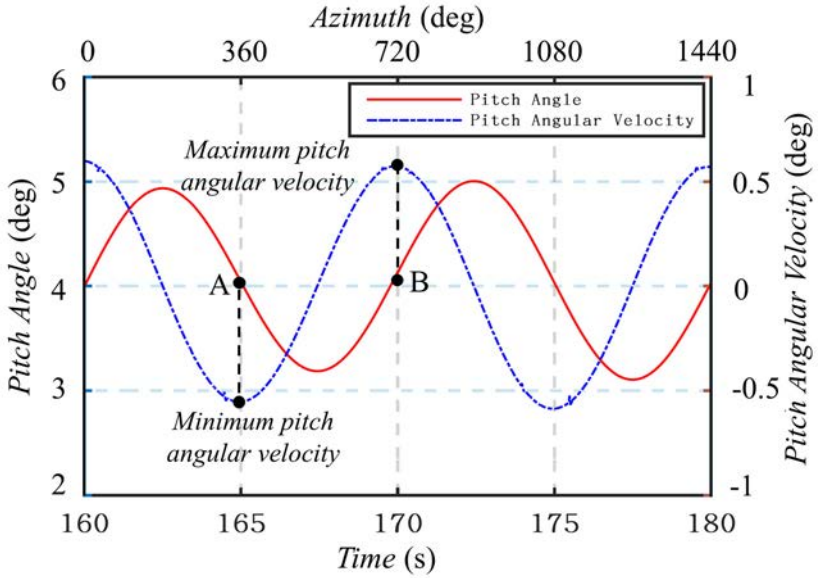
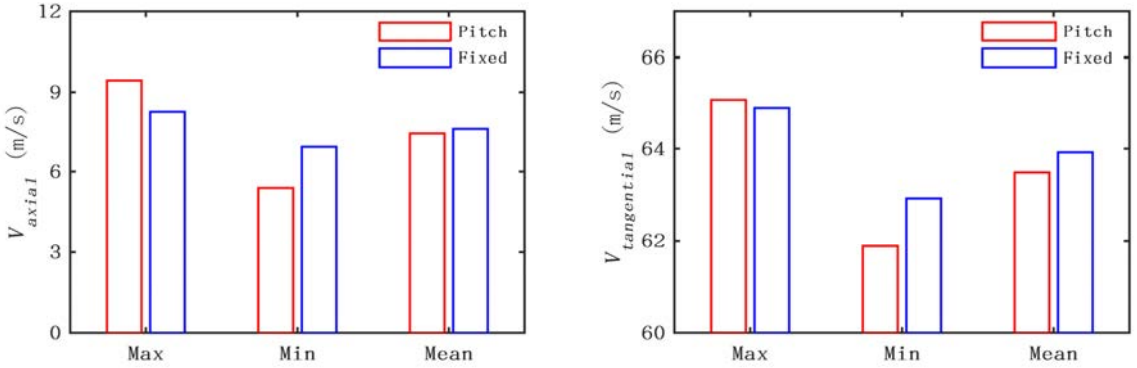



**Fig. 6 Pitch responses of the floating support platform under operating wind and wave conditions.**

Due to the additional relative wind speed induced by the pitch responses, the amplitudes of the axial wind speed and the tangential wind speed experiencing by the turbine blades both become larger. The pitch angular velocity for the floating support platform shown in Fig. 6 may explain this phenomenon. The relative wind speed reaches the maximum value when the floating

platform moves forward passing the equilibrium position (Point A shown in the Fig. 6) corresponding to the minimum value of pitch angular velocity, resulting from that the directions of platform pitch motion and wind speed are opposite. By contrast, the minimum relative wind speed corresponds to the maximum pitch angular velocity (Point B shown in the Fig. 6).




**(a) axial wind speed** **(b) tangential wind speed**

**Fig. 7 Characteristic values of relative wind speed experiencing by blade #1 at a typical blade cross section *r*/*R*=0.8 during the rotation of wind turbine.**

The change of relative wind speed will cause the variation of the local attack angle and furthermore influence the lift and drag forces acting on the rotating blades, which leads to the unsteady characteristics of aerodynamic loads including the rotor power and thrust. It is easy to find out that the variation period of local attack angle is about 5s in the fixed case while it is about 10s is the in the pitch case. And averaged value of the attack angle for the rotating blade affected by the platform pitch motion decreases by 7% compared to that of the fixed wind turbine aerodynamics. Moreover, the same changing tendency can also be found in the variation of lift coefficient respected to the azimuth angle shown in Fig. 9. The rotor power and thrust of the wind turbine for different simulation conditions are compared in the Fig. 10. It can be observed that the averaged aerodynamic responses including rotor power and thrust for rotating turbine with platform pitch motion are obviously smaller than those with fixed platform. The rotor power and thrust in the pitch case averaged in 160s~180s are decreased by nearly 11% and 8%, respectively. It indicates that the pitch responses of floating support platform under operating wind and wave conditions have adverse effects on the aerodynamic power output of the wind turbine. In addition, the oscillating rotor power and thrust amplitudes of the wind turbine with platform pitch motion are about 83% (3.94MW) and 41% (272kN) of the corresponding averaged aerodynamic loads. This dramatic change of aerodynamic load will significantly alter the forces acting on the rotating blades with a result of rapidly increased fatigue loads and instability problem, which may cause severe damage to critical system and structures and reduce the service life of turbine blades. As presented in Fig. 11, the characteristic values of bending moments acting on the blade root during the rotation of turbine blades are plotted to show how the platform pitch responses influence the fatigue loads. It can be seen that the axial bending moment responses are obviously greater than that of the tangential bending moment due to the structure characteristics of turbine blades. Moreover, the bending moments along axial direction and tangential direction for the rotating blades with platform pitch motion both have significantly larger variation amplitudes compared to those with fixed platform. The variation amplitude of tangential bending moment in pitch case is nearly 2.86 times of that in fixed case, and this ratio increases up to 3 times for the axial bending moment. Therefore, complicated control strategies are supposed to apply in the FOWT system to suppress the motion responses of floating platform.

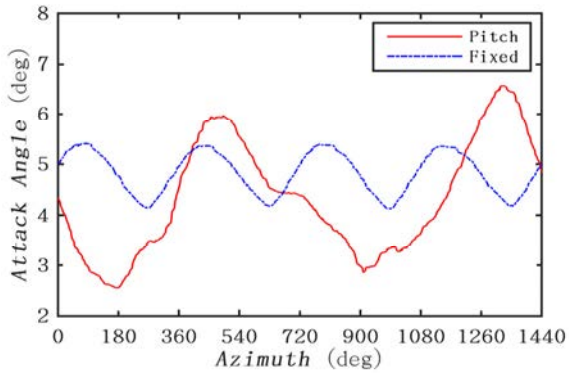



**Fig.8 Comparison of the attack angle at *r*/*R*=0.8 of the rotating blade.**

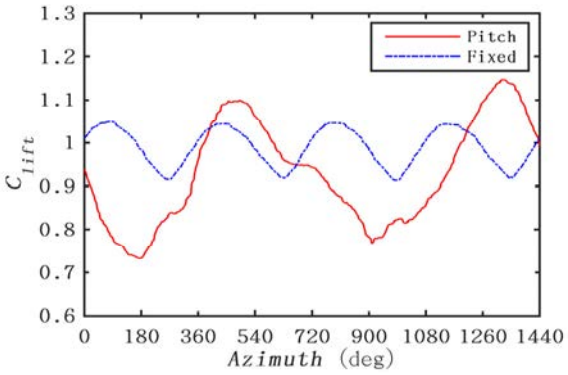
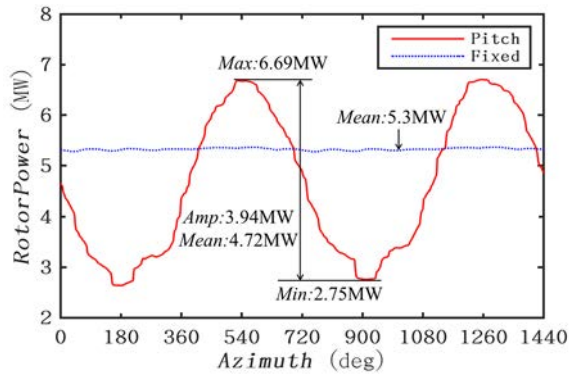
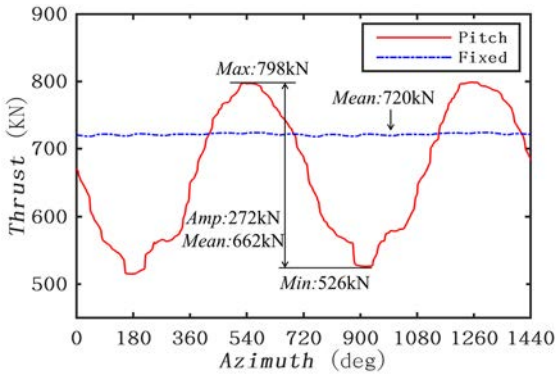



**Fig. 9 Comparison of lift coefficients of the blade section at *r*/*R*=0.8.**

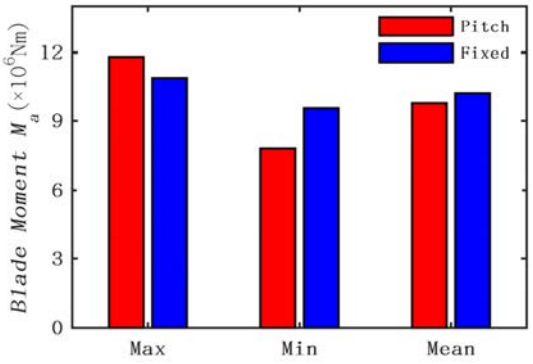



**(a) rotor power**

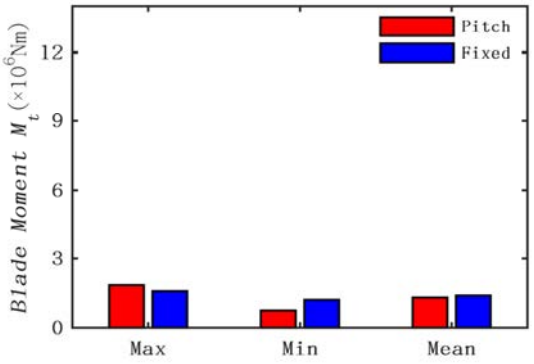



**(b) thrust**

**Fig. 10 Comparison of the aerodynamic loads for wind turbine with fixed and pitch degree-of-freedom platform.**




**(a) axial bending moment**



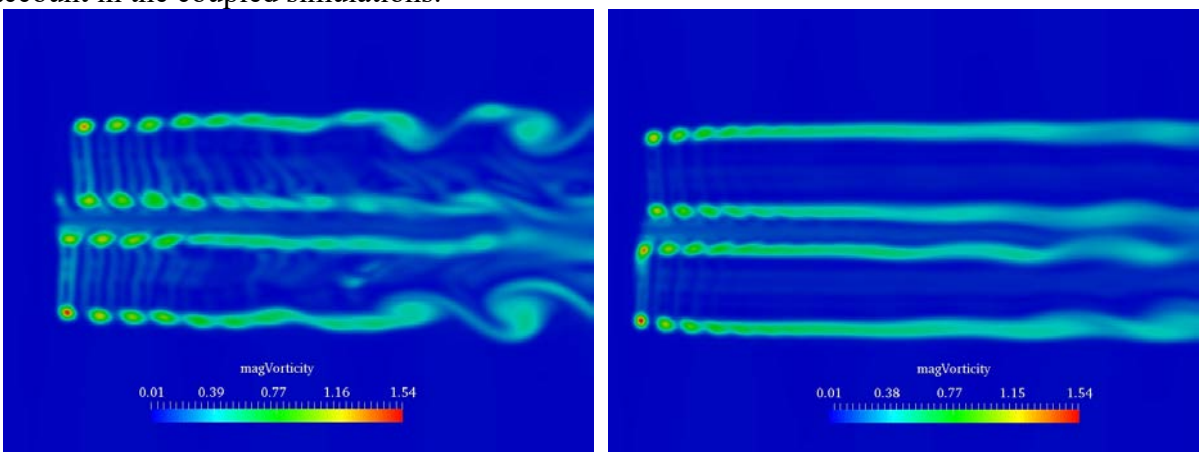

**(b) tangential bending moment**

**Fig. 11 Comparison of time-averaged blade root bending moments defined at blade-aligned coordinate system.**

*Wake field characteristics*

The computational fluid dynamic method is utilized in the present work to investigate the influence of pitch responses of floating support platform under operating wind and wave conditions on the unsteady aerodynamic characteristics of FWOT. Modified body forces model UALM is employed to model the unsteady aerodynamics of wind turbine, which is an effective way to achieve the coupled aero-hydrodynamic simulations for the FOWT with affordable computational resources. Detailed wake information and flow field characteristics for different

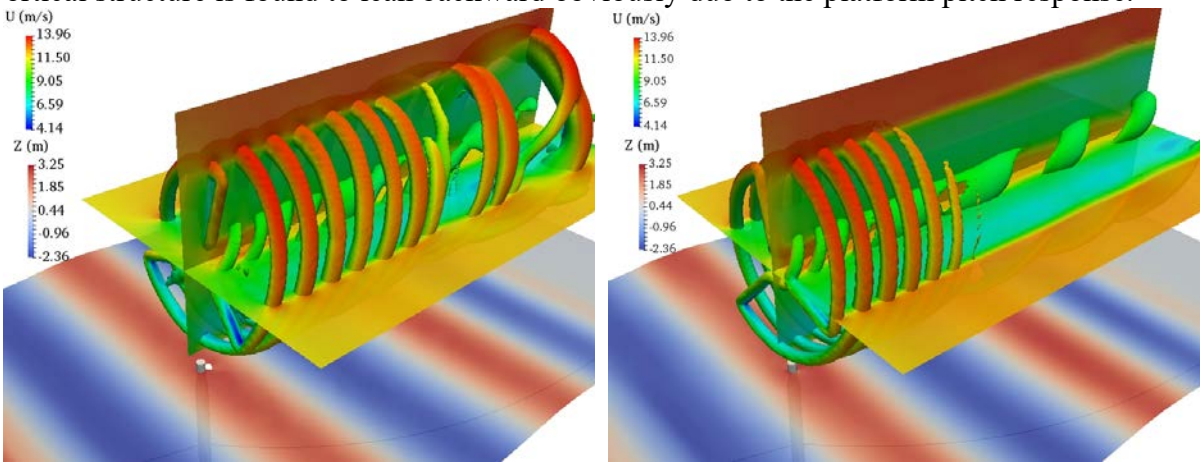
simulation conditions are compared and analyzed to explore the effects of platform pitch motion on the wake field. It should be noted here that the hub, cabin and tower are not taken into account in the coupled simulations.




**(a) pitch case** **(b) fixed case**

**Fig. 12 Visualization of instantaneous (*t*=170s) vorticity counter at horizontal plane with the height of hub center (z=90m).**

As illustrated in Fig. 12, the instantaneous vorticity at the reference horizontal plane with the height of hub center for different simulation conditions are presented. It can be seen that clear vorticities generated from the blade tip and root are captured and the periodic vortex shedding phenomenon is also observed. Duo to additional velocity induced by the cyclic pitch motion of floating support platform, the distance between two adjacent wake vortices in pitch case is larger than that with fixed platform. In fixed case, the tip vortices quickly merge with the adjacent vortex during the development of wake flow. Whereas the vortices generated from the rotating blades with platform pitch motion spear further before they are merged in the downstream of wake field. In addition, it can be obviously found that the wake filed maldistribution is more serious due to the influence of platform pitch motion, which indicates the downstream FOWT will suffer more complicated and unsteady inflow condition. Moreover, the iso-surface plot of the second-order invariant of velocity gradient tensor Q colored by the wind velocity are illustrated in Fig. 13 to visualize the vortices. The vortices generated from the wind turbine with fixed platform are rapidly dissipated in the downstream wake filed. And the vortical structure is found to lean backward obviously duo to the platform pitch response.




**(*a*) Pitch case** **(*b*) Fixed case**

**Fig. 13 Illustration of coupled simulations for the FOWT where the wake vortex is counted by Q=0.01 and colored by wind speed.**

## Conclusions

In this study, computational fluid dynamic method with modified body force model UALM is employed to simulate the coupled aero-hydrodynamic characteristics of the FOWT with realistic platform pitch responses under combined wind and wave condition. The application of UALM for unsteady aerodynamics of wind turbine makes the coupled CFD analysis for FOWT much more effectively compared with conventional method that consider the actual blade surface. Pitch responses of the floating support platform, unsteady aerodynamic characteristics including the attack angle, relative wind speed, aerodynamic loads, blade root bending moments and detailed wake information are obtained and discussed to investigate the strong interactions between the unsteady aerodynamics and platform pitch motion for FOWT system in operating state. It can be found that the average platform pitch response under operating wind-wave loads is nearly 4 degrees and the pitch amplitude is almost half of average value, about 2 degrees. Due to the contribution of wind loads, both the average and oscillating amplitude of pitch responses increase remarkably. And this cyclical pitch motion in return amplified the unsteady aerodynamic characteristics of FWOT by altering the relative wind speed. The change of relative wind speed causes the variation of the local attack angle and furthermore influence the lift and drag forces acting on the rotating blades, which leads to the periodical change of aerodynamic loads including the rotor power and thrust. The oscillating rotor power and thrust amplitudes of the wind turbine with platform pitch motion are about 83% (3.94MW) and 41% (272kN) of the corresponding averaged aerodynamic loads. And the average rotor power and thrust in the pitch case are decreased by nearly 11% and 8% respectively, indicating the platform pitch responses may have adverse effects on the power output. Furthermore, the dramatic change of aerodynamic loads significantly alters the forces acting on the rotating blades with a result of rapidly increased fatigue loads. The oscillating amplitudes of axial and tangential bending moments at the rotating blade root with realistic platform pitch motion are both nearly 3 times of those with fixed platform. In the view of wake characteristics, the tip vortices are clearly captured and the wake filed maldistribution is more serious due to the influence of platform pitch motion, which leads to more complicated and unsteady inflow condition for the downstream FOWT. Thus, complicated control strategies are supposed to apply in the FOWT system to suppress the motion responses of floating platform. Moreover, the influence of other degree-of-freedom platform motions such as surge and yaw on the unsteady aerodynamic performance of the FOWT in operating wind-wave conditions will be discussed in the future.

## Acknowledgement

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

### References

[1] Wang, X., Zeng, X., Li, J., Yang, X., and Wang, H. (2018) A review on recent advancements of substructures for offshore wind turbines, Energy Conversion and Management 158, 103-119.

[2] Lei, H., Zhou, D., Bao, Y., Chen, C., Ma, N., and Han, Z. (2017) Numerical simulations of the unsteady aerodynamics of a floating vertical axis wind turbine in surge motion, Energy, 127, 1-17.

[3] Rodrigues, S., Restrepo, C., Kontos, E., Pinto, R. T., and Bauer, P. (2015) Trends of offshore wind projects. Renewable and Sustainable Energy Reviews 49, 1114-1135.

[4] Sun, X., Huang, D., and Wu, G. (2012) The current state of offshore wind energy technology development, Energy 41(1), 298-312.
[5] Butterfield, S., Musial, W., Jonkman, J., and Sclavounos, P. (2007) Engineering challenges for floating offshore wind turbines (No. NREL/CP-500-38776), National Renewable Energy Lab.(NREL), Golden, CO (United States).
[6] Shen, X., Chen, J., Hu, P. (2018) Study of the unsteady aerodynamics of floating wind turbine, Energy 145, 793-809.
[7] Sebastian, T. The aerodynamics and near wake of an offshore floating horizontal axis wind turbine, PhD Thesis, University of Massachusetts, USA, 2012.
[8] Leble, V., and Barakos, G. (2017) 10-MW wind turbine performance under pitching and yawing motion, Journal of Solar Energy Engineering 139(4), 041003.
[9] Tran, T. T., and Kim, D. H. (2015) The platform pitching motion of floating offshore wind turbine: A preliminary unsteady aerodynamic analysis, Journal of Wind Engineering and Industrial Aerodynamics 142, 65-81.
[10] Tran, T., Kim, D., and Song, J. (2014) Computational fluid dynamic analysis of a floating offshore wind turbine experiencing platform pitching motion, Energies 7(8), 5011-5026.
[11] Tran, T. T., and Kim, D. H. (2015) The aerodynamic interference effects of a floating offshore wind turbine experiencing platform pitching and yawing motions, Journal of Mechanical Science and Technology 29(2), 549-561.
[12] Wen, B., Dong, X., Tian, X., Peng, Z., Zhang, W., and Wei, K. (2018) The power performance of an offshore floating wind turbine in platform pitching motion, Energy 154, 508-521.
[13] Wen, B., Tian, X., Zhang, Q., Dong, X., Peng, Z., Zhang, W., and Wei, K. (2019) Wind shear effect induced by the platform pitch motion of a spar-type floating wind turbine, Renewable Energy 135, 1186-1199.
[14] Hu, H., Khosravi, M., and Sarkar, P. (2016) An experimental investigation on the aeromechanic performance and wake characteristics of a wind turbine model subjected to pitch motions, In 34th Wind Energy Symposium (p. 1997).
[15] Rockel, S., Camp, E., Schmidt, J., Peinke, J., Cal, R., and Hölling, M. (2014), Experimental study on influence of pitch motion on the wake of a floating wind turbine model, Energies 7(4), 1954-1985.
[16] Huang, Y., Cheng, P., and Wan, D.C. (2019). Numerical analysis of a floating offshore wind turbine by coupled aero-hydrodynamic simulation. Journal of Marine Science and Application.
[17] Cheng, P., Huang, Y., and Wan, D.C. (2019) A numerical model for fully coupled aero-hydrodynamic analysis of floating offshore wind turbine, Ocean Engineering 173, 183-196.
[18] Sorensen, J. N., and Shen, W. Z. (2002) Numerical modeling of wind turbine wakes, Journal of Fluids Engineering 124(2), 393.
[19] Jonkman, J., Butterfield, S., Musial, W., and Scott, G. (2009) Definition of a 5-mw reference wind turbine for offshore system development, Office of Scientific & Technical Information Technical Reports, 1–75.
[20] Jonkman, J., and Musial, W. (2010) Offshore code comparison collaboration (oc3) for IEA wind task 23 offshore wind technology and deployment, Office of Scientific & Technical Information Technical Reports 303, 275-3000.

# Aerodynamic Analysis of the Wind Turbine by Two Different Numerical Methods

**Ruicheng Zhang, Decheng Wan***

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

**Abstract**

With the rapid development of wind power industry in recent years, the aerodynamic performance of wind turbines has attracted more attention, on account of its significant effects on the power generating capacity. Two of numerical simulation methods are selected in the present work to study aerodynamic performance of wind turbine. One is the actuator line model (ALM), the other is the overset grid method. In this study, the aerodynamic performance of a NREL 5-MW baseline wind turbine under rated wind speed of 11.4m/s is studied with these two methods. The time history of aerodynamic power and thrust, contours of axial direction wind velocity and vorticity, and wake vortex structures can be obtained and analyzed. Both methods can obtain relatively accurate simulation results. ALM method can reduce the number of grids and calculation time effectively. Overset grid method can obtain more accurate power and thrust forecasting due to simulation of real blades and hub. In addition, detailed flow field characteristics including the pressure distribution on the blade surface can be obtained with overset grid method.

**Keywords:** Aerodynamic loads, Wake field characteristics; Actuator line model; Overset grid method.

## Introduction

The rapid development of global economic has led to increasing demand for energy in each country. The situation of the energy crisis is becoming more and more serious. As a kind of clean and renewable energy with huge reserves, wind energy has great potential for development and utilization. Wind energy has become one of the fastest growing renewable energy sources in the world, and wind energy technology has received extensive attention of many countries. Wind power is mainly used by converting wind energy into electric energy through wind turbines. Accurate prediction of wind turbine aerodynamic performance and wake field characteristics is very important for the early economic evaluation of wind farms.

In the past few years, many researchers have studied on the aerodynamic performance of wind turbines and proposed several methods, including the Blade Element Momentum theory (BEM), potential flow theory and CFD numerical simulation methods. The Blade Element Momentum theory (BEM) is one of the most classical methods for calculating the aerodynamic load of a wind turbine [1]. The Blade Element Momentum theory divides the wind turbine blade into a number of micro-segments that do not interfere with each other. The three-dimensional

aerodynamic characteristics of the wind wheel can be obtained by integrating the aerodynamic characteristics of the blade elements in the radial direction. The numerical simulation results based on BEM theory have basically met the needs of practical engineering applications. BEM has the advantages of simplicity and ease of application, but it cannot study the details of the flow field, nor can it explain the three-dimensional effect of the blade and the stall delay effect [2]. The potential low theory introduces the aerodynamic model of the three-dimensional potential flow into the calculation of the wind turbine [3]. Although more detailed aerodynamic performance of the wind turbine is obtained, the viscosity of the flow field around the blade and the phenomenon of flow separation are not considered.

Computational Fluid Dynamics (CFD) is a technique developed by the rapid development of computers. It numerically solves the Navier-Stokes equations describing the conservation of viscous incompressible fluid momentum, which can accurately describe the complex flow field around the wind turbine, simulate the actual motion of the fluid in the field and obtain more complete flow field information. Choi et al. [4][5] used the CFD to numerically simulate the wind farm of two and three wind turbines, and studied the influence of the distance between the wind turbines on the power output and wake field characteristics of the wind farm. Yuwei et al. [6] used the DES method combined with dynamic overlapping grid technology to simulate the aerodynamic performance of wind turbines. Churchfield et al. [7][8] used SOWFA software to analyze the wake characteristics of the wind turbine using the large eddy simulation method.

However, the CFD method has disadvantages of the difficulty of meshing, long calculation time and high hardware requirements. Reducing the amount of calculation is a key issue in applying the CFD method. Therefore, the actuator line model (ALM), which combines BEM theory and CFD method, has been proposed. ALM do not need to establish the real rotor geometry model. ALM uses the virtual actuator line to replace the real blade structure, which can avoid solving the boundary layer of the blade surface and further reduce the calculation time. The ALM is very research-worthy and has attracted the attention of many researchers. Troldbrg and Sørensen [9] used the actuation line model to numerically simulate a three-blade wind turbine, obtained detailed information on the wake region, and fond a good agreement after comparing the power output curve with the experimental data. Mikkelsen et al. [10] used ALM combined with CFD technology to study the aerodynamic power output and wake characteristics of a wind farm with three wind turbines and obtained satisfactory simulation results.

Another widely used method of simulating the aerodynamic performance of wind turbines is the overset grid method. The overset grid method allows unconstrained relative displacement between multiple independent grids, and can achieve unconstrained six-degree-of-freedom motion of the object, so it is suitable for solving dynamic problems [11][12]. Naoe-FOAM-os-SJTU is a CFD numerical solver for marine and offshore engineering based on open source toolbox OpenFOAM combined with overset grid technology. The solver introduces the overlay overlap mesh technology to solve the topology constraint relationship between the object and the mesh, and can realize the six-degree-of-freedom unconstrained motion between multi-level objects that cannot be processed by the traditional dynamic mesh technology.

In this study, the ALM based on the OpenFOAM is used to study the aerodynamic performance of a NREL 5-MW baseline wind turbine. The overset grid method is also applied to investigate the aerodynamics of full-scale wind turbine model by naoe-FOAM-os-SJTU slover. The simulation results obtained from these two different numerical methods are compared to illustrate the advantages of different numerical methods.

**Numerical Method**

*Actuator line model (ALM)*

The actuating line model was first proposed by Sørensen and Shen [13][14]. Its main idea is to replace the real blade with a virtual, volumetric actuating line to avoid solving the boundary layer of the blade surface, thus reducing the difficulty of the meshing and the computation time. The actuator line model discretizes the blades in the radial direction into micro-segments that do not interfere with each other, called blade elements. The lift force and drag force of each blade element can be calculated as:

$$L = \frac{1}{2} C_l(\alpha) \rho U_{rel}^2 c dr \tag{1}$$

$$D = \frac{1}{2} C_d(\alpha) \rho U_{rel}^2 c dr \tag{2}$$

Where, $C_l(\alpha)$ is the lift coefficient, $C_d(\alpha)$ is the drag coefficient, $\alpha$ is the local angle of attack, $\rho$ is the density, $U_{rel}$ is the air flow rate relative to the blade, $c$ is the chord length, and $dr$ is the blade element width.

The relative velocity can be calculated by the local velocity vector relationship of the rotating blades. According to the velocity triangle in the figure1, the relative velocity can be expressed by the following formula:

$$U_{rel} = \sqrt{U_z^2 + (\Omega r - U_\theta)^2} \tag{3}$$

Where, $U_z$ is the axial velocity, $U_\theta$ is the tangential velocity, and $\Omega$ is the rotational velocity of the blade.

The angle of attack is calculated from the geometric relationship:

$$\alpha = \phi - \gamma \tag{4}$$

Where, $\varphi = tan^{-1}((U_z/(\Omega r - U_\theta))$, $\gamma$ is the blade pitch angle.

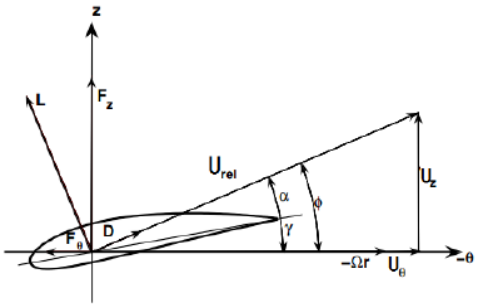



**Figure1. Cross-sectional aero foil element**

After obtaining the relative velocity and the attack angle, the lift force and drag force of each blade element can be calculated by:

$$f = (L, D) = \frac{1}{2}\rho U_{rel}^2 c(C_i\overrightarrow{eL} + C_d\overrightarrow{eD}) \quad (5)$$

The volume force generated by each actuator element is discrete and cannot directly act on the flow field, otherwise it will cause numerical oscillations, so it needs to be smoothed. The expression of the Gaussian smoothing function is as follows:

$$\eta_\epsilon(d) = \frac{1}{\varepsilon^3\pi^{3/2}} exp\left[-\left(\frac{d_i}{\epsilon}\right)^2\right] \quad (6)$$

After smoothing, the volume force at a point (x, y, z) in the flow field can be calculated by:

$$f_{\epsilon,i}(x, y, z, i) = f \otimes \eta_\varepsilon = \sum_{i=1}^{N} f(x_i, y_i, z_i, i) \frac{1}{\varepsilon^3\pi^{3/2}} exp\left[-\left(\frac{d_i}{\epsilon}\right)^2\right] \quad (7)$$

Where, $d_i$is the distance from the point $(x, y, z)$ in the calculation domain to the *i*th actuator point$(x_i, y_i, z_i)$, and $\varepsilon$ is the Gaussian smoothing parameter.

*Overset grid method*

The overset grid method is to mesh each part of the object separately, then embed them in another set of grids. After the overlapping areas between the meshes are subjected to pre-processing such as tunneling, the mesh outside the calculation domain (such as the mesh cells located inside the surface of the object) will be dug out and excluded from the calculation, and the interpolation relationship is established in the remaining overlapping area. The interpolation relationship is calculated by the DCI data obtained by the SUGGAR program, and allows data exchange between overset grid to achieve the overall calculation of the flow field.

The process of solving DCI can be divided into four steps. The first step is to mark the grid outside the calculation domain as hole cells and exclude them from the calculation. The second step is to search for the donor cell and provide interpolation information for the interpolated cell. The third step is to calculate the interpolation coefficient (weight coefficient). The fourth step is to optimize the overlap area.

$$\phi_I = \sum_{i=1}^{n} \omega_i \cdot \phi_i \quad (8)$$

Where $\phi_I$ is the value of a variable $\phi$ of the fringe cell, $\phi_i$ is the value for the *i*th donor cell, $\omega_i$ is the weight coefficient.

*Governing Equation*

The governing equation used in the ALM and AMI is the RANS equations. The RANS algorithm treats the turbulence with irregular random pulsation characteristics as laminar flow, introduces the concept of pulsation that reflects the turbulence characteristics in the NS equation, and averages it over time to get the RANS equation whose expression is:

$$\frac{\partial U}{\partial X_i} = 0 \quad (9)$$

$$\frac{\partial U_i}{\partial t} + \frac{\partial (U_i U_j)}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\left(v\frac{\partial U_i}{\partial x_j} - \overrightarrow{u_i' u_j'}\right) + \frac{1}{\rho} f_\epsilon \quad (10)$$

Where, *U* is the flow field velocity, $\rho$ is the fluid density, $p$ is the flow field pressure, $v$ is the kinematic viscosity, and $f_\epsilon$ is the volumetric force in the actuator line model.

Solving the governing equation requires the use of a turbulence model to close the equation to

achieve the numerical solution of the flow field. The k-ω SST turbulence model which is suitable for simulating the aerodynamic performance of wind turbines is used in this paper.

**Simulation Setup**

The wind turbine model used in this paper is NERL-5MW wind turbine. The main specifications of NERL-5MW turbine are given in the Table 1.

**Table 1 Specifications of NERL-5MW turbine**

| | |
|---|---|
| Rating | 5MW |
| Rotor Orientation, Configuration | Upwind, 3 Blades |
| Control | Variable Speed, Collective Pitch |
| Drivetrain | High Speed, Multiple-Stage Gearbox |
| Rotor, Hub Diameter | 126m, 3m |
| Hub Height | 90m |
| Cut-in, Rated, Cut-out Wind Speed | 3m/s, 11.4m/s, 25m/s |
| Cut-in, Rated Rotor Speed | 6.9rpm, 12.1rpm |
| Rated Tip Speed | 80m/s |
| Overhang, Shaft Tilt, Precone Angle | 5m, 5°, 2.5° |

The computation domain is sketched in Figure 2. The height and width of the whole computation domain are both 400m. The length of the domain is 1000m.The distance from the inlet to the wind turbine is 300m. The height of hub is 200m.

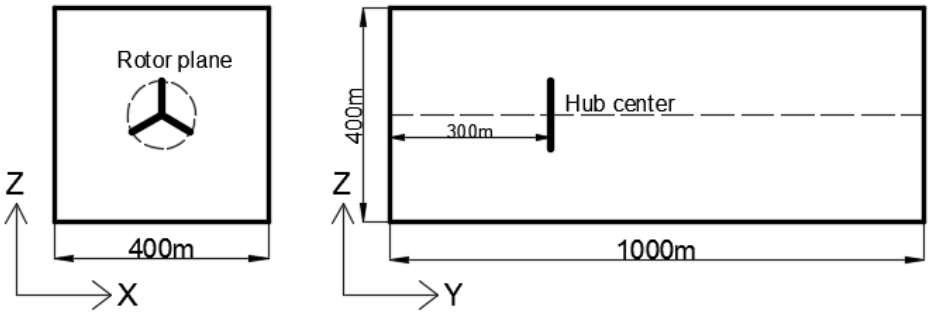
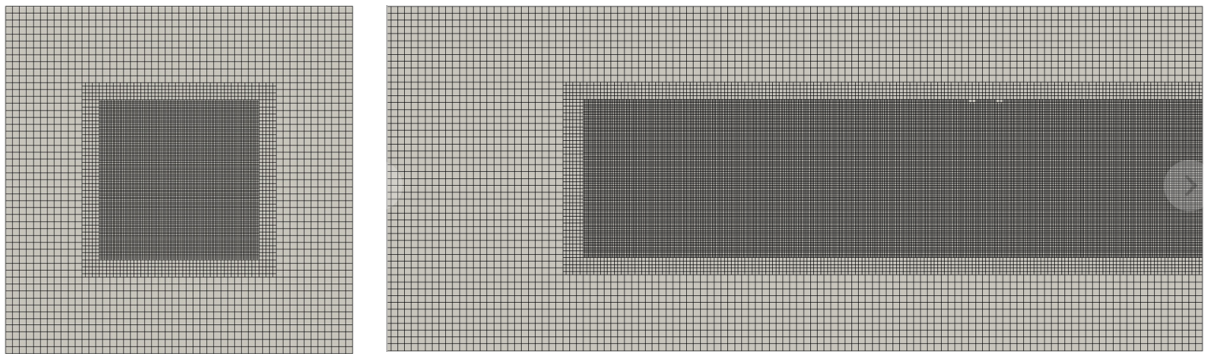



**Figure 2. Computational domain**

In this paper, the simulations conducted with ALM method and Overset Grid method share same background mesh. In simulation with ALM, a refined region covered the wind turbine blades and wake field with three levels refinement is provided as sketched in Figure 3. The total number of mesh is about 350 million. In simulation with overset grid method, the impeller of wind turbine is meshed and put into cylindrical background grid, and then embedded in the whole computational domain, as sketched in Figure 4. The total number of mesh is about 390 million.



**(a)cross section** **(b)lengthwise section**

**Figure 3. Grid Structure with ALM method**

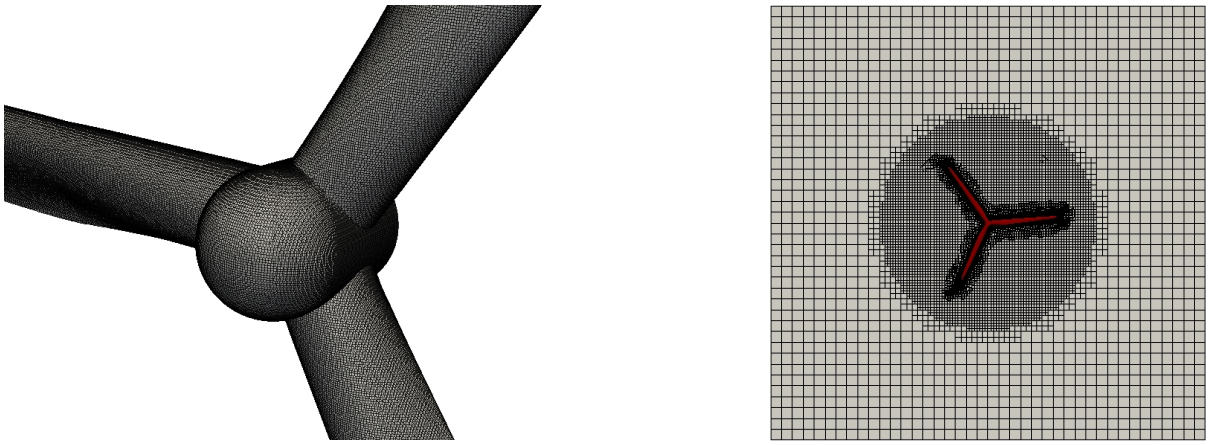


**Figure 4. Grid structure with overset grid method**

The inlet boundary adopts free flow boundary conditions, the wind speed is 11.4m/s constantly. The outlet boundary is applied to Dirichlet pressure condition, the pressure is equal to the atmospheric pressure. The top boundary and the bottom boundary are applied to the sliding conditions. The symmetric boundary conditions are set on the left boundary and the right boundary.

## Results and Discussions

*Aerodynamic power and thrust*

The aerodynamic power and thrust of the wind turbine can be obtained by using both two simulation methods. Figure 5 shows the time history of aerodynamic power of turbine rotor in two simulations. The aerodynamic power of turbine is over predicted with ALM method than that with overset grid method. This is because the effect of the flow analysis phenomenon on the blade surface on aerodynamic power is not considered with ALM method. Figure 6 shows the time history of aerodynamic thrust of turbine rotor in two simulations. The aerodynamic power of thrust is underestimate with ALM method than that with overset grid method. This results from the presence of turbine hub accounted in overset grid method. In addition, data fluctuations over time can be observed on the curves of overset grid in both Figure 5 and Figure 6, and the period of the fluctuations equals to 1/3 of the rotating period of turbine rotor. This is because the three blades alternately cut in the area with higher wind speed under the effect of wind shear.

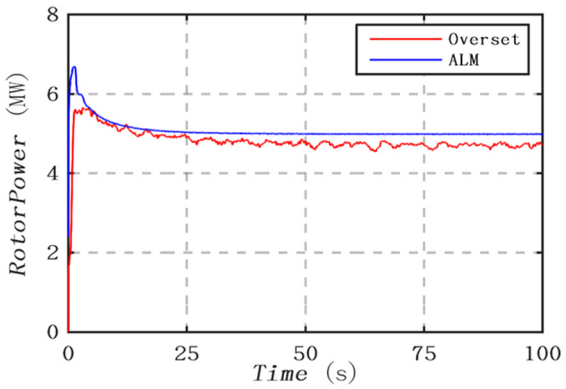



**Figure 5. Aerodynamic power**

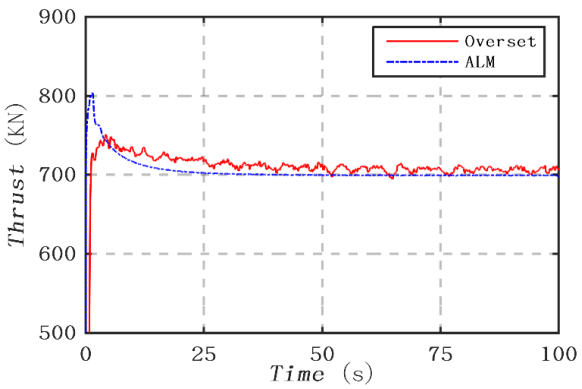
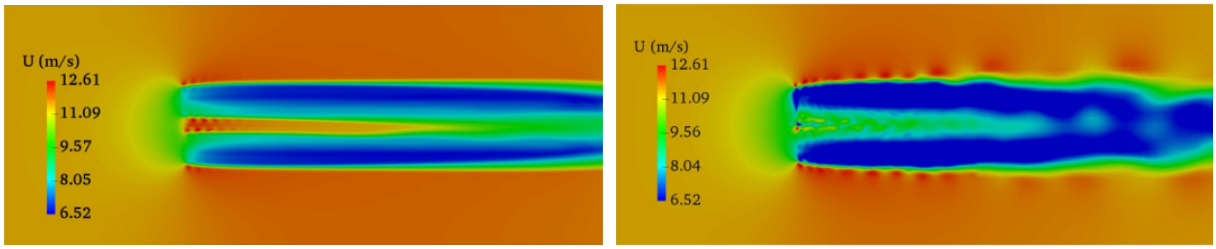



**Figure 6. Aerodynamic thrust**

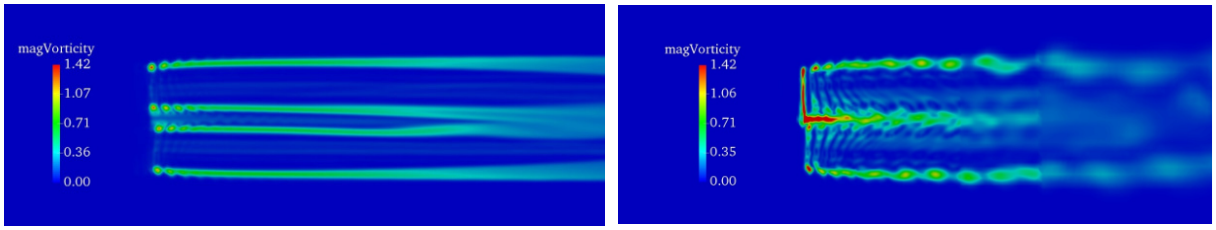
*Wake Field*

Through the post-processing software Paraview for visual processing, both simulation methods can clearly show the wake field information of wind turbine. Figure 7 shows axial direction wind velocity counters at the height of the center of rotor in horizontal plane. Figure 8 shows Vorticity at height z=0 with two simulation methods. Since the overset grid method meshes the real blades and hub of turbine, the wake information obtained is more abundant. It can seen wake velocity decrease greatly after wind passes through the turbine. In the Figure 7(a), along the direction of the inflow, at the 600m behind the hub(about 5 times the diameter of the turbine rotor), the wake velocity has recovered and the color there shows green. While in the Figure 7(b), there is still significant velocity speed loss. This is because overset grid method takes the effect of the hub into account.




**(a) ALM** **(b)Overset grid**

**Figure 7. Contours of axial direction wind velocity in horizontal plane at z=0**

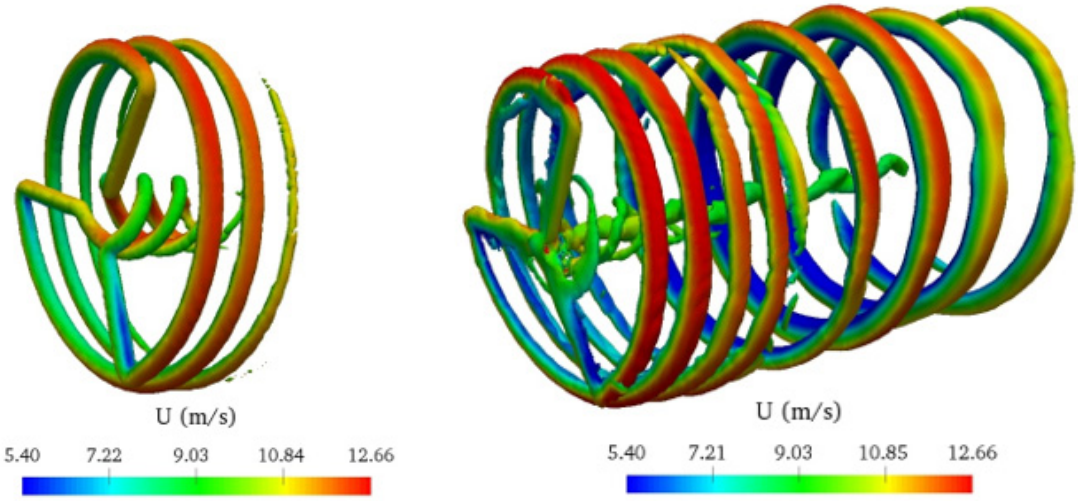



**(a) ALM** **(b)Overset grid**

**Figure 8. Contours of vorticity in horizontal plane at height z=0**

Figure 9 shows the wake vortex structure of wind turbine at moment of 100s with two simulation methods. The wake vortex structure is visualized by the contour of the second

invariant of the velocity gradient tensor $Q$. Distinct blade tip vortices and root vortices are observed with both simulations. The wake vortex structure dissipates faster with ALM method than that with overset grid method, and three cycles of tip vortex can be observed in the Figure 9(a), while eight cycles of tip vortex can be observed in the Figure 9(b).

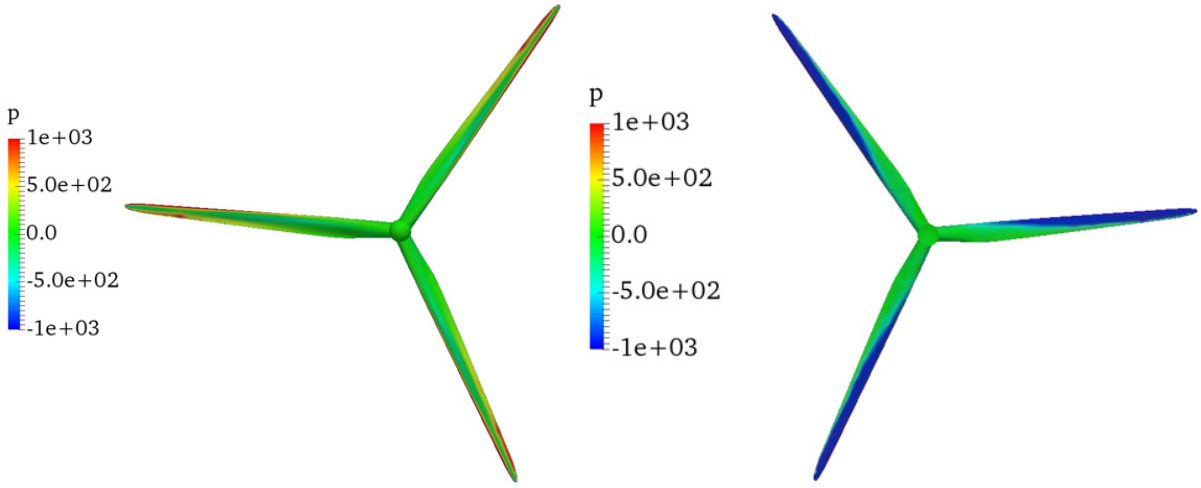



**(a) ALM** **(b) Overset grid**

**Figure 9. Wake vortex structure with two simulation methods**

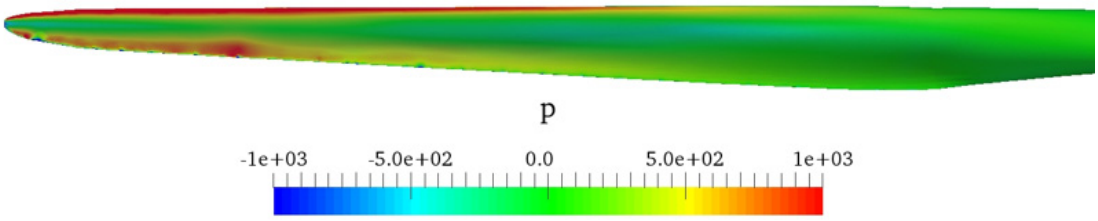
*Pressure Distribution*

Overset grid method can provide more detailed flow information near the blades surface due to the simulation of real blades and hub of wind turbine. Figure 10 shows variation of pressure distribution on the blades surface at moment of 100s. In the Figure 11, the pressure distribution on tip sections are enlarged for better observation. The pressure distribution on the blades is mainly concentrated on the tips, and the pressure on the windward side is larger than the leeward side.

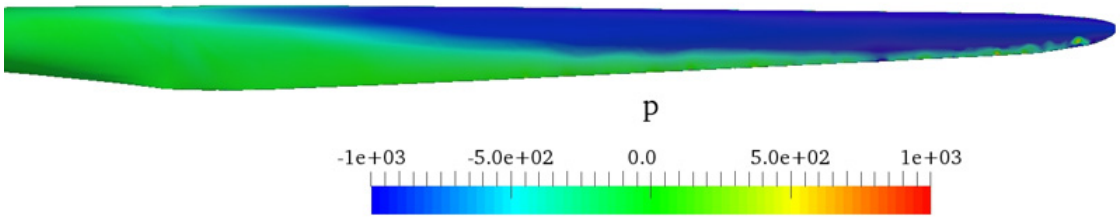



**(a) Front** **(b) Back**

**Figure 10. Variation of pressure distribution on the blades surface**




**(a) Front**




**(b) Back**

**Figure 11. Variation of pressure distribution on the tip section**

## Conclusions

In this study, aerodynamic and wake field simulations of the NREL 5MW wind turbine are conducted with ALM method and overset grid method. With these two methods, the time history of aerodynamic power and thrust, contours of axial direction wind velocity and vorticity, and wake vortex structures are obtained and compared. Both methods can obtain relatively accurate simulation results to analyze of aerodynamic characteristics and wake characteristics of wind turbine. The ALM method replaces the real blades with actuator elements, which can reduce the number of grids and calculation time effectively. Compared with overset grid method, the aerodynamic power of turbine is over predicted, and the aerodynamic thrust of turbine is underestimate with ALM method. In addition, overset grid method can reflect fluctuations in aerodynamic power and thrust output with time. This indicates that more accurate power and thrust forecasting still needs to consider the effect of the blades and hub. Overset grid method simulates real blades of wind turbine, therefore it can provide more detailed flow information near the blades surface, such as variation of pressure distribution on the blades surface, which is helpful to further understand the complex flow phenomena around the wind turbine and optimal design for the wind turbine.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

## References

[1] Lanzafame, R., and M. Messina. "Fluid dynamics wind turbine design: Critical analysis, optimization and

application of BEM theory." Renewable energy 32.14 (2007): 2291-2305.

[2]Shen, Wen Zhong, et al. "Tip loss corrections for wind turbine computations." *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 8.4 (2005): 457-475.

[3]WHALE J, FISICHELLA C, SELIG M. Correcting inflow measurements from J3 AWTS using a lifting-surface code[J]. Urbana, 1999, 51:61

[4] Choi, Nak, et al. "CFD Study on Aerodynamic Power Output Changes with Inter-Turbine Spacing Variation for a 6 MW Offshore Wind Farm." *Energies* 7.11 (2014): 7483-7498.

[5] Choi, Nak Joon, et al. "Numerical study on the horizontal axis turbines arrangement in a wind farm: Effect of separation distance on the turbine aerodynamic power output." *Journal of Wind Engineering and Industrial Aerodynamics* 117 (2013): 11-17.

[6] Li, Yuwei, et al. "Dynamic overset CFD simulations of wind turbine aerodynamics." Renewable Energy 37.1 (2012): 285-298.

[7] Churchfield, Matthew, et al. "A large-eddy simulation of wind-plant aerodynamics." *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*. 2012.

[8] Churchfield, Matthew J., et al. "A numerical study of the effects of atmospheric and wake turbulence on wind turbine dynamics." *Journal of turbulence* 13 (2012): N14.

[9] Sørensen J N, Shen W Z. Computation of wind turbine wakes using combined Navier-Stokes/actuator-line Methodology[C]//1999 European Wind Energy Conference and Exhibition. 1999: 156-159.

[10] Mikkelsen, Robert, et al. "Analysis of power enhancement for a row of wind turbines using the actuator line technique." *Journal of Physics: Conference Series*. Vol. 75. No. 1. IOP Publishing, 2007.

[11] SUGGAR: a general capability for moving body overset grid assembly." *17th AIAA computational fluid dynamics conference*. 2005.

[12] Carrica, P. M., et al. "Large-scale DES computations of the forward speed diffraction and pitch and heave problems for a surface combatant." *Computers & Fluids* 39.7 (2010): 1095-1111.

[13] Sorensen J N, Shen W Z. Numerical modeling of wind turbine wakes[J]. Journal of fluids engineering, 2002, 124(2): 393-399.

[14] Troldborg, Niels, Jens N. Sørensen, and Robert Mikkelsen. "Actuator line simulation of wake of wind turbine operating in turbulent inflow." *Journal of physics: conference series*. Vol. 75. No. 1. IOP Publishing, 2007.

# Study on the change of entangling structure of molecular chains during the tensional and swelling process of elastomeric gel

**†*Isamu Riku, Masashi Ueda, Tomoki Sawada and Koji Mimura**

Department of Mechanical Engineering, Osaka Prefecture University, Japan

†*riku@me.osakafu-u.ac.jp

## Abstract

A nonaffine model is proposed for elastomeric gel to account for the change of the entangling structure of molecular chains during the tensional and swelling process, in which the change in the number of polymer chains per unit volume, $N$, is depending on the first invariant of right Cauchy-Green tensor, $I_1$, and on the volume of the gel, $J$, separately. It is found that the free swelling process may lead to a larger change of the entangling structure of molecular chains rather than the tensional process does. Moreover, the various combination of the effect of each process on the nonaffine movement of molecular chains may lead to different interesting mechanical responses of the gel, such as yield.

**Keywords:** Entangling structure, Elastomeric gel, Computational model, Tensional and swelling process

## Introduction

The elastomeric gel, consisting of cross-linked polymer molecules and discrete solvent molecules, have a high permeability to small molecules and undergo reversible volume change by exuding or absorbing solvent in response to a wide range of stimuli, such as light, temperature, pH, ionic strength and chemical reactions. As a result, the gel has been developed for diverse applications and used as smart materials in sensors and actuators [1-3].

To characterize how mechanical constraint affects the amount of swelling, and how chemical processes generate forces, several nonlinear field theories have been developed [4-7]. Flory and Rehner proposed a statistical mechanical model for the network of polymer molecules and indicated that the swelling capacity of the gel is diminished by the application of an external stress [8]. Recently, Cai et al. [9] employed an alternative approach, in which a set of equations of state is developed based on two assumptions: the amount of the solvent in the gel varies when the gel changes volume, but remains constant when the gel changes shape; the Helmholtz free energy of the gel is separable into the contributions of stretching the network and mixing the polymer and the solvent. However, in these studies, the computational models are just phenomenological ones and the change of entangling structure of molecular chains during the swelling process has not been accounted for explicitly.

Therefore, in this study, we focus on the discussion of the mechanical behavior of elastomeric gel based on a nonaffine molecular chain network model (nonaffine model) [10], which was originally developed for the orientation hardening of amorphous polymers and may account for the change in the entanglement situation for the physical linkages during the deformation processes. After formulating the dependence of the number of polymer chains per unit volume on the first invariant of right Cauchy-Green tensor and the swelling ratio of the elastomeric gel, a computational model is proposed and employed to investigate the mechanical response of the gel during constrained swelling process.

## Conditions of Equilibrium

The derivation of the conditions of equilibrium for the dry polymer and the solvent has been done by Cai et al. [9]. It is convenient to consider that, in the reference state, a block of network of dry polymers is a unit cube, and contains no solvent and subject to no applied forces. In the current state, the network is submerged in a solvent-containing environment, and the six faces of the block are subject to applied forces. When the network, the solvent, and the applied forces equilibrate, the network absorbs $C$ number of solvent molecules, and deforms homogeneously into the shape of a parallelepiped. When the deformation gradient of the network is expressed by $\mathbf{F}$, the ratio of volume of the swollen gel and that of the dry network is determined as $J = \det \mathbf{F}$. Subject to applied forces, the gel changes shape readily, but the volumes of individual polymer chains or solvent molecules remain nearly unchanged. As an idealization, it is assumed that the volume of the gel is a function of the concentration of the solvent:

$$J = f(C) \tag{1}$$

That is, the amount of the solvent in the gel varies when the gel changes volume, but remains constant when the gel changes shape.

Eq. (1) determines the concentration of solvent, $C$, once the deformation gradient is known. Consequently, the nine components of the deformation gradient $\mathbf{F}$ specify the state of the gel. Let $W$ be the Helmholtz free energy of the gel in the current state. The Helmholtz free energy of the gel can be taken to be a function of the nine components of the deformation gradient, $\mathbf{F}$, and is assumed to be separable into contributions from stretching the network and mixing the polymer and solvent [8]:

$$W = W_{stretch}(\mathbf{F}) + W_{mix}(J) \tag{2}$$

The free energy due to the stretching of the network, $W_{stretch}(\mathbf{F})$, is a function of the deformation gradient, and depends on the density of crosslinks. The free energy due to the mixing of the polymer and the solvent, $W_{mix}(J)$, is a function of the concentration of the solvent in the gel, but is independent of the density of crosslinks. Eq. (1) and Eq. (2) form the basis for the model of ideal elastomeric gels. In equilibrium, the change of the Helmholtz free energy of the composite vanishes and one can obtain that

$$\sigma_{ij} = F_{jk} \frac{\partial W_{stretch}(\mathbf{F})}{J \partial F_{ik}} - \Pi(J)\delta_{ij}, \tag{3}$$

where $\sigma_{ij}$ is the true stress, which is defined by imagining a small cube cut from the parallelepiped, $\delta_{ij}$ is the Kronecker delta and $\Pi(J)$ is the osmotic pressure due to mixing the network and the solvent.

## Free Energy Functions

In the original Flory-Rehner model [8], specific functions are adopted for $W_{stretch}(\mathbf{F})$ and $W_{mix}(J)$. The free energy associated with stretching the network is given by the Gaussian-chain model [8]:

$$W_{stretch}(\mathbf{F}) = \frac{1}{2} N k_B T (F_{ik} F_{ik} - 3 - 2 \log J), \tag{4}$$

where $N$ is the number of polymer chains per unit volume, $k_B$ is Boltzmann constant and $T$ is the temperature. It is obviously that the variation of the effect of the swelling ratio, $J$, on the elastic modulus of the gel has been considered. However, this model is just a

phenomenological one and the change of entangling structure of molecular chains during the swelling process of gel has not been accounted for explicitly.

Therefore, in this study, based on the molecular chain network theory [11], the microstructure of the gel is assumed to consist of long molecular chains which are randomly distributed in space. A single chain, which consists of several segments containing monomers, is defined by two linkages which are assumed to be chemically or physically entangled points of molecular chains. The physical links are corresponding to the pseudo entanglement points and may change depending on deformation. On the other hand, chemical links are permanent and preserve the entanglement situation. The physical links may suffer breakdown during the thermodynamic process [12]. The decrease in the number of entangled points due to deformation causes an decrease in the number of polymer chains per unit volume, $N$, an enhanced extensibility, and a reduction in the stiffness of the material, i.e., softening. To account for the effect of the nonaffine movement of the polymer chain on the deformation behavior of the gel, we employ the simplest version of the nonaffine model [10] to accommodate the change in the number of polymer chains per unit volume, $N$, depending on the first invariant of right Cauchy-Green tensor, $I_1$, and on the volume of the gel, $J$, separately as

$$\frac{N}{N_0}=1-\alpha\left[1-\left(\frac{I_1-I_1^{\max}}{3-I_1^{\max}}\right)^{m_2}\right], \tag{5}$$

$$\frac{N}{N_0}=1-\beta\left[1-\left(\frac{J-J^{\max}}{1-J^{\max}}\right)^{m_2}\right], \tag{6}$$

where $N_0$ is the number of polymer chains per unit volume of the gel in the reference state, $\alpha$, $\beta$, $m_1$, $m_2$ are the parameters for identification, $I_1^{\max}$ and $J^{\max}$ are the limit value of $I_1$ and $J$, separately.

As chemomechanical interactions are material-specific and can be very complex, in this study, we just consider some special cases, in which the discussion of the effect of nonaffine movement of the polymer chains can be done without any assumption of the function $W_{mix}(J)$.

**Stress-Stretch Relations**

Inserting Eq. (4) into Eq. (3), we obtain that

$$\sigma_{ij}=\frac{Nk_BT}{J}\left(F_{jK}F_{iK}-\delta_{ij}\right)-\Pi(J)\delta_{ij}. \tag{7}$$

In relating the model to experiments, we describe the deformation of the gel in the coordinates of principal stretches. Let $\lambda_1$, $\lambda_2$, $\lambda_3$ be the principal stretches of the gel, so that $\mathbf{F}=\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)$, $J=\lambda_1\lambda_2\lambda_3$ and $I_1=\lambda_1^2+\lambda_2^2+\lambda_3^2$. Rewrite Eq. (7) in terms of the principal stretches and principal stresses:

$$\sigma_i=\frac{Nk_BT}{J}\left(\lambda_i^2-1\right)-\Pi(J). \tag{8}$$

*Reference State*

Consider the elastomeric gel is in the reference state, i.e. a block of network of dry polymers, where the swelling ratio $J$ is always equal to unit. When the dry block is subject to a uniaxial tensile force, the relation between the stretch $\lambda$ and the stress $\sigma_T$ on the tensile direction can be rewritten based on Eq. (8) as:

$$\sigma_T = Nk_BT\left(\lambda - \lambda^{-2}\right), \tag{9}$$

which can also be used to define the shear modulus of the gel in the reference state, $G_0 = N_0 k_B T$.

*Free Swelling State*

Submerged in the solvent-containing environment but subject to no applied forces, the elastomeric gel attains a state of equilibrium, the free swelling state, characterized by an isotropic swelling ratio, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda = J^{-1/3}$. When a gel in the free swelling state is subject to a uniaxial tensile force, in a short time the solvent in the gel has no time to redistribute, so that the concentration of the solvent in the gel remains fixed, and the gel behaves like an incompressible material. Therefore, the relation between the additional stretch $\tilde{\lambda}$, which is defined referred to the free swelling state, and the stress $\sigma_T$ on the tensile direction can be rewritten based on Eq. (8) as:

$$\sigma_T = J^{-1/3}Nk_BT\left(\tilde{\lambda} - \tilde{\lambda}^{-2}\right), \tag{10}$$

which can also be used to define the shear modulus of the gel in the free swelling state, $G = J^{-1/3}Nk_BT$.

*Constrained Swelling State*

Submerged in the solvent-containing environment but subject to constrained displacement at one direction, $\lambda_3 = 1$, the elastomeric gel attains a state of equilibrium, the constrained swelling state, characterized by an bidirectional swelling ratio, $\lambda_1 = \lambda_2 = \lambda = J^{-1/2}$. The relation between the swelling ratio $\lambda$ and the stress $\sigma_T$ on the constrain direction can be rewritten based on Eq. (8) as:

$$\sigma_T = Nk_BT\left(\lambda^{-2} - 1\right). \tag{11}$$

**Results**

In Fig. 1(b), the theoretical predictions are compared with experimental data [13] for the elastomeric gel under uniaxial tensile. Even though the affine model can reproduce the experimental result at the initial deformation stage very well, the overestimation of the stress increases greatly at the large deformation stage. Therefore, we suppose that the number of polymer chains per unit volume of the gel, $N$, in Eq. (9) changes together with the uniaxial deformation according to Eq. (5), in which the value of parameters has been identified as: $\alpha = 0.21$, $I_1^{\max} = 27.2$. It is obviously that the proposed nonaffine model can reproduce the experimental data successfully. On the other hand, from Fig. 1(a), it can be understood that the dramatic change of the entangling structure of molecular chains during the uniaxial tensile process of gel may occur mostly at the early deformation stage, not the later deformation stage.

In Fig. 2(b), the theoretical predictions are compared with experimental data [14] for the elastomeric gel under different swelling ratios. Here, we suppose that the number of polymer chains per unit volume of the gel, $N$, in Eq. (10) changes together with the swelling ratio according to Eq. (6), in which the value of parameters has been identified as: $\beta = 0.8$, $J^{\max} = 27.3$. Quite similar to that shown in Fig. 1(b), it is obviously that the proposed nonaffine model can reproduce the experimental data successfully. However, compared with the minimum value of $N/N_0$ shown in Fig. 1(a), that value shown in Fig. 2(a) decreases remarkably, which means that the free swelling process may lead to a larger change of the entangling structure of molecular chains rather than the tensional process does.

In Fig. 3, the theoretical prediction of the mechanical resistances of the elastomeric gel during the constrained swelling process. As the tensional process and the swelling process occur simultaneously in this case, we estimate the effect of each process on the change of the entangling structure of molecular chains with the weighted parameters, $A$ and $B$, separately. When the value of $A$ increases, the mechanical resistance of the gel increases and gets close to that predicted by affine model. Interestingly, when the value of $B$ is equal to unit, the stress-stretch relation of the gel shows a yield region, which is left to our future study.

## Conclusions

In this study, a nonaffine model is proposed for elastomeric gel to account for the change of the entangling structure of molecular chains during the tensional and swelling process. It is found that the free swelling process may lead to a larger change of the entangling structure of molecular chains rather than the tensional process does. Moreover, the various combination of the effect of each process on the nonaffine movement of molecular chains may lead to different interesting mechanical responses of the gel, such as yield.

### References


[1] Liu, Z. and Clavert, P. (2000) Multilayer hydrogels as muscle-like actuators, *Advanced Materials* **12**, 288–291.

[2] Beebe, D. J., Moore, J. S., Bauer, J. M., Yu, Q., Liu, R. H., Devadoss, C. and Jo, B. H. (2000) Functional hydrogel structures for autonomous flow control inside microfluidic channels, *Nature* **404**, 588–590.

[3] Cai, S. Q., Lou, Y. C., Ganguly, P., Robisson, A. and Suo, Z. G. (2010) Force generated by a swelling elastomer subject to constrain, *Journal of Applied Physics* **107**, 103535.

[4] Hong, W., Zhao, X. H., Zhou, J. X. and Suo, Z. G. (2008) A theory of coupled diffusion and large deformation in polymeric gels, *Journal of the Mechanics and Physics of Solids* **56**, 1779–1793.

[5] An, Y. H., Solis, F. J. and Jiang, H.Q. (2010) A thermodynamic model of physical gels, *Journal of the Mechanics and Physics of Solids* **58**, 2083–2099.

[6] Hui, C. Y. and Muralidharan, V. (2005) Gel mechanics: A comparison of the theories of Biot and Tanaka, Hocker, and Benedek, *Journal of Applied Physics* **123**, 154905.

[7] Chester, S. A. and Anand, L. (2010) A coupled theory of fluid permeation and large deformations for elastomeric materials, *Journal of the Mechanics and Physics of Solids* **58**, 1879–1906.

[8] Flory, P. J. and Rehner, J. (1943) Statistical mechanics of cross-linked polymer networks II. swelling, *Journal of Applied Physics* **11**, 521.

[9] Cai, S. Q. and Suo, Z. G. (2012) Equations of state for ideal elastomeric gels, *A Letters Journal Exploring The Frontiers of Physics* **97**, 34009.

[10] Riku, I. and Mimura, K. (2010) Computational characterization on mechanical behavior of polymer electrolyte membrane based on nonaffine molecular chain network model, *International Journal of Mechanical Sciences* **52**, 287–294.

[11] Tomita, Y., Adachi, T. and Tanaka, S. (1998) Modeling and application of constitutive equation for glassy polymer based on nonaffine network theory, *European Journal of Mechanics - A/Solids* **16**, 745–755.

[12] Sakai, T. (2017) *Physics of Polymer Gels*, University of Tokyo Press.

[13] Katashima, T., Chung, U. and Sakai, T. (2015) Effect of swelling and deswelling on mechanical properties of polymer gels, *Macromolecular Symposia* **358**, 128–139.

[14] Matsunaga, T., Sakai, T., Akagi, Y., Chung, U. and Shibayama, M. (2009) Structure characterization of Tetra-PEG gel by small-angle neutron scattering, *Macromolecules* **42**, 1344–1351.

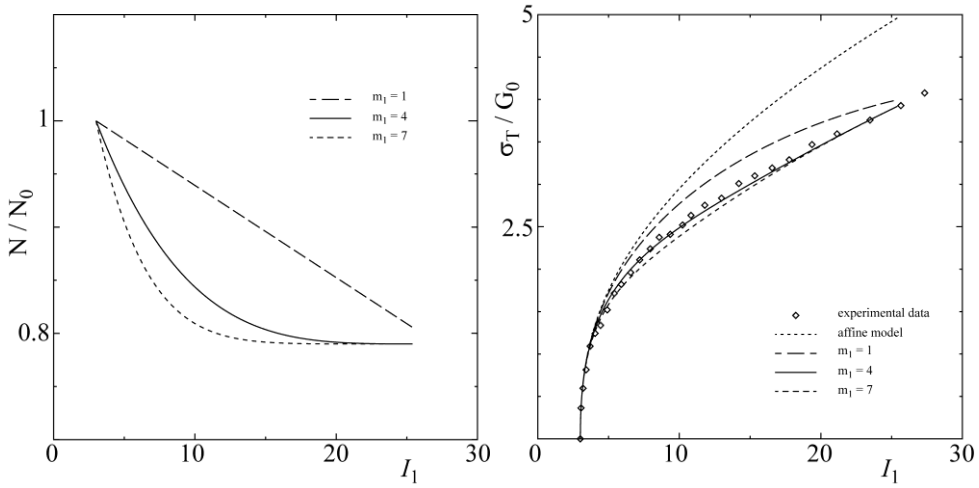



**Figure 1. Characteristics of the gel in the reference state**
**(a) variations of the entangling structure; (b) uniaxial tensile responses**

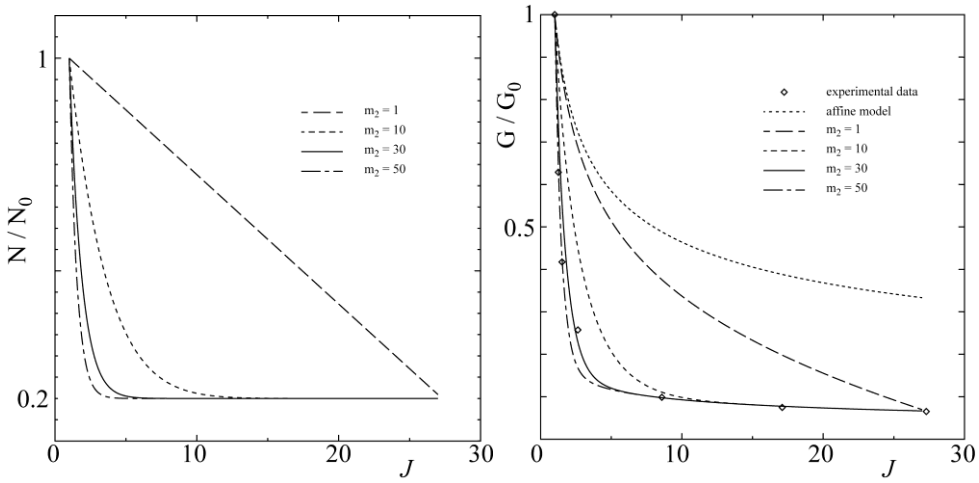



**Figure 2. Characteristics of the gel in the free swelling state**
**(a) variations of the entangling structure; (b) variations of shear modulus**

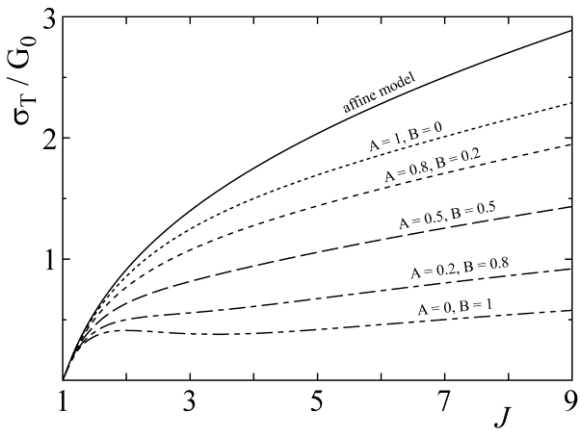



**Figure 3. Mechanical resistances of the gel during the constrained swelling process**

# Nonlinear analysis of two-way beam string structures

***Minhee Seo[1], Seunghye Lee[1], Thi Thanh Dieu Do[1], Thi Mai Linh Phi[1], and †Jaehong Lee[1]**

[1]Deep Learning Architecture Research Center, Sejong University, Korea.

*Presenting author: seomini28@naver.com
†Corresponding author: jhlee@sejong.ac.kr

## Abstract

In this study, structures subjected to bi-directional loading on roof systems can be called as two-way beam string structures. The two-way beam string structure is composed of two types of cables which have different pressure and directions of cable. The cables are used to force the beam through struts supporting both positive and negative loads. This can be solved by introducing proper pre-tension forces to the structure because it prevents slacking of the cables. In addition, it examined how the behavior of the structure varies with respect to changing the angle of the cable. As a result, the evaluation is performed with this system that controls structural behavior by applying the tensile stress of the cables. The parametric studies are performed through nonlinear analysis of displacement control by changing the length and angle of the members. Moreover, the structure is implemented in the ABAQUS finite element package to verify the accuracy and validity of the results.

**Keywords:** Beam string system, Finite element model, High strength cable, Abaqus

## Introduction

In large span spatial structures, one of the major issues is how to handle the horizontal thrust made by vertical loading economically [1]. Tension structures can be solved the issue due to their larger load carrying capacities than typical rigid beam or truss structures. There are two types of the tension structures along the usage of the tension elements, namely, membrane structures and string structures. The string structures can be further classified into two groups: thoroughbred and hybrid tension structures [2]. A beam string structure, which was first suggested in 1984 [3], is a typical type of the hybrid tension structures. In the last three decades, a series of research for the beam string structure has been reviewed mostly in Japan and China. It performed an analytical method in beam string structures [4]. The non-linear finite element method appeared to obtain the buckling load. It introduced an experimental study in Shanghai Yuanshen Areana [5]. The optimum design, namely, the rise-span ratio, beam and string sections, and the pre-stressing force were optimized by the ANSYS program and experimental work. More recently, it developed a new type foldable long-span retractable roof based on the beam string structure [6]. Most of those studies are mainly focused on one-cable beam string structures or string-beam coupled systems without struts. The beam string structure is composed of upper beams and lower strings which are used to stiffen the beam through struts. Because of the advantages in which the pre-tensioned strings elicit pre-camber of the upper beam, the maximum moment and deflection can be greatly decreased. In conclusion, the beam string structure has been widely used in large span roofs of arenas [2,5,7], airport terminals [8], public halls [9], etc. The major purpose of the beam string structure is to carry gravity-load. In contrast, because of their characteristics of light mass and

high flexibility, the structures have completed progressively more sensitive to wind loading than most large span spatial structure.

**Parametric study**

A two-way beam string structure consists of the beam, three struts, and cables. The beams were of the same size with H−150X150X7X10 and a length of 4000 mm. The steel grade was SM490B. The beams had a yield strength of 325 MPa with elastic modulus of 200 GPa. Next, the cables were used as the sagging cable and the arch-shaped cable. The sagging cable has a tensile strength of 1,860 MPa and diameter is the 17.8 mm, while the diameter of arch-shaped cable is 12.59 mm. At last, strut was made by a pair of steel channels whose material was SS400. The cross-sectional information was 2C−125X65X6X8. At that time, a two-way beam string structure with 1860 MPa of cable was named BSS1.

The first variable is the number of struts. When struts were added to existing structures, it was able to check how values of results are changed. Therefore, it is to make a total of five struts by adding one strut each next to the post on both ends, which is the structure BSS1-$S_1$. Another is a structure that has a total of five struts by adding one strut each side of the center strut, which is BSS1-$S_2$. The last BSS1-$S_3$ has a total of seven struts between the two end posts by adding struts to all areas.

In this section, the second variable is cable angle control. When the cable is formed with the center of the circle at both ends of the upper beam and the center of the circle, the angle of theta is formed. The L (length of as the upper beam) and the H (length from the upper beam to the arch-shaped cable) are fixed, and the h of strut length and the c of cable length are indicated by the below formula. If the angle of the cable is different, t hen the length of each cable and the length of the strut are also different. Consequently, the angles of the sagging cable and arch-shaped cable were controlled to make the various structures.

On this occasion, the angle of the sagging cable is top of angle and the name is $\theta_T$. On the other hand, the angle of the arch-shaped cable is bottom of angle and the name is $\theta_B$. Moreover, the angle of the cable is marked N according to the angle which is $\theta_N$ . In additions, it was changed the length of cables and struts as the angle of cable increased and decreased. On this occasion, the length of the sagging cable is $C_T$. Contrarily, the length of the arch-shaped cable is $C_B$. Similarly, the length of the strut to the sagging cable is $H_T$ and the length of the strut to the arch-shaped cable is $H_B$. Therefore, the length of each cable and strut presented. Depending on the types of cable angle control, the structure appeared to change.

In the positive pressure, the shape of the structure varies according to the ratio of $L/H_T$ , it was called the shallow type when the ratio of $L/H_T \geq 7.5$, it was called the moderate type when it was the ratio of $7.5 > L/HT \geq 3.5$, it was called deep type when the ratio of $L/HT < 3.5$. On the other hand, in the negative pressure, it was called the shallow type when the ratio of L/HT  6.3, it was called the moderate type when it was the ratio of $6.3 > L/HT \geq 3.2$, it was called deep type when the ratio of $L/HT < 3.2$.

According to the formula, the radius of top circle RT and the radius of bottom circle $R_B$ are computed as

$$RT = RB = L/2sin\theta T = L/2sin\theta B \quad (1)$$

2 sin $\theta_T$ 2 sin $\theta_B$ The length of the strut at the center of the beam $H_T$ is computed as The length of the strut at the ends of the beam $H_B$ is computed as The length of the sagging cable $C_T$ and the length of the arch-shaped cable $C_B$ are computed as

**Analytical result and discussion**

In this study, it decided to add struts between posts on both ends of the original structure to see how the results change. To increase the number of struts in the original structure, it consisted of three types of cable. It appeared that BSS1-S1, BSS1-S2, and BSS1-S3. The results showed that BSS1-S3, which added struts to all areas, had slightly larger deflection values against load in the positive pressure. Accordingly, three struts were the ideal for this structure. Initially, the number of struts was added to both ends, the center, and all areas. However, the load-deflection curve showed little change. Therefore, it appeared that the three were the most ideal layout.

In this section, it is a curve of the ultimate load values for all structures by dividing into positive and negative pressures. Consequently, the load increased as the angle multiplied. This is the most optimized structure with an angle of 60 degrees when considering ultimate loads.

Second, it is a curve of the initial stiffness values for all structures by dividing into positive and negative pressures. Similarly, the initial stiffness value of all structures was accurately compared by dividing the initial stiffness values by the length of each cable as shown Fig. 16. The result was the largest value at 70 degrees.

The stress values on the top and bottom of the beam and the cable are plotted. First, the stress of all members is reduced as the angle increases in the positive pressure. Among these, it is illustrated two types of stress diagrams by selecting a structure with a cable angle of 30 and 80 degrees. Subsequently, a stress diagram of 80 degrees showed that the stress value at the bottom of the beam was almost zero. The curve of the next negative pressure shows that the stress value is reduced as the angle increases.

Therefore, it calculated the stress sharing ratio of the top and bottom of the beam and the cable with these two curves. In conclusion, the stress divided 60% of the cable and 40% of the beam when the angle was 10 degrees in the positive pressure. Contrarily, the stress sharing ratio of the beam was close to zero as the angle went over 50 degrees. This is proof that beam is not doing the role. In the negative pressure, the stress sharing ratio of the top and bottom of the beam was approximately the same depending on the angle. The beam's ratio was about 40% when the angle was 10 degrees, but it is found that the ratio dropped to about 20% when the angle was 90 degrees.

**Table 1. Specification of structures.**

| Notation | BSS1 |
|---|---|
| Number of beams | 1 |
| Number of struts | 3 |
| Number of posts | 2 |
| Number of sagging cable | 1 |
| Number of arch-shaped cable | 1 |
| Length of beam (mm) | 4000 |
| Beam | |
| Material | SM490B |

| Size | H−150X150X7X10 |
|---|---|
| Sagging cable | |
| Yield stress (MPa) | 1860 |
| Size | ϕ17.8 |
| Arch-shaped cable | |
| Yield stress (MPa) | 1860 |
| Size | ϕ12.59 |
| Strut | |
| Material | SS400 |
| Size | 2C−125X65X6X8 |

**Table 2. Stress sharing ratio of the top and bottom of beam and the sagging cable**

| $\theta_T$ | $L/H_T$ | Top(%) | Bottom(%) | S.Cable(%) |
|---|---|---|---|---|
| 10 | 22.9 | 21 | 17.4 | 61.6 |
| 15 | 15.2 | 21 | 16.6 | 62.4 |
| 20 | 11.3 | 20.5 | 14.5 | 65 |
| 25 | 9 | 19.3 | 11.7 | 69 |
| 30 | 7.5 | 17.9 | 9.2 | 72.9 |
| 35 | 6.3 | 15.8 | 6.6 | 77.6 |
| 40 | 5.5 | 14.1 | 4.7 | 81.3 |
| 45 | 4.8 | 12.4 | 2.7 | 85 |
| 50 | 4.3 | 11.5 | 1.3 | 87.2 |
| 55 | 3.8 | 10.7 | 0 | 89.3 |
| 60 | 3.5 | 9.5 | 0 | 90.5 |
| 65 | 3.1 | 9.4 | 0 | 90.6 |
| 70 | 2.9 | 9.6 | 0 | 90.4 |
| 75 | 2.6 | 9.9 | 0 | 90.1 |
| 80 | 2.4 | 10.7 | 0 | 89.3 |
| 85 | 2.2 | 12 | 0 | 88 |
| 90 | 2 | 14 | 0 | 86 |

**Table 3. Stress sharing ratio of the top and bottom of beam and the arch-shaped cable**

| $\theta_B$ | $L/H_B$ | Top(%) | Bottom(%) | A.Cable(%) |
|---|---|---|---|---|
| 10 | 14.5 | 21.6 | 21.6 | 56.8 |
| 15 | 11 | 21.9 | 22.2 | 56 |
| 20 | 8.8 | 21.4 | 22.5 | 56.1 |
| 25 | 7.4 | 20.5 | 22 | 57.6 |
| 30 | 6.3 | 19.2 | 20.8 | 60 |

| | | | | |
|---|---|---|---|---|
| 35 | 5.5 | 17.4 | 19.3 | 63.3 |
| 40 | 4.8 | 16 | 18 | 66 |
| 45 | 4.3 | 14.8 | 16.4 | 68.9 |
| 50 | 3.9 | 13.6 | 15.4 | 71 |
| 55 | 3.5 | 12.3 | 13.6 | 74 |
| 60 | 3.2 | 11.6 | 13 | 75.5 |
| 65 | 2.9 | 11 | 12.2 | 76.9 |
| 70 | 2.7 | 10.3 | 11.5 | 78.2 |
| 75 | 2.4 | 10.6 | 10.9 | 78.5 |
| 80 | 2.2 | 10.8 | 11.4 | 77.8 |
| 85 | 2.1 | 11.3 | 11.5 | 77.3 |
| 90 | 1.9 | 12.5 | 12.7 | 74.8 |

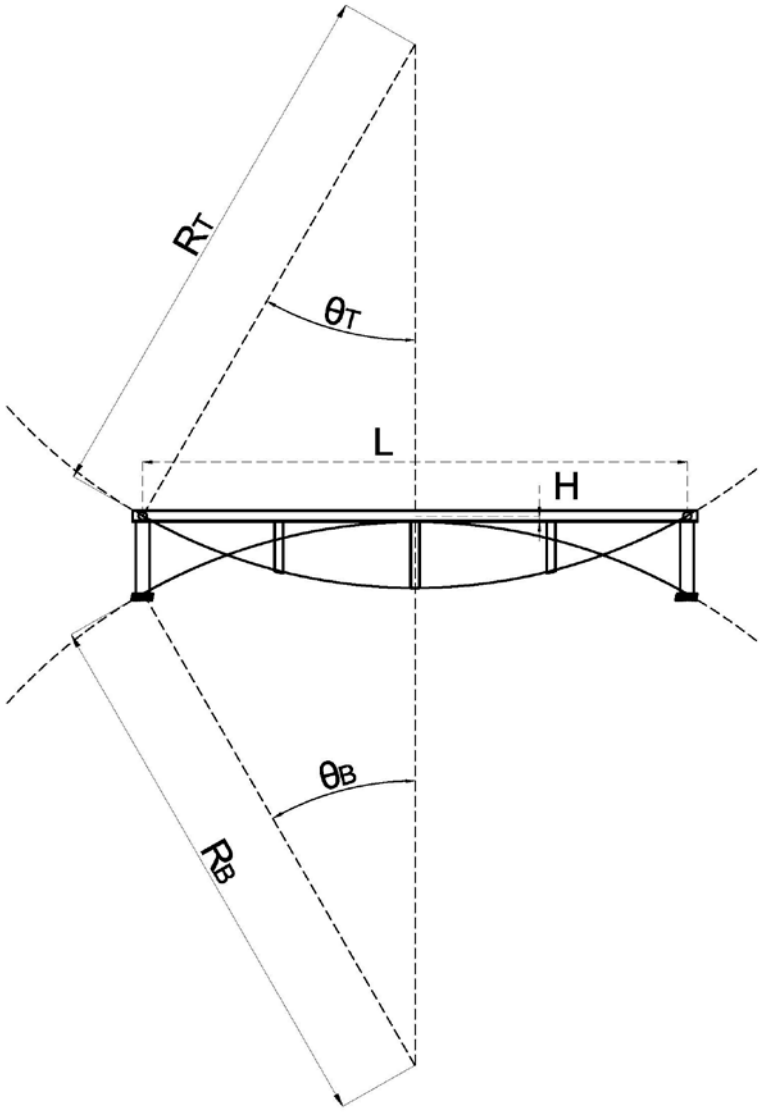



**Figure 1. Proposed two-way beam string structures with cable angle control at the top and bottom of the beam.**

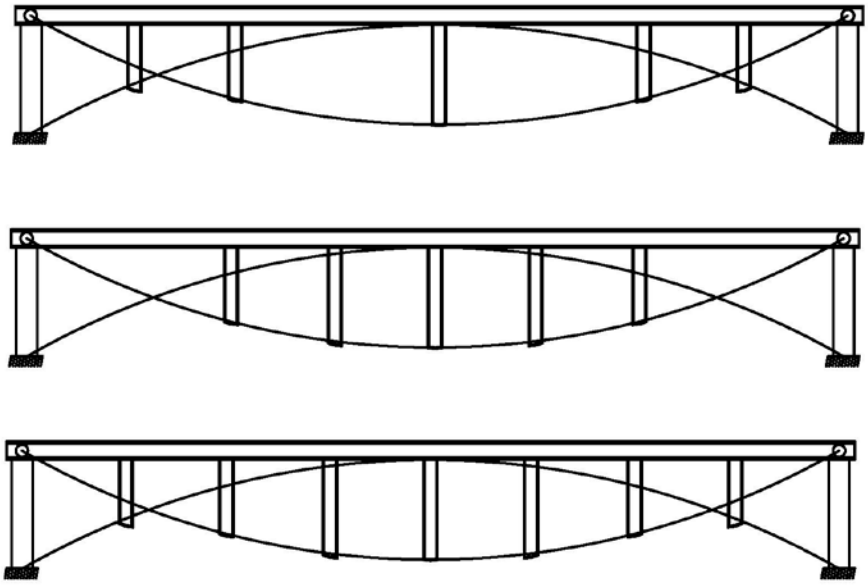


**Figure 2. Proposed two-way beam string structures with respect to the three types of struts.**

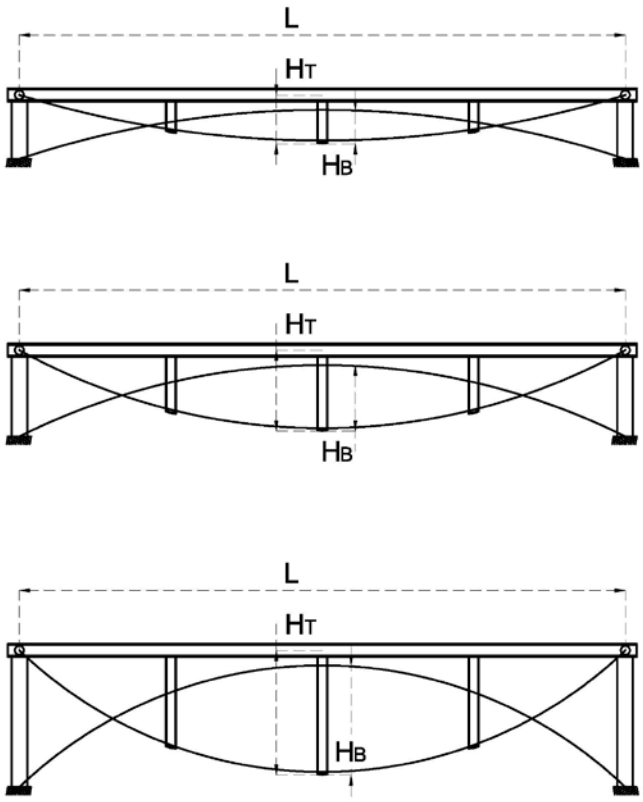



**Figure 3. Proposed two-way beam string structures with respect to shallow, moderate, and deep types depending on the cable angle control.**

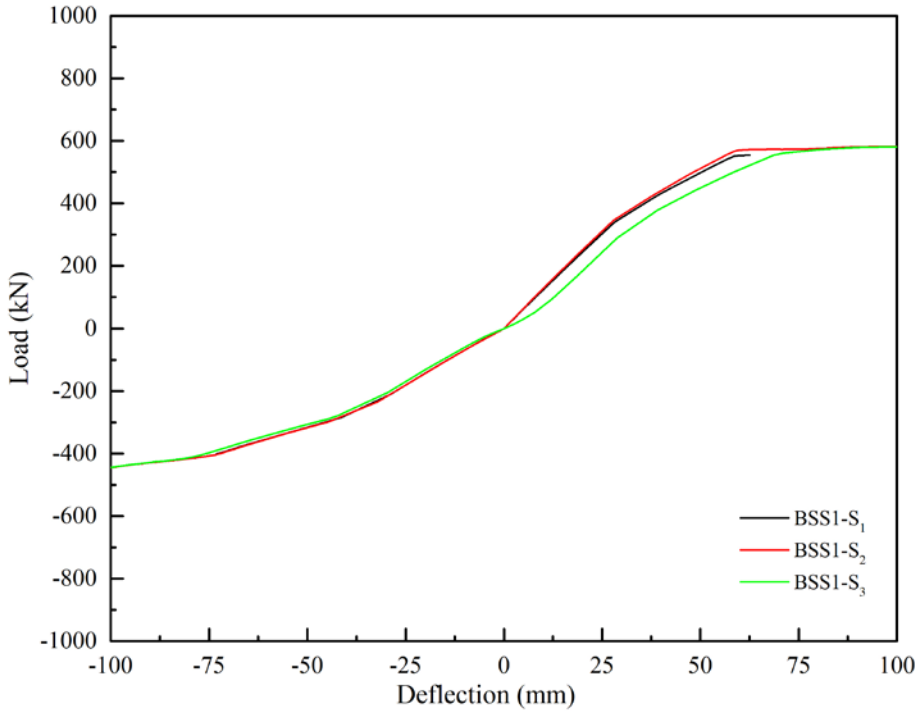



**Figure 4. Load-deflection curve of two-way beam string structures with respect to the three types of struts.**

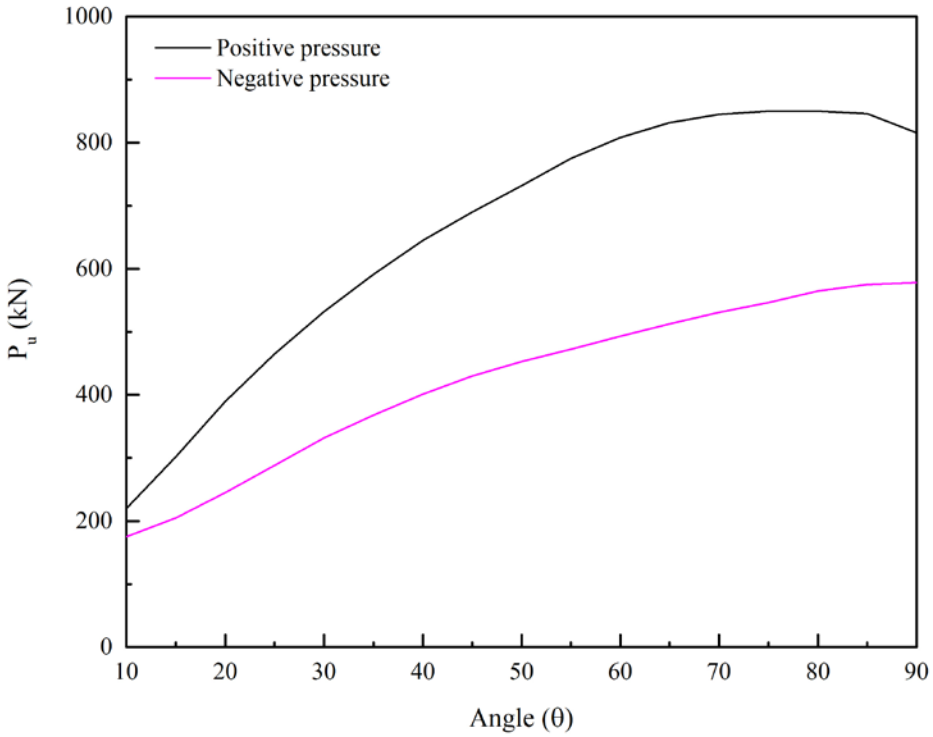



**Figure 5. Pu with respect to the angle**

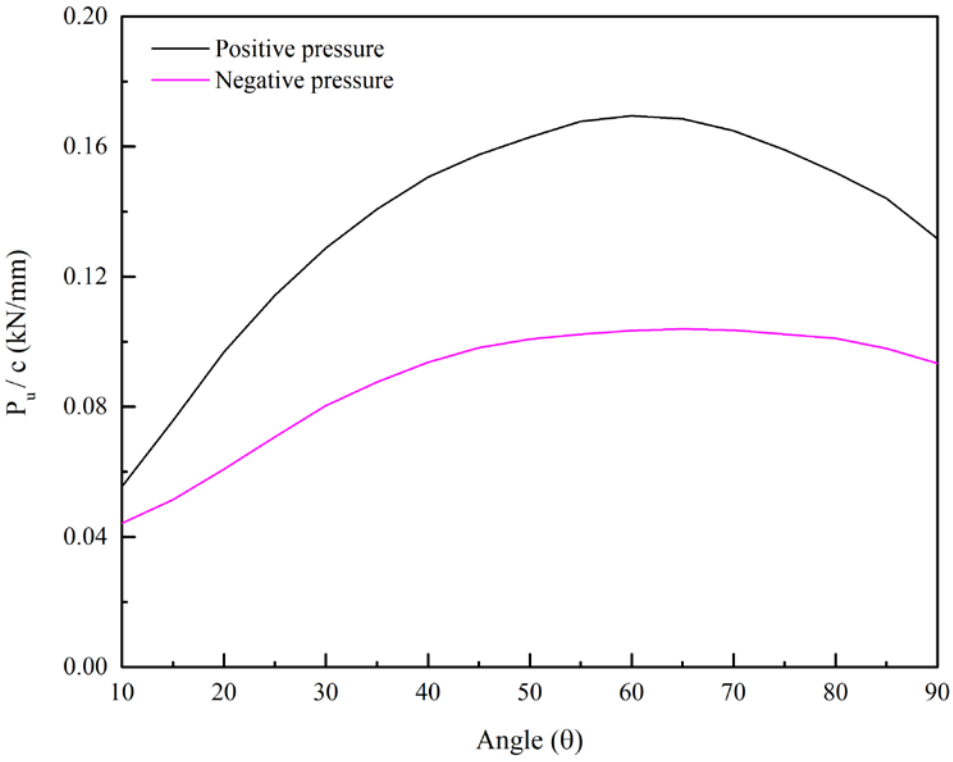



**Figure 6. Normalized Pu with respect to the angle**

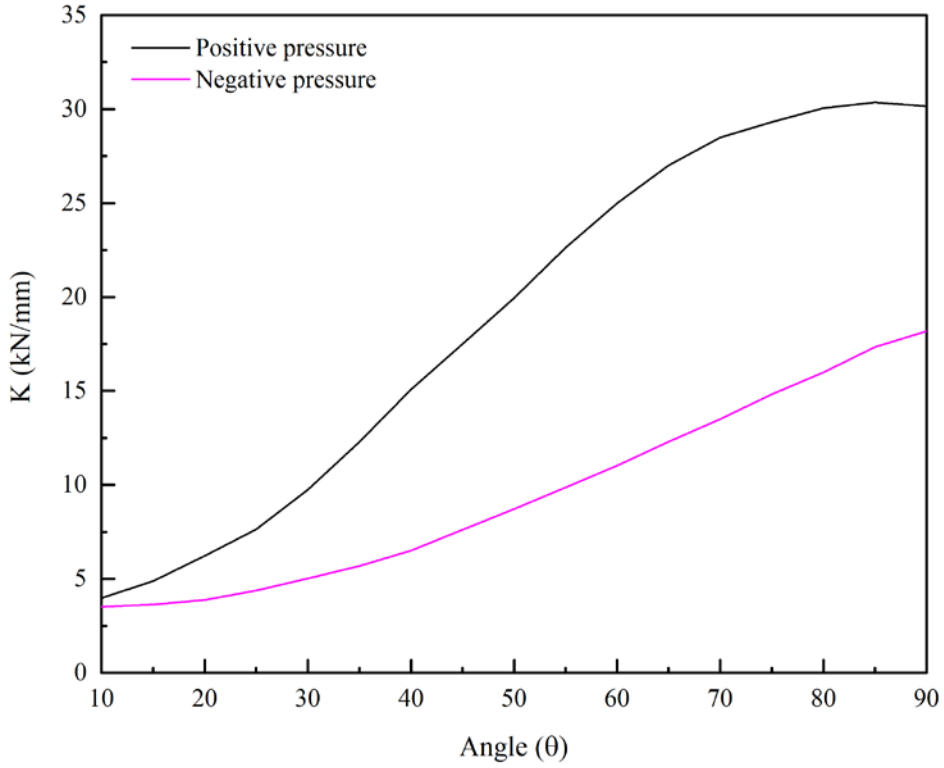



**Figure 7. K with respect to the angle**

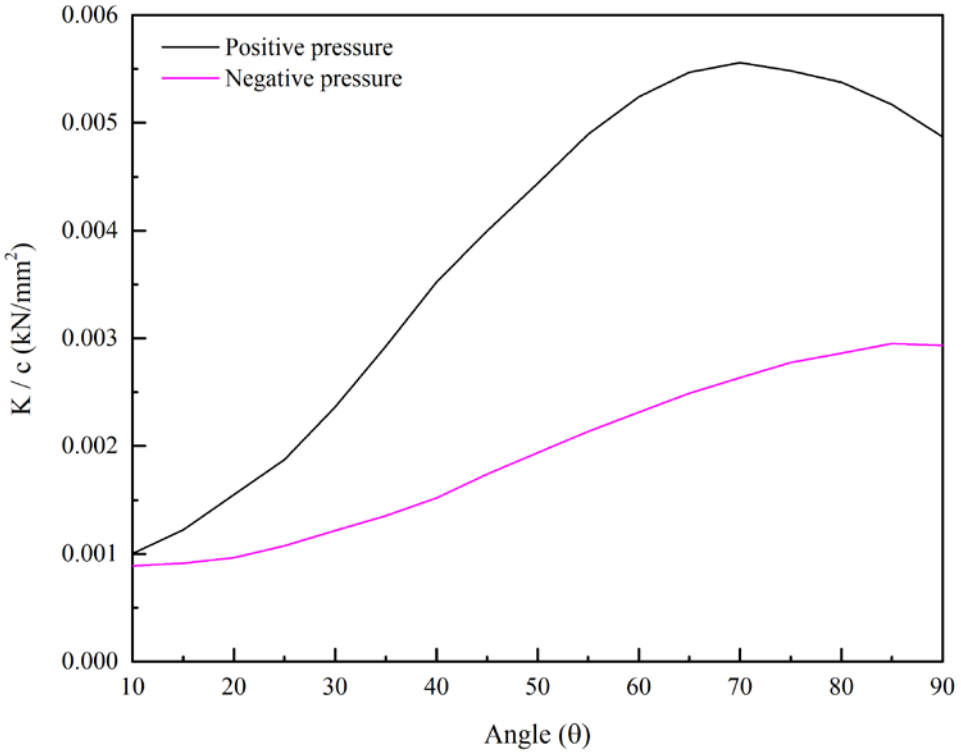



**Figure 8. Normalized K with respect to the angle**

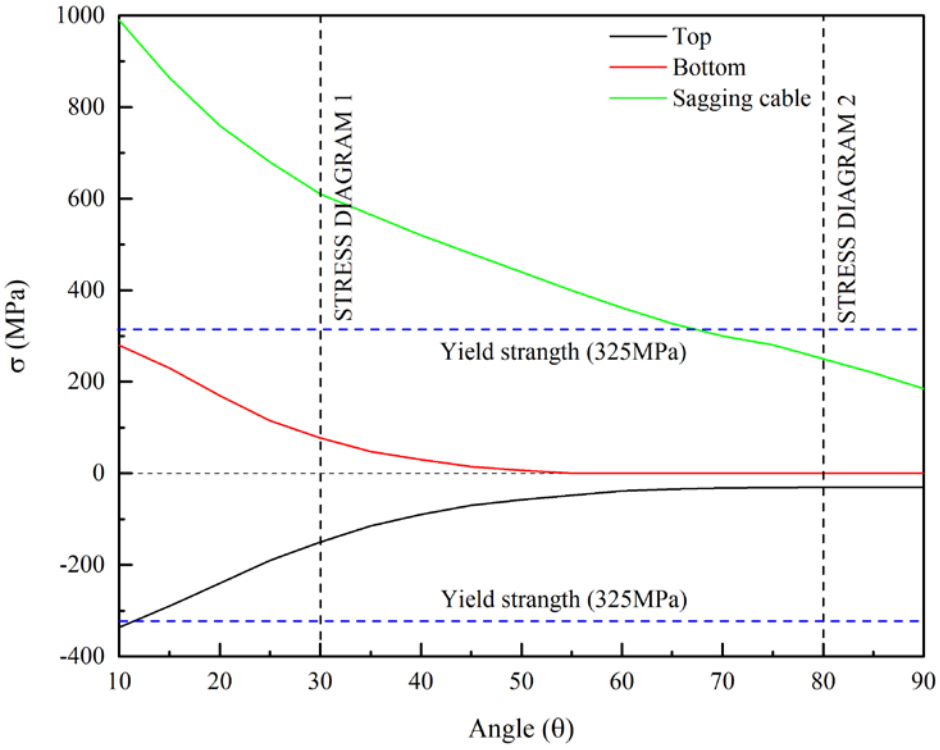



**Figure 9. Stress values with respect to the angle in the positive pressure**

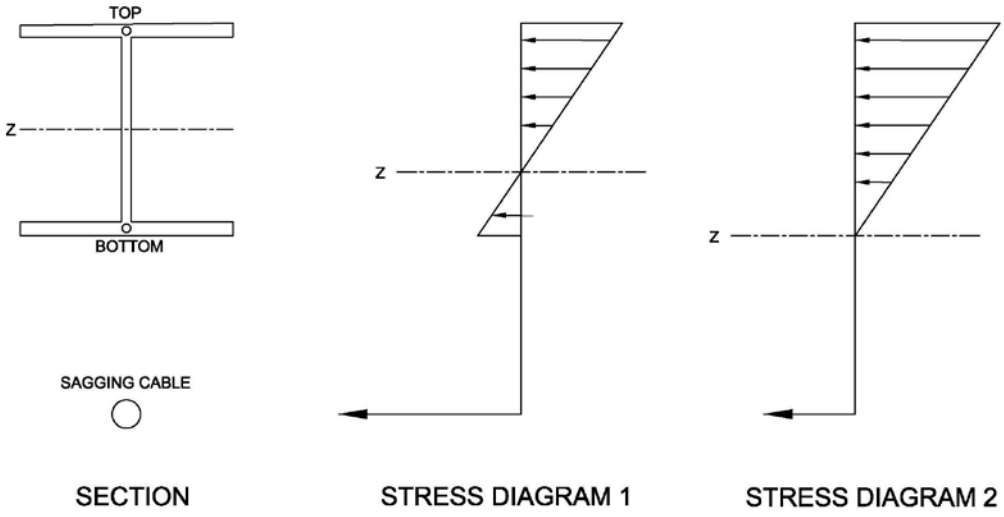



**Figure 10. Stress distribution of cross-section in the positive pressure**

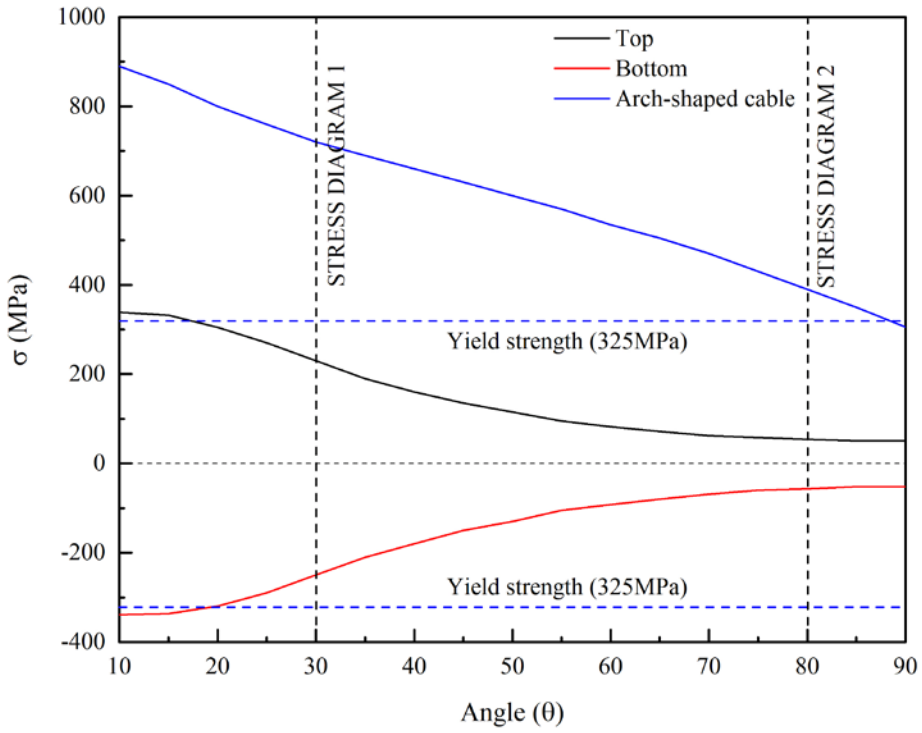



**Figure 11. Stres values with respect to the angle in the negative pressure**

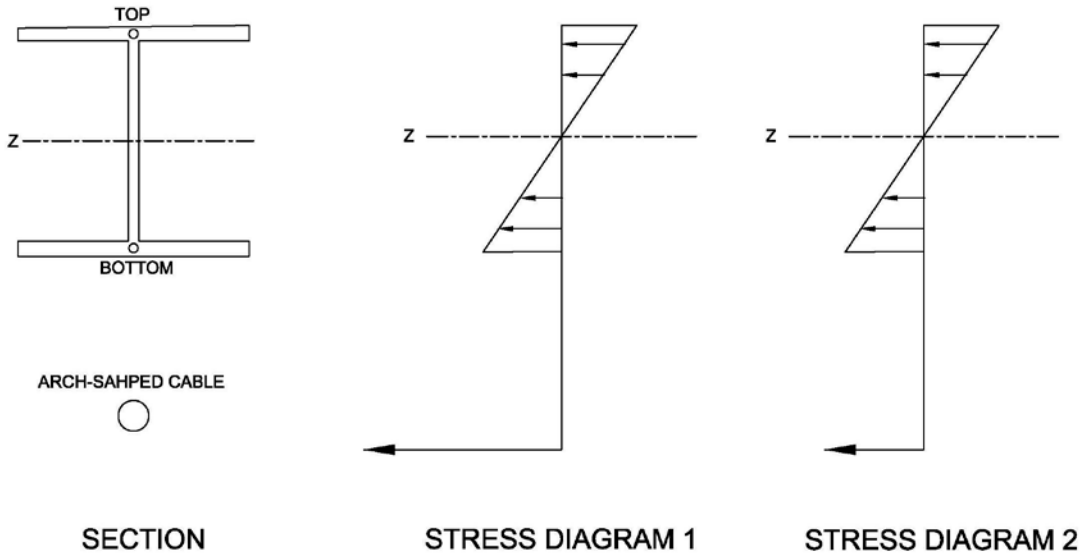



**Figure 12. Stress distribution of cross-section in the negative pressure.**

## Conclusions

From this previous study, a two-way beam string structure was produced that targeted at a two-way roof system. In this study, it performed how the behavior of the structure changes with respect to various variables. The propriety of the proposals was verified by means of reliability analysis. Consequently, the most optimal structure was made through various parametric studies in this study.

-The three struts were the most ideal except for the posts on both ends. Because the performance of the structure was not enhanced when the load-deflection curve was checked despite the addition of struts to the original structure.

-As considering ultimate loads, the load increased as the angle multiplied. However, the structure with cable angle of 60 degrees were most advantageous as a result of the normalized curve.

-As considering initial stiffness values, the structure with cable angle of 70 degrees were also most advantageous.

-When considering the stress of top and bottom of the beam and the cables, the larger the angle, the closer to 0 the stress of the beam is.

-According to the angle of cable increases, struts and cables became longer, which was disadvantageous for economic and aesthetic values.

-Thus, the structure with an angle of 30 to 35 degrees was founded to be the most optimal structure.

**Acknowledgements**

This research was supported by grants (NRF-2018R1C1B6004751, NRF-2018R1A2A1A05018287, NRF-2017R1A4A1015660) from NRF (National Research Foundation of Korea) funded by MEST (Ministry of Education and Science Technology) of Korean government.

**References**

[1] Hosozawa, O., Shimamura, K., and Mizutani, T. (1999). The role of cables in large span spatial structures: introduction of recent space structures with cables in Japan. Engineering structures, 21(8), 795-804.
[2] Saitoh, M., and Okada, A. (1999). The role of string in hybrid string structure. Engineering structures, 21(8), 756-769.
[3] Saitoh, M., and Tosiya, K. (1985). A study on structural behaviors of beam string structure. In Summaries of technical papers of annual meeting architectural institute of Japan (Vol. 1, pp. 280-284). Tokyo, Japan, B.
[4] Wu, M. (2008). Analytical method for the lateral buckling of the struts in beam string structures. Engineering Structures, 30(9), 2301-2310.
[5] Xue, W., and Liu, S. (2009). Design optimization and experimental study on beam string structures. Journal of Constructional Steel Research, 65(1), 70-80.
[6] Cai, J., Feng, J., and Jiang, C. (2014). Development and analysis of a long- span retractable roof structure. Journal of Constructional Steel Research, 92, 175-182.
[7] Zhengxing, L. W. S. J. G. (2003). Research on prestress stretching control of a large-span truss string structure [J]. Journal of Southeast University (Natural Science Edition), 5, 013.
[8] Yiyi, C., Zuyan, S., Xianzhong, Z., Yangii, C., Dasui, W., Chengyong, G., and Hongyu, C. (1999). Experimental study on a full-scale roof truss of Shanghai Pudong International Airport Terminal. Journal of Building Structures, 20(2), 9-17.
[9] Sun, W., Yang, S., and Chen, R. (2003). Stiffness performance of trussstring structure of Guangzhou International Convention and Exhibition Center. J South China Univ Technol (Nat Sci Ed), 31(11), 33-36.

# Application of the R-function Theory for the Bending Problem of Shallow Spherical Shells with Concave Boundary

***†Hong Yuan[1], Shanqing Li[1]**

[1] MOE Key Laboratory of Disaster Forecast and Control in Engineering, Institute of Applied Mechanics, Jinan University, Guangzhou 510632, China

*Presenting author: tyuanhong@jnu.edu.cn
†Corresponding author: tyuanhong@jnu.edu.cn

## Abstract

The R-function theory is applied to describe a shallow spherical shells on Winkler foundation with concave boundary, and then a quasi-Green's function is established by using the fundamental solution and the normalized boundary equation. The quasi-Green's function satisfies the homogeneous boundary condition of the problem. The differential equation of the problem is reduced to two simultaneous Fredholm integral equations of the second kind by the Green's formula. The singularity of the kernel of the integral equation is overcome by choosing a suitable form of the normalized boundary equation. A comparison with the ANSYS finite element solution shows a good agreement, and it demonstrates the feasibility and efficiency of the present method.

**Keywords:** Green's function, R-function, integral equation, bending of shallow spherical shell, concave boundary

## Introduction

As a kind of structural forms, the shells and plates are widely used in various fields, such as, in the large-span roof, the underground foundation engineering, the hydraulic engineering, the large container manufacturing, the aviation, the shipbuilding, the missiles, the space technology, the chemical industry, and so on. Only few problems of the shells and plates with a regular geometric boundary and a simple differential equation can be solved with an analytical or a half analytical method. For most these problems with a geometry of arbitrary shape and a complex boundary condition, only numerical methods can be used to solve the problems, such as the boundary element method[1], the finite element method[2] and the finite difference method[3].

In the present paper, the R-function theory and the quasi-Green's function method (QGFM) proposed by Rvachev [4] are utilized. The bending problem of simply supported dodecagon shallow spherical shells on Winkler foundation with concave boundary is studied. The governing differential equation of the problem is decomposed into two simultaneous differential equations of lower order by utilizing an intermediate variable. A quasi-Green's function is established by using the fundamental solution and the boundary equation of the problem. This function satisfies the homogeneous boundary condition of the problem, but it does not satisfy the fundamental differential equation. The key point of establishing the quasi-Green's function consists in describing the boundary of the problem by a normalized equation $\omega = 0$ and the domain of the problem by an inequality $\omega > 0$. There are multiple choices for the normalized boundary equation. Based on a suitably chosen normalized boundary equation, a new normalized boundary equation can be established such that the singularity of

the kernel of the integral equation is overcome. For any complicated domain, a normalized boundary equation can always be found according to the R-function theory. Thus, the problem can always be reduced to two simultaneous Fredholm integral equations of the second kind without the singularity. Using the R-function theory, Li and Yuan described successfully the rectangular, trapezoidal, triangular and parallelogrammic domains of plates[5][6] and shallow spherical shells[7][8]. For the first time, the R-function theory is applied to describe the dodecagon domain of the shallow spherical shells with concave boundary. The numerical example demonstrates the feasibility and efficiency of the present method. The R-function theory can be used to describe any more complex domains of the plates and shells.

**Fundamental equations**

The governing differential equations of the bending problem of simply supported shallow spherical shells on Winkler foundation[9] can be expressed as follows.

$$\nabla^4\varphi(\mathbf{x})-\frac{Eh}{R}\nabla^2 w(\mathbf{x})=0\,,\qquad \mathbf{x}\in\Omega \tag{1}$$

$$D\nabla^4 w(\mathbf{x})+\frac{1}{R}\nabla^2\varphi(\mathbf{x})+kw(\mathbf{x})=P_Z\,,\quad \mathbf{x}\in\Omega \tag{2}$$

where $\nabla^4=(\partial^2/\partial x_1^2+\partial^2/\partial x_2^2)^2$ is the biharmonic operator, $\varphi$ is the stress function, $w$ is the radial deflection of the shell, $R$ is the radius of curvature of the shell, $k$ is the elastic coefficient of the foundation, $\mathbf{x}=(x_1,x_2)$, $\Omega$ is the domain of the trapezoid of shallow spherical shells in Cartesian coordinates, $P_Z$ is the radial load; and $D=Eh^3/(12(1-\nu^2))$ is the flexural rigidity of the shell, in which $h$ is the thickness of the shell, and $E$ and $\nu$ are Young's modulus and Poisson's ratio, respectively.
The simply supported boundary conditions can be written as.

$$w\big|_\Gamma=\nabla^2 w\big|_\Gamma=\varphi\big|_\Gamma=\nabla^2\varphi\big|_\Gamma=0 \tag{3}$$

where $\nabla^2=\partial^2/\partial x_1^2+\partial^2/\partial x_2^2$ is the Laplace operator, and $\Gamma=\partial\Omega$ is the boundary of the domain $\Omega$.Making use of Eqs.(1) and (3), we can easily obtain.

$$\nabla^2\varphi=wEh/R \tag{4}$$

Substituting Eq.(4) into Eq.(2) yields.

$$D\nabla^4 w+wEh/R^2+kw=P_Z \tag{5}$$

To decompose Eq.(5), let us introduce the following intermediate variable.

$$M=(M_1+M_2)/(1+\nu) \tag{6}$$

where $M_1=-D(\partial^2 w/\partial x_1^2+\nu\partial^2 w/\partial x_2^2)$ and $M_2=-D(\partial^2 w/\partial x_2^2+\nu\partial^2 w/\partial x_1^2)$.
Then, substituting Eq.(6) into Eq.(5), we obtain the following two simultaneous differential equations of second rank.

$$\nabla^2 M=-P_Z+wEh/R^2+kw \text{ and } \nabla^2 w=-M/D\,,\quad \mathbf{x}\in\Omega \tag{7}$$

The displacement and the bending moment should be equal to zero along the simply supported boundary of shallow spherical shells on Winkler foundation, which can be written as.

$$w = 0 \text{ and } M = 0, \quad \mathbf{x} \in \Gamma \tag{8}$$

**Integral equations**

Let $\omega = 0$ be the normalized boundary equation of the first-order on the boundary $\Gamma$, i.e.[4]

$$\omega(\mathbf{x}) = 0, \ |\nabla \omega| = 1, \quad \mathbf{x} \in \Gamma \text{ and } \omega(\mathbf{x}) > 0, \quad \mathbf{x} \in \Omega \tag{9}$$

The quasi-Green's function can be established as follows.

$$G(\mathbf{x}, \boldsymbol{\xi}) = \frac{1}{2\pi} \ln r - \frac{1}{2\pi} \ln R_1 \tag{10}$$

where $r = \|\boldsymbol{\xi} - \mathbf{x}\| = \sqrt{(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2}$ and $R_1 = \sqrt{r^2 + 4\omega(\boldsymbol{\xi})\omega(\mathbf{x})}$, in which $\mathbf{x} = (x_1, x_2)$ and $\boldsymbol{\xi} = (\xi_1, \xi_2)$. Obviously, the quasi-Green's function $G(\mathbf{x}, \boldsymbol{\xi})$ satisfies the following condition.

$$G(\mathbf{x}, \boldsymbol{\xi})\big|_{\xi \in \partial\Omega} = 0 \tag{11}$$

To reduce the boundary value problems Eqs.(7) and (8) into the integral equations, the following Green's formula of sets of function $C^2(\Omega)$, i.e., $U$ and $V \in C^2(\Omega \bigcup \Gamma)$, is applied.

$$\int_{\Omega} (V\nabla^2 U - U\nabla^2 V) \, \mathrm{d}_{\xi} \Omega = \int_{\Gamma} (V \frac{\partial U}{\partial n} - U \frac{\partial V}{\partial n}) \, \mathrm{d}_{\xi} \Gamma \tag{12}$$

From Eqs.(7), (8), (11)and (12), and noticing that $(1/2\pi)\ln r$ is the fundamental solution[10] of the Laplace operator, then the following integral equations are obtained.

$$w(\mathbf{x}) = -\frac{1}{D} \int_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) M(\boldsymbol{\xi}) \, \mathrm{d}_{\xi} \Omega + \int_{\Omega} w(\boldsymbol{\xi}) K(\mathbf{x}, \boldsymbol{\xi}) \, \mathrm{d}_{\xi} \Omega \tag{13}$$

$$M(\mathbf{x}) = \int_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) [-P_Z(\boldsymbol{\xi}) + \frac{Eh}{R^2} w(\boldsymbol{\xi}) + k w(\boldsymbol{\xi})] \mathrm{d}_{\xi} \Omega + \int_{\Omega} M(\boldsymbol{\xi}) K(\mathbf{x}, \boldsymbol{\xi}) \, \mathrm{d}_{\xi} \Omega \tag{14}$$

where

$$K(\mathbf{x}, \boldsymbol{\xi}) = [R_1^2 \omega(\mathbf{x}) \nabla^2 \omega + 4\omega(\mathbf{x})\omega - 4(\mathbf{r} \cdot \nabla \omega)\omega(\mathbf{x}) - 4\omega^2(\mathbf{x})(\nabla \omega)^2] / \pi R_1^4 \tag{15}$$

here $\omega = \omega(\boldsymbol{\xi})$, $\nabla = \nabla_{\xi}$; and $\mathbf{r} = (\xi_1 - x_1)\mathbf{i} + (\xi_2 - x_2)\mathbf{j}$, in which $\mathbf{i}$ and $\mathbf{j}$ denote unit vectors in $x_1$ and $x_2$ directions, respectively.

$K(\mathbf{x}, \boldsymbol{\xi})$ in Eq.(15) appears discontinuous only if $R = 0$, i.e., both $\mathbf{x} = \boldsymbol{\xi}$ and $\omega = 0$ come into existence. Actually, when $\mathbf{x} = \boldsymbol{\xi}$, Eq.(15) can be reduced to.

$$K(\mathbf{x}, \boldsymbol{\xi})\big|_{\mathbf{x}=\boldsymbol{\xi}} = [1 + \omega\nabla^2 \omega - (\nabla \omega)^2] / 4\pi\omega^2 \tag{16}$$

To make the kernel of the integral equations $K(\mathbf{x},\xi)\in C(\Omega\cup\partial\Omega)$, A normalized boundary equation will be constructed to ensure the continuity of $K(\mathbf{x},\xi)$ in the following. It can be easily testified that.

$$\omega=[3\omega_0+\omega_0^2\nabla^2\omega_0-\omega_0(\nabla\omega_0)^2]/2 \tag{17}$$

where $\omega_0=0$ is the normalized equation on the boundary $\Gamma$, i.e., $\omega_0$ satisfies Eq.(9). Obviously, equation $\omega$ is also a normalized boundary equation of the first-order.
Based on a suitably chosen normalized boundary equation $\omega_0=0$, a new normalized boundary equation $\omega=0$ can be constructed by using Eq.(17), which ensure the continuity of the integral kernel $K(\mathbf{x},\xi)$ in the integral domain.
To obtain the numerical results of the boundary problem, the integral domain $\Omega$ is divided into several subdomains $\Omega_i(i=1,2,.....,N)$, and in each subdomain, a rectangular quadrature formula is applied. Thus, the integral equations (13) and (14) can be discretized into the linear algebraic equations. Then, the radial deflection $w(\mathbf{x})$ can be obtained by solving the algebraic equations.

**Numerical example**

We investigate a simply supported dodecagon shallow spherical shell on Winkler foundation with the planform shown in Fig.1. Take $a$ =60, $b=80$, $c=40$ and $d$ =30. The following reference parameters are used: the radius of curvature of the shell $R$ =200, the thickness of the shell $h$ =2, Poisson's ratio $\nu=0.3$, Young's modulus $E=2.1\times10^6$, the elastic coefficient of the foundation $k$ =200, and the radial load $P_Z$ =70. According to the R-function theory[4], a normalized boundary equation of the first rank $\omega_0=0$ can be constructed from the following equation:

$$\omega_0=(\omega_1\wedge_0\omega_2)\wedge_0(\omega_1\vee_0\omega_2) \tag{18}$$

where $\omega_1=(a^2-x_1^2)/2a\geq0$ is the vertical band limited by straight lines $x_1=\pm a$;
$\omega_2=(b^2-x_2^2)/2b\geq0$ is the horizontal band limited by straight lines $x_2=\pm b$;
$\omega_3=(c^2-x_1^2)/2c\geq0$ is the vertical band limited by straight lines $x_1=\pm c$;
and $\omega_4=(x_2^2-d^2)/2d\geq0$ is the outer part of the band limited by straight lines $x_2=\pm d$.
The Boolean operations $\vee_\alpha$, $\wedge_\alpha$ (disjunction and conjunction), which correspond to the union $\cup$ and intersection $\cap$. These R-operations are defined as follows[1]:

$$X\wedge_\alpha Y=\frac{1}{1+\alpha}(X+Y+\sqrt{X^2+Y^2-2\alpha XY}),\ X\vee_\alpha Y=\frac{1}{1+\alpha}(X+Y-\sqrt{X^2+Y^2-2\alpha XY}),$$

where the parameter $\alpha$ varies within $-1<\alpha\leq1$. For example, if the value $\alpha$ is equal to zero,then the whole domain can be presented as Eq.(18). $\omega_1,\omega_2,\omega_3$ and $\omega_4$ are a normalized equation of the first rank. $\omega_1=0$, $\omega_2=0$, $\omega_3=0$ and $\omega_4=0$ denote various parts of the boundary of the dodecagon shallow spherical shell on Winkler foundation, respectively. The radial deflection curves of line $x_2=0$ and line $x_1=0$ for different $k$ and different $R$ by the QGFM and by the ANSYS finite element method (FEM) are shown in Figs.2-5 for a comparison, respectively; a good agreement is observed between the two methods.

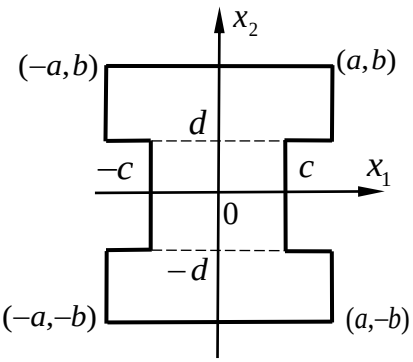



**Fig.1 Dodecagon planform**

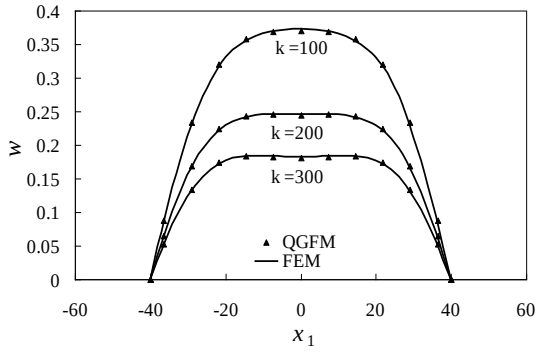



**Fig.2 The deflection curve of line $x_2=0$ in Fig.1 for different $k$**

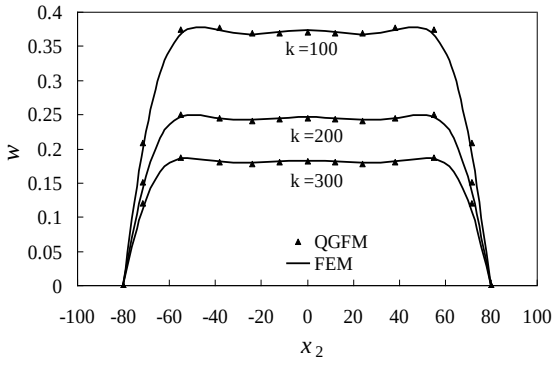



**Fig.3 The deflection curve of line $x_1=0$ in Fig.1 for different $k$**

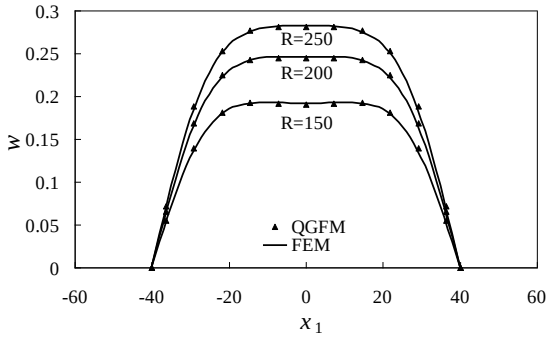



**Fig.4 The deflection curve of line $x_2=0$ in Fig.1 for different $R$**

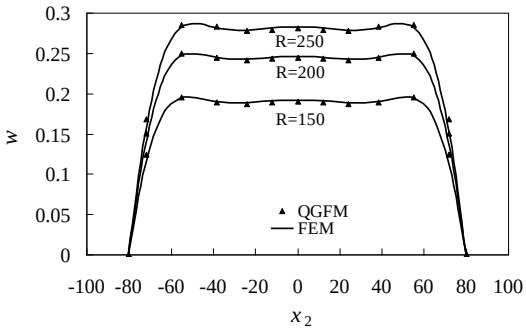



**Fig.5 The deflection curve of line $x_1=0$ in Fig.1 for different $R$**

## Conclusions

In the present paper, the R-function theory is applied to describe a shallow spherical shells on Winkler foundation with concave boundary, and it is used to construct a quasi-Green's function. Compared with the FEM solution, the numerical results of the QGFM demonstrate its feasibility, efficiency and rationality. The R-function theory can also be used to effectively solve various boundary value problems of the plates and shells by constructing a trial function

that satisfies the complex boundary shape and by combining with the other method of weighted residuals such as the variational method[11] and the spline-approximation[12].

## Acknowledgements

The authors gratefully acknowledge the financial support provided by "the Fundamental Research Funds for the Central Universities" of China (No.11611769, No. 21611455)

## References

[1] Kong, F. Z., Zheng, X. P. and Yao, Z. H. (2005) Numerical simulation of 2D fiber-reinforced composites using boundary element method. *Applied Mathematics and Mechanics (English Edtion)* **26**(11), 1515-1522 .

[2] Sestak, I. and Jovanovic, B. S. (2010) Approximation of thermoelasticity contact problem with nonmonotone friction. *Applied Mathematics and Mechanics (English Edtion)* **31**(1),77-86

[3] Zhao, W. J., Chen, L. Q. and Jean, W. Z. (2006) A finite difference method for simulatting transverse vibrations of an axially moving viscoelatic string, *Applied Mathematics and Mechanics (English Edtion)* **27**(1), 23-28

[4] Rvachev, V.L. (1982) Theory of R-function and Some of its Application, *Kiev: Nauk Dumka*, 415-421. (in Russian)

[5] Shan-qing Li, Hong Yuan. (2011) Quasi-Green's function method for free vibration of clamped thin plates on Winkler foundation, *Applied Mathematics and Mechanics (English Edtion)* **32**(3), 265-276.

[6] Hong Yuan, Shan-qing Li, and Ren-Huai Liu. (2007) Green quasifunction method for vibration of simply-supported thin polygonic plates on Pasternak foundation, *Applied Mathematics and Mechanics (English Edtion)***28** (7): 847-853.

[7] Shan-qing Li, Hong Yuan. (2010) Quasi-Green's function method for free vibration of simply-supported trapezoidal shallow spherical shell, *Applied Mathematics and Mechanics (English Edtion)* **31**(5), 635-642.

[8] Shanqing Li, Hong Yuan. (2010) Green quasifunction method for free vibration of simply-supported trapezoidal shallow spherical shell on Winkler foundation, *Acta Mechanica Solida Sinica* **23**(4), 370-376.

[9] Paliwal,D.N. and Sinha,S.N. (1986) Static and dynamic behaviour of shallow spherical shells on winkler foundation, *Thin-Walled Structures* **4**(6), 411-422.

[10] Ortner, V.N. (1980) Regularisierte faltung von distributionen. Teil 2: Eine tabelle von fundamentallocunngen, *Zeitschrift für angewandte Mathematik und Physik* **31**(1), 155-173.

[11] Kurpa, L.V., Shmatko,T. and Timchenko, G. (2010) Free vibration analysis of laminated shallow shells with complex shape using the R-functions method, *Composite Structures* **93**(1), 225-233.

[12] Awrejcewicz, J., Kurpa, L.V. and Osetrov, A. (2011) Investigation of the stress-strain state of the laminated shallow shells by R-functions method combined with spline-approximation, *Zeitschrift für Angewandte Mathematik und Mechanik* **91**(6), 458-467.

# Effects of hydrodynamic pressure on the seismic responses of underground vertical shafts

***Bu Zhang[1] and †Zhiyi Chen[1,2]**

[1]Department of Geotechnical Engineering, Tongji University, Shanghai 200092, China
[2]Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Shanghai 200092, China

*Presenting author: zhangbu@tongji.edu.cn
†Corresponding author: zhiyichen@tongji.edu.cn

## Abstract

More and more deep drainage and storage pipeline systems play an important role in preventing urban flooding and waterlogging in urban 'sponge cities'. As the main structures in this systems, shafts are used to collect, reserve and transport the water. The effect of inner water on seismic response of a deep shaft is studied with three different levels of inner water: full, half, and empty. The hydrodynamic pressure is simulated with the CAS (coupled acoustics and structure) method which can consider not only the impulsive pressure effect, but also the convective pressure effect. Results present that the seismic displacement responses of shaft under different water levels are nearly consistent, while the shaft inner forces increase with the increase of the inner water level.

**Keywords:** Vertical shaft, Inner water, Coupled acoustics and structure method, Seismic Response

## 1. Introduction

With the construction and development of urban cities, especially the city resilience in controlling and preventing the city flood in heavy rainstorm, the construction of the sponge city gradually becomes a focus. The underground reservoir shaft, as a main underground structure and the important part of the sponge city project, is used to store up and discharge the water. Due to the non-renewable characteristic of underground space and the great difficulties in restore of the underground structure, the safety of reservoir shaft under earthquake excitation has become a rising concern.

Many researchers have studied the hydrodynamic distribution formulation of the aboveground tank under seismic excitation [1-3]. They assumed the tank as a rigid body with only translation motion. The added mass method was proposed to simulate the effect of hydrodynamic impulsive pressure on the above ground tank [4-6], which are usually neglected the effect of the connective pressure.

Simplified analytical method [7], experiment test [8] and numerical method [9-11] was used to explore the seismic performance of shallow underground tanks with inner water. And the results showed that the effect of the sloshing of inner water can be neglected, which is the same as the above ground tanks. However, as a vertical embedded structure, the shaft not only behaves with translation motion but also with rotational motion under earthquake excitation. Obviously, the effect of hydrodynamic pressure on the underground reservoir, especially on the shaft inner forces and dynamic displacements, should be further researched.

In order to simulate the dynamic responses of shaft inner water, the hydrodynamic pressure needs to be simulated accurately. The CEL (Coupled Euler and Lagrange), SPH (Smoothed Particle Hydrodynamic), and CAS (Coupled Acoustics and Structure) methods all can be used to simulate the dynamic behavior of inner water in detail. For the SPH method [12-14], the water is simulated as uniform distributed partials. However, the particles need to be intensive in order to obtain the accurate results. For the CEL method [15-17], the shaft is modeled as Lagrange element and the inner water is modeled with the Euler element; in order to simulate the water motion, the Euler elements need to be refined, which would increase the element number especially in 3D model. For the CAS method [17, 18], the Acoustics element with pressure DOF can be used to obtain the pressure of inner water, and the calculation efficiency is higher than the above two methods. Taking the calculation efficiency into account, the CAS method is efficient in simulating the sloshing and hydrodynamic behavior of inner water and shafts. Thus the CAS method is adopted in this paper.

In this paper, three cases with full water, half water, and no water have been simulated and the effects on the shaft displacement and inner forces distribution pattern has been investigated with CAS method.

## 2. Numerical modelling

### *2.1 Site and drainage shaft*

Here in order to carry out parameter sensitivity analyses, a series of 3D dynamic time history analyses of shafts are performed with the large-scale commercial finite element software *ABAQUS* [19]. The model is shown in Fig. 1a. The model is 400 m in the horizontal directions, which is more than five times of the shaft diameter to prevent boundary-waves reflection. Because the wave velocity of the bottom soil layer is greater than 600 m/s, the bottom soil layer is considered as the bedrock; hence, the model is 120 m in the vertical direction. The depth of the shaft is 60 m and the baseboard thickness is 4 m. The thickness of shaft side wall is uniform along depth. Fig. 1b shows the model of inner water which simulated by the AC3D8R element. Fig. 1c shows the central cross-section of the shaft, in which the lining is simulated by S4R element and the baseboard is simulated by C3D8R element. The parameters of the shaft in drainage system are listed in Table 1.

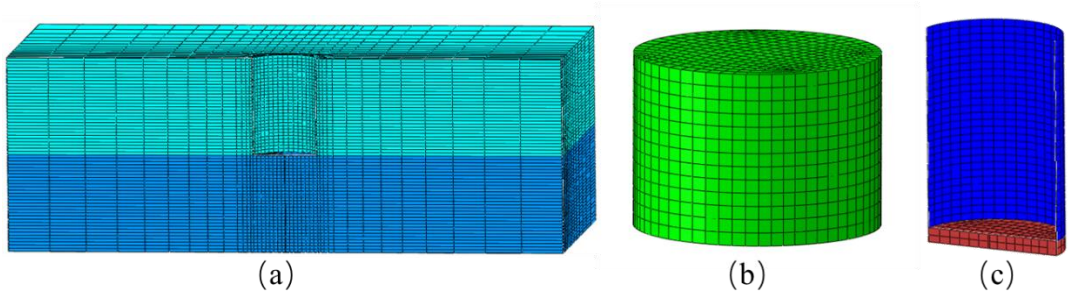



**Figure 1. Schematic diagram of vertical section of the three-dimensional model: (a) vertical profile of surrounding soil; (b) inner water and (c) vertical profile of shaft**

**Table 1. Geometrical parameters of the shaft and surrounding soils**

| Name | Depth $D$ (m) | Density $\rho$ (kg/m$^3$) | Shear-wave velocity $v_s$ (m/s) | Poisson's ratio $u$ | Elasticity modulus $E$ (MPa) |
|---|---|---|---|---|---|
| Soil-1 | 30 | 1850 | 250 | 0.3 | 300 |
| Soil-2 | 30 | 1900 | 300 | 0.3 | 520 |
| Shaft | 60 | 2500 | - | 0.2 | 34500 |

*2.2 Inner water*

The inner water is simulated by the Acoustics element AC3D8 with pressure DOF. In this model, three different water levels are considered. The height of water level in each case is 0, 28m, 56m, respectively. The detail parameters of inner water are listed in Table 2.

**Table 2. Parameters of inner water**

| Calculation Cases | Water Depth $H$ (m) | Density $\rho_w$ (kg/m$^3$) | Dynamic viscosity $\mu$(N sec/m$^2$) | Bulk modulus $K$ (MPa) |
|---|---|---|---|---|
| No water | 0 | | | |
| Half water | 28 | 1000 | 0.001 | 2140.4 |
| Full water | 56 | | | |

*2.3 Boundary conditions*

MPC boundary conditions are obtained for lateral boundary conditions in the dynamic procedure, allowing it to move as a free field. The boundary at the model top is free. The vertical degree of the model base is fixed, while the input seismic motion is applied in horizontal direction.

For the inner water, the impedance boundary conditions are applied for the free surface in order to simulate the effect of the sloshing effect. The acoustic particle velocity in the outward normal direction of fluid surface, $\dot{u}_{out}$, is related to the pressure as well as the change rate with time of pressure as follows:

$$\dot{u}_{out} = \frac{1}{k}\frac{\partial p}{\partial t} + \frac{1}{c}p \tag{1}$$

where, $p$ is the acoustic pressure, $1/k$ is the proportionality coefficient between the pressure and the displacement in normal direction to the surface, and $1/c$ is the proportionality

coefficient between the pressure and the velocity in normal direction to the surface ( $1/k=1/\rho_w g$, $1/c=0$) [17-19].

*2.4 Input seismic motion*

An artificial earthquake motion for Shanghai is input horizontally from the model base. Fig. 2 shows the time history of the seismic excitation signal and its Fourier spectrum. The ground motion is uniform without taking the wave travelling effects into account. The ground motion has a duration of 20 s with a maximum value of 0.1g. The frequency is mostly between 0 and 10 Hz.

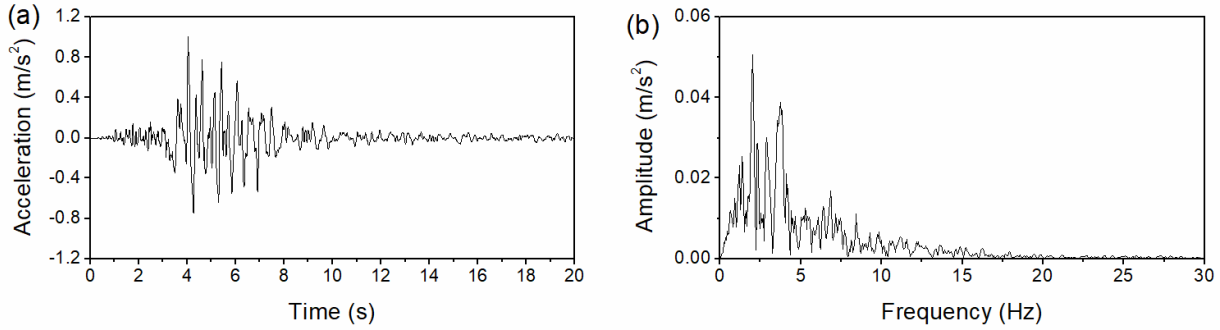



**Figure 2. (a) Acceleration time history curve and (b) Fourier amplitude of the Shanghai artificial earthquake motion**

## 3. Results and discussion

*3.1 Sloshing displacement*

In the FEM model, the acoustic pressure is obtained directly at the top surface of the liquid and the sloshing displacement, *h*, can be obtained by using the correlation formulation [17]:

$$h = \frac{p}{\rho_w g} \tag{2}$$

where, $p$ is the hydrodynamic pressure of free surface, $\rho_w$ is the density of water, $g$ is the gravity acceleration.

Fig. 3 shows the sloshing displacement time history response of the water free surface. A probe is placed on the initial free surface connecting the shaft right side wall. Compared with the shaft sloshing displacement, the sloshing tendency is nearly the same. It can be seen that maximum sloshing displacement amplitude is 10 cm for the full water case, while the maximum sloshing displacement amplitude is 6 cm for the half water case. The magnitude of the sloshing displacement is relatively small compared with the inner water height.

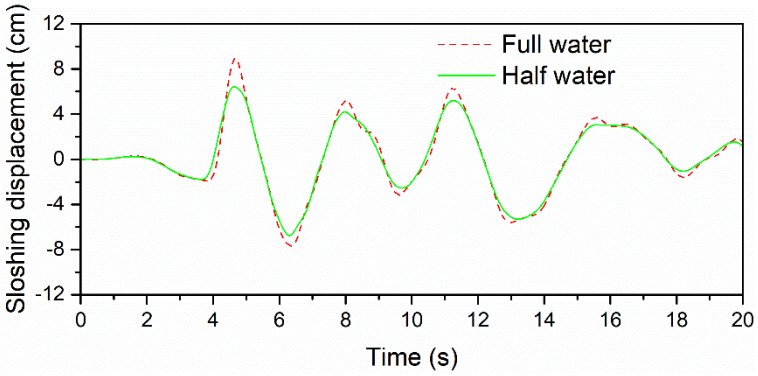



**Figure 3. Sloshing displacement time-history responses of the shaft subjected to earthquake ground motions**

*3.2 Distribution of hydrodynamic pressure*

The maximum hydrodynamic pressure distribution pattern is extracted at the time t=6.84s, when the magnitude of the shaft bottom acceleration reaches the maximum. In order to represent the distribute pattern, Fig. 4 shows the distribution of normalized hydrodynamic pressure along the height of the shaft. From Fig. 4 it can be seen that the normalized hydrodynamic pressure in both the full water and half water cases are nearly the same, while the negative magnitude at the top in the full water case indicates the effect of the convective pressure.

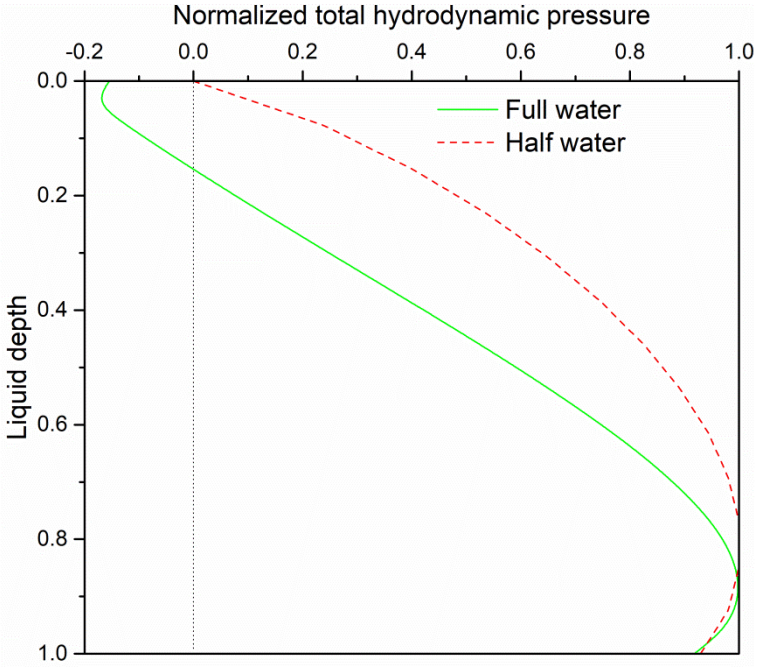



**Figure 4. Normalized hydrodynamic pressure in shaft**

*3.3 Shaft displacements*

In order to investigate the effect of inner water on shaft displacement, displacement history curves of the shaft bottom and top in three cases are compared in Fig. 5. It can be seen from Fig. 5a, the shaft bottom displacement curves in three cases are nearly the same, so are the shaft top displacement curves (Fig. 5b). It can be seen that effect of inner water on the shaft's

dynamic displacement can be neglected. At the same time, it also indicates that the shaft's dynamic displacement is mainly determined by the surrounding soils.

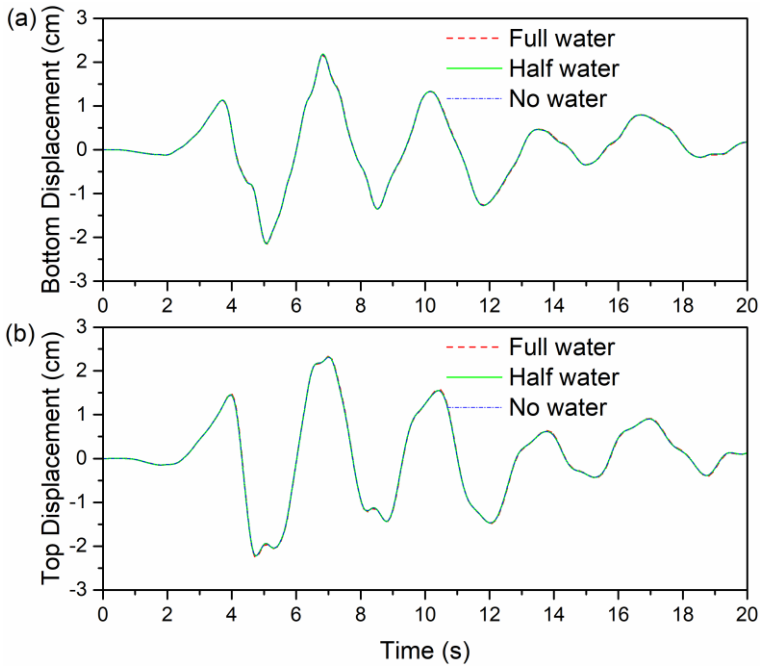



**Figure 5. Shaft displacement time-history responses of the shaft: (a) shaft bottom; (b) shaft top**

*3.4 Shaft inner forces*

Fig. 6 shows the envelope distribution pattern of the circumferential bending moment, $M_{cs}$, and the circumferential axial force, $T_{cs}$, along the shaft depth in the cross-section of the shaft hoop. Calculation results of the three cases show that as the shaft inner water level increases, the circumferential inner forces of the shaft decrease also increases. This is because as the shaft water level increases, the hydrodynamic pressure increases, and the seismic load on the shaft lining increases finally.

Fig. 7 shows the envelope distribution pattern of the vertical bending moment, $M_{as}$, and the vertical axial force, $T_{as}$, along the shaft depth. Calculation results of the three cases show that as the shaft inner water level increases, the vertical inner forces along the depth increase. This is also because as the shaft water level increases, the hydrodynamic pressure increases, and the seismic load on the shaft lining increases finally.

From the perspective of the inner forces of the structure, the magnitude of the inner forces are nearly the same and there is no obvious effects on the shaft inner forces. As the shaft inner water level increases, the inner forces of the shaft lining increase with the shaft inner water level.

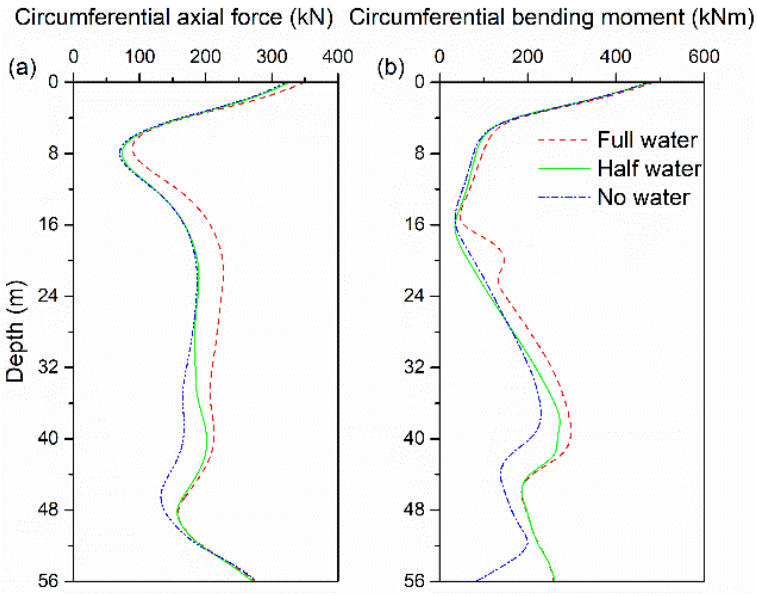



**Figure 6. Envelopes of maximum internal forces in circumferential direction: (a) axial force; (b) bending moment**

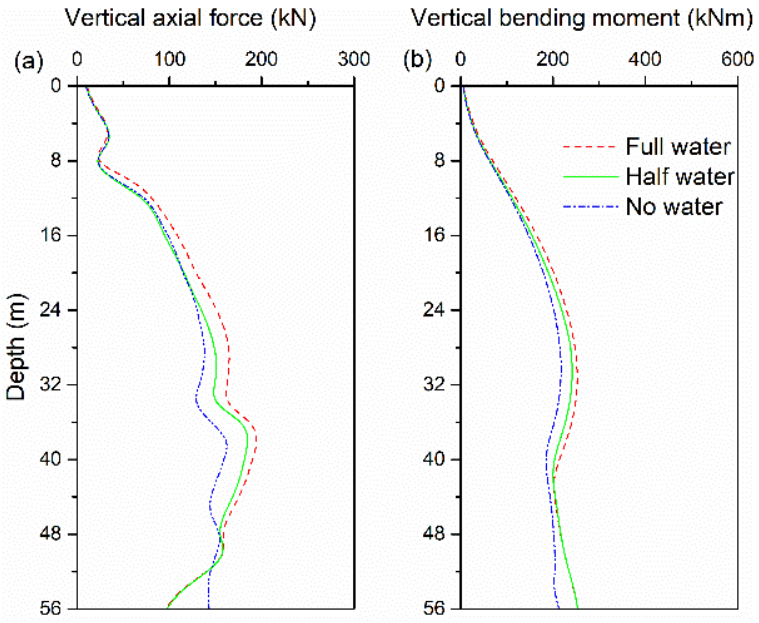



**Figure 7. Envelopes of maximum internal forces in vertical direction: (a) axial force; (b) bending moment**

## 4. Conclusions

The hydrodynamic effects of inner water on the shaft seismic responses are investigated in this paper through 3D CAS dynamic time history analysis with different inner water levels. The main conclusions are listed as follows:

- The inner water sloshing displacement is small, which is not the key factor in the seismic design of the underground shaft.

- The shaft dynamic displacement is not effected obviously by the different inner water levels. This also indicates that the shaft dynamic response is controlled by surrounding soils.

- The shaft dynamic inner force distribution pattern is mainly determined by surrounding soils, meanwhile the magnitude of the hydrodynamic pressure is nearly the same. Its effect can be neglected compared with the effect of surrounding soils.

## Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 51778464), the Fundamental Research Funds for the Central Universities and "Shuguang Program" supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission. All supports are gratefully acknowledged.

## References

[1] Housner, G. W. (1963) The dynamic behavior of water tanks, *Bulletin of the seismological society of America*, 53(2), 381-387.
[2]Veletsos, A. S.and Shivakumar, P. (1997) Tanks containing liquids or solids. *Computer Analysis and Design of Earthquake Resistant Structures, A Handbook*, 725-774.
[3] Westergaard, H. M. (1933) Water pressures on dams during earthquakes, *Trans. ASCE* **95**, 418-433.
[4] Virella, J. C., Suarez, L. E. and Godoy, L. A. (2005) Effect of pre-stress states on the impulsive modes of vibration of cylindrical tank-liquid systems under horizontal motions, *Modal Analysis* **11**(9), 1195-1220.
[5] Virella, J. C., Godoy, L. A. and Suárez, L. E. (2006) Dynamic buckling of anchored steel tanks subjected to horizontal earthquake excitation, *Journal of Constructional Steel Research* **62**(6), 521-531.
[6] Buratti, N. and Tavano, M. (2014) Dynamic buckling and seismic fragility of anchored steel tanks by the added mass method, *Earthquake Engineering & Structural Dynamics* **43**(1), 1-21.
[7] Livaoglu, R. (2008) Investigation of seismic behavior of fluid–rectangular tank–soil/foundation systems in frequency domain, *Soil Dynamics and Earthquake Engineering* **28**(2), 132-146.
[8] Khanmohammadi, M., Rad, P. L. and Ghalandarzadeh, A. (2017) Experimental study on dynamic behavior of buried concrete rectangular liquid storage tanks using shaking table, *Bulletin of Earthquake Engineering* **15**(9), 3747-3776.
[9] Livaoglu, R., Cakir, T., Dogangun, A. and Aytekin, M. (2011) Effects of backfill on seismic behavior of rectangular tanks, *Ocean Engineering* **38**(10), 1161-1173.
[10] Cakir, T. and Livaoglu, R. (2012) Fast practical analytical model for analysis of backfill-rectangular tank-fluid interaction systems, *Soil Dynamics and Earthquake Engineering* **37**, 24-37.
[11] Kianoush, M. R. and Ghaemmaghami, A. R. (2011) The effect of earthquake frequency content on the seismic behavior of concrete rectangular liquid tanks using the finite element method incorporating soil–structure interaction, *Engineering Structures* **33**(7), 2186-2200.
[12] Ibrahim, R. A. (2005). Liquid sloshing dynamics: theory and applications, Cambridge University Press.
[13] Liu, M. B. and Liu, G. R. (2010) Smoothed particle hydrodynamics (SPH): an overview and recent developments, *Archives of computational methods in engineering* **17**(1), 25-76.
[14] Shao, J. R., Li, H. Q., Liu, G. R. and Liu, M. B. (2012) An improved SPH method for modeling liquid sloshing dynamics, *Computers & Structures* **100**, 18-26.
[15] Mittal, V., Chakraborty, T. and Matsagar, V. (2014) Dynamic analysis of liquid storage tank under blast using coupled Euler-Lagrange formulation, *Thin-Walled Structures* **84**, 91-111.
[16] Tippmann, J., Prasad. S. and Shah, P. (2009) 2-D tank sloshing using the coupled Eulerian-Lagrange (CEL) capability of ABAQUS®/Explicit. *2009 SIMULIA Customer Conference*, Dassault Systemes Simulia Corp.
[17] Rawat, A., Mittal, V., Chakraborty, T. and Matsagar, V. (2019) Earthquake induced sloshing and hydrodynamic pressures in rigid liquid storage tanks analyzed by coupled acoustic-structural and euler-lagrange methods, *Thin-Walled Structures* **134**, 333-346.
[18] Virella, J. C., Prato, C. A. and Godoy, L. A. (2008) Linear and nonlinear 2d finite element analysis of sloshing modes and pressures in rectangular tanks subject to horizontal harmonic motions, *Journal of Sound and Vibration* **312**(3), 442-460.
[19] ABAQUS, (2010) ABAQUS: theory and analysis user's manual, version 6.10. Dassault Systèmes SIMULIA, Providence, RI, USA.

# Comparative studies on regularization penalties for structural damage detection

**Ziwei Luo and *†Ling Yu**

MOE Key Laboratory of Disaster Forecast and Control in Engineering, School of Mechanics and Construction Engineering, Jinan University, Guangzhou, China.

*Presenting and †Corresponding author: lyu1997@163.com

## Abstract

Structural damage detection (SDD) is an essential link to structural safety in the field of structural health monitoring (SHM). With the development of SHM technologies, higher requirements are necessary for the safety of structures. Therefore, many SDD methods have been emerging in the last decades. Due to the ill-posedness of SDD problems, regularization techniques are introduced to locate structural damages and quantify severities of damages with a higher accuracy. However, the influence of regularization penalties on SDD results is lack of consideration to date. In this study, based on the model updating technique, an intensive study is proposed to investigate the effect of different regularization penalties in structural damage patterns. First-order sensitivity analysis is chosen to establish the identified equation. Considering structural damage patterns, three regularization penalties, i.e. the $l_2$ norm, $l_1$ norm and $l_{1/2}$ norm penalties are adopted for this comparative study. The SDD problem is converted into a kind of optimization problems by defining an objective function with different regularization penalties, and they are finally solved by the particle swarm optimization (PSO) algorithm. The spring-mass model and cantilever beam are taken as examples in numerical simulations for comparative studies. The illustrated results show that there are significant effects on the SDD results using different regularization penalties. The $l_2$ norm penalty is more suitable for structural model updating. The $l_1$ norm penalty has positive effect on identifying structural damages for contiguous zones, and the $l_{1/2}$ norm penalty has higher accuracy for noncontiguous damage identification than the $l_1$ norm penalty, which provides a potential tool for SDD onsite in the SHM field.

**Keywords:** Structural health monitoring (SHM), structural damage detection (SDD), regularization, norm penalty, structural damage patterns, sensitivity analysis, particle swarm optimization (PSO).

## Introduction

In recent years, more and more scholars have been devoting to ensure the safety of in-service structures [1]-[6]. As an effective way for monitoring long-term properties and states in the service life of structures, structural health monitoring (SHM) uses measured structural responses to estimate the change in structural states. Structural damage detection (SDD) is a vital step in the SHM field and is applied to locate and quantify damages of structures.

Modal-based method [1] is one kind of SDD methods to detect damages by modal parameters. The common modal parameters for SDD are frequencies and mode shapes. Effectively utilizing both frequencies and mode shapes for SDD has been proposed in many methods, and sensitivity analysis is a common technique to establish the identified equations in these methods. Cawley and Adams [2] proposed a SDD method using sensitivity analysis and frequencies. However, frequencies are global structural properties, and they are not sensitive to local damages. To overcome this shortcoming, mode shapes, local structural properties and so on, were introduced to the SDD method by Chen et al [3]. Li et al [4] improved SDD methods with the advantages of frequencies and mode shapes to improve accuracy of SDD results. However, accurate results are not obtained due to the ill-posedness of SDD problems.

Regularization techniques are common approaches to deal with the ill-posed problem. Many scholars have introduced regularization techniques into SDD. Tikhonov regularization method (referred to as $l_2$ norm regularization method) is a classical regularization technique, and has been used in some SDD studies. For example, Li and Law [5] presented an adaptive Tikhonov regularization method for solving the nonlinear model updating problem. SDD results obtained by the $l_2$ norm regularization method do not match the sparsity property of actual damages. Lasso regularization method (referred to as $l_1$ norm regularization method) based on modal updating techniques with natural frequencies and mode shapes was proposed by Hou et al. [6]. Moreover, the $l_q$ ($0 < q < 1$) norm regularizations have been verified to obtain sparser solution than the $l_1$ norm regularization. The experimental study [7] has been conducted to show that the $l_{1/2}$ norm regularization can be taken as a representation among the $l_q$ norm regularizations, and the $l_{1/2}$ norm regularization was introduced to SDD [8].

To compare performances of regularization techniques, Zhang and Xu [9] gave comparative studies between the $l_2$ norm and the $l_1$ norm regularization on damage detection. This study showed that the $l_1$ norm regularization exhibited superiority over the $l_2$ norm regularization for SDD. Sparsity of solutions and appropriate scenarios of regularization methods is diverse when different properties of norm penalties are used. Single damage pattern is unable to reflect effects of norm penalties, so more damage patterns will be considered in this paper.

On the other hand, swarm intelligence algorithms (SI) are evolutionary algorithms to promote the progress and development of scientific research. Due to their advantages of solving optimization problems, SI-based algorithms have been widely used in SDD, such as firefly algorithm [10], genetic algorithm [11], artificial bee colony algorithm [12] and so on. As one of SI-based algorithms, particle swarm optimization (PSO) was proposed by Kennedy and Eberhart [13] and widely applied in many fields due to its simplicity and easy implementation. PSO has been introduced into SDD for obtaining the optimal solutions [14]-[15].

In this study, comparative studies on SDD with different regularization methods are conducted. Sensitivity analysis is adopted to establish the relationship between structural damages and modal parameters. The $l_2$ norm penalty, the $l_1$ norm penalty and the $l_{1/2}$ norm penalty are selected to define objection functions respectively. The PSO algorithm is utilized

to solve these objection functions. To compare the appropriate scenarios of regularization methods, a spring-mass model and a cantilever beam are simulated.

## Theoretical background

*Sensitivity Analysis*

Structural frequencies and mode shapes are affected by the change in structural physical parameters. The sensitivity-based dynamic analysis method defines the first-order sensitivity analysis equation based on the correlation between structural modal parameters and physical parameters. It can detect damage locations and quantify damage severities based on finite element model and the sensitivity equation.

In this study, it is assumed that structural damages only cause the change in stiffness, and it is described by the change in elastic modulus. Thus, the global stiffness matrix of a *n*-element structure can be expressed as:

$$\mathbf{K} = \sum_{j=1}^{n} \alpha_j \mathbf{K}_j \left(0 \le \alpha_j \le 1\right) \tag{1}$$

where $\mathbf{K}_j$, $\alpha_j$ represent the $j$th element stiffness matrix and damage reduction factor, respectively.

The first-order sensitivity analysis equation based on derivatives of frequencies and mode shapes can be expressed as follows:

$$\begin{bmatrix} \frac{\partial \upsilon_1}{\partial \alpha_1} & \cdots & \frac{\partial \upsilon_1}{\partial \alpha_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \upsilon_m}{\partial \alpha_1} & \cdots & \frac{\partial \upsilon_m}{\partial \alpha_n} \\ \frac{\partial \boldsymbol{\varphi}_1}{\partial \alpha_1} & \cdots & \frac{\partial \boldsymbol{\varphi}_1}{\partial \alpha_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \boldsymbol{\varphi}_l}{\partial \alpha_1} & \cdots & \frac{\partial \boldsymbol{\varphi}_l}{\partial \alpha_n} \end{bmatrix} \boldsymbol{\Delta\alpha} = \mathbf{S}\boldsymbol{\Delta\alpha} \approx \boldsymbol{\Delta\mathbf{f}} = \begin{Bmatrix} \boldsymbol{\Delta\mathbf{f}}_\upsilon \\ \boldsymbol{\Delta\mathbf{f}}_\varphi \end{Bmatrix} \tag{2}$$

where $\boldsymbol{\Delta\mathbf{f}}_\upsilon = \{\Delta\upsilon_1, \Delta\upsilon_2, \ldots, \Delta\upsilon_m\}^T$ and $\boldsymbol{\Delta\mathbf{f}}_\varphi = \{\boldsymbol{\Delta\varphi}_1, \boldsymbol{\Delta\varphi}_2, \ldots, \boldsymbol{\Delta\varphi}_l\}^T$ are the difference of frequencies and mode shapes, respectively. $\Delta\upsilon_p = \upsilon_{up} - \upsilon_{dp}$ ($p$ = 1, 2, …, $m$), $\upsilon_{up}$ and $\upsilon_{dp}$ are the $p$th frequencies in undamaged and damaged structures, respectively. $m$ is the order of frequencies. $\boldsymbol{\Delta\varphi}_q = \boldsymbol{\varphi}_{uq} - \boldsymbol{\varphi}_{dq}$ ($q$ = 1, 2, …, $l$), $\boldsymbol{\varphi}_{uq}$ and $\boldsymbol{\varphi}_{dq}$ are the $p$th mode shapes in undamaged and damaged structures, respectively. $l$ is the order of mode shapes. $\boldsymbol{\Delta\alpha} = \{\Delta\alpha_1, \Delta\alpha_2, \ldots, \Delta\alpha_n\}^T$ is the change in damage reduction factors. $\mathbf{S}$ is the first-order sensitivity matrix.

The least square method can be used to solve Eq. (2):

$$J_{\mathrm{LS}}(\boldsymbol{\Delta\alpha}) = \arg\min_{\boldsymbol{\Delta\alpha}} \frac{1}{2}\left\|\mathbf{S}\boldsymbol{\Delta\alpha} - \boldsymbol{\Delta\mathbf{f}}\right\|_2^2 \tag{3}$$

However, it cannot obtain a stable result for inverse problem due to the ill-conditioned matrix S and noise [16]. That is to say, due to the influence of noise, the solution of Eq. (2) is ill-posed. In this study, the influence of noise can be reduced and a stable solution can be obtained by combining sensitivity analysis with regularization methods.

*Tikhonov Regularization*

The principle of regularization methods is to replace the original ill-posed problem with an approximate well-posed problem whose solution equals original solution approximately. So Li and Law [5] introduced the $l_2$ norm regularization method into SDD for improving the identification precision. The $l_2$ norm regularization method [17] is a popular regularization method. It adds a quadratic penalty to Eq. (3):

$$J_2(\boldsymbol{\Delta\alpha}) = \arg\min_{\boldsymbol{\Delta\alpha}\in\mathbf{R}^P} \left\{\frac{1}{2}\left\|\mathbf{S}\boldsymbol{\Delta\alpha} - \boldsymbol{\Delta\mathbf{f}}\right\|_2^2 + \lambda\left\|\boldsymbol{\Delta\alpha}\right\|_2^2\right\} \tag{4}$$

where real solution and noise can be balanced by the regularization parameter ($\lambda > 0$). Regularization term controls the norm of solution. $P$ is the dimension of $\boldsymbol{\Delta\alpha}$.

*Sparse Regularization*

The $l_0$ norm regularization method is an original definition of sparse regularization:

$$\arg\min_{\mathbf{X}} \left\{\frac{1}{2}\left\|\mathbf{X} - \mathbf{B}\right\|_2^2 + \lambda\left\|\mathbf{X}\right\|_0\right\} \tag{5}$$

where, $\left\|\mathbf{X}\right\|_0$ represents $l_0$ norm of vector $\mathbf{X}$.

The $l_0$ norm regularization method recovers sparse vector precisely, but it is a NP-hard problem. The $l_1$ norm regularization method, which was first proposed by Tibshirani [18] in 1996, can be used to approximately replace the $l_0$ norm regularization. The $l_1$ norm regularization method obtains sparse coefficient vector because coefficients with small absolute value will set to be zero:

$$\arg\min_{\mathbf{X}} \left\{\frac{1}{2}\left\|\mathbf{X} - \mathbf{B}\right\|_2^2 + \lambda\left\|\mathbf{X}\right\|_1\right\} \tag{6}$$

where, $\left\|\mathbf{X}\right\|_1$ represents $l_1$ norm of vector $\mathbf{X}$.

Eq. (6) can be solved by the soft-thresholding [19]:

$$\mathbf{X} = soft\left(\mathbf{B}, \lambda\right) = \begin{cases} \mathbf{B} + \lambda, & \mathbf{B} < -\lambda \\ 0, & \left|\mathbf{B}\right| \le \lambda \\ \mathbf{B} - \lambda, & \mathbf{B} > \lambda \end{cases} \tag{7}$$

To obtain sparser solution, the $l_q$ norm regularizations were proposed after the $l_1$ norm regularization. Some studies focused on selecting the best value of $q$. They showed that the $l_q$ norm regularizations gain sparser solutions as the $q$-value decreases. There are no significant different performances when $0 < q \leq 1/2$, so the $l_{1/2}$ norm regularization is a representative of the $l_q$ norm regularizations [7]. The $l_{1/2}$ norm regularization method is more sparsity and viability than the $l_1$ norm regularization method, the $l_{1/2}$ norm regularization is defined as:

$$\arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \left\| \mathbf{X} - \mathbf{B} \right\|_2^2 + \lambda \left\| \mathbf{X} \right\|_{1/2}^{1/2} \right\} \tag{8}$$

where, $\left\| \mathbf{X} \right\|_{1/2}$ represents $l_{1/2}$ norm of vector $\mathbf{X}$.

Compared with Eq. (6), a generalized shrinkage-thresholding operator [20] is given for the $l_{1/2}$ norm regularization:

$$\mathbf{X} = gsoft\left(\mathbf{B}, \lambda\right) = \begin{cases} 0, & \left|\mathbf{B}\right| \leq \tau^{GST}\left(\lambda\right) \\ \operatorname{sgn}\left(\mathbf{B}\right) s^{GST}\left(\left|\mathbf{B}\right|; \lambda\right) & \left|\mathbf{B}\right| > \tau^{GST}\left(\lambda\right) \end{cases} \tag{9}$$

where the thresholding $\tau^{GST}$ is given by:

$$\tau^{GST}\left(\lambda\right)\Big|_{q=\frac{1}{2}} = \left[2\lambda\left(1-q\right)\right]^{\frac{1}{2-q}} + \lambda q\left[2\lambda\left(1-q\right)\right]^{\frac{q-1}{2-q}} = \frac{3}{2}\lambda^{\frac{2}{3}} \tag{10}$$

and the $s^{GST}(|\mathbf{B}|;\lambda)$ can be calculated by the following formula:

$$s^{GST}\left(\left|\mathbf{B}\right|; \lambda\right) - \mathbf{B} + \frac{\lambda}{2}\left[ s^{GST}\left(\left|\mathbf{B}\right|; \lambda\right)\right]^{-\frac{1}{2}} = 0 \tag{11}$$

For the SDD problem, structural damages occur in few locations, so the coefficient vector $\boldsymbol{\Delta\alpha}$ is a sparse vector. By respectively adding the $l_1$ norm and the $l_{1/2}$ norm regularization into Eq. (3), the following equations are obtained:

$$J_1(\boldsymbol{\Delta\alpha}) = \arg\min_{\boldsymbol{\Delta\alpha}\in\mathbf{R}^P} \left\{ \frac{1}{2} \left\| \mathbf{S}\boldsymbol{\Delta\alpha} - \boldsymbol{\Delta}\mathbf{f} \right\|_2^2 + \lambda \left\| \boldsymbol{\Delta\alpha} \right\|_1 \right\} \tag{12}$$

$$J_{1/2}(\boldsymbol{\Delta\alpha}) = \arg\min_{\boldsymbol{\Delta\alpha}\in\mathbf{R}^P} \left\{ \frac{1}{2} \left\| \mathbf{S}\boldsymbol{\Delta\alpha} - \boldsymbol{\Delta}\mathbf{f} \right\|_2^2 + \lambda \left\| \boldsymbol{\Delta\alpha} \right\|_{1/2}^{1/2} \right\} \tag{13}$$

Compared Eq. (12) with Eq. (6), it can be found that the objective functions in these two equations are different, so Eq. (12) cannot be directly solved by Eq. (6). Similarly, Eq. (13) cannot be directly solved by Eq. (7). To solve Eqs. (12) and (13), in this study, the PSO algorithm is introduced, by combining PSO algorithm with Eq. (6) and Eq. (7) respectively, the SDD results can be obtained by using Eqs. (12) and (13).

*Particle Swarm Optimization*

PSO is a heuristic algorithm, and it is inspired by group behavior of birds. With information sharing system of the bird flock, PSO simulates their foraging process in space. The solution process becomes orderly from unorderly, so an optimal solution can be obtained. PSO is simple and low computational cost compared to other novel heuristic algorithms. A solution of the optimization problem is a particle of search space. Velocity of each particle decides its direction and moving step. The best positions of individual particles $\mathbf{p}_{\text{best}}$ and the previous best solution of the entire warm $\mathbf{g}_{\text{best}}$ are used to update particle positions. That is to say, particles update their positions and velocities according to the following equations:

$$\mathbf{x}_j\left(t+1\right)=\mathbf{x}_j\left(t\right)+\mathbf{v}_j\left(t+1\right) \tag{14}$$

$$\mathbf{v}_j\left(t+1\right)=\mathbf{v}_j\left(t\right)+c_1\cdot r_1\cdot\left(\mathbf{p}_{best_j}\left(t\right)-\mathbf{x}_j\left(t\right)\right)+c_2\cdot r_2\cdot\left(\mathbf{g}_{best_j}\left(t\right)-\mathbf{x}_j\left(t\right)\right) \tag{15}$$

where $\mathbf{x}_j$ and $\mathbf{v}_j$ are the position and velocity of the $j$th particle, respectively. $t$ is the iterative number. $r_1$ and $r_2$ are uniformly distributed random numbers in the range of [0,1], and the cognitive coefficient $c_1$ and the social coefficient $c_2$ are equal to 2.

In this study, Eqs. (4), (12) and (13) are optimization problems, so they can be solved by PSO. To obtain accurate solutions and reduce computation time, Eqs. (7) and (9) are added into each particle to solve Eqs. (12) and (13), respectively.

**Numerical Simulations**

To compare the appropriate scenarios of different regularization methods, a 2-DOF spring-mass model and a cantilever beam with the dimension of 0.7 m×0.05 m×0.01 m are adopted to simulate damages.

*Spring-mass model*

A 2-DOF spring-mass model is shown in Fig. 1. The stiffness and the mass of each DOF are 150 kN/m and 100 kg, respectively. The first two frequencies of the structure are 3.8096Hz and 9.9736Hz, respectively.

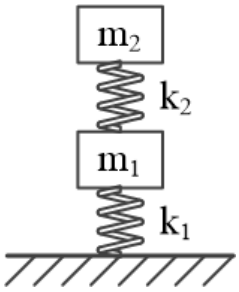



**Figure 1. Spring-mass model**

Gaussian white noise is the ideal model for analyzing additive noise in channels, so it is used to discuss the effect of noise on SDD results. In this study, measurement noise is considered

to be related to the change in frequencies and mode shapes. Frequencies and mode shapes with noise are defined as follows:

$$\upsilon_i = \upsilon_i^a + \varepsilon_\upsilon R_i \left(\frac{\left\|\boldsymbol{\upsilon}_s - \boldsymbol{\upsilon}_e\right\|_2}{m}\right) \tag{16}$$

$$\varphi_{ij} = \varphi_{ij}^a + \varepsilon_\varphi R_{ij} \left(\frac{\left\|\boldsymbol{\varphi}_s - \boldsymbol{\varphi}_e\right\|_2}{lN}\right) \tag{17}$$

where, $\upsilon_i^a$ and $\upsilon_i$ ($i$ = 1, 2, …, $m$) are the $i$th frequency and the one with noise, respectively. $\varphi_{ij}^a$ and $\varphi_{ij}$ ($i$ = 1, 2, …, $l$; $j$ = 1, 2, …, $N$) are the $j$th element of the $i$th mode shape and the one with noise, respectively. $N$ is the length of each mode shape vector. $\boldsymbol{\upsilon}_s$ and $\boldsymbol{\upsilon}_e$ are vectors of frequencies in undamaged and damaged structures, respectively. $\boldsymbol{\varphi}_s$ and $\boldsymbol{\varphi}_e$ are rearranged vectors of mode shapes in undamaged and damaged structures, respectively. $\varepsilon_\upsilon$ and $\varepsilon_\varphi$ are noise levels of frequencies and mode shapes, respectively. $R_i$ and $R_{ij}$ are random numbers subjected to the standard normal distribution.

As shown in Table 1, four damage scenarios are considered to compare properties of the $l_2$ norm regularization, the $l_1$ norm regularization and the $l_{1/2}$ norm regularization. Sketch maps of cost functions and norm penalties are offered to illustrate the appropriate scenarios of different regularization methods.

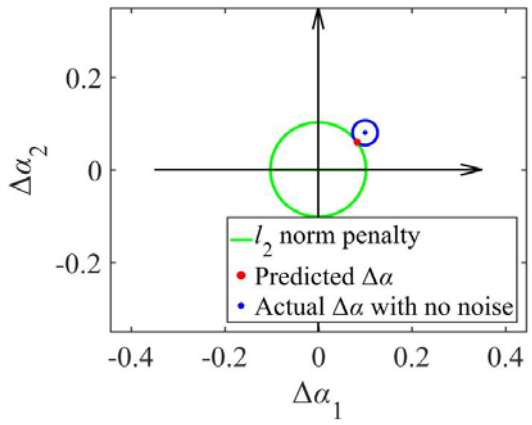
The SDD results are shown in Figs. 2-7. Where, (a) represents a sketch map of cost functions and norm penalties under different scenarios. The blue lines are constant value lines of cost functions and their radii are same in Figs. 2-3. The green lines are constant value lines of norm penalties and their junctions of coordinating axis are same in Figs. 4-6. The black points are possible values of actual $\boldsymbol{\Delta\alpha}$ with noise and the black dotted lines are possible constant value lines of cost functions.

**Table 1. Damage scenarios for spring-mass model**

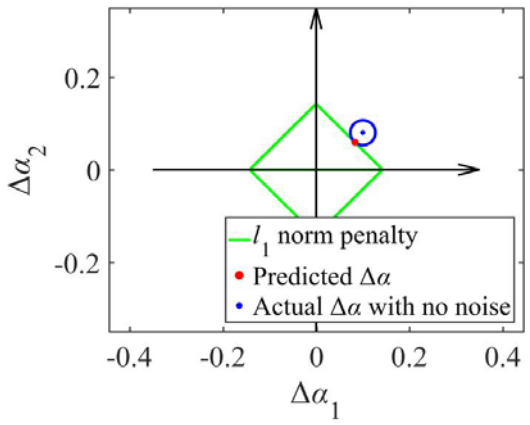
| Scenario no. | Damage degrees @ damaged elements | Noise levels | Values of $m$ and $l$ | Norm penalties | $\lambda$ | Predicted damage numbers |
|---|---|---|---|---|---|---|
| 1 | 10%@E1, 8%@E2 | 0% | $m$=$l$=2 | $l_2$ | 0.0040 | 2 |
| | | | | $l_1$ | 0.0273 | |
| | | | | $l_{1/2}$ | 0.1000 | |
| 2 | 10%@E1, 2%@E2 | 0% | $m$=$l$=2 | $l_{1/2}$ | 0.4300 | 1 |
| 3 | 10%@E2 | 1% | $m$=2, $l$=0 | $l_2$ | 0.0017 | 2 |
| | | | | $l_1$ | 30 | 1 |
| | | | | $l_{1/2}$ | 859.1 | 1 |
| 4 | 10%@E2 | 10% | $m$=2, $l$=0 | $l_1$ | 11.1970 | 2 |

Same SDD results can be identified when each DOF is damaged, as shown in Fig. 2. To analyze properties of norm penalties, the radii of cost functions are kept same and their

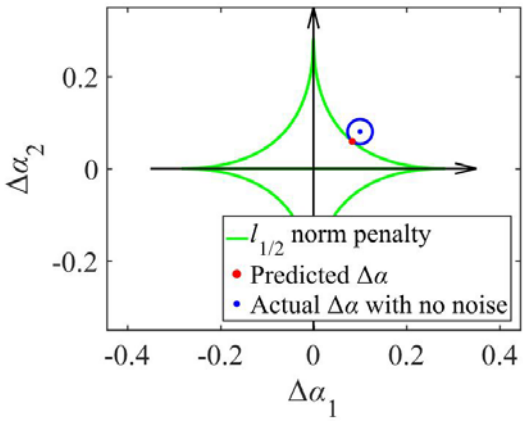
regularization parameters are different. It shows that the $l_2$ norm regularization, the $l_1$ norm regularization and the $l_{1/2}$ norm regularization can detect multiple damages in scenario 1.



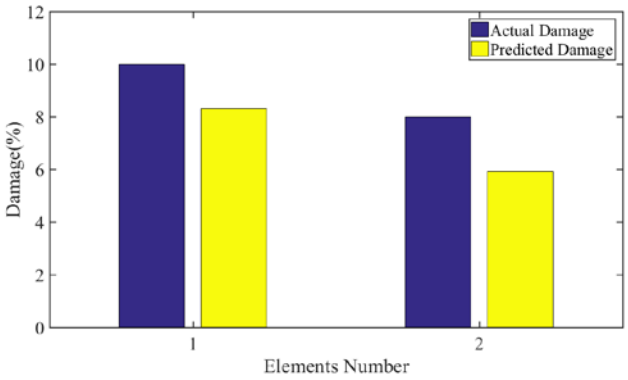

(a) Sketch map by adding $l_2$ norm penalty




(b) Sketch map by adding $l_1$ norm penalty

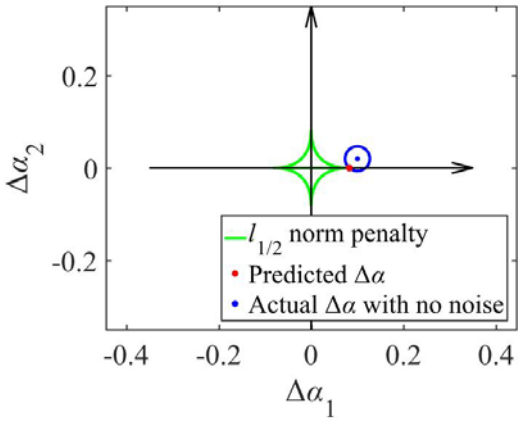



(c) Sketch map by adding $l_{1/2}$ norm penalty

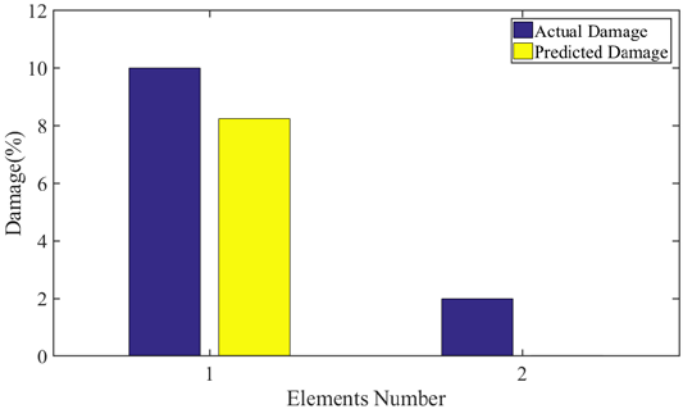



(d) Identified damage results

**Figure 2. SDD results for spring-mass model in scenario 1**




(a) Sketch map



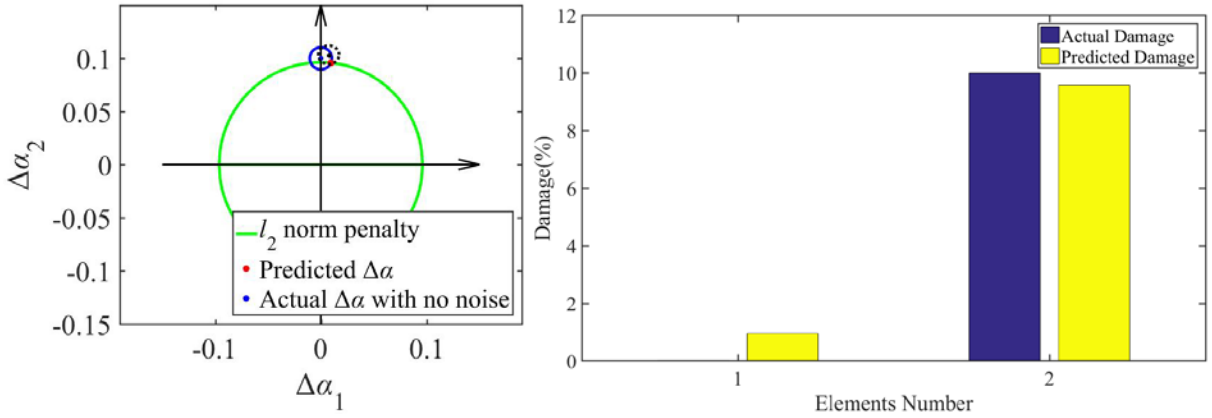

(b) Identified damage results

**Figure 3. SDD results for spring-mass model in scenario 2 by adding $l_{1/2}$ norm penalty**

By adding the $l_{1/2}$ norm penalty, the SDD result in scenario 1 are compared with that in scenario 2. Different damage degrees lead to different identified results of damage locations even if same damage locations are assumed. In scenario 2, great difference of damage degrees

between element 1 and element 2 is given, and the $l_{1/2}$ norm penalty makes the SDD results sparse. Distinct from the sketch maps in Figs. 2(c) and 3(a), the cost function and the $l_{1/2}$ norm penalty intersect at coordinate axis. This is the reason why single damage is identified in scenario 2.

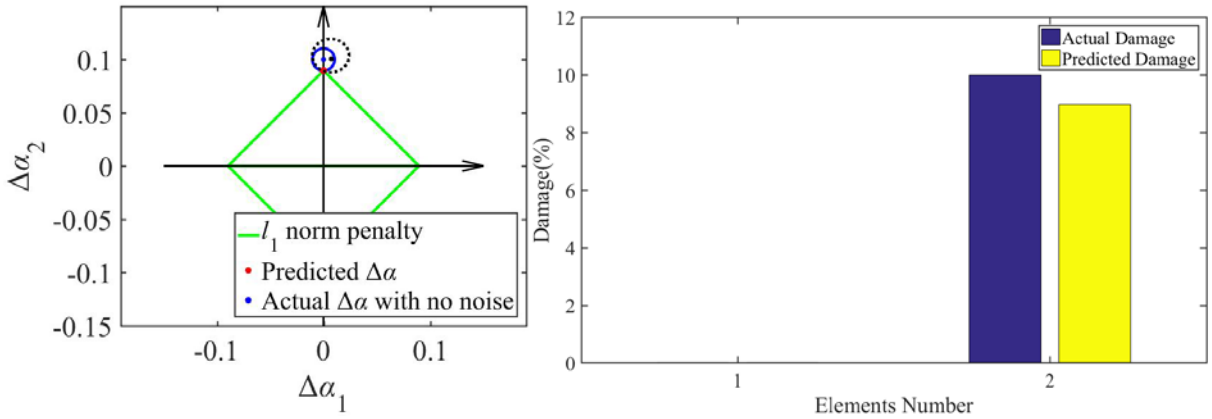
In scenario 3, multiple damages are identified when the $l_2$ norm penalty is added in the objective function, but one damage is identified when other two norm regularizations are used. Due to the influence of noise, the identified result $\boldsymbol{\Delta\alpha}$ will go away from the coordinate axis without adding penalty. Therefore, when the penalty is added, the junction between the contour lines of cost functions and the $l_2$ norm penalty is not on the axis. Under these circumstances, sparse solution cannot be obtained by adding the $l_2$ norm regularization.




(a) Sketch map (b) Identified damage results

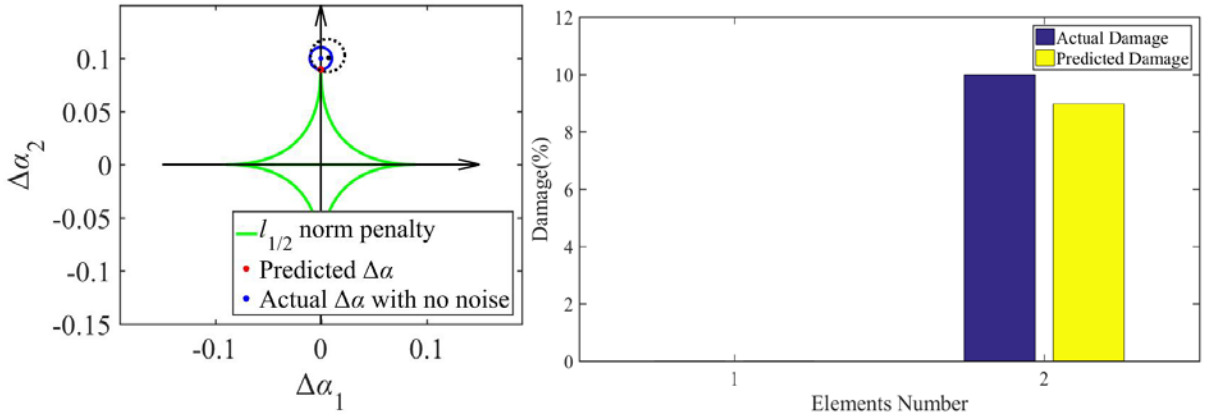
**Figure 4. SDD results for spring-mass model in scenario 3 by adding $l_2$ norm penalty**

SDD results are similar in scenario 3 when the $l_1$ norm and the $l_{1/2}$ norm penalty are respectively used. Different norm penalties give different spaces for the objective functions. Sparse solutions can be obtained due to the angles of the $l_1$ norm and $l_{1/2}$ norm penalties, which is different from the $l_2$ norm penalty. As a result, horned spaces of norm penalties are more beneficial to obtain sparse results. On the other hand, Figs. 2(a) and 4(a) shows that dense results are easy to be obtained by using the $l_2$ norm regularization.



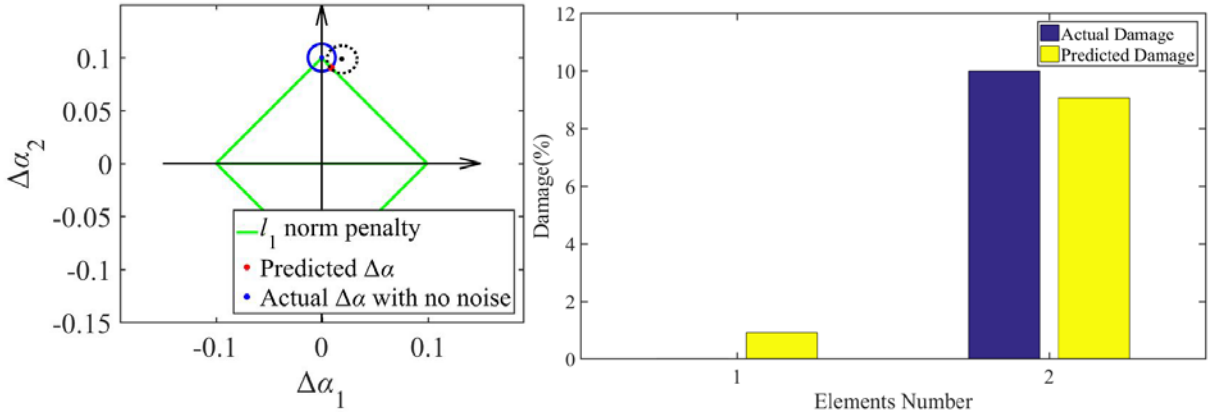

(a) Sketch map (b) Identified damage results

**Figure 5. SDD results for spring-mass model in scenario 3 by adding $l_1$ norm penalty**




(a) Sketch map (b) Identified damage results

**Figure 6. SDD results for spring-mass model in scenario 3 by adding $l_{1/2}$ norm penalty**




(a) Sketch map (b) Identified damage results

**Figure 7. SDD results for spring-mass model in scenario 4 by adding $l_1$ norm penalty**

Compared Figs. 5 and 7, it shows that the $l_1$ norm regularization does not obtain a sparse solution when the noise level increases, and great biases is produced due to the influence of noise. The junction between the contour lines of cost functions and the $l_1$ norm penalties is not on the axis. It indicates that both dense and sparse results may be identified by using the $l_1$ norm regularization.

By comparing Figs. 3, 5, 6 and 7, it shows that the curvatures of norm penalties will affect the SDD results. The sharp change in curvature increases the availability of getting sparse results. Thus, the $l_{1/2}$ norm regularization is more suitable for application in detecting sparse damages.

Some brief conclusions are summarized as follows: firstly, the $l_2$ norm regularization easily obtains a dense result, so it may have a good performance in the application of model updating. Secondly, dense or sparse solutions may be identified by the $l_1$ norm regularization, so it is more suitable to detect contiguous damages than the $l_{1/2}$ norm regularization. Thirdly, the $l_{1/2}$ norm regularization gets sparser results with a high probability than the $l_1$ norm regularization, so it is suitable for detecting noncontiguous damages.

*Cantilever beam*

To select appropriate regularization methods for different problems, a cantilever beam is simulated. As shown in Fig. 8, the cantilever beam is divided into ten elements. Numbers in circles represent element numbers. The elastic modulus is $2.01\times10^{11}$ N/m$^2$ and the density is 7800 kg/m$^3$ for each element.

As shown in Table 2, seven damage scenarios are considered to select appropriate regularization methods for different scenarios. Four structural damage patterns, i.e. model updating, contiguous damages, noncontiguous damages and composite damages are offered to distinguish properties of different regularization methods. SDD results are shown in Figs. 9-15.

**Table 2. Damage scenarios for cantilever beam**

| Scenario no. | Damage degrees @ damaged elements | Noise levels | Values of $m$ and $l$ | Norm penalties | $\lambda$ | Damage patterns |
|---|---|---|---|---|---|---|
| 1 | 3%@E1-E10 | 15% | $m$=$l$=9 | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.3 | model updating |
| 2 | 0.8%@E1, E9<br>0.9%@E3, E5<br>1.1%@E4, E8<br>1%@E2, E6, E7, E10 | 10% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.2 | |
| 3 | 5%@E4, 13%@E5<br>7%@E6 | 15% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.15 | contiguous damages |
| 4 | 6%@E7, 4%@E8<br>8%@E9 | 10% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.2 | |
| 5 | 12%@E2, 6%@E4<br>10%@E6 | 15% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.09 | noncontiguous damages |
| 6 | 5%@E3, 10%@E5<br>7%@E9 | 10% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.09 | |
| 7 | 10%@E4, 5%@E7<br>14%@E8, 6%@E9 | 15% | | $l_2$<br>$l_1$<br>$l_{1/2}$ | 0.16 | composite damages |

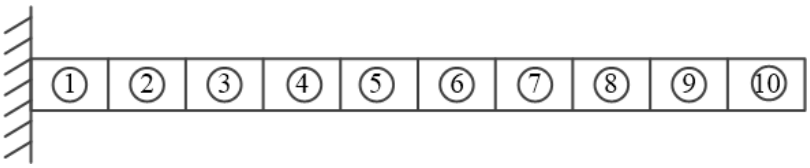
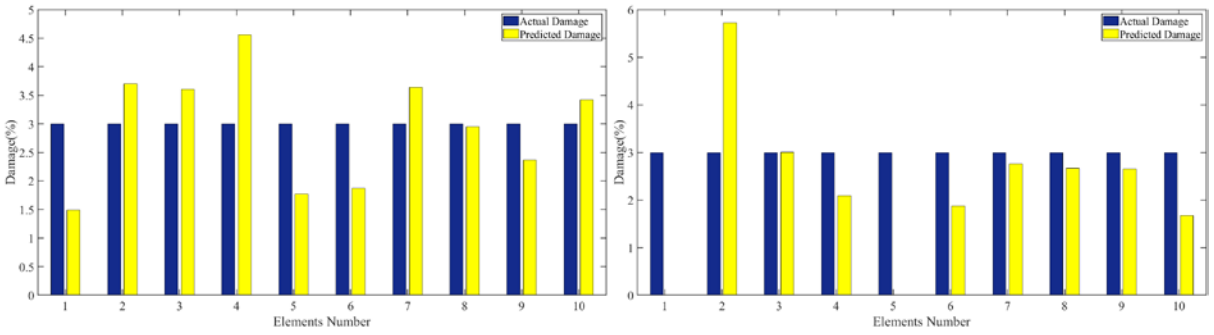
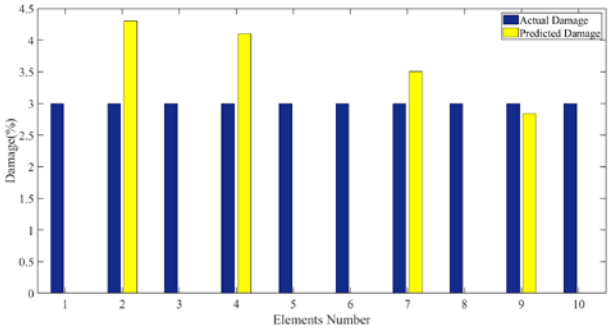



**Figure 8. 10-element cantilever beam**




(a) (b)

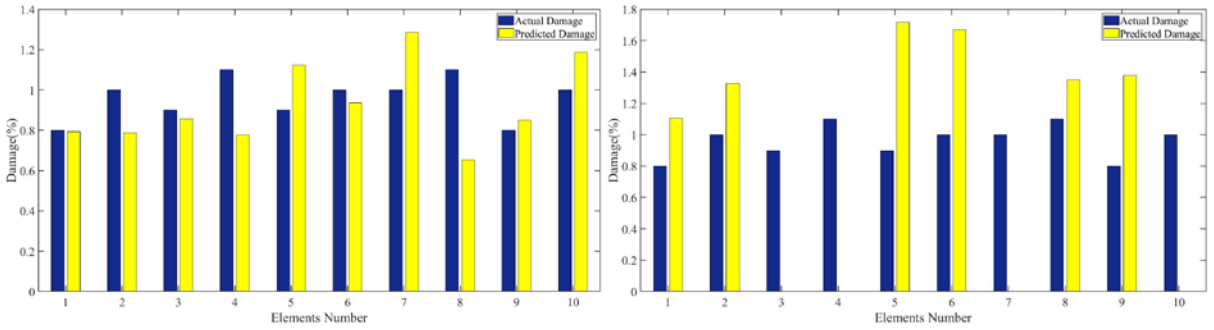



(c)

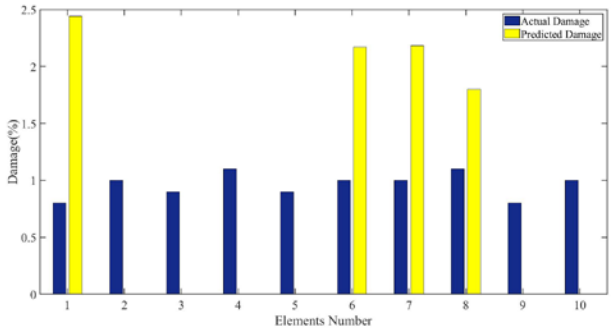
**Figure 9. SDD results for cantilever beam in scenario 1 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

In scenarios 1 and 2, dense results are obtained when the $l_2$ norm penalty is added into the objection functions, and sparse results are obtained by other regularization methods. Moreover, the solution by adding the $l_{1/2}$ norm regularization is sparser than that by adding the $l_1$ norm regularization. It can be concluded that the $l_2$ norm regularization is more suitable for application in model updating.

As mentioned above, the $l_1$ norm regularization is able to detect sparse and dense damages. It has good performances in scenarios 3 and 4. Damage locations can be effectively identified and quantify the damage degrees. The $l_{1/2}$ norm regularization can only detect two damage locations with a lower precision in these scenarios. It shows that the $l_1$ norm regularization has the ability to identify contiguous damages.

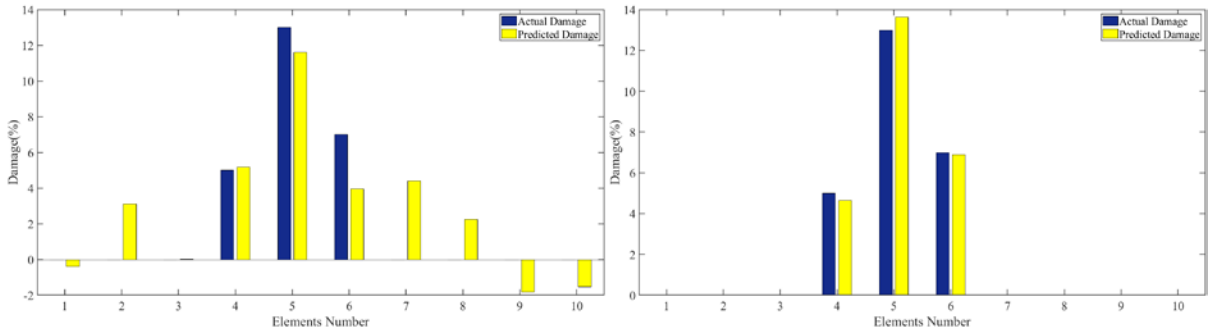



(a) (b)

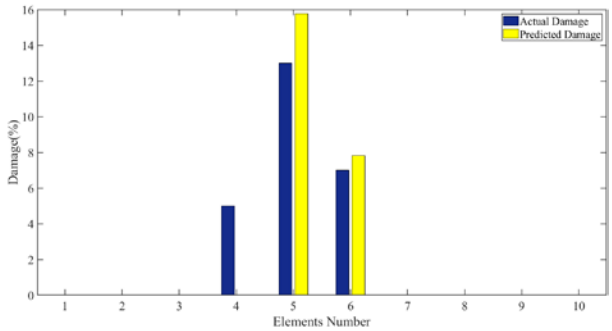



(c)

**Figure 10. SDD results for cantilever beam in scenario 2 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**




(a) (b)




(c)

**Figure 11. SDD results for cantilever beam in scenario 3 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

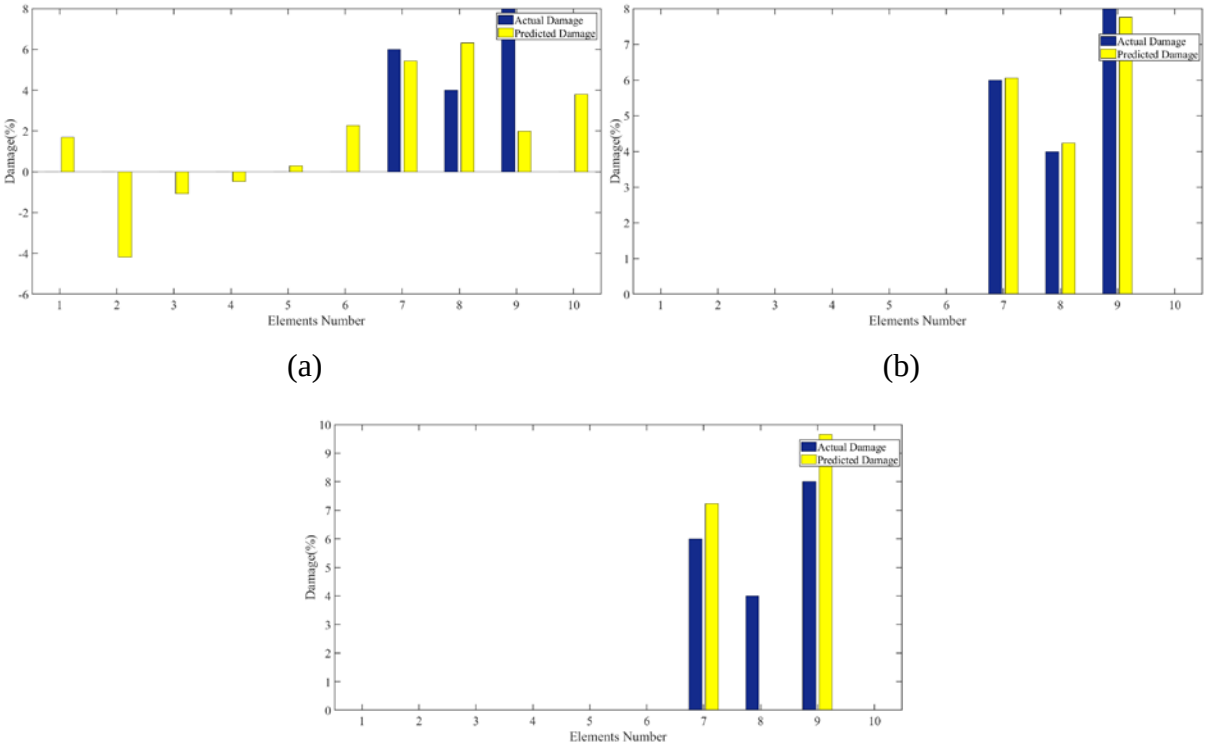



**Figure 12. SDD results for cantilever beam in scenario 4 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

Different from first two scenarios, exact damage results can be identified by adding the $l_{1/2}$ norm penalty to objection functions in scenarios 5 and 6. As indicated in the second conclusion, the $l_1$ norm regularization can identify contiguous damages. It may lead to misjudging near the actual damages. In this pattern of scenarios, the $l_{1/2}$ norm regularization can make good use of its advantage which makes the solution sparser. It is suitable for detecting noncontiguous damages.

For contiguous and noncontiguous damages, the $l_2$ norm regularization performs badly to detect them. Damage locations are misjudged in scenarios 3-7. The $l_2$ norm penalty is different from the $l_1$ norm and $l_{1/2}$ norm penalties. In the solving process, each element of the solution is not equal to zero by using the $l_2$ norm regularization. So the $l_2$ norm regularization is unreasonable for detecting sparse damages.

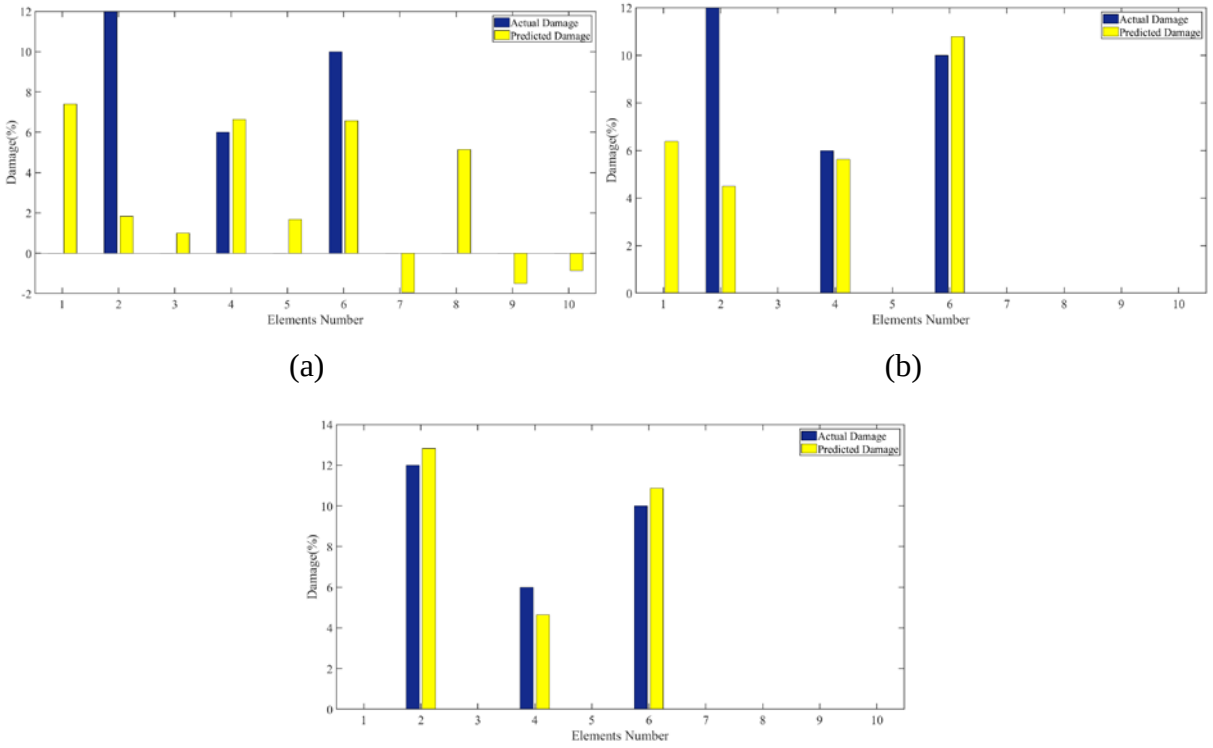



**Figure 13. SDD results for cantilever beam in scenario 5 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

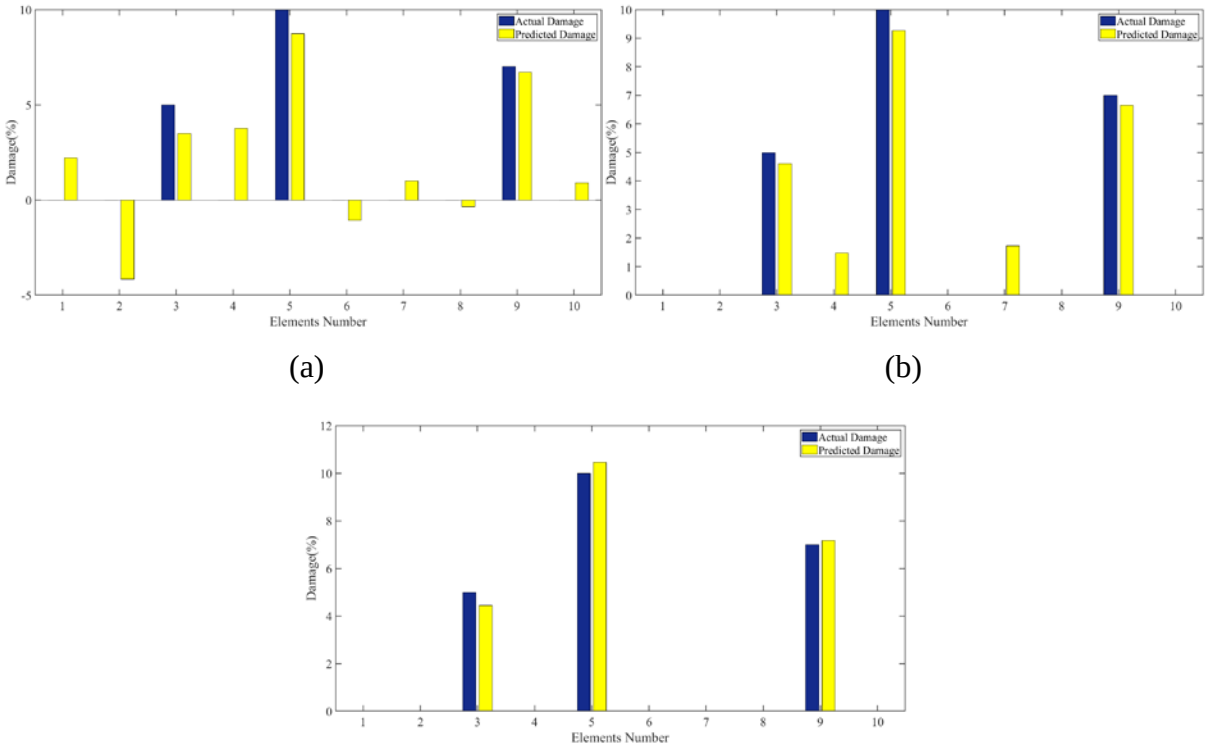



**Figure 14. SDD results for cantilever beam in scenario 6 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

Combining contiguous and noncontiguous damages, composite damages are set in scenario 7. SDD results by adding the $l_1$ norm penalty are more accurate than ones by adding the $l_{1/2}$ norm penalty, but it does not mean that the $l_1$ norm regularization is always able to obtain good results for composite damages, because only one damage location is not detected by the $l_{1/2}$ norm regularization. More future work should be done to verify their abilities in the further studies.

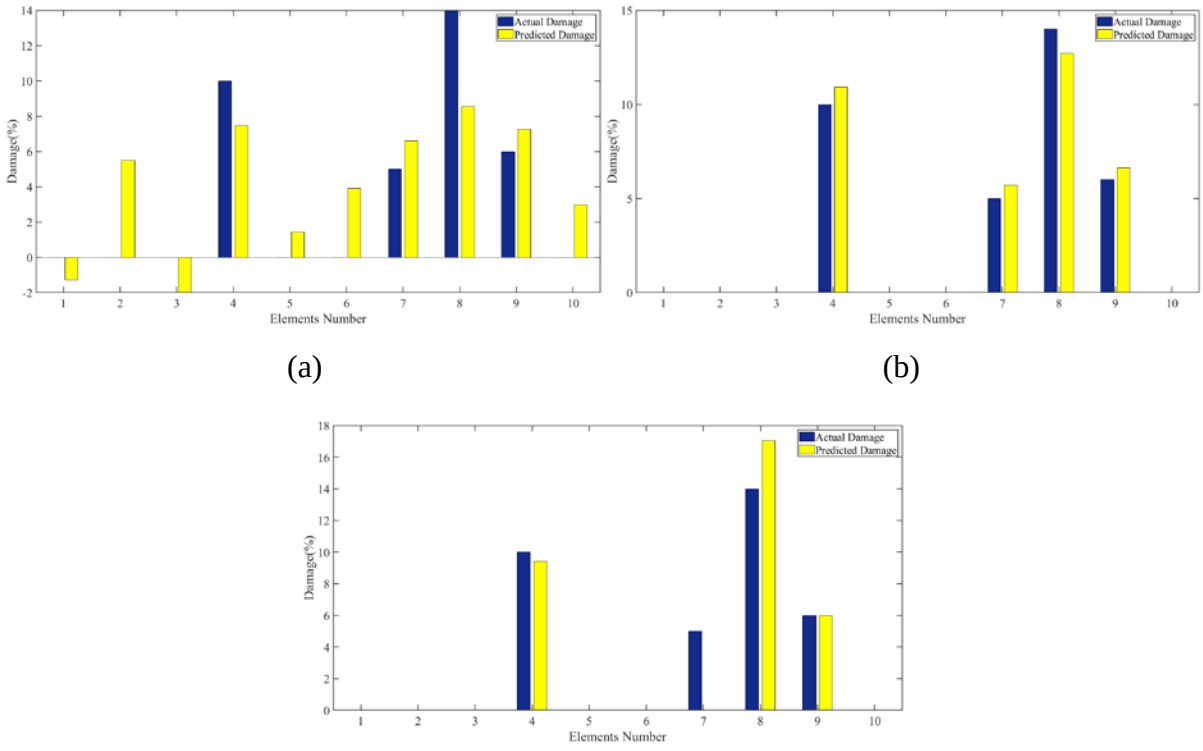



**Figure 15. SDD results for cantilever beam in scenario 7 by adding different norm penalties: (a) $l_2$ norm penalty (b) $l_1$ norm penalty (c) $l_{1/2}$ norm penalty**

## Conclusions

In this study, based on sensitivity analysis method, different norm regularization methods, i.e. $l_2$ norm, $l_1$ norm and $l_{1/2}$ norm penalties, are compared to distinguish their abilities of detecting structural damages. Objective functions are defined by adding different norm penalties, and these functions are solved by the particle swarm optimization (PSO). A 2-DOF spring-mass model and a cantilever beam are simulated to analyze properties of the $l_2$ norm regularization, the $l_1$ norm regularization and the $l_{1/2}$ norm regularization, respectively. Sketch maps of cost functions and norm penalties under different scenarios are drawn to describe their relationship with predicted solution intuitively. Due to different application scopes, diverse damage scenarios are given in two numerical simulation models. Some main conclusions can be made as follows:

1) Dense solutions can be obtained when $l_2$ norm penalty is used, so the $l_2$ norm

regularization has a good performance in the application of model updating.

2) Dense solutions or sparse solutions can be identified by the $l_1$ norm regularization which depends on the deviation of actual solution from the coordinate axis. Sparse results can be obtained when the deviation is small. Otherwise, dense results will be obtained. So the $l_1$ norm regularization has ability to detect contiguous damages.
3) A dense result will be obtained due to the influence of noise. The $l_{1/2}$ norm regularization can address this issue effectively. Comparing with the $l_1$ norm regularization, the $l_{1/2}$ norm regularization can obtain sparser solution, and it is suitable for detecting noncontiguous damages.

## Acknowledgments

The project is jointly supported by the National Natural Science Foundation of China with Grant Numbers 51678278 and 51278226.

### References

[1] Das, S., Saha, P. and Patro, S. K. (2016) Vibration-based damage detection techniques used for health monitoring of structures: a review, *Journal of Civil Structural Health Monitoring* **6**, 477-507.
[2] Cawley, P. and Adams, R. D. (1979) The location of defects in structures from measurements of natural frequencies, *Journal of Strain Analysis* **14**, 49-57.
[3] Chen, J. C. and Garba, J. A. (1988) On-orbit damage assessment for large space structures, *AIAA Journal* **26**, 1119-1126.
[4] Li, H., Huang, Y., Ou, J. and Bao, Y. (2011) Fractal dimension-based damage detection method for beams with a uniform cross-section, *Computer-Aided Civil and Infrastructure Engineering* **26**, 190-206.
[5] Li, X. Y. and Law, S. S. (2010) Adaptive Tikhonov regularization for damage detection based on nonlinear model updating, *Mechanical Systems and Signal Processing* **24**, 1646-1664.
[6] Hou, R., Xia, Y. and Zhou, X. (2017) Structural damage detection based on $l_1$ regularization using natural frequencies and mode shapes, *Structural Control and Health Monitoring* **25**, e2107.
[7] Xu, Z. B., Guo, H. L., Wang, Y. and Zhang, H. (2012) Representative of $L_{1/2}$ regularization among $L_q$ ($0 < q \leq 1$) regularizations: an experimental study based on phase diagram, *Acta Automatica Sinica* **38**, 1225-1228.
[8] Luo, Z. W. and Yu, L., PSO-based sparse regularization approach for structural damage detection, *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Yong, L., Liang, Z., Guoyong, C., Guoqing, X., Kenli, L. and Lipo, W., Eds., Guilin, China, 2017, 1034-1040.
[9] Zhang, C. D. and Xu, Y. L. (2015) Comparative studies on damage identification with Tikhonov regularization and sparse regularization, *Structural Control and Health Monitoring* **23**, 560-579.
[10] Pan, C. D. and Yu, L. (2014) Moving force identification based on firefly algorithm, *Advanced Materials Research* **919-921**, 329-333.
[11] Ullah, I., Sinha, J. K. and Pinkerton, A. (2013) Vibration-based delamination detection in a composite plate, *Mechanics of Advanced Materials and Structures* **20**, 536-551.
[12] Ding, Z. H., Huang, M. and Lu, Z. R. (2016) Structural damage detection using artificial bee colony algorithm with hybrid search strategy, *Swarm and Evolutionary Computation* **28**, 1-13.
[13] Kennedy, J. and Everhart, R. (2002) Particle swarm optimization, *Int Conf Neural Netw* **4**, 1942-1948.
[14] Wei, Z., Liu, J. and Lu, Z. (2017) Structural damage detection using improved particle swarm optimization, *Inverse Problems in Science and Engineering* **26**, 792-810.
[15] Gokdag, H. and Yildiz, A. R. (2012) Structural damage detection using modal parameters and particle swarm optimization, *Materials Testing* **54**, 416-420.
[16] Hansen, P. C., Nagy, J. G. and O'leary, D. P. (2006) *Deblurring images: matrices, spectra, and filtering*, Society for Industrial and Applied Mathematics, USA.
[17] Tikhonov, A. N. and Arsenin, V. Y. (1977) *Methods for solving ill-posed problems*, John Wiley and Sons, USA.
[18] Tibshirani, R. (1996) Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267-288.
[19] Selesnick, I. W. (2010) Sparse signal restoration, *Polytech-nic Uni*, 1-16.
[20] Zuo, W., Ren, D., Zhang, D., Gu, S. and Zhang, L. (2016) Learning iteration-wise generalized shrinkage-thresholding operators for blind deconvolution, *IEEE Transactions on Image Processing* **25**, 1751-1764.

# Implementation of a conservative cut-cell method for the simulation of two-phase cavitating flows

***Yiannis-Panayiotis G. Vrionis[1], Konstantinos D. Samouchos[1], and Kyriakos C. Giannakoglou[1]**

[1]Parallel CFD & Optimization Unit, National Technical University of Athens, Greece.
[*]Presenting and Corresponding author: vrionis.yiannis@gmail.com

## Abstract

A cut-cell method utilizing 2D cartesian meshes with embedded boundaries is employed to simulate steady-state, turbulent and cavitating flows over isolated hydrofoils. The multi-phase Reynolds-Averaged Navier-Stokes equations for a homogeneous mixture are extended with an additional transport equation accounting for the liquid volume fraction and the Kunz cavitation model. The mean-flow equations are appropriately preconditioned to render the system hyperbolic in space and time and to alleviate numerical stiffness due to the low speed of sound of the mixture. The standard $k-\varepsilon$ turbulence model is implemented. A one-layer submesh is generated to solve Spalding's composite wall function on the embedded boundaries. Mesh generation employs a hierarchical quadtree-based data structure resulting in a fast and memory-efficient process. Cut-cells are constructed by intersecting the discretized geometry with the Cartesian mesh, creating thus faces where boundary conditions are imposed. The capabilities of the cut-cell solver are demonstrated over two hydrofoils featuring mid-chord and leading-edge cavitation. Results show satisfactory agreement with numerical and experimental data.

**Keywords:** Cut-cell method; Cavitation; CFD; Multi-phase flow.

## Introduction

Immersed Boundary Methods (IBMs) enjoy a high level of mesh generation automation and flexibility when simulating complex flows with moving or stationary geometries. Initially developed by [1], IBMs focused on making flow solvers to accurately predict inviscid [2], viscous [3] and, more recently, turbulent [4][5] flows. The mesh generation process, employing a hierarchical tree-based data structure, demonstrates high efficiency in terms of CPU cost and storage requirements [6].

IBMs can be classified as continuous and discrete, [7]. In the continuous IBMs, source terms are introduced into the flow equations to simulate the effect of solid boundaries. Since the source terms are not imposed directly on the geometry shape, the effect of the source terms is smeared across multiple neighbouring cells. Difficulties arise in the conservation of mass, momentum and energy in cells intersecting the solid boundaries, [8]. In contrast, in discrete IBMs, solid boundaries are represented by sharp interfaces and boundary conditions are imposed, without altering the governing PDEs. This representation, however, introduces temporal discontinuity, for the cells without time history, and additional actions are required to simulate moving geometries. Among the discrete IBMs, the cut-cell method reshapes finite volumes using their intersections with the geometry. Very small finite volumes may appear near the interface during reshaping that can cause numerical instabilities [9], if not properly treated. However, the

discrete representation of the geometry, employed by the cut-cell method, guarantees the satisfaction of local and global conservation laws.

In fluid flows, the inception of cavitation occurs when the static pressure of the liquid drops below the vapor pressure. Sheet, bubble, vortex or supercavitation may appear. In some applications, such as high-speed underwater bodies, cavitation is beneficial for drag reduction, while in some others, such as propellers, undesirable since it can reduce their performance and durability. The ability to predict and control its inception is, thus, of importance. As a result, cavitation, and its forms, has been studied experimentally in cases, such as the Venturi channel [10], the flow over hydrofoils [11] and cylindrical head-forms [12], and extensive research has been dedicated to the development of numerical models [13][14][15][16] capable of simulating cavitating flows. These models are classified into two categories, namely, interface tracking and interface capturing methods. In interface tracking methods, the liquid/vapor interface is explicitly tracked and acts as an internal boundary, whereas, the interface capturing methods provide the liquid/vapor interface as part of the solution procedure. Differences of the interface capturing methods lie in the set of governing equations satisfied. In one-fluid models, the conservation laws for the mixture are solved and the phase change is controlled by a state law, such as the barotropic mixture law. In two-fluid models, the conservation laws are solved for each fluid with appropriate mass transfer rates at their interface, whereas hybrid models, such as that employed in the present work, are based on the conservation laws for the mixture and a phasic transport equation coupled with a cavitation model to control the phase change rates. In cases, such as ventilated cavitating flows, where more than two phases are considered, the models can be extended either by modifying their state law [17], including additional conservation laws or by including additional transport equations [13] to account for additional phases. Several formulations for the cavitation model have been proposed, mostly relying on empirical formulas, see [13] and [15], or bubble dynamics such as the Full Cavitation Model [16].

Previous work combining IBMs and cavitating flows is limited to the use of one-fluid models. [18] employed a cut-cell method coupled with a modified Tait law, to simulate weakly compressible cavitating flows through a closing fuel injector control valve. Recently, [17] and [19] presented computational studies on cavitating flows for an external gear pump and a projectile impacting a water jet, both using a continuous IBM.

The present work employs a hybrid model in conjunction with the cut-cell method to simulate cavitating flows. The effects of the rapidly varying finite volumes near the solid boundary, as well as their impact on the mass transfer model are investigated. A hybrid model has been selected as it may provide high flexibility and capture baroclinic vorticity generation [20]. The proposed implementation allows the prediction of cavitating flows while also maintaining the benefits of employing a cut-cell method. Studies of turbulent flows over isolated hydrofoils are presented.

## Mathematical Model

The multi-phase Reynolds-Averaged Navier-Stokes (RANS) equations are adopted to simulate steady-state cavitating flows. A homogemeous mixture comprises liquid and vapor, both considered to be incompressible. The mixture is described by the momentum and phasic continuity equations. The continuity equation for the vapor phase is replaced by the one for the mixture, rather than using separate continuity equations for each phase. The retained mixture and liquid continuity equations are enriched with source terms to simulate mass transfer associated with the cavitation. From the numerical point of view, pseudo-time derivatives are introduced to

render the system hyperbolic in space and time and alleviate numerical stiffness created by the low speed of sound of the mixture.

The Jones and Launder $k-\varepsilon$ turbulence model [21], coupled with the Spalding's composite wall function [22] technique, is employed. The turbulence model takes into account the mixture densities and viscosities, with the turbulent variables referring to the mixture.

*Governing equations*

In vector form, the system of the preconditioned multi-phase mean-flow and turbulence equations for the mixture, is written as:

$$\begin{aligned} \Gamma\frac{\partial \mathbf{Q}}{\partial \tau} \; + \; \frac{\partial \mathbf{F_j}}{\partial x_j} - \frac{\partial \mathbf{F_j^v}}{\partial x_j} - \mathbf{S} = \mathbf{0} \\ \frac{\partial \mathbf{Q_T}}{\partial \tau} \; + \; \frac{\partial \mathbf{F_{T_j}}}{\partial x_j} - \frac{\partial \mathbf{F^v_{T_j}}}{\partial x_j} - \mathbf{S_T} = \mathbf{0} \end{aligned} \tag{1}$$

with $\mathbf{Q}=[\, p\ \mathbf{u}\ a_l\,]^T$ the mean-flow state vector, $\mathbf{Q_T}=[\, \varrho_m k\ \ \varrho_m \varepsilon]^T$ the turbulent variables state vector, $\tau$ the pseudo-time and $x_j$ the cartesian directions. The convective, $\mathbf{F_j}, \mathbf{F_{T_j}}$, viscous, $\mathbf{F_j^v}, \mathbf{F^v_{T_j}}$ fluxes and source vectors, $\mathbf{S}, \mathbf{S_T}$ are given by:

$$\mathbf{F_j} = \begin{bmatrix} u_j \\ \varrho_m u_j u_i + \delta^i_j\, p \\ a_l u_j \end{bmatrix}, \mathbf{F_j^v} = \begin{bmatrix} 0 \\ (\mu_m + \mu_{m,t}) \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \\ 0 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} (\dot{m}^+ + \dot{m}^-) \left( \frac{1}{\varrho_l} - \frac{1}{\varrho_v} \right) \\ \mathbf{0} \\ (\dot{m}^+ + \dot{m}^-) \, \frac{1}{\varrho_l} \end{bmatrix} \tag{2}$$

$$\mathbf{F_{T_j}} = \begin{bmatrix} \varrho_m k u_j \\ \varrho_m \varepsilon u_j \end{bmatrix}, \; \mathbf{F^v_{T_j}} = \begin{bmatrix} \left( \mu_m + \frac{\mu_{m,t}}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \\ \left( \mu_m + \frac{\mu_{m,t}}{\sigma_\varepsilon} \right) \frac{\partial \varepsilon}{\partial x_j} \end{bmatrix}, \; \mathbf{S_T} = \begin{bmatrix} \mathbb{P} - \varrho_m \varepsilon \\ (C_1 \mathbb{P} - C_2 \varrho_m \varepsilon) \, \frac{\varepsilon}{k} \end{bmatrix} \tag{3}$$

where $p,\ \mathbf{u}\ \left(=[\, u,\ v]^T\right),\ a,\ \varrho,\ \mu,\ k,\ \varepsilon$ refer to pressure, velocity vector, volume fraction, density, viscosity coefficient, turbulent kinetic energy and turbulence dissipation rate, respectively. Subscripts $m,\ l,\ v$ refer to the mixture, liquid and vapor phase, respectively. $\delta$ is the Kronecker delta. Quantities $\dot{m}^{\pm}$ are associated with the mass transfer model rates characterizing both phase changes; for them, the Kunz cavitation model [13] is implemented. The evaporation term $m^-$ is a function of the pressure and the liquid volume fraction whereas the condensation term $m^+$ is a function of the liquid volume fraction, $a_l$,

$$\dot{m}^- = \frac{C_{dest} \varrho_v a_l \; min\,[0, p - p_v]}{\left( \frac{1}{2} \varrho_l U_\infty^2 \right) t_\infty}, \qquad \dot{m}^+ = \frac{C_{prod} \varrho_v a_l^2 \, (1 - a_l)}{t_\infty} \tag{4}$$

The empirical time rate constants $C_{dest}$ and $C_{prod}$ are case-dependent and their values are non-dimensionalized with respect to the mean-flow time scale, $t_\infty = \frac{L}{U_\infty}$. $L$ is the characteristic length scale, in the case of hydrofoils their chord length, and $p_v$ is the vapor pressure.

The preconditioning matrix $\Gamma$ takes the form [13]

$$\Gamma = \begin{bmatrix} \frac{1}{\varrho_m \beta^2} & 0 & 0 & 0 \\ 0 & \varrho_m & 0 & u\Delta\varrho \\ 0 & 0 & \varrho_m & v\Delta\varrho \\ \frac{a_l}{\varrho_m \beta^2} & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

with $\beta$ the pseudo-compressibility parameter and $\Delta\varrho = \varrho_l - \varrho_v$. The mixture density and molecular viscosity are based on a local volume-averaging:

$$\varrho_m = a_l \varrho_l + (1 - a_l)\varrho_v \tag{6}$$

$$\mu_m = a_l \mu_l + (1 - a_l)\mu_v \tag{7}$$

The turbulence production term is $\mathbb{P} = \mu_{m,t} \left( \frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) \frac{\partial u_j}{\partial x_i}$ and the mixture turbulent viscosity coefficient is $\mu_{m,t} = C_\mu \varrho_m \frac{k^2}{\varepsilon}$. The turbulence model constants are: $C_\mu = 0.09,\ C_1 = 1.44,\ C_2 = 1.92$. Spalding's wall function formula, resolves both the viscous and the log layer through a single formula,

$$y^+ \left(u^+\right) = u^+ + e^{-\kappa B} \left( e^{\kappa u^+} - 1 - \kappa u^+ - \frac{1}{2} \left(\kappa u^+\right)^2 - \frac{1}{6} \left(\kappa u^+\right)^3 \right) \tag{8}$$

with $u^+ = \frac{u_t}{u_\tau},\ y^+ = \frac{u_\tau \Delta}{\nu}$ and $u_\tau,\ \Delta,\ \nu, \kappa$ represent the friction velocity, distance from the wall, kinematic viscosity and von Karman constant, respectively.

## Discretization and Numerical Solution

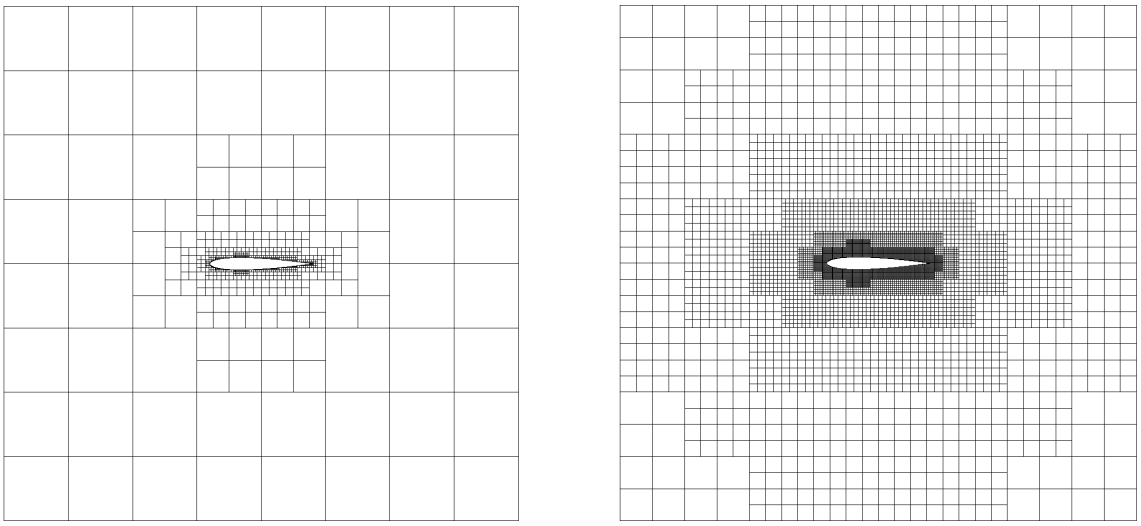
To solve the homogeneous multi-phase flow equations, a cell-centered, finite-volume scheme on unstructured meshes is employed. Eqs. (1) are integrated and discretized with second-order accuracy in space. The resulting linearized system is solved using a block Gauss-Seidel method. The implementation is able to run in parallel on distributed memory systems by partitioning the mesh into subdomains, communicating using the MPI protocol.

### *Generation of a Geometry Adapted Computational Mesh*

The generation of an adapted computational mesh with embedded boundaries is fully automated [23]. Initially, a uniform mesh is generated and cells intersected with the body surface are identified. Then, a recursive algorithm is employed to refine all intersected cells until the user-defined minimum cell volume has been reached, without allowing the refinement levels of two adjacent cells to differ by more than one. Once the maximum refinement level is reached at the intersected cells, the fluid part of the cells is retained, whereas the solid part is discarded, thus creating the cut-cells. Occasionally, cut-cells with volumes noticeably smaller than originally can be created that could stiffen the numerical solution of the flow equations [9]. To avoid convergence difficulties due to these cells, algorithms that merge cut-cells with volumes less than a user-defined threshold value with one of their immediate neighbours are employed [9]. Merged cells are differently shaped finite volumes, treated identically to all other cut-cells.

This procedure produces Cartesian meshes that are generally unsuitable for CFD simulations with large flow gradients present, e.g the viscous boundary layers and the cavity interface, since the cells in the immediate vicinity of solid walls abruptly change refinement levels, Fig. 1a.

Therefore, the refinement levels across the computational domain are smoothed to increase mesh resolution close to solid walls, by employing a progressive refinement algorithm. The algorithm, based on the distance of each cell from the geometry and a sigmoid function, computes the maximum allowable volume of each cell. Cells exceeding this value are refined. In the absence of smoothing, the refinement levels of neighbouring cells, result in increased numerical errors near the geometry. The mesh generation and adaptation process is based on a hierarchical quadtree data structure [9] and follows stages illustrated in Fig. 1.



**(a)** **(b)**

**Figure (1): Stages of mesh generation. Computational mesh of the NACA0012 hydrofoil (a) before and (b) after implementing the progressive refinement algorithm.**

*Discretization of the Governing Equations*

Equations (1) are integrated, over a finite volume $\Omega$, yielding:

$$\Gamma \frac{V}{\Delta\tau}\,\Delta\mathbf{Q} + \int_{\partial\Omega} \mathbf{F_j} n_j \, dS - \int_{\partial\Omega} \mathbf{F_j^v} n_j \, dS - \int_{\Omega} \mathbf{S} dV = \mathbf{0} \tag{9}$$

where $\mathbf{n}$ the unit normal vector to the faces of each volume. The discretization of the turbulence model equations is similar, thus, omitted.

The discretization of the convective fluxes in Eq. (9) is based on the Roe scheme [24], using a second-order linear reconstruction. The state variable gradients at each cell centre are estimated by a linear least-squares method using all face neighbours. In the presence of strong flow gradients, the extrapolation procedure is limited to provide a monotone-preserving TVD scheme. The pseudo-time step $\Delta\tau$ is determined by stability criteria.

The inviscid flux, $\mathbf{\Phi}$, at each cell face reads

$$\begin{aligned} \mathbf{\Phi} &= \frac{1}{2}\left(\mathbf{F^L} + \mathbf{F^R} - \Gamma_{Roe}\left|\hat{A_{Roe}}\right|\left(\mathbf{Q}^- - \mathbf{Q}^+\right)\right) \\ \mathbf{Q}^{\pm} &= \mathbf{Q^{L/R}} + \psi\,\nabla\mathbf{Q}\Delta\mathbf{x} \end{aligned} \tag{10}$$

$$\left|\hat{A_{Roe}}\right| = \left|\Gamma_{Roe}^{-1} A_{Roe}\right|, \qquad |A_{Roe}| = M_{Roe}\,|\Lambda_{Roe}|\,M_{Roe}^{-1} \tag{11}$$

In the equations above, $\psi$ is the Barth-Jespersen limiter [25] value, $M,\ M^{-1}$ are the right and left eigenvectors of the preconditioned Jacobian matrix, $\Gamma^{-1}A$, and $\Lambda$ is the diagonal eigenvalue matrix, $diag\,(\Lambda) = [\mathbf{u},\ \mathbf{u}+c,\ \mathbf{u}-c,\ \mathbf{u}]^T$. The pseudo-sound, $c$, is computed as $c = \sqrt{(u_n + \beta^2)}$ and $u_n$ is the normal to the face velocity, [13]. Superscripts $+, -$ refer to the reconstructed variables, while $L/R$ to the adjacent to the face cells. Subscript $(Roe)$ denotes Roe-averaged variables, defined as

$$\varphi_{Roe} = \frac{\sqrt{\varrho^+}\varphi^+ + \sqrt{\varrho^-}\varphi^-}{\sqrt{\varrho^+} + \sqrt{\varrho^-}}, \quad \varphi \in (\mathbf{u}, a_l, k, \varepsilon) \tag{12}$$

The velocity gradients at each cell face, necessary for the viscous fluxes, are computed using central differences. At the faces of cut-cells or those separating cells with different refinement levels, the barycenters of the corresponding cells are not aligned with the face midpoint and, thus, undergo a non-orthogonality correction.

For the solution procedure, the Jacobians $A = \frac{\partial \mathbf{F}}{\partial \mathbf{Q}}$ are computed analytically. Source terms are linearized in order to promote diagonal dominance.

*Implementation of Wall Functions*

A one-layer submesh, Fig. 2, with constant normal distance from the solid boundary is generated, as in [26]. Cells dimension $(\Delta\mathbf{x})$ can be computed based on their refinement level, by exploiting the quadtree data structure [9]. Then, the normal distance is computed $\Delta = 1.6min(\Delta\mathbf{x})$, and is the same for all cut-cells, since all cut-cells reach the same level. The normal distance ensures that the submesh nodes reside outside the (possibly merged) cut-cell. Eq. (8) is solved to extract the required wall shear stress and turbulent variables boundary conditions, by employing the submesh. Finally, the velocity gradients of the cut-cells are replaced by the ones provided by the Spalding formula [26]. The submesh node corresponding to the solid-

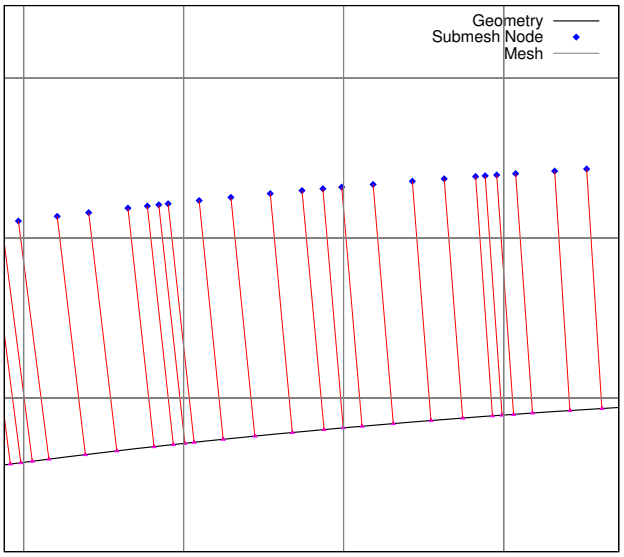



**Figure 2: One-layer submesh used to implement the wall function technique. Each submesh node is connected with a solid face of the cut-cell.**

node line closest to each cut-cell barycenter is located and used to compute the friction velocity $u_\tau$. Each cut-cell is associated with a submesh node on which Eq. (8) is solved. The tangent to

the wall velocity $u_t$ at that node is linearly reconstructed from the nearest cell to compute $u_\tau$, iteratively. The shear stress is computed separately at each solid face to account for cases where the solid face direction changes inside a cut-cell, providing a more accurate computation of the corresponding solid face viscous flux.

*Numerical Treatment of Cut-Cells - Application of Boundary Conditions*

A blow-up view of the area close to the interface with a solid body is presented in Fig. 3a. Two cut-cells, a merged (blue) and non-merged (brown), are included for illustative purposes. A merged cell is created when a slave cell (the smaller one) attaches to a master cell to create a new bigger cell. A master cell accomodates one slave cell at most to prevent the formation of non-convex finite volumes. The sizes of the faces of a slave cell are compared to determine which neighbour should become its master cell; the one with the largest common face is selected as master cell.

Fig. 3b shows a cut-cell finite volume along with two fluxes (*4*, *5*) due to the presence of the solid boundaries. For brevity, the following refer to inviscid flows where the no-penetration $(\mathbf{u}\cdot\mathbf{n}=\mathbf{0})$ condition is applied. For all internal faces, (*1*, *2*, *3*, *6*), fluxes are cast in Cartesian coordinates, e.g. flux through face *2* becomes $\mathbf{F}\cdot\hat{\mathbf{n}}=[u,\ \varrho_m u^2+p, \varrho_m uv,\ 0]^T$. On solid faces, boundary conditions are weakly imposed. Flow variables are extrapolated at the midpoint of each solid face and the appropriate flux, $\mathbf{F}^w=[0, p_{face}\ \mathbf{n},\ 0]^T$ is imposed.

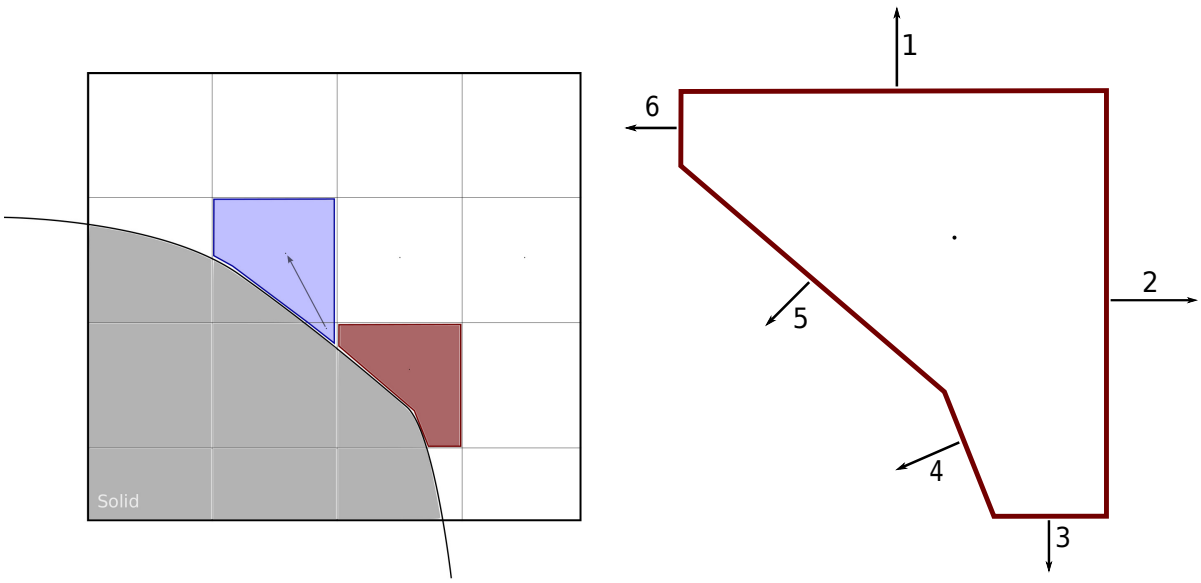



**Figure (3): (a) Areas near the solid boundary of a mesh with embedded boundaries where a merged (blue) and a non-merged (brown) cut-cell can be seen. (b) Enlarged view of a cut-cell; normal vectors represent fluxes. Solid wall conditions are imposed along faces *4* and *5*.**

## Results and Discussion

The capabilities of the cut-cell solver in the simulation of sheet cavitation are demonstrated in numerical studies of two isolated hydrofoils, for different cavitation numbers, and results are

compared with published data, [11][14][27]. Sheet cavitation has been experimentally observed to exhibit a quasi-steady behaviour and the experimental data [11] obtained regarding sheet cavitation are time-averaged. Therefore, it is a common practice, e.g [14][27][28], to perform simulations using a steady-state solver. The cavitation number is defined as $\sigma = \frac{p_\infty - p_v}{\frac{1}{2}\varrho_l U_\infty^2}$ and the pseudo-compressibility parameter as $\beta^2 \approx 10 U_\infty^2$.

*Cavitating Flow around the NACA0012 hydrofoil*

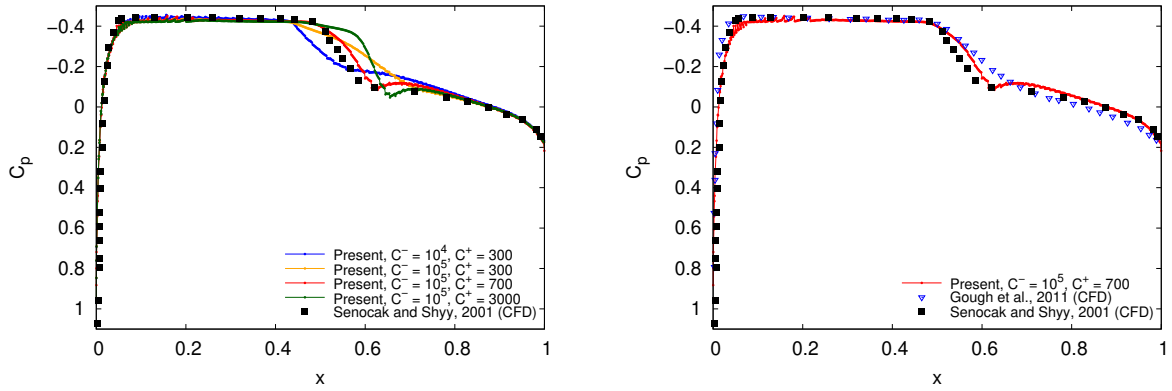
The first case considers mid-chord sheet cavitation on a NACA$0012$ hydrofoil, based on the numerical study presented in [14]. [14] employed the same cavitation model in a pressure-based (SIMPLE) algorithm, with a density ratio equal to $\frac{\varrho_l}{\varrho_v} = 1000$, which is also valid here. The Reynolds number is $Re_c = 2 \cdot 10^6$ based on the chord and the infinite flow angle is $1^\circ$. With a cavitation number equal to $\sigma = 0.42$, a thin layer of vapor is created at the mid-chord of the hydrofoil that exhibits no re-entrance jet/recirculation. A mesh with $42K$ cells is generated with sufficient progressive refinement near the solid body and an average $\overline{y}^+ \approx 90$ of the first cell barycenters off the wall (Fig. 5a) to capture the thin vapor layer.

In the literature [14][28], it is reported that the empirical time rate constants $(C_{prod},\ C_{dest})$ of the mass transfer model are case dependent and different values are proposed and used by different codes for the same test cases. In the current implementation, the proposed time rate constants also produced unrealistic results. A parametric study was conducted to calibrate the dimensionless mass transfer time rates constants, $C^+ = \frac{C_{prod}}{t_\infty},\ C^- = \frac{C_{dest}}{t_\infty}$.

Fig. 4a shows pressure distributions obtained during this study with several pairs of constants and their impact on the converged solution. In the results presented hereby, the pressure inside the cavity is equal to $p_v$, but this is not the case if $C^-$ becomes lower. The pressure gradients aft the cavity are influenced by the correct balance between the two source terms. Too small values of $C^-$ shorten the cavity length, while too big values of $C^+$ steepen the pressure gradients at the cavity closure region.

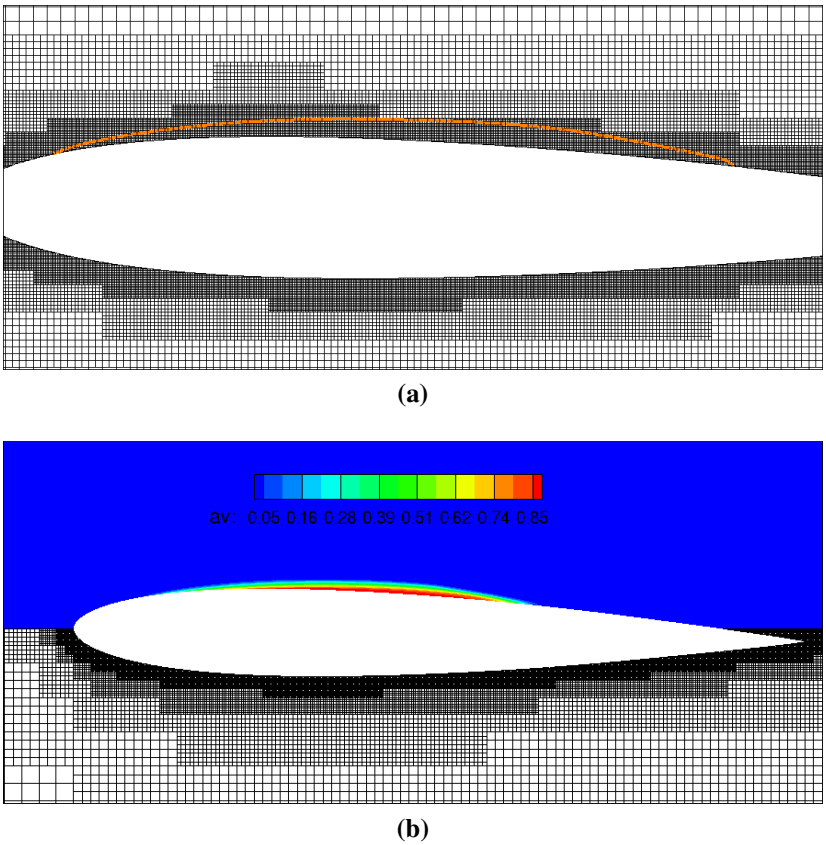
Fig. 4b compares the results of the surface pressure profile, with the selected time rate constants, of the current implementation with two numerical results obtained on body-fitted mesh solvers under the same conditions and mass transfer model, though with different time rates. Differences are limited to the cavity closure region.

In Fig. 5b, the vapor volume fraction field and, therefore, the cavity is presented. A thin layer of vapor is generated along the suction side of the hydrofoil. The maximum vapor volume fraction inside the cavity $a_{v_{max}}$ approaches $0.9$. The liquid-vapor interface is spread across multiple cells. This might be due to the smooth pressure gradients observed at the cavity closure, which implies that the condensation source term is weak enough to avoid an instantaneous phase change and make it appear gradually. Fig. 5a shows a close-up view of the area near the cavity. By plotting the liquid volume fraction iso-line $a_l = 0.99$ over the mesh, it can be seen that the cavity resides inside the finest refinement levels of the mesh.




**(a)** **(b)**

**Figure 4: Surface pressure suction side profile over a NACA0012 hydrofoil, $\alpha$=1°, $Re_c$=$2 \cdot 10^6$ and $\sigma$=0.42. (a) Parametric study to find the best value set of $C^+$ and $C^-$. (b) Comparison with the numerical studies of [14][28] and the best value-set of time constants found.**




**(a)**

**(b)**

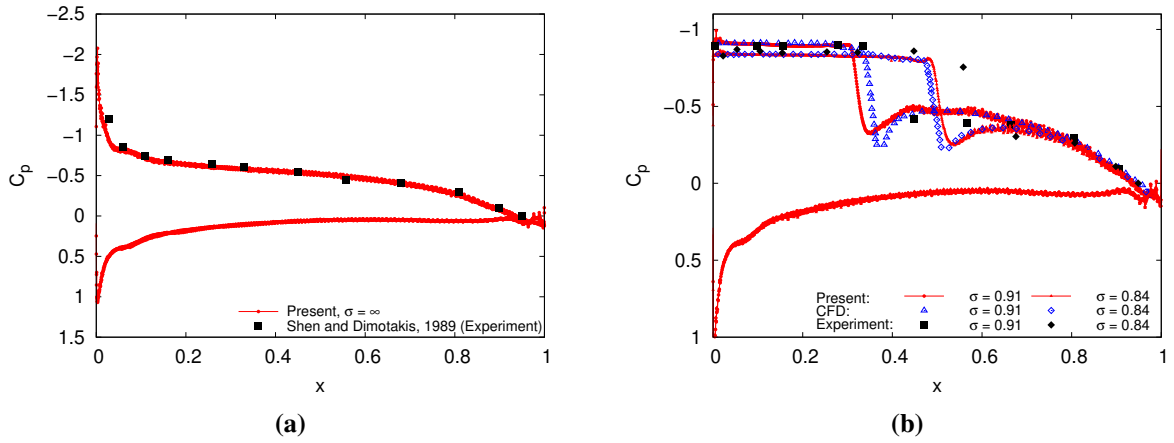
**Figure 5: NACA0012 hydrofoil: (a) Liquid volume fraction iso-line (orange) $a_l$=0.99 which determines the size of the cavity along with the smoothed mesh refinement near the geometry. (b) Iso-areas of the vapor volume fraction $a_v$.**

*Cavitating Flow around the NACA66(MOD) hydrofoil*

In this section, the cavitating flow over a NACA 6-series hydrofoil is presented. The hydrofoil has been experimentally investigated in [11] for leading edge and mid-chord cavitation

and time-averaged pressure distributions are available in the literature [11]. The cambered NACA66(MOD) hydrofoil has a camber ratio of $0.02$, a mean line of $0.8$, a thickness ratio of $0.09$ and a chord length of $0.1524m$. The presented study focuses on steady-state simulations of flows with an infinite flow angle of $\alpha = 4^{\circ}$, a Reynolds number $Re_c = 2 \cdot 10^6$, based on the chord length, a density ratio $\frac{\varrho_l}{\varrho_v} = 1000$ and three cavitation numbers, $\sigma = \infty$, $\sigma = 0.91$ and $\sigma = 0.84$. A cartesian mesh with $80K$ cells is generated with sufficient mesh resolution near the geometry and $\overline{y}^+ \approx 50$. The best value-set of the two time rate constants was found to be $C^- = 10^6$, $C^+ = 9000$ after a parametric study. Furthermore, for the cavitation numbers $\sigma$=$0.91,\ 0.84$ the surface pressure distributions are also compared with the numerical results of [27]. The latter employed a commercial CFD solver and calibrated the mass transfer rates for the same conditions over a NACA66(MOD) hydrofoil.

The first case, $\sigma = \infty$, is the non-cavitating case and is included to validate the cut-cell solver with the experimental data of [11]. In the other cases, with $\sigma$=$0.91,\ \sigma$=$0.84$, vapor is generated since the pressure decreases below the pressure vapor and cavities are observed. Fig. 6a shows the surface pressure distributions of the non-cavitating case ($\sigma$=$\infty$) and the experimental data from [11]. The surface pressure distribution compares favourably with the time-averaged experimental data. In Fig. 6b, the surface pressure distribution in the case with $\sigma = 0.91$ is presented and compared with measurements [11] and other numerical results [27]. The results obtained with the cut-cell method show that the pressure inside the cavity remains constant and agrees with both numerical and experimental data. Furthermore, comparing with the numerical results, an underprediction of the cavity length is also observed. Fig. 7a, shows the vapor vol-




**(a)** **(b)**

**Figure 6: Surface pressure distribution over the NACA66(MOD), (a) The non-cavitating ($\sigma$=$\infty$) case (b) Cavitating cases of $\sigma$=$0.91$ and $\sigma$=$0.84$.**

ume fraction with the computational mesh employed. In Fig. 7c the pressure field is presented. The constant pressure inside the cavity can be seen and its value is approximately equal to the negative of the cavitation number $\sigma$=$0.91$, as expected.

By further reducing the cavitation number, a larger cavity is created. In Fig. 6b, the surface pressure distribution for $\sigma = 0.84$ is compared with experimental time-averaged data [11] and numerical results [27]. Both numerical results produce the appropriate pressure inside the cavity and capture the same pressure recovery trend. Fig. 7b shows the vapor volume fraction on the suction side of the hydrofoil. The pressure flow field is shown in Fig. 7d. Finally, comparing Fig. 7(a)-(b), it can be seen that decreasing the cavitation number increases the size of the vapor cavity created.

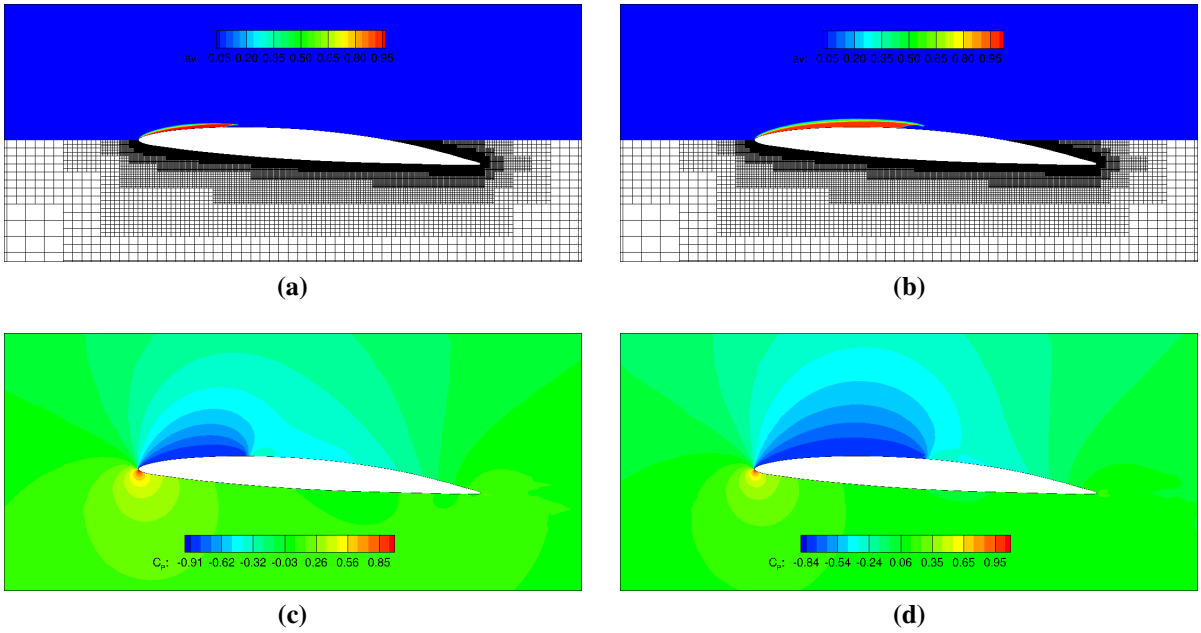



**Figure 7: Cavitating NACA66(MOD) hydrofoil. $\sigma$=0.91 (left) and $\sigma$=0.84 (right). (a) - (b) Vapor volume fraction along with the computational mesh used and (c) - (d) Pressure coefficient iso-areas.**

Compared to the experimental data, both numerical solvers underpredict the cavity length at $\sigma = 0.84$. The differences noted, could be attributed to the additional multi-phase turbulent effects which dominate in these areas. Addressing the additional turbulent effect in cavitating flows is still an ongoing research area and new findings could improve numerical predictions. However, this is beyond the scope of the present paper.

The use of flow-based adaptation is common in transonic and supersonic flows [23], where shock waves are present and require increased mesh resolution locally. IBMs, employing hierarchical data structures, make flow-based adaptation techniques easy to implement and is considered one of their main advantages over their body-fitted counterparts. Hence, a natural next step is the addition of flow-based refinement in the multi-phase solver as steep flow gradients also characterize cavitating flows. The choice and definition of the adaptation sensor, that identifies the areas of interest and implements the refinement algorithm, require special attention. An obvious choice may be the normalized $\|\nabla a_l\|$, where large gradients are present only near the cavity interface, see in Fig. 8. The pressure gradients could also be considered as an adaptation sensor since the mass transfer models are closely related to the field pressure.

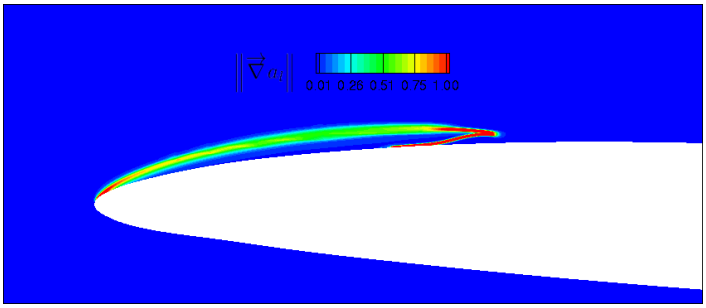



**Figure 8: Iso-areas of the normalized $\|\nabla a_l\|$ over the cavitating NACA66 hydrofoil.**

## Conclusions

A cut-cell solver for a homogeneous mixture model has been implemented to simulate steady, turbulent, cavitating flows, based on the Kunz model. The cut-cells and the one-layer submesh that help to employ the Spalding's composite wall function model for the $k - \varepsilon$ turbulence model, together with the merging of the small cut-cells, are shown to handle the large density gradients present in cavitating flows.

To assess the programmed method, two hydrofoil cases featuring mid-chord and leading edge sheet cavitation have been selected and comparisons with experimental and numerical data are presented. The cavitation model implemented proved to be sensitive to the two time rate constants which should, therefore, be carefully selected to avoid nonphysical solutions or the divergence of the numerical scheme. The automatic mesh generation and refinement, even in complex geometries, offered by the cut-cell method, renders it an appealing alternative to CFD solvers employing body-fitted meshes. This is particularly important in case of moving bodies, which is not the case in this paper; to summarize, the programmed software simulates both 2D and 3D flows, with stationary or moving bodies (see [29], presented in the same conference).

## Acknowledgement

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

### References

[1] C. S. Peskin, "Flow Patterns Around Heart Valves: A Numerical Method," *Journal of Computational Physics*, vol. 10, pp. 252–271, 1972.

[2] D. Clarke, M. Salas, and H. Hassan, "Euler Calculations for Multielement Airfoils Using Cartesian Grids," *AIAA Journal*, vol. 24, no. 3, pp. 353–358, 1986.

[3] T. Ye, R. Mittal, H. S. Udaykumar, and W. Shyy, "An Accurate Cartesian Grid Method for Viscous Incompressible Flows with Complex Immersed Boundaries," *Journal of Computational Physics*, vol. 156, pp. 209–240, 1999.

[4] F. Capizzano, "Turbulent Wall Model for Immersed Boundary Methods," *AIAA Journal*, vol. 49, pp. 2367–2381, 2011.

[5] M. J. Berger and M. J. Aftosmis, "An ODE-Based Wall Model for Turbulent Flow Simulations," *AIAA Journal*, vol. 56, pp. 700–714, 2018.

[6] M. J. Aftosmis, M. J. Berger, and J. E. Melton, "Robust and Efficient Cartesian Mesh Generation for Component-Based Geometry," *AIAA Journal*, vol. 36, pp. 952–960, 1998.

[7] R. Mittal and G. Iaccarino, "Immersed Boundary Methods," *Annual Review of Fluid Mechanics*, vol. 37, no. 1, pp. 239–261, 2005.

[8] J. Kim, D. Kim, and H. Choi, "An Immersed-Boundary Finite-Volume Method for Simulations of Flow in Complex Geometries," *Journal of Computational Physics*, vol. 171, pp. 132–150, 2001.

[9] H. Ji, F. S. Lien, and E. Yee, "A new Adaptive Mesh Refinement Data Structure with an Application to Detonation," *Journal of Computational Physics*, vol. 229, pp. 8981–8993, 2010.

[10] B. Stutz and J.-L. Reboud, "Two-phase Flow Structure of Sheet Cavitation," *Physics of Fluids*, vol. 9, pp. 3678–3686, 1997.

[11] Y. Shen and P. E. Dimotakis, "The Influence of Surface Cavitation on Hydrodynamic Forces," in *Proc. 22nd ATTC.*, (St. John's, Newfoundland, Canada), pp. 44–53, 1989.

[12] J. S. Rouse, Hunter and McNown, "Cavitation and Pressure Distribution: Head Forms at Zero Angle of Yaw," in *Studies in Engineering Bulletin*, (Iowa City), p. 32, State University of Iowa, 1948.

[13] R. F. Kunz, D. A. Boger, D. R. Stinebring, T. S. Chyczewski, J. W. Lindau, H. J. Gibeling, S. Venkateswaran, and T. R. Govindan, "A Preconditioned Navier-Stokes Method for Two-phase Flows with Application to Cavitation Prediction," *Computers and Fluids*, vol. 29, pp. 849–875, 2000.

[14] I. Senocak and W. Shyy, "Numerical Simulation of Turbulent Flows with Sheet Cavitation," in *CAV2001: Fourth International Symposium of Cavitation*, (California Institute of Technology, Pasadena, CA, U.S.A.), pp. 20–23, 2001.

[15] C. L. Merkle, J. Feng, and P. E. O. Buelow, "Computational Modeling of the Dynamics of Sheet Cavitation," in *Proceedings of the 3rd International Symposium on Cavitation*, (Grenoble, France), 1998.

[16] A. K. Singhal, M. M. Athavale, H. Li, and Y. Jiang, "Mathematical Basis and Validation of the Full Cavitation Model," *Journal of Fluids Engineering*, vol. 124, no. 3, p. 617, 2002.

[17] M.-G. Mithun, P. Koukouvinis, I. K. Karathanassis, and M. Gavaises, "Numerical Simulation of Three-phase Flow in an External Gear Pump Using Immersed Boundary Approach," *Applied Mathematical Modelling*, vol. 72, pp. 682–699, 2019.

[18] F. Örley, V. Pasquariello, S. Hickel, and N. A. Adams, "Cut-element based Immersed Boundary Method for Moving Geometries in Compressible Liquid Flows with Cavitation," *Journal of Computational Physics*, vol. 283, pp. 1–22, 2015.

[19] E. Stavropoulos Vasilakis, N. Kyriazis, P. Koukouvinis, M. Farhat, and M. Gavaises, "Cavitation Induction by Projectile Impacting on a Water Jet," *International Journal of Multiphase Flow*, vol. 114, pp. 128–139, 2019.

[20] I. Senocak, "Computational Methodology for the Simulation of Turbulent Cavitating Flows. PhD thesis, University of Florida, 2002.

[21] W. Jones and B. Launder, "The Prediction of Laminarization with a Two-equation Model of Turbulence," *International Journal of Heat and Mass Transfer*, vol. 15, pp. 301–314, 1972.

[22] D. B. Spalding, "A Single Formula for the Law of the Wall," *Journal of Applied Mechanics*, vol. 28, no. 3, p. 455, 1961.

[23] K. Samouchos, S. Katsanoulis, and K. C. Giannakoglou, "Unsteady Adjoint to the Cut-cell Method Using Mesh Adaptation on GPUs," in *ECCOMAS Congress*, (Crete Island, Greece), 2016.

[24] P. Roe, "Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes," *Journal of Computational Physics*, vol. 43, pp. 357–372, 1981.

[25] T. Barth and D. Jespersen, "The Design and Application of Upwind Schemes on Unstructured Meshes," in *27th Aerospace Sciences Meeting*, (Reston, Virigina), AIAA, 1989.

[26] M. J. Berger and M. J. Aftosmis, "Progress Towards a Cartesian Cut-Cell Method for Viscous Compressible Flow," in *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, 2012.

[27] M. Morgut and E. Nobile, "Numerical Predictions of Cavitating Flow Around Model Scale Propellers by CFD and Advanced Model Calibration," *International Journal of Rotating Machinery*, vol. 2012, 2012.

[28] H. Gough, A. L. Gaitonde, and D. P. Jones, "A Dual-time Central-difference Interface-capturing Finite Volume Scheme Applied to Cavitation Modelling," *International Journal for Numerical Methods in Fluids*, vol. 66, pp. 452–485, 2011.

[29] K. Samouchos, D. Kapsoulis, X. Trompoukis, K. Giannakoglou, "Shape Optimization of 3D Diaphragm Pumps using the Continuous Adjoint Approach to the Cut-Cell Method," in *ICCM2019*, (Singapore), 2019.

# A GPU based acceleration of Finite Element and Isogeometric analysis

***Utpal Kiran, Vishal Agrawal, Deepak Sharma and †Sachin Singh Gautam**

Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Assam, India.

*Presenting author: ukiran@iitg.ac.in
†Corresponding author: ssg@iitg.ac.in

## Abstract

In this paper, a Graphics Processing Unit (GPU) based novel parallelization scheme is proposed to reduce the extensive computational cost associated with the finite element (FE) and isogeometric analysis (IGA) simulations of linear and non-linear problems. An innovative parallelization strategy is introduced that achieves fine-grain parallelism and is suitable for GPU. The proposed strategy is implemented within the MatLab® programming environment for the GPU based FE and IGA simulations. It, thus, avoids the need for specialized programming languages like CUDA/C++, which require in-depth knowledge for their implementation. The capabilities and performance of the proposed strategy are examined by solving both linear and non-linear problems. The results demonstrate that the proposed strategy achieves a considerable improvement in the assembly and computation of global tangent matrices over both the single core CPU and multicore CPU based computations. A maximum speedup of 41.4 times over single core CPU and 10.8 times over multicore CPU is achieved for linear problem. For non-linear problem of strip peeling from an adhesive substrate, a maximum speedup of 12.3 times is obtained in comparison to multicore CPU based computation. The proposed strategy can be easily incorporated within the existing codes with little modification.

**Keywords:** Parallel Programming, GPU computing, IGA, Nonlinear FEA, MatLab.

## Introduction

Finite Element Analysis (FEA) is one of the most popular numerical methods used for the solution of a variety of problems governed by the partial differential equations [1]. It is known that FEA involves a large amount of computation to find the nearly exact solution of the problem. Moreover, the computational efforts increase substantially if the analysis includes a large number of finite elements. It, as a resultant, leads to a significant amount of simulation time even on fast modern computers. However, this may not be a desirable choice in the analysis community.

In 2005, Hughes et. al [2] introduced the Isogeometric analysis (IGA) technique to circumvent the above mentioned issue associated with FEA. Unlike FEA, IGA doesn't need to perform expensive mesh generation of the Computer Aided Design (CAD) model as it directly enables the transition of CAD generated model into the analysis framework. Thus, reducing the execution time of IGA based simulation significantly compared to FEA [3].

However, the computational time in case of IGA solver for large-scale problems can still be high [3]. The large simulation time in scientific applications is often reduced by using parallel computers. It involves decomposing a large-scale problem into a smaller number of parts and solving them in parallel over multiple processors. Recently, GPU based parallel computing has achieved great success in accelerating time-consuming scientific applications [4]. The

GPU is specialized hardware designed to handle parallel and independent data task in a very efficient manner. It is a massively threaded processor having thousands of simpler cores instead of few powerful cores like in CPU. A large number of cores in GPU bring the performance of a mini computer cluster to the desktop computer at very low capital cost, low cooling cost and low power consumption [5]. There are numerous applications accelerated successfully by GPU in various fields including aerospace, defense, finance & economics, oil & gas, and computer games [6]. However, writing code for a GPU requires knowledge of parallel programming strategy and specific programming language like CUDA C/C++. Since CUDA C/C++ is a lower level programming language, it demands a great amount of time and effort from the user. On the other hand, scripting languages like MatLab® have become more popular in the scientific community. MatLab® provides an integrated computing environment, which supports the effortless development of code, easier and faster debugging, visualization, and a large number of built-in functions. The ease of code development in MatLab® comes at the cost of lesser flexibility and reduced control to the programmers, which may lead to sub-optimal performance. In the current work, a novel parallelization strategy is introduced which provides a possibility to achieve accurate result at a considerably lower computational cost compared to standard sequential computation approaches. The execution time is further reduced significantly by the use of GPU.

The previous efforts to accelerate FEA on MatLab® have focused on efficient vectorization techniques or parallel computation on multicore CPU through `parfor` or `spmd` construct of parallel computing toolbox [7]. A parallel implementation for coupled electro-mechanical finite element analysis of micro-electro-mechanical (MEMS) device is found in [8]. The work shows the use of `parfor` loop to calculate element stiffness matrices in parallel over 40 MatLab® workers set up to reduce simulation time from 60 hours to 2 hours. A significant amount of reduction in FEA assembly time is achieved by vectorization of code in [9]. In another work [10], a comparative analysis of multicore parallelization and GPU parallelization is done. The implementation uses `parfor` and `spmd` construct for CPU and `arrayfun` function wrapper for GPU. The result shows GPU based algorithm performing poorly than other two. However, the authors believe that with efficient vectorization GPU based parallelization can achieve better performance.

Most of the previous work on acceleration of FEM using GPU is found to be based on CUDA C/C++. The most detailed study of GPU implementation of finite element assembly process is presented in [11]. The authors show the speedup of several folds in assembly for lower order as well as higher order elements. An efficient implementation of numerical integration on GPU is found in [12]. The authors show speedup of 7× over the efficient CPU implementation for quadrilateral element. In [13], a novel interaction-wise strategy for assembly of stiffness matrix in IGA is presented. The proposed strategy achieves speedup up to 54× over single core CPU implementation. The GPU based integration strategy of B-spline basis function in IGA is found in [14]. The above-mentioned works along with many others in literature [15] signify the effectiveness of GPU in accelerating FEM simulation. However, to the best of authors' knowledge no literature exists that discusses the capabilities of MatLab® to use GPU to accelerate FEM.

FEM consists of a sequence of computationally expensive steps like evaluation of local matrices (mass and stiffness), assembly of local matrices into global matrix and solution of assembled system of equations [1]. GPU based computing has been found to be very effective in accelerating almost every step of FEM [15]. The solution of linear system of equation often dominates the simulation time and, therefore, it must be implemented with an optimized linear solver [16][17]. However, the time consumed in element matrix creation and their

assembly to global matrix cannot be ignored, particularly, in nonlinear problems. In nonlinear problems, a large number of time steps are required to reach final solution. Within each time step, there are Newton-Raphson iterations that require revaluation of element stiffness matrices and their reassembly [1]. Thus, an optimum implementation of this step can lead to significant amount of reduction in simulation time.

The objective of the current work is to accelerate the evaluation and assembly of mass and tangent matrices by making use of GPUs through parallel computing toolbox of MatLab®. A MatLab® code can be made to run on GPUs with the minimum amount of changes requiring far less development effort than the language like CUDA C/C++. The function wrappers provided by MatLab® like `bsxfun`, `pagefun`, and `arrayfun` have been used for numerical integration. Assembly to global matrix is done by `sparse` function of the MatLab®. First, an efficient GPU parallel strategy for FEM analysis for 2D elasticity problems is proposed and compared with sequential and CPU parallel (`parfor`) strategy. The proposed strategy is further used to accelerate IGA based nonlinear analysis of a strip peeling problem. The present work aims to utilize the computational power of GPU for FEA and IGA while keeping the development effort minimum by implementing it within the MatLab® environment. The outcome of this study is expected to help people in academic and industry accelerate their FEA based simulation code in MatLab®.

The paper is organised as follows. In next section, problem formulation is presented using IGA. Thereafter, the parallel implementation of FEM is explained along with the data structure. In second last section, results of the numerical experiments done to evaluate the performance of proposed strategy are presented. The last section concludes the paper.

## Problem Formulation

This section is divided into the two subsections. In the first, a continuum based formulation of adhesion model and its weak formulation is briefly overviewed. In the second, finite element and the NURBS based discretization of the continuum is presented.

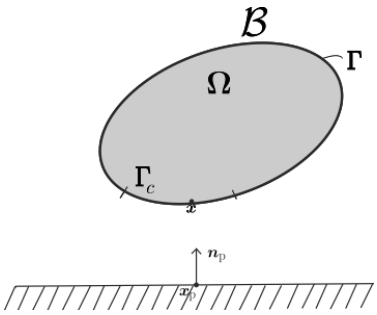



Fig. 1 Contact kinematics of a deformable body and a rigid substrate pair in their current configuration.

### *Model for Adhesive Contact Problem*

Consider a body $\mathcal{B}$ in its current configuration having domain $\Omega$, surface $\Gamma$, and the part of its boundary where contact takes places as $\Gamma_c$. The schematic arrangement of the body interacting with the rigid substrate is illustrated in Fig. 1. The governing equation for the quasi-static frictionless adhesive contact problem is given by the equilibrium between the work done by the internal, external, and contact forces. For the admissible variation of interaction potential

function $\delta\boldsymbol{\varphi}$, the weak form for the adhesive contact for a deformable body is given by the following statement [18]:

$$\int_{\Omega} \text{grad}(\delta\boldsymbol{\varphi}) : \boldsymbol{\sigma}\, \mathrm{d}\Omega - \int_{\Gamma_c} \delta\boldsymbol{\varphi} \cdot \boldsymbol{t}_c\, \mathrm{d}\Gamma - \delta\Pi_{\text{ext}} = 0, \quad \forall\, \delta\boldsymbol{\varphi} \in \mathcal{V} \qquad (1)$$

where $\mathcal{V}$ is the space of kinematically admissible variation function $\delta\boldsymbol{\varphi}$, $\boldsymbol{\sigma}$ is the Cauchy stress tensor, and $\boldsymbol{t}_c$ is the contact traction over $\boldsymbol{\Gamma}_c$. For the evaluation of contact tractions, penalty method based regularization of constitutive equation for the contact surface traction is used. First, based on the unique projection of slave points on the master surfaces, the unit normal $\mathbf{n}_{\boldsymbol{p}}$ and minimum gap $r_s$ between the contact surfaces are determined, see Fig. 1. Using the definition of the normal gap, the contact traction can be written as:

$$\boldsymbol{T}_{\text{c}}(\boldsymbol{x}_s) = \begin{cases} -\varepsilon_{\text{n}}\, r_s\, \mathbf{n}_p, & r_s < 0 \\ 0, & r_s \geq 0 \end{cases} \qquad (2)$$

where $\varepsilon_{\text{n}}$ represents the penalty parameter. For the van der Waals adhesion model and the contact traction $\boldsymbol{T}_{\mathbf{c}}$ is obtained by integrating the Lennard-Jones interaction potential four time [18] and is given as

$$\boldsymbol{T}_{\text{c}}(\mathbf{x}_{\text{s}}) = \boldsymbol{T}_c(r_s)\mathbf{n}_p,$$

$$\text{where} \quad \boldsymbol{T}_c(r_s) = \frac{A_H}{2\pi {r_o}^3}\left[\frac{1}{45}\left(\frac{r_o}{r_s}\right)^9 - \frac{1}{3}\left(\frac{r_o}{r_s}\right)^3\right] \cdot \mathbf{n}_p\,. \qquad (3)$$

Here, $A_H$, and $r_o$ denote the Hamaker's constant, and the equilibrium spacing of interacting particles of contacting bodies, respectively.

*FE discretized weak formulation*

Within the FEA, the domain of the body $\mathcal{B}$ is discretized into $n^e$ number of elements such that $\Omega = \sum_{e=1}^{n_e} \Omega^e$ and the displacement field $\boldsymbol{u}^e$, its variation $\delta\boldsymbol{u}^e$ for a standard finite element $\Omega^e$ is given by the summation of product of Lagrange basis function and field variables as

$$\boldsymbol{u}^e = \sum_{i=1}^{n_n} N_i \boldsymbol{u}_i = \mathbf{N}\mathbf{u}\,,\ \delta\boldsymbol{u}^e = \sum_{i=1}^{n_n} N_i \delta\boldsymbol{u}_{\boldsymbol{i}} = \mathbf{N}\delta\mathbf{u}, \qquad (4)$$

where $\boldsymbol{u}_i$ represents the displacement vector of node $i$, and $n_n$ denotes the total number of nodes in an element $\Omega^e$. $\mathbf{N}$ is the basis function matrix: $\mathbf{N} = \left[N_1\boldsymbol{I}, N_2\boldsymbol{I}, \ \ldots, N_{n_n}\boldsymbol{I}\right]$, where $N_i$ represents the Lagrangian basis function associated to node $i$, and $\boldsymbol{I}$ is the identity tensor in $\mathbb{R}^{\mathbf{2}}$. Following the Galerkin approach, the initial configuration $\boldsymbol{X}$, and current configuration $\boldsymbol{x}$ of body $\mathcal{B}$ are described in a likewise manner as in Eq. (4). In the context of IGA, NURBS basis functions used for the discretization of the geometry are employed for the determination of solution field. The displacement field $\boldsymbol{u}$, its variation $\delta\boldsymbol{u}$, and the current configuration of the geometry $\boldsymbol{x}$ is represented in terms of the NURBS basis functions $R_{i,p}(\xi,\eta)$, i.e. $N_i$ is replaced by $R_{i,p}(\xi,\eta)$ in Eq. (4). The reader is referred to [19] for the detailed description of the implantation of IGA into the finite element code structure. The discretized weak form Eq. (1) for the adhesive contact can be cast into the following matrix form [18]:

$$\delta\mathbf{u}^{\text{T}}[\mathbf{f}_{\text{int}} + \mathbf{f}_{\text{c}} - \mathbf{f}_{\text{ext}}] = \mathbf{0}, \quad \forall\ \ \delta\mathbf{u}^{\text{T}} \in \mathcal{V} \qquad (5)$$

where $\mathbf{f}_{\text{int}}$, $\mathbf{f}_{\text{c}}$, and $\mathbf{f}_{\text{ext}}$, are the vectors for internal, contact, and externally applied forces, respectively. Internal force vector is described by the constitutive relation of the material

model. In the present work, a Neo-Hookean hyperelastic material model is used and the Cauchy stress is determined by the following expression [20]

$$\boldsymbol{\sigma} = \frac{\lambda}{J}\ln J\boldsymbol{I} + \frac{\mu}{J}\left(\boldsymbol{F}\boldsymbol{F}^{T} - \boldsymbol{I}\right) \tag{6}$$

where $\lambda$ and $\mu$ are Lamé's constants, and $J$ denotes the determinant of deformation gradient tensor $\boldsymbol{F}$. The contact contribution $\mathbf{f}_{\mathrm{c}}$ over the contact surface $\boldsymbol{\Gamma}_c$ of an element $\Omega^e$ can be computed by the following expression

$$\mathbf{f}_{\mathrm{c}} = \sum_{e=1}^{n_e} \mathbf{f}_{\mathrm{c}}^{\mathrm{e}}, \qquad \text{where} \quad \mathbf{f}_{\mathrm{c}}^{\mathrm{e}} = -\int_{\boldsymbol{\Gamma}_c^e} \mathbf{N}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{c}}\, \mathrm{d}\Gamma. \tag{7}$$

**Parallel implementation of FEA**

The proposed strategy has been implemented entirely in MatLab® environment using the parallel computing toolbox. The parallel computing toolbox provides various ways to run code on the GPU. The simplest way is to use built-in function enhanced to work on GPU with `gpuArray` type of input data. Since the built-in functions are not always sufficient, we have written our own MatLab® functions and used them with function wrappers to implement our strategy on the GPU. The user-defined MatLab® functions can be used without any function wrappers but it may launch multiple CUDA kernels even for simpler function. The function wrapper like `arrayfun` compiles multiple operations of a function into single GPU kernel and therefore provides better performance. However, the GPU function wrappers have some restrictions. Only those user-defined functions that contain element wise operations can be used. The `arrayfun` wrapper allows the function to take arrays/matrix as input but indexing into the array is not allowed.

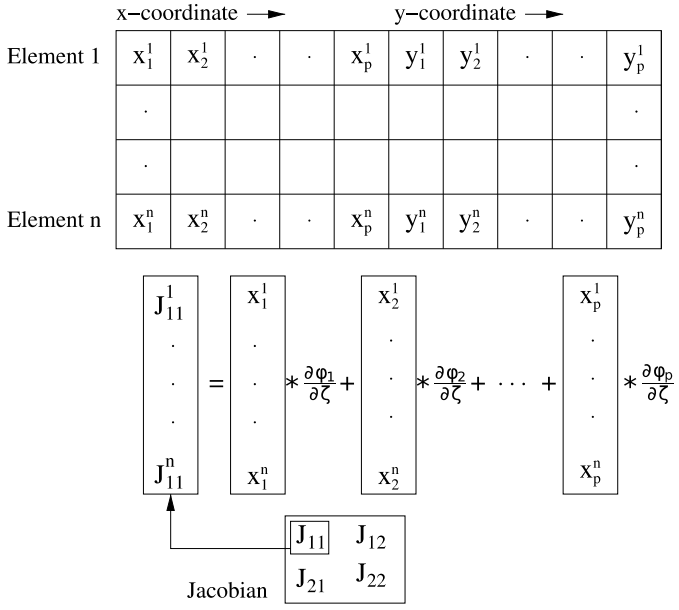



Fig. 2. Calculation of Jacobian.

Vectorization of code is critical to achieve better performance by the GPU parallelization. Our proposed strategy is based on a vectorization scheme in which computation at the elemental level with scalar variable is converted into computation at mesh level with arrays. It

enables us to easily implement our strategy by functions with element wise operations and use them with `arrayfun` function wrapper. In our implementation, the numerical integration is done by looping over the Gauss points. For each Gauss point, Jacobian is calculated for transformation to the reference coordinates. Since we are doing calculation at the mesh level, Jacobian is calculated for each element of the mesh simultaneously. The data structure and procedure for Jacobian calculation is shown in Fig 2. Here, nodal coordinates are reordered and stored in the matrix with each column containing an individual coordinate for all elements. The derivative of the basis function in reference coordinate is pre-computed for all the Gauss points and reordered to facilitate the computation of Jacobian. As shown in Fig. 2, the calculation of an entry of Jacobian is done by multiplying coordinate values with basis function derivatives for a particular Gauss point. This evaluates to an array that contains an entry of Jacobian matrix for all the elements of the mesh. The other entries of the Jacobian matrix are calculated in the similar way. The computation of determinant for all the element of the mesh can be done simply by element wise operations over array of entries of Jacobian matrix. The computation of inverse of Jacobian is done by co-factor calculation. This can again be done in parallel for all the Jacobian matrices by element wise operations.

The inverse of Jacobian is multiplied with derivative of shape function in reference coordinates to calculate shape function derivative in physical coordinates. Since we are working with arrays, the above product produces shape function derivative for all the elements of the mesh. To facilitate the computation of element stiffness matrix, the derivative of shape function is stored as shown in Fig. 3. Each column of the matrix contains derivative of particular shape function for all the elements of the mesh. The evaluation of element stiffness matrix is done by computing each individual entry of the matrix simultaneously for

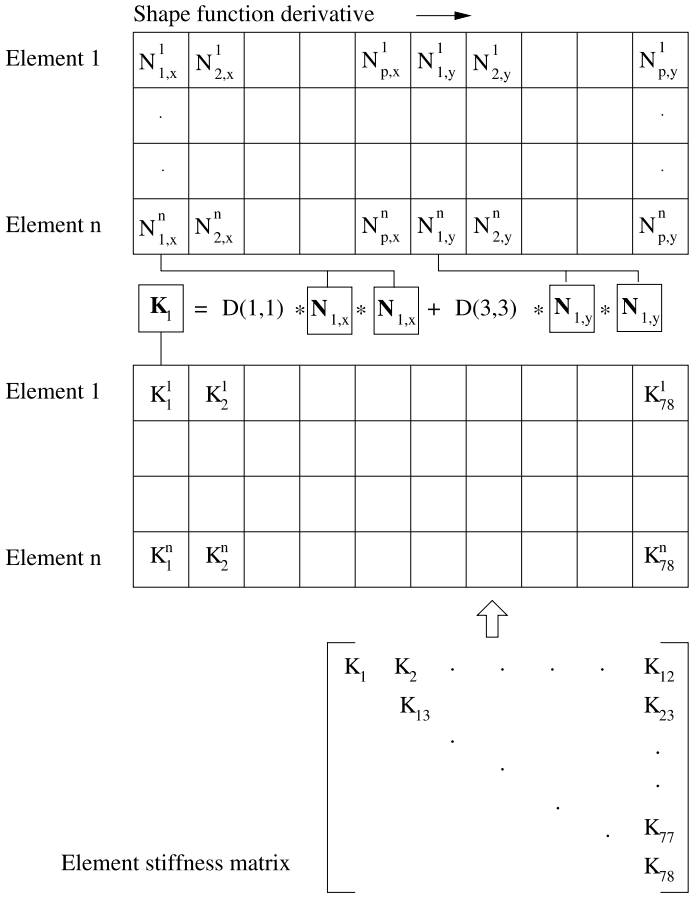



Fig. 3. Calculation of element stiffness matrix.

all the elements. The calculation and storage pattern is shown in Fig. 3. Since the elemental matrix is symmetric, we calculate only the unique entries.

All the matrices involved in the computation are stored in GPU memory. The assembly to global matrix is done by `sparse` function of the MatLab®. The off-diagonal entries are assembled first, so that it can be transposed to generate the symmetric part of the matrix. The on-diagonal entries are added later to complete the global matrix. The `sparse` function requires the row and column indices of the values to be assembled. This can be calculated beforehand using mesh connectivity. We pre-compute the row and column indices and reorder them according to the storage arrangement of element stiffness matrix. The global matrix is assembled on GPU.

**Results and Discussion**

To evaluate the performance of the proposed strategy two different problems are solved. The first one is linear elastic two dimensional (2D) cantilever beam problem with concentrated load at the tip and the second one is a strip peeling problem. The performance is compared among a CPU sequential, CPU parallel, CPU vectorized and GPU implementation. The CPU sequential approach uses a loop over each element of the mesh for element matrix computation and assembly. The CPU parallel implementation uses `parfor` construct of the parallel computing toolbox to utilize multiple processors on the CPU for parallel computation of element matrix computation and assembly. The CPU vectorized implementation is based on the proposed vectorization scheme but uses the CPU for computation.

The machine used for the numerical experiment consists of Intel Xeon® E5-2650 processor having 2.2GHz of clock speed and a NVIDIA Tesla K40c GPU with 2880 cores clocked at 745MHz. The proposed strategy has been implemented on MatLab® R2016a using the parallel computing toolbox.

*2D Cantilever beam*

A 2D cantilever beam with concentrated load at tip is taken as shown in Fig. 4. The geometric and material parameters are given as: length (L) – 10 m, breadth (B) – 1 m, Young's modulus (E) - 210 GPa, Poisson's ratio (ν) - 0.3 and end load (P) - $10^5$ N. Linear quadrilateral element with two degrees of freedom (DOF) per node is used to discretize the domain.

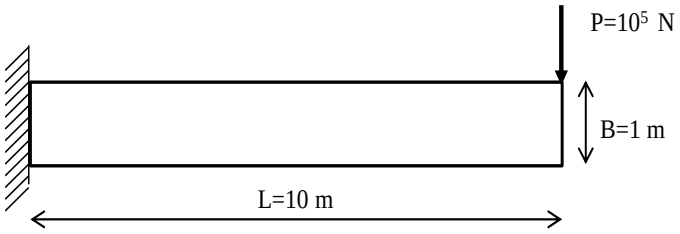



**Fig. 4**. A 2D cantilever beam with end load.

Table 1. Finite element mesh.

| Mesh | Elements | Nodes | Degrees of freedom |
|---|---|---|---|
| Mesh 1 | 900 000 | 903 301 | 1 806 602 |
| Mesh 2 | 1 600 000 | 1 604 401 | 3 208 802 |
| Mesh 3 | 2 500 000 | 2 505 501 | 5 011 002 |
| Mesh 4 | 3 600 000 | 3 606 601 | 7 213 202 |

The problem is solved for different level of mesh refinement keeping the aspect ratio of the elements same. The finite element mesh with varying level of refinement is shown in Table 1. Structured mesh has been used for the purpose of numerical experiment but the calculation is performed for each of the elements treating them as unstructured. Fig. 5 shows the comparison of numerical integration and assembly time as a function of mesh sizes. It can be observed that the proposed GPU implementation obtains least timings for all of the mesh sizes. The CPU vectorized implementation achieves significantly less time than CPU parallel strategy (using 12 workers), highlighting the effectiveness of the proposed vectorization scheme. The speedups of the GPU implementation over all other implementations are shown in Fig. 6. A maximum of 41.4× (Mesh 1) speedup is obtained over CPU sequential and 10.8× (Mesh 1) over CPU parallel implementations. The GPU code could achieve speedup of only 3.6× - 3.8× over CPU vectorized implementation, which shows that the proposed vectorization of FEM is able to scale very effectively also on the CPU.

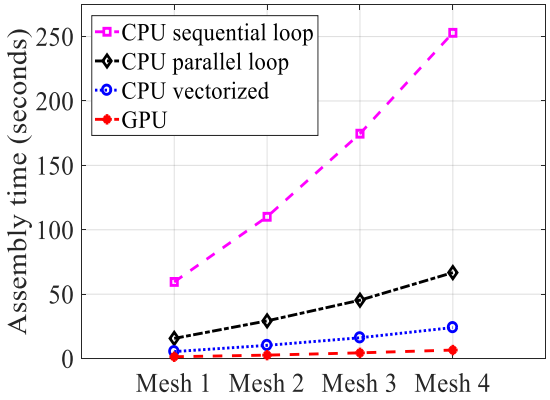



**Fig. 5.** Numerical integration and assembly time.

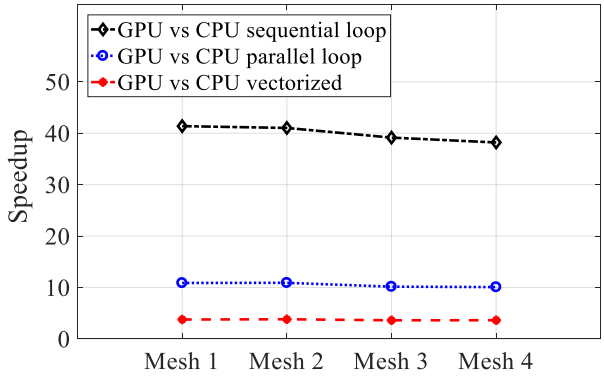



**Fig. 6.** Speedup in numerical integration and assembly

The linear system of equations given by FEM can be solved by either direct sparse solver or by iterative solvers. The mldivide(\) operator in MatLab® implicitly chooses the best algorithm (direct solver) depending upon the type of input matrix. On CPU, mldivide performs much better than the iterative sparse solvers. Since mldivide operator is not supported on GPU for sparse matrices, we compared mldivide on CPU with iterative solver on the GPU. For Mesh 3 in Table 1, the pcg (preconditioned conjugate gradient) function takes 295 seconds on GPU, whereas mldivide takes 25.5 seconds to solve the system of equations on CPU. This prompts us to adopt a strategy in which assembly, numerical integration is done on GPU, and solution of linear system of equations takes place on CPU by the mldivide operator. We call this as GPU + CPU strategy. Fig. 7 shows the comparison of overall execution time. The GPU + CPU strategy achieves 2.2× speedup over CPU parallel strategy, 5.6× speedup over sequential strategy and 1.4× over CPU vectorized strategy for the finest mesh.

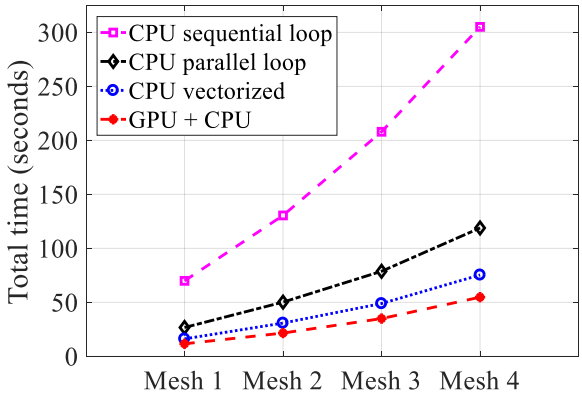



Fig. 7. Overall execution time for cantilever beam problem.

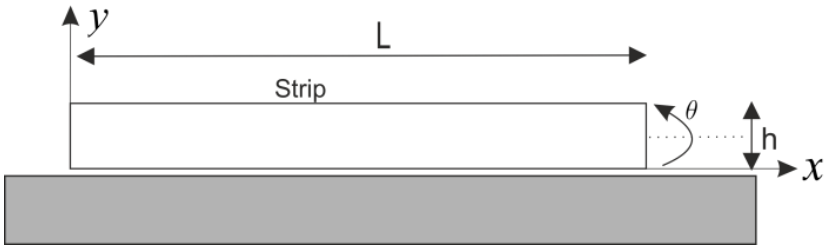



Fig. 8. The geometrical set-up of peeling problem.

*Strip peeling problem*

We consider the peeling of a deformable strip (having length $L = 200Lo$, height $h = 10Lo$ with $Lo = 1$ nm) adhering to a flat, rigid substrate. The schematic arrangement of this problem is shown in Fig. 8. An isotropic, nonlinear Neo-Hookean material model with $E = 1$ GPa, and $\nu = 0.2$, under plane strain conditions is used. It is considered that the adhesive contact forces are present at the 75% of the bottom length of strip (from $x = 0$ to $x = 150$ Lo) and are calculated using Eq. (3) with $r_o = 0.4$ nm and $A_H = 10^{-19}$ J. At one end, a rotation angle $\theta$ is applied in such a manner that it yields a constant moment during the peeling process and a rotation step size $\Delta\theta = 0.1^o$ is chosen for the simulation. The strip is discretized with 240 x 12, 320 x 16, 400 x 20, 480 x 24, 640 x 32, and 720 x 36 number of elements along each direction and corresponding discretization is referred as mesh 1, Mesh 2, Mesh 3, Mesh 4, Mesh 5, and Mesh 6, respectively.

In Fig. 9, the comparison of numerical integration and assembly time is done. A considerable amount of reduction in assembly timings can be observed for the proposed strategies. For Mesh 6, the integration and assembly time reduces from 21876 seconds to 1780 seconds. The CPU vectorized code achieves speedup in the range 5.1× - 4.1× over CPU parallel strategy. The GPU based strategy achieves speedup in the range 2.4× - 12.3× over CPU parallel and 0.47× - 2.9× over CPU vectorized. The device set up time and data communication time dominates in GPU +CPU strategy for smaller mesh size which results into inferior performance compared to CPU vectorized for Mesh 1 and Mesh 2. When the mesh size increases, effectiveness of GPU becomes more apparent. The comparison of total execution time is shown in Fig. 10. The GPU + CPU strategy takes the least amount of time and reduces the total simulation time from 49343.3 seconds to 28452 seconds for Mesh 6, which is remarkable.

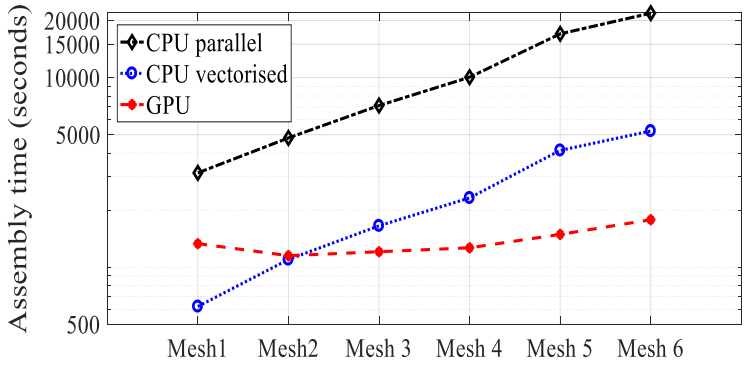



Fig. 9. Numerical integration and assembly time for strip peeling problem.

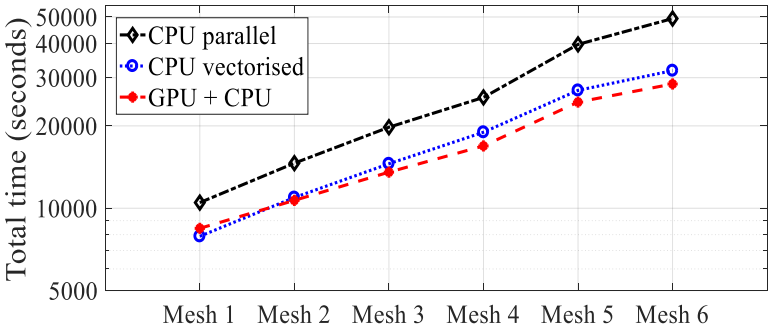



**Fig. 10.** Total execution time for strip peeling problem.

## Conclusions

In this work, a novel vectorization strategy is presented to implement FEA and IGA on GPU using MatLab® environment. The proposed vectorization strategy accelerates elemental tangent matrix evaluation and their assembly by performing the required computation at mesh level rather than at element level. For linear elastic cantilever beam problem, the GPU based strategy is found to be 10.1× faster than CPU parallel (parfor) and 38.2× faster than sequential strategy in numerical integration and assembly for 7.2 million DOF. For strip peeling problem, a maximum speedup of 12.3× is achieved over CPU parallel strategy in numerical integration and assembly. This leads to total reduction in overall simulation time from 13.7 hours to 7.9 hours. Based on numerical simulation done in this paper, we find that a GPU is very effective in accelerating the evaluation of element tangent matrix and their assembly to global matrix in MatLab®. In future, we would like to extend this work to accelerate evaluation of contact forces on GPU.

**References**


[1] Bathe, K. J. (1996) *Finite element procedures*, Prentice Hall of India, New Delhi, India.

[2] Hughes, T.J.R., Cottrell, J. A. and Bazilevs, Y. (2005) Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer methods in applied mechanics and engineering* **194,** 4135-4195.

[3] Cottrell, J. A., Hughes, T. J. R. and Bazilevs, Y. (2009) *Isogeometric Analysis: Toward Integration of CAD and FEA*, Wiley.

[4] Mittal, S. and Vetter, J. S. (2015) A survey of CPU-GPU heterogeneous computing techniques, *ACM Computing Surveys (CSUR)* **47,** 69.

[5] Huang, S., Xiao, S. and Feng, W., On the energy efficiency of graphics processing units for scientific computing, *In Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, IEEE 2009, 1-8.

[6] NVIDIA, *Popular GPU-accelerated Applications Catalog* 2017.

[7] Martin, J. and Sharma, G. (2009) MATLAB: A Language for parallel computing, *International Journal of Parallel Programming* **37,** 3-36.

[8] Hosagrahara, V., Tamminana, K. and Sharma, G., Accelerating Finite Element Analysis in MATLAB with Parallel Computing, *The MathWorks News and Notes*, 2010. https://www.mathworks.com/tagteam/66859_91826v00_FEM_final.pdf.

[9] Cuvelier, F. Japhet, C. and Scarella, G. (2016) An efficient way to assemble finite element matrices in vector languages, *BIT Numerical Mathematics* **56**, 833-864.

[10] Simkus, A. and Turskiene, S. (2013) Parallel computing for the finite element method in MATLAB, *Computational Science and Techniques* **1**, 214-221.

[11] Cecka, C., Lew, A. J. and Darve, E. (2011) Assembly of finite element methods on graphics processor, *International Journal of Numerical Methods in Engineering* **85,** 640-669.

[12] Zhang, J. and Shen, D., GPU based implementation of finite element method for elasticity using CUDA, *in High Performance Computing and Communications 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC EUC), 10th International Conference on IEEE*, 2013, 1003-1008.

[13] Karatarakis, A., Karakitsios, P. and Papadrakakis, M. (2014) GPU accelerated computation of the isogeometric analysis stiffness matrix, *Computer Methods in Applied Mechanics and Engineering* **269**, 334-355.

[14] Woźniak, M. (2015) Fast GPU integration algorithm for isogeometric finite element method solvers using task dependency graphs, *Journal of Computational Science* **11**, 145-152.

[15] Georgescu, S., Chow, P. and Okuda, H. (2013) GPU acceleration for FEM-based structural analysis, *Archives of Computational Methods in Engineering* **20**, 111-121.

[16] Elman, H., Sylvester, D. and Wathen, A. (2014) *Finite elements and fast iterative solvers*, Second Ed. Oxford University Press, Oxford, UK.

[17] Jung, J. H. and Bae, D. S. (2017) An implementation of direct linear equation solver using a many-core CPU for mechanical dynamic analysis, *Journal of Mechanical Science and Technology* **31**, 4637-4645.

[18] Sauer, R. A., and Li, S. (2007) An atomic interaction-based continuum model for adhesive contact mechanics, *Finite Elements in Analysis and Design* **43**, 384-396.

[19] Agrawal, V., and Gautam, S. S., IGA: A Simplified Introduction and Implementation Details for Finite Element Users. *Journal of The Institution of Engineers (India): Series C* (https://doi.org/10.1007/s40032-018-0462-6).

[20] Bonet, J. and Wood, R. (2008) *Nonlinear Continuum Mechanics for Finite Element Analysis*, Cambridge University Press, Cambridge. (doi:10.1017/CBO9780511755446)

# Improving the Accuracy of Stress Intensity Factors Obtained by the Scaled Boundary Finite Element Method on Hierarchical Meshes

***Adrian Egger[1], Savvas Triantafyllou[2], and Eleni Chatzi[1]**

[1]Institute of Structural Engineering, ETH Zurich, Switzerland
[2]Centre for Structural Engineering and Informatics, The University of Nottingham, UK

*Presenting and corresponding author: egger@ibk.baug.ethz.ch

## Abstract

In this work, the unique properties of the scaled boundary finite element method (SBFEM), a semi-analytical numerical method, which introduces a scaling center in each element's domain, are exploited to improve the accuracy of computed generalized stress intensity factors (gSIFs) on hybrid balanced quadtree-polygon (QT) meshes. The gSIFs are extracted by harnessing the semi-analytical solution in radial direction. This is achieved by placing the scaling center of the element containing the crack at the crack tip. Taking an analytic limit of this element's stress field as it approaches the crack tip, delivers an expression for the singular stress field. By applying the problem specific boundary conditions, the geometry correction factor is obtained and the gSIFs are then evaluated based on their formal definition.
Computation of the gSIFs by SBFEM permits exceptional accuracy, however, when combined with QT meshes employing linear elements, this does not always hold. Nevertheless, it has been shown that crack propagation schemes are highly effective even given very coarse discretizations, utilizing criteria than only rely on the ratio of mode I to mode II gSIFs. The absolute values of the gSIFs may still be subject to undesirable errors. Hence, we propose a post-processing scheme, which minimizes the error resulting from the approximation space of the cracked element. Thus, the errors in the calculation of the gSIFs is only limited to the discretization error of the quadtree mesh. This is achieved by h- and/or p-refinement of the cracked element, which elevates the amount of modes present in the solution. The resulting numerical description of the element is highly accurate, with the main error source now stemming from its boundary displacement solution. Numerical examples show, that this post-processing procedure can significantly improve the accuracy of the computed gSIFs with negligible computational cost even on coarse meshes resulting from QT decompositions.

**Keywords:** SBFEM, hierarchial meshes, gSIFs, crack propagation, LEFM.

## Introduction

The need for robust, economical and high-fidelity numerical methods to simulate failure processes in linear elastic fracture mechanics (LEFM) arises as as consequence of sustainable design, mandating lighter, stronger and more resilient structures in, e.g., the aerospace, automotive, and construction industries [7,21,23]. To this end, several numerical methods have been investigated. Some more notable representatives include: The finite element method, boundary element method, extended finite element method, material point method, meshless methods, peridynamics and recently also the scaled boundary

finite element method, which all consider a discrete crack representation. As of late, methods considering diffuse crack representations, e.g., variants of the phase field and thick level set method have been gaining traction, due to their ability to incorporate complex crack behaviour such as nucleation, propagation, branching, merging and arrest [12, 13]. For a comprehensive treatment, we kindly refer the reader to [4].
Although currently the extended finite element method (XFEM) [8] is arguably the most popular method, pending widespread adoption in industry and academia alike, SBFEM [17] offers similar capabilities and additional benefits in LEFM scenarios. SBFEM, which introduces a scaling center within the element domain and, where applicable, at the crack tip, permits an analytic solution in radial direction and thus the gSIFs can be efficiently and effectively evaluated in post-processing as the limit of the singular stress field as one approaches the crack tip [16, 18]. The benefits include the accurate determination of gSIFs at bi-material notches [2] and the fact the no a priori knowledge is required of the order of singularity. Further, SBFEM's polygon underpinning permits direct integration with QT meshes, which eliminates issues with hanging nodes, results in a limited amount of precomputable elements realizations and offers a level of adaptivity around domain features [1, 9, 11, 22].
However, the accurate calculation of gSIFs requires sufficient angular resolution of the singular stress field. This issue is traditionally circumvented by refining an area around the crack tip with subsequent homogenization into a cracked macro element. In conjunction with the balancing operation performed on the QT meshes, this introduces a significant amount of degrees of freedom (DOFs) thus increasing the computational toll of analysis. Although the direction of crack propagation is accurately determined by the ratio of gSIFs, the load-deflection curve can exhibit significant errors. Hence, we aim to increase the accuracy of the calculated gSIFs, utilizing the same global mesh as previously for analysis. This is achieved by enforcing the external boundaries of the cracked element to be compatible with the linear shape functions, yet internally permit the use of arbitrary higher order approximations to model the singular stress field.
This remainder of the paper is structured as follows: First the problem statement and the pertinent theory are summarized. Next, the proposed method is detailed. Subsequently, two numerical examples are investigated. They serve to validate the proposed method and form the basis for the discussion. Finally, conclusions are drawn.

## Problem Statement and Pertinent Theory

*Problem Statement*

To formulate the LEFM problem, we consider a two dimensional cracked domain $\Omega$ (Fig. 1). The boundary $\Gamma = \Gamma_0 \cup \Gamma_u \cup \Gamma_t \cup \Gamma_c$ comprises regions of free surface conditions ($\Gamma_0$), prescribed displacements $\bar{u}$ ($\Gamma_u$) and prescribed traction $\bar{t}$ ($\Gamma_t$), where applicable. The strong form with associated boundary conditions follows as:

$$\nabla \cdot \{\sigma\} + \{b\} = \{0\} \quad \text{in} \quad \Omega \tag{1a}$$
$$\{u\} = \{\bar{u}\} \quad \text{on} \quad \Gamma_u \tag{1b}$$
$$\{\sigma\} \cdot \{n\} = \{\bar{t}\} \quad \text{on} \quad \Gamma_t \tag{1c}$$
$$\{\sigma\} \cdot \{n\} = \{0\} \quad \text{on} \quad \Gamma_c^0 \tag{1d}$$

where $\{\sigma\}$ denotes the Cauchy stress tensor, $\{n\}$ the unit outward normal to the boundary, $\{b\}$ the applied body force per unit volume, $\{u\}$ the displacement field and $\nabla$ the gradient operator.

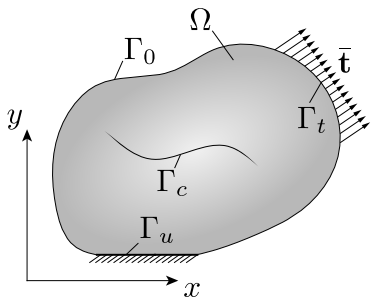



**Figure 1: Cracked Body and boundary conditions.**

The stress $\{\sigma\}$ and strain field $\{\epsilon\}$, given small deformations and linear elastic material behaviour, follow from the modulus of elasticity $E$ and Poisson ratio $\nu$:

$$\{\epsilon\} = \nabla_s\{u\} \quad \text{and} \quad \{\sigma\} = [D]\{\epsilon\} \tag{2}$$

for which $\nabla_s$ is the 2D symmetric gradient operator and $[D]$ the elasticity tensor:

$$[D] = \frac{E}{1-\nu^2}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix}, \quad \text{for plane stress.} \tag{3}$$

*Summary of SBFEM Theory*

SBFEM [17] sets itself apart from other numerical methods by the introduction of a scaling center. Only one scaling center may be present per polygonal element, termed a subdomain. One notable restriction is imposed: The complete boundary must be visible from the scaling center, i.e., star-convexity. A new polar-like reference system is introduced with radial coordinate $\xi$ and local tangential coordinate $\eta$ (Fig. 2).

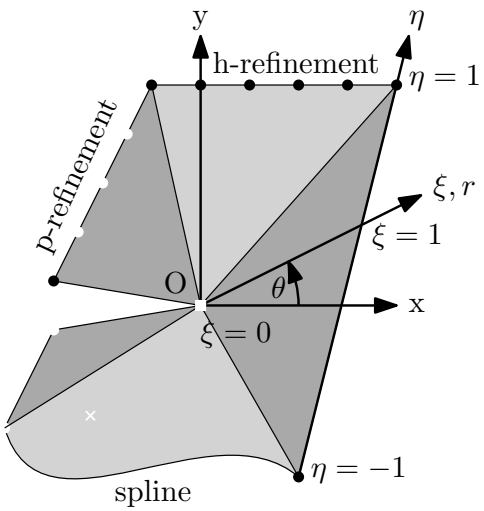



**Figure 2: Subdomain with scaled boundary coordinates $\xi$ and $\eta$.**

The radial coordinate, with origin at the scaling center 0 and a value of 1 on the boundary such that $0 \leq \xi \leq 1$ is kept analytic throughout the analysis. This reduces the dimensionality of the problem by one. In 2D only the boundary remains, which is discretized in the conventional finite element sense, into independent line elements. Each line element possesses its own natural coordinate $-1 < \eta < 1$ in tangential direction. Mapping between

Cartesian $(x, y)$ and scaled boundary coordinates $(x(\xi,\eta), y(\xi,\eta))$ requires an analytic, associated with $\xi$, and interpolatory, associated with $\eta$, component: Any point within the domain can be expressed by scaling of a point $\{x_b\}, \{y_b\}$ on the boundary by a fraction of $\xi$ in radial direction and an interpolation using the conventional finite element shape function $[N(\eta)]$ along the natural coordinate.

$$x(\xi,\eta) = \xi x_b(\eta) = \xi[N(\eta)]\{x_b\} \tag{4}$$
$$y(\xi,\eta) = \xi y_b(\eta) = \xi[N(\eta)]\{y_b\} \tag{5}$$

An iso-parametric representation is adopted and the displacements are analogously mapped:

$$\{u(\xi,\eta)\} = [N_1(\eta)[I], ..., N_n(\eta)[I]]\{u(\xi)\} \tag{6}$$

The amount of degrees of freedom (DOFs) per line element is denoted by $n$, while $[I]$ is a $2x2$ identity matrix and $\{u(\xi)\}$ represent nodal displacement functions in radial direction, which need to be determined by SBFEM analysis.
Applying the transformation to scaled boundary coordinates, the stresses may be written as [19]:

$$\{\sigma(\xi,\eta)\} = [D]([B^1(\eta)]\{u(\xi)\}_{,\xi} + [B^2(\eta)]\{u(\xi)\}/\xi) \tag{7}$$

where the strain-displacement relation is described by partitioning the linear operator into $[B^1(\eta)]$ and $[B^2(\eta)]$ [17]. Deriving the weak form involves recasting the strong form in scaled boundary coordinates and applying the standard techniques in tangential direction [3, 5, 20]. This gives rise to two equations:

$$[E^0]\xi^2\{u(\xi)\}_{,\xi\xi} + ([E^0] - [E^1] + [E^1]^T)\xi\{u(\xi)\}_{,\xi} - [E^2]\{u(\xi)\} = 0 \tag{8}$$
$$\{P\} = [E^0]\xi\{u\}_{,\xi} + [E^1]^T\{u\} \tag{9}$$

The scaled boundary finite element equation in displacements (Eqn. 8) governs the system response within the domain, while Eqn. 9 defines the behavior on the boundary. Three coefficient matrices $[E^0], [E^1], [E^2]$ arise, which bear strong similarity to stiffness matrices in conventional FEM: They are calculated element-wise with subsequent assembly on subdomain level. The vector or nodal forces and displacements on the boundary are given as $\{P\}$ and $\{u\} = \{u(\xi = 1)\}$ respectively. Typically, the solution to the set of homogeneous Euler-Cauchy differential equations is constructed as a power series:

$$\{u(\xi)\} = [\Psi^{(u)}]\xi^{-[S]}\{c\} = \sum_{i=1}^{n}[\Psi_i^{(u)}]\xi^{-[S_i]}\{c_i\} \tag{10}$$

The transformation matrix $[\Psi]$ and block diagonal real Schur form $[S]$ arise from rewritting the quadratic eigen-problem described by combining Eqns. 8 and 9 as a system of first order differential equations:

$$\xi \begin{Bmatrix} \{u(\xi)\} \\ \{q(\xi)\} \end{Bmatrix}_{,\xi} = -[Z] \begin{Bmatrix} \{u(\xi)\} \\ \{q(\xi)\} \end{Bmatrix} \tag{11}$$

with the Hamiltonian coefficient matrix Z defined by

$$Z = \begin{bmatrix} [E^0]^{-1}[E^1]^T & -[E^0]^{-1} \\ -[E^2] + [E^1][E^0]^{-1}[E^1]^T & -[E^1][E^0]^{-1} \end{bmatrix} \tag{12}$$

It can be shown that Eqn. 11 decouples the block-diagonal Schur decomposition [15].

$$[Z][\Psi] = [\Psi][S] \tag{13}$$

The modes comprise the columns of the transformation matrix, while the corresponding eigenvalues are contained within the diagonal blocks of the real Schur form. Having doubled the problem size by considering a system of first order differential equations, the bounded response is striped from the unbounded one by sorting $[S]$ and $[\Psi]$ and partitioning according to sign:

$$[S] = diag([S_{neg}], [S_{pos}]) \tag{14}$$

$$[\Psi] = \begin{bmatrix} [\Psi_{neg}^{(u)} & \Psi_{pos}^{(u)}] \\ [\Psi_{neg}^{(q)} & \Psi_{pos}^{(q)}] \end{bmatrix} \tag{15}$$

Substituting the bounded component of the displacement solution (Eqn. 10) into the equation governing the boundary response (Eqn. 9), the formulation of the subdomain stiffness matrix arises, which permits the calculation of the displacement field using conventional FEM methods.

$$K_{bounded} = [\Psi_{pos}^{(q)}][\Psi_{neg}^{(u)}]^{-1} \tag{16}$$

The final expression of the stresses is obtained by substituting Eqn. 10 into Eqn. 7:

$$\{\sigma(\xi, \eta)\} = \sum_{i=1}^{n} [\Psi_{\sigma i}(\eta)] \xi^{-[S_i]-[I]} \{c_i\}] \tag{17}$$

where stress mode $[\Psi_{\sigma i}(\eta)]$ arise from its corresponding displacement mode $[\Psi_i^{(u)}]$:

$$[\Psi_{\sigma i}(\eta)] = [D](-[B^1(\eta)][\Psi_i^{(u)}][S_i] + [B^2(\eta)][\Psi_i^{(u)}]) \tag{18}$$

*Generalized Stress Intensity Factors*

Since the general solution to the SBFEM equation is constructed from a power series, the singular modes are readily identified: By inspection of $S_i$, any $-1 < real(\lambda) < 0$ will result in a singularity at $\xi = 0$. By placing the scaling center at a crack tip, this feature is exploited in calculating the gSIFs (Fig. 2). By including a double node at the crack mouth, two additional modes, i.e., the singular modes, arise, whose eigen-vectors resemble the mode I and mode II fracture cases. The singular stress field is extracted from the general solution (Eqn. 17), where the superscript $^{(s)}$ denotes the singular quantities:

$$\{\sigma^{(s)}(\xi, \eta)\} = [\Psi_{\sigma}^{(s)}(\eta)] \xi^{-([S^{(s)}]-[I])} \{c^{(s)}\} \tag{19}$$

Only the components $\{\sigma^{(s)}(r, \theta)\} = (\sigma_{\theta}^{(s)}(r, \theta), \tau_{r\theta}^{(s)}(r, \theta))^T$ are retained, which correspond to mode I and II cracks.

$$\begin{Bmatrix} \sigma_{\theta}^{(s)}(r, \theta) \\ \tau_{r\theta}^{(s)}(r, \theta) \end{Bmatrix} = \frac{1}{\sqrt{2\pi L}} \xi^{-[\tilde{S}^{(s)}(\theta)]} \begin{Bmatrix} K_I(\theta) \\ K_{II}(\theta) \end{Bmatrix} \tag{20}$$

Comparing Eqn. 20 to the gSIFs formal definition [18] permits their evaluation as:

$$\begin{Bmatrix} K_I(\theta) \\ K_{II}(\theta) \end{Bmatrix} = \sqrt{2\pi L} \{\sigma^{(s)}\} \tag{21}$$

*Hierarchical Meshes*

Hierarchical meshes arising from quadtree decompositions traditionally suffer from hanging nodes, whose treatment typically incurs computational, algorithmic or implementational overhead. Leveraging the polygon-underpinnings of SBFEM alleviates all issues commonly associated with hanging nodes, resulting in analysis ready meshes. To avoid irregularity of the mesh, a balancing operation is performed, which limits the amount of unique subdomain realization to 16 and enables precomputation. Hence, this approach garners considerable attention [1, 10, 11].
Strong and weak discontinuities are introduced by clipping the QT mesh [9]. Contrary to the XFEM, double nodes are introduced. Crack tips require special treatment: A double node is introduced where the crack enters an element and the scaling center is placed at the crack tip (Fig. 2). Crack tips do not require discretization, however, they are assumed straight. For crack propagation calculations, sufficiently accurate SIFs are required. Unfortunately, the elements typically encountered on QT meshes do not permit sufficient resolution of the singular stress field. Hence, a region around the crack tip is first locally refined and then homogenized (Fig. 3, bottom left). Imposing a suitable criteria to determine the critical crack propagation angle in conjunction with a user specified crack propagation increment $\Delta a$, determines the crack tip location in the subsequent step (Fig. 3, bottom right). Due to the balancing operation, the homogenization step affects large portions of the mesh, introduces many spurious DOFs in the process and forces a substantial system update across iterations. The increased accuracy with which the SIFs are calculated is attributed to two affects: Mesh refinement about the singularity and improving the approximation space of the crack tip element.

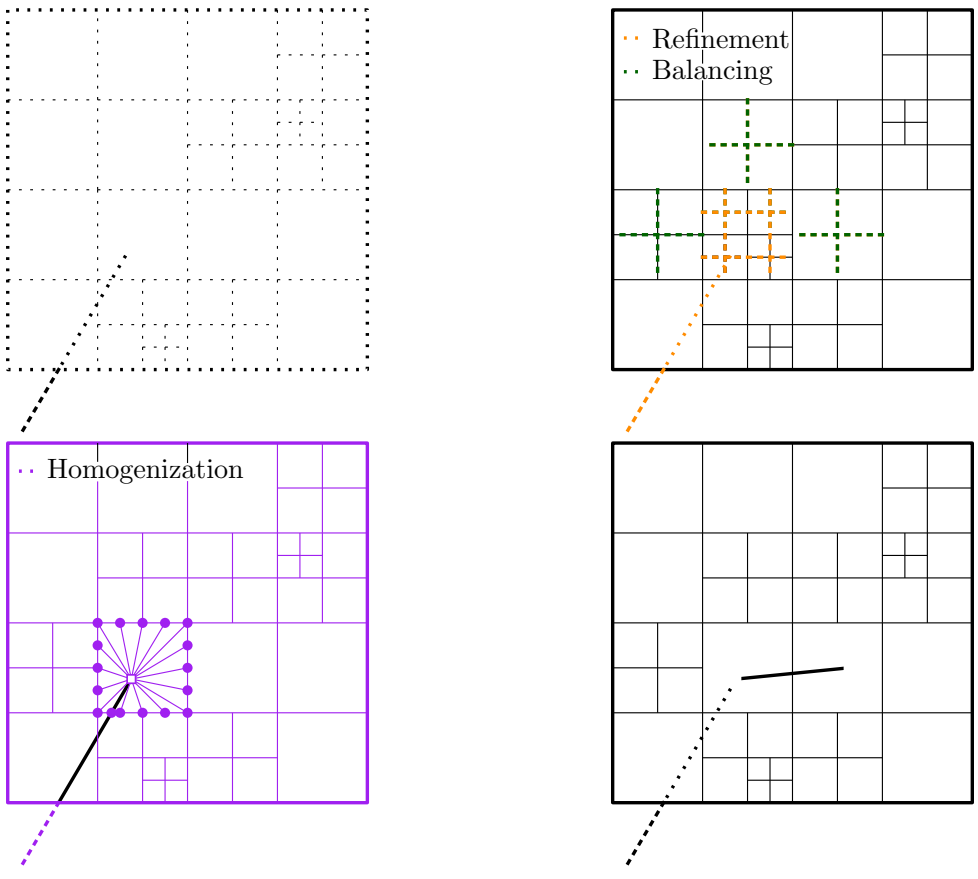



**Figure 3: Stages in crack propagation by SBFEM on QT meshes.**

**Proposed Method**

By inspecting the expression of the singular stresses (Eqn. 19), the accuracy of the gSIFs (Eqn. 21) can be improved in two ways:

1. Finding a better approximation for the integration constants $[c]$ stemming from the displacement solution of the domain.

2. Enhancing the displacement and thus stress field approximation $[\Psi]$ and $[S]$ within the cracked element by escalating the amount of DOFs present.

The contribution of the first source can be adjusted by h- or p-refinement on the mesh, either locally or globally. However, this is a costly procedure due to, e.g., remeshing, reassembly followed by solution, in the understanding that a significant amount of DOFs are newly introduced. Considering the convergence rate of the linear elements employed, this is deemed a sub-optimal approach from a computational resources standpoint. The second source allows for enhancement directly in post-processing and permits certain insight into how close we might be to the exact solution, by contrasting the gSIFs calculated using the traditional approach to those using the improved scheme.
The steps comprising the improved scheme are:

1. **Perform analysis**
   A displacement solution for the given mesh is sought.

2. **Create a high order cracked element**
   The existing cracked element is replaced by one elevated by h- and/or p-refinement. New nodes, lying in-between existing QT mesh nodes, are introduced in the process.

3. **Impose linearized displacements on all DOFs**
   To guarantee compatibility with the surrounding QT mesh, linear displacements are enforced between its nodes, on the boundary of the elevated cracked element.

4. **Back calculate $[c]$ for cracked element**
   Since the cracked element comprises one subdomain, with all DOFs situated on the boundary, where the prescribed displacements are imposed, the calculation of the stiffness matrix is not necessary. Only the block-diagonal Schur decomposition is required. The integration constants are obtained as $[c] = [\Psi^{(u)}_{neg}]^{-1}\{u\}$ [17]. Typically, refined crack elements comprising 100 DOFs have sufficiently converged. Since the original already contains 10-30 DOFs, this does not noticeably impact computation time.

5. **Extract singular stress modes**
   The identification and extraction of the singular stress modes remains unchanged.

6. **Calculate gSIFs**
   The gSIFs are calculated given Eqn. 21, however, now based on the quantities originating from the elevated cracked element.

7. **Contrast original with improved gSIFs**
   By contrasting the values for the gSIFs calculated by both methods, we can gauge the quality of the original analysis. If the discrepancy in gSIFs is deemed too large, refinement of the original mesh may be warranted.

## Numerical Examples

*Edge cracked square plate in mode II*

An edge cracked square plate subjected to a plane stress state is examined (Fig. 4). Along the boundary, the analytic solution [6] of the near-tip crack field is prescribed given $K_{II} = 1$ (Tab. 1), imposing pure mode II loading. The material properties are $E = 200$ [N/mm$^2$], $v = 0.3$ and the side length is $L = 1$ [mm].

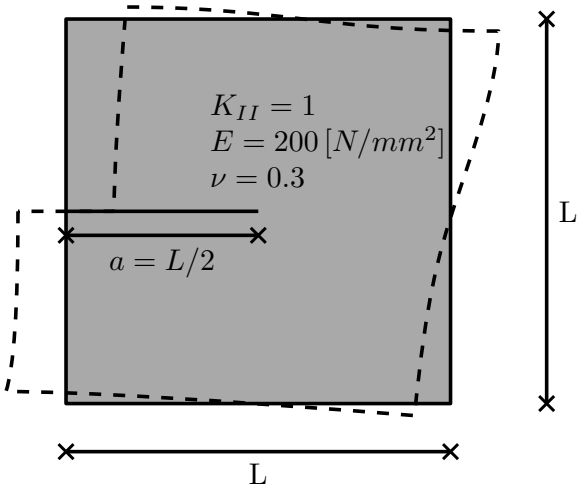



**Figure 4: Experimental setup for numerical example A.**

**Table 1: Analytic solution of the near-tip crack field.**

| | Exact displacement solution for mode II fracture |
|---|---|
| $u_x$ | $\frac{K_{II}}{2\mu}\sqrt{r/(2\pi)}sin\frac{\theta}{2}\left(\kappa + 1 + 2cos^2\frac{\theta}{2}\right)$ |
| $u_y$ | $-\frac{K_{II}}{2\mu}\sqrt{r/(2\pi)}cos\frac{\theta}{2}\left(\kappa - 1 - 2sin^2\frac{\theta}{2}\right)$ |

Three cracked element discretizations are considered, as they arise typically on QT meshes, prior to the refinement and homogenization steps (Fig. 3). The exact solution is applied to the QT mesh nodes identified in Fig. 5. The remaining nodes resulting from internal element refinement are restricted to move as a linear combination of their neighbouring QT mesh nodes, enforcing compatibility.

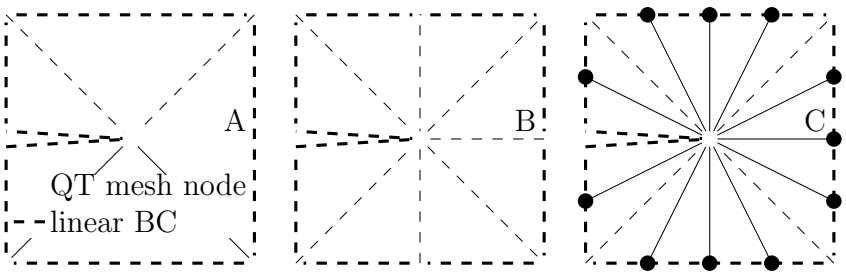



**Figure 5: Typical element types A-C arising from QT meshes.**

The convergence behaviour is investigated in Fig. 6. Due to the over-constraining of the boundary, deviation in convergence behaviour is expected, i.e., by enforcing linearly

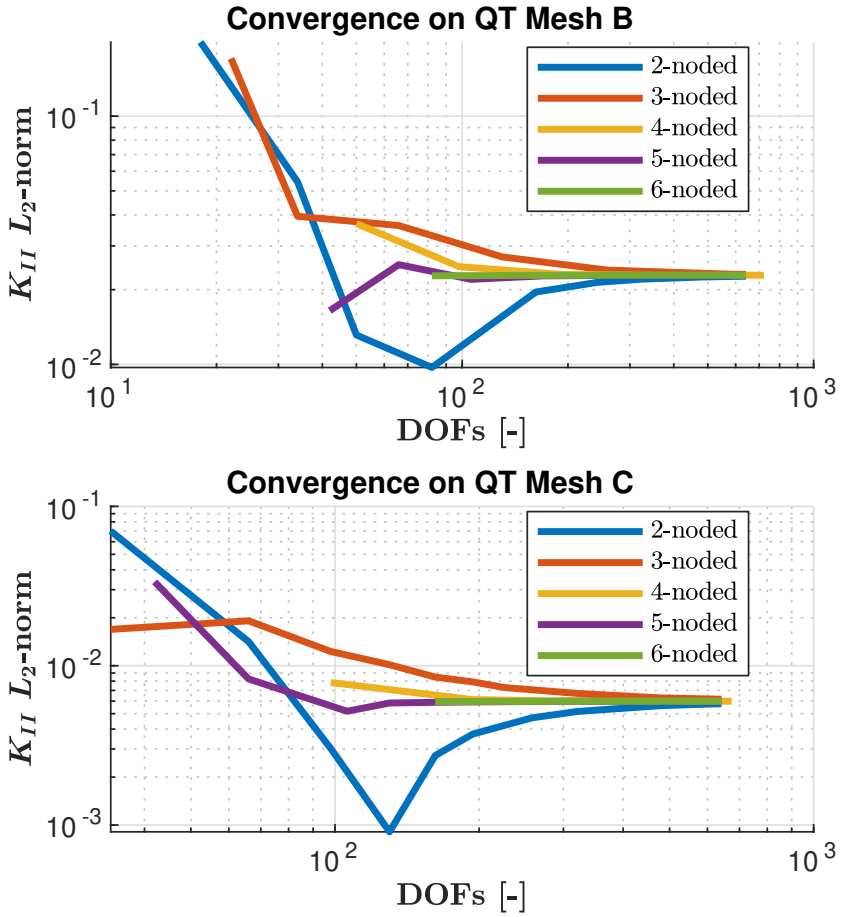



**Figure 6: Convergence behaviour of $K_{II}$ on QT meshes employing n-noded elements as boundary discretization.**

dependant displacement boundary conditions between QT mesh nodes, an effective stress state is imposed that differs mildly from the exact solution. In this example, the gSIFs are therefore slightly overestimated, which is evident in the asymptotic behaviour on all three meshes. Nevertheless, the error in $K_{II}$ is significantly reduced for all QT meshes, as can be seen in Tab. 2. The diverging results for the case of 2-noded elements is explained by the examination of the absolute values. While the higher order elements approach the asymptotic solution from the high end, the two-noded elements approach from the lower end, thus crossing the imposed $K_{II} = 1$ in the process. The obtainable accuracy is naturally limited to the asymptotic case by the artificially imposed boundary conditions. Therefore, the expected convergence behaviour is disrupted accordingly. In this specific example, the convergence behaviour of even-noded elements was observed to behave predictably, while the odd-noded elements exhibited slight oscillatory behaviour. It can be observed that the computed values for $K_{II}$ remain practically stable when more than 100 DOFs are employed within the cracked element. Preference should be given to higher order elements due to their enhanced convergence properties. On a contemporary laptop employing unoptimized Matlab code, this entails calculations completing in less than 0.1s. Tab. 2 provides the results for $K_{II}$ calculated on the original QT meshes and contrasts them to the asymptotic solutions for the refined crack elements, given a high fidelity SBFEM solution. Assuming a user-specified accuracy tolerance, the improved method of calculating the gSIFs facilitates a decision criterion on whether global mesh

refinement is required. The difference in achieved accuracy for each QT mesh is primarily attributed to the manner in which the enforced displacements on the boundary conform to the exact field solution. Deviations therein are reflected in the integration constants $[c]$ (Eqn. 19) and comprise the remaining error in the asymptotic solution. In a practical application of this method, the displacement solution on the QT mesh nodes results from the FEM solution of the global QT mesh. Since the method proposed to improve the calculation of gSIFs does not increase the amount of global DOFs, the overall analysis does not differ. Solely on the cracked element that is refined, "virtual" nodes are introduced that, however, never manifest in calculation of system displacements as their values are predetermined due to the linearized BCs. In principle, a more accurate representation of the displacement modes and eigenvalues is achieved inside the cracked element, while the boundary constraints remain unchanged. This explains the counter intuitive results of the asymptotic error for QT mesh type A for which smaller errors are observed than for the QT mesh type B: For this specific loading, i.e., mode II excitation, on average, the enforced displacement field on the boundary results in a closer approximation of the exact stress field. If mode I excitation where, however, considered, for which the right boundary of the cracked element exhibits concave behaviour, QT mesh Type A results in a 15% error, since it is not able to reproduce such displacement behaviour. Between QT meshes type B and C we observe convergence as expected, as QT mesh C can represent all displacement modes of QT mesh B, while also introducing additional ones.

**Table 2: Convergence of gSIFs to imposed solution.**

| **Method** | | $\boldsymbol{K_{II}}$ | **error [-]** | **error [%]** |
|---|---|---|---|---|
| imposed | | 1 | - | - |
| (v) hi-fi | | 1.00000000524414 | $\approx 0$ | $\approx 0$ |
| (iii) QT mesh | A | 1.01547261490198 | 0.0155 | 1.55 |
| | B | 0.80116357970950 | 0.1988 | 19.88 |
| | C | 0.92998436271287 | 0.0700 | 7.00 |
| (iv) asympt. | A | 1.01124862611779 | 0.0113 | 1.12 |
| | B | 1.02286467268900 | 0.0287 | 2.29 |
| | C | 1.00598223218854 | 0.0060 | 0.60 |

*Edge cracked square plate in mode I*

An edge cracked square plate subject to a plane stress state is examined (Fig. 7). The bottom edge of the plate is fully clamped, while on the top edge forced displacements $u_y = 1$ and $u_x = 0$ are applied. The Young's modulus, Poisson's ratio, fracture energy, crack propagation increment and side length are given as $E = 200$ [N/mm$^2$], $v = 0.3$, $G_c = 2.7$ [N/mm$^2$], $\Delta a = 0.025$ [mm] and $L = 1$ [mm] respectively.
First, the load-deflection curves obtained from several methods are compared: (i) XFEM, (ii) traditional QT SBFEM, (iii) SBFEM on QT mesh types A-C, (iv) the newly proposed SBFEM method and (v) a high fidelity SBFEM solution comprising one subdomain with h- and p-refinement. The load-deflection curves (Fig. 8) are constructed using the following procedure:

1. Calculate the critical stress intensity factor $K_c$ from the E-modulus and the critical

energy release rate $G_c$:

$$K_c = \sqrt{EG_c} \quad \text{for plane stress} \tag{22}$$

2. Formulate the equivalent stress intensity factor $K_{eq}$:

$$K_{eq} = \sqrt{K_I^2 + K_{II}^2} \tag{23}$$

This implies that the crack propagates as soon as $K_{eq} \leq K_c$. Hence, the gSIFs and the sum of the reaction forces are stored at each iteration.

3. Determine the load factor such that crack propagation initiates, i.e., as the ratio $K_c/K_{eq}$. The effective displacements and loads at each crack increment step are calculated by scaling the initially imposed values by the load factor.

4. Employ the minimum strain energy density criterion (or equivalent) to calculate the crack propagation angle [14]:

$$\theta_c = 2\tan^{-1}\left[\frac{-2K_I/K_{II}}{1+\sqrt{1+8(K_I/K_{II})^2}}\right] \tag{24}$$

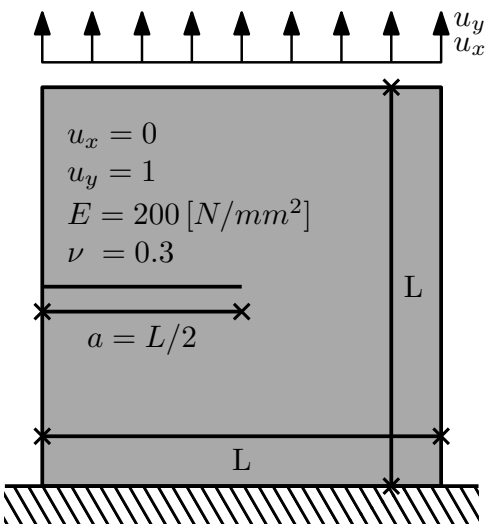



**Figure 7: Experimental setup for numerical example A.**

For this simple example, the crack paths for all methods coincide, i.e., they propagate in direct extension of the imposed crack. Investigating the load-deflection behaviour (Fig. 8), two reference solutions are constructed: a high fidelity SBFEM solution (v) employing 568 DOFs and an XFEM solution (i) employing a domain discretization of 161x161 element (52'488 DOFs). As expected, the traditional approach (ii) utilizing 1024 DOFs mirrors the reference XFEM solution. Similarly, the correct load-deflection response is obtained by the novel approach (iv), however, employing only 880 DOFs, i.e., on the same mesh as the deficient method (iii). The difference in DOFs is attributed to the balancing operation following refinement around the crack tip, which propagates across the domain.
Given this specific geometry, an exact solution for the crack tip singularity is available, i.e., a square root singularity, which we exploit for benchmarking purposes. By inspection

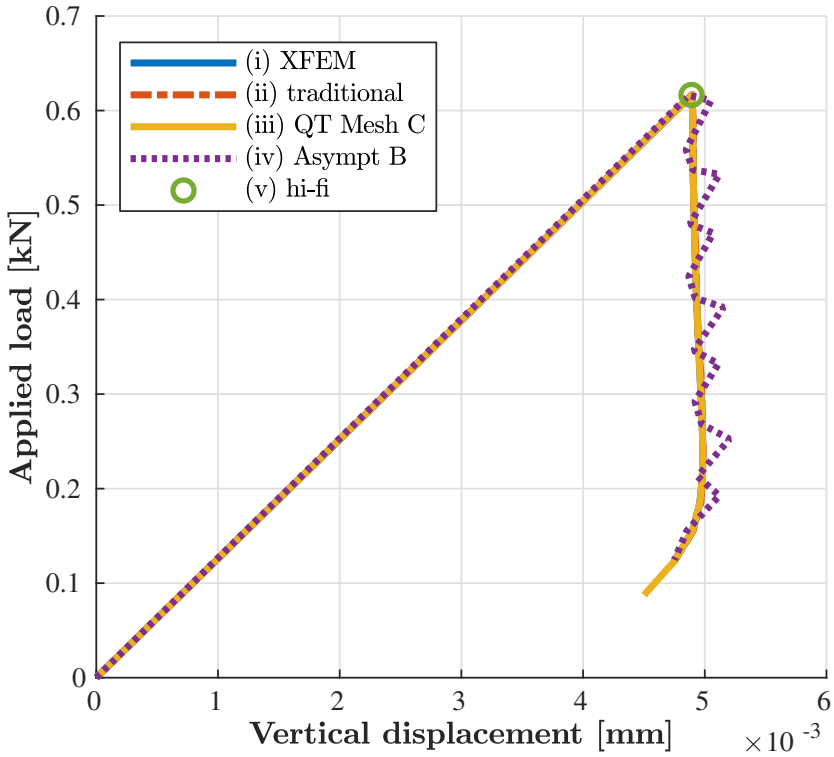



**Figure 8: Load deflection curves for methods (i) - (v).**

of Eqn. 19, this is the case when all eigenvalues contained in $[S^{(s)}]$ are equal to 0.5. Since the singular eigenvalues depend only on the boundary discretization, a SBFEM element's ability to reproduce the exact solution is purely a reflection of its approximation space. Hence, by leveraging hp-elements, the error in calculated gSIFs, due to the choice of approximation space, can be minimized so that the remaining error stems only from the surrounding mesh discretization. Tab. 3 summarizes the results for the SBFEM based methods given different levels of refinement. As expected, the hi-fi (v) solution is able to approximate the exact solution to nearly machine precision even with very few DOFs. Both the proposed method (iv) and traditional (ii) approaches deliver approximations accurate to several significant figures. For method (iv) 3- and 5-noded elements, denoted by 3n and 5n respectively are considered. For method (ii), studies were conducted using either {2,3,6,12} linear elements per long side of the cracked domain. It can be observed that the proposed method (iv) significantly outperforms the traditional (ii) approach at similar internal discretization levels. The elements, as they are employed in method (iii), perform poorly and report significantly fewer accurate digits.
We compare $K_I$ obtained by our proposed method (iv) and method (iii) to the hi-fi reference solution (v) (Tab. 4). Since both methods employ the same mesh, we isolate the gains of proposed method (iv). Indeed, for mesh types A and B we observe an improvement in accuracy of the calculated gSIFs by approximately and order of magnitude. For mesh C, however, we approach a limit given by the discretization error of the underlying mesh and therefore the error cannot be reduced to such a degrees as with the previous meshes. Nevertheless, accuracy $< 1\%$ is readily achieved, indicating a sufficiently accurate solution for most SIF-based applications.

**Table 3: Convergence of eigenvalues to square root singularity.**

| Method | | DOFs | $\lambda_1$ | $\lambda_2$ |
|---|---|---|---|---|
| exact | | - | 0.5 | 0.5 |
| (ii) trad. | 2 | 34 | 0.502106496308655 | 0.502106496308665 |
| | 3 | 66 | 0.500547555626861 | 0.500547555626900 |
| | 6 | 162 | 0.500088619101087 | 0.500088619101087 |
| | 12 | 322 | 0.500022191325667 | 0.500022191325716 |
| (iii) QT mesh | A | 12 | 0.543331260622274 | 0.487073508787698 |
| | B | 18 | 0.506300843546734 | 0.506300843547260 |
| | C | 34 | 0.501432879576478 | 0.501432879577241 |
| (iv) asympt. | 3n | 66 | 0.499993111156218 | 0.499991972174270 |
| | 3n | 98 | 0.499998547338467 | 0.499998547338467 |
| | 3n | 130 | 0.499999524753473 | 0.499999524753473 |
| | 5n | 42 | 0.500958236174165 | 0.500555777249317 |
| | 5n | 68 | 0.500001822342609 | 0.500001760606780 |
| | 5n | 106 | 0.500000389900685 | 0.500000214938575 |
| (v) hi-fi | | 578 | 0.499999999998063 | 0.500000000000369 |

**Table 4: Convergence of gSIFs to high-fidelity solution.**

| Method | | $K_I$ | error [-] | error [%] |
|---|---|---|---|---|
| (iii) QT mesh | A | 18.824826991544300 | 3.4427 | 22.38 |
| | B | 17.144699325716516 | 1.7626 | 11.46 |
| | C | 15.767196554450349 | 0.3851 | 2.50 |
| (iv) asympt. | A | 15.714940415023673 | 0.3328 | 2.16 |
| | B | 15.629945166079814 | 0.2478 | 1.61 |
| | C | 15.457607616329554 | 0.0755 | 0.49 |
| (v) hi-fi | | 15.382113483624098 | - | - |

## Conclusion

This paper demonstrates that the accuracy of gSIFs calculated on hierarchical meshes can be significantly enhanced by implementing local refinement of the cracked element, while constraining the boundary displacements to conform with the surrounding mesh. The error stemming from the element's ability to accurately represent the singular stress field is alleviated and the remaining deviation is attributed to the discretization error introduced by the mesh. The proposed method permits the use of significantly coarse discretizations of the domain without the need for artificial refinement about the crack tip to obtain comparable accuracy of the gSIFs. This benefit is compounded, since the balancing operation employed on the hierarchical mesh, to arrive at a finite number of precomputable element realizations, is eliminated. Numerical examples have demonstrated that the use of higher order elements and approximately 100 DOFs in the refined element produce accurate results, while retaining high computational efficiency. The limitations of this approach stem from the linear boundary discretization, imposing artificial constraints on

the solution. The use of higher order elements, such as cubic line elements on hierarchical meshes could minimize this issue and require the development of targeted implementations for more involved domain geometries.

## Acknowledgment

This research was performed under the auspices of the Swiss National Science Foundation (SNSF), Grant # 200021_153379, A Multiscale Hysteretic XFEM Scheme for the Analysis of Composite Structures. The authors would like to thank the group of Prof. Song from UNSW for guidance and code on SBFEM. Further, we would like to extend our thanks to Prof. Devloo from UNICAMP for his illuminating discussions. Finally, we extend our gratitude to Dr. Agathos for allowing access to his XFEM codes.

### References

[1] Chen, X., Luo, T., Ooi, E., Ooi, E., and Song, C. A quadtree-polygon-based scaled boundary finite element method for crack propagation modeling in functionally graded materials. *Theoretical and Applied Fracture Mechanics 94* (apr 2018), 120–133.

[2] Chowdhury, M. S., Song, C., and Gao, W. Highly accurate solutions and Pad approximants of the stress intensity factors and T-stress for standard specimens. *Engineering Fracture Mechanics 144* (aug 2015), 46–67.

[3] Deeks, A. J., and Wolf, J. P. A virtual work derivation of the scaled boundary finite-element method for elastostatics. *Computational Mechanics 28*, 6 (jun 2002), 489–504.

[4] Egger, A., Pillai, U., Agathos, K., Kakouris, E., Chatzi, E., Aschroft, I. A., and Triantafyllou, S. P. Discrete and phase field methods for linear elastic fracture mechanics: A comparative study and state-of-the-art review. *Applied Sciences 9*, 12 (June 2019), 2436.

[5] Hu, Z., Lin, G., Wang, Y., and Liu, J. A Hamiltonian-based derivation of Scaled Boundary Finite Element Method for elasticity problems. *IOP Conference Series: Materials Science and Engineering 10* (jun 2010), 012213.

[6] Janssen, M. *Fracture mechanics*. DUP Blue Print, Delft, 2002.

[7] Liu, P., and Zheng, J. Recent developments on damage modeling and finite element analysis for composite laminates: a review. *Materials & Design 31*, 8 (2010), 3825–3834.

[8] Mos, N., Dolbow, J., and Belytschko, T. A finite element method for crack growth without remeshing. *International Journal for Numerical Methods in Engineering 46*, 1 (Sept. 1999), 131–150.

[9] Ooi, E., Man, H., Natarajan, S., and Song, C. Adaptation of quadtree meshes in the scaled boundary finite element method for crack propagation modelling. *Engineering Fracture Mechanics 144* (aug 2015), 101–117.

[10] Ooi, E. T., Natarajan, S., Song, C., and Ooi, E. H. Dynamic fracture simulations using the scaled boundary finite element method on hybrid polygonquadtree meshes. *International Journal of Impact Engineering 90* (apr 2016), 154–164.

[11] Ooi, E. T., Natarajan, S., Song, C., and Ooi, E. H. Crack propagation modelling in concrete using the scaled boundary finite element method with hybrid polygonquadtree meshes. *International Journal of Fracture 203*, 1-2 (jan 2017), 135–157.

[12] Ravi-Chandar, K. Dynamic fracture of nominally brittle materials. *International Journal of Fracture 90*, 1-2 (1998), 83–102.

[13] Ravi-Chandar, K., and Knauss, W. An experimental investigation into dynamic fracture: Iii. on steady-state crack propagation and crack branching. *International Journal of fracture 26*, 2 (1984), 141–154.

[14] Sih, G. Strain-energy-density factor applied to mixed mode crack problems. *International Journal of fracture 10*, 3 (1974), 305–321.

[15] Song, C. A matrix function solution for the scaled boundary finite-element equation in statics. *Computer Methods in Applied Mechanics and Engineering 193*, 23-26 (jun 2004), 2325–2356.

[16] Song, C. Evaluation of power-logarithmic singularities, T-stresses and higher order terms of in-plane singular stress fields at cracks and multi-material corners. *Engineering Fracture Mechanics 72*, 10 (July 2005), 1498–1530.

[17] Song, C. *The scaled boundary finite element method: intorduction to theory and implementation.* John Wiley & Sons, Hoboken, New Jersey, 2018.

[18] Song, C., Tin-Loi, F., and Gao, W. A definition and evaluation procedure of generalized stress intensity factors at cracks and multi-material wedges. *Engineering Fracture Mechanics 77*, 12 (aug 2010), 2316–2336.

[19] Song, C., and Wolf, J. P. The scaled boundary finite-element methodalias consistent infinitesimal finite-element cell methodfor elastodynamics. *Computer Methods in Applied Mechanics and Engineering 147*, 3-4 (aug 1997), 329–355.

[20] Wolf, J. P. *The scaled boundary finite element method.* J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, USA, 2003.

[21] Xu, P., Zheng, J., and Liu, P. Finite element analysis of burst pressure of composite hydrogen storage vessels. *Materials & Design 30*, 7 (2009), 2295–2301.

[22] Yang, Z. Fully automatic modelling of mixed-mode crack propagation using scaled boundary finite element method. *Engineering Fracture Mechanics 73*, 12 (aug 2006), 1711–1731.

[23] Zheng, J., and Liu, P. Elasto-plastic stress analysis and burst strength evaluation of al-carbon fiber/epoxy composite cylindrical laminates. *Computational Materials Science 42*, 3 (2008), 453–461.

# Exploring Topology Optimization on Hierarchical Meshes by Scaled Boundary Finite Element Method

***Adrian Egger[1], Albert Saputra[2], Savvas Triantafyllou[3], and Eleni Chatzi[1]**

[1]Institute of Structural Engineering, ETH Zurich, Switzerland
[2]Centre for Infrastructure Engineering and Safety, University of New South Wales, Australia
[3]Centre for Structural Engineering and Informatics, The University of Nottingham, UK

*Presenting and corresponding author: egger@ibk.baug.ethz.ch

## Abstract

In this contribution, we aim at accelerating topology optimization by recasting the forward problem into a form that directly interacts with the structural solver: The polytope nature of SBFEM elements is exploited on quad-/octree meshes to alleviate issues associated with hanging nodes. Furthermore, a balancing operation applied to the mesh results in a manageable number of precomputable element configurations, which significantly accelerates the forward analysis. The analysis mesh for each optimization iteration is obtained via automated image-based decomposition of the design variables.
A number of benefits arise from this combination of methods. The ease with which higher-order elements may be incorporated, coupled with the use of unstructured meshes, combats the formation of checker-boarding. Second, computational effort only arises where required by the problem definition, since adaptivity is automatically provided. Third, numerical examples in both 2D and 3D indicate that the amount of degrees of freedom present during analysis is reduced by more than an order or magnitude.

**Keywords:** Topology optimization, SBFEM, Hierarchical meshes.

## Introduction

The adoption of Topology Optimization (TO) into engineering practice stems from the ever-increasing need for the sustainable, economical and accountable use of resources across diverse sectors, e.g., construction, aerospace and automotive. The solution of a computationally expensive forward problem is typically fundamental to TO of continuum structures [4], which aims at defining an optimal structural layout subject to constraints. The conventional approach employing a fine grid of design variables, with values of 0 and 1 corresponding to void and solid, respectively, and whose layout remains constant across all iterations, does not reflect the nature of the evolving topology and leads to an excessive computational toll. Several adaptive schemes have been proposed to alleviate this computational burden [8, 19, 27]. In [17] tree-based meshing strategies are explored. Unfortunately, therein the treatment of hanging nodes severely affects the computational efficiency, which the herein proposed scheme remedies.
This work is motivated by the adoption of numerical methods capable of treating polytope-elements to alleviate issues commonly associated with hanging nodes. The proposed scheme is suited for fusion with, but not limited to, the boundary element method [14], polygonal finite element method [26], conforming shape functions [7] or virtual element method [3]. Here, we employ the scaled boundary finite element method (SBFEM) [24], a semi-analytical numerical method that permits the treatment of star-convex polytopes by introducing a scaling center in every element, thereby retrieving an analytical solution in radial direction, while only necessitating discretization of element boundaries. This

attribute permits the construction of hp-elements. The combination of higher order elements and unstructured meshes combats the formation of checker-boarding [4]. Although a mixed SBFEM has been employed for solving TO problems of incompressible materials [12], it follows the conventional fine-grid approach to TO and limits its treatment to linear elements and 2D applications.

The SBFEM has proven itself as a remarkably versatile tool in automatic image-based stress analysis [15, 21]. Such hierarchical meshes arising from tree-like image decompositions drastically reduce the amount of degrees of freedom (DOFs) present, which accelerates the solution of the forward problem by alleviating computational effort and memory requirements. Image decomposition techniques, within the context of TO, produce fewer DOFs when material transition zones are eliminated. Hence, bi-directional evolutionary structural optimization (BESO) [10], sequential element rejection and admission (SERA) [2], solid isotropic microstructure with penalization (SIMP) combined with grayscale filters [1, 13] and level-set methods [16], for example, represent suitable algorithms. Since image decomposition operates on the design variables to produce analysis-ready meshes at each iteration, this family of techniques only requires interchanging the solver of the forward problem. Hence, incorporation into existing code bases requires minimal modification. Furthermore, as tree-like image decomposition techniques [5] utilize the uniformity of an element as a criterion for subdivision, multi-material TO [20] may be incorporated through extension to color-aware tree-like decompositions [11].

This work is structured as follows: The pertinent theory is provided first. This entails the salient features of TO and automated image segmenting into analysis-ready hierarchical meshes. These meshes, consisting of significantly fewer DOFs than traditional fine-grid approaches, are employed during the solution of the forward problem and contribute significantly to the reduction in computational cost. Subsequently, the proposed scheme is introduced in detail. Its novelty is i) the fusion of TO with automated image segmenting schemes, ii) the use of polytope-elements constructed by SBFEM to alleviate issues associated with hanging nodes, iii) color-encoding of domain and analysis parameters into the input images and iv) the extension to SBFEM-powered 3D TO. Based on three numerical examples, the proposed scheme is then verified, showcased and discussed. Novelties i) and ii) are demonstrated to significantly reduce computational effort and memory requirements, when compared to traditional fine-grid approaches. Finally, we offer conclusions and directions for future work.

## Theory

*Topology Optimization for Compliance Problems*

In this paper we consider TO with focus on compliance minimization when subjected to a volume constraint, i.e., maximizing the structural stiffness when only a fraction of the original volume is available:

$$\begin{aligned} \min_{\mathbf{x}} &: c(\mathbf{x}) = \mathbf{U}^{\mathbf{T}}\mathbf{K}\mathbf{U} = \sum_{e=1}^{N} g(x_e)\mathbf{u}_{\mathbf{e}}^{\mathbf{T}}\mathbf{k}_{\mathbf{0}}\mathbf{u}_{\mathbf{e}} \\ \text{subject to} &: V(\mathbf{x})/V_0 = f \\ &: \mathbf{K}\mathbf{U} = \mathbf{F} \\ &: \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \end{aligned} \tag{1}$$

where the set of $N$ design variables $\mathbf{x}$ spans the design domain and specifies the material distribution. Three conditions are imposed, constraining the optimal solution. First, each

element of $\mathbf{x}$, denoted by $x_e$, must fall within limit values of 0 and 1, which correspond to void and solid regions respectively. Second, a user prescribed volume fraction f follows as the ratio of current volume $V(\mathbf{x})$ to initial volume $V_0$. Third, the displacement field $\mathbf{U}$, required to compute the compliance $c$, follows from the solution of the forward problem in 3D elastostatics. $\mathbf{K}$ denotes the corresponding stiffness matrix and $\mathbf{F}$ the load vector respectively. The compliance can either be calculated globally or locally, as a summation of element contributions. The elemental nodal displacements are denoted by $\mathbf{u_e}$ and the corresponding stiffness matrix, calculated with Young's modulus equal to one, i.e., a solid element, is given as $\mathbf{k_0}$. $g(x_e)$ is a function, which typically scales $\mathbf{k_0}$ according to the specific realization of its design variable $x_e$ to account for intermediate material properties, i.e., Young's modulus.

*Segmenting Images into Analysis-ready Hierarchical Meshes*

Tree-based image decomposition techniques typically operate on gray-scale input then output regions, which fulfill a user specified homogeneity criterion. In this paper, we limit our focus to quadtrees (2D) and octrees (3D), which follow the same underlying principles: The region is bisected parallel to the Cartesian axes, if the spread in gray-scale values of any contained pixels exceeds a user-specified threshold. In 2D this results in splitting and replacing the region by four equally sized blocks, explaining the prefix "quad". If this process is repeated often enough, visually, a tree-like structure emerges (Fig. 1), which clarifies the suffix. For pixel-based applications, bisection requires the

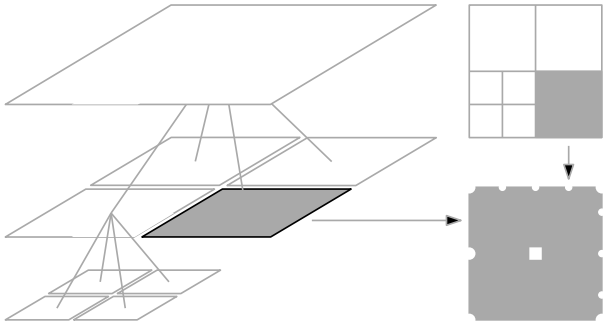


**Figure 1: Example of image decomposition by quadtree algorithm with sample SBFEM polytope element in gray.**

resolution of the input image to comprise a power of two. Since this is generally not the case, images are padded to the next power of 2, where applicable. Each homogeneous block is represented by a single element. Assuming nodes in each element corner, the image decomposition is transformed into a mesh representation (Fig. 2). Due to the irregularity of the mesh, challenges arise during analysis: Constructing appropriate elements and satisfying sufficient discretization. Traditionally, balancing such meshes, i.e., enforcing at maximum at 2:1 ratio of adjacent blocks, tempers most issues. However, hanging nodes (Fig. 2, red) still persist. One option to alleviate this issue is the adoption of a polytope-based finite element variant, such as the SBFEM. Since, for example, only 16 possible element realizations exist in 2D, precomputation may even be exploited in the process.

*The Forward Problem for 3D Elastostatics*

We consider a three dimensional domain $\Omega$, whose boundary $\Gamma = \Gamma_0 \cup \Gamma_u \cup \Gamma_t$ comprises regions of free surface conditions ($\Gamma_0$), prescribed displacements $\bar{u}$ ($\Gamma_u$) and prescribed

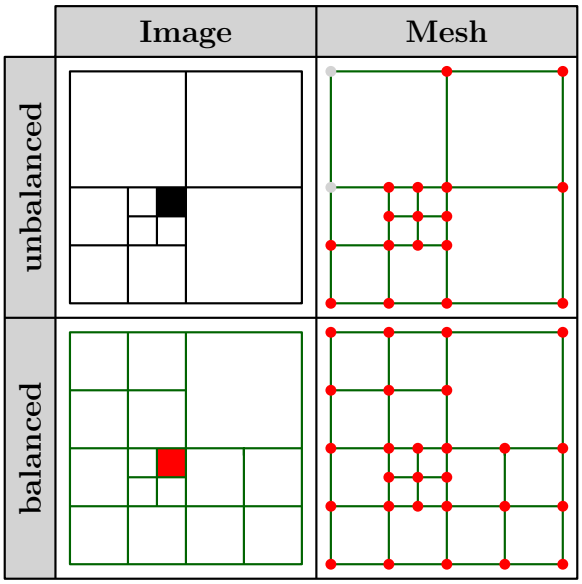



**Figure 2: Segmentation of image with resulting mesh for unbalanced and balanced quadtree decompositions. Hanging nodes in red and additional elements introduced due to balancing in green.**

traction $\bar{t}$ ($\Gamma_t$). The strong form with boundary conditions may be stated as:

$$\nabla \cdot \sigma + b = 0 \quad \text{in} \quad \Omega \tag{2a}$$
$$u = \bar{u} \quad \text{on} \quad \Gamma_u \tag{2b}$$
$$\sigma \cdot n = \bar{t} \quad \text{on} \quad \Gamma_t \tag{2c}$$

The Cauchy stress tensor, unit outward normal to the boundary and applied body force per unit volume are denoted by $\sigma$, $n$ and $b$ respectively. $u$ is the displacement field and $\nabla$ the linear gradient operator. Imposing small deformations and linear elastic material behaviour, the stress $\sigma$ and strain $\epsilon$ fields are dependent on the modulus of elasticity $E$ and Poisson's ratio $\nu$:

$$\epsilon = \nabla_s u \quad \text{and} \quad \sigma = D\epsilon \tag{3}$$

where $\nabla_s$ is the symmetric gradient operator and $D$ the 6x6 elasticity tensor.

*The Scaled Boundary Finite Element Method in 3D Elastostatics*

In this section, a brief summary of the scaled boundary finite element method in 3D is given. For a more elaborate derivation and detailed explanations, the readers may consult references [24, 25].

The salient features of an SBFEM analysis are illustrated on the problem domain described by the volume $V$ depicted in Fig. 3. V comprises the volume spanned by the scaling center $O$ and the 2D surface element, describing the boundary. One minor constraint is required: The domain must remain star-convex, i.e., the entire surface must be visible from the scaling center. The introduction of this scaling center is accompanied by a transition from a Cartesian reference system into one resembling polar coordinates. In radial direction the analytic variable $\xi$ is introduced, while for each tangential direction, $\eta$ respectively $\zeta$ represent the local coordinates on the boundary. Therefore, each surface element may be

described by 2D interpolation shape functions $\mathbf{N}(\eta,\zeta)$ formulated in natural coordinates $-1 \leq \eta \leq 1$ and $-1 \leq \zeta \leq 1$. The interior of the domain is constructed by scaling the boundary $(x, y, z)$ along the dimensionless radial coordinate $0 \leq \xi \leq 1$, which originates at the scaling center and ends on the boundary. The mapping of points employing the newly introduced scaled boundary coordinate system is therefore given as:

$$\hat{x}(\xi,\eta,\zeta) = \xi x(\eta,\zeta) = \xi \mathbf{N}(\eta,\zeta)\,\mathbf{x}, \tag{4a}$$
$$\hat{y}(\xi,\eta,\zeta) = \xi y(\eta,\zeta) = \xi \mathbf{N}(\eta,\zeta)\,\mathbf{y}, \tag{4b}$$
$$\hat{z}(\xi,\eta,\zeta) = \xi z(\eta,\zeta) = \xi \mathbf{N}(\eta,\zeta)\,\mathbf{z}. \tag{4c}$$

We denote the vectors of nodal coordinates of a surface element by $\mathbf{x}, \mathbf{y}, \mathbf{z}$ respectively. The set $(\xi,\eta,\zeta)$ is termed the scaled boundary coordinates in the three-dimensional domain.

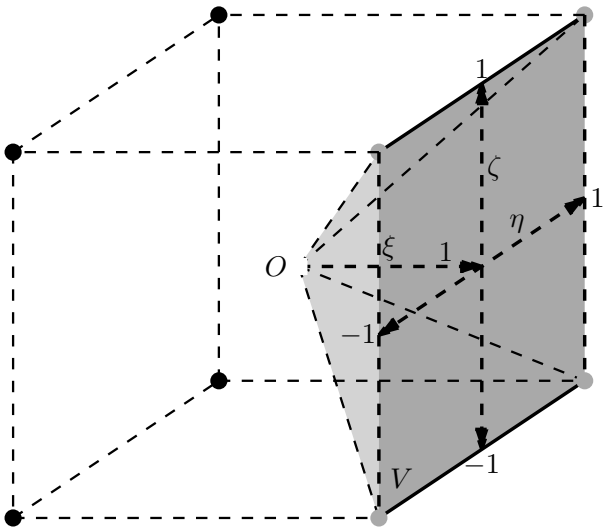



**Figure 3: Three-dimensional coordinates for a scaled boundary finite element.**

Similarly, the iso-parametric mapping of the displacements $\mathbf{u}(\xi,\eta,\zeta)$ at a point $(\xi,\eta,\zeta)$ comprises an analytic $(\xi)$ and interpolatory $(\eta,\zeta)$ component:

$$\mathbf{u}(\xi,\eta,\zeta) = \left[u_x(\xi,\eta,\zeta), u_y(\xi,\eta,\zeta), u_z(\xi,\eta,\zeta)\right]^{\mathrm{T}} = \mathbf{N}^{\mathrm{u}}(\eta,\zeta)\,\mathbf{u}(\xi), \tag{5}$$

where $\mathbf{u}(\xi)$ represents an analytic displacement function along $\xi$, unique to each node on the boundary. These displacements functions are determined during the SBFEM solution. The interpolation shape function $\mathbf{N}^{\mathrm{u}}(\eta,\zeta)$ in Eq. (5) are defined analogously to the conventional FEM:

$$\mathbf{N}^{\mathrm{u}}(\eta,\zeta) = \left[N_1(\eta,\zeta)\,\mathbf{I}, N_2(\eta,\zeta)\,\mathbf{I}, \ldots, N_n(\eta,\zeta)\,\mathbf{I}\right], \tag{6}$$

where $n$ denotes the amount of nodes of the surface element and $\mathbf{I}$ is the $3 \times 3$ identity matrix. Expressing the strains in scaled boundary coordinates requires splitting the linear differential operator into components $\mathbf{B}_1$ and $\mathbf{B}_1$, whose combined effect mimics the original transformation:

$$\boldsymbol{\varepsilon}(\xi,\eta,\zeta) = \mathbf{B}_1(\eta,\zeta)\,\mathbf{u}(\xi)_{,\xi} + \frac{1}{\xi}\mathbf{B}_2(\eta,\zeta)\,\mathbf{u}(\xi), \tag{7}$$

The stresses are obtained conventionally by pre-multiplying the strains by the constitute matrix:

$$\boldsymbol{\sigma}\left(\xi,\eta,\zeta\right)=\mathbf{D}\left(\mathbf{B}_1\left(\eta,\zeta\right)\mathbf{u}\left(\xi\right)_{,\xi}+\frac{1}{\xi}\mathbf{B}_2\left(\eta,\zeta\right)\mathbf{u}\left(\xi\right)\right). \tag{8}$$

The weak form for each subdomain may be derived by applying several methods [6,9,29]. Neglecting body loads and surface tractions, two equations arise (Eqs. 9 and 10):

$$\mathbf{E}_0\xi^2\mathbf{u}\left(\xi\right)_{,\xi\xi}+\left(2\mathbf{E}_0-\mathbf{E}_1+\mathbf{E}_1^{\mathrm{T}}\right)\xi\mathbf{u}\left(\xi\right)_{,\xi}+\left(\mathbf{E}_1^{\mathrm{T}}-\mathbf{E}_2\right)\mathbf{u}\left(\xi\right)=0, \tag{9}$$

The coefficient matrices $\mathbf{E}_0,\mathbf{E}_1,\mathbf{E}_2$ bare a striking similarity to conventional FEM stiffness matrices, both in their structure and in that they are calculated for each element individually, with subsequent assembly for each subdomain. The internal nodal forces modes at the boundary $\mathbf{q}\left(\xi\right)$ are derived as:

$$\mathbf{q}\left(\xi\right)=\xi\left(\mathbf{E}_0\xi\mathbf{u}\left(\xi\right)_{,\xi}+\mathbf{E}_1^{\mathrm{T}}\mathbf{u}\left(\xi\right)\right). \tag{10}$$

The scaled boundary finite element equation is solved by the matrix function solution proposed by Song [23]. Therein, the quadratic eigen-problem is recast into a system of first-order differential equations in $\xi$, at the expense of doubling the amount of unknowns:

$$\xi\left\{\begin{array}{c}\xi^{0.5}\mathbf{u}\left(\xi\right)\\ \xi^{-0.5}\mathbf{q}\left(\xi\right)\end{array}\right\}_{,\xi}=-\mathbf{Z}\left\{\begin{array}{c}\xi^{0.5}\mathbf{u}\left(\xi\right)\\ \xi^{-0.5}\mathbf{q}\left(\xi\right)\end{array}\right\}, \tag{11}$$

with the Hamiltonian coefficient matrix $\mathbf{Z}$ defined as:

$$\mathbf{Z}=\left[\begin{array}{cc}\mathbf{E}_0^{-1}\mathbf{E}_1^{\mathrm{T}}-0.5\mathbf{I} & -\mathbf{E}_0^{-1}\\ -\mathbf{E}_2+\mathbf{E}_1\mathbf{E}_0^{-1}\mathbf{E}_1^{\mathrm{T}} & -\left(\mathbf{E}_1\mathbf{E}_0^{-1}-0.5\mathbf{I}\right)\end{array}\right]. \tag{12}$$

The Schur decomposition with subsequent block-diagonalization is employed to avoid numerical deficiencies in the solution and decompose Z to satisfy:

$$\mathbf{ZV}=\mathbf{VS}, \tag{13}$$

where $\mathbf{S}$ and $\mathbf{V}$ are the real Schur form and the transformation matrix, respectively. In order to strip the bounded from the unbounded response, the diagonal blocks containing the eigen-values of $\mathbf{S}$ are sorted in ascending order and the columns of $\mathbf{V}$, which contain the associated eigen-modes, are reordered accordingly. The following partitioning is devised:

$$\mathbf{S}=\left[\begin{array}{cc}\mathbf{S}_{11} & 0\\ 0 & \mathbf{S}_{22}\end{array}\right], \tag{14a}$$

$$\mathbf{V}=\left[\begin{array}{cc}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{array}\right]. \tag{14b}$$

$\mathbf{S}_{11}$ contains all eigenvalues with negative sign, i.e., $\mathrm{Re}\left(\lambda(\mathbf{S}_{11})\right)<0$, which can be shown to correspond to the bounded domain solution. The general solutions for the displacements and internal nodal forces for the bounded domain is sought in the form of a power series and can be expressed as:

$$\mathbf{u}\left(\xi\right)=\mathbf{V}_{11}\xi^{-\mathbf{S}_{11}-0.5\mathbf{I}}\mathbf{c}, \tag{15a}$$

$$\mathbf{q}\left(\xi\right)=\mathbf{V}_{21}\xi^{-\mathbf{S}_{11}+0.5\mathbf{I}}\mathbf{c}. \tag{15b}$$

The integration constants $\mathbf{c}$, whose values are determined in post-processing, follow from the effective boundary conditions of the problem. Formulating Eq. (15) at the boundary ($\xi = 1$), the nodal displacements $\mathbf{u} = \mathbf{u}(\xi = 1)$ and the nodal forces $\mathbf{q} = \mathbf{q}(\xi = 1)$ can be expressed as:

$$\mathbf{u} = \mathbf{V}_{11}\mathbf{c}, \tag{16a}$$

$$\mathbf{q} = \mathbf{V}_{21}\mathbf{c}, . \tag{16b}$$

Since the static stiffness matrix $\mathbf{K}$ is defined as $\mathbf{q} = \mathbf{K}\mathbf{u}$, $\mathbf{K}$ of a subdomain is determined by substituting Eq. (16):

$$\mathbf{K} = \mathbf{V}_{21}\mathbf{V}_{11}^{-1}. \tag{17}$$

Upon assembly of all individual subdomain stiffness matrices into a global stiffness matrix, the displacement solution is sought according to conventional FEM procedures.

**Proposed Scheme**

The proposed scheme primarily entails a drop-in replacement for calculating the displacement field. By exploiting hierarchical image decomposition techniques, regions with homogeneous material properties are identified and represented by a single element, such that coarser, adaptive meshes result with significantly fewer DOFs than with conventional grid approaches. Although remeshing is required for each iteration, the implemented decomposition techniques are economical and the resulting computational toll is easily recuperated by solving a forward problem with significantly fewer DOFs, even on modestly sized example domains. The steps differ slightly between 2D and 3D, since in 3D the precomputation procedure employed for 2D would result in 4096 unique element realizations, whose construction requires building a substantial library and accompanying algorithmic logic. Instead, the unique element realizations for each 3D mesh are determined, which are computed once and subsequently cloned for the remaining mesh. Empirically, only a fraction of possible element realizations exist simultaneously on a mesh.
The proposed scheme comprises the following steps:

1. **Precompute**
   For the 2D case, the 16 possible element realizations are precomputed with Young's modulus equal to one.

2. **Initialize and begin TO loop**
   Only the filter must be prepared. Conversely to conventional approaches, remeshing at each iteration is required, rendering the preparation of sparse stiffness matrix assembly vectors [1] obsolete.

3. **Calculate displacement field**
   The grid of design variables is fed as a gray-scale image to the decomposition algorithm, which outputs an analysis-ready hierarchical mesh. Color-encoded regions are automatically recognized and resolved. Their inscribed operations are then applied. Once the stiffness properties of each element are identified, the displacement field is calculated analogous to the conventional FEM.

4. **Determine compliance**
   The compliance at each iteration is calculated as the product of the system displacement field and the force vector.

5. **Determine sensitivities**
   The sensitivities of the design variables are evaluated element-wise, by iterating over each subdomain. Since the subdomains are of variable size, the calculated sensitivities must be normalized per unit volume.

6. **Filter sensitivities**
   Standard mesh-independency filtering techniques may be applied as necessary.

7. **Design variable pro-/demotion**
   Solving the optimization problem, for example by optimality criterion (OC) approach identifies, which design variables to promote or demote, i.e., assign or subtract material. While B/ESO, SERA and level-set approaches result in black-and-white outputs, SIMP-based approaches introduce intermediate material distributions. This is rectified by either employing a Heavyside projection [1] or a gray-scale filter [13, 22]. The updated design variables form the input for the subsequent iteration.

8. **Export of results**
   Upon completing the analysis, the hierarchical mesh is thresholded to yield the optimized system geometry. Simple methods permit exporting the hierarchical mesh to STL format for subsequent additive manufacturing.

In order to further accelerate the analysis procedure, which is constrained by the solution of the forward problem, so-called hard-kill variants have been proposed. These differ from standard soft-kill approaches in how they treat void elements: Soft-kill approaches assign a very small stiffness, typically $10^{-9}$, which impacts the conditioning of the numerical problem, while hard-kill approaches disregard such elements entirely. Not all problems, however, are amenable to this approach, since multiple independent substructures potentially arise during analysis, leading to numerical instabilities.

## Numerical Examples

Three numerical examples are examined in this paper:

1. A thick cantilever subject to a point load at mid-height.
2. An L-shaped bracket with prescribed material distributions and multiple load cases.
3. The 3D wheel.

The first example verifies the proposed method, while the second showcases the extended capabilities, by color-encoding system and analysis information directly into the input image. Having thoroughly discussed the 2D behaviour, we extend the analysis to the 3D wheel problem. For each of the numerical examples, we couple the proposed method with a different TO variant, e.g., B/ESO, SERA and SIMP. One could have equally chosen to employ a level-set based method as an alternative to the ones listed prior.

### *Thick Cantilever*

A thick cantilever subject to a point load at mid-height (Fig. 4) is considered. The width and height are discretized by 512 and 256 pixels respectively. The prescribed volume fraction f is chosen as 0.4. The penalty exponent of both SIMP and BESO approaches is equal to 3. Following [1], a filter utilizing Matlab's built-in `conv2` function is implemented with a radius of 16 pixels. For the BESO, the evolutionary volume ratio parameter is set

to 0.1. Elements arising during quadtree decompositon are limited in size to $\leq 32$ pixels. For the first iteration the discretization of the conventional grid is adopted. This is necessary, since the initial homogeneous material distribution would lead to a too coarse discretization, biasing the calculated sensitivities.

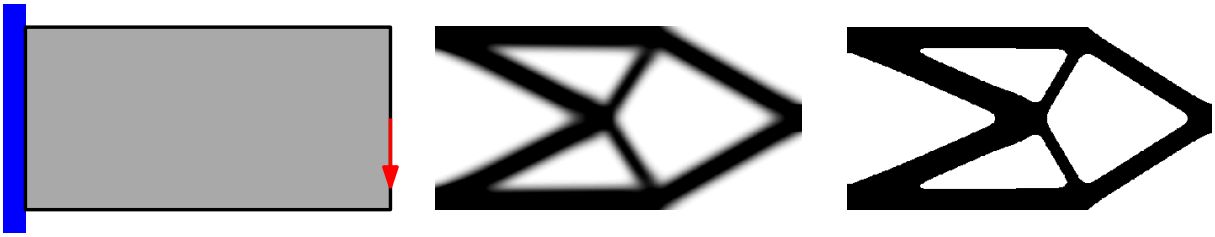


**Figure 4: From left to right: Thick cantilever setup, SIMP reference solution and BESO baseline.**

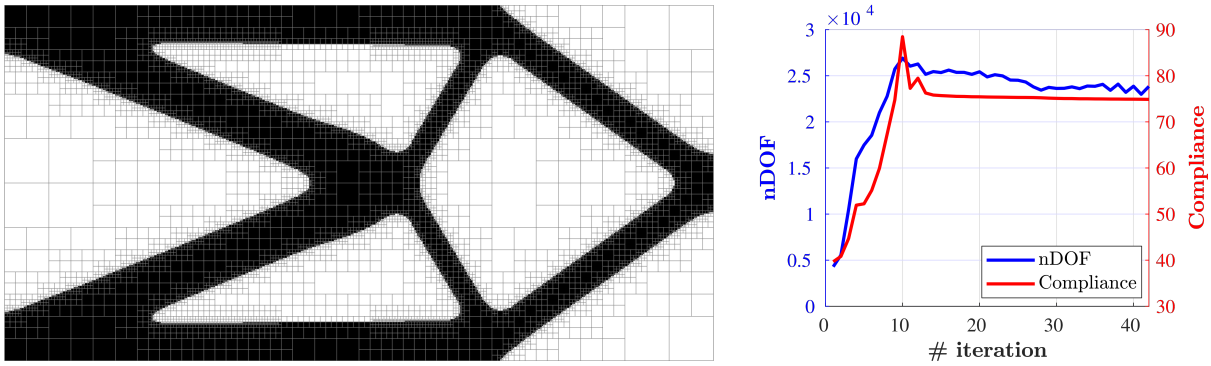



**Figure 5: BESO+QT optimized topology with evolution of DOFs and compliance across iterations.**

In order to verify the proposed scheme, we investigate an elementary case: TO schemes, which result in black-and-white output, i.e., do not posses noticeable transitions in material distributions, permit the adaptive meshing algorithm to develop its full potential and minimize the DOFs required for analysis. Therefore, we contrast our scheme, which we term BESO+QT, to a BESO baseline [10]. Since, BESO follows a heuristic approach to TO, and it can therefore be mislead to local minima under certain circumstances [18], we first supply a SIMP reference solution [1]. Both the reference solution and the baseline depict the same resulting topology (Fig. 4). The BESO+QT (Fig. 5) is indistinguishable from the BESO baseline. The difference in compliance (Tab. 1) resulting from the SIMP and BESO approaches stems from the presence of transition material in the SIMP, leading to a slightly more flexible structure and therefore higher compliance. Given a higher value for the penalization parameter, this difference diminishes. The discrepancy in compliance between BESO and BESO+QT is attributed to the discretization: A coarse discretization results in a stiffer structure and therefore lower compliance. This is evident, since the amount of DOFs present during analysis is reduced by more than one order of magnitude. This in turn significantly alleviates the computational burden and associated memory requirements (Fig. 5, DOFs evolution). The two additional iterations required to reach the stopping criterion, i.e., a 5% increase, represents a negligible difference to the base line case. In this contrived example, in which all three methods share a common implementation, differing only in the method of solving the forward problem, a glimpse

of the computational potential of the proposed scheme is possible: For this specific numerical example, an analysis concludes almost 5x faster, when employing the proposed scheme.

**Table 1: A comparison of results for SIMP reference solution, BESO baseline and proposed BESO+QT.**

| Method | nIt. | Compliance | DOFs | time [s] |
|---|---|---|---|---|
| SIMP | 49 | 87.6 | 263'682 | 110 |
| BESO | 40 | 75.6 | 263'682 | 98 |
| BESO+QT | 42 | 74.8 | 23'846 | 20 |

*Modified L-bracket*

A modified L-bracket setup is considered (Fig. 6). In this example the SERA is employed, which remedies the drawbacks of the BESO scheme [2], while maintaining black-and-white solutions. For this analysis, only a color-encoded input image is provided. The proposed scheme automatically recognizes significant regions and their associated operations during the automated decomposition phase. The colors blue, red, green, white and gray correspond to boundary conditions, loadings, solid, void and domain pixels respectively. Input images are easily constructed by small scripts or obtained from, e.g., medical imaging applications. Especially from the user perspective, manipulating analysis parameters by color-encoding simplifies the overall process and permits direct visual verification of the input prior to analysis. Further, it facilitates testing of variants. In this example, the domain is discretized by 512 pixels in each direction. A volume fraction of 0.3 is specified. The `conv2`-type filter is employed with radius equal 12 pixels. The SERA parameters PR, SR and B are chosen as 0.03, 1.3 and 0.003 respectively. All quadtree elements are of size $\leq 128$.

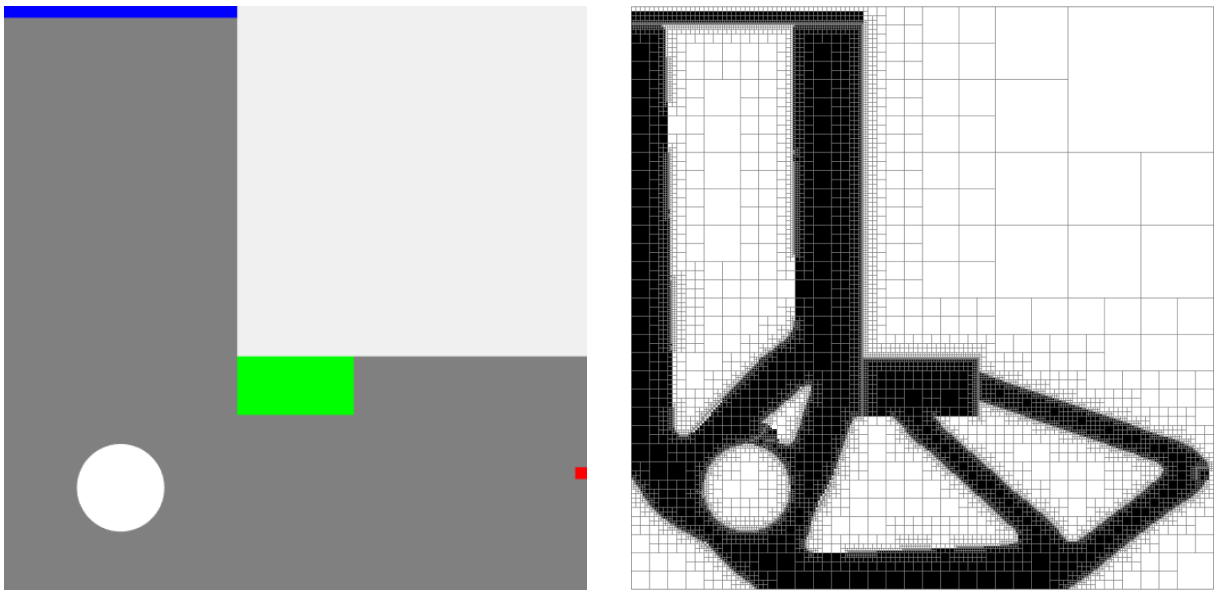


**Figure 6: Topology optimization of an L-bracket via automated image-based analysis. Input image (left) and resulting topology (right).**

For more involved geometries, gratuitous discretization may arise due to slight misalignment with respect to the optimal quadtree meshing strategy. This is apparent surrounding the blue and red regions (Fig. 6), where a pertubation by 1 pixel triggers excessive refinement to accurately capture the domain's geometry. To demonstrate this general case, the L-bracket is treated as is by the quadtree decomposition, which results in 32'694 DOFs, while a shifted, scaled and therefore better aligned domain geometry results in 26'554 DOFs respectively. The conventional method employing the fine grid discretization treats 526'338 DOFs. In this example, the required DOFs are reduced by over an order of magnitude. For such small examples, a typical forward analysis completes in $\leq 0.6$ seconds on a modest desktop computer running in serial.

*3D-Wheel*

In this example we extend SBFEM-powered TO to 3D problems. The established 3D wheel problem is studied (Fig. 7): The width, depth and height are discretized by 80, 80 and 40 pixels respectively. All four corners at the bottom edge are fully restrained. A point load is applied in downward direction in the middle of the bottom surface. A volume fraction of 0.075 is sought. The penalty parameters for SIMP and gray-scale filter are chosen as 3 and 2 respectively. The filter radius is given as 3 pixels. All octree elements are $\leq 16$ pixels in size. A sensitivity filter is employed. A Young's modulus of $E = 10^{-9}$ denotes material voids. The analysis concludes after 200 iterations.

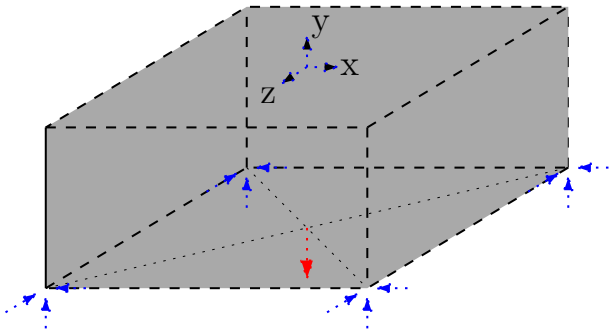



**Figure 7: Problem domain of the 3D wheel benchmark.**

The SIMP with gray-scale filter is utilized. The addition of the gray-scale filter is crucial to obtaining computational efficiency, since regions of transitional material distributions, which impact the proposed hierarchical meshing techniques, are minimized. The preconditioned conjugate gradients method (pcg) is employed to obtain the solution of the forward problem. Scaling is performed to combat the conditioning issues due to the treatment of void elements [28] with secondary preconditioning by incomplete Cholesky decomposition. This significantly reduced the amount of iterations required for convergence. A strict convergence tolerance of $10^{-9}$ was specified to contain any divergence from the reference solution to the adaptive discretization choice. In Fig. 8 the final topology is depicted along with a sample of the octree mesh.
It is demonstrated, that the amount of DOFs and number of non-zero entries in the stiffness matrix (nnz) is significantly reduced (Fig. 9) by employing the proposed scheme, especially after the initial phase in the TO process. Although up to 4096 unique element realization are possible on the octree, only a handful exist at each iteration (Fig. 9). Computational effort is alleviated by only computing the element stiffness matrices for the unique elements and cloning the remainder. However, the overwhelming computational burden still resides in the solution of the forward problem.

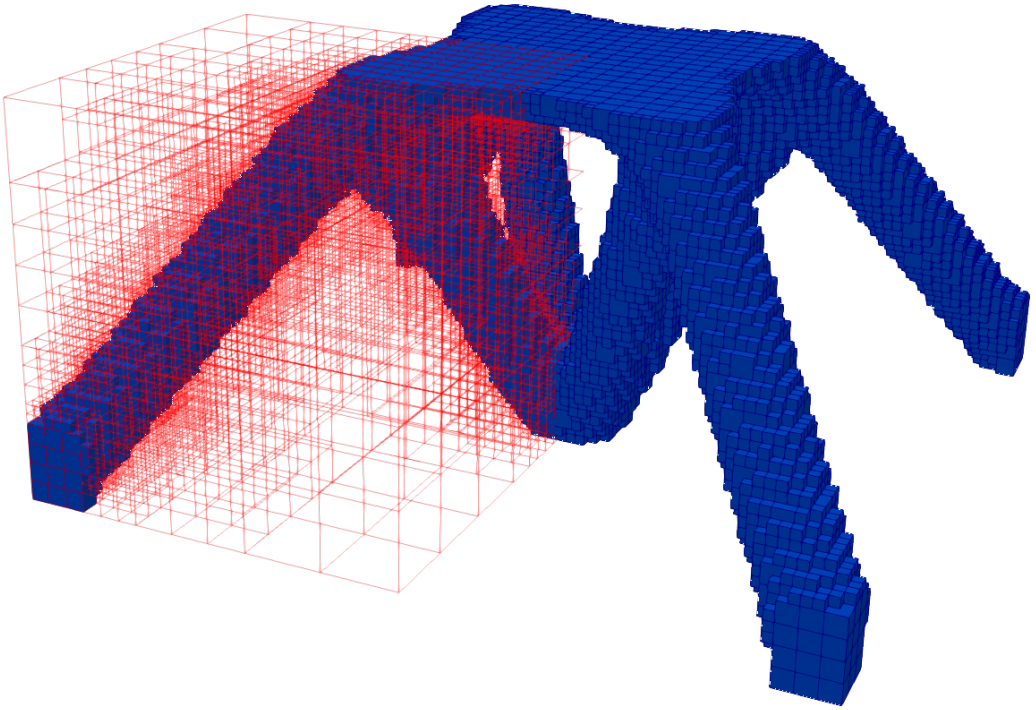


**Figure 8: Resulting topology for 3D wheel benchmark in blue with sample octree mesh in red.**

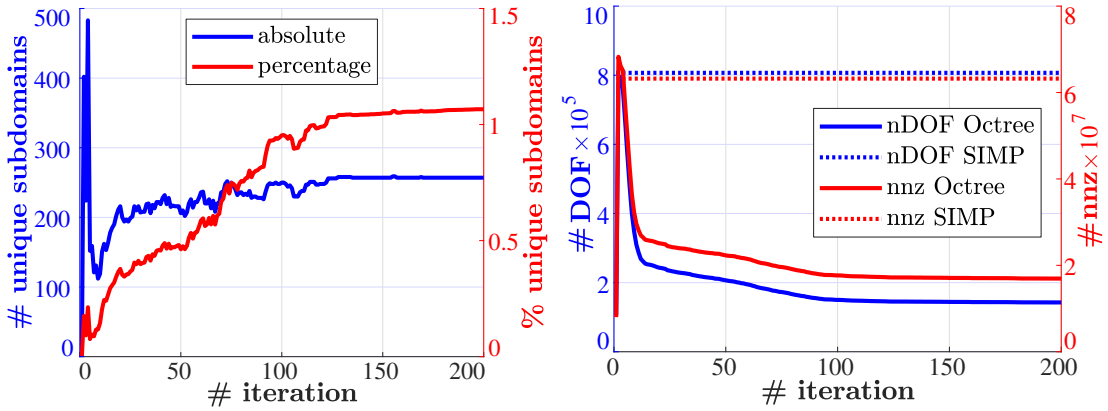



**Figure 9: Evolution of hierarchical mesh quantities across iterations.**

The final topology is readily exported to STL format by operating on the octree, greatly facilitating subsequent additive manufacturing procedures.

## Conclusions

This paper advances the current state-of-the-art in TO by employing hierarchical meshes coupled with polytope-based numerical methods, which alleviate issues commonly associated with hanging nodes. Enforcing a balancing operation on the mesh further limits the amount of feasible element realizations, which is exploited through precomputation of element stiffness matrices and cloning of element properties. Moreover, we propose a novel and intuitive scheme for interacting with the analysis parameters by color-encoding input images, which the decomposition techniques automatically translate into analysis-ready meshes. Further, we extend SBFEM-powered TO to 3D problems and demonstrate an ef-

fective means for exporting results to STL format for subsequent additive manufacturing (3D printing).
TO relying on hierarchical meshes is shown to require only a fraction of the DOFs demanded by traditional grid approaches, significantly reducing the computational toll. The three investigated numerical examples consistently demonstrate a remarkable reduction in required DOFs and memory requirements, without incurring a perceivable loss of accuracy.
Considering the generality of this approach in handling color-coded input, an extension to multi-material TO presents an intriguing direction for future work.

## Acknowledgements

This research was performed under the auspices of the Swiss National Science Foundation (SNSF), Grant # 200021_153379, A Multiscale Hysteretic XFEM Scheme for the Analysis of Composite Structures.

### References

[1] Andreassen, E., Clausen, A., Schevenels, M., Lazarov, B. S., and Sigmund, O. Efficient topology optimization in MATLAB using 88 lines of code. *Structural and Multidisciplinary Optimization 43*, 1 (Jan. 2011), 1–16.

[2] Ansola Loyola, R., Querin, O. M., Garaigordobil Jimnez, A., and Alonso Gordoa, C. A sequential element rejection and admission (SERA) topology optimization code written in Matlab. *Structural and Multidisciplinary Optimization 58*, 3 (Sept. 2018), 1297–1310.

[3] Beiro Da Veiga, L., Brezzi, F., Cangiani, A., Manzini, G., Marini, L. D., and Russo, A. BASIC PRINCIPLES OF VIRTUAL ELEMENT METHODS. *Mathematical Models and Methods in Applied Sciences 23*, 01 (Jan. 2013), 199–214.

[4] Bendse, M. P., and Sigmund, O. *Topology Optimization*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[5] de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. *Computational Geometry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[6] Deeks, A. J., and Wolf, J. P. A virtual work derivation of the scaled boundary finite-element method for elastostatics. *Computational Mechanics 28*, 6 (jun 2002), 489–504.

[7] Gupta, A. K. A finite element for transition from a fine to a coarse grid. *International Journal for Numerical Methods in Engineering 12*, 1 (1978), 35–45.

[8] Hoshina, T. Y. S., Menezes, I. F. M., and Pereira, A. A simple adaptive mesh refinement scheme for topology optimization using polygonal meshes. *Journal of the Brazilian Society of Mechanical Sciences and Engineering 40*, 7 (July 2018), 348.

[9] Hu, Z., Lin, G., Wang, Y., and Liu, J. A Hamiltonian-based derivation of Scaled Boundary Finite Element Method for elasticity problems. *IOP Conference Series: Materials Science and Engineering 10* (jun 2010), 012213.

[10] Huang, X., and Xie, Y. M. *Evolutionary Topology Optimization of Continuum Structures: Methods and Applications*. John Wiley & Sons, Ltd, Chichester, UK, Apr. 2010.

[11] Lan, L. T. T., and Boucher, A. Simplified quadtree image segmentation for image annotation.

[12] Li, C., and Tong, L. Topology optimization of incompressible materials based on the mixed SBFEM. *Computers & Structures 165* (Mar. 2016), 24–33.

[13] Liu, K., and Tovar, A. An efficient 3d topology optimization code written in Matlab. *Structural and Multidisciplinary Optimization 50*, 6 (Dec. 2014), 1175–1196.

[14] Liu, Y. J., Mukherjee, S., Nishimura, N., Schanz, M., Ye, W., Sutradhar, A., Pan, E., Dumont, N. A., Frangi, A., and Saez, A. Recent Advances and Emerging Applications of the Boundary Element Method. *Applied Mechanics Reviews 64*, 3 (Mar. 2012), 030802.

[15] Man, H., Song, C., Natarajan, S., Ooi, E. T., and Birk, C. Towards Automatic Stress Analysis using Scaled Boundary Finite Element Method with Quadtree Mesh of High-order Elements. *arXiv:1402.5186 [math]* (Feb. 2014). arXiv: 1402.5186.

[16] Otomori, M., Yamada, T., Izui, K., and Nishiwaki, S. Matlab code for a level set-based topology optimization method using a reaction diffusion equation. *Structural and Multidisciplinary Optimization 51*, 5 (May 2015), 1159–1172.

[17] Panesar, A., Brackett, D., Ashcroft, I., Wildman, R., and Hague, R. Hierarchical remeshing strategies with mesh mapping for topology optimisation: Hierarchical remeshing strategies with mesh mapping for topology optimisation. *International Journal for Numerical Methods in Engineering 111*, 7 (Aug. 2017), 676–700.

[18] Rozvany, G. I. N. A critical review of established methods of structural topology optimization. *Structural and Multidisciplinary Optimization 37*, 3 (Jan. 2009), 217–237.

[19] Salazar de Troya, M. A., and Tortorelli, D. A. Adaptive mesh refinement in stress-constrained topology optimization. *Structural and Multidisciplinary Optimization 58*, 6 (Dec. 2018), 2369–2386.

[20] Sanders, E. D., Pereira, A., Aguil, M. A., and Paulino, G. H. PolyMat: an efficient Matlab code for multi-material topology optimization. *Structural and Multidisciplinary Optimization 58*, 6 (Dec. 2018), 2727–2759.

[21] Saputra, A., Talebi, H., Tran, D., Birk, C., and Song, C. Automatic image-based stress analysis by the scaled boundary finite element method: AUTOMATIC IMAGE-BASED STRESS ANALYSIS BY THE SCALED BOUNDARY FEM. *International Journal for Numerical Methods in Engineering 109*, 5 (feb 2017), 697–738.

[22] Sigmund, O. Morphology-based black and white filters for topology optimization. *Structural and Multidisciplinary Optimization 33*, 4-5 (Feb. 2007), 401–424.

[23] Song, C. A matrix function solution for the scaled boundary finite-element equation in statics. *Computer Methods in Applied Mechanics and Engineering 193*, 23-26 (jun 2004), 2325–2356.

[24] Song, C. *The scaled boundary finite element method: intorduction to theory and implementation.* John Wiley & Sons, Hoboken, New Jersey, 2018.

[25] Song, C., and Wolf, J. P. The scaled boundary finite-element methodalias consistent infinitesimal finite-element cell methodfor elastodynamics. *Computer Methods in Applied Mechanics and Engineering 147*, 3-4 (aug 1997), 329–355.

[26] Tabarraei, A., and Sukumar, N. Adaptive computations on conforming quadtree meshes. *Finite Elements in Analysis and Design 41*, 7-8 (Apr. 2005), 686–702.

[27] Wang, S., de Sturler, E., and Paulino, G. H. Dynamic Adaptive Mesh Refinement for Topology Optimization. *arXiv:1009.4975 [cs, math]* (Sept. 2010). arXiv: 1009.4975.

[28] Wang, S., Sturler, E. d., and Paulino, G. H. Large-scale topology optimization using preconditioned Krylov subspace methods with recycling. *International Journal for Numerical Methods in Engineering 69*, 12 (Mar. 2007), 2441–2468.

[29] Wolf, J. P. *The scaled boundary finite element method.* J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, USA, 2003.

# Data Management Challenges of Exascale Scientific Simulations: A Case Study with the Gyrokinetic Toroidal Code and ADIOS

**†Lipeng Wan[1], Kshitij V. Mehta[1], Scott A. Klasky*[1], Matthew Wolf[1], H. Y. Wang[2], W. H. Wang[2], J. C. Li[2], and Zhihong Lin[2]**

[1]Computer Science and Mathematics Division, Oak Ridge National Laboratory, USA.
[2]Department of Physics and Astronomy, University of California Irvine, USA.

*Presenting author: klasky@ornl.gov
†Corresponding author: wanl@ornl.gov

## Abstract

The push towards exascale computing and the recent introduction of multi-petascale supercomputers have enabled science applications to run complex simulations. However, the gap between compute and I/O has grown wider, even as applications seek to generate and persist increasing amounts of data. Optimizing I/O is challenging and remains a bottleneck at scale. In this paper, we present initial I/O performance results of running Gyrokinetic Toroidal Code (GTC) on Summit, a 200 Petaflop system at Oak Ridge National Laboratory. To manage the complex data in GTC, we use ADIOS, an I/O and data management middleware that provides a rich set of APIs to manage and interact with scientific data. We discuss optimizations performed to obtain improvements in I/O performance and identify a set of challenges that will drive the design and development of next generation data management libraries.

**Keywords:** Computational, GTC, ADIOS, I/O, Scientific Data, Fusion Simulation

## Introduction

With the recent push towards exascale computing, the arrival of systems such as Summit[12], Sierra[22], and Sunway TaihuLight[21] has broken the 100 Petaflop peak performance barrier. This has paved the way for science applications to run more complex simulations. Supercomputer architects are putting more focus on making systems more power efficient, in addition to increasing the memory per core on a node. However, the evolution of I/O technologies has not kept up with other architectural innovations, and the gap between compute capability and I/O bandwidth continues to grow wider.

Science applications that look towards exploiting the compute capabilities of modern systems tend to generate increasing amounts of data. It is common for applications running at full scale to write half the size of system memory for checkpointing. For example, the Chimera molecular modeling and simulation system [15] creates 160 TB checkpoints on the Titan supercomputer at Oak Ridge National Laboratory. On Summit, which has 512 GB of main memory per node, checkpoint data would be close to a Petabyte in size.

The Gyrokinetic Toroidal Code (GTC)[1] is a well-known particle-in-cell application that simulates the movement of magnetic particles in a confined fusion plasma. It generates a variety of data that differ in volume and velocity. Checkpoint and particle data generated over the course of a running simulation are projected to be in the range of a few Petabytes. Management of this data is a daunting task; both the file system capacity and the available I/O bandwidth are limiting factors. The GPFS parallel file system on Summit provides a theoretical peak bandwidth of 2.2 TB/s. Writing a Petabyte of data at near peak bandwidth would take over 500 seconds. The use of non-volatile memory, such as the local NVME chips

on Summit nodes, can help in reducing the time to write bursty, high volume data, but the efficient utilization of NVME is not straight-forward. In combination with high volume and high frequency particle data, I/O can easily become a bottleneck. In addition, projects on Summit have a quota of only 100 TB on the parallel file system, which means the application cannot store all the data it generates to persistent storage. Thus, applications need to make smart decisions about how to write data and what data must be preserved.

ADIOS (Adaptable I/O System) [20] is a well-known I/O framework that is used across many science domains for optimizing I/O and data management. ADIOS provides a self-describing data format and a log-based data organization. As part of its data model, ADIOS generates metadata to preserve the logical ordering of a file. It provides a rich set of APIs to tune various parameters related to I/O to obtain optimal performance. As part of this work, the I/O component of GTC has been adapted to utilize ADIOS. ADIOS provides optimized ways to store data which are comprised of physics variables recorded over thousands of timesteps. In this paper, we present preliminary results of using GTC with ADIOS on Summit. We discuss the optimizations performed in ADIOS that lead to performance improvements in GTC. We highlight the challenges that are faced with increasing data volumes and metadata overhead, and discuss the design of a new metadata format in ADIOS for exascale data. The findings of our study provide valuable insight into the I/O characteristics of scientific simulation codes and help us estimate the areas where bottlenecks can occur when applications are run at exascale. Moreover, we also investigate and evaluate some potential approaches to improving the efficiency of scientific data management and discuss the future research directions.

## Background

### *Summit*

Summit is an IBM system located at the Oak Ridge Leadership Computing Facility [12]. With a theoretical peak double-precision performance of approximately 200 PF, it is currently rated as the fastest supercomputer in the world as of this writing [13]. The basic building block of Summit is the IBM Power System AC922 node. Each of the approximately 4,600 compute nodes on Summit contains two IBM POWER9 processors and six NVIDIA Volta V100 accelerators and provides a theoretical double-precision capability of approximately 40 TF. Each POWER9 processor is connected via dual NVLINK bricks, each capable of a 25GB/s transfer rate in each direction. Nodes contain 512 GB of DDR4 memory for use by the POWER9 processors and 96 GB of High Bandwidth Memory (HBM2) for use by the accelerators. Additionally, each node has 1.6TB of non-volatile memory that can be used as a burst buffer. Summit mounts a POSIX-based IBM Spectrum Scale parallel file system called Alpine. It consists of 77 IBM Elastic Storage Server (ESS) GL4 nodes and has a maximum capacity of 250 PB. Each IBM ESS GL4 node is constituted by two dual-socket IBM POWER9 storage servers, and a 4X EDR InfiniBand network for up to 100 Gbit/sec of networking bandwidth. The maximum performance of the final production system is about 2.5 TB/s for sequential I/O and 2.2 TB/s for random I/O.

### *Gyrokinetic Toroidal Code*

As a particle-in-cell code, GTC [1] tracks individual charged particles in a Lagrangian frame in a continuous phase-space, whereas the moments (number density, charge density and current density etc) of particle distribution of different species (thermal ion, thermal electron, fast ion, fast electron, etc.) are simultaneously computed on a stationary Eulerian field mesh. The electromagnetic fields are then solved on the field mesh using proper combinations of Poisson equation, Ampere’s law, Faraday’s law and force-balance equations with finite

difference and finite element methods. This field mesh is also used to interpolate the local electromagnetic fields at the particle positions in phase-space. GTC has been extensively applied to study collisional transport [2], energetic particle transport [3], microturbulence [4], Alfven eigenmodes [5], kink modes, and tearing modes in fusion plasmas.

GTC carried out the first fusion production simulations at tera-scale on the Seaborg computer at NERSC in 2001 [6] and at peta-scale in 2008 as an early science application on the Jaguar computer at OLCF [7] GTC is a key production code in the DOE SciDAC ISEP Center [8] and one of the two fusion codes selected by the Center for Accelerated Application Readiness (CAAR) [9], a DOE ASCR program to prepare prominent codes across all DOE supported research portfolio for the emerging exascale computers. GTC is also one of the production codes in the Summit acceptance benchmark suite.

GTC uses MPI domain decomposition, particle decomposition, and OpenMP shared memory partitioning to scale up to millions of CPU cores to take advantage of the memory hierarchy. Thanks to closed collaborations with computational scientists through ASCR CAAR and SciDAC ISEP projects, GTC has been ported to GPU-based supercomputers including Titan and Summit early access systems at OLCF. The computationally expensive particle and field subroutines are fully ported and optimized on GPU using OpenACC and CUDA. Using realistic fusion simulation parameters [10], GTC shows near-ideal weak scaling performance up to the full capacity of Titan computer with GPU. GTC has recently demonstrated good scalability on more than 20% of Summit and achieved an unprecedented speed of one trillion particle pushes in 2 seconds wall-clock time using 928 nodes on Summit [11].

Preparing GTC for large simulations that generate large volumes of data requires special consideration. GTC generates a combination of diagnostics data, field data, checkpoint-restart files, and particle data. These data components differ vastly in terms of their output frequency, number of variables contained within, and size. Table 1 provides a summary of the various data items along with some of their important characteristics in a typical large simulation.

**Table 1: Summary of the different types of data generated by GTC**

| **Data Type** | **Frequency of Output** | **Size of the data per output** | **Number of variables** |
|---|---|---|---|
| Diagnostics | Every timestep or every few timesteps | Megabytes | 50 - 100 |
| Snapshot | Once every hour | Megabytes up to many Gigabytes | 50 - 100 |
| Checkpoint | Once every hour | Terabytes, potentially up to Petabytes | < 20 |
| Restart | Once | Terabytes, potentially up to Petabytes | < 20 |
| Field data | Every timestep | Gigabytes | < 20 |
| Particles | Every timestep | Few hundred terabytes up to petabytes | < 20 |

In general, diagnostics data are high velocity data that are written frequently, and read back multiple times for analysis, both online as well as offline. Checkpoint data are write-once; that is, they are written for resilience purposes with the intention of simulation restarts. Checkpoint and particle data are high volume and writing them usually consumes significant resources. Furthermore, checkpoint and restart data are high variety data, as their underlying variables are comprised of scalars, vectors, and multi-dimensional arrays.

GTC traditionally uses POSIX I/O, where data are written to binary files. Based on historical knowledge, multiple processes writing data to the same file generally yields poor performance. To circumvent this commonly known N-1 write pattern, all high-volume data in GTC are written such that each process writes its data to an independent file. For many outputs such as field data, data are written into new files per process per output timestep. However, such an approach typically suffers from two issues. One, the metadata overhead for parallel file systems on supercomputers is a massive challenge and often forms the bottleneck for intensive I/O. At extreme scales, this will be one of the most important challenges. Two, a logical file being split into multiple sub-files to improve I/O performance puts additional burden on the developers to ensure that data is read back correctly. Inspecting such data offline can be highly prohibitive without specialized tools.

For the different types of outputs in GTC, optimizing I/O is challenging as there is no 'one-size-fits-all' approach that can be used to optimize I/O. Different strategies need to be adopted for different components. At the exascale, the sheer size of the data can be an issue as many Petabytes of data can be generated over the course of a few hours.

*ADIOS*

ADIOS is an I/O framework which provides a simple and flexible way for scientists to describe the data in their code. ADIOS provides highly optimized I/O routines that allow users to read and write data in an optimal fashion for the target architecture. In the ADIOS design, variables and steps are first-class concepts. ADIOS provides the API to define a variable which may be a simple scalar or a global array partitioned amongst processes. It provides the ability to write variables to a file and append "steps" to it, so that a single file can contain information about a physical quantity as it evolves over time. ADIOS stores data in a proprietary, log-based, file format named BP (current version 3). For every process that writes data to a file, it creates an independent sub-file and writes metadata in the global file container to reconstruct the original order of the file. The log-based data organization allows ADIOS to write each process's output into a separate chunk of file or aggregate several processes' outputs into a smaller number of files, which can maximize the I/O bandwidth. These operations are kept transparent from the end user. The self-describing nature of the file allows users to inspect the file outside the scope of the running simulation using pre-bundled tools that come with the library.

The ADIOS API is designed as a publish-subscribe library. Applications can write data directly to the underlying storage or publish data so that it can be read by processes that subscribe to it. This allows various applications to couple through in-memory transports, which plays an important role in providing the ability to analyze data in situ. This is done through the use of built-in ADIOS "engines" that provide users with a way to select how data must be published. For example, there are engines to read/write data from/to underlying storage, communicate with other applications through in-memory data exchange, as well as perform a wide area network transfer of data to remote sites. ADIOS also provides the ability to transform data through various compression and reduction methods. Users may set these

options directly in their code or in an XML file. Furthermore, ADIOS provides ways to tune parameters for different engines and operations to optimize I/O performance.

**Data Management in GTC using ADIOS**

In this section, we discuss our efforts at improving I/O performance in GTC through the use of ADIOS and present results from experiments run on the Summit supercomputer.

At large scale, GTC suffers from issues arising due to sub-optimal I/O patterns and the metadata overhead on the file system. To overcome these limitations, ADIOS is now being used to manage data in GTC. The core design features of ADIOS lead to performance improvements out of the box for various data output by GTC. As timesteps can be appended to existing files, new files per output timestep are no longer required. Additionally, the number of writers, and thus the number of sub-files for different outputs in GTC has been tuned to obtain good performance by alleviating the metadata overhead on the file system. Table 2 lists the number of output files that were created by the original POSIX I/O version of GTC and the ADIOS version of GTC for a run that simulated 10,000 timesteps using 3072 processes on Summit. No particle data was generated was generated for these experiments.

**Table 2: Number of output files created by GTC when using POSIX vs. using ADIOS.**

| **Output filename** | **Number of files in the POSIX I/O module** | **Corresponding ADIOS output file container** | **Number of files created by the ADIOS module** |
|---|---|---|---|
| equilibrium.out | 1 | equilibrium.bp | 1 |
| data1d.out | 1 | data1d.bp | 1 |
| history.out | 1 | history.bp | 1 |
| snapshot#.out | 10,000 | snapshot.bp | 1 |
| phi#.out | 32,000 | phi3d.bp | 2048 |
| restart1.# | 3072 | restart1.bp | 2048 |
| restart2.# | 3072 | restart2.bp | 2048 |

It can be seen that the POSIX I/O version of GTC creates almost 50,000 files, whereas the ADIOS version creates just over 6000 files on storage with 4 writers per node, with further potential to reduce the number of sub-files. This leads to significant improvements in the overall I/O performance. This is further demonstrated by the performance improvements in writing snapshot data, as shown in Figure 1. Recall from Table 1 that snapshot data are written to a new file at every output step. Using ADIOS, snapshot data are appended to an existing file which leads to 50x improvement over the POSIX version of the application.

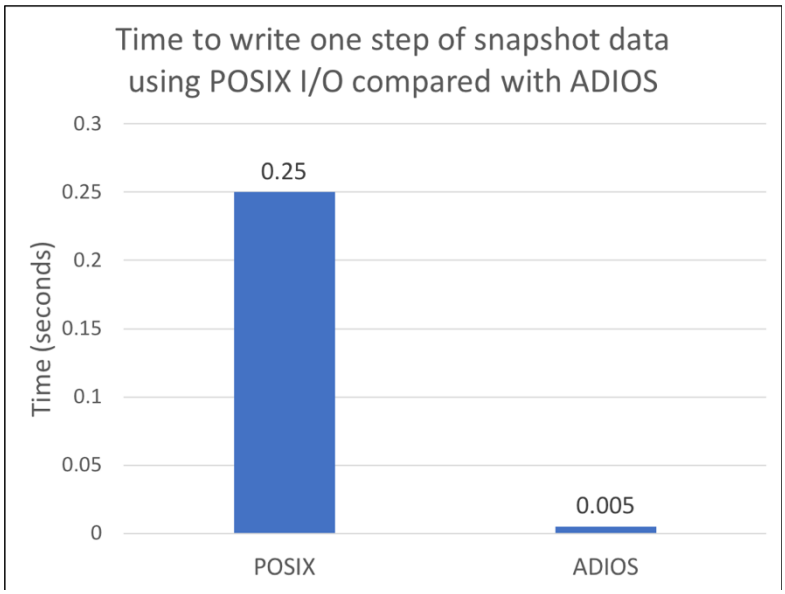



**Figure 1: Comparison of the time taken to write one snapshot of GTC on Summit to the GPFS parallel file system for POSIX vs. ADIOS**

Figure 2 shows the I/O performance for writing checkpoint data from GTC to the GPFS parallel file system using ADIOS. The simulation was run on 512 nodes with 6 processes per node, for a total of 3072 processes. The aggregate size for each checkpoint was 2.6 Terabytes. For our tests, we compute 5000 simulation timesteps and write 50 checkpoints. We show results for varying number of writers per node in ADIOS. The figure shows that a peak performance of 2.27 TB/s is observed when we use 6 writers per node. Using 6 writers per node displays high variability in I/O performance with an average bandwidth of 1.1 TB/s, whereas using 4 writers per node shows a more consistent average of 1.5 TB/s.

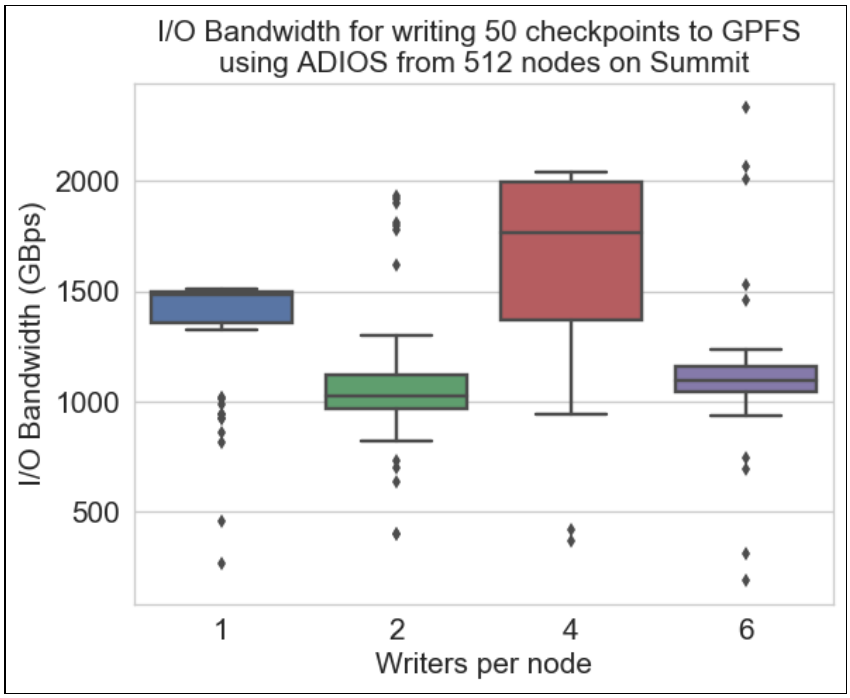



**Figure 2: Using ADIOS to write checkpoint data from GTC to the GPFS parallel file system on Summit. Results show the effect of tuning ADIOS options to control the number of sub-files created on the parallel file system for checkpoint data.**

I/O variability is common on large-scale computing facilities since the I/O bandwidth is shared by hundreds of running jobs simultaneously. Previous studies [16] [17] have shown significant I/O performance variability on DoE's Titan [18] and Cori [19] supercomputers. This can be mitigated to an extent by the use of the local non-volatile memory on Summit nodes. An in-depth study to model the performance of NVME for use in GTC and effective ways to utilize it is an important future work for this research.

## Metadata Optimizations in ADIOS for Large Data

An ADIOS file is a collection of sub-files created by writer processes and metadata information required to recreate the file in its original intended form. With increasingly large output data, the overhead of this metadata is no longer negligible. This will be an important consideration as we move towards the exascale. In this section, a study of the metadata overhead in ADIOS for different types of output data is discussed, along with a design of the next generation file format in ADIOS.

*Factors that affect metadata overhead*

Experiments have been performed to study the impact of the data size, number of variables, number of processes writing data, and the number of timesteps generated by GTC. To study the impact of data volume, the simulation was run on 64 compute nodes on Summit (6 processes per node) for 100 simulation timesteps. In each step, the size of the variables is varied, while the number of variables is kept constant. As shown in Figure 3, an increase in data size leads to an increase in the overall write time but does not show a significant increase in metadata overhead. This is expected, as ADIOS generates metadata for every write issued to the underlying file by every writer process. Thus, for a constant number of write operations, the metadata overhead remains constant. The increase in overhead seen in the figure is attributed to variability in the I/O bandwidth as the size of the metadata is small. However, from the figure, we can observe that if the size of GTC output data is relatively small, the metadata overhead dominates the total I/O overhead.

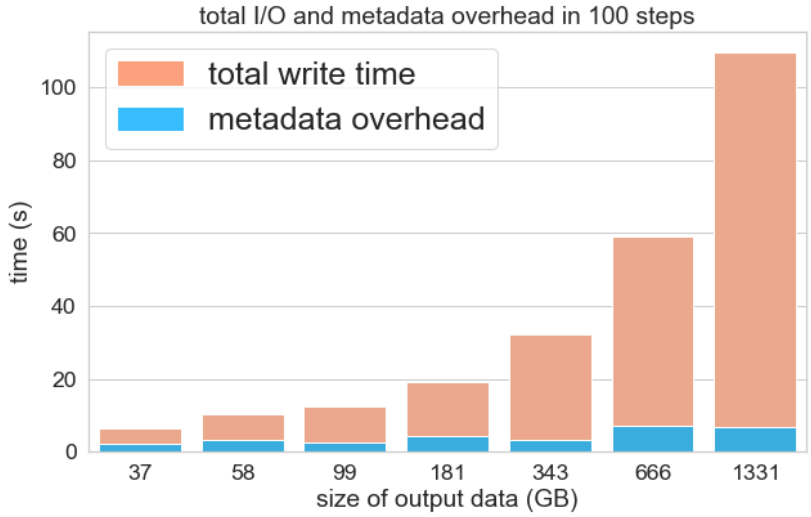



**Figure 3: Impact of increasing data volume on metadata overhead**

At larger scales, an increase in output data also corresponds to an increase in the number of physical quantities written by the application. Figure 4 shows the impact of an increase in the number of variables generated by GTC. Increasing the number of variables causes a

significant increase in the metadata overhead. The amount of metadata generated increases with an increase in the number of variables, as does the time complexity of constructing and writing those metadata items.

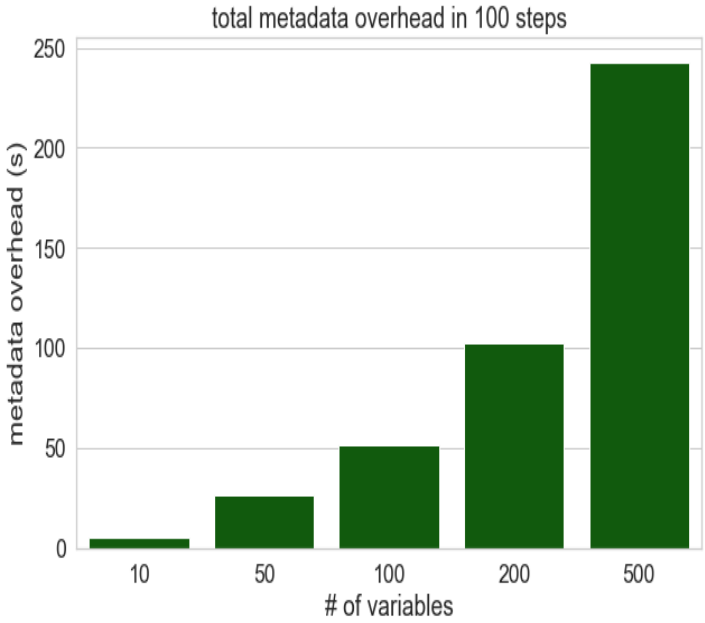



**Figure 4: Impact of increasing number of variables on metadata overhead**

In order to fully utilize the available I/O bandwidth of parallel file systems, ADIOS allows MPI processes of large-scale simulation runs to write data to separate files, which requires metadata to associate data chunks with processes, and recording the offset of the data chunk in the global file. To study the impact of increasing writers, experiments were run with a constant number of variables and timesteps, but varying number of MPI processes. Figure 5 shows that increasing the number of writers shows almost a linear increase in the metadata overhead.

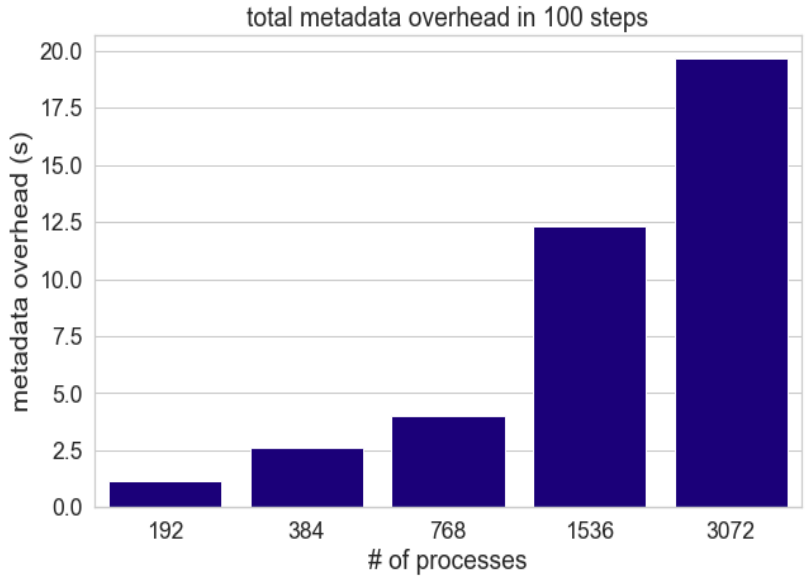



**Figure 5: Impact of increasing the number of writer processes on metadata overhead**

Recall that applications can append "steps" to an ADIOS file. In addition to metadata associated with variables and writer processes, ADIOS also needs to maintain metadata about the steps in the file. At large scale, GTC can run tens of thousands of time steps. To study the impact of increasing number of timesteps in the simulation, experiments were run with a fixed number of variables, output data sizes, and MPI process count, while increasing the number of timesteps simulated by the application. Figure 6 shows that the average metadata overhead per simulation step increases with an increase in the total number of simulation steps. This is because when ADIOS constructs the metadata at each simulation step, it needs to reorganize and serialize metadata of current and all previous steps in memory, and write the serialized metadata to file.

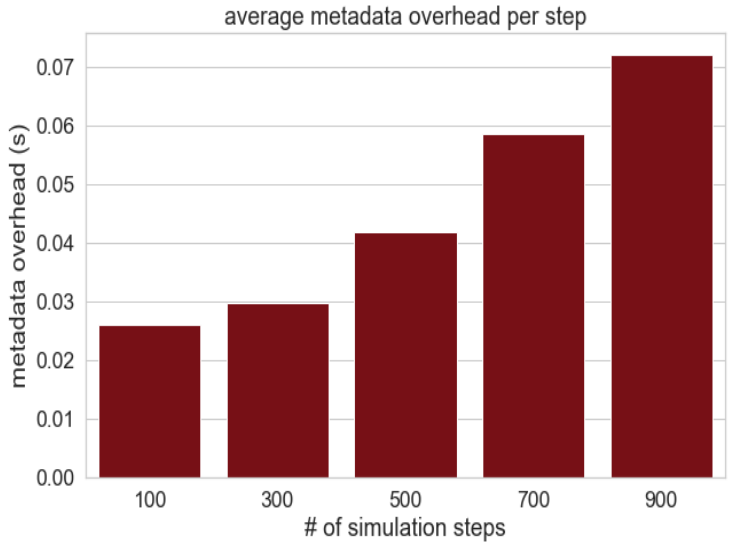



**Figure 6: Impact of increasing number of timesteps in GTC on the metadata overhead**

When combined with a large number of variables, process counts, and simulation timesteps, the metadata overhead in ADIOS can grow quickly. To alleviate the pressure of increasing number of processes, users can tune ADIOS parameters to reduce the number of writers. However, the overhead due to increasing timesteps is a primary concern as reconstructing the full metadata for every step written to an ADIOS file can lead to an exponential increase in metadata overhead. At extreme scale, this is projected to be a major limiting factor in obtaining good I/O performance. In the next section, we discuss performance improvements obtained through a re-designed metadata format for ADIOS that mitigates this issue.

*BP4: The next generation ADIOS file format with improved metadata capabilities*

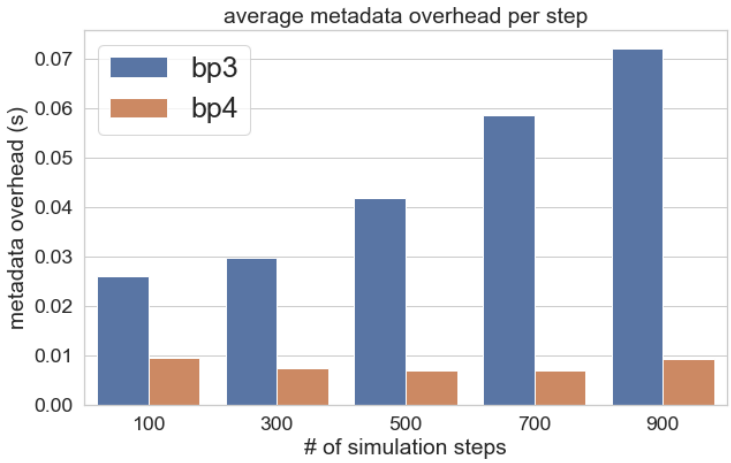



**Figure 7: Per-step metadata overhead of current and new metadata construction approach**

The BP4 file format is the next generation file format for ADIOS that is targeted towards exascale simulations that generate high volumes of data. Emphasis has been put on redesigning the metadata schema to optimize for increasing timesteps. The central concept is the introduction of an index table that stores the metadata offsets that represent timesteps in the global metadata file. Using an index table removes the need to sort the metadata and parse it serially to retrieve information when timesteps are appended to a file. Figure 7 shows that the metadata overhead per step with BP4 is constant for an increasing number of timesteps. Consequently, the total metadata overhead during the entire simulation run is also significantly reduced.

## Conclusions

In this paper, we use Gyrokinetic Toroidal Code (GTC) as a concrete example to demonstrate the data management challenges that arise when we run pre-exascale scientific simulations. We use ADIOS to optimize I/O and data management in GTC. Initial experiments on the Summit supercomputer show a peak performance in excess of 2 TB/s for writing checkpoint data to the GPFS parallel file system. An improvement of 50X is obtained for writing snapshot data in GTC with ADIOS. A novel file format named BP4 for ADIOS is introduced with the objective of reducing metadata overhead at extreme scale. Preliminary results show significant improvement in metadata performance of ADIOS with the new file format. Optimizing I/O on leadership class machines is challenging and further research is required to efficiently utilize the evolving complex storage hierarchy that includes non-volatile memory along with new parallel file systems.

### References

[1] Z. Lin, T. S. Hahm, W. W. Lee, W. M. Tang, and R. B. White, Turbulent Transport Reduction by Zonal Flows: Massively Parallel Simulations, *Science* 281, 1835 (1998).
[2] Z. Lin, W. M. Tang, and W. W. Lee, Neoclassical Transport in Enhanced Confinement Toroidal Plasmas, *Phys. Rev. Lett*. 78, 456 (1997).
[3] Wenlu Zhang, Zhihong Lin, and Liu Chen, Transport of Energetic Particles by Microturbulence in Magnetized Plasmas, *Phys. Rev. Lett*. 101, 095001 (2008).
[4] H. S. Xie, Y. Xiao, and Z. Lin, New Paradigm for Turbulent Transport Across a Steep Gradient in Toroidal Plasmas, *Phys. Rev. Lett*. 118, 095001 (2017).
[5] H. S. Zhang, Z. Lin, and I. Holod, Nonlinear Frequency Oscillation of Alfven Eigenmodes in Fusion Plasmas, *Phys. Rev. Lett*. 109, 025001 (2012).

[6] Z. Lin, T. S. Hahm, S. Ethier, and W. M. Tang, Size Scaling of Turbulent Transport in Magnetically Confined Plasmas, *Phys. Rev. Lett*. 88, 195004 (2002).
[7] Yong Xiao and Zhihong Lin, Turbulent Transport of Trapped Electron Modes in Collisionless Plasmas, *Phys. Rev. Lett*. 103, 085004 (2009).
[8] https://www.scidac.gov/projects/2018/fusion-energy-sciences/isep.html
[9] https://www.olcf.ornl.gov/caar/
[10] Zhixuan Wang, Zhihong Lin, Ihor Holod, W. W. Heidbrink, Benjamin Tobias, Michael Van Zeeland, and M. E. Austin, Radial Localization of Toroidicity-Induced Alfven Eigenmodes, *Phys. Rev. Lett.* 111*,* 145003 (2013).
[11] Wenlu Zhang, Wayne Joubert, Peng Wang, Matthew Niemerg, Bei Wang, William Tang, Sam Taimourzadeh, Lei Shi, Jian Bao, and Zhihong Lin, Heterogeneous Programming and Optimization of Gyrokinetic Toroidal Code Using Directives, *Lecture Notes in Computer Science* 11381, 3–21 (2019). (WACCPD 2018 Workshop, Dallas).
[12] https://www.olcf.ornl.gov/for-users/system-user-guides/summit/summit-user-guide/#system-overview
[13] top500.org
[14] C. S. Chang, S. Klasky, J. Cummings, R. Samtaney, A. Shoshani, L. Sugiyama, D. Keyes, S. Ku, G. Park, S. Parker, N. Podhorszki, H. Strauss, H. Abbasi, M. Adams, R. Barreto, G. Bateman, K. Bennett, Y. Chen, E. D. Azevedo, C. Docan, S. Ethier, E. Feibush, L. Greengard, T. Hahm, F. Hinton, C. Jin, A. Khan, A. Kritz, P. Krsti, T. Lao, W. Lee, Z. Lin, J. Lofstead, P. Mouallem, M. Nagappan, A. Pankin, M. Parashar, M. Pindzola, C. Reinhold, D. Schultz, K. Schwan, D. Silver, A. Sim, D. Stotler, M. Vouk, M. Wolf, H. Weitzner, P. Worley, Y. Xiao, E. Yoon and D. Zorin, Toward a First-Principles Integrated Simulation of Tokamak Edge Plasmas*, Journal of Physics: Conference Series,* Volume 125, Number 1 (2008).
[15] http://www.cgl.ucsf.edu/chimera/about.html
[16] Lipeng Wan, Matthew Wolf, Feiyi Wang, Jong Youl Choi, George Ostrouchov and Scott Klasky, Comprehensive Measurement and Analysis of the User-Perceived I/O Performance in a Production Leadership-Class Storage System, *IEEE 37th International Conference on Distributed Computing Systems* (2017).
[17] Lipeng Wan, Matthew Wolf, Feiyi Wang, Jong Youl Choi, George Ostrouchov and Scott Klasky, Analysis and Modeling of the End-to-End I/O Performance on OLCF's Titan Supercomputer, *IEEE 37th International Conference on Distributed Computing Systems* (2017).
[18] https://www.olcf.ornl.gov/olcf-resources/compute-systems/titan/
[19] https://www.nersc.gov/users/computational-systems/cori/
[20] Lofstead, J. F., Klasky, S., Schwan, K., Podhorszki, N., & Jin, C. (2008, June). Flexible IO and integration for scientific codes through the adaptable IO system (ADIOS). In *Proceedings of the 6th international workshop on Challenges of large applications in distributed environments* (pp. 15-24). ACM.
[21] Fu, H., Liao, J., Yang, J. et al. Sci. China Inf. Sci. (2016) 59: 072001. https://doi.org/10.1007/s11432-016-5588-7
[22] https://computation.llnl.gov/computers/sierra

# Energy response characteristics of a subway station structure based on probability density evolution method

***Zhiqian Liu [1] and † Zhiyi Chen [1,2]**

[1] Department of Geotechnical Engineering, Tongji University, Shanghai 20092, China
[2] Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Shanghai 200092, China

*Presenting author:1510165@tongji.edu.cn
†Corresponding author: zhiyichen@tongji.edu.cn

## Abstract

Energy-based seismic design has been widely used in ground building structures. The vibration characteristics of underground structures are quite different from those of ground structures due to the constraints of surrounding soils. When energy-based seismic design is applied to underground structures, the energy response characteristics of underground structures should be defined first. Based on the probability density evolution method (PDEM), this paper studies energy responses (including plastic energy dissipation, elastic strain energy, and structural kinetic energy) of a multi-story subway station structure from the perspective of stochastic analysis. It is found that the change of kinetic energy, elastic deformation energy, and plastic energy dissipation is consistent with the trend of the intensity change of the earthquake motion. The distribution range of PDF of kinetic energy, elastic deformation energy and plastic energy dissipation becomes wider when the earthquake intensity is larger. Plastic energy does not dissipate all the input energy of the station structure, and the proportion of plastic energy dissipation is less than 1/3.

**Keywords:** Probability density evolution method, Subway station structure, Energy response, Stochastic earthquake

## 1 Introduction

In the 1995 Kobe earthquake, the subway station structure represented by Daikai Subway Station suffered serious damage [1][2], which gives rise to the seismic performance study of subway station structures [3][4].

Since Housner [5] put forward the energy analysis method of structural seismic response, the energy-based seismic design method has gradually been accepted and developed rapidly [6-11]. The response of structures to earthquakes is a process of energy input and dissipation. However, due to the constraints of surrounding soils around, the vibration characteristics of underground structures are not obvious, and the energy input and dissipation characteristics need to be further studied. Considering the randomness of ground motions, the energy input and dissipation characteristics of underground structures can be obtained more comprehensively by using the stochastic analysis method.

Based on the probability density evolution method (PDEM) and the idea of equivalent extreme value events, this paper studies the energy response of a multi-story subway station structure under stochastic earthquake from the perspective of stochastic analysis.

## 2 Numerical model

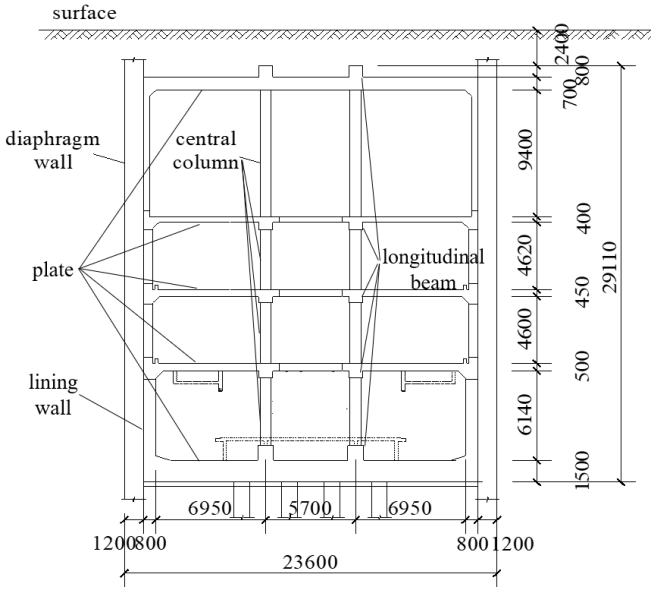



**Figure 1. Standard cross section of the subway station, dimensions in mm**

Figure 1 gives the standard cross section size of the multi-story subway station structure analyzed in this paper, which located in Shanghai, China. The width and height of the standard section of the our-story three-span island station is 23.6m and 29.1m, respectively, with buried depth of the roof of 3.2m. The inner lining wall is connected to the diaphragm wall through the embedded parts, thus both of them bear forces together and form the side wall. The longitudinal distance between columns is 8 m. According to "Chinese Code for Seismic Design of Urban Rail Transit Structures" [12], the shear wave velocity of the soil is 150m/s.

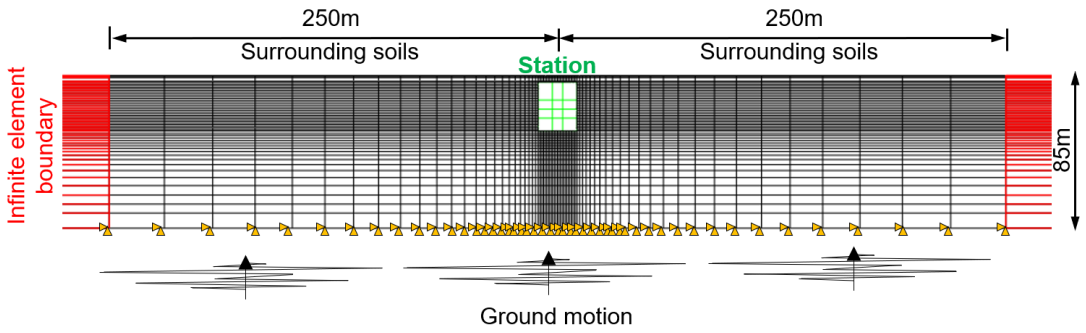



**Figure 2. Numerical model of the underground subway station with surrounding soils**

The plane-strain dynamic analysis model is established in the finite element code ABAQUS [13], which is shown in Fig. 2. Mohr-Coulomb constitutive model with Rayleigh damping is applied to the soil. The Poisson's ratio, internal friction angle, and cohesion is 0.3, 15˚, and 20kPa, respectively. Beam element, B21, is chosen for the structure. Central columns are made of C45 concrete and other parts of the structure are made of C35 concrete [14]. The concrete damaged plasticity model is adopted to better simulate the dynamic response of the structure, and the calculation of damage parameters has been done in previous studies [15][16]. Idealized elastic-plastic model was selected for rebar. Rebar used in central columns

and other parts were HRB400 and HRB335, respectively, whose yield strength is 400MPa and 335MPa, respectively. The soil-structure interaction is defined by the Coulomb friction law, and the coefficient of friction is assumed to be 0.4.

In this paper, the depth of soil is 85m, and the single side width of the soil mass is 250 m, which is more than 3 times the structural width specified in "Chinese Code for Seismic Design of Buildings" [17]. To reduce the boundary effect, infinite element boundary is adopted as the lateral soil boundary [13], as shown in Fig. 2.

## 3 Process of PDEM

*4.1 Numerical analysis process of PDEM*

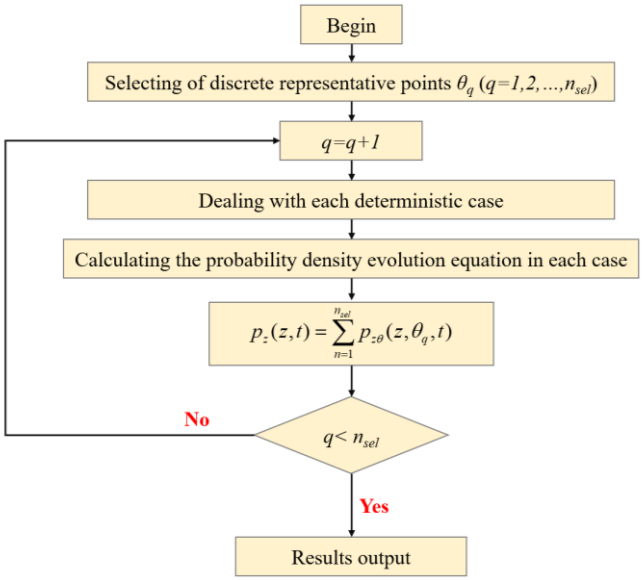



**Figure 3. Basic steps of PDEM**

Based on the principle of probability conservation, the PDEM is proposed by Li and Chen [14], and a relatively complete system of PDEM has been formed After more than ten years of development. The basic steps of stochastic response analysis of structures using PDEM are shown in Fig. 3. For more detailed contents of the method, please refer to the references [18] [19] and previous study [20].

In the numerical solution process of PDEM, the stochastic process of earthquake motion should be discrete in the probability space, then a certain number of stochastic earthquake motion samples are obtained. In this paper, Spectral representation - Random function method, which is proposed by Liu et al. [21], is used to simulate the random process of ground motion. 254 representative points are selected in the probability space, after that 254 earthquake motion samples are generated for the dynamic time history analysis. Relevant parameters are determined according to "Chinese Code for Seismic Design of Buildings" [17] for generating the ground motion samples. Figure 4 gives 3 typical ground motion samples.

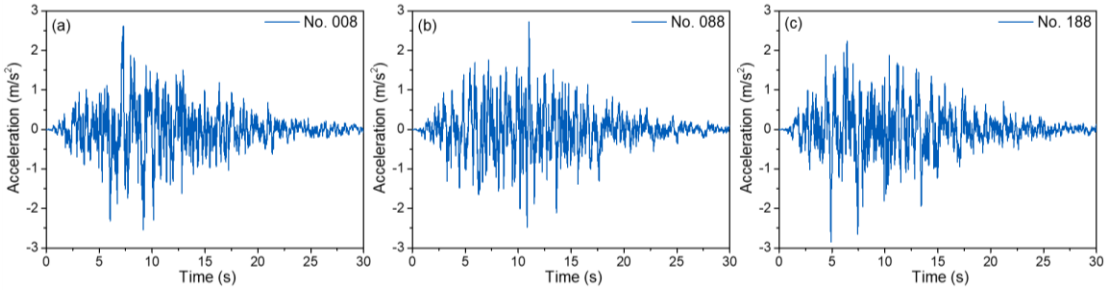



**Figure 4. Typical acceleration time history samples**

## 5 Results and analysis

The total input energy of the structure consists of three parts [6], as shown in the energy balance equation given by Eq. (1):

$$W_e + W_p + W_h = E \tag{1}$$

where, $W_e$ is elastic vibration energy, $W_p$ is cumulative plastic energy dissipation and $W_h$ is damping energy dissipation. Elastic vibration energy can be divided into kinetic energy $W_{ek}$ and elastic deformation energy $W_{es}$, as shown in Eq. (2):

$$W_e = W_{ek} + W_{es} \tag{2}$$

Damping of underground structures during earthquake is not clear, so this paper does not consider the damping of subway station structure at present. The kinetic energy, elastic deformation energy, and cumulative plastic energy of the structure are mainly studied.

### *4.1 Stochastic response analysis of energy response*

Figure 5 (a-c) shows the probability density function (PDF) evolution contours of kinetic energy, elastic deformation energy and plastic energy dissipation of the structure, respectively, during earthquake.

The PDF evolution of kinetic energy with time is shown in Fig. 5 (a). From Fig. 5 (a), it can be seen that the PDF distribution of kinetic energy is highly concentrated near 0 kJ in the initial stage of the stochastic process (i.e. 0 < T < 2 s). This is because the amplitude of ground motion is very small at the beginning, which leads to the weak motion of structure. As time goes on, the amplitude of ground motion increases gradually, and the distribution range of PDF of structural kinetic energy is also increased. The PDF with the widest distribution range of structural kinetic energy is in the period of 7 < T < 17 s, which is within the period of strong amplitude of ground motion (i.e. 3 < T < 17 s). After 17 s, the amplitude of ground motion decreases gradually, and the distribution range of PDF of structural kinetic energy becomes narrower, which indicates that the magnitude of ground motion amplitude is closely related to structural kinetic energy. It is noteworthy that the distribution of PDF is not completely reduced to zero after the earthquake, which indicates that the structure is still in motion.

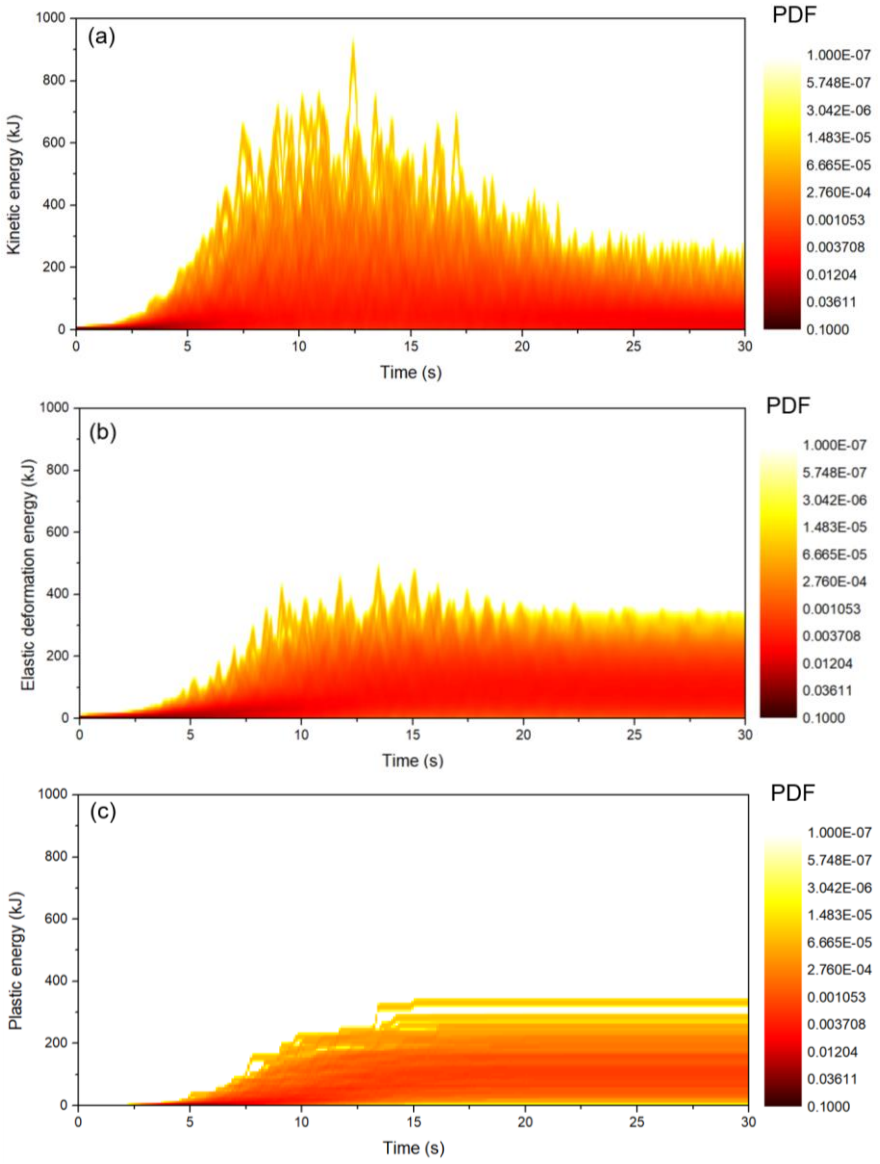



**Figure 5. PDF evolution contours of (a) kinetic energy; (b) elastic deformation energy and (c) plastic energy dissipation of the structure**

The PDF evolution of elastic deformation energy with time is shown in Fig. 5 (b). Similar to PDF of structural kinetic energy, it can be seen from Fig. 5 (b) that PDF of structural elastic deformation energy is highly concentrated in the vicinity of 0 kJ at the initial stage of ground motion, indicating that the elastic deformation of the structure is very small at this time. With the increase of ground motion amplitude, the distribution range of elastic deformation energy PDF becomes wider. In the period of large amplitude of ground motion ($3 < T < 17$ s), the PDF distribution range of elastic deformation energy is obviously narrower than that of kinetic energy at the same time. In the later period of earthquake ($T > 17$ s), the PDF distribution range of elastic deformation energy is narrower, but it basically changes little. This is because the structural deformation is controlled by the deformation of surrounding soil. The surrounding soil appears obvious deformation in the earthquake, which makes the structure store higher elastic deformation energy.

The PDF evolution of plastic energy dissipation with time is shown in Fig. 5 (c). The plastic energy dissipation increases monotonously with time. From Fig. 5 (c), it can be seen that the

time period when the PDF distribution range of plastic energy dissipation of the structure changes is mainly in the range of 3 < T < 17 s, which is consistent with the time period when the earthquake amplitude is large. When T < 3 s and T > 17 s, the PDF distribution of plastic energy dissipation hardly changed. This indicates that plastic energy dissipation mainly occurs in the period of large earthquake amplitude.

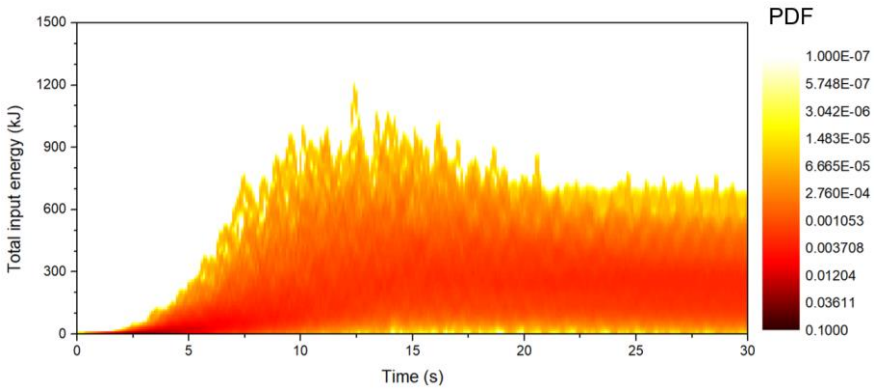



**Figure 6. PDF evolution contours of total input energy of the structure**

Figure 6 shows the PDF evolution contours of total input energy of the structure during earthquake. The total input energy is the sum of kinetic energy, elastic strain energy and plastic energy dissipation at the moment. From Fig. 6, it can be seen that the PDF distribution range of total input energy experienced a process of broadening (T < 15 s) and narrowing (T > 15 s). This indicates that plastic energy dissipation does not dissipate all the input energy of the station structure, and the rest of the energy probably dissipates into the soil through soil-structure interaction.

*4.2 Extreme value analysis of energy response*

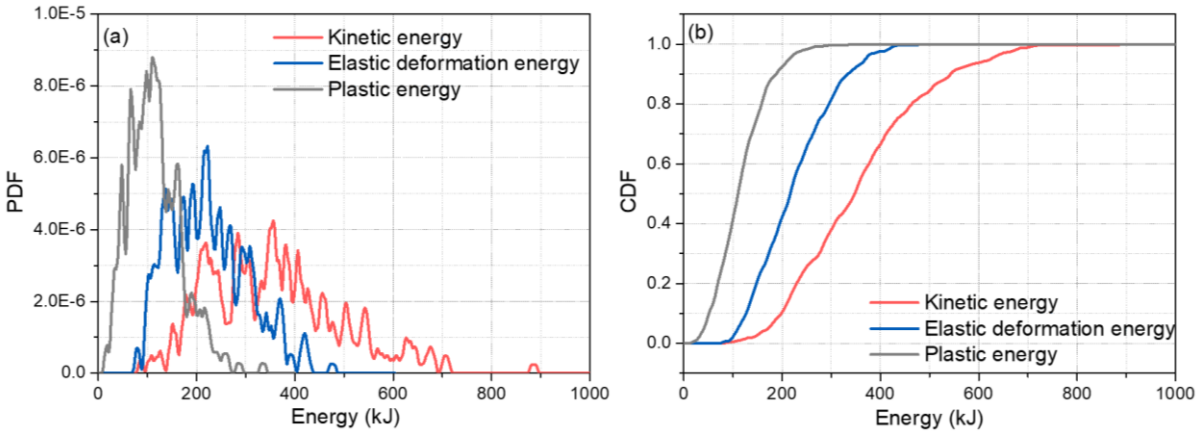



**Figure 7. PDF and CDF curves of maximum energy during earthquake**

Based on PDEM and the equivalent extreme event thought, the corresponding equivalent extremum events are constructed to calculate the distribution of extreme energy value of the structure under stochastic earthquake motion. Figure 10 gives the probability density function (PDF) and the cumulative probability function (CDF) of the maximum energy during earthquake. The extreme energy value corresponding to 50% CDF in Fig. 10 (b) is taken as

representative value, and the corresponding representative extreme values of kinetic energy, elastic deformation energy and plastic energy dissipation are 348 kJ, 216 kJ, 110 kJ, respectively. It is indicated that plastic energy dissipation does not dissipate all the input energy of the station structure, and the proportion of plastic energy dissipation is less than 1/3. The rest of the energy probably dissipates into the soil through soil-structure interaction.

In terms of energy balance, two aspects can be considered to give advices to the seismic design of subway station structure. On the one hand, the total input energy of the structure should be controlled by improve surrounding soils. On the other hand, the proportion of plastic energy dissipation of structural components should be reduced by means of increasing the energy dissipated to the soil and reasonably arranging the energy dissipation components such as shear plate dampers (SPDs).

## 6 Conclusions

Based on PDEM and finite element elastoplastic dynamic time history analysis, the energy response characteristics of a multi-story subway station structure under stochastic earthquake motion are studied in the presented paper. The main conclusions are as follows:
(1) In a seismic process, the change of kinetic energy, elastic deformation energy, and plastic energy dissipation of subway station structure is closely related to earthquake intensity. The distribution range of PDF of kinetic energy, elastic deformation energy and plastic energy dissipation becomes wider when the earthquake intensity is larger.
(2) Plastic energy dissipation does not dissipate all the input energy of the station structure, and the proportion of plastic energy dissipation is less than 1/3. The rest of the energy probably dissipates into the soil through soil-structure interaction.

## Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 41472246, 51778464), Key laboratory of Transportation Tunnel Engineering (TTE2014-01), and “Shuguang Program” supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission. All supports are gratefully acknowledged.

### References

[1] Nakamura, S., Yoshida, N., and Iwatate, T. (1996) Damage to Daikai subway station during the 1995 Hyogoken-Nambu earthquake and its investigation, *Japan Society of Civil Engineers*, Committee of Earthquake Engineering, 287-295.
[2] Yamato, T., Umehara, T., Aoki, H., Nakamura, S., Ezaki, J., and Suetomi, I. (1996) Damage to Daikai subway station, Kobe rapid transit system and estimation of its reason, *Proceedings, technical conference on the great Hanshin-Awaji Earthquake*, JSCE, Tokyo, 247-254. (in Japanese)
[3] Chen, G., Wang, Z., Zuo, X., Du, X., and Gao, H. (2013) Shaking table test on the seismic failure characteristics of a subway station structure on liquefiable ground, *Earthquake Engineering & Structural Dynamics* **42**(10), 1489-1507.
[4] Khani, S., and Homami, P. (2014) Seismic performance of underground shallow subway stations in soft soil, *Journal of Engineering Geology* **8**(1), 1983-2002.
[5] Housner, G. W. (1959) Behavior of structures during earthquakes. *Journal of the Engineering Mechanics Division* **85**(4), 109-130.
[6] Akiyama, H. (1985) Earthquake-resistant limit-state design for buildings. *Univ of Tokyo Pr*.
[7] Fajfar, P., Vidic, T., and Fischinger, M. (1989) Seismic demand in medium- and long-period structures. *Earthquake Engineering & Structural Dynamics* **18**(8), 1133-1144.
[8] Fajfar, P. (1992) Equivalent ductility factors, taking into account low-cycle fatigue. *Earthquake Engineering & Structural Dynamics* **21**(10), 837-848.

[9] Fajfar, P., and Gašperšič, P. (1996) The N2 method for the seismic damage analysis of RC buildings. *Earthquake Engineering & Structural Dynamics* **25**(1), 31-46.
[10] Shen, J., and Akbaş, B. (1999) Seismic energy demand in steel moment frames. *Journal of Earthquake Engineering* **3**(04), 519-559.
[11] Chou, C. C., and Uang, C. M. (2000) Establishing absorbed energy spectra—an attenuation approach. *Earthquake engineering & structural dynamics* **29**(10), 1441-1455.
[12] GB50909 (2014) *Code for Seismic Design of Urban Rail Transit Structures*, China Planning Press, Beijing, China.
[13] ABAQUS, Users Manual V. 6.14. (2015) Dassault Systemes Simulia Corp., Providence, RI.
[14] GB50010 (2010) *Code for Design of Concrete Structure*, China Architecture & Building Press, Beijing, China.
[15] Chen, Z. Y., Chen, W. and Zhang, W. (2014) Seismic performance evaluation of multi-story subway structure based on pushover analysis, *Advances in Soil Dynamics and Foundation Engineering* 444-454. ASCE.
[16] Chen, Z. Y., Chen, W., Zhang, W. and Lou, M. L. (2016) Effects of Axial Compression Ratio of Central Columns on Seismic Performance of a Multi-Story Underground Structure, *International Journal of Computational Methods* **13**(04), 1641014.
[17] GB50011 (2010) *Code for Seismic Design of Buildings*, China Architecture & Building Press, Beijing, China.
[18] Li, J. and Chen, J. (2008) The principle of preservation of probability and the generalized density evolution equation, *Structural Safety* **30**(1), 65-77.
[19] Li, J. and Chen, J. (2009) *Stochastic dynamics of structures*, John Wiley & Sons.
[20] Liu Z. Q., and Chen Z. Y. (2018) Failure modes analysis of a multi-story subway station under stochastic earthquake based on probability density evolution method. *Proceedings of the International Conference on Computational Methods*, Italy, **5**, 629-637.
[21] Liu Z.J., Zeng B. and Wu L.Q. (2015) Simulation of non-stationary ground motion by spectral representation and random functions, *Journal of Vibration Engineering* **28**(3), 411-417. (in Chinese)

# Isolation effect analysis on friction pendulum bearings in underground station structures

***Peng Jia [1] and †Zhiyi Chen [1,2]**

[1]Department of Geotechnical Engineering, Tongji University, Shanghai 200092, China

[2]Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Shanghai 200092, China

*Presenting author: 919526724@qq.com

†Corresponding author: zhiyichen@tongji.edu.cn

## Abstract

The previous earthquake surveys show that the central columns are the weak parts in subway stations during the earthquakes. Reducing the seismic responses of the central columns can ensure the safety of the stations. Setting friction pendulum bearings at the top of the central columns may be a good strategy. In this paper, based on a subway station in Shanghai, the friction pendulum bearings is simulated detailedly and the isolation effect of the friction pendulum bearings is studied with the three-dimensional dynamic time history analysis method. The results show that the friction pendulum bearings can effectively reduce the shearing force of the central columns and slightly reduce the bending moment during the earthquake. At the same time, they does not cause a significant increase in deformation of the side walls. It is stated that the friction pendulum bearings are also effective in the underground station structures.

**Keywords:** Underground subway station, Friction pendulum bearings, Seismic Response

## 1. Introduction

In 1987, Zayas et al. developed a friction pendulum bearing at the University of California, Berkeley [1]. At present, more than 10 kinds of friction pendulum bearings have been developed. The most common friction pendulum bearing is the single curved surface friction pendulum bearing, which is mainly composed of a slider, a lower base and an upper base.

Friction pendulum bearings have been widely used in above-ground structures, but rarely in underground structures [2][3]. Since the Kobe earthquake in Japan caused severe damage to the Dakai Station in 1995, the seismic capacity of underground structures has gotten more and more attention. Previous studies have shown that the damage of the underground station structure is mainly caused by the damage of the central columns [4]. Strengthening the seismic performance of the central columns can effectively improve the seismic capacity of the underground station. Setting friction pendulum bearings at the top of the central columns may be a good strategy.

## 2. Fine simulation of FPB

At present, the most common friction pendulum bearing on the ground is the single curved friction pendulum bearing. It is mainly composed of three parts, which are an upper base, a lower base and a slider. During the earthquake, the slider is mainly subjected to the vertical pressure, horizontal force transmitted by the structure, support force and friction on the concave surface. According to the force's equilibrium condition, Eq. (1) can be get that:

$$F = \frac{WD}{R\cos\theta} + \frac{f}{\cos\theta} \tag{1}$$

When the angle of rotation is relatively small, it can be approximately simplified as Eq. (2).

$$F = \frac{WD}{R} + f = \frac{WD}{R} + \mu W \operatorname{sgn}(\dot{\theta}) \tag{2}$$

where, F is horizontal force transmitted by the structure, W is vertical pressure, R is the radius of the concave surface, D is the relative displacement of the slider, and μ is the friction coefficient of the sliding surface [5].

The ABAQUS finite element software is used to carry out the refined simulation [6]. The finite element model of the friction pendulum bearing is shown in Figure 1. The concave radius is 2m and the friction coefficient of the sliding surface is 0.1.

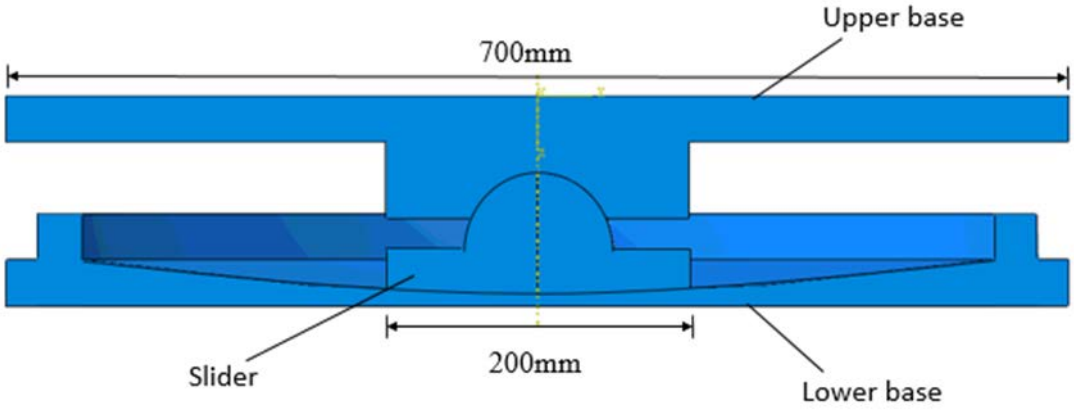



**Figure 1. Fine simulation of FPB**

Applying a vertical pressure of 2000kN to the friction pendulum bearing. The force-displacement curve of the finite element model is compared with the result calculated by the Eq. (2). As shown in Figure 2, the two curves are roughly coincident, indicating that the refined finite element model can simulate the mechanical characteristics of the friction pendulum bearing well.

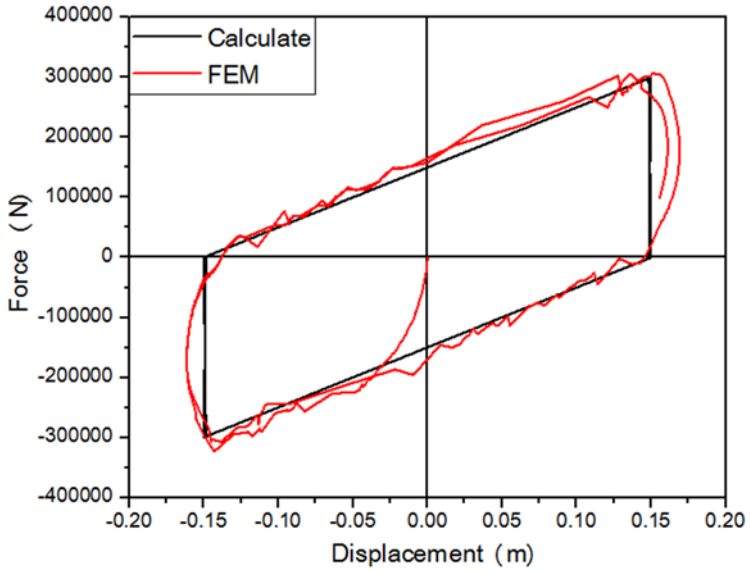



**Figure 2. The force-displacement curve of FPB**

## 3. Numerical Simulation

*3.1 Finite element model*

A cross section of the station is shown in Figure 3. The station is 21.84m wide and 13.1m high. The cross section dimensions of the upper and lower central columns are both 0.7m×0.7m, and the interval between the central columns is 5.8m.

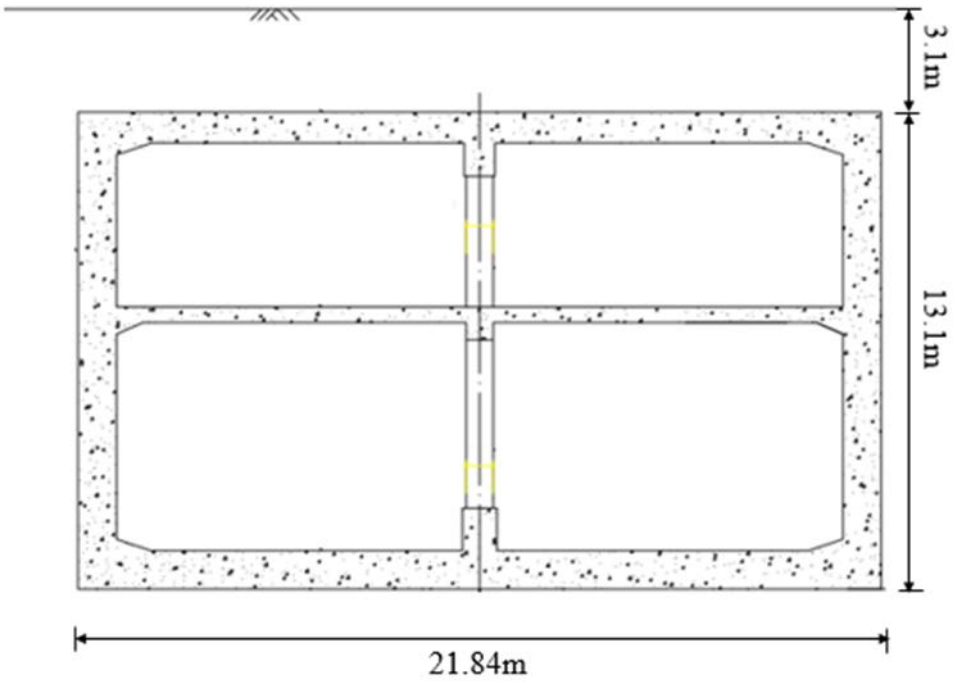



**Figure 3. The cross section of the station**

The ABAQUS finite element software was used to establish a three-dimensional finite element model of the station and soil layers. The dimensions of the station model are established according to the cross section shown in Figure 3. According to the detailed geotechnical survey report of the project, the soil can be divided into ten layers. The thickness

and basic physical properties of each layer are shown in Table 1. The overall size of the model is 80m long, 60m high and 5.8m wide, since the interval between the central columns in the station is 5.8m, as shown in Figure 4.

**Table 1. The basic physical properties of each soil layer**

| Soil number | Soil texture | Thickness（m） | Density（g/m3） | Internal friction angle(°) | Cohesion (kPa) |
|---|---|---|---|---|---|
| 1 | Artificial fill | 1.3 | 1900 | 15 | 20 |
| 2-1 | Yellowish dark brown silty clay | 1.1 | 1920 | 31.3 | 9.5 |
| 2-2 | Yellow gray silty clay | 1.1 | 1800 | 33.8 | 15.1 |
| 3 | Gray mucky silty clay | 3.4 | 1740 | 28.3 | 5.3 |
| 4 | Gray mucky clay | 8.2 | 1670 | 24.9 | 7.2 |
| 5-1 | Gray clay | 1.6 | 1740 | 29.7 | 10 |
| 6 | Dark green silty clay | 4.4 | 1950 | 29.1 | 31.3 |
| 7 | Olive drag-gray sandy silt | 6.9 | 1820 | 31.1 | 2 |
| 8-1 | Gray clay | 15 | 1770 | 32.5 | 8.1 |
| 8-2 | Gray silty clay and silty sand interbedded soil | 17 | 1840 | 28.1 | 8 |

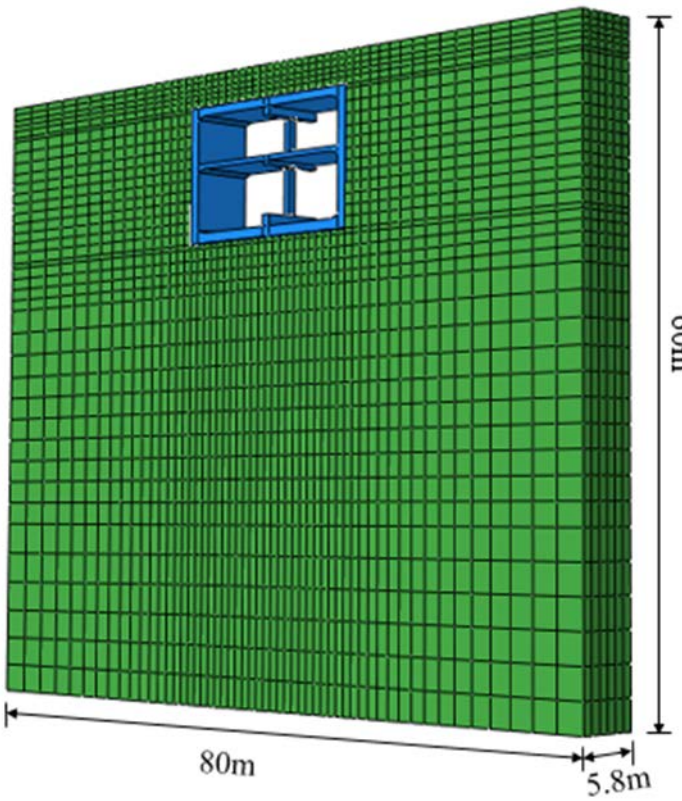



**Figure 4. The finite element model of the station and soil**

The boundary conditions of the ground are fixed at the bottom, and an equal displacement boundary is adopted to the four side boundaries, so that each layer elements have the same horizontal displacement. The equal displacement constraint can avoid the reflection of seismic

waves at the boundary [7][8].

*3.2 Constitutive models*

The station is a reinforced concrete structure. To simplify the analysis, the material is equivalent to a homogeneous material, and the material parameters are approximated to the concrete's. The density of the material is 2500kg/m3, the Poisson's ratio is 0.2, and the modulus of elasticity is 31.5Gpa.

In order to simulate the dynamic properties of the soil, an equivalent linear viscoelastic model with damping is used. The dynamic elastic modulus of the soil layers can be calculated according to the Eq. (3)

$$G = \rho V_s^2$$

$$E = 2(1+\sigma)G \tag{3}$$

where, $V_s$ is the shear wave velocity, ρ is the soil density, σ is the Poisson's ratio, and G is the equivalent dynamic shear modulus of the soil. The variation of the shear wave velocity along the depth is in accordance with the law of exponential or logarithm. For the Shanghai region, the exponential form can be selected, as shown in Eq.(4).

$$V_s = aH^b \tag{4}$$

where, H is the depth of soil layers, a and b are the calculation parameters. For the area where the station is located, a and b are 79.03 and 0.3437 [9]. The soil layer parameters obtained by calculation are shown in Table 2.

**Table 2. The parameters of each soil layer**

| Soil number | Depth (m) | Shear wave velocity (m/s) | Shear modulus (Gpa) | Poisson's ratio | Elastic modulus (Gpa) |
|---|---|---|---|---|---|
| 1 | 1.3 | 86 | 14.2 | 0.32 | 37.5 |
| 2-1 | 2.4 | 107 | 21.9 | 0.32 | 57.8 |
| 2-2 | 3.5 | 122 | 26.6 | 0.34 | 71.3 |
| 3 | 6.9 | 153 | 41 | 0.38 | 113.2 |
| 4 | 15.1 | 201 | 67.4 | 0.4 | 188.8 |
| 5-1 | 16.7 | 208 | 75.2 | 0.35 | 203.2 |
| 6 | 21.1 | 225 | 99.1 | 0.29 | 255.6 |
| 7 | 28 | 248 | 112.3 | 0.29 | 289.8 |
| 8-1 | 43 | 288 | 146.7 | 0.33 | 390.2 |
| 8-2 | 60 | 323 | 191.7 | 0.32 | 506.2 |

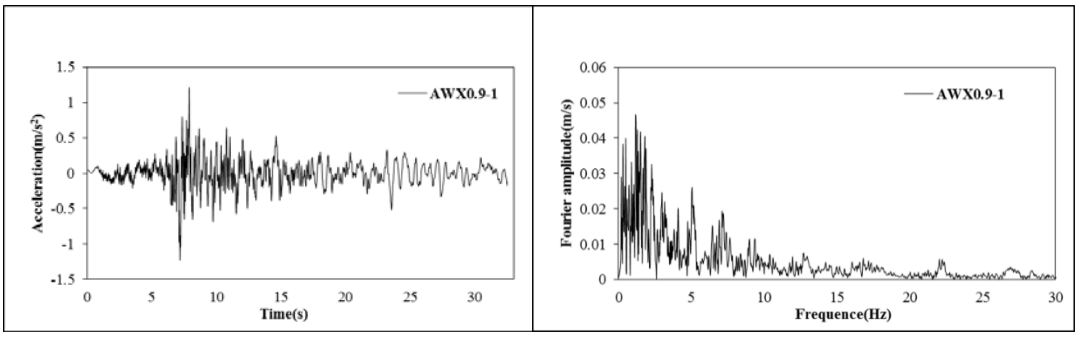
The damping of the soil is Rayleigh damping. Using the modal analysis function of ABAQUS, the first-order mode frequency of the site is 0.36Hz, and the second-order mode frequency is 1.08Hz. Rayleigh damping coefficient can be calculated according to the Eq. (5).

$$\xi = \frac{\alpha}{2} \times \frac{1}{\omega_1} + \frac{\beta}{2} \times \omega_2 \tag{5}$$

where, $\omega_1$ and $\omega_2$ are the main mode frequencies of the soil, and the damping ratio $\xi$ is 0.05. It is calculated that the Rayleigh damping coefficient $\alpha = 0.1698$, $\beta = 0.0129$.

*3.3 Ground motion characteristics*

In order to study the seismic response of the station, the seismic waves are input in the Y direction at the bottom of the model. The Shanghai artificial wave is selected as the input ground motion. The acceleration time history curve and the Fourier spectrum of the seismic wave are shown in the Figure4.

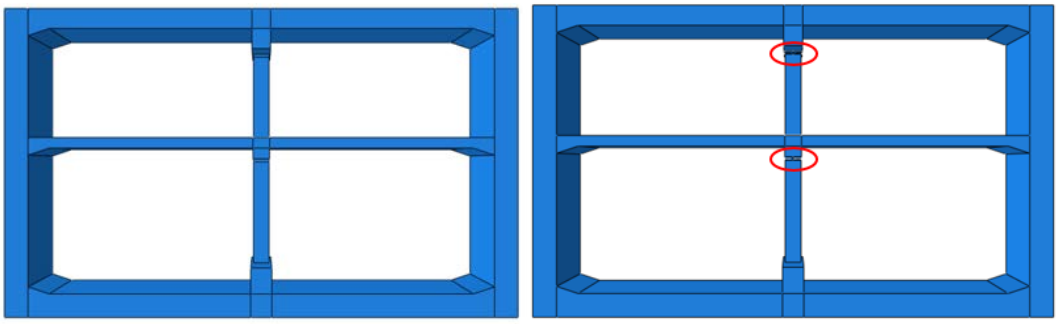



**(a)** **(b)**

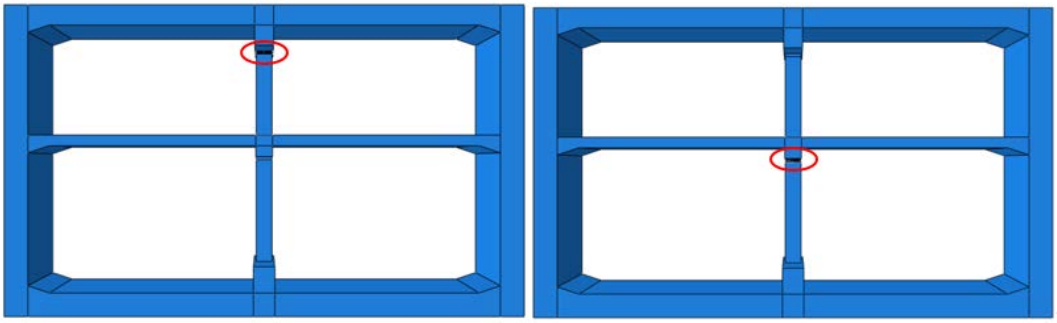
**Figure 4. Shanghai artificial wave characteristics: (a) Acceleration time history curve and (b) Fourier spectrum**

*3.4 Cases introduction*

As the station is a frame structure with two floors and two spans, the friction pendulum bearings are placed on the top of the central columns, so there are four arrangements, as shown in figure5. That is, the friction pendulum bearings are not arranged, the friction pendulum bearings are arranged in both floors, the friction pendulum bearings are arranged only on the upper floor and the friction pendulum bearings are arranged only on the lower floor. And the amplitude of seismic wave acceleration is 0.1g and 0.5g.



**(a) None** **(b) Both**



**(c) Up** **(d) Down**

**Figure 5. Four different arrangements: (a) FPB are not arranged, (b) FPB are arranged in both floors, (c) FPB are arranged in upper floor, (d) FPB are arranged in lower floor**

## 4. Isolation effect of friction pendulum bearings

*4.1 Seismic response of the central columns*

The destruction of the underground stations is mainly caused by the destruction of the central columns. The bending and shearing is the major failure form of the central columns. So the maximum bending moment and shearing force at the bottom of central columns are extracted as the indexes of dynamic response during the earthquake. In different cases, the shearing force at the bottom of the central columns in two floors is shown in Table 3.

**Table 3. The shearing force in central columns in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor (kN) | Upper floor (kN) | Lower floor (kN) | Upper floor (kN) |
| none | 178 | 213 | 854 | 1105 |
| both | 87 | 110 | 427 | 390 |
| up | 187 | 81 | 911 | 366 |
| down | 88 | 257 | 429 | 1307 |

The isolation effect of friction pendulum bearings is defined as Eq. (6)

$$\gamma = (R_0 - R_f) / R_0 \tag{6}$$

where, $R_f$ is dynamic response of central columns with friction pendulum bearings and $R_0$ is the response without friction pendulum bearings. In each case, the isolation effect on the shearing force of the central columns is shown in the Table.4 and Figure 6.

**Table 4. The isolation effect on shearing force in central columns in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor | Upper floor | Lower floor | Upper floor |
| none | - | - | - | - |
| both | 51.1% | 48.4% | 50.0% | 64.7% |
| up | - | 62.2% | - | 66.9% |
| down | 50.4% | - | 49.8% | - |

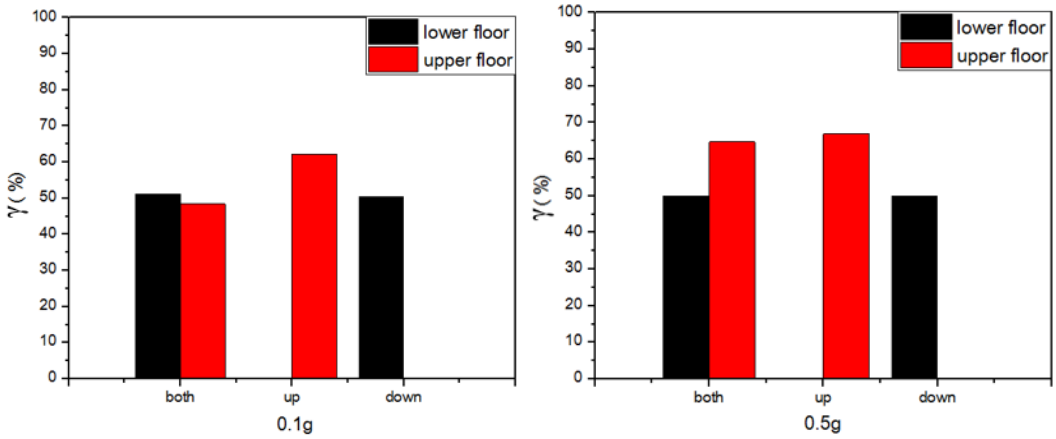



**Figure 6. The isolation effect on shearing force in central columns in different cases**

It can be found from the results that the friction pendulum bearings can effectively reduce the maximum shearing force on the central columns during the earthquake. When the friction pendulum bearings are placed on one of the two floors, the maximum shearing force during the earthquake in the central columns of this floor is significantly reduced. The maximum shearing force in the other floor's central columns is slightly increased. And in different cases, the bending moment at the bottom of the central columns in two floors is shown in Table 5.

**Table 5. The bending moment in central columns in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor (kN*m) | Upper floor (kN*m) | Lower floor (kN*m) | Upper floor (kN*m) |
| none | 360 | 275 | 1780 | 1322 |
| both | 295 | 260 | 1427 | 1046 |
| up | 369 | 220 | 1840 | 990 |
| down | 299 | 374 | 1450 | 1787 |

In each case, the isolation effect of the isolation bearing on the bending moment of the central columns is shown in the Table 6 and Figure7.

**Table 6. The isolation effect on bending moment in central columns in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor | Upper floor | Lower floor | Upper floor |
| none | - | - | - | - |
| both | 18.1% | 5.5% | 19.8% | 20.9% |
| up | - | 20.0% | - | 25.2% |
| down | 17.0% | - | 18.6% | - |

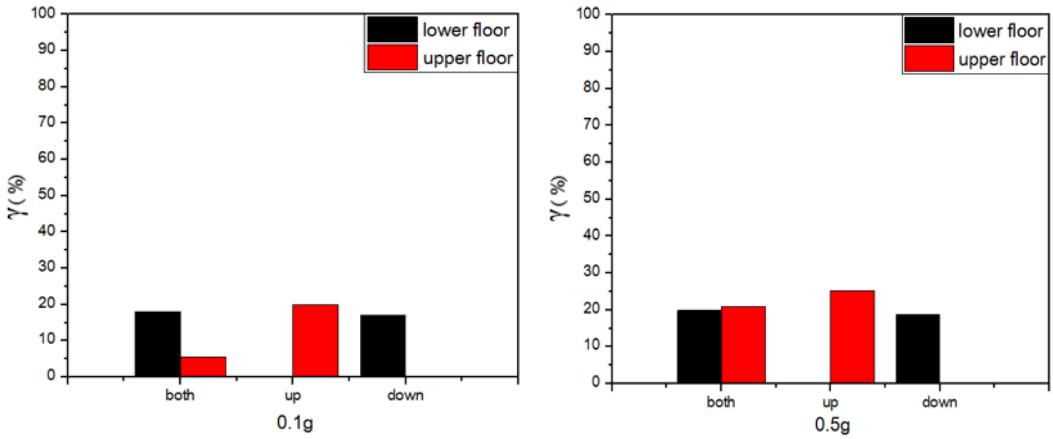



**Figure 7. The isolation effect on bending moment in central columns in different cases**

It can be found from the results that the friction pendulum bearings can reduce the maximum bending moment on the central columns during the earthquake. The isolation effect on the bending moment is not obvious compared with the shearing force. When the friction pendulum bearings are placed on one of the two floors, the maximum bending moment in the other floor's central columns is slightly increased.

*4.2 Seismic response of the side walls*

The station studied in this paper is a two-story and two-span frame structure. The deformation of the side walls during the earthquake can be judged by inter-layer displacement and inter-layer displacement angle. In different cases, the inter-layer displacement and inter-layer displacement angle are shown in Table 7 and Table 8.

**Table 7. The inter-layer displacement in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor (mm) | Upper floor (mm) | Lower floor (mm) | Upper floor (mm) |
| none | 5.16 | 3.41 | 25.38 | 17.56 |
| both | 5.26 | 3.49 | 26.11 | 17.72 |
| up | 5.17 | 3.53 | 25.43 | 18.12 |
| down | 5.22 | 3.37 | 25.81 | 17.23 |

**Table 8. The inter-layer displacement angle in different cases**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor | Upper floor | Lower floor | Upper floor |
| none | 1/1198 | 1/1488 | 1/243 | 1/289 |
| both | 1/1176 | 1/1452 | 1/237 | 1/286 |
| up | 1/1194 | 1/1436 | 1/243 | 1/280 |
| down | 1/1183 | 1/1505 | 1/239 | 1/294 |

In each case, compared with the station without friction pendulum bearings, the percentage increment of the inter-layer displacement and the inter-layer displacement angle after the friction pendulum bearings are set is shown in the Table 9.

**Table 9. The percentage increment of the inter-layer displacement**

| Case | 0.1g | | 0.5g | |
|---|---|---|---|---|
| | Lower floor | Upper floor | Lower floor | Upper floor |
| none | - | - | - | - |
| both | 1.9% | 2.5% | 2.8% | 0.9% |
| up | - | 3.6% | - | 3.2% |
| down | 1.2% | - | 1.7% | - |

It can be found from Table 9 that the friction pendulum bearings can effectively reduce the shearing force and slightly reduce the bending moment of the central columns. And they will only increase the interlayer displacement and the interlayer displacement angle by 1-4%. That is, they will not cause a significant increase in the internal force and deformation of the side wall.

## 5. Conclusion

In this paper, through the refined finite element model, the isolation effect of the friction pendulum bearings in the underground structure is studied, and the following conclusions are obtained.

(1) The friction pendulum bearings can effectively reduce the shearing force received by the central columns in the station, and the reduction can exceed 60%.

(2) The friction pendulum bearings effectively reduce the shearing force of the central columns in the station, but the isolation effect on the bending moment is not obvious.

(3) The reduction of the internal force of the central columns on which the friction pendulum bearings are arranged does not cause a significant increase in the deformation of the side walls.

This paper proves that friction pendulum bearing is also effective in the underground station. And they can significantly improve the stations' earthquake resistance

## References

[1] Zayas V, Low S and Mahin S. The FPS earthquake resisting system. (1987) Technical Report UCB/EERC-87 /01, University of California at Berkeley.

[2] Paolo M. Calvi, Gian Michele Calvi. (2018) Historical development of friction-based seismic isolation systems. *Soil Dynamics and Earthquake Engineering* **106**, 14-30.

[3] P. Scott Harvey Jr., Karah C. Kelly. (2016) A review of rolling-type seismic isolation: Historical development and future directions. *Engineering Structures* **125**, 521-531.

[4] Chao Ma, De-Chun Lu, Xiu-Li Du, Cheng-Zhi Qi, Xiang-Yang Zhang. (2019) Structural components functionalities and failure mechanism of rectangular underground structures during earthquakes. *Soil Dynamics and Earthquake Engineering* **119**, 265-280.

[5] P. Castaldo, B. Palazzo, P. Della Vecchia. (2015) Seismic reliability of base-isolated structures with friction pendulum bearings. *Engineering Structures* **95**, 80-93.

[6] ABAQUS. Standard User's Manual, vols. Ⅰ&Ⅱ. (2009) Hibbit, Karlsson & Sorensen Inc. Pawtucket, Rhode Island.

[7] A. E. Kampitsis, E. J. Sapountzakis, S. K. Giannakos, N. A. Gerolymos. (2013) Seismic soil–pile–structure kinematic and inertial interaction—A new beam approach. *Soil Dynamics and Earthquake Engineering* **55**, 211-224.

[8] A. M. MARSHALL, R. FARRELL, A. KLAR, R. MAIR. (2012) Tunnels in sands: the effect of size, depth and volume loss on greenfield displacements. *Geotechnique* **62(5)**, 385-399.

[9] Fei Gao, Xiaogang Sun. (2005) Characteristics analysis of shear wave velocity of foundation ground in Shanghai region. *Shanghai Geology* **2**, 27-29.

# An improved reflectance prediction model for halftone printing dot based on Monte Carlo method

**Honghao Liu[1], *Kaixing Zhang[1], Xianxi, Liu[1], Jinxing Wang[1], Heow Pueh Lee [2]**

1 College of Mechanical and Electronic Engineering，ShanDong Agricultural University
Taian 271018, China;
2 Department of Mechanical Engineering, National University of Singapore, Singapore 117575, Singapore
* Corresponding author. E-mail address: kaixingzhang@sdau.edu.cn

## Abstract

The improved reflectance prediction model for printing dot was established with Monte Carlo method. Reflectance model is a useful approach to predict and control printing quality, which was widely used in color duplication field. The Monte Carlo prediction model for printing dot is a simulation model, satisfying industrial virtual reality needs, which could simulates reflectance data as well as dot shape image. This paper mainly focuses on improving the existing reflectance model for printing dot with simulation program, expanding dot simulation area from one single dot to nine dots. The results showed that the reflectance value closely approached to the measured value with high dot area in particular and the correlation coefficient improved to 0.924. One hundred thousand times was shown to be the optimal number of simulated photon under regular printing condition. The expanding simulation area could significantly eliminate interactive influence among the dots. The photon distribution and dot shape could be sufficiently simulated through this model, providing a reference for quality control. The results indicate that the improved reflectance model increased reflectance accuracy and could further promote optical study about color duplication.
**Key words**: Tracking photon path, Monte-Carlo method, Reflectance model

## 1. Introduction

Printing works consisting of paper and colorful ink, almost existing widely, is a necessary media for distribution of text and image. Although the internet has become a powerful media gradually, prints still accounts for 60 percent contribution on information dissemination. Whether information can be exactly described and disseminated, depends on the quality of printing works with art picture in particular. Before actual print with machine, several procedures must be made including color separation, image screening and dot selection. In a typical printing process, printing material, printing speed and machine pressure could raise the difference. High quality prints must be the optimal combination of all influence factors. Because accomplishing one printing job is a large project, which requires huge energy and material. Hence it is not effective to judge the final quality with real printed work. For several decades, relevant researchers have been working on an approach to predict, judge and adjust printing quality before actual print without consuming large energy, and therefore promoting energy conservation. The printing model is a good choice. The printing model that can predict the printing reflectance before actual print is widely used in printing domain.

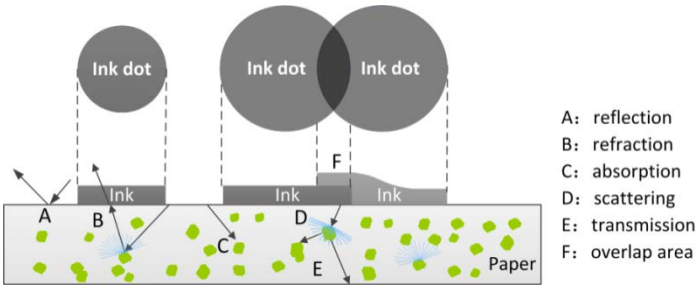



**Fig 1 The distribution and interactive results of different dots.**

The first printing model that can calculate total the reflectance is the Murray-Davies model [1], which only paid attention to the direct reflection of paper and ink surface. The printing woks have various layers with different reflective characters and multiple reflections between different layers cannot be ignored [2]. Although the Clapper-Yule model [3] that was improved from the Murray-Davies model taken multiple reflection into consideration by introducing correction parameters *N*, it did not perform better and had low stability in that the correction parameters *N* was obtained by experience which cannot be quantified objectively [4]. The point spread function model that used the optical radiation theory to calculate reflectance could achieve better performance, but it would need complicated calculation [5]. Relevant studies such as DT2002 and Kubelka-Munk model all improved printing model accuracy to some extant [6-9]. Not until the Monte-Carlo model was introduced into printing field, printing prediction model only output single data. Apart from reflectance prediction, the Monte-Carlo model [10-12] that was widely used in finance, mathematic and medical fields can simulate printing dot shape, and providing a new approach to adjust printing quality. The first Monte-Carlo model used in printing field was proposed by Robert Beuc, which successfully simulated paper surface in 2009 [13]. Based on that, Damir Modric accomplished single printed line simulation in 2012. Besides, relevant studies [14] with Monte-Carlo method achieved different dot shape and photon distribution simulation, which extended the printing model researches into micro field.

The aim of this paper is to propose an improved printing reflectance model with Monte-Carlo method by expanding the simulation area, further promoting printing quality control and optical study.

## 2 Improving reflectance model

The Monte-Carlo method proposed by Metropolis is essentially a probabilistic method [15]. It assumes that a sample can reflect population attributes as long as sample number is sufficient. Monte-Carlo method has been applied in various simulation fields that has random properties. The comprehensive phenomenon of random problem is affected by various factors. It is quite difficult to study the affluence of each factor through doing real experience. Using computer simulation technology with Monte-Carlo method could solve random problem effectively. As the requirement of environment protection increases, prediction model of Monte-Carlo method attracts increasing attentions of relevant researchers all over the world [16].

Compared with other printing reflectance model, the reasons why Monte–Carlo model perform better rely on that it take different factors into consideration such as, paper property, ink attribute and various light phenomenon. The color sense that finally formed in brain is the comprehensive results between light of prints surface and visual nerve. There are almost no differences of visual nerve for healthy people, therefore the attributes of light coming from prints surface becomes the main symbol to predict prints quality. The light that finally coming

into eye from prints surface contain different parts including direct reflectance, scattering parts and refraction parts showed by Fig 1, all of which can be simulated with Monte-Carlo reflectance model. It has been proven that light is composed of photon. The basic principle of Monte-Carlo used in this study is to simulate light propagation trajectory in prints. When the number of photon trajectory satisfy statistical requirement, reflectance information of prints can be obtained through statistic calculation.

The first simulation procedure proposed by Modric [13] within the Monte-Carlo method accomplished modeling of light scattering in paper. That work built the foundation for the following study. The simulation procedure mainly contains four steps, initialization, direction change, energy record and shape design. In the initialization step, light location and moving distance of photon was initialized. The light that reflected from prints surface is assumed to be parallel light with the location *(x, y, z)*. The moving distance also called step-size is a significant parameter deciding the accuracy of reflectance model. The simulated photon cannot be reflected by different layers in paper before leaving prints if step-size is too long. Photon moving distance in prints is a random variable. It follows Beer law and decided by scattering coefficient $\sigma_s$ and absorption coefficient $\sigma_\alpha$ shown by formula 1. Parameter $\varepsilon$ is random value ranging from *0* to *1*.

$$s = \frac{-\ln \varepsilon}{\sigma_a + \sigma_s} \tag{1}$$

For printing reflectance model, scattering phenomenon play an important role in calculating optical dot gain. In the simulation process, the scattering phenomenon in prints involves photon direction change caused by the interaction between paper coating and photon. Three phase functions can be used to describe photon direction change including, Henyey-Greenstein function， Mie function and exponential cosine function [17,18]. Henyey-Greenstein function is short to H-G function. It is widely used in simulation domain for its calculation superiority presented by formula 2. The parameter *β, g* are deflection angle and anisotropic coefficient separately.

$$\cos\beta = \begin{cases} \dfrac{1}{2g}\left\{1+g^2-\left[\dfrac{1-g^2}{1-g+2g\varepsilon}\right]\right\} & g \neq 0 \\ 2\varepsilon - 1 & g = 0 \end{cases} \tag{2}$$

Each photon of light is supposed to be carried with energy *w* to satisfy the statistical requirement. Photon energy will decrease with step moving and ink absorption. It is necessary to point out that photon energy variation only refers to statistic energy decrease in this study, with which printing information can be obtained eventually. After each step movement, photon energy can be recorded *w'* shown by formula 3.

$$w' = \frac{\sigma_s}{\sigma_a + \sigma_s} w \tag{3}$$

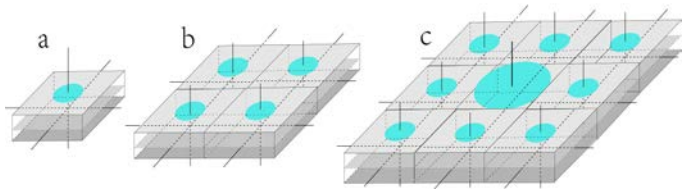



**Fig 2 The simulated area of different number of photons.(a) one dot.(b) four dots.(c) nine dots**

The design of printing dot shape is the final step. Whether photon will go through ink area depends on photon location and dot shape. In the case of round shape with radius *r*, photon has to pass ink layer before arriving to paper layer if radius *r* is greater than the distance from center to photon location*(x, y, z)*. The relationship between screening line *L* and radius *r* is presented by formula 4.

$$r = \frac{2.54}{L}\sqrt{\frac{a}{\pi}} \tag{4}$$

For printing project, dot shape is an important parameter. According to digital screening method, there are two types of dot. These are the *FM* dot and the *AM* dot [19]. Relevant studies have built reflectance model for one single printing dot using Monte-Carlo method, in which the reflectance accuracy would still need improvement especially in large dot area. Single round dot shape is shown in Fig 2(a). As the dot area grows larger, the interaction of dot will affect model accuracy. Modeling for four dots may be an improved approach, while it cannot eliminate interactive influence completely. The only revolution is expanding simulation area from one dot to nine dots as shown in Fig (c). Based on the above theory and the former study [14], an improved reflectance prediction model was established by expanding the simulated area from one dot to nine dots.

## 3 Results and discussion

The Monte-Carlo model for printing dot can simulate photon propagation. Apart from outputting reflectance value, the Monte-Carlo model can provide dot and photon distribution image which broadens the evaluation of quality control. The improved reflectance model for printing dot was built by expanding the simulation area. For computational model, the final aim is to provide a simulation value for designer to improve product quality. More attention should be paid on effectiveness and applicability

### 3.1 Optimization of simulation time

Visible light is a certain portion of the whole spectrum. Wave particle duality is the main attributes of light. The Monte-Carlo model for printing dot only concern light particle attributes. Therefore light can be regarded as a stream of particles. For printing model with Monte-Carlo method, the process of building the model is to simulate the moving path of photon in paper. The photon number required for statistic calculation, determining model effectiveness, has to be optimized. For medical field, Monte-Carlo model are commonly used to simulate radiation particles treatment [20]. Excessive dose will do harm to the body. There will be no efficacy if it is not enough. Besides, radiation treatment varies with each individual according to physical conditions. Similarly, it is necessary to confirm the required number of simulation photon in order to build an effective model for printing dot.

Comparative experiment was conducted with in Matlab 2016. The simulated results of different number of photon are presented in Fig 3. One of the refreshing aspects of Monte-Carlo model is directly outputting simulated results in the form of image. Fig 3(a) illustrate that single printing line becomes increasingly distinct with more simulated photons. Meanwhile, more time is required. Therefore, the compromise has to be made between accuracy and effectiveness. The required time with different number of photon is shown in Fig 3(b) showing as exponential function shape, which indicates that simulation number should not exceed 20 thousands. According to the simulated image and reflectance deviation, 10 thousand times is the optimized. Fig.3(c) show the 3D simulated result with 10 thousand photons.

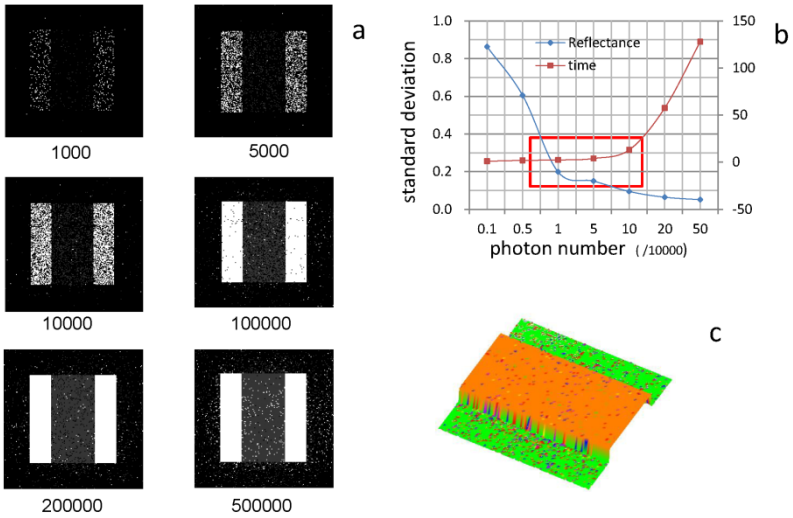



**Fig.3 The simulation results with different number of photon（a）the simulation of single printing line with different number of photon ranging from one thousand to fifty thousands，（b）standard deviation of reflectance and requiring time (c) 3-D simulated results of single printing line with ten thousand photons**

3.2 Reflectance simulation

The reflectance model that can provide a reference value before actual printing is a useful tool to make quality control. This is because printing is related to systems engineering, which needs several different technological processes. Any negligence in a process step could cause big effect. For a simple image made by digital camera, three main processes must be conducted to obtain high quality duplication. Digital screening is the first process for any original prints. That is because that ink cannot reflect different tone if it cover on paper uniformly. The ink that covered on the duplication is not continues. The ink distribute separately in the form of different dot shapes after digital screening. Under the characteristics of low pass filter of human eye, dots with different size could present multiple color tone.

Besides, digital screening can be divided into two types involving frequency and amplitude modulation screening according to screening theory. Both of them have advantage and drawback. The frequency screening can satisfy the tendency of fast print and print on demand with digital printing press. Although frequency screening is developing continually in recent year, amplitude screening still dominate in printing field for its excellent stability. The printing press will print sheets with received screening information. Through measure and analysis, the reflectance value can be obtained. In the case of amplitude screening, Fig.4 shows the main process of duplication. The traditional printing process is relatively

complicated. With the assistance of computer science especially for virtual reality, printing model can simulate printing quality. Fig.5 presents the comparative reflectance value of printing dot.

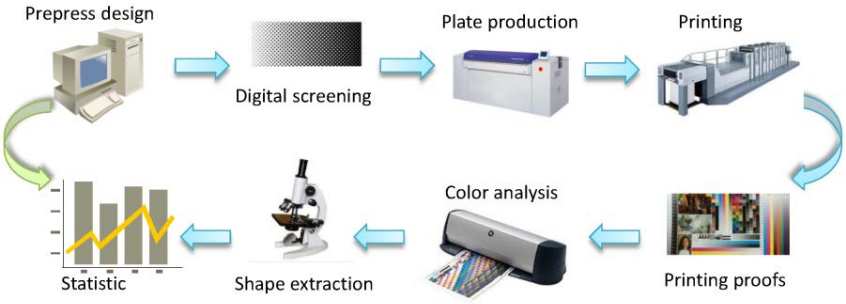



**Fig.4 The main process of printing duplication**

Using Monte-Carlo method and optical parameters of printing material, the printing model can smoothly simulate reflectance of printing dot with computer. Fig 5 showed the compared results of original model and the improved model. For a round dot in Fig.5(a,b), there is no apparent difference in the range of 0 to 60 percent area. But above 60 percent, the simulated value approach much close to the measured value. In the unit square area, the round dot will expand beyond the edge of screening unit when its area is bigger than 60 percent, which means that more interaction between dots occur gradually. As shown in Fig.1, bigger dot will cover on neighborhood dot and the overlap area could change photon moving path. In addition to that, printing ink is a kind of viscous fluid [21]. When ink approach on paper, it will spread around the center. Simultaneously, it also could permeate into the paper under the pressure of capillary effect of fibers [22]. Therefore, it is difficult to keep a round shape and to spread evenly. The original model could not take into account the overlap factors sufficiently. The comparative results in large dot area proved that the modified model indeed made an improvement.

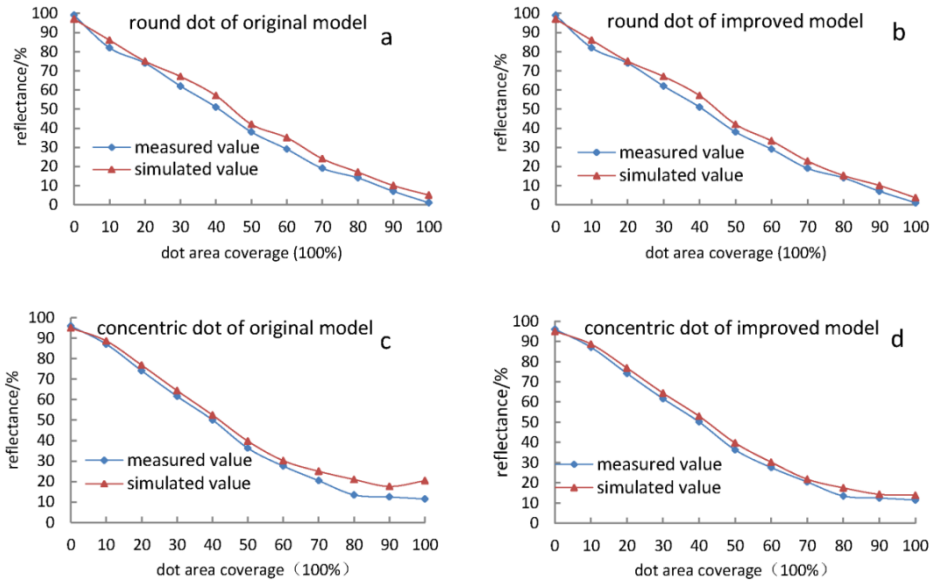



**Fig 5. The reflectance simulation results of original model and modified model. (a) round dot with original model.(b) round dot with modified model (c) concentric dot with original model (d) concentric dot with modified model. The simulated paper was 157g/m$^2$ coated paper. The measured value was obtained by IC-Plate and X-rite530.**

Concentric dot that has double ring structure raises increasing attention in recent year for two reasons. Firstly, printing works with concentric dot appears more saturated, which means using concentric dot can save printing ink to some extent. Secondly, concentric dot is good at duplicating warm tone, which is widely used to depict face figure. The only drawback exists in its special structure, costing much time in printing plate production. Moreover, reflectance character of concentric dot is unique as shown in Fig.5(c), indicating almost no change above 80 percent. This maybe caused by the compromise of ink overlap region and blank gap between double rings. The prototype model had not focused on the overlap ink region and just calculated within single dot unit neglecting interactive effect, which would not achieve ideal results for large dot area. The improved model showed better result in Fig.5(d). The correlation coefficient between the measured and the simulated value improved from 0.875 to 0.924. With simulated reflectance value, printing worker can adjust dot compensation to print exquisite duplication.

3.3 Photon distribution

The improved reflectance model can effectively record photon location when photon leaves paper and propagates into the air. Learning from geometric optical theory, three types of photon can be defined according to photon position. If photons leave from the upper or lower surface, those portions are called as reflection and transmission separately. The absorption refers to the photon whose supposed energy exhausted in paper. The simulated result of 9 round dots with one thousand photons is illustrated in Fig.6. The nine dots are the minimal unit for studying interactive effect between dots. Besides, one thousand photons are enough to describe the main characters of photon distribution.

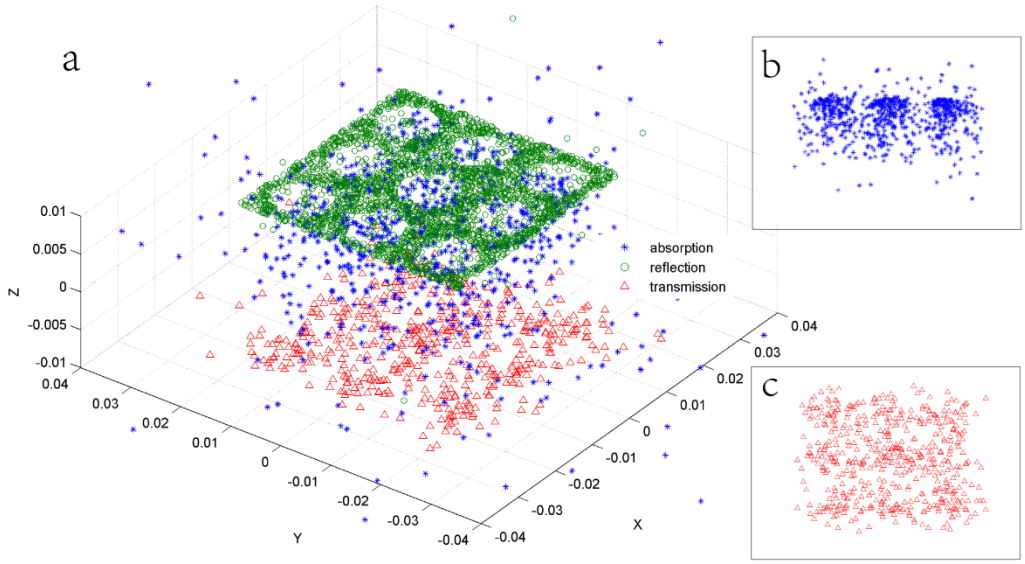



**Fig. 6. The photon distribution of nine round dots Parameters include photons numbers = 1000, dot area = 70%,and 157 g/m$^2$ coated paper.(a) the 3-D distribution of simulated photons when photons escape prints.(b) the side view of the absorbed photons.(c) the top view of transmitted photons.**

The reflection part domain the whole photons in Fig.6(a), which proved that most of photons were reflected by paper surface and only a small portion of photons have opportunity to get into the paper. The blank reign among reflected photons shows round shape corresponding to original screening state. The side view in Fig.6(b) shows regions where photons are absorbed completely. The absorbed photons mainly locate in the middle layer of the paper, providing useful reference for paper making process. If paper making factory needs low transparency

paper, more light absorbing material such as coating and filler should be concentrated in the middle layer. The transmission part of photons did not show apparent shape character in the top view Fig.6(c). This is because photons change its original way after several collision against fiber or filler. The distribution information outputted by reflectance model provides a new approach to research printing dot optical attributes and also make it possible to take targeted measure to instruct paper making production.

## 4 Conclusions

In this paper an improved reflectance prediction model for printing dot was proposed with the Monte Carlo method. Based on the prototype model, this paper mainly focuses on improving reflectance model accuracy by expanding dot simulation area from one single dot to nine dots. The results showed that the simulated value closely approached to the measured value with high dot area in particular and the correlation coefficient improved from 0.875 to 0.924. One hundred thousand photon packets proved to be the optimal number of simulated photon under regular printing condition. Apart from outputting reflectance value, the improved Monte-Carlo model can simulate dot shape as well as photon distribution, which make it possible to further promote optical study about color duplication. Future works could focus on simulate reflectance of various colors and finally realize spectrum simulation.

## Acknowledgment

This work was supported by the project of national key research and development (2017YFD0700102) and the funds of Shandong “Double Tops” with grant number of SYL2017XTTD14.

## References

[1] Arney, J. S., Engeldrum, P. G., and Zeng, H. (1995). An expanded Murray-Davies model of tone reproduction in halftone imaging, *Journal of imaging science and technology*,*39*(6), 502-508.
[2] Huntsman, J. R. (1987). A new model of dot gain and its application to a multilayer color proof, *Journal of Printing Science and Technology*,*24*(3), 189-202.
[3] Hébert, M., and Hersch, R. D. (2011). Yule–Nielsen based recto–verso color halftone transmittance prediction model, *Applied Optics*,*50*(4), 519-525.
[4] Ke, N., He, X., Wang, Y., and Zhang, Y. (2014). Improving Clapper–Yule model of the reflectance prediction by the path branching factor depending on the screen frequency of color halftone imaging, *Optik-International Journal for Light and Electron Optics*,*125*(20), 6242-6244.
[5] Balasubramanian, R. (1999). Optimization of the spectral Neugebauer model for printer characterization, *Journal of Electronic Imaging*,*8*(2), 156-167.
[6] Yang, L. (2010). Probabilistic spectral model of color halftone incorporating substrate fluorescence and interface reflections, *JOSA A*,*27*(10), 2115-2122.
[7] Edström, P. (2004). Comparison of the DORT2002 radiative transfer solution method and the Kubelka-Munk model, *Nordic Pulp & Paper Research Journal*,*19*(3), 397-403.
[8] Džimbeg-Malčić, V., Barbarić-Mikočević, Ž., and Itrić, K. (2012). Kubelka-Munk theory in describing optical properties of paper (II), *Tehnički vjesnik*,*19*(1), 191-196.
[9] Neuman, M., Coppel, L. G., and Edström, P. (2011). Point spreading in turbid media with anisotropic single scattering,*Optics express*,*19*(3), 1915-1920.
[10] Oshima, S., and Sankai, Y. (2011). Development of Red Blood Cell–Photon Simulator for Optical Propagation Analysis in Blood using Monte Carlo Method, *IEEE Transactions on Information Technology in Biomedicine*,*15*(3), 356-363.
[11] Zoia, A., Brun, E., Damian, F., and Malvagi, F. (2015). Monte Carlo methods for reactor period calculations, *Annals of Nuclear Energy*,*75*, 627-634.
[12] Lutsyshyn, Y. (2015). Fast quantum Monte Carlo on a GPU, *Computer Physics Communications*,*187*, 162-174.
[13] Modrić, D., Bolanča, S., and Beuc, R. (2009). Monte Carlo modeling of light scattering in paper, *Journal of Imaging Science and Technology*,*53*(2), 20201-1.

[14] Wang, Q., Yu, Y., Wang, T., and Liu, H. (2016). Research on a new reflectance prediction model of frequency modulated dots,*Optik,127*(22), 10539-10545.
[15] Walter, J. C. and Barkema, G. T. (2015). An introduction to Monte Carlo methods, *Physica A: Statistical Mechanics and its Applications,418*, 78-87.
[16] Reif, R., A'Amar, O., and Bigio, I. J. (2007). Analytical model of light reflectance for extraction of the optical properties in small volumes of turbid media, *Applied optics,46*(29), 7317-7328.
[17] Kattawar, G. W. (1975). A three-parameter analytic phase function for multiple scattering calculations, *Journal of Quantitative Spectroscopy and Radiative Transfer,15*(9), 839-849.
[18] Tancrez, M., and Taine, J. (2004). Direct identification of absorption and scattering coefficients and phase function of a porous medium by a Monte Carlo technique, *International Journal of Heat and Mass Transfer,47*(2), 373-383.
[19] He, Z., and Bouman, C. A. (2004). AM/FM halftoning: digital halftoning through simultaneous modulation of dot size and dot density, *Journal of Electronic Imaging,13*(2), 286-303.
[20] Yadavalli, V. K., Russell, R. J., Pishko, M. V., McShane, M. J., & Coté, G. L. (2005). A Monte Carlo simulation of photon propagation in fluorescent poly (ethylene glycol) hydrogel microsensors, *Sensors and Actuators B: Chemical,105*(2), 365-377.
[21] Fatkhullina D, Zhukova E. Study of ink optical properties by ATR spectroscopy, Conference Proceedings. 2013, 205-211.
[22] Alava, M., and Niskanen, K. (2006). The physics of paper, *Reports on progress in physics,69*(3), 669.

# The employment of porous media model to simulate lung microwave ablation

**Xiang Gao, *Qun Nan**

College of Life Science and Bioengineering, Beijing University of Technology, China

* Corresponding author: nanqun@bjut.edu.cn

## Abstract

Purpose: Microwave ablation is one of the minimally invasive local treatment techniques for lung tumors. But lung tissue contains plenty of air. In order to study the problem that poor ablation or excessive ablation in the treatment of lung tumors, the lung was used as a porous media model for numerical simulation, and compared with the traditional biological tissue model.

Methods: In this paper, based on finite element method, COMSOL Multiphysics is used to simulate numerically. Two-dimensional microwave antenna model, porous media lung model and biological tissue lung model are constructed. Coupled calculation of electromagnetic field, temperature field and flow field is carried out to obtain the temperature distribution of lung tissue. The Pennes bio-heat transfer model is used in the temperature field. The Brinkman equation is used for the flow field of the porous media model. The ablation frequency was 2450 MHz, the ablation power was 30 W, 40 W and 50 W, and the ablation time was 6 min.

Results: The ablation area is approximately spherical. When the power was 30W, 40W and 50W, the ablation area of porous media lung models were 25.9mm, 31.0mm and 35.1mm in transverse diameter, 44.8mm, 50.5mm and 54.4mm in longitudinal diameter, and the maximum temperature were 121℃, 149℃, 176℃, respectively. The ablation area of traditional biological lung models were 40.5mm, 46.9mm and 51.9mm in transverse diameter, 56.3mm, 62.1mm and 66.7mm in longitudinal diameter, and the maximum temperature were 183℃, 232℃, 280℃, respectively. Compared with the biological tissue model, the porous media model is 29% smaller in ablation area and 36% smaller in maximum temperature. The simulation results of porous media model are more close to the results of ex-vivo experiments.

Conclusions: In the traditional biological tissue model, lung ablation does not consider the influence of air, so its ablation area is often large. The porous media model is more suitable for lung tissue with sufficient air. The method combines the electromagnetic field emitted by microwave, the flow field of air in lung tissue and the heat transfer field of porous media. It can accurately simulate the real-time distribution of temperature field in lung tissue under microwave irradiation, which plays an important role in real-time monitoring of temperature field in the process of microwave ablation of lung tissue.

**Keywords:** Microwave ablation; Numerical simulation; Lung; Porous media

## 1. Introduction

Lung cancer is one of the most common malignant tumors, ranking first in the causes of cancer incidence and death. Currently, resection is still the mainstream treatment. However, the limitations of resection surgery are very large, only about 20% of patients are suitable for surgery [1][2]. Microwave ablation for lung cancer is the use of microwave antenna puncture into lung tumors by using medical imaging equipment, such as ultrasound imaging, magnetic resonance imaging and computed tomography imaging. Under the action of microwave electric field, the tumor tissue produces high temperature above 60 ℃ in a short time, which causes the degeneration and necrosis of the tumor tissue. So as to achieve the purpose of treatment of tumors [3][4].

But lung tissue contains enough air, so the conductivity and thermal conductivity of lung tissue are very small, the temperature is difficult to control. Therefore, some large tumors and irregular tumors will have poor curative effect. It can also cause serious complications due to excessive ablation, which limits the application of this technology in clinical practice [5].

In this study, lung tissue is regarded as porous media, and the porous media model is used for simulation calculation, and compared with the traditional biological tissue model, in order to establish a more realistic simulation model.

## 2. Methods

### 2.1 Geometric model

COMSOL Multiphysics multi-physical field coupling analysis software is used in the simulation, and finite element method is used in the calculation.

The microwave antenna used in this study is a commonly used microwave antenna in clinic. Its frequency is 2450 MHz. Its geometric model structure is shown in Fig. 1 and Table 1 [6]. In this study, microwave antenna model, porous media lung model and biological tissue lung model are constructed. Because the action area of microwave antenna is axisymmetric, it is simplified to an axisymmetric two-dimensional. At the same time, assuming that the lung is isotropic homogeneous organization, the microwave antenna is inserted into the lung tissue for 10 cm. The two-dimensional model is constructed as shown in Figure 2.

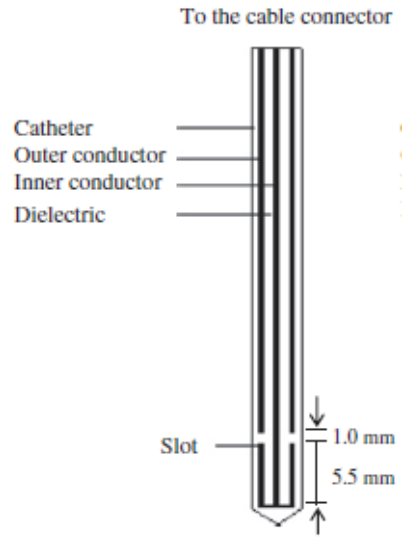



**Figure 1. Geometric model of microwave antenna**

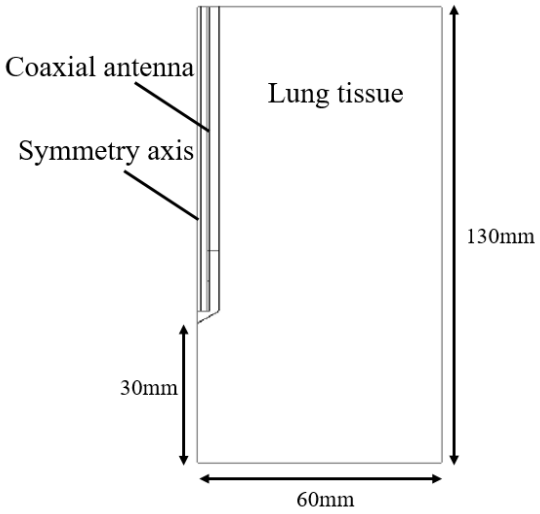



**Figure 2. Two-Dimensional Simulated Geometric Model**

**Table 1. Dimensions of microwave antenna[6]**

| Materials | Dimensions (mm) |
|---|---|
| Inner conductor | 0.135 (radial) |
| Dielectric | 0.335 (radial) |
| Outer conductor | 0.460 (radial) |
| Catheter | 0.895 (radial) |
| Slot | 1.000 (wide) |

### 2.2 Bio-heat equation

The Pennes bio-heat equation (Eq. (1)) governs heating transfer during the thermal ablation [7].The equation is as follows:

$$\rho c\frac{\partial T}{\partial t} = \nabla \cdot (k\nabla T) + \omega_b C_b(T_b - T) + q_m + q_r \quad (1)$$

In the formula, $T$(℃)、ρ(kg/m3)、$C$(J/kg·℃) and $k$( W/m·℃) are the temperature, density, specific heat and heat conductivity of tissues, $\omega_b$ is the blood perfusion rate(kg/m3·s), $C_b$ and $T_b$ are the blood specific heat and blood temperature, $q_m$ is the heat production rate of tissue metabolism, and $q_r$ is the external heat source.

In the porous media lung model, the flow field is calculated by using Brinkman flow equation coupled with temperature field and electromagnetic field.

## 3. Results and Discussion

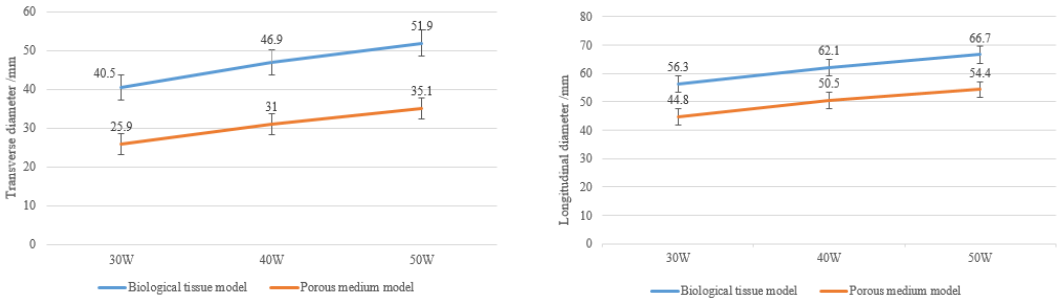
The temperature distribution and the data analysis of the ablation area at different ablation power (30W, 40W, 50W) were studied. The calculation time was 360 s. The temperature boundary was 60 ℃. The ablation area parallel to microwave antenna was longitudinal and the ablation area perpendicular to the microwave antenna was transverse. The experimental data and temperature distribution at different ablation power are shown in Table 2.

**Table 2. Experimental data**

| Input powers | 30W | | 40W | | 50W | |
|---|---|---|---|---|---|---|
| | Biological tissue model | Porous medium model | Biological tissue model | Porous medium model | Biological tissue model | Porous medium model |
| Transverse diameter /mm | 40.5 | 25.9 | 46.9 | 31.0 | 51.9 | 35.1 |
| Longitudinal diameter /mm | 56.3 | 44.8 | 62.1 | 50.5 | 66.7 | 54.4 |
| Highest temperature/ ℃ | 183.2 | 121 | 231.9 | 149 | 280.7 | 176 |

As can be seen from Table 2, with the increase of ablation power, the range of ablation temperature field increases gradually, the maximum temperature rises continuously, and the ablation areas are ellipsoidal. When the power was 30W, 40W and 50W, the ablation area of porous media lung models were 25.9mm, 31.0mm and 35.1mm in transverse diameter, 44.8mm, 50.5mm and 54.4mm in longitudinal diameter, and the maximum temperature were 121℃, 149℃, 176℃, respectively. The ablation area of traditional biological lung models were 40.5mm, 46.9mm and 51.9mm in transverse diameter, 56.3mm, 62.1mm and 66.7mm in longitudinal diameter, and the maximum temperature were 183℃, 232℃, 280℃, respectively. Compared with the biological tissue model, the porous media model is 29% smaller in ablation area and 36% smaller in maximum temperature. As shown in Figure 3, although the bio-tissue model is higher than the porous media model in terms of ablation zone range and maximum temperature field, the growth trend of ablation zone length with power of the two models is highly consistent. But comparing the two simulation results with

the experimental results in vitro, the results obtained by the porous media model are closer to the ablation data in vitro [8].




(a)Transverse diameter (b) Longitudinal diameter

**Figure 3. Changes in the length of the lung tissue with power**

## 4. Conclusions

In the traditional biological tissue model, lung ablation does not consider the influence of air, so its ablation area is often large. The porous media model is more suitable for lung tissue with sufficient air. The method combines the electromagnetic field emitted by microwave, the flow field of air in lung tissue and the heat transfer field of porous media. It can accurately simulate the real-time distribution of temperature field in lung tissue under microwave irradiation, which plays an important role in real-time monitoring of temperature field in the process of microwave ablation of lung tissue.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (31771021) and Jiangsu Science and Technology Project (BE2017758).

**References**

[1] F Bray, J Ferlay, I Soerjomataram, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.[J]. Ca Cancer J Clin, 2018, 68, 394–424.
[2] Carrafiello G , Mangini M , Bernardi I D , et al. Microwave ablation therapy for treating primary and secondary lung tumours: technical note[J]. La Radiologia Medica, 2010, 115(6):962-974.
[3] Ahmed M, Solbiati L, Brace CL, et al. Image-guided tumor ablation: standardization of terminology and reporting criteria-a 10-year update[J]. Radiology, 2014, 273(1): 241-260.
[4] Brace C L . Microwave Tissue Ablation: Biophysics, Technology, and Applications[J]. Critical Reviews in Biomedical Engineering, 2010, 38(1):65-78.
[5] Huang W, Ma X S, Liu Y E. Application of Microwave Ablation for Lung Cancer [J]. Chinese Journal of Clinical Thoracic and Cardiovascular Surgery, 2015(3):265-268.
[6] Rattanadecho P , Keangin P . Numerical study of heat transfer and blood flow in two-layered porous liver tissue during microwave ablation process using single and double slot antenna[J]. International Journal of Heat and Mass Transfer, 2013, 58(1-2):457-470.
[7] Pennes H H . Analysis of tissue and arterial blood temperatures in the resting human forearm. 1948.[J]. Journal of Applied Physiology, 1998, 85(1):5-34.
[8] Luo G C. Experimental study on 2450Mhz microwave ablation for poreine lung in vitro [D]. Taishan Medical University, 2014.

# Approximate calculation of certain type of statically indeterminate truss

***Janusz Rębielak[1]**

Faculty of Architecture, Cracow University of Technology, Poland.

*Presenting author: j.rebielak@wp.pl

## Abstract

The paper presents examples of approximate calculations of force values in members of a certain type of truss, which is at the same time an internal and external statically indeterminate system. Static calculations are carried out for two selected forms of trusses by the application of the two-stage method of computations of such structural systems. The two-stage method, due to the application of the principle of superposition, makes possible the calculation of such complex trusses by the means of e.g. the Cremona's method. In this two-stage method the static calculations are done in two stages. In each stage a statically determinate truss system is considered, a pattern of which is defined by removing from the basic truss a suitable number of members. The number equals the degree of statically indeterminacy of the basic truss. In the paper are presented results of calculations of statically indeterminate trusses carried out by the two-stage method, which moreover are the external statically indeterminate systems. There are also presented results of calculations of the same trusses done by means of a suitable computer software together with the comparison of outcomes obtained in two different methods.

**Keywords:** Statically indeterminate system, Superposition principle, Calculus of vectors, Cremona's method, Approximate solution.

## Introduction

The values of forces in members of the statically indeterminate trusses have to be determined by a suitable computational method, which among others takes into consideration the stiffness differences between members connected to the truss nodes. For this purposes are applied such methods like for instance, the force method, the displacement method, the iteration methods like the method of successive approximations, and the finite elements method, etc. [1]-[8]. Mathematic concepts of these methods are adapted in modern types of various computer software [9][10]. The two-stage method has been invented during initial static analysis of a certain type of a tension-strut structure [11]-[13]. If the basic structure is overloaded then a certain number of members are excluded form process of the force transmission. Number of these members is equal to the statically indeterminacy of the basic truss system. The point of the two-stage method is to carry out static calculations in two independent stages for statically determinate trusses, shapes of which are determined by removing from area of the basic truss the number of members equal to its statically indeterminacy. In each stage it is considered an appropriate statically determinate truss, therefore values of forces in its members can be calculated by means of e.g. the Cremona's method. In the both stages are considered suitable trusses having the same clear span and construction depth like the basic truss, but they are appropriately loaded by forces having half values in comparison to values of forces applied to the basic truss. Final values of forces acting in all members of the basic truss are resultants of the force values determined in each stage for the suitable member. The calculation procedure of the two-stage method is justified by rules of calculus of vectors and principle of superposition. It is one of the approximate methods of calculation of the statically indeterminate systems [14].

## Definition of research problem

The correctness of the basic theoretic assumption of the two-stage method has been verified by calculating the internal statically indeterminate trusses supported on two supports, while the one is the pivot bearing and the second one is the pivot sliding bearing. It implies that there are

only three unknown bearing reactions. Two forms of trusses were computed in order to prove feasibility study that the two-stage method can be useful for calculations of the internal statically indeterminate trusses being at the same time the external statically indeterminate trusses. The first one is the vertically positioned truss loaded by the means of horizontally applied forces and supported by two pivot bearings. The second one is the horizontally located truss loaded by the means of vertically and optionally applied forces, and which is supported by two pivot bearings. All the calculations are carried out for the same geometric, structural and load conditions. Results obtained in this way are compared with outcomes gained by the application of suitable computer software for calculation of the force values in a truss having the same geometric, structural and load conditions.

## First subject of static calculation and analysis

The assumed form of the basic truss is built by the usage of three square strut modules located vertically on each other, see Fig. 1. The truss is supported by two pivot bearings A and B and it is loaded by three concentrated forces F horizontally applied to nodes of its left vertical chord. All calculations have been made for the single load forces F, each of value equal to 1.00 kN, applied to nodes of a basic truss. It was assumed that the construction depth of the truss tower "H" equals 1.00 meter, while its height "L" is equal to 3.0 meters. The number of nodes was assumed to denote by symbol "w", while symbol "p" to define number of the members. The condition for the internal statically determinacy of the plane truss is determined by the following equation:

$$p = 2 \cdot w - 3 \tag{1}$$

The considered basic truss system is built by number of nodes w = 11, what implies that the statically determinate truss created by the means of this number of nodes has to be constructed by the following number of members:

$$19 = 2 \cdot 11 - 3 \tag{2}$$

The basic truss system is built by the number of members p = 22, what implies that the calculated structure is the threefold statically indeterminate system. It implies further that in order to create an appropriate statically determined system it is necessary to exclude 3 members from area of the basic truss.

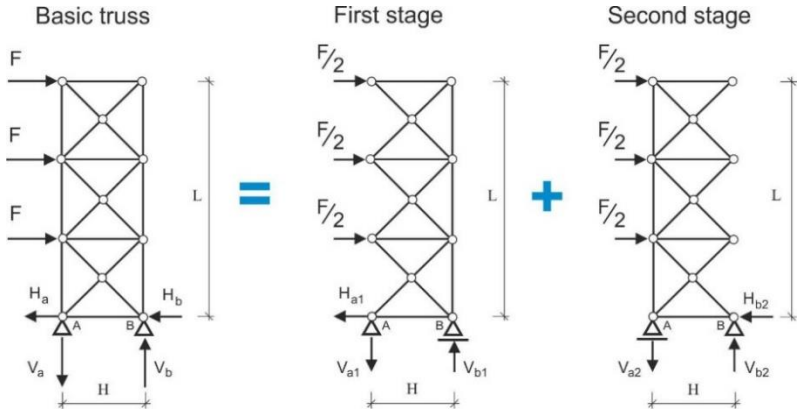



**Figure 1. The concept of the two-stage method applied for calculation of a vertically positioned statically indeterminate truss**

According to the rules of the two-stage method in its first stage one should remove three members, for instance from the left vertical chord of the basic truss, and then to apply load forces of half values (F/2) to suitable nodes of this chord. In the second stage it is necessary to remove three members from the right vertical cord and, like previously, to apply load forces of half values (F/2) having the same senses, like in the basic truss, to corresponding nodes of the left vertical chord. Because the basic truss is supported by two pivot bearings A and B, therefore it is also the external statically indeterminate system. That is why similar operations have to be

undertaken regarding the changes of statuses of the supports. In the first stage it is proposed to keep support A as the pivot bearing and consider support B as the pivot sliding bearing, see Fig. 2. In the second stage support A is considered as the pivot sliding bearing, while support B remains the pivot bearing, see Fig. 3. Values of forces acting in members of the basic truss are resultants of forces calculated in corresponding members at each stage. The concept of the two-stage method is compatible with the rules of calculus of vectors, principle of superposition and respects the three fundamental conditions of equilibrium presented below:

$$\sum_{i=1}^{n} F_{ix} = 0 \tag{3}$$

$$\sum_{i=1}^{n} F_{iy} = 0 \tag{4}$$

$$\sum_{i=1}^{n} M_{i} = 0 \tag{5}$$

The results of the calculations of the basic truss, having a tower configuration, see Fig. 1, obtained in each stage of the two-stage method by application of the Cremona's method, are presented respectively in Fig. 2 and in Fig. 3. Final values of the forces, defined in this method, in all members of the basic truss are shown in Fig. 4a.

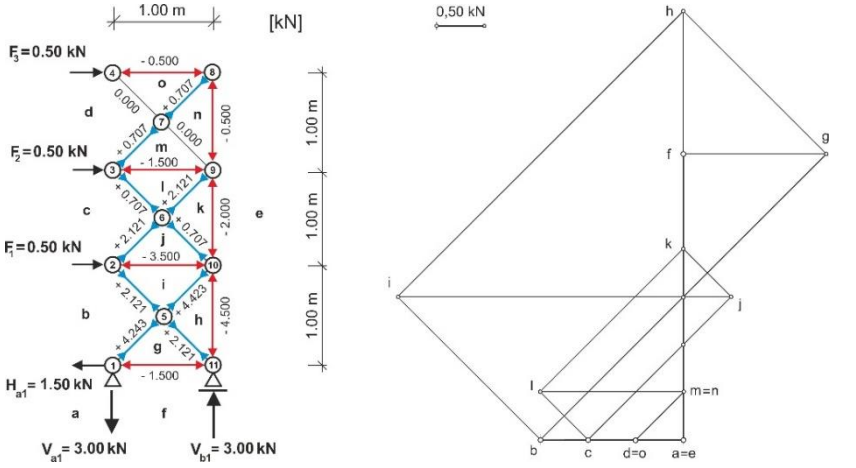



**Figure 2. The values of the forces determined in the first stage of calculations for the truss of a tower configuration with suitable Cremona's polygon of forces**

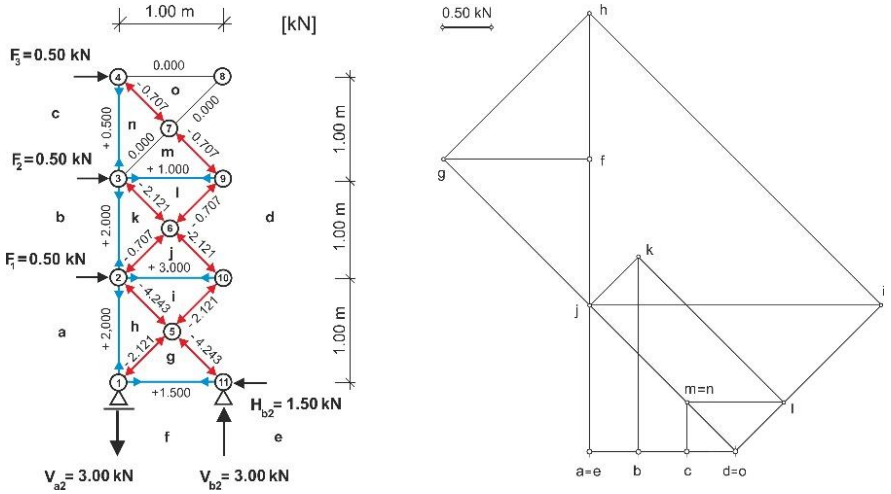



**Figure 3. The values of the forces determined in the second stage of calculations for the truss of a tower configuration with suitable Cremona's polygon of forces**

The same vertical, tower configuration of the basic truss has been subjected to the static calculation carried out by the application of the Autodesk Robot Structural Analysis Professional 2019. The computer software is considered to be the very precise tool for calculation of the force values acting in members of the statically indeterminate systems. Static calculations were made by the assumption that the truss consists of the steel tubular members having diameter of 30.00 mm, the thickness of the section equals to 4.00 mm and the steel material has the Young's modulus equal to 210 GPa. The results of the computer calculations of the same truss system are presented in Fig. 4b.

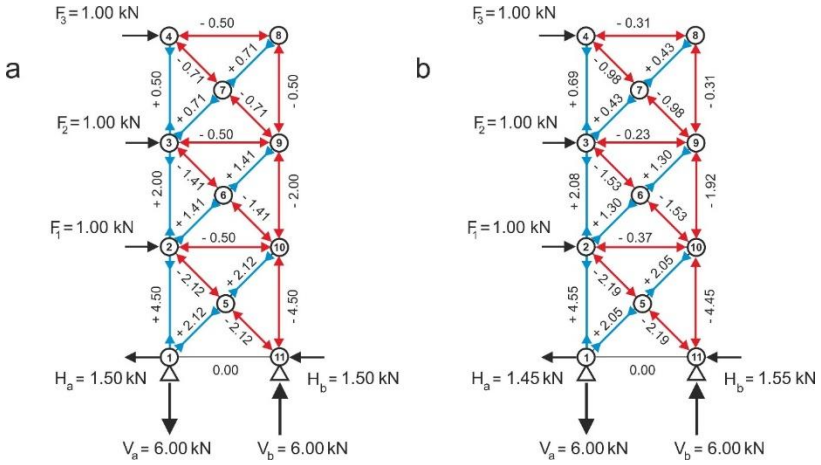



**Figure 4. Comparison of the values of forces in members for the tower truss configuration calculated, a) in the two-stage method, b) by the means of computer software**

In the two-stage method the final values of forces acting in particular members are calculated according to the rules of the calculus of forces and to principle of superposition. For instance the final value of a compressive force in the cross brace located between node 3 and node 6 is equal to -1.41 kN, see Fig. 4a. It is a resultant of the tensile force +0.707 kN determined in the first stage in corresponding member located between nodes of the same numbers, see Fig. 2, and the compressive force -2.121 kN determined in the second stage, see Fig. 3. In similar way the final force is defined in e.g. vertical member placed between node 3 and node 2. In the first stage of calculation this member has been rejected from the basic truss, see Fig. 2, therefore it is assumed, that in this case the appropriate force value equals 0.000 kN. In the second stage the value of tensile force defined in corresponding member is equal to +2.000 kN, see Fig. 3. Therefore the final force value in member located between nodes 2 and 3 equals +2.000 kN, see Fig. 4a.

From the comparison of the force values gained in both methods for the same truss members follows that in general the results are congruent to each other. For instance value of a compressive force determined in the member placed between node 3 and node 9 by application of the two-stage method is equal to -0.50 kN. The force value defined in the same member by application of the computer software mentioned above equals -0.23 kN. The relative difference equals up to 54% of the bigger value. One can notice a smaller differentiation between values of forces calculated in two different methods for the same member e.g. in vertical member placed between node 2 and node 1. The value of the tensile force calculated for this member in the two-stage method equals +4.50 kN, while by applying of the computer software it is equal to +4.55 kN. In this case the relative difference is equal to only ca. 1% of the bigger value. It is worthy to notice that values of all types of the suitable bearing reactions determined in both compared methods are the same or they are of very approximated values.

## Second subject of static calculation and analysis

The scheme of the basic truss presented in Fig. 5 has been selected as an object of further comparative investigation in order to estimate the usefulness of the two-stage method for calculation of all types of the planar internal and external statically indeterminate systems. The basic truss is of similar structure like the previous one but it is located horizontally and

supported by two supports, both being the pivot bearings. That is why four unknown bearing reactions have to be considered in these supports. It implies that the basic truss, like the previous one, is the threefold internal indeterminate system and once-fold external indeterminate system. Moreover the truss is loaded by the forces, which may be applied at any directions, what is represented by the direction of force $F_1$. This force is inclined at an angle of 45 degrees towards the horizontal line. According to the rules of the two-stage method presented above, in its first stage it is calculated truss, which form is obtained by removing three members of e.g. the bottom chord from the pattern of the basic truss. The shape of truss considered in the second stage is a result of the cancelation of three members of the top chord from the basic truss. In both stages the calculated truss is loaded by forces of the half values applied in appropriate way to the suitable truss system. The results of calculations obtained in the first stage are shown in Fig. 6. The values of the forces defined in the second stage of calculations are presented in Fig. 7. The final values of forces calculated in the two-stage method are presented in Fig. 8a. The results obtained by the application of suitable computer software are shown in Fig. 8b.

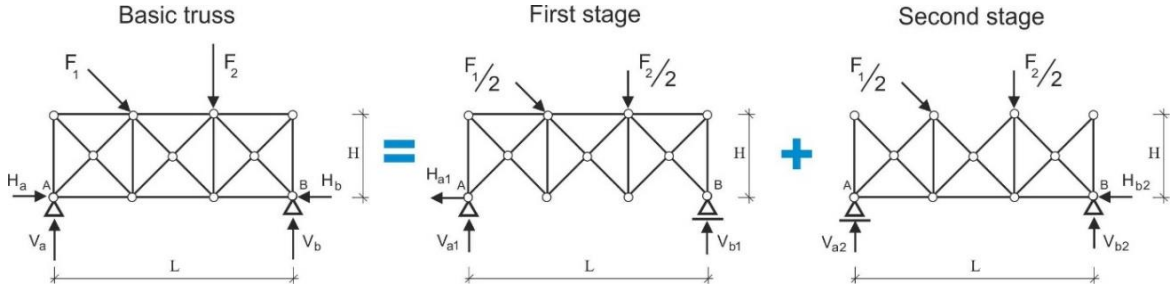



**Figure 5. The concept of the two-stage method applied for the calculation of horizontally positioned statically indeterminate truss supported in two pivot bearings**

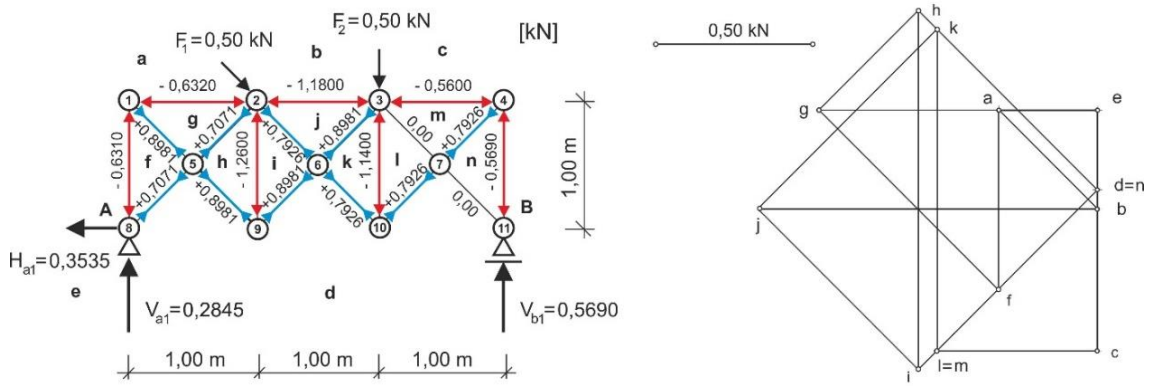



**Figure 6. The values of the forces defined in the first stage of the calculations for horizontally positioned truss with suitable Cremona's polygon of forces**

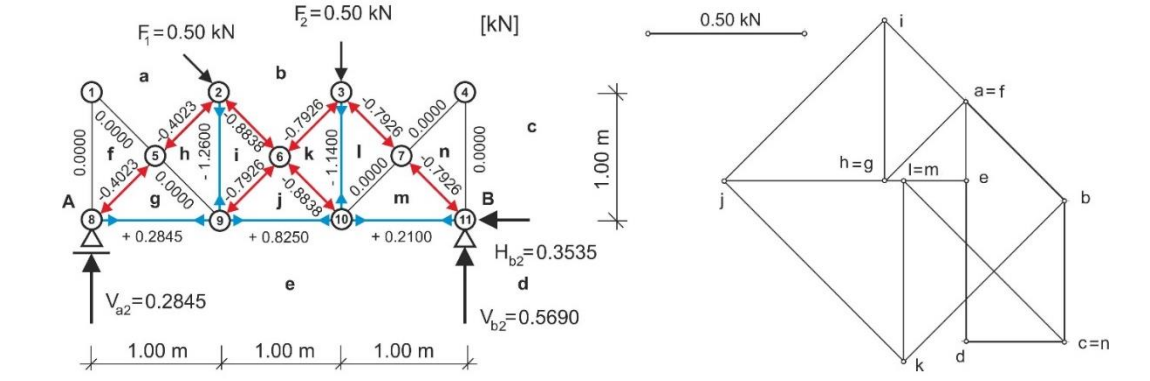



**Figure 7. The values of the forces defined in the second stage of the calculations for horizontally positioned truss with suitable Cremona's polygon of forces**

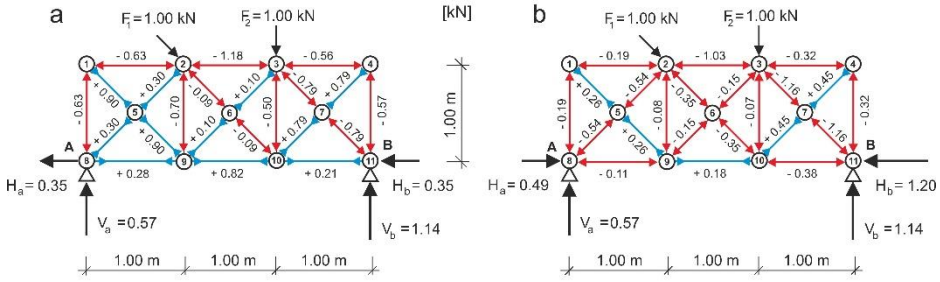



**Figure 8. Comparison of the values of the forces in the members of horizontally positioned truss calculated, a) in the two-stage method, b) by application of computer software**

One can notice significant differences of the values and even of senses of the force determined in the same members in two compared methods of calculations. For instance, by application of the two-stage method value of tensile force in a member located between nodes 3 and 6 equals +0.10 kN, while by applying of the computer software it was defined as a compressive force of value equal to -0.15 kN. Much smaller difference is noticed between compressive force value defined in a member placed between nodes 3 and 2 by the help of the two-stage method, which equals -1.18 kN, while the outcome of computer software defines it as a compressive force having value of -1.03 kN. One can observe the substantial differentiations in values of forces calculated by the application of compared methods in numerous members of the truss. For example, in the cross brace placed between nodes 2 and 5 in the two-stage method it is calculated as a tensile force of value equal +0.30 kN, while in the same member a force calculated by means of computer software is defined as the compression force of the value equal to -0.54 kN. The values and senses of the vertical components of the bearing reactions calculated in both methods are equal. To the most significant differences one has to count the differentiation of values and senses of the horizontal components of the bearing reactions. These reactions, defined in the two-stage method, are of equal values and both have same senses, see Fig. 8a, what is directly determined by the basic principles of this method. Because the investigated truss is also the once-fold external indeterminate system, therefore the real horizontal components of the bearing reactions are of the values and senses presented in Fig. 8b. It implies that the horizontal components of these reactions have to be calculated in another way by application of the two-stage method.

## The calculation of the horizontal components of reaction in two stages

General schemes of a computation procedure proposed for the calculation of approximate force values of this type of the bearing reactions in the basic truss system are presented in Fig. 9.

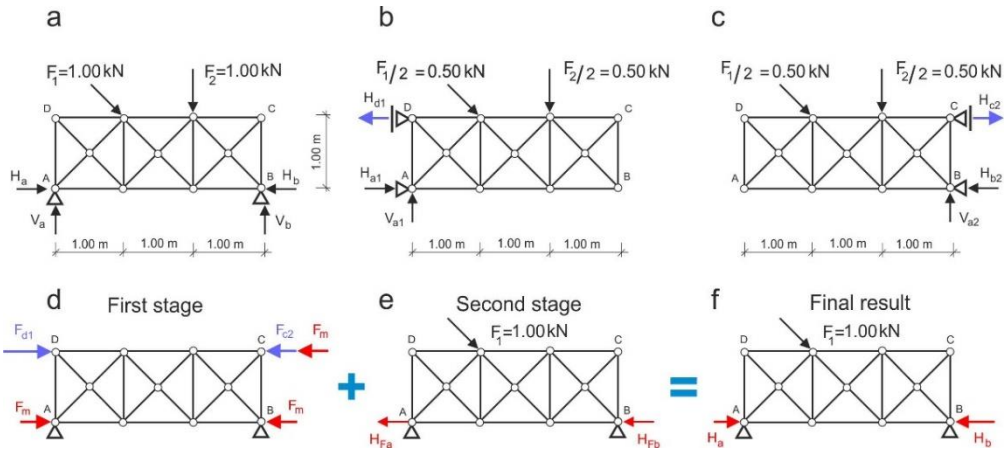



**Figure 9. Schemes of the computation procedure of calculation of horizontal components of bearing reactions in two stages**

Taking into account general static principles of the two-stage method it was assumed, that the basic truss, see Fig. 9a, can be considered as suitable composition of two respectively supported cantilever trusses having the same internal structure and loaded in the same way like the basic truss. In the first step the truss is considered as cantilever system supported in nodes A and D, see Fig. 9b. Because corner node D in the basic truss is not the support node therefore in the first step it is treated as the sliding pivot bearing, in which only horizontal component of reaction $H_{d1}$ can exist. In this case the value of the horizontal reaction is computed in the following way:

$$\sum M_A = 0 \tag{6}$$

$$\frac{F1x}{2} \cdot 1.00m + \frac{F1y}{2} \cdot 1.00m + \frac{F2}{2} \cdot 2.00m - H_{d1} \cdot 1.00m = 0 \tag{7}$$

$$\frac{\sqrt{2}}{4}\ kN \cdot 1.00m + \frac{\sqrt{2}}{4}\ kN \cdot 1.00m + 0.5\ kN \cdot 2.00m - H_{d1} \cdot 1.00m = 0 \tag{8}$$

$$H_{d1} \approx 1.7071067\ kN \tag{9}$$

In the second step of the proposed procedure the truss is also considered as the cantilever system, but this time supported in nodes B and C, see Fig. 9c. Because node C in the basic truss is also not the support node, therefore in this stage, like previously, it is considered as being supported in pivot sliding bearing where only horizontal component of reaction $H_{c2}$ can exist. Its value will be defined in the way presented below:

$$\sum M_B = 0 \tag{10}$$

$$-\frac{F1y}{2} \cdot 2.00m + \frac{F1x}{2} \cdot 1.00m - \frac{F2}{2} \cdot 1.00m + H_{c2} \cdot 1.00m = 0 \tag{11}$$

$$-\frac{\sqrt{2}}{4}\ kN \cdot 2.00m + \frac{\sqrt{2}}{4}\ kN \cdot 1.00m + 0.5\ kN \cdot 1.00m + H_{c2} \cdot 100m = 0 \tag{12}$$

$$H_{c2} \approx 0.8535533\ kN \tag{13}$$

The values of horizontal components of the bearing reactions $H_{a1}$ and $H_{b2}$, determined in this calculation in support A and B, have to be omitted, what is justified by the first condition of equilibrium (3). Value of the force $H_{d1}$ is bigger than the force value $H_{c2}$. Difference of values of these two is called $F_m$ and its value equals:

$$F_m = H_{d1} - H_{c2} = 1.7071067\ kN - 0.8535533\ kN \tag{14}$$

$$F_m = 0.8535534\ kN \tag{15}$$

Members located between nodes D and C in the top chord of the basic truss are subjected to act of compressive forces, compare Fig. 8. Therefore force $F_{d1}$, see Fig. 9d, has to have the same value like the force $H_{d1}$ but its sense must be inversed and it is applied to the corner node D. Similar operation one should make in corner node C. Horizontal force $F_{c2}$ is applied to this node, its value is equal to value of force $H_{c2}$, but its sense is oppositely directed. Force $F_m$ has to be appropriately applied to the top chord in order to keep the force balance in this chord. From the first condition of equilibrium (3) follows, that in the first stage force $F_m$ has to be applied to node C, having the same sense like force $F_{c2}$, see Fig. 9d. The condition of equilibrium of the whole considered structure justifies suitable application of two oppositely directed horizontal forces $F_m$ to the two bearing nodes A and B. From the basis of the same condition it follows that horizontal components of the bearing reactions $H_{Fa}$ and $H_{Fb}$ are of the same values, as well as their senses, like it is shown in Fig. 9d. When the horizontal component of the load force $F_1 = \sqrt{2}/2$ kN $\approx 0.7071067$ kN then the absolute values of the bearing reactions $H_{Fa}$ and $H_{Fb}$ are equal to 0.3535533 kN. Final values of horizontal components of bearing reactions acting in these supports, see Fig. 9f, will be the resultants of horizontal reactions calculated in the first stage, see Fig. 9d, and calculated in the second stage, see Fig. 9e.

$$H_a = F_m - H_{Fa} = 0.8535534\ kN - 0.3535533\ kN \approx 0.50\ kN \quad (16)$$

$$H_b = F_m + H_{Fb} = 0.8535534\ kN + 0.3535533\ kN \approx 1.21\ kN \quad (17)$$

The values of horizontal components of bearing reactions computed in this way are almost of the same values as if they were calculated by the application of the computer software, compare Fig. 8b, for the same type of the internal and external statically indeterminate truss. The procedure presented above applies the rules of calculus of forces as well as the principle of superposition. That is why it can be considered as an integral part of the two-stage method of calculation of statically indeterminate trusses.

### Another way of calculation of the same basic statically indeterminate truss

Application of the two-stage method in a direct way gives in result not exact values and senses of horizontal bearing reactions, compare Fig. 8a. In order to recognize features of the two-stage method it has been applied for calculation of the same basic truss, see Fig. 9a, for values of horizontal bearing reactions determined in two stages described above. Intermediate results are presented in Fig. 10 and in Fig. 11. Final force values are shown in Fig. 12.

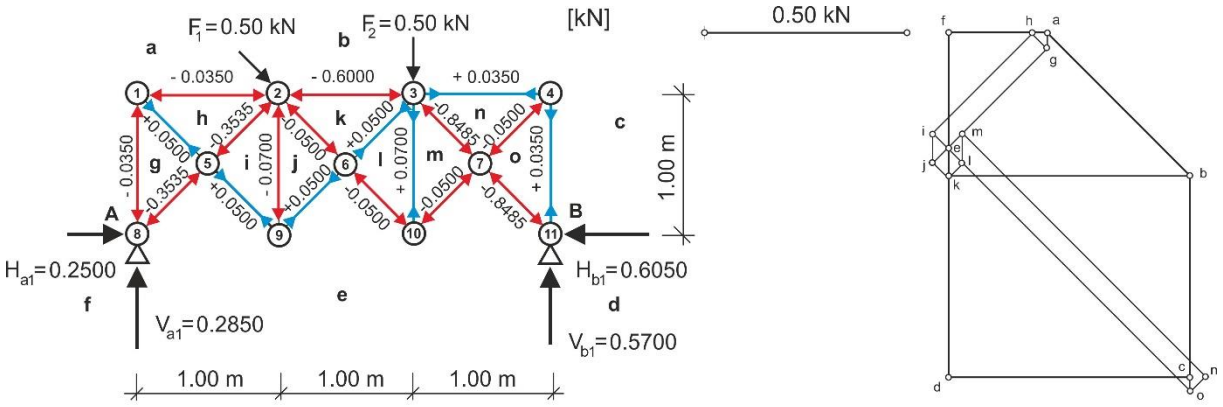



**Figure 10. Force values computed in the first stage of calculations for values of bearing reactions of basic truss estimated in two stages together with Cremona's polygon of forces**

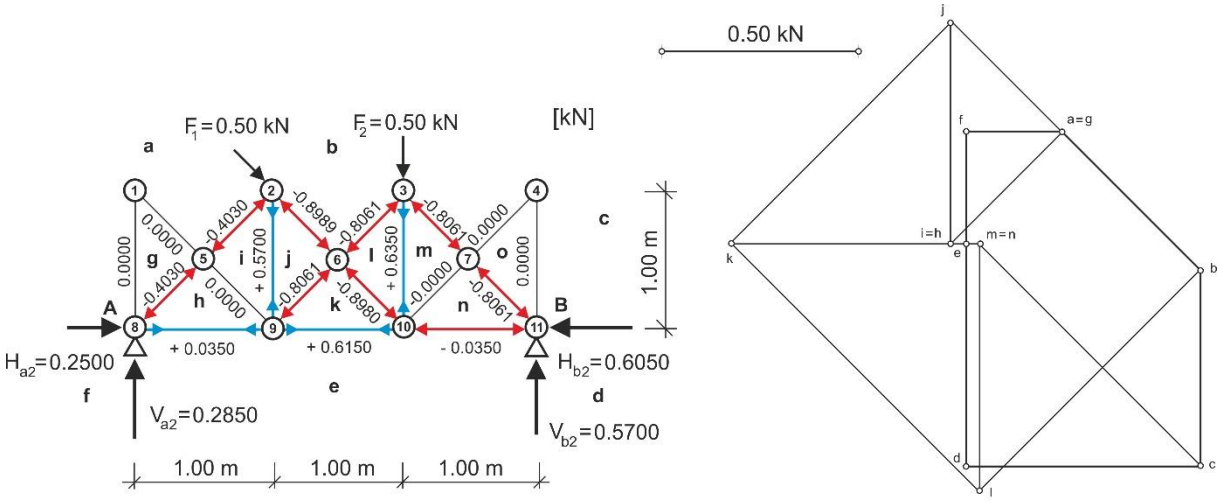



**Figure 11. Force values defined in the second stage of calculations for values of bearing reactions of basic truss estimated in two stages together with Cremona's polygon of forces**

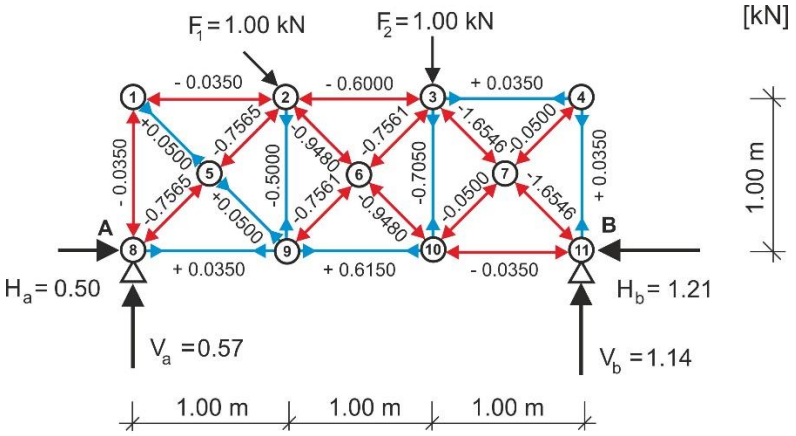



**Figure 12. Values of forces determined in members of the basic truss as a result of alternative static calculations**

Values of forces calculated in members of the investigated basic truss by application of the Autodesk Robot Structural Analysis Professional 2019 are considered as the exact results. In comparison with them the force values determined in the two-stage method for the same members by taking into account the approximate values of the real horizontal components of the bearing reaction, while $H_a$ = 0.50 kN and $H_b$ = 1.21 kN and being almost equal to the exact values, are considerably different. The degree of differentiation is even bigger than in the results of calculations carried out by the direct application of the simple rules of the two-stage method. For instance compressive force value computed by means of the computer software in the member located between nodes 2 and 3 equals -1.03 kN, see Fig. 8b. In the same member due to the application of the simple kind of two-stage method, value of compressive force is defined as equal to -1.18 kN, see Fig 8a, while due application of the two-stage method with taking into account real values of horizontal components of bearing reactions, the calculated value of compressive force equals -0.60 kN, see Fig. 12. Similar remarks refer also regarding various senses of the vector forces calculated in both compared kinds of the two-stage method. From comparative analysis of the presented results calculated for the horizontal positioned trusses loaded by forces applied at optional directions to the truss nodes follows, that somewhat better approximate force values to the exact values of forces acting in members of the truss, one can obtain by application of the simple kind of the two-stage method, see Fig. 8a. Values of the horizontal components of the bearing reactions should then be calculated separately in the way described above, the procedure of which is presented in schemes shown in Fig. 9.

## Conclusions

The two-stage method is an approximate method of calculation of the statically indeterminate trusses because in both stages it applies the rules, which are appropriate for the calculation procedures of the statically determinate trusses. The degree of approximation of the obtained force values to the values of forces defined by means of the exact methods in general is good enough when the two-stage method is applied for calculation of the inner indeterminate trusses. One can notice significantly differences between the exact and appropriate force values calculated for members, where are acting the smallest forces, especially having absolute values close to zero. However one should be aware that members subjected to the act of such forces are designed mostly according to instructions of the building codes. In these cases the cross-sectional areas of such members are much larger than they are determined directly on basis of the results of static calculations. The accuracy of the force values calculated by application of the two-stage method can be significantly improved by taking into account the differences between the stiffness of members connected in the truss nodes. It can be made by defining a set of appropriate coefficients defining differences of members connected to each particular node. When the two-stage method is applied for calculation of the external and internal statically indeterminate truss and if directions of the applied load forces are parallel to the line determined by positions of two pivot bearings, then the approximate force values are almost in exact accordance with outcomes gained by usage of suitable computer software. If this method is used in the computation processes of such trusses loaded by forces applied at optional directions

then one can notice quite big differences between the force values defined in this way and the values of forces calculated by application of an exact computer method. Various possible applications of the two-stage method for calculations of different types of statically indeterminate trusses are planned to be subjects of further research in order to estimate more closely the features and the practical usefulness of this method for static computations.

**References**


[1] Timoshenko, S.P. (1966) *History of strength of materials*, Arkady, Warszawa, - in Polish.
[2] Allen E., Zalewski W. and Boston Structures Group (2010) *Form and forces. Designing efficient, expressive structures*, John Wiley & Sons, Hoboken, New Jersey.
[3] Kolendowicz, T. (1993) *Theory of structures for architects*, Arkady, Warszawa, - in Polish.
[4] Hibbeler, R.C. (1995) *Structural analysis*, Prentice Hall.
[5] Niezgodziński, M.E. and Niezgodziński, T. (1979) *Strength of materials*, State Scientific Publishing House, Warszawa, - in Polish.
[6] Przewłócki, J. and Górski, J. (2006) *Basis of theory of structures*, Arkady, Warszawa, - in Polish.
[7] Dyląg, Z. and Krzemińska-Niemiec, E. and Filip, F. (1989) *Theory of structures*, State Scientific Publishing House, Warszawa, - in Polish.
[8] Cywiński, Z. (1976) *Theory of structures in problems*. *Vol. II. The rudiments of statically indeterminate systems*, State Scientific Publishing House, Warszawa - Poznań - in Polish.
[9] Makowski, Z.S. (1981) *Analysis, design and construction of double-layer grids*, Applied Science Publishers, London.
[10] Zienkiewicz, O.C. and Taylor R.L. (2000) *The finite element method*, Oxford Press, UK.
[11] Rębielak, J. (2014) A two-stage method for an approximate calculation of statically indeterminate trusses, *Journal of Civil Engineering and Architecture*, **78**, 567-572.
[12] Rębielak, J. (2018) Simple method of approximate calculation of statically indeterminate trusses, *International Journal of Computational Methods*, Vol. 15, Issue 1, DOI: 10.1142/S0219876218400261.
[13] Rębielak, J. (2018) Examples of applications of two-stage method in calculations of statically indeterminate trusses, *International Journal of Computational Methods*, Vol. 15, No. 5, DOI: 10.1142/S0219876218440097.
[14] Tan, Z-Q., Jiang, X-D., He, Y-S., Ban, S-H., Xu, R. and Xi R-Q. (2018) Generalized variational principles for solutions of statically indeterminate trusses under mechanical-thermal-assembly loadings, *Journal of Engineering Mechanics*, Volume 144, Issue 1, 04017145:1-5.

# Application of PDS-FEM to the Simulation of Dynamic Crack Propagation and Supershear Rupture

**†Lionel Quaranta[1,2], Lalith Wijerathne[1,2], Tomoo Okinaka[3] and Muneo Hori[1,2]**

[1]Department of Civil Engineering, University of Tokyo, Japan

[2]Earthquake Research Institute, Japan

[3]Department of Civil and Environmental Engineering, Kindai University, Japan

†Corresponding author: lionel@eri.u-tokyo.ac.jp

## Abstract

Classical dynamic crack propagation problems are simulated with PDS-FEM, which uses a simple and effective particle-based approach for failure, without the need of a complex treatment of the crack. Two Hamiltonian based formulations are proposed for accurate time integration: the traditional displacement-momentum form and the strain-momentum form, for which consistent momentum conserving and symplectic time integration schemes are derived. Numerical results are verified and validated by comparing with high speed photoelastic observations of a dynamic mode-I crack captured with a 1Mfps camera. Our results show that both methods are successful in reproducing the crack patterns observed experimentally for standard 2D and 3D problems, as well as the variation of the stress distribution around the crack tip during the propagation. The two methods appear numerically indifferentiable although the displacement-based method offers a significantly better computational performance.

We also show that our approach can be applied to the simulation of super-shear rupture in earthquakes. The fault is represented by a crack under mode II loading, and the frictional contact follows a classical slip-weakening law. We could successfully reproduce both the intersonic and sub Rayleigh regimes, in good agreement with the expected rupture velocity ranges for different seismic factors.

**Keywords**: Brittle failure, dynamic crack propagation, Hamiltonian, PDS-FEM, supershear

## Introduction

Predicting the potential impact of extreme events, such as near field supershear earthquakes, on critical infrastructures is a challenging task. It involves shock waves and large-scale dynamic crack propagation simulations, which are computationally demanding, especially in three dimensions. Such simulations require a high degree of accuracy to reproduce a crack path and a crack tip speed consistent with what is observed experimentally, since both are extremely sensitive to the variations of the stress distribution in the material.

An accurate and relatively light numerical method is required to conduct large scale crack propagation simulations, like supershear rupture earthquakes. Adapted advanced discretized methods such as XFEM with level-set [1]-[3] or discontinuous Galerkin methods [4] have been widely used and refined in recent years, leading to a more accurate representation of crack nucleation and propagation. However, most of them involve high numerical overheads, especially in 3D problems, making those difficult to be adapted to large scale simulations. Most 3D studies focus their efforts on the accurate reproduction of the crack path, where the crack is often defined by either cohesive elements/interactions or level-set functions. Recently, Seagraves *et al.* [5] simulated the dynamic crack propagation caused by the impact on brittle plates, where the crack evolution as a function of time is showed to be quite consistent with the experiment. Bede *et al.* [6] conducted a similar simulation for a notched beam fracture modes.

We propose an alternative approach based on PDS-FEM [7]-[10], which uses a simple and effective particle-based approach for failure, without the need of a complex treatment of the crack. Two Hamiltonian based formulations were proposed [11] for accurate time integration:

the traditional displacement-momentum form and the rather new strain-momentum form, for which numerical accuracy and computational efficiency were not yet estimated. In this study the two methods are implemented and compared, both in terms of the accuracy in dynamic crack propagation applications and computational performance.
We also demonstrate that our method can be applied to the simulation of supershear earthquake scenarios. The existence of supershear rupture was first theoretically predicted by Burridge *et al.*[12], and although this result could be subsequently reproduced numerically, in particular by Andrews[13] and Das and Aki[14], actual observations are relatively recent. Rosakis *et al.*[15,16] experimentally reproduced a supershear rupture of Homalite-100 plates, and visualized the characteristic shock wave pattern (Mach cone) with photoelastic fringes. Although rare, several occurrences of supershear fault rupture in earthquakes have been identified, including the 1979 Imperial Valley[17], the 1999 Izmit and Duzce[18], 2001 Kunlun[19], 2002 Denali[20], and more recently the 2018 Palu Earthquake[21]. Understanding the conditions leading to supershear rupture, and the influence on ground motion close and far from the fault, is of particular importance as it will allow the evaluation of the impact on infrastructures, which may be significantly different from that of conventional earthquakes.

Section 1 presents the analytical expressions of the two Hamiltonian formulations; displacement-momentum and strain-momentum forms. Section 2 reintroduces the PDS-FEM formalism and derives the corresponding discretized Hamiltonian system for the two methods, and section 3 describes the time integration schemes. Section 4 presents the verification of the two methods with known solutions for a 1D and a 3D problem. Section 5 describes the validation of our method by comparison with photoelastic fringe patterns of a propagating crack, captured with a 1Mfps camera. Section 6 compares the two Hamiltonian formulations in term of numerical accuracy and computational efficiency. Section 7 presents some preliminary results of supershear rupture simulations using the above methods.

## 1. Two Hamiltonian formulations of elastic wave propagation

Lagrangian based formulations are often used in continuum mechanics to solve dynamic problems with FEM based numerical methods. Classically, the Lagrangian is written:

$$L(\boldsymbol{v},\boldsymbol{u}) = \int_V l(\boldsymbol{v},\boldsymbol{u})dV = \int_V \frac{1}{2}\rho \boldsymbol{v}^2 dV - \int_V \frac{1}{2}\boldsymbol{\nabla u}\colon \boldsymbol{C}\colon \boldsymbol{\nabla u}\, dV$$

According to the Hamilton's principle, the Euler-Lagrange equation can be derived as:

$$\frac{\partial \dot{l}}{\partial \boldsymbol{v}} = \frac{\partial l}{\partial \boldsymbol{u}}$$

By using a Legendre's transform:

$$l(\boldsymbol{v},\boldsymbol{u}) = \boldsymbol{p}.\boldsymbol{v} - h(\boldsymbol{p},\boldsymbol{u})$$

an alternative formulation using the Hamiltonian operator can be used based on the pair of variables momentum $\boldsymbol{p}$ and displacement $\boldsymbol{u}$, leading to the dynamic system:

$$\begin{cases} \dot{\boldsymbol{p}} = -\dfrac{\partial h}{\partial \boldsymbol{u}} \\ \dot{\boldsymbol{u}} = \dfrac{\partial h}{\partial \boldsymbol{p}} \end{cases} \quad (1)$$

where $h$ is the Hamiltonian density.

An alternative is to use the strain $\boldsymbol{\epsilon}$ as a variable instead of the displacement. By doing so, the variable of interest for most mechanical problems, especially when fracture is involved, is directly computed through the formulation. This analytically equivalent system, when used for numerical computation, would reduce the number of floating point operations to obtain the strain/stress distribution, potentially leading to an improved accuracy. Considering the Lagrangian:

$$L(\boldsymbol{v},\boldsymbol{\epsilon}) = \int_V l(\boldsymbol{v},\boldsymbol{\epsilon})dV = \int_V \frac{1}{2}\rho \boldsymbol{v}^2 dV - \int_V \frac{1}{2}\boldsymbol{\epsilon}:\boldsymbol{C}:\boldsymbol{\epsilon}\, dV,$$

we can obtain the equivalent strain based Euler-Lagrange equation:

$$\frac{\partial \dot{l}}{\partial \boldsymbol{v}} = -\nabla.\frac{\partial l}{\partial \boldsymbol{\epsilon}},$$

from which we can directly derive the Hamiltonian system (see [11] for more details of the derivation):

$$\begin{cases} \dot{\boldsymbol{p}} = \nabla.\dfrac{\partial h}{\partial \boldsymbol{\epsilon}} \\ \dot{\boldsymbol{\epsilon}} = sym\left\{\nabla \dfrac{\partial h}{\partial \boldsymbol{p}}\right\} \end{cases} \qquad (2)$$

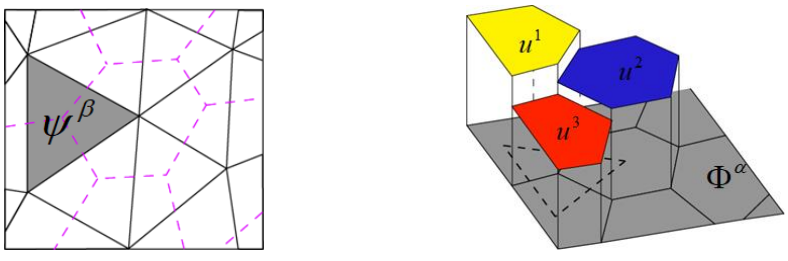
The set of equations (1) and (2) are the strong forms of the $(\boldsymbol{p},\boldsymbol{u})$ and $(\boldsymbol{p},\boldsymbol{\epsilon})$ based Hamiltonian formulations.

To numerically solve the above two Hamiltonian systems, we derive a discretized system in space and time, which is detailed in the two next sections.

## 2. Space discretization in PDS-FEM

### *2.1 Model and consistency condition*

For the spatial discretization, we use a model of FEM based on Particle Discretization Scheme (PDS) as proposed in [7], due to its advantages in modeling cracks. PDS uses conjugate tessellations to approximate functions and their derivatives respectively. The Delaunay tessellation (indexed with $\beta$) is defined by a tetrahedron mesh, and the conjugate Voronoi tessellation (indexed with $\alpha$), is constructed as indicated schematically in Fig. 1. In 2D, a Voronoi element or "particle" is essentially derived from joining the mass centers of the neighboring triangle elements.




(a) Delaunay tessellation (b) Voronoi tessellation

**Figure 1: Conjugate tessellations in 2D**

In the model formulation, non-overlapping functions are used to approximate the functions and their derivatives respectively:

$$\boldsymbol{f}(\boldsymbol{x}) \approx \boldsymbol{f^d}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha=1}}^{\boldsymbol{N^\alpha}} \sum_{\boldsymbol{n=0}}^{\boldsymbol{|P^\alpha|}} \boldsymbol{f^{\alpha n}} P^{\alpha n}(\boldsymbol{x})\, \varphi^{\alpha}(\boldsymbol{x})$$

$$\boldsymbol{\nabla_{,i} f}(\boldsymbol{x}) \approx \boldsymbol{g^d}(\boldsymbol{x}) = \sum_{\boldsymbol{\beta=1}}^{\boldsymbol{N^\beta}} \sum_{\boldsymbol{m=0}}^{\boldsymbol{|Q^\beta|}} \boldsymbol{g_i^{\beta m}} Q^{\beta m}(\boldsymbol{x})\, \psi^{\beta}(\boldsymbol{x}),$$

where $\varphi^{\alpha}$ and $\psi^{\beta}$ are respectively the characteristic functions of Voronoi and Delaunay elements. The displacement field is discretized on the Voronoi tessellation while the strain and stress distribution are discretized on the Delaunay tessellation.

PDS obtains bounded and consistent approximations for derivatives, $\boldsymbol{g}^d$, over the Delaunay tessellations by minimizing the following measure of error:

$$\min_{g^{\beta m}} \int_{\Psi^\beta} (\boldsymbol{g^d} - \nabla \boldsymbol{f^d})^2 dV$$

which leads to:

$$\boldsymbol{g}^{\beta m} = \boldsymbol{B}^{\beta \alpha m n} \otimes \boldsymbol{f}^{\alpha n}$$

With

$$B_i^{\beta \alpha m n} = \sum_{m'} \left(I^{\beta -1}\right)^{m,m'} \int_V \frac{\partial (P^{\alpha n} \phi^{\alpha})}{\partial x_i} Q^{\beta m} \psi^{\beta} dV$$

$$I^{\beta m m'} = \int_{\Psi^\beta} Q^{\beta m} Q^{\beta m'} dV$$

The complete derivation of the vector $\boldsymbol{B}^{\beta \alpha m n}$ is provided in [10]. Then, applying it to the strain-displacement relationship, the consistency condition is written:

$$\boldsymbol{\epsilon}^{\beta m} = sym\{\boldsymbol{B}^{\beta \alpha m n} \otimes \boldsymbol{u}^{\alpha n}\} \qquad (3)$$

For the ease of computer implementation, we use the Voigt notation and express the above in matrix notation as follows:

$$\epsilon^{\beta m} = \tilde{B}^{\beta \alpha m n} . u^{\alpha n} \qquad (4)$$

*2.2 Discretized Lagrangian and Hamiltonian forms*

In this model, a discretized Lagrangian can be written $L(v^{\alpha n}, u^{\alpha n})$ or $L(v^{\alpha n}, \epsilon^{\beta m})$ depending on the formulation, and the corresponding Euler-Lagrange equations can be written as:

$$\frac{d}{dt}\frac{\partial L}{\partial v^{\alpha n}} = \frac{\partial L}{\partial u^{\alpha n}} \qquad \text{and} \qquad \frac{d}{dt}\frac{\partial L}{\partial v^{\alpha n}} - \sum_{\beta, m} {}^{t}\tilde{B}^{\beta \alpha m n} . \frac{\partial L}{\partial \epsilon^{\beta m}} = 0 \qquad (5)$$

By applying the Legendre transform:

$$L(v^{\alpha n}, u^{\alpha n}) = p^{\alpha n} . v^{\alpha n} - H(p^{\alpha n}, u^{\alpha n}) \qquad \text{or} \qquad L\left(v^{\alpha n}, \epsilon^{\beta m}\right) = p^{\alpha n} . v^{\alpha n} - H(p^{\alpha n}, \epsilon^{\beta m})$$

with

$$p^{\alpha n} = \frac{\partial L}{\partial v^{\alpha n}} = \rho \sum_{n'=0}^{|P^\alpha|} v^{\alpha n'} I^{\alpha n n'}, \qquad I^{\alpha n n'} = \int_{\Phi^\alpha} P^{\alpha n} P^{\alpha n'} dV, \tag{6}$$

we can obtain the following two equivalent Hamiltonian systems:

$$\begin{cases} \dot{p}^{\alpha n} = -\dfrac{\partial H}{\partial u^{\alpha n}} \\ \dot{u}^{\alpha n} = \dfrac{\partial H}{\partial p^{\alpha n}} \end{cases} \tag{7}$$

$$\begin{cases} \dot{p}^{\alpha n} = -\displaystyle\sum_{\beta}^{N^\beta} \sum_{m}^{|Q^\beta|} {}^t\tilde{B}^{\beta \alpha m n} . \frac{\partial H}{\partial \epsilon^{\beta m}} \\ \dot{\epsilon}^{\beta m} = \displaystyle\sum_{\alpha}^{N^\alpha} \sum_{n}^{|P^\alpha|} \tilde{B}^{\beta \alpha m n} . \frac{\partial H}{\partial p^{\alpha n}} \end{cases} \tag{8}$$

The derivation for the strain formulation is detailed in Appendix A.

### 2.3 Discretized Hamiltonian for continuum mechanics

Using the classical form of the Lagrangian in continuum mechanics (excluding for now the boundary conditions whose treatment is detailed in appendix B) and applying the Legendre's transform, we can write explicitly the Hamiltonian as:

$$H(p^{\alpha n}, u^{\alpha n}) = T(p^{\alpha n}) + U(u^{\alpha n})$$

Where, using Einstein's summation convention:

$$\begin{cases} T(p^{\alpha n}) = \dfrac{1}{2\rho} W^{\alpha n n'} \, {}^t p^{\alpha n} . p^{\alpha n'} \\ U(u^{\alpha n}) = \dfrac{1}{2} I^{\beta m m'} \, {}^t u^{\alpha n} . \, {}^t\tilde{B}^{\beta \alpha m n} . c^\beta . \tilde{B}^{\beta \alpha' m' n'} . u^{\alpha' n'} \end{cases}$$

with

$$W^\alpha = I^{\alpha^{-1}}$$

Similarly, the Hamiltonian can be written in term of strain:

$$H\left(p^{\alpha n}, \epsilon^{\beta m}\right) = T(p^{\alpha n}) + U(\epsilon^{\beta m})$$

where

$$U\left(\epsilon^{\beta m}\right) = \frac{1}{2} I^{\beta m m'} \, {}^t\epsilon^{\beta m} . c^\beta . \epsilon^{\beta m'}$$

We can then replace the Hamiltonian derivatives in the Eq. 7 and 8 to obtain the Hamiltonian systems in displacement-momentum form as:

$$\begin{cases} \dot{p}^{\alpha n} = -\displaystyle\sum_{\alpha'}^{N^\alpha} \sum_{n'}^{|P^\alpha|} K^{\alpha n \alpha' n'} . u^{\alpha' n'} \\ \dot{u}^{\alpha n} = \dfrac{1}{\rho} \displaystyle\sum_{n'}^{|P^\alpha|} W^{\alpha n n'} . p^{\alpha n'} \end{cases}, \tag{9}$$

where

$$K^{\alpha n \alpha' n'} = \sum_{\beta}^{N^\beta} \sum_{m,m'}^{|Q^\beta|} I^{\beta m m'} \; {}^t\tilde{B}^{\beta \alpha m n} . c^\beta . \tilde{B}^{\beta \alpha' m' n'},$$

and in strain-momentum form as:

$$\begin{cases} \dot{p}^{\alpha n} = -\sum_{\beta}^{N^\beta} \sum_{m,m'}^{|Q^\beta|} I^{\beta m m'} \; {}^t\tilde{B}^{\beta \alpha m n} . c^\beta . \epsilon^{\beta m'} \\ \dot{\epsilon}^{\beta m} = \sum_{\alpha}^{N^\alpha} \sum_{n,n'}^{|P^\alpha|} \frac{1}{\rho} W^{\alpha n n'} \tilde{B}^{\beta \alpha m n} . p^{\alpha n'} \end{cases} . \qquad (10)$$

*2.4 Crack treatment*

The above derived system of governing equations is valid for time invariant topologies, and some adaptations are necessary to simulate dynamic crack propagation. As shown in Fig. 1b, the displacement approximated with PDS is discontinuous at the Voronoi boundaries, $\partial\Phi^\alpha$. PDS uses these discontinuities to model a crack propagating along a Voronoi boundary by nullifying appropriate mechanical contributions. In the 3D case, each tetrahedral Delaunay element contains 6 Voronoi surfaces along which the cracks can appear (see Fig. 2).

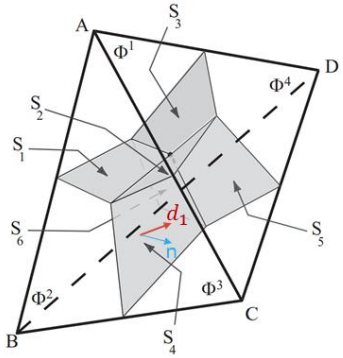



**Figure 2: Surfaces of interaction**

If the maximum principle stress is greater than a given value $\sigma_s$, the surface whose normal $n$ is closer to the principal stress direction $d_1$ (corresponding to the surface with highest tensile stress) is broken, nullifying its contribution to the strain $\epsilon^\beta$, by modifying the coefficient $B_i^{\beta \alpha m n}$:

$$\int_V \frac{\partial (P^{\alpha n} \phi^\alpha)}{\partial x_i} Q^{\beta m} \psi^\beta dV$$

$$= \int_{\Psi^\beta \cap \Phi^\alpha} \frac{\partial P^{\alpha n}}{\partial x_i} Q^{\beta m} dV + \int_{\Psi^\beta \cap \partial\Phi^\alpha - s} P^{\alpha n} Q^{\beta m} n_i \, dV + \underbrace{\int_s P^{\alpha n} Q^{\beta m} n_i \, dV}_{\to 0}$$

The first right-hand term corresponds to the volume component of the gradient (null for rigid particles with constant displacement, unchanged by the crack). The second and third right-hand terms correspond to the surface component coming from the interaction between the particles, for the unbroken and broken surfaces respectively. The latter contribution is set to zero.

For the sake of simplicity, the maximum principal stress criterion is used for the simulations presented in this paper. However, at the crack tip, larger elements have a lower average stress value and so are less prone to crack, making this criterion mesh dependent, although it doesn't affect the crack path as long as the crack prone region is homogeneously refined.

## 3. Time integration schemes

Out of the large collection of algorithms for the time integration of Hamiltonian systems, we chose to use the Störmer-Verlet integration scheme, which is a second order variational integrator (see for instance [22] section VI for reference):

$$\begin{cases} p_{n+1/2} = p_n - \dfrac{dt}{2}\dfrac{\partial H}{\partial u}(p_{n+1/2}, u_n) \\ u_{n+1} = u_n + \dfrac{dt}{2}\left(\dfrac{\partial H}{\partial p}(p_{n+1/2}, u_n) + \dfrac{\partial H}{\partial p}(p_{n+1/2}, u_{n+1})\right) \\ p_{n+1} = p_{n+1/2} - \dfrac{dt}{2}\dfrac{\partial H}{\partial u}(p_{n+1/2}, u_{n+1}) \end{cases}$$

In particular, the symplectic property of the flow $(p, u)$ in time is ensured, which can be written:

$$\left(\frac{\partial\,(p_{n+1}, u_{n+1})}{\partial\,(p_0, u_0)}\right)^T J \left(\frac{\partial\,(p_{n+1}, u_{n+1})}{\partial\,(p_0, u_0)}\right) = J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

where *I* is the identity matrix.

For the strain formulation, such symplectic property cannot be defined directly as the strain (dimension 6) and the momentum (dimension 3) don't belong to the same vector spaces. However, it is possible to express the above symplectic property in term of the strain and momentum variables. Using Eq. 4, we can write:

$$\frac{\partial \epsilon}{\partial u'} = \tilde{B} \qquad \text{and} \qquad \frac{\partial u'}{\partial \epsilon} = \tilde{B}^{-1},$$

with $u'$ the displacement field excluding the Dirichlet boundary conditions and $\tilde{B}$ the corresponding reduced matrix built with the blocks $\tilde{B}^{\beta\alpha}$. $\tilde{B}^{-1}$ is the left inverse of $\tilde{B}$. If $\phi$ is the flow of the Hamiltonian system in $(p, u)$ and $\psi$ the flow of the Hamiltonian system in $(p, \epsilon)$, we can write

$$\frac{\partial \phi}{\partial (p_0, u_0)} = \underbrace{\begin{pmatrix} I_{3N^\alpha} & 0 \\ 0 & \tilde{B}^{-1} \end{pmatrix}}_{A^{-1}} \frac{\partial \psi}{\partial (p_0, \epsilon_0)} \underbrace{\begin{pmatrix} I_{3N^\alpha} & 0 \\ 0 & \tilde{B} \end{pmatrix}}_{A}.$$

Therefore, we can express the symplectic property from the flow $\psi$:

$$\left(A^{-1}\frac{\partial \psi}{\partial (p_0, \epsilon_0)}A\right) J \left(A^{-1}\frac{\partial \psi}{\partial (p_0, \epsilon_0)}A\right) = J$$

Following this property, we can derive an algorithm equivalent to the classical Störmer-Vertlet scheme for $(p, \epsilon)$ (Einstein convention is used for the summation):

$$\begin{cases} p_{n+1/2} = p_n - \dfrac{dt}{2}\ {}^t\tilde{B}.\dfrac{\partial H}{\partial \epsilon}(p_{n+1/2}, \epsilon_n) \\ \epsilon_{n+1} = \epsilon_n + \dfrac{dt}{2}\tilde{B}.\left(\dfrac{\partial H}{\partial p}(p_{n+1/2}, \epsilon_n) + \dfrac{\partial H}{\partial p^\alpha}(p_{n+1/2}, \epsilon_{n+1})\right) \\ p_{n+1} = p_{n+1/2} - \dfrac{dt}{2}\ {}^t\tilde{B}\dfrac{\partial H}{\partial \epsilon}(p_{n+1/2}, \epsilon_{n+1}) \end{cases}$$

By ensuring the conservation of the momentum, the average conservation of the total energy, and the symplecticity of the two systems, these time integration schemes can maintain accuracy and stability in long-time integration.

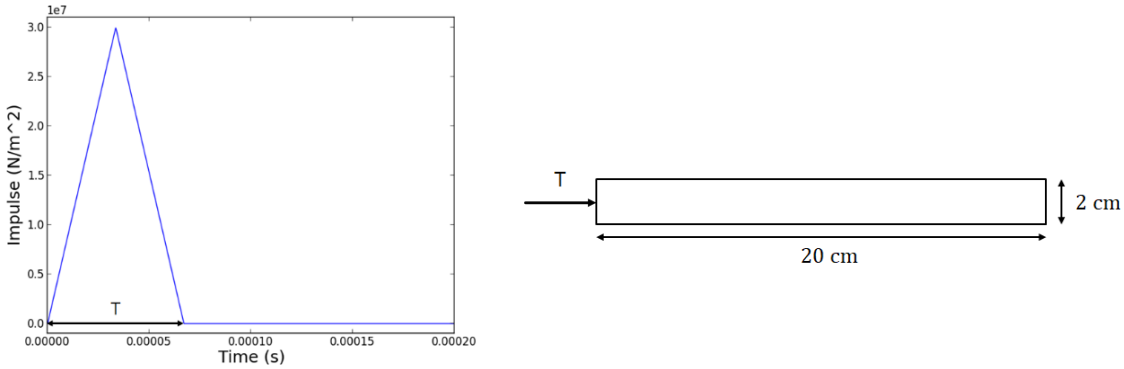
Since our Hamiltonians are separable, the integration scheme computation becomes explicit (see Eq. 9 and 10). In particular, the time integration scheme can be written in the matrix form as:

$$\begin{cases} \boldsymbol{p_{n+1/2} = p_n - \frac{dt}{2} K_1 q_n} \\ \boldsymbol{q_{n+1} = q_n + dt\, K_2 p_{n+1/2}} \\ \boldsymbol{p_{n+1} = p_{n+1/2} - \frac{dt}{2} K_1 q_{n+1}} \end{cases}$$

where $\boldsymbol{q_n}$ represents either $\boldsymbol{u_n}$ or $\boldsymbol{\epsilon_n}$, and $\boldsymbol{K_1}$, $\boldsymbol{K_2}$ are two matrices derived from the expanded Hamiltonian form.

## 4. Verification of dynamic crack propagation simulations

In order to verify the application of above schemes to simulate dynamic crack propagation simulations, we compare the results of the simulation with the known solutions of a 1D spallation problem and a 3D torsion problem.




(a) Impulse shape (b) Geometrical setting

**Figure 3: Problem settings for the spallation test**

### *4.1 1D spallation problem*

One mode of material failure during high velocity impact is spallation. As an example, an impact on one end of a bar creates a compressive wave which is then inverted and reflected at the opposite end, resulting in a high tensile stress wave likely to generate a crack. In this verification test, we consider the problem settings shown in Fig. 3, and compare the position of the crack and the time of occurrence with the corresponding analytical solution. We chose a simple stress criterion with the strength $\sigma_c = 29.5$MPa, slightly lower than the amplitude of the input wave as we observe a small loss of amplitude with the propagation. Obviously the crack opens when the tensile stress, propagating backwards after reflection on the right tip, exceeds this limit value. It can be shown that the maximum tensile stress occurs at $t = L/c + 3T/4 = 1.65\ 10^{-4}s$ and $x = L - cT/4 = 0.171\ m$ where $L = 0.2m$ is the length of the bar, $c = 1.75\ 10^{-3}m.s^{-1}$ the wave velocity and $T = 6.7\ 10^{-5}s$ the impulse duration. The numerical simulations produced the crack at the above theoretically expected time and location (see Fig. 4), with reasonably small errors of $\pm\ 5.0 \times 10^{-7}s$ and $-\ 1.0 \times 10^{-3}$m, respectively.

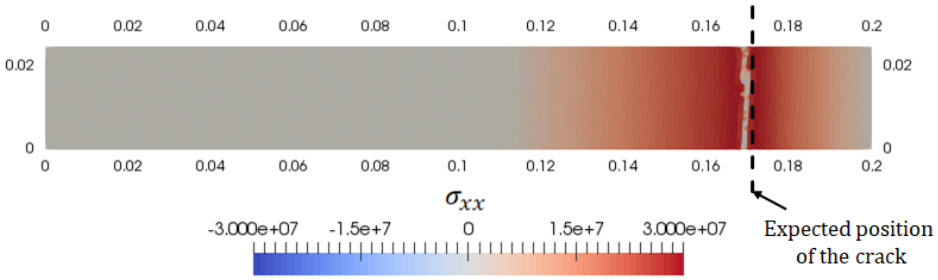



**Figure 4: Simulated crack position and stress distribution for spallation**

*4.2 Torsional cracks in a cylindrical specimen*

To verify that the above schemes can reproduce the 3D geometry of dynamically growing cracks, we consider the classical experiment of torsion of a cylinder bar with the standard dimensions (in mm, see Fig. 5) and the following material properties:

- Young Modulus: 3.0 GPa
- Tensile strength: 30 MPa
- Density: 1.25 kg.m$^{-3}$

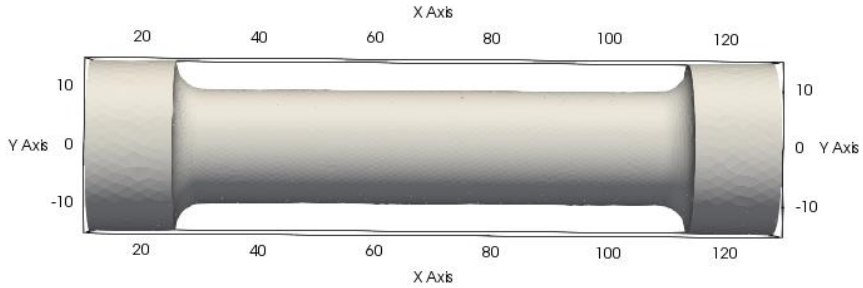



**Figure 5: Geometrical setting for the torsion simulation**

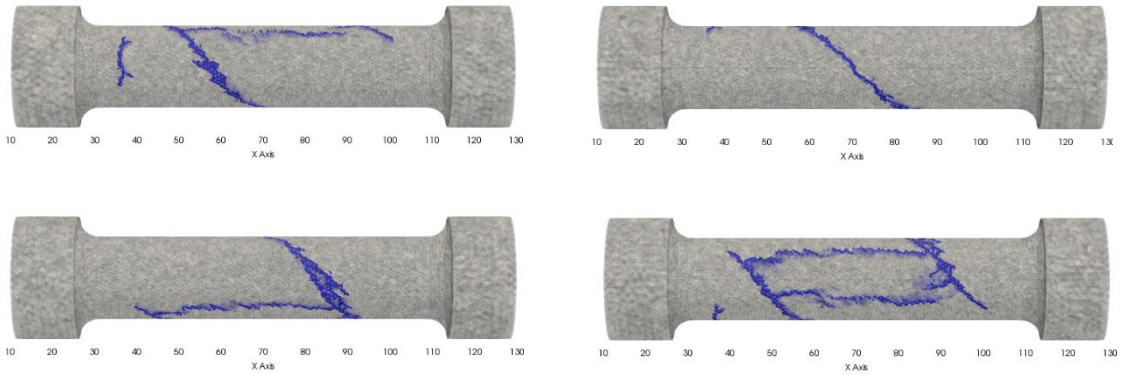


**Figure 6: Trace of the crack on the surface from 4 different angles**

The right end of the bar ($x \geq 115mm$) is fixed at all time (i.e. $p = 0$), while the left end of the bar ($x \leq 25m$) is rotated at the angular speed $\dot{\theta} = 3° \, s^{-1}$. In order to save computation time, we first solve the static problem for a rotation of angle $\theta = 6.3°$, which brings the principal stress on the surface very close to the tensile strength, and then proceed to the dynamic loading. In this problem, anywhere on the surface of the cylinder is primed to crack, as the shear stress is uniformly distributed. So as to control where the crack will initiate, in particular to avoid having a crack propagating near the two extremities, we generate a small initial crack on the surface near the middle of the bar. The trace of the crack on the surface after completed fracture is shown in Fig. 6.

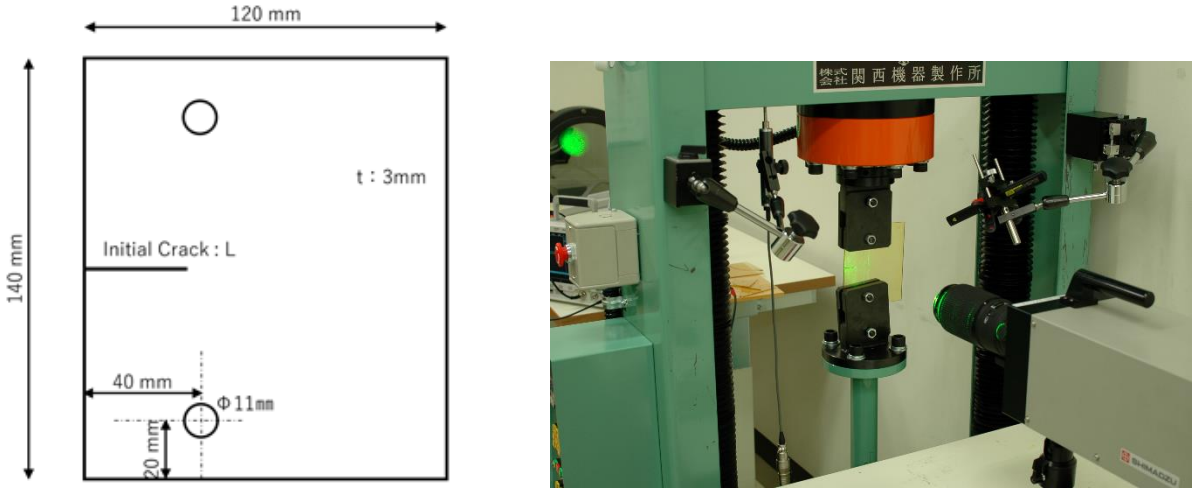
Though we lack experimental observations of the crack propagation speed, we observe that the crack begin to open at the expected load, and that the total crack path corresponds to the classical spiral shape oriented at 45° observed during torsion test, albeit a few branches. Additional experimental measures of secondary crack patterns inside ruptured material would be needed to verify this observation.

## 5. Validation

### *5.1 Experimental setting*

The PDS-FEM implementations for dynamic crack propagation simulations are validated by comparing with high speed photoelastic observations of a dynamic mode-I crack captured with a 1Mfps camera. These photoelastic fringes characterize the stress distribution in the material, specifically the difference between the two principal stresses. The experimental setting is shown in Fig. 7, and the Epoxy resin (CY232 & HY951) sample has the following properties:

- Young modulus: 3.84 GPa
- Poisson ratio: 0.3
- Static photoelastic constant: 0.53 mm/N (measured by calibration with a Brazilian test)




(a) Geometry of the sample (b) Experimental setting without analyzer

**Figure 7: Experimental setting**

Two bars are inserted in the holes of the plate, and then respectively pulled up and down. The pulling phase is displacement controlled, and the corresponding applied force loading is also measured. In the experiment conducted, the initial crack was 1 mm wide, and 20.5 mm long. The displacement is increased at a speed of 0.5 mm/min until the onset of crack propagation. Compared to the few hundred μs period required for the crack to move across the sample, our external loading can be considered quasi-static and the positions of the bars fixed during the duration of the propagation. As the initial displacement required to bring the bars and the holes just in contact cannot be accurately estimated from the available experimental observations, we use the equivalent force load in our simulations. In the experiment, the crack initiated at 0.54 kN. The same load is used to calculate the static solution before the crack initiation, which will be used as the initial condition of the dynamic problem. The stress criterion for rupture is chosen to be the maximum value for which the crack propagates in the simulation (corresponding to 40 MPa for tetrahedral elements of average size 0.2 mm). We used $0^{th}$ order PDS-FEM discretization, which corresponds to constant values of the variables in each element.

*5.2 Numerical model*

In order to accurately reproduce the crack tip stress field, we refined the mesh as follows (also, see Fig. 8):

- Most outer elements ($|y - 0.07| > 0.01$ or $z > 0.045$): average size 1 mm.
- Intermediate layer ($0.005 < |y-0.07| < 0.01$): average size 0.5 mm, refinement required for the observation of the fringe patterns, but not very fast variation of stress
- Inner band ($|y - 0.07| < 0.005$): average size of 0.2 mm, refinement required to capture well the stress distribution around the crack tip, all along its propagation.

In total, the mesh consists of around $3\times10^6$ elements and $5\times10^5$ nodes. 288 MPI processes are used for the computation.

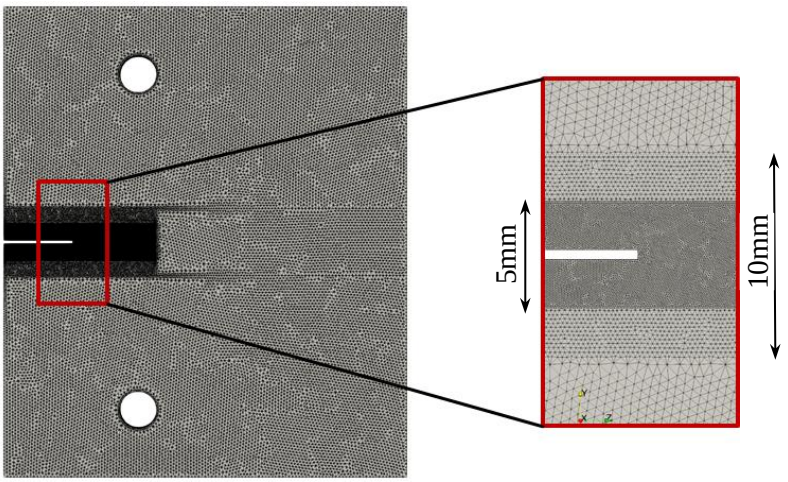



**Figure 8: Mesh refinement**

*5.3 Observations*

*Crack path*

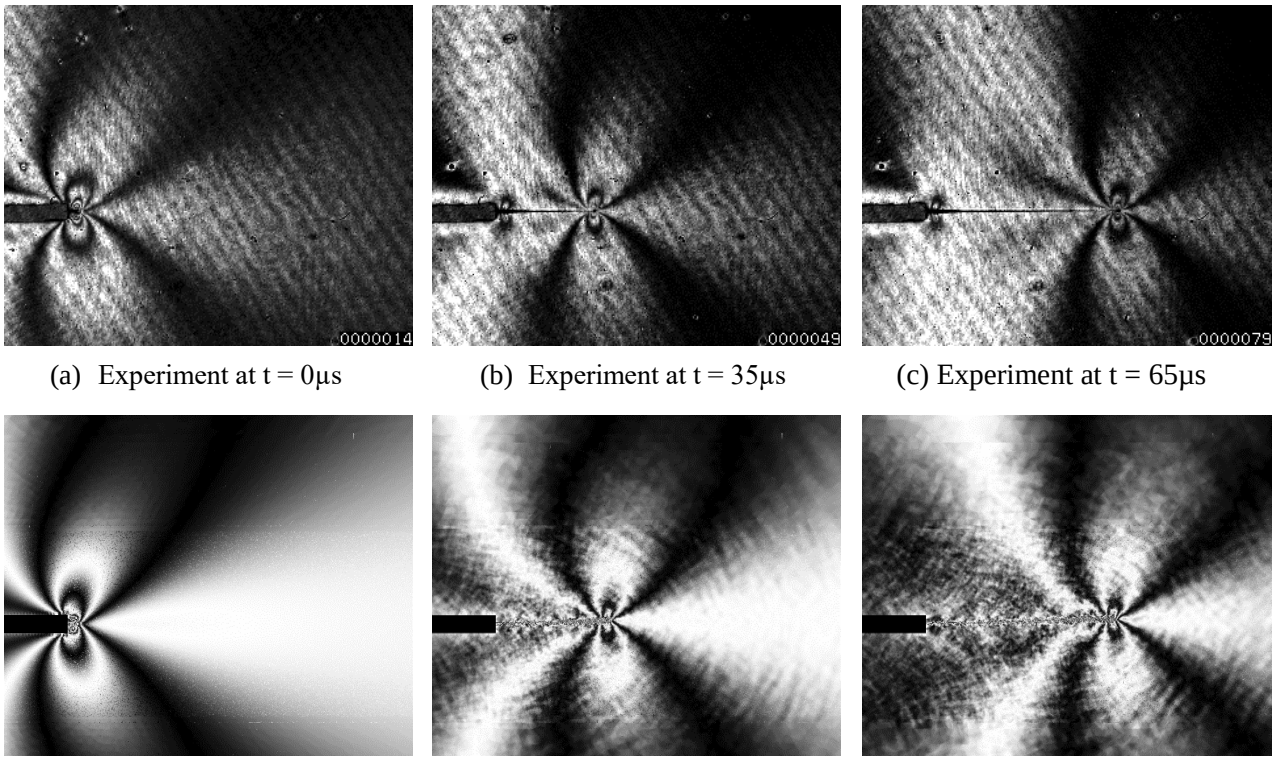
The horizontal crack path observed in the experiment is correctly reproduced in the simulation.

*Stress distribution*

The propagation of the crack and the fringe patterns are captured with a high speed camera, with a capturing speed of 1Mfps. Figure 9 compares the observed fringe patterns with those corresponding to the crack propagation simulation; $(p, \epsilon)$ form is used.

Fringe patterns, which characterize the difference of principal stresses in the material, are very sensitive to the change of stress distribution. In this experiment, each fringe corresponds to an increment of 3.95 MPa of the principal stresses difference. As seen in Fig. 9, the numerical results are in good agreement with those observed in the experiment. In particular, the amplitude of the main fringes decreases the same amount (the fringes get closer to the crack tip) when the propagation begins.

In the simulation results, we can clearly observe a wave emanating from the moving crack, which is completely absent in the experimental observations. This is most likely due to our failure criterion. As the stress is considered constant inside an element, a surface rupture releases a relatively high amount of energy (corresponding to the strain energy of the entire element), creating a source point for a stress wave, whose amplitude might decrease with a more refined mesh (for which the rupture is more "continuous"). This wave could also be smoothened if higher order functions are considered and the stress release limited to the zone




(a) Experiment at t = 0μs (b) Experiment at t = 35μs (c) Experiment at t = 65μs

(d) Simulation at t = 0μs (e) Simulation at t = 35μs (f) Simulation at t = 65μs

**Figure 9: Comparison of the fringe patterns at three stages; (a)-(c) are experimental while (d)-(f) are numerical. Each fringe corresponds to an increase of 3.95 MPa**

surrounding the surface. Also, the camera recording might have a smoothing effect as the assumed wave propagates between two frames and only its average value can be recorded.

*Crack propagation speed*

For $t \leq 20\ \mu s$, the propagation speed in the simulation is almost twice faster than in the experiment. For $t > 20\ \mu s$, the crack tip position in the simulation is about $5\ \mu s$ ahead of the crack tip position measured in the experiment until $t = 60\ \mu s$, after which the position of the crack in both the simulation and the experiment are in fairly good agreement, showing that the average speed of the simulated crack is consistent with the experiment. The record of crack tip position is detailed in table 1.

**Table 1: Comparison of the crack tip position**

| Time [μs] | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|---|
| Experiment [mm] | 0.5 | 1.6 | 3.0 | 4.6 | 6.2 | 7.8 | 9.6 | 11.3 | 13.5 |
| Simulation [mm] | 0.6 | 3.3 | 4.9 | 5.9 | 7.0 | 8.6 | 9.6 | 11.3 | 13.3 |

## 6. Comparison of the two formulations

Although the strain-based formulation and the displacement-based formulation presented in sections 2 and 3 are analytically equivalent, they lead to different numerical schemes. Therefore, they may produce stress and displacement distributions with different degrees of accuracy. In particular, in the strain formulation, the strain is computed directly by the dynamic system, though additional calculations are required to extract the displacement distribution. Therefore,

it is expected that the strain formulation would be more accurate for the stress distribution while the displacement formulation would be more accurate for the displacement distribution.

In this section, we compare quantitatively computational efficiency and numerical accuracy of these two formulations. For this comparison, we use the dynamic crack propagation problem presented in the section 5. Since this problem is very sensitive to changes in the stress/displacement distribution which could particularly alter the direction and the speed of the crack, it is an ideal problem to identify potential numerical differences. Figure 10 shows the maximal magnitude of the difference between stress distributions in the whole sample along time.

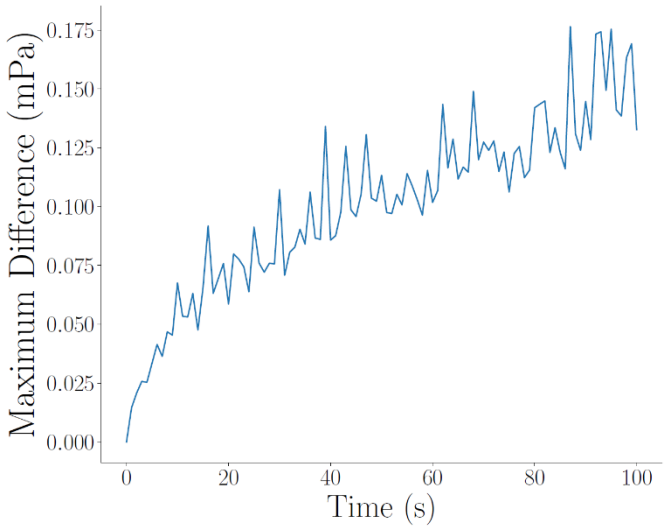



**Figure 10: Comparison of the stress distribution between the two formulations**

We observe that the maximum difference, although increasing with time, remains very insignificant compared to the relevant level of stress in the sample (order of a few MPa). Similar results can be observed for the displacement formulation although we don't show it in this article. Therefore, on a numerical accuracy perspective, both formulations can be considered equivalent.

On a computational perspective, however, there is a significant difference. The main computational task in the respective dynamic systems are the matrix-vector multiplications. However, as the mesh contains significantly more tetrahedron elements than nodes, and that strain vectors have at least 6 components while displacement only have 3 components, it is clear that strain formulation involves much more floating point operation in the main computation. Also, additional overhead is required to compute the strain change due to the rupture of an element in the strain formulation. The memory usage and average computation time for a single iteration are compared in Table 2. In order to avoid additional overheads related to parallel computing, like communication time, the compared values are obtained from one of the MPI process (with $2 \times 10^4$ elments and $4 \times 10^3$ nodes), out of the 288 MPI processes used for the computation.

**Table 2: Comparison of computational performance**

| | Disp formulation | Strain Formulation |
|---|---|---|
| Single process memory usage (MB) | 6.1 | 14.8 (+35.3 crack strain overhead) |
| Single iteration simulation time (s) | 0.0304 | 0.0857 |

We observe that the strain formulation is significantly more computationally demanding, both in memory usage and computation time. Further, as shown above, both formulations essentially lead to a near identical numerical accuracy. Therefore, we conclude that it is more relevant to use the displacement formulation in general.

## 7. Application to supershear rupture simulation

### *7.1 Theoretical considerations*

In this section, we apply the developed numerical method to the simulation of an idealized supershear rupture in earthquakes. This part should be understood as an illustration of the potential capabilities of this method rather than a full study of supershear rupture.

Supershear rupture is defined as a crack propagation faster than the S-wave velocity, producing a characteristic shear wave Mach cone following the crack tip. The possible occurrence of this kind of rupture was first theoretically predicted by Burridge *et al.*[12], who identified three steady crack propagation regimes:

- $v_r < c_R$: sub Rayleigh regime
- $c_s < v_r < \sqrt{2}c_s$: unstable intersonic (supershear) rupture
- $\sqrt{2}c_s < v_r < c_P$: stable intersonic (supershear) rupture

The zone $c_R < v_r < c_s$ is forbidden as it leads to a negative fracture energy.

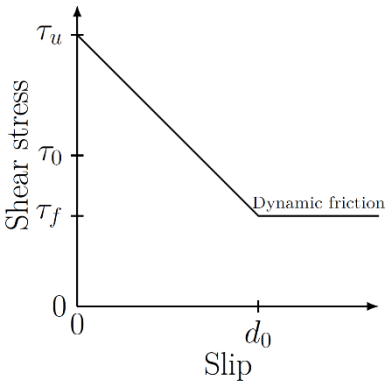



**Figure 11: Linear Slip-weakening law**

In the case of earthquakes, fault "rupture" corresponds to the part of the fault where slipping occurs, where there is a discontinuity of displacement across the fault. For the slipping, the linear slip-weakening friction law used by Andrews [13], shown in Fig. 11, is a good first approximation of the earthquake slipping law, and has been widely used for the simulation of supershear rupture. The initial stress in the material is noted $\tau_0$. When a perturbation makes the stress exceed the maximum cohesive strength $\tau_u$ (static frictional stress), slipping starts and the maximum frictional stress decreases linearly until a minimum $\tau_f$ corresponding to the dynamic frictional stress.

Although diverse factors, like fault roughness, local material heterogeneities, asperities etc., can influence the occurrence of supershear rupture, for the ideal scenario of a perfectly straight plane fault in an otherwise homogenous material, the seismic factor, introduced by Das and Aki [14]:

$$s = \frac{\tau_u - \tau_0}{\tau_0 - \tau_f}$$

characterizes the crack propagation regimes. In particular, for a 2D problem, supershear propagation is theoretically possible only for $s < 1.77$ [13]. In our simulation, we don't

observe supershear rupture for $s \geq 1.1$, which might be partly due to wave defocusing in 3D plates compared to an actual 2D setting.

*7.2 Numerical problem setting*

We simulate the rupture in a crust layer of dimensions $10\ km \times 4\ km\ \times\ 0.1\ km$, centered around a fault plane extending all along the layer (see Fig. 12). The parameters for the slip weakening law are chosen arbitrarily, which is sufficient for this qualitative study, as they influence only the slip-weakening zone length and the seismic factor. We use the following material properties corresponding to the Earth crust:

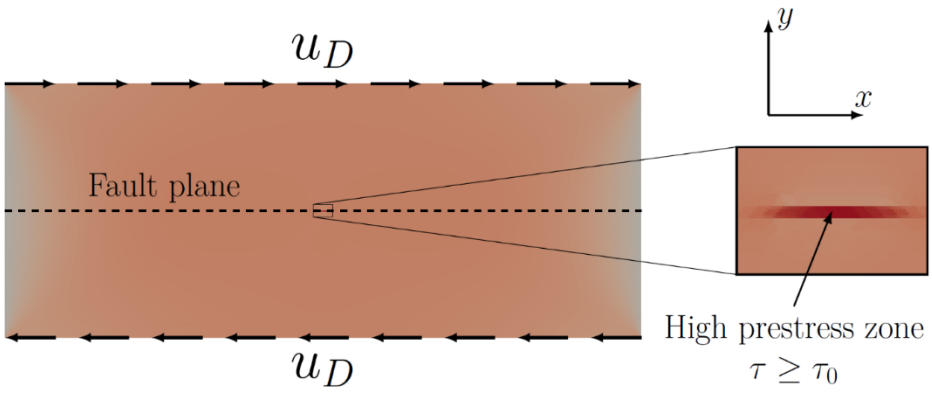



**Figure 12: Supershear problem setting**

- Young modulus: 75 GPa
- Poisson ratio: 0.3
- Density: 2.5 $10^3$ kg.m$^{-3}$
- $\tau_u = 100MPa$, $\tau_f = 0MPa$
- $d_0 = 0.2\ m$
- $\tau_0 = 44\ MPa$ $(s = 1.27)$ for sub-Rayleigh, $\tau_0 = 58\ MPa$ $(s = 0.7)$ for supershear

The corresponding P-wave and S-wave velocities are, respectively, $c_p = 5477m.s^{-1}$ and $c_S = 3162m.s^{-1}$.

At $t = 0$, the fault is perfectly cohesive and the stress/displacement is continuous across the fault (no broken element on the fault). We solve a static problem with Dirichlet boundary conditions to obtain the initial stress distribution. The fault parallel component $u_D$ is fixed and the two other components are set to 0 on the Dirichlet boundaries.

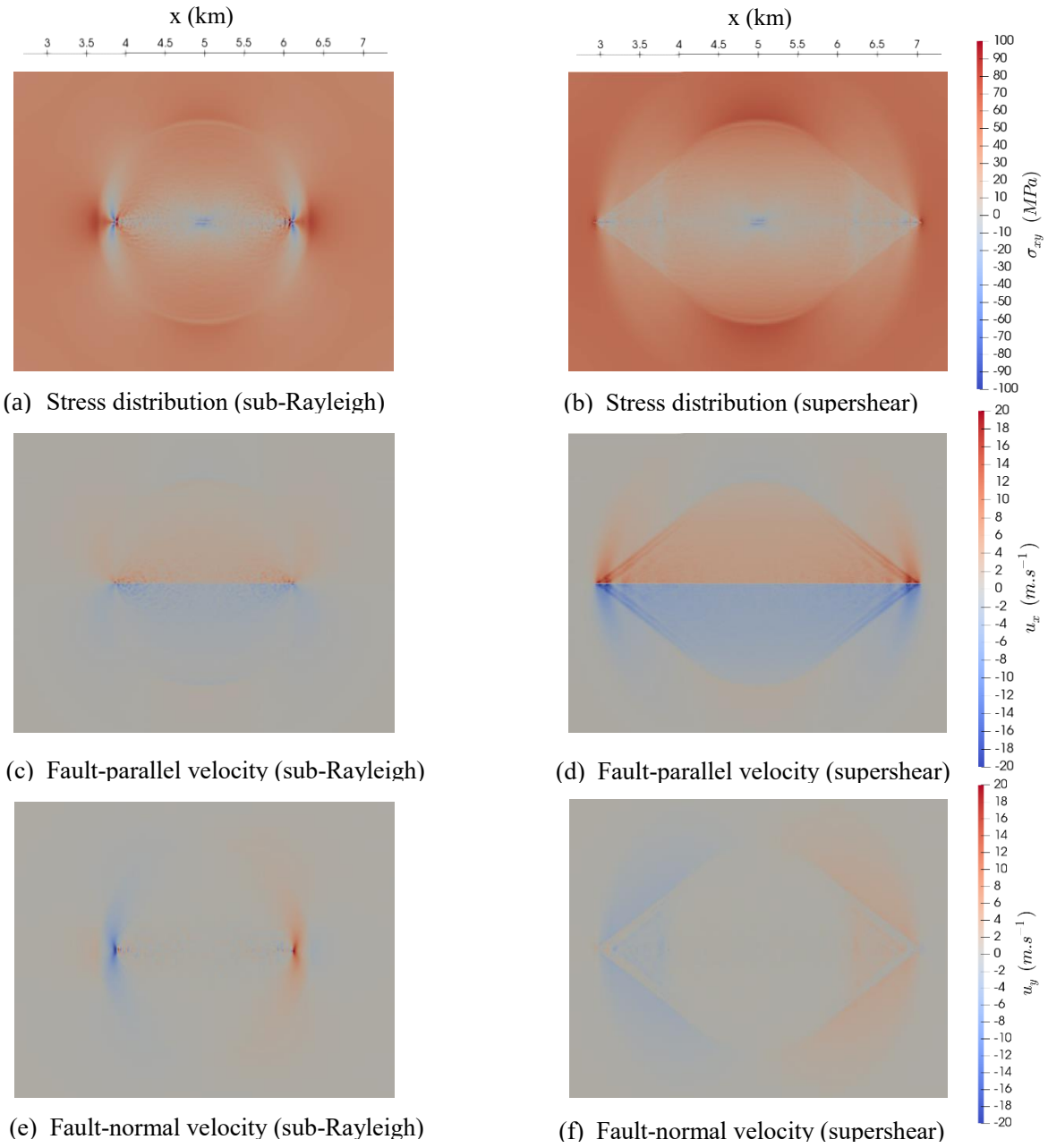
At $t > 0$, if the stress exceeds the maximum shear stress in an element, the inner fault surface is broken and frictional forces are mobilized and distributed on the nodes on the fault surfaces, according to the slip-weakening friction law.

*7.3 Results and observations*

Stress and velocity distributions at $t = 0.4s$ on the surface for sub-Rayleigh and supershear regimes are shown in Fig. 13.

In the sub-Rayleigh scenario, we observe the classical wave pattern, with a clear S-wave front progressing ahead of the crack which propagates at $c_R$. The P-wave front is however not clearly observable. In the supershear scenario, we can clearly observe the P-wave front which is just ahead of the crack which propagates at a speed close to $c_P$. The Mach cone, characteristic of a shock wave propagation, is also clearly visible. Fault parallel velocity is significantly higher in

the supershear case, especially far from the fault. On the contrary, fault-normal velocity is higher for the sub-Rayleigh case near the crack tip, and dissipates rapidly far from the fault.




(a) Stress distribution (sub-Rayleigh)

(b) Stress distribution (supershear)

(c) Fault-parallel velocity (sub-Rayleigh)

(d) Fault-parallel velocity (supershear)

(e) Fault-normal velocity (sub-Rayleigh)

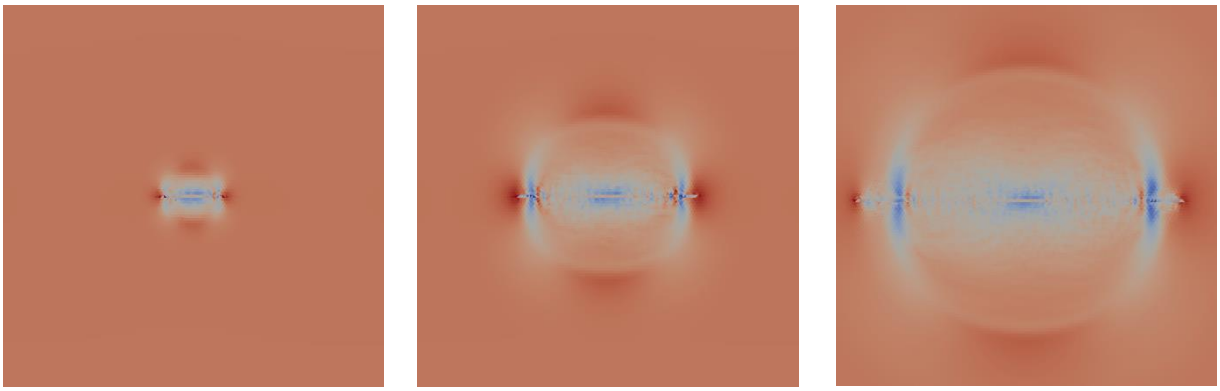
(f) Fault-normal velocity (supershear)

**Figure 13: Stress and velocity distributions for sub-Rayleigh and supershear rupture at t=0.4s**

### *7.4 Transition phenomenon*

In the supershear rupture scenario, the rupture doesn't propagate immediately at intersonic speed. This transition from the initial sub-Rayleigh propagation to supershear rupture, in a homogenous material, is classically explained by the Burridge-Andrews mechanism [13]. At first, the crack propagate behind the shear wave, at a speed close to $c_R$. The P-wave and S-wave propagating ahead generate a zone of high shear stress ahead of the crack tip, eventually creating a secondary crack, also called daughter crack, if the prestress is sufficiently high. When the main crack merges with the daughter crack, there is a jump of the rupture velocity, thereby

starting the intersonic propagation. This phenomenon can be observed in our simulation, and is detailed in Fig. 14.



(a) Sub-Rayleigh propagation ($t = 0.05s$)

(b) Apparition of daughter crack ($t = 0.1s$)

(c) Transition to supershear rupture ($t = 0.25s$)

**Figure 14: Burridge-Andrews mechanism observed in the simulation**

## Concluding remarks

We developed two Hamiltonian-based formulations in the frame of PDS-FEM, and derived consistent time integration schemes which are momentum conserving and symplectic. The separable nature of the Hamiltonian in continuum mechanics ensures that the computation can be done explicitly and thus is relatively cheap computationally. The results of the two verification tests and the validation test show a good accordance with the analytical solution when available, and with the experimental measures in term of crack path, propagation speed and stress distribution around the propagating crack tip. The results also showed that both formulations are numerically indifferentiable, although the displacement formulation proves to have a significantly lower computational cost. This paper also demonstrates that this numerical method can be applied to the simulation of supershear earthquakes, and that characteristic wave profiles and transition mechanism were qualitatively reproduced for both sub-Rayleigh and supershear fault rupture. We are improving computational capabilities of a parallel program with the aim of conducting large scale simulation of supershear earthquake scenarios with the models of actual faults.

## Appendix A: Euler-Lagrange equation of the strain formulation

For the sake of simplicity, the indexes *m* and *n* defining the functions used for interpolation are omitted in the following derivation of the Hamilton's principle, as their inclusion is straightforward. Einstein summation is assumed in the following.
By using the Hamilton's principle, and using Eq. 3:

$$
\begin{aligned}
0 = \delta\left(\int_T L\left(\boldsymbol{v}^{\boldsymbol{\alpha}}, \boldsymbol{\epsilon}^{\boldsymbol{\beta}}\right) dt\right) &= \int_T \left(\frac{\partial L}{\partial \boldsymbol{v}^{\boldsymbol{\alpha}}} \cdot \boldsymbol{\delta \dot{u}}^{\boldsymbol{\alpha}} + \frac{\partial L}{\partial \boldsymbol{\epsilon}^{\boldsymbol{\beta}}} : \boldsymbol{\delta \epsilon}^{\boldsymbol{\beta}}\right) dt \\
&= \int_T \left(-\frac{\partial \dot{L}}{\partial \boldsymbol{v}^{\boldsymbol{\alpha}}} \cdot \boldsymbol{\delta u}^{\boldsymbol{\alpha}} + \frac{\partial L}{\partial \boldsymbol{\epsilon}^{\boldsymbol{\beta}}} : sym\{\boldsymbol{B}^{\boldsymbol{\beta\alpha}} \otimes \boldsymbol{\delta u}^{\boldsymbol{\alpha}}\}\right) dt \\
&= \int_T \left(-\frac{\partial \dot{L}}{\partial \boldsymbol{v}^{\boldsymbol{\alpha}}} \cdot \boldsymbol{\delta u}^{\boldsymbol{\alpha}} + \frac{\partial L}{\partial \epsilon_{ij}^{\beta}} B_i^{\beta\alpha} \delta u_j^{\alpha}\right) dt \\
&= \int_T \left(-\frac{\partial \dot{L}}{\partial \boldsymbol{v}^{\boldsymbol{\alpha}}} \cdot \boldsymbol{\delta u}^{\boldsymbol{\alpha}} + \frac{\partial L}{\partial \boldsymbol{\epsilon}^{\boldsymbol{\beta}}} \boldsymbol{B}^{\boldsymbol{\beta\alpha}} \cdot \boldsymbol{\delta u}^{\boldsymbol{\alpha}}\right) dt
\end{aligned}
$$

Note that only one integration by part is needed, instead of the two needed for the general analytical expression.

By writing $\frac{\partial L}{\partial \boldsymbol{\epsilon^{\beta}}}$ in vector form following Voigt notations: $\left(\frac{\partial L}{\partial \epsilon_{11}^{\beta}}, \frac{\partial L}{\partial \epsilon_{22}^{\beta}}, \frac{\partial L}{\partial \epsilon_{33}^{\beta}}, \frac{\partial L}{\partial \epsilon_{23}^{\beta}}, \frac{\partial L}{\partial \epsilon_{13}^{\beta}}, \frac{\partial L}{\partial \epsilon_{12}^{\beta}}\right)$, and rewriting $\boldsymbol{B^{\beta\alpha}}$ in the adapted matrix form:

$$ {}^{t}\tilde{B}^{\beta\alpha} = \begin{pmatrix} B_1^{\beta\alpha} & 0 & 0 & 0 & B_3^{\beta\alpha} & B_2^{\beta\alpha} \\ 0 & B_2^{\beta\alpha} & 0 & B_3^{\beta\alpha} & 0 & B_1^{\beta\alpha} \\ 0 & 0 & B_3^{\beta\alpha} & B_2^{\beta\alpha} & B_1^{\beta\alpha} & 0 \end{pmatrix} $$

and finally replacing them in the above integral, the vector form Euler-Lagrange equation (Eq. 5) can be derived.

## Appendix B: Boundary conditions

*Dirichlet BC*

Let $u_D$ be the value of displacement imposed on the boundary.
We will set the coefficients $u^{\alpha n}$ such as:

$$ \frac{\partial}{\partial u^{\alpha n}} \int_{\Gamma_D} \left| P^{\alpha' n'} u^{\alpha' n'} - u_D \right|^2 ds = 0 \Rightarrow \sum_{n'} I_S^{\alpha n n'} u^{\alpha n'} = \int_{\Gamma_D} u_D p^{\alpha n} \, ds $$

where $I_S^{\alpha n n'} = \int_{\Gamma_D} P^{\alpha n} P^{\alpha n'} ds$.

We note $W_S^{\alpha} = (I_S^{\alpha})^{-1}$, $u^{\alpha n} = \sum_{n'} W_S^{\alpha n n'} \int_{\Gamma_D} u_D p^{\alpha n'} \, ds$

This inverse is ensured to exist only for the 0th order (when $n = n' = 0$), as otherwise several combinations of coefficients may be solution to the boundary problem. In particular, the integration of first degree polynomials on a surface can lead to null columns in the matrix. Therefore, for the boundary only, only the 0th order terms are computed while the other components are set to 0.

The momentum corresponding condition is derived using Eq. 6}:

$$ p^{\alpha n} = \rho \sum_{n'} I^{\alpha n n'} \sum_{n"} W_S^{\alpha n n"} \int_{\Gamma_D} \dot{u_D} p^{\alpha n"} \, ds $$

*Neumann BC*

We consider a Lagrangian (with *u* as an additional variable for the strain-based formulation, not necessary for the displacement-based one):

$$ L(\boldsymbol{v}, \boldsymbol{\epsilon}, \boldsymbol{u}) = \int_V \frac{1}{2} \rho \boldsymbol{v}^2 dV - \frac{1}{2} \boldsymbol{\epsilon} : \boldsymbol{C} : \boldsymbol{\epsilon} \, dV + \int_{\Gamma_N} T_N . u \, ds $$

In PDS-FEM formalism, applying the Hamilton's principle leads to the updated Euler-Lagrange equation:

$$ \frac{d}{dt} \frac{\partial L}{\partial v^{\alpha n}} - \sum_{\beta, m} {}^{t}\tilde{B}^{\beta \alpha m n} . \frac{\partial L}{\partial \epsilon^{\beta m}} - \int_{\Gamma_N} T_N . P^{\alpha n} \, ds = 0 $$

adding the term $\int_{\Gamma_N \cap \Phi^{\alpha}} T_N . P^{\alpha n} \, ds$ to the calculation of $\dot{p}^{\alpha n}$ in the Hamiltonian systems.

## References

[1]Ferté, G., Massin, P., & Moes, N. (2016). 3D crack propagation with cohesive elements in the extended finite element method. *Computer Methods in Applied Mechanics and Engineering*, *300*, 347-374.
[2]Chen, L., Rabczuk, T., Bordas, S. P. A., Liu, G. R., Zeng, K. Y., & Kerfriden, P. (2012). Extended finite element method with edge-based strain smoothing (ESm-XFEM) for linear elastic crack growth. *Computer Methods in Applied Mechanics and Engineering*, *209*, 250-265.
[3] Moës, N., Gravouil, A., & Belytschko, T. (2002). Non-planar 3D crack growth by the extended finite element and level sets—Part I: Mechanical model. *International journal for numerical methods in engineering*, *53*(11), 2549-2568.
[4] Cockburn, B. (2003). Discontinuous galerkin methods. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik: Applied Mathematics and Mechanics*, *83*(11), 731-754.
[5] Seagraves, A., & Radovitzky, R. (2015). Large-scale 3D modeling of projectile impact damage in brittle plates. *Journal of the Mechanics and Physics of Solids*, *83*, 48-71.
[6] Bede, N., Ožbolt, J., Sharma, A., & İrhan, B. (2015). Dynamic fracture of notched plain concrete beams: 3D finite element study. *International Journal of Impact Engineering*, *77*, 176-188.
[7] Hori, M., Oguni, K., & Sakaguchi, H. (2005). Proposal of FEM implemented with particle discretization for analysis of failure phenomena. *Journal of the Mechanics and Physics of Solids*, *53*(3), 681-703.
[8] Pal, M. K., Wijerathne, L., Hori, M., & Ichimura, T. (2015). Simulation of cracks in linear elastic solids using higher order Particle Discretization Scheme-FEM. *Journal of Japan Society of Civil Engineers A2 (Applied Mechanics) 71*(2), I_327-I_337.
[9] Pal, M. K., Wijerathne, L., Hori, M., Ichimura, T., & Tanaka, S. (2014). Implementation of finite element method with higher order particle discretization scheme. *Journal of Japan Society of Civil Engineers A2 (Applied Mechanic)*, *70*(2), I_297-I_305.
[10]Wijerathne, M. L. L., Oguni, K., & Hori, M. (2009). Numerical analysis of growing crack problems using particle discretization scheme. *International journal for numerical methods in engineering*, *80*(1), 46-73.
[11] Hori, M., Wijerathne, M. L. L., Riaz, M., & Ichimura, T. (2018). Rigorous derivation of hamiltonian from lagrangian for solid continuum. *Journal of Japan Society of Civil Engineers A2 (Applied Mechanics) 6*(1), 1-11.
[12] Burridge, R. (1973). Admissible speeds for plane-strain self-similar shear cracks with friction but lacking cohesion. *Geophysical Journal International*, *35*(4), 439-455.
[13] Andrews, D. J. (1976). Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research*, *81*(32), 5679-5687.
[14] Das, S., & Aki, K. (1977). A numerical study of two-dimensional spontaneous rupture propagation. *Geophysical journal international*, *50*(3), 643-668.
[15] Rosakis, A. J., Samudrala, O., & Coker, D. (1999). Cracks faster than the shear wave speed. *Science*, *284*(5418), 1337-1340.
[16] Xia, K., Rosakis, A. J., & Kanamori, H. (2004). Laboratory earthquakes: The sub-Rayleigh-to-supershear rupture transition. *Science*, *303*(5665), 1859-1861.
[17] Archuleta, R. J. (1984). A faulting model for the 1979 Imperial Valley earthquake. *Journal of Geophysical Research: Solid Earth*, *89*(B6), 4559-4585.
[18] Bouchon, M., Bouin, M. P., Karabulut, H., Toksöz, M. N., Dietrich, M., & Rosakis, A. J. (2001). How fast is rupture during an earthquake? New insights from the 1999 Turkey earthquakes. *Geophysical Research Letters*, *28*(14), 2723-2726.
[19] Bouchon, M., & Vallée, M. (2003). Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake. *Science*, *301*(5634), 824-826.
[20] Ellsworth, W. L., Celebi, M., Evans, J. R., Jensen, E. G., Kayen, R., Metz, M. C., ... & Stephens, C. D. (2004). Near-field ground motion of the 2002 Denali Fault, Alaska, earthquake recorded at Pump Station 10. *Earthquake spectra*, *20*(3), 597-615.
[21] Socquet, A., Hollingsworth, J., Pathier, E., & Bouchon, M. (2019). Evidence of supershear during the 2018 magnitude 7.5 Palu earthquake from space geodesy. *Nature Geoscience*, *12*(3), 192.
[22] Hairer, E., Lubich, C., & Wanner, G. (2006). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations* (Vol. 31). Springer Science & Business Media.

# On the heat conduction in micro-periodic laminate: uncertainty of material properties

***Piotr Ostrowski and Jarosław Jędrysiak**

Department of Structural Mechanics, Lodz University of Technology, Poland

*Presenting and corresponding author: piotr.ostrowski@p.lodz.pl

## Abstract

An object of our consideration is a two-phase laminate of micro-periodic structure in one of directions. That periodicity is represented by distinguished unit cell of diameter $l$ much smaller when compared with other composite dimensions. In the remaining directions laminate has uniform structure. Each component has isotropic material properties, however theirs values are uncertain from assumption, e.g. ratios of conductivities $\omega = k_2 / k_1$ and specific heats $\chi = c_2\rho_2 / c_1\rho_1$ are random variables of known probabilistic distribution.

By any heat flow, transversally to the laminae, there can be two phenomena observed: jump of the gradient of temperature field on interfaces, and apparent temperature oscillations around averaged temperature, with (local) maximum value on interfaces. That function of temperature oscillations depend strongly on parameters $\omega$ and $\chi$, what makes it new random variable, cf. [1]. In this presentation we are going to answer if the maximum of these oscillations is also Gaussian random variable. Steady state as well as transient state will be considered, and all obtained results will be presented.
The govern equation of heat transfer is described by well-known Fourier's law. Easy to see that for considered structure coefficients in the PDE are discontinuous and highly oscillating. Therefore, a tolerance averaging technique (cf. Woźniak et al. [2]) is used in order to get averaged model equations of constant coefficients. These new equations will be used in further numerical (Monte-Carlo) simulations.

**Keywords:** Heat transfer, laminate, uncertain material properties, tolerance averaging technique

## Introduction

The object of our consideration is a multi-layered heat conductor, build of two different materials distributed alternately along $x := x_1$ axis, Fig. 1. Composite layout is *l*-periodic in such a way that each interval $\Delta \subset \mathbb{R}$ of length $l > 0$ consists of two components, and the first one is of volume ratio $\gamma \in [0,1]$. These components, numbered by $i = 1$ or 2, are called sub-laminae and have isotropic material properties: $k_i$ - conductivity, $c_i$ - specific heat, and $\rho_i$ - density.

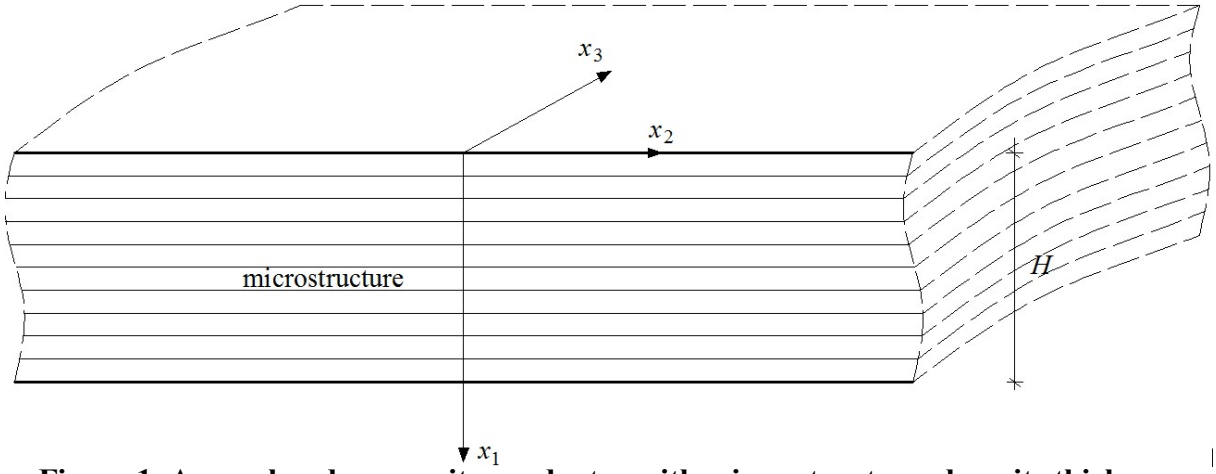



**Figure 1. An analysed composite conductor with micro-structure along its thickness**

The govern equation of the heat transfer in a conductor under consideration is a partial differential equation (PDE) having, due to micro-structure, highly oscillating and discontinuous coefficients. The exact solution to the steady or transient state of heat transfer is reachable for uniform conductor, but this is not the case. Naturally, one can build a system of PDEs, written for every single sub-layer, satisfying appropriate continuity and boundary conditions, but still, one may find a number of problems in resolvability. Even if we narrow our restrictions to one-directional heat flow, perpendicular to the layers, it would not relax our problem entirely. Therefore, we shall use in this paper the tolerance averaging technique, cf. Woźniak et al. [2]-[3], which leads to the system of differential equations with continuous coefficients and, what is crucial, for special cases (steady state) to the same solution as the exact one.

Number of publications have been devoted to the problem of heat transfer in micro-heterogeneous structures and its modelling, for which differential equations are of discontinuous coefficients. The most popular averaging approaches are based on the asymptotic homogenisation cf. Jikov et al. [4]. For instance, homogenisation theory was realised by Matysiak and Yevtushenko [5] or Matysiak and Perkowski [6] by using a concept of micro-local parameters. To another approaches, but still popular, belong RVE methods. We can mention here for example works of Han et al. [7] or Bayat and Gaitanaros [8].

A new class of problems is recently investigated. Namely the uncertainty effect of physical or geometrical properties on the overall composite behaviour. For instance, probabilistic homogenization with Monte-Carlo simulation in fiber composites is investigated by Kamiński in [9]. Dynamics of micro-periodic composite rod with uncertain parameters under a moving random load is presented by Mazur-Śniady et al. [10]. A refined averaged theory of a rigid heat conductor with a micro-periodic structure is used by Ignaczak and Baczyński [11] to solve a one-dimensional heat conduction problem in a periodically layered plate.

Averaged model equations used in presented paper describe heat transfer in considered composite, and theirs origin we find in tolerance averaging technique. This particular method is commonly used to other then heat conduction problems: e.g. to dynamics of shells [12] and plates [13]-[15], to thermoelasticity [16]-[17].

There are two general goals we would like to present. The first one deals with application of tolerance averaging technique in order to have equations of constant coefficients describing well the heat conduction problem in considered laminate. The second one is to investigate the measure of magnitude of temperature oscillations and how these oscillations depend on uncertainty of material properties.

**Preliminaries**

Let $\Xi \subset \mathbb{R}^3$ be a bounded region in Euclidean space occupied by the conductor under consideration, wherein Cartesian coordinate system $Ox_1x_2x_3$ is introduced. Denote by $\Omega := (0,H)$, $H>0$, a bounded and regular region in $\mathbb{R}$ assigned to the micro-structure. Hence, the region of the composite can be expressed as $\Xi = \Omega \times \Pi$, where $\Pi := (0,L)\times(0,B)$, for $B,L>0$, is a bounded subspace of $\mathbb{R}^2$. In other words, the conductor has for every $x_1 \in [0,H]$ invariant homogeneous structure and material properties, see Fig. 2.

Through this paper points from $\Omega$ are denoted by Latin letters $x \equiv x_1$ or $y \equiv y_1$, while from $\Xi$ by $\mathbf{x} = (x_1,x_2,x_3)$. Points from $\Pi$ are denoted by $\mathbf{z} = (x_2,x_3)$, and the time coordinate by $t$. Gradient operators used in this contribution are $\partial = (\partial_1,0,0)$, $\overline{\nabla} = (0,\partial_2,\partial_3)$ and $\nabla = \partial + \overline{\nabla}$, where $\partial_i = \partial / \partial x_i$, $i = 1,2,3$, stand for partial derivatives. Dots over the function name stand for the time derivatives.

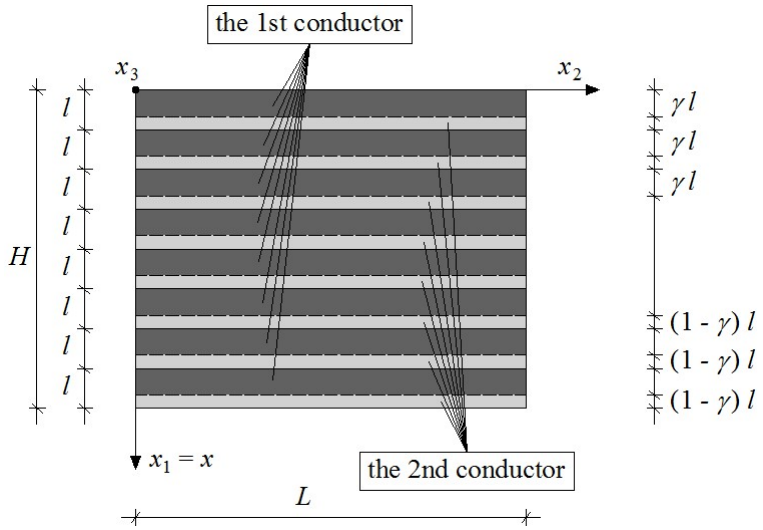



**Figure 2. Laminar layout of micro-structure**

The heat conduction equation is based in this contribution on the well-known Fourier's law

$$\nabla \cdot (\mathbf{K} \cdot \nabla \Theta) = c\rho\, \dot{\Theta}, \tag{1}$$

where $\mathbf{K} = [k_{ij}]$, $k_{ij} : \Xi \to \mathbb{R}$, stands for the second order conductivity tensor, $c : \Xi \to \mathbb{R}$ is a specific heat and $\rho : \Xi \to \mathbb{R}$ is the density. Since material properties are isotropic we have $k_{ij} = k \cdot \delta_{ij}$ for some $k>0$, where $\delta_{ij}$ is Kronecker's delta.

Equation (1) has to be satisfied for every $\mathbf{x} \in \Xi$ and $t \in (t_0,t_1)$ by continuous function of temperature $\Theta : \Xi \times (t_0,t_1) \to \mathbb{R}$. Easy to see that direct description leads to the system of PDEs with highly oscillating coefficients, and it might be far to complicated to solve it in engineering applications. Even for a unidirectional problem

$$\begin{aligned} &\partial \cdot (\mathbf{K} \cdot \partial \Theta) = c\rho\, \dot{\Theta}, \\ &\Theta(0,t) = \Theta_1, \quad \Theta(H,t) = \Theta_2 \quad \text{for } t \in (t_0,t_1), \\ &\Theta(x,t_0) = \Theta_0 \quad \text{for } x \in (0,H), \end{aligned} \tag{2}$$

where the number of independent variables decreases to two, $x$ and $t$.

Denote the ratios

$$\omega = \frac{k_2}{k_1} \quad \text{and} \quad \chi = \frac{c_2}{c_1}\frac{\rho_2}{\rho_1} \tag{3}$$

as inhomogeneity parameters. They take only positive values, and equal one for homogeneous material.

In order to give appropriate motivation to our studies, consider a following problem of heat transfer in the laminate build of $n$ two-component layers: find continuous function $\Theta:[0,H]\to\mathbb{R}$ satisfying boundary conditions, $\Theta=\Theta_1$ at $x=0$ and $\Theta=\Theta_2$ at $x=H$. Easy to prove that general solution to Eq. (2) for the case of steady state (time derivative vanishes and the problem is independent of time $t$) can be decomposed into the sum of averaged temperature

$$\Theta_{avg}(x) = \Theta_1 + (\Theta_2 - \Theta_1)\cdot\frac{x}{H},\ x\in[0,H], \tag{4}$$

and oscillating temperature $\Theta_{osc}$. The last one is a "saw-type" function, oscillating around zero value, having local extrema on interfaces and depending explicitly on parameter $\omega$, e.g. for a special distribution of sub-laminae we have

$$\sup_{x\in[0,H]}\left|\Theta_{osc}(x)\right| = \left|\Delta\Theta\right|\cdot\frac{\gamma\cdot(1-\gamma)}{2n}\cdot\frac{|\omega-1|}{1+\gamma\cdot(\omega-1)}, \tag{5}$$

where $\Delta\Theta=\Theta_2-\Theta_1$. Part of this function, $h(\omega)=\frac{\gamma\cdot(1-\gamma)}{2}\cdot\frac{|\omega-1|}{1+\gamma\cdot(\omega-1)}$, is depicted in Fig. 3 where we can see how strong values of $\Theta_{osc}$ depend on $\omega$.

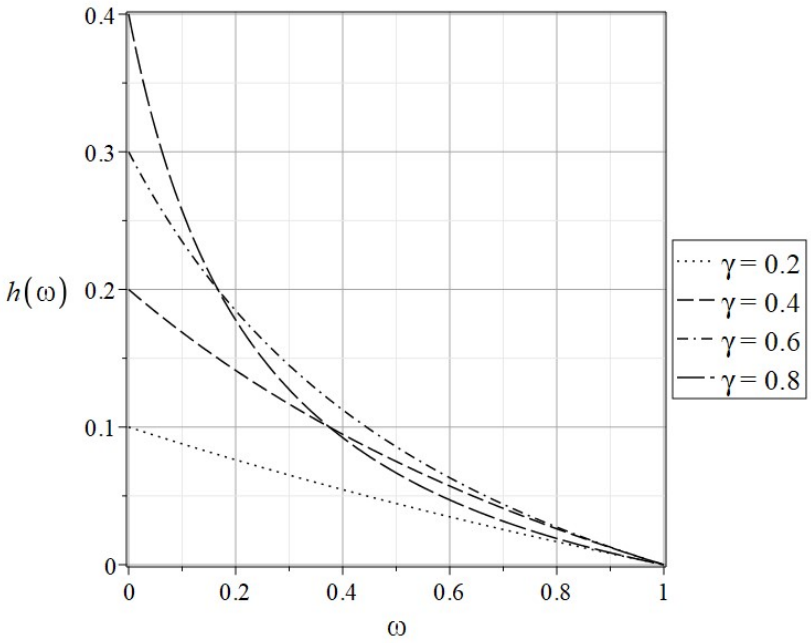



**Figure 3. Plot of function $h(\omega)$**

One can also observe that limits

$$\lim_{\omega\to 0^+} h(\omega)=\frac{\gamma}{2} \text{ and } \lim_{\omega\to +\infty} h(\omega)=\frac{1-\gamma}{2} \tag{6}$$

exist and are finite.

We can imagine now that the magnitude of $\Theta_{osc}$ is sensitive, in transient state as well as in steady state, to differences of material properties. The bigger differences we have, the bigger magnitude of oscillations we get. In other words, function $\Theta_{osc}$ on interfaces depends on parameters $\omega$ and $\chi$.

In this paper we assume that parameters $\omega$ and $\chi$ are uncertain, i.e. they are random variables of known probability distribution, lognormal to be precise. Hence, there exist $\mu_\omega$, $\mu_\chi>0$ and $\sigma_\omega$, $\sigma_\chi\geq 0$ such that $\omega\sim\Lambda\left(\mu_\omega,\sigma_\omega^2\right)$ and $\chi\sim\Lambda\left(\mu_\chi,\sigma_\chi^2\right)$. This denotation means that median values of $\omega$, $\chi$ are respectively $\exp(\mu_\omega)$, $\exp(\mu_\chi)$, while expected values are $\exp(\mu_\omega+\sigma_\omega^2/2)$, $\exp(\mu_\chi+\sigma_\chi^2/2)$. For the sake of simplicity, initial-boundary conditions to Eq. (1) are imposed in such a way to assure unidirectional (along $x_1$ axis) heat transfer problem.

**Modelling concepts**

Throughout this paper, by $H^r(\Xi)$, $r\geq 0$, we shall understand a Sobolev space of functions which are, together with theirs weak derivatives to the *r*th order, $L^2$-measurable on $\Xi$. Function space $H^r(\Delta)$ denotes the space of all $\Delta$-periodic functions which are $H^r(V)$ on any arbitrary compact subset $V\subset\mathbb{R}$. Easy to see that $H^0$ means the same what $L^2$, however there is no equilibrium between $H^r(\Xi)$ and $H^r(V)$ for $V=\Xi$. All essentially bounded functions on $X\subset\mathbb{R}^m$, $m=1,2,3$, are denoted by $L^\infty(X)$.

Let $n\in\mathbb{N}$ be the number of two-component layers of common width $l=H/n$. Each layer (cf. Fig. 2) consists of two sub-layers made of different material. The first one, called „conductor 1", is of width $d_1=\gamma\cdot l$, where $\gamma\in[0,1]$ is fixed. Second sub-layer, called „conductor 2", has therefore width $d_2=(1-\gamma)\cdot l$. Easy to see that uniform conductor provides $\gamma\in\{0,1\}$.

Fix for a moment $x\in\bar{\Omega}$. Since the composite is periodic, cf. Fig. 2, representative volume element $\Delta=(-l/2,l/2)\subset\mathbb{R}$, called further *a unit cell*, can be simply distinguished. To every cell $\Delta$ we can assign a local coordinate system $Oy$, and the cell with a centre at $x$ is denoted by $\Delta(x)=x+\Delta$. Note, that above representation of cell $\Delta$ is not the only one, it is only an example. It can be any *l*-length surrounding of 0, i.e. $\Delta=(a,a+l)$ for $a\in[-l/2,0]$.

In order to derive averaged model equations we apply the tolerance averaging technique, which is based mainly on the concept of tolerance and in-discernibility relation. Its definition is given below.

*Definition* 1. Let $\varepsilon$ stands for an arbitrary positive number. We say that numbers $a,b\in\mathbb{R}$ are in tolerance relation $a\approx b$ if and only if $|a-b|\leq\varepsilon$. Parameter $\varepsilon$ is called *the tolerance parameter*.

The general modelling procedures, basic definitions and theorems of this technique can be found in the book by Woźniak and Wierzbicki [2] or by Ostrowski [18]. We will mention here some basic concepts of this technique, but first we introduce a notion of maximum oscillation of continuous function $f$ on $\Delta(x)$ as follows

$$[f]_x=\Omega_\Delta\ni x\mapsto\sup_{y\in\Delta(x)}f(y)-\inf_{y\in\Delta(x)}f(y)\in\mathbb{R}, \tag{7}$$

which is very helpful in constructing definition of slowly-varying function.

*Definition* 2. Function $F\in C^r(\Xi)$ is called *the slowly varying function of* r*th order*, with respect to cell $\Delta$ and tolerance parameter $\varepsilon$, if for every $p=0,1,...,r$ following conditions hold

$$\forall(x,\mathbf{z})\in\Xi\quad\left[\partial_1^pF(\cdot,\mathbf{z})\right]_x\approx 0. \tag{8}$$

Set of all *r*th order slowly varying functions with respect to the cell $\Delta$ and tolerance parameter $\varepsilon$ is denoted by $SV_\varepsilon^r(\Xi,\Delta)$.

Another important definition in tolerance modelling is the definition of the mean value operator

$$\langle f\rangle(x,\cdot):=\frac{1}{l}\cdot\int_{\Delta(x)}f(y,\cdot)\,dy. \tag{9}$$

which can be applied to any locally integrable function $f\in L^1_{loc}(\Xi)$. In applications we restrict ourselves to essentially bounded functions, i.e. $f\in L^\infty(\Xi)$.

The last definition related to tolerance averaging technique is a *periodic-like function* [2], whose name was after years changed into *tolerance periodic function*.

*Definition* 3. Function $f\in L^\infty(\Xi)\cap H^r(\Xi)$ will be called *the tolerance periodic function of* r*th order*, with respect to the cell $\Delta$ and tolerance parameter $\varepsilon$, if for every $p=0,1,...,r$ and every $x\in\Omega_\Delta$ there exists periodic approximation $f_x\in L^\infty_{per}(\Xi)\cap H^r(\Xi)$ of function $f$ such that

$$\forall\mathbf{z}\in\Pi\quad\left[\left\langle\partial_1^pf-\partial_1^pf_x\right\rangle(\cdot,\mathbf{z})\right]_x\approx 0. \tag{10}$$

Set of all *r*th order tolerance periodic functions with respect to the cell $\Delta$ and tolerance parameter $\varepsilon$ will be denoted by $TP_\varepsilon^r(\Xi,\Delta)$.

The space $L^\infty_{per}(\Xi)$ mentioned above is a set of all essentially bounded functions defined on $\Xi$ which are periodic, in particular $\Delta$-periodic. By the notion of $\partial_1^0f$ we shall understand $f$, and by $O(\cdot)$ Landau's symbol is denoted. The tolerance parameter $\varepsilon$ related to any tolerance periodic function can be determined only a posteriori.

Another class of functions possessing special properties is a class of fluctuation shape functions.

*Definition* 4. Function $g \in C^0(\Xi)$ is called *the fluctuation shape function* of weight $\rho \in L^\infty(\Xi)$, if following conditions hold

(a) $\langle \rho g \rangle \approx 0$ on $\Xi$,

(b) $\partial_1 g$ is piecewise continuous,

(c) $g \in O(l)$.

Set of all fluctuation shape functions of weight $\rho$ is denoted by $FS_\varepsilon^\rho(\Xi, \Delta)$.

For further simplifications we consider only these functions $g \in FS_\varepsilon^1(\Xi, \Delta)$ which satisfy $\langle g \rangle \equiv 0$. An example of such function is shown in Fig. 4.

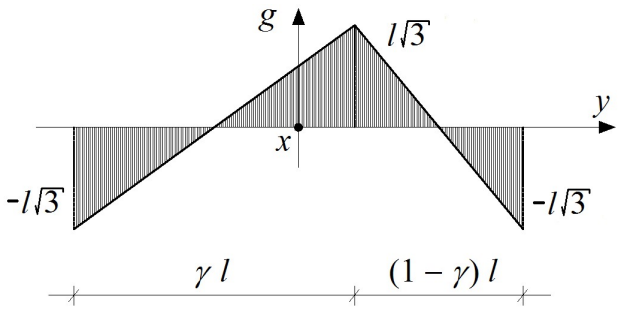



**Figure 4. Fluctuation shape function**

The last part of this section deals with theorem as a conclusion implied from all presented definitions and theirs properties.

*Theorem* 1 (Tolerance Averaging Approximations). For every $g \in FS_\varepsilon^1(\Xi, \Delta)$, $F \in SV_\varepsilon^1(\Xi, \Delta)$, $f \in TP_\varepsilon^0(\Xi, \Delta)$, $\varphi \in L^1(\Omega)$ and $k \in L_{per}^\infty(\Xi)$ the following proposition hold

(a) $\nabla \langle \varphi \rangle = \langle \nabla \varphi \rangle$,

(b) $\langle k \nabla(gF) \rangle \approx \langle k \partial g \rangle F + \langle kg \rangle \bar{\nabla} F$,

(c) $\langle kF \rangle \approx \langle k \rangle F$,

(d) $\langle g \nabla(kf) \rangle \approx \bar{\nabla} \langle gkf \rangle - \langle kf \partial g \rangle$.

Proof of this theorem is given inter alia by Ostrowski in monograph [18].

**Model equations**

In this section we derive averaged model equations describing heat conduction in micro-structured laminate. We must assume first that temperature field $\Theta$ and its time derivative $\dot{\Theta}$ are tolerance periodic functions, i.e. $\Theta \in TP_\varepsilon^1(\Xi, \Delta)$ and $\dot{\Theta} \in TP_\varepsilon^0(\Xi, \Delta)$. Secondly, by tolerance averaging technique we impose on temperature field *micro-macro decomposition* as follows

$$\Theta(\mathbf{x},t)=\theta(\mathbf{x},t)+g(\mathbf{x})\cdot\psi(\mathbf{x},t) \tag{11}$$

for every $\mathbf{x}\in\bar{\Xi}$ and $t\in[t_0,t_1]$, where $\theta(\cdot,t),\psi(\cdot,t)\in SV_\varepsilon^1(\Xi,\Delta)$. This assumption states that $\theta$ and $\psi$ are unknown slowly varying functions for every $t\in[t_0,t_1]$. Function $g\in FS_\varepsilon^1(\Xi,\Delta)$ is *a priori* given, dependent on the micro-structure size parameter *l*, fluctuation shape function as depicted in Fig. 4. Function $\theta$ is called *the averaged temperature* in a medium, while $\psi$ stands for *the temperature oscillation amplitude*.

About functions of material properties we assume that each component has isotropic properties, i.e. $k_{ij}=k\cdot\delta_{ij}$ for some $k>0$, where $\delta_{ij}$ stands for Kronecker's delta. Moreover, let $k,c,\rho\in L_{per}^\infty(\Xi)$.

It is obvious that (1) may not be satisfied by decomposition (11) everywhere on $\Xi$ and for every $t\in[t_0,t_1]$. Nevertheless, we expect from residuum function

$$\Re=\nabla\cdot(\mathbf{K}\cdot\nabla\Theta)-c\rho\,\dot{\Theta} \tag{12}$$

to satisfy on its domain some orthogonal conditions, namely

$$\langle\Re\rangle=0 \text{ and } \langle\Re g\rangle=0. \tag{13}$$

Bearing in mind all properties from Theorem 1 and by omission all terms $O(\varepsilon)$, $O(l)$, we conclude to the final averaged model equations

$$\begin{aligned}&\nabla\cdot\left(\langle\mathbf{K}\rangle\cdot\nabla\theta+\langle\mathbf{K}\cdot\partial g\rangle\cdot\psi\right)=\langle c\rho\rangle\,\dot{\theta},\\&\bar{\nabla}\cdot\left(\langle\mathbf{K}gg\rangle\cdot\bar{\nabla}\psi\right)-\langle\partial g\cdot\mathbf{K}\cdot\partial g\rangle\cdot\psi-\langle\mathbf{K}\cdot\partial g\rangle\cdot\nabla\theta=\langle c\rho gg\rangle\,\dot{\psi}.\end{aligned} \tag{14}$$

The above system has continuous, for periodic structure even constant, coefficients in contrast to equations from the direct description (1) which has discontinuous and highly oscillating ones. System (14) represents equations for the averaged temperature $\theta$ and the temperature fluctuation amplitude $\psi$, and together with micro-macro decomposition (11) constitutes the tolerance model (TM) of the heat conduction in considered laminated conductor.

Along with micro-macro decomposition (11) came two unknown functions, $\theta$ and $\psi$, instead of one $\Theta$. Thus, we need to formulate somehow, based on the known, conditions for these new functions. Let $\Theta_0(\cdot)=\Theta(\cdot,t_0)$ on $\Xi$ be the initial temperature, while $\Theta_1(\cdot)=\Theta(0,\cdot)$ and $\Theta_2(\cdot)=\Theta(H,\cdot)$ be on $\bar{\Pi}\times[t_0,t_1]$ the temperature on the top and bottom surface, respectively. On the remaining boundary surfaces we assume that they are subjected to thermal isolation. Conditions for averaged temperature and temperature oscillation amplitude we can evaluate as

$$\theta_i=\langle\Theta_i\rangle \text{ and } \psi_i=\frac{\langle\Theta_i g\rangle}{\langle gg\rangle}, \tag{15}$$

for every $i=0,1,2$. Easy to see that $\theta_i(\cdot,t)$ and $\psi_i(\cdot,t)$ are for every $t\in[t_0,t_1]$ constant functions if $\Theta_i(\cdot,t)$ is *l*-periodic. In particular, $\psi_i(\cdot,t)\equiv 0$ iff $\Theta_i$ is constant in *x*.

Suppose now $\Theta_0, \Theta_1$ and $\Theta_2$ are constant functions in theirs domain and material properties are isotropic for each of component, i.e. $k_1, c_1, \rho_1$ for the first phase and $k_2, c_2, \rho_2$ for the second phase. The rest of boundary surfaces are thermally isolated. So formulated Cauchy's problem assures unidirectional heat flow in a media (along *x*-axis) and the tolerance model equations (14) reduce to a simpler form

$$\begin{aligned} &\langle k \rangle \cdot \partial_1^2 \theta + \langle k\, \partial_1 g \rangle \cdot \partial_1 \psi - \langle c\rho \rangle \dot{\theta} = 0, \\ &\langle k\, \partial_1 g \rangle \cdot \partial_1 \theta + \langle k\, \partial_1 g\, \partial_1 g \rangle \cdot \psi + \langle c\rho g g \rangle \dot{\psi} = 0. \end{aligned} \tag{16}$$

and the initial-boundary conditions

$$\theta_i = \Theta_i \text{ and } \psi_i = 0, \tag{17}$$

for every $i = 0,1,2$. Averaged coefficients

$$\begin{aligned} &\langle k \rangle = k_1 \cdot \big(\omega + (1-\omega)\cdot\gamma\big), && \langle c\rho \rangle = c_1\rho_1 \cdot \big(\chi + (1-\chi)\cdot\gamma\big), \\ &\langle k\, \partial_1 g \rangle = k_1\sqrt{12}\cdot(1-\omega), && \langle c\rho g g \rangle = l^2 \langle c\rho \rangle, \\ &\langle k\, \partial_1 g\, \partial_1 g \rangle = \frac{12k_1}{\gamma\cdot(1-\gamma)}\cdot\big(1+(\omega-1)\cdot\gamma\big), \end{aligned} \tag{18}$$

are in this case constant and depend explicitly on parameters $\omega$ or $\chi$, cf. (18), that play the role of random variables. If $\omega = \chi = 1$ then we deal with uniform conductor. For further applications we introduce the dimensionless spatial $\xi = x/H$ and time $\tau = t/3600[s]$ coordinates.

If we neglect time derivatives in (16) and make all functions as time independent, then we obtain description to the steady state of heat conduction. Temperature oscillation amplitude $\psi$ depends then explicitly on averaged temperature

$$\psi = -\frac{\langle k\, \partial_1 g \rangle}{\langle k\, \partial_1 g\, \partial_1 g \rangle} \cdot \partial_1 \theta, \tag{19}$$

while $\theta$ must satisfy

$$\partial_1^2 \theta = 0 \tag{20}$$

under already known boundary conditions. What is most interesting, the obtained solution satisfies all continuity conditions, including heat flux across interfaces. Ergo, by tolerance averaging technique we get the exact solution, without solving large system of equations, and without solving any eigenvalue problem.

## Monte-Carlo simulation

The exact solution for the direct description of the heat transfer problem (1) in micro-periodic laminate exists, however it usually needs complex algebraic calculations. For example, by *n* cells in our two-phase laminate we have $2n-1$ interfaces where continuity of temperature and heat flux field should be assured. That makes $4n-2$ equations plus two boundary conditions. By the use of tolerance averaging technique we obtain system of PDEs (or ODEs) but of constant coefficients, wherein the number of equations depends on the number of terms in micro-macro decomposition (11). But that is much smaller then $4n$, usually it is only two.

**Table 1. Reference material properties**

| | Component 1 (steel) | Component 2 (aluminium) |
|---|---|---|
| $k\left[W\,m^{-1}K^{-1}\right]$ | 58 | 200 |
| $c\left[J\,kg^{-1}K^{-1}\right]$ | 500 | 920 |
| $\rho\left[kg\,m^{-3}\right]$ | 7800 | 2700 |

*Example* 1. Let us consider periodic laminate of thickness $H=1[m]$, consisting of $n=20$ two-component layers. Hence, the thickness of a single layer is $l=5[cm]$. Volume fraction of the first component is fixed at $\gamma=0.25$. Let $\Theta_0=0[°C]$ for $t_0=0[s]$ and $\Theta_1=1[°C]$, $\Theta_2=0[°C]$ for $x=0[m]$ and $x=H$, respectively. These imply $\theta_0=0[°C]$ for $\tau_0=0$ and $\theta_1=1[°C]$, $\theta_2=0[°C]$ for $\xi=0$ and $\xi=1$, respectively. Material properties are given in Tab. 1, and thus parameters $\omega=3.448$ and $\chi=0.637$ are fixed. The considered time range for this example is one hour, $t_1=3600[s]$, and it provides range of $[0,1]$ for $\tau$.

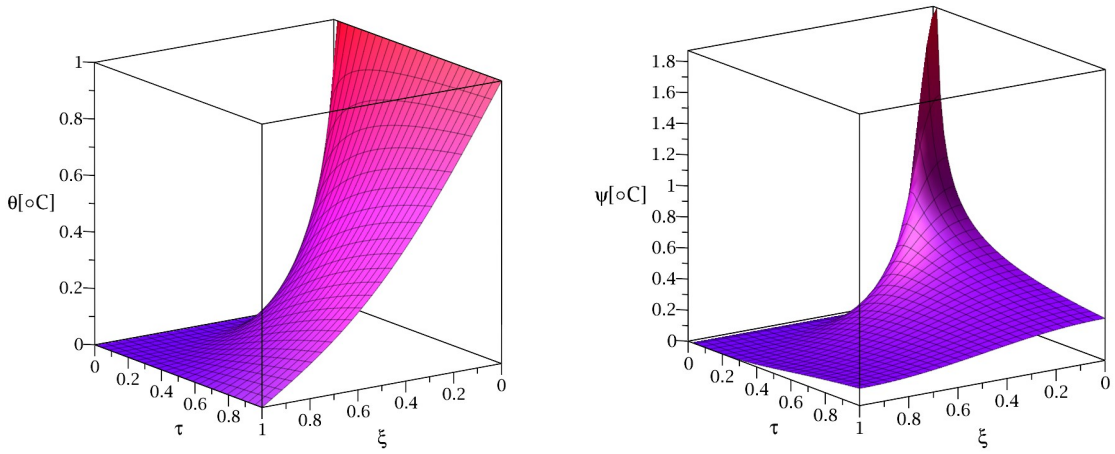



**Figure 5. The averaged and the temperature oscillation amplitude varying in time during one hour for fixed material properties and $\gamma=0.25$**

System (16) were solved under assumption that time derivative of $\psi$ can be omitted (asymptotic model, cf. [1]), i.e. instead of $(16)_2$ we have (19). Fig. 5 depicts changes in time of averaged temperature $\theta$ and temperature oscillation amplitude $\psi$. There is apparent convergence to the steady state, and the maximum oscillation amplitude appears close to top surface ($\xi=0$).

These results were prepared only for certain values of material properties. Suppose now that $k_1, c_1, \rho_1$ are fixed while $\omega$ and $\chi=\omega-2.811$ are uncertain parameters. To be precise, we will investigate the impact of randomness of parameter $\omega$ on the magnitude of temperature oscillations

$$\Psi(x,t)=\sup_{y\in\Delta(x)}\left|g(y)\cdot\psi(y,t)\right|,\ x\in\Omega_\Delta,\ t\in[t_0,t_1]. \tag{21}$$

All following simulations are restricted to the case of fixed spatial coordinate $\xi = 1/n$, but the whole analysis could be simply transferred for any value of $\xi$.

In the next example we expect to find and investigate an effect of variables $\omega$ and $\chi = \omega - 2.811$ on the function $\Psi = \Psi(\omega, \xi, \tau)$, under given above geometry and initial-boundary conditions, but for various values of $\gamma$.

*Example* 2. We postulate that $\Psi(\omega, \xi, \tau)$, given by Eq. (21), is for $\xi = 1/n$ and every $\tau \in [0,1]$ a new random variable with unknown probability distribution. To variable $\omega$, we say that it is of lognormal distribution with parameters: mean value $\mu = \mu_\omega$ and standard deviation $\sigma = \sigma_\omega$. Moreover, we assume that $\sigma = \nu \cdot \mu$ for $\mu = 3.448$ and $\nu = \nu(\omega) > 0$. All statistical characteristics, like expected value $E(\Psi)$, standard deviation $\sigma(\Psi)$, skewness $\beta(\Psi)$, kurtosis $\kappa(\Psi)$ and coefficient of variation $\nu(\Psi)$, will be determined in order to qualify $\Psi$ to Gaussian distributed variable. Subsequent numerical experiment is based on the Monte Carlo simulation for $N = 1000$ probe values.

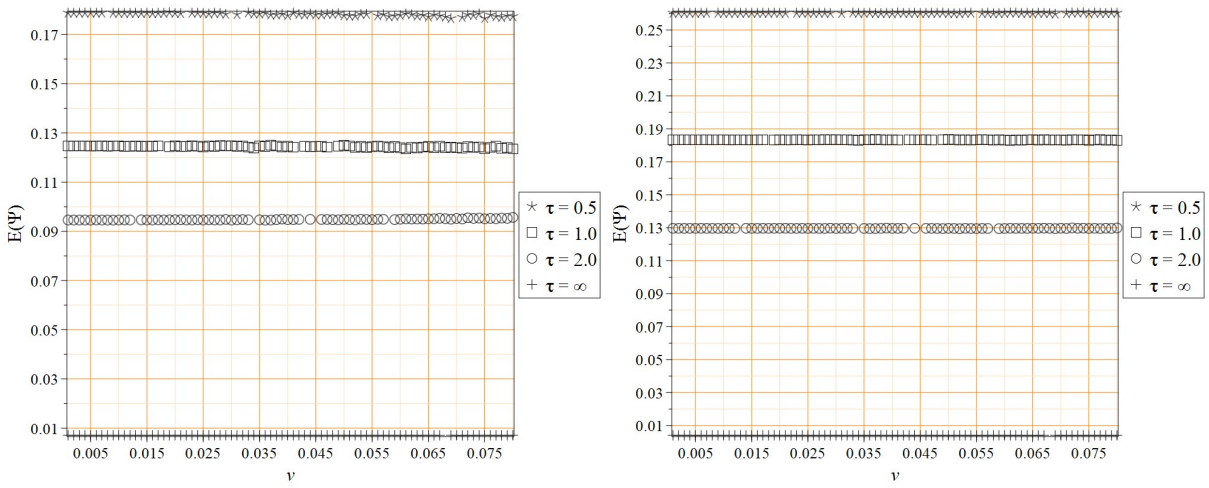



**Figure 6. The expected values of $\Psi$ against $\nu$ varying in time for $\gamma = 0.25$ and $\gamma = 0.75$**

In Fig. 6 we see how the magnitude of temperature oscillation vary in time, and theirs values are not necessary negligibly small when compared with total temperature. It seems that they are not affected with parameter $\nu$.

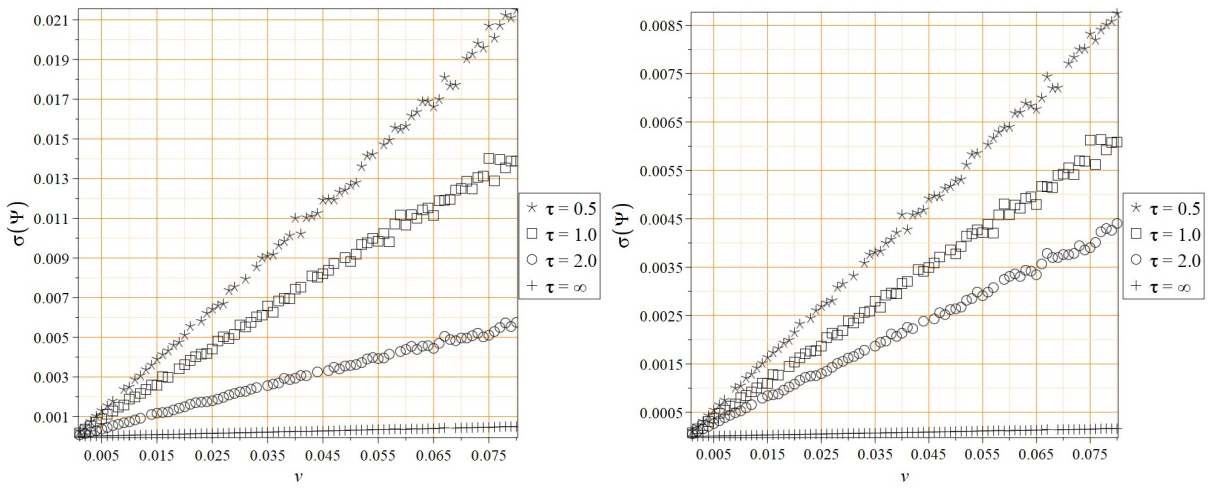


**Figure 7. The standard variations of $\Psi$ against $\nu$ varying in time for $\gamma = 0.25$ and $\gamma = 0.75$**

Standard deviations depend on parameter $\nu$ for sure, but that dependence is almost linear. The highest standard deviations we get for first moment of time, then they drop down.

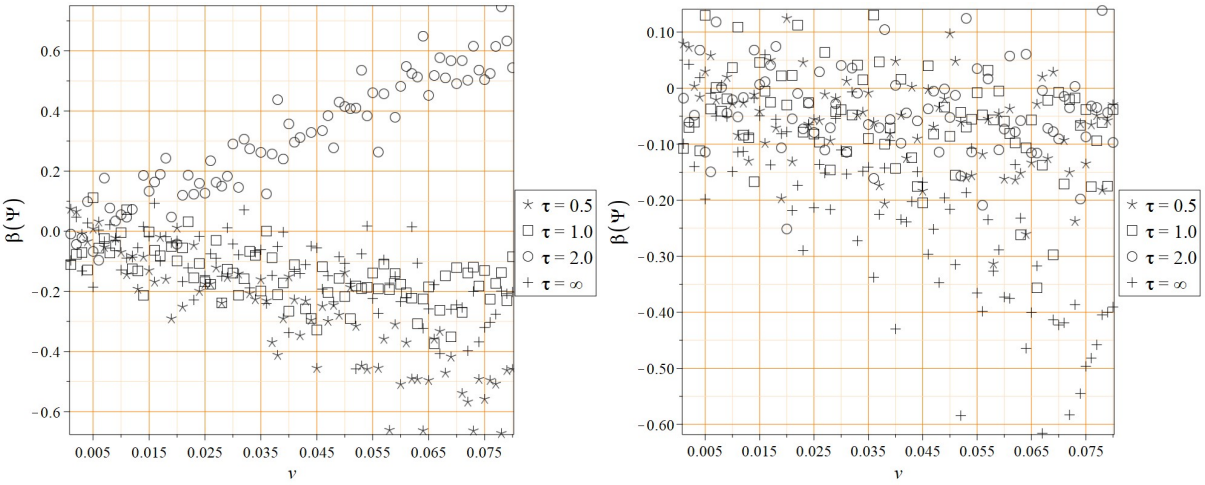



**Figure 8. The skewness's of $\Psi$ against $\nu$ varying in time for $\gamma = 0.25$ and $\gamma = 0.75$**

Skewness is first statistical parameter that says a lot about character of randomness. If its value is close to zero then we may suspect that this is Gaussian distribution. In Fig. 8 we can see that there is no typical pattern for skewness's but theirs values are sufficiently small.

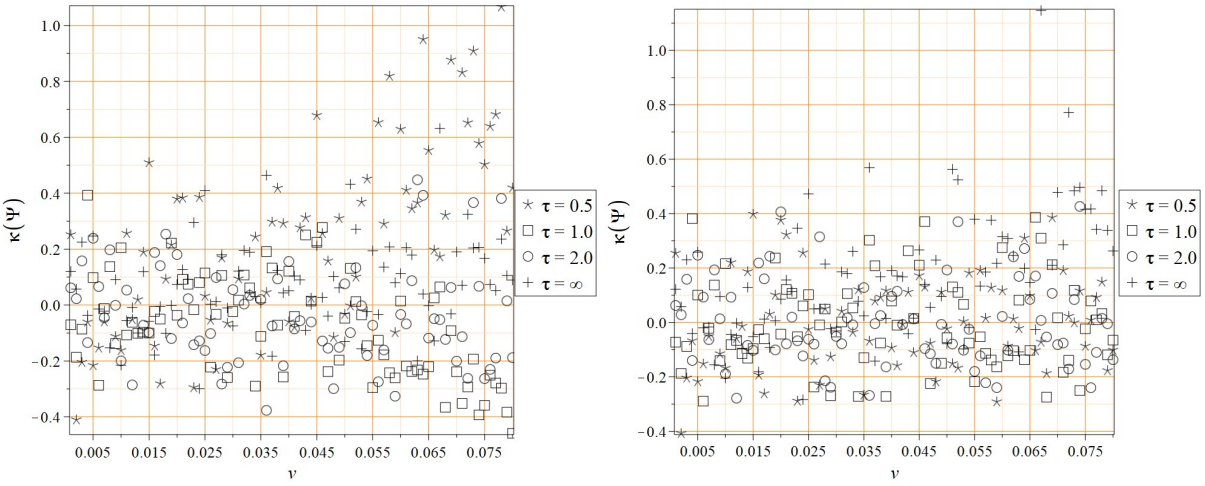



**Figure 9. The kurtoses of $\Psi$ against $\nu$ varying in time for $\gamma = 0.25$ and $\gamma = 0.75$**

The second statistical parameter is kurtosis, Fig. 9, determining whether $\Psi$ is Gaussian variable. The closer value to zero the better alignment we get. As well as skewness's, results are kind of chaotic but also sufficiently small.

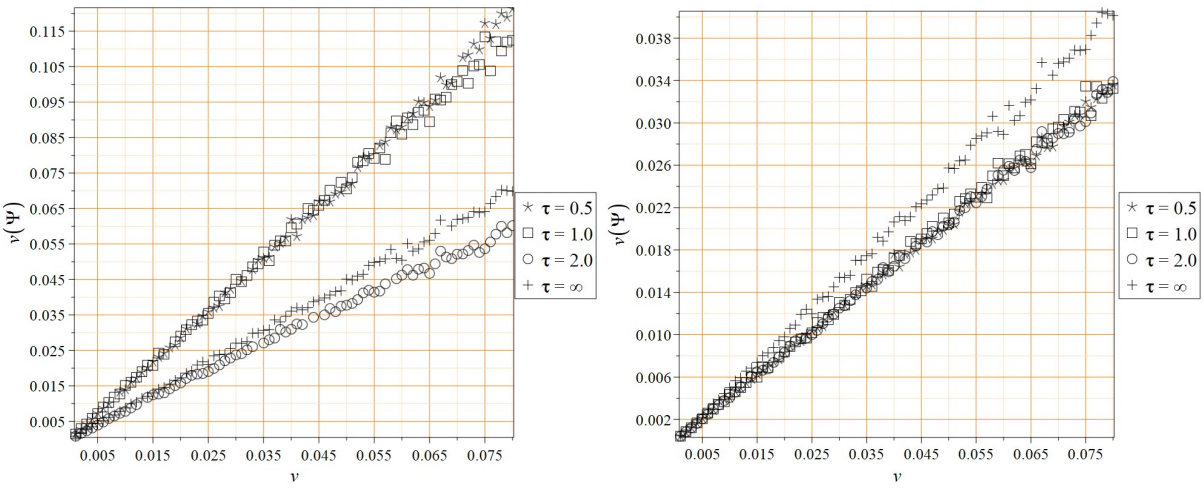



**Figure 10. The coefficients of variation of $\Psi$ against $\nu$ varying in time for $\gamma = 0.25$ and $\gamma = 0.75$**

The most interesting parameter is the ratio of standard deviation and expected value, called coefficient of variation, Fig. 10, because it says a lot about relative dispersion of expected values.

## Summary

We assumed that ratio of conductivities $\omega$ (ratio of specific heats $\chi$ depends explicitly on $\omega$) is a random variable of lognormal distribution. That distribution was considered in many variations, i.e. for different values of $\nu$. Its change affects naturally the function of magnitude of temperature oscillation $\Psi$, which is also a random variable. Statistical characteristics were calculated and Shapiro-Wilk's test for normality was made for each case. Only those results, for which the test gave positive answer, were presented in Figs 6-10. Plots of coefficient of variation reveal interesting conclusion: estimation of $\Psi$ might be of higher probability then made for ratio $\omega$. But this is only from specific moment of time: $\nu(\Psi) \le \nu$ for $\tau > 2$ regardless from $\gamma$.

### References

[1] Ostrowski, P. and Jędrysiak, J. (2017) Heat conduction in periodic laminates with probabilistic distribution of material properties, *Heat and Mass Transfer* **53**(4), 1425–1437.
[2] Woźniak, Cz and Wierzbicki, E. (2000) *Averaging techniques in thermomechanics of composite solids*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa, Poland.
[3] Woźniak, Cz. et al. [Eds] (2008) *Thermomechanics of microheterogeneous solids and structures. Tolerance averaging approach*, Lodz University of Technology Press, Łódź, Poland.
[4] Jikov, V. V, Kozlov, S. M. and Oleinik, O. A. (1994) *Homogenization of differential operators and integral functionals*, Springer-Verlag, Berlin, Heidelberg.
[5] Matysiak, S. J. and Yevtushenko, A. A. (2008) Influence of microstructures on temperature distribution for some mixed problem of composite half-space, *International Communications in Heat and Mass Transfer* **35**, 704–709.
[6] Matysiak, S. J. and Perkowski, D. M. (2014) Temperature distributions in a periodically stratified layer with slant lamination, *Heat and Mass Transfer* **50**(1), 75–83.
[7] Han, X., Xu, Ch., Xie, W. and Meng, S. (2019) Multiscale computational homogenization of woven composites from microscale to mesoscale using data-driven self-consistent clustering analysis, *Composite Structures* **220**, 760–768.
[8] Bayat, A. and Gaitanros, S. (2019) Elastic wave propagation in open-cell foams, *Journal of Applied Mechanics* **86**(5), 051008 (11 pages).
[9] Kamiński, M. (1999) Monte-Carlo simulation of effective conductivity for fiber composites, *International Communications in Heat and Mass Transfer* **26**(6), 791–800.

[10] Mazur-Śniady, K., Śniady, P. and Zielichowski-Haber, W. (2009) Dynamic response of micro-periodic composite rods with uncertain parameters under moving random load, *Journal of Sound and Vibrations* **320**(1-2), 273–288.
[11] Ignaczak, J. and Baczyński, Z. F. (1997) On a refined heat conduction theory for microperiodic layered solids, *Journal of Thermal Stresses* **20**(7), 749–771.
[12] Tomczyk, B. and Szczerba, P. (2018) A new asymptotic-tolerance model of dynamic and stability problems for longitudinally graded cylindrical shells, *Composite Structures* **202**, 473–481.
[13] Marczak, J. and Jędrysiak, J. (2018) Some remarks on modelling of vibrations of periodic sandwich structures with inert core, *Composite Structures* **202**, 752–758.
[14] Wirowski, A., Michalak, B. and Gajdzicki, M. (2015) Dynamic modelling of annular plates of functionally graded structure resting on elastic heterogeneous foundation with two modules, *Journal of Mechanics* **31**(5), 493–504.
[15] Świątek, Ma., Domagalski, Ł. and Jędrysiak, J. (2019) Free vibrations spectrum of periodically inhomogeneous Rayleigh beams using the tolerance averaging technique, *Journal of Theoretical and Applied Mechanics* **57**(1), 141–154.
[16] Pazera, E. and Jędrysiak, J. (2018) Effect of microstructure in thermoelasticity problems of functionally graded laminates, *Composite Structures* **202**, 296–303.
[17] Ostrowski, P., Thermoelasticity in a two-phase hollow cylinder with longitudinally graded material properties, *Shell Structures: Theory and Applications, Vol. 3*, Gdańsk, Poland, 2014, Pietraszkiewicz, W. and Górski, J., Eds, 133–136, CRC Press Taylor & Francis Group.
[18] Ostrowski, P. (2017) *Tolerance modelling of thermomechanics in microstructured media*, Lodz University of Technology Press, Łódź, Poland.

# Accuracy verification of a 2D adaptive mesh refinement method using backward facing step flows of low Reynolds numbers

**†Zhenquan Li[1], and Miao Li[2]**

[1]School of Computing and Mathematics, Charles Sturt University, Australia.
[2]CSU Engineering, Charles Sturt University, Australia.

*Presenting author: jali@csu.edu.au
†Corresponding author: jali@csu.edu.au

## Abstract

Identifying accurate centers of vortices of fluid flow is one of the accuracy measures for computational methods. After verifying the accuracy of the 2D adaptive mesh refinement (AMR) method by the benchmarks of 2D lid-driven cavity flows, this paper shows the accuracy verification by the benchmarks of 2D backward facing step flows. The AMR method refines a mesh using the numerical solutions of the Navier-Stokes equations calculated on the mesh by an open source software Navier2D which implemented a vertex centered finite volume method (FVM) using the median dual mesh to form control volumes about each vertex. The accuracy of the refined meshes is shown by the centers of vortices given in the benchmarks being held within the twice refined cells. The accuracy is also shown by the comparison between vortex center locations calculated from the linearly interpolated numerical solutions and those obtained in the benchmark. The AMR method is proposed based on the qualitative theory of differential equations, and it can be applied to refine a mesh as many times as required and used to seek accurate numerical solutions of the mathematical models including the continuity equation for incompressible fluid or steady-state fluid flow with low computational cost.

**Keywords:** adaptive mesh refinement, finite volume method, backward facing step flow

## Introduction

The AMR is a computational approach to increase the accuracy of numerical solutions of differential equations with low computational cost. A big number of papers on AMRs and their applications have been published [1]. Some common AMR methods take local truncation errors as a refinement criterion (e.g. Almgren et al. [2]). Other common AMR methods include h-refinement (e.g. Lohner [3]), p-refinement (e.g. Bell et al. [4]) or r-refinement (e.g. Miller et al. [5]), and different combinations of the above (e.g. Capon et al. [6]). These AMR methods aim to obtain a balance between the accuracy and the computational cost in finding numerical solutions of differential equations.

We introduced AMR methods for calculating accurate 2D (Li [7]) and 3D numerical velocity fields (Li [8]) based on a theory derived from Theorem 1.14 in the book by Ye et al. [9]. The theorem states that a 2D vector field has no one sided limit cycles if it satisfies the continuity equation. In other words, all trajectories of a vector field are closed curves in bounded domains if the vector field satisfies the continuity equation. A vector field which satisfies continuity equation is called divergence free field. The benchmarks (e.g. Erturk et al. [10]) confirm the theorem numerically. The AMR method refines a mesh based on the velocity fields calculated numerically on it. The refinement can be performed as many times as required. The more the

refinements, the less the area on which the linearly interpolated velocity field is not equivalent to a divergence free vector field is.

Locating singular points, and drawing accurate asymptotic lines (planes) and closed streamlines of a calculated velocity field are commonly used to measure the accuracy of computational methods. We demonstrated the accuracy of the AMR methods using examples of analytical velocity fields by comparing the exact results from the analytical velocity fields with the corresponding results from the numerical velocity fields that take vectors of the analytical velocity fields at nodes of the meshes: the singular points and asymptotic lines for 2D [11]; the singular points and asymptotic plane for 3D [12]; and closed streamlines (Li [11]-[12]). We also demonstrated the accuracy of the 2D AMR method using numerical velocity fields of 2D steady incompressible lid-driven cavity flows (Lal et al. [13]). We obtained numerical velocity fields of the Navier-Stokes equations with the boundary conditions using a second order collocated FVM (GSFV) with a splitting method for time discretization (Faure et al. [14]). We applied the AMR method once to the initial meshes based on the numerical velocity fields calculated by GSFV on them, and estimated the singular point locations using the centers of refined cells in the corresponding vortex regions. The estimated locations are accurate by comparing with the corresponding benchmarks.

Mesh refinement is necessary for calculating accurate numerical solutions since different levels of vortices requires different densities of mesh nodes (Li [15]). The same conclusion was obtained by Armaly et al. [16]. We conducted a study starting from relatively coarse initial meshes and demonstrating that the centers of vortices were held within the refined cells of once refined meshes (Li et al. [17]). We also verified the accuracy of the AMR method by applying the method twice to the initial meshes. Li et al. [1] applied the AMR method twice to the initial meshes and the twice refined meshes show that centres of the vortices are held within the twice refined cells. Li [18] considered flow past a square cylinder over symmetrical domain but the streamlines drawn on the initial mesh are not symmetrical. The symmetry of streamlines on the refined meshes are improved significantly after applying the AMR method once on the initial meshes.

This paper demonstrates the accuracy of the 2D AMR method proposed by Li [7] using the benchmarks for 2D backward facing step flows. The backward facing step flows have features of separation, reattachment, recirculation and shear layers in the flow region. We conduct study with constant boundary and initial conditions at the inlet channel and apply the AMR method twice to the initial meshes. We compare the profiles of the exact horizontal component of the velocity field and the profiles obtained numerically after the flow is well developed at the step [19]. Finally we show the differences between calculated locations of all detachment, reattachment and centres of vortices and the corresponding benchmarks [19].

## Governing equations

This paper considers the 2D incompressible or steady backward facing step flow. The governing equations are as follows:

$$\nabla \cdot \boldsymbol{V} = 0$$

$$\frac{\partial \boldsymbol{V}}{\partial t} + \boldsymbol{V} \cdot \nabla \boldsymbol{V} = \frac{1}{\rho} \nabla P + \nu \nabla^2 \boldsymbol{V}$$

where $\boldsymbol{V} = (u, v)$ is the velocity field, $\nu$ is the kinematic viscosity, and $P$ is the pressure. The governing equations are implemented in MATLAB, named Navier2D by Darren Engwirda [20], and it is used to calculate the velocity field $\boldsymbol{V}$ on triangular meshes numerically.

The computational domain is illustrated in the top figure of Figure 1. The height of the step is designated as $h$. We choose the length of the upstream channel as $250h$ since the initial conditions for both velocity components in this study are constants at the inlet so the $u$ profile can be developed well enough for long channel. The length of the downstream channels of the step is $50h$.

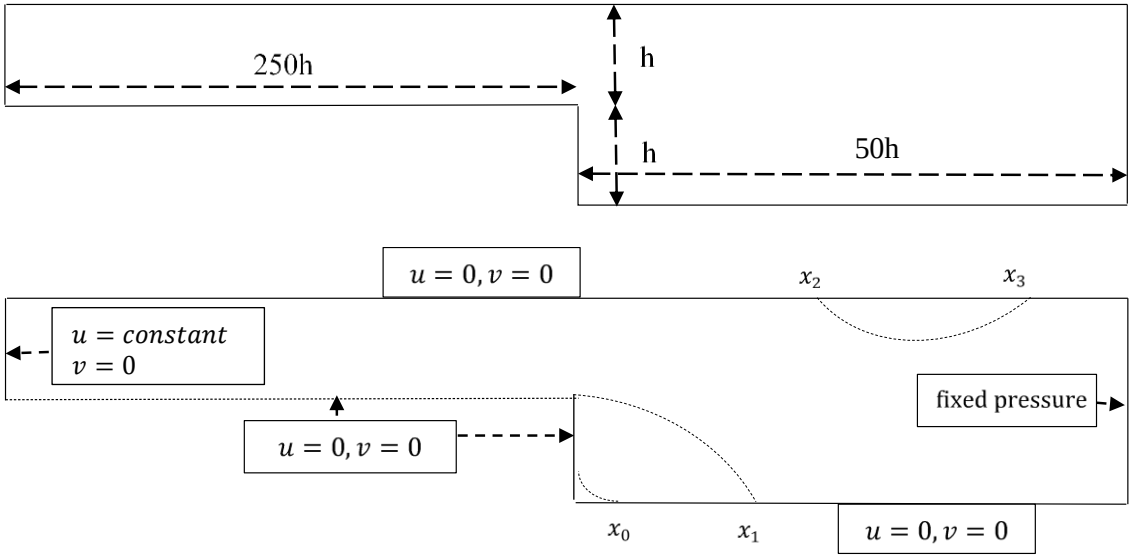



**Figure 1. Computational domain and boundary conditions**

The boundary conditions and the locations of detachments and reattachment are shown in the bottom figure of Figure 1. The detachments and reattachments $y_0$, $x_0$, $x_1$, $x_2$ and $x_3$ are shown in blue, red, cyan, magenta and green dots in the figures in Section 3. The locations of centers of vortices are shown by *. The computational domain is normalized using $h=1$. The outputs in terms of detachments, reattachments and locations of vortices are also normalized and compared with the results in Erturk [19].

## Review of AMR Method

This section summarizes the 2D AMR method proposed by Li [7] based on the theory developed from qualitative theory of differential equations [9].

Assume that $\boldsymbol{V}_l = \boldsymbol{AX} + \boldsymbol{b}$ is a vector field on a triangle obtained by linearly interpolating the vectors at the three vertexes of the triangle, where

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is a matrix of constants,

$$\boldsymbol{b} = \begin{pmatrix} b'_1 \\ b'_2 \end{pmatrix}$$

is a vector of constants and

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is a vector of spatial variables. The continuity equation for $\boldsymbol{V}_l$ and a steady flow or an incompressible fluid is

$$\nabla \cdot \boldsymbol{V}_l = trace(\boldsymbol{A}) = 0. \tag{1}$$

Let $f$ be a scalar function depending on spatial variables only. Substituting $f\boldsymbol{V}_l$ into the vector field $\boldsymbol{V}$ of the continuity equation $\nabla \cdot \boldsymbol{V} = 0$ obtains a differential equation. Solving the differential equation for $f$ for the four different Jacobian forms of the coefficient matrix $\boldsymbol{A}$ gives the expressions of $f$ as shown in Table 1. In Table 1, $(y_1, y_2)^T = \boldsymbol{V}^{-1}\boldsymbol{X}$ and $(b_1, b_2)^T = \boldsymbol{V}^{-1}\boldsymbol{b}$ where $\boldsymbol{V}$ satisfies $\boldsymbol{AV} = \boldsymbol{VJ}$ and $\boldsymbol{J}$ is one of the Jacobian matrices in Table 1. Vectors $\boldsymbol{V}_l$ and $f\boldsymbol{V}_l$ produce same streamlines if $f \neq 0, \infty$ (refer to Section 2.2 of [11]). The introduction of functions $f$ reduces the number of refined cells in refined meshes dramatically [21].

**Table 1. Jacobean matrices and corresponding expressions of $f$ ($C \neq 0$)**

| Case | Jacobean | $f$ |
|---|---|---|
| 1 | $\begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} (0 \neq r_1 \neq r_2 \neq 0)$ | $\dfrac{C}{\left(y_1 + \dfrac{b_1}{r_1}\right)\left(y_2 + \dfrac{b_2}{r_2}\right)}$ |
| 2 | $\begin{pmatrix} r_1 & 0 \\ 0 & 0 \end{pmatrix} (r_1 \neq 0)$ | $\dfrac{C}{y_1 + \dfrac{b_1}{r_1}}$ |
| 3 | $\begin{pmatrix} r & 1 \\ 0 & r \end{pmatrix} (r \neq 0)$ | $\dfrac{C}{\left(y_2 + \dfrac{b_2}{r}\right)^2}$ |
| 4 | $\begin{pmatrix} \mu & \lambda \\ -\lambda & \mu \end{pmatrix} (\mu \neq 0, \lambda \neq 0)$ | $\dfrac{C}{\left(y_1 + \dfrac{\mu b_1 - \lambda b_2}{\mu^2 + \lambda^2}\right)^2 + \left(y_2 + \dfrac{\lambda b_1 + \mu b_2}{\mu^2 + \lambda^2}\right)^2}$ |

The conditions (MC)(MC is the abbreviation of mass conservation) are the functions $f$ in Table 1 not equaling zero or infinity at any point on the triangular domains.

We review the algorithm of AMR method for quadrilateral meshes [17]. The algorithm can also be used to a triangular mesh after a subdivision scheme for a triangle is defined. We describe the algorithm of AMR method into two parts:

- cell refinement algorithm - describes how to use the conditions (MC) to refine a quadrilateral cell in a given mesh.
- the algorithm of AMR method.

The AMR is an infinite process. To avoid an infinite refinement of a mesh, we choose a pre-specified threshold number of refinements $T$ based on the accuracy requirements. The algorithm of cell refinement is:

**Step 1** Subdivide a quadrilateral cell into two triangles. If $\boldsymbol{V}_l$ satisfies Eq. (1) on both triangles, no refinement for the cell is required. Otherwise, go to Step 2;

**Step 2** Apply the conditions (MC) to both of the triangles. If the conditions (MC) are satisfied on both triangles, no refinement for the cell is required. Otherwise, we subdivide the cell into a number of small cells such that the lengths of all sides of the small cells are truly reduced (e.g. connecting the mid-points of opposite sides of a quadrilateral by line segments produces four small quadrilaterals and the lengths of the sides of the four small quadrilaterals are truly reduced).

The algorithm of AMR method is:

**Step 1** Evaluate the numerical velocity field for a given initial mesh;

**Step 2** Refine all cells of the initial mesh one by one using the above algorithm of cell refinement;

**Step 3** Take the refined mesh as initial mesh and go to Step 1 until a satisfactory numerical velocity field is obtained or the threshold number $T$ is reached.

In this paper, we subdivide a quadrilateral cell by connecting the mid-points of two opposite sides of a quadrilateral [1] and set $T = 2$, that is, we subdivide the cells on which one of the MC conditions is satisfied at most twice.

**Accuracy Verifications by Comparison with Benchmarks**

The accuracy is examined by using numerical velocity fields with the residuals less than $10^{-8}$ for both $u$ and $v$. The initial mesh has a step size of 0.1 in both $x$ and $y$ directions. The initial velocity field is $u = v = 0$. The accuracy of the AMR method depends fully on the accuracy of the numerical velocity fields calculated by Navier2D. The profiles of $u$ of the numerical velocity fields show the accuracies of the calculated fields (refer to Figs. 2, 3, 5, 6, 8, 9, 11, and 12). The refined meshes show the accuracy of the AMR method.

The Reynolds number is defined as $Re = \frac{UD}{\upsilon}$ where $U$ is the inlet mean velocity or in other words two thirds of the maximum inlet horizontal component of the velocity field and the $D$ is the hydraulic diameter of the inlet channel which is equivalent to twice the inlet channel height $h$ [19]. However, as you will see, the Reynolds numbers $Re$ are not same using the above two definitions for each of the cases considered in this section. Therefore, the $Re$ in the following subsections are estimates. In this study, we make the maximum horizontal component of the numerical velocity fields at the step approximately 1.5 by choosing appropriate boundary condition at the inlet. We set different CFL numbers in Navier2D for different Reynolds numbers. The CFL condition is a necessary condition for convergence and stability of a numerical method. We choose suitable CFL numbers to obtain reliable numerical solutions.

$Re = 100$

Figs. 2 and 3 show the comparisons between the profile of the horizontal component $u$ of the exact analytical velocity and the calculated ones on the initial and once refined meshes at the step. From these figures, we understand the differences between the exact $Re$ and calculated ones. The difference between the profiles of analytical $u$ and the calculated one on the initial mesh is smaller than that on the once refined mesh. However, the profile on the once refined mesh has better agreement with the profile generated using 500 uniform cells on the inlet

channel upstream of the step with length $20h$ and width $h$, and the analytical $u$ profile as the inlet boundary condition (Figure 3 [19]). The $u$ profiles for low Reynolds numbers are slightly different from the corresponding exact analytical profiles [19].

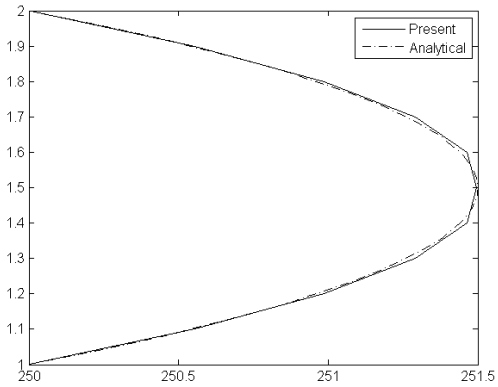



**Figure 2. Comparison of the profiles of $u$ at the step for $Re = 100$ between the current study and the analytical solutions on initial mesh**

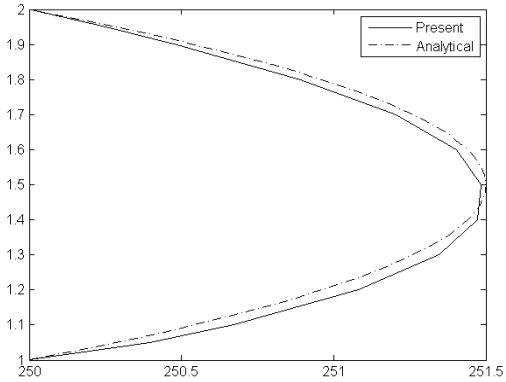



**Figure 3. Comparison of the profiles of $u$ at the step for $Re = 100$ between the current study and the analytical solutions on once refined mesh**

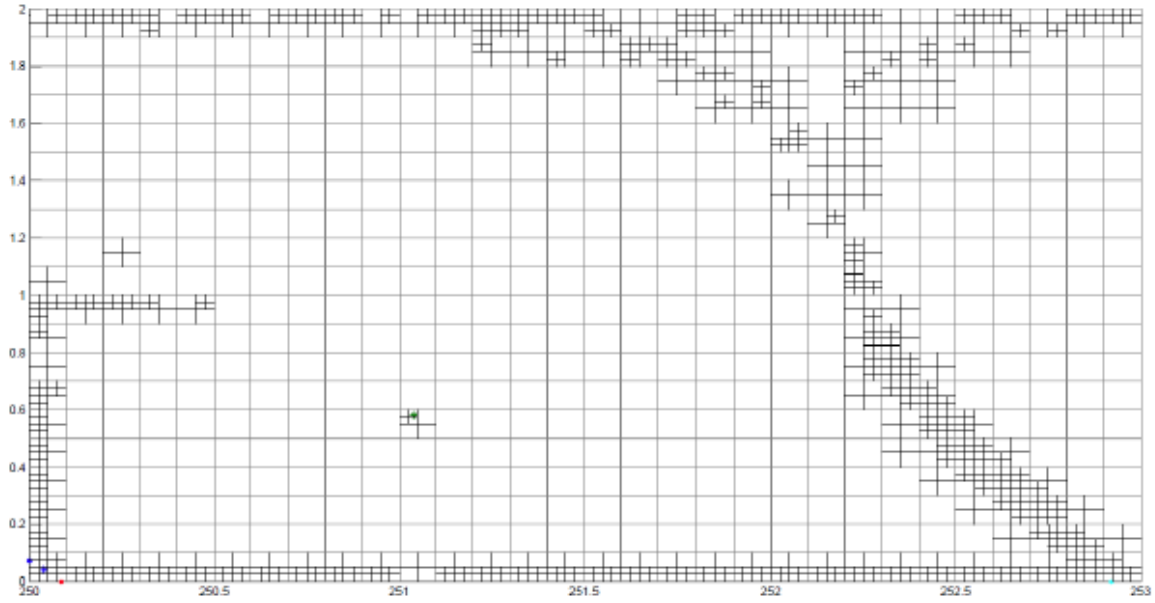



**Figure 4. Twice refined mesh for $Re = 100$ with locations of detachment, reattachments and vortex centre**

The once refined and twice refined meshes are shown in Figure 4. The locations of all detachment, reattachment and centers of vortices are held within the twice refined cells.

$Re = 200$

The $u$ profiles for $Re = 200$ are shown in Figs. 5 and 6. They present similar patterns to those for $Re = 100$ in Figs. 2 and 3 but bigger differences between the analytical and calculated $u$ profiles at the step for both cases.

The locations of all detachment, reattachment and centers of vortices are held within the twice refined cells as shown in Figure 7.

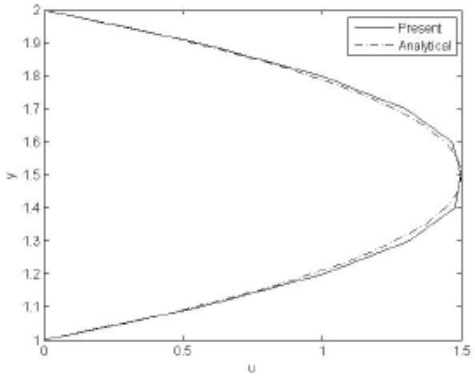



**Figure 5. Comparison of the profiles of $u$ at the step for $Re = 200$ between the current study and the analytical solutions on initial mesh**

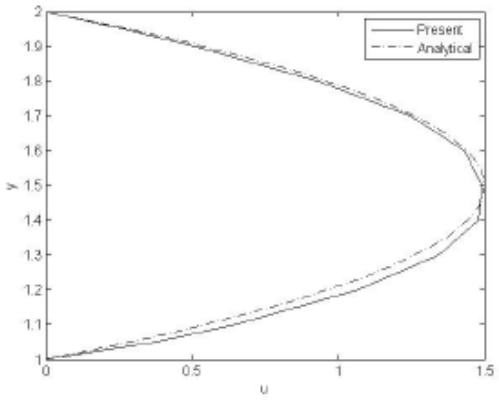



**Figure 6. Comparison of the profiles of $u$ at the step for $Re = 100$ between the current study and the analytical solutions on once refined mesh**

$Re = 400$

The $u$ profiles for $Re = 400$ are shown in Figs. 8 and 9. Once again, similar patterns to those for $Re = 100$ are obtained with even bigger differences between the analytical and calculated $u$ profiles for both cases.

The locations of all detachment, reattachment and centres of vortices below the middle line of the channel and $x_2$ are held within the twice refined cells but the locations of $x_3$ and the centre of vortex $x_2$-$x_3$ are held within the once refined cells only as shown in Figure 10. Finer initial mesh or more accurate computational velocity field is required for $Re = 400$ from the $u$ profiles.

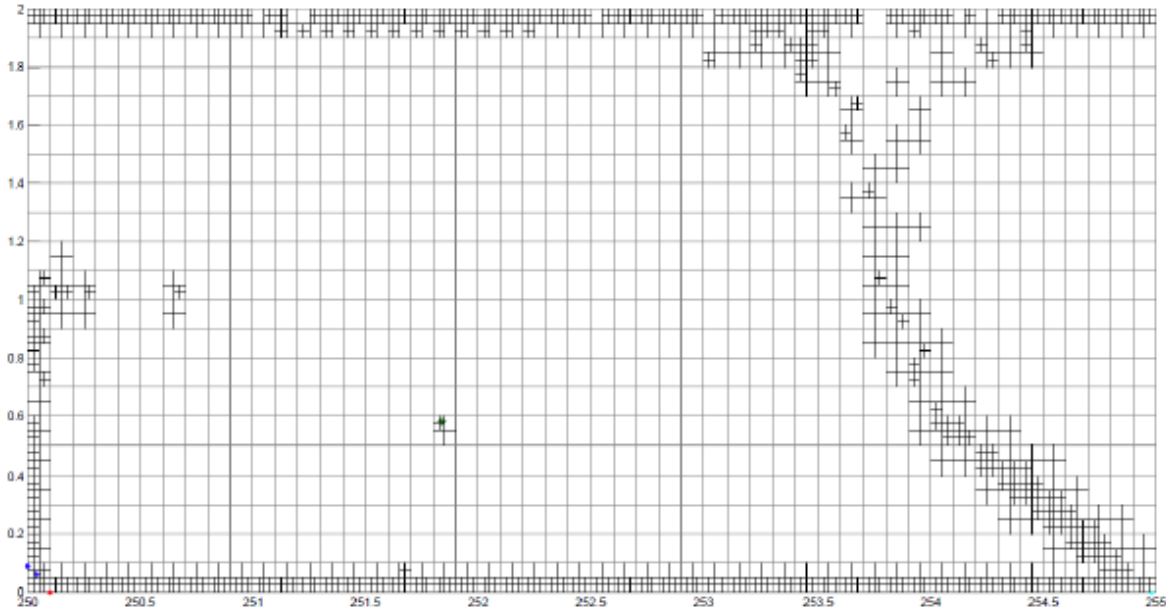


**Figure 7. Twice refined mesh for $Re = 200$ with locations of detachment, reattachments and vortex centre**

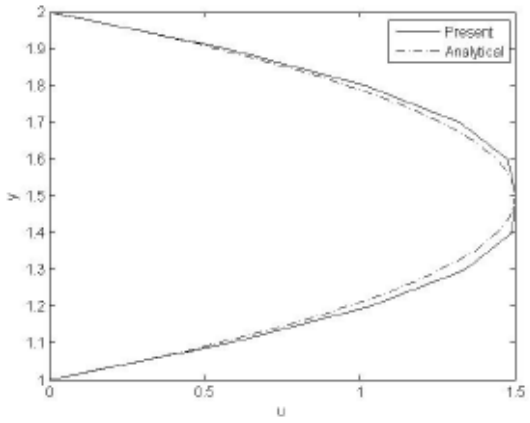



**Figure 8. Comparison of the profiles of $u$ at the step for $Re = 400$ between the current study and the analytical solutions on initial mesh**

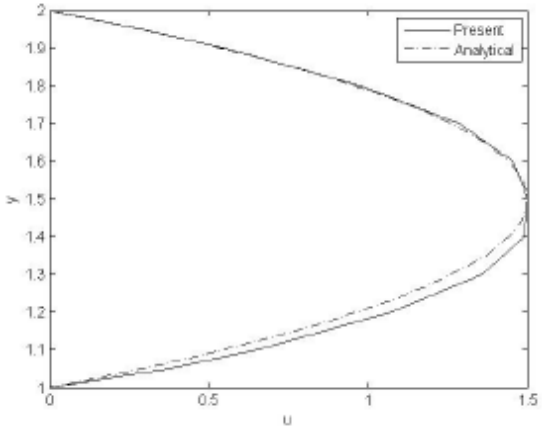



**Figure 9. Comparison of the profiles of $u$ at the step for $Re = 400$ between the current study and the analytical solutions on once refined mesh**

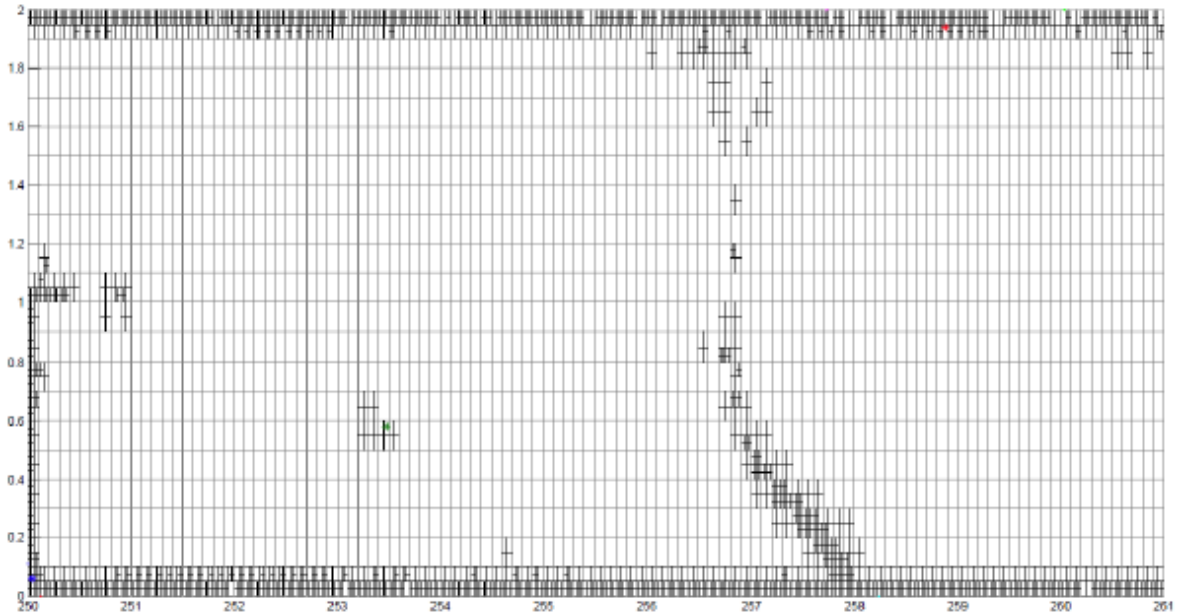


**Figure 10. Twice refined mesh for $Re = 400$ with locations of detachment, reattachments and vortex centre**

$Re = 800$

Figs. 11 and 12 show the largest differences between the analytical and calculated $u$ profiles for all Reynolds numbers considered. The differences may be not acceptable in practice. Finer initial mesh or more accurate computational velocity field is required for better outcomes.

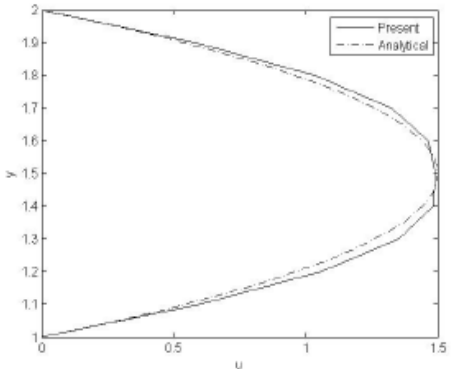



**Figure 11. Comparison of the profiles of $u$ at the step for $Re = 800$ between the current study and the analytical solutions on initial mesh**

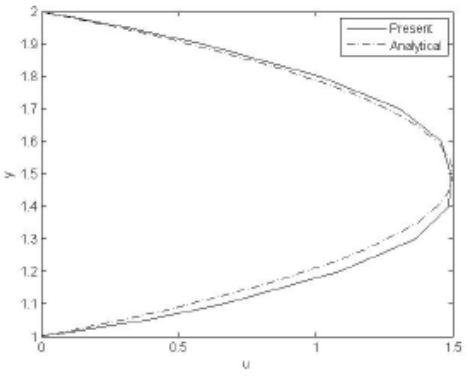



**Figure 12. Comparison of the profiles of $u$ at the step for $Re = 800$ between the current study and the analytical solutions on once refined mesh**

The locations of $y_0$, $x_1$, and the centre of primary vortex are held within the twice refined cells but $x_0$ and center of vortex $y_0$-$x_0$ are held within the once refined cells as shown in Figure 13. The locations of $x_2$ and $x_3$ are held within the twice refined cells but the center of vortex $x_2$-$x_3$ is held within a cell without any refinements as shown in Figure 14.

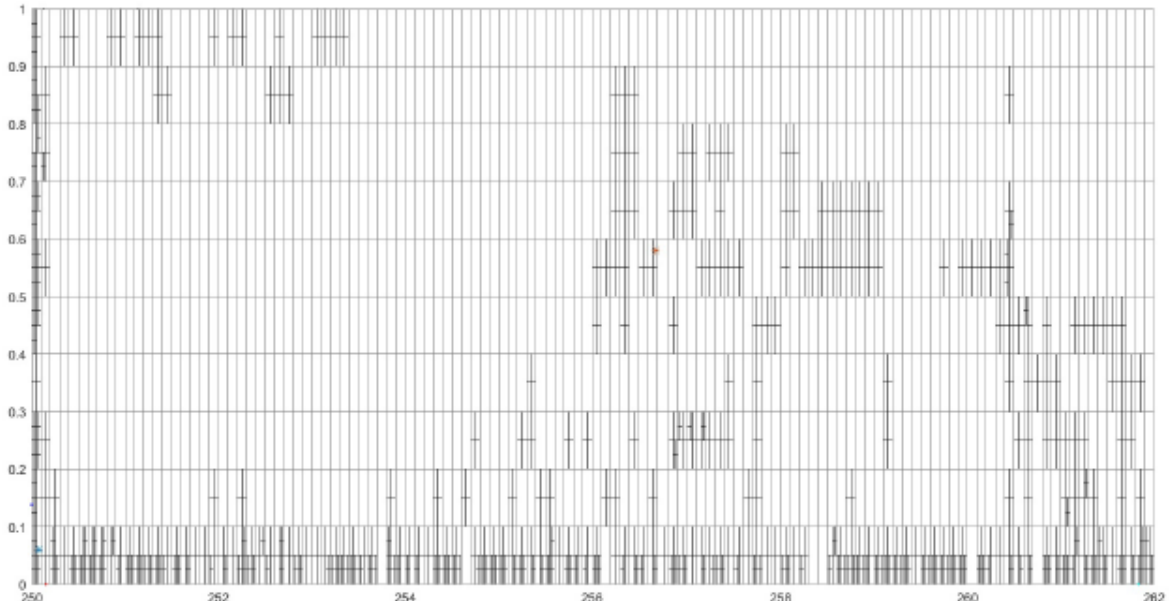



**Figure 13. Twice refined mesh for $Re = 800$ with locations of detachment, reattachment and vortex centre for $y$ in [0, 1]**

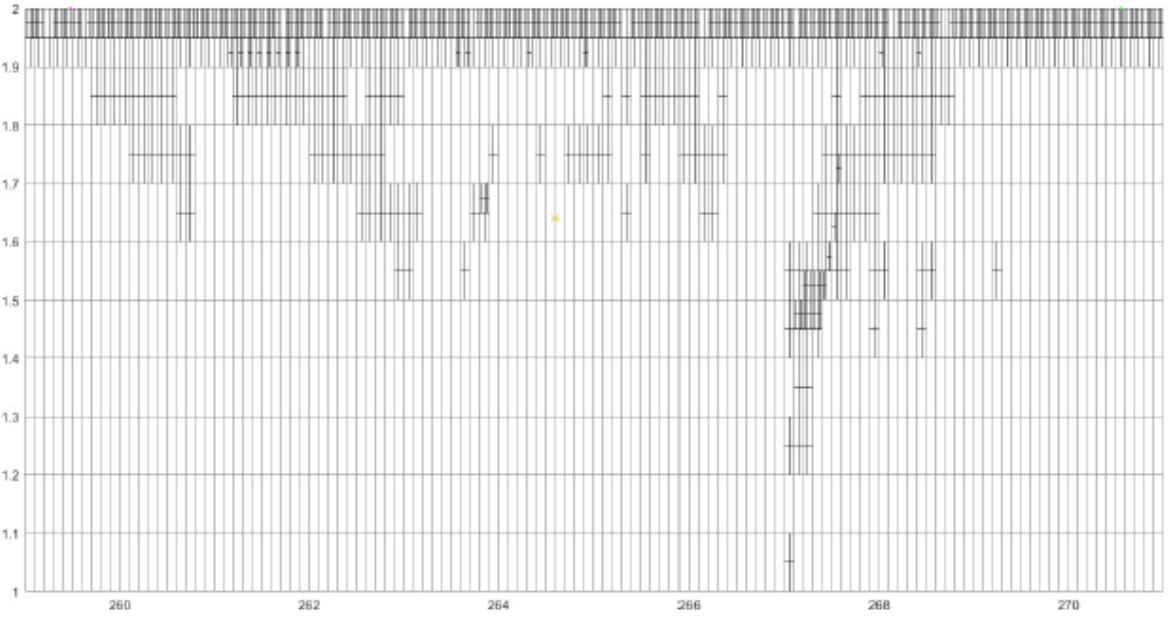



**Figure 14. Twice refined mesh for $Re = 800$ with locations of detachment, reattachment and vortex centre for $y$ in [1, 2]**

### Estimated Locations of the Centres of Vortices

Table 2 shows the coordinates of vortex centers from the benchmark [20] and the corresponding coordinates for $Re = 100$, $200$, $400$, and $800$ calculated from the linearly interpolated velocity fields $\boldsymbol{V}_l$ on the initial and once refined meshes. The first row in the table are the

estimated centres using initial mesh, the second row are the estimated centers using the once refined mesh and the last row are benchmarks for each category. There are a number of singularities around point (13.8705, 1.6494) in the MATLAB outputs for vortex $x_2$-$x_3$ and case $Re = 800$ on the once refined mesh with one example illustrated in Table 2 below.

**Table 2. Estimated locations of the centers of vortices**

| Vortex type | Reynolds numbers | | | |
|---|---|---|---|---|
| | Re = 100 | Re = 200 | Re = 400 | Re = 800 |
| $y_0$–$x_0$ | - | - | - | - |
| | - | (0.0500, 0.0000) | (0.0525, 0.0587) | (0.0651, 0.0649) |
| | (0.04,0.04) | (0.04, 0.06) | (0.04, 0.06) | (0.08, 0.06) |
| $x_0$–$x_1$ | (1.0404, 0.5759) | (1.8413,0.5885) | (3.4877,0.5927) | (6.6667,0.5875) |
| | (1.0466, 0.5731) | (1.8440,0.5824) | (3.5101, 0.5814) | (6.4437,0.5834) |
| | (1.04,0.58) | (1.84, 0.58) | (3.48, 0.58) | (6.68, 0.58) |
| $x_2$–$x_3$ | - | - | - | (13.8743,1.6537) |
| | - | - | (8.5595, 1.9490) | (13.8705,1.6494) |
| | - | - | (8.88,1.94) | (14.60, 1.64) |

## Conclusions

We applied the AMR method twice to the initial meshes based on the numerical solutions of 2D backward facing step flow produced by Navier2D. The accuracy of the AMR method shown is sufficient for $Re = 100$ and $200$, and is acceptable for $Re = 400$ and $800$ based on the differences between the profiles of the horizontal component of the velocity fields at the step. The conclusion is that one setting for the domain, initial mesh and the residual error cannot achieve similar accurate numerical velocity fields for different Reynolds numbers. Therefore, the further research topics may include:

- Use longer lengths of upstream and downstream channels of the step for bigger Reynolds numbers.
- Use finer initial meshes when Reynolds number increases.
- Calculate more accurate numerical velocity fields when Reynolds number increases.

### References


[1] Li, Z., and Wood, R. (2017) Accuracy verification of a 2D adaptive mesh refinement method for incompressible or steady flow, *J. Comput. Appl. Math.* **318**, 259--265.

[2] Li, Z. (2017) Computational complexity of the algoriithm for a 2D adaptive mesh refinement method using li-driven cavity flows, *Comput. Therm. Sci.* **9**, 395--403.

[3] Almgren, A., Bell, J., Colella, P., Howell, L. and Welcome, M. (1998) Conservative adaptive projection method for the variable density incompressible Navier--Stokes equations, *J. Comput. Phys.* **142**, 1--46.

[4] Lohner, R. (1987) An adaptive finite element scheme for transient problems in CFD, *Comput. Methods in Appl. Mech. Eng.* **61**, 323--338.

[5] Bell, J., Berger, B., Saltzman, J., and Welcome, M. (1994) Three dimensional adaptive mesh refinement for hyperbolic conservation laws, *SIAM J. Sci. Comput.* **15**, 127--138.

[6] Miller, K. and Miller, R. (1981) Moving finite elements, Part I, *SIAM J. Numer. Anal.* **18**, 1019--1032.

[7] Capon, P. and Jimack, P., An adaptive finite element method for the compressible Navier--Stokes equations, *Numerical Methods for Fluid Dynamics*, 1995, K. Morton and M. Baines, Eds, 327--333. Clarendon Press.

[8] Li, Z. (2008) An adaptive two-dimensional mesh refinement method based on the law of mass conservation, *J. Flow Visual. Image Process* **15**, 17--33.

[9] Li, Z. (2007) An adaptive three-dimensional mesh refinement method based on the law of mass conservation, *J. Flow Visual. Image Process* **14**, 375--395.

[10] Ye, Y. and Others (1986) Theory of Limit Cycles, Providence, Rhode Island: American Mathematical Society Press.

[11] Erturk, E., Corke, T.C. and Gökcöl (2005) Numerical solutions of 2-D steady incompressible driven cavity flow at high Reynolds numbers, *Int. J. Numer. Methods Fluids* **48**, 747--774.
[12] Li, Z. (2006) An adaptive streamline tracking method for two-dimensional CFD velocity fields based on the law of mass conservation, *J. Flow Visual. Image Process.* **13**, 1--14.
[13] Li, Z. (2006) An adaptive streamline tracking method for three-dimensional CFD velocity fields based on the law of mass conservation, *J. Flow Visual. Image Process.* **13**, 359--376.
[14] Lal, R. and Li, Z. (2015) Sensitivity analysis of a mesh refinement method using the numerical solutions of 2-D steady incompressible driven cavity flow, *J. Math. Chem.* **53**, 844--867.
[15] Faure, S., Laminie, J. and Temam, R. (2008) Colocated finite volume schemes for fluid flows, *Commun. Comput. Phys.* **4**, 1--25.
[16] Li, Z. (2014) Accuracy analysis of a mesh refinement method using benchmarks of 2-D lid-driven cavity flows and finer meshes, *J. Math. Chem.* **52**, 1156--1170.
[17] Armaly, B., Durst, F., Pereir, J. and Schönung, B. (1983) Experimental and theoretical investigation of backward-facing step flow, *J. Fluid Mech*.**127**, 473--496.
[18] Li, Z. and Wood, R. (2015) Accuracy analysis of an adaptive mesh refinement method using benchmarks of 2-D steady incompressible lid-driven cavity flows and coarser meshes, *J. Comput. Appl. Math*. **275**, 262--271.
[19] Li, Z. (2017) Analysis of 2D unsteady flow past a square cylinder at low Reynolds numbers with CFD and a mesh refinement method, *WSEAS Trans. on Fluid Mech*. **12**, 150--157.
[20] Erturk, E. (2008) Numerical solutions of 2-D steady incompressible flow over a backward-facing step, Part I: High Reynolds number solutions, *Comput. Fluids* **37**, 633--655.
[21] Engwirda, D. (2006) Navier-Stokes Solver (Navier2d), MATLAB Central File Exchange. Last accessed May 2006.
[22] Li, Z. and Mallinson, G., Simplification of an existing mass conservative streamline tracking method for two-dimensional CFD velocity fields, *GIS and Remote Sensing in Hydrology, Water Resources and Environment*, 2004, Y. Chen, K. Takara, Cluckies I, F.H. DS., Eds, 269—275, IAHS Press, Wallingford, UK.
[23] Li, Z. (2002) A mass conservative streamline tracking method for two-dimensional CFD velocity fields, *J. Flow Visual. Image Process.* **9**, 75--87.

# Optimization Research of Fertilizer Guiding Mechanism Based on the Discrete Element Method

**†Hongjian Zhang[2], Yufeng Li[2], *Jinxing Wang[1,2], Kaixing Zhang[1,2], Shuangxi Liu[1,2], Zhen Wang[1,2], Guikun Chu[2], Zekun Quan[2]**

[1]Shandong Provincial Key Laboratory of Horticultural Machineries and Equipments, Shandong Agricultural University, China

[2]College of Mechanical and Engineering, Shandong Agricultural University, China

*Presenting author: zhanghongji_an@163.com

†Corresponding author: jinxingw@163.com

## Abstract

In view of the shortcomings of low uniformity and high discrete of fertilization in orchard, the Discrete Element Method was used to study the fertilization mechanism of the orchard ditching fertilizer applicator. Firstly, on the basis of studying the existing linear fertilizer guiding mechanism, a concave and convex fertilizer guiding mechanism was proposed, and a virtual simulation model of three fertilizer guiding mechanisms was established. Secondly, the basic parameters of granular organic fertilizers were analyzed, and the kinetic model was established by combining the kinetic analysis during the falling process. Finally, the uniformity of fertilization was evaluated by discrete coefficient and was set as the target. The discrete element simulation optimization experiments were carried out on the concave, linear, convex and different curvature radius of the fertilizer guiding mechanism, and the optimal concave fertilizer guiding mechanism and its optimal radius of curvature were optimized. Simulation and field experiments show that among the three kinds of fertilizer guiding mechanisms, the discrete coefficient of the convex fertilizer guiding mechanism is the smallest and the uniformity of fertilization is the highest; in the convex fertilizer guiding mechanism with different curvature radius, the coefficient of discrete is the smallest and the uniformity of fertilization is the highest when the radius of curvature is 600mm; after optimization, the discrete coefficient was reduced from 0.51 to 0.26, and the uniformity of fertilization was increased by 49.02%. This provides scientific basis for design optimization of orchard ditching fertilizer applicator.

**Keywords:** Fertilizer Guiding Mechanism; DEM; Simulation optimization; Discrete coefficient; Uniformity

## Introduction

Fertilizer is the grain of fruit trees. Fertilization of fruit trees is the key operation link in fruit tree production. The quality of fertilization directly affects the absorption of nutrients in fruit

trees. Rational fertilization is the basis for high quality and high yield of fruit trees [1]-[3]. Fruit trees fertilizer predominantly are inorganic fertilizer and organic fertilizer: Inorganic fertilizer is mainly chemical fertilizer, characterized by strong fertility and fast fertilizer efficiency, but it is easy to cause soil structure change, soil organic matter content decline, tree growth and fruit quality reduction; application of organic fertilizer can help improve soil physicochemical properties, improve leaf physiology, increase fruit yield and improve fruit quality [4]-[6]. Fertilization uniformity is a key factor affecting the effect of fertilization operations, and is an important indicator for evaluating the performance and effectiveness of work tools [7]-[10]. Patterson [11] and the like studied the theoretical motion model and motion mechanism of various fertilizer granules, simulated and tested the fertilizer granule distribution and fertilization uniformity, and analyzed the average error between the theoretical value and the actual measured value, which laid the foundation for the mathematical theoretical model of the fertilizer application machine. P. Van Liedekerke [12] and the like used the discrete element method to start the flow of granules from the container, and the plate and the inclined disk are used to simulate the flow of granules affected by the rotating disk. Artur Przywara [13] and the like studied the influence of the structure and operating parameters of the centrifugal disc spreader on the spatial distribution of the fertilizer, determined the rotational speed of the disc, the feed position of the fertilizer on the disc, the blade angle on the disc and the type of fertilizer has an effect on the spatial distribution of fertilizer. Villette [14]and the like proposed a new method for simultaneously measuring the horizontal flow and vertical flow distribution of a disc-type fertilizer, paving the way for studying the distribution of fertilizer granules and testing the uniformity of fertilization. Aphale [15] and the like calculated and tested the trajectories of various fertilizer granules, and obtained the average error between the theoretical value and the actual measured value under different disk rotational speed conditions, which provided a theoretical basis for the test. Zhang Tao [16] and so on simulated the fertilization process of the fertilizer-discharging mechanism through discrete elements, formulated design variables and optimization targets, developed motion simulation software and parameterized platform, optimized the structural parameters of the fertilizer-discharging mechanism, and improved the uniformity of fertilization. Chen Xiongfei [17] and so on designed a two-stage spiral fertilizer device, and established a mathematical model of the single-circle discharge amount of the fertilizer-fertilizing spiral. It was determined through experiments that the two-stage spiral fertilizer-removing device had better fertilizer-removing effects for various forms and better adaptability with fertilizer. Yang Xinlun [18] and so on established the mathematical model of the blade type fertilizer mechanism, formulated the design variables and optimization objectives, developed the motion simulation software and the parameterization platform, and optimized the structural parameters of the blade type fertilizer removal mechanism. Yuan Wensheng [19] and so on designed a scoop wheel fertilizer discharge device and established the three-dimensional model for simulation of fertilizer, tested cavitation resistance and uniformity of the amount of fertilizer per hole fertilizer apparatus, and studied the fertilizer discharge effect of the fertilizer discharger under different rotation speeds of the fertilizer components. Lv Jinqing [20] and so on carried out simulation tests on the screw-type fertilizer discharge mechanism under different working speed conditions, accurately analyzed the mathematical relationship between the rotational speed and the displacement, realized the optimal design of the screw-type fertilizer discharge

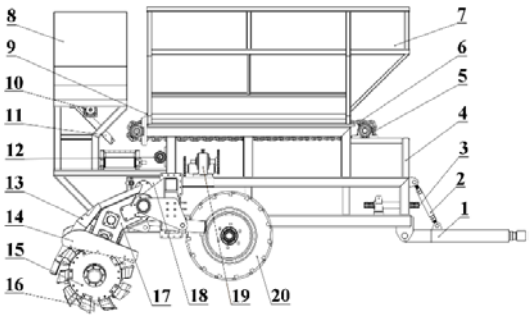
mechanism, and improved the uniformity of fertilization.

After comprehensively analyzing the research status at home and abroad, and the effect of fertilizer-discharging mechanism on fertilization uniformity, a conclusion has been made that, at present, the research on the effect of fertilizer-dispensing institutions on fertilization uniformity is mature, but the research on the influence of fertilizer guiding mechanism on fertilization uniformity is still in its infancy. Based on the study of the influence of the fertilizer guiding mechanism on the uniformity of fertilization, this paper optimizes the existing fertilizer guiding mechanism and provides a theoretical basis for the design and processing of the orchard ditching fertilizer machine.

## Orchard Double-row Ditching Fertilizer Structure Design

*Machine Structure*

Double-row orchards ditching fertilizer overall structure is shown in Fig. 1 and Fig. 2, mainly configured by the rack, fertilizer box, transmission, ditching mechanism, mechanism of fertilizer, fertilizer guiding mechanism, and soil covering machine, and the technical parameters of the whole machine are shown in Table 1.



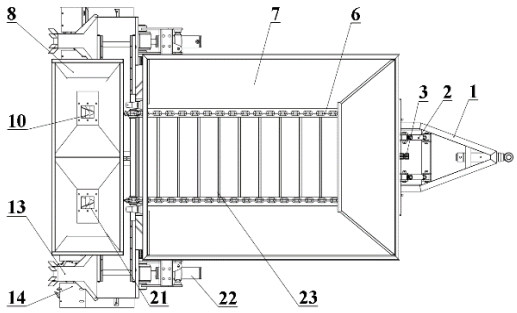




**Figure 1. Main view of the machine Figure 2. Vertical view of the machine**

1 traction rack 2 adjust pull pipe 3 drive shaft 4 rack 5 base fertilizer sprocket 6 O-chain 7 base fertilizer box 8 fertilizer box 9 base fertilizer outlet 10 auger 11 conveyor board 12 side transmission box 13 fertilizer guiding mechanism 14soil cover 15 trench cutter disc 16 trench cutter 17 trench transmission box 18 main transmission box 19 middle transmission box 20 wheel 21 fertilizer outlet 22 hydraulic cylinder 23 base fertilizer scraper

**Table 1. Main technical parameters**

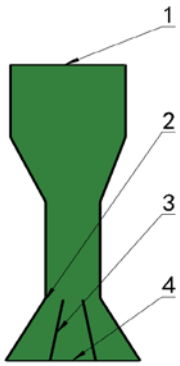
| Item | Parameters | Units |
|---|---|---|
| Supporting power | ≥58 | kw |
| Outline size | 3820×2040×2510 | mm |
| Fertilization depth | 0~50 | mm |
| Fertilizer application amount | 0.5~2 | $kg.m^{-1}$ |
| Base fertilizer application amount | 3~5 | $kg.m^{-1}$ |

| | | |
|---|---|---|
| Fertilizer box volume | 650 | L |
| Base fertilizer box volume | 2500 | L |

*Fertilizer Guiding Mechanism*

The fertilization methods in the orchard mainly include the general application of the whole garden, the application of the crown, the application of the ring ditch, the application of the strip ditch, the application of the radial ditch and the application of the hole. For new orchards with wide line spacing and wide plant spacing, strip-shaped furrow fertilization is often used. The strip-shaped furrow fertilization method requires strips of 1000~2000mm long, 300~400mm wide and 300~500mm deep between the rows of fruit trees, then fertilize and cover the soil [21]-[22].

Because the amount of fertilization in the orchard is large, in order to ensure the smooth fall of the fertilizer, the fertilizer guiding mechanism of the existing orchard trenching and fertilizing machine is mostly a trough-shaped structure, as shown in Fig. 3. Among them, the fertilizer-in port of the fertilizer guiding mechanism is connected with the fertilizer-out port of the fertilizer-discharging mechanism, and the fertilizer-out port of the fertilizer guiding mechanism is equal to the groove width. The fertilizer plug-in plate in the fertilizer guiding mechanism divides the fertilizer-transfer port and the fertilizer-out port into three parts.




1 Fertilizer-in port 2 Fertilizer-transfer port 3 Fertilizer plug-in plate 4 Fertilizer-out port

**Figure 3. Diagrammatic sketch of fertilizer guiding mechanism**

**Fertilizer Granules Motion Analysis**

*Determination of Parameters of Fertilizer Granules Characteristics*

The fertilizer granules used in the experiment were granular organic fertilizers. 100 granular organic fertilizers were randomly selected as the parameter measurement samples, which were placed on blank A4 paper and measured by image processing technology to obtain the accurate size of the granular organic fertilizer.

First, a black square of 10mm×10mm was fabricated on the red A4 paper as an area calibration, and blank A4 paper with area calibration was used as the background plate. Next, the parameter measurement was randomly scattered on the background board, and the original image of granular organic fertilizers was obtained. The color difference between the test sample and the background plate was used, and a single original image was divided to obtain the threshold value of the other image, as shown in Fig. 4. Finally, according to the area calibration on the background plate in the binary image, the equivalent diameter and circularity of the granular organic fertilizer were calculated, and the minimum equivalent diameter, the minimum circularity, the maximum equivalent diameter and the maximum circularity, the average equivalent diameter and the average circularity of the statistical parameter measurement sample were calculated, as shown in Table 2. The granular organic fertilizer circularity distribution diagram was drawn, as shown in Fig. 5.

$$\text{Circularity } \varnothing = 4\pi S / C^2 \tag{1}$$

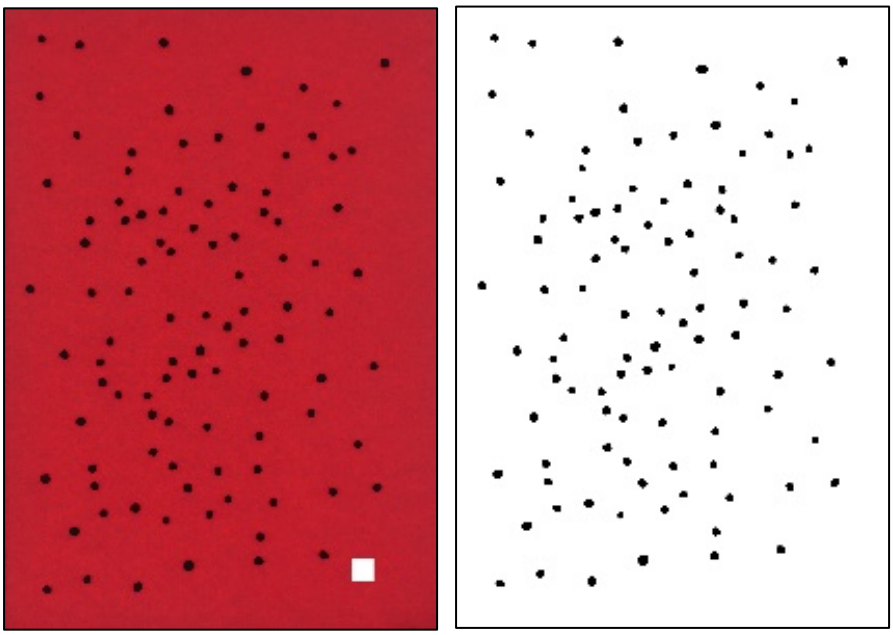
$$\text{Equivalent diameter } d = C/\pi \tag{2}$$

$$\text{The average diameter } \bar{D} = D/N \tag{3}$$

$$D = \sum_{i=1}^{N} d_i \tag{4}$$

In equations (1), (2), (3), (4):

*S*-the area of granular organic fertilizer; *C*- the circumference of granular organic fertilizer; *N*-the amount of granular organic fertilizer



(a) Original image (b) Binary image

**Figure 4. Original image and binary image of granular organic fertilizer**

**Table 2. Parameter of granular organic fertilizer**

| Granular organic fertilizer | Circumference $C$(mm) | Equivalent diameter $\overline{D}$(mm) | Circularity $Ø$ |
|---|---|---|---|
| Average value | 9.16 | 2.92 | 0.93 |
| Minimum value | 7.60 | 2.42 | 0.90 |
| Maximum value | 11.92 | 3.79 | 0.97 |

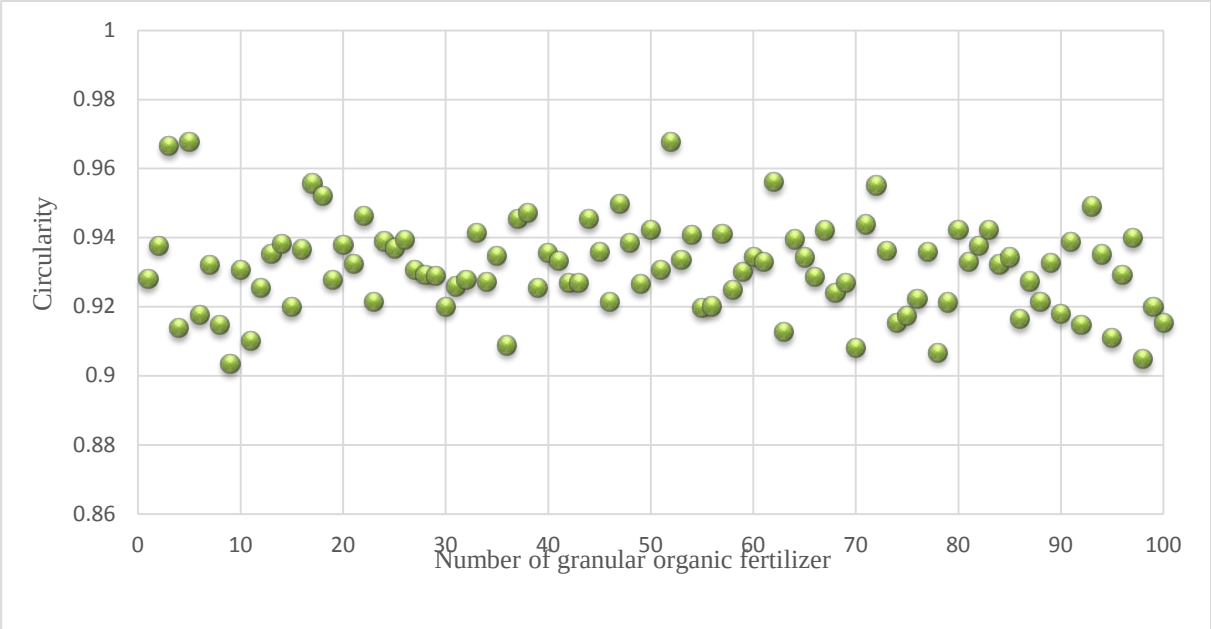



**Figure 5. Circularity distribution of granular organic fertilizer**

It can be seen from Table 2 and Fig. 4 that the circularity of the 100-parameter measurement samples is 0.93, and the concentration is between 0.90 and 0.97. As the circularity of parameter measurement sample gets closer to 1, the closer the parameter measurement sample is to the sphere [23]. Since the parameter measurement sample has a high circularity distribution, the granular organic fertilizer is assumed to be spherical during the simulation.

*The Equation of Motion of Fertilizer Granules in the Fertilizer Guiding Mechanism*

Granular organic fertilizers fall into the fertilizer guiding mechanism by means of fertilizer discharge mechanism, with a certain initial velocity $v_p$. After the granular organic fertilizer enters the fertilizer guiding mechanism, it is subjected to gravity $G_p$, Buoyancy $F_{fp}$, and air resistance $F_{zp}$. After the interaction, it finally falls into the ditch of the orchard trenching and fertilizing machine. The specific process is shown in the Fig. 6 [24].

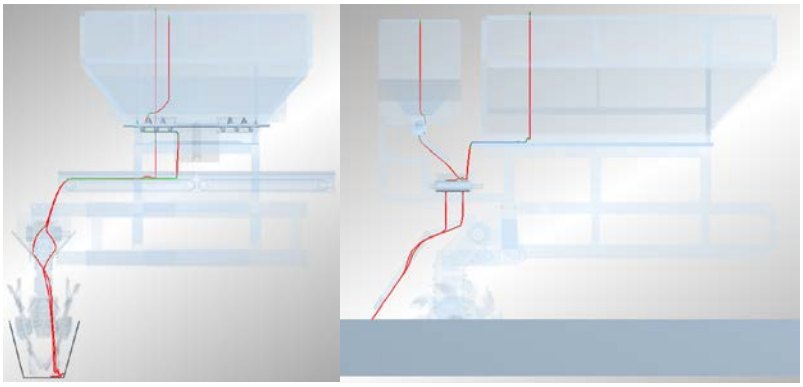


**Figure 6. Falling process of fertilizer**

$$\text{Gravity } G_p = \rho_p V_p; \tag{5}$$

$$\text{Buoyancy } F_{fp} = \rho_a V_p g; \tag{6}$$

$$\text{Air resistance } F_{zp} = \frac{1}{2} K \rho_p S_p V_p^{\,2}; \tag{7}$$

$$\text{Air resistance coefficient } K = \frac{3}{8} C_D \rho_a \frac{1}{\rho_p r_p}; \tag{8}$$

There is a correlation between drag coefficient $C_D$ and Reynolds number $R_e$: $R_e = 2\frac{r_p v_p \rho_p}{\eta_a}$; (9)

In equations (5), (6), (7), (8), (9):

$\rho_p$-granular organic fertilizer density; $V_p$-volume of granular organic fertilizer; $\rho_a$-air density; $\eta_a$-aerodynamic viscosity; $v_P$-granular organic fertilizer velocity; $S_p$-granular organic fertilizer frontal area; $r_p$-granular organic fertilizer radius; g-gravity acceleration;

Assuming that the positive direction of the Z axis is opposite to the direction of gravity of the granular organic fertilizer, the equation of motion of the granular organic fertilizer in the X, Y, and Z directions [25]-[26] is

$$\text{X direction } \frac{d^2x}{dt^2} = -K v_{Px} \sqrt{v_{Px}^2 + v_{Py}^2 + v_{Pz}^2}; \tag{10}$$

$$\text{Y direction } \frac{d^2y}{dt^2} = -K v_{Py} \sqrt{v_{Px}^2 + v_{Py}^2 + v_{Pz}^2}; \tag{11}$$

$$\text{Z direction } \frac{d^2z}{dt^2} = -K v_{Pz} \sqrt{v_{Px}^2 + v_{Py}^2 + v_{Pz}^2}; \tag{12}$$

In equations (10), (11), (12):

$v_{Px}$-the velocity component in the X direction; $v_{Py}$-the velocity component in the Y direction; $v_{Pz}$-the velocity component in the Z direction;

**Fertilization Uniformity Evaluation**

The granular organic fertilizer passes through the fertilizer guiding mechanism and finally falls into the ditch of the orchard ditching fertilizing machine. In order to evaluate the uniformity of fertilization, the distribution range of fertilizer in a ditch is selected as the sampling area, and meshing is performed according to 15 rows and 15 columns, as shown in Fig. 7. Among them, the width of the sampling area is the groove width, and the cell grid size is 20mm×20mm.

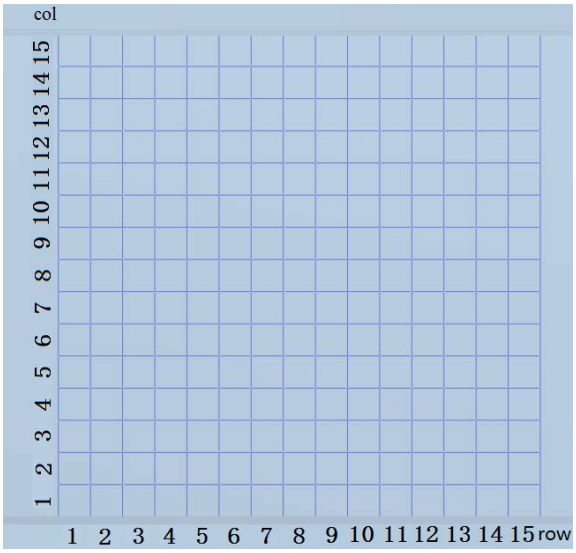



**Figure 7. Mesh partition of sampling regions**

The discrete coefficient CV is used as a measure of the uniformity of fertilizer distribution. The equations for calculating the discrete coefficient is:

$$CV=\frac{S}{\bar{q}}; \quad (13)$$

$$S=\sqrt{\frac{\sum_{i=1}^{n}(q_i-\bar{q})^2}{n-1}}; \quad (14)$$

$$\bar{q}=\sqrt{\frac{\sum_{i=1}^{n}(q_i)^2}{n}}; \quad (15)$$

In equations (13), (14), (15):

*S*-standard deviation; $\bar{q}$-The average number of samples in the unit area of the mesh granular organic fertilizer; *n*-total number of cell grids in the entire sampling area; $q_i$-the number of samples in *the i-th* unit lattice region granular organic fertilizer;

To better reflect the distribution of granular organic fertilizer in the entire sampling area, the discrete coefficients of the $1^{th}$ row, and the $15^{th}$ row are selected as a reflection of the edge area of the distribution of organic fertilizer, discrete coefficients of $5^{th}$ row, $8^{th}$ row, $11^{th}$ row are used as a parameter to reflect the uniformity of granular organic fertilizer distribution in the middle region.

**Experiment and Analysis**

Firstly, the fertilizer guiding mechanism was introduced into the EDEM simulation software and the relevant parameters were set. In order to obtain the effect of the granular organic fertilizer passing through the fertilizer guiding mechanism into the ditch, an open groove of 1000mm in length, 300mm in width and 400mm in depth was set to simulate the ditching of the ditching and fertilizing machine. The open groove was located directly below the export of the fertilizer guiding mechanism, wherein the axis was in the same vertical plane as the central axis of the fertilizer guiding mechanism, and the bottom of the opening groove was 200mm away from the export of the fertilizer guiding mechanism. Secondly, the fertilizer guiding mechanism was set as a random granules generator, so that granular organic fertilizer was randomly generated at the opening port in the simulation process. Among them, the granular organic fertilizer was randomly generated with an average equivalent diameter of 2.92mm as the mean value, a minimum equivalent diameter of 2.42mm, and a maximum equivalent diameter of 3.79mm. Finally, the motion characteristics of the assembly was set up. According to the principle of relative motion, the motion of the assembly was converted into the relative motion of the geometry, that is, the open groove moved in a reverse direction with respect to the fertilizer guiding mechanism at a speed of 0.25m/s [27].

In the simulation experiment, in order to ensure the accuracy of the test data under different structural parameters, the total simulation time was set to 2s, the fixed time step was set to 20%, and the target storage interval was set to 0.04s. The material of the mechanism was set to steel, and the relevant material parameters in the test were shown in Table 3. The dynamic friction coefficient and the static friction factor between the material granules and the different materials were determined by the shear box method and the bevel method respectively. The dynamic and static friction factor measurement test of the material granules and each material was repeated three times, and the test results were averaged; The free fall was utilized. The method was used to determine the collision recovery coefficient between material granules and different materials. The collision recovery coefficient determination test of material granules and the same material was repeated three times, and the test results were

averaged. The contact mechanical parameters between different materials were shown in Table 4.

**Table 3. Parameter of materiel**

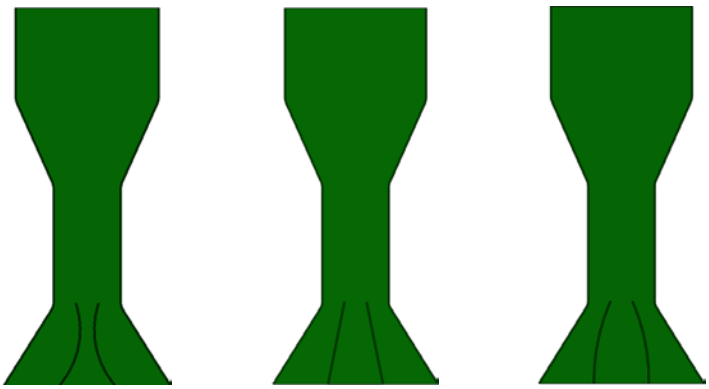
| Parameter | Granular organic fertilizer | Fertilizer guiding mechanism | Land |
|---|---|---|---|
| Poisson ratio | 0.25 | 0.45 | 0.50 |
| Shear modulus/Pa | $1\times10^{7}$ | $1\times10^{6}$ | $1\times10^{8}$ |
| Density/（$kg\cdot m^{-3}$） | 1300 | 3500 | 1200 |

**Table 4. Parameter of contact mechanics between materials**

| Parameter | Granular organic fertilizer—Granular organic fertilizer | Granular organic fertilizer—Fertilizer guiding mechanism | Granular organic fertilizer—Land |
|---|---|---|---|
| Recovery coefficient | 0.10 | 0.45 | 0.02 |
| Static friction factor | 0.30 | 0.30 | 1.25 |
| Dynamic friction factor | 0.25 | 0.20 | 1.25 |

*Test and Analysis of Different Shapes of Fertilizer Plug-in Plate*

In the fertilizer guiding mechanism, the fertilizer plug-in plate was one of the important factors affecting the uniformity of fertilization. The fertilizer plug-in plate of the existing linear fertilizer guiding mechanism was a linear type, and on this basis, a concave and a convex fertilizer guiding mechanism are proposed. Among them, the fertilizer plug-in plate of the concave fertilizer guiding mechanism was a concave curve type, and the fertilizer plug-in plate of the convex fertilizer guiding mechanism was a convex curved type, and the three fertilizer plug-in plates basically covered most guiding fertilizer insertion boards, which covered the main possible types of fertilizer plug-in plates, namely the concave, linear, and convex fertilizer guiding mechanisms, shown in Fig. 8.



(a). concave type (b). linear type (c). convex type

**Figure 8. Diagrammatic sketch of three fertilizer guiding mechanisms**

The shape of the fertilizer plug-in plate in the fertilizer guiding mechanism was taken as a single factor variable, and other parameters were unchanged. The simulation tests were carried out on the concave, linear and convex fertilizer guiding mechanisms respectively. Among them, the radius of curvature of the fertilizer plug-in plate of the concave and the linear fertilizer guiding mechanism was set to 350mm.

At 1s, the velocity directions of 330 granular organic fertilizers at the fertilizer outlets of the concave, linear and convex types were measured, as shown in Table 5. Among them, it was represented by -1~1 in the EDEM simulation software. Based on the horizontal speed, the entire speed plane was equally divided into six parts at intervals of 30°, as shown in Fig. 9. The speed direction in Table 5 through $\angle\alpha=\frac{v_{fi}}{2}\times180°$ ( 16 ) was converted to the speed plane of Fig. 9, and the amount of granular organic fertilizer in each part was counted. The velocity distribution map of different parts of granular organic fertilizer was drawn, as shown in Fig. 10.

In equation (16):

"$\angle\alpha$"-the angle between the velocity direction of the granular organic fertilizer and the horizontal velocity direction; "$v_{fi}$"-the speed direction of the *i*-th granular organic fertilizer;

**Table 5. Velocity direction of granular organic fertilizer**

| Item | Velocity direction of granular organic fertilizer in concave fertilizer guiding mechanism | Velocity direction of granular organic fertilizer in linear fertilizer guiding mechanism | Velocity direction of granular organic fertilizer in convex fertilizer guiding mechanism |
|---|---|---|---|
| Minimum value | -0.98 | -0.98 | -0.98 |
| Maximum value | 0.99 | 0.99 | 0.99 |
| Average value | 0.18 | 0.34 | 0.13 |
| Variance | 0.35 | 0.40 | 0.30 |

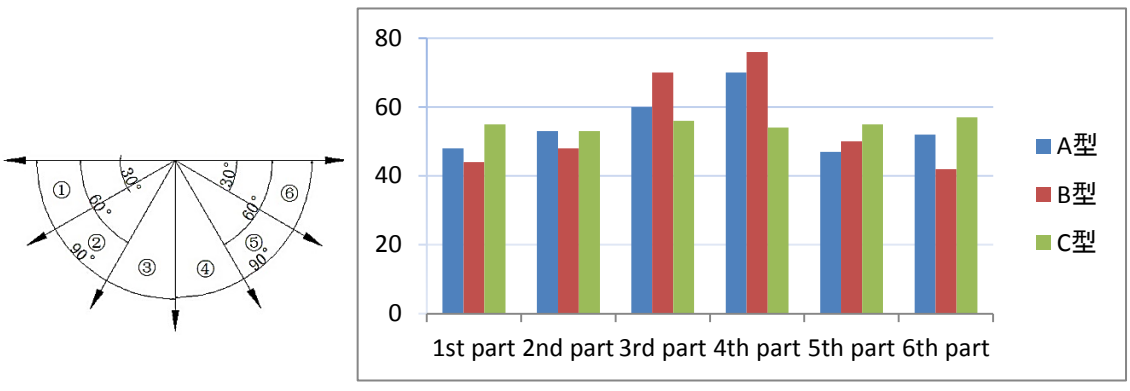


**Figure 9. Partition of velocity direction Figure 10. Distribution of velocity direction in different parts**

It can be seen from Fig. 10 that in the concave and linear fertilizer guiding mechanism, the granule direction of the granular organic fertilizer is distributed in the third and fourth parts, and the number of granules is higher than the first, second, fifth and sixth parts; In the convex fertilizer guiding mechanism, the granule direction distribution of the granular organic fertilizer is almost equal in the number of granules in each part. The granular organic fertilizer in the first and sixth parts of the velocity direction is concentrated on both sides of the bottom of the ditch when falling into the bottom of the ditch; the granular organic fertilizer in the third and fourth parts of the velocity direction is concentrated in the bottom of the ditch when falling into the middle part of the bottom of the ditch; the granular organic fertilizer in the second and fifth parts of the velocity direction is concentrated in the middle of the ditch when it falls into the bottom of the ditch. In the concave and linear fertilizer guiding mechanism, when the granular organic fertilizer falls into the bottom of the ditch, it is concentrated in the middle part of the ditch, and the distribution on both sides is less, resulting in uneven distribution of granular organic fertilizer; in the convex fertilizer guiding mechanism, the velocity direction of the granular organic fertilizer is distributed evenly in all parts, and the distribution after falling into the bottom of the groove is relatively uniform.

In order to further evaluate the uniformity of fertilization, the discrete coefficients of the edge region and the middle region in the sampling area were calculated and counted. The results of the discrete coefficients of the edge regions of different fertilizer guiding mechanisms are shown in Table 6. The discrete coefficients of the middle regions of different fertilizer guiding institutions are shown in Table 7.

**Table 6. CV of different fertilizer guiding mechanisms in edge region**

| Edge region | CV of concave fertilizer guiding mechanism | CV of linear fertilizer guiding mechanism | CV of convex fertilizer guiding mechanism |
|---|---|---|---|
| 1st row | 0.57 | 0.43 | 0.36 |
| 15th row | 0.50 | 0.54 | 0.43 |

**Table 7. CV of different fertilizer guiding mechanisms in middle region**

| Middle region | CV of concave fertilizer guiding mechanism | CV of linear fertilizer guiding mechanism | CV of convex fertilizer guiding mechanism |
|---|---|---|---|
| 5th row | 0.43 | 0.50 | 0.29 |
| 8th row | 0.42 | 0.52 | 0.32 |
| 11th row | 0.42 | 0.49 | 0.28 |

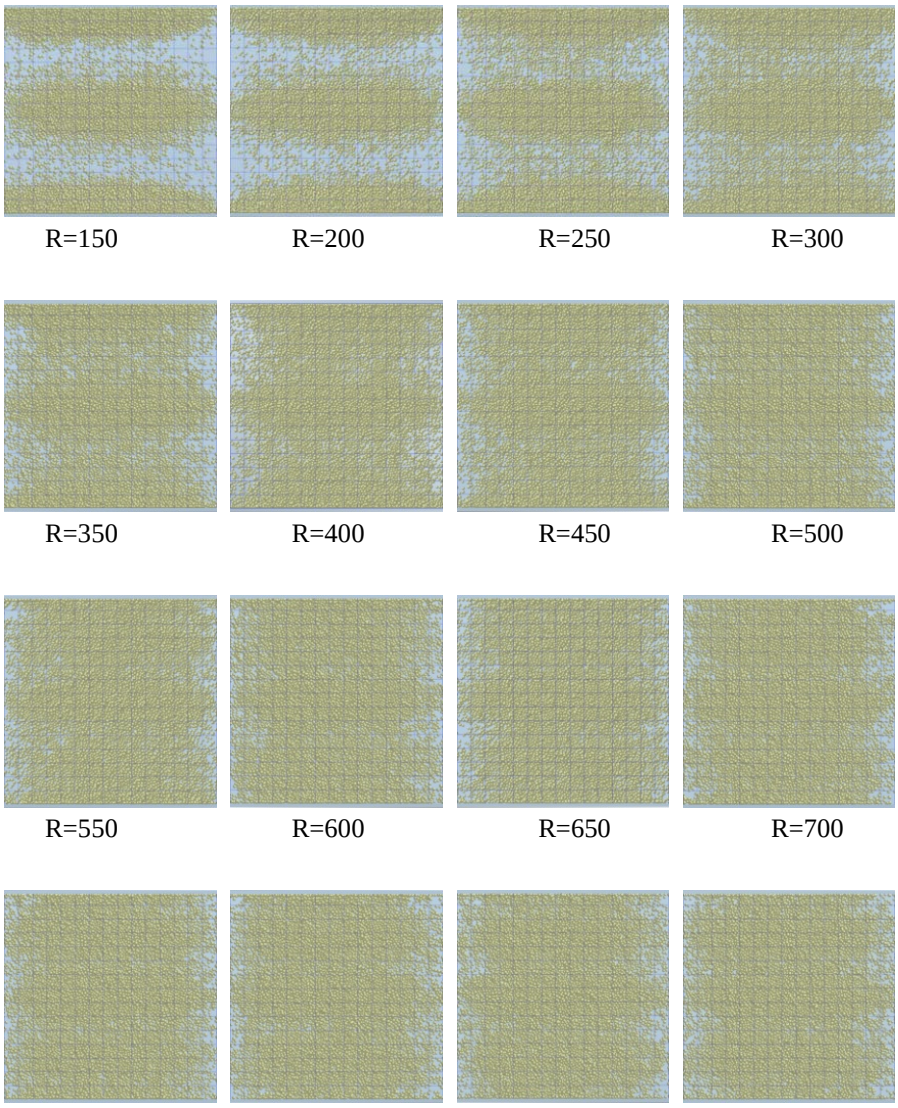
From the discrete coefficients of each region in the sampling area, it can be seen that in the three fertilizer guiding mechanisms of concave, linear and convex that, the discrete coefficients of the edge region and the middle region of the convex fertilizer guiding

mechanism are lower than the concave and linear guides. The discrete coefficient of each part of the fertilizer guiding mechanism has the highest uniformity of fertilization.

The shape of the fertilizer plug-in plate changes the speed direction of the granular organic fertilizer, thereby affecting the uniformity of fertilization. The convex fertilizer plug-in plate is superior to the linear and concave fertilizer plug-in plate, and the convex fertilizer guiding mechanism is optimal.

*Test and Analysis of the Fertilizer Plug-in Plate with Different Curvature Radius*

In order to further determine the influence of the fertilizer plug-in plate with different curvature radius on the fertilization uniformity in the convex fertilizer guiding mechanism, the radius of curvature of the fertilizer plug-in plate was taken as a single factor variable, with a radius of curvature of 150mm~800mm at intervals of 50mm. The simulation experiment was carried out on the fertilizer plug-in plate, and the simulation results are shown in Fig. 11. Separate the discrete coefficients of the edge region and the middle region at different curvature radius were analyzed, as shown in Table 8.



R=750 R=800 R=850 R=900

**Figure 11. Simulation result under different curvature radius**

**Table 8. CV under different curvature radius**

| Curvature radius | Edge region | | Middle region | | |
|---|---|---|---|---|---|
| | 1st row | 15th row | 5th row | 8th row | 11th row |
| 150 | 0.99 | 1.02 | 0.81 | 0.82 | 0.86 |
| 200 | 0.73 | 0.82 | 0.67 | 0.66 | 0.67 |
| 250 | 0.64 | 0.54 | 0.52 | 0.54 | 0.54 |
| 300 | 0.53 | 0.50 | 0.44 | 0.39 | 0.44 |
| 350 | 0.36 | 0.44 | 0.29 | 0.32 | 0.28 |
| 400 | 0.30 | 0.32 | 0.25 | 0.24 | 0.30 |
| 450 | 0.21 | 0.38 | 0.25 | 0.21 | 0.23 |
| 500 | 0.23 | 0.37 | 0.19 | 0.16 | 0.19 |
| 550 | 0.18 | 0.31 | 0.15 | 0.21 | 0.16 |
| 600 | 0.19 | 0.29 | 0.20 | 0.18 | 0.15 |
| 650 | 0.21 | 0.39 | 0.16 | 0.15 | 0.19 |
| 700 | 0.24 | 0.36 | 0.21 | 0.19 | 0.17 |
| 750 | 0.28 | 0.28 | 0.17 | 0.20 | 0.21 |
| 800 | 0.25 | 0.37 | 0.14 | 0.18 | 0.17 |

It can be seen from Table 8 that when the radius of curvature is in the range of 150mm~550mm, the discrete coefficient of the edge region and the middle region is gradually decreased; when the radius of curvature is in the range of 550mm~800mm, the discrete coefficient of the edge region and the middle region does not change much; the radius of curvature of the fertilizer plug-in plate affects the uniformity of fertilization within a certain range.

To further describe the relationship between the radius of curvature and the uniformity of fertilization, a cubic polynomial was used to fit the discrete coefficient curves of the edge region and the middle region.

First, at different radius of curvature, the average of the discrete coefficients of the first row and the 15th row is taken as the feature point of the discrete coefficient of the edge region; secondly, at the different radius of curvature, the discrete coefficient of the fifth row, the eighth row, the 11th row is averaged and used as the feature point of the middle region discrete coefficient. Finally, the cubic polynomial fitting is performed on the feature points of the edge region and the middle region to obtain the discrete coefficient fitting curve between the edge region and the middle region. The results of the discrete coefficient fitting are shown in Fig. 12.

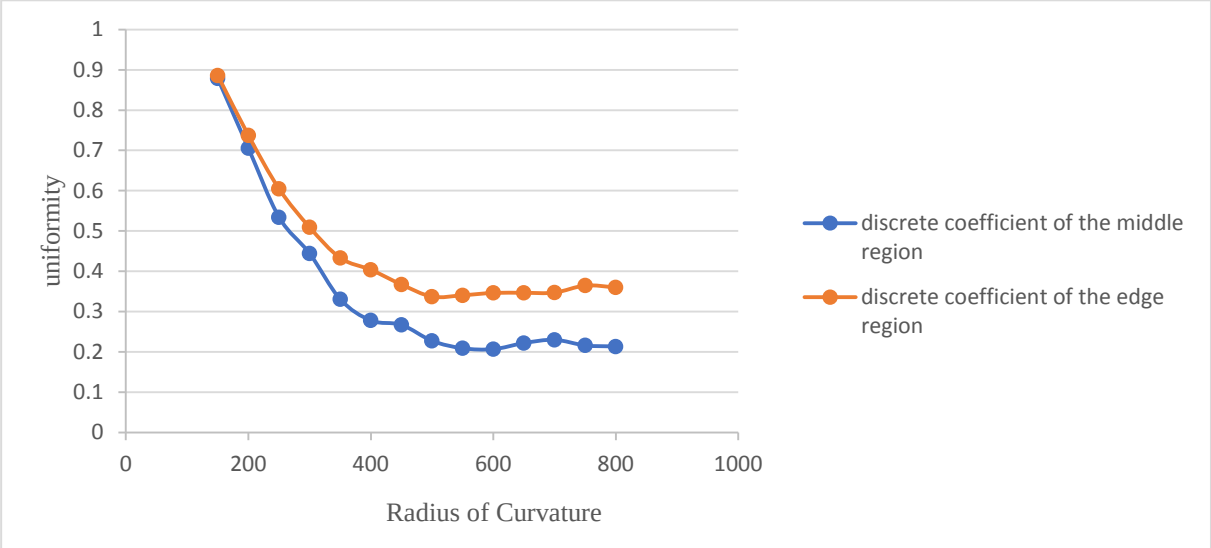



**Figure 12. Fitting curve of CV**

It can be seen that the edge region fitting discrete coefficient obtains a minimum value when the radius of curvature is 566mm, and the middle region fitting discrete coefficient obtains a minimum value when the radius of curvature is 596mm. In order to ensure that the discrete coefficients of edge and middle regions are minimized, and combined with processing technology and cost, the optimal curvature radius is finally determined to be 600 mm.

In the convex fertilizer guiding mechanism, the radius of curvature of the fertilizer plug-in plate affects the uniformity of fertilization. Under the processing conditions, when the radius of curvature of the fertilizer plug-in plate is 600mm, the cost is low and the uniformity of fertilization is high.

**Field Trials**

In mid-June 2017, a field trial was conducted in the experimental farm of Henghe in Shandong Province, as shown in Fig. 13. The farm was a large-scale standardized planting of orchard, with a row spacing of 2.5m and a plant spacing of 1m. The Plant was 3 years old and grows well. The test site was loam, the soil had an absolute moisture content of 23.3%, the soil firmness was 76.0kPa, and the terrain was flat, which provided favorable conditions for the smooth progress of the test. The test prototype was a two-row ditching and fertilizing machine for the orchard. The fertilizer guiding mechanism was a linear fertilizer guiding mechanism before optimization and an external convex fertilizer guiding mechanism with a radius of curvature of 600mm. In order to ensure the accuracy of the experimental data, three regions were randomly selected, and repeated verification tests of three discrete coefficient measurements were performed for each region, and the discrete coefficient curves before and after optimization were plotted, as shown in Fig. 14.

(a). Field operation (b). Fertilization effect in sampling area before and after optimization

**Figure 13. Field Test**

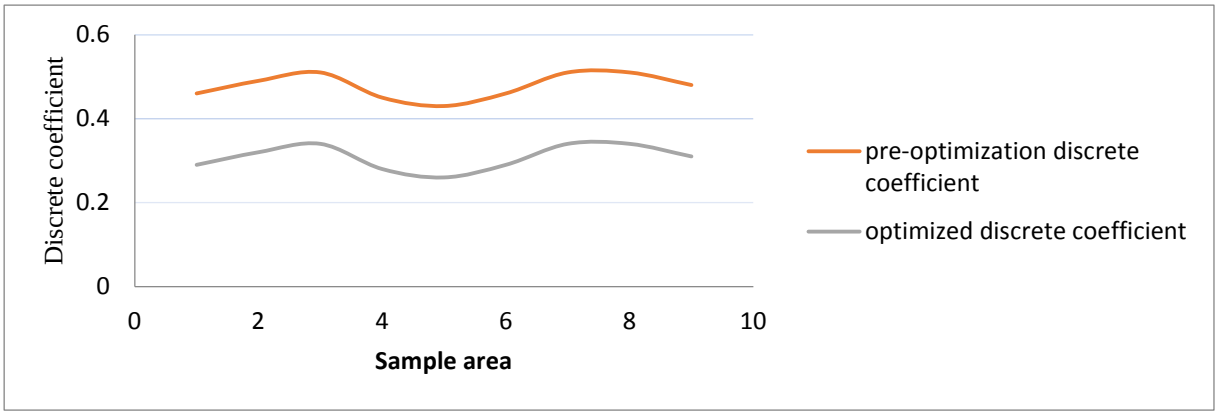



**Figure 14. CV curve before and after optimization in different regions**

The experimental results show that the optimized discrete coefficient curve is below the pre-optimization discrete coefficient curve. The optimized discrete coefficient is significantly smaller than the pre-optimization discrete coefficient. The average value of the discrete coefficient of the sampled area before optimization is 0.51. The average value of the discrete coefficient of the sampled area after optimization is 0.26. Fertilization uniformity increased by 49.02%.

After optimizing the parameters of the fertilizer guiding mechanism, under the conditions of the same ditch depth and the speed, the discrete coefficient of the sampling area is reduced, the uniformity of fertilization is improved, and the technical requirements for the ditching and fertilization of the orchard are satisfied.

## Conclusion

(1) Based on the existing linear fertilizer guiding mechanism, two other fertilizer guiding mechanisms, a concave fertilizer guiding mechanism and a convex fertilizer guiding

mechanism are proposed.

(2) By using the shape of the fertilizer plug-in plate as a single factor variable, the three kinds of fertilizer guiding mechanisms of concave, linear and convex type are simulated respectively to determine the convex fertilizer plug-in plate of the convex fertilizer guiding mechanism has the best shape and its uniformity of fertilization is the highest. Among them, the shape of the fertilizer plug-in plate affects the uniformity of fertilization by changing the velocity direction of the granular organic fertilizer.

(3) By using the radius of curvature of the fertilizer plug-in plate as a single factor variable, the simulation experiment is carried out on the convex fertilizer guiding mechanism with the radius of curvature of 150mm~800mm, and the polynomial is used to fit the discrete coefficient of the edge region and the middle region to determine the curve. The convex fertilizer guiding mechanism has an optimum radius of curvature of 600 mm. According to the field test, the discrete coefficient after optimization was reduced from 0.51 to 0.26, and the uniformity of fertilization was increased by 49.02%.

**References**

[1] Shi, Y. Y., Chen M., Wang, X. C., Morice, O. O., Li, C. G. and Ding, W. M. (2018) Design and Experiment of Variable-rate Fertilizer Spreader with Centrifugal Distribution Cover for Rice Paddy Surface Fertilization. *Transactions of the Chinese Society for Agricultural Machinery* 49, 86-93+113.

[2] Zheng, X. C., Lu, H. J., Che, J. X., Zhai. B. N., Zhao, Z. Y. and Wang, Y. L. (2011) Investigation of present yield and fertilization on Fuji apple in Baishui County. *Journal of Northwest A&F University* 39, 145-151.

[3] Han, D. Y., Lv, Z. Q., Cui, F. F. and Shen, X. (2010) Design of the Fertilizing Machine for Fruit Trees. *Journal of Agricultural Mechanization Research* 32, 65-68.

[4] Ma, C., Meng, H. W., Kan, Z. and Qi, J. T. (2017) Design of Jujube Harvest Test Device Based on Self-excited Vibration and Force Compensation. *Journal of Agricultural Mechanization Research* 39, 12-17.

[5] Gao, J. S., Xu, M. G., Wang, B. R., Qin, D. Z., Wen, S. L. and Shen, H. P. (2005) Combined fertilization of chemical and organic fertilizers in a long-term position experiment. *Chinese Agricultural Science Bulletin* 21, 211-214.

[6] Wu, N., Wu, D. S., Li H. and Liang, H. (2016) Calibration of Manure Spreaders. *Journal of Agricultural Mechanization Research* 9, 255-259.

[7] Ma, J., Yan, H. J. and Wang, Ch. Y. (2016) Effect of end gun on/off on variable rate fertigation uniformity and its improvement for center pivot irrigation system. *Journal of Hydraulic Engineering* 12, 1577-1584.

[8] Zhou, Z., Fu, Z. T., Wang, X. and Zheng, L. J. (2009) Experiment of fertigation uniformity of drip fertigation machine. *Transactions of the Chinese Society of Agricultural Engineering* 25, 7-13.

[9] Li, M., Zhang, T., Dong, X. H., Wang, C., Niu, Z. J., Ge, C. and Wei, L. J. (2016) Parameter optimization on scraper fertilizer feed unit of 3ZSP-2 type sugarcane intertillage fertilizer applicator-cum-hiller. *Transactions of the Chinese Society of Agricultural Engineering* 32, 36-42.

[10] Zhao, L. Y. *Study on the Text System of Virtual Instrument for Precision Seed-Metering Device*, MS Thesis, Nanjing: Nanjing Agricultural University, 2002.

[11] Patterson, D. E. and Reece, A. R. The theory of the centrifugal distributor. I: Motion on the disc, near-centre feed. (1962) *Journal of Agricultural Engineering Research* 7, 232-240.

[12] Liedekerke, P. Van., Tijskens, E., Dintwa, E., Rioual, F., Vangeyte, J. and Ramon, H. (2009) *Powder*

*Technology*, 190, 348-360.

[13] Przywara, Artur. The Impact of Structural and Operational Parameters of the Centrifugal Disc spreader on the Spatial Distribution of Fertilizer. 2015 *Agriculture and Agricultural Science Procedia* 7,215~222.

[14] Villette, S., Piron, E., Martin, R., Courreau, D.,Miclet, D. and Gée, C. (2010) Impact recording system to characterize centrifugal spreading. *Book of Abstracts*.

[15] Aphale, A., Bolander, N., Park, J.,Shaw, L., Svec, J. and Wassgren, C. Granular fertilizer particle dynamics on and off a spinner spreader. *Biosystems Engineering*, 2003, 319-329.

[16] Zhang, T., Liu, F., Liu, Y. Q., Zhao, M. Q., Zhang, S., Li, N., Li, L. F. and Lv, B. (2015) Discrete Element Simulation of Outer Groove Wheel Type Fertilizer Discharging Device Capacity Analysis. *Journal of Agricultural Mechanization Research* 9, 198-201.

[17] Chen, X. F., Luo, X. W., Wang, Z. M., Zhang, M. H., Hu, L., Yang, W. W., Zeng, S., Zang, Y., Wei, H. D. and Zheng. L. (2015) Design and experiment of fertilizer distribution apparatus with double-level screws. *Transactions of the Chinese Society of Agricultural Engineering* 31, 10-16.

[18] Yang, X. L., Wang, J. W., Wang, J. F. and Zhou, W. (2015) Parameters Optimization Design and Analysis of Vaned Fertilizer Distributing Device. *Journal of Agricultural Mechanization Research* 7, 160-163.

[19] Yuan, W. S., Li, K., Jin, C. Q., Hu, M. J. and Zhang, W. Y. (2018) Design and Experiment of Hill Placement Fertilizer Applicator. *Journal of Agricultural Mechanization Research* 40, 145-149.

[20] Lv, J. Q., Wang, Z. M., Sun, X. S., Li, Z. H. and Guo, Z. P. (2015) Design and Experimental Study of Feed Screw Potato Planter Propulsion. *Journal of Agricultural Mechanization Research* 6, 194-196.

[21] Liu, B., Xiao, H. R., Song, Z. Y. and Mei, S. (2017) Present State and Trends of Fertilizing Machine in Orchard. *Journal of Agricultural Mechanization Research* 39, 263-268.

[22] Qi, W. Z., Wang, J. X., Liu., S. X., Wang, Y. L., Wang, Z. and Zhao, G. X. (2019) Design and Experiment of Control System for Rice and Wheat Variable Rate Fertilizer Applicator. *Journal of Agricultural Mechanization Research* 41, 72-79.

[23] Yuan, J., Liu, Q. H., Liu, X. M., Zhang, T. and Zhang, X. H. (2014) Simulation of Multi-fertilizers Blending Process and Optimization of Blending Cavity Structure in Nutrient Proportion of Variable Rate Fertilization. *Transactions of the Chinese Society for Agricultural Machinery* 45, 125-132.

[24] Inns, F. M. and Reece, A. R. (1962) The theory of the centrifugal distributor II: motion on the disc, off-centre feed. Journal of Agricultural Engineering Research 7, 345-353.

[25] Cunningham, F. M. (1963) Performance characteristics of bulk spreaders for granular fertilizer. *Transactions of the ASAE* 6,108-114.

[26] Pitt, R. E., Farmer, G. S., Walker, L. P. (1982) Approximating equations for rotary distributor spread patterns. *Transactions of the ASAE* 25,1544-1552.

[27] Yuan, J., Liu, Q. H., Liu, X. M., Zhang, T. and Zhang, X. H. (2014) Granular Multi-flows Fertilization Process Simulation and Tube Structure Optimization in Nutrient Proportion of Variable Rate Fertilization. *Transactions of the Chinese Society for Agricultural Machinery* 11, 81-87.

# Study of blood flow velocity in epicardial microwave ablation of atrial fibrillation

**Zhang Meng, *NAN Qun**

College of Life Science and Bioengineering, Beijing University of Technology, China

* Corresponding author: nanqun@bjut.edu.cn

## Abstract

Purpose: Atrial Fibrillation is the most common clinically cardiac arrhythmia. At present, microwave ablation has been widely used in the treatment of atrial fibrillation, but the effect of microwave ablation on blood flow velocity is not enough, and blood flow velocity is closely related to the ablation effect of atrial fibrillation. In order to study this problem, the temperature field of microwave ablation in atrial fibrillation was numerically simulated.

Methods: This research is based on the finite element method and COMSOL Multiphysics software for numerical simulation. In this study, a three-dimensional fat-myocardium-blood model is constructed, involving numerical simulation of electromagnetic field, temperature field and flow field. The microwave frequency was 2450 MHz, and fat thickness was 1 mm, myocardial thickness was 3 mm. The ablation power was 100 Wand the ablation time was 15s.The blood convective heat transfer coefficient were 0 W/($m^2$·℃), 1417 W/($m^2$·℃), 3350 W/($m^2$·℃), 7100 W/($m^2$·℃).

Results: In this simulation, the ablation area is hemispherical. When the blood convective heat transfer coefficient were 0 W/($m^2$·℃), 1417 W/($m^2$·℃), 3350 W/($m^2$·℃), 7100 W/($m^2$·℃), the maximum temperature were 192℃, 186℃, 184℃, 183℃, the transverse width were 7.5 mm, 7.2 mm, 7 mm and 6.8 mm, and the axial length were 3.7 mm, 3.4 mm, 3.3 mm, 3.2 mm respectively. The maximum temperature, transverse width and axial length decrease with the increasing convective heat transfer coefficient.

Conclusions: A thin cryogenic layer forms between the Myocardium and the blood . The maximum temperature of the ablation zone decreases with the increasing convective heat transfer coefficient, and the position of the effective ablation zone (⩾50℃) moves to the fat layer.

**Keywords:** microwave ablation; numerical simulation; blood; convective heat transfer coefficient

## 1. Introduction

Atrial fibrillation(AF) is the most common persistent arrhythmia in clinic. It ranks the second in arrhythmia and has a high fatality and disability. At present, drug therapy is the most commonly method to cure AF. But 10% to 15% of patients have side effects after taking the drug [1] [2]. In non-drug treatment, surgical treatment has become the main means, but it has a greater risk. In recent years, minimally invasive ablation, such as microwave ablation, radiofrequency ablation[3] and cryoablation[4], has been widely used on the basis of maze surgery. Microwave ablation is a kind of high-frequency electromagnetic wave, which

generates heat by inducing the vibration of water molecules in tissues and heating tissues directly, independent of the current passing through the probe. microwave ablation can achieve greater ablation depth, especially in the treatment of ventricular tachycardia. Microwave energy attenuates very little when it passes through fat and blood, so it has strong tissue penetration.

But the heart is rich in blood, and the blood flow varies greatly in different heart locations. the greater the velocity of blood flow, the greater the heat transfer coefficient of blood convection, and the more heat it takes away.

In order to reflect the epicardial ablation, this study constructed a fat-myocardium-blood model to explored the influence of blood flow velocity, and simulated the temperature field atrial fibrillation.

## 2. Methods

### 2.1 Geometric model

The numerical simulation of this study is based on COMSOL Multiphysics Software.

The structure of microwave antenna comes from the literature[5]. Based on the basic requirements of electromagnetic wave simulation, the material of antenna is defined as copper, and the part between inner and outer conductors was set as Teflon. The structure of THE microwave antenna was shown in Figure 1. The antenna, fat, myocardium and blood model were established. The front end of the microwave antenna was perpendicular to adipose tissue. The thickness of fat was 1 mm, the thickness of myocardium was 3 mm, and the thickness of blood was 2 cm. The model was shown in Figure 2. In this study, the convective heat transfer coefficients at the boundary of blood and myocardium are applied to simulate the blood flow velocity. The coefficients of different atrial parts were shown in Table 1.

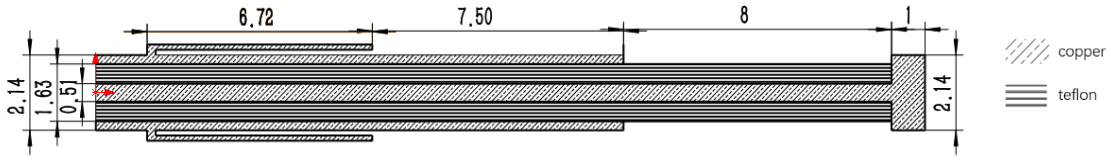



**Figure 1. Structure of microwave ablation antenna**

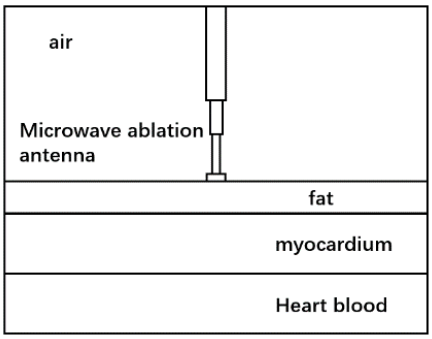



**Figure 2. Section drawing of numerical simulation model for microwave ablation of epicardial atrial fibrillation**

**Table 1. Convective heat transfer coefficient in different applied location[6][7]**

| position | Convective heat transfer coefficient |
|---|---|
| within 10mm above the mitral valve on the lateral wall | 1417W/(m$^2$·℃) |
| CS，Ventricular AP | 3350 W/(m$^2$·℃) |
| AV node，Atrial AP，RV outflow VT | 7100 W/(m$^2$·℃) |

### 2.2 Bio-heat equation

During ablation of atrial fibrillation, the thermoelectric coupling effect is controlled by Pennes' bio-heat transfer equation[8]. The formula is as follows

$$\rho C\frac{\partial T}{\partial t} = k\nabla^2 T + q - Q_b + Q_m \tag{1}$$

T(℃), ρ（kg/m$^3$）, C（J/kg·K）, K (W/m·K) and Q (W/m$^{3)}$ are temperature, density, specific heat, thermal conductivity and power density, respectively. $Q_b$ and $Q_m$ represent blood perfusion and metabolic heat, respectively. They are neglected for far less than that generated by microwave ablation .

## 3. Results and Discussion

The temperature distribution and data of ablation zone were studied under different convective heat transfer coefficients (0 W/(m$^2$·℃), 1417 W/(m$^2$·℃), 3350 W/(m$^2$·℃), 7100 W/(m$^2$·℃). The ablation power was 100W. And the heating time was 15s. The axial ablation region is parallel to the ablation region of the microwave antenna, and the transverse ablation region is perpendicular to the ablation region of the microwave antenna. According to the tissue and physical characteristics, 50℃ is the threshold temperature for the formation of ablation blockade line caused by myocardial injury. Table 2 shows the experimental data at different heat transfer coefficient.

**Table 2. Experimental data**

| convective heat transfer coefficient W/(m$^2$·℃) | 0 | 1417 | 3350 | 7100 |
|---|---|---|---|---|
| transverse width/mm | 7.5 | 7.2 | 7 | 6.8 |
| axial length/mm | 3.7 | 3.4 | 3.3 | 3.2 |
| maximum temperature/℃ | 192 | 186 | 184 | 183 |

As can be seen from Figure 3, with the increase of heat transfer coefficient, the maximum temperature deceases, while the lateral width and axial length decrease too. When the convective heat transfer coefficient was 0 W/(m$^2$·℃), 1417 W/(m$^2$·℃), 3350 W/(m$^2$·℃) and 7100 W/(m$^2$·℃), the transverse width of the ablation zone was 7.5mm, 7.2mm, 7mm and 6.8mm, and the axial length was 3.7mm, 3.4mm, 3.3mm and 3.2mm, respectively. The maximum temperature of the ablation zone was 192℃, 186℃, 184℃ and 183℃,

respectively. The presence of blood flow lowers the maximum temperature, decreases both the axial length and the transverse width, but the amplitude is very small.

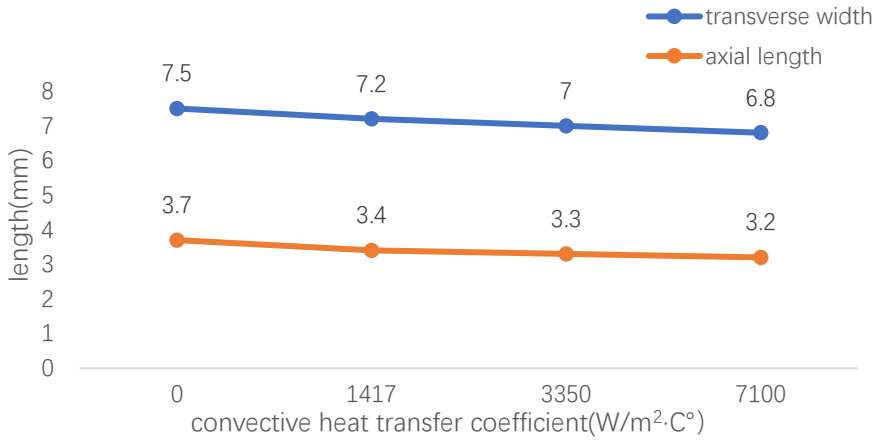



**Figure 3. Changes of cardiac tissue length with convective heat transfer coefficient**

## 4. Conclusion

This study simulated the effect of blood flow on microwave ablation of atrial fibrillation in a simple three-layer myocardial model. The results showed that the maximum temperature of ablation region decreased with the increase of convective heat transfer coefficient, and the transverse width and axial length decreased, but the amplitude was small.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (31771021) and Jiangsu Science and Technology Project (BE2017758).

**References**

[1] PATEL M R, MAHAFFEY K W, GARG J, et al. Rivaroxaban Versus Warfarin inNonvalvular Atrial Fibrillation[J]. New England Journal of Medicine, 2011, 365(10): 883-891.

[2] WILBER D J. Radiofrequency Catheter Ablation of Cardiac Arrhythmias: Basic Concepts and Clinical Applications[M]. Futura Pub, 1994.

[3] Pérez, Juan J, González-Suárez, Ana, Berjano E . Numerical analysis of thermal impact of intramyocardial capillary blood flow during radiofrequency cardiac ablation[J]. International Journal of Hyperthermia, 2017: 1-25.

[4] Michael H , Gerald F , Michael S , et al. Computer simulation of cardiac cryoablation: Comparison with in vivo data[J]. Medical Engineering & Physics, 2013, 35(12): 1754-1761.

[5] Chiu H M , Mohan A S , Weily A R , et al. Analysis of a novel expanded tip wire (ETW) antenna for microwave ablation of cardiac arrhythmias[J]. IEEE Transactions on Biomedical Engineering, 2003, 50(7): 890-899.

[6] LAI Y C, CHOY Y B, HAEMMERICH D, et al. Lesion Size Estimator of Cardiac Radiofrequency Ablation at Different Common Locations with Different Tip Temperatures[J]. Biomedical Engineering, IEEE Transactions on, 2004, 51(10): 1859-1864.

[7] BERJANO E J, HORNERO F. Thermal-electrical Modeling for Epicardial Atrial Radiofrequency

Ablation[J]. Biomedical Engineering, IEEE Transactions on, 2004, 51(8): 1348-1357.

[8] Jiang Y , Zhao J , Li W , et al. A coaxial slot antenna with frequency of 433MHz for microwave ablation therapies: design, simulation, and experimental research[J]. Medical & Biological Engineering & Computing, 2017.

# Development of a window system with acoustic metamaterial for air and noise control

***Gioia Fusaro[1,2], Xiang Yu[1], †Fangsen Cui[1] and †Jian Kang[2,3]**

[1]Institute of High Performance Computing, A*Star, Singapore
[2]Acoustics Group, School of Architecture, University of Sheffield, UK
[3]Institute for Environmental Design and Engineering (IEDE), The Bartlett, University College London, UK

*Presenting author: gfusaro1@sheffield.ac.uk

† Corresponding author: cuifs@ihpc.a-star.edu.sg and j.kang@ucl.ac.uk

## Abstract

To improve window performances in reducing noise and allowing for air exchange, most current approaches focus on techniques such as double glazed and ducted designs, generally leading to bulky designs, visually non-optimised, and with narrow-banded frequency. In this research, window systems based on acoustic metamaterials (AMMs) are developed, and both natural air ventilation and acoustic performances are evaluated. The systems incorporate bistable auxetic metamaterials and acoustic origami metacage designs which are particularly interesting for their reconfigurable and deployable nature. Several design cases with different design features are examined, and a specific design is then selected for a parametric analysis using Finite Element Method (FEM) aiming to optimise the acoustical performance. It is demonstrated that significant improvement in acoustic performance can be obtained in terms of Transmission Loss (TL). The use of AMMs could lead to designs with manifold merits over traditional windows, including compact size with deployability, easy reconfigurability and installation, and thus paving new direction in ventilation window design.

**Keywords:** Acoustic metamaterials, Ventilation windows, FEM, auxetic metamaterials, Transmission Loss.

## 1. Introduction

Conventional acoustic techniques allow controlling sound wave propagation for a limited range of frequency, due to the device shape and bulky configuration (1,2). Metamaterials can be very versatile thanks to its advantages in acoustic properties related to its physical size (3,4). Two specific kinds of metamaterials are particularly interesting from the geometrical point of view: origami metamaterials and bistable auxetic metamaterials. The first metamaterial changes the spatial and the acoustic range of efficacy, while assuming different physical sizes by folding. The second metamaterial has the capacity of keeping a permanent, consistent volume, which may allow openings thanks to their well-known negative Poisson's ratio (5,6). Recent studies have associated such techniques with acoustic performances for mechanical devices improvement (7,8). Actual research is still seeking a significant impact on the combination of noise reduction and natural ventilation, in addition to architectonical sustainable solutions (9).
This study presents a novel acoustic design approach based on two acoustic metamaterials, which enable the natural air ventilation while reducing noise transmission significantly. The two mechanisms (origami and auxetic) are applied to already tested acoustic structure to increase the dynamicity in noise reduction and ventilation capacity (9). A metacage will be implemented by origami system, while a metasurface will be implemented with an auxetic mechanism. These designs allow expansion and compression of the geometries, which surround

(Design 1) or face (Design 2) a sound source. Finite Elements Method (FEM) simulations are performed with a frequency range between 100 and 5000 Hz, in order to attest the effectiveness of the designs. In the first case, two extreme configurations are tested (folded and unfolded configuration). The possibility of using a transparent material to realise both the models give hints for also achieving natural lightning.

## 2. Methodology

### 2.1. *Geometrical settings*

FEM simulation is employed to investigate the acoustic characteristics of both models. The boundary conditions and simulation set-ups are detailed in this section.
Two design models are proposed to enable noise reduction and natural ventilation between two separate spaces. For the origami metamaterial (Design 1) the acoustic performance is tested with a monopole sound source in the inside of the mechanism, aiming to screen the radiation towards the outer space. In the bistable metasurfaces (Design 2) the acoustic mechanism works with a surface sound source facing the structure which reduces noise propagation towards space behind it. In both cases, the sound wave and the air are meant to pass through a duct characterised by a number of cavities connected to it. This mechanism is supposed to create a resonance effect and influence significantly the sound wave propagation and so also the TL. In Design 1, the duct system is followed by apertures on the edges of the origami starred points (See Figure *1*). The origami metacage has two different configurations, folded and unfolded, which are both tested through the numerical method to evaluate their acoustic performances. In Design 2, the air and sound wave passes from one space to the other through the openings that are created from the negative Poisson's ratio displacements. As in the previous case, the openings are supported acoustically from a certain number of cavities, which in this case, connect the opening, as shown in Figure 2. Two configurations are tested for Design two. A different rotation angle characterises them as10° and 5°, so that the aperture towards which the cavities face is wider accordingly.

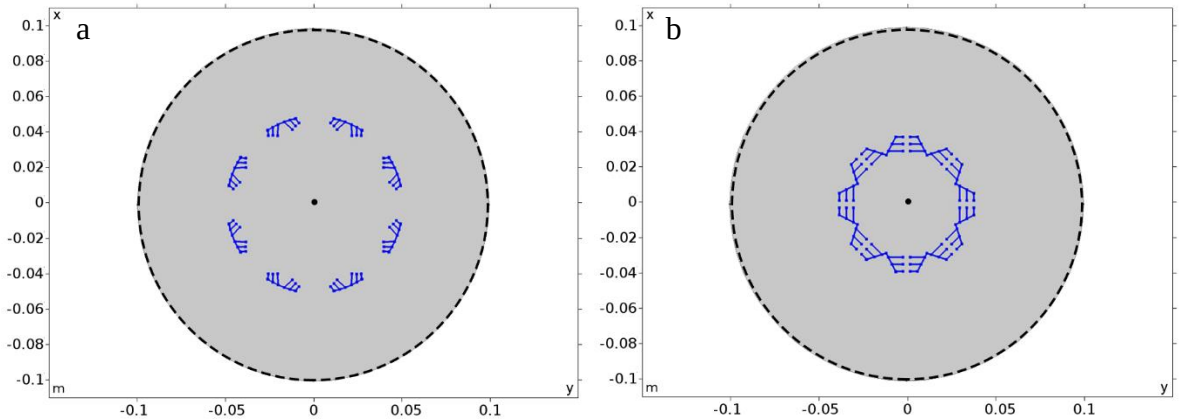



Figure 1. Geometrical configurations of Design 1 unfolded (a) and folded (b) and boundary conditions: central point source, interior sound hard boundaries (blue), and cylindrical free wave radiation (dashed line).

About the geometry in Design 1, it is mainly composed of a deployable system that can achieve two configurations: folded and an unfolded. Indeed, the origami structure allows the valley and mountain folds to go from eight points star shape (with 0.025 m length of each point's side) to a circular shape of 0.256 m diameter. Internally, each point is characterised by an opening 0.052 m each long (52% of perforation ratio of the entire boundary structure) and two cavities with

0.008 m depth created by three layers built starting from the perimeter surface of the metacage. These layers are modelled so they can leave at the centre of the point a resulting duct width 0.008 m which allows the air and acoustic wave to flow freely. When they turn with the structure to assume the unfolded configuration, they have direction perpendicular to the centre with an angular difference of +30° (see Figure 1).

Design 2 is an auxetic metasurface generated from the coupling of two layers, each one made by the repetition and connection of 4x4 basic squared units (each one 0.01 $m^2$ wide and 0.02 m thick). Figure 2 shows how this unit is repeated and connected with the others through hinges applied on the four edges of each one. From the negative Poisson's ratio displacements, apertures are generated in between the units, which allow the air and sound wave propagation to pass through them. This lead to an opening ratio of 30% for the 10° configuration and of 15% for the 5° one. Each opening has a 50% of perimeter surface removed to allow the resulting cavities to work as a resonator (see Figure 2). Design 2 is characterised by two layers (so four cavities facing the aperture, two per two of the blocks composing it). The openings which result from the Negative Poisson's ratio effect are 10° and 5° wide (see Figure *2*a-c), due to each block rotation (see Figure 2). About the cavities, a further investigation on both Design 1 and 2 is done afterwards, involving the relationship between TL and scale changing of both the models. For Design 1two bigger model are analysed, having respectively 0.4 and 0.8 m diameter in the folded configuration (the original one is 0.2 m). While thickness changing is done on Design 2 to investigate if a 4-layer or 8-layer model (so for a total metasurface thickness of 0.08 or 0.16 m) would affect TL.

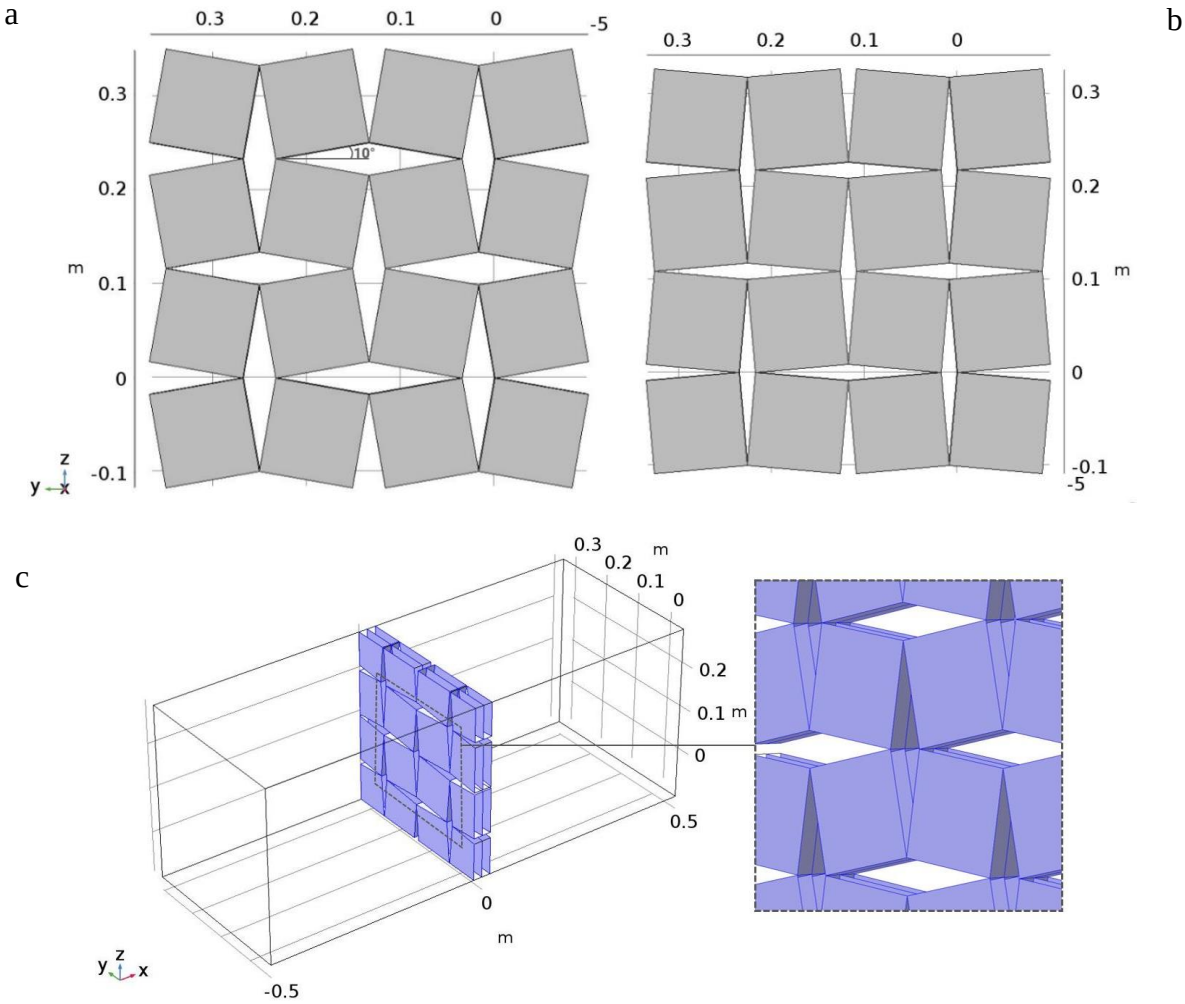



Figure 2. Geometrical configurations of Design 2, frontal view of 10° configuration (a) and 5° configuration (b). Schematic of interior sound hard boundaries, highlighted in blue (c).

### 2.2. *Boundary Conditions and Study Settings*

The Acoustics module of a commercial FEM software, Comsol Multiphysics, is used to implement the numerical model. For Design 1, a monopole point source is placed at the centre of the origami metacage with volume flow rate is 0.01 $m^2/s$. At the outer boundary, cylindrical wave radiation is defined to simulate free outgoing waves without reflection. The simulation domain is filled with air, where air density and sound speed at room temperature are used. The walls of the metacage and material cells are set as interior sound hard boundaries, as depicted in Figure 1. Sound transmission through walls of the metacage and possible viscous-thermal effect in the narrow resonator channels are neglected in this study. In Design 2, a plane wave radiation is applied to one of the ends of the 3D boundary volume (incident pressure = 1 Pa). This is a parallelepiped centred with the analysed geometry, having a length of 1 m (x-axis), and a width and depth of 0.38 each (y and z-axis). The opposite end of the boundary volume is characterised with air impedance.
For Design 1, the TL is calculated mathematically within the simulation software, from the averaged SPL at the outlet boundary (dashed line in Figure 1) and the monopole source SPL (=130 dB), to compare the acoustic response in the unfolded and folded state. In Design 2, TL is calculated by the reduction of sound power through the metamaterial interface (in decibel). An increase in the TL curve will thus indicate less efficient sound transmission because sound energy is more confined in the two systems (Design 1 and 2). The mesh size is determined according to the FEM criterion, where at least six nodes are used to simulate a wavelength in air. The dimensions and the complexity of the geometric problems have defined two different frequency ranges of application. So for Design 1, to reach 5000 Hz, the maximum allowed element size is thus 343/6/5000=0.0114 m. Indeed, the study is a frequency domain analysis from 100 Hz to 5000 Hz with a step size of 10 Hz. The meshes characterisation of Design 2 instead, has a maximum allowed element size of 343/6/3000=0.0114 m. Although this model results very complex and, since the convergence of results is proved, simplification is needed. So the maximum allowed element size is increased at 343/6/2000=0.0285 m. The study has a frequency domain that goes from 100 Hz to 3000 Hz with a step size of 10 Hz. In the results, the TL and SPL distribution are shown linearly and superficially within the simulation frequencies.

### 2.3. *Parametric studies*

The acoustic effectiveness of different metamaterials, is tested through a parametric study in both models. The 2D parameters, ‘cavities thickness’ and ‘duct width’, took in consideration three configurations for the first one (a= 0.006-0.008-0.010 m) and three for the second one (b= 0.006-0.008-0.010 m). In Design 1 each side has two cavities positioned towards the centre (upper section and lower section), delimited by layers which start from the sides and extend towards the middle of it for respectively 0.008 m, 0.012 m, and 0.016 m, and cavities width as 0.08 m (see Figure 3). In Design 2, the parametrization is performed with straight sides and geometry defined by cavities width and layers length. In this case, the layers’ length is set the same for all respectively 0.04 m, 0.06 m, and 0.08 m, and cavities width as 0.01 m, 0.02 m, and 0.03 m (see Figure 3). (see Figure 3). The parametrisation is set to see if there are any interesting correlations between the cavities or the duct’s width and the consecutive TL behaviour.

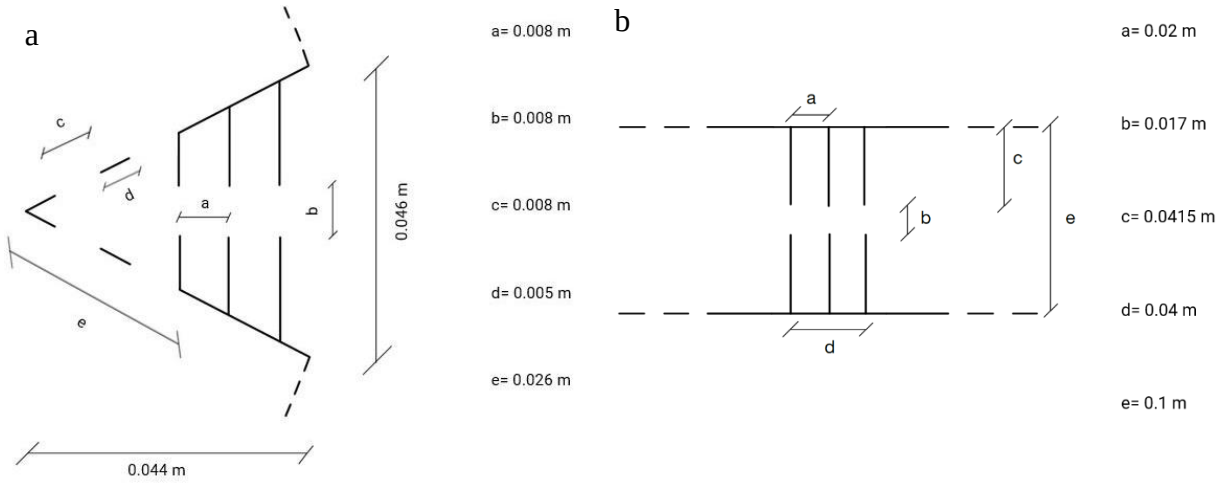



Figure 3. Schematic and dimensions of the metamaterial unit formed in the folded state for Design 1 (a), and of Design 2 (b).

## 3. Numerical results of Design 1

### 3.1. *Design 1 (Origami metacage with apertures)*

Figure 4 first shows the simulation results of Design 1 in the folded and unfolded state. The TL is between 8 dB and 30 dB, where some variations can be observed due to the resonance of the circular enclosure. From the TL graph (Figure 4a), both the folded and unfolded configuration effects are analysed. For the folded one the TL reduce significantly at low frequencies (average of 20 dB of TL), while in medium frequency it loose efficacy, and from 2500 Hz to go on, an increasing sinusoidal behaviour starts, with a TL average of 18 dB and a TL peak of 22 dB at 3900 Hz. Figure 4b and 4c highlight the confinement effect of the SPL at the different TL peak frequencies for both unfolded and folded configurations. In both graphs, it is evident how the unfolded state has a slightly higher acoustic impact on the sound wave confinement.

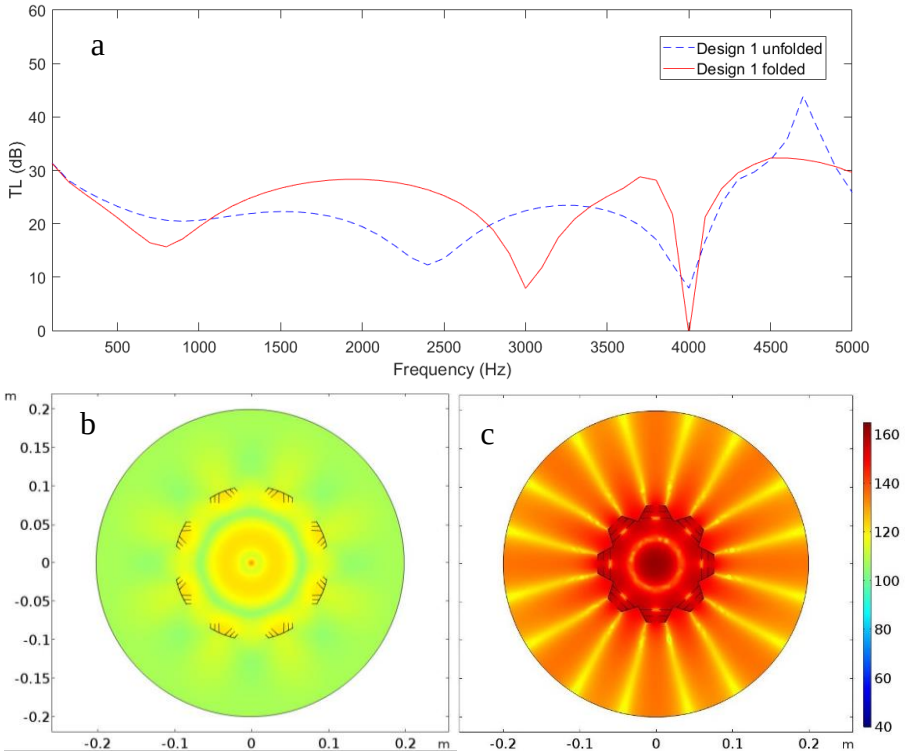



Figure 4. TL (a) and SPL distribution graph for 0.2 m Design 1 at 3000 Hz of the unfolded (b) and 3900Hz of the folded configuration (c).

3.2. *Parametric study on cavities' dimensions ratio*

Figure 5 shows the average behaviours according to two parameters: cavities thickness and duct width. From the nine combinations of the three per three options of two different variables (see schematic in Figure 3): $a_1=b_1=0.006$ m, $a_2=b_2=0.008$ m, $a_3=b_3=0.01$ m. From the results, it is clear that either the cavities or the central duct width change do not affect the acoustic metasurface performance. So the configuration $a_2$ and $b_2$ can be set as standard ($a_2=b_2=0.008$ m) to guarantee a significant sound reduction performance and sufficient ventilation at the same time. Indeed, the acoustic and airwave propagation from the inside to the outside of the metacage and vice versa is guaranteed by the resulted duct of width 0.008 and cavities total thickness of 0.016 m.

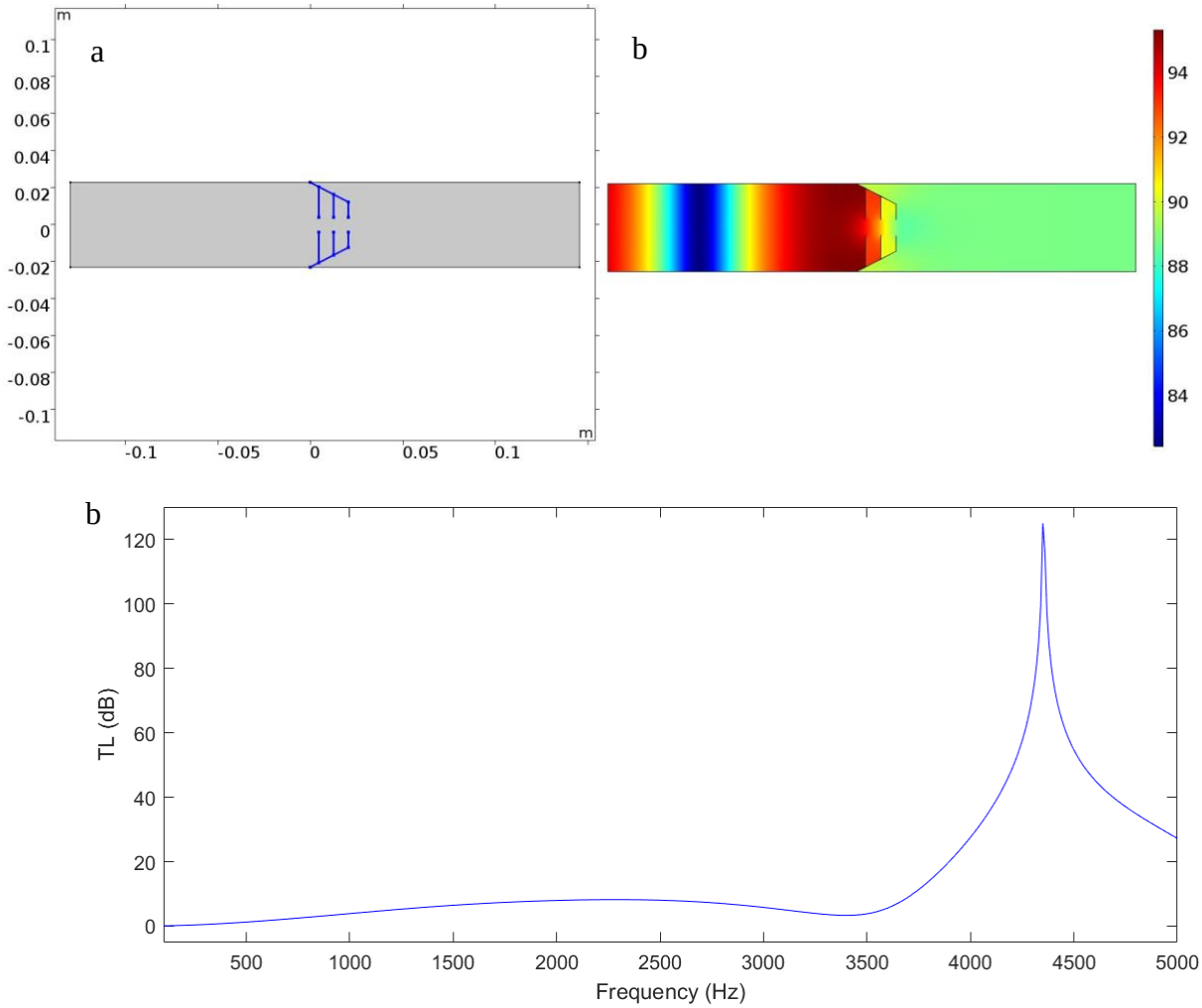



Figure 5. TL Parametrisation for conical duct as the one in Design 1: (a) geometrical setting (internal boundary in blue), (b) SPL distribution at peak frequency 4400 Hz, (c) TL.

3.3. *Comparison of the Different Scaled Models*

The TL increase is correlated with the dimension of the device. For the sake of completeness, wider samples of Design 1 are built and analysed through the same acoustic simulation settings. So results will be presented, comparing the performance of the original model with those of diameter equal to 0.4 and 0.8 m.

From Figure 6, it is clear that, as expected, the increasing of the dimensions (two and four times bigger in this case), causes a shift of the TL peak towards lower frequencies. In particular, for the folded configuration, in the 0.4 m model, the peak is at 4000 Hz (Figure 6) with TL of 83dB. This phenomenon happens consistently and progressively with the increasing of the models,

and it is demonstrated by the other study, with dimensions ten times bigger than the original ones. For the 0.8 m models, the peak is at 2000 Hz where SPL is 89dB.
From Figure 6, the effectiveness of the origami metacage it is demonstrated in this frequency range, and the contribution of the folded configuration in this process is proved.

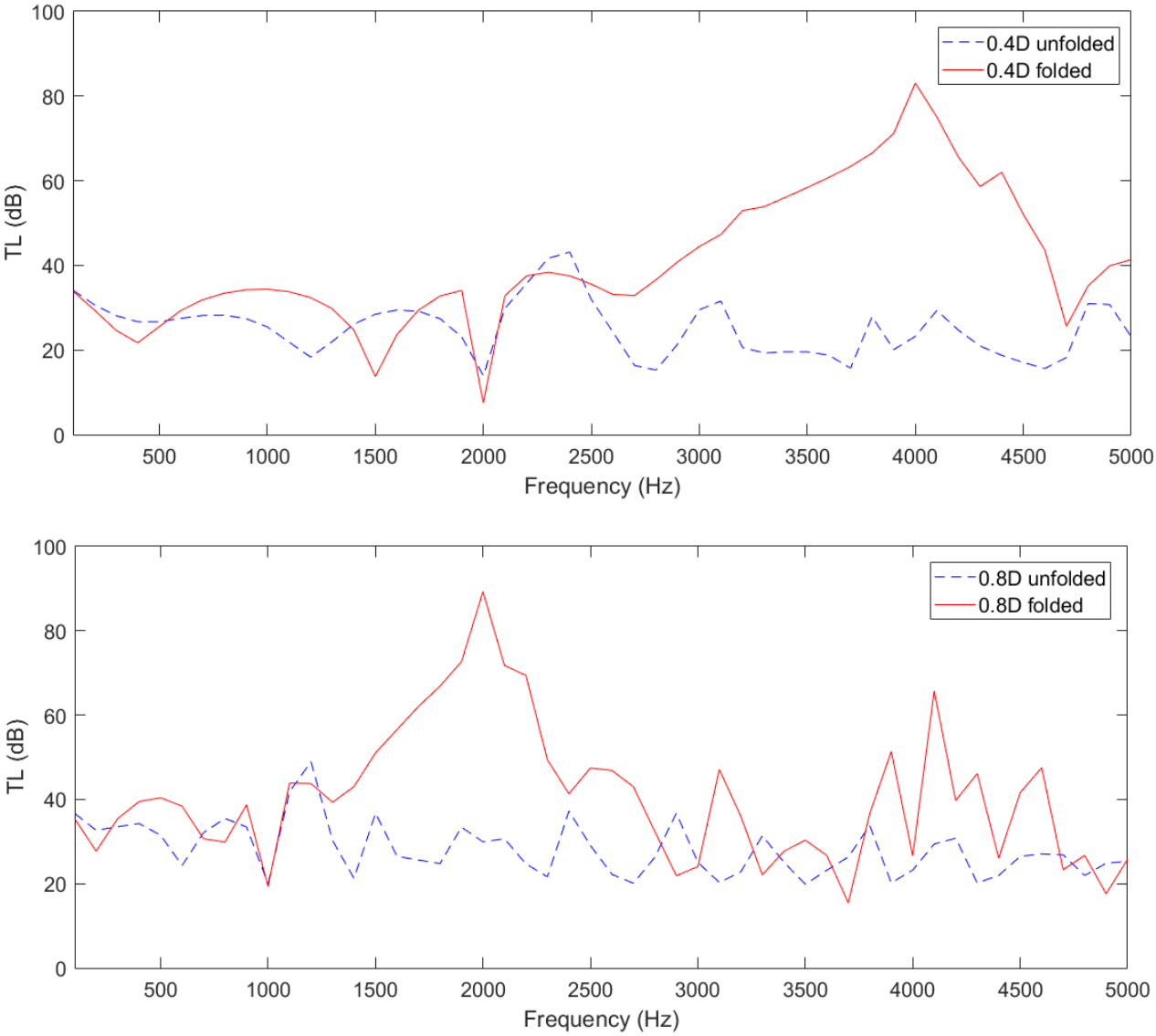



Figure 6. TL comparison of Design 1 (unfolded and folded) with different diameters dimensions (0.2, 0.4, and 0.8 m).

## 4. Numerical results of Design 2

### 4.1. *Design 2 (Bistable Auxetic)*

Figure 7 shows the simulation results of Design 2. For the SPL distribution, a slice graph is placed at the middle of the metasurface height, to compare the effect of the two models (Design 1 and Design 2). In both configurations (10° and 5° rotating angles), the TL behaviours are very similar but shifted on lower results for the 10° one. Overall, the TL is between 0 dB and 51 dB, where some variations can be observed due to the resonance of the openings of the cavities composing the duct. The TL has a sinusoidal behaviour at low-medium frequencies (500-1500 Hz, average of 7 dB), while in the upper-medium frequency range (1500-3000 Hz) it increases in efficacy, with a TL average of 23 dB and a TL peak of 51dB at 1700 Hz. Figure 7b highlights the confinement effect of the SPL at the TL peak frequency for both 10° and 5° configurations.

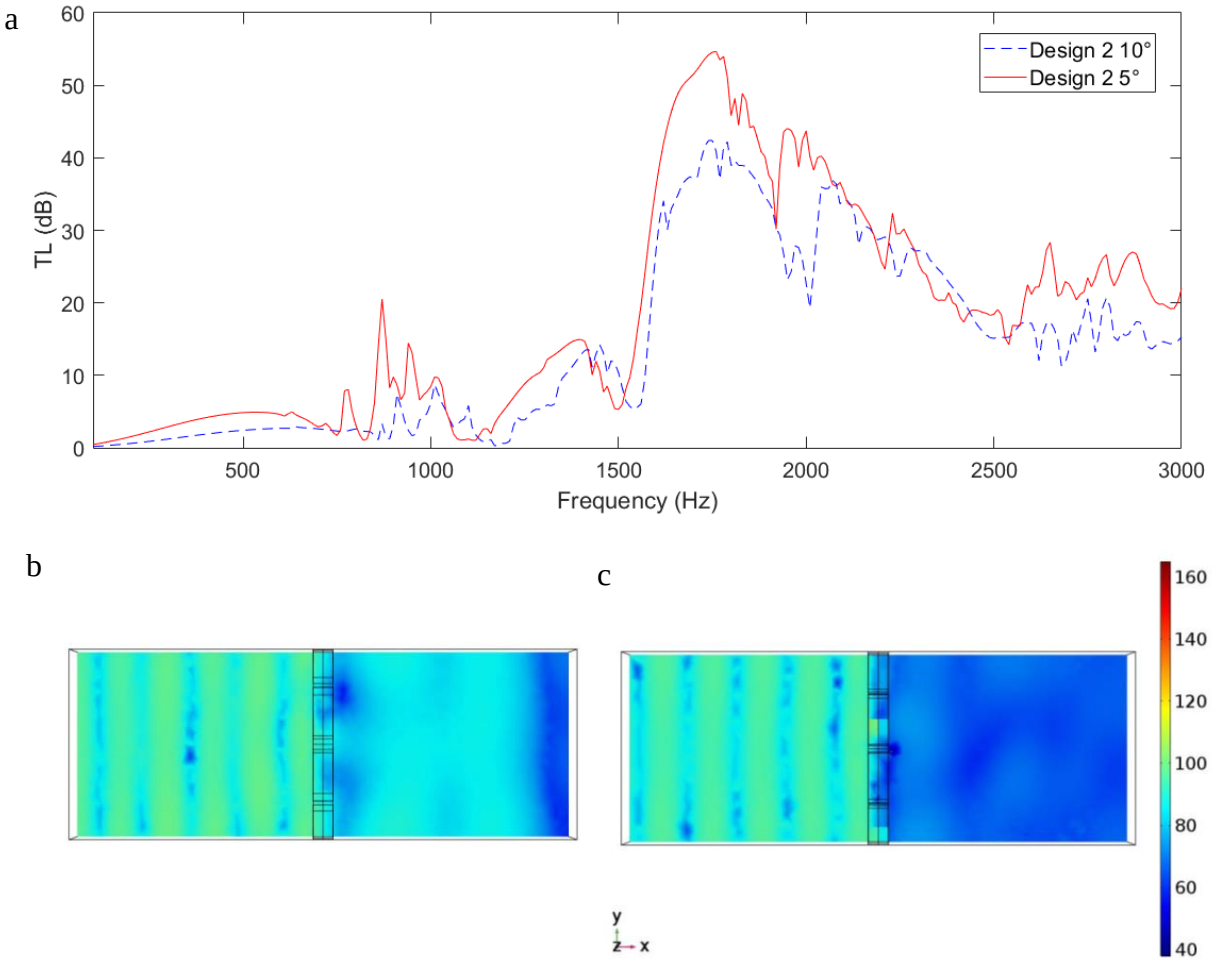



Figure 7. TL (a) and SPL distribution graph for Design 2 10° at 1800 Hz (b) and Design 2 5° at 1700 (c).

4.2. *Parametric study on cavities' dimension's ratio*

Figure 8 shows the average behaviours according to the two variables considered: cavities thickness and duct width. From the nine combinations of the three per three different options (see schematic in Figure 3): $a_1=b_1=0.01$ m, $a_2=b_2=0.02$ m, $a_3=b_3=0.03$ m. From the results, it is clear that either the cavities or the central duct width change do not affect the acoustic metasurface performance. So the configuration $a_2$ and $b_2$ can be set as standard ($a_2=b_2=0.02$ m) to guarantee a significant sound reduction performance and sufficient ventilation at the same time. Indeed, the acoustic and airwave propagation from the inside to the outside of the metacage and vice versa is guaranteed by the resulted duct of width 0.017 and cavities total thickness of 0.04 m.

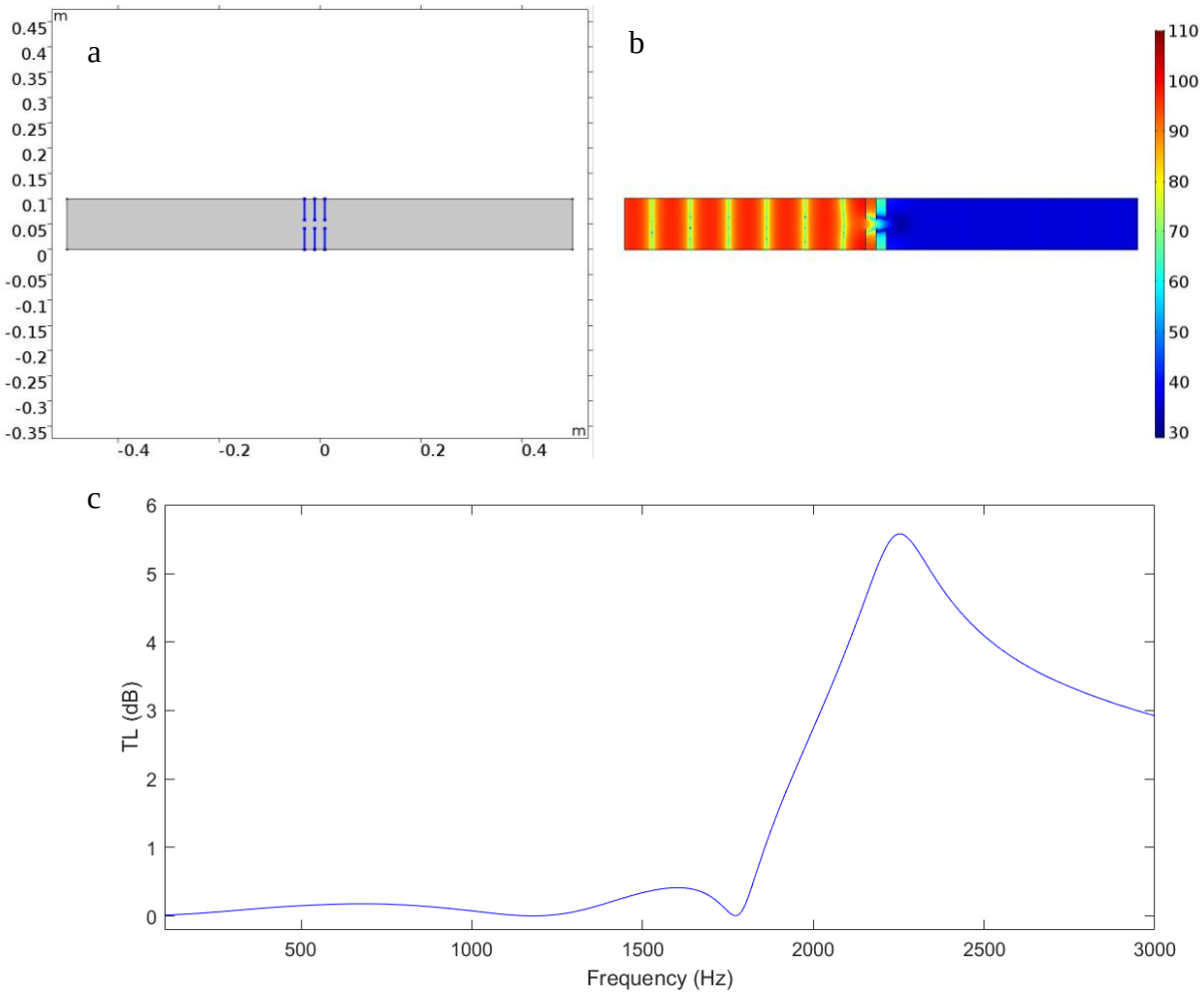



Figure 8. TL Parametrisation for straight duct reproducing the one in Design 2: (a) geometrical setting (internal boundary conditions in blue), (b) SPL distribution at peak frequency 2300 Hz, (c) TL.

4.3. *Comparison of the Different Scaled Models*

Differently from Design 1, the TL increase is not significantly correlated with the dimension of the device. For the sake of completeness, thicker samples of Design 2 are built and analysed through the same acoustic simulation settings. The results are presented in Figure 9, comparing the performance of the original model with those of overall thickness equal to 0.08 and 0.16 m. From Figure 9, it is clear that there is no shift of the TL peak while increasing the metasurface thickness. In particular, an interesting result is the different tendency of 10° and 5° configuration. While the TL associated with the first one decreases its amplitude around the peak frequency range (1500-2500 Hz), the second one increase significantly. The peak is always around 1700 Hz, but the amplitude is different. For the four-layer models, the TL peak is indeed 67dB (1760 Hz) for the 5° configuration, and it is averagely 38 dB for the 10° one (much spread and less concentrated than the previous one). The eight-layer model TL peak is lower again for the 10° configuration (averagely 40 dB), while increases with a maximum peak of 77 dB for 5° model at 1740 Hz.

Generally, from Figure 9, the effectiveness of the auxetic metasurface appears not to be connected significantly with the thickness increase. A part from the isolated peak, the graph shows a similar overall behaviour of TL related to the model with 2, 4, and 8 layers of Design 2. This means that, differently from Design 1, a two-layer model might be enough to allow natural ventilation and reduce noise on a broader frequency band.

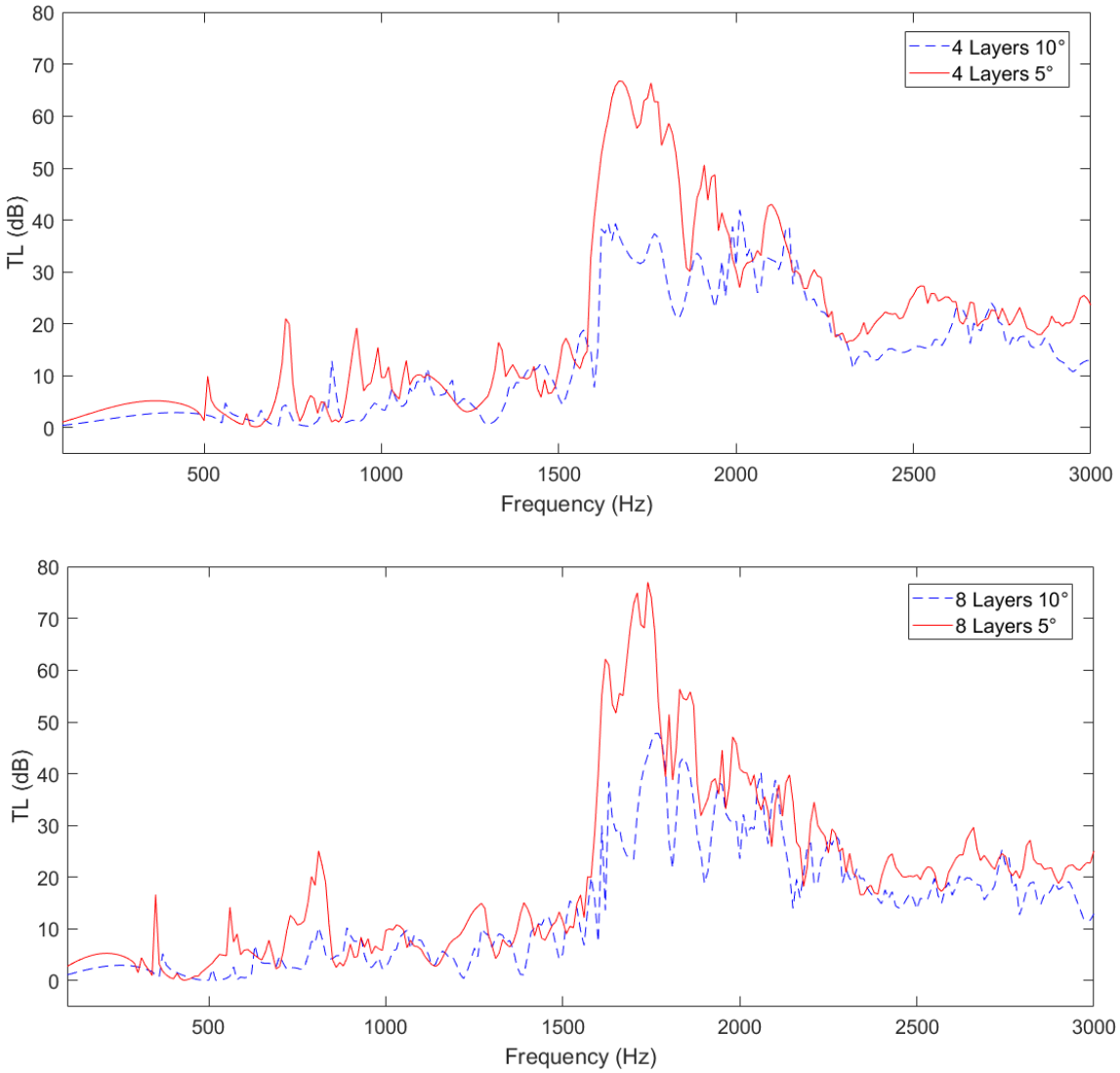



Figure 9. TL comparison of Design 2 (unfolded and folded) with different diameters dimensions (0.2, 0.4, and 0.8 m).

## 5. Discussion on possible design applications

The main aims of this research are achieved, and further numerical improvement or experimental study on the model will give more completeness to the research. So now new possibilities are open for devices' design which aim noise reduction together with natural ventilation. The proposed geometry may be embedded in a window design. The resulting TL broad peaks of 54, 67, and 77 dB (two, four and eight-layer model) might result of some impact in a situation where the application area is affected by high-level noise. An example for Design 1 could be to embed a system of origami metacage ducts in window frames or using the structure itself with increased width and a transparent back panel to allow also light exchange between two environments. The structure of Design 2 could be used as a transparent panel for windows enclosures. A further parametrisation and validation work will follow to test the actual building feasibility of the prototypes and determine whether the use of transparent materials might affect their performance or allow a new generation of tunable window systems.

## 6. Conclusions

The acoustic characteristics of the proposed acoustic metamaterials with a unique reconfigurable mechanism have been investigated. Different configurations with specific ventilation designs have been tested to assess the noise reduction and estimate the ventilation volume between the two areas separated by the devices: the inside and the outside for Design 1 (52% of opening ratio) and front and back for Design 2 (from 15% to 30% of opening ratio). Both models show high peaks in the TL due to the effective silencing effect provided by each metamaterial unit in front of the ventilation apertures. In different effective ways, the frequency and bandwidth of the effective region are related to the geometric parameters and scales of the systems. The potential of the proposed devices to be used in ventilation window systems is proved since natural ventilation is possible without any additional element. Better ventilation and noise reduction in the desired frequency range can be achieved by developing further numerical models for optimising the devices in terms of size and shape.

## REFERENCES

1. Lam B, Elliott S, Cheer J, Gan WS. Physical limits on the performance of active noise control through open windows. Appl. Acoust. 2018;137:9–17.
2. Kang J, Brocklesby MW. Feasibility of applying micro-perforated absorbers in acoustic window systems. Appl. Acoust. 2005 Jun 1;66(6):669–89.
3. Yu X, Lau S-K, Cheng L, Cui F. A numerical investigation on the sound insulation of ventilation windows. Appl. Acoust. 2017; 117:113–121
5. Yu X, Lu Z, Cheng L, Cui F. On the sound insulation of acoustic metasurface using a sub-structuring approach. J Sound Vib. 2017;401:190–203.
6. Yu X, Fang H, Cui F, Cheng L, Lu Z. Origami-inspired foldable sound barrier designs. J Sound Vib. 2019;442:514–26.
7. Babaee S, Overvelde JTB, Chen ER, Tournat V, Bertoldi K. Reconfigurable origami-inspired acoustic waveguides. Sci Adv. 2016;2(11):e1601019–e1601019.
8. Shen C, Xie Y, Li J, Cummer SA, Jing Y. Acoustic metacages for sound shielding with steady air flow. J Appl Phys. 2018;123(12).
9. Ma G, Sheng P. Acoustic metamaterials: From local resonances to broad horizons. Sci Adv. 2016;2(2):e1501595–e1501595.

# Natural convection in a nanofluid-filled enclosure with partially heated on side wall

**†Vai Kuong Sin, *Pengxiang Sui**

[1]Department of Electromechanical Engineering, University of Macau, Macau SAR, China.
[2]Department of Electromechanical Engineering, University of Macau, Macau SAR, China

*Presenting author: YB77437@um.edu.mo
†Corresponding author: vksin@um.edu.mo

## Abstract

The buoyant flow and heat transfer in the enclosure filled with nanofluid is investigated numerically in this paper. The heated source is located on the left side-wall with constant heat flux, the right side-wall kept constant temperature and the top and bottom walls of the enclosure are insulated. The governing equations are solved by using the finite element method. The influences of the variable Rayleigh number, volume fractions, length of the heated source and different types of nanofluids are studied. Results are performed as the streamline and isotherm plots as well as the variation of the local and mean Nusselt number. It is observed from the result that the increase of Rayleigh number enhances buoyant flow and causes increasing mean Nusselt number on the heated source. In addition, the increasing heated source length investigated significantly effect on maximum temperature along the heated source. It is shown that the increase of volume fraction improves the thermal conductivity of the nanofluid. Finally, the nanofluid has a better cooling performance by compared with pure water.

**Keywords:** Natural convection, Nanofluid, Partially heated source, Rayleigh number

## Introduction

Natural convection in the enclosure has wide applications in industry and engineering, such as electronic cooling system, biological sciences, material science and heat exchangers. Enhancement of heat transfer in natural convection is a necessary topic from saving energy and improving efficiency.

Majority of the present studies on thermal properties of the nanofluid presence the nanoparticles in the fluids can increase the effective thermal conductivity and enhance the heat transfer characteristics. Nanofluid is a dilute suspension of solid nanoparticles with a size typically of 1-100 nm dispersed in the liquid. The small solid nanoparticles at low volume fraction in liquid changes in physical properties, such as density, dynamic viscosity and specific heat, which can enhance the thermal conductivity and increase in critical heat flux in boiling heat transfer over the base-fluid value [1]-[4]. The nanofluids are a new class of heat transfer fluids by consisting of different nanoparticles. Masuda et al. [5] investigated different nanofluids CuO-water , $Al_2O_3$-water , $SiO_2$-water and $TiO_2$-water enhanced the thermal conductivity of nanofluids at a small volume fraction. This thermal property enhancement phenomenon was also reported by Eastman et al. [6].

Many studies of natural convection in enclosure are considered under the assumption of heated by constant temperature or heated by constant heat flux. Oztop and Abu -Nada [7] employed the finite volume method to investigate heat transfer and fluid flow due to the buoyancy forces in a partially heated rectangular enclosure for different types of nanofluids. The results indicated heat transfer enhancement by using nanofluids and was more pronounced at low aspect ratios than at high aspect ratios and the mean Nusselt number increased as increasing the volume fraction of nanoparticles for the entire range of Rayleigh number. Another study of natural convection in a square cavity with partially active side-walls filled with Cu-water nanofluid is investigated by Sheikhzadeh et al. [8]. The active heated sources are on both vertical side-walls of the cavity at the constant temperature. The effects of locations of the active heated sources, the Rayleigh number and the nanoparticle's volume fraction were studied and the results showed that the average Nusselt number increases with increasing both the Rayleigh number and the volume fraction of the nanoparticles.

Cheikh et al. [9] studied natural convection in air-filled square enclosure heated with the constant heat flux from below and cooled from above for different thermal boundary conditions. The simulations are performed for two different lengths of the heated sources and several Rayleigh number. The results are presented by streamline and isotherm plots and the maximum temperature along the heated source surface. Aminossadati and Ghasemi [10] also investigated a numerical study of natural convection cooling of the constant heat flux source on the bottom wall of the enclosure. Various Rayleigh number, location, and geometry of the heated source, different types of nanofluids and volume fractions are studied and the results showed nanofluids can enhance cooling performance, especially at low Rayleigh number, and the maximum temperature at heated source is based on the length and location of heated source, and type of nanofluids.

Despite several numerical studies and experimental which investigated natural convection in enclosure with different thermal boundary conditions. There is no study has been reported in the literature which investigated the enclosure heated by constant heat flux on one side-wall and cooling at constant temperature on the other side-wall. This problem may be encountered in a number of electronic cooling systems. In this paper, the aim of present study is to study fluid flow and heat transfer in the enclosure at a constant heat flux on the left side-wall and cooled from the right side-wall by constant temperature. The results are showed by streamline and isotherm plots with different Rayleigh number and volume fractions, the maximum temperature along heated source and mean Nusselt number also would be analyzed and discussed in this paper.

## Problem Description

Figure 1 shows the physical model of partially heated source enclosure with the length L. The heated source with constant heat flux $q''$ is located on the left side-wall of enclosure and the length of heated source is e. The distance of heated source center from the bottom wall is d. The right side-wall is kept at the constant low temperature $T_c$ and the top and bottom walls are thermally insulated. The enclosure is filled with nanofluid, which is assumed incompressible and flow as laminar within the enclosure. The thermo-physical properties of base fluid and various nanofluids are given in Table 1 [11]. Also, the nanofluid's thermo-physical properties are assumed constant except for its density variation which is depended on the Boussinesq approximation.

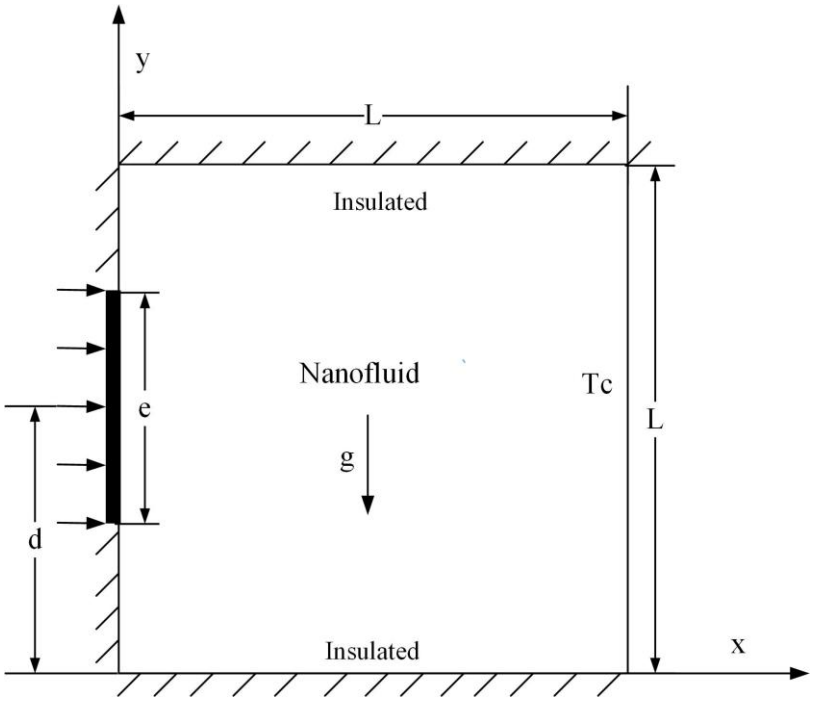



**Figure 1. Geometry and coordinate system of the enclosure**

**Table 1. Thermo-physical properties of the base fluid and various nanoparticles**

| | Density<br>$\rho\ (kg/m^3)$ | Specific Heat<br>$C_p\ (J/kgK)$ | Thermal Conductivity<br>$k\ (W/mK)$ | Thermal Expansion Coefficient<br>$\beta \times 10^{-5}\ (1/K)$ |
|---|---|---|---|---|
| Pure Water | 997.1 | 4179 | 0.613 | 21 |
| Copper | 8933 | 385 | 401 | 1.67 |
| Alumina | 3970 | 765 | 40 | 0.85 |
| Titanium Oxide | 4250 | 686.2 | 8.9538 | 0.9 |

**Governing Equation**

The steady-state natural convection in terms of continuity, momentum and energy equations for buoyancy-driven flow within the enclosure are given by:

Continuity equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \qquad (1)$$

where *u* and *v* are the velocity components in x and y direction

Momentum equations

$$\rho_{nf}\left(u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}\right)=-\frac{\partial p}{\partial x}+\mu_{nf}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right) \quad (2)$$

$$\rho_{nf}\left(u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}\right)=-\frac{\partial p}{\partial x}+\mu_{nf}\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+g(\rho\beta)_{nf}(T-T_C) \quad (3)$$

where $\rho_{nf}$ is the density of nanofluid, $\mu_{nf}$ is the dynamic viscosity of nanofluid, p is the pressure, β is the thermal expansion coefficient.

Energy equation

$$u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=\alpha_{nf}\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}\right) \quad (4)$$

where T is the temperature and α is the fluid thermal diffusivity.

The effective density of nanofluid is given as

$$\rho_{nf}=(1-\emptyset)\rho_f+\emptyset\rho_s \quad (5)$$

where $\rho_f$ is the density of fluid, $\rho_s$ is the density of solid nanoparticle and $\emptyset$ is the volume fraction.

The effective heat capacitance of the nanofluid can be determined by

$$(\rho C_p)_{nf}=(1-\emptyset)(\rho C_p)_f+\emptyset(\rho C_p)_s \quad (6)$$

where $C_p$ is the specific heat.

The effective thermal expansion coefficient of the nanofluid given as

$$(\rho\beta)_{nf}=(1-\emptyset)(\rho\beta)_f+\emptyset(\rho\beta)_s \quad (7)$$

The dynamic viscosity of the nanofluid given by the Brinkman [12] as

$$\mu_{nf}=\frac{\mu_f}{(1-\emptyset)^{2.5}} \quad (8)$$

The effective thermal conductivity of the nanofluid can be approximated by the Maxwell-Garnetts [13] as

$$k_{nf} = k_f \left[ \frac{k_s + 2k_f - 2\emptyset(k_f - k_s)}{k_s + 2k_f + \emptyset(k_f - k_s)} \right] \quad (9)$$

where $k_f$ and $k_s$ are the thermal conductivities of base fluid and nanoparticle.

The thermal diffusivity of the nanofluid is expressed by

$$\alpha_{nf} = \frac{k_{nf}}{(\rho C_p)_{nf}} \quad (10)$$

Equations (1) to (4) can be converted to dimensionless forms, and the dimensionless variables are (dimensionless coordinate X and Y, dimensionless velocity U and V, dimensionless pressure P, reference temperature different ΔT and dimensionless temperature θ)

$$X = \frac{x}{L}, \quad Y = \frac{y}{L}, \quad U = \frac{uL}{\alpha_f}, \quad V = \frac{vL}{\alpha_f}, \quad P = \frac{pL^2}{\rho_{nf}{\alpha_f}^2}, \quad \Delta T = \frac{q''L}{k_f}, \quad \theta = \frac{(T-T_c)k_f}{q''L} \quad (11)$$

The dimensionless governing equations of continuity, momentum and energy equations are given as follows

Dimensionless continuity equation

$$\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y} = 0 \quad (12)$$

Dimensionless momentum equations

$$U\frac{\partial U}{\partial X} + V\frac{\partial U}{\partial Y} = -\frac{\partial P}{\partial X} + \frac{\mu_{nf}}{\rho_{nf}\alpha_f}\left(\frac{\partial^2 U}{\partial X^2} + \frac{\partial^2 U}{\partial Y^2}\right) \quad (13)$$

$$U\frac{\partial V}{\partial X} + V\frac{\partial V}{\partial Y} = -\frac{\partial P}{\partial X} + \frac{\mu_{nf}}{\rho_{nf}\alpha_f}\left(\frac{\partial^2 V}{\partial X^2} + \frac{\partial^2 V}{\partial Y^2}\right) + \frac{(\rho\beta)_{nf}}{\rho_{nf}\beta_f} RaPr\theta \quad (14)$$

Dimensionless energy equation

$$U\frac{\partial \theta}{\partial X} + V\frac{\partial \theta}{\partial Y} = -\frac{\alpha_{nf}}{\alpha_f}\left(\frac{\partial^2 \theta}{\partial X^2} + \frac{\partial^2 \theta}{\partial Y^2}\right) \quad (15)$$

In above equations, the Rayleigh number and Prandtl number are defined by

$$Ra = \frac{gq''L^4\beta_f}{{v_f}^2 k_f} \times Pr, \qquad Pr = \frac{v_f}{\alpha_f} \quad (16)$$

where $v_f$ is the kinematic viscosity of fluid.

The boundary conditions for equations (12) – (15) in dimensionless form are as follows

$$\begin{aligned}
&\mathrm{U}=\mathrm{V}=0, \quad \frac{\partial\theta}{\partial X}=0 && \text{for } \mathrm{X}=0 \text{ and } 0\le \mathrm{Y}\le \mathrm{D}-0.5\mathrm{E} \text{ and } \mathrm{D}+0.5\mathrm{E}\le \mathrm{Y}\le 1\\
&\mathrm{U}=\mathrm{V}=0, \quad \frac{\partial\theta}{\partial X}=-\frac{k_f}{k_{nf}} && \text{for } \mathrm{X}=0 \text{ and } \mathrm{D}-0.5\mathrm{E}\le \mathrm{Y}\le \mathrm{D}+0.5\mathrm{E}\\
&\mathrm{U}=\mathrm{V}=0, \quad \frac{\partial\theta}{\partial Y}=0 && \text{for } \mathrm{Y}=1 \text{ and } 0\le \mathrm{X}\le 1\\
&\mathrm{U}=\mathrm{V}=0, \quad \theta=0 && \text{for } \mathrm{X}=1 \text{ and } 0\le \mathrm{Y}\le 1\\
&\mathrm{U}=\mathrm{V}=0, \quad \frac{\partial\theta}{\partial Y}=0 && \text{for } \mathrm{Y}=0 \text{ and } 0\le \mathrm{X}\le 1
\end{aligned} \qquad (17)$$

The local Nusselt number of heated source is calculated from the following equations

$$Nu_{local}=\frac{hL}{k_f} \qquad (28)$$

$$h=\frac{q''}{T_{local}-T_c} \qquad (19)$$

where $T_{local}$ is the local temperature along the heated source.

The local Nusselt number by using dimensionless variables are given by

$$Nu_{local}=\frac{1}{\theta_s(X)} \qquad (20)$$

The mean Nusselt number is determined by integrating $\mathrm{Nu}_{local}$ along the heated source on the left side-wall as

$$Nu_m=\frac{1}{E}\frac{k_{nf}}{k_f}\int_{Y_1}^{Y_2}\frac{1}{\theta_s(X)}dx \qquad (21)$$

where E is the dimensionless length of heated source (e/L), D is the dimensionless distant of heated source center from the bottom wall (d/L), and $Y_1=D-0.5E$ and $Y_2=D+0.5E$.

## Validation

COMSOL Multiphysics software is used to solve the governing equation, which employs Boussinesq term and dimensionless parameters for buoyant flow driven with the laminar flow and the heat transfer coupled interfaces. The present simulation results are also validated against the numerical results by Cheikh et al. [9] and S.M. Aminossadati et al. [10]. Figure 2 presents geometry of the air-filled square enclosure in study of Cheikh. Both two vertical side-walls cooled at constant temperature $T_c$ and the top and the bottom wall are insulated. The enclosure heated by the heated source from below. Table 2 gives the value of the mean Nusselt number along the heated source on the bottom wall of the enclosure, and the

maximum temperature within enclosure for various Rayleigh number and two different lengths of the heated source.

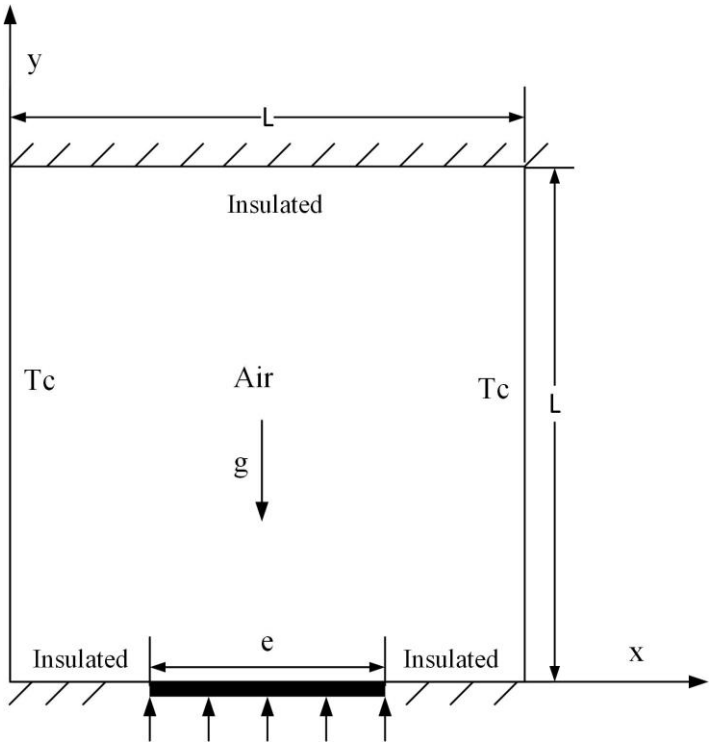



**Figure 2. Geometry and coordinate system of the enclosure in study of Cheikh** [9]

**Table 2. Comparisons of the present simulation results with the other studies**

| | | | Rayleigh Number | | | | |
|---|---|---|---|---|---|---|---|
| | | | $10^3$ | $10^4$ | $10^5$ | $10^6$ | $10^7$ |
| E=0.2 | $Nu_m$ | Present work | 5.9487 | 5.9808 | 7.6353 | 12.1218 | 19.9414 |
| | | Cheikh et al | 5.9152 | 5.9467 | 7.5805 | 12.0390 | 19.9801 |
| | | Aminossadati et al | 5.9228 | 5.9539 | 7.5910 | 12.0624 | 20.0195 |
| | $T_{max}$ | Present work | 0.1818 | 0.1815 | 0.1484 | 0.1042 | 0.0734 |
| | | Cheikh et al | 0.1819 | 0.1815 | 0.1484 | 0.1040 | 0.0730 |
| | | Aminossadati et al | 0.1819 | 0.1815 | 0.1484 | 0.1040 | 0.0729 |
| E=0.8 | $Nu_m$ | Present work | 3.5618 | 3.8156 | 6.3079 | 9.9082 | 16.6032 |
| | | Cheikh et al | 3.5532 | 3.8047 | 6.2942 | 9.9160 | 16.7432 |
| | | Aminossadati et al | 3.5551 | 3.8060 | 6.2944 | 9.9159 | 16.6779 |
| | $T_{max}$ | Present work | 0.3642 | 0.3634 | 0.2493 | 0.1702 | 0.1162 |
| | | Cheikh et al | 0.3642 | 0.3635 | 0.2494 | 0.1701 | 0.1163 |
| | | Aminossadati et al | 0.3642 | 0.3635 | 0.2495 | 0.1700 | 0.1160 |

Finally, for the various Rayleigh number considered. The maximum difference between the mean Nusselt number and the maximum temperature obtained by the present results and the above results are 0.72% and 0.68%, respectively.

**Grid Independence Study**

In order to select a proper grid density, various grids are employed to simulate the fluid flow and heat transfer within the enclosure at E=0.4, D=0.5 and Ra $= 10^6$. The nanofluid inside enclosure is Cu-water with ∅=0.1. The four different grid sizes are present in Table 3. The Grid size 3 changed grid shape from rectangular to triangular, that decreases the thickness of the boundary layer computation at higher Rayleigh number. Also the Grid size 3 increases grid element in two vertical side-walls, which due to the buoyant flow occur mainly on the two vertical side-walls. Figure 5 shows the result of local Nusselt number along the left heated source surface. The Grid size 3 has smoother curve of local Nusselt number than others especially at two end of the heated source.

**Table 3. Different gird sizes for grid independent study**

| Grid size | grid elements in each horizontal side-wall | grid elements in each vertical side-wall | grid shape | Total grid elements |
|---|---|---|---|---|
| Grid size 1 | 20 | 20 | rectangular | 400 |
| Grid size 2 | 40 | 40 | rectangular | 1600 |
| Grid size 3 | 80 | 100 | triangular | 19116 |
| Grid size 4 | 120 | 120 | rectangular | 14400 |

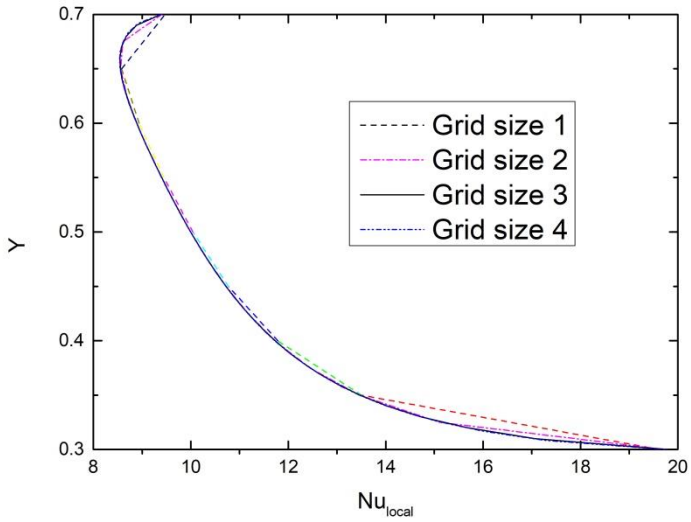



**Figure 3. The local Nusselt number along the left heated source for various grid sizes (Cu-water, ∅ $= 0.1$, Ra $= 10^6$, E=0.4, D=0.5)**

Table 4 presents the result of the maximum temperature within the enclosure for different Rayleigh numbers and grid sizes. For each Rayleigh number, the value of maximum temperature within enclosure increases from Grid size 1 to Grid size 4, but the value between Grid size 3 and Grid size 4 are almost the same. Finally, the simulations show that the Grid size 3 is sufficiently fine to describe the buoyant flow inside the enclosure.

**Table 4. The maximum temperature within enclosure for different Rayleigh numbers and grids (Cu-water, $\emptyset = 0.1$, E=0.4, D=0.5)**

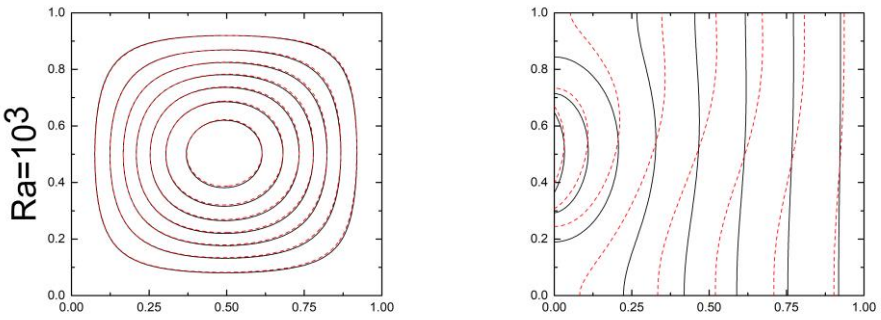
| Grid size | $Ra = 10^3$ | $Ra = 10^4$ | $Ra = 10^5$ | $Ra = 10^6$ | $Ra = 10^7$ |
|---|---|---|---|---|---|
| Grid size 1 | 0.37399 | 0.31321 | 0.19211 | 0.11654 | 0.06976 |
| Grid size 2 | 0.37441 | 0.31331 | 0.19223 | 0.11688 | 0.07235 |
| Grid size 3 | 0.37441 | 0.31329 | 0.19247 | 0.11702 | 0.07244 |
| Grid size 4 | 0.37442 | 0.31331 | 0.1925 | 0.1171 | 0.07255 |

## Results and Discussion

The nanofluid-filled enclosure is studied by three different lengths (E=0.2, E=0.4 and E=0.8) and different types of nanofluids and various volume fractions. The Prandtl number is chosen to be 6.2 for all simulations.

Figure 4 shows the streamlines (on the left) and isotherms (on the right) for Cu-water ($\emptyset$=0.2) nanofluid (plotted by solid line) and pure water (plotted by dashed line) at various Rayleigh numbers when E=0.4 and D=0.5. Since the left and right vertical walls of the enclosure are located heated and cooled sources, which causes buoyant flow along this two side-walls. As the Rayleigh number increases, the buoyant flow strength also increases that causes the boundary layers to become more distinguished. When Rayleigh number equal to $10^6$, Cu-water presents an oval shaped rotating cell in the right-center as shown Figure 4(d). For Ra=$10^7$, Figure 4(e) presents the oval shaped rotating cell break up into two small rotating cells, one flows near the right side-wall and another was observed near the left side-wall.

Figure 4(a) and (c) shows the isotherm is distributed gather near the heated source on the left side-wall and tends to be paralleled near the right cooled side-wall. Equation (17) presents $\frac{\partial \theta}{\partial X} = -\frac{k_f}{k_{nf}}$ depends on the ratio of thermal conductivity along the heated source surface (for X=0 and D-0.5E≤Y≤D+0.5E). The value of $\frac{\partial \theta}{\partial X}$ equal to -1 and -0.5897 for pure water and Cu-water nanofluid, which means the tangent angle of isotherm to the X direction along heated source are $-45^{\circ}$ and $-30.528^{\circ}$ for pure water and Cu-water nanofluid, respectively. Cu-water's isotherms present shaper than pure water.




(a)

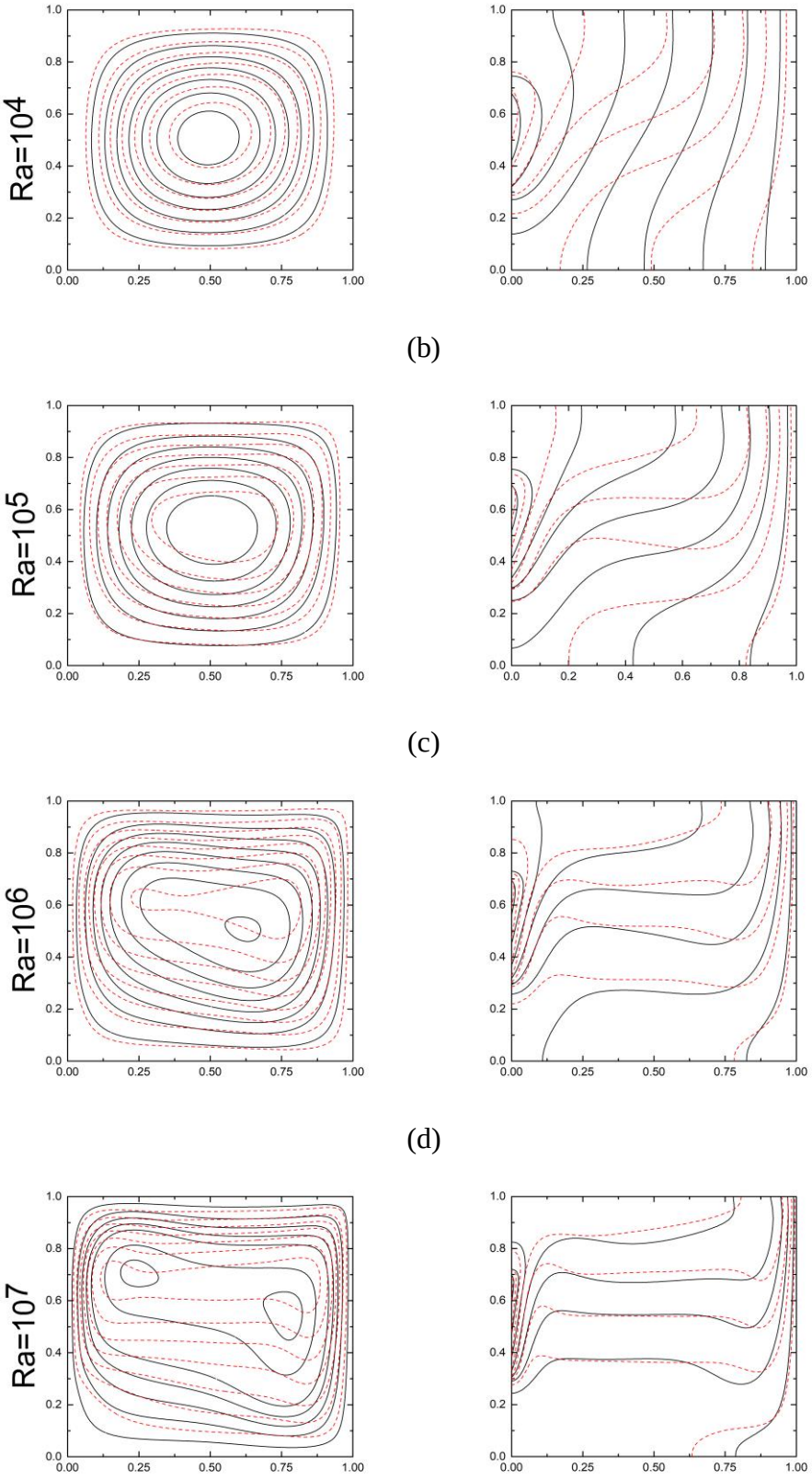



**Figure 4. Streamline (on the left) and isotherm (on the right) of pure water (plotted by dashed line) and Cu-water (Ø=0.2) nanofluid (plotted by solid line) at various Rayleigh numbers (E=0.4, D=0.5, (a)Ra=$10^3$, (b)Ra=$10^4$, (c)Ra=$10^5$, (d)Ra=$10^6$, (e)Ra=$10^7$ )**

The result of local Nusselt number along the heated source for different volume fraction is shown in Figure 5. It is noted that the maximum value of the local Nusselt number is presented as $\emptyset = 0.2$ and the minimum value as shown for $\emptyset = 0$. As the increasing volume fraction, the value of the local Nusselt number also increases that due to the enhancement of convection within the enclosure.

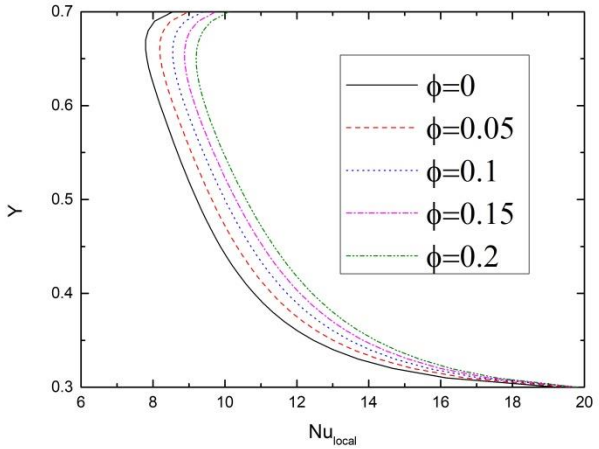



**Figure 5. The local Nusselt number on the heated source for various volume fraction (Cu-water, Ra=$10^6$, D=0.5, E=0.4)**

Figure 6 presents the result of the maximum surface temperature on the heated source for various volume fraction and Rayleigh number. The surface temperature on heated source is not uniform and the maximum value is reduced by increasing volume fraction for all range Rayleigh number. The maximum temperature decreases much more rapidly for a higher volume fraction. This reduction is due to the higher volume fraction improves the thermal conductivity of the nanofluid and shifts the heat transfer from conduction to convection inside the enclosure.

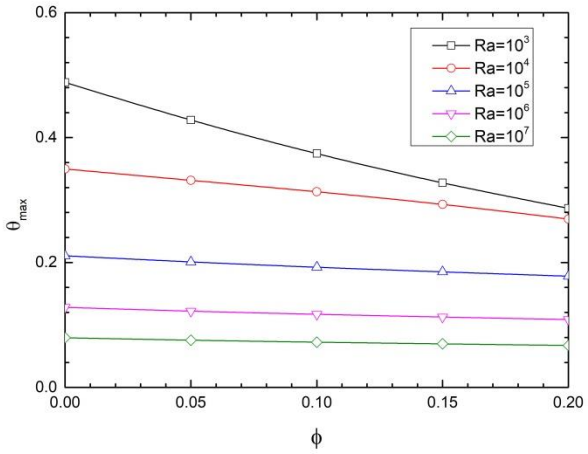



**Figure 6. The maximum temperature on the heated source for various volume fraction and Rayleigh number (Cu-water, D=0.5, E=0.4)**

Figure 7 exhibits the streamlines (on the left) and isotherms (on the right) of Cu-water (Ø=0.1) nanofluid at various heated source's length. The comparison between 0.8E heated source length (plotted by solid line) and 0.2E heated source length (plotted by dashed line) are presented that counter rotating cells are formed in clockwise direction for both situations. As the increasing heated source length, the cell boundaries become more distinguished on side-walls, which mean the buoyant flow strength increases in the enclosure. Isotherms show that temperature gradients near 0.8E heated source become more sever along both vertical side-walls that can be explained as the higher heat transport rate is generated by increasing heated source.

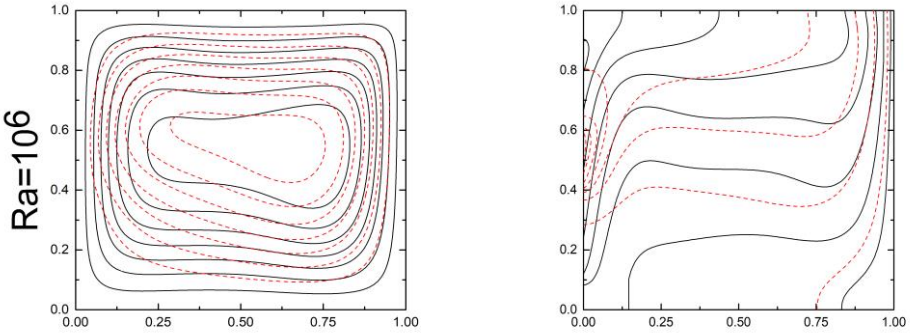



**Figure 7. Streamlines (on the left) and isotherms (on the right) of Cu-water (Ø=0.1) nanofluid for different heated source length, E=0.2 (plotted by dashed line) and E=0.8 (plotted by solid line) (D=0.5, Ra=$10^6$)**

Figure 8 shows the results of the maximum temperature at heated source for different heated source lengths. As increases the heated source's length, the maximum temperature also increases for all range Rayleigh number. It is due to the longer heated source length causes higher heat flux transfer. But the maximum temperature decreases as Rayleigh number increases that can be explained as more heat flux away from heated source surface by stronger flow strength.

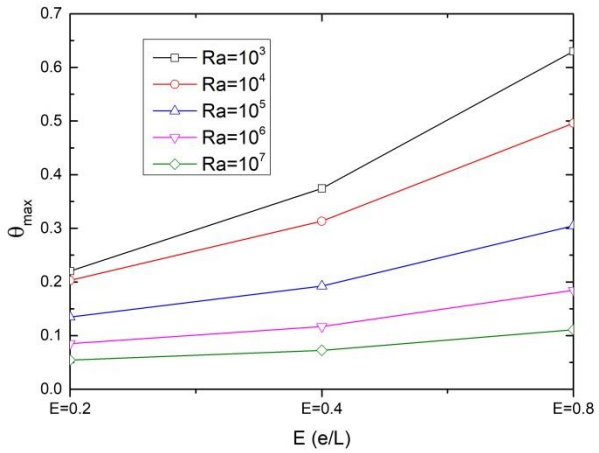



**Figure 8. The maximum temperature on the heated source for various heated source lengths and Rayleigh number (Cu-water nanofluid, Ø=0.1, D=0.5)**

In order to study the fluid flow inside the enclosure, Figure 9 presents the result of U-velocity (component velocity in x-direction) in mid-section centerline of enclosure (from (0.5, 0) to (0.5, 1)) for various heated source lengths. The strength of fluid flow enhances near two vertical side-walls and descends in the center of the enclosure, which due to stronger buoyant flow near the two side-walls. The U-velocity increases by increasing the length of heated source because of higher heat generation rates between left and right side-walls.

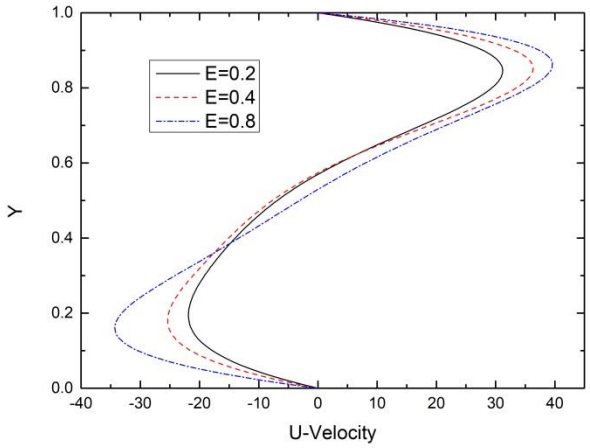



**Figure 9. U-velocity at mid-section V (0.5, Y) of the enclosure for various heated source lengths (Cu-water, Ø=0.1, D=0.5, Ra=$10^6$)**

Figure 10 presents the result of local Nusselt number along the heated source for different types of nanofluids. Cu has the highest local Nusselt number value and the lowest value is pure water. Compared between three different nanofluids, $TiO_2$ has the lowest local Nusselt number because it has the lowest value of thermal conductivity as shown in Table 1. On the other hand, Cu has the highest value of thermal conductivity that indicates the local Nusselt number increases by increasing thermal conductivity.

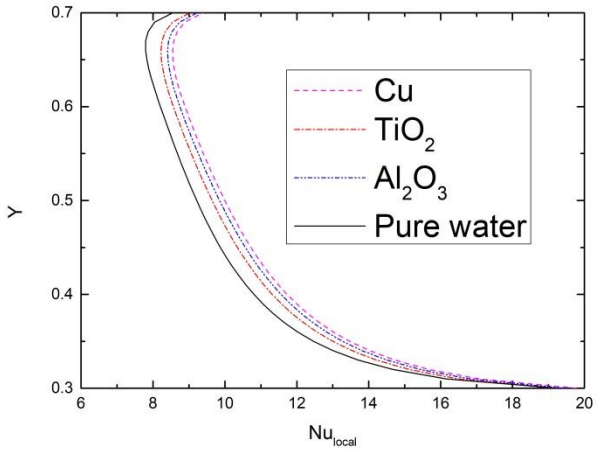



**Figure 10. The local Nusselt number on the heated source for various types of nanofluids (D=0.5, E=0.4, Ra=$10^6$, Ø=0.1)**

The result of mean Nusselt number on the heated source for different types of nanofluids are shown in Figure 11. The mean Nusselt number increases with increasing volume fractions and the lowest value of mean Nusselt number was generated by $TiO_2$ for all range Rayleigh number as the same situation as the local Nusselt number results. That also due to $TiO_2$ has lowest thermal conductivity value compared with the other nanoparticles. Figure 11(f) exhibits the different of mean Nusselt number become larger as increasing Rayleigh number that due to high Rayleigh number causes stronger heat transfer within the enclosure. Moreover, the value of thermal conductivity of $Al_2O_3$ is approximately one tenth of Cu (Table 1), but $Al_2O_3$ has low thermal diffusivity that causes the mean Nusselt number of $Al_2O_3$ is lower than that for Cu at high Rayleigh number but almost the same at low Rayleigh number.

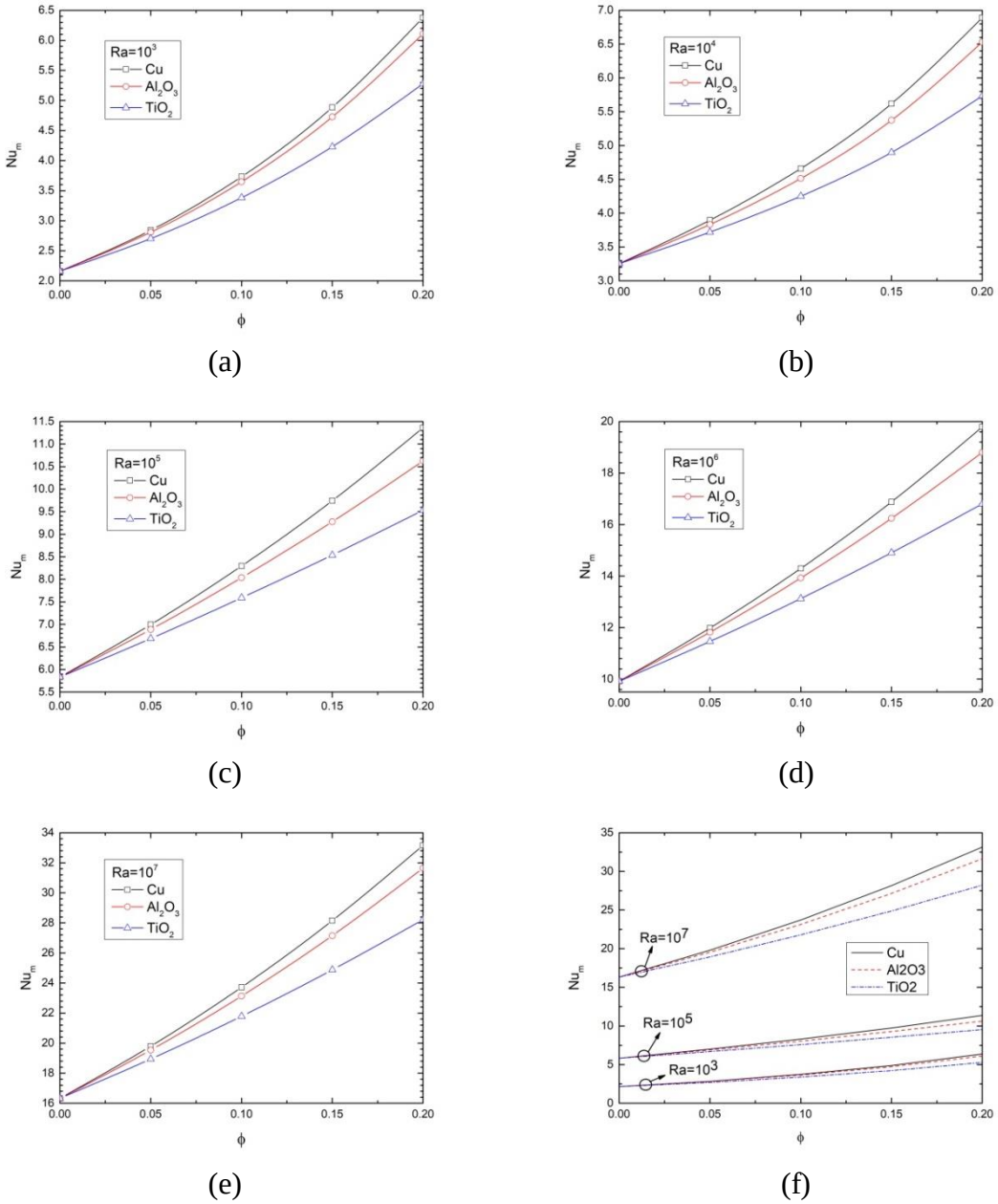



**Figure 11. The mean Nusselt number on the heated source for different types of nanofluids as various volume fractions and Rayleigh number (D=0.5, E=0.4, (a)Ra=$10^3$, (b)Ra=$10^4$, (c)Ra=$10^5$, (d)Ra=$10^6$, (e)Ra=$10^7$, (f)Ra=$10^3$, $10^5$ *and* $10^7$ )**

Figure 12 shows result of V-velocity at mid-section centerline of the enclosure (from (0, 0.5) to (1, 0.5)) for various types of nanofluids. The V-velocity indicated a parabolic variation near two vertical side-walls and almost zero at X from 0.4 to 0.6 that means the fluid flow almost stopped in the center of the enclosure. The V-velocity of nanofluids is lower than the pure water that can be explained as nanoparticle's suspension affects the buoyant flow. To clarify the effects of nanofluid on reduction of temperature along heated source surface, Table 4 displays the maximum temperature on the heated source for various nanofluids as different volume fractions. Compared with the pure water, nanofluids provide temperature reduction on the heated source surface and Cu has the best cooling performance.

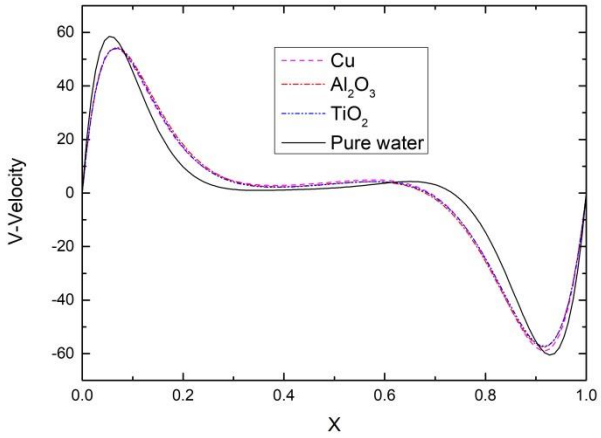



**Figure 12. V-velocity at mid-section V(X, 0.5) of the enclosure for different types of nanofluids (D=0.5, E=0.4, Ra=$10^6$ and Ø=0.1)**

**Table 4. The maximum temperature on heated source (E=0.4, D=0.5) for pure water and different types of nanofluids at various Rayleigh number and volume fraction**

| | | Rayleigh Number | | | | |
|---|---|---|---|---|---|---|
| | | Ra=$10^3$ | Ra=$10^4$ | Ra=$10^5$ | Ra=$10^6$ | Ra=$10^7$ |
| | Pure water | 0.488 | 0.350 | 0.211 | 0.128 | 0.080 |
| Ø=0.05 | Cu | 0.428 | 0.332 | 0.201 | 0.122 | 0.076 |
| | $Al_2O_3$ | 0.431 | 0.335 | 0.203 | 0.123 | 0.076 |
| | $TiO_2$ | 0.439 | 0.339 | 0.205 | 0.125 | 0.077 |
| Ø=0.1 | Cu | 0.374 | 0.313 | 0.192 | 0.117 | 0.072 |
| | $Al_2O_3$ | 0.379 | 0.319 | 0.196 | 0.119 | 0.073 |
| | $TiO_2$ | 0.393 | 0.327 | 0.201 | 0.122 | 0.075 |
| Ø=0.15 | Cu | 0.328 | 0.293 | 0.185 | 0.113 | 0.070 |
| | $Al_2O_3$ | 0.333 | 0.301 | 0.191 | 0.115 | 0.071 |
| | $TiO_2$ | 0.352 | 0.313 | 0.197 | 0.119 | 0.073 |
| Ø=0.2 | Cu | 0.287 | 0.270 | 0.178 | 0.109 | 0.067 |
| | $Al_2O_3$ | 0.293 | 0.278 | 0.186 | 0.113 | 0.069 |
| | $TiO_2$ | 0.315 | 0.295 | 0.194 | 0.117 | 0.072 |

## Conclusions

Natural convection in a nanofluid-filled enclosure with partially heated source on the left side-wall has been numerically simulated. The effect of heat transfer and fluid flow within the enclosure by various volume fractions, various heated source lengths and different types of nanofluids were investigated. The results show as

- The increase of Rayleigh number causes increasing mean Nusselt number on heated source and enhances the fluid flow and temperature gradient on the two vertical side-walls.
- Increasing the length of heated source enhances heat generation rates, which causes both the maximum temperature on the heated source surface and flow strength within the enclosure increased.
- The increase of volume fractions improves the thermal conductivity of the nanofluid and shifts more heat flux from conduction to convection, that observed by the increasing local Nusselt number and maximum temperature on the heated source.
- Nanofluids (Cu, $Al_2O_3$ and $TiO_2$) reduce the fluid flow strength inside the enclosure and the maximum temperature on heated source surface. The best cooling performance by using Cu-water because it has the highest value of thermal conductivity among the three nanofluids.

**Reference**

[1] Eastman J.A., Phillpot S.R., Choi S.U.S., Keblinski P.. (2004) THERMAL TRANSPORT IN NANOFLUIDS, *Annual Review of Materials Research, Vol.34*, 219–246.
[2] Das S.K., Choi S.U.S., Patel H.E.. (2006) Heat Transfer in Nanofluids-A Review, *Journal Heat Transfer Engineering, volume27,2006-Issue 10 Journal*, 3–19.
[3] Kleinstreuer C. and Feng Y.. (2011) Experimental and theoretical studies of nanofluid thermal conductivity enhancement: a review, *A SpringgerOpen Journal*, 1–13.
[4] Keblinski P., Eastman J.A., Cahill D.G.. (2005) Nanofluid for thermal transport, *Materialstoday,Volume8, Issue6*, 36–44.
[5] Masuda H., Ebata A., Teramea K. and Hishinuma N.. (1993) Alteration of thermal conductivity and viscosity of liquid by dispersing ultra-fine particles, *Netsu Bussei 1993,4*, 227-223.
[6] Eastman J.A., Choi U.S., Li S., Thompson L.J., Lee S.. (1997) Enhanced thermal conductivity through the development of nanofluid, *In Nanophase and Nanocomposite Materials II*, 3-11.
[7] Oztop H.F. and Abu-Nada E.. (2008) Numerical study of natural convection in partially heated rectangular enclosures filled with nanofluid, *International Journal of Heat and Fluid Flow 29 (2008)*, 1326-1336.
[8] Sheikhzadeh G.A., Arefmanesh A., Kheirkhah M.H., Abdollahi R.. (2010) Natural convection of Cu-water nanofluid in a cavity with partially active side walls, *European Journal of Mechanics B/Fluids 30 (2011)*, 166-176.
[9] Cheikh N.B., Beya B.B., Lili T.. (2007) Influence of thermal boundary conditions on natural convection in a square enclosure partially heated from below, *International Communications in Heat and Mass Transfer 34 (2007)*, 369-379.
[10] Amoinossadati S.M. and Ghasemi B.. (2009) Natural convection cooling of a localized heat source at the bottom of a nanofluid-filled enclosure, *European Journal of Mechanics B/Fluid 28 (2009)*, 630-640.
[11] Abu-Nada E., Masoud Z., Hijazi A.. (2008) Natural convection heat transfer enhancement in horizontal concertric annuli using nanofluids, *Int. Comm. Heat mass Transfer 35 (5) (2008)*, 657-665.
[12] Brinkman H.C.. (1952) The viscosity of concentrated suspensions and solution, *H. Chem. Phys. 20*, 571-581.
[13] Maxwell J. (1904) A Treatise on Electricity and Magnetism, *Second ed. Oxford University Press, Cambridge,UK*.

# Exploiting Symmetry in Elemental Computation and Assembly Stage of GPU-Accelerated FEA

***Subhajit Sanfui[1], †Deepak Sharma[1]**

[1]Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Assam, India

*Presenting author: s.sanfui@iitg.ac.in
†Corresponding author: dsharma@iitg.ac.in

## Abstract

Finite element analysis (FEA) is one of the most popular numerical methods for solutions of boundary value problems in partial differential equations (PDEs). Owing to the computational load of handling fine meshes and complex geometries in many real-world problems coupled with the data-parallel and throughput-intensive nature of FEA computation, it is considered a highly suitable candidate for Graphics processing units (GPU) based acceleration. This work aims to accelerate the elemental stiffness calculation and assembly stages of FEA on a single GPU by exploiting the symmetric nature of the elemental and global stiffness matrix. The key idea of the implementation is to design GPU kernels that calculate and assemble only the lower triangular part of the local stiffness matrix into the global stiffness matrix. This leads to (a) reduced FLOP count and (b) reduced memory storage and access, which, in turn results in an overall reduced execution time. The elemental stiffness matrix is computed by assigning each GPU thread to one entry of the elemental stiffness matrix. The assembly stage is performed in two steps for computing the indices and values of the global stiffness matrix. For handling the race condition in assembly, mesh coloring and atomics-based approaches are implemented. The results from the proposed implementation is compared to a standard GPU-based version of the implementation that computes and assembles the complete elemental stiffness matrices. Hexahedral meshes with up to three million nodes are tested for the performance analysis. The proposed implementation is found to be approximately twice as fast as the standard implementation for both elemental calculation and assembly stages.

**Keywords:** Computation, Finite Element Method, Graphics Processing Units, Symmetry

## Introduction

Finite Element Method (FEM) is a numerical method for approximating solutions of boundary value problems for partial differential equations (PDEs). It is extensively used in fields such as mechanical engineering, civil engineering, electrical engineering, medical applications. Due to its natural advantages such as flexibility, adaptability and ease of implementation even for complex geometries, it has become an integral part of a large variety of specializations. In industries like aviation, automotive and construction, FEM is usually an inherent part of the design process [1].

Although finite element analysis is widely popular in several fields, it often suffers from a high computational intensity, especially for real-world problems. This is primarily caused by the resolution of system of equations performed in the solver step [2]. Several applications [3][4] that make use of FEM, have reported it to be the most time consuming or computationally expensive part of the whole applications especially for larger and more complex geometries. Due to these reasons, many practical applications see the need for

efficient parallelization of FEA. Among the several steps in a complete FEA, the numerical solver, global matrix assembly and elemental computation are the three most time-consuming and computationally expensive in nature. In the present work, the focus is kept on the elemental calculation and assembly into a global stiffness matrix by parallelization on a single GPU. For the numerical solver stage, multiple optimized libraries exist that can provide solution to the linear system of equation after application of proper boundary conditions.

Numerical Integration, as reported by many researchers, is a suitable candidate for GPU acceleration. For low order finite elements, the numerical integration becomes comparatively easy, whereas for high order elements, the computation turns equally tedious. This is because higher order elements necessitate higher number of Gauss points, which in turn increases the memory requirements while also increasing the amount of computation. It was demonstrated in [5], that using a $10^{th}$ order Gaussian Quadrature, the numerical integration step requires 73% - 83% and 87% of the total time of matrix generation step on a CPU and GPU respectively. For low orders of approximations, on the other hand, values of shape functions and their derivatives may be pre-computed and stored for reuse since the necessary storage requirements would be sufficiently small. Such an implementation, sadly, for higher order elements is unrealizable with present hardware. The first work dedicated toward numerical integration on the GPU is by using the Gauss-Legendre Quadrature Method [6]. Authors demonstrated the complete scalability of the numerical integration process on the GPU. One limitation of this study was that it was conducted entirely in single-precision, which can pose serious problems for convergence in the solver stage. This is specifically true for iterative solvers such as CG solver, which are known to be notoriously sensitive to round-off errors. A maximum speedup of 20x was achieved for third order approximation on an NVIDIA GeForce GTX8800 compared to an AMD X2 at 2.4 GHz. Authors also concluded that the massive amount of parallelism was not fully realized due to the insufficient memory resources in individual threads. This finding was later supported by Dziekonski et al. [7], where several strategies on efficient generation and assembly of large finite-element matrices were presented, while maintaining the desired level of accuracy in numerical integration. Authors used higher-order curvilinear elements making the integration step more compute intensive. Again the method of Gauss Quadrature was used for analysis. A speedup of 2.5x was achieved on an NVIDIA Tesla C2075 over two 12-core AMD Opteron 6174 at 2.2 GHz. Among more recent works, Banas [8][9] addressed the problem of implementing numerical integration that is portable across several GPU architectures and different orders of approximation. This is in general difficult to achieve because of vastly varying memory size, memory hierarchy and computational resources available to the programmer across different GPU vendors such as NVIDIA, AMD and Intel. The OpenCL implementation in [8] achieved a maximum speedup of 4x tested on four GPUs with different architectures (NVIDIA GeForce GTX 580, Nvidia Tesla M2075, AMD HD5870 and AMD HD7950). The portable OpenCL implementation by Banas [9], targeting numerical integration for only first-order approximation, achieved a maximum speedup of approximately 9x when tested on a Tesla K20m, compared to an Intel Xeon E5 2620 CPU. Authors provided details on several optimization aspects for implementations on different target hardware, while also indicating the factors limiting the performance for different problem types on different architectures. Apart from these, many works concentrating on applications of FEA have also implemented numerical integration on the GPU [3][4][10]. However, in these cases, no relevant details on the implementation of numerical integration was provided.

There has been only a handful of works that target the assembly stage of FEA on the GPU, despite of the inherent parallelism in this step. This is primarily because of this step being

significantly less compute-intensive than the matrix-solver step. In other words, even a small to medium speedup in the linear solver stage would benefit the FEA process more than a decent to good speedup in the assembly step. The first work concerning assembly on GPU is by Filipovic et al. [11], where a speedup of 15x was achieved on an NVIDIA GeForce GTX280 compared to an Intel Core2Quad Q9550 CPU at 2.83 GHz. One important observation by the authors was that using one single kernel for the entire computation results in massive under-utilization of the GPU. Cecka et al. [2] has studied several aspects of assembly on an unstructured mesh using single precision arithmetic. Different algorithms for efficient use of global, shared and local memory available on the GPU along with methods to achieve memory coalescing are introduced by the authors. Four different implementations were analyzed for assembly. These are respectively, assembly by elements using graph-coloring, by NZ using shared memory, by NZ using local memory and by NZ using global memory. Among these, the first two were shown to be the most efficient. A speedup of about 35x was achieved for the best implementation on an NVIDIA GTX8800 compared to an Intel Core2 Quad Q9450 CPU at 2.66 GHz. Later, Dziekonski et al. [5] used an NVIDIA Tesla C2075 to accelerate both numerical integration and matrix assembly and achieved combined speedups of 81x and 19x over single and multi-threaded implementations respectively on a 12-core Opteron 6174. Markall et al. [12] have discussed several assembly strategies on multi-core and many-core architectures. Among more recent studies, Dinh and Marechal [13] studied a real-time FEM implementation on GPU, where a sorting-based implementation of parallel global assembly was performed. An implementation based on the principle of dividing the GPU assembly with standard sparse formats was presented by Sanfui and Sharma [14]. The authors used structured meshes with brick elements to demonstrate the advantage of workload division at the assembly stage. The implementation divided the assembly operation into a separate symbolic and a numeric kernel. Later, Zayer et al. [15] accelerated assembly of sparse matrices by modifying the assembly stage as a matrix-matrix multiplication with the aim to remove any CPU or GPU-based preprocessing. This approach enabled them to reduce storage and movement of data on the GPU. Among more recent works, Kiran et al. [16] presented a warp-based assembly approach for hexahedral elements in single precision where the numerical integration and assembly were performed in the same kernel. An implicit finite element model with cohesive zones and collision response was accelerated using CUDA by Gribanov et al. [17]. For handling the race condition in assembly, instead of coloring the elements, *atomicAdd* function was used to resolve it at the hardware level.

There has been several attempts to exploit properties of the stiffness matrix arising in the FEA to *further* accelerate it on the GPU. This essentially translates to two separate goals: Reducing the memory and/or reducing the total FLOP count of the application. For example, [18] implemented the SpMV stage of the FE solver on GPUs based on a FEA specific prefetching strategy to obtain a 3x speedup over the *traditional* SpMV implementation on the GPU. Another example is by [19], where the author exploited the typical sparsity pattern of global stiffness matrix depending on the number of degrees of freedom per node to reduce memory requirements and computation, thereby attaining a 18% to 51% performance improvements over the standard CuSPARSE library. The present work is in line of these work as we try to exploit the symmetry property of the local and global stiffness matrix to achieve those two goals. The local matrix generation stage, assembly stage and the SpMV stage are implemented on the GPU, each using two versions: One that exploits symmetry and one that does not. In the next two sections the methodology of implementations for the elemental calculations and matrix assembly are presented. This is followed by the results and discussion, after which the concluding remarks are presented.

**Symmetry in FEA**

*Symmetry in Elemental Computation*

In order to take advantage of the symmetry, a specialized kernel is written which is responsible for computing only the lower triangular part of the elemental stiffness matrix. In this kernel, each thread block is assigned to one element of the mesh and each thread of that thread block is assigned to one entry of the elemental stiffness matrix. A standard version of the kernel is also implemented where the entire elemental matrix is computed for comparison.

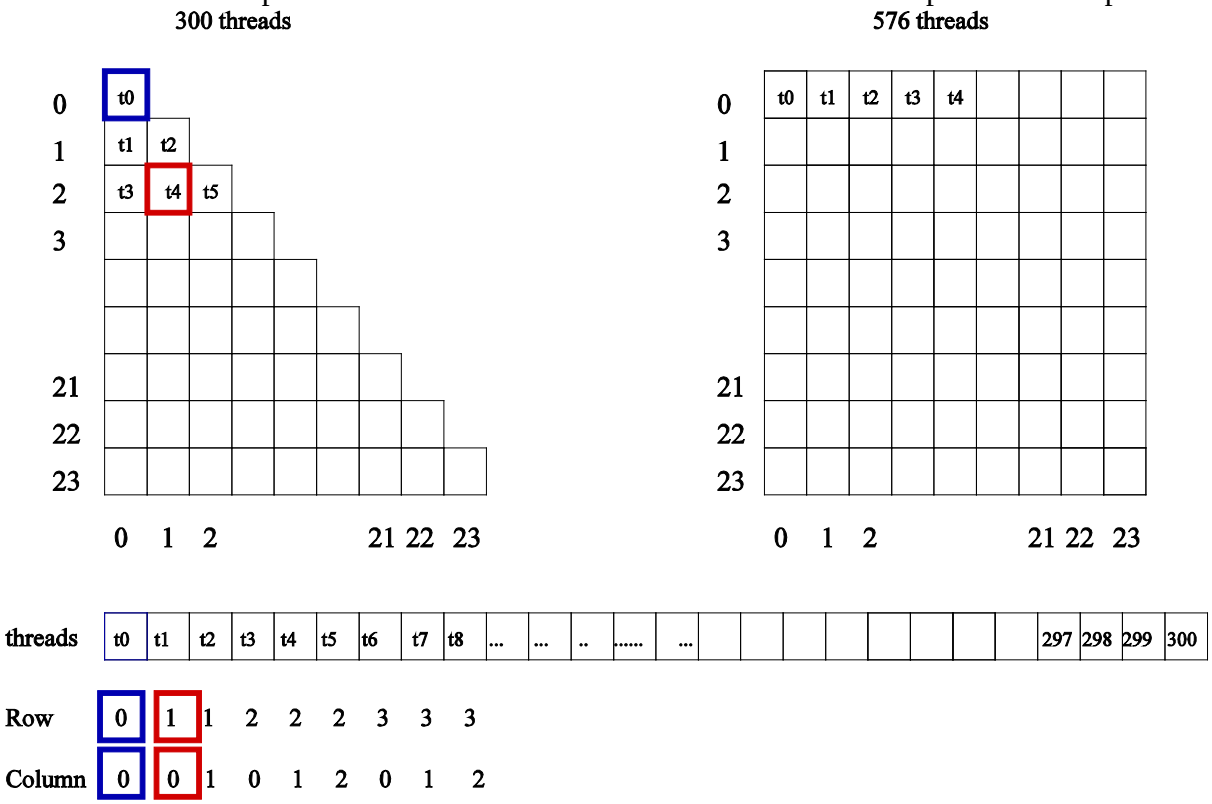



**Figure 1. Distribution of threads in symmetric and non-symmetric kernel**

As shown in figure 1, for a one entry per node GPU kernel, each block needs to have exactly 300 threads (shown in the left) as compared to 576 in the standard version (shown in the right). Since each warp has 32 threads, this corresponds to 10 and 18 warps respectively for the two versions. It should be noted here that the last warp in the symmetric version only has 12 threads. Although it is usual practice to take the number of threads as an exact multiple of the warp size, taking 320 threads per block and having 20 threads of the last warp as idle does not improve the results. So the block size is kept at 300.

Each of the threads in the thread block needs to know what row and column number it needs to calculate the elemental stiffness entry for. For example, in figure 1, the thread t4 needs to know that it is responsible for computing $K^e[2][1]$. Similarly, t10 needs to know that it is responsible for computing $K^e[4][0]$. It is important to note here that for a CPU-based sequential implementation of the same is not a major issue. A simple conditional statement with counters can take care of the issue. But, there is a different scenario for parallel versions on the GPU. This is because inside the kernel all the 300 threads need to know their target indices simultaneously. An approach could be where the same while loop is run for each thread, but with (i < threadIdx.x) as the termination condition. This will have the while loop

running for 0, 1, 2, 3... iterations for t0, t1, t2, t3 and so on respectively. The first problem with this is that this will create an uneven load on threads. Second and most importantly, this will create massive branch divergence within each warp due to the conditional statements. To counter this we have implemented a numbering scheme that computes the values of the target index from the thread index alone. The row and column index for each thread is computed using equations (1) and (2) respectively.

$$m = (\sqrt{1 + 8 \times threadIdx.x} - 1)/2 \tag{1}$$

$$n = threadIdx.x + m - (m + 1)(m + 2)/2 + 1 \tag{2}$$

It is important to note that all the divisions and the square root is performed as integer calculations. This means that if a floating point value is obtained at some point, the fractional part is dropped. The values for m and n with corresponding threadIdx.x up to 39 is shown in table 1.

**Table 1: Value of m and n with threadIdx.x**

| threadIdx.x | m | n | threadIdx.x | m | n | threadIdx.x | m | n | threadIdx.x | m | n |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 10 | 4 | 0 | 20 | 5 | 5 | 30 | 7 | 2 |
| 1 | 1 | 0 | 11 | 4 | 1 | 21 | 6 | 0 | 31 | 7 | 3 |
| 2 | 1 | 1 | 12 | 4 | 2 | 22 | 6 | 1 | 32 | 7 | 4 |
| 3 | 2 | 0 | 13 | 4 | 3 | 23 | 6 | 2 | 33 | 7 | 5 |
| 4 | 2 | 1 | 14 | 4 | 4 | 24 | 6 | 3 | 34 | 7 | 6 |
| 5 | 2 | 2 | 15 | 5 | 0 | 25 | 6 | 4 | 35 | 7 | 7 |
| 6 | 3 | 0 | 16 | 5 | 1 | 26 | 6 | 5 | 36 | 8 | 0 |
| 7 | 3 | 1 | 17 | 5 | 2 | 27 | 6 | 6 | 37 | 8 | 1 |
| 8 | 3 | 2 | 18 | 5 | 3 | 28 | 7 | 0 | 38 | 8 | 2 |
| 9 | 3 | 3 | 19 | 5 | 4 | 29 | 7 | 1 | 39 | 8 | 3 |

Figure 2 demonstrates the flow of work in the standard and the symmetric kernel. On the left side of the figure, the changes from the standard kernel are highlighted in red. As discussed the first change is in the kernel launch parameter. Where the standard version require 576 threads per thread block for eight-noded hexahedral mesh, the symmetric version require only 300. In the next change the values of m and n are computed from equations (1) and (2). These values are stored in per thread registers for faster access. Following this a *__syncthreads()* is called for barrier synchronization. Following this in both approaches 24 threads collaborate to compute the Jacobian matrices. After another synchronization is performed, 8 threads collaborate to compute the Jacobian inverse and determinants for both the approaches. After this another barrier synchronization is performed. Finally in case of the symmetric kernel, the value of the entry is calculated using the values of m and n and the values are stored in the elemental matrix in parallel.

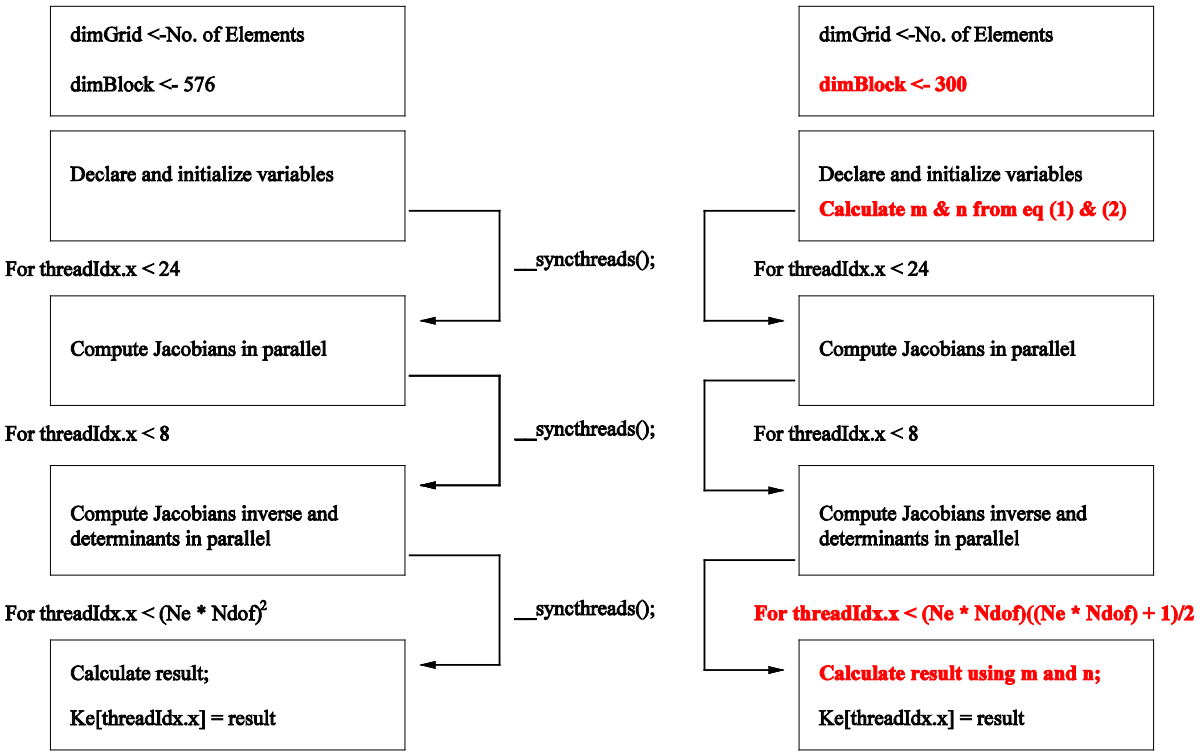



**Figure 2. Flow of work in standard (left) and symmetric (right) kernel**

*Symmetry in the Assembly Phase*

Unlike in the CPU, assembly of the local matrices on the GPU is considered to be a tedious and complex task on the GPU. This is why often *Assembly-free* methods are preferred on the GPU. Markall et al. [12] has commented in his work that assembly-free methods are more suitable for GPUs, whereas, assembly is more suitable for the CPUs.

We have implemented assembly of the local stiffness matrices on the GPU using both standard and symmetric versions. The key idea behind exploiting symmetry in the assembly stage is that to obtain the lower (or upper) triangular part of the global stiffness matrix, one only needs to assemble the lower (or upper) triangular part of all the local stiffness matrix. The idea is shown in figure 3, where the example of a simply supported beam is discretized using 4 elements in 2D. The four elemental stiffness matrices are shown with their lower triangular part highlighted. It is shown that during assembly, if the lower triangular part of the local matrix is assembled, the resulting global stiffness matrix is also lower triangular in nature, thus obviating the need to compute the upper half of the matrix.

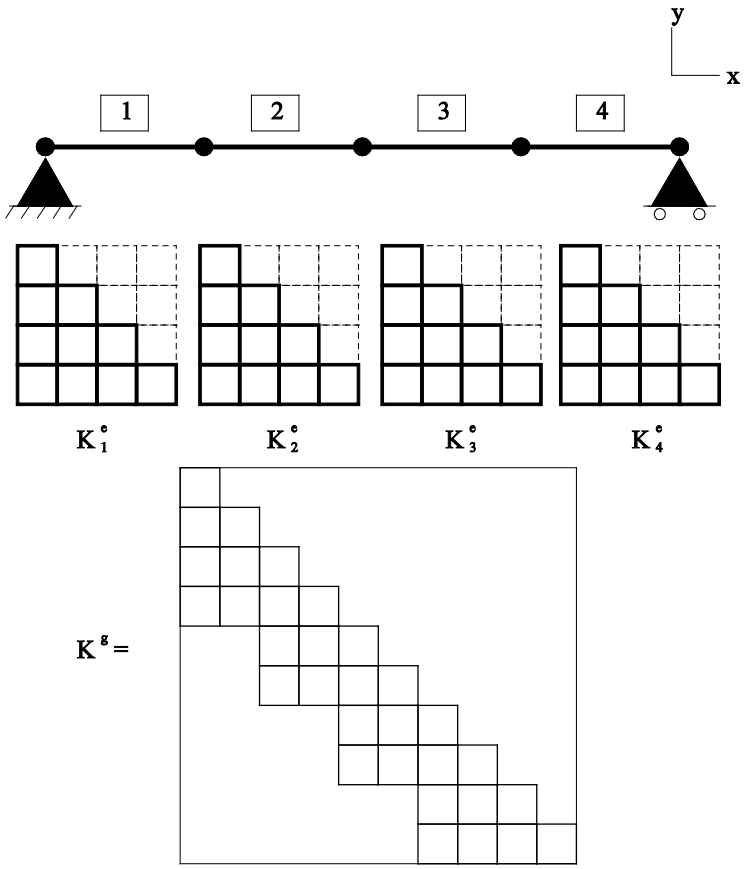



**Figure 3. Assembly in the symmetric kernel for only the lower triangular part of the elemental stiffness matrix**

GPU implementations of global stiffness assembly has the inevitable problem of race condition due to the parallel nature of the algorithm. We have implemented two strategies to counter the data race. The first one using *atomic operations*, has the disadvantage that it cannot use double precision. This creates a major issue in the linear solver stage because the method of conjugate gradients is considered to be notoriously sensitive to round-off errors. However, it should be mentioned here that on the latest GPUs from NVIDIA such as the Pascal P100 with compute capability 6.X, a version of atomic operations with double precision is introduced. In the second approach, a coloring scheme is applied. The key idea is to partition the elements into several different colors such that no two elements of the same color share a common node. Then the algorithm is run for the different colors in a sequential manner. The assembly implementation with coloring is run entirely in double precision.

**Results and Discussion**

For performance analysis of the proposed algorithms based on symmetry exploitation, a workstation with Intel Xeon ES1650 (6 core, 3.2 GHz) processor, 12 GB RAM, and NVIDIA K40c GPU is used. The GPU has 12 GB of global memory with 15 multiprocessors and 192 cores per multiprocessor. A cantilever beam with an end load meshed with eight-noded hexahedral elements (HEX8) are considered for the analysis.

The symmetry exploiting version of the code is tested and compared with the standard implementation that computes the entire matrix. Figure 4(a) shows the variation of the wall

clock time versus the number of nodes for both symmetric and standard version. It can be seen that both application show the same trend of linearly increasing execution time. It can also be observed the symmetric version takes considerably lower time than the standard version especially at higher node numbers. Figure 4(b) shows the variation of the GFLOP/s count with the number of nodes for both version. Both the curves exhibit the trend of increasing GFLOP/s with increasing node numbers at lower mesh size and a more or less stable state for higher mesh sizes. However, it can be seen that for a mesh size of more than 1,000,000, there is an approximately 25% increase in the GFLOP/s count in the symmetric version over the standard implementation. Figure 5 shows the variation of speedup of the symmetric version over the standard implementation with increasing number of nodes. It can be seen that the speedup increases rapidly up to a mesh size of 1,500,000 nodes. After this point the speedup increases very slowly with increasing node numbers.

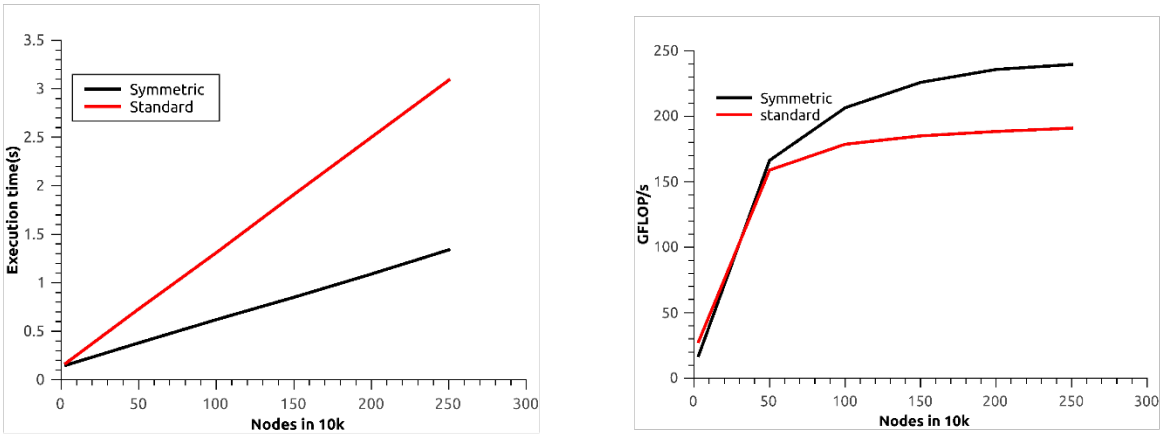



**Figure 4. Wall clock time (a) and GFLOP/s count (b) for local matrix generation vs number of nodes in 10,000 in symmetric and standard versions.**

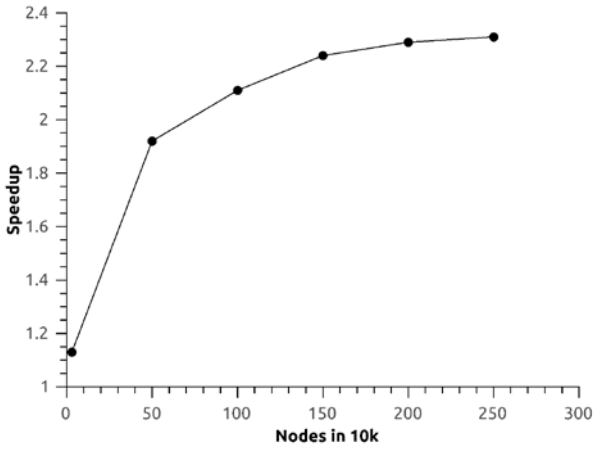



**Figure 5. Comparison of speedup of symmetric version compared to atomics version with number of nodes**

Two different versions, one symmetric and one standard, has been implemented for both coloring and atomics approach of assembly. It should be mentioned that due to lack of double precision support for atomics in the testing hardware only single precision has been implemented for atomics implementations. Figure 6(a) and figure 6(b) show the comparison of wall clock time with the number of nodes for Atomics and coloring respectively. Both symmetric and standard implementation is plotted in both of the figures. Although both figures 6(a) and figure 6(b) show similar time, it should be noted that figure (b) shows time using single precision arithmetic whereas, figure (a) uses double precision. Again a similar increasing trend is seen in the execution time for both the plots. The difference between the symmetric and standard implementation time increases as the mesh size is increased. The comparison of speedup in the entire proposed assembly operation using atomics and coloring over the standard implementation using coloring is plotted for different mesh sizes in figure 7. It can be observed that the speedup for the coloring approach is slightly higher than in case of the atomics approach. Also in both cases the speedup value increases with increasing mesh size and becomes stable at approximately 2x for coloring and 1.7x for atomics.

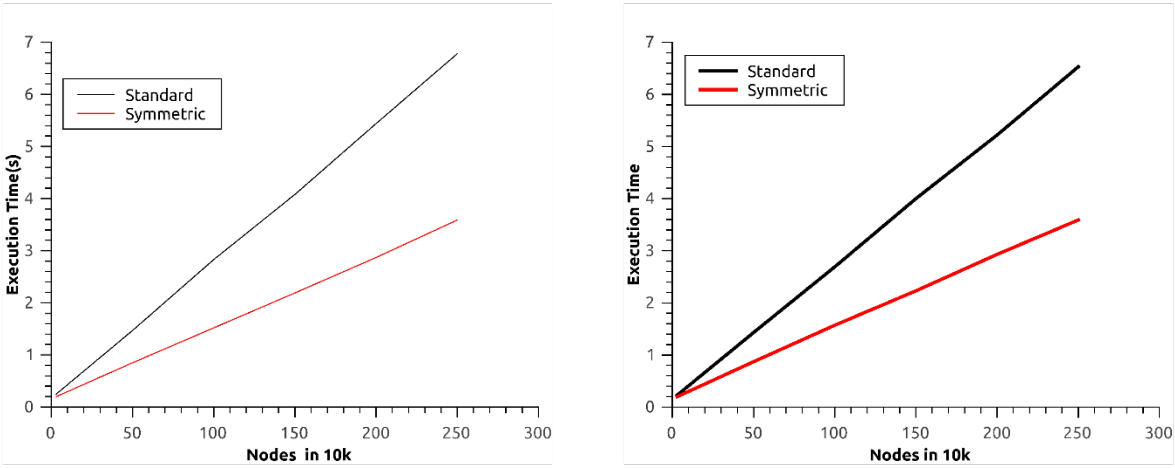



**Figure 6. Wall clock time comparison for Coloring (A) and atomics (B) version with nodes in 10,000**

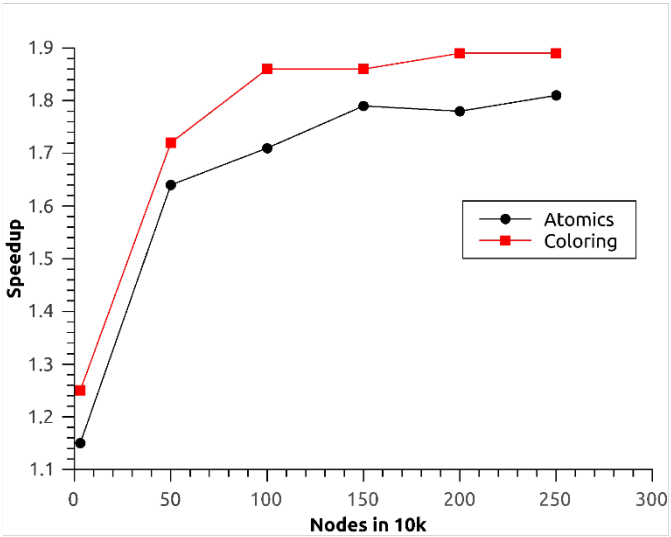



**Figure 7. Comparison of speedup of symmetric version compared to atomics version with number of nodes**

## Concluding Remarks

An implementation for accelerating FEA on the GPU based on the exploitation of the symmetry in the local and global matrix is presented. The implementation covers the elemental computation and assembly stage of a typical FEA. By exploiting symmetry in the local matrix generation stage, the execution time can be reduced by an amount of more than two. Furthermore there is also the benefit of lower storage space required, as only the symmetric part of the matrix is recorded. The symmetric version in the local matrix generation stage performs significantly better than the standard implementation. However, after the mesh size of approx. 1,500,000 nodes, the speedup is seen to be varying by little. The GFLOP/s count is 25% higher in the symmetric version than in the standard version of local matrix generation for higher node numbers. For handling race condition, the coloring method is the only viable option since atomics is incompatible with double precision, which is very important for the CG method. In the assembly stage as well approximately two speedup is obtained for the symmetric implementation over the standard one using both atomics and coloring. However, the speedup is slightly higher in case of coloring.

**References**


[1] Zienkiewicz, O. C., Taylor, R. L., Nithiarasu, P., & Zhu, J. Z. (1977). *The finite element method* (Vol. 3). London: McGraw-hill.

[2] Cecka, C., Lew, A. J., & Darve, E. (2011). Assembly of finite element methods on graphics processors. *International journal for numerical methods in engineering, 85*(5), 640-669.

[3] Schmidt, S., & Schulz, V. (2011). A 2589 line topology optimization code written for the graphics card. *Computing and Visualization in Science, 14*(6), 249-256.

[4] Komatitsch, D., Michéa, D., & Erlebacher, G. (2009). Porting a high-order finite-element earthquake modeling application to NVIDIA graphics cards using CUDA. *Journal of Parallel and Distributed Computing, 69*(5), 451-460.

[5] Dziekonski, A., Sypek, P., Lamecki, A., & Mrozowski, M. (2012). Finite element matrix generation on a GPU. *Progress In Electromagnetics Research, 128*, 249-265.

[6] Macioł, P., Płaszewski, P., & Banaś, K. (2010). 3D finite element numerical integration on GPUs. *Procedia Computer Science, 1*(1), 1093-1100.

[7] Dziekonski, A., Sypek, P., Lamecki, A., & Mrozowski, M. (2012). Accuracy, memory, and speed strategies in GPU-based finite-element matrix-generation. *IEEE Antennas and Wireless Propagation Letters, 11*, 1346-1349.

[8] Banaś, K., Płaszewski, P., & Macioł, P. (2014). Numerical integration on GPUs for higher order finite elements. *Computers & Mathematics with Applications*, 67(6), 1319-1344.

[9] Banaś, K., Krużel, F., & Bielański, J. (2016). Finite element numerical integration for first order approximations on multi-and many-core architectures. *Computer Methods in Applied Mechanics and Engineering, 305*, 827-848.

[10] Reguly, I. Z., & Giles, M. B. (2015). Finite element algorithms and data structures on graphical processing units. *International Journal of Parallel Programming, 43*(2), 203-239.

[11] Filipovic, J., Peterlik, I., & Fousek, J. (2009, July). GPU Acceleration of equations assembly in finite elements method-preliminary results. *In SAAHPC: Symposium on Application Accelerators in HPC*.

[12] Markall, G. R., Slemmer, A., Ham, D. A., Kelly, P. H. J., Cantwell, C. D., & Sherwin, S. J. (2013). Finite element assembly strategies on multi-core and many-core architectures. *International Journal for Numerical Methods in Fluids, 71*(1), 80-97.

[13] Dinh, Q., & Marechal, Y. (2015). Toward real-time finite-element simulation on GPU. *IEEE Transactions on Magnetics, 52*(3), 1-4.

[14] Sanfui, S., & Sharma, D. (2017, February). A two-kernel based strategy for performing assembly in FEA on the graphics processing unit. In *2017 International Conference on Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS)* (pp. 1-9). IEEE.

[15] Zayer, R., Steinberger, M., & Seidel, H. P. (2017, September). Sparse matrix assembly on the GPU through multiplication patterns. In *2017 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1-8). IEEE.

[16] Kiran, U., Sharma, D., & Gautam, S. S. (2018). GPU-Warp based Finite Element Matrices Generation and Assembly using Coloring Method. *Journal of Computational Design and Engineering.*

[17] Gribanov, I., Taylor, R., & Sarracino, R. (2018). Parallel implementation of implicit finite element model with cohesive zones and collision response using CUDA. *International Journal for Numerical Methods in Engineering*, *115*(7), 771-790.

[18] Dehnavi, M. M., Fernández, D. M., & Giannacopoulos, D. (2010). Finite-element sparse matrix vector multiplication on graphic processing units. *IEEE Transactions on Magnetics*, *46*(8), 2982-2985.

[19] Altinkaynak, A. (2017). An efficient sparse matrix-vector multiplication on CUDA-enabled graphic processing units for finite element method simulations. *International Journal for Numerical Methods in Engineering*, *110*(1), 57-78.

# Analysis of Cracked Body Strengthened by Adhesively Bonded Patches by BEM-FEM Coupling

***Binh Viet Pham[1], Thai Binh Nguyen[2], and †Jaroon Rungamornrat[3]**

[1]Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand.
[2]Faculty of Civil Engineering, Ho Chi Minh City University of Technology, Vietnam National University Ho Chi Minh City, 740128, Vietnam
[3]Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand.

*Presenting author: phamvietbinh872426@gmail.com
†Corresponding author: jaroon.r@chula.ac.th

## Abstract

This paper presents an efficient numerical technique capable of handling the stress analysis of three-dimensional cracked bodies strengthened by adhesively bonded patches. The proposed technique is implemented within the framework of the coupling of the weakly singular boundary integral equation method and the standard finite element procedure. The former is applied to efficiently treat the elastic body containing cracks whereas the latter is adopted to handle both the adhesive layers and patches. The approximation of the near-front relative crack-face displacement is enhanced by using local interpolation functions that can capture the right asymptotic behaviour. This also offers the direct calculation of the stress intensity factors along the crack front. A selected set of results is reported to demonstrate the capability of the proposed technique and the influence of various parameters on the performance of the strengthening.

**Keywords:** Boundary integral equation method, Bonded patches, Cracked body, Finite element method, Stress intensity factors

## Introduction

Sustainability and integrity of engineering structures are ones among various crucial issues that must be properly integrated in the design procedure and the subsequent maintenance stage. It has been known that the presence of cracks/flaws/damages is one of the major causes of subsequent failures of components, parts of or the entire structures leading finally to the loss of their functions. For such reason, the strengthening of damaged/cracked structures at the earliest time as soon as they are detected during inspections is considered essential. Furthermore, the control of cracked structures after being strengthened is also one of vital tasks to evaluate the selected strengthening method so that those strengthened structures can maintain their bearing capacity.

Replacement of cracked or damaged parts can be time-consuming and expensive, and, in addition, requires a high level of expertise. It is usually more cost effective to strengthen those damaged components by using patching techniques. One of those methods, with adhesively bonded repairs, has been widely employed in practices. Many theoretical and empirical investigations have demonstrated the advantages of this particular method relative to other existing techniques for strengthening cracked structures due mainly to its cost effectiveness (e.g., [1-6]). In addition to their high stiffness and strength, the patches are structurally efficient and induce much less damages to the strengthened structures.

In the modeling point of view via a theory of linear elasticity and linear elastic fracture mechanics, the stress intensity factors along the crack front can be significantly reduced after

the cracked body is strengthened by attaching a patch over the cracked region (e.g., [6-19]). This is due to the fact that some of externally applied loads exerted to the cracked bodies are either shared by or transferred directly to the patches via the adhesive shear layers rendering the enhancement of stiffness near the patching region and the reduction of relative crack-face displacements. As a direct consequence, the growth of the cracks can be delayed or even ceased if the strengthening is properly designed. Evidences from past studies have indicated that various parameters including the dimensions and material properties of the strengthened structure, the patches, and the adhesive layers significantly affects the efficiency and effectiveness of the strengthening (e.g., [6-19]). The full investigation to understand the role of those parameters is considered essential and can assist designers/engineers in the optimization of their designs. While experimental studies offer an excellent means to establish a set of results reflecting real responses, the methods themselves consume a significant amount of resources and are quite limited to test settings. In particular, to assess efficiency and influence of the patches in the experiments, a large set of testing specimens including un-patched and patched cracked structures for various specimen configurations must be considered. An alternative approach is to adopt computer-based simulations via reliable mathematical models to perform such extensive parametric study. It is remarked, however, that to accomplish such an important task within a broad and general framework (e.g., three-dimensional settings, large-scale and complex cases), powerful and computationally efficient tools are prerequisite.

On the basis of an extensive literature survey, most of existing studies toward the development of computational techniques to assist the analysis and simulations of cracked components repaired by adhesively bonded patches were focused extensively on the two-dimensional framework and quite specific to certain settings such as the repair configurations and types of repaired structures. The enhancement of existing modeling tools to be capable of handling more complex and general scenarios such as fully three-dimensional and large scale problems is challenging and requires further extensive investigations.

## Problem Formulation

Consider a three-dimensional, finite body containing both embedded and surface-breaking cracks and strengthened or reinforced by adhesively bonded patches as shown schematically in Figure 1. The cracked body is made of a homogeneous, generally anisotropic, linearly elastic material. The ordinary boundary of the cracked body, denoted by $S_0$, consists of a surface $S_u$ on which the displacement $\boldsymbol{u}^{bu}$ is prescribed, a surface $S_t$ on which the traction $\boldsymbol{t}^{bt}$ is known a priori, and a flat or planar surface $S_a$ on which the patch is attached. The surface of displacement discontinuity used to describe the crack in the undeformed state (i.e., stress-free state) is represented by a pair of geometrically identical surfaces, denoted by $S_c^+$ and $S_c^-$, and, in the present study, the attention is restricted only to the case that the crack surface is subjected to the point-wise self-equilibrated tractions; i.e., the prescribed tractions $\boldsymbol{t}^{b+}, \boldsymbol{t}^{b-}$ acting respectively to the surfaces $S_c^+, S_c^-$ satisfy the condition $\boldsymbol{t}^{b+} + \boldsymbol{t}^{b-} = \boldsymbol{0}$. Each patch is made of a homogeneous, linear elastic material and fully adhered to the cracked body on the surface $S_a$ by means of an adhesive bonding material. The prescribed traction exerted to the surface of the patch, opposite to the attached surface, is denoted by $\boldsymbol{t}^{p0}$. In the present study, the thickness of the patches and adhesive layers is assumed relatively small in comparison with the characteristic dimension of the repaired body; as a result, the bending stiffness of the patches can be considered negligible whereas only the shear resistance is treated for the adhesive layers.

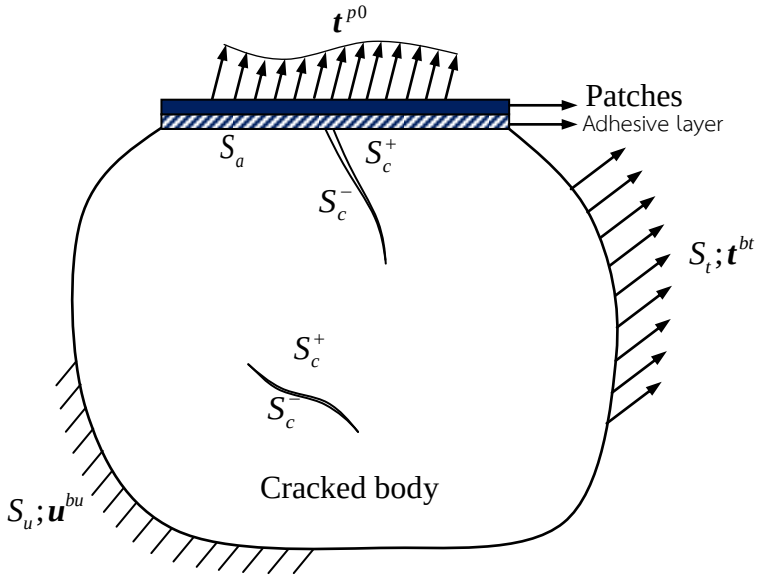



**Figure 1. Schematic of a body containing embedded and surface-breaking cracks and strengthened by adhesively bonded patches**

Since the thickness of each patch is sufficiently small and the bending effect can be ignored, its responses can then be properly modeled by a two-dimensional, plane-stress, linear elasticity theory. The final governing equation formulated in a local, two-dimensional, Cartesian coordinate system containing the patch by taking the in-plane displacement vector $\bar{\boldsymbol{u}}^{*p}$ as the primary unknown is given by

$$h_p \boldsymbol{L}^T \boldsymbol{C}^{*p} \boldsymbol{L} \bar{\boldsymbol{u}}^{*p} + \bar{\boldsymbol{s}}^{*pa} + \bar{\boldsymbol{s}}^{*p0} = \boldsymbol{0} \tag{1}$$

where $\boldsymbol{L}$ is a conventional, two-dimensional differential operator transforming $\bar{\boldsymbol{u}}^{*p}$ into a vector containing independent in-plane strain components; the superscript "$T$" denotes the matrix transpose; $h_p$ is the thickness of the patch; $\bar{\boldsymbol{s}}^{*p0}$ is a vector containing local components of the prescribed shear traction $\boldsymbol{s}^{p0}$ (i.e., the tangential component vector of the prescribed traction $\boldsymbol{t}^{p0}$ on the plane of the patch), $\bar{\boldsymbol{s}}^{*pa}$ is a vector containing local components of the unknown shear traction exerted by the adhesive layer, and $\boldsymbol{C}^{*p}$ is the elastic in-plane modulus matrix for the plane-stress case. An alternative weak-form of (1) can be readily established via a standard weighted residual technique and used as the basis in the discretization procedure.

From the assumption that the adhesive layer can transfer only shear across its thickness, the state of strain for the entire layer can be completely described by the out-of-plane shear strain. Since the thickness of the adhesive layer is infinitesimal in comparison with its planar dimensions, it is legitimate to assume that the out-of-plane shear strain components are uniform across the thickness or, equivalently, the in-plane displacement varies linearly across the thickness. The out-of-plane shear stress induced within the layer $\bar{\boldsymbol{\sigma}}^{*a}$ is then governed by

$$\bar{\boldsymbol{\sigma}}^{*a} = \frac{G_a}{h_a}(\bar{\boldsymbol{u}}^{*ap} - \bar{\boldsymbol{u}}^{*ab}) \tag{2}$$

where $h_a$ is the thickness of the adhesive layer, $G_a$ denotes the elastic shear modulus of the adhesive material, and $\bar{\boldsymbol{u}}^{*ap}$ and $\bar{\boldsymbol{u}}^{*ab}$ are values of the in-plane displacement at its interfaces connecting to the patch and the cracked body, respectively. Similarly, the weak-form statement of (2) can also be formulated by the weighted residual technique.

For the cracked body, the theory of linear elasticity with the absence of the body force is adopted and the key governing equations are formulated within the framework of boundary integral equations. In particular, the following pair of weakly singular, weak-form boundary integral equations for the displacements and tractions, proposed by Rungamornrat and Mear (2008a), is utilized to form a system of integral equations governing the unknown data on the boundary and crack surface:

$$\begin{aligned}\frac{1}{2}\int_{S_0}\tilde{\tau}_p(\boldsymbol{y})u_p^b(\boldsymbol{y})dS(\boldsymbol{y}) &= \int_{S_0}\tilde{\tau}_p(\boldsymbol{y})\int_{S_0}U_j^p(\boldsymbol{\xi}-\boldsymbol{y})t_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y}) \\ &+\int_{S_0}\tilde{\tau}_p(\boldsymbol{y})\int_{S}G_{mj}^p(\boldsymbol{\xi}-\boldsymbol{y})D_m v_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y}) \\ &-\int_{S_0}\tilde{\tau}_p(\boldsymbol{y})\int_{S}n_i(\boldsymbol{\xi})H_{ij}^p(\boldsymbol{\xi}-\boldsymbol{y})v_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y})\end{aligned} \tag{3}$$

$$\begin{aligned}-\frac{1}{2}\int_{S}\tilde{u}_k(\boldsymbol{y})\tau_k^b(\boldsymbol{y})dS(\boldsymbol{y}) &= \int_{S}D_t\tilde{u}_k(\boldsymbol{y})\int_{S}C_{mj}^{tk}(\boldsymbol{\xi}-\boldsymbol{y})D_m v_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y}) \\ &+\int_{S}D_t\tilde{u}_k(\boldsymbol{y})\int_{S_0}G_{tk}^j(\boldsymbol{\xi}-\boldsymbol{y})t_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y}) \\ &+\int_{S}\tilde{u}_k(\boldsymbol{y})\int_{S_0}n_l(\boldsymbol{y})H_{lk}^j(\boldsymbol{\xi}-\boldsymbol{y})t_j^b(\boldsymbol{\xi})dS(\boldsymbol{\xi})dS(\boldsymbol{y})\end{aligned} \tag{4}$$

where $S \equiv S_0 \cup S_c^+$ denotes the total boundary of the cracked body; $\tilde{\tau}_p$ is any sufficiently smooth test function defined on the ordinary boundary $S_0$; $\tilde{u}_k$ is any sufficiently smooth test function defined on the total boundary $S$; $u_p^b$ and $t_j^b$ are components of the displacement and traction on the ordinary boundary $S_0$ of the cracked body; $n_i$ are components of the outward unit normal vector to the total boundary $S$; $D_m = n_i\varepsilon_{ism}\partial/\partial\xi_s$ or $D_m = n_i\varepsilon_{ism}\partial/\partial y_s$ denotes the surface differential operator; $U_j^p$, $C_{mj}^{tk}$, $G_{mj}^p$, and $H_{ij}^p$ are known fundamental solutions (see details of development and explicit expressions in Rungamornrat and Mear (2008a)); and $v_j^b$ and $\tau_k^b$ are data defined by

$$v_j^b(\boldsymbol{\xi}) = \begin{cases} u_j^b(\boldsymbol{\xi}), & \boldsymbol{\xi}\in S_0 \\ \Delta u_j^b(\boldsymbol{\xi}), & \boldsymbol{\xi}\in S_c^+ \end{cases} \tag{5}$$

$$\tau_k^b(\boldsymbol{\xi}) = \begin{cases} t_k^b(\boldsymbol{\xi}), & \boldsymbol{\xi}\in S_0 \\ \Delta t_k^b(\boldsymbol{\xi}), & \boldsymbol{\xi}\in S_c^+ \end{cases} \tag{6}$$

in which $\Delta u_j^b = u_j^+ - u_j^-$ denotes the relative crack-face displacement and $\Delta t_k^b(\boldsymbol{\xi}) = t_k^{0+} - t_k^{0-}$ denotes the jump in the crack-face traction. In particular, for the self-equilibrated crack-face tractions, it yields $\Delta t_k^b(\boldsymbol{\xi}) = 2t_k^{0+}$. To form a system of integral equations governing all

unknown data on the boundary and the crack surface, the displacement boundary integral equation (3) is applied to the surface $S_u$ with $\tilde{\tau}_p \equiv 0$ on $S_t \cup S_a$ whereas the traction boundary integral equation (4) is adopted for the remaining surface $S_t \cup S_a \cup S_c^+$ with $\tilde{u}_k \equiv 0$ on $S_u$.

A system of governing equations for the whole repaired cracked body shown in Figure 1 can now be obtained by combining the weak-form equations governing all patches, the weak-form equations governing all adhesive layers, and those governing the cracked body together with the continuity of the displacement and the traction along all material interfaces. The final system contains the following unknown functions: the shear stress within the adhesive layers $\bar{\boldsymbol{\sigma}}^{*a}$, the in-plane displacement of the patch $\bar{\boldsymbol{u}}^{*p}$, the displacement $\boldsymbol{u}^{ba}$ on the surface $S_a$, the displacement $\boldsymbol{u}^{bt}$ on the surface $S_t$, the traction $\boldsymbol{t}^{bu}$ on the surface $S_u$, and the relative crack-face displacement $\Delta\boldsymbol{u}^b$.

## Numerical Implementations

To discretize the governing weak-form equations for the patches and the adhesive layers, a standard finite element procedure for two-dimensional problems (e.g., [21-23]) is adopted. The unknown shear stress within the adhesive layer $\bar{\boldsymbol{\sigma}}^{*a}$, the unknown in-plane displacement of the patch $\bar{\boldsymbol{u}}^{*p}$, the unknown displacement on the surface of the cracked body $\boldsymbol{u}^{ba}$ and all involved test functions are approximated using standard basis functions constructed locally on a finite element mesh consisting of standard, isoparametric, $C^0$-elements.

To discretize the weakly-singular, weak-form integral equations governing the cracked body, Galerkin-based procedure similar to that proposed by Rungamornrat and Mear (2008b) is implemented. Due to the weakly singular feature of all involved integrals, both the trial and test functions can be approximated by a set of continuous basis functions constructed locally on a finite element mesh. In particular, standard isoparametric $C^0$ elements are employed everywhere in the solution discretization except in a local region of the crack surface adjacent to the crack front where special crack-tip elements, originally proposed by Li *et al*. (1998) and used later by Rungamornrat and Mear (2008b) to treat cracks in anisotropic media, are adopted. Element shape functions of such special crack-tip elements were properly enriched to contain the square-root-type behavior and accurately capture the near-front relative crack-face displacement (also see details in Yates et al. (2010) and Rungamornrat *et al*. (2019) for the structure of the near-front elastic field). Special quadrature rules proposed by Xiao (1998) are implemented to handle both weakly singular and nearly singular integrals and the efficient interpolation-based algorithm similar to that employed by Rungamornrat and Mear (2008b) is adopted to calculate all involved fundamental solutions for generally anisotropic materials.

The final system of linear algebraic equations resulting from the discretization of the governing equations of the patches, the adhesive layers, and the cracked body is solved by a selected efficient linear solver. The stress intensity factors along the crack front are then extracted directly from the solved relative crack-face displacement data together with the properties of the special crack-tip elements via the post-process formula proposed by Rungamornrat and Mear (2008b).

## Numerical Results

To verify the implemented technique and also provide a set of results from a preliminary parametric study on the strengthening of cracked bodies, the following representative problem is chosen in numerical simulations. Consider a cube of an isotropic linearly elastic material that occupies the region $[-w,w]\times[-w,w]\times[-w,w]$ in space and contains a penny-shaped

crack of radius $a$ as shown schematically in Figure 2. The crack lies on a plane $x_3 = 0$ with its center located at point $(0.4w, 0, 0)$. The crack front can be parametrized in terms of the angular position $\theta \in [0, 2\pi]$ by

$$x_1 = 0.4w - a\cos\theta, \quad x_2 = a\sin\theta, \quad x_3 = 0 \tag{7}$$

The cube is loaded by a uniform normal traction $t_3 = \sigma_0$ on the face $x_3 = w$ and the uniform normal traction $t_3 = -\sigma_0$ on the face $x_3 = -w$. To strengthen the cracked body, a patch of uniform thickness $h_p$ is bonded to its entire face $x_1 = w$ by the adhesive layer of uniform thickness $h_a$. In the numerical study, the aspect ratio $a/w = 0.5$ and Young's modulus and Poisson's ratio given in Table 1 are considered and three meshes shown in Figure 3 are adopted.

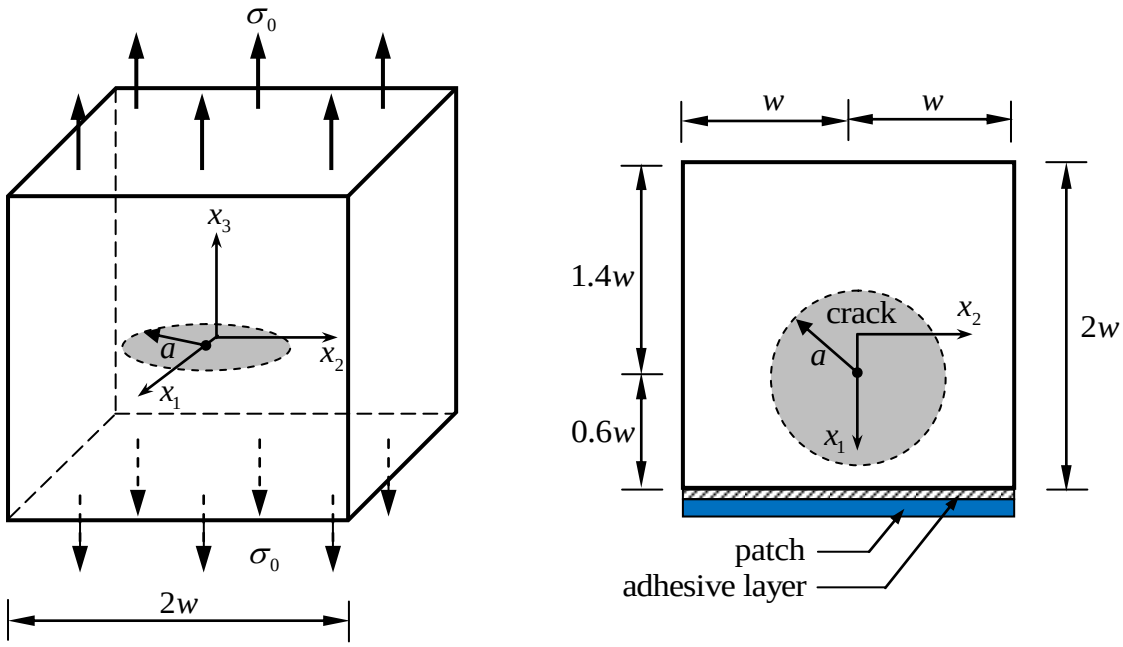



**Figure 2. Schematic of cube material containing near-surface penny-shaped crack and strengthening by adhesively bonded patch**

**Table 1. Young's modulus and Poisson's ratio for cracked body, patch, and adhesive layer used in parametric study**

| Materials | Young's modulus ($\times 10^6$ psi) | Poisson's ratio |
|---|---|---|
| Cracked body | 2.0 | 0.25 |
| Patch | 17.4 | 0.25 |
| Adhesive layer | 0.1 | 0.33 |

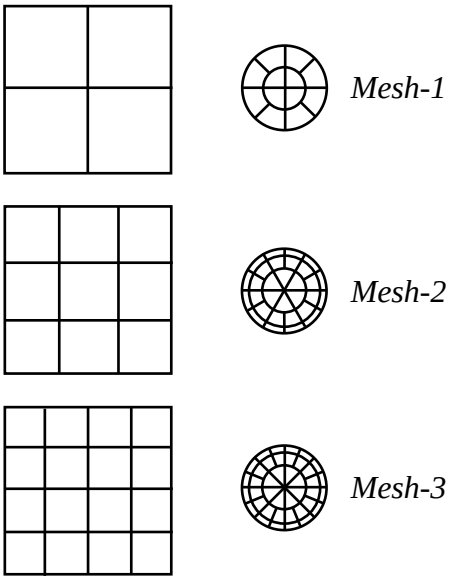



**Figure 3. Three meshes adopted in analysis (only mesh of each face of cube is shown and it is identical to those for patch and adhesive layer)**

Due to the symmetry and the loading condition considered, only the mode-I stress intensity factor ($K_I$) is non-zero. The normalized $K_I$ obtained from the three meshes are reported in Figure 4 along with those generated by ABAQUS for the cases with and without the strengthening. It can be concluded from this set of results that numerical solutions converge as the mesh is refined and the good agreement between the converged and reference solution (with the difference within a fraction of one percent) is observed. Note in particular that relatively coarse meshes such as the Mesh-1 and Mesh-2 can also yield quite accurate results; this is due mainly to the use of special crack-tip elements in the approximation of the near-front relative crack-face displacement. After the cracked body is strengthened by the adhesively bonded patch, the stress intensity factor is significantly reduced especially in the region near the bonded patch.

After fully tested, the proposed technique can be further applied to study the influence of various strengthening parameters (e.g., thickness of the adhesive layer and thickness of the patch) on the strengthening performance. For instance, to explore the influence of the patch thickness on the reduction of the stress intensity factor of the crack in the representative problem, simulations can be carried out for different values of $h_p$ while all other parameters remain fixed. A plot of the normalized mode-I stress intensity factor resulting from such simulations are reported in Figure 5, as examples, for $h_p/w = 0.00, 0.01, 0.02, 0.03, 0.04$. Besides the expected reduction of the stress intensity factor as the patch thickness increases (due to the increase in the stiffness after the strengthening), this piece of information is potentially useful in the selection of the patch thickness to confine the stress intensity factor below the tolerance or to prevent the subsequent crack growth. Similarly, the influence of the thickness of the adhesive layer on the response after the strengthening can also be investigated by carrying out simulations for various values of $h_a$ while maintaining all other parameters. Results shown in Figure 6 are for the representative cracked body with three different values of the thickness of the adhesive layer (i.e., $h_a / w = 0.001, 0.005, 0.01$). It is evident that as the thickness of the adhesive layer increases, the apparent stiffness of the cracked body after strengthening tends to decrease.

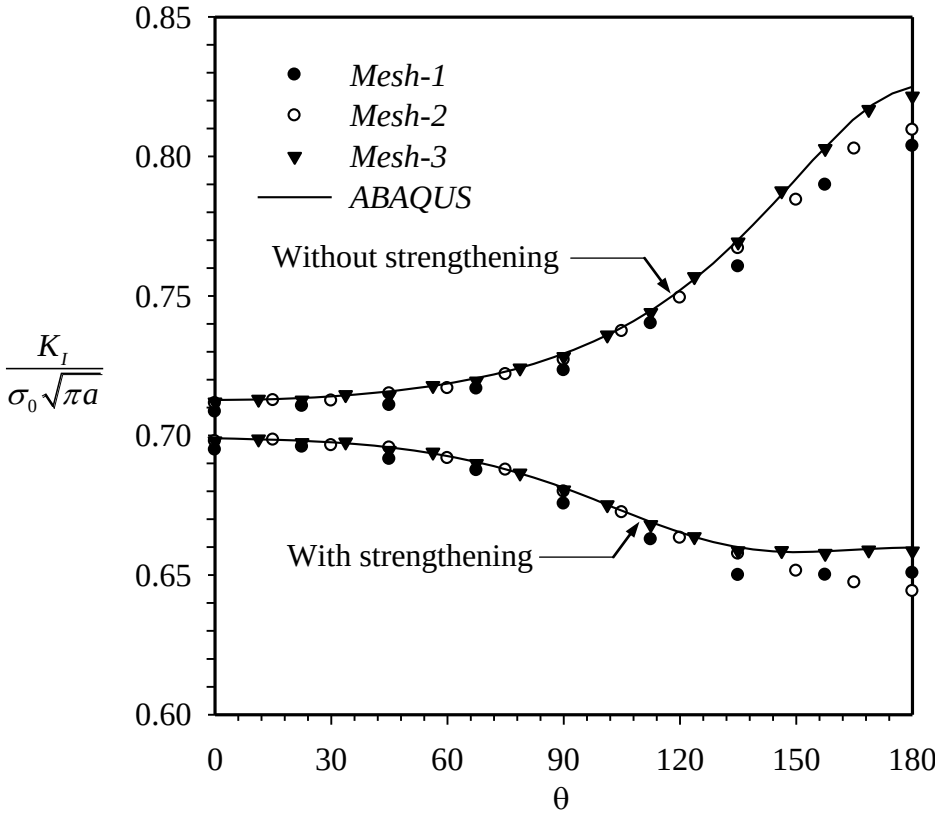



**Figure 4. Normalized mode-I stress intensity factors of near-surface penny-shaped crack in cube of material under uniform normal traction $\sigma_0$ on its upper and lower faces. Results for the case of strengthening are reported for $h_a / w = 0.001$ and $h_p / w = 0.01$.**

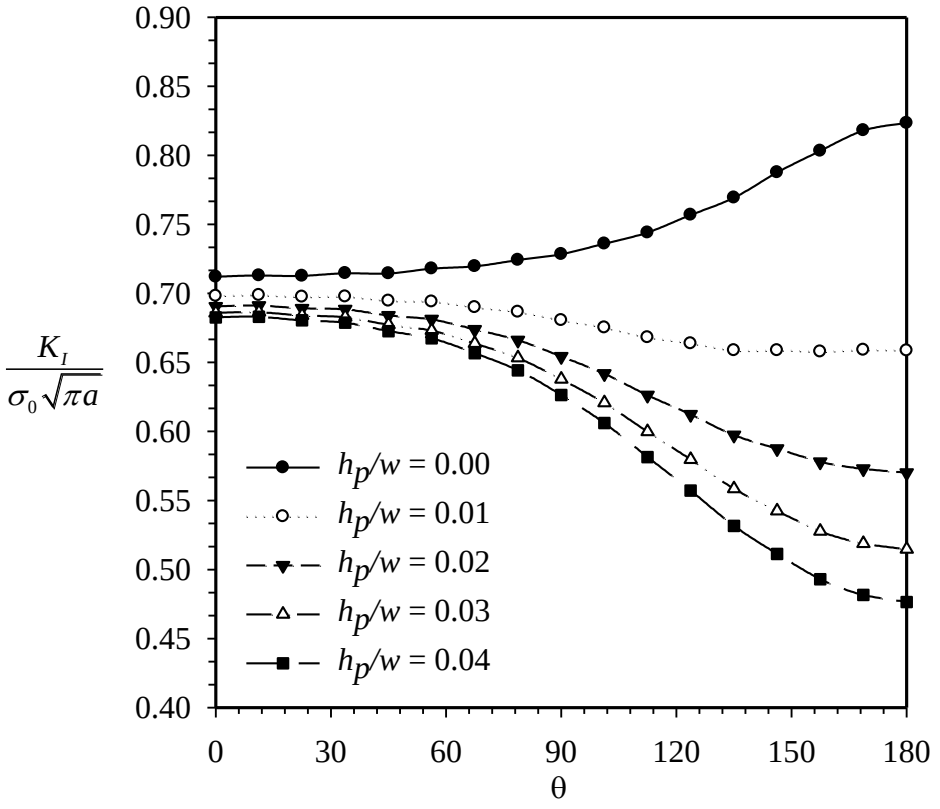



**Figure 5. Influence of thickness of patch on normalized mode-I stress intensity factors for near-surface penny-shaped crack in cube of material**

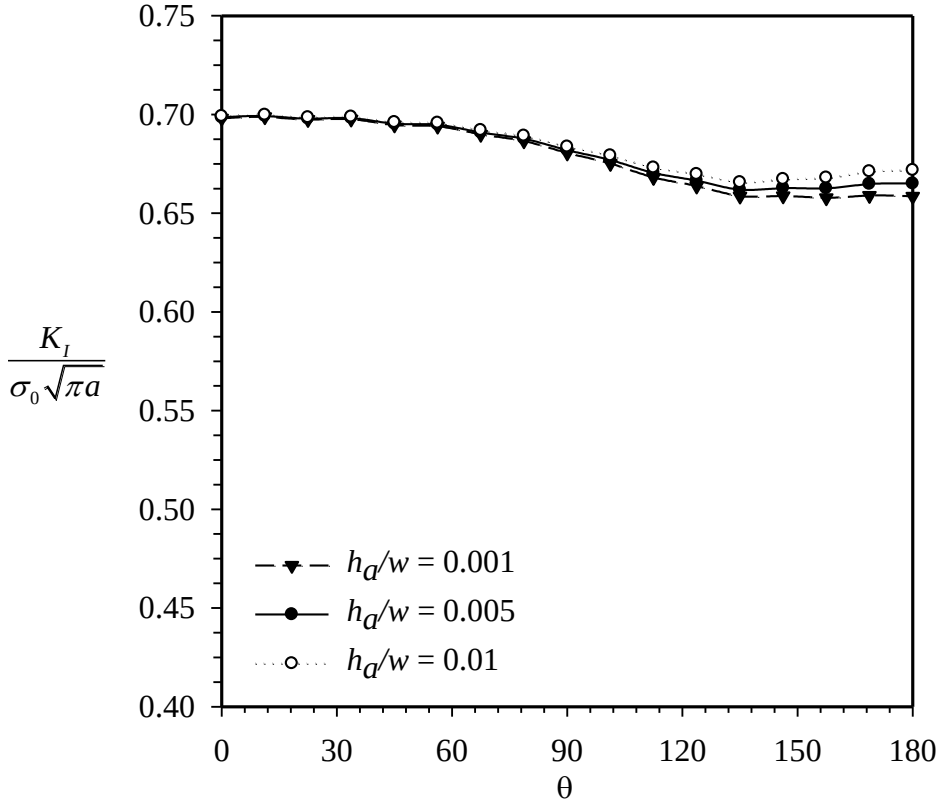



**Figure 6. Influence of thickness of adhesive layer on normalized mode-I stress intensity factors for near-surface penny-shaped crack in cube of material**

## Conclusion and Remarks

The efficient and accurate BIE-FE coupling technique has been successfully implemented for the analysis of three-dimensional cracked bodies strengthened by adhesively bonded patch. The boundary integral equation method has been adopted to efficiently treat the cracked body whereas the standard finite element method has been utilized to handle both the adhesive layer and the patch. The near-front approximation of the relative crack-face displacement has been enhanced by means of using special crack-tip elements and this allows relatively coarse meshes to be employed in the discretization while still yielding sufficiently accurate fracture data along the crack front. Results from a numerical study have indicated that numerical solutions obtained from the proposed technique possess the good convergence behaviour and are of excellent agreement with reliable benchmark solutions. In addition, the preliminary parametric study has shown that the stress intensity factor along the crack front is significantly reduced as the thickness of the patch increases while the reverse trend has been observed for the adhesive layer.

## Acknowledgements

The authors gratefully acknowledge the supports provided by the JICA Project for AUN/SEED-Net and The Thailand Research Fund (Grant No. RSA5980032).

### References

[1] Arendt C and Sun CT. Bending effects of unsymmetric adhesively bonded composite repairs on cracked aluminum panels, 1994.

[2] Sun CT, Klug J, Arendt C. Analysis of cracked aluminum plates repaired with bonded composite panels. AIAA Journal, 34(2), pp. 369-74, 1996.

[3] Baker AA, Jones R. Bonded repair of aircraft structures. Dordrecht: Martinus Nijhoff, 1988.
[4] Rose LRF. An application of the inclusion analogy for bonded reinforcements. International Journal of Solids and Structures, 17(8), pp. 827-838, 1981.
[5] Jones R. Bonded repair of damage. Journal of the Aeronautical Society of India, 1984.
[6] Young A, Rooke DP, Cartwright DJ. Analysis of patched and stiffened cracked panels using the boundary element method. International journal of solids and structures, 29(17), pp. 2201-2216, 1992.
[7] Liu HB, Zhao XL, Al-Mahaidi R. Boundary element analysis of CFRP reinforced steel plates. Composite Structures, 91(1), pp. 74-83, 2009.
[8] Pisa CD, Aliabadi MH. Boundary element analysis of stiffened panels with repair patches. Engineering Analysis with Boundary Elements, 56, pp. 162-175, 2015.
[9] Salgado NK, Aliabadi MH. The application of the dual boundary element method to the analysis of cracked stiffened panels. Engineering Fracture Mechanics, 54(1), pp. 91-105, 1996.
[10] Salgado NK, Aliabadi MH. The boundary element analysis of cracked stiffened sheets, reinforced by adhesively bonded patches. International journal for numerical methods in engineering, 42(2), pp. 195-217, 1998.
[11] Sekine H, Yan B, Yasuho T. Numerical simulation study of fatigue crack growth behavior of cracked aluminum panels repaired with a FRP composite patch using combined BEM/FEM. Engineering Fracture Mechanics, 72(16), pp. 2549-2563, 2005.
[12] Jiann-Quo T, Kam-Lun S. Analysis of cracked plates with a bonded patch. Engineering Fracture Mechanics, 40(6), pp. 1055-1065, 1991.
[13] Alaimo A, Milazzo A, Orlando C. Boundary elements analysis of adhesively bonded piezoelectric active repair. Engineering fracture mechanics, 76(4), pp. 500-511, 2009.
[14] Useche J, Sollero P, Albuquerque EL, Palermo L. Boundary element analysis of cracked thick plates repaired with adhesively bonded composite patches. SDHM: Structural Durability & Health Monitoring, 4(2), pp. 107-116, 2008.
[15] Wen PH, Aliabadi MH, Young A. Stiffened cracked plates analysis by dual boundary element method. International journal of fracture, 106(3), pp. 245-258, 2000.
[16] Wen PH, Aliabadi MH, Young A. Boundary element analysis of flat cracked panels with adhesively bonded patches. Engineering fracture mechanics, 69(18), pp. 2129-2146, 2002.
[17] Wen PH, Aliabadi MH, Young A. Boundary element analysis of curved cracked panels with adhesively bonded patches. International journal for numerical methods in engineering, 58(1), pp. 43-61, 2003.
[18] Widagdo D, Aliabadi MH. Boundary element analysis of cracked panels repaired by mechanically fastened composite patches. Engineering analysis with boundary elements, 25(4-5), pp. 339-345, 2001.
[19] Yu QQ, Zhao XL, Chen T, Gu XL, Xiao ZG. Crack propagation prediction of CFRP retrofitted steel plates with different degrees of damage using BEM. Thin-Walled Structures, 82, pp. 145-158, 2014
[20] Rungamornrat J, Mear ME. Weakly-singular, weak-form integral equations for cracks in three-dimensional anisotropic media. International Journal of Solids and Structures, 45(5), 1283-1301, 2008a.
[21] Hughes TJ. The finite element method: linear static and dynamic finite element analysis, Courier Corporation, 2012.
[22] Bathe KJ. Finite element procedures, Klaus-Jurgen Bathe, 2006.
[23] Zienkiewicz OC, Taylor RL. The finite element method: solid mechanics, Butterworth-heinemann, 2000.
[24] Rungamornrat J, Mear ME. A weakly-singular SGBEM for analysis of cracks in 3D anisotropic media. Computer Methods in Applied Mechanics and Engineering, 197(49-50), pp.4319-4332, 2008b.
[25] Li S, Mear ME, Xiao L. Symmetric weak-form integral equation method for three-dimensional fracture analysis. Computer Methods in Applied Mechanics and Engineering, 151(3-4), pp.435-459, 1998.
[26] Yates JR, Zanganeh M, Tai YH. Quantifying crack tip displacement fields with DIC. Engineering Fracture Mechanics, 77, pp. 2063-2076, 2010.
[27] Rungamornrat J, Sukulthanasorn N, Mear ME. Analysis for T-stress of cracks in 3D anisotropic elastic media by weakly singular integral equation method. Computer Methods in Applied Mechanics and Engineering, 347, pp. 1004-1029, 2019.
[28] Xiao L. Symmetric weak-form integral equation method for three-dimensional fracture analysis. Ph.D. Dissertation, University of Texas at Austin, USA, 1998.

# Wrinkle generation mechanism during draw bending forming

**†*Xia Zhu[1], Keiji Ogi[1] and Nagatoshi Okabe[1]**

[1]Department of Mechanical Engineering, Faculty of Engineering, Ehime University, JP

*Presenting author: zhu.xia.mx@ehime-u.ac.jp
†Corresponding author: zhu.xia.mx@ehime-u.ac.jp

## Abstract

The purpose of this research is to focus on wrinkles, which are the main processing limit factor of draw bending, and to elucidate the mechanism of the generation. The forming experiments provided data for verifying the validity of the analytical model and the limits of crease formation. In addition, the effect of the friction coefficient between the pipe and the bending die on the occurrence of wrinkles was confirmed by finite element analyses. Furthermore, by analyzing the deformation behavior inside the material during the draw bending process, the mechanism of the wrinkle was clarified. The following findings were obtained from this study. First, wrinkles occur not at the bend but at the straight pipe part on the raw pipe side. Next, the coefficient of friction between the pipe and the bending die promotes wrinkles and the growth of the generated wrinkles. Finally, the wrinkles are formed by the transition of the flexure due to the drawing-rolling phenomenon in the draw bending forming.

**Keywords:** Draw bending forming, Wrinkles, Finite-element method

## Introduction

Metal pipes, such as plumbing, have been used for fluid transportation for a long time and form a transportation path together with fittings, such as elbows, within the structure. However, in view of the demands for energy saving and cost reduction in recent years, reduction of joints and thinning of pipes are required. Therefore, a method has been adopted in which the pipe is processed by draw bending processing, and the number of joints used is suppressed by forming a bent portion at an arbitrary position [1-5]. Here, when the wall thickness of the pipe becomes thin, problems such as high flattening rate, wrinkles, and plastic buckling occur. A wrinkle is considered undesirable except for special processing, such as wrinkle bending, and the wrinkle is classified as defective at the stage of wrinkle generation. Compared to plastic buckling, wrinkles are more likely to occur, which is a major factor in the processing limit of rotary drawing and bending.

The purpose of this research is to determine the state inside the material using finite-element analysis and to improve the performance of draw bending forming by clarifying the mechanism of wrinkle generation. An analytical model of rotational drawing was made by using the general-purpose nonlinear finite-element analysis software MSC Marc 2016, and the analytical results were compared with experimental results to verify the validity of the model. Furthermore, the mechanism of wrinkle generation was investigated.

## Verification of analytical model validity

Fig. 1 shows a process schematic view of the draw bending process. Here, $R$ is the bending radius of a bending die, $\omega$ is the rotational speed of a bending die, and $\theta$ is the bending angle of a bending die. Prior to investigating the mechanism of wrinkle generation, the validity of the analytical model was verified. In this study, plastic deformation behaviors of the material and the friction coefficient between the pipe and the bending die were focused on as parameters

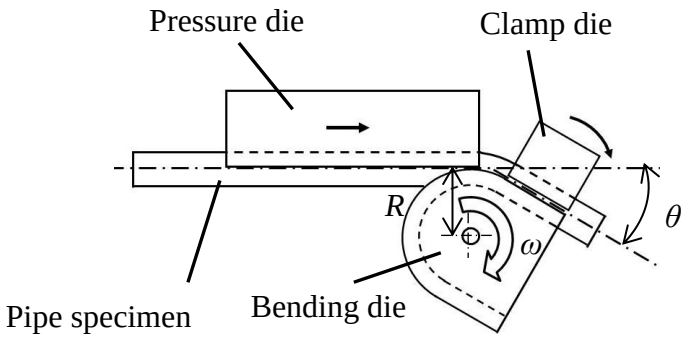



**Fig. 1 Schematic diagram of draw bending forming**

**Table1 Numbers of element and minimum element size**

| | Numbers of elements | Minimum element size |
|---|---|---|
| Axial direction | 65<br>(Bend part divided in half again) | 1 mm |
| Circumferential direction | 18 | 0.698 mm (10°) |
| Radial direction (thickness) | 6 | 0.1667 mm |

affecting the wrinkles generated during the draw bending forming.
Under the processing conditions in which no wrinkling occurs, the influence of the friction coefficient is smaller to the plastic deformation behavior. Therefore, the frictional force generated between the pipe and bending die is small under the processing condition in which wrinkles do not occur, even in actual processing. Therefore, to verify the validity of the material properties used for the analysis, the flattening ratio $f = (D_{max} - D_{min}) / D_0$ of the bent portion of the pipe specimen after processing was investigated under the processing condition in which wrinkles do not occur. Here, $D_{max}$, $D_{min}$, $D_0$ are the major axis, minor axis, and outer diameter of the original pipe in the cross section of the bent portion.

*Experimental conditions*

A SUS 304 (Japanese Industrial Standards, JIS G 3448：2012) seamless steel pipe with an outside diameter $D_0$ = 10 mm, which is widely used for pipes, was used, with a wall thickness $t_0$ = 1.0 mm. A bending radius $R$ of the bending die was fixed at 20 mm, the rotational angular velocity $\omega$ was fixed at 120 rpm, the bending angle $\theta$ was changed from 15 ° to 120 ° at intervals of 15 °, and processing experiments were conducted.

*Analytical conditions*

As shown in Fig. 1, a bending die, a clamping die, and a pressure die were defined as rigid bodies, and the pipe was defined as an elastoplastic body with twenty-node three-dimensional isoparametric elements in the analytical model. In consideration of symmetry, a 1/2 model was used. $D_0$ and $t_0$ are identical to the experimental conditions, and the axial length is 130 mm. Table 1 shows numbers of element and the minimum mesh size. An ideal condition was set

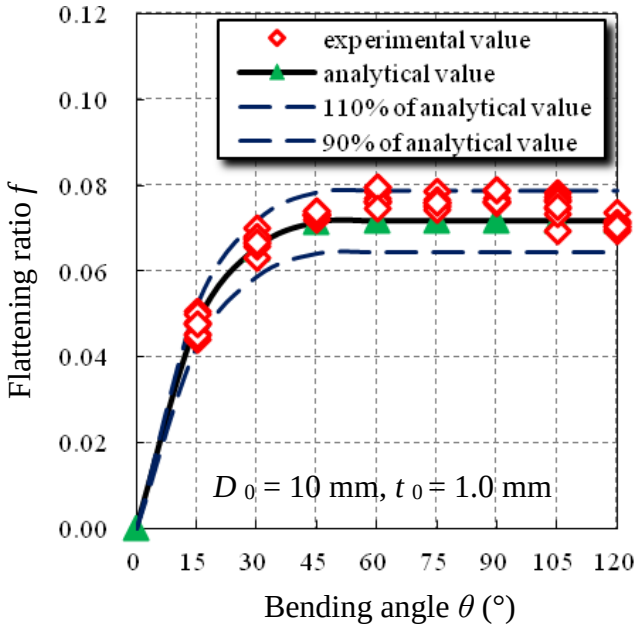



**Fig. 2 Comparison of flattening ratio for experiment and analysis results**

when the coefficient of friction $\mu$ between the pipe and the bending dice was 0.0. Young's modulus $E$ = 193 GPa, Poisson's ratio $\nu$ = 0.29, and yield stress $\sigma_y$ = 320 MPa were used as material properties. For the plastic deformation behavior, the relationship between the true stress σ and the true strain $\varepsilon$ in the plastic region is shown in Eq. 1, and an approximation formula of the Ludwik type [6] is used.

$$\sigma = \sigma_y + K_P\left(\varepsilon - \varepsilon_y\right)^n \tag{1}$$

Where $\varepsilon_y$ is the yield strain, $K_p$, and n is the material constant. In this study, $K_p$ = 1600 MPa and $n$ = 0.85 were used.

*Evaluation of the validity of the analytical model*

Figure 2 shows flattening ratios $f$ obtained by experiments and analyses. Increments of ±10% of the analysis value are indicated by a broken line. It is suggested that the average value of the experimental results is within ±10% of the analysis results. From the Fig.2, it is judged that the analysis model can sufficiently reproduce the draw bending process. Therefore, this analytical model was used in examining the wrinkle occurrence mechanism to be described later.

## Investigation on mechanism of wrinkles

*Criterion for wrinkles*

Wrinkles are believed to be due to buckling that occurs on the inside (the compression side) of the bend. However, in the actual bending process, shear stress, axial force, etc. are added to the bending moment, and in the rotational pull bending process, the machined part changes dynamically as the piping material is drawn into the bending die as the process progresses. Here, we focused on the borderline of wrinkle. Table 2 shows the dimensions of the pipe specimens and bending dice. In addition to the conditions in Table 2, experiments were conducted with $\omega$ = 7000 ° / min and $\theta$ = 120 °, and the occurrence of wrinkles at the bend was investigated by

**Table 2 Dimensions of the pipe specimens and bending dies**

| Initial outer diameter $D_0$ (mm) | Initial thickness $t_0$ (mm) | Bending radius $R$ (mm) |
|---|---|---|
| 10 | 0.5, 0.8, 1.0, 1.2 | 15, 20, 25 |
| 14 | 0.5, 0.8, 1.0, 1.2, 1.5 | 20, 28, 35 |

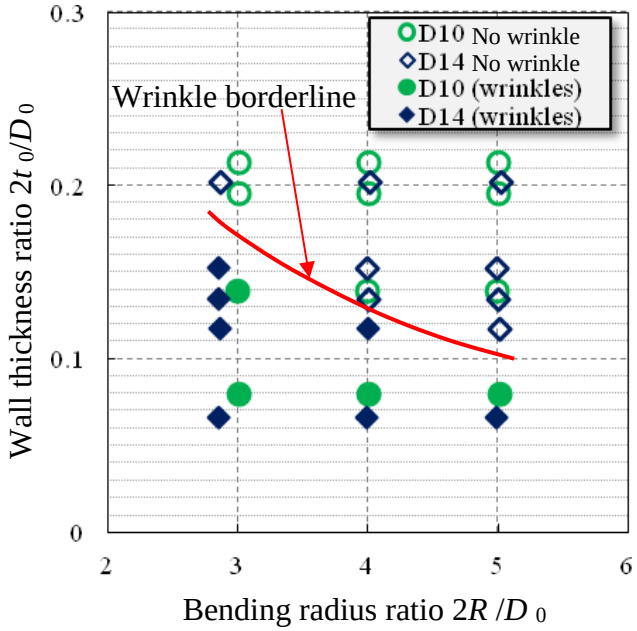



**Fig. 3 Wrinkle borderline**

the forming experiments. Figure 3 shows the bending radius ratio $2R / D_0$ and the thickness ratio $2t_0 / D_0$ in which the bending radius $R$ and the thickness $t0$ are made nondimensional by the radius $D_0 / 2$ of the piping material, respectively. The experimental conditions where wrinkles did not occur are indicated by hollow markers, and the experimental conditions where wrinkles occurred are indicated by solid markers, a boundaries of wrinkles in this experiment are shown by solid lines in the figure. From Fig.3, even if the dimensions of the material and the bending die are different, if the ratio of $D_0$, $t_0$ and $R$ is the same, the wrinkle borderline is the same.

On the other hand, simulation analysis was performed on a thin-walled pipe with $D_0 = 10$ mm, $t_0 = 0.5$ mm ($2R / D_0 = 4.0$, $2t_0 / D_0 = 0.1$, $R = 20$ mm), and an example of the result is shown in Fig. 4. As the forming process progressed, wrinkles were generated not from the bend of the pipe but from the straight pipe part on the raw pipe side. This was the same even at the stage where the first wrinkle occurred under the same conditions. Therefore, it was found that wrinkles were generated not at the bend but at the straight pipe section on the base pipe side.

*Effect of friction coefficient on wrinkles*

The borderline of the presence or absence of wrinkles was investigated by FEM analysis assuming that the coefficient of friction $\mu = 0$, 2.0, and is shown by the broken line in Fig.5. The borderline of $\mu = 2.0$ is almost in agreement with the experimental result shown by the solid

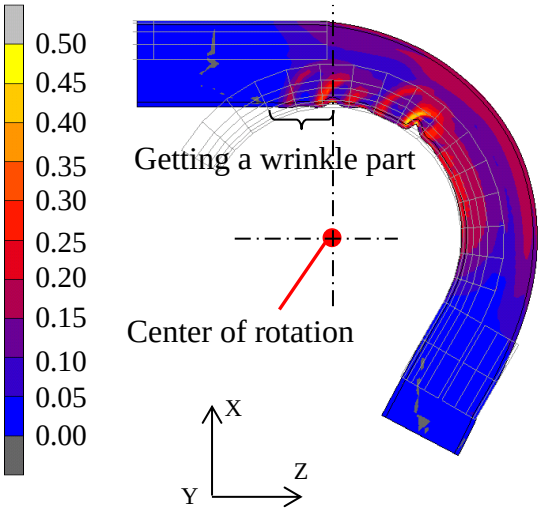



**Fig. 4. Equivalent plastic strain contour ($D_0$ = 10 mm, $t_0$ = 0.5 mm).**

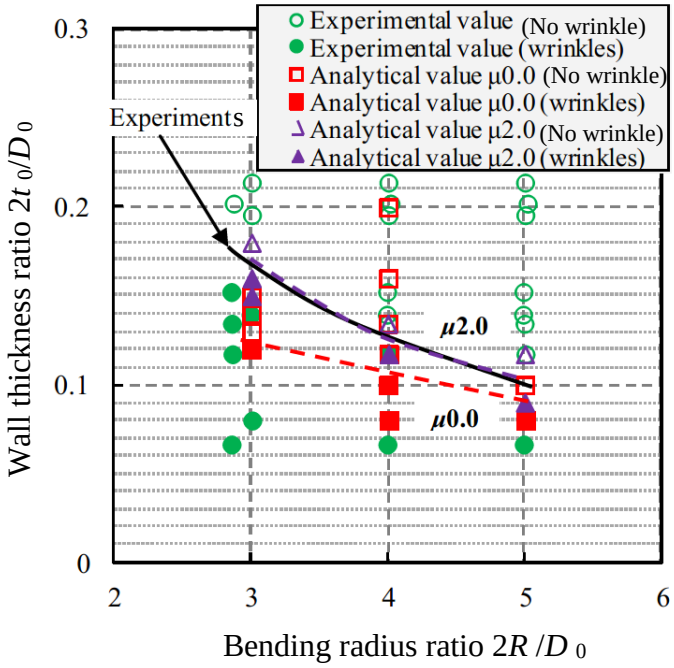



**Fig.5 Comparision of borderline curves for initiation of wrinkles**

line. The smaller the value of $\mu$, the lower the crease of creases, and the less the creases become. In addition, the smaller the bending radius of the bending die, the more sensitive the borderline of crease formation to friction, and the more it changes with the size of the coefficient of friction.

*Mechanism of wrinkles*

This study has clarified that the wrinkles occur not inside the bend but inside the straight pipe. Therefore, using the contact point between the pipe and the bending die at $\theta$ = 0 ° as the reference point, the displacement $\Delta x$ of the bending inner surface of the pipe in the x-axis direction (see Fig. 4) was investigated from the reference point to the base pipe side. Fig. 6 shows the x-axis displacement $\Delta x$ for each bending angle at $D_0$ = 14 mm, $t_0$ = 0.5 mm, $R$ = 28

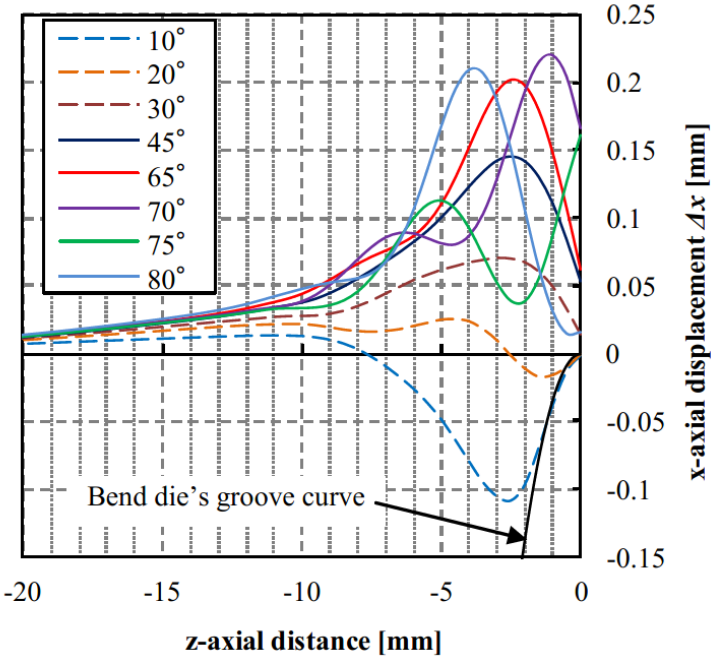



**Fig.6 x-axial displacement ($2R / D_0 = 4.0$, $2t_0 / D_0 = 0.1$, $\mu = 2.0$)**

mm ($2R / D_0 = 4.0$, $2t_0 / D_0 = 0.071$), and $\mu = 2.0$ obtained by analysis. The broken line is $\Delta x$ at the stage where wrinkles have not occurred.

Immediately after the start of bending, deformation caused by the Hertzian contact occurred, which is also referred to as the drawing-rolling phenomenon, in the draw bending processing. The deflection to the side of the bending die due to the drawing-rolling phenomenon changes with the progress of processing, becomes maximum around the bending angle $\theta = 10°$ to $15°$, and then decreases. Furthermore, with the increase of the bending angle $\theta$, the deformed portion due to the drawing-rolling phenomenon is made uniform by the contact with the bending die, and the straight pipe portion on the raw pipe side begins to bend outward on the opposite side of the bending die.

The deflection converges to a constant value under processing conditions where wrinkles do not occur. On the other hand, under the processing conditions where wrinkles occurred, the deflection continued to increase and eventually became wrinkles. In addition, in this study, it was confirmed that wrinkles occurred in the region from the flexible portion to the straight pipe portion, which was slightly deviated from the maximum flexible portion, but not from the maximum flexible portion.

In the experiment, wrinkles occurred, but in the analysis of $\mu = 0.0$, deformation similar to that in the case of wrinkles was observed even under processing conditions in which no wrinkles occurred. However, under this condition, the deformation was equalized by the bending die as the processing progressed, and no wrinkles were finally confirmed. From these facts, it was found that the wrinkles are formed by the transition of the flexure due to the drawing-rolling phenomenon.

## Conclusions

The following findings were obtained from this study.

Wrinkles occur not at the bend but at the straight pipe part on the raw pipe side.

The coefficient of friction between the pipe and the bending die promotes the occurrence of wrinkles and the growth of the generated wrinkles.

The deflection of piping starting from the drawing-rolling phenomenon converges to a constant value in the processing condition where wrinkles do not occur.
Wrinkles are formed by the transition of flexures due to the drawing-rolling phenomenon in rotary bending.

## References

[1] The Japan Plastic Processing Society, *Tube Forming – Secondary Processing of Tubular Materials and Product Design* –, Corona Company, Tokyo, 1992, pp. 39 – 40.
[2] Shigeki Mori, ken–ichi Manabe, Hisashi Nishimura and Kimio Tamura, Deformation of Clad Tubes in Rotary Draw Bending, *J. of JSTP*, **33**–378 (1992) 862 – 867.
[3] Osamu Sonobe, Yuji Hashimoto, Koji Suzuki, Kei Sakata and Ken–ichi Kawai, Experimental and Analytical Studies of Deformation Behavior of ERW Tubes in Rotary Draw Bending, *J. of JSTP*, **51** – 589 (2010) 121 – 125.
[4] Shuji Sakaki, Yusuke Okude and Shoichiro Yoshihara, Working Limit for Several Types of Cross Section in Draw Bending, *J. of JSTP*, **53** – 614 (2012) 246 – 250.
[5] Masabumi Yuhara, Makoto Hoshino and Osamu Wada, Draw Bending of Tube Using Ultrasonic-Vibration Plug, *J. of JSTP*, **53** – 618 (2012) 646 – 650.
[6] Fusahito Yoshida, *Foundations of Elastic Plastic Dynamics*, First edition, 12 press, Kyoritsu Publishing, Tokyo, 2008, p. 116.

# Equivalent Circuit Modeling and Analysis Study for Vortex-induced Aerodynamic Energy Harvesting

***Jinda Jia[1,2], Xiaobiao Shan[1], †Tao Xie[1] and and †Yaowen Yang[2]**

[1]State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin, China

[2]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

*Presenting author: jiajinda@hit.edu.cn

†Corresponding author: xietao@hit.edu.cn; cywyang@ntu.edu.sg

## Abstract

Low-speed wind energy can be captured to power wireless sensors or remote micro electro mechanical systems by utilizing vortex-induced vibration phenomenon of a cylinder tip connected to a piezoelectric cantilever. Researchers have developed several configurations to study the behavior of vortex-induced energy harvesters. However, these theoretical analysis or experimental prototypes only consider the interface circuit as a pure resistance. The interface circuit for practical usage may be more complex or the powered wireless sensor needs direct voltage input. The solve the complex interface circuit problem, this paper presents an equivalent circuit modeling and analysis study for vortex-induced energy harvesting. The mechanical parameters are substituted by standard circuit elements. The nonlinear vortex-induced aerodynamic force is represented by self-defined function sources. The total equivalent circuit is simulated by circuit simulation software. The simulation results are consistent to our previous experimental results, which can verify the accurateness of the presented equivalent circuit model.

**Keywords:** equivalent circuit modeling; energy harvest; vortex-induced vibration

## Introduction

Wireless sensing nodes or electronic instruments have been used in structural health monitoring, medical health examination and micro electro mechanical systems, which promote the research of energy harvesting from the ambient environment or vibrations [1-3]. Different energy convention mechanisms including electrostatic [4,5], electromagnetic [6,7] and piezoelectric [8-10] mechanisms. Energy harvesting via piezoelectric mechanism has attracted much attention because of easy construction, clean and high output power density. Quite numbers of researchers focus on energy harvesting from base excitations [11-13]. However, the stable base excitation exists rarely in nature and thus limits its usage situations.

Wind exists widely in nature and has the potential to be utilized for energy harvesting to drive sensors in inaccessible mountain, canyon or seabed. In terms of energy harvesting from fluid flows, Akaydin et al. [14] proposed a piezoelectric energy harvester consisted of a cantilevered beam with attachment cylinder and the results showed a non-rigid bonding model had a better agreement with experiments than non-rigid bonding model. Wang et. al [15] studied a d31 mode piezoelectric energy harvester generating voltage from pressure oscillation in pressure chamber and the energy harvester could generate 2.2 V. Li et. al [16] tested a bio-inspired piezo-leaf architecture using flexible piezoelectric materials and a single leaf had 2 $mW/cm^3$ power density. Weinstein et al. [17] studied the interactions between a fin and downstream vortex shedding. The results showed the addition of the fin could make a significant improvement output power from a piezoelectric energy harvester. Gao et. al [18] presented an upright energy harvester with cylinder extension. The experimental results

showed the piezoelectric energy harvester generated higher voltage in turbulent flow because of additional contribution in the lock-in region than in laminar flow. Dai [19,20] studied the output power from both vortex-induced vibration and base excitation. Among other different shapes of attachments, Liu et al. [21] proposed a Y-type three-blade bluff body and the experiment showed its superiority than a square prism. Abdelkefi et al. [22] proposed an energy harvester with an equilateral triangle cross-section and found that the minimum transverse displacement amplitude resulted in maximum harvested power. Yang et al. [23] compared the output abilities of energy harvesters with different cross sections. The experimental results demonstrated the superiority of the square-sectioned tip. However, most researchers only consider the external circuit as a pure resistance because external circuits are difficult to be modeled in mechanical systems.

Therefore, this paper puts forward a novel modeling method for vortex-induced vibration energy harvester according to the principle of equivalent circuit, which can handle complex external circuit. The mechanical parameters in practical situations are replaced by electrical circuit element. The piezoelectric coefficient parameter is replaced by an ideal transformer. The vortex-induced force is represented by oscillation circuit. The equivalent circuit model is simulated in analog software for circuits. The results of an equivalent circuit model are consistent with our previous experimental results, which can prove the validity of the modeling method.

**Conventional Modeling of Vortex-induced Vibration Energy Harvesting**

As Fig. 1 shows, the vortex-induced vibration piezoelectric energy harvester consists of a piezoelectric cantilever beam with cylinder extension. The cylinder attachment undergoes periodic pressure when the wind flows through it, and there occurs vibrations when the frequency is near to the natural frequency of energy harvester. The strains in the piezoelectric layer generate output voltage across the interface circuit. Here we directly use the governing equation through the nonlinear distributed parameter model by our previous work [24], as shown in Eqs. (1-3).

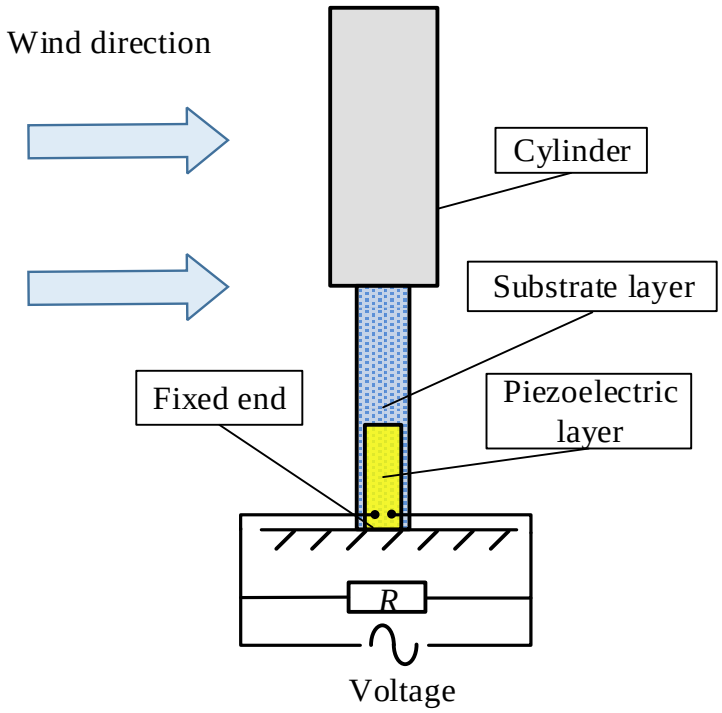



**Figure 1. Composition of piezoelectric energy system [24]**

$$\ddot{r}(t)+(2\zeta\omega+\eta)\dot{r}(t)+(\omega^2+\mu)r(t)+\theta V(t)=\frac{C_{L0}\rho_f DU^2}{4}\left(L_c\phi_{12}(L_s)+\frac{L_c^{\,2}}{2}\phi'_{12}(L_s)\right)q(t) \tag{1}$$

$$C_P\frac{dV_{RL}(t)}{dt}+\frac{V_{RL}(t)}{R}-\theta\dot{r}(t)=0 \tag{2}$$

$$\ddot{q}(t)+\varepsilon\omega_f\left(q^2(t)-1\right)\dot{q}(t)+\omega_f{}^2 q=\frac{A}{D}\left(\phi_{12}(L_s)+\frac{L_c}{2}\phi'_{12}(L_s)\right)\ddot{r}(t) \tag{3}$$

where $r(t)$ is model coordinate; $\zeta$ is the damping coefficient; $\omega$ is the natural frequency of energy harvester; $\eta$ is fluid resistance coefficient; $\mu$ the coefficient of gravity effect; $\theta$ is piezoelectric coefficient; $V(t)$ is output voltage; $C_{L0}$ is lift coefficient; $D$ and $L_c$ is diameter and length of cylinder, respectively; $C_p$ is capacitance of piezoelectric sheet; $\rho_f$ and $U$ are the density and velocity of air flows, respectively; $\varepsilon$ and $A$ are constants; $\omega_f$ is frequency of vortex; $q(t)$ is a parameter used to describe the behavior on the near wake of the cylinder.

**Equivalent Circuit Modeling**

In this part, the equivalent circuit method is proposed. Equivalent electrical components for vortex-induced force and mechanical parts of energy harvesting system are listed in Table 1.

**Table 1. Analogy between mechanical domain and electrical domain**

| Mechanical parameters | Electrical parameters |
|---|---|
| $r(t)$ | Charge: $Q_1(t)$ |
| $dr(t)/dt$ | Current: $i_1(t)$ |
| $q(t)$ | Charge: $Q_2(t)$ |
| $dq(t)/dt$ | Current: $i_2(t)$ |
| 1 | Inductance: $L_1$, $L_2$ |
| $2\zeta\omega+\eta$ | Resistance: $R_1$ |
| $1/(\omega^2+\mu)$ | Capacitance: $C_1$ |
| $1/\omega_f^2$ | Capacitance $C_2$ |
| $\frac{C_{L0}\rho_f DU^2}{4}\left(L_c\phi_{12}(L_s)+\frac{L_c^{\,2}}{2}\phi'_{12}(L_s)\right)q(t)$ | Voltage: $V$(A3) |
| $\frac{A}{D}\left(\phi_{12}(L_s)+\frac{L_c}{2}\phi'_{12}(L_s)\right)\ddot{r}(t)$ | Voltage: $V$(A2) |
| $-\theta$ | Turn ratio: $N1$ |

Eqs. (1-3) can be rewritten by replacing electrical parameters into mechanical parameters as Eqs. (4-6).

$$L_1\ddot{Q}_1(t)+R_1\dot{Q}_1(t)+C_1Q(t)-NV(t)=V(\mathrm{A3}) \tag{4}$$

$$C_P\frac{dV_{RL}(t)}{dt}+\frac{V_{RL}(t)}{R}+N\dot{Q}_1(t)=0 \tag{5}$$

$$L_2\ddot{Q}_2(t)+V(\mathrm{A1})/I(\mathrm{V5})\dot{Q}_2(t)+C_2Q(t)=V(\mathrm{A2}) \tag{6}$$

## Model validation

In this part, the equivalent circuit model is validated [24]. The pure resistance is chosen as the external circuit for comparison with our previous experimental results. According to the analogy between the mechanical and electrical domains, the parameters of the equivalent circuit model are obtained from Table 1. Fig. 2 shows the circuit diagram of equivalent circuit when wind velocity is 4.2 m/s. Fig. 3 shows time-voltage diagram across *RL* when wind velocity is 4.2 m/s. As time goes on, the voltage fluctuates about 3.5 seconds and then forms a steady periodic fluctuation.

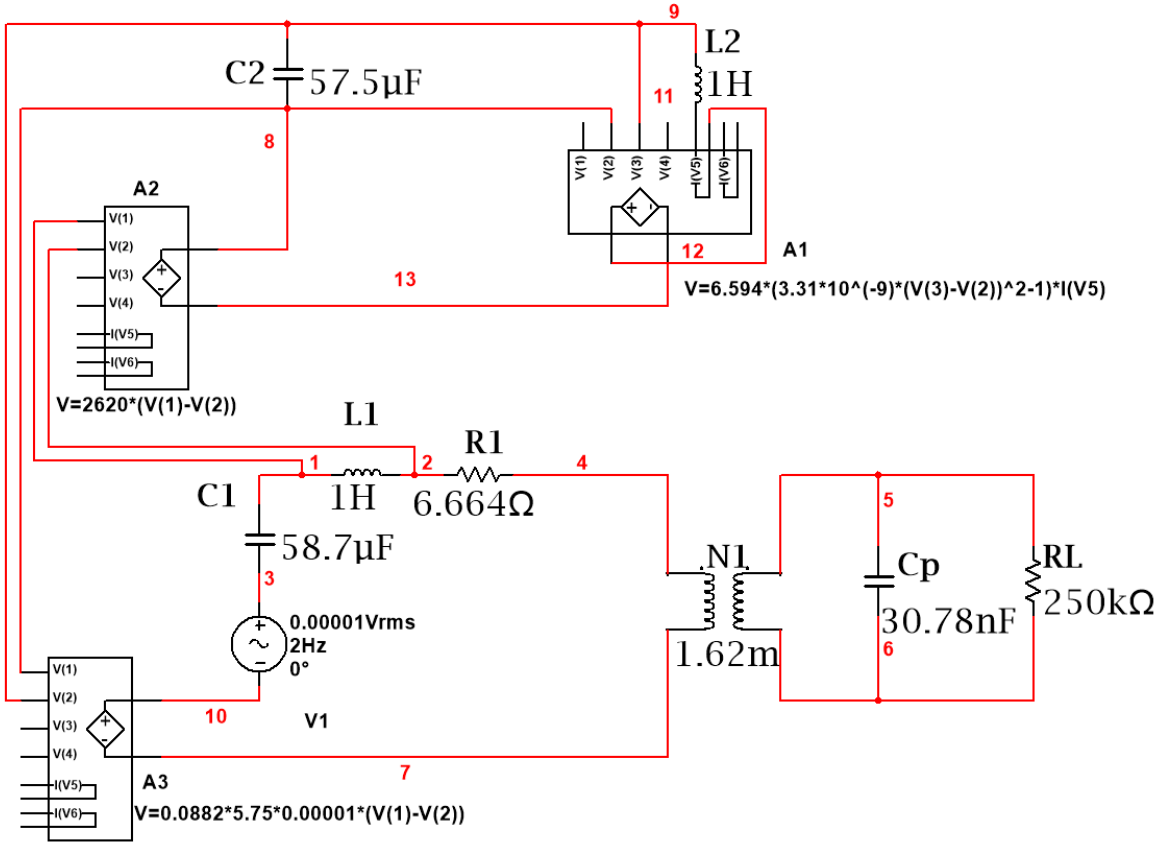



**Figure 2. Circuit diagram of equivalent circuit when wind velocity is 4.2 m/s**

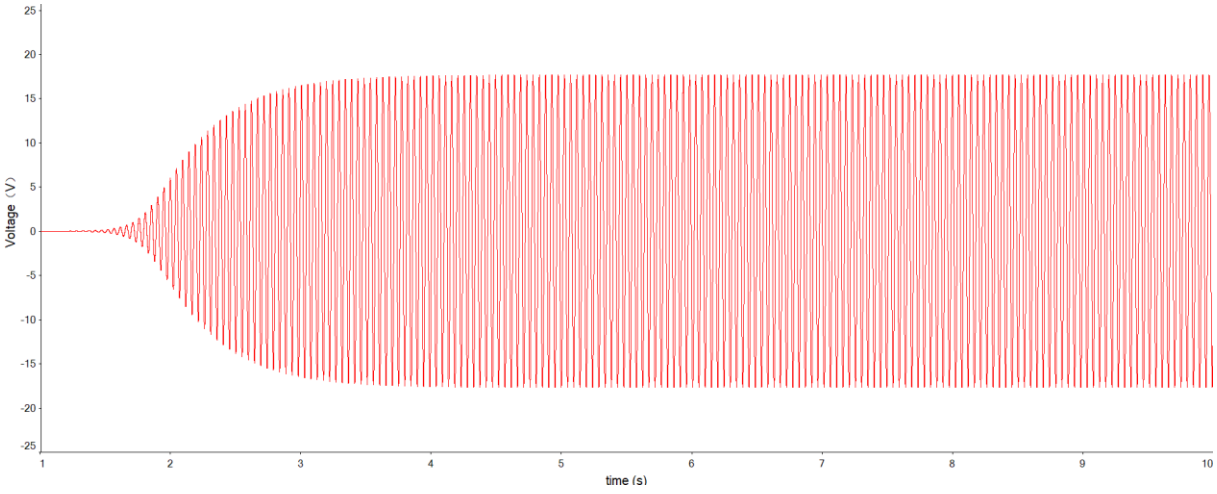



**Figure 3. Time-voltage diagram**

To investigate the effect of external resistance *RL* on energy harvesting ability, a serious of external resistance is chosen to calculate output power *P*. Fig. 3 shows the output power from circuit simulation and experimental versus load resistance *RL* when the wind velocity is 4.2 m/s. The circuit simulation shows that the maximum power is 626.58 μW at 250 kΩ while 605.0 μW at 250 kΩ from circuit result. The optimum resistance from experimental results is consistent with that from circuit simulation.

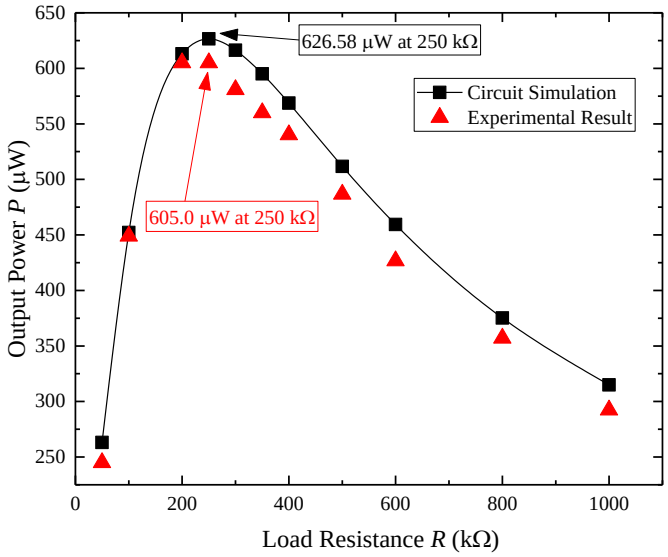



**Figure 4. Output power from circuit simulation and experimental versus load resistance**

For further investigating the influence of wind speed *U*, the output power of load resistance at various wind speeds are calculated through circuit simulation. Both circuit simulation and experimental resistance is chosen to 250 kΩ, which is the optimal resistance from fig. 3. Figure 4 indicates the relationship between output power and wind speed from circuit simulation and experimental result. The maximum output power *P* is 635.04 μW when the wind speed *U* is 4.2 m/s, which is obtained in the experiment. The maximum average output power calculated by circuit simulation is 628.91 μW when the wind speed *U* is 4.3 m/s. The maximum output power and optimal wind velocity from circuit simulation matches well with that from experiment.

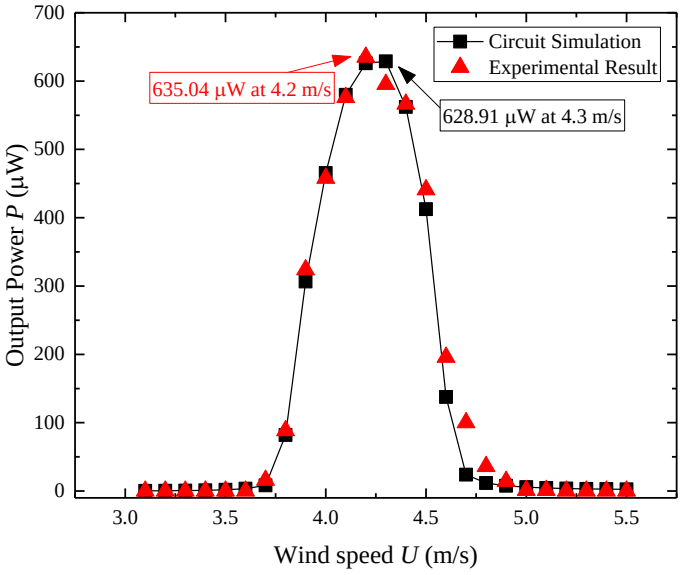



**Figure 5. Time-voltage diagram**

## Conclusions

This paper investigates the equivalent circuit modeling method for vortex-induced piezoelectric energy harvesting. First the analogy between mechanical domain and electrical domain is presented and used to replace the aerodynamic fluid-solid-electricity governing equations by equivalent circuit equations. Next, the equivalent circuit is simulated in circuit simulation software. The output voltage becomes steady sine wave in a few seconds. Finally, the effects of external load resistance and wind velocity are studied and compared with experimental results. The comparative results illustrate that the optimal load resistance, the optical wind velocity and maximum output power from circuit simulation are consistent with that from experimental results, which validities the accurateness of the equivalent circuit simulation. This work provides basic simulation method for designing external circuit for vortex-induced piezoelectric energy harvesting.

### References

1. Zuo, L.; Scully, B.; Shestani, J.; Zhou, Y. Design and characterization of an electromagnetic energy harvester for vehicle suspensions. *Smart Materials and Structures* **2010**, *19*.
2. Matiko, J.W.; Grabham, N.J.; Beeby, S.P.; Tudor, M.J. Review of the application of energy harvesting in buildings. *Measurement Science and Technology* **2014**, *25*.
3. Amin Karami, M.; Inman, D.J. Powering pacemakers from heartbeat vibrations using linear and nonlinear energy harvesters. *Appl Phys Lett* **2012**, *100*.
4. Crovetto, A.; Wang, F.; Hansen, O. Modeling and optimization of an electrostatic energy harvesting device. *Journal of Microelectromechanical Systems* **2014**, *23*, 1141-1155.
5. Khan, F.U.; Qadir, M.U. State-of-the-art in vibration-based electrostatic energy harvesting. *J Micromech Microeng* **2016**, *26*.
6. Foisal, A.R.M.; Hong, C.; Chung, G.S. Multi-frequency electromagnetic energy harvester using a magnetic spring cantilever. *Sensors and Actuators a-Physical* **2012**, *182*, 106-113.
7. Dias, J.A.C.; De Marqui, C.; Erturk, A. Hybrid piezoelectric-inductive flow energy harvesting and dimensionless electroaeroelastic analysis for scaling. *Appl Phys Lett* **2013**, *102*.
8. Liu, H.; Zhong, J.; Lee, C.; Lee, S.-W.; Lin, L. A comprehensive review on piezoelectric energy harvesting technology: Materials, mechanisms, and applications. *Applied Physics Reviews* **2018**, *5*.
9. Nechibvute, A.; Chawanda, A.; Luhanga, P. Piezoelectric energy harvesting devices: An alternative energy source for wireless sensors. *Smart Materials Research* **2012**, *2012*, 1-13.
10. Lin, X.J.; Zhou, K.C.; Zhang, X.Y.; Zhang, D. Development, modeling and application of piezoelectric fiber composites. *Transactions of Nonferrous Metals Society of China* **2013**, *23*, 98-107.
11. Erturk, A.; Inman, D.J. On mechanical modeling of cantilevered piezoelectric vibration energy harvesters. *Journal of Intelligent Material Systems and Structures* **2008**, *19*, 1311-1325.

12. Zhang, Y.L.; Wang, T.Y.; Luo, A.X.; Hu, Y.S.; Li, X.; Wang, F. Micro electrostatic energy harvester with both broad bandwidth and high normalized power density. *Applied Energy* **2018**, *212*, 362-371.

13. Abdelkefi, A.; Barsallo, N.; Tang, L.; Yang, Y.; Hajj, M.R. Modeling, validation, and performance of low-frequency piezoelectric energy harvesters. *Journal of Intelligent Material Systems and Structures* **2013**, *25*, 1429-1444.

14. Akaydin, H.D.; Elvin, N.; Andreopoulos, Y. The performance of a self-excited fluidic energy harvester. *Smart Mater Struct* **2012**, *21*.

15. Wang, D.A.; Ko, H.H. Piezoelectric energy harvesting from flow-induced vibration. *J Micromech Microeng* **2010**, *20*.

16. Li, S.G.; Yuan, J.P.; Lipson, H. Ambient wind energy harvesting using cross-flow fluttering. *J Appl Phys* **2011**, *109*.

17. Weinstein, L.A.; Cacan, M.R.; So, P.M.; Wright, P.K. Vortex shedding induced energy harvesting from piezoelectric materials in heating, ventilation and air conditioning flows. *Smart Materials and Structures* **2012**, *21*.

18. Gao, X.T.; Shih, W.H.; Shih, W.Y. Flow energy harvesting using piezoelectric cantilevers with cylindrical extension. *Ieee T Ind Electron* **2013**, *60*, 1116-1118.

19. Dai, H.L.; Abdelkefi, A.; Wang, L. Theoretical modeling and nonlinear analysis of piezoelectric energy harvesting from vortex-induced vibrations. *Journal of Intelligent Material Systems and Structures* **2014**, *25*, 1861-1874.

20. Dai, H.L.; Abdelkefi, A.; Wang, L. Piezoelectric energy harvesting from concurrent vortex-induced vibrations and base excitations. *Nonlinear Dynam* **2014**, *77*, 967-981.

21. Liu, F.R.; Zou, H.X.; Zhang, W.M.; Peng, Z.K.; Meng, G. Y-type three-blade bluff body for wind energy harvesting. *Appl Phys Lett* **2018**, *112*.

22. Abdelkefi, A.; Yan, Z.M.; Hajj, M.R. Modeling and nonlinear analysis of piezoelectric energy harvesting from transverse galloping. *Smart Materials and Structures* **2013**, *22*.

23. Yang, Y.W.; Zhao, L.Y.; Tang, L.H. Comparative study of tip cross-sections for efficient galloping energy harvesting. *Appl Phys Lett* **2013**, *102*.

24. Jia, J.; Shan, X.; Upadrashta, D.; Xie, T.; Yang, Y.; Song, R. Modeling and analysis of upright piezoelectric energy harvester under aerodynamic vortex-induced vibration. *Micromachines (Basel)* **2018**, *9*, 667.

# Numerical study of the solute dispersion in microchannel with interphase transport

**Wenbo Li[1]*, Wenyao Zhang[1], Huang Deng[1], and Qian Fang[1], Cunlu Zhao[1]†**

[1] Key Laboratory of Thermo-Fluid Science and Engineering of MOE, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, P.R. China

*Presenting author: wenboli817@gmail.com

†Corresponding author: mclzhao@xjtu.edu.cn

## Abstract

The dispersion phenomenon in the pressure-driven microchannel flow of solute with interphase transport is numerically studied using a transient 2D model. The interphase mass transport between the mobile phase and the stationary phase of the microchannel plays a critical role in the process of solute dispersion. The traditional studies of the solution dispersion with the interphase transport are based on the moment analysis which actually uses the 1D model and cannot provide detailed understandings of the solute dispersion in the microchannel. In this work, the 2D numerical model enables a more detailed characterization of the transient evolution of the solute dispersion in a microchannel by predicting the 2D transient concentration contours of the solute. The model characterizes the effect of interphase mass transport on the solute dispersion with two parameters, i.e., the partition coefficient ($K$) and the kinetic mass transfer rate ($k_f$). From the 2D concentration contours, we observe that the solute is more dispersed in the mobile phase along the microchannel axis and also moves slower with $K$ increasing, which indicates significant dispersion and retention of solute in the microchannel respectively. For a smaller $k_f$ which means a high mass transfer resistance between the mobile and stationary phases, the solute concentration in mobile phase is more dispersed and also the corresponding concentration profile along the channel axis is more skewed. It is also noticed that a thicker layer of stationary phase ($d_f$) corresponding to a larger solute capacity of the stationary phase, causes more significant retention of solute in the microchannel. The results from this study provide a straightforward picture of the solute dispersion phenomenon in microchannel with interphase transport and are of high relevance to practical applications such as chromatography and microfluidics.

**Keywords:** Microchannel; Dispersion; Interphase Transport; Concentration Contour

## Introduction

With the rapid development of microfluidic lab-on-a-chip (LOC) technology, study of fluids and solute transport in microchannels have received widespread attention since miniaturization of fluid channels in microfluidic LOC devices leads to new problems for the fluid and mass transport. The dispersion of solutes in fluids flows in microchannels is one of the most widely studied topic due to its key roles in the trace detection of samples, chromatography, multiphase microfluidics and soil remediation[1-3], etc.

The study of solute dispersion has been a constant focus since Taylor who investigated analytically and experimentally the solute dispersion in the water flowing in a tube due to the hydrodynamic convection and molecular diffusion [4-6]. Basing on Taylor's studies, Aris [7] presented a theoretical study of the solute dispersion by means of moments analysis. The works by Taylor and Aris form the theoretical basis for the solution dispersion, their approach is jointly termed as the Taylor-Aris theory. Later Aris studied the solute dispersion in two coaxial layers of immiscible fluids flowing in an annulus. In addition to the traditional solution dispersion mechanisms of hydrodynamic convection and molecular diffusion, the interphase exchange between two fluids was also taken into account as a new mechanism. The consideration of this new mechanisms is due to a number of applications such as distillation and partition chromatography in which interphase mass transport has a crucial contribution [8]. In such researches, moments are adopted as the theoretical tool because of their ability in providing good statistical descriptions of the solute concentration distribution in the term of time. Kučera [9] derived explicit moments expressions as part of his study on non-equilibrium chromatography considering the longitudinal diffusion in the mobile phase, the radial diffusion inside the porous grains of the packing material, the finite rate of mass transfer through the boundary. Grushka [10] related the moments of a chromatographic peak to the determinative experimental parameters, i.e., partition coefficients, column length, etc. J. A. Jonsson proposed that considering the solute dispersion, the median of the chromatographic peak should be taken as the best measurement of retention time[11] and gave the moments to specifically study the dispersion and to determine the isotherm[12]. The moment analysis not only provides insightful understanding of fundamental characteristics of solution dispersion in channel flows, but also is practically useful in determining the diffusivity of solutes in specific solutions and the partition coefficient of solute between two phases [13-16] by the inverse chromatography.

Another popular method to study solute dispersion in microchannel flow is based on the concept of height equivalent to a theoretical plate (H.E.T.P.) [17-19]. Different from the moment analysis which focuses only on the statistical description of solution dispersion on time scale, H.E.T.P describes the instant equilibrium of solution dispersion on length scale. The introduction of H.E.T.P to the solute dispersion study is due to its wide use as an indicator for separation performance in the chromatography. Various investigations performed fundamental study of the solution dispersion in chromatography using the H.E.T.P. and particularly discussed the effects of various factors, such as solute diffusivity and phase ratio, on the

H.E.T.P. [20-22]. Fabrice presented a comprehensive summary for the solution dispersion studies based on the H.E.T.P. in the field of chromatography [23]. Recently, Beauchamp investigated the solute dispersion in both short and long capillary with the H.E.T.P under the slip boundaries and reached the conclusion that only in tube of very small diameters the use of slip flow boundary to reduce chromatography dispersion is suitable [24].

The aforementioned moment analysis and H.E.T.P methods focus on different aspects of the solute dispersion, but are all the simplified models of the more general convection-diffusion theory of mass transport. As has been reviewed previously, there has been significant development in the moment analysis and H.E.T.P. for investigating solute dispersion. Yet, these two methods are derived from the general convection-diffusion theory with a cross-section average treatment, and thus cannot provide detailed information of the solute concentration distribution in the whole channel domain. Especially, the details of the solute exchange between two phases (mobile phase and stationary phase) are missing from these simplified models. The present work is to study the solute dispersion in a pressure-driven microchannel flow with interphase transport with a full numerical model based on the convection-diffusion theory. The model is able to give an intuitive understanding of the solution transport with the numerically predicted transient concentration contours in the microchannel. More interestingly, the interphase transport of solute is to be studied in an unprecedented detail. Our numerical analyses also systematically address the effects of various model parameters on the solution dispersion characteristics.

## Methods

### *Mathematical model*

Considering a circular straight microchannel whose inner wall holds an extremely thin layer of static polymer liquid, gases flow in the microchannel without causing the flow of liquid. The gases are termed as the mobile phase and the polymer liquid as the stationary phase.

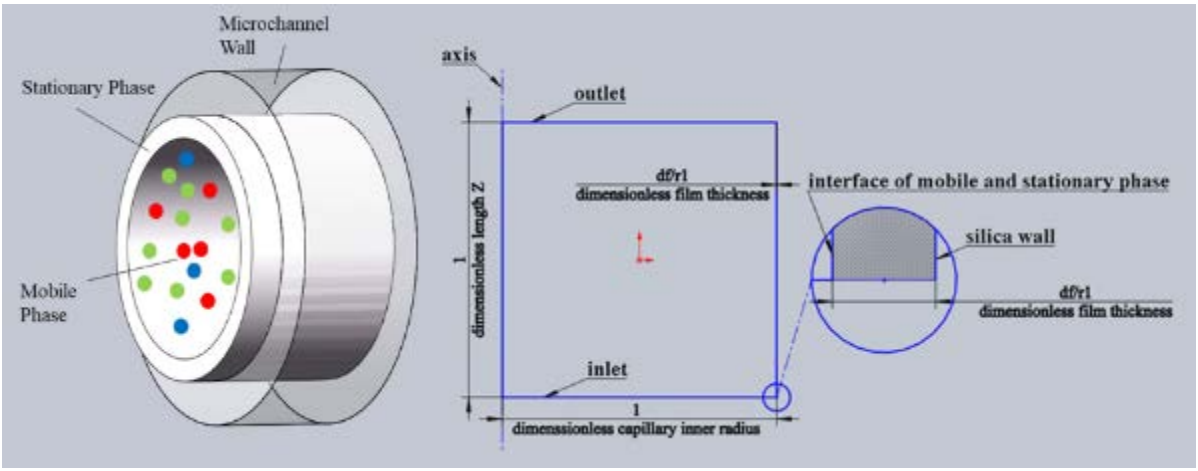



Figure 1. Schematic of physical model and computational domain

The physical model and the corresponding computational domain of this study are showed in figure 1. The mobile phase is mixed of the carrier gas inert to the stationary phase and the injected gaseous solute. In this study, the mobile phase is incompressible and under isothermal condition. The diffusions along both axial and radial conditions are considered and the diffusivities of solute in both mobile and stationary phase are concentration independent. The stationary phase is homogenous and has constant thickness along axial direction of the microchannel. What's more, no chemical reaction occurs during the flow through the microchannel. Due to the magnitudes of the diameter and the flow rate in microchannel, the velocity distribution of mobile phase obeys Poiseuille's law. Considering that the solute is dilute, the mass convection-diffusion equation is only performed on solute. Based on above assumptions and equations from results of Aris[8], and with dimensionless parameters introduced as following,

$$c_m = \left(\frac{c_1}{c_0}\right) \quad Z = \left(\frac{z}{L}\right) \quad c_s = \left(\frac{c_2}{c_0}\right) \quad T = \left(\frac{U_1 t}{L}\right) \quad R = \left(\frac{r}{r_1}\right)$$

the dimensionless PDEs, I.C. and B.C.s of this study are:

$$\frac{\partial c_m}{\partial T} + 2[1-R^2]\frac{\partial c_m}{\partial Z} = \frac{h_0}{Pe_1}\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{\partial c_m}{\partial R}\right) + \frac{1}{Pe_1 h_0}\frac{\partial^2 c_m}{\partial Z^2} \qquad 0 < R < 1 \tag{1}$$

$$\frac{\partial c_s}{\partial T} = \frac{h_0}{Pe_2}\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{\partial c_s}{\partial R}\right) + \frac{1}{Pe_2 h_0}\frac{\partial^2 c_s}{\partial Z^2} \qquad 1 < R < \frac{r_1 + d_f}{r_1} \tag{2}$$

Initial condition is set that no solute is in microchannel before injection:

$$c_m(R,Z,T) = c_s(R,Z,T)/K = 0 \qquad at\ T = 0, 0 < Z < 1 \tag{3}$$

At the inlet of the microchannel, the Dirac delta function is adopted to express the instantaneously uniform injection.

$$c_m(R,Z,T) = \frac{U_1}{L}\delta(T) \qquad at\ Z = 0 \tag{4}$$

At the interface of the mobile and the stationary phases, a kinetic equation is used to govern the interphase transport complying with the mass conservation law. The direction of interphase transport is decided by the value of $c_m$-$c_s/K$. The partition coefficient, $K$, physically brings solute to bear a linear limit concentration relationship between the mobile phase and the stationary phase. $K_f$ represents the dimensionless mass transport rate at the interface of the mobile and the stationary phase. When $K_f$ goes infinite the boundary condition (5) degenerates to the ideal one[14].

$$\frac{h_0}{Pe_1}\frac{\partial c_m}{\partial R} = \frac{h_0}{Pe_2}\frac{\partial c_s}{\partial R} = -K_f(c_m - c_s/K) \qquad at\ R = 1 \tag{5}$$

Symmetric condition is set at the axis of microchannel.

$$\frac{\partial c_m}{\partial R} = 0 \qquad at\ R = 0 \tag{6}$$

And no reaction or adsorption occurs between solute and microchannel wall.

$$\frac{\partial c_s}{\partial R} = 0 \qquad at\ R = \frac{r_1 + df}{r_1} \tag{7}$$

where $r_1$ is the inner radius and $L$ the length of microchannel and $d_f$ is the thickness of the stationary phase. $c_0$ is the injection concentration of the solute and $U_1$ the mean velocity of the mobile phase. The concentrations are $c_1$ and $c_2$ while the diffusivities of solute are $D_1$ and $D_2$, respectively in the mobile phase and the stationary phase.

The definition of dimensionless parameters are $Pe_1=r_1u_1/D_1$, $Pe_2=r_1u_1/D_2$, $K_f = k_f t_0/r_1$, $h_0=L/r_1$, where $t_0=L/U_1$represents the time for carrier gas to flow through the microchannel. Here one should distinguish the Peclet number $Pe_1$ from $Pe_2$ which only takes the symbol of Peclet number, and the physically meaning of $Pe_2$ is the ratio of the convection in the mobile phase over the solute diffusivity in the stationary phase.

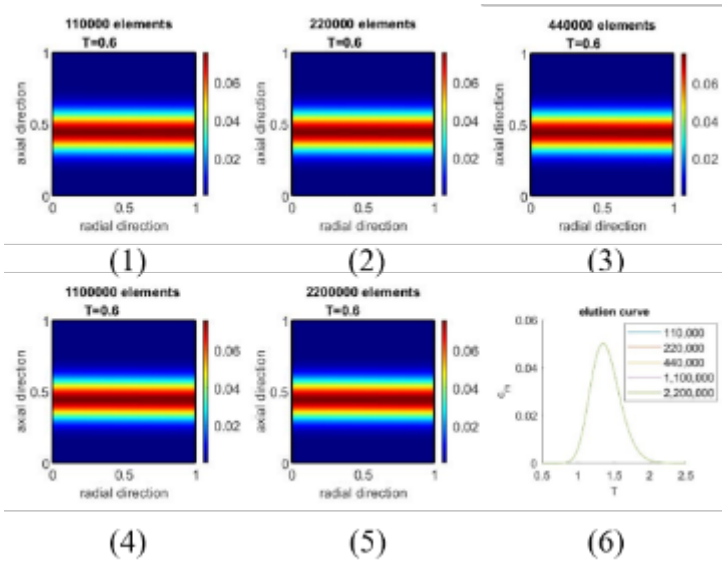



Figure 2. Mesh independence verification

*Model verification*

The finite element method is adopted for the computation operated by COMSOL Multiphysics 5.4 in this study. The mesh is constructed quadrilateral in the computational domain in figure 1. The distribution of mesh elements is symmetry in both axial and radial directions while near the interface of the mobile and the stationary phase the mesh is more compact. To conduct a

transient computation, the relative tolerance in solver is set 0.001 with respect to time. Figure 2 shows the verification of mesh independence with varying total mesh elements from 110,000 to 2,200,000. From the contours and the concentration profile in terms of time at the outlet of the microchannel, namely, the elution curve, of different cases, the mesh independence is verified. We take 1,100,000 as the number of mesh elements to perform computation and set $t_0 = 40,\ h_0 = 100$ constant in all the computational cases. Values of other parameters are set based on practicability. Pawlisch[14] once gave a computational result in his paper and we recurred his result with the 1-D model as the reference. Then we computed the dimensionless 2-D model before-mentioned under the same condition of Pawlisch's result for contrast. The results and comparation are shown in figure 3. The picture shows good agreement of two models indicating that the dimensionless 2-D model is correct.

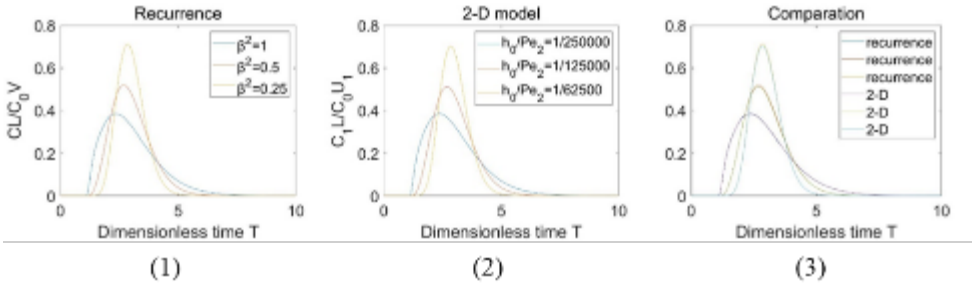



Figure 3. Verification of proposed 2-D model by Recurrence of Pawlisch's Case

**Results and discussion**

Solute dispersion in microchannel with interphase transport is affected by three simultaneous process, i.e., the hydrodynamic convection/molecule diffusion in the mobile phase[4], the molecule diffusion in stationary phases and the interphase transport between the two phases[8]. The discussion in this section consists of three subsections. In the first subsection, we show the general influence of interphase transport by comparing the concentration contours of solute dispersion with and without interphase transport. In the second subsection, we use concentration distribution contours to illustrate the impact of $Pe_1$ and $Pe_2$ on dispersion. In the third section, we discuss the influence of $K$, $K_f$ and $d_f/r_1$ on dispersion. In all figures of contours shown below, the time-varying contours of the same parameter are placed in the same row while the parameter-varying contours of the same time are placed in the same column.

*Dispersion with interphase transport*

In this subsection, impacts of interphase transport on solute dispersion is discussed without loss of generality. Figure 4 shows two series of contours of which (a) is related to the solute dispersion without interphase transport and (b) is related to the solute dispersion with interphase transport. The two cases share the same parameters that $Pe_1 = 100$ while only the case (b) has a set of parameters, i.e., $Pe_2 = 10000, K = 100, K_f = 10^6$. It should be noted here that the value of $K_f$ indicates extremely large rate of interphase transport which can be regarded that

$c_m = c_s$ is tenable at the interface of two phases all along. The contours show clearly that interphase transport results in the enhancement of solute dispersion and solute retention in microchannel.

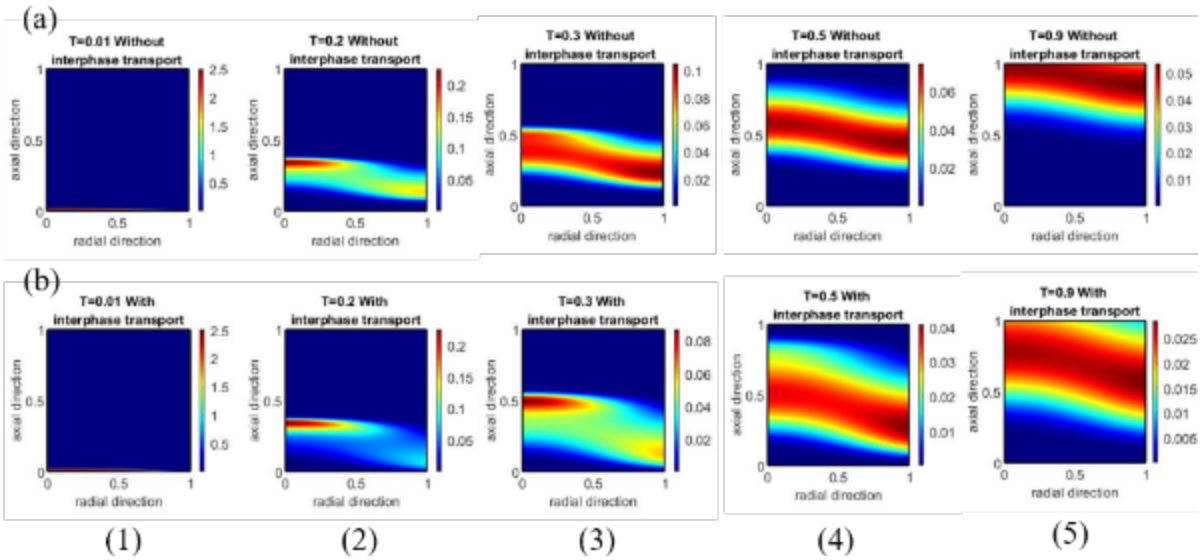



Figure 4. Contrast contours of solute dispersion without and with interphase transport

At the inlet of the microchannel, injected solute is not dispersed yet (a1, b1). With the proceeding of flow, the Poiseuille's law acts on the solute that a parabolic but uneven concentration distribution can be observed in (a2) and (b2) whose contours in main stream are almost the same. However, solute dispersion caused by interphase transport show up that the solute concentration is smaller and the concentration distribution band is wider in (b2) near the interface than that in (a2) near the wall. The concentration gradient near the wall decreases slowly compared to that in the main stream thus a concentration gradient pointing from the wall to the main stream formed as shown in (a3). The diffusion of solute into the stationary phase causes the decrease of solute concentration in the mobile phase near the interface since with a $Pe_1$ equals 100, diffusion in mobile phase is not strong enough to compensate the depletion of solute into the stationary phase. Consequently, the solute is more dispersed and the transport of solute into the stationary phase delays the formation of concentration band in the mobile phase with the contrast of (a3) and (b3). The concentration band is more even and narrower in (a4) than that in (b4). Once the parabolic solute concentration band is formed, longitudinal concentration gradient becomes significant thus the dispersion evolves broader in axial direction till the outlet of the microchannel as shown in (a5) and (b5). However, the concentration band in (b5) moves much slower and is also wider than the band in (a5). In sum, the existence of interphase transport causes the solute moves with a smaller velocity and disperses much stronger in the microchannel.

*Impacts of $Pe_1$ and $Pe_2$*

The radial motion of solute consists of three parts, i.e., diffusion in the mobile and the stationary phases and the interphase transport. The direction of solute radial motion is controlled by the difference of solute concentrations at the interface between the mobile phase and the stationary phase, namely, the value of $c_m$-$c_s$/K. Each of the three parts has a resistance to solute diffusion[25]. Based on the mass conservation law, the solute diffusion from one phase to the other is analogous to the electric current through three tandem different resistances. We set $K = 100, K_f = 10^6$ (infinitely large interphase transport rate) to investigate the impacts of $Pe_1$ and $Pe_2$ on the solute dispersion.

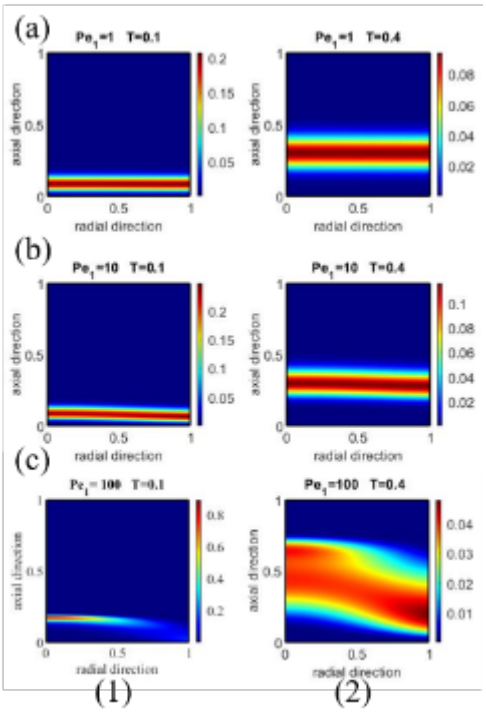



Figure 5. Concentration distribution contours of solute with different $Pe_1$

Firstly, figure 5 shows the concentration contours of different $Pe_1$ with $Pe_2$=$10^5$. In row (a1) and (b1), the concentration band is of plug shape. In (c1) $Pe_1$ equals 100, the concentration band is of parabolic shape. This is the evidence of solute dispersion by a stronger convection. Since the impact of diffusivity in the mobile phase will appear later with increase of $Pe_1$, the concentration band is narrower and more bent in column (1). Later with the proceeding of flow the concentration band reforms dramatically wide in the case of $Pe_1$=100. The concentration bands in (a2) and (b2) keep their shape and moves with weaker dispersion. Consequently, a much small or much big $Pe_1$ will both cause a stronger solute dispersion while a much big one is worse for the trace detection. It takes longer with larger $Pe_1$ for diffusion in the mobile phase to compensate the depletion of solute due to interphase transport into the stationary phase and thus the reformation of concentration band is slower and the reformed band is dramatically

wide. This indicates an optimal concentration $Pe_1$ to be chosen which is consistent with the result of H.E.T.P.[17] to choose an optimal flow rate.

Secondly, Figure 6 gives a comparation of the solute dispersion of different $Pe_2$ and constant $Pe_1$=100. It can be seen from (1) and (2) that with increasing of $Pe_2$ from $10^5$ to $10^6$, the concentration band has little difference. However, with increasing of $Pe_2$ from $10^6$ to $10^7$, the concentration band is wider and one can observe a relatively high concentration in mainstream (3) rather than near the interface (2). The big $Pe_2$ indicates a relatively slow diffusion in the stationary phase so it takes longer for solute to change the diffusion direction from into the stationary phase to out of the stationary phase. Additionally, in (3) the concentration distribution is wider at the upstream of the most concentrated part than that at the downstream. This is because that the point satisfying $c_m=c_s/K$ moves upstream as a result of the big $Pe_2$. In sum, the change of $Pe_2$ below $10^6$ has neglectable influence on solute dispersion while increasing of $Pe_2$ from $10^6$ to $10^7$ causes a stronger solute dispersion and an asymmetric concentration distribution. This can be verified by the elution curve shown in figure 7.

Figure 8 shows a series of solute concentration contours ordered chronologically with $Pe_1$=100 and $Pe_2=10^7$. The contours in the enlargement part of figures are depicted with the replacement of $c_s$ by $c_s/K$ for sake of observation. As aforementioned, $Pe_2$ is the ratio of the convection in the mobile phase over the solute diffusivity in the stationary phase. Due to a big $Pe_1$, when T=0.1 the solute flows in the mobile phase with a parabolic but uneven distribution as shown in the contour (1) and the concentration of the solute left in the stationary phase is satisfied with the relationship $c_m<c_s/K$. In other words, this part of solute is lagged by the stationary phase. When the concentration band forms as shown in (2), the solute diffuse into the stationary phase at the downstream where $c_m>c_s/K$ and out of the stationary phase at the upstream where $c_m<c_s/K$. This indicates that there would be a point satisfying the condition $c_m=c_s/K$. With the proceeding of flow, the concentration band broadens and the difference of $c_m$ and $c_s/K$ vanishes gradually as shown in (3) and (4). By far, the process of solute dispersion with interphase transport has been clearly illustrated.

*Impacts of K, $K_f$ and $d_f/r_1$*

This subsection will discuss the impacts of $K$, $K_f$ and $d_f/r_1$ with $Pe_2=10^5$. The partition coefficient, $K$, indicates the limit of interphase transport and the dimensionless kinetic mass transfer rate, $K_f$, represents the ratio of interphase transport rate over the diffusion rate in terms of $r_1/t_0$. Both of the two parameters characterize the interphase transport. The ratio of $d_f$ over $r_1$ represents the relative thickness of the stationary phase. With a constant diffusion rate in the stationary phase, it takes longer time to reach an even radial distribution in the stationary phase under the condition of a larger $d_f/r_1$.

Firstly, figure 9 shows the contours of different $K$ with $Pe_1$=1 and $K_f=10^6$. With $K$ increasing from 100 in row (a) to 400 in row (d), the concentration band moves slower gradually. The

solute velocity is proportional to the fraction of solute in the mobile phase[26]. As a result, the solute with a larger $K$ will be retained in the microchannel longer. Solute diffuses into the stationary phase more with the $K$ increasing.

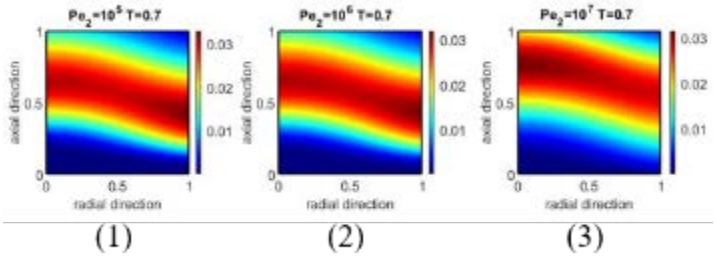



Figure 6. Concentration distribution contours of solute with different $Pe_2$

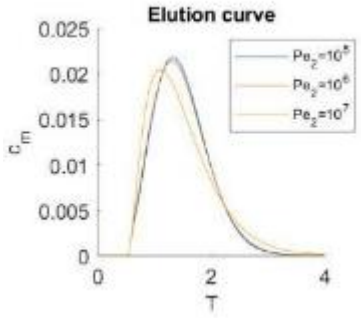



Figure 7. Elution curve of solute with different $Pe_2$

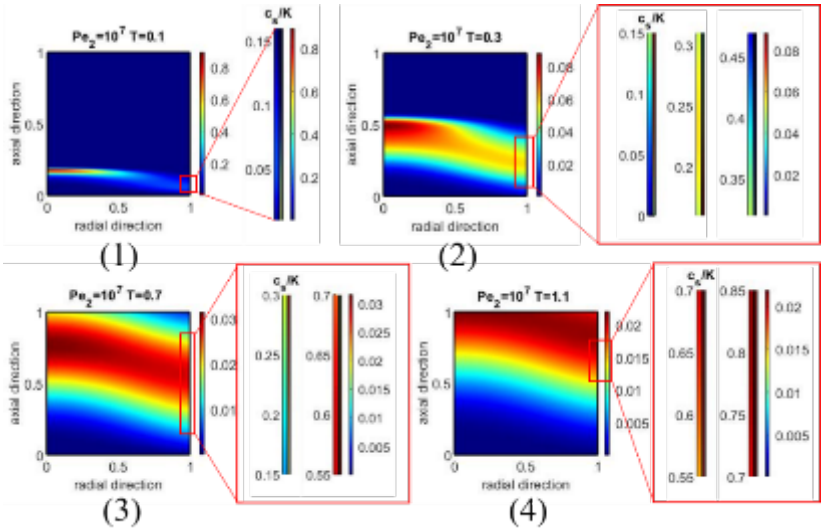



Figure 8. Evolution concentration contours of solute dispersion with $Pe_2$=$10^7$

Consequently, the solute velocity in the mobile phase decreases thus the solute band in the stationary phases get wider. The wider band in stationary phase causes a wider solute concentration band in the mobile phase. In sum, solute of stronger affinity to the station phase, namely, larger partition coefficient $K$, will be dispersed wider and stay longer in the

microchannel. This can be verified by the elution curve in Figure 10(1). Partition coefficient is the key factor that makes solutes separation in microchannel feasible.

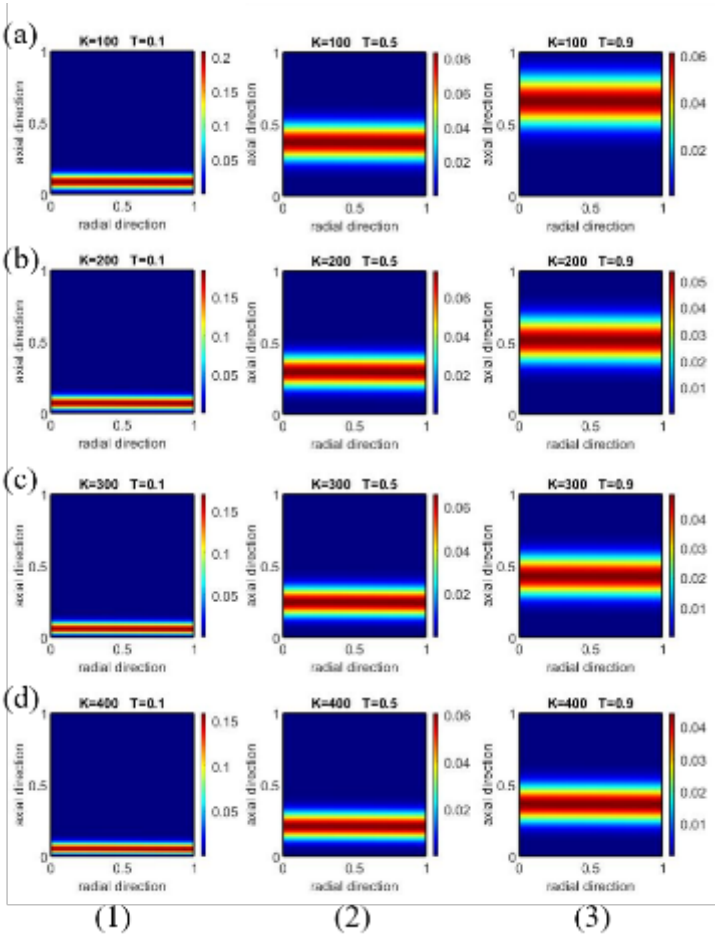



Figure 9. Concentration distribution contours of solute with different *K*

Secondly, figure 11 shows the concentration contours of solute of different $K_f$ with $Pe_1$=1 and $K$=100. In this study $K_f = k_f t_0/r_1$, we assume $t_0$ and $r_1$ constant to specifically investigate the impact of $k_f$ on the solute dispersion. From the enlargement part of contours in column (1), it can be observed that with increase of $K_f$, the difference between $c_m$ and $c_s/K$ goes smaller. In column (2) with $K_f$ increasing from 1.6 to 160, the solute concentration band becomes narrower but the band widths of (a2) and (b2) differs to a larger extent than those of (b2) and (c2). However, the solute concentration band in (c2) and (d2) have few differences in width and value. In addition, (a2) tells that a small $K_f$ causes the asymmetry of solute concentration band with respect to the most concentrated point. However, the change of $K_f$ does not cause the change of the retention time of solute in microchannel. This can be verified by the elution curve in figure 10 (2). The combinative function of $K$ and $K_f$ results in the retention and asymmetric solute dispersion in the microchannel. As aforementioned, there would be an ideal point satisfying $c_m=c_s/K$ at the interface. Solute diffuses into the stationary phase at the downstream ($c_m>c_s/K$) of the ideal point while the solute diffuses out of the stationary phase at the upstream of the ideal point. The position of the ideal point is the most concentrated point of solute band

under the ideal condition, namely, $K_f$ is infinitely large. Under the real condition that $K_f$ is finite, the ideal point moves to the upstream of the most concentrated point.

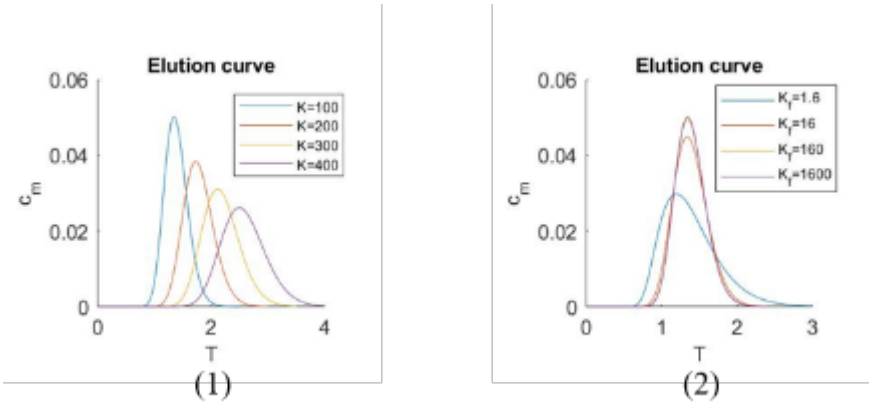



Figure 10. Elution curves of different $K$ in (1) and of different $K_f$ in (2)

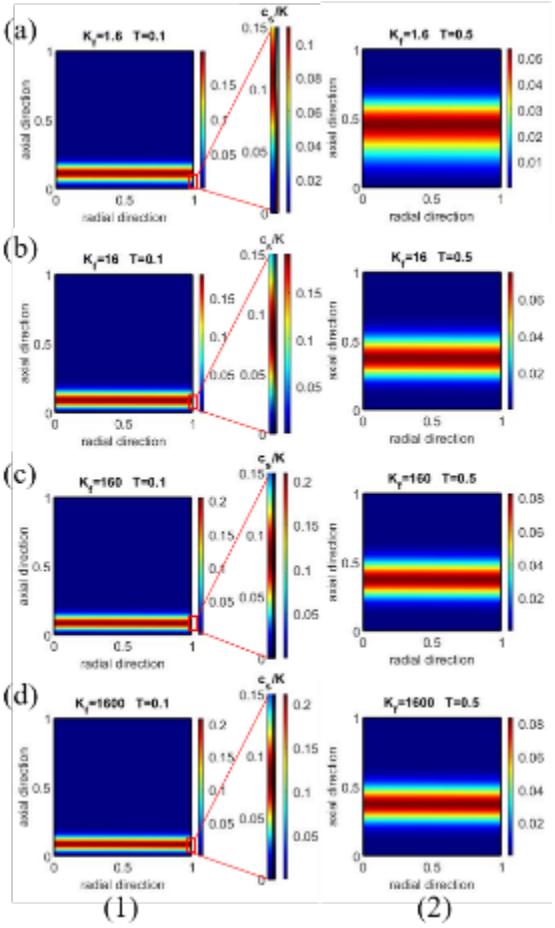



Figure 11. Concentration distribution contours of solute with different $K_f$

In the other word, the condition, $c_m=c_s/K$, at the interface is delayed thus the diffusion of solute out of the stationary phase is delayed. Consequently, the solute concentration band is tailed at the upstream and the elution curve is left-leaning as shown in figure 10 (2). In sum, for a smaller $K_f$ which means a high mass transfer resistance between the mobile and stationary phases, the solute is more dispersed in the mobile phase and also the corresponding concentration band

becomes asymmetric. As $K_f$ is beyond 160, the further increase of $K_f$ affects the solute dispersion marginally.

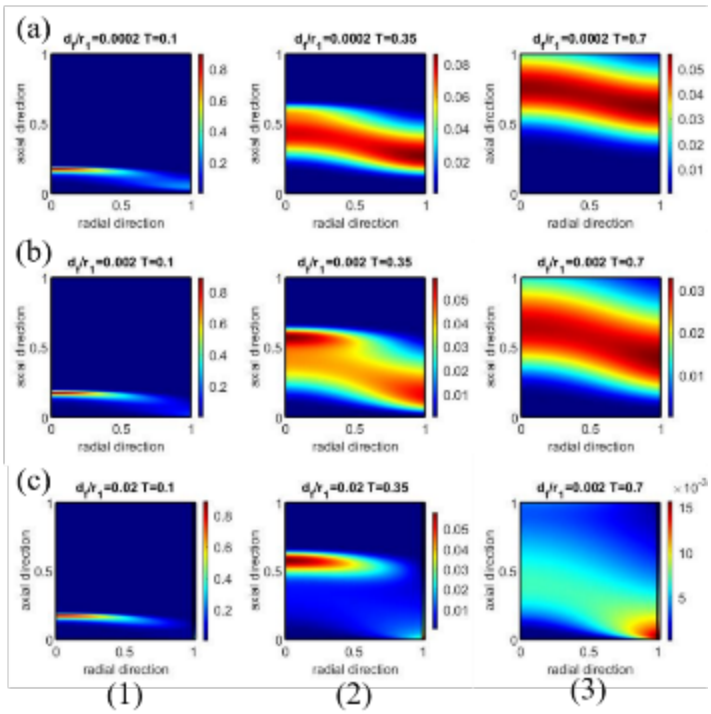



Figure 12. Concentration distribution contours of solute with different $d_f/r_1$

Thirdly, figure 12 shows the concentration contours of solute of different $d_f/r_1$ with $Pe_1$=100 and $K_f$=1600. We assume $r_1$ constant to specifically investigate the influence of $d_f$. In column (1) when the solute dispersion is mainly caused by convection, the solute concentration in mobile phase near the interface increases as $d_f/r_1$ decreases. And when the solute band forms, the solute concentration band goes wider as $d_f/r_1$ increases from 0.0002 of (a2) to 0.002 of (b2). However, the band in (c2) is far from formation due to the extremely strong retention of solute caused by relatively thick stationary phase. Additionally, in (a2) according to the contour the solute has begun to diffuse out of the stationary phase while in (b2) not yet. Contours show that at the same dimensionless time $T$=0.7, when the concentration are distributed evenly, the most concentrated point of solute band in (b3) is further from the outlet of the microchannel than that of solute band in (a3). Consequently, the increase of the $d_f/r_1$ also causes the delay of solute from flow out of the microchannel. This is consistent with the mathematical expression of the first moment[14]. However, in (c3) the solute in main stream has flow out of the microchannel, which indicates that with a relatively thick stationary phase, the retention of solute is so strong that the solute band is not formed in the microchannel. In sum, as $d_f/r_1$ increases the solute capacity of the stationary phase increases and consequently, on one hand the solute is more significantly retained in the microchannel, but on the other hand the solute band may be not able to form in the microchannel.

**Conclusions**

A transient 2-D numerical model based on the convection-diffusion theory is formulated to study the solute dispersion in pressure-driven microchannel flow with the interphase transport. The solute dispersion in the mobile phase is shown to be affected by three processes, i.e., the hydrodynamic convection/molecule diffusion in the mobile phase, the molecule diffusion in stationary phases and the interphase transport between the two phases. In the present analysis, various nondimensional parameters are defined to facilitate the discussion of effects of the three processes on the solution dispersion. Specifically, $Pe_1$ represents the combined effect of hydrodynamic convection and molecule diffusion in the mobile phase, $Pe_2$ represents the effect of molecule diffusion in stationary phase, $K$ and $K_f$ represent the effect of interphase transport.

From the 2D transient concentration contours, we observe that with increase of $Pe_1$ from 1 to 10, the solute concentration band becomes narrower and bent slightly. With the $Pe_1$ further increasing to 100, and the concentration band becomes extremely irregular. As $Pe_2$ is below $10^6$, $Pe_2$ has negligible effect on the solute dispersion. However, with the increase of $Pe_2$ from $10^6$ to $10^7$, the solute shows a significant increase of dispersion in mobile phase. With $K$ increasing from 100 to 400, the solute becomes more dispersed in the mobile phase along the microchannel axis and also shows prolonged retention in the microchannel. For a smaller $K_f$ which means a high mass transfer resistance between the mobile and stationary phases, the solute is more dispersed in the mobile phase and also the corresponding concentration band becomes asymmetric. As $K_f$ is beyond 160, the further increase of $K_f$ affects the solute dispersion marginally.

It is also noticed that a thicker layer of stationary phase ($d_f$) corresponding to a larger solute capacity of the stationary phase, on one hand causes a more significant retention of solute in the microchannel, but on the other hand leads to no formation of the solute band.

## References

[1] Chakraborty, D., Bose, N., Sasmal, S., Dasgupta, S., Maiti, T. K., Chakraborty, S., et al. (2012) Effect of dispersion on the diffusion zone in two-phase laminar flows in microchannels, *Analytica Chimica Acta*, **710**, 88-93.

[2] Datta, S., and Ghosal, S. (2009) Characterizing dispersion in microfluidic channels, *Lab Chip*, **9**, 2537-2550.

[3] Wang, P., and Chen, G. Q. (2016) Solute dispersion in open channel flow with bed absorption, *Journal of Hydrology*, **543**, 208-217.

[4] Taylor, G. (1953) Dispersion of soluble matter in solvent flowing slowly through a tube, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **219**, 186-203.

[5] Taylor, G. (1954) The Dispersion of Matter in Turbulent Flow through a Pipe, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **223**, 446-468.

[6] Taylor Geoffrey, I. (1954) Conditions under which dispersion of a solute in a stream of solvent can be used to measure molecular diffusion, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **225**, 473-477.

[7] Aris, R. (1956) On the dispersion of a solute in a fluid flowing through a tube, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **235**, 67-77.

[8] Aris, R., and Taylor Geoffrey, I. (1959) On the dispersion of a solute by diffusion, convection and exchange between phases, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **252**, 538-550.

[9] Kučera, E. (1965) Contribution to the theory of chromatography: Linear non-equilibrium elution chromatography, *Journal of Chromatography A*, **19**, 237-248.

[10] Grushka, E. (1972) Chromatographic peak shapes. Their origin and dependence on the experimental parameters, *The Journal of Physical Chemistry*, **76**, 2586-2593.

[11] Jonsson, J. A. (1981). *The median of the chromatographic peak as the best measure of retention time* (Vol. 14).

[12] Jaulmes, A., Vidal-Madjar, C., Gaspar, M., and Guiochon, G. (1984) Study of peak profiles in nonlinear gas chromatography. 2. Determination of the curvature of isotherms at zero surface coverage on graphitized carbon black, *The Journal of Physical Chemistry*, **88**, 5385-5391.

[13] Tihminlioglu, F., and Danner, R. P. (1999) Application of inverse gas chromatography to the measurement of diffusion and phase equilibria in polyacrylate–solvent systems, *Journal of Chromatography A*, **845**, 93-101.

[14] Pawlisch, C. A., Macris, A., and Laurence, R. L. (1987) Solute diffusion in polymers. 1. The use of capillary column inverse gas chromatography, *Macromolecules*, **20**, 1564-1578.

[15] Ronco, N. R., Menestrina, F., Romero, L. M., and Castells, C. B. (2019) Determination of gas–liquid partition coefficients of several organic solutes in trihexyl (tetradecyl) phosphonium dicyanamide using capillary gas chromatography columns, *Journal of Chromatography A*, **1584**, 179-186.

[16] Navarro-Tovar, G., Moreira, J., Valades-Pelayo, P., and Lasa, H. (2014). *Diffusion and Equilibrium Adsorption Coefficients of Aromatic Hydrocarbon Species in Capillary Columns* (Vol. 12).

[17] Giddings, J. C. (1961) Plate Height Contributions in Gas Chromatography, *Analytical Chemistry*, **33**, 962-963.

[18] Giddings, J. C. (1962) Plate Height Theory of Programmed Temperature Gas Chromatography, *Analytical Chemistry*, **34**, 722-725.

[19] Giddings, J. C. (1962) Liquid Distribution on Gas Chromatographic Support. Relationship to Plate Height, *Analytical Chemistry*, **34**, 458-465.

[20] Frey, G. L., and Grushka, E. (1996) Numerical Solution of the Complete Mass Balance Equation in Chromatography, *Analytical Chemistry*, **68**, 2147-2154.

[21] van Deemter, J. J., Zuiderweg, F. J., and Klinkenberg, A. (1995) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography, *Chemical Engineering Science*, **50**, 3869-3882.

[22] Knox, J. H. (2002) Band dispersion in chromatography—a universal expression for the contribution from the mobile zone, *Journal of Chromatography A*, **960**, 7-18.

[23] Gritti, F., and Guiochon, G. (2012) Mass transfer kinetics, band broadening and column efficiency, *Journal of Chromatography A*, **1221**, 2-40.

[24] Beauchamp, M. D., and Schure, M. R. (2019) Simulation and theory of open-tube dispersion in short and long capillaries with slip boundaries and retention, *Journal of Chromatography A*, **1588**, 85-98.

[25] Khan, M. A. (1960) Non-Equilibrium Theory of Gas-Liquid Chromatography, *Nature*, **186**, 800-801.

[26] Giddings, J. C. (1965). *Dynamics of Chromatography*.

# Resonant frequencies of a radial field piezoelectric diaphragm

***Xingxu Zhang[1,2], Xiaobiao Shan[1], †Tao Xie[1] and †Jianmin Miao[2].**

[1]State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China;
[2]School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798, Singapore.
*Presenting author: 16B908021@stu.hit.edu.cn

†Corresponding author: xietao@hit.edu.cn; mjmmiao@ntu.edu.sg.

## Abstract

This paper presents a circular piezoelectric diaphragm which has a radial polarization distribution by using inter-circulating electrodes. An equation for calculating the resonant frequency of the radial field piezoelectric diaphragm is obtained based on the thin plate elastic theory combined with the Rayleigh-Ritz method. The finite element analysis (FEA) was used to predict the resonant frequency of the radial field piezoelectric diaphragm. To verify the theoretical analysis, the impedance characteristic of the radial field piezoelectric diaphragm was measured using an experimental method. The results obtained from theoretical analysis were in good agreement with those from the FEA and experimental results. The effect of geometrical changes to the first resonance frequency of the diaphragm is also described. The calculated results were found to be in good agreement with the FEA results. The results indicate that the resonant frequency of the radial field piezoelectric diaphragm decreases from 16.43 kHz to 11.92 kHz when the diameter increases from 9.2 mm to 10.8 mm.

**Keywords:** Radial field diaphragm, Piezoelectric transducer, Resonant frequency, Rayleigh-Ritz method.

## Introduction

Circular piezoelectric diaphragms are commonly employed for the purpose of sensing and actuating, such as electroacoustic devices [1], micromotors [2,3], microfluidic pumps [4] and so on. They are also widely used as energy harvesters [5,6]. Cui et al. [7] proposed a multi-layer piezoelectric structure working as an actuator for valveless pumps. They established an analytical model and verified it by finite element analysis. Kim et al. [8] analyzed a clamped circular piezoelectric plate with different electrode patterns to enhance the power generation of energy harvesters. Papila et al. [9] addressed the design of a clamped circular piezoelectric composite diaphragm with oppositely polarized piezoceramic patches. Smyth et al. [10] modeled a micromachined piezoelectric ultrasonic transducer with circular and ring electrodes. Piezoelectric elements in these transducers work in $d_{31}$ mode. However, because the $d_{33}$ parameter is usually two times that of the $d_{31}$ parameter, the PZT elements which work in $d_{33}$ mode are more efficient when the transducers have the similar geometrical structure under the same pressure [11].

For the circular diaphragm structure, it is generally difficult to do in-plane poling due to the much bigger scale in the radial direction. Bryant et al. [12,13] designed equivalent $d_{33}$ mode piezoelectric diaphragms poled by interdigitated ring electrodes or inter-circulating electrodes. Hong et al. [14-16] designed micromachined radial field piezoelectric diaphragms with interdigitated ring electrodes. Wang et al. [17] proposed radial field piezoelectric diaphragms with interdigitated ring electrodes using bulk PZT materials. Shen et al. [18,19] designed piezoelectric diaphragms with spiral electrodes to perform in-plane poling. For piezoelectric transducers, the fundamental resonance is generally the most important one [19]. However, most of the studies utilized finite element analysis and experiments to characterize the radial field piezoelectric diaphragm. Theoretical analysis of calculating the fundamental resonant frequency of the radial field piezoelectric diaphragms has been rarely reported.

Hence, in this paper, an equation for calculating the fundamental resonant frequency of the radial field piezoelectric diaphragms was established. Finite element analysis was used to analyze the vibration mode. Experiments were conducted to verify the theoretical analysis. Besides, the influence of geometrical parameters of the diaphragms was also discussed.

## Theoretical Analysis

Figure 1 shows a schematic diagram of a piezoelectric diaphragm. Figure 2 shows a schematic diagram of the electrode pattern on the surfaces of PZT with polarization orientation. A circular piezoelectric wafer was bonded with a polymethyl methacrylate (PMMA) substrate which had a through hole. There are inter-circulating electrodes on the surfaces of the piezoelectric wafer. When a poling electric field is applied to the electrodes, the result is a radial polarization distribution. Hence, a radial field piezoelectric diaphragm is obtained. The diaphragm has a diameter of $d_1$ and thickness of $h$. Its diameter is much larger than its thickness. Because the thickness of the electrodes is as small as 400 nm, the mass and stiffness of the electrodes have little influence on the system dynamic properties and they will be ignored during the theoretical analysis.

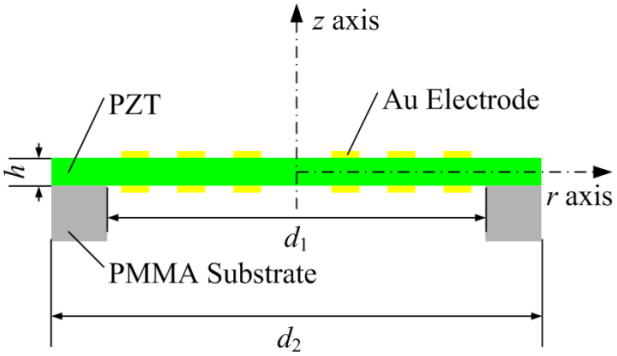



**Figure 1. Schematic diagram of a piezoelectric diaphragm**

Here, a cylindrical coordinate system was adopted in order to conveniently conduct theoretical analysis due to the circular diaphragm structure. When an electric field is applied to the electrodes, the induced strain in the piezoelectric diaphragm will lead to a flexural deformation. The deformation is much smaller than its thickness. According to the Kirchhoff-Love thin plate elastic theory, the strains and stresses in some directions in the

piezoelectric element are given as

$$S_z = S_{zr} = S_{\theta z} = S_{r\theta} = 0 \tag{1}$$

$$T_z = 0 \tag{2}$$

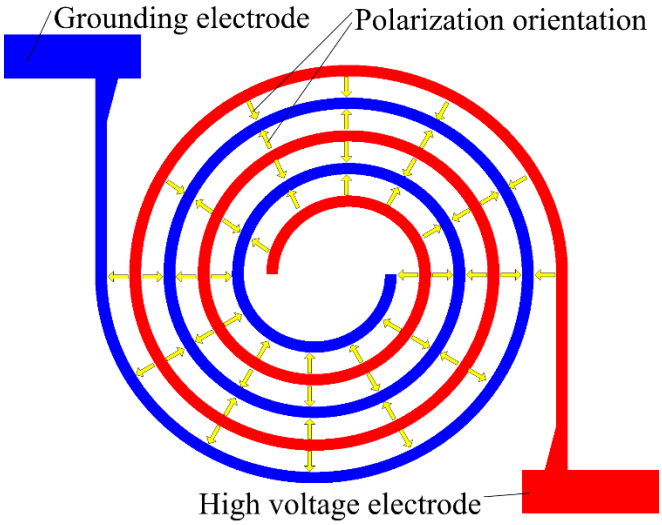



**Figure 2. Electrode pattern on the PZT with polarization orientation**

Based on the d-type piezoelectric equations, the constitutive equations are obtained as [20,21]

$$S_\theta = s_{11}^E T_\theta + s_{13}^E T_r + d_{31} D_r \tag{3a}$$

$$S_r = s_{13}^E T_\theta + s_{33}^E T_r + d_{33} D_r \tag{3b}$$

$$D_r = d_{31} T_\theta + d_{33} T_r + \varepsilon_{33}^T E_r \tag{3c}$$

where $S_\theta$ and $S_r$ are the circumferential and radial strains of the piezoelectric element, respectively; $T_\theta$ and $T_r$ are the circumferential and radial stresses of the piezoelectric element, respectively; $D_r$ and $E_r$ are the electric displacement and electric field, respectively; $s_{11}^E$, $s_{13}^E$, and $s_{33}^E$ are the compliance coefficients at constant electric field; $d_{31}$ and $d_{33}$ are the piezoelectric coefficients; $\varepsilon_{33}^T$ is the permittivity of the piezoelectric element at constant stress. According to Eq. (3), the stresses in the piezoelectric element can be described as

$$T_\theta = \lambda_1 S_\theta + \lambda_2 S_r + \lambda_3 D_r \tag{4a}$$

$$T_r = \lambda_2 S_\theta + \lambda_4 S_r + \lambda_5 D_r \tag{4b}$$

$$E_r = \lambda_6 S_\theta + \lambda_7 S_r + \lambda_8 D_r \tag{4c}$$

where $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$, $\lambda_6$, $\lambda_7$, and $\lambda_8$ are a series of constants. They are expressed as follows

$$\lambda_1 = \frac{s_{33}^E}{s_{11}^E s_{33}^E - s_{13}^E s_{13}^E} \tag{5a}$$

$$\lambda_2 = \frac{-s_{13}^E}{s_{11}^E s_{33}^E - s_{13}^E s_{13}^E} \tag{5b}$$

$$\lambda_3 = \frac{s_{13}^E d_{33} - s_{33}^E d_{31}}{s_{11}^E s_{33}^E - s_{13}^E s_{13}^E} \tag{5c}$$

$$\lambda_4 = \frac{s_{11}^E}{s_{11}^E s_{33}^E - s_{13}^E s_{13}^E} \tag{5d}$$

$$\lambda_5 = \frac{s_{13}^E d_{31} - s_{11}^E d_{33}}{s_{11}^E s_{33}^E - s_{13}^E s_{13}^E} \tag{5e}$$

$$\lambda_6 = \frac{-d_{31}\lambda_1 - d_{33}\lambda_2}{\varepsilon_{33}^T} \tag{5f}$$

$$\lambda_7 = \frac{-d_{31}\lambda_2 - d_{33}\lambda_4}{\varepsilon_{33}^T} \tag{5g}$$

$$\lambda_8 = \frac{1-d_{31}\lambda_3 - d_{33}\lambda_5}{\varepsilon_{33}^T} \tag{5h}$$

There is no free charge on the surfaces of the piezoelectric diaphragm, hence the charge equation of electrostatics is [22]

$$\frac{1}{r}\frac{\partial}{\partial r}(rD_r) = 0 \tag{6}$$

Based on the thermodynamic equilibrium principle, $D_\mathrm{p}$, which is the strain energy density of an infinitesimal volume element in piezoelectric material, is written as

$$D_\mathrm{p} = \frac{1}{2}T_r S_r + \frac{1}{2}T_\theta S_\theta \tag{7}$$

By substituting Eq. (4) into Eq. (7), the $D_\mathrm{p}$ is obtained as

$$D_\mathrm{p} = \frac{1}{2}\lambda_1 S_\theta^2 + \frac{1}{2}\lambda_4 S_r^2 + \lambda_2 S_\theta S_r + \frac{1}{2}(\lambda_3 S_\theta + \lambda_5 S_r)D_r \tag{8}$$

The strains in the diaphragm has the relationship with the curvature of the mid-plane as

$$S_r = z\mu_r \tag{9a}$$

$$S_\theta = z\mu_\theta \tag{9b}$$

where $z$ is the distance from the natural surface; $\mu_\mathrm{r}$ and $\mu_\theta$ are the curvatures of mid-plane, and they are defined as [23]

$$\mu_r = -\frac{\partial^2 w}{\partial r^2} \tag{10a}$$

$$\mu_\theta = -\frac{1}{r}\frac{\partial w}{\partial r} \tag{10b}$$

where $w$ is the displacement in the $z$-axis direction. The diaphragm is in the status of simple harmonic vibration

$$w = w_z \sin(\omega t) \tag{11}$$

where $w_z$ is the vibration mode displacement function, $t$ is time, $\omega$ is angular frequency.
By substitution of Eq. (9) into Eq. (8) and integration in the thickness direction, the unit area of the strain energy in the piezoelectric element is given by

$$u_p = \int_z D_p \mathrm{d}z = \frac{h^3}{24}\left(\frac{1}{2}\lambda_1\mu_\theta^2 + \lambda_2\mu_\theta\mu_r + \frac{1}{2}\lambda_4\mu_r^2\right) \tag{12}$$

The total strain energy of the piezoelectric diaphragm can be obtained by an area integration

$$U = \iint_s u_p \mathrm{d}s = \frac{\pi h^3}{12}\int_0^{d_1/2}\left(\frac{1}{2}\lambda_1\mu_\theta^2 + \lambda_2\mu_\theta\mu_r + \frac{1}{2}\lambda_4\mu_r^2\right)r\mathrm{d}r \tag{13}$$

The kinetic energy of the radial field piezoelectric diaphragm can be denoted as

$$T = \frac{1}{2}\iint_s \left(\frac{\partial w}{\partial t}\right)^2 \mathrm{d}s = \rho\pi h\int_0^{d_1/2}\left(\frac{\partial w}{\partial t}\right)^2 r\mathrm{d}r \tag{14}$$

where $\rho$ is the density of the PZT.
The electric energy of the radial field piezoelectric diaphragm can be calculated as

$$U^{\mathrm{E}} = \iiint_V E_r D_r \mathrm{d}V = \int_0^{2\pi}\mathrm{d}\theta\int_0^{d_1/2}\mathrm{d}r\int_{-h/2}^{h/2} E_r D_r r\mathrm{d}z \tag{15}$$

Based on the Rayleigh-Ritz method and the symmetrical structure, the functional analysis of the radial field piezoelectric diaphragm can be given as [23]

$$L = U_{\max} - T_{\max} - U_{\max}^{\mathrm{E}} \tag{16}$$

where $U_{\max}$ is the maximal strain energy, $T_{\max}$ is the maximal kinetic energy, $U_{\max}^{\mathrm{E}}$ is the maximal electric energy, respectively, of the radial field piezoelectric diaphragm.
By substituting Eqs. (10), (13), (14), (15) into Eq. (16), the functional analysis is obtained as

$$\begin{aligned} L = {} & \frac{\pi h^3}{12}\int_0^{d_1/2}\left[\frac{1}{2}\lambda_1\left(-\frac{1}{r}\frac{\partial w_z}{\partial r}\right)^2 + \lambda_2\frac{\partial w_z^2}{\partial r^2}\frac{1}{r}\frac{\partial w_z}{\partial r} + \frac{1}{2}\lambda_4\left(-\frac{\partial w_z^2}{\partial r^2}\right)^2\right]r\mathrm{d}r \\ & -\rho\pi h\omega^2\int_0^{d_1/2} w_z^2 r\mathrm{d}r - 2\pi h\int_0^{d_1/2}\lambda_8 D_r^2 r\mathrm{d}r \end{aligned} \tag{17}$$

When the diaphragm has a clamped boundary, the boundary condition can be written as

$$w_z = 0 \tag{18a}$$

$$\frac{\mathrm{d}w_z}{\mathrm{d}r} = 0 \tag{18b}$$

where $r = d_1/2$.
The approximate vibration mode displacement function can be written as [8]

$$w_z = A(\frac{1}{4}d_1^2 - r^2)^2 \tag{19}$$

By substituting Eq. (19) into Eq. (17), then solve the equation

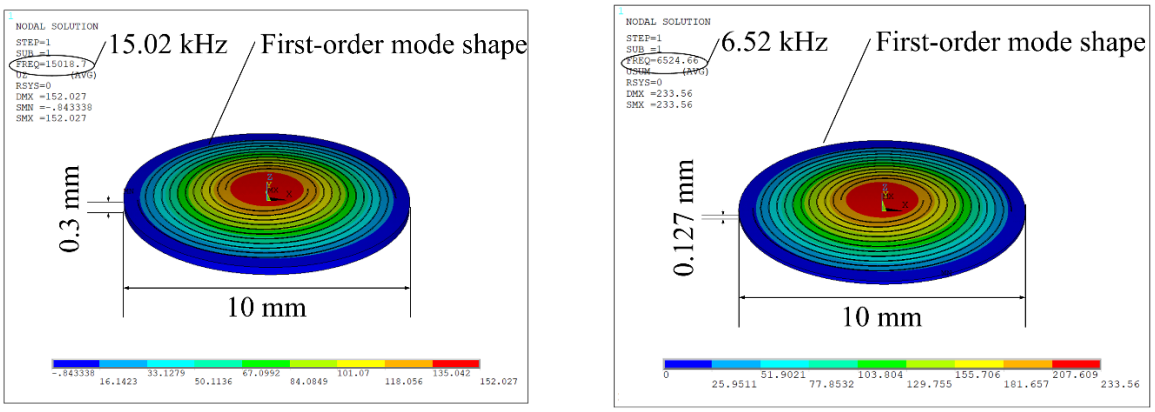
$$\frac{\mathrm{d}L}{\mathrm{d}A} = 0 \tag{20}$$

Finally, the fundamental resonant angular frequency $\omega$ will be obtained.

**Finite Element Analysis**

Finite element method was also employed here to analyze the influence of geometrical parameters on the fundamental resonant frequency. The element SOLID 226 was selected to conduct the modal analysis. A circular diaphragm model with inter-circulating electrodes was established and the edge of the circular diaphragm was set as clamped boundary. The radial polarization distribution was released by rotating the local element coordinate systems after comparing the poling electric field strength with the coercive field strength. The PZT5A parameters are given below as: $s_{11}^{E}$=15.4×10$^{-12}$ m$^2$/N, $s_{12}^{E}$=-4.8×10$^{-12}$ m$^2$/N, $s_{13}^{E}$=-8.4×10$^{-12}$ m$^2$/N, $s_{33}^{E}$=15.4×10$^{-12}$ m$^2$/N, $s_{44}^{E}$=47.8×10$^{-12}$ m$^2$/N, $s_{66}^{E}$=40.4×10$^{-12}$ m$^2$/N, $d_{31}$=-191×10$^{-12}$ C/N, $d_{33}$=430×10$^{-12}$ C/N, $d_{15}$=590×10$^{-12}$ C/N, $\varepsilon_{11}$=1780$\varepsilon_0$, $\varepsilon_{33}$=1950$\varepsilon_0$, $\varepsilon_0$=8.854×10$^{-12}$ F/m, where $s_{ij}^{E}$, $d_{ij}$ and $\varepsilon_{ij}$ are the elastic compliance, piezoelectric and permittivity constants, respectively, where $i$, $j$ ($i$, $j$ = 1, 2, 3, 4, 5, 6) denote tensor notation. Diaphragm models with different diameters and thicknesses were built and their fundamental resonant frequencies were calculated.

Figure 3 shows some modal analysis simulation results. Figure 3a shows the fundamental resonant vibration mode of the piezoelectric diaphragm with a diameter of 10 mm and thickness of 0.3 mm. Figure 3b shows the fundamental resonant vibration mode of the piezoelectric diaphragm with a diameter of 10 mm and thickness of 0.127 mm. From Figure 3, one can know that the fundamental resonant frequency will increase with the increase of thickness. More detailed results will be discussed in next section.




**(a)** **(b)**

**Figure 3. Modal analysis simulation results: (a) Fundamental resonant frequency of the piezoelectric diaphragm with diameter of 10 mm and thickness of 0.3 mm: $f_1$ = 15.02 kHz, (b) Fundamental resonant frequency of the piezoelectric diaphragm with diameter of 10 mm and thickness of 0.127 mm: $f_1$ = 6.52 kHz.**

## Experiments and Results

Inter-circulating electrodes on the surfaces of PZT were fabricated using micro-fabrication technology in a clean room. Figure 4 shows a schematic diagram of the fabrication procedure. Photolithography, magnetron sputtering and lift-off process were conducted. More details were described in [24].

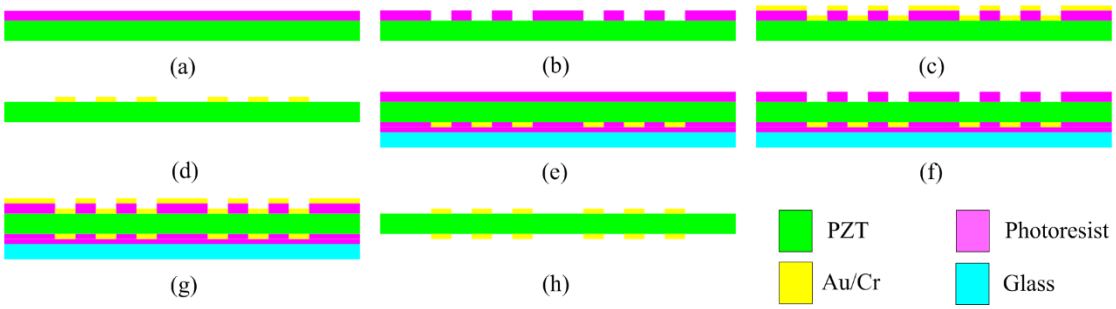



**Figure 4. Schematics of the device fabrication process. (a) Photoresist spin coating; (b) Photolithography; (c) Magnetron sputtering; (d) Lift off process; (e) Backside photoresist spin coating; (f) Backside alignment photolithography; (g) Backside magnetron sputtering; (h) Lift off process**

Then the PZT wafer was bonded with a PMMA substrate. Wire connection and poling process were conducted. Hence a radial field piezoelectric diaphragm was obtained. The diaphragm has a diameter of 10 mm and thickness of 0.3 mm. An impedance analyzer (Agilent 4294A, Agilent Technologies Inc., Santa Clara, CA, USA) was utilized to measure the resonant frequency of the diaphragm.

Figure 5 shows a photograph of the impedance and phase spectrum testing. Figure 6 shows the impedance and phase spectrum of the diaphragm. The frequencies of the theoretical analysis, finite element analysis and experiment results are 13.91 kHz, 15.02 kHz and 14.36 kHz, respectively. The results of theoretical analysis were in good agreement with those of

finite element analysis and experiment results.

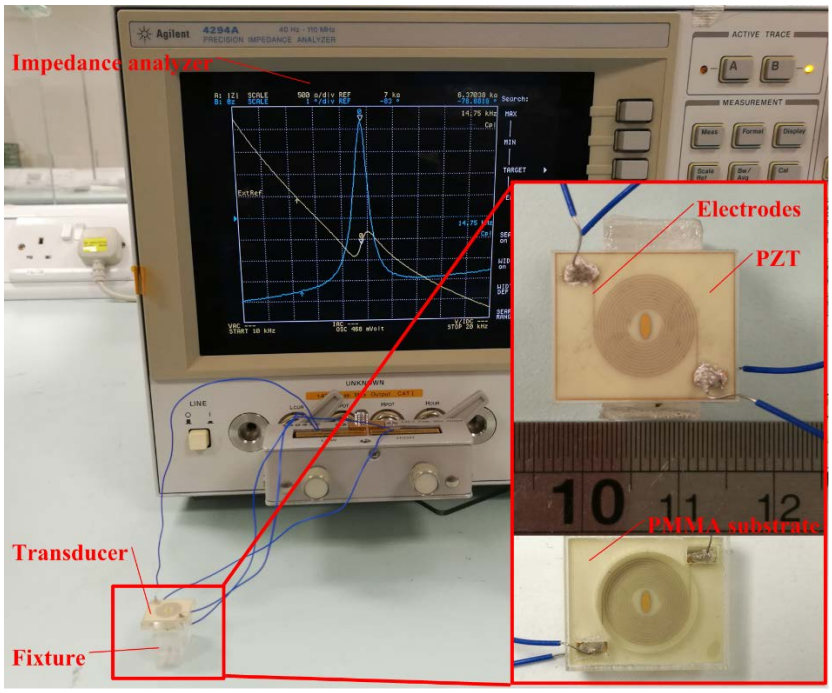



**Figure 5. Impedance and phase spectrum testing**

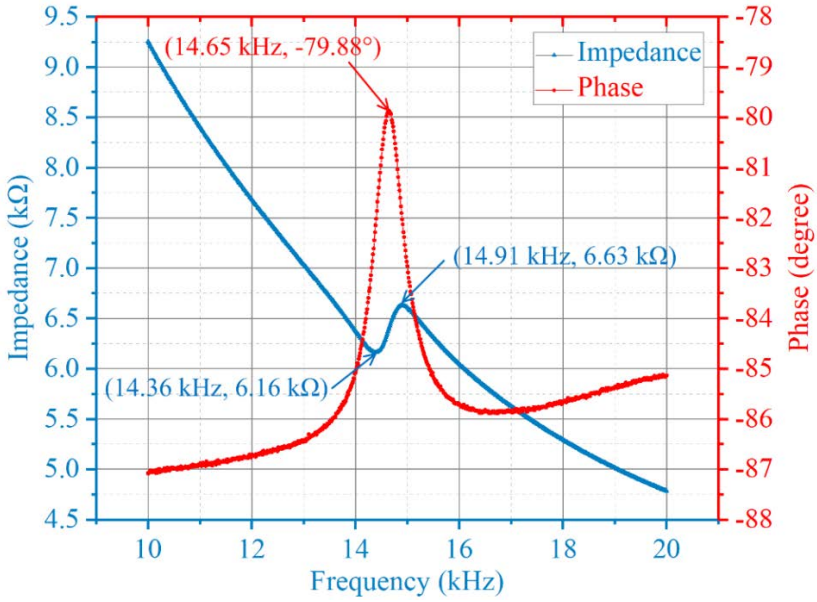



**Figure 6. Impedance and phase spectrum of the diaphragm**

As mentioned in previous sections, the diameter of the piezoelectric diaphragm has a profound effect on the fundamental resonant frequency. Figure 7 shows the results obtained from the theoretical analysis and finite element analysis with different diameters. From Figure 7, one can know that the theoretical results are in good agreement with the finite element analysis results. The fundamental resonant frequency will decrease from 16.43 kHz to 11.92 kHz when the diameter increases from 9.2 mm to 10.8 mm.

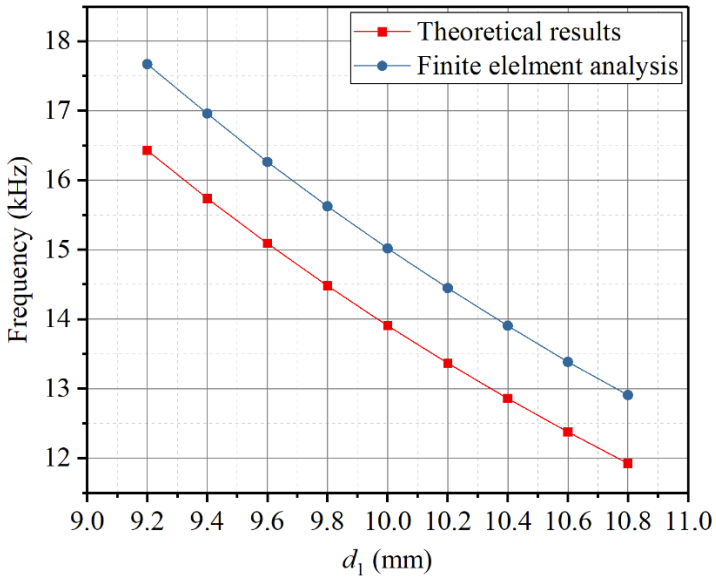



**Figure 7. Frequency-diameter curves of the piezoelectric diaphragm**

## Conclusion

Based on the Kirchholf-Love thin plate elastic theory and piezoelectric equations combined with the Rayleigh-Ritz method, an equation for calculating the fundamental resonant frequency of a radial field piezoelectric diaphragm was obtained. Finite element analysis (FEA) was conducted to calculate the resonant vibration mode. Prototype was fabricated by microfabrication technology in a clean room and its impedance spectrum was examined by an impedance analyzer. The theoretical analysis results were in good agreement with the FEA and experimental results. The effect of geometrical parameters on the fundamental resonant frequency was also analyzed. The results showed that the fundamental resonant frequency of the radial field piezoelectric diaphragm decreases from 16.43 kHz to 11.92 kHz when the diameter increases from 9.2 mm to 10.8 mm.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 51677043) and the China Scholarship Council (Grant No. 201706120130).

### References

1. Prasad, S.A.; Gallas, Q.; Horowitz, S.B.; Homeijer, B.D.; Sankar, B.V.; Cattafesta, L.N.; Sheplak, M. Analytical electroacoustic model of a piezoelectric composite circular plate. *AIAA journal* **2006**, *44*, 2311-2318.
2. Deshpande, M.; Saggere, L. An analytical model and working equations for static deflections of a circular multi-layered diaphragm-type piezoelectric actuator. *Sensors and Actuators A: Physical* **2007**, *136*, 673-689.
3. Wright, R.; Mo, C.; Clark, W.W. Effect of electrode pattern on the performance of unimorph piezoelectric diaphragm actuators. In Proceedings of Smart Structures and Materials 2005: Smart Structures and Integrated Systems; pp. 42-50.
4. He, X.; Xu, W.; Lin, N.; Uzoejinwa, B.; Deng, Z. Dynamics modeling and vibration analysis of a piezoelectric diaphragm applied in valveless micropump. *Journal of Sound and Vibration* **2017**, *405*, 133-143.

5. Yang, Y.; Wang, S.; Stein, P.; Xu, B.-X.; Yang, T. Vibration-based energy harvesting with a clamped piezoelectric circular diaphragm: analysis and identification of optimal structural parameters. *Smart Materials and Structures* **2017**, *26*, 045011.
6. Mo, C.; Radziemski, L.J.; Clark, W.W. Analysis of piezoelectric circular diaphragm energy harvesters for use in a pressure fluctuating system. *Smart Materials and Structures* **2010**, *19*, 025016.
7. Cui, Q.; Liu, C.; Zha, X.F. Modeling and numerical analysis of a circular piezoelectric actuator for valveless micropumps. *Journal of intelligent material systems and structures* **2008**, *19*, 1195-1205.
8. Kim, S.; Clark, W.W.; Wang, Q.-M. Piezoelectric energy harvesting with a clamped circular plate: analysis. *Journal of intelligent material systems and structures* **2005**, *16*, 847-854.
9. Papila, M.; Sheplak, M.; Cattafesta III, L.N. Optimization of clamped circular piezoelectric composite actuators. *Sensors and Actuators A: Physical* **2008**, *147*, 310-323.
10. Smyth, K.; Bathurst, S.; Sammoura, F.; Kim, S.-G. Analytic solution for N-electrode actuated piezoelectric disk with application to piezoelectric micromachined ultrasonic transducers. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **2013**, *60*, 1756-1767.
11. Shen, Z.; Liu, S.; Miao, J.; Woh, L.S.; Wang, Z. Spiral electrode d33 mode piezoelectric diaphragm combined with proof mass as energy harvester. *Journal of Micromechanics and Microengineering* **2015**, *25*, 035004.
12. Bryant, R.G.; Effinger, R.T.; Aranda, I.; Copeland, B.M.; Covington, E.W. Active piezoelectric diaphragms. In Proceedings of Smart Structures and Materials 2002: Active Materials: Behavior and Mechanics; pp. 303-315.
13. Bryant, R.G.; Effinger Iv, R.T.; Aranda Jr, I.; Copeland Jr, B.M.; Covington Iii, E.W.; Hogge, J.M. Radial field piezoelectric diaphragms. *Journal of intelligent material systems and structures* **2004**, *15*, 527-538.
14. Hong, E.; Krishnaswamy, S.; Freidhoff, C.; Trolier-McKinstry, S. Micromachined piezoelectric diaphragms actuated by ring shaped interdigitated transducer electrodes. *Sensors and Actuators A: Physical* **2005**, *119*, 521-527.
15. Hong, E.; Trolier-McKinstry, S.; Smith, R.; Krishnaswamy, S.V.; Freidhoff, C.B. Vibration of micromachined circular piezoelectric diaphragms. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **2006**, *53*, 697-706.
16. Hong, E.; Trolier-McKinstry, S.; Smith, R.L.; Krishnaswamy, S.V.; Freidhoff, C.B. Design of MEMS PZT circular diaphragm actuators to generate large deflections. *Journal of Microelectromechanical systems* **2006**, *15*, 832-839.
17. Wang, Z.; Miao, J.; Tan, C.W.; Xu, T. Fabrication of piezoelectric MEMS devices-from thin film to bulk PZT wafer. *Journal of electroceramics* **2010**, *24*, 25-32.
18. Shen, Z.; Olfatnia, M.; Miao, J.; Wang, Z. Displacement and resonance behaviors of a piezoelectric diaphragm driven by a double-sided spiral electrode. *Smart materials and structures* **2012**, *21*, 055001.
19. Shen, Z.; Lu, J.; Tan, C.W.; Miao, J.; Wang, Z. d33 mode piezoelectric diaphragm based acoustic transducer with high sensitivity. *Sensors and Actuators A: Physical* **2013**, *189*, 93-99.
20. Lin, S.; Fu, Z.; Zhang, X.; Wang, Y.; Hu, J. Radially sandwiched cylindrical piezoelectric transducer. *Smart Materials and Structures* **2012**, *22*, 015005.
21. Salowitz, N.P.; Kim, S.-J.; Kopsaftopoulos, F.; Li, Y.-H.; Chang, F.-K. Design and analysis of radially polarized screen-printed piezoelectric transducers. *Journal of Intelligent Material Systems and Structures* **2017**, *28*, 934-946.
22. Chen, W.; Lu, C.; Yang, J.; Wang, J. A circular cylindrical, radially polarized ceramic shell piezoelectric transformer. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **2009**, *56*, 1238-1245.
23. Yuan, J.-b.; Xie, T.; Shan, X.-b.; Chen, W.-s. Resonant frequencies of a piezoelectric drum transducer. *Journal of Zhejiang University-SCIENCE A* **2009**, *10*, 1313-1319.
24. Zhang, X.; Shan, X.; Shen, Z.; Xie, T.; Miao, J. A New Self-Powered Sensor Using the Radial Field Piezoelectric Diaphragm in d33 Mode for Detecting Underwater Disturbances. *Sensors* **2019**, *19*, 962.

# Topological optimization of microstructures with isogeometric analysis

***Wei Zeng[1], Jie Gao[2], and †Zhen Luo[2]**

[1] School of Information Engineering, East China Jiaotong University, China.

[2] School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Australia.

*Presenting author: zengweixigua@163.com

†Corresponding author: zhen.luo@uts.edu.au

## Abstract

This paper will develop a topology optimization method for computational design of architected microstructures based on isogeometric analysis (IGA). The NURBS (non-uniform rational B-splines) is applied to represent the geometry in both the design and analysis processes, as well as interpolate a material density distribution function (DDF) with the desired smoothness and continuity to represent the material layout in the design domain. An isogeometric topology optimization formulation is then formulated based on the DDF, with the energy-based homogenization method (EBHM) to evaluate the effective properties of the microstructure. Several numerical examples are used to demonstrate the effectiveness of the proposed method for computational design of micro-structured periodic composite structures.

**Keywords:** Topology optimization; Isogeometric analysis; Microstructures

## Introduction

Architected materials with a series of periodically distributed microstructures, a kind of rationally artificial materials, are featured with the superior performance, such as the higher specific stiffness and strength, the better fatigue strength and improved corrosion-resistance and etc. [1, 2]. It is known that the performance of architected materials is mostly dependent on the microstructural information, namely the configuration, rather than the constituent properties. Hence, how to develop a rational design framework for architected materials has accepted enormous attentions in recent years.

Topology optimization has made remarkable progress in creating architected materials with new properties [3], which can be viewed as a numerically iterative procedure to optimize material layout in a given design domain, under the specified objective function and constraint(s) [4], Several topology optimization methods have been developed, like the homogenization method [4], the Solid Isotropic Material with Penalization (SIMP) method [5,6], the Evolutionary Structural Optimization (ESO) method [7] and the level set method (LSM) [8-10]. Since an inverse homogenization method was proposed for the architected materials [11], topology optimization combined with the homogenization method has become more and more popular for the design of architected materials with the specific properties [12, 13] and even more advanced topological designs [14,15].

Although the research on how to obtain architected materials has been extensively studied in recent years, only a limited number of works are devoted to obtaining architected materials with the low density. In this paper, we aim to develop an effective and efficient isogeometric topology optimization (ITO) method for the rational design of the low-density architected materials. Firstly, most of the previous works are studied based on the conventional finite element method (FEM). However, the FEM is also one factor to influence the effectiveness of the topology optimization for the design of architected materials. This is because: (1) The finite element mesh is just an approximation of the original shape of the design domain; (2) The lower-order (C0) continuity of the responses between the neighboring finite elements; (3) The lower efficiency to achieve a finite element mesh with the high quality. Isogeometric analysis (IGA) [16] has attracted much interests, due to its favorable features in numerical analysis, such as the consistency between the computer-aided design (CAD) model and the computer-aided engineering (CAE) model, and the high-order continuity between different elements. Secondly, in the developed ITO method, a sufficiently smooth and continuous DDF is constructed to represent the topological changes during the optimization. Thirdly, the IGA is applied to numerical implement the energy-based homogenization method. Finally, the corresponding isogeometric topology optimization is developed for the design of architected materials with the low-density. Several numerical examples are tested to show the effectiveness and efficiency.

## NURBS-based IGA

### (1) NURBS

An example of a square modelled by NURBS is shown in **Figure 1**. The NURBS basis functions are linearly combined with a series of control points plotted with the red color to construct the geometrical model shown in **Figure 1** (*b*), and the mathematical form of the NURBS surface $\mathbf{S}(\xi,\eta)$ is given as:

$$\mathbf{S}(\xi,\eta)=\sum_{i=1}^{n}\sum_{j=1}^{m}R_{i,j}^{p,q}(\xi,\eta)\mathbf{P}_{i,j} \tag{1}$$

where $n$ and $m$ are the numbers of control points in two parametric directions, and $\xi$ and $\eta$ denote the corresponding parametric directions. $p$ and $q$ are the polynomial orders. The detailed information for the square is listed below **Figure 1**. $\mathbf{P}_{i,j}$ correspond to the $(i,j)_{th}$ control point. It should be noted that control points are not necessarily on the structural design domain. $R$ are the bivariate NURBS basis functions, and which are constructed by the B-spline basis functions, as:

$$R_{i,j}^{p,q}(\xi,\eta)=\frac{N_{i,p}(\xi)M_{j,q}(\eta)\omega_{ij}}{\sum_{\hat{i}=1}^{n}\sum_{\hat{j}=1}^{m}N_{\hat{i},p}(\xi)M_{\hat{j},q}(\eta)\omega_{\hat{i}\hat{j}}} \tag{2}$$

where $\omega_{ij}$ is the positive weight for the $(i,j)_{th}$ control point $\mathbf{P}_{i,j}$. $N_{i,p}$ and $M_{j,q}$ are the univariate B-spline basis functions in two parametric directions, respectively. The B-spline

basis function is defined by the Cox-de-Boor formula [16], and the recursive formula in $\xi$ direction with a non-decreasing knot vector $\Xi = \{\xi_1, \xi_2, \cdots, \xi_{n+p+1}\}$ is defined as:

$$\begin{cases} N_{i,0}(\xi) = \begin{cases} 1 & if\ \xi_i \le \xi_{i+1}, \\ 0 & otherwise \end{cases} & p = 0 \\ N_{i,p}(\xi) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_{i,p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1,p-1}(\xi), & p \ge 1 \end{cases} \tag{3}$$

It is noted that the fractions with the form 0/0 in Eq. (3) are defined as zero. Similarly, the basis functions $M_{j,q}$ in the $\eta$ direction are also defined by Eq. (3) with the knot vector. The NURBS basis functions of the square in two parametric directions are respectively displayed in **Figure 1** (*d*) and (*e*). The bivariate basis functions are also plotted in **Figure 1** (*f*).

We can easily see that the NURBS basis functions are featured with several important properties: (1) **Nonnegativity**: $N_{i,p}(\xi) \ge 0$; (2) **Local support**: the support of each basis function $N_{i,p}$ is contained in the interval $[\xi_i, \xi_{i+p+1}]$; (3) **Partition of unity**: for an arbitrary knot span $[\xi_i, \xi_{i+1}]$, $\forall \xi \in [\xi_i, \xi_{i+1}]$, $\sum_{j=i-p}^{i} N_{j,p}(\xi) = 1$; (4) **Continuity**: The continuity between knot spans is equal to $C^{p-k}$ where $k$ is the multiplicity of the knots.

**(2) Numerical discretization in the IGA**

The NURBS basis functions are firstly applied to parametrize the structural domain, and then construct the space for structural responses. As far as the latter, the key principle is that the continuous solution space is approximately defined by a linear combination of all NURBS basis functions with the nodal responses on control points. The mathematical formula of the space keeps the same as the geometrical model in Eq. (1), while control coefficients correspond to the structural responses on control points, expressed as:

$$\mathbf{x}(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} R_{i,j}^{p,q}(\xi, \eta) \mathbf{x}_{i,j} \tag{4}$$

where $\mathbf{x}$ is the field of structural responses in design domain, and $\mathbf{x}_{i,j}$ is the structural response on the control point $(i,j)_{th}$. Considering the linearly elastic in IGA, the system stiffness matrix is obtained by assembling the element stiffness matrix which is calculated by the Gauss quadrature method, as:

$$\mathbf{K}_e = \sum_{i=1}^{3} \sum_{j=1}^{3} \{ \mathbf{B}^T(\xi_i, \eta_j) \mathbf{D} \mathbf{B}(\xi_i, \eta_j) |\boldsymbol{J}_1(\xi_i, \eta_j)| |\boldsymbol{J}_2(\xi_i, \eta_j)| \omega_i \omega_j \} \tag{5}$$

where $\mathbf{B}$ is the strain-displacement matrix calculated by the partial derivatives of NURBS basis functions with respect to parametric coordinates.

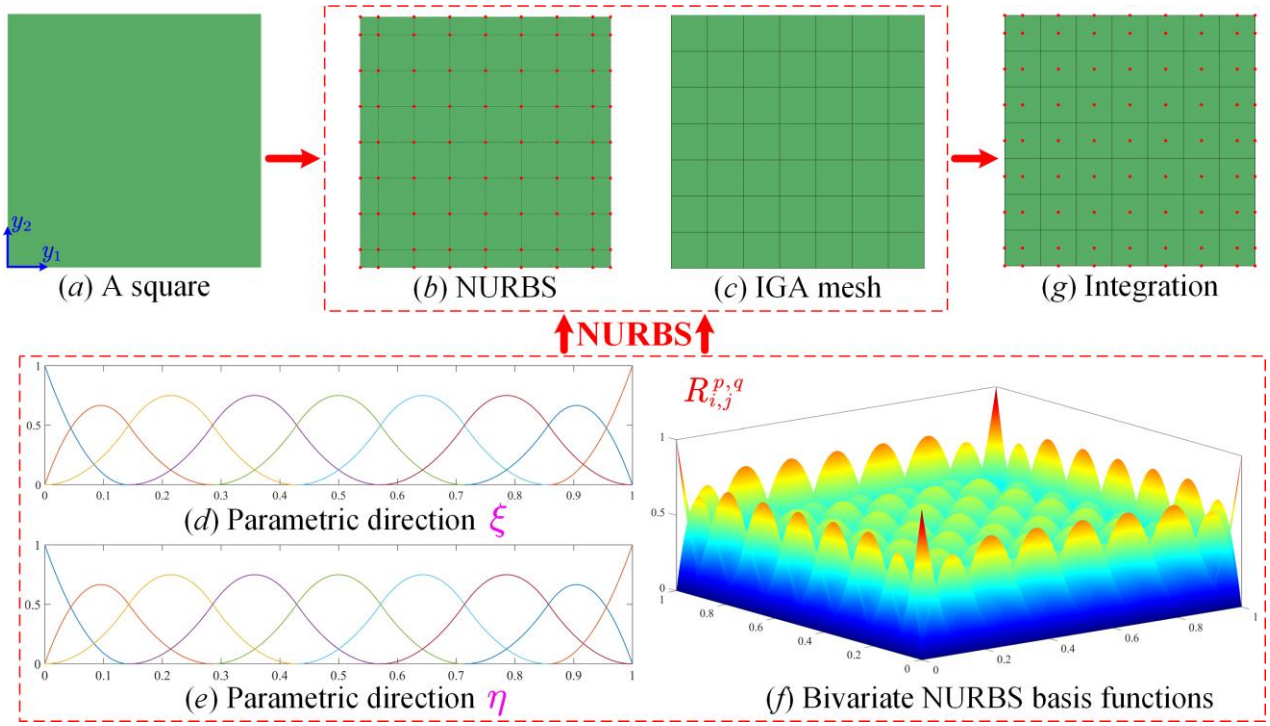



**Figure 1.** NURBS-based IGA for a square: $\Xi = \{0,0,0,0.1429,\cdots,0.8517,1,1,1\}$, $\mathcal{H} = \{0,0,0,0.1429,\cdots,0.8517,1,1,1\}$; $n = m = 9$; $p = q = 2$.

**IGA-based EBHM**

The principle of the homogenization is that the macroscopic effective properties of the bulk material are determined by using the information from the microstructure, where the microstructure is described in the coordinate system **y**. Considering the linear elasticity, only the first-order variation term with respect to the parameter expansion $\epsilon$ is considered. The effective elastic tensor of the bulk material $D_{ijkl}^{H}$ can be computed as:

$$D_{ijkl}^{H} = \frac{1}{|\Omega|}\int_{\Omega}\left(\varepsilon_{pq}^{0(ij)} - \varepsilon_{pq}\left(u^{ij}\right)\right) D_{pqrs}\left(\varepsilon_{rs}^{0(kl)} - \varepsilon_{rs}(u^{kl})\right) d\Omega \tag{6}$$

where $|\Omega|$ is the area (2D) or volume (3D) of the microstructure, and $D_{pqrs}$ is the locally varying elastic property. $\varepsilon_{pq}^{0(ij)}$ is the linearly independent unit test strain field, containing three components in 2D and six in 3D. $\varepsilon_{pq}\left(u^{ij}\right)$ denotes the unknown strain field in the microstructure, which is solved by the following linear elasticity equilibrium equation with **y**-periodic boundary conditions (PBCs):

$$\int_{\Omega}\varepsilon_{pq}\left(u^{ij}\right)D_{pqrs}\varepsilon_{rs}\left(\delta u^{ij}\right)d\Omega = \int_{\Omega}\varepsilon_{pq}^{0(ij)}D_{pqrs}\varepsilon_{rs}\left(\delta u^{ij}\right)d\Omega,\ \ \forall \delta u \in H_{per}(\Omega,\mathbb{R}^{d}) \tag{7}$$

where $\delta u$ is the virtual displacement in the microstructure belonging to the admissible displacement space $H_{per}$ with **y**-periodicity, and $d$ denotes the dimension of material

microstructure.

The homogenization is numerically performed by discretizing and solving Eq. (8) using the finite element method (FEM), and the utmost importance is the imposing of the PBCs on material microstructure. As an alternative method, the EBHM with a simplified periodic boundary formulation [13] is developed. Here, the numerical analysis of material microstructure is performed by IGA. In IGA, the displacement field in material microstructure is approximately expressed by a combination of the NURBS basis functions with the displacements at control points:

$$\mathbf{u} = \sum_{i=1}^{n}\sum_{j=1}^{m} R_{i,j}^{p,q}(\xi,\eta)\mathbf{u}_{i,j} \tag{8}$$

where $\mathbf{u}_{i,j}$ denote the displacements of the $(i,j)_{th}$ control point. As we can see, NURBS basis functions are linearly combined with nodal displacements to approximate the displacement field in the microstructure. In the application of the EBHM to evaluate material effective properties, the displacement field in material microstructure needs to satisfy the PBCs, and a general form is expressed as:

$$\mathbf{u}_k^+ - \mathbf{u}_k^- = \varepsilon(\mathbf{u}_0)\Delta k \tag{9}$$

where $k$ denote the normal direction of the structural boundary. $\mathbf{u}_k^+$ indicate the displacements of points at the structural boundary with the normal direction $k$, and the normal direction is in the positive direction of the coordinate axis. $\mathbf{u}_k^-$ correspond to the displacements of points at the opposite structural boundary. $\Delta k$ is the scale of the material microstructure along the direction of $k$.

**ITO formulation for architected materials**

Before developing the DDF, the definition of nodal densities assigned to control points needs to satisfy two basic conditions: (1) non-negativity; and (2) the strict bounds ranging from 0 to 1. Meanwhile, the Shepard function is used to improve the overall smoothness of nodal densities, so as to make sure the smoothness of the DDF. The corresponding mathematical model is given as:

$$\mathcal{G}(\rho_{i,j}) = \sum_{i=1}^{\mathcal{N}}\sum_{j=1}^{\mathcal{M}} \psi(\rho_{i,j})\rho_{i,j} \tag{10}$$

where $\mathcal{G}(\rho_{i,j})$ is the smoothed nodal density assigned to the $(i,j)_{th}$ control point, and $\rho_{i,j}$ is the initial nodal density. $\mathcal{N}$ and $\mathcal{M}$ are the numbers of nodal densities located at the local support area of the current nodal density in two parametric directions. $\psi(\rho_{i,j})$ is the Shepard function of the $(i,j)_{th}$ nodal density.

Assuming that the DDF in the structural domain is denoted by $\mathcal{X}$, the DDF is constructed by the NURBS basis functions with a linear combination of the smoothed nodal densities, expressed as:

$$\mathcal{X}(\xi,\eta)=\sum_{i=1}^{n}\sum_{j=1}^{m}R_{i,j}^{p,q}(\xi,\eta)\mathcal{G}\left(\rho_{i,j}\right) \tag{11}$$

It can be seen that the DDF has the same mathematical formula for NURBS in Eq. (1). The key difference is the physical meaning of control coefficients.

Here, the objective function for the topology optimization of architected materials with the low-density, which is defined by a function of the homogenized elastic tensor, given as:

$$\begin{cases} Find: \boldsymbol{\rho}\ \left\{\left[\rho_{i,j}\right]_{2\mathrm{D}}\ \ \left[\rho_{i,j,k}\right]_{3\mathrm{D}}\right\} \\ Min: J(\mathbf{u},\mathcal{X})=f\left(D_{\hat{i}\hat{j}\hat{k}\hat{l}}^{H}(\mathbf{u},\mathcal{X})\right) \\ S.t: \begin{cases} G(\mathcal{X})=\frac{1}{|\Omega|}\int_{\Omega}\mathcal{X}(\boldsymbol{\rho})v_0\,d\Omega-V_0\le 0 \\ a(\mathbf{u},\delta\mathbf{u})=l(\delta\mathbf{u}),\ \forall\delta\mathbf{u}\in H_{per}(\Omega,\mathbb{R}^d) \\ 0<\rho_{min}\le\boldsymbol{\rho}\le 1,(i=1,2,\cdots,n;j=1,2,\cdots,m;k=1,2,\cdots,l) \end{cases} \end{cases} \tag{12}$$

where $\boldsymbol{\rho}$ denotes the nodal densities assigned to control points, working as the design variables. $J$ is the objective function. $d$ is the spatial dimension of materials. $G$ is the volume constraint, in which $V_0$ is the maximum value and $v_0$ is the volume fraction of the solid. $\mathcal{X}$ is the DDF. $\mathbf{u}$ is the unknown displacement field in material microstructure, which have to satisfy the PBCs given in the above. $\delta\mathbf{u}$ is the virtual displacement field belonging to the admissible displacement space $H_{per}$ with **y**-periodicity, which is calculated by the linearly elastic equilibrium equation. $a$ and $l$ are the bilinear energy and linear load functions, as:

$$\begin{cases} a(\mathbf{u},\delta\mathbf{u})=\int_{\Omega}\varepsilon(\mathbf{u})\left(\mathcal{X}(\boldsymbol{\rho})\right)^{\gamma}\mathbf{D_0}\boldsymbol{\varepsilon}(\delta\mathbf{u})\,d\Omega \\ l(\delta\mathbf{u})=\int_{\Omega}\boldsymbol{\varepsilon^0}\left(\mathcal{X}(\boldsymbol{\rho})\right)^{\gamma}\mathbf{D_0}\boldsymbol{\varepsilon}(\delta\mathbf{u})\,d\Omega \end{cases} \tag{13}$$

It should be noted that the elastic tensor is assumed to be an exponential function with respect to the DDF, and $\gamma$ is the penalization parameter. $\mathbf{D_0}$ is the constitutive elastic tensor of the basic material.

**Numerical Examples**

In this section, several numerical examples are provided to demonstrate the effectiveness and efficiency of the ITO method. In all examples, the Young's moduli $E_0$ and the Poisson's ratio $v_0$ for the basis material are defined as 1 and 0.3, respectively. In the numerical analysis, 3×3

(2D) or 3×3×3 (3D) Gauss quadrature points are chosen in an IGA element. For numerical simplicity, the dimensions of material microstructures in all directions are set to be 1. The penalty parameter is set as 3. Considering 2D materials, the structural design domain is a square with 1×1, shown in **Figure 1**. Here, NURBS surface is applied to parametrize the design domain, where the quadratic NURBS basis functions are chosen and the knot vectors are set as: $\Xi = \mathcal{H} = \{0,0,0,0.01,\cdots,0.99,1,1,1\}$. The corresponding IGA mesh for the design domain has 100×100 elements, and 101×101 (10202) control points are contained in the NURBS surface. In all examples, the maximum material consumption $V_0$ for different cases is defined as 10%.

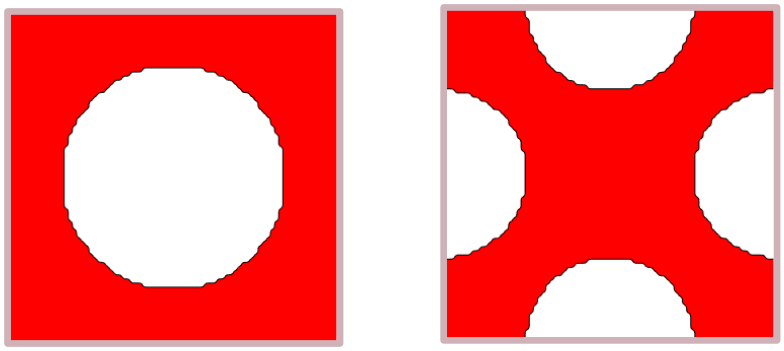


**Figure 2.** Initial design 1 and Initial design 2

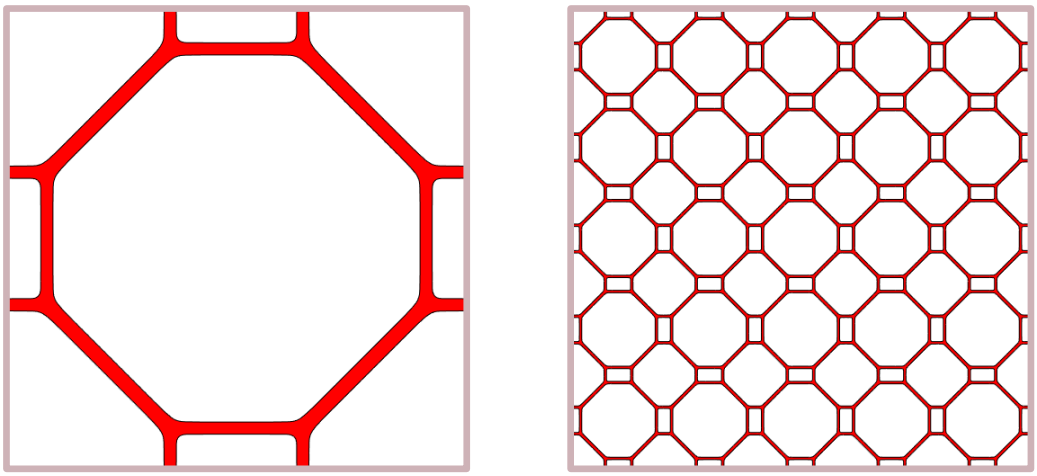


**Figure 3.** The optimized design 1

In this example, we study the effectiveness of the defined formulation on topology optimization of architected materials with the maximum bulk modulus in an extremely low volume fraction. As shown in Fig. 2, two different initial designs are defined and then discussed into two different cases, respectively. As shown in Figs. 3 and 4, the optimized results of the low-density architected materials with the maximum bulk modulus are provided. It can be easily seen that the optimized results are very similar to the known lattice structures, but the current design is obtained from a rational design using the isogeometric topology optimization framework.

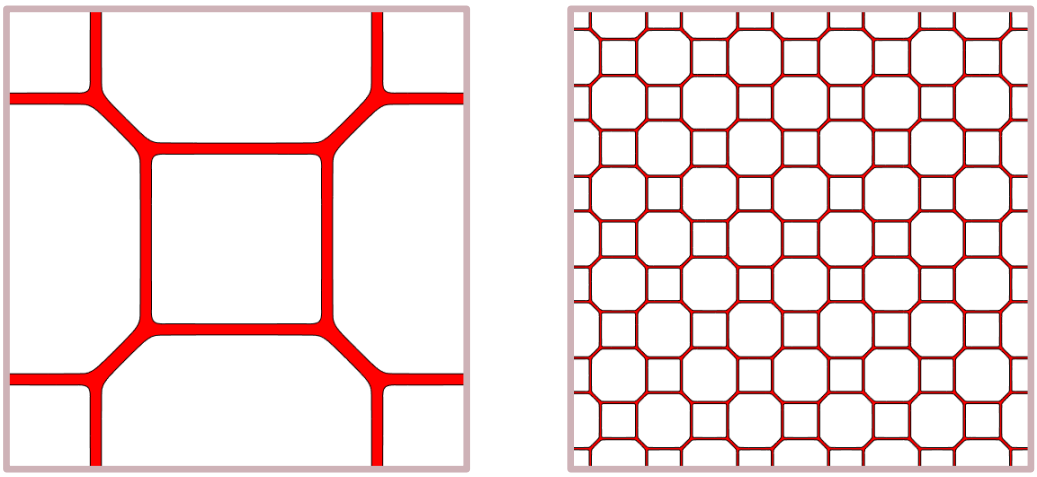


**Figure 4.** The optimized design 2

Moreover, we also perform the discussions of the ITO method on the optimization of architected materials with the maximum shear modulus. In this example, the parameters are consistent with the above example, and the objective function is defined based on the shear modulus. Two different cases are both discussed with two distinct initial designs, and the optimized topologies of the low-density architected materials with the maximum shear modulus are displayed in Figs. 5 and 6. As we can see, the optimized design with the bars in $45^o$ can provide the sufficient stiffness for affording the load, which shows the effectiveness of the current work. Additionally, the optimized low-density architected materials are also analogous to the known lattice materials, which has gained the extensive applications in the aerospace engineering. Hence, we can easily see that the current ITO method has the capability to seek the low-density architected materials with the sufficiently stiffness.

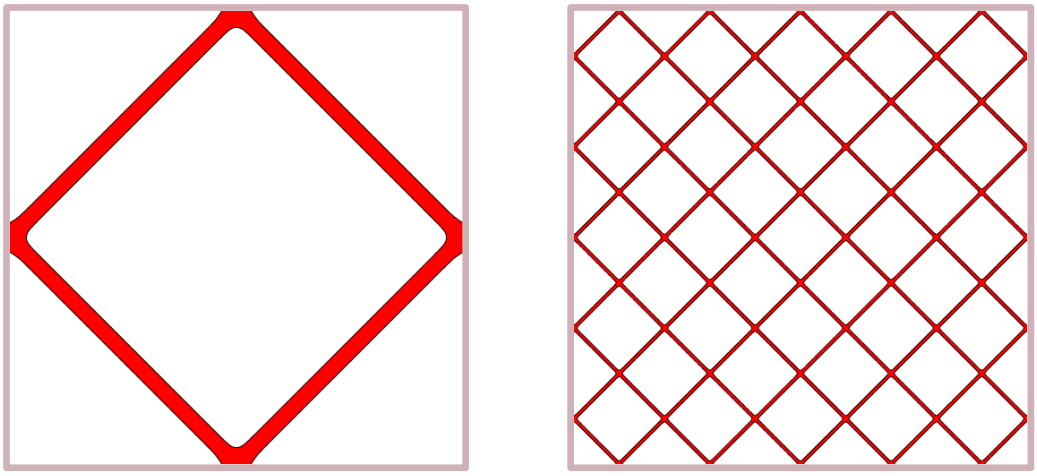


**Figure 5.** The optimized design 1

## Conclusions

In this paper, an effective and efficient ITO method is developed for the rational design of the low-density architected materials with the optimal stiffness, where a DDF with the desired smooth and continuous is constructed to represent the structural topology and IGA is applied to solve the displacement responses in microstructures. The EBHM to predict the macroscopic

effective properties is numerical implemented by the IGA, with the consideration of the periodic boundary conditions. Several numerical examples are given to show the basic features and effectiveness of the proposed ITO method. We can obtain the optimized low-density architected materials, very similar to the lattice materials.

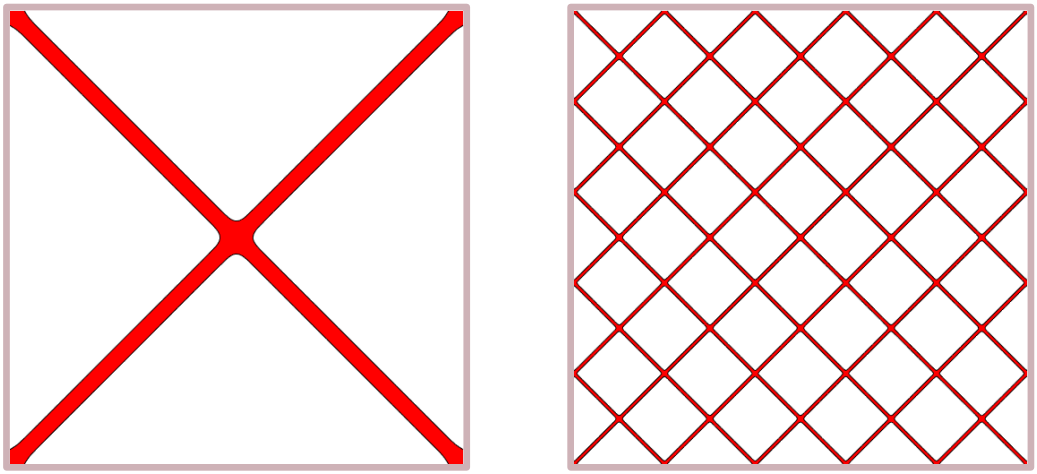


**Figure 6.** The optimized design 2

## References

[1] Gibson, L.J. and Ashby, M.F. (1999) Cellular solids: structure and properties, Cambridge university press.

[2] Ashby, M.F., Evans, A., Fleck, N.A. and et al. (2000) Metal Foams: A Design Guide. Butterworth-Heinemann.

[3] Osanov, M., and Guest, J.K. (2016) Topology optimization for architected materials design, *Annual Review of Materials Research* 46, 211-233.

[4] Bendsøe, M.P. and Kikuchi, N. (1988) Generating optimal topologies in structural design using a homogenization method, *Computer Methods in Applied Mechanics and Engineering* 72, 197-224.

[5] Zhou, M. and Rozvany, G. (1991) The COC algorithm, Part II: Topological, geometrical and generalized shape optimization, *Computer Methods in Applied Mechanics and Engineering* 89, 309-336.

[6] Bendsøe, M.P. and Sigmund, O. (1999) Material interpolation schemes in topology optimization, *Archive of Applied Mechanics* 69, 635-654.

[7] Xie, Y.M. and Steven, G.P. (1993) A simple evolutionary procedure for structural optimization, *Computers & Structures* 49, 885-896.

[8] Sethian, J.A. and Wiegmann, A. (2000) Structural boundary design via level set and immersed interface methods, *Journal of Computational Physics* 163, 489-528.

[9] Wang, M.Y., Wang, X.M. and Guo, D.M. (2003) A level set method for structural topology optimization, *Computer Methods in Applied Mechanics and Engineering* 192, 227–246.

[10] Allaire, G. Jouve, F. and Toader A-M. (2004) Structural optimization using sensitivity analysis and a level-set method, *Journal of Computational Physics* 194, 363-393.

[11] Sigmund, O. (1994) Materials with prescribed constitutive parameters: an inverse homogenization problem, *International Journal of Solids and Structures* 31, 2313-29.

[12] Gao, J., Li, H., Luo. Z. and et al. (2018) Topology optimization of micro-structured materials featured with the specific mechanical properties, *International Journal of Computational Methods* 15, 1850144.
[13] Gao, J., Li, H., Gao, L. and Xiao, M. (2018) Topological shape optimization of 3D micro-structured materials using energy-based homogenization method, *Advances in Engineering Software* 116, 89-102.
[14] Li, H., Luo, Z., Gao, L. and Qin Q. (2018) Level set-topology optimization for concurrent design of structures with multi-patch microstructures, *Computer Methods in Applied Mechanics and Engineering* 331, 536-561.
[15] Xia, L. and Breitkopf, P. (2014) Concurrent topology optimization design of material and structure within nonlinear multiscale analysis framework, *Computer Methods in Applied Mechanics and Engineering* 278, 524-542.
[16] Cottrell J.A., Hughes T.J. and Bazilevs Y. (2009) Isogeometric analysis: toward integration of CAD and FEA, John Wiley & Sons.

# The performance of ES-FEM with automatic triangle mesh adaptation in engineering mechanics applications

***Le Hoang Vu, †Sawekchai Tangaramvong,**

Applied Mechanics and Structures Research Unit
Department of Civil Engineering, Chulalongkorn University, Thailand.

*Presenting author: lehoangvu.chula.11793@gmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

This paper theoretical studies the performance of automatic adaptive edge-based smoothed finite element method (ES-FEM) for the solutions of elastic in-plane engineering mechanics problems. An ES-FEM adopting a strain smoothing technique over the edges of a generic triangular mesh presents coarse mesh accuracy in numerical analysis results. The automatic refinement of model construction from coarse-to-fine of meshes associated with high intensity of stresses and vice versa for others with low stress intensity. The specific L2-norm error, in a similar fashion to Zienkiewicz–Zhu, estimator indicates the difference between numerical von Mises stress solutions and recovery stresses underpinning the structural model with applied forces. A number of benchmarks, i.e. especially those subjected to stress singularity and/or incompressibility conditions, have been adopted for the comparisons in views of solution accuracy and computational robustness between ES-FEM and some standard isoparametric finite element model. Significant improvement in the computing efficiency and hence solution convergence has been clearly evidenced as when the ES-FEM analysis was encoded with the automatic model adaptation and vectorization within MATLAB environment, simultaneously.

**Keywords:** ES-FEM, Automatic Adaptive Mesh, Posterior Error Estimator, Vectorization.

## Introduction

Smoothed finite element methods (S-FEMs) have been successfully applied in solving many engineering mechanics problems. [1] combined the finite element method (FEM) to some of the meshfree techniques. Besides the information at nodes on each element S-FEM models consider nodal unknowns of neighboring elements to construct smoothed strain field to enhance stability, convergence and accuracy of the solutions. The diversity of two-dimensional S-FEMs which are edge-, nodal- and cell-based S-FEMs with different properties is created by applying the strain smoothing technique from [2]. With the smoothed strain field, these models exhibit desirable properties and work well with a general n-side polygonal elements, especially for the three-node triangular (T3) meshes. Among the above S-FEM models, the Edge-based Smoothed Finite Element Method (ES-FEM) have arisen as the most outstanding S-FEM model which possesses so many advantages such as stable both spatially and temporally, much more accurate compared with many available FEMs. The ES-FEM creates models with close-to-exact stiffness so that it is efficient for solving both static and dynamic problems.

For mechanics problems associated with physically instabilizing stress singularity, the standard model construction has experienced the low accuracy of results using standard finite elements. A process of iterative mesh reconstruction will automatically decide where high

density of mesh are required to obtain a proper mesh distribution in each step of analysis. A layer of singular five-node elements will be applied around the crack-tip to capture the theoretical occurrence of unbounded stresses. Modifications in error assessment procedure for these elements are presented to accurately converge the stress response results over the local areas of structures considered.

The method [3] employed standard Delaunay triangulation procedure without any error indicator, considered as a geometric dependent refinement. In addition, a singular ES-FEM [4] adopted a recovery-based error indicator in an energy norm to predict accurately singular stress field around re-entrant corners. The applications of the node-based smoothed finite element method (NS-FEM) using the similar recovery-based error function were described in [5]. where it demonstrated clearly the good convergence capability and upper-bound strain energy solutions over iterative mesh reconstruction processes.

However, it is uncommon that the researchers interested in adaptive analysis showed the results in terms of runtime or computing resources in the past. The ES-FEM-T3 models incorporating a simple yet effective recovery-based error function of von Mises stresses were encoded within a vector-oriented MATLAB environment. The newest node bisection algorithm was adopted to the automatic AMR procedures. In essence, a parent triangle element is sub-divided into several children triangle elements along the longest edge to eliminate hanging nodes. The effciency of the present models is tested by using the two problems, from which successfully solved by [6]-[7], in providing solutions of such challenging problems under the presence of stress singularity and discontinuity field. Moreover, the proposed analysis framework also describes a significant reduction of computing resources as compared to an uniform model construction strategy.

The following section includes a brief description of ES-FEM-T3 formulations and displacement interpolation within a singular element. Then, an overview of the recovery-based stress adaptive mesh algorithm incorporated with the longest-edge refinement technique is provided. In the next Section, two numerical examples subjected to the challenging of elastic stress singularity and discontinuity are given to illustrate applications of the developed analysis scheme. They also highlight the accuracy and robustness of the proposed ES-FEM-T3 framework. The main conclusion and suggestion for future research are drawn in the end part.

## Brief description of the ES-FEM model

*An Overview of The ES-FEM Using Triangular Elements*

A set of $N_s = N_{eg}$ for both ”non-overlap” and “no-gap” smoothing domains $\Omega_k^s$ will fill in the whole problem domain $\Omega = \cup_{k=1}^{N_s} \Omega_k^s$ and $\Omega_i^s \cap \Omega_j^s = \varnothing,\ i \neq j$. The strains of generic ES-FEM-T3 elements are smoothed over the smoothing domains by connecting between two endpoints of the edge and central points of neighboring elements.
The smoothing strain operator over the edge-based smoothing domain $\Omega_k^s$ can be defined by:

$$\bar{\varepsilon}_k = \int_{\Omega_k^s} \tilde{\varepsilon}(x) W(x) d\Omega \tag{1}$$

where $\tilde{\varepsilon}(x)$ is the compatible strain field in FEM-T3, and $W(x)$ is a weight function as

$$W(x)=\begin{cases}1/A_k^s, & x\in\Omega_k^s \\ 0, & x\notin\Omega_k^s\end{cases} \quad \& \int_{\Omega_k^s} W(x)d\Omega=1 \tag{2}$$

A Green's divergence theorem is applied and yields the following smoothed strains:

$$\bar{\varepsilon}_k=\frac{1}{A_k^s}\int_{\Omega_k^s}\tilde{\varepsilon}(x)d\Omega=\int_{\Omega_k^s}L_d\left[\bar{u}(x)\right]W(x)d\Omega=\frac{1}{A_k^s}\int_{\Gamma_k^s}L_n(x)\bar{u}(x)d\Gamma \tag{3}$$

where $L_n(x)$ is a matrix of containing the outward normal vectors to the boundary $\Gamma_k^s$ as depicted in Fig. 1, $A_k^s=\int_{\Omega_k^s}d\Omega=\frac{1}{3}\sum_{j=1}^{N_e^k}A_j$ is the area of smoothing domain $\Omega_k^s$, $N_e^k$ is the number of elements containing edge k.

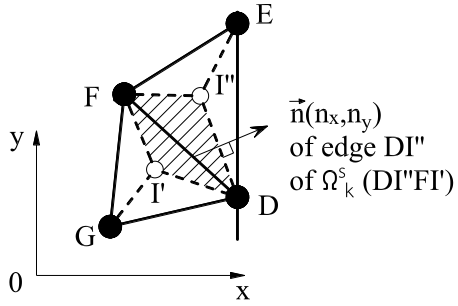



**Figure 1. Outward normal to the edge DI'' under the smoothing domain DI''-FI' in x-y coordinates**

The smoothed strain–displacement matrix $\bar{B}_i(x,y)$ of node i-th is constructed by having

$$\bar{B}_i=\frac{1}{A_k^s}\int_{\Gamma_k^s}L_n(x)N_i(x)d\Gamma=\begin{bmatrix}\bar{b}_{ix} & 0\\ 0 & \bar{b}_{iy}\\ \bar{b}_{iy} & \bar{b}_{ix}\end{bmatrix} \text{ with } \bar{b}_{ix(y)}=\frac{1}{A_k^s}\int_{\Gamma_k^s}n_{x(y)}(x)N_i(x)d\Gamma \tag{4}$$

One single Gauss's integration point is applied for each q-th segment $\Gamma_{k,q}^s$ of the boundary $\Gamma_k^s$

$$\bar{b}_{ix(y)}=\frac{1}{A_k^s}\sum_{p=1}^{n_\Gamma^s}n_{x(y),\mathrm{p}}(x)N_i\left(x_p^{Gauss}\right)\mathrm{l}_p \tag{5}$$

where $n_\Gamma^s$ is the total number of boundary segments $\Gamma_{k,q}^s\in\Gamma_k^s$, $x_p^{Gauss}$ are the coordinates of a Gauss's point on the boundary segment $\Gamma_{k,q}^s$, $n_{x(y),p}$ and $l_p$ denote the unit normal and the length of the boundary q-th segment $\Gamma_{k,q}^s$.

*An ES-FEM Formulation Using A Layer of Singular Elements*

Problems with a re-entrant corner as introduced by [8] have the singular stress field of arbitrary order. The power singular term $\lambda$, depends only on the vertex angles of $\pi<\phi\le 2\pi$. [9] and [10] theoretically showed the occurrence of elastic stress singularities at angular corners resulting from various BCs rather than only the free-free BC as normally encountered in crack problems. The $\lambda$ (in the term $\sigma\sim r^{\lambda-1}$) is interpolated from the graph as provided in

[9] or computed from the characteristic equations in [10] depending on the value of vertex angle and the BCs on the two radial edges.

The linear interpolation used in standard finite elements cannot reproduce such a singular field. The most widely used technique to simulate this kind of stress singularity is the so-called (quadratic) 6-node crack-tip element in which the additional midpoint are shifted by a quarter edge-lengths toward the crack-tip. The singularity is then achieved nicely by the well-known iso-parametric mapping procedure [11]. In the present singular ES-FEM-T3 method, however, no mapping is needed and only the shape function values (not the derivatives) are required. Making use of this important feature of ES-FEM-T3, the stress singularity at the crack tip can be created by a simple point interpolation method with extra basis functions of proper fractional order polynomials.

The domain with a horizontal opening crack is discretized using a layer of five-noded triangular elements that contribute to the crack-tip and standard three-node triangular elements in the remaining area as in Fig. 2a). Only an additional node on each edge that directly connected to the singular point is added in general at any point Fig. 2b) that can produce a proper order of stress singularity near the crack tip.

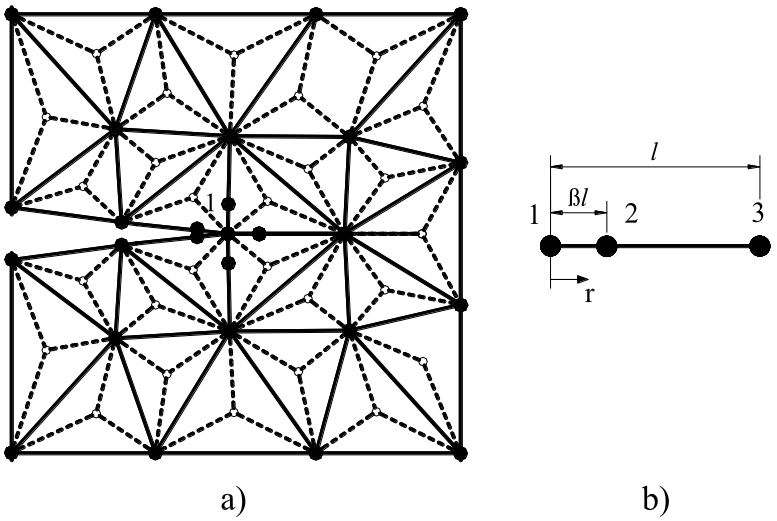



**Figure 2. a) Triangular mesh with layer of five-node elements in the ES-FEM-T3,**

**b) Additional node at arbitrary location in radial coordinate originated at the crack-tip**

The displacement field, u, along the crack-tip element edge is approximated using:

$$u(x) = P^T(x)c = c_0 + c_1 r + c_2 r^{\lambda} \tag{6}$$

where $0 \le r \le l$ is the radial coordinate originated at the crack-tip (node 1 in Fig. 2b), and $c_i (i = 0,1,2)$ are the coefficients yet to be determined, $1/2 \le \lambda < 1$ is a singularity parameter.

The unknown coefficients $c$ can be obtained by substituting the radial coordinates of nodes into Eq. (6). Then, we replace it in the same equation to get the matrix form as given follows:

$$u(x) = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 \end{bmatrix} \begin{Bmatrix} u_1 & u_2 & u_3 \end{Bmatrix}^T = \Phi(x) d \tag{7}$$

where $\Phi_i (i = 1,2,3)$ are shape functions corresponding to the three nodes on the edge that directly connected to the crack-tip and is defined as follows:

$$\Phi_1 = 1 + \frac{r^\lambda(\beta - 1) + l^{\lambda-1}(1-\beta^\lambda)r}{(\beta l)^\lambda - \beta l^\lambda}, \Phi_2 = \frac{r^\lambda - l^{\lambda-1}r}{(\beta l)^\lambda - \beta l^\lambda}, \Phi_3 = \frac{l^{\lambda-1}\beta^\lambda r - \beta r^\lambda}{(\beta l)^\lambda - \beta l^\lambda} \tag{8}$$

and $u_i (i = 1,2,3)$ are the nodal displacements, $l$ is the length of the element edge, and $0 \le \beta \le 1$ is the propotion of the edge 1-2 over the edge 1-3.

For fracture mechanics problems with in-line crack faces, $\lambda = 1/2$, and simply take $\beta = 1/4$, the shape functions become

$$\Phi_1 = 1 + 2\frac{r}{l} - 3\sqrt{\frac{r}{l}},\ \Phi_2 = -4\frac{r}{l} + 4\sqrt{\frac{r}{l}},\ \Phi_3 = 2\frac{r}{l} - \sqrt{\frac{r}{l}} \tag{9}$$

These $\Phi_i (i = 1,2,3)$ satisfy all the basic properties of a shape function such as linear reproducibility, Partition of Unity, Kronecker Delta properties and can actually produce the singularity of stress field with the power of 1/2 near the crack-tip.

In the radial direction, the displacement field takes the enriched form as given in Eq. (6), while in the tangential direction it is assumed to be linearly dependent to ensure the compatibility along the two-node edge of crack-tip elements. Now, we consider the five-node element 1-4-2-3-5 and two layers of singular smoothing domains are good enough to ensure both stability and accuracy in approximating the singular term around the crack-tip as depicted in Fig. 3a)

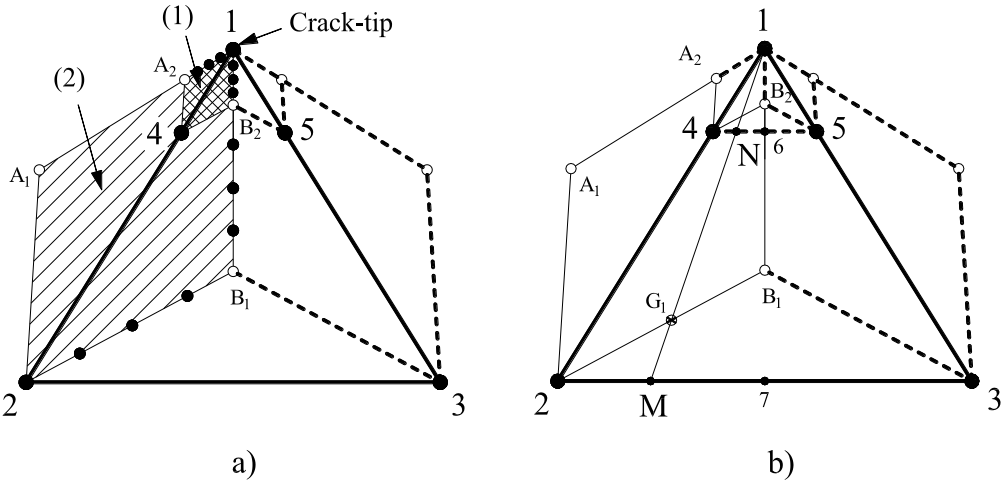



**Figure 3.**

**a) Two layers of singular smoothing domains using three Gauss-point interpolation,**

**b) Additional node at arbitrary location in radial coordinate originated at the crack-tip**

Along the arbitrary radial line 1-N-M, displacement is obtained using the Eq. (6) as

$$u = \Phi_1 u_1 + \Phi_2 u_N + \Phi_3 u_M \tag{10}$$

and

$$u_N = \left(1 - \frac{l_{N-4}}{l_{5-4}}\right)u_4 + \frac{l_{N-4}}{l_{5-4}}u_5,\quad u_M = \left(1 - \frac{l_{M-2}}{l_{3-2}}\right)u_2 + \frac{l_{M-2}}{l_{3-2}}u_3 \tag{11}$$

Similar triangle rule leads to $\frac{l_{N-4}}{l_{5-4}} = \frac{l_{M-2}}{l_{3-2}} = \alpha$ . Substituting Eq. (11) into Eq. (10) yields

$$u = \Phi_1 u_1 + (1-\alpha)\Phi_3 u_2 + \alpha\Phi_3 u_3 + (1-\alpha)\Phi_2 u_4 + \alpha\Phi_2 u_5 \tag{12}$$

In matrix form: $u = \begin{bmatrix} N_1 & N_2 & N_3 & N_4 & N_5 \end{bmatrix} \begin{Bmatrix} d_1 & d_2 & d_3 & d_4 & d_5 \end{Bmatrix}^T = Nd$ (13)

where $N$ is the matrix of shape functions of the singular element.

The smoothed strain–displacement matrix of each layer of singular smoothing domain is

$$\bar{B}_i^a = \begin{bmatrix} \bar{b}_{ix}^a & 0 \\ 0 & \bar{b}_{iy}^a \\ \bar{b}_{iy}^a & \bar{b}_{ix}^a \end{bmatrix} \text{ with } \bar{b}_{ix(y)}^a(x_k) = \frac{1}{A_k^{s,a}} \int_{\Gamma_k^{s,a}} N_i(x) n_{x(y)}^{k,a}(x) d\Gamma \tag{14}$$

where $A_k^{s,a}$, $\Gamma_k^{s,a}$ is the area and the boundary of the a-th layer of the singular SD $\Omega_k^{s,a}$, respectively, the shape functions $N_i(x)$ in Eq. (12) are adopted in this case, $n_{x(y)}^{k,a}$ is the unit normal vector of the boundary segment $\Gamma_k^{s,a}$

Similarly, we apply the Gauss integration along the segments of boundary $\Gamma_k^{s,a}$ , then

$$\bar{b}_{ix(y)}^a = \frac{1}{A_k^{s,a}} \sum_{p=1}^{n_\Gamma^s} \left[ \sum_{b=1}^{n_{Gauss}} w_{p,\mathrm{b}} n_{p,x(y)}^{k,a}(x) N_i\left(x_{p,b}^{Gauss}\right) \right] \tag{15}$$

where $n_{Gauss} = 3$ is the number of Gauss-points on each boundary segment Fig. 3a), $w_{p,\mathrm{b}}$ is the corresponding weight coefficient of the Gauss-points, $x_{p,b}^{Gauss}$ is the b-th Gauss-point of the pth boundary segment of $\Gamma_{k,p}^{s,a}$ and $n_\Gamma^s$ is the number of boundary segments of $\sum_{p=1}^{n_\Gamma^s} \Gamma_{k,p}^{s,a} = \Gamma_k^{s,a}$

In this model, the Galerkin Weak Form is employed in a similar fashion to standard FEM-T3. However, the formulations of a stiffness matrix are associated with the smoothing domains, rather than individual members by

$$\bar{K} = \sum_{k=1}^{N_s} \bar{K}^k \tag{16}$$

and the nodal load vector $\tilde{f}$ is the same as that of the corresponding standard FEM-T3 as given below

$$\tilde{f}^e = \sum_{N_e} \int_{\Omega^e} N^T b^e d\Omega + \sum_{N_e} \int_{\Gamma_N^e} N^T t_N^e d\Gamma_N^e + \sum_{i=1}^{N_{node}} p_i \tag{17}$$

**Adaptive Procedure**

*Recovery Based Strain Error Indicator*

The global ZZ-type error indicator, $\eta_Z$ , is the sum of the local ZZ-type error indicators from all individual elements in L2-norm, namely

$$\eta_Z = \left( \sum_{l=1}^{N_e} \eta_l^2 \right)^{1/2} \tag{18}$$

$$\eta_l = \left\| \bar{\sigma}_v - \bar{\sigma}_v^R \right\|_{L^2(\Omega_l)} = \left[ \int_{\Omega_l} \left( \bar{\sigma}_v - \bar{\sigma}_v^R \right)^T \left( \bar{\sigma}_v - \bar{\sigma}_v^R \right) d\Omega \right]^{1/2} \tag{19}$$

where $\bar{\sigma}$ is the numerical solution of stress and is area-weight averaged using the stresses from the two layers of singular SDs that connected to the crack-tip by

$$\bar{\sigma}_i = \frac{A_k^{s,1} \bar{\sigma}_i^1 + A_k^{s,2} \bar{\sigma}_i^2}{A_k^s} \tag{20}$$

and $\bar{\sigma}^R$ is the recovery field of stress that is continuous over the entire domain and converges to an exact solution for a sufficiently fine mesh in ES-FEM-T3 obtained by

$$\bar{\sigma}^R = \sum_{i=1}^{n_n^e} N_i(x,y) \bar{\sigma}^R(i) \tag{21}$$

with $N_i(x,y)$ is the shape function at the i-th node (the same with standard FEM-T3), $\bar{\sigma}^R(i)$ is the vector containing the nodal stresses of an element (presenting the area-weight averaged stresses within smoothing singular ES-FEM-T3 domains see Fig. 4a,b,c)

$$\bar{\sigma}^R(\mathrm{i}) = \frac{1}{\mathrm{A}_i^{ns}} \sum_{k=1}^{n_s^i} \bar{\sigma}_k \, \mathrm{A}_k \tag{22}$$

in which $n_s^i$ is the number of smoothing domains $\Omega_k^s$ around the i-th node, $\mathrm{A}_i^{ns} = \sum_{k=1}^{n_s^i} A_k$ is the total area of all the smoothing domains sharing the i-th node, and $\mathrm{A}_k$ is the area of the k-th smoothing domain sharing the i-th node

Note that: (1) the layer of singular SDs that close to the node will be considered case b) & c) Fig. 4. (2) for the standard ES-FEM-T3, the procedure simply take the case a) Fig. 4.

The formulation for a direct determination of the recovery-based error indicator is given by

$$\eta_l^2 = \| \bar{\sigma}_v - \bar{\sigma}_v^R \|_{L^2(\Omega_l)}^2 = \| \bar{\sigma}_v - \sum_{i=1}^{3} N_i(x,y) \bar{\sigma}_v^R(i) \|_{L^2(\Omega_l)}^2 = \sum_{q=1}^{3} \| \bar{\sigma}_{v,q} - \sum_{i=1}^{3} N_{i,q}(x,y) \bar{\sigma}_{v,q}^R(i) \|_{L^2(\Omega_{l,q})}^2 \tag{23}$$

where $\Omega_{l,q}$ is the q-th sub-smoothing domain (sub-SD) of the l-th element and $\bar{\sigma}_{v,q}^R(i)$ is the recovery nodal von Mises stress at the i-th node of the q-th sub-SD.

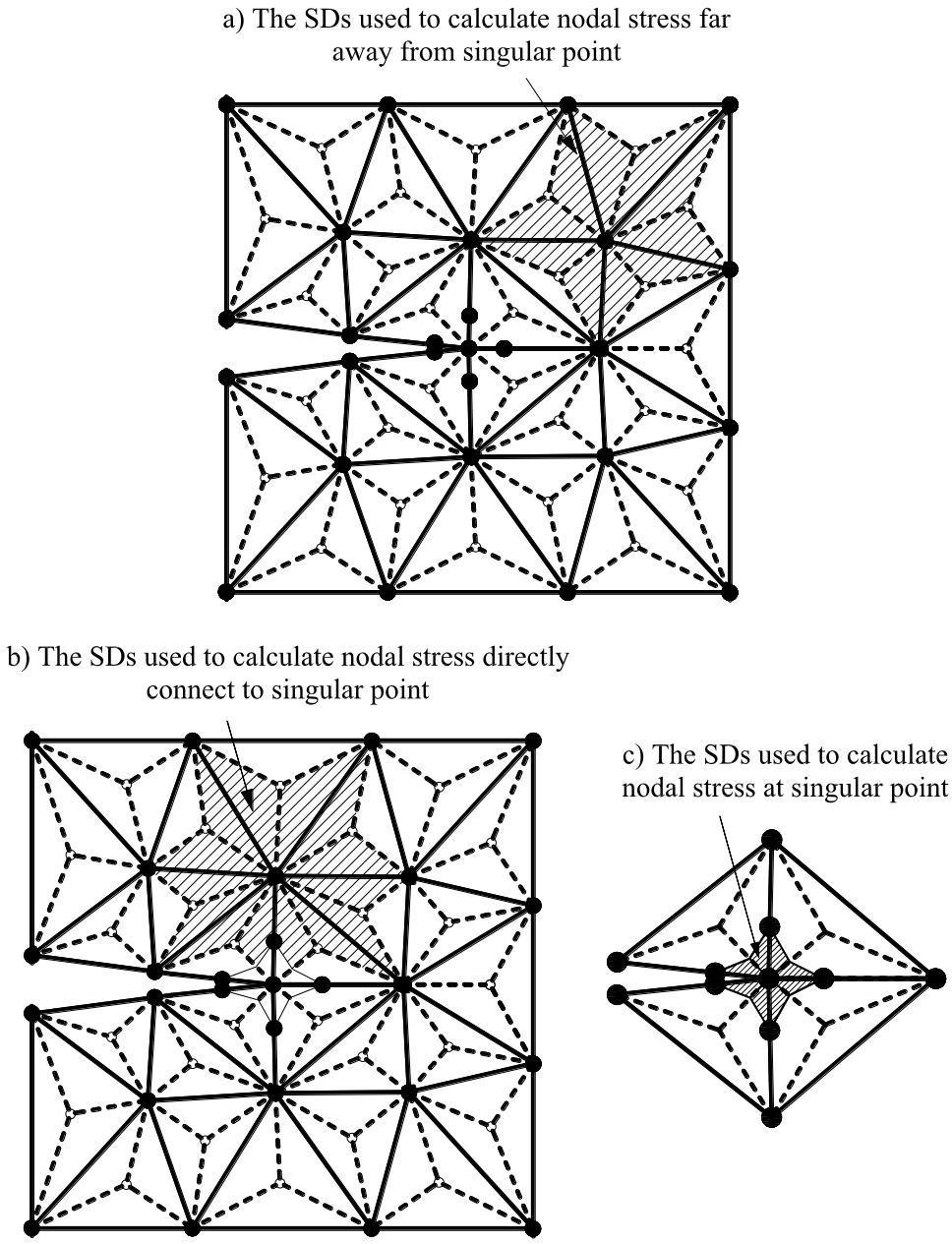



**Figure 4. The smoothing domains used to calculate the nodal stress for nodes in the singular ES-FEM-T3**

For each triangle element, there are three sub-SDs with constant smoothed stress and the recovery-based error indicator can be computed from their summation.
From the partition of unity property

$$\sum_{i=1}^{3} N_{i,q}(x,y) = 1 \tag{24}$$

$$\begin{aligned} \Rightarrow \eta_l^2 &= \sum_{q=1}^{3} \| \sum_{i=1}^{3} N_{i,q}(x,y)(\bar{\sigma}_{v,q}(i) - \bar{\sigma}_{v,q}^{R}(i)) \|_{L^2(\Omega_{l,q})}^2 \\ &= \sum_{q=1}^{3} \left[ \sum_{i,j=1}^{3} \bar{r}_{l,\mathrm{q,i}} . \bar{r}_{l,\mathrm{q,j}} \int_{\Omega_{l,q}} N_{i,q}(x,y) N_{j,q}(x,y) d\Omega \right] \end{aligned} \tag{25}$$

where $\bar{r}_{l,\mathrm{q,i}} = \bar{\sigma}_{v,q}(i) - \bar{\sigma}^{R}_{v,q}(i)$ at i-th node

The shape function for each sub-SD satisfies

$$\int_{\Omega_{l,q}} N_{i,q}(x,y) N_{j,q}(x,y) d\Omega = \begin{cases} A_{l,q}/6 = A_l/18 & i = j \\ A_{l,q}/12 = A_l/36 & i \neq j \end{cases} \tag{26}$$

where $A_{l,q}$ is the area of the q-th sub-SD (covering one-third area) of the l-th element, and hence

$$\eta_l^2 = \sum_{q=1}^{3} (\bar{r}^2_{l,\mathrm{q},1} + \bar{r}^2_{l,\mathrm{q},2} + \bar{r}^2_{l,\mathrm{q},3} + \bar{r}_{l,\mathrm{q},1}.\bar{r}_{l,\mathrm{q},2} + \bar{r}_{l,\mathrm{q},1}.\bar{r}_{l,\mathrm{q},3} + \bar{r}_{l,\mathrm{q},2}.r_{l,\mathrm{q},3}).(A_l/18) \tag{27}$$

*Refinement Strategy*

Using the element refinement indicators, $\eta_l$, the well-known Dorfler criterion [12] defines the elements $\Omega^e \in \Omega_M$ for refinement, where the minimal set $\Omega_M \in \Omega$ satisfies

$$\theta \sum_{\Omega_l \in \Omega} \eta_l^2 \leq \sum_{\Omega_l \in \Omega_M} \eta_l^2 \text{ with } \theta \in (0,1) \tag{1}$$

A new mesh $\Omega'$ is generated from the refinement of at least the marked elements so-called M-Group $\Omega_M \in \Omega$ to reduce the total numerical error in the whole domain.

*Newest Vertex Bisection Algorithm*

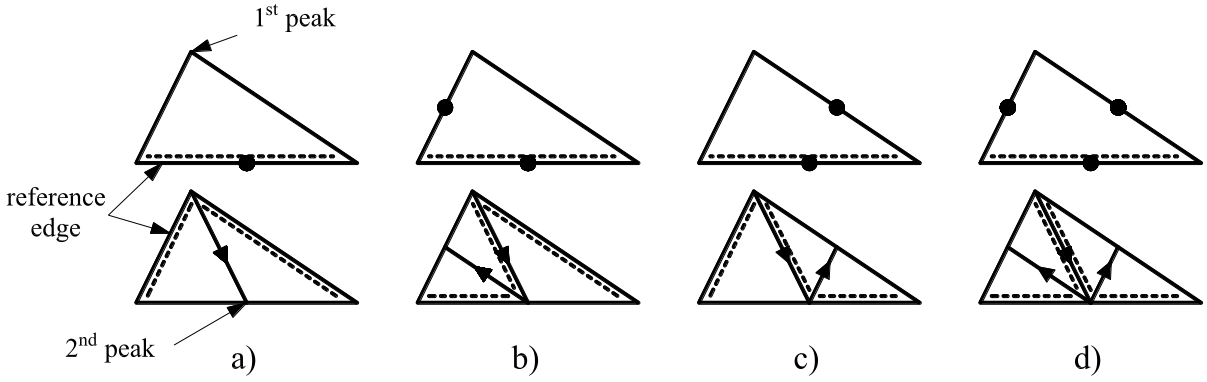



**Figure 4. Four basic longest-edge mesh refinement patterns.**

The Newest Vertex Bisection strategy [13], chooses to divide the parent element along the longest (reference) edge. This eliminates the capacity of producing triangles with smaller angles and the problem associated with hanging nodes [14]. Four typical types of partitioning a parent element into so-called child elements as in Fig. 4 are formed of lines connecting the newest (peak) vertex to the mid-point of the longest (reference) edge.

Refinement procedure:
(1) At least one reference edge (or dash-line) is marked for refinement.
(2) The newest vertex (midpoint) of reference edge becomes the peak for the next refinement step.
(3) The 2-nd refinement step is then implemented on the two other marked edges, if any from the newest peak.
* The same refinement procedure for each adaptive iteration.

## Illustrative Examples

Two problems with difficulties under the presence of discontinuity and singularity stress field are tested. The ES-FEM-T3 model were applied and encoded within a MATLAB environment using built-in function and vector language. A layer of five-node singular elements is employed in crack problems to validate the accuracy and robustness of the proposed analysis framework.

*Example 1: Prandtl's Punch*

A plane-strain Prandtl's punch problem with flexible foundation is drawn in Fig. 5. A total of uniformly distributed loads of 10 represents for the footing length of 2. The material properties employed were: $E = 10^4$, $v = 0.25$ and $t = 1$. Due to the symmetry of both geometry and loading configurations, only half of the structure is modeled. The characteristic discrete structural model in Fig. 5b) contains 256 simple triangle elements.

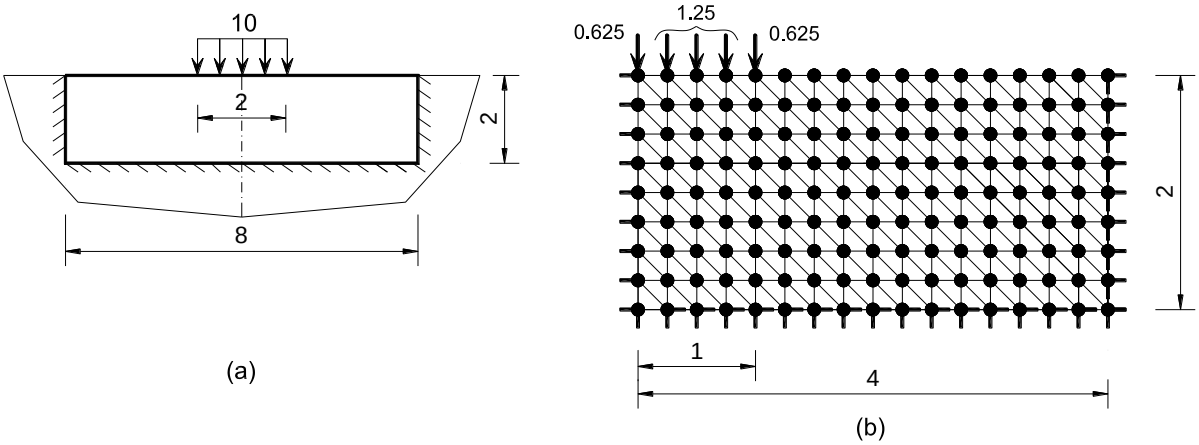



**Figure 5. Example 1: Prandtl's punch (a) geometry and loading, (b) characteristic ES-FEM-T3 model, where thick solid lines denote nodal restrained directions.**

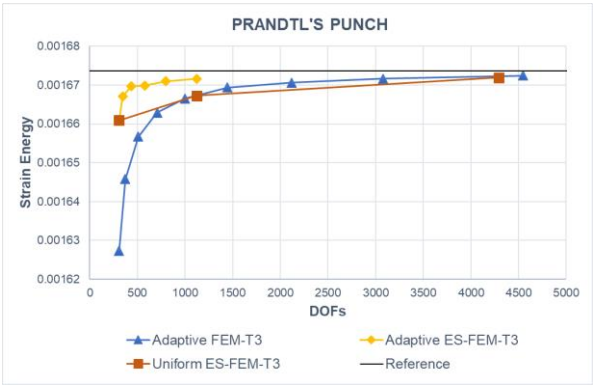



**Figure 6. Example 1: Convergence of strain energy results from various analysis methods.**

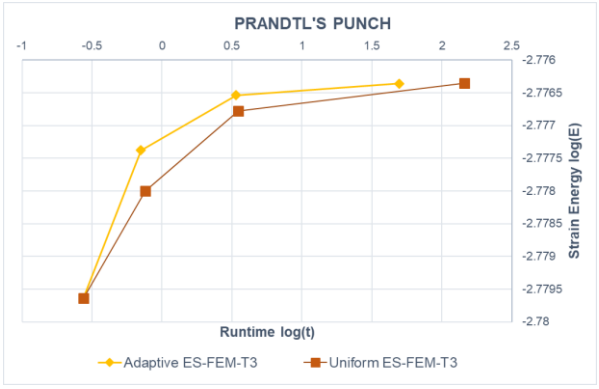



**Figure 7. Example 1: Convergence rate of runtime versus strain energy between adaptive ES-FEM-T3 and uniform ES-FEM-T3.**

The proposed automatic adaptive ES-FEM-T3 analysis approach was successfully processed to obtain elastic strain energy response solutions. The analysis results computed are plotted with their associated DOFs in Fig. 6, where those of some other standard FEMs, namely FEM-T3 with mesh adaptive scheme and ES-FEM-T3 with uniform mesh refinement. It is clear that all methods yield the solutions converged to the reference value at the sufficient fine

numbers of discrete elements. In addition, the proposed automatic adaptive ES-FEM-T3 approach provided the fast-converged strain energy solutions as compared to ES-FEM-T3 with uniform mesh refinement. The computing times as required for successfully converging the results by automatic mesh adaptive recovery-based strain algorithm, as displayed in Fig. 7, were less than those from standard uniform mesh refining technique.

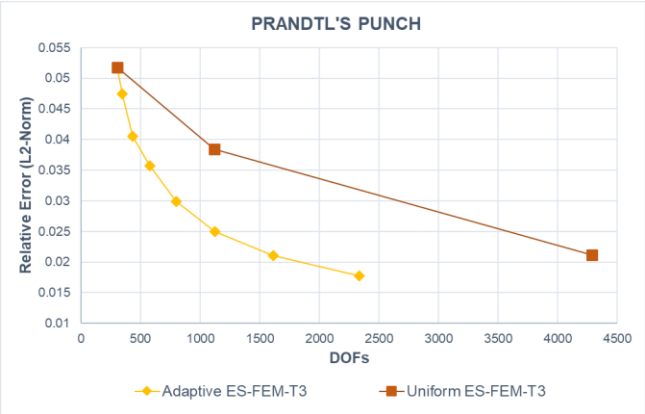
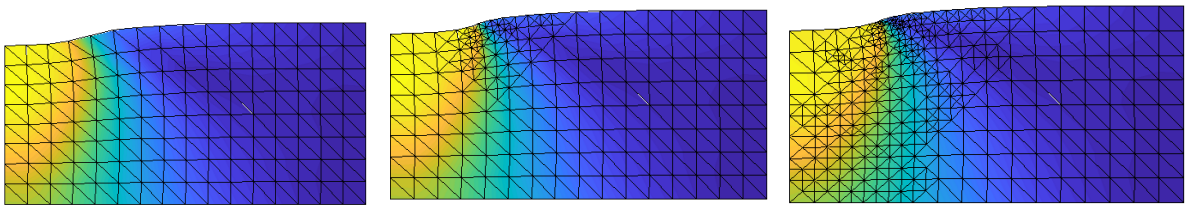



**Figure 8. Example 1: Convergence of relative error results from adaptive ES-FEM-T3 and uniform ES-FEM-T3.**



**(a) 256 ele. (error 5.17%) - (b) 379 ele. (error 4.18%) - (c) 758 ele. (error 3.14%)**

**Figure 9. Example 1: Automatic adaptive meshes with the corresponding contour line of von Mises stress distributions.**

The present adaptive mesh implementation dramatically reduces the recovery-based relative error compared to a slow reduction using the normal uniform refinement strategy. The values from adaptive ES-FEM-T3 fast approach the zero value, while those from uniform ES-FEM-T3 is still at a very high value in Fig. 8. The von Mises stress results corresponding to automatically adaptive meshes are depicted in Fig. 9. It illustrates the mesh refinements localizing over strong discontinuity areas of applied load.

*Example 2: Double-Edge Notched Specimen*

The second example considers a plane strain double-edge notched specimen as depicted in Fig. 10 subjected to the total uniform lateral load of 1.44 on both end edges. The material properties of E = 70, v = 0.3 and t = 1 were adopted. Because of symmetry in both x-axis and y-axis, only a quarter of the specimen without undue loss of accuracy was analyzed. The initial characteristic ES-FEM-T3 with 150 elements is displayed in Fig. 10b)

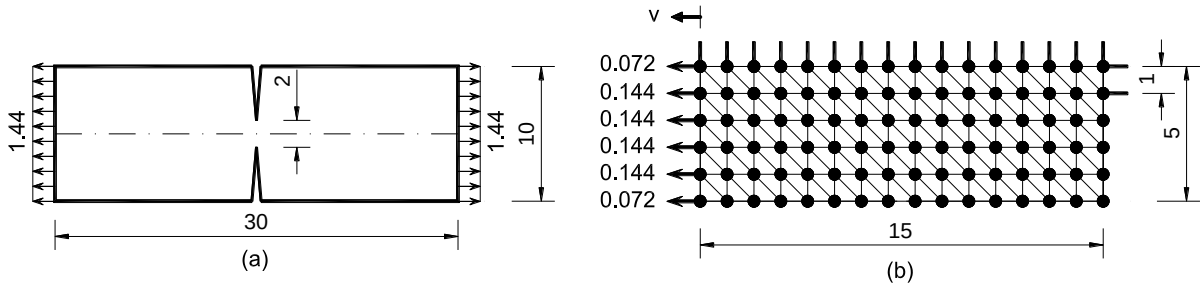



**Figure 10. Example 2: Double-edge notched specimen (a) Initial mesh and loading (b) ES-FEM-T3 model, where thick solid lines denote nodal restrained directions.**

The ES-FEM-T3 incorporated the automatic adaptive scheme adopting recovery-based strain error functions. The lateral displacements v and strain energy responses (displayed in Figs.11 and 12, respectively) were successfully computed for various mesh refinements, and compared with those found from other standard FEM-T3 with the similar adaptive mesh algorithm.

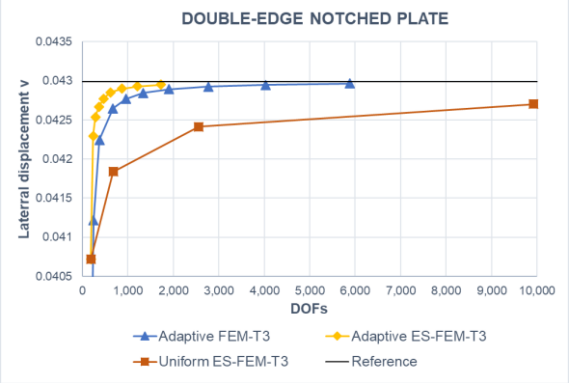



**Figure 11. Example 2: Convergence of lateral displacement results for various automatic adaptive mesh algorithms.**

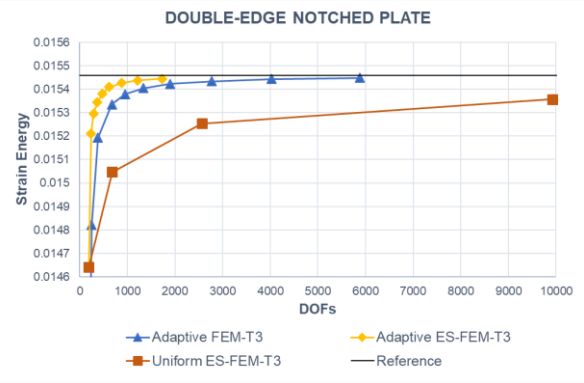



**Figure 12. Example 2: Convergence of strain energy results for various automatic adaptive mesh algorithms**

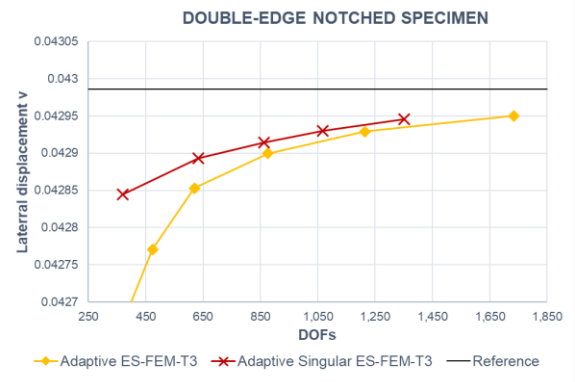



**Figure 13. Example 2: Convergence of lateral displacement results from adaptive ES-FEM-T3 and adaptive singular ES-FEM-T3**

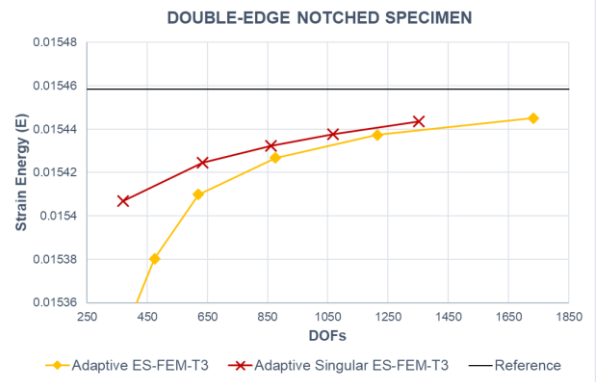



**Figure 14. Example 2: Convergence of strain energy results from adaptive ES-FEM-T3 and adaptive singular ES-FEM-T3**

After adopting a layer of singular five-node elements around the crack-tip, the results obtained from the adaptive singular ES-FEM-T3 converge with a more reasonable computing efforts (DOFs) in terms of both the lateral displacements v and strain energy solutions (as depicted in Figs. 13 & 14, respectively). In addition, both models incorporated the same automatic adaptive scheme adopting recovery-based strain error functions converge to the reference values.

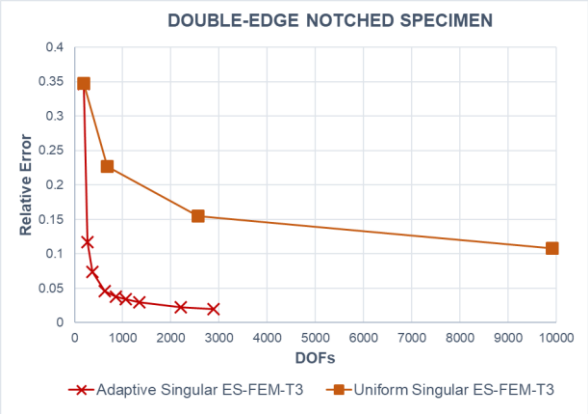



**Figure 15. Example 2: Convergence of relative error results from adaptive singular ES-FEM-T3 and uniform singular ES-FEM-T3**

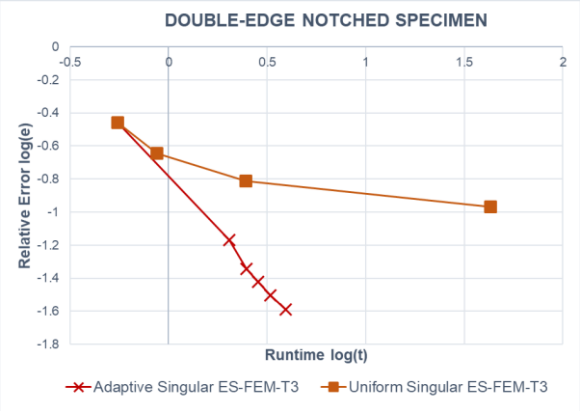
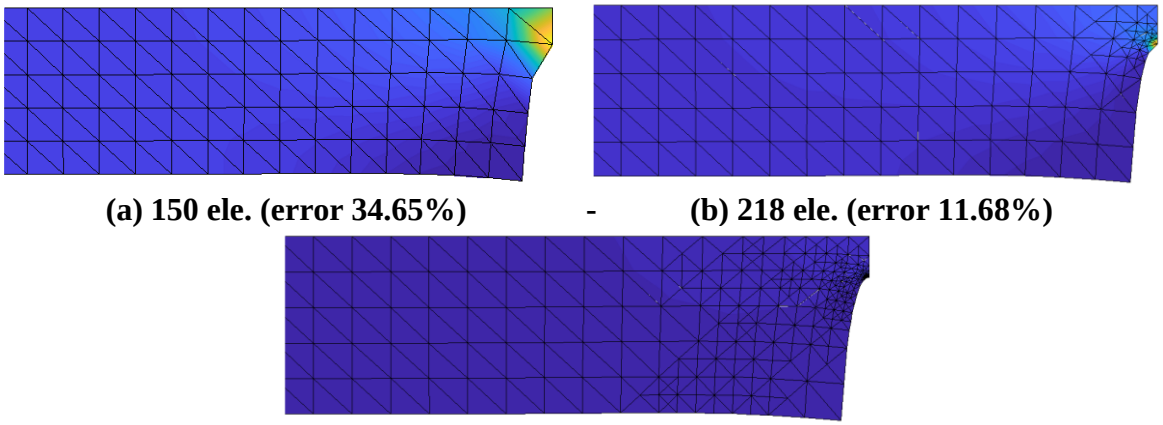



**Figure 16. Example 2: Convergence rate of runtime versus relative error between adaptive singular ES-FEM-T3 and uniform singular ES-FEM-T3.**

Within the framework of singular ES-FEM-T3, the present adaptive mesh implementation dramatically reduces relative error with less computing efforts (i.e. both DOFs and runtime as in Figs. 15 & 16) compared to standard model adopting uniform refinement strategy. Relative errors from adaptive singular ES-FEM-T3 navigates to the zero value, while those obtained from uniform singular ES-FEM-T3 are still at a very high value (i.e. less than 5% can be obtained with about 500 DOFs as shown in Fig. 15. It is clear that the convergence rate in a log scale equals 1:1 for adaptive singular ES-FEM-T3 and around 1:3 for uniform singular ES-FEM-T3 from Fig. 16.



**(a) 150 ele. (error 34.65%)** - **(b) 218 ele. (error 11.68%)**

**(c) 568 ele. (error 4.57%)**

**Figure 17. Example 2: Automatic adaptive meshes with the corresponding contour line of von Mises stress distributions.**

The mesh discretized patterns in Fig. 17 as expected progressively refined over the concentrated stress singularity and discontinuity areas.

## Conclusions

The novel ES-FEM-T3 and singular ES-FEM-T3 approaches employed automatic AMR algorithm to efficiently and accurately provide the response solution of elastic structures. The automatic AMR adopted the newest vertex bisection algorithm and recovery-based error function in L2-norm. A number of numerical examples (including both benchmarks and practical in-plane structures) were successfully solved using the proposed analysis scheme. Three of which are given in this study. These illustrate robustness of the proposed analysis method, in which the adaptive singular ES-FEM-T3 approach provided the superconvergence of elastic response solutions as compared to the other models for crack problems and the adaptive ES-FEM-T3 approach for general discontinuity problems (i.e. discontinuity applied load). The computed results agreed well with all reference values, and thus evidenced the computational advantages in yielding the close-to-exact solutions for modest computing resources.

A nontrivial extension of the proposed analysis framework is to apply into contact or nonlinear fracture problems (i.e. using cohesive fracture model). A new error estimation (i.e. the simple splitting the normal and singular parts of stress field, similar to [15]) that is suitable for the five-node singular elements is also our interest.

## Acknowledgement

The financial support from Chulalongkorn University through Applied Mechanics and Structure Research Unit is acknowledged. The first author would also like to thank the ASEAN Full-Scholarship.

### References

1. Liu, G., K. Dai, and T.T. Nguyen, *A smoothed finite element method for mechanics problems.* Computational Mechanics, 2007. **39**(6): p. 859-877.
2. Chen, J.S., et al., *A stabilized conforming nodal integration for Galerkin mesh-free methods.* International journal for numerical methods in engineering, 2001. **50**(2): p. 435-466.
3. Nourbakhshnia, N. and G. Liu, *A quasi-static crack growth simulation based on the singular ES-FEM.* International Journal for Numerical Methods in Engineering, 2011. **88**(5): p. 473-492.
4. Nguyen-Xuan, H., et al., *An adaptive singular ES-FEM for mechanics problems with singular field of arbitrary order.* Computer Methods in Applied Mechanics and Engineering, 2013. **253**: p. 252-273.
5. Nguyen-Thoi, T., et al., *Adaptive analysis using the node-based smoothed finite element method (NS-FEM).* International Journal for Numerical Methods in Biomedical Engineering, 2011. **27**(2): p. 198-218.
6. Tangaramvong, S., F. Tin-Loi, and C. Song, *A direct complementarity approach for the elastoplastic analysis of plane stress and plane strain structures.* International Journal for Numerical Methods in Engineering, 2012. **90**(7): p. 838-866.
7. Vicente da Silva, M. and A. Antao, *A non-linear programming method approach for upper bound limit analysis.* International Journal for Numerical Methods in Engineering, 2007. **72**(10): p. 1192-1218.
8. Szabó, B., B.A. Szabo, and I. Babuška, *Finite element analysis*. 1991: John Wiley & Sons.
9. Williams, M., *Stress singularities resulting from various boundary conditions in angular corners of plates in extension.* Journal of applied mechanics, 1952. **19**(4): p. 526-528.
10. Seweryn, A. and K. Molski, *Elastic stress singularities and corresponding generalized stress intensity factors for angular corners under various boundary conditions.* Engineering Fracture Mechanics, 1996. **55**(4): p. 529-556.
11. Zienkiewicz, O.C., et al., *The finite element method: solid mechanics*. Vol. 2. 2000: Butterworth-heinemann.
12. Dörfler, W., *A convergent adaptive algorithm for Poisson's equation.* SIAM Journal on Numerical Analysis, 1996. **33**(3): p. 1106-1124.
13. Rivara, M.C., *Algorithms for refining triangular grids suitable for adaptive and multigrid techniques.* International journal for numerical methods in Engineering, 1984. **20**(4): p. 745-756.
14. CHEN, L. *Short implementation of bisection in MATLAB*. in *Recent Advances In Computational Sciences: Selected Papers from the International Workshop on Computational Sciences and Its Education*. 2008. World Scientific.
15. Ródenas, J.J., et al., *A recovery-type error estimator for the extended finite element method based on singular+ smooth stress field splitting.* International Journal for Numerical Methods in Engineering, 2008. **76**(4): p. 545-571.

# Heat transfer in periodic laminated layer – Robin boundary conditions

***Ewelina Pazera and Piotr Ostrowski**

Department of Structural Mechanics, Lodz University of Technology, Poland

*Presenting and corresponding author: ewelina.pazera@p.lodz.pl

## Abstract

In this note, the problem of heat conduction in periodic laminated layer is considered. This layer is characterized by a microstructured composition and the microstructure is realized as a uniform distribution of the cells. The Robin boundary conditions, which are analyzed in this work, are combined with the convective heat exchange and there is an analytical solution for a homogeneous layer and this type of boundary conditions. To consider the heat conduction issue in presented laminated layer the tolerance averaging technique is used. The equations, obtained by using this technique, are solved by using finite difference method. As the results, the distributions of the temperature are obtained. The algorithm, which is created to obtain the distribution of the temperature can be verified by using the results from the analytical solution for a homogeneous structure and the Robin boundary conditions.

**Keywords:** Heat conduction, Robin boundary conditions, mathematical modelling, composite, microstructure

## Introduction

The two-dimensional issue in periodic laminated layer is considered in this work. Every cell of this layer is made of two different materials and the proportion between the first and the second material in the cell is constant. The thickness of the cells is also constant and denoted by $\Delta$, what is shown in Fig. 1.

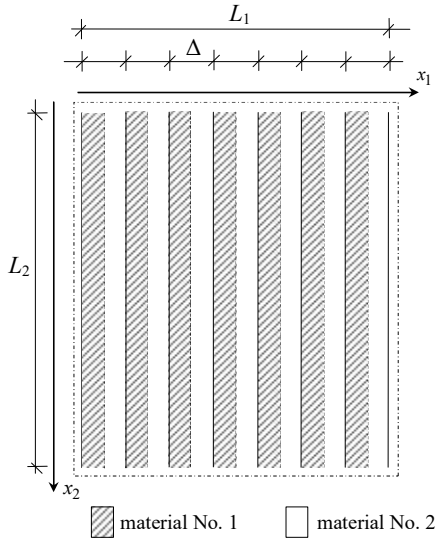



**Figure 1. The cross-section of considered layer**

The various issues related to this type of structures are considered in relation to micromechanical models with idealized geometry.

The Robin boundary conditions are analyzed in this note, there is an analytical solution for a homogeneous layer and this type of boundary conditions [13] and it is possible to obtain the distribution of the temperature according to the Eq. (1):

$$\theta = 2\sum_{n=1}^{\infty} \frac{\left(h^2+\alpha_n^2\right)\cos\left(\alpha_n x\right)\left\{\alpha_n \cosh\left(\alpha_n\left(L_2-y\right)\right)+h\sinh\left(\alpha_n\left(L_2-y\right)\right)\right\}}{\left[\left(h^2+\alpha_n^2\right)L_1+h\right]\left\{\alpha_n\cosh\left(\alpha_n L_2\right)+h\sinh\left(\alpha_n L_2\right)\right\}} \int_0^{L_1} f(x)\cos\left(\alpha_n x\right)dx, \quad (1)$$

where $L_1$, $L_2$ are dimensions along directions $x_1$, $x_2$, $h$ is the quotient of heat transfer coefficient and the thermal conductivity, and $\alpha_n$ are the solutions of the Eq. (2):

$$\alpha \cdot \mathrm{tg}\left(\alpha L_1\right) = h. \quad (2)$$

The solution is limited to the finite number $n$ equals 20 and shown in Fig. 2. As a material and geometry, it was assumed $L_1$=$L_2$=1 [m] and steel.

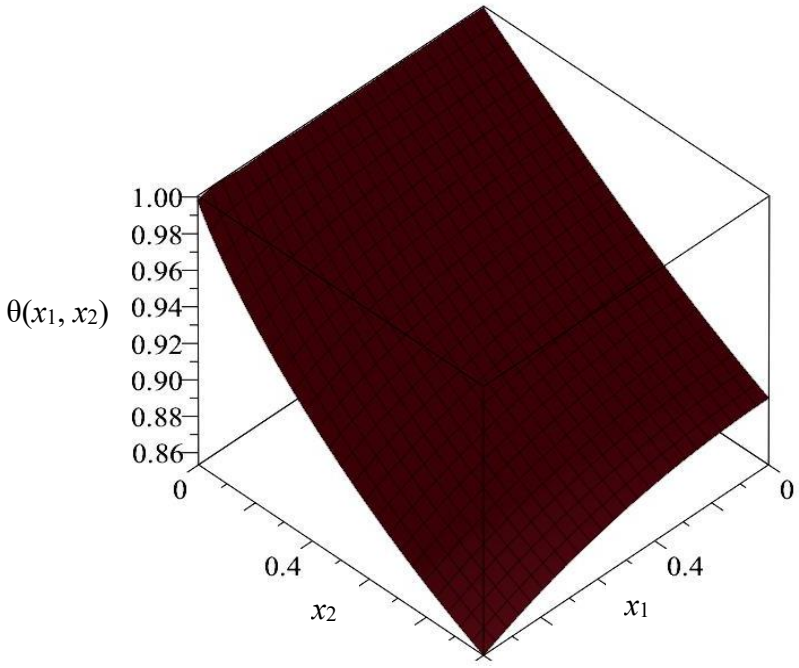



**Figure 2. The temperature for homogeneous layer**

To analyze the laminated layer, where the distribution function of material properties is periodic, the tolerance averaging technique is used [13]. This technique gives us a possibility to take into account the effect of the microstructure size. The tolerance modelling is expanded and applied in many publications to analyze various issues concerning both periodic and functionally graded structures. Among them are thermal issues [13]-**Błąd! Nie można odnaleźć źródła odwołania.** and dynamic problems [13]-**Błąd! Nie można odnaleźć źródła odwołania.**.

The main aim of this work is to obtain the equations of the tolerance model with the macrotemperature and the fluctuations amplitudes of the temperature as unknowns.

## Modelling procedures

The stationary heat conduction issue for laminated layer can be described by Eq. (3):

$$\nabla \cdot \left(\mathbf{K} \cdot \nabla\theta\right) = 0, \quad (3)$$

where **K** is a tensor of conductivity, wherein components are denoted by $k_{ij}$.

The main aim of the application of the tolerance modelling is to replace the system of differential equations (3) with non-continuous coefficients, by equations, where the coefficients are slowly-varying. The basic assumption of the tolerance modelling is the micro-macro decomposition, where the temperature θ (the main unknown) can be expressed as a sum of the averaged part ϑ (the macrotemperature) and the oscillating part, according to the Eq. (4):

$$\theta(x_1,x_2) = \vartheta(x_1,x_2) + g(x_1)\psi(x_1,x_2). \tag{4}$$

On the other hand the oscillating part can be defined as a product of the know fluctuation shape function $g$ and the fluctuation amplitudes of the temperature ψ (the new basic unknown). In this work the fluctuation shape function is assumed as a saw-type function. The second assumption of the tolerance modelling is the periodic approximation of some derivatives of function of the temperature, where some terms can be treated as negligibly small. Additionally, the tolerance averaging technique introduces some new concepts, among them the tolerance-periodic and slowly-varying function.

By using the micro-macro decomposition to the Eq. (3), using the orthogonalisation method, formulating the residuum function of temperature and the condition, which have to be fulfilled by this function, by doing appropriate averaging and transformations, the equations of the tolerance model for considered laminated layer are obtained in the form of Eqs (5):

$$\begin{aligned} &\nabla\cdot\left(\langle\mathbf{K}\rangle\cdot\nabla\vartheta + \langle\mathbf{K}\cdot\partial g\rangle\psi\right) = 0, \\ &\langle\partial g\cdot\mathbf{K}\cdot\partial g\rangle\psi + \langle\mathbf{K}\cdot\partial g\rangle\cdot\nabla\vartheta = 0. \end{aligned} \tag{5}$$

**Example**

Let $L_1$=$L_2$=1 [m]. The problem under consideration was a stationary heat conduction issue for laminated layer characterized by periodic structure of size Δ=$L_1$/20. For both sublayers the material properties (steel and aluminum) were defined and the constant distribution function of material properties was assumed ($v_1$=0.5). Based on Eqs (5) and by using the assumption of the asymmetrical character of the fluctuation shape function, the equations of the tolerance model for considered issue are in the form of Eqs (6):

$$\begin{aligned} &\partial_1\left(\langle k_{11}\rangle\partial_1\vartheta + \langle k_{11}\partial g\rangle\psi\right) + \partial_2\left(\langle k_{22}\rangle\partial_2\vartheta\right) = 0, \\ &\partial_2\left(\langle k_{22}gg\rangle\partial_2\psi\right) - \langle k_{11}\partial g\rangle\partial_1\vartheta - \langle k_{11}\partial g\partial g\rangle\psi = 0. \end{aligned} \tag{6}$$

The boundary conditions were assumed as follows: known temperature on the upper edge of the laminate ($\theta\big|_{x_2=0} = \theta_u$), thermally isolated left edge ($q_1\big|_{x_1=0} = q_t \Rightarrow \dfrac{\partial\vartheta}{\partial x_1}\Big|_{x_1=0} = 0$), the Robin boundary conditions on the right edge ($x_1$=$L_1$) according to the Eq. (7) and on the bottom edge ($x_2$=$L_2$) according to the Eq. (8):

$$-k_{11}\frac{\partial\vartheta}{\partial x_1} = H\left(\vartheta - \vartheta_e\right), \tag{7}$$

$$-k_{22}\frac{\partial\vartheta}{\partial x_2} = H\left(\vartheta - \vartheta_e\right), \tag{8}$$

where $\vartheta_e$ is the external temperature and $H$ is the heat transfer coefficient. In this note the external temperature is assumed to be equal zero. Then the boundary conditions for the fluctuation amplitudes of the temperature were assumed as: the known fluctuation amplitudes on the upper edge ($\psi\big|_{x_2=0}=0$), on the left edge ($\psi\big|_{x_1=0}=0$), on the right edge ($\psi\big|_{x_1=L_1}=0$) and the term on the bottom edge ($x_2$=$L_2$) following Eq. (8):

$$\frac{\partial\psi}{\partial x_2}+\frac{\langle Hgg\rangle}{\langle k_{22}gg\rangle}\cdot\psi=0. \tag{8}$$

To solve the equations of the tolerance model (Eqs (6)), the finite difference method was used. Along both directions (parallel and perpendicular to the laminas) the grid nodes distribution was uniform. By using this method the set of non-homogeneous discretized equations was obtained with the macro-temperature and the fluctuation amplitudes of the temperature as unknowns in the form of Eq. (9):

$$K\cdot X=Q, \tag{9}$$

where $K$ is a matrix of coefficients, $X$ is a vector of unknowns ranked alternately at individual points, and $Q$ is a vector of free terms.

The results were shown in Fig. 3 in the form of plots of the total temperature.

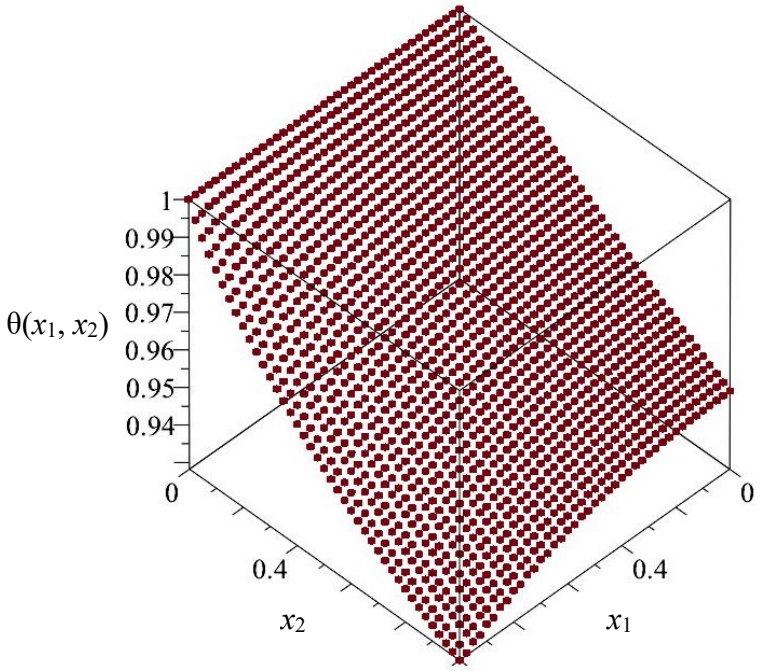



**Figure 3. The total temperature**

## Conclusions

By using the tolerance averaging technique it is possible to replace the system of differential equations with non-continuous coefficients, by the equations where the coefficients are constant or slowly-varying. By using the equations of the tolerance model it is possible to take into account the effect of the microstructure size in thermal problems.

### References

[1] Carslaw, H.S. and Jaeger, J.C. (1959) Conduction of heat in solids, At the Clarendon Press, Oxford.

[2] Woźniak, Cz., Michalak, B. and Jędrysiak, J. (2008) Thermomechanics of heterogeneous solids and structures. Tolerance Averaging Approach, Publishing House of Lodz University of Technology, Lodz.
[3] Ostrowski, P. and Michalak, B. (2015) The combined asymptotic-tolerance model of heat conduction in a skeletal micro-heterogeneous hollow cylinder, *Composite Structures* **134**, 343–352.
[4] Ostrowski, P. and Michalak, B. (2016) A contribution to the modelling of heat conduction for cylindrical composite conductors with non-uniform distribution of constituents, *International Journal of Heat and Mass Transfer* **92**, 435–448.
[5] Pazera, E. and Jędrysiak, J. (2015) Thermoelastic phenomena in transversally graded laminates, *Composite Structures* **134**, 663–671.
[6] Pazera, E. and Jędrysiak, J. (2018) Effect of microstructure in thermoelasticity problems of functionally graded laminates, *Composite Structures* **202**, 296–303.
[7] Pazera, E. and Jędrysiak, J. (2018) Thermomechanical analysis of functionally graded laminates using tolerance approach, *AIP Conference Proceedings* **1922**, 140001.
[8] Domagalski, J. and Jędrysiak, J. (2016) Geometrically nonlinear vibrations of slender meso-periodic beams. The tolerance modeling approach, *Composite Structures* **136**, 270–277.
[9] Jędrysiak, J. and Pazera, E. (2016) Vibrations of non-periodic thermoelastic laminates, *Vibrations in Physical Systems* **27**, 175–180.
[10] Jędrysiak, J. (2017) Tolerance modelling of free vibration frequencies of thin functionally graded plates with one-directional microstructure, *Composite Structures* **161**, 453–468.
[11] Jędrysiak, J., Domagalski, Ł., Marczak, J. and Pazera, E. (2018) Tolerance modelling of nonstationary problems of microheterogeneous media and structures, *Vibrations in Physical Systems* **29**, 2018001.
[12] Marczak, J. and Jędrysiak, J. (2015) Tolerance modelling of vibrations of periodic three-layered plates with inert core, *Composite Structures* **134**, 854–861.
[13] Pazera, E. and Jędrysiak, J. (2018) Effect of temperature on vibrations of laminated layer, *Vibrations in Physical Systems* **29**, 2018032.
[14] Tomczyk, B. and Szczerba, P. (2017) Tolerance modelling of vibrations of periodic three-layered plates with inert core, *Composite Structures* **162**, 365–373.

# European tall building with a height of 200 m and an irregular form – aerodynamic analysis methods

**†Joanna Pietrzak[1] and Wiesław Rokicki[2]**

[1]Department of Structure Design, Construction and Technical Infrastructure, Faculty of Architecture, Warsaw University of Technology, Poland.
[2]Department of Structure Design, Construction and Technical Infrastructure, Faculty of Architecture, Warsaw University of Technology, Poland.

†Corresponding author: joanna.pietrzak@pw.edu.pl

## Abstract

A tall building with an irregular form requires optimisation due to the impact of wind loads. The aerodynamic analyses should be carried out at an early concept stage. The decision which wind analysis method to choose is a key issue. This issue will be discussed on the example of an irregularly shaped skyscraper in Warsaw. The aim of this research was to determine and compare the results of three methods. The three methods were: the analytical analysis according to the European building codes (Eurocode 1), wind tunnel tests (a simulation in an aerodynamic tunnel at the Institute of Aeronautics and Applied Mechanics of the Warsaw University of Technology) and Computational Fluid Dynamics (numerical wind tunnel flow simulation in FLUENT). The possibility of examining the air flow around the building, wind pressure distribution on facades and net force values were analysed.

The methods differ in their accuracy and types of the obtained results. The Eurocode noticeably lacks methods for tackling irregular forms. Because the influence of the form of the building is treated very generally and the aerodynamic interference with the surrounding buildings is not taken into account, the results from the Eurocode calculations are characterized by a large safety factor. The results of a wind tunnel test are much more accurate. Values of pressures, forces and moments can be measured, but the presentation of the results requires statistical and/or graphic processing. Currently, it is the only method combining the accuracy and reliability of the obtained results. However, the precise, time-consuming tunnel tests should be conducted on the final form as the final verification of the adopted architectural and structural solutions. At the concept stage the architects need tools to quickly estimate the air flow and wind effects on the building. Computer simulations are easier and cheaper to conduct than tunnel tests. Moreover, the results of computational simulations can become the basis for understanding the air flow around the building. This method would be used more frequently if it did not require verification of its results.

**Keywords:** skyscraper, tall building, wind influence, Eurocode, wind tunnel testing, computational fluid dynamics

## Introduction

In the context of obtaining an original architectural form and reducing construction costs the possibility of eliminating the adverse impact of wind raises investors' and designers' interest in wind engineering. This field which is particularly important in the case of tall buildings. It should be noted, however, that the influence of wind is strongly conditioned not only by the height of the building but also its shape, and the surrounding buildings. The unique shape and the vicinity of each skyscraper, especially the proximity of dense urban fabric, is associated with the need to conduct precise aerodynamic analyses [9].

The forms of contemporary European high-rise buildings become more varied. It happens so, among others, due to the resignation from simple forms, characterized by regular, repetitive floors. An effective form of a skyscraper requires the cohesion of a functional and spatial solution. As a basis for the design of a skyscraper, one should strive to limit stresses in structural elements [10],

and in skyscrapers the magnitude and distribution of forces are affected to a large extent by its form [2]. The interaction between the wind and form determines the air flow and distribution of the wind pressure on facades. Optimizing the shape of a skyscraper may lead among others to i.e. an optimization of the load-bearing structure or the technical solution of the facade. Aerodynamic optimization analyses at the stage of developing the architectural concept are of particular importance in achieving more effective and economical spatial solutions. Aerodynamic optimization helps to reduce adverse impact of wind in the context of obtaining more rational design of load-bearing structures and the reduction of construction costs. A balanced approach to design, and often above all economic conditions, raise interest in reducing the adverse impact of wind.

In order to optimize the form and structure of a skyscraper, it is necessary to look for tools that allow for an accurate understanding of the wind effects. With the buildings forms becoming more complex and irregular the ability to accurately calculate and predict aerodynamic phenomena relies on the chosen method of aerodynamic analysis. In Europe, for buildings taller than 200 m, wind tunnel tests should be recommended. Designer can also chose to use computational methods, that are developing dynamically. When designing buildings of less than 200 meters (which account for over 96% of European tall buildings), wind loads can be calculated according to design standards described in Eurocode 1. There is a lack of scientific studies which would analyse both the effects of calculating the wind loads in accordance with Eurocode 1 and on the basis of the results obtained in the tunnel test and numerical simulations.

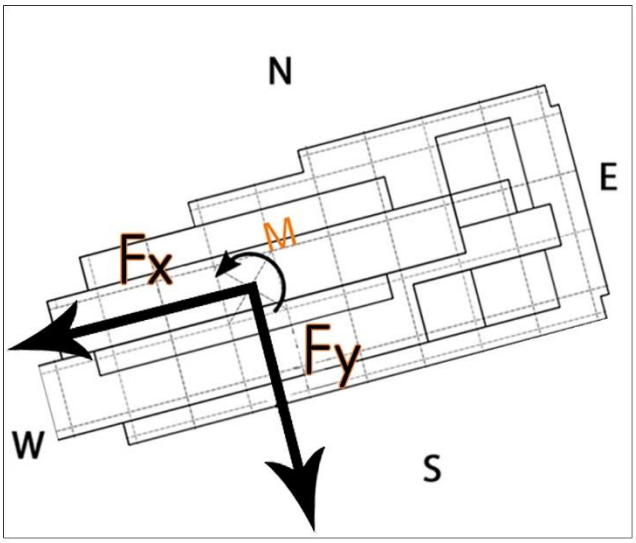



**Figure 1. Analysed building - relation to the directions of the world and the coordinates system**

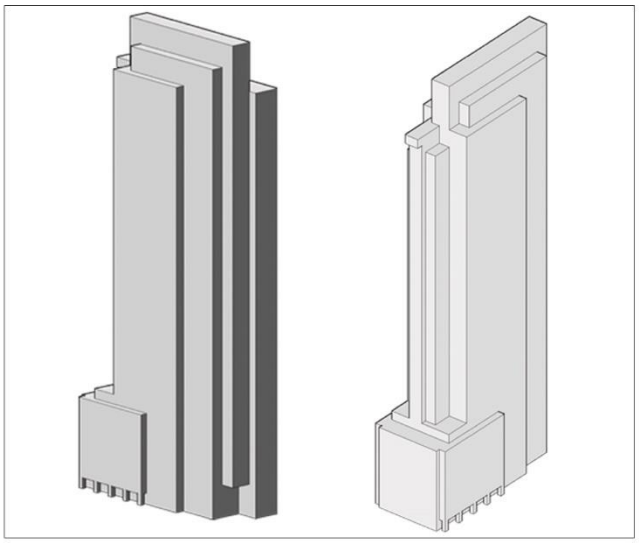


**Figure 2. Model of the analysed building**

## Subject of the study

In particular, the forms which cause unfavourable aerodynamic phenomena, such as the non-aerodynamic, asymmetric, irregular buildings, should be carefully analysed[11]. Detailed analyses are also required for objects located in an urban space, in a context causing difficult to predict, variable and asymmetrical wind effects. This is evident in the example of the analysed building - a skyscraper constructed in the centre of Warsaw with a height of almost 200 m and an irregular, asymmetrical shape and geometry based on a right angles. The skyscraper was shaped from slender solids with different heights and widths (Fig. 2). The planned ground floor area is over 2,500 $m^2$. Functionally and spatially the building has been divided into: a base part with a height of 10 floors

(41 m) and a dominant with a height of 36 floors. The shape of the floor plan is similar to an elongated rectangle with the proportions of sides 1:3.5 (ca. 85 x 35 m in the base part, 65 x 30 m in the tower part).

The analysed skyscraper is located in the centre of Warsaw, in the intensive building zone. In the surroundings of the building there are tall office buildings, residential buildings and low shopping centres. From the north-east side (Fig. 1), a complex of multi-storey buildings is located, from the east there are mainly low and medium-rise buildings. On the south side there are low-rise buildings. On the north side there is a high-rise building complex, and another one is built from the west.

The research compared the results obtained for 10 selected floors and all facades [12][14]. The results presented in the article were limited to 4 selected floors (some of them were given only for the 25$^{th}$ floor) and the southern façade (Fig. 3).

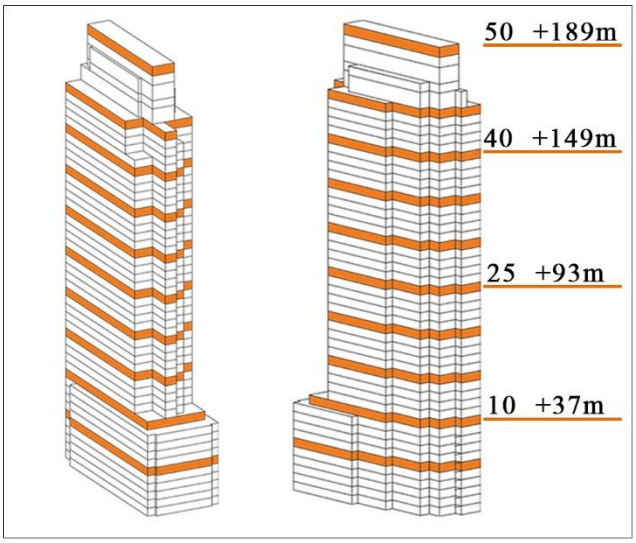



**Figure 3. Floors selected for the analyses and presented in the article**

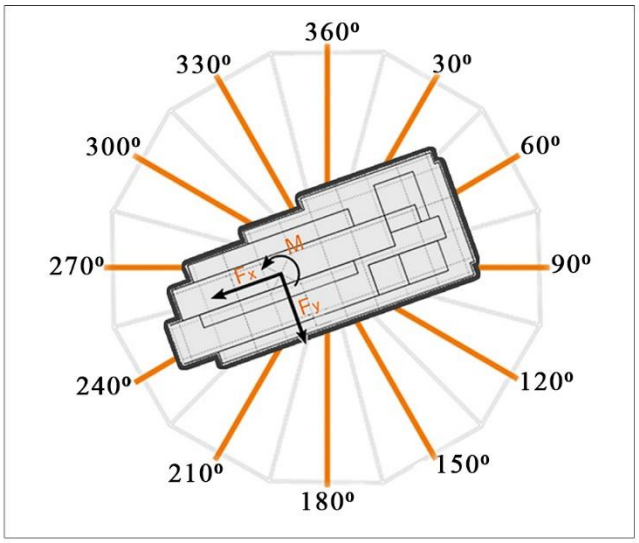



**Figure 4. Analysed wind directions**

## Study Description

The task of wind engineering is to provide methods and tools for testing the wind-form-structure relationships and to find the criteria for deciding which procedure to choose [9]. Currently, engineers use three methods, and the purpose of this research was to compare them. The first method – based on design codes and standards, uses analytical methods in accordance with applicable regulations, recommendations or other similar documents. The second - empirical (experimental) method includes simulations conducted in a wind tunnel. The third method - a fully computational analysis, uses known mathematical models to define the impact of wind [7]. This group comprises primarily digital computational analyses.

According to European design standards analyses have been carried out for 12 wind directions (Fig. 4). The presented results of wind impact on the given object are: qualitative comparison of pressure distributions on facades (for 3 methods), peak pressures (calculated using code procedure and measured in a tunnel), comparison net forces transferred to the structure on the selected floor's levels. The possibility of analysing the air flow around the building was also examined.

## Methods

*PN-EN 1991-1-4: 2008 Eurocode 1*

The first analysis included the analysis of the wind loads according to PN-EN 1991-1-4:2008 Eurocode 1 [13][17][18]. The code procedure did not take into account the detailed configuration of the surrounding buildings, only a very generally defined class of terrain. In accordance with the national annex to the Eurocode, 1st wind zone and terrain category IV were assumed. In all methods the characteristics of the atmospheric boundary layer were reproduced by adopting the standard mean velocity and turbulence intensity profiles . In order to accurately represent the actual structure of the wind at the ground-level, the results of climate analysis developed at the Faculty of Power and Aeronautical Engineering, WUT were used.

**Figure 5. Model of the analysed building and the surrounding in the wind tunnel**

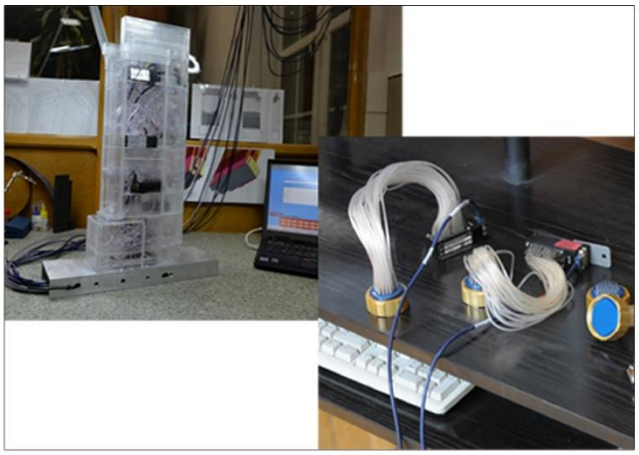


**Figure 6. Pressure sensors – installed inside the model and connected to quick-release couplings**

*Wind tunnel testing*

The empirical method consisted of experimental research in the wind tunnel at the Institute of Aeronautics and Applied Mechanics of the Warsaw University of Technology. The wind tunnel is a closed-loop tunnel measuring 2.60 x 2.25 x 11.00 m. Passive methods were used to map the atmospheric boundary layer characteristic in the tunnel. The analysed building together with the neighbouring buildings (within a radius of 500 m) was mapped in a 1:350 scale (Fig. 5). The model contained all the designed tall buildings, concepts of which were known at the time of the experiment. The tests were carried out using rigid models of buildings which allow the measurement of wind pressure on walls by means of pressure sensors (Fig. 6) and measurements of resultant forces and aerodynamic moments using aerodynamic balance [7]. The signals from the measurements were subjected to numerical processing, the aim of which was to obtain pressure and force values on the entire surface of the model.

*Computational Fluid Dynamics*

The computational method included simulation performed in the Ansys Fluent program, which is used by engineers at the Institute of Aeronautics and Applied Mechanics of the Warsaw University of Technology and is widely described in the scientific literature as a tool to analyse the impact of wind on cuboid high-rise buildings with proportions enabling precise determination of wall boundary layer separation points (e.g. [5][6]). The simulation was carried out according to the recommendations of [1][3][4][8][13][15][16].

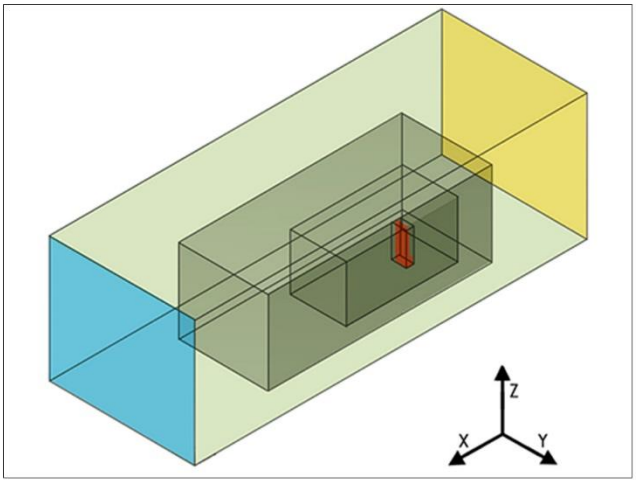



**Figure 7. The computational domain - boundary conditions (description in the article)**

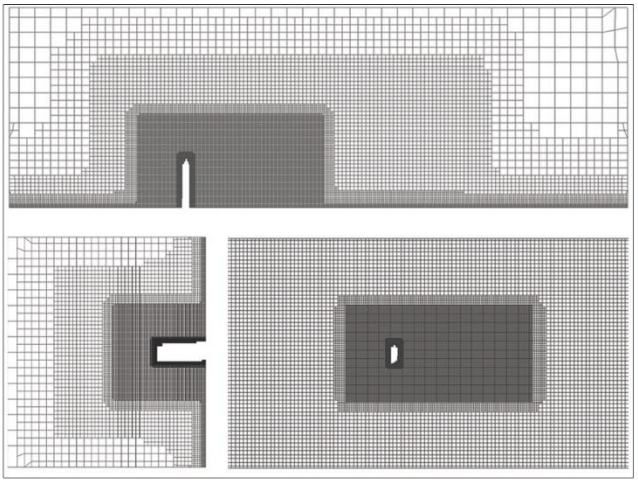


**Figure 8. The structural grid generated to discretize the computational domain**

The goal of the computational analysis was to recreate the conditions of the wind tunnel experiment. The computing domain simulated the dimensions of the actual tunnel. The boundary conditions had been assigned to the appropriate surfaces limiting the computational domain (Fig. 7): inflow (the plane marked in yellow) and outflow (blue). The remaining boundaries of the domain (green) have been given the boundary condition of the wall *without slipping*. The grid compaction areas (grey) were modelled as surfaces fully permeable to the fluid (*internal* condition) (Fig. 7 and 8). The vertical profiles of mean wind velocity and turbulence intensity were set at the inflow. To obtain the results more similar to the results of the experiment, it was decided to choose the improved $k - \varepsilon$ *realizable* model from the RANS group [1][16]. A standard wall function was used to model the boundary layer. The flow was modelled as laminar. A *pressure based solver* was used. The *SIMPLE* algorithm, based on the *segregated* method, was used to solve the equations describing the flow. The finite volume method was used to discretize the model. *Standard* method recommended by Fluent producers was used for interpolation of pressure. The *upwind* method was used to discretize equations: moments, kinetic energy of turbulence and turbulence kinetic energy dissipation. The course of residual values was monitored until their convergence reached the value of $1e^{-5}$.

In order to check the numerical calculations the global net forces acting on the analysed model were compared with the forces measured in tunnel test. The comparison of the results showed their certain convergence. The results obtained in the computational analysis reflected the nature of the occurring aerodynamic phenomena and were then used for qualitative analyses. However, they were not included in more detailed quantitative analyses.

## Results

*Wind pressure distributions on facades*

The first analysis was aimed at comparing the results obtained with the 3 methods (Fig. 9). Due to the limitation of the impact of the surroundings, the analysis was limited to the comparison of pressure distributions on the south facade with wind from the same direction. The results obtained in the Eurocode analysis are characterized by significant inaccuracy. However, after adopting appropriate assumptions, in a tunnel test and in a computational analysis one can get similar precision of the results, which can then be the basis for a detailed optimization of a complex building. The numerical calculations have been quite accurate as to the reproduction of the qualitative nature of the phenomenon. The zones of pressure and suction on facades and their

changes have a similar distribution. One can also notice how the results obtained from the norms are simplified when compared to the exact simulation of real conditions. The differences in pressure distribution resulting from irregularities of the form are not visible in the design code results. Moreover, the changes in pressure values are a very big simplification in relation to reality.

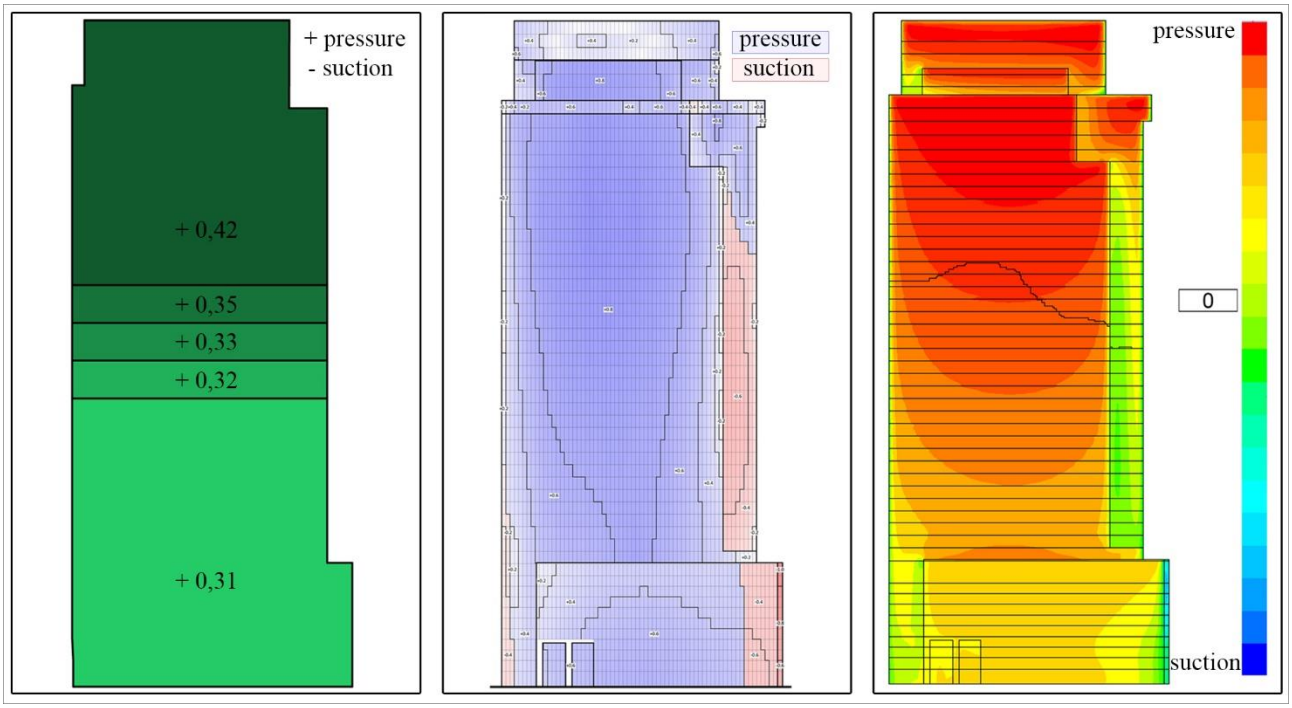



**Figure 9. Maps of the wind pressure distributions on the southern façade, obtained according to: Eurocode procedure, wind tunnel testing and computational simulations. 160° wind direction**

*Peak pressure envelopes*

Since similar results are obtained both in the analytical method and tunnel tests, further analyses have been limited to the comparison of the methods giving extreme results, i.e. the Eurocode and the wind tunnel testing.

First, the peak pressures envelopes for selected floors were analysed (Fig. 10). It can be observed that the effect of suction is particularly important. Practically all the corners achieve much higher values than the flat sections. For short sections of the façades, the analytical method does not reflect differences in the suction volume. The distribution of suction force for long facades is also different. In terms of pressure, according to analytical method its values remain the same along the length of the facade, while in the tunnel results there is significant differentiation.

*Peak pressure*

For long facades – the northern and southern (Fig. 11) one the peak pressure values obtained from the calculations are even twice as large as those measured in the wind tunnel (Table 1 and 3). In the case of the shorter (eastern and western) facades and corners (Fig. 12) the results are similar (Table 2 and 4). Strict norms regarding the edges and corners of the building have been confirmed experimentally.

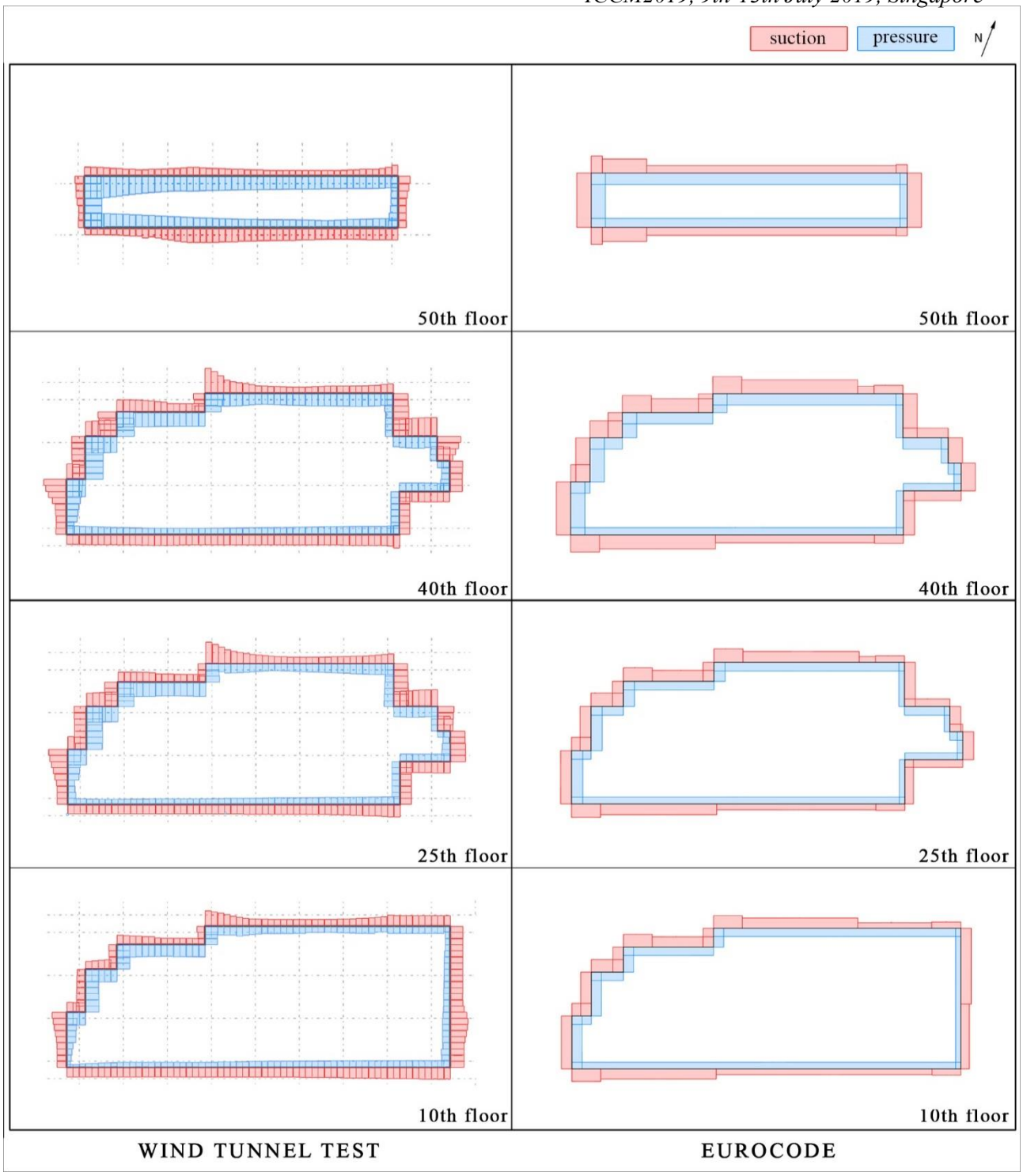



**Figure 10. Peak pressure envelopes according to wind tunnel tests results and Eurocode procedure obtained for floors: 10th, 25th, 40th and 50th**

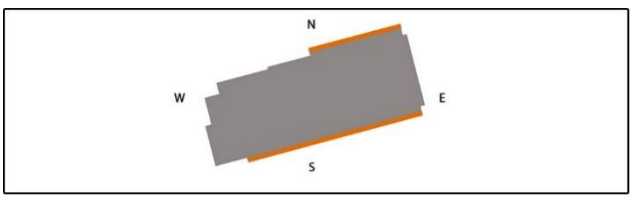



**Figure 11. Facades for which the results are presented in Tables 1 and 3.**

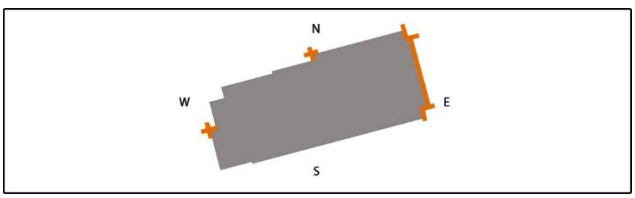



**Figure 12. Facades and corners for which the results are presented in Tables 2 and 4**

**Table 1. Peak pressure values for the northern facade**

| | Pressure [kPa] | | Suction [kPa] | |
|---|---|---|---|---|
| | **Euro code** | **Tunnel Test** | **Euro code** | **Tunnel Test** |
| **10th floor** | 0,83 | 0,37 - 0,53 | 0,55 - 1,02 | 0,41 - 0,60 |
| **25th floor** | 0,85 | 0,40 – 0,62 | 0,57 – 1,05 | 0,37 – 0,73 |
| **40th floor** | 1,12 | 0,64 – 0,75 | 0,74 - 1,38 | 0,38 – 0,70 |
| **50th floor** | 1,12 | 0,61 – 1,00 | 0,74 - 1,38 | 0,27 – 0,44 |

**Table 2. Peak pressure values for the western facade**

| | Pressure [kPa] | | Suction [kPa] | |
|---|---|---|---|---|
| | **Euro code** | **Tunnel Test** | **Euro code** | **Tunnel Test** |
| **10th floor** | 0,50 | 0,27 - 0,50 | 0,83 – 1,02 | 0,73 – 1,03 |
| **25th floor** | 0,52 | 0,35 – 0,55 | 0,83 – 1,05 | 0,80 – 1,00 |
| **40th floor** | 0,68 | 0,39 – 0,54 | 1,09 – 1,38 | 0,39 – 0,54 |
| **50th floor** | 0,68 | 0,40 – 0,50 | 1,38 | 0,35 – 0,66 |

**Table 3. Peak pressure values for the southern facade**

| | Pressure [kPa] | | Suction [kPa] | |
|---|---|---|---|---|
| | **Euro code** | **Tunnel Test** | **Euro code** | **Tunnel Test** |
| **10th floor** | 0,65 | 0,20 – 0,39 | 0,57 - 1,02 | 0,55 – 0,69 |
| **25th floor** | 0,67 | 0,29 – 0,44 | 0,59 – 1,05 | 0,54 – 0,70 |
| **40th floor** | 0,88 | 0,36 – 0,48 | 0,78 - 1,38 | 0,60 - 0,65 |
| **50th floor** | 0,88 | 0,35 – 0,66 | 0,78 - 1,38 | 0,36 – 0,70 |

**Table 4. Peak pressure values for the chosen corners on the 25$^{th}$ floor**

| | Pressure [kPa] | |
|---|---|---|
| | **Eurocode** | **Tunnel Test** |
| **W corner** | 1,05 | 1,09 |
| **N corner** | 1,29 | 1,27 |
| **N-E corner** | 0,63 | 0,58 |
| **S-E corner** | 0,63 | 0,73 |

*Net force values*

The impact of the selected test method on the design of an irregular structure of the building is better reflected by the analysis of the forces acting on the load-bearing structure. They can be considered, among others in the form of net forces e.g. separately for each floor. The Tables 5 and 6 present the components $F_x$ and $F_y$ of the net forces for the selected floors. The observed differences in the results are significant. Usually the results obtained from the standard calculations are 2-3 times larger than those measured in the tunnel (e.g. $F_y$ component force for the 340° direction), and in extreme cases the differences are 7-10 fold (70° direction - $F_x$ component force for the 50$^{th}$ floor and $F_y$ component force for 10$^{th}$ floor) and even the forces have an opposite vector direction (component force $F_x$ for the 340° direction). The analyses confirm the hypothesis that the results obtained with different methods are divergent and the choice of method has a significant impact on the adopted technical solutions.

**Table 5. The components Fx and Fy of the net forces for the selected floors - 340° wind direction**

| 340° DIRECTION ( N ) | Fx [kN] | | Fy [kN] | |
|---|---|---|---|---|
| | Euro-code | Tunnel Test | Euro-code | Tunnel Test |
| **10th floor** | -98,06 | 13,80 | 330,25 | 175,80 |
| **25th floor** | -72,88 | 26,50 | 245,46 | 75,90 |
| **40th floor** | -66,43 | 15,30 | 223,72 | 74,80 |
| **50th floor** | -70,79 | -4,90 | 211,93 | 66,10 |

**Table 6. The components Fx and Fy of the net forces for the selected floors - 70° wind direction**

| 70° DIRECTION ( E ) | Fx [kN] | | Fy [kN] | |
|---|---|---|---|---|
| | Euro-code | Tunnel Test | Euro-code | Tunnel Test |
| **10th floor** | 75,25 | 26,30 | -106,58 | -10,20 |
| **25th floor** | 55,93 | 16,00 | -79,21 | -11,30 |
| **40th floor** | 50,98 | 15,10 | -72,20 | -25,50 |
| **50th floor** | 43,62 | 6,80 | -68,17 | -28,50 |

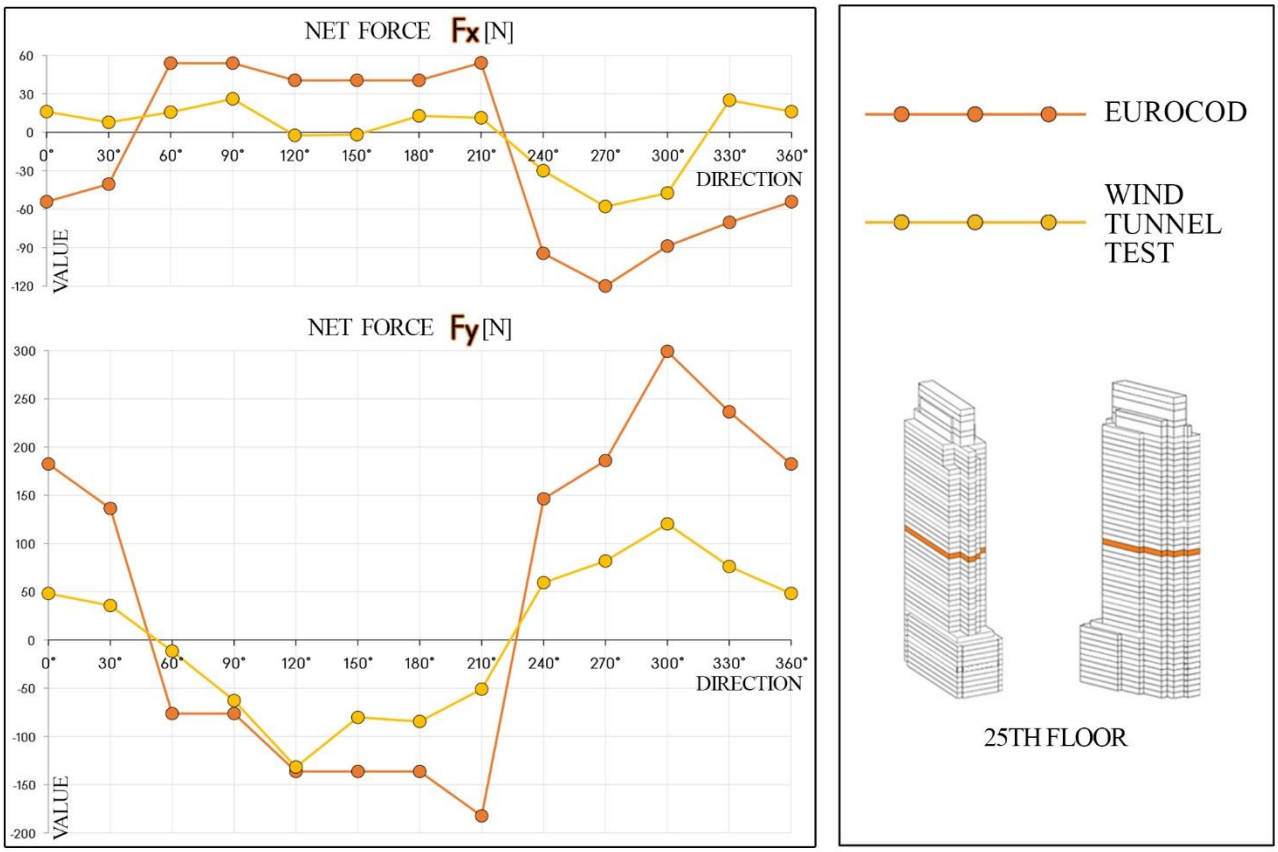



**Figure 13. The components Fx and Fy of the net forces for the 25th floor, depending on wind direction**

*The impact of the surrounding*

Analysing the chart of the component net forces values, e.g. for 25$^{th}$ floor, the influence of the surroundings on the obtained results can be noticed (Fig. 13). The most varied are the results for the northern (300°, 330° and 0°/360°) and western (210°, 240° and 270°) directions, which results, among others, from close proximity to high-rise buildings. On the other hand, the smaller

differences of resultant forces for the south-eastern wind directions (120°, 150° and 180°) result from the lack of significantly tall objects in the immediate vicinity, however, setting the building perpendicular to the wind direction results in some variation of the obtained results. For the winds from the east (60° and 90° directions), the results are even more convergent. There are no tall objects from this side, and the wind flow is parallel to the longer side of the analysed building.

There are also some similarities between results from both methods. One can notice a difference in the stiffness of a building with a plan similar to an elongated rectangle - the components $F_y$ (parallel to the shorter side) reach much higher values than the components $F_x$ (parallel to the long side).

*Air flow around the building*

Detailed results of pressure and forces values do not always show what they result from. To explain the obtained results it is important to understand the aerodynamic phenomena occurring around the building. The precise data on this subject is provided only by digital computational methods. In addition, this methods allow for quick changes and analysis of many variants. Due to their graphical form (Fig. 14) of presentation they can be more understandable and useful for architects. The obtained results can be used for general optimization of the building form at the conceptual stage.

Figure 15 shows wind speed distributions obtained in the Fluent program. On the presented visualizations one can notice an increase in the value of the wind velocity vector as a function of height, differences resulting from: the lack of symmetry of the building (flow around the model, the shape of the wake), the irregularity of the form (the boundary layer separation point and turbulences), the changes of wind direction.

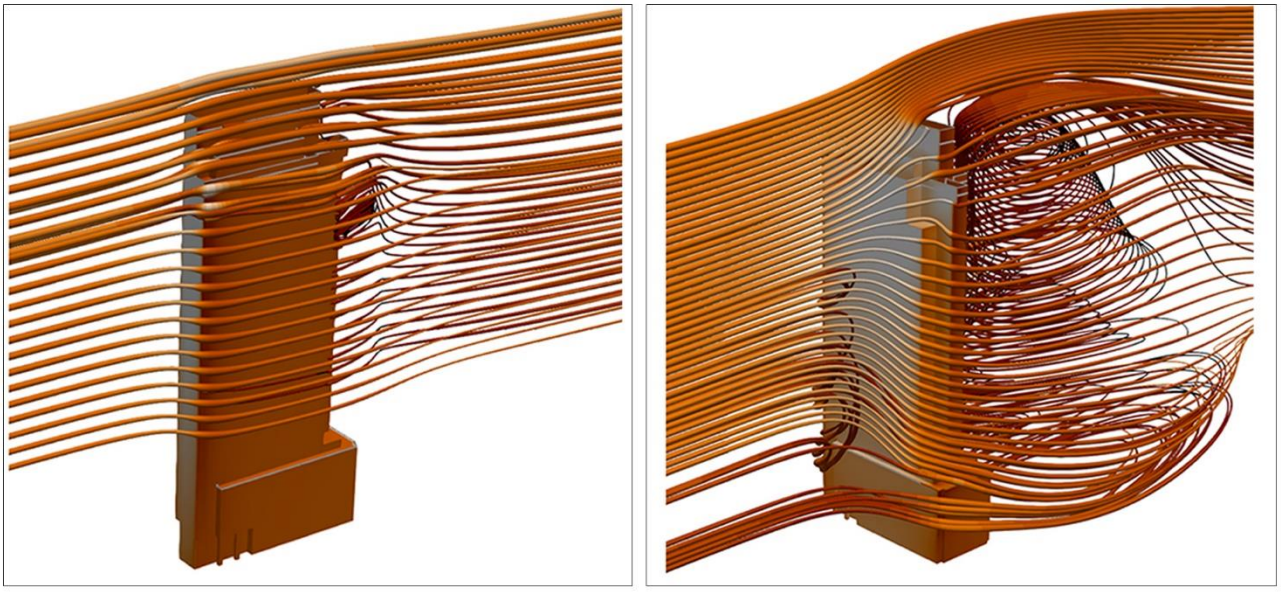


**Figure 14. Air flow around the building depending on the wind direction**

**Summary**

The analysed methods differ in accuracy and types of obtained results.

*PN-EN 1991-1-4: 2008 Eurocode 1*

The calculations according to design codes do not require access to a laboratory, special software or a lot of time, and financial expenses. However, analytical procedures have been developed to

analyse simple, basic geometries, and the results obtained in them are not always reliable. In the standard calculations, the effects of the wind on the building are examined, not the character of the flow. The influence of the surrounding is not taken into account and the results obtained lead to a too large safety factors being incorporated. A simplified representation of the wind's influence as to the value and spatial distribution also results in imprecision and overestimation of the calculated quantities.

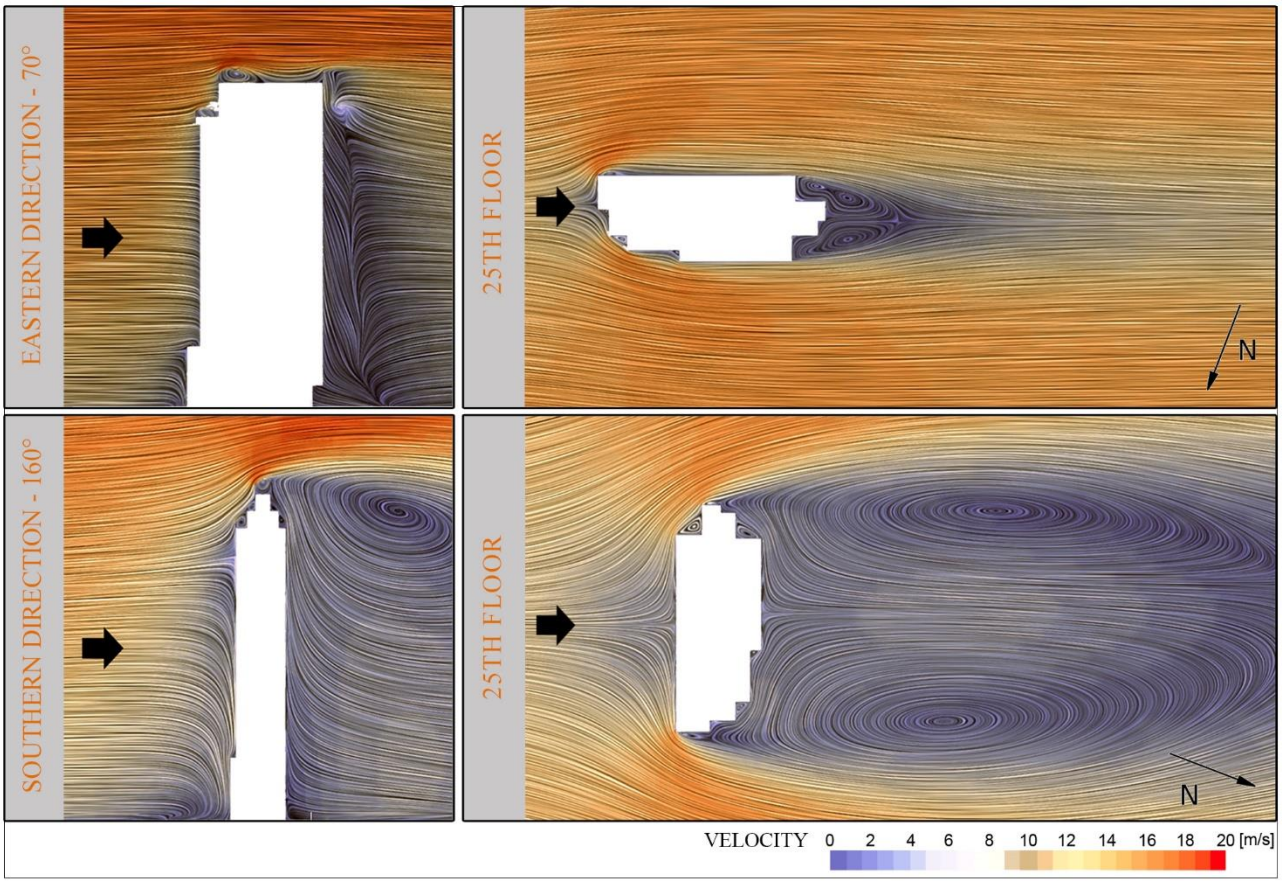



**Figure 15. Wind velocity map obtained from computational analyses in Ansys Fluent program in the planes of cross-sections and 25$^{th}$ floor - wind direction parallel to the plane mapped**

*Wind tunnel testing*

The wind tunnel testing allows to study objects with an unusual geometry located in a complex environment. Accuracy of the results obtained in the experiment allows for a relatively precise determination of pressure distribution and calculation of resultant forces transmitted to the structure. However, the measurements refer only to discrete points, usually located on the surface of the building model. Presentation of the results requires statistical and/or graphical processing, e.g. the obtained results can be interpolated to the distributions on whole facades. With a sufficiently large number of measurement points, the result can be very precise. Using the aerodynamic balance, one can measure the magnitudes of resultant forces and moments that accurately reflect reality. However, this method is time-consuming and requires large financial expenses.

*Computational Fluid Dynamics*

Less time-consuming and cheaper analysis can be carried out using computational methods. The CFD numeric programs used in these methods, such as Ansys Fluent, are used to reproduce real conditions or wind tunnel tests. The results of the computational simulation help the designers

understand the nature of the flow in the entire domain, as well as to simulate the measurements performed using the aerodynamic balance. Thanks to the flow simulations, it is possible to collect detailed data impossible to measure in the tunnel and to visualize the occurring phenomena in a relatively simple way. This is especially important in enabling the designers to understand the qualitative aspect of the occurring phenomena. In order to perform quantitative analyses, it is necessary to generate data in a tabular form and then subject it to processing.

Computer simulations also enable relatively fast variant testing and would probably be used much more often if it was not necessary to verify the assumptions and the results. Appropriate definition of boundary conditions and adoption of preliminary assumptions are generally based on the experience of aerodynamic experts. Checking the correctness of the results is problematic if we do not compare them with measurements in reality or in a tunnel simulation. In addition, an appropriate tool for the analyzed geometry should be selected, because the turbulent flow models used in CFD programs have been calibrated to a certain type of task, e.g. Fluent is dedicated to simulating the air flow around cuboid bodies.

## Conclusions

Optimization of the form of a tall building due to the impact of wind requires as accurate as possible recognition of the magnitude of the loads. The key issue here is the choice of the wind analysis method, because the results return different values.

Although the wind standards in Europe can be applied to buildings with a height of up to 200 m, the procedures described are sufficient only for the calculation of wind loads for a simple object with a regular shape. For buildings with a complex, irregular geometry, the standard procedures do not specify a more precise procedure. Because the Eurocode methods do not analyse numerous factors as accurately, the safety factors are far larger than necessary. It should be noted, however, that design in accordance with the Eurocode standards is a safe procedure, and accepting loads lower than standard provisions should always be justified by detailed analyses.

Precise determination of loads for irregularly shaped buildings located in the vicinity of tall buildings becomes possible after tunnel tests or numerical simulations using various CFD programs (after verifying the reliability of results).

Currently, tunnel tests are the only method that combines the accuracy and reliability of the results obtained. However, taking into account scientific and technical progress, we can expect further development of computational methods in the field of CFD. Thanks to the simulation of the wind flow the designers can collect detailed data on the flow of air masses around a building, impossible to measure in the tunnel test. Also when comparing different concepts CFD allows to effectively re-examine the modified model. Accurate understanding of the nature of the phenomena occurring around the designed skyscraper allows for effective optimization of its form and supporting structure. Then by interpreting the results graphically they can be more easily understood by designers.

Optimization of an irregular form due to wind loads requires a very accurate recognition of aerodynamic interactions. Aerodynamic analysis should be considered at the conceptual stage when changes in the geometry of the building are possible. The results of standard calculations are so imprecise that they do not constitute a good basis for spatial and structural optimization. At this stage, complicated, expensive and long-lasting tunnel tests will also be of limited use. Only the digital tools for computational analysis can be used to quickly estimate the basic flow characteristics. Simple analyses of the flow around a building do not require interdisciplinary cooperation if the architect has the right tools which can support the design process. The obtained

results may even be slightly inaccurate. Precise model tests may be used to obtain the skyscraper's final form and constitute the final verification of the adopted architectural and structural solutions.

**References**


[1] Bardina, J., Huang, P. and Coakley, T. (1997). Turbulence modeling validation, testing and development. *NASA Technical Memorandum, 110446*. Retrieved from: ntrs.nasa.gov/search. jsp?R=19970017828.

[2] Billington, D. (1983). *The tower and the bridge: The new art of structural engineering*, Princeton University Press, Princeton.

[3] Blocken, B., Stathopoulos, T. and Carmeliet, J. (2007) CFD simulation of the atmospheric boundary layer: wall function problems. *Atmospheric Environment*, *41*(2), 2007, 238-252.

[4] Blocken, B., Janssen, W. and Hooff, T., van (2012). CFD simulation for pedestrian wind comfort and wind safety in urban areas: General decision framework and case study for the Eindhoven University campus. *Environmental Modelling & Software*, 30, 15-34.

[5] Boda, D. and Banda, L. (2011). *Wind effects on typical tall structures*. Saarbrücken: Lambert Academic Publishing.

[6] Chakraborty, S. and Dalui, S. (2015). *Wind effects on "+" plan shaped buildings. An experimental and numerical investigation.* Saarbrücken: Lambert Academic Publishing.

[7] Flaga, A. (2008). *Inżynieria wiatrowa. Podstway i zastosowania.* Warszawa: Arkady.

[8] Franke, J., Hellsten, A., Schlünzen, H. and Carissimo, B. (ed.). (2007). *Best practice guideline for the CFD simulation of flows in the urban environment.* Brussels: COST office.

[9] Irwin, P., Denoon, R. and Scott, D. (2013). *Wind Tunnel Testing of High-Rise Buildings: An output of the CTBUH Wind Engineering Working Group*. Chicago: Council on Tall Buildings and Urban Habitat, Routledge / Taylor and Francis Group.

[10] Kahn, F. R. (1982). The rise and fall of structural logic in architecture. *Chicago Architectural Journal*, 2.

[11] Lawson, T. (2001). *Building aerodynamics*. Londyn: Imperial College Press.

[12] Pietrzak, J., Rutkowski M. and Wrona M. (2017). *Analizy normowe oraz badania modelowe oddziaływania wiatru na budynek wysoki o* nieregularnej formie na przykładzie projektowanego wieżowca w Warszawie. Unpublished manuscript, Politechnika Warszawska, Wydział Architektury, Warszawa.

[13] PN-EN 1991-1-4:2008 Eurokod 1: oddziaływania na konstrukcje. Część 1-4, Oddziaływania ogólne: oddziaływania wiatru.

[14] Rokicki, W., Pietrzak, J. and Wrona, M. (2018). Wpływ obciążeń od wiatru na budynek wysoki o nieregularnej formie - badania modelowe. *Przestrzeń, Ekonomia, Społeczeństwo, 12*(2), 179-192.

[15] Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M. and Shirasawa, T. (2008). AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10-11), 1749-1761.

[16] Wilcox, D. C. (1998). *Turbulence Modeling for CFD. La* Canada, CA: DCW Industries

[17] Żurański, J. A. (2005). *Wpływ warunków klimatycznych i terenowych na obciążenie wiatrem konstrukcji budowlanych*. Warszawa: ITB.

[18] Żurański, J. A. and Gaczek, M. (2011). *Oddziaływania klimatyczne na konstrukcje budowlane według Eurokodu 1. Komentarz z przykładami obliczeń.* Warszawa: Instytut Techniki Budowlanej.

# Structural Design of Nonuniform Metasurface to Mimic the Deformation Behavior of Human Skin at Highly-Stretched Joint Area

**†*Yafeng Han[1], Changmeng Liu[1], Shuyuan Ma[1], Jiping Lu[1], Ying Liu[1], and Wenfeng Lu[2]**

[1]School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

[2]Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, 117975, Singapore

†Corresponding author & *Presenting author: hanyafeng@bit.edu.cn

## Abstract

Developing new generation nonuniform metasurface with skin-like stretchability and conformability has become the research hotspot in many areas, such as bio-sensoring, medical bandages, wearable devices and soft robotics. In this study, a computational method was proposed to design nonuniform 2D metasurface, which could perform a similar deformation as the target skin surface during the rotation of joint. Considering the nonuniform auxetic deformation behavior of the skin around joint area, the designed metasurface should also have the same Poisson's ratio (PR) distribution. By mapping unit cells with different PR to the target surface, the obtained structure would have the same mechanical property and deformation behavior as the covered skin. With the unique capabilities of additive manufacturing (AM), the generated 2D metasurface can be easily fabricated. Especially, new generation wearable electronics with superior conformability was developed based on this nonuniform structure. Besides, the proposed method is also promising to be applied to design novel biomaterials, such as wearable electronics, smart bandage, skin scaffold, *etc.*

**Keywords:** Nonuniform metasurface, Poisson's ratio distribution, nonuniform auxetic deformation, additive manufacturing

## 1. Introduction

Thin film structures are used in diverse technological applications such as stretchable electronics, soft robotics, smart bandages, wearable devices and living devices. In all these areas, achieving a "human-skin-like" deformation behaviour is the main design objective of the manufactured film structure. Mainly, researches have tried to solve this problem from two aspects: soft materials or stretchable structures.

To obtain better comfortability, many researchers in stretchable electronics area have dedicated to develop new materials with better stretchability. As early as 1994, Garnier, Hajlaoui, Yassar and Srivastava [1] had fabricated a field-effect transistor from organic materials with printing technology. This flexible device can be bended or twisted without causing much difference on the conductivity. From then on, organic electronics have proven useful in a number of applications, such as light patterning techniques and organic semiconductors [2]. The strain sustainability of organic electronics, despite much better than common wafer-based devices, is still not enough to conform the complex 3D surface of human body [3].

Changing the structure of the film metasurface is another way to achieve better stretchability. The most commonly applied strategy is converting the traditional all solid surface into "wavy"

structures [4]. In the research [5] of Kim et al., several different circuit designs were introduced to construct the "epidermal electronic system". While, all these circuit patterns were developed based on the significant stretchability of "wavy" structure. Someya et al. [6] manufactured a networks of pressure and thermal sensors with conductive rubber. Similarly, the impressive skin-like conformability and flexibility was gained from the "wavy" layout of the material. Besides, many fabric-based electronics [7,8] also have shown good conductivity during deformation. The reason is that the conductive wires enabled in fabrics are always at a loose- knit state, which is still a nonconstant "wavy" structure essentially. To eliminate this deformation property difference between substrate and conductive wires, Jeong, Kim, Cho and Hong [4] introduced the "wavy" structure to the construction of both substrate and electrode. This wavy substrate was fabricated through casting, and base material was chosen as PDMS for better flexibility. Consequently, the deformation behaviour of both substrate and silver electrodes can be maintained as uniform, which will dramatically decrease the sheer stress at the connection areas. Similarly, a spiral shape wiring system was proposed in the study of Sawada et cl. [9], and both substrate and conductive wires were manufactured into same structure. The only difference is that the spiral shaped substrate was made of photosensitive polyimide and exposed with copper foil. Resistance of the constructed stretchable wiring network changed little, even after 200 cycles of 20% stretching.

Compared with stretchability, conformability is a more important property for wearable devices or medical bandages. Especially, some joint areas of human body have very complex 3D surfaces, and skin around these areas always bear extremely high stretching. Besides, the skin at joint area is always unevenly stretched at all directions, which means this is a nonuniform auxetic deformation. To mimic such a unique deformation behaviour, negative Poisson's ratio (NPR) structures became necessary to construct the suitable wearable electronics [10]. Yang, Choi and Kamien [11] utilized fractal cuts to develop super conformable materials that could be applied to design stretchable electronics. After the proposed fractal cut process, a common silicone rubber sheet can be easily stretched and conformed to a spherical surface. In Vogiatzis, Ma, Chen and Gu's study [12], topology optimization method was applied to design NPR structures with optimized conformability. The optimized 2D metasurface can conform to the complex surface of a human face, which provided great potential for the structural design of wearable electronics. However, all these conformable NPR structures were uniformly constructed. Considering the nonuniform auxetic deformation of the skin, there will beinevitable deformation behaviour difference between the thin film and human skin. This difference will definitely cause sheer stress at the connection area between substrate and skin surface, which will reduce the stability of the device and may even cause detachment of the wearable electronic. To make sure the designed metasurface have the same deformation behaviour as the target skin surface, substrate structures with nonuniform PR distribution [13– 15] can be introduced.

In this study, a novel method was proposed to design and manufacture thin film structures with customized deformation behaviour, which can conform with the highly-stretched skin surface at human joint areas. By analysing the deformation behaviour of the skin during the rotation of joint, elongation (along the rotation surface) and expandation (vertical to the rotation surface) values of each small subsurface were measured. With this deformation data of all subsurfaces, strain and PR distribution of the target skin were calculated. To mimic the nonuniform PR distribution of the skin, the structure of substrate was organized by connecting unit cells with different PR values. Tuneable PR of unit cell was obtained by changing the interior angle of re- entrant honeycomb structure, which was still a mutation of "wavy" structures. Skin around a person-specific elbow was chosen as the target surface. After deformation analysis and structural construction, finite element analysis (FEA) was

conducted to validate the conformability of the designed nonuniform substrate. With PolyJet printed flexible nonuniform metasurface, stretching test was carried out to measure deformation behaviour, and both computational and experimental tests have shown a perfect conformability.

## 2. Materials and methods

### 2.1 Materials

To enhance the stretchability and flexibility of the elastomer film, rubber-like material Agilus30 was selected. This newly developed material is one of the most flexible materials available for Objet Connex 260 (Stratasys Inc., Edina, MN, USA). As tested in [16], this material can be assumed as linear elastic when the applied strain is smaller than 100%, and the material Young's modulus is 423MPa and Poisson's ratio is 0.34.

### 2.2 General processes of the proposed method

In this study, a structural design method was proposed to develop metasurface structure with nonuniform PR distribution to conform with skin surfaces around highly stretched joint areas. Particularly, the method was mainly composed of six steps as shown in Figure 1. Firstly, 3D scanning was conducted to get the point cloud data of the skin surface around joint. The joint surface was scanned at two positions: fully-straightened and fully-bended. The fully-straightened position was set as initial state with assumed strain and deformation of zero. On the other hand, the fully-bended position was the objective state. From the initial state to the objective state, the deformation behaviour of the skin can be analysed. The skin surface was descriptive into finite subsurfaces to analyse how the strain and deformation was distributed over the surface area. By setting as joint bending direction the applied strain direction, PR of each subsurface could be calculated easily. To mimic the nonuniform auxetic deformation of the skin, the PR distribution in the substrate should also be nonuniform correspondingly. As been done in our previous research [17], re-entrant honeycomb structure was introduced to achieve tuneable PR by manipulating the interior angle of each unit cell. The size of unit cell was defined based the geometry of subsurface to make sure that each subsurface was covered by one unit cell. Considering the relative large deformation, geometrically nonlinear FEA was applied to get the relationship between unit cell's PR and the interior angle. With this relationship, Unit cells with the same PR values were mapped to cover the all the subsurfaces of the skin. Both computational and experimental tests were conducted to validate the conformability of the design substrate structure. Two comparison structures, uniform NPR anduniform PPR, were also tested to demonstrate whether the generated nonuniform structure had better conformability.

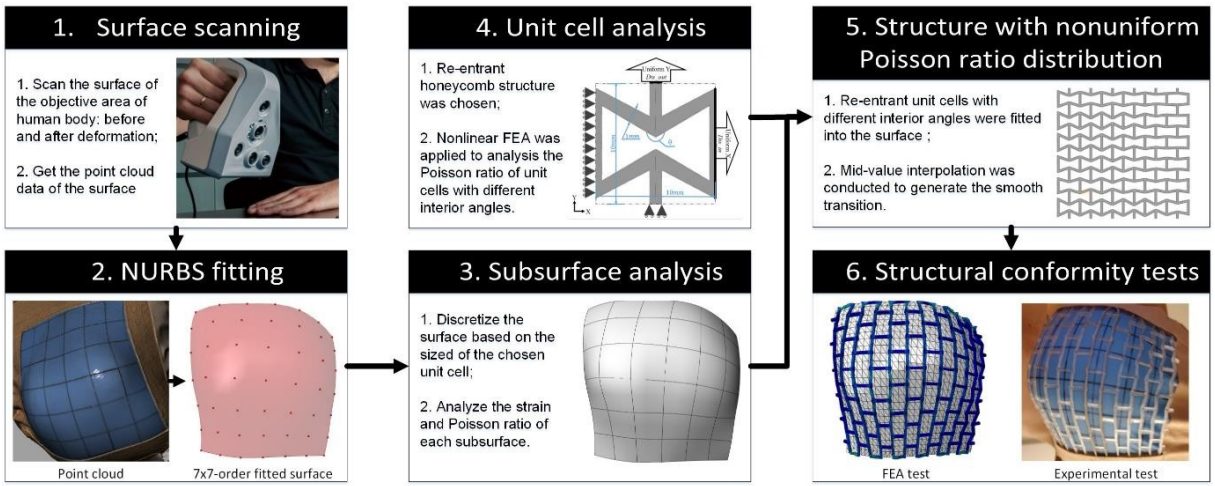



**Figure 1. General process of the proposed design and manufacturing method**

### 2.3 Skin surface scanning and analysis

Commonly, human skin is in the form of complex 3D surface. It is extremely challenging to model the such a complex geometry manually, especially with the consideration of the slight difference between each individual. To customize the metasurface structure for each person-specific design, 3D scanning was utilized in this study to collect the geomaterial data of the target skin surface. An ArtecTM Space Spider (Artec Inc., Luxembourg) was applied to conduct the 3D scanning of the skin surface. This portable/handheld scanner has outstanding capabilities to render complex geometry, sharp edges and thin ribs.

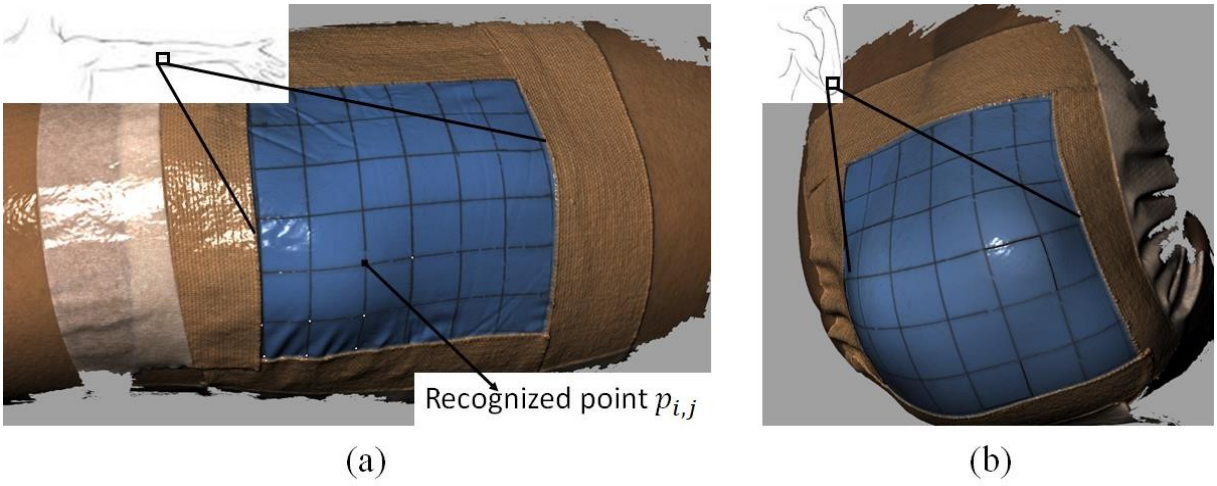



**Figure 2. Scanning results of the skin surface around a person-specific joint area: (a) straightened state, (b) bended state.**

Figure 2 shows the scanning results of the skin surface around a person-specific joint area. For the convenience of feature recognition, a blue rubber film with marked uniform grid was attached to the target area. Considering the difficulty of analysing the point cloud data, curvilinear surface fitting was necessary. Nonuniform Rational B-splines (NURBS) are mathematical models usually used in computer graphics [18], which was chosen to mathematically represent the scanned surface. However, it's not possible to fit NURBS surface to the whole point cloud directly because of the massive separated point data and requirement of NURBS surface approximation. To make the surface approximation more efficiently computer aided feature recognition was introduced to find the coordinates of all intersection points "+" in the scanning results, as shown in Figure 2a. By fitting a NURBS

surface that crossed all the recognized points, geometrically analysis could be processed. Detailedly, for a 3D NURBS with the order of $(p, q)$, the mathematical definition is given in Equation (1).

$$S(u,v) = \frac{\sum_{x=0}^{n}\sum_{y=0}^{m} N_{x,p}(u)N_{y,q}(v)w_{x,y}C_{x,y}}{\sum_{x=0}^{n}\sum_{y=0}^{m} N_{x,p}(u)N_{y,q}(v)\,w_{x,y}}$$
$$0 < u, v < 1 \qquad (1)$$
$$p \leq n; q \leq m$$

where, $n$ and $m$ denote the number of control points at direction of $u$ and $v$. To construct a NURBS surface with the order of $(p, q)$, the number of control points must be larger or equals to the order. $N_{x,p}$ are the non-rational B-spline basis functions defined on the knot vectors. $C_{x,y}$ is the coordinate of control point, and $w_{x,y}$ is the weight of the point. In this paper, to obtain a linear solution, all the weights were set to 1. With the coordinates of all recognized points $P_{i,j}$, the NURBS surface can be approximated in the least-squares sense:

$$\min \sum_{i=1}^{k}\sum_{j=1}^{l}(P_{i,j} - S(u_i, v_j)) \qquad (2)$$

where, $k$ and $l$ are the number of recognized points on the surface at the direction of $u$ and $v$ respectively. With a total number of fitting points equals $k \times$ , a 3D NURBS was approximately constructed. Figure 3a shows a 7 × 7 order NURBS that fitted from 7 × 7 points on the scanned skin surface of the bended elbow (Figure 2b). The fitted NURBS had a significant advantage that it can be evenly subdivided into any number of subsurfaces, with mathematical interpolating process. As demonstrated in Figure 3b & 3c, the fitted NURBS was divided into smaller subsurfaces and each subsurface still could represent the deformed geometry of a smaller skin area.

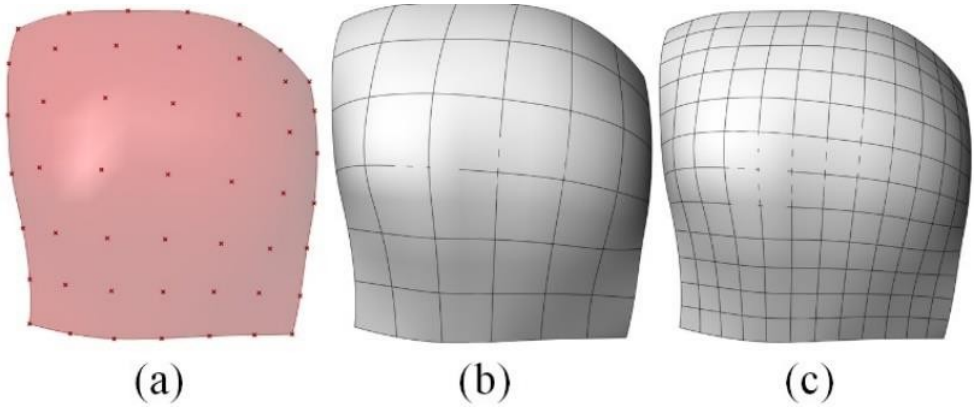



**Figure 3. NUMBS fitting and dividing. (a) a NUMBS fitted from 7 × 7 points on the surface. (b) the fitted NUMBS was divided into 6 × 6 subsurfaces. (c) the fitted NUMBS was divided into 12 × 12 subsurfaces.**

To analysis the deformation behaviour of the target skin area, the straightened joint was set as the initial state (Figure 2a). As this state, the skin was assumed as non-stretched and the side-lengths of all square subsurfaces were kept as uniform with the value of $l_{int}$, as shown in Figure 4a. The side-lengths of each square would keep changing along with the bending of the joint. When the maximum bending was applied as suggested in Figure 2b, the skin would also be stretched to the maximum extent. The side-lengths of each subsurface of the deformed skin were measured based on the interpolation of the NURBS. As marked in Figure 4b, the length

of each edge of subsurface $SubS_{w,z}$ became different to each other.

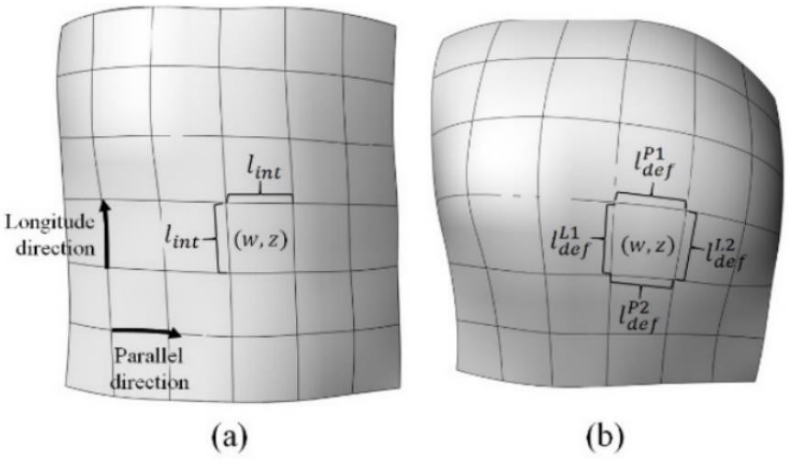



**Figure 4. Measuring the side-length of each subsurface: (a) initial state, (b) target state.**

The length change of each edge represented the elongation (along longitude direction) and expandation (along parallel direction) of the subsurface. To simplify the deformation analysis, the deformation of the subsurface was assumed as axisymmetric. Therefore, average elongation values were used to calculate the applied strain of each subsurface, as demonstrated in Equation (3). Besides, Poisson's ratio of the subsurface could also be derived from the side-lengths of the deformed subsurface (Figure 4b) with Equation (4).

$$\varepsilon_{w,z} = (l_{def}^{L1} + l_{def}^{L2} - 2l_{int}) \Big/ 2l_{int} \tag{3}$$

$$\nu_{w,z} = (l_{def}^{P1} + l_{def}^{P2} - 2l_{int}) \Big/ (l_{def}^{L1} + l_{def}^{L2} - 2l_{int}) \tag{4}$$

As shown in Figure 5, $(w,z)$ is the number of the analysed subsurface. $l_{int}$ denotes the initial length of all subsurface edges. Respectively, $l_{def}^{L}$ and $l_{def}^{P}$ represent the elongation and expandation the subsurface after the bending of joint. Even though the calculation of both strain and PR was based on the assumption of axisymmetric deformation, the influence could be ignored if the size of subsurface was small enough. By analysing all the subsurfaces one by one, the nonuniform strain and PR distribution of the target skin surface was demonstrated.

### 2.4 Unit cell analysis and structure generation

To mimic the nonuniform deformation behaviour of the skin around joint area, the designed metasurface should also possess a similar PR distribution to generate no sheer stress over the connection area of substrate and human skin. As the basic strategy of the proposed method, the nonuniform of PR distribution was achieved by connecting unit cells with different PR into an integrated structure. The PR of re-entrant honeycomb structure can be easily modified from negative to positive by just changing the interior angle [17]. To analyse the PR of unit cells with different interior angle $\theta$ and under different strain, geometrical nonlinear FEA was conducted. Agilus30 was selected as the base material, and the load and boundary conditions of the analysis is demonstrated in Figure 5. The unit cell can cover a square area with side-length of $lcell$ = $10mm$. By applying a $+X$ direction displacement $Dis\_in$ on the left boundary of the unit cell, the output displacement $Dis\_out$ was measured. The applied strain $\varepsilon cell$ and PR $\nu cell$ of the unit cell was calculated with Equation (5) & (6).

$$\varepsilon_{cell} = Dis\_in / l_{cell} \tag{5}$$

$$\nu_{ce\square\square l} = Dis\_out/is\_out \quad (6)$$

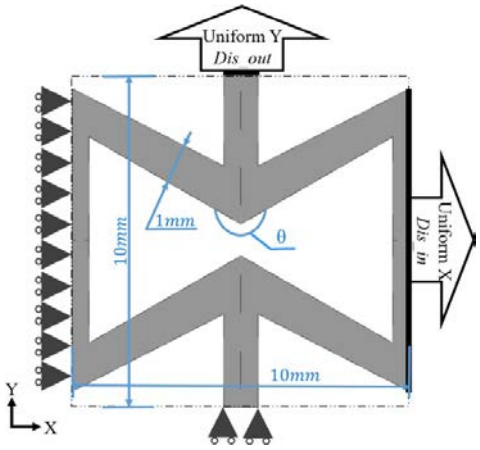



**Figure 5. Load and boundary conditions for the nonlinear FEA of unit cell**

Thirteen unit cells with different interior angle, form $\theta = 120°$ to $\theta = 240°$, were analysed. Because geometrical nonlinearity was considered, PR of each unit cell was different with the change of applied strain, as shown in Figure 7. With the summarized "ν-$\theta$-ε" relationship in Figure 7, the nonuniform substrate structure was generated by assigning the corresponding unit cell to mimic the properties of each subsurface, letting $\varepsilon_{cell} = \varepsilon_{w,z}$ and $\nu_{cell} = \nu_{w,z}$.

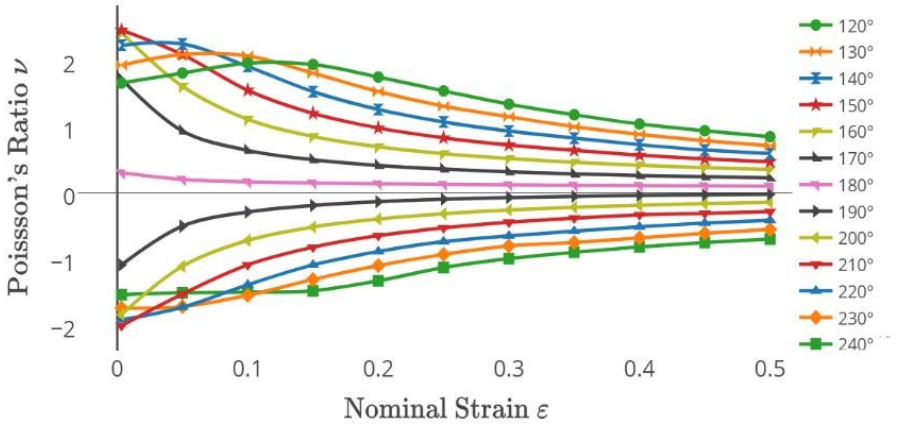



**Figure 6. The relationship between applied strain, PR value and unit cell's interior angel**

## 3. Numerical and experimental validation

To validate the proposed structural design method, a more challenging task, designing a nonuniform metasurface structure that suitable for a person-specific elbow, was introduced. The skin surface of the person-specific elbow is shown in Figure 2. A $60 \times 60mm^2$ area was chosen as the target surface to the attach the film structure. With the surface scanning and analysis method discussed in Section 2.3, the point cloud data of the scanned target surface was fitted into a 7 × 7 order NUMBS (Figure 3a). The reason why such a relatively low order was chosen to represent this NUMBS is that the precision of our equipment can only guarantee the stretchability and conductivity of the structure with relatively large unit cell (Figure 5).

For each subsurface of the target NUMBS, length of each edge was measured. Based on Equation (3) & (4), applied strain and PR of the subsurface can be obtained. Figure 7b shows the length of each edge for subsurface $SubS_{1,1}$. With Equation (3) and (4), we could get the

applied strain ($\varepsilon_{1,1} = 0.175$) and PR ($\nu_{1,1} = -0.543$) of the subsurface $SubS_{1,1}$. From the resulted "ν-$\theta$-ε" relationship summarized in Figure 6, the corresponding unit cell (with $\varepsilon_{cell} = 0.175$ & $\nu_{cell} = -0.543$) has the interior angle of $\theta_{cell} = 206.8°$, as shown in Figure 7c. As a result, this unit cell will be mapped to the position of $SubS_{1,1}$. One by one form $SubS_{1,1}$ to $Sub\square\square_{6,6}$, every subsurface were analysed and unit cell with the same deformation behaviour was mapped. The stretched strain and PR of all subsurfaces were summarized in Appendix 1. Besides, the corresponding interior angle of unit cell were also listed in the appendix. By connecting all the unit cells, the nonuniform structure of substrate were generated.

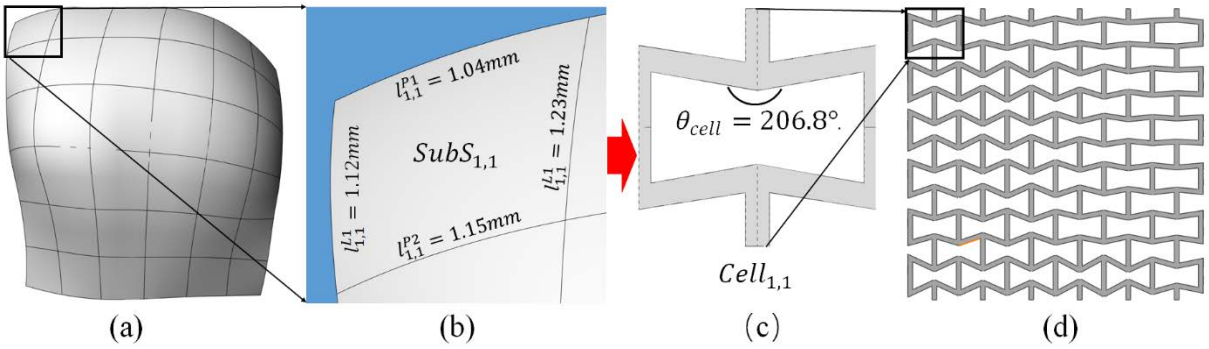



**Figure 7. The construction process of the nonuniform substrate. (a) shows the geometry of target NUMBS. (b) & (c) demonstrates the detailed generation process of the unit cell $Cell_{1,1}$. (d) is the layout of the final substrate structure.**

First, nonlinear FEA was conducted to test the conformability of the generated substrate. The setup of the test is illustrated in Figure 8a. The fitted NUMBS was defined as rigid body, and user-defined displacements were applied on both ends of the flexible substrate. All these displacements were defined through user-subroutines in Abaqus, and the objective was to make sure two ends of substrate were exactly connected with top and bottom edges of the target NUMBS. The dynamic-explicit solver was utilized to solve this nonlinear problem. As comparisons, structures with both uniform PPR and uniform NPR were also analysed. All these three FEA results were showed in Figure 8. To quantitatively evaluate the conformality of each stretched substrate, reciprocal of the maximum distance between the substrate and the edge of objective surface was introduced. As listed in Table 1, the substrate with nonuniform PR distribution had the best conformality to cover the target skin surface.

**Table 1. FEA resulted maximum distance and conformability of all three structures.**

| Substrate structure | Uniform PPR | Uniform NPR | Nonuniform |
|---|---|---|---|
| Maximum distance ($mm$) | 9.51 | 3.12 | 0.92 |
| Conformability ($mm^{-1}$) | 0.039 | 0.321 | 1.097 |

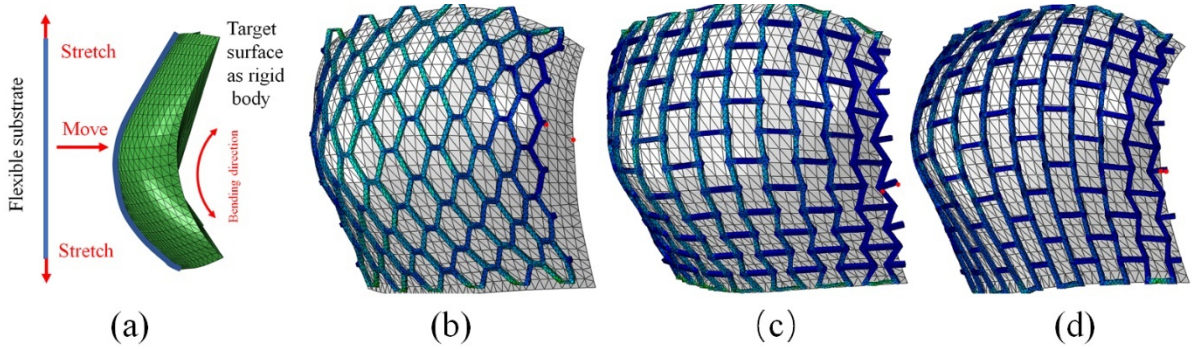



**Figure 8. (a) illustrates the initial state and boundary conditions of the dynamic FEA test. Respectively, (b), (c) and (d) show the FEA results of uniform PPR, uniform NPR and nonuniform substrates.**

Experimental tests were also carried out to validate both the conformability and conductivity of the designed nonuniform structure. Still, uniform NPR and PPR structure were tested as comparison. The deformation results of all three substrates were demonstrated in Figure 9, and measured maximum distance and resulted conformability were listed in Table 2. Same as the FEA test, the deformation behaviour of the designed nonuniform substrate wes more alike to the target skin. With large gap areas or out-of-bound areas (rad circle areas in Figure 9), it needs extra forces be applied on both free boundaries for uniform structures to obtain better conformation. However, this will bring sheer stress at the substrate-skin connection area, which will decrease the stability of the wearable devices dramatically. Therefore, the nonuniform structure, with a similar deformation behaviour to the target skin, is the most preferred design for wearable electronics at highly-stretched joint areas. Moreover, this nonuniform structure can exactly conform with the target surface without causing any wrinkle on the skin, which is impossible for those uniform structures, such as bulking structures, open-mesh structures, sponge structures *etc.*

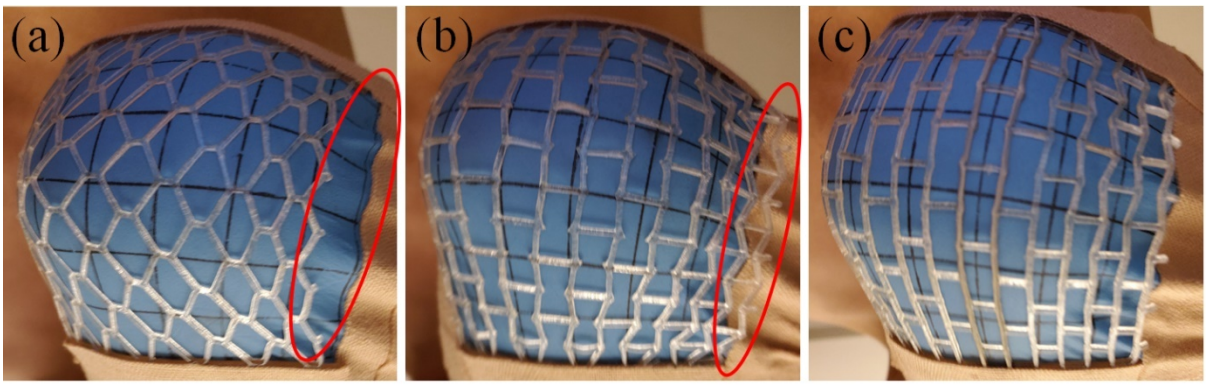



**Figure 9. Experimental tests of the conformability of three different structures: (a) uniform PPR, (b) uniform NPR, and (c) nonuniform.**

**Table 2. Experimentally resulted maximum distance and conformability of all three structures.**

| Substrate structure | Uniform PPR | Uniform NPR | Nonuniform |
|---|---|---|---|
| Maximum distance ($mm$) | 7.02 | 5.60 | 1.32 |
| Conformability ($mm^{-1}$) | 0.142 | 0.179 | 0.758 |

## 4. Conclusions

In this paper, a novel method was proposed to develop metasurface structure for highly-

stretched skin around joint areas. Both structural design and manufacture processes were included in this novel method. The PR distribution in the designed thin film structure was organized base on the analysis of the target skin surface. By mapping re-entrant unit cells with different PR to the subsurface areas with the same mechanical properties, the generated structure could mimic the deformation behaviour of the target skin. As a result, the sheer stress at the connection areas between substrate and the attached skin would be eliminated, which is the main challenge for those common film structures with uniformly distributed material. This property brings good flexibility and conformability to the designed metasurface. The proposed structural design method can be introduced to solve problems in different areas. One of our future works is to develop patient-specific skin scaffold that can perfectly conformed to the 3D complex surface of human body. Soft robotics and 4D printing are also the possible application areas.

## References

[1] F. Garnier, R. Hajlaoui, A. Yassar, P. Srivastava, All-Polymer Field-Effect Transistor Realized by Printing Techniques , Sci. . 265 (1994) 1684–1686. doi:10.1126/science.265.5179.1684.

[2] E. Menard, M.A. Meitl, Y. Sun, J.U. Park, D.J.L. Shir, Y.S. Nam, S. Jeon, J.A. Rogers, Micro- and nanopatterning techniques for organic electronic and optoelectronic systems, Chem. Rev. 107 (2007) 1117–1160. doi:10.1021/cr050139y.

[3] J.H. Ahn, J.H. Je, Stretchable electronics: Materials, architectures and integrations, J. Phys. D. Appl. Phys. 45 (2012). doi:10.1088/0022-3727/45/10/103001.

[4] J. Jeong, S. Kim, J. Cho, Y. Hong, Stable stretchable silver electrode directly deposited on wavy elastomeric substrate, IEEE Electron Device Lett. 30 (2009) 1284–1286. doi:10.1109/LED.2009.2033723.

[5] D.-H. Kim, N. Lu, R. Ma, Y.-S. Kim, R.-H. Kim, S. Wang, J. Wu, S.M. Won, H. Tao, Islam, K.J. Yu, T. Kim, R. Chowdhury, M. Ying, L. Xu, M. Li, H.-J. Chung, H. Keum, M. McCormick, P. Liu, Y.-W. Zhang, F.G. Omenetto, Y. Huang, T. Coleman,

[6] J.A. Rogers, Epidermal Electronics , Sci. . 333 (2011) 838–843. doi:10.1126/science.1206157. T. Someya, Y. Kato, T. Sekitani, S. Iba, Y. Noguchi, Y. Murase, H. Kawaguchi, T. Sakurai, Conformable, flexible, large-area networks of pressure and thermal sensors with organic transistor active matrixes, Proc. Natl. Acad. Sci. 102 (2005) 12321– 12325. doi:10.1073/pnas.0502392102.

[7] Q. Wu, J. Hu, A novel design of wearable thermoelectric generator based on 3D fabric structure, Smart Mater. Struct.,. at press (2017). https://doi.org/10.1088/1361- 665X/aa5694.

[8] W. Zeng, L. Shu, Q. Li, S. Chen, F. Wang, X.M. Tao, Fiber-based wearable electronics: A review of materials, fabrication, devices, and applications, Adv. Mater. 26 (2014) 5310–5336. doi:10.1002/adma.201400633.

[9] S. Sawada, Y. Tomita, K. Hirano, H. Morita, T. Ichiryu, M. Nomura, K. Kawakita, Novel wiring structure for 3D-conformable devices, 2016 Int. Conf. Electron. Packag. ICEP 2016. (2016) 124–128. doi:10.1109/ICEP.2016.7486795.

[10] X. Ren, R. Das, P. Tran, T.D. Ngo, Y.M. Xie, Auxetic metamaterials and structures: a review Auxetic metamaterials and structures: a review, Smart Mater. Struct. (2018). http://iopscience.iop.org/article/10.1088/1361-665X/aaa61c/pdf.

[11] S. Yang, I.S. Choi, R.D. Kamien, Design of super-conformable, foldable materials via fractal cuts and lattice kirigami, MRS Bull. 41 (2016) 130–137. doi:10.1557/mrs.2016.5.

[12] P. Vogiatzis, M. Ma, S. Chen, X.D. Gu, Computational design and additive manufacturing of periodic conformal metasurfaces by synthesizing topology optimization with conformal mapping, Comput. Methods Appl. Mech. Eng. 328 (2018) 477–497. doi:10.1016/j.cma.2017.09.012.

[13] Y. Han, W.F. Lu, Structural design of wearable electronics suitable for highly-stretched joint areas, Smart Mater. Struct. 27 (2018). doi:10.1088/1361-665X/aadf05.

[14] Y. Han, W. Lu, Optimizing the deformation behavior of stent with nonuniform Poisson's ratio distribution for curved artery, J. Mech. Behav. Biomed. Mater. (2018).

[15] Y. Han, W. Lu, Evolutionary design of nonuniform cellular structures with optimized Poisson's ratio distribution, Mater. Des. 141 (2017) 384–394. doi:10.1016/j.matdes.2017.12.047.

[16] S. Akbari, A.H. Sakhaei, K. Kowsari, B. Yang, A. Serjouei, Z. Yuanfang, Q. Ge, Enhanced multimaterial 4D printing with active hinges, Smart Mater. Struct. 27 (2018). doi:10.1088/1361-665X/aabe63.

[17] Y. Han, W. Lu, Evolutionary design of nonuniform cellular structures with optimized Poisson's ratio

distribution, Mater. Des. 141 (2018) 384–394. doi:10.1016/j.matdes.2017.12.047.
[18] L. Piegl, W. Tiller, The NURBS Book , (1997). doi:10.1007/978-3-642-59223-2.

# Influence of Surface and Couple Stresses on Response of Surface-loaded Elastic Half-plane

***Toan Minh Le[1], †Jaroon Rungamornrat[1], and Tinh Quoc Bui[2]**

[1]Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.
[2]Department of Civil and Environmental Engineering, School of Environment and Society, Tokyo Institute of Technology, Tokyo, Japan.

*Presenting author: minhtoanxdbk@gmail.com
†Corresponding author: Jaroon.R@chula.ac.th

## Abstract

A mathematical model integrating both Gurtin-Murdoch surface elasticity and consistent couple stress theories is proposed to simulate the simultaneous effects of the surface energy and the material microscopic structure on the mechanical response of an elastic half-plane under arbitrarily surface normal loadings. The displacement-based governing equations for the bulk material and the top material layer are established and then solved, via the method of Fourier integral transform together with the prescribed boundary conditions, to obtain the closed form solution of the elastic field in the transform space. To obtain solutions in the physical space, an efficient quadrature is adopted to evaluate all involved integrals associated with the Fourier transform inversion. A selected set of results is reported and they have indicated that both the surface and couple stresses significantly influence the elastic field within the bulk when the size of the loading region is comparable to the internal length scales of the surface material and the bulk.

**Keywords:** Elastic half-plane, Surface stresses, Couple stresses, Surface elasticity, Couple stress elasticity.

## Introduction

In past several decades, micro- and nano-technologies have received increasingly growing attention due to their vast applications in various disciplines. In the field of material sciences and engineering, understanding the fundamental characteristics and mechanical behavior of materials at those tiny scales is considered essential in the design procedure and fabrication of micro- and nano-scale devices and systems such as MEMS and NEMS. Unfortunately, many difficulties and challenges arise in the study of small-size objects since the response at those scales is significantly complex, generally size-dependent, and mostly influenced by various actions such as the surface free energy [1]-[3], existing defects and flaws, and material microscopic structures [4]. Experimental-based approaches are ones of the most popular candidates extensively and successfully employed to investigate the physical phenomena in a tiny scale (e.g., [5]-[8]). While results gained from those approaches have been found reliable and closely resembling the actual response, tests can only be performed in fully equipped laboratories and significant amount of resources associated with sophisticated testing setups and procedures and preparations of specimens must also be paid. This therefore renders purely experimental-based approaches less cost efficient in comparison with those combined with theoretical-based simulations. The latter, once properly equipped with physically admissible and sufficiently validated mathematical models, can be used not only to obtain the first-order prediction of the actual phenomena but also to assist the reduction of the number of cases to be considered in the experiments.

It has been well recognized that classical size-independent theories in continuum mechanics adopted specifically for simulating mechanical response of macro-scale problems have failed to simulate situations when the external length scale (e.g., size of loading region, crack length, contact length, etc.) is comparable to the internal length scale of materials (e.g., granular distance, lattice parameters). Attempts have been devoted to modify/enhance existing continuum-based mathematical models by integrating the influences observed in a small-scale before used in the simulations. Several continuum-based models have been proposed and successfully employed to capture the size-dependent behavior due to the presence of both surface-free energy and microstructures of constituting materials. For instance, the model established by Gurtin and his colleagues (e.g., [9][10]), called the theory of surface elasticity, has been successfully utilized to capture the surface stress effects. Due to its mathematical simplicity and capability in handling small-scale influence, the Gurtin-Murdoch model has become popular and extensively applied to investigate various problems in mechanics, e.g., thin films [11][12], thin plates [13], dislocations [14], nano-scaled elastic layer [15]-[17], half-space [18][19], and layered elastic half-space [20]. To handle the influence of material microstructures, various theories have been considered including the Cosserat theory [21], the couple stress theory [22]-[24], the strain gradient elasticity theory [25]-[27], the modified couple stress theory [28], and the consistent couple stress theory [29]. During the past decades, these theories have been extensively employed in the simulations and modeling of nano/micro-structured systems, especially for small-scaled beams and plates [30]-[32] and the size-dependent contact problems of elastic solids [33]-[35].

Towards the modeling of micro-/nano-scale layered media, results from an extensive literature survey have indicated that most of existing studies considered separately either the effect of the surface-free energy or the influence of the microscopic structures of constituting materials (e.g., [13]-[20], [33]-[38]). Applications of both Gurtin-Murdoch surface elasticity and the couple stress theory to simultaneously handling those small-scale influences, especially within the framework of surface and contact mechanics, have not been well recognized and this, as a consequence, leaves a significant gap of knowledge for further investigations. The idea of integrating both the surface-free energy and the micro/nano-structure of the bulk material in the modeling has been found in the study of size-dependent responses of nano-scale structures such as nano-beams (e.g., [39]) and nano-plates (e.g., [40]-[42]). This set of investigations not only confirms the applicability of the two theories but also provides the useful basis and ingredients essential for the extension to treat nano-scale problems of interest.

The present study aims to investigate the mechanical response of an elastic half-plane loaded on its surface by taking the influence of both surface and couple stresses into account. Gurtin-Murdoch surface elasticity theory is employed together with the consistent couple stress theory to form the underlying mathematical model and the analytical solution of elastic fields is obtained via the method of Fourier integral transform and a selected efficient numerical quadrature. The complete elastic fields under the simultaneous effects of surface stresses and couple stresses within the half-plane are thoroughly studied.

## Problem Formulation

Consider a linearly elastic half-plane loaded by an arbitrarily distributed normal traction $p$ over the length $2a$ on the top surface as shown schematically in Fig. 1. For convenience, a

reference Cartesian coordinate system $\{x, y, z; O\}$ is chosen such that the origin is at the center of the loading region and the $x$ , $y$ , and $z$ -axes direct rightward, downward, and normal to the half-plane, respectively. The bulk material is assumed homogeneous and its response is described by the consistent couple stress theory (e.g., [29]) whereas the material layer at the top surface of the half-plane is governed by the theory of surface elasticity (e.g., [9][10]). In the present study, it is assumed that the body force and the body couple are negligible and the plane-strain deformation prevails.

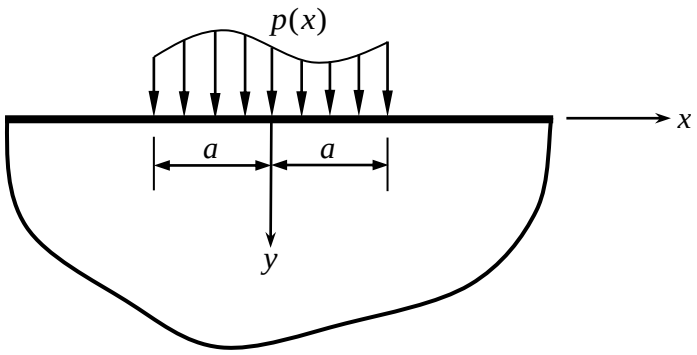



**Figure 1. Schematic of an elastic half-plane loaded on its surface by normal traction**

For the bulk medium, basic field equations (i.e., equilibrium equations, constitutive laws, and kinematics) from the consistent couple stress theory (e.g., [29]) when specialized to the two-dimensional body under the plane strain condition are given by

$$\frac{\partial \sigma_{xx}}{\partial x}+\frac{\partial \sigma_{yx}}{\partial y}=0,\ \ \frac{\partial \sigma_{xy}}{\partial x}+\frac{\partial \sigma_{yy}}{\partial y}=0,\ \ \frac{\partial \mu_{xz}}{\partial x}+\frac{\partial \mu_{yz}}{\partial y}+\sigma_{xy}-\sigma_{yx}=0 \tag{1}$$

$$\begin{aligned}
&\sigma_{xx}=\frac{2\mu}{1-2\nu}\left[(1-\nu)\varepsilon_{xx}+\nu\varepsilon_{yy}\right],\ \ \sigma_{yy}=\frac{2\mu}{1-2\nu}\left[\nu\varepsilon_{xx}+(1-\nu)\varepsilon_{yy}\right],\\
&\sigma_{xy}=2\mu\varepsilon_{xy}-2\mu\ell^2\Delta\omega,\ \ \sigma_{yx}=2\mu\varepsilon_{xy}+2\mu\ell^2\Delta\omega,\\
&\mu_{xz}=-8\mu\ell^2\kappa_{xz},\ \ \mu_{yz}=-8\mu\ell^2\kappa_{yz}
\end{aligned} \tag{2}$$

$$\begin{aligned}
&\varepsilon_{xx}=\frac{\partial u_x}{\partial x},\ \ \varepsilon_{yy}=\frac{\partial u_y}{\partial y},\ \ \varepsilon_{xy}=\frac{1}{2}\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right),\\
&\omega=\frac{1}{2}\left(\frac{\partial u_y}{\partial x}-\frac{\partial u_x}{\partial y}\right),\ \ \kappa_{xz}=-\frac{1}{2}\frac{\partial \omega}{\partial x},\ \ \kappa_{yz}=-\frac{1}{2}\frac{\partial \omega}{\partial y}
\end{aligned} \tag{3}$$

where $\{\sigma_{xx},\sigma_{yy},\sigma_{xy},\sigma_{yx}\}$ are the force-stress components; $\{\mu_{xz},\mu_{yz}\}$ are the couple-stress components; $\{\varepsilon_{xx},\varepsilon_{yy},\varepsilon_{xy}\}$ and $\omega$ are the infinitesimal strain components and the rotation about the $z$ -axis, respectively; $\{\kappa_{xz},\kappa_{yz}\}$ represent the curvature components; $\{u_x,u_y\}$ are the in-plane displacement components; $\mu$ and $\nu$ are the elastic shear modulus and Poisson's ratio of the bulk material, respectively; $\ell$ is a length-scale parameter in the couple stress elasticity; and $\Delta$ denotes two-dimensional Laplacian operator. It is worth noting that the presence of the couple stresses renders the force-stress tensor non-symmetric and this is in contrast to the classical linear elasticity. Note also that by setting the parameter $\ell=0$, one can readily recover the classical case.

For the material layer at the top surface of the half-plane, basic equations governing its response can be established from Gurtin-Murdoch surface elasticity theory [9][10] and, when specialized to this particular case, they are given by

$$\frac{\partial \sigma^s{}_{xx}}{\partial x} + t_x^s = 0, \ \frac{\partial \sigma^s{}_{xy}}{\partial x} + t_y^s + p = 0 \tag{4}$$

$$\sigma^s{}_{xx} = \tau^s + \kappa^s \varepsilon_{xx}^s, \ \sigma^s{}_{xy} = \tau^s \frac{\partial u_y^s}{\partial x} \tag{5}$$

$$\varepsilon_{xx}^s = \frac{\partial u_x^s}{\partial x} \tag{6}$$

where the superscript '*s*' is utilized to designate the surface quantities; $\{\mu^s, \lambda^s\}$ and $\tau^s$ denote surface Lamé constants and the residual surface tension, respectively; $\kappa^s = 2\mu^s + \lambda^s$ denote the surface elastic modulus; and $\{t_x^s, t_y^s\}$ are tractions acting to the bottom side of the material layer induced by the underlying bulk material.

**Solution Procedure**

A set of displacement-based governing equations for the bulk material can be readily obtained by properly combining all basic field equations, Eq. (1)-(3), and its general solution can be established in a closed-form via the method of Fourier integral transform (e.g., [43]). The final expression for the in-plane displacements $\{u_x, u_y\}$ is given by

$$u_x(x,y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[ Ai|\xi| e^{-|\xi|y} - Bi(\kappa - |\xi| y) e^{-|\xi|y} + Ci\zeta e^{\frac{-\zeta}{\ell} y} \right] e^{-i\xi x} d\xi \tag{7}$$

$$u_y(x,y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[ A\xi e^{-|\xi|y} + B\xi y e^{-|\xi|y} + C\ell\xi e^{\frac{-\zeta}{\ell} y} \right] e^{-i\xi x} d\xi \tag{8}$$

where $i = \sqrt{-1}$ denotes an imaginary number; $\xi$ is a transform parameter; $\zeta = \sqrt{1 + \xi^2 \ell^2}$ ; $\kappa = 3 - 4\nu$ ; and $A, B, C$ are arbitrary unknown functions of $\xi$ to be determined from boundary conditions. The general solution for the rotation $\omega$, the force stresses $\{\sigma_{xx}, \sigma_{yy}, \sigma_{xy}, \sigma_{yx}\}$, and the couple stresses $\{\mu_{xz}, \mu_{yz}\}$ can be obtained from Eq. (7)-(8) together with Eq. (2)-(3).

To form sufficient conditions for determining the unknown functions *A, B, C*, the continuity of the displacements and tractions at the interface of the bulk and the material layer is enforced along with the surface equations, Eq. (4)-(6). For the case of a constant residual surface tension, the following set of three boundary conditions is obtained:

$$\tau^s \left. \frac{\partial^2 u_y}{\partial x^2} \right|_{y=0} + \sigma_{yy}\big|_{y=0} + p(x) = 0 \tag{9}$$

$$\kappa^s \left.\frac{\partial^2 u_x}{\partial x^2}\right|_{y=0} + \left.\sigma_{yx}\right|_{y=0} = 0 \tag{10}$$

$$\left.\mu_{yz}\right|_{y=0} = 0 \tag{11}$$

It is worth noting that the couple traction boundary condition Eq. (11) results directly from that the material layer cannot resist the bending moment. By substituting Eq. (7)-(8) and the general solution for the force and couple stresses into Eq. (9)-(11), it gives rise to a system of three linear algebraic equations:

$$\begin{bmatrix} \dfrac{\tau^s\xi^3}{\mu} + 2\xi|\xi| & 2\xi(2\nu-1) & \dfrac{\tau^s\ell\xi^3}{\mu} + 2\xi\zeta \\ i\xi^2(\Lambda|\xi|+2) & -i\left[\Lambda\xi^2(3-4\nu)+4|\xi|(1-\nu)\right] & i\xi^2(\Lambda\zeta+2\ell) \\ 0 & 8i\ell^2\xi^2(\nu-1) & 2i\zeta \end{bmatrix} \begin{Bmatrix} A \\ B \\ C \end{Bmatrix} = \frac{1}{\mu} \begin{Bmatrix} \hat{p}(\xi) \\ 0 \\ 0 \end{Bmatrix} \tag{12}$$

where $\Lambda = \kappa^s / \mu$ is defined as the material length-scale parameter corresponding to the presence of the surface stresses and $\hat{p}(\xi)$ is the Fourier transform of the prescribed normal traction $p(x)$ given by

$$\hat{p}(\xi) = \int_{-\infty}^{+\infty} p(x) e^{i\xi x} dx \tag{13}$$

By solving the system of linear equations, Eq. (12), the unknown functions *A*, *B*, *C* are obtained in a closed-form as functions of $\xi$, two length-scale parameters $\Lambda$ and $\ell$, and transformed traction $\hat{p}(\xi)$. Substituting these functions into the general solutions gives the integral expressions for the displacements, rotation, force-stress and couple-stress fields at any location within the half-plane. Interestingly, it can be observed that all elastic field quantities contain both $\Lambda$ and $\ell$ representing the length-scale parameters corresponding to the presence of the surface stresses and couple stresses, respectively. Therefore, it is anticipated that the influence of simultaneous effects of the surface and couple stresses can be captured in predicted solutions as well as the size-dependent behavior.

## Numerical Results and Discussion

To obtain numerical results for elastic field within the bulk material, standard Gaussian quadrature is adopted to efficiently evaluate all involved integrals resulting from Fourier integral inversion. In the numerical study, following material properties associated with Silicon [44] (i.e., $\nu = 0.33$, $\mu = 40.23$ GPa, $\lambda = 78.08$ GPa, $\mu^s = 2.78$ N/m, $\lambda^s = 4.49$ N/m, $\tau^s = 0.61$ N/m) and the length-scale parameter associated with the presence of couple stresses $\ell = 50$ nm [45] are chosen. A representative surface load, corresponding to a uniformly distributed normal pressure $p(x) = p_0$ over the length $2a$, is chosen and its Fourier transform is given explicitly by

$$\hat{p}(\xi) = \frac{2\sin(a\xi)}{\xi} p_0 \tag{14}$$

Numerical results for the force stresses and couple stresses of an elastic half-plane under the uniformly distributed normal pressure $p(x) = p_0$ and the simultaneous influence of both surface and couple stresses are reported in Fig. 2 and Fig. 3, respectively. The stress distributions along the positive *x*-direction of the surface-loaded half-plane are considered at different normalized depths, $y/a \in \{0.2; 0.4; 0.8\}$ with $a = 0.5\Lambda$ and $\Lambda = 0.25$ nm. Results represented by the dash lines denote the classical solutions corresponding to $\ell = \Lambda = 0$. It is worth noting that while the classical solutions are independent of the length scale $\Lambda$, the use of $\Lambda$ in the normalization is only for the purpose of comparison.

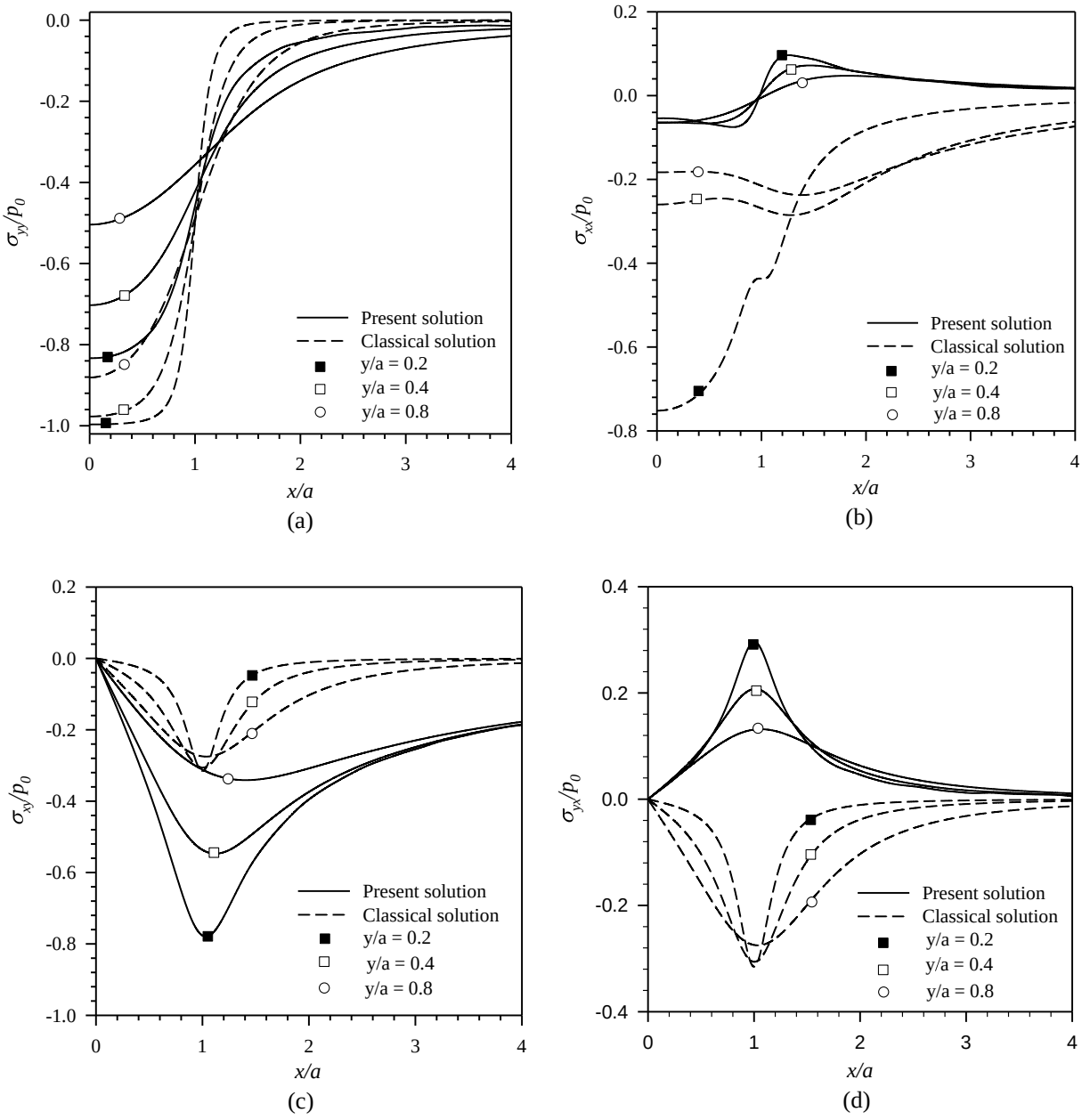



**Figure 2. Normalized force stresses of elastic half-plane under uniformly distributed normal traction at various depths**

Although the normalized vertical stress $\sigma_{yy}/p_0$ and the shear stress $\sigma_{xy}/p_0$ from the proposed model and the classical solution display similar trends for all values of the normalized coordinate $x/a$ as displayed in Fig. 2(a) and Fig. 2(c), the magnitude of the normalized stresses $\sigma_{yy}/p_0$ and $\sigma_{xy}/p_0$ with the influence of the surface and couple stresses are lower and higher, respectively, than those of the classical case. In contrast, the normalized horizontal stress $\sigma_{xx}/p_0$ and the shear stress $\sigma_{yx}/p_0$ possess different characteristic in comparison with the classical solutions. In addition, the non-symmetric character of the force-stress tensor is confirmed by results shown in Fig. 2(c) and Fig. 2(d). Variations of the couple stresses $\mu_{xz}$, $\mu_{yz}$ within the elastic half-plane at various depths are also displayed in Fig. 3. For the classical case, the couple stresses within the bulk vanish identically. It is also evident from this set of results that solutions predicted by the proposed model exhibit the significant departure from the classical solutions.

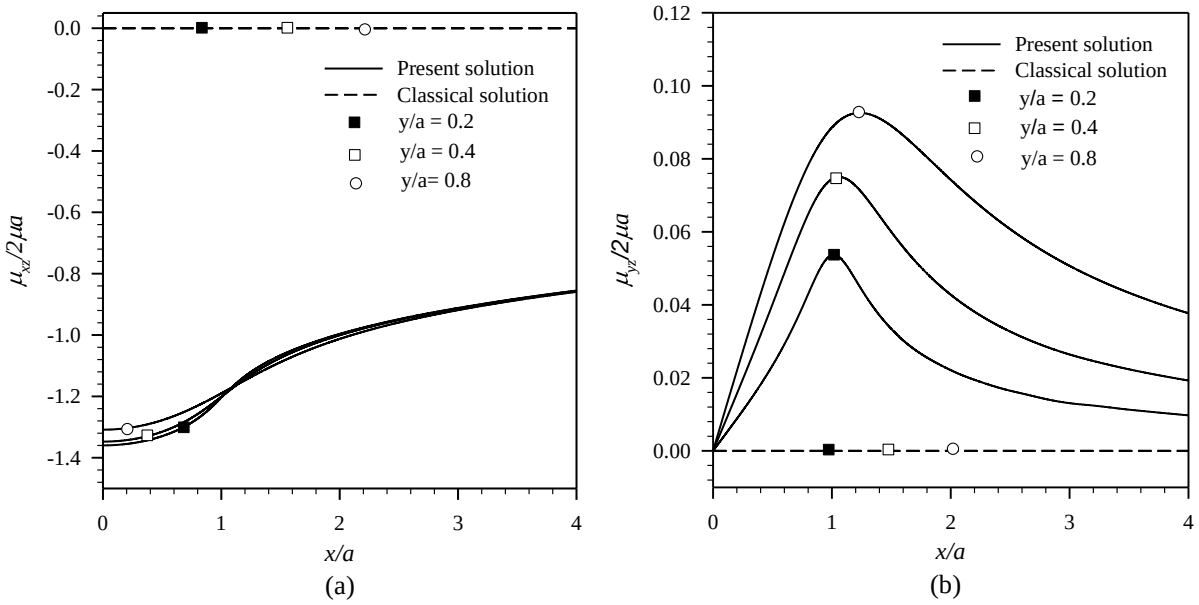



**Figure 3. Normalized couple stresses of elastic half-plane under uniformly distributed normal traction at various depths**

## Conclusion and Remarks

An elastic half-plane under the plane-strain deformation and loaded on its surface by the normal traction has been studied by integrating the influence of both surface stresses and couple stresses. Gurtin-Murdoch surface elasticity theory and the consistent couple stress theory have been used in the problem formulation and the closed-form integral expressions of the elastic field have been derived via the method of Fourier integral transform. An efficient quadrature has been adopted to evaluate all involved integrals resulting from the Fourier transform inversion. A set of preliminary results has indicated the significant influence of the surface and couple stresses on the behavior of predicted solutions; in particular, the obvious deviation from the classical solutions has been observed. The size-dependence behavior of predicted responses and the effects of the two material length scales are also of key interest and still under investigation. It should be remarked that the analytical solutions established in

the present study can be used either as benchmark results in the verification process or as the fundamental solutions in the formulation of related problems such as contact and indentation problems.

## Acknowledgements

The authors gratefully acknowledge the support provided by CU Scholarship for ASEAN Countries 2017 and Thailand Research Fund (Grants No. RSA5980032).

## References

[1] Dingreville, R., Qu, J. and Cherkaoui, M. (2005) Surface free energy and its effect on the elastic behavior of nano-sized particles, wires and films, *Journal of the Mechanics and Physics of Solids* **53**(8), 1827-1854.
[2] Cammarata, R. C. (1994) Surface and interface stress effects in thin films, *Progress in Surface Science* **46**(1), 1-38.
[3] Cammarata, R. C. (1997) Surface and interface stress effects on interfacial and nanostructured materials, *Materials Science and Engineering: A* **237**(2), 180-184.
[4] Maranganti, R. and Sharma, P. (2007) A novel atomistic approach to determine strain-gradient elasticity constants: Tabulation and comparison for various metals, semiconductors, silica, polymers and the (Ir) relevance for nanotechnologies, *Journal of the Mechanics and Physics of Solids* **55**(9), 1823-1852.
[5] Wong, E. W., Sheehan, P. E. and Lieber, C. M. (1997) Nanobeam mechanics: elasticity, strength, and toughness of nanorods and nanotubes, *Science* **277**(5334), 1971-1975.
[6] Peng, B., Locascio, M., Zapol, P., Li, S., Mielke, S. L., Schatz, G. C. and Espinosa, H. D. (2008) Measurements of near-ultimate strength for multiwalled carbon nanotubes and irradiation-induced crosslinking improvements, *Nature Nanotechnology* **3**, 626-631.
[7] Mao, S. X., Zhao, M. and Wang, Z. L. (2003) Nanoscale mechanical behavior of individual semiconducting nanobelts, *Applied Physics Letters* **83**(5), 993-995.
[8] Li, X., Gao, H., Murphy, C. J. and Caswell, K. K. (2003) Nanoindentation of silver nanowires, *Nano Letters* **3**(11), 1495-1498.
[9] Gurtin, M. E. and Murdoch, A. I. (1975) A continuum theory of elastic material surfaces, *Archive for Rational Mechanics and Analysis* **57**(4), 291-323.
[10] Gurtin, M. E. and Murdoch, A. I. (1978) Surface stress in solids, *International Journal of Solids and Structures* **14**(6), 431-440.
[11] He, L. H., Lim, C. W. and Wu, B. S. (2004) A continuum model for size-dependent deformation of elastic films of nano-scale thickness, *International Journal of Solids and Structures* **41**(3-4), 847-857.
[12] Huang, D. W. (2008) Size-dependent response of ultra-thin films with surface effects, *International Journal of Solids and Structures* **45**(2), 568-579.
[13] Lu, P., He, L. H., Lee, H. P. and Lu, C. (2006) Thin plate theory including surface effects, *International Journal of Solids and Structures* **43**(16), 4631-4647.
[14] Intarit, P., Senjuntichai, T. and Rajapakse, R. K. N. D. (2010) Dislocations and internal loading in a semi-infinite elastic medium with surface stresses, *Engineering Fracture Mechanics* **77**(18), 3592-3603.
[15] Intarit, P., Senjuntichai, T., Rungamornrat, J. and Rajapakse, R. K. N. D. (2011), Surface elasticity and residual stress effect on the elastic field of a nanoscale elastic layer, *Interaction and Multiscale Mechanics* **4**(2), 85-105.
[16] Zhao, X. J., *Surface loading and rigid indentation of an elastic layer with surface energy effects*, Master Thesis, The University of British Columbia, Canada, 2009.
[17] Rungamornrat, J., Tuttipongsawat, P. and Senjuntichai, T. (2016) Elastic layer under axisymmetric surface loads and influence of surface stresses, *Applied Mathematical Modelling* **40**(2), 1532-1553.
[18] Pinyochotiwong, Y., Rungamornrat, J. and Senjuntichai, T. (2013) Rigid frictionless indentation on elastic half space with influence of surface stresses, *International Journal of Engineering Science* **71**, 15-35.
[19] Zhou, S. and Gao, X.-L. (2013) Solutions of half-space and half-plane contact problems based on surface elasticity, *Zeitschrift für angewandte Mathematik und Physik ZAMP* **64**(1), 145-166.
[20] Tirapat, S., Senjuntichai, T. and Rungamornrat, J. (2017) Influence of surface energy effects on elastic fields of a layered elastic medium under surface loading, *Advances in Materials Science and Engineering* **2017**, 7530936.
[21] Cosserat, E. and Cosserat, F. (1909) *Théorie des corps déformables*, A. Herman et Fils, Paris.

[22] Mindlin, R. D. and Tiersten, H. F. (1962) Effects of couple-stresses in linear elasticity, *Archive for Rational Mechanics and Analysis* **11**(1), 415-448.
[23] Mindlin, R. D. (1963) Influence of couple-stresses on stress concentrations, *Experimental Mechanics* **3**(1), 1-7.
[24] Toupin, R. A. (1962) Elastic materials with couple-stresses, *Archive for Rational Mechanics and Analysis* **11**(1), 385-414.
[25] Mindlin, R. D. (1965) Second gradient of strain and surface-tension in linear elasticity, *International Journal of Solids and Structures* **1**(4), 417-438.
[26] Mindlin, R. D. and Eshel, N. N. (1968) On first strain-gradient theories in linear elasticity, *International Journal of Solids and Structures* **4**(1), 109-124.
[27] Chen, S. and Wang, T. (2001) Strain gradient theory with couple stress for crystalline solids, *European Journal of Mechanics - A/Solids* **20**(5), 739-756.
[28] Yang, F., Chong, A. C. M., Lam, D. C. C. and Tong, P. (2002) Couple stress based strain gradient theory for elasticity, *International Journal of Solids and Structures* **39**(10), 2731-2743.
[29] Hadjesfandiari, A. R. and Dargush, G. F. (2011) Couple stress theory for solids, *International Journal of Solids and Structures* **48**(18), 2496-2510.
[30] Ma, H. M., Gao, X. -L. and Reddy, J. N. (2011) A non-classical Mindlin plate model based on a modified couple stress theory, *Acta Mechanica* **220**(1-4), 217-235.
[31] Ma, H. M., Gao, X. -L. and Reddy, J. N. (2008) A microstructure-dependent Timoshenko beam model based on a modified couple stress theory, *Journal of the Mechanics and Physics of Solids* **56**(12), 3379-3391.
[32] Beni, Y. T., Mehralian, F. and Razavi, H. (2015) Free vibration analysis of size-dependent shear deformable functionally graded cylindrical shell on the basis of modified couple stress theory, *Composite Structures* **120,** 65-78.
[33] Gourgiotis, P. and Zisis, T. (2016) Two-dimensional indentation of microstructured solids characterized by couple-stress elasticity, *The Journal of Strain Analysis for Engineering Design* **51**(4), 318-331.
[34] Song, H. X., Ke, L. L. and Wang, Y. S. (2017) Sliding frictional contact analysis of an elastic solid with couple stresses, *International Journal of Mechanical Sciences* **133**, 804-816.
[35] Song, H., Ke, L., Wang, Y., Yang, J. and Jiang, H. (2018) Two-dimensional frictionless contact of a coated half-plane based on couple stress theory, *International Journal of Applied Mechanics* **10**(5), 1850049.
[36] Muki, R. and Sternberg. E. (1965) The influence of couple-stresses on singular stress concentrations in elastic solids, *Zeitschrift für angewandte Mathematik und Physik ZAMP* **16**(5), 611-648.
[37] Zisis, T., Gourgiotis, P. A., Baxevanakis, K. P. and Georgiadis, H. G. (2014) Some basic contact problems in couple stress elasticity, *International Journal of Solids and Structures* **51**(11-12), 2084-2095.
[38] Wang, Y., Shen, H., Zhang, X., Zhang, B., Liu, J. and Li, X. (2018) Semi-analytical study of microscopic two-dimensional partial slip contact problem within the framework of couple stress elasticity: Cylindrical indenter, *International Journal of Solids and Structures* **138**, 76-86.
[39] Gao, X. -L. and Mahmoud, F. F. (2014) A new Bernoulli–Euler beam model incorporating microstructure and surface energy effects, *Zeitschrift für angewandte Mathematik und Physik ZAMP* **65**(2), 393-404.
[40] Shaat, M., Mahmoud, F. F., Gao, X. -L. and Faheem, A. F. (2014) Size-dependent bending analysis of Kirchhoff nano-plates based on a modified couple-stress theory including surface effects, *International Journal of Mechanical Sciences* **79**, 31-37.
[41] Gao, X. -L. and Zhang, G. Y. (2016) A non-classical Kirchhoff plate model incorporating microstructure, surface energy and foundation effects, *Continuum Mechanics and Thermodynamics* **28**, 195-213.
[42] Liu, S., Yu, T., Lich, L. V., Yin, S. and Bui, T. Q. (2019) Size and surface effects on mechanical behavior of thin nanoplates incorporating microstructures using isogeometric analysis, *Computers and Structures* **212**, 173-187.
[43] Sneddon, I. N. (1951) *Fourier Transforms*, McGraw-Hill, New York.
[44] Miller, R. E. and Shenoy, V. B. (2000) Size-dependent elastic properties of nanosized structural elements, *Nanotechnology* **11**(3), 139-147.
[45] Rahimi, Z., Rezazadeh, G. and Sadeghian, H. (2018) Study on the size dependent effective Young modulus by EPI method based on modified couple stress theory, *Microsystem Technologies* **24**(7), 2983–2989.

# Learning with Navigation Feature: Quantitative Risk Analysis for the Navigation of Autonomous Ships

Wei Chian Tan[1], Kie Hian Chua[2], and Yanling Wu[3]

[1]Nanyang Technological University
[2]Technology Centre for Offshore and Marine, Singapore
[3]University of Science and Technology of China
[3]Corresponding author: yanlwu@ustc.edu.cn

**Abstract**

This work presents a data-driven approach for the automated risk estimation of the voyage of a vessel or ship. While the industry is moving from a compliance-based framework with existing rules to a risk-based one, there is also a need to monitor the risk of a vessel from the perspective of the navigation. This is of even higher importance for the case of autonomous ships. Built based on the state-of-the-art mathematical representation, the navigation feature, each existing voyage is transformed into a corresponding series of points in $d$-dimensional space. During the stage of pre-processing, given a set of historical Automatic Identification System (AIS) data, those records that belong to the same vessel within a certain period of time are taken as a voyage and mapped to the corresponding space of the navigation feature. After the pre-processing and during the online monitoring, the current trajectory of the vessel is transformed into the corresponding representation in the same way. Based on a nearest-neighbour search scheme, the distance from the nearest neighbour is taken as the risk of the current voyage. In other words, the deviation from the closest route in the historical data is taken as the risk. The developed method has demonstrated encouraging performance on a set of challenging historical AIS data from the Australian Maritime Safety Authority, covering three regions in the Australian territory, namely the Bass Strait, the Great Australian Bight and the North West.
**Keywords:** Autonomous ships, navigation planning, historical AIS data, machine learning.

## 1 Introduction

Autonomous ships have attracted significant amount of attention from the marine and offshore industry lately. Compared to conventional manned vessels, the huge potential of these autonomous vessels towards a new level of operational efficiency makes them a promising candidate as a technological solution for the next step of the industry.

As each ship and offshore platform is a massive engineering system by itself, the need of ensuring the seaworthiness of the vessel or platform is of ultimate importance. Usually performed by an independent third party such as a classification society [10], the process of engineering a ship is required to be certified and classed. However, a trend or movement from a compliance-based framework to a risk-based one has been observed in the domain. As the complexity of the massive engineering systems nowadays is going beyond the capability of a binary (pass or

fail) evaluation, there needs to be a way to ensure the reliability of these systems, beyond a compliance according to established rules. A risk-based approach can come in as an effective alternative, as it does not only take the part of compliance into a consideration, but also looking at the risk of the system, sub-systems and components quantitatively. This would give the engineers and the relevant authorities a sense on the level of confidence in terms of performance and reliability of the system.

Navigation at sea typically requires compliance with applicable rules and regulations including but not limited to Convention on the International Regulations for Preventing Collision at Sea, 1972 (COLREGs) [11] from the International Maritime Organisation [12]. These are essentially the rules of traffic at sea, globally and regionally. As an inappropriate movement of a vessel can lead to risky situations such as two vessels being too close and a potential collision, ensuring a real-time compliance by all the vessels is of significant importance for the safety and the smoothness of operations of all.

Hence, while the industry is moving from a compliance-based system towards a risk-based one, together with the recent focus on autonomous vessels, there is a need for an approach to do automated risk estimation and monitoring of the voyage of each vessel on a (near-) real-time basis. This work aims to address the issue through a historical data-driven approach. Based on a set of historical AIS (Automatic Identification System) data, the correspondence among the AIS points which belong to the same vessel is established to turn the collection of points into a voyage. After a list of historical voyages is obtained, given the current location of the vessel, the distance from the nearest point of a historical voyage in the space of navigation feature is returned as the risk estimation of the current voyage at the moment. The developed method has demonstrated encouraging performance on an existing set of historical AIS data from the Australian Maritime Safety Authority [3], covering three regions in Australian territory, namely the Bass Strait, the Great Australian Bight and the North West [4] [1]. Figure 1 shows the three regions studied, and figure 2 to 4 show the visualisation of the historical AIS data for each of the regions.

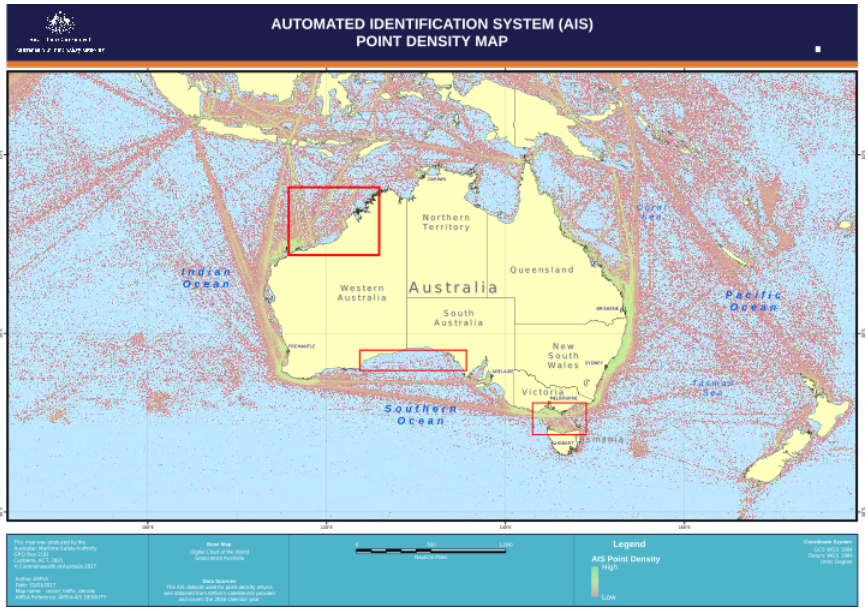



Figure 1: Maritime regions studied in this work [4].

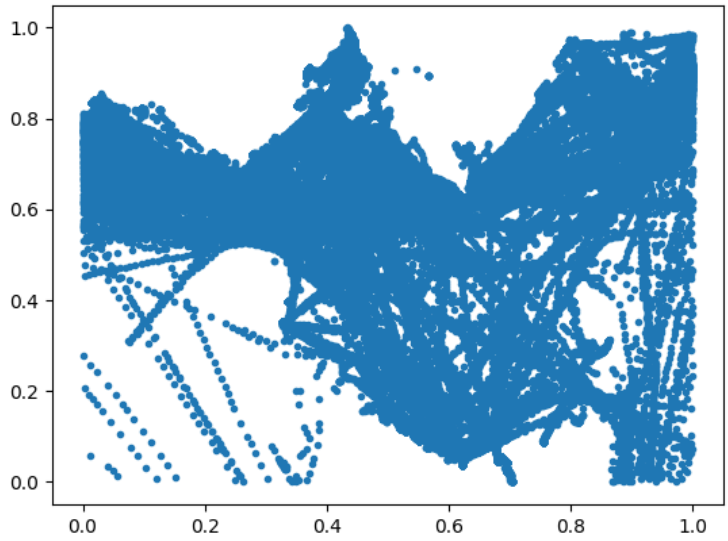



Figure 2: A visualisation of the historical AIS data for the Bass Strait [4].

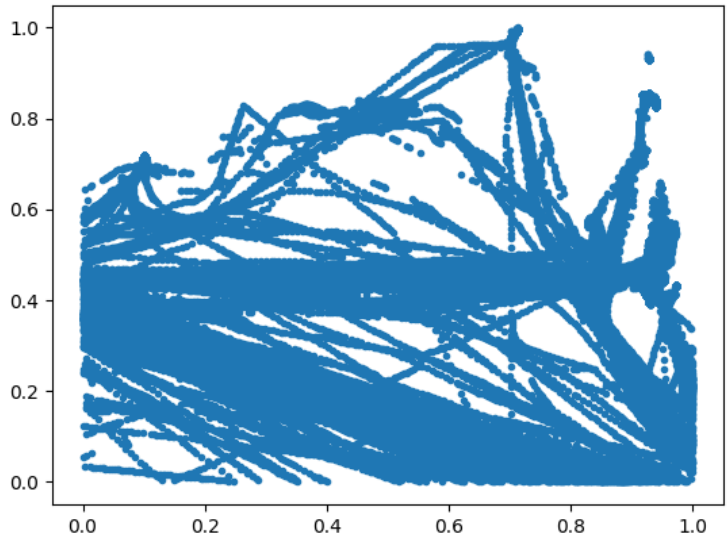



Figure 3: A visualisation of the historical AIS data for the Great Australian Bight [4].

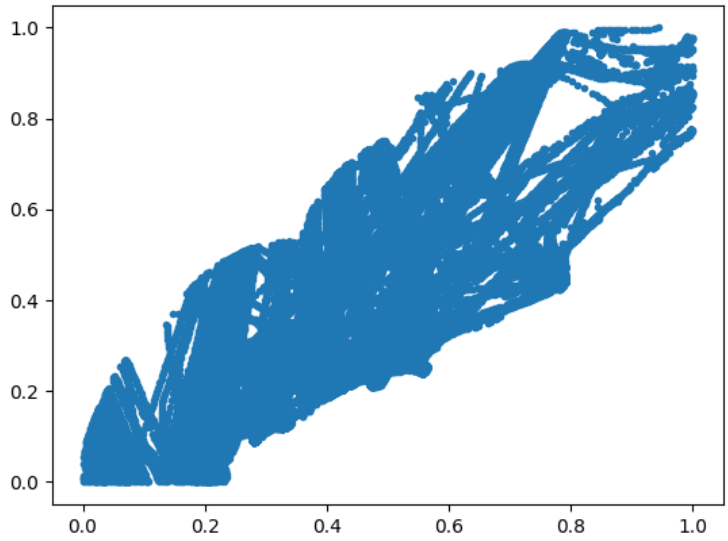



Figure 4: A visualisation of the historical AIS data for the North West [4].

## 2 Related work

There is a significant volume of existing work on the domain of autonomous ships, covering things like situations at high seas [9], path generation [22] to collision detection and avoidance [13] [14]. In addition, the aspect of connectivity of autonomous ships has been looked into [6]. An interesting idea to augment the sensing capability of an autonomous ship via the support from an unmanned aerial system has also been explored [8]. Along the topic of landing a quadrotor on the deck of an autonomous ship, an invariant ellipsoid method has been developed [16].

Closely relevant to the aim of this work, risk based methods have been seen in maritime for different purposes such as autonomous systems [21] and unmanned merchant ships [15]. Similarly, an approach of learning using corrosion feature with non-linear (Support Vector Machines) SVM [5] has been introduced to determine the potential corrosion mechanisms which may happen based on a set of design and operating conditions in an automated manner [19].

In a recent approach to augment the existing navigation planning with the use of historical AIS data, a weighted nearest-neighbour search in the space of ship and navigation feature has been introduced for the retrieval of a suitable route for an upcoming voyage from a database of historical routes [20].

As the industry typically deals with massive engineering systems such as ships or offshore platforms, the importance of compliance according to established applicable rules has been put at the first place. A Histogram of Connectivity and linear SVM based approach has been introduced for the evaluation of the piping design of a ship [17]. On the same topic, by fine-tuning pre-trained deep convolutional neural networks, the hypothesis that common visual features learnt can be reused for ship design has been validated to a significant extent [18].

Based on the discussion above, there appears to be a gap between the state-of-the-art and the aim of this work, which is to do an automated estimation of the risk of a voyage based on historical AIS data.

## 3 Methodology

Details of the main method developed in this work are presented in this section.

### 3.1 Representation: Navigation Feature

Before the idea of doing a risk-based estimation for a voyage can be achieved, there is a need to establish a representation or a feature space. This starts with linking those AIS points that belong to the same vessel as a voyage. Next, for each voyage identified in the database, a representation is established accordingly.

Similar to previous work, the idea of navigation feature [20] is adopted as the representation here. Assuming that there are $d$ attributes that need to be taken into consideration, each of these attributes constitutes a dimension and this leads to a navigation feature for that set of attributes. In other words, a set of $d$ navigation attributes is transformed into a point $\mathbf{x} \in R^d$. For each point in a voyage, a navigation feature is established and this process is repeated for all the points in the voyage. The process of forming navigation features for a voyage is illustrated in algorithm 1 and 2.

The current form of navigation feature has the advantage of being flexible in terms of attributes to be taken into consideration. Depending on the actual situation, the variables such as the speed or the operating status of the engine on-board can be included for a better analysis. This can be very helpful when certain signals become unreliable and these sources can just be excluded from the analysis meanwhile. Four variables are taken into consideration in this work, including the longitude, the latitude, the speed and the course, hence the analysis is done in four-dimensional space.

**Algorithm 1** Formation of a Navigation Feature.

1: Given $d$ attributes
2: Initialise $\mathbf{x} \in R^d$
3: **for** $0 \leq i < d$ **do**
4: $\quad \mathbf{x}_i = i$-th attribute
5: **end for**
6: return $\mathbf{x}$

**Algorithm 2** Pre-processing of a voyage.

1: Given a voyage with $N_p$ points
2: **for** $0 \leq i < N_p$ **do**
3: $\quad$ Form $i$-th navigation feature for the $i$-th point, $\mathbf{f}_i \in R^d$
4: $\quad$ Store $\mathbf{f}_i$
5: **end for**

### 3.2 Risk-based estimation via nearest-neighbour search

The idea of doing risk estimation of the current voyage based on a list of historical voyages in the space of navigation feature is introduced in this section. Based on a database of historical voyages or AIS data, the deviation of the current voyage from the closest historical one is determined as the risk. Theoretically, the risk is determined by the distance between the query and the closest point in the space of navigation feature.

The process is mainly divided into two stages, including the pre-processing and the online estimation. The stage of pre-processing involves the process of establishing a correspondence among all the AIS points which come from the same vessel, registering them as a voyage and repeating the process for all the points in the database. These points are mapped to the space of navigation feature subsequently. This is summarised in Algorithm 3. During the online retrieval, given the current AIS location of the voyage, the information is transformed into the same representation as a navigation feature, which is used as the query to search for the nearest neighbour from the database of registered voyages subsequently. The distance to the nearest neighbour is returned as the risk estimation. This is summarised in Algorithm 4.

**Algorithm 3** Pre-processing.

1: **for** each voyage in the historical data **do**
2: Establish the series of navigation features and store
3: **end for**

**Algorithm 4** Risk estimation with nearest-neighbour based search.

1: Given a navigation feature as the query
2: Search for the closest points in the historical voyages
3: Return the distance between the query and the nearest neighbour

## 4 Experimental study

The details of the experimental study carried out are presented in this section.

### 4.1 Setup

The implementation of algorithm 1 to 4 was done in Python. Each online retrieval took less than five seconds. The three existing sets of historical data were separated randomly into two equal sets, one for training (or establishing the database during pre-processing) and another for testing.

### 4.2 Results

Table 1 shows the performance of the developed algorithms. The results are reported in the form of the mean and the standard deviation, for the distance of the query point to the closest voyage.

In addition, as a visualisation towards the method for monitoring the real-time risk of a voyage, figure 5 to 7 show an example of the risk estimation for the case of the Bass Strait, the Great Australian Bight and the North West. The red point in each of the figures refers to the query

point, green points refer to the closest voyage found and blue points are the AIS data in the training set.

Table 1: Performance of the risk estimation of a voyage.

| Dataset | Distance ($\sigma \pm \nu$) |
|---|---|
| Bass Strait | $0.0067 \pm 0.012$ |
| Great Australian Bight | $0.0095 \pm 0.014$ |
| North West | $0.0046 \pm 0.0064$ |

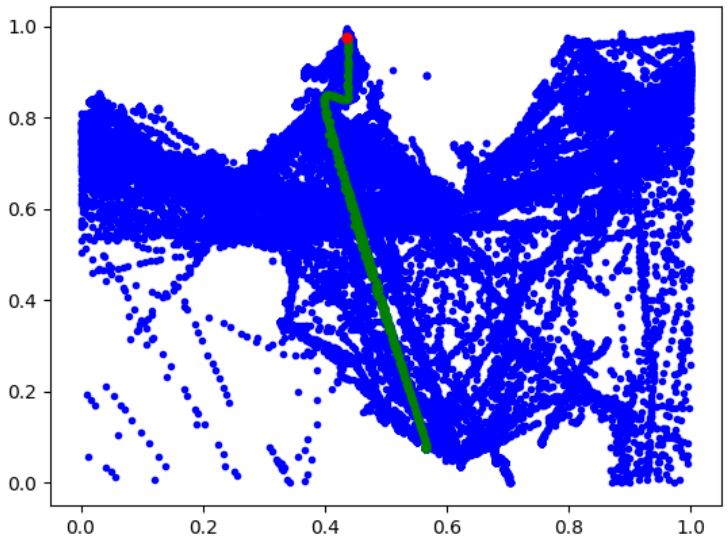



Figure 5: An example of the risk estimation for the Bass Strait.

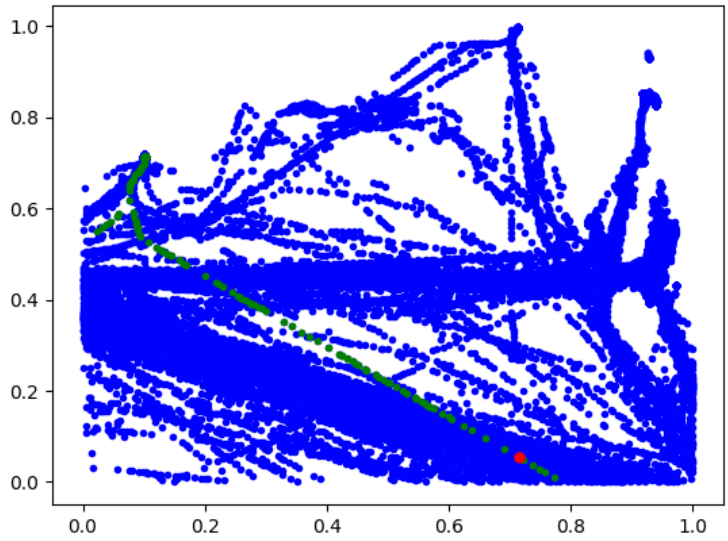



Figure 6: An example of the risk estimation for the Great Australian Bight.

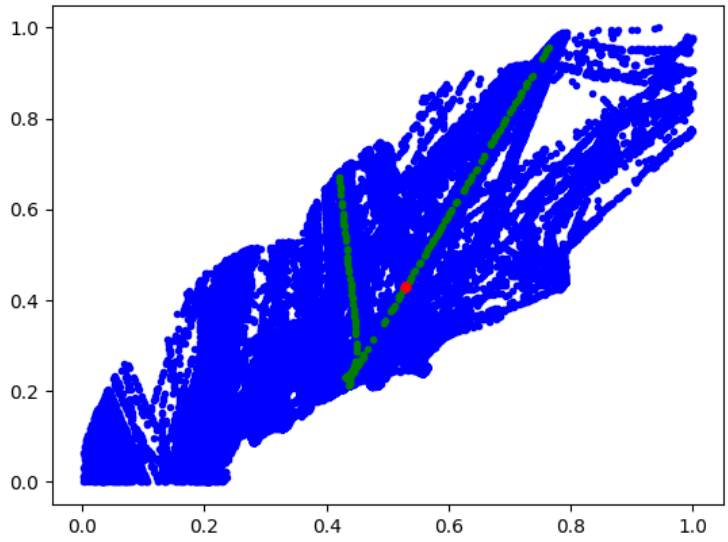



Figure 7: An example of the risk estimation for the North West.

## 5 Discussion

As demonstrated in the previous section, the developed method has demonstrated encouraging performance on the three existing sets of historical AIS data covering three Australian regions.

As one may notice, the distance of the query to the nearest neighbour found is returned as the estimated risk. There has not been a clear lower and upper bound determined for the navigational risk. A necessary step to take next is to identify a suitable lower and upper bound for

the variable. This can be done by a mapping of the risk to a range between 0 and 1, a limit could be determined from the historical AIS points.

While the developed method has demonstrated encouraging performance, there is a need to test further. In particular, it would be necessary to compare the risk estimation with those from experienced mariners before the further development and deployment to the industry. This work serves as the first step towards the possibility of a risk-based navigation for autonomous ships, a further understanding of the requirements from the relevant industries and the authorities would also be an essential step.

Besides, there is a need to consider more parameters when the data is available. These include and are not limited to weather conditions and the time. Mathematically, this translates into a representation that moves to a higher dimensional space, covering more parameters. Achieving a balance among the various parameters or dimensions would also be necessary. While the current version of the work maps the historical data in the range of 0 to 1 individually before further processing, a better or a more suitable form could benefit the developed algorithms significantly.

The interaction between the vessels is not taken into consideration in the current version of the work. In fact, for a complete assessment, the risk coming from the other vessels especially in terms of being too close and hence a collision would need to be considered. For example, having two vessels which are relatively close especially and high estimated risks could be a good indication.

Next, while the computations so far were relatively fast, the computational complexity is still of linear basis. For a real-time monitoring or operation, an improvement to a (near-) constant complexity in terms of the time for an online search is preferred. Recent breakthroughs in achieving a reasonably close search with some additional assumptions such as approximate near-neighbour search [2] and Locality Sensitive Hashing [7] may help.

## 6 Conclusion

In summary, this work presents an approach for the risk estimation of a navigation. Starting by mapping the list of historical AIS points into a list of corresponding voyages, each of the points in a voyage is transformed as a point in the space of navigation feature. After the pre-processing, given a query in the form of a navigation feature, a search for the nearest neighbour from the list of historical voyages is performed and the distance to the nearest neighbour in the list of the pre-processed historical voyages is returned as the estimated risk at the moment. The developed method has demonstrated encouraging performance on a challenging dataset of historical AIS data covering three regions in Australia, namely the Bass Strait, the Great Australian Bight and the North West.

## Acknowledgement

The support from ”National Natural Science Foundation of China” (Grant No. 11621202) and ”the Fundamental Research Funds for the Central Universities” (Grant No. WK2090000009) for the author Yanling Wu are gratefully acknowledged.

## References


[1] https://creativecommons.org/licenses/by-nc/3.0/au/.

[2] Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '06, pages 459–468, Washington, DC, USA, 2006. IEEE Computer Society.

[3] Australian Maritime Safety Authority.

[4] Australian Maritime Safety Authority. https://www.operations.amsa.gov.au/spatial/.

[5] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.

[6] M. Höyhtyä, J. Huusko, M. Kiviranta, K. Solberg, and J. Rokka. Connectivity for autonomous ships: Architecture, use cases, and research challenges. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 345–350, Oct 2017.

[7] Ke Jiang, Qichao Que, and Brian Kulis. Revisiting kernelized locality-sensitive hashing for improved large-scale image retrieval. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[8] T. A. Johansen and T. Perez. Unmanned aerial surveillance system for hazard collision avoidance in autonomous shipping. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1056–1065, June 2016.

[9] O. Levander. Autonomous ships on the high seas. *IEEE Spectrum*, 54(2):26–31, February 2017.

[10] International Association of Classification Societies. http://www.iacs.org.uk/.

[11] International Maritime Organisation. Convention on the international regulations for preventing collisions at sea, 1972 (colregs).

[12] International Maritime Organisation. International maritime organisation.

[13] Lokukaluge P. Perera and C. Guedes Soares. Collision risk detection and quantification in ship navigation with integrated bridge systems. *Ocean Engineering*, 109:344 – 354, 2015.

[14] Riccardo Polvara, Sanjay Sharma, Jian Wan, Andrew Manning, and Robert Sutton. Obstacle avoidance approaches for autonomous navigation of unmanned surface vehicles. *Journal of Navigation*, 71(1):241–256, 2018.

[15] Ørnulf Jan Rødseth and Hans-Christoph Burmeister. Risk assessment for an unmanned merchant ship. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 9(3):357–364, 2015.

[16] C. K. Tan, J. Wang, Y. C. Paw, and F. Liao. Autonomous ship deck landing of a quadrotor using invariant ellipsoid method. *IEEE Transactions on Aerospace and Electronic Systems*, 52(2):891–903, April 2016.

[17] W. C. Tan, I. Chen, S. J. Pan, and H. K. Tan. Analysis with histogram of connectivity: For automated evaluation of piping layout. *IEEE Transactions on Automation Science and Engineering*, 15(1):381–392, Jan 2018.

[18] W. C. Tan, I. Chen, D. Pantazis, and S. J. Pan. Transfer learning with pipnet: For automated visual analysis of piping design. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pages 1296–1301, Aug 2018.

[19] W. C. Tan, P. C. Goh, K. H. Chua, and I. Chen. Learning with corrosion feature: For automated quantitative risk analysis of corrosion mechanism. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pages 1290–1295, Aug 2018.

[20] W. C. Tan, C. Weng, Y. Zhou, K. H. Chua, and I. . Chen. Historical data is useful for navigation planning: Data driven route generation for autonomous ship. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7478–7483, May 2018.

[21] Ingrid Bouwer Utne, Asgeir J. Sørensen, and Ingrid Schjølberg. Risk management of autonomous marine systems and operations, 2017.

[22] Y. Wang, S. Wang, and M. Tan. Path generation of autonomous approach to a moving ship for unmanned vehicles. *IEEE Transactions on Industrial Electronics*, 62(9):5619–5629, Sep. 2015.

# Non-Equilibrium Two-Phase Flow Computations by a Mixture Model

**Eric Goncalves[1], †Dia Zeidan[2] and Lukas van Gemmeren, Peer Ueberholz, Peter Farber, Jens Gräbel[3]**

[1]ISAE-ENSMA, Institut Pprime, UPR 3346 CNRS, Poitiers, France
[2]School of Basic Sciences and Humanities, German Jordanian University, Amman, Jordan
[3]Institute for Modelling and High Performance Computing, Niederrhein University of Applied Sciences, Reinarzstraße 49, 47805 Krefeld, Germany
†Corresponding and presenting author: dia.zeidan@gju.edu.jo

## ABSTRACT

This work presents an assessment of the capabilities of mixture processes to solve two-phase flows developing shocks and discontinuities. The mixture equations are based on a two-phase flow model with full non-equilibrium processes expressed in conservative form. This mixture model is computed by using Godunov-type finite volume methods. A four shock waves test problem is simulated to highlight the performance of the proposed mixture model for one-dimensional compressible two-phase flows. The results from the simulations appear to be qualitatively in agreement with those available in literature.

**Keywords:** Gas-liquid, Non-equilibrium, Mixture equations, Shock waves, Simulation

## Introduction and Equations

The demand for two-phase fluid flow computations and their understanding is growing because of their attractive challenges in both basic research and engineering applications. Two-phase flows investigations have relied comprehensively on either the two-fluid model or homogeneous mixture model types. However, both types are always expensive due to their theoretical and physical nature in addition to their own difficulties [1, 2]. Alternatively, two-phase flows such as gas and liquid can be formulated in terms of mixture parameters of state. In this approach, mixture models are based on non-linear partial differential equations and are able to describe dynamically the evolution of non-equilibrium behaviour between the different phases. See [3, 4] and references therein. These models assume knowledge of several parameters, for instance, the relative motion between phases, and compressibility for all phases can be taken into account. These models also are practical because they contain more information about both the different phases and their combination, i.e., the mixture flow. In this framework, the mixture model consists of equations for the conservation of mixture mass $(\rho)$, conservation of mixture momentum $(\rho u)$, conservation of mixture energy $(\rho E)$ and a gas void fraction equation $(\rho\alpha)$, a gas mass fraction equation $(\rho c)$ and a relative velocity $(u_r)$ between the gas and liquid. These are written in a compact vector form as [5]:

$$\frac{\partial \mathbb{U}}{\partial t} + \frac{\partial \mathbb{F}(\mathbb{U})}{\partial x} = \mathbb{S}, \quad t > 0, \quad x \in \mathbb{R}. \tag{1}$$

where

$$\mathbb{U} = \begin{pmatrix} \rho \\ \rho\alpha \\ \rho u \\ \rho c \\ u_r \\ \rho E \end{pmatrix} \quad \text{and} \quad \mathbb{F}(\mathbb{U}) = \begin{pmatrix} \rho u \\ \rho u\alpha \\ \rho u^2 + P + \rho c(1-c)u_r^2 \\ \rho uc + \rho c(1-c)u_r \\ uu_r + \frac{1-2c}{2}u_r^2 + \psi(P) \\ \rho uE + Pu + \rho c(1-c)u_r\Big(uu_r + \frac{1-2c}{2}u_r^2 + \psi(P)\Big) \end{pmatrix}. \quad (2)$$

In the above, $\mathbb{U}$ is the vector of conservation variables; $\mathbb{F}$ and $\mathbb{S}$ are the mathematical flux function in the $x$-direction and the vector of source terms assumed to be at this time, respectively. Other notations denote $\rho$ = mixture density, $u$ = mixture velocity, $u_r$ = relative velocity, $E$ = mixture energy, $P$ = mixture pressure, $c$ = gas mass fraction and $\alpha$ = gas void fraction which satisfies $\alpha + (1-\alpha) = 1$. The term $\psi$ links the two phases through the relative velocity equation. The above system has the common form of a conservation law and fulfill an addition conservation law [6, 7]:

$$\frac{\partial}{\partial t}(\rho S) + \frac{\partial}{\partial x}(\rho S u) = Q, \quad (3)$$

which is the conservation of mixture entropy and $Q$ is the entropy introduction. Furthermore, each phase has its own thermodynamics properties described through different equations of state. An often employed equation of state in simulating realistic two-phase problems is the stiffened equations of state (EoS) [8]. This is due to its simplicity and ability to capture strong and weak shock-waves in addition to its resemblance to other equations of state. Thus, the EoS is given by:

$$P_j = K_j \left(\frac{\rho_j}{\bar{\rho}_j}\right)^{\gamma_j} \exp\left(\frac{S_j}{c_{j,v}}\right) - \bar{P}_j,$$

where $\gamma_j, K_j, c_{j,v}, \bar{\rho}_j$ and $\bar{P}_j$ are characteristic constants of the thermodynamic behaviour of each phase and $S_j$ is the entropy of the different phases [6, 7]. This equation becomes the EoS for the gas phase if $\bar{P}$ is set to zero. It is worth noting that the above mixture model is different from the homogeneous relaxation model and the homogeneous equilibrium model. Certainly, the model in hand processes advantages, for instance, the well-posedness and conservativity natures makes of interest to different applications.

## Computations and Results

The model equations (1)- (3) constitute a non-linear hyperbolic system written in a conservation form which can be solved by any numerical method of interest. These equations are solved by means of Godunov-type approach where the hyperbolic conservative left hand side is integrated using finite volume, high-resolution, shock-capturing methods. In a finite volume Godunov-type approach, there are mainly central and upwind intercell flux computations which are carried out by a discretization of a spatial computational domain and time computational domain of interest, respectively. In the context of Godunov-type centred methods, this discretization in processes without relaxation takes the following form [9]:

$$\mathbb{U}_i^{n+1} = \mathbb{U}_i^n - \frac{\Delta t}{\Delta x}\left(\mathbb{F}_{i+\frac{1}{2}}^n - \mathbb{F}_{i-\frac{1}{2}}^n\right), \quad (4)$$

where $\mathbb{U}_i^n$ denotes the integral average of the solution $\mathbb{U}$ and $\mathbb{F}$ is the numerical flux function which is a scheme-dependent function of the conservative variables. This is the Slope-Limited Centered (SLIC) scheme where the solution of the Riemann problem is fully numerical rather than analytical as in upwind methods. The SLIC scheme is a second-order in time and space and Total Variation Diminishing (TVD) using any limiter of interest. For further work on the SLIC scheme for fluid flow problems see, for example, [9]. To illustrate the type of outputs which the model equations produces for two-phase flow problems, a benchmark test is considered from the literature on the basis of the Riemann problem. Further, the SLIC scheme is employed for the resolution of shock waves problem presented in [7]. For this test problem, CFL $= 0.9$, SUPERBEE limiter along with transmissive boundary conditions are considered in the computational domain of $[-10, 10]$. Finally, simulation results are evaluated by comparing them with other numerical methods that do not depend on the structure of the Riemann problem. Numerical results are presented in figures 1. The results are displayed for three different numerical methods, namely, the Lax-Friedrichs which avoids solving the Riemann problem at every cell interface, First-Order Centered (FORCE) and Total Variation Diminishing (TVD) SLIC methods with $100$ coarse grid cells and compared with the reference solution which is provided on a very fine mesh of $5000$ cells. Clearly, the results agree well with the reference solutions and with those presented earlier in [7]. It is worth nothing that the relative velocity and different EoS have strong effect on the complete wave structure of this shock waves problem. Finally, the mixture equations together with the mixture entropy provided similar wave structure by the different numerical methods without any source terms effects.

## Concluding Remarks

A non-equilibrium fully compressible mixture model is presented and simulated with Riemann problem based methods. Four shock waves problem is tested with liquid water and vapour using real and ideal equations of state. Simulation results show that the mixture formulations together with the mixture entropy can accurately resolve the left and right shocks as well as contact discontinuities. It is observed that the model in hand agree well with the calculus of the eigenstructure of the system by providing six waves. This demonstrate the capabilities of mixture formulations to resolve two-phase flow discontinuities when using the mixture entropy equation. Ongoing and future research will include the simulation of rarefaction wave propagation using a temperature gradient.

## Acknowledgment

This project has received funding from the Scientific Research Support Fund, Ministry of Higher Education & Scientific Research through the German Jordanian University (Project Number Bas/1/05/2016).

## References

[1] Ishii M and Hibiki T, *Thermo-fluid Dynamics of Two-Phase Flow*, Springer, 2011.

[2] Yadigaroglu G, Hewitt GF, Introduction to Multiphase Flow: Basic Concepts, Applications and Modelling, Zurich Lectures on Multiphase Flow, Springer; 2018.

[3] Godunov SK and Romenski E, *Elements of Continuum Mechanics and Conservation Laws*, Kluwer Academic/Plenum Publishers, 2003.

[4] Baer M, Nunziato JW. A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. Int J Multiphase Flows 1986; 12: 861-889.

[5] Zeidan D, Romenski E, Slaouti A, Toro EF. Numerical study of wave propagation in compressible

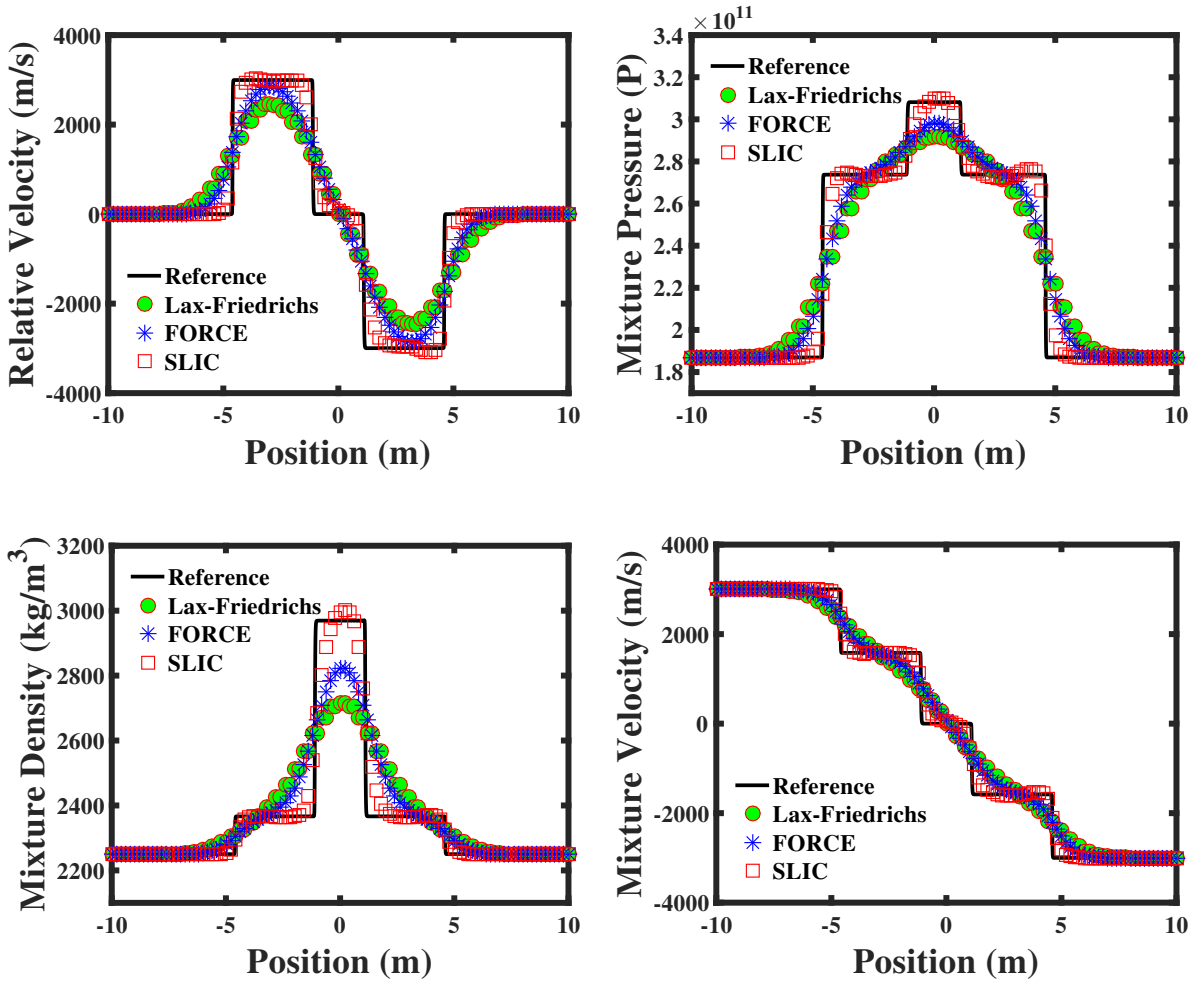



**Figure 1: Shock-tube problem of [7] at time $t = 9.0 \times 10^{-5} s$ ms. The TVD SLIC, FORCE and Lax-Friedrichs methods are compared with the reference solution results. Coarse meshes, symbols, are provided on $100$ cells and very fine meshes of $5000$ cells for the solid lines. The waves seen from left to right, two left shock and two right shock waves separated by a multiple contact discontinuity for the mixture velocity $u$.**

two-phase flow. Int J Numer Meth Fluids 2007; 54 (4): 393-417.

[6] Romenski E, Resnyansky AD, Toro EF. Conservative hyperbolic formulation for compressible two-phase flow with different phase pressures and temperatures. Quart Appl Math 2007; 65: 259-279.

[7] La Spina G, de'Michieli Vitturi M. High-Resolution Finite Volume Central Schemes for a Compressible Two-Phase Model. SIAM J Sci Comput 2012; 34(6), B861-B880.

[8] Cowperthwaite M. Relationships between incomplete equations of state. J.Franklin Inst 1969; 285: 379-387.

[9] Toro EF. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer, 2009.

# Morphological Rank-Space Segmentation of Clumped Filaments in Fluorescence Microscopy Imagery

**Yongjian Yu[1] and Jue Wang[2]**

[1]Axon Connected, LLC, Earlysville, VA, USA
[2]Department of Mathematics, Union College, Schenectady, NY, USA
Corresponding author: wangj@union.edu

## Abstract

We present a morphological rank-space segmentation method for automatically detecting and segmenting connected filamentous structures with application to unsupervised analysis of microscopic vaginal flora samples. This approach begins with a binarization process with adaptivity to local noise and object orientation. A novel morphological rank-space is proposed to decompose the connected filaments through morphological thinning, axial linking, labeling, and selection, resulting in a discrete rank-space representation. A ribbon delineation method is devised to reconstruct the boundaries of filaments through the rank-space. Our approach has been successfully applied to detect and segment the pseudohyphae for diagnosis of fungus Candida vaginitis.

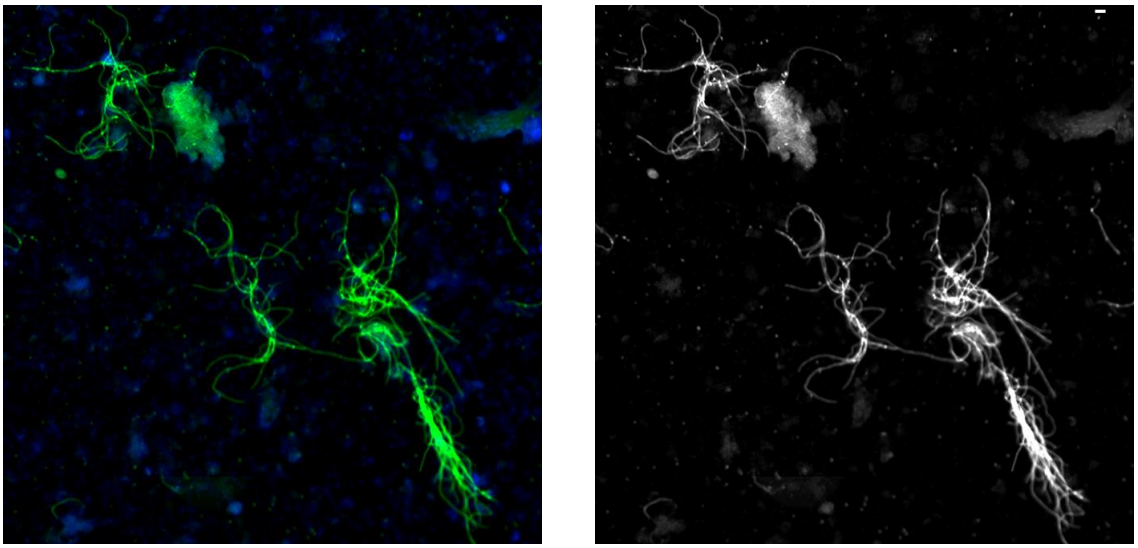
**Keywords:** Segmentation, morphology, rank-space, fluorescent microscope, filament, pseudohyphae

## 1. Introduction

Candida albicans is a pleiomorphic fungus and is the most common cause of vaginal fungal infections. Approximately 75% of women have at least one Candida vaginitis (CV) during their lifetimes while nearly half have at least two [4]. Candida overgrowth appears as mats of yeasts measuring 3 to 5 μm in diameter intermingled with pseudohyphae [6], that consist of invasive filaments comprising chains of conjoined, elongated yeast cells. Pseudohyphal growth is regarded as a defined developmental state and important fungal pathogen for CV [12]. The multiband fluorescent microscope is used to provide multi-band digital scans of sample microorganisms that have been pre-processed with immunofluorescent labeling or staining techniques. The fluorescence of the specific dyed microorganisms under the excitation of lights of various wavelengths better reveals the structures of cells of interest. Sample yeast cells and pseudohyphae are shown in Figure 1.

Manual finding and quantification of the pseudohyphae for CV screening and diagnosis is prohibitively time consuming. It is practically desired to develop automated techniques for segmenting the pseudohyphae from microscopic vaginal flora images towards providing visual evidence of fungal pathogens.

Unsupervised segmentation and quantification of the pseudohyphal structures is essential to perform automated high-throughput analysis, in order to compute, for instance, the distributions of lengths, widths, curvatures, budding scars, etc. The challenge for such unsupervised methods attributes to the fact that the pseudohyphae vary in shapes and sizes, form clumps, and the images often contain fiber contamination.



(a) (b)

**Figure 1.** A representative fluorescent image of yeast and pseudohyphae. (a) 470 nm green and 395 nm blue fluorophores; the former specifies the yeast and pseudohyphal cells and the latter indicates the nuclei (DNA). (b) The green channel.

The global thresholding methods are not appropriate for finding individual pseudohypha from multicellular clumps of pseudohyphae (colonies). The image decomposition approach through curvelet and wavelet transform is applied for segmenting actin filaments [2]. This method is inappropriate for extracting highly curved filaments due to poor computational efficiency as the so-called curvelet transform is actually the size-varying block-wise ridgelet transform through the scale space. Applying Hessain-based scale-space enhancement filters [5, 8] results in disconnections due to the low filter response along the bifurcation regions. These filters have been combined with tracking [3], bifurcation enhancement and suppression of non-tubular structures [15]. All enhancement filter methods require further decomposition and delineation. Other developments include fuzzy shape representation [1] and fuzzy connectedness [7]. An extensive review can be found in [10]. To date, not a single method is capable of successfully segmenting filamentous structures from every imaging modality and every organ.

In this paper, we explore innovative techniques using mathematical morphology and domain knowledge for automatically detecting and segmenting the pseudohyphae, or filamentous structures in general, from microscopic images. We present a paradigm composed of adaptive binarization, clue extraction, morphological rank-space transform, and rank-controlled ribbon contouring of filaments. The binarization is less sensitive to noise and object orientation. The new rank-space allows user specifications to be incorporated into the decomposition; a higher ranked filament is segmented with more integrity at places such as junctions. We demonstrate using clinical data that our system achieves visually agreeable results that enable the feasibility for providing assessment evidence for quick and accurate CV screening.

The rest of the paper is organized as follows. In section 2, we outline the framework for automatically detecting, segmenting, and quantifying pseudohyhae in the microscopic scans of vaginal flora samples. In section 3, we demonstrate the efficacy of our method using clinical data. Conclusions are drawn in section 4.

## 2. Methodology

Figure 2 depicts the overall structure of the process for segmenting connected filaments. The details of each block are described below.

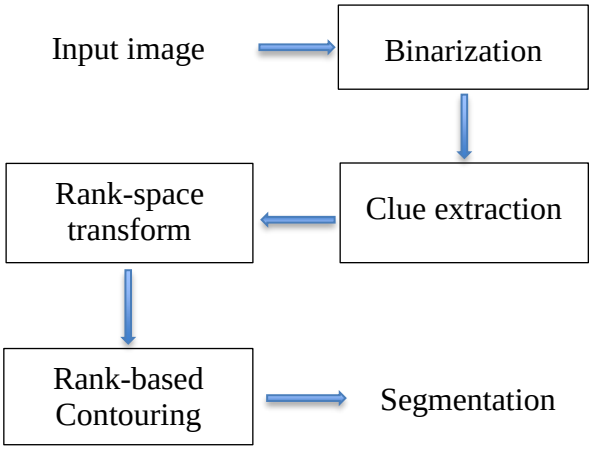



**Figure 2.** Flowchart of processing algorithms.

### 2.1 Binarization

The vaginal flora sample is labeled with the Caza Health® antibody package and illuminated with lights. A black and white digital camera captures and grabs the data. The image contains mainly the yeast, budding yeast, pseudohyphae and fibers (possibly from the swab used to collect the sample). The fluorescent microscopic scan must first undergo a series of pre-processing to correct image artifacts caused by channel cross-talk, LED non-uniform illumination, background auto-fluorescence, and so forth. The green channel (see Figure 1) captures most effectively the fluorescence emitted by the yeast and pseudohyphae.

The binarization process masks the interest areas of pseudohyphal colonies, resulting in a binary ridge image. This procedure adopts local adaptive thresholding algorithms to achieve the robustness to the local noise variation. The detection technique needs to be sensitive to the object orientation, as well as lowering the impact of interfering objects present in the test window upon the noise statistics. It is accomplished through the following steps. We slide $M$ tilted line-segment windows over the image domain; at each point, we collect $N$ reference samples per window to compute local statistics as a function of the window tilt angle. We then select the window that is most orthogonal to the pseudohyphal axis direction, and implement the adaptive threshold test, as described by

$$I - Z_{\perp} > C \cdot \left(1 + \left(\frac{Z_{\perp}}{Z_{max}}\right)^{r}\right), \tag{1}$$

where $I$ is the green channel image with intensity normalized to [0, 1]. $Z_{\perp}$ corresponds to the test statistics obtained in the window that is most orthogonal to the pseudohyphal segment,

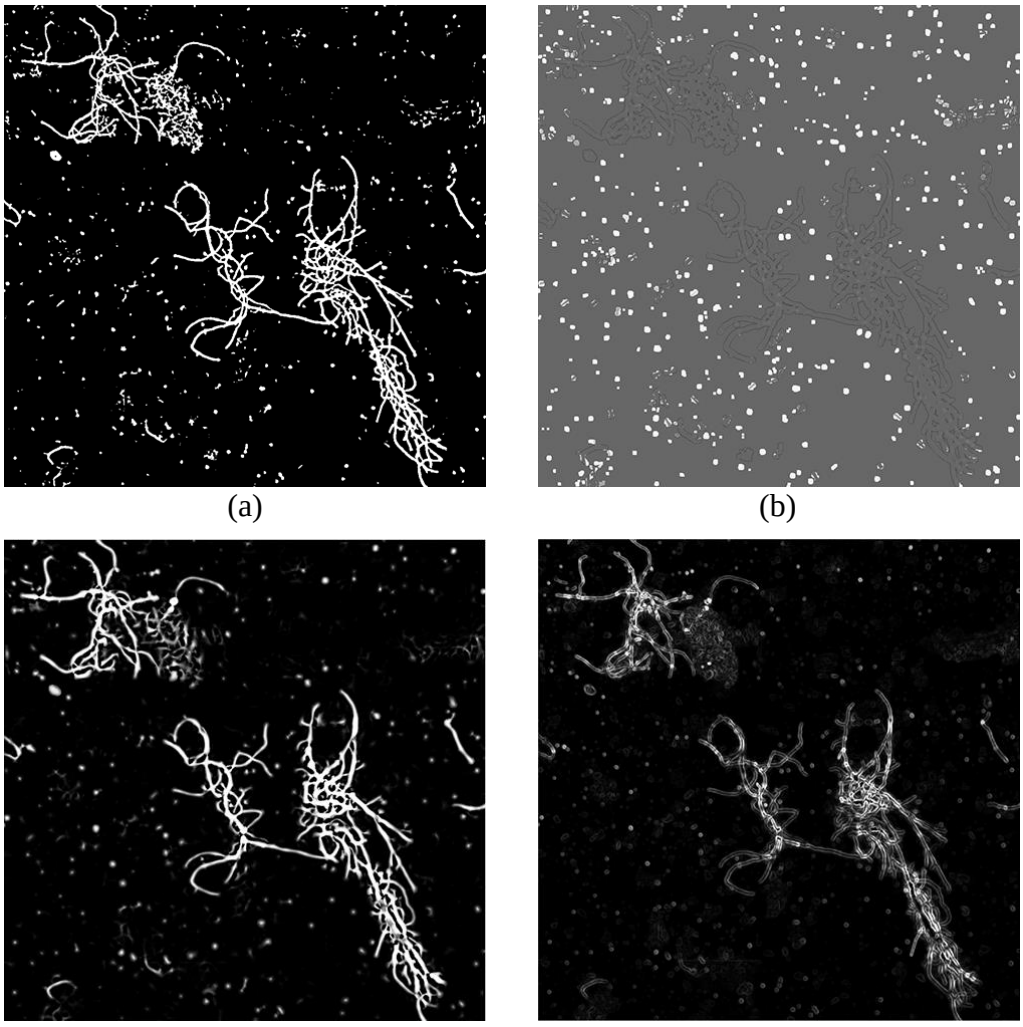
$$Z_{\perp} := \min_{1 \le k \le M} Z_k,$$

$$Z_{max} = \max_{1 \le k \le M} Z_k;$$

$Z_k$ is the mean intensity in the $N$-pixel line segment window $W_k$ centered at position $(i, j)$ in the image domain with a tilt angle of $\frac{k-1}{M}\pi$ from the $x$-axis, *i.e.*,

$$Z_k = \text{mean}\{I_{ij} \in W_k\}.$$

The $r$ in Eq. (1) is a positive real factor; $C$ is a threshold that is proportional to the noise standard deviation. It is set to maintain an acceptable filament detection probability for low SNR with a consistent false detection rate. The orientation adaptivity manifests through the multiplier of $C$. For elongated structures, $Z_{max} > Z_{\perp}$, and the right-hand side of Eq. (1) is approximately equal to $C$. For round objects, $Z_{max} = Z_{\perp}$, and the right-hand side is doubled $C$ so that the round or close to round ones are not being detected as pseudohyphae. In this study, we choose the parameters experimentally: $M$ = 8~16, $N$ = 15, $C$ = 0.01~0.05, and $r$ = 2. The resulting binary ridge mask is shown in Figure 3(a).



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

**Figure 3.** (a) Binary ridge mask, (b) Spotness map, (c) Ridgeness map, (d) Gradient magnitude of Figure 1(b).

## 2.2 Spotness and ridgeness features

We adopt a newly devised *spotness* feature [14] and *ridgeness* feature for providing segmentation criteria for the next section. They are illustrated in Figure 3 (b) and (c). The spotness appears bright on compact structures such as yeast while showing dark boundaries on elongated structures such as pseudohyphae. Thus, the spotness feature is tailored to discriminating between yeast and pseudohyphae. The spotness image is a rotational morphological top-hat transform of an input image $I$, defined as

$$p = I - \max\{\gamma_{i,B}(I)\}_{i=0}^{K-1}, \tag{2}$$

where $\gamma_{i,B}(I) := R_{-\theta_i}\left(R_{\theta_i}(I) \circ B\right)$; $R_{\theta_i}$ is the rotation w.r.t. the image center by an angle $\theta_i := i\frac{\pi}{K}$; $K$ is the number of angles. The operator $\circ$ denotes opening by a horizontal line structure element $B$ of $l$-pixel long and 1-pixel wide. The second term in Eq. (2) suppresses spot structures of size less than $l$ while preserving the interiors and augmenting the borders of filamentous structures of dimension longer than $l$, regardless of orientations. In our study, we set $K$ = 36 and $l$ = 20.

The ridgeness feature is based on the multi-resolution enhancement filter response [5]. For an input image $I$, the Hessian matrix $H$ is defined as

$$H_{ij}(x,\sigma) = \sigma^2 I(x) * \frac{\partial^2}{\partial x_i \partial x_j} G_\sigma(x),\ i,j = 1,2, \tag{3}$$

where $G_\sigma(x)$ is the 2D Gaussian of standard deviation $\sigma$ and * denotes convolution. The eigenvalues of $H$ are sorted, $|\lambda_1| \leq |\lambda_2|$. The larger intensity variation takes place along the second eigenvector, corresponding to the direction across the intensity boundary. To enhance the elongated more than the spherical structures, a ridgeness indicator function is computed as

$$R = \max_{\sigma \in D} e^{-\frac{(\lambda_2/\lambda_1)^2}{2\alpha^2}} \left(1 - e^{-\frac{\sqrt{\lambda_1^2+\lambda_2^2}}{2\beta^2}}\right), \tag{4}$$

where $\alpha$ and $\beta$ are sensitivity parameters. The set $D$ is preset according to the expected range of filament width. $R$ lies between 0 and 1. We set $\alpha$ = 2, $\beta$ = 8, and $D$ = {1,2,3,4}.

### 2.3 Segmentation of connected filaments

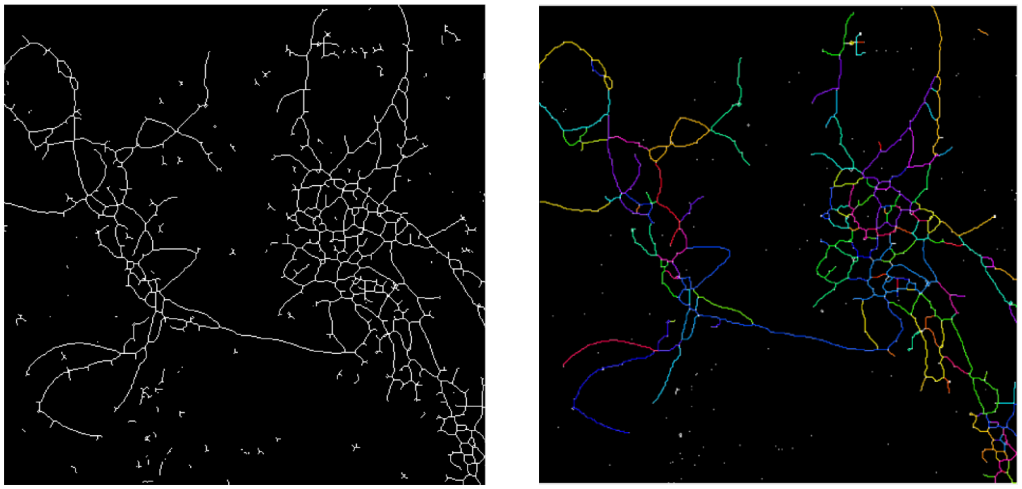
The pseudohyphae tend to clump together in the developed state of Candida infection. In order to assess the infection by taking the geometrical and other measurements of individual pseudohypha, it is necessary to split and segment the clumped cells. Our segmentation algorithm comprises two steps: analysis and synthesis. In the first step, the binary mask of ridges is decomposed into distinct, thin ridge markers of descending rank-scale orders; each marker represents a disjoint cell axis (or central line). The synthesis of the rank-scale space restores the filamentous shapes while preventing the merging of the boundary contours of split ridge markers.

The analysis procedure is described as follows. We first extract appropriate pseudohypha-like ridge markers by morphological thinning and pruning of the binarized yeast mask map, and then link the 1-pixel wide ridge elements to form chains of connected ridges. Finally, a morphological rank-space representation of pseudohypha-like filaments is generated by conditioned, scale-rank prioritized reconstruction and filtering based on significance metrics and thresholds derived from clinical relevance.

The thinning algorithm removes pixels on the boundaries of connected components without breaking them apart. This operation is repeated until the image is stable, followed by morphological pruning that removes the end points of lines without removing small objects completely.

The linking algorithm takes the thinned binary mask of ridges and ridge direction angle image (obtained from multi-resolution enhancement filter), and outputs a list of labeled ridge chains, as shown in Figure 4. The ridge-linking algorithm is outlined below:

1) Start a row-wise raster scanning of the binary ridge map from the upper left corner, *i.e.*, the image origin.
2) Follow the ridge with 8-connectivity to the end. Search through its neighbors within a preset distance for the points that are most compatible with respect to the ridge direction.
3) A chain of ridges either terminates at a joint or branch to grow if its continuity is stronger than that of other chains connected to the same joint.
4) Flip the tracked pixels, and track the rest of the ridge segment, if any, to the end of the ridge chain.
5) Finally make sure that the starting point of each ridge chain is closer to the image origin than its end.



(a) (b)

**Figure 4.** (a) Zoomed-in map of thinned ridges representing the axial lines of the filamentous pseudohyphae. (b) Segmented ridge central lines with random color-labeling.

The morphological rank-space filament clump decomposition depends on the priority criteria that are devised specifically according to application. We elaborate in this paper the morphological multi-rank space with focus on segmenting pseudohyphae. This paradigm is applicable to general curved filament segmentation tasks.

Three metrics are measured for each ridge marker: the length, average fluorescence intensity, and average ridgeness. The prioritizing algorithm performs as follows:

1) Each ridge chain receives three separate scores $\{t_i\}_{i=1,2,3}$, respectively, according to its index in the descending order of length, average fluorescent intensity, and average ridgeness.
2) A total score $t = \sum_{i=1}^{3} t_i$ is assigned to each ridge chain. The ridge chains are then sorted again in the ascending order based on their total scores.
3) The first $P$ ridge chains are selected from the final sorted list. The number $P$ depends on a predefined maximum number of pseudohyphae.

With these metrics, the long, bright ridges (or filaments) receive a higher priority than short, dark, non-filamentous ones (such as edges). We define the rank of a ridge to be its index of the descending total scores. In the rank-space, the originally crowded ridge markers are split

and redistributed into multiple rank subspaces. The markers with higher priority reside sparsely in the high-rank subspace.

In the synthesis process, the exterior boundaries of the disjoint filamentous structures are reconstructed in a bottom-up manner through the rank-space. Starting from ridge marker(s) with the highest rank, we delineate the boundary contours of filaments that have the same rank *s*, then append them to the output sequence indexed with ranks $(1, 2, \dots, s-1)$. During this process, contours do not merge. The algorithm is provided below.

1) Compute the gradient magnitude $g(x)$ of the green channel image. An example is shown in Figure 3(d).
2) Remove 1 pixel at the joint end of each line marker of rank *s* resulting in a trimmed ridge marker. Then compute a rank-based binary influence zone $B_s(x)$, which is given by the watershed transform of the distance map of the union of trimmed ridge markers at rank *s* and the already-segmented filament ribbons from rank 1 to $s-1$.
3) At rank *s*, loop through each point along the curve marker, and search in the direction perpendicular to the curve, within the influence zone $B_s(x)$, to locate the maximum gradient magnitude (boundary) points on both sides of the curve marker. The boundary contour of the marker is described by a marker curve $\{x_{s,i}; i = 1, \dots, l_s\}$, where $x_{s,i}$ represents the row and column index of the $i^{\text{th}}$ ridge point along the central line of rank *s*; $l_s$ is the length of that marker. The variable radius is given by

$$\hat{r}_{s,i} = \min\left\{\max_{r_i}\{g(x_i^+)\}, \max_{r_i}\{g(x_i^-)\}\right\}, \tag{5}$$

where $x_i^{\pm} = x_i \pm r_i \frac{(x_i - x_{i-1})^{\perp}}{|x_i - x_{i-1}|}$; the symbol $\perp$ denotes transverse of a vector. For simplicity the subscript *s* is omitted. The geometry is illustrated in Figure 5.
4) Append $\{(x_{s,i}, \hat{r}_{s,i}); i = 1, \dots, l_s\}$ to the previously reconstructed contours.
5) Compute the average width and minimum spotness for each segmented filament. The combination of length, average fluorescence intensity, average ridgeness, average width, and minimum spotness fully characterizes each segmented pseudohyphal suspect.
6) Finally, identify the pseudohyphal filaments by means of length thresholding or width thresholding.

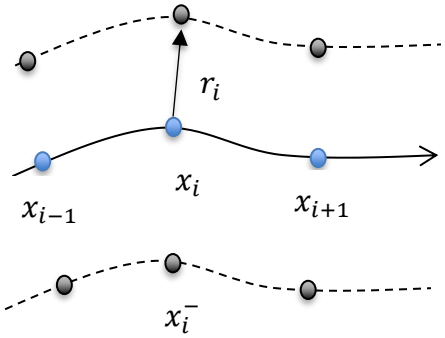



**Figure 5.** Geometry of filament segmentation.

## 3. Results

We evaluate the morphological rank-space segmentation method using two sets of images. In the first test, images of size 1024 by 1024 pixels were acquired using a Nikon inverted

microscope, charge-coupled device camera and 4X objective lens. The proposed method is applied to a selection of the images. The magnified results are shown in Figure 6.

We compare our method to the marker-controlled watershed transform [9] and morphological multiscale decomposition (MSD) [11, 13]. Our approach outperforms the other methods in terms of preserving the integrality of long filaments and providing visually agreeable splitting of clumped cells. An average ribbon width is computed and adopted per pseudohypha display for easy visualization in comparison. The actual ribbon size distribution can be used for pseudohyphae quantification. The 15-scale MSD as applied to the binary ridge detection breaks the long continuous filament into multiple segments. The watershed transform fails with the same markers as used in our method.

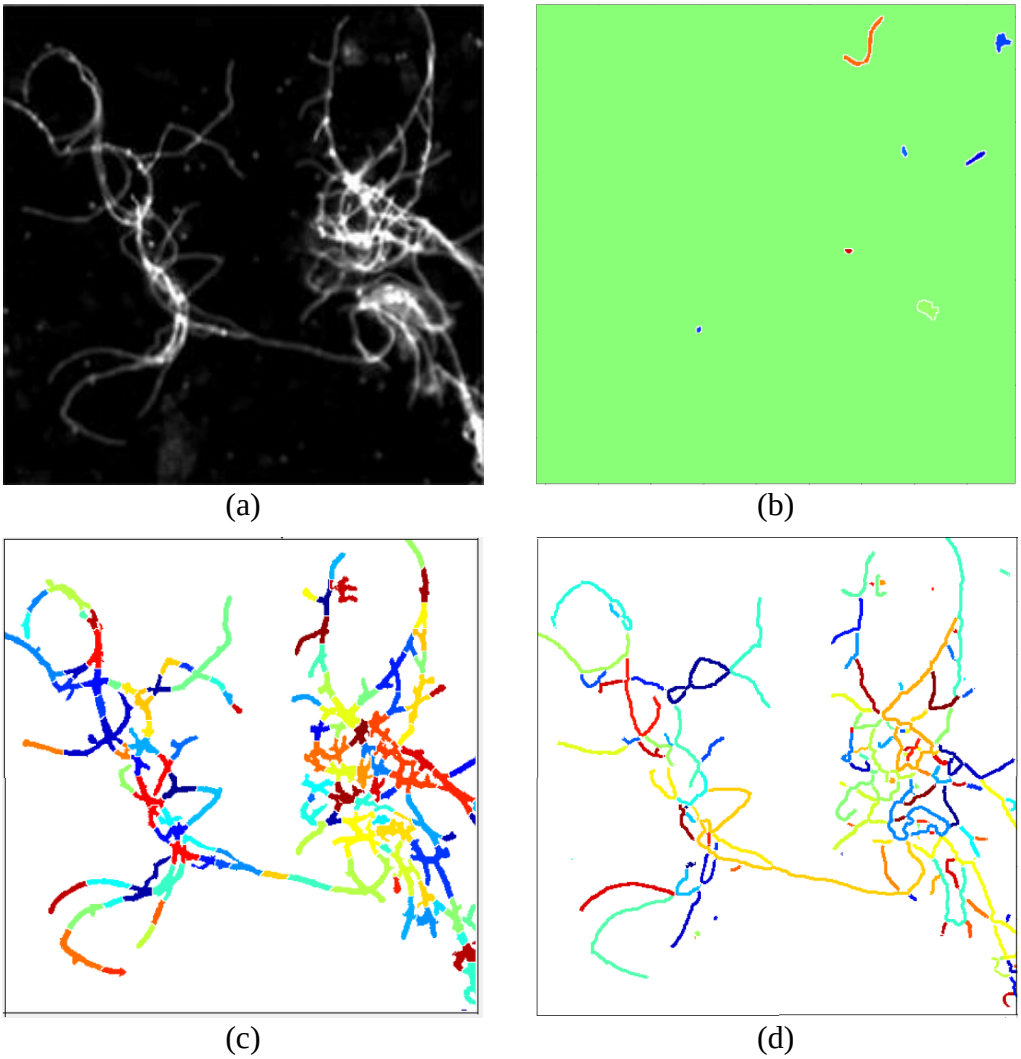
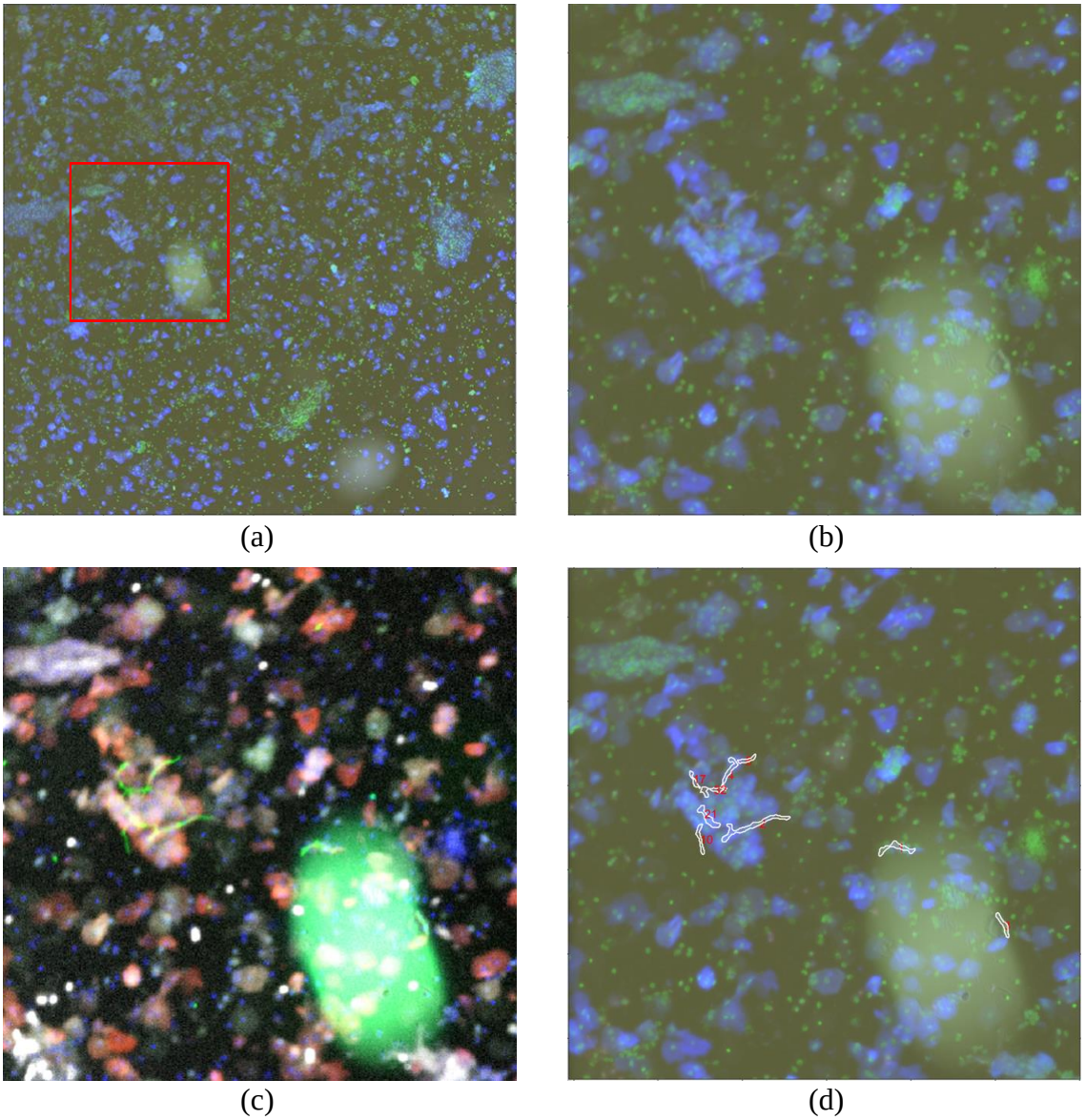



**Figure 6.** Method comparison. (a) Input image, (b) Marker based watershed, (c) 15-scale MSD, (d) Our method.

In the second test, we apply the method to clinical vaginal samples from Discovery Life Sciences (DLS), Inc. The samples are prepared using Axon Pac and scanned using nCyte microscope. The image size is 2048 by 2048 pixels. One of the test images is shown in Figure 7(a). In this fluorescent image, the pseudohyphae are highlighted in red, the nuclei (DNA) in green, the epithelial tissue cells in blue, and trich markers in white. Our analysis is performed

for the entire image. For detail viewing, magnified displays of the region in the red box are illustrated in Figures 7(b)-(d).

A color-enhanced view is shown in Figure 7(c). In this view, different color encoding allows the pseudohyphae to be highlighted in green so that the viewer can see the ground truth more clearly. The blue dots are nuclei (DNA); the red are epithelial tissue cells; the white blobs are trich markers. In Figure 7(d), the pseudohyphae detected and segmented by our method are highlighted as white contours and overlaid on the original image, along with their count numbers. This provides key information for CV diagnosis. A comparison of Figure 7(d) with 7(c) indicates that our algorithm achieves consistent performance.



(a) (b)

(c) (d)

**Figure 7.** Test on DLS clinical vaginal sample scan. (a) 2048 by 2048 input image. (b) Magnified view of the red box in (a). The pseudohyphae are highlighted in red, nuclei (DNA) in green, epithelial cells in blue, and trich markers in white. (c) Color-enhanced view. The pseudohyphae are highlighted in green, nuclei (DNA) in blue, epithelial cells in red, and trich markers in white. (d) The pseudohyphae detected and segmented by our method are overlaid as white contours, along with their count numbers in red.

## 4. Conclusions

We have demonstrated a generic morphological rank-space decomposition technique for segmenting connected filamentous structures. This approach enables a user-specific interest priority and natural scales to be integrated in a novel segmentation paradigm. Our method is evaluated using clinical fluorescent microscopic images of vaginal samples. It is computationally efficient. The results are visually expected, consistent and promising. They provide clues for diagnosis and aid in quantitative Candida vaginitis screening.

## References

[1] G. Agam, S.G. Armato and C. Wu, "Vessel tree reconstruction in thoracic CT scans with application to nodule detection", *IEEE Trans. Med. Imag.*, vol. 24, no. 4, pp. 486-499, 2005.

[2] M. Alisocha-Perez, C. Benadiba, K. Goossens, S. Kasas, G. Dietler, R. Willaert and H. Sahli, "A robust actin filaments image analysis framework", *PLOS Computational Biology*, 1005063, 2016.

[3] S. Aylward and E. Bullitt, "Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction", *IEEE Tans. Med. Imag.*, vol. 21, no. 2, pp. 61-75, 2002.

[4] M.E. Egan and M.S. Lipsky, "Diagnosis of vaginitis", *Am. Fam. Physician*, vol. 62, no. 5, pp. 1095-104, 2000.

[5] A.F. Frangi, W.J. Niessen, K.L. Vincken and M.A. Viergever, "Multiscale vessel enhancement filtering", *Proc. Med. Image Comput. Assist. Interv.*, vol. 1496, pp. 130-137, 1998.

[6] J. Guarner and M.E. Brandt, "Histopathologic diagnosis of fungal infections in the 21st century", *Clin. Microbiol. Rev.*, vol. 24, no. 2, pp. 247-280, 2011.

[7] J.N. Kaftan, A.P. Kiraly, A. Bakai, M. Das, C.L. Novak and T. Aach, "Fuzzy pulmonary vessel segmentation in contrast enhanced CT data", *Proc. SPIE Medical Imaging 2008: Image Processing*, vol. 6914, 2008.

[8] K. Krissian, G. Malandain, N. Ayache, R. Vaillant and Y. Trousset, "Model based detection of tubular structures in 3D images", *Comp. Vis. Image Understand.*, vol. 80, no. 2, pp. 130-171, 2000.

[9] C. Rambabu and I. Chakrabarti, "An efficient immersion-based watershed transform and its prototype architecture", *J. Syst. Archit.*, vol. 53, no. 4, pp. 210-226, 2007.

[10] D.R. Rina *et. al.*, "Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study", *Med. Image Anal.*, vol. 18, no. 7, pp. 1217-1232, 2014.

[11] O. Schmitt and M. Hasse, "Morphological multiscale decomposition of connected regions with emphasis on cell clusters", *Comput Vis. Image Underst.*, vol. 113, no. 2, pp. 188-210, 2009.

[12] V. Veses and N.A.R. Gow, "Pseudohypha budding patterns of Candida albicans", *Medical Mycology*, vol. 47, no. 3, pp. 268-275, 2009.

[13] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review", *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234-263, 2016.

[14] Y. Yu and J. Wang, "Automated enumeration and classification of bacteria in fluorescent microscopy imagery", *Proc. IEEE LSC*, 2018.

[15] C. Zhou, H.P. Chan, B, Sahiner, L.M. Hadjiiski, A. Chughtai, S. Patel, J. Wei, J. Ge, P.N. Cascade and E.A. Kazerooni, "Automated multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications", *Med. Phys.*, vol. 34, no. 12, pp. 4567-4577, 2007.

# Chemical structure and adhesion of DLC film with amorphous SiC interlayer

**†*Xia Zhu[1], Hiromichi Toyota[1], Daiki Muranaka[1],**

**Yutaroh Kimura[1] and Ryoya Shiraishi[1]**

[1]Department of Mechanical Engineering, Faculty of Engineering, Ehime University, JP

*Presenting author: zhu.xia.mx@ehime-u.ac.jp
†Corresponding author: zhu.xia.mx@ehime-u.ac.jp

## Abstract

This study was performed to improve the adhesiveness of a diamond-like carbon (DLC, a-C:H) multilayer film with an a-SiC interlayer. DLC (a-C: H) multilayer film with an a-SiC interlayer were deposited, and residual stresses the influence of the residual stresses on the adhesion of the film were investigated. The following conclusions can be drawn from the results presented herein: first, the a-SiC interlayer affects the DLC film structure. When the thickness of the interlayer t <= 0.25 m μm, the proportion of sp 3 bonds in the DLC film is approximately constant and about 10%. In the case of t> 0.25 μm, the proportion of sp3 bonds in the DLC film tends to decrease slightly. Second residual compressive stress exists in the DLC film, but residual tensile stress or compressive residual shrinkage stress exists in the a-SiC interlayer. Third, the residual stress of the DLC film is not due to the peeling failure of the film. Fourth, the residual tensile stress of the a-SiC interlayer does not cause the peeling failure of the film, but a part of the compressive residual stress causes the local buckling of the interlayer film and may also cause the peeling failure of the film. Finally, the strength and rigidity of the a-SiC interlayer dominate the adhesion between the a-SiC-DLC multilayer film and the substrate.

**Keywords:** Diamond-like carbon film, Amorphous silicon carbide interlayer, Raman spectral analysis, I (D) / I (G) ratio, FWHM (G), residual stress

## Introduction

Diamond-like carbon (DLC) is an amorphous carbon material containing a mixture of sp2 hybridized graphite, and sp3 hybridized diamond. DLC exhibits good abrasion resistance and high chemical safety due to its diamond content and flatness/low counterbody aggressiveness due to the presence of graphite. Thus, DLC is expected to be useful as a coating material [1]. However, a large obstacle for applications of DLC film is its low adhesiveness to various substrates, which is caused by two main factors: intrinsic residual stress and stable carbon bonds [2]. Several reports [3] have detailed attempts to prevent layer separation and improve the adhesiveness of DLC thin layers by forming an interlayer of silicon carbide (SiC) with high substrate adhesiveness, and then covering the interlayer with the DLC. However, uncertainties remain regarding the optimal conditions for forming a SiC thin film and DLC/a-SiC stacked thin layer. In particular, no reports have been published to date regarding the effects of an a-SiC thin film on the structure and film thickness. The goal of this study is to improve the adhesiveness of a DLC/a-SiC layered film formed by ion plating. In previous studies, the formation of a hydrogen-containing DLC (a-C: H) /a-SiC layered thin film was achieved, and the effects of an a-SiC interlayer on its structure and mechanical properties were determined.

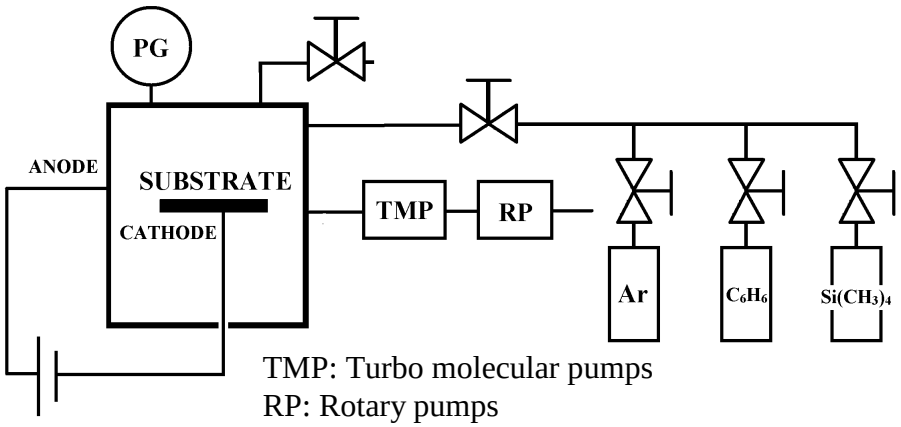



Fig. 1 Schematic diagram of the experimental setup for the ion plating method with introduction of reactive gas

## Film Deposition and Evaluation Methods

*Experimental Apparatus for Film Deposition and Evaluation Methods*

The ion plating methodwas used to form the a-SiC interlayer and DLC/a-SiC stacked layer. A schematic of the experimental apparatus is shown in Fig. 1, consisting of a vacuum chamber, rotary pump, turbo molecular pump, high-voltage power device, gas supply system, negative electrode substrate, and positive electrode chamber inner wall. Using rotary and turbo molecular pumps, it was possible to reduce the pressure in the chamber to 1.0 x 10-3 Pa. We built a power device consisting of a variable pressure system/bridge rectifying circuit that was capable of outputting a maximum of 500 V. The gas supply system could supply argon (Ar), benzene (C6H6), and tetramethylsilane (TMS, Si (CH3)4).
A stainless steel SUS304 (JIS G 4304, width 10 x length 30 x thickness 1.5 mm) substrate was pulverized and ultrasonically cleaned in an acetone solution. Subsequently, the substrate was fastened inside the chamber and the pressure inside the chamber was reduced using rotary and turbo molecular pumps. Inert gas was introduced into the chamber and the pressure adjusted. Voltage was applied to turn the inert gas into plasma and the substrate was washed by sputtering. To form the Si-C layer, tetramethylsilane was introduced into the chamber, turned to plasma, and deposited onto the substrate as an interlayer. For the DLC layer, benzene was introduced into the chamber, turned into plasma, and deposited on the substrate.
In this experiment, argon/TMS (mixing ratio of 1:1) gas was introduced, an interlayer was formed for 0 to 3 min, and a-SiC films 0.05 to 0.45 μm in thickness were formed. Subsequently, argon and benzene (mixing ratio of 1:1) were introduced, and a DLC film with a thickness of 2 μm was formed in 30 min.

*Chemical Structure Analysis on film*

Visible Raman spectral analysis is effective for analyzing the structure of DLC films [4-10]. Typical Raman spectra were observed for the prepared DLC when as shown in Fig. 2 [11]. From Fig. 2, the Raman spectrum of the DLC film could be divided into a D peak near 1350 cm-1 arising from the stretching and contracting of the six membered ring, and a G peak between 1500 and 1600 $cm^{-1}$ originating from the vibrations of the sp2 carbons on the chain and ring. The G-peak position and the FWHM (G) were used as parameters to qualitatively evaluate the degree of amorphous nature of the film (or crystallization). In addition, the hydrogen density of the film was evaluated using the photoluminescence (PL) of the Raman

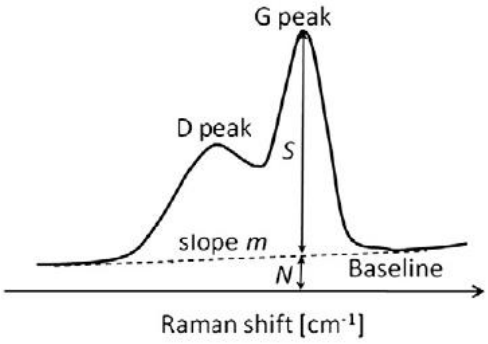



Fig. 2 Raman peak obtained for the DLC film [11]

curve. The baseline slope m [8] of the Raman spectrum and the ratio (log N/S) [9] of the photoluminescence component N to the strength of the Raman scattering at the G peak position were used as Raman parameters. In this study, a Raman spectrometer was used to evaluate the structure of the film with an oscillating Ar ion laser (514 nm), laser output of 1.0 mW, and an analysis range of 20 μm Φ.

*Calculation of residual stress in film*

The average residual stress $\sigma_i$ of a film can be calculated from the curvature of the substrate (substrate deformation method), that is, the radius of curvature using the following Stoney equation (1) [12].

$$\sigma_i = -\frac{E_s t_s^2}{3(1-\nu_s)t_F} \cdot \frac{1}{R} \tag{1}$$

Where, E s and ν s are the Young's modulus and Poisson's ratio of the substrate, t s and t F are the substrate thickness and thin film thickness, respectively, and R is the curvature radius of substrate warpage. The residual stress can be determined by measuring the curvature R of the specimen on which DLC is deposited. Here, assuming that the substrate has a convex shape upward, the curvature of the R is defined as positive, and the negative $\sigma_i$ indicates that the residual stress is a compressive stress. In order to measure R, a contact-type surface roughness tester was used.

In this research, the DLC film deposited is a film of two layers from the a-SiC intermediate layer and the DLC film, and the formula (1) cannot calculate the residual stress of each layer. Therefore, the residual stress was calculated using the residual stress calculation equation (2) corresponding to the film of the two-layer structure proposed by Nakamura et al. [13].

$$\sigma_F = \sigma_i \cdot \frac{t_i}{t_F} - \frac{E_S t_s^2}{3(1-\nu_s)t_F} \cdot \frac{1}{R} \tag{2}$$

where, $\sigma_i$ is a residual stress of the intermediate layer, and $\sigma_F$ is a residual stress of the outermost layer. Therefore, the residual stress of the intermediate layer can be calculated by equation (1), the radius of curvature of the substrate of the thin film having a two-layer structure can be measured, and the residual stress of the outermost layer can be calculated using equation (2). In the case of the test piece with only the a-SiC intermediate layer, there was a case that it

did not have a clear arc shape because the deformation of the substrate was partially waved. The stress value was set to zero for such specimens.

*Evaluation of adhesion between film and substrate*

During the film deposition, a residual stress generated inside the film reduces the adhesion between the film and the substrate, and in some cases, the film peels off the substrate. In this study, the adverse effect of the residual tensile stress on film adhesion is evaluated by a stress intensity factor. On the other hand, the adverse effect of the compressive residual stress is evaluated by a buckling load. Assuming that the mode of a crack surface displacement due to tensile residual stress is the opening mode, the stress intensity factor $K_I$ is expressed by equation (3) [14].

$$K_{\mathrm{I}} = \sigma \sqrt{\pi a} \tag{3}$$

where, σ is an average stress in the film, and a is a length of the latent crack in the film.
In the a-SiC intermediate layer, it is an internal penetration crack with the length $a = 1 / 2t_{a\text{-}SiC}$, and in the DLC film, it is a surface penetration crack with the length $a = t_{DLC}$. If the stress intensity factor is large, the risk of film peeling due to crack growth due to the tensile stress increases. On the other hand, even if a compressive stress is applied to the crack surface, crack propagation does not occur, so the stress intensity factor in that case is set to zero. However, the compressive stress may cause the film to buckle and also cause the film to peel off the substrate.
The Euler buckling load Pc of a beam fixed at both ends is expressed by the equation (4) [15].

$$P_{\mathrm{c}} = \frac{4\pi^2 EI}{l^2} \tag{4}$$

Where, E is an elasticity modulus, I is a second moment of area, and l is a length of the beam. In this study, the elastic modulus of a-SiC interlayer and DLC film were set to 80 GPa and 100 GPa, respectively. In addition, l was 10 μm in consideration of the surface roughness of the substrate. As the compressive load on the cross section of the film due to residual the compressive stress approaches the Euler buckling load, buckling of the film is more likely to occur.

## Results and Discussion

*Changes in chemical structure*

Generally, it is possible to gain information regarding the structure of the film and chemical bonding of the carbon atoms from the position of the G-band, the behavior of the I (D)/I (G), and FWHM(G), which can be obtained from the results of the Raman spectrum. Ferrari et al. [6] developed a model wherein upon introduction of defects into graphite to create disorder, an amorphization trajectory can be classified into three stages according to strength (Fig. 3). Specifically, from the monocrystalline to nanocrystal graphite, this model describes the series of disordering processes, with a third stage leading to the formation of tetrahedral amorphous carbon (ta-C) with $sp^3$ bonding, via amorphous carbon (a-C), which contains mainly $sp^2$ bonds. These structural and chemical bonding state changes can be observed in the visible (514 nm) Raman spectrum. The G-peak is a parameter depends on the length of the $sp^2$ bonds between the carbon atoms in the DLC film, and shifts to higher frequencies as the bond lengths shorten.

Therefore, the DLC film where the G-peak is located at high frequency indicates a higher degree of graphite crystallization. In contrast, the FWHM (G) indicates the degree of amorphous character of the graphite. When various bond lengths exist, as in DLC films with marked amorphous character, the G peak appears over a broad range. Therefore, the G peak becomes broadens for the DLC film overall, and, as a result, the FWHM (G) increases. The G peak position shifts to higher frequencies as the amorphous character of the film increases. Generally,

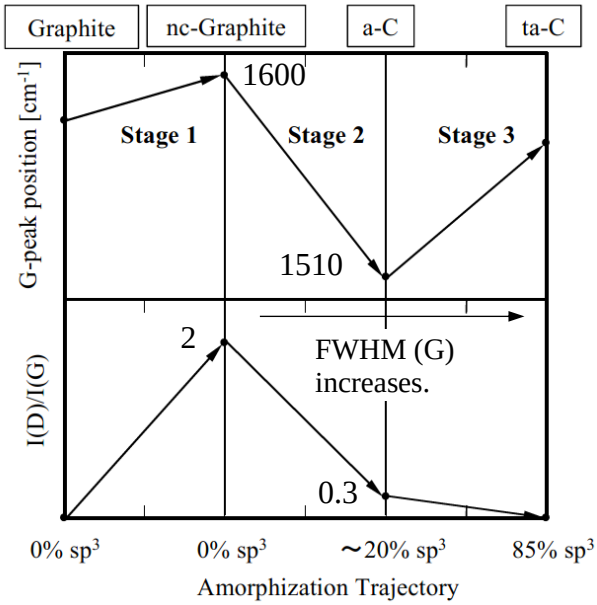



Fig. 3 Amorphization trajectory, showing a schematic variation of the G peak position, *I*(D)/*I*(G) ratio and *FWHM* (G) [6]

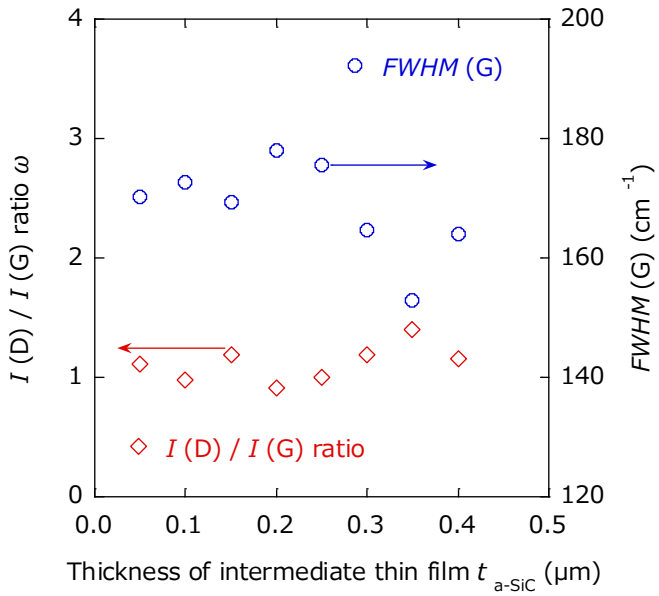



Fig. 4 *I* (D) / *I* (G) ratio and *FWHM* (G) for each interlayer thickness

amorphous films contain a high degree of sp3 bonds, which is positively correlated with FWHM (G) and hardness [7]. This model of structural changes from stages 1 to 3 shown in Fig. 3 was studied by correlation with the Raman spectra of the DLC films.
Figs. 4 and 5 respectively show I (D) / I (G) ratio, FWHM (G) and Raman shift at G peak position values for each interlayer thickness. When the intermediate layer is thin (t <= 0.25), I (D) / I (G), FWHM (G) and Raman shifts ω at G peak positions are almost constant, with some variations. As the intermediate layer becomes thicker, it tends to increase in I (D) / I (G) and

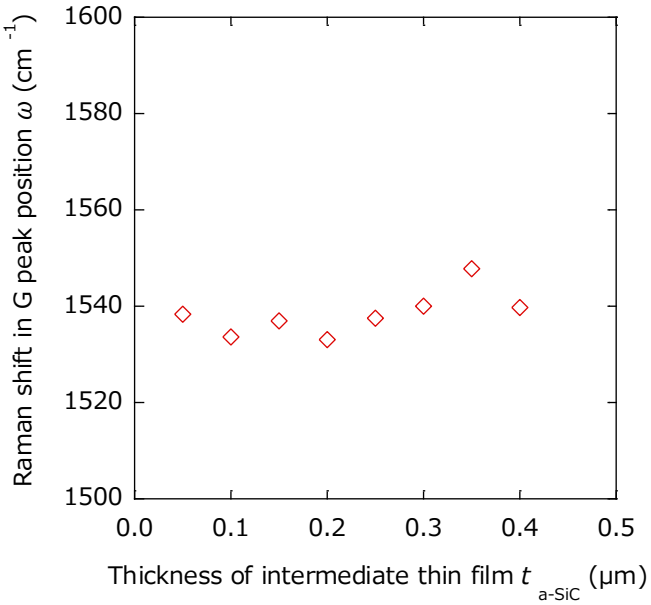



Fig. 5 Raman shift in G peak position for each interlayer thickness

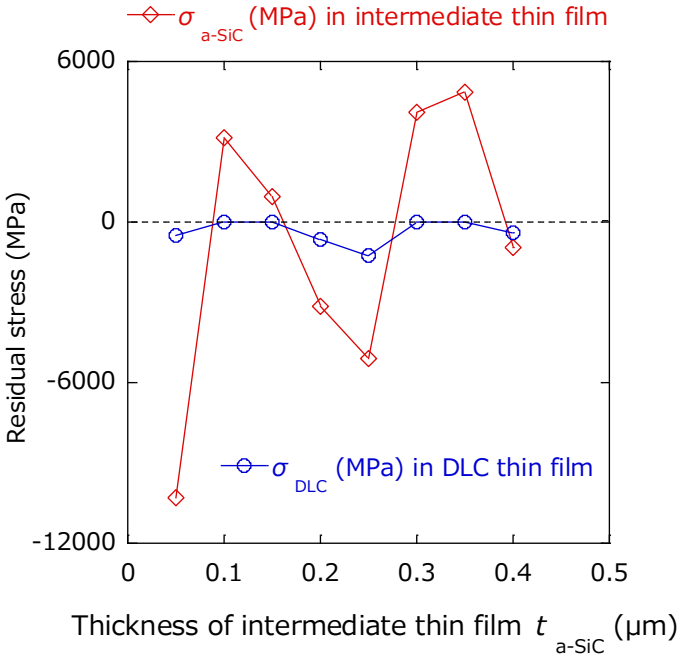



Fig. 6 Changes of residual stress in a-SiC intermediate film and DLC film.

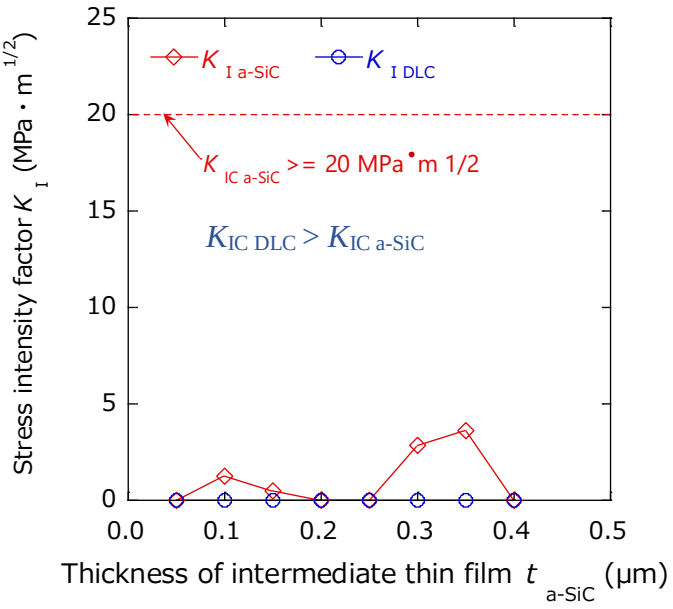



Fig. 7 Changes of residual stress in a-SiC intermediate film and DLC film

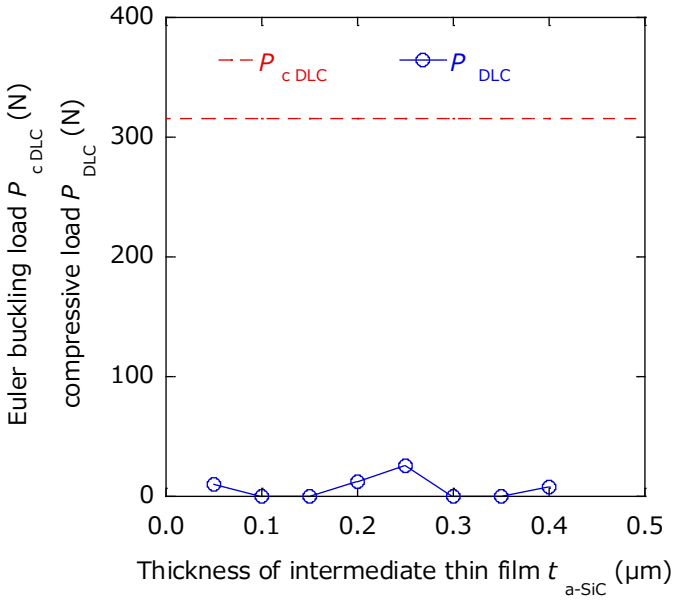



Fig. 8 Euler buckling loads and compressive loads in DLC film

Raman shifts ω, but decreases in FWHM (G). Compared to FIG. 3, the proportion of sp3 bonds in the DLC film is approximately 10%.

*Changes in residual stress in a-SiC fime and DLC film*

Fig.6 shows the residual stress σ a-SiC in the a-SiC thin film and the residual stress σ DLC in the DLC thin film by the intermediate layer thickness. In any thickness of the a-SiC interlayer, the intermediate thin film had a considerably higher residual stress than the DLC thin film.
The stress intensity factor in the residual stress field is shown in Fig.7. Since no tensile stress is applied to the DLC film, the stress intensity factor is zero. Therefore, even if there is a latent

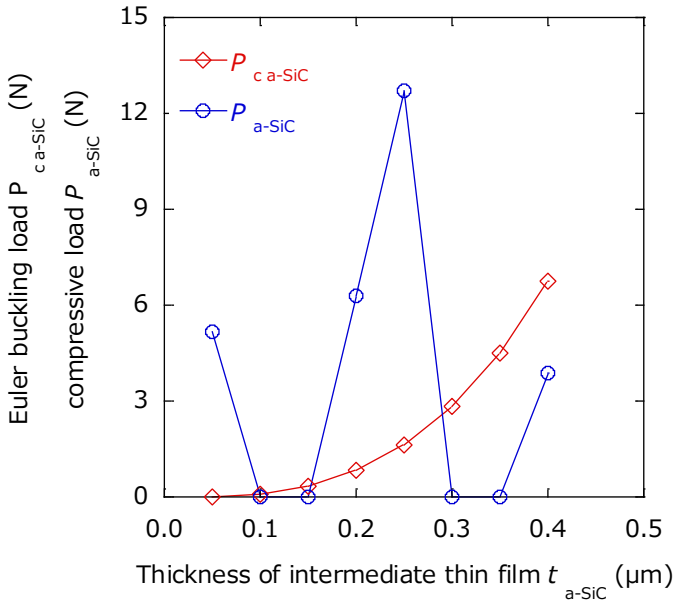



Fig. 9 Euler buckling loads and compressive loads in a-SiC film

surface penetration crack of the DLC film, the crack does not become the origin of the peel fracture of the film. On the other hand, although the a-SiC intermediate thin film has a larger tensile stress than the DLC film, the maximum stress intensity factor does not exceed the fracture toughness value, so even if there is a latent penetration crack in the film, the crack does not progress.

Figures 8 and 9 show the Euler buckling load and compressive residual stress of both films applied to the cross section of the film. In the case of a DLC film, no buckling occurs because the compressive load due to residual stress is much smaller than the buckling load.

On the other hand, at t a-SiC = 0.05, 0.2, 0.25, 0.4 μm, compressive residual stress is applied to the a-SiC intermediate thin film, and local buckling of the film does not occur only at t = 0.4 μm, and in other cases, local buckling of the membrane occurs.

In the experimental results of film depositions, peeling failure of the film was confirmed only at t = 0.05 μm. From the results, in the case of t = 0.05, the buckling of the whole film was caused by the local buckling of the interlayer. On the other hand, in the case of t = 0.2 and 0.25, local buckling occurred in the film but it did not reach the entire film.

From the above, the strength and rigidity of the a-SiC interlayer greatly affect the adhesion of the multilayer film than the DLC film. The reduction of the residual stress of the a-SiC interlayer can improve the film adhesion.

In the future, using quantum chemical calculation program Gaussian and finite element method program MSC. Marc, we will investigate the mechanism of interfacial peeling failure of the film and find the optimum deposition conditions for improving the adhesion of the film.

## Conclusion

DLC (a-C: H) multilayer film with an a-SiC interlayer were deposited, and residual stresses the influence of the residual stresses on the adhesion of the film were investigated. The following conclusions can be drawn from the results presented herein:

(1) The thickness of the a-SiC interlayer affects the DLC film structure. When the thickness of the interlayer $t_{a\text{-}SiC}$ <= 0.25 m μm, the proportion of sp 3 bonds in the DLC film is approximately

constant and about 10%. In the case of $t_{a\text{-}SiC} > 0.25$ μm, the proportion of sp3 bonds in the DLC film tends to decrease slightly.
(2) In the film deposition process, residual compressive stress exists in the DLC film, but residual tensile stress or compressive residual shrinkage stress exists in the a-SiC interlayer.
(3) The residual stress of the DLC film is not due to the peeling failure of the film.
(4) The residual tensile stress of the a-SiC interlayer does not cause the peeling failure of the film, but a part of the compressive residual stress causes the local buckling of the interlayer film and may also cause the peeling failure of the film.
(5) The strength and rigidity of the a-SiC interlayer dominate the adhesion between the a-SiC-DLC multilayer film and the substrate.

**References**

[1] J. Robertson, (1992) Properties of Diamond-Like Carbon, Surf. Coat. Technol. **50**, 185-195.
[2] K. Akari, (2002) Improvement in Adhesion of DLC Coating, J. Japanese Soc. Tribol. **47**, 809-814.
[3] Z. Sun, C. H. Lin, Y. L. Lee, J. R. Shi, B. K. Tay and X. Shi, (2000) Effectsn on deposition and mechanical properties of diamond-like carbon film using inert gases in methane plasma, Thin Solid Films 377-378, 198-202.
[4] Goujaro, S., Vandenbulcke, L., Tawil, H., (1994) the Oxidation Behaviour of Two-and Three-Dimensional C/SiC Thermostructural Materials Protected by Chemical-Vapour-Deposition Polylayers Coatings J. Mater. Sci. **29**, 6212-6220.
[5] R. E. Shroder, R. J. Nemanich & J. T. Glass, (1990) Analysis of the Composite Structures in Diamond Thin Films by Raman Spectroscopy, Phys. Rev. B **41**, 3738-3748
[6] A. C. Ferrari & J. Robertson, (2000) Interpretation of Raman Spectra of Disordered and Amorphous Carbon," Phys. Rev. B **61**, 14095-14017
[7] C. Casiraghi, A. C. Ferrari & J. Robertson, (2005) Raman Spectroscopy of Hydrogenated Amorphous Carbon, Phys. Rev. B **72**, 085401-085411.
[8] G. Adamopoulos, J. Robertson, N. A. Morrison & C. Godet, (2004) Hydrogen Content Estimation of Hydrogenated Amorphous Carbon by Visible Raman Spectroscopy, J. Appl. Phys. **96**, 6348 -6358.
[9] K. Miura, M. Nakamura, (2008) Analysis of Hydrogen Concentration in DLC Films by Raman Spectroscopy", J. Surf. Fin. Soc. Japan **59**, 203-205.
[10] J. Choi, T. Hibi, A. Furuno, M. Kawaguchi, T. Kato, (2017) Evaluation and Prediction of Mechanical Properties of DLC Films by Raman Spectroscopy (Part 2), J. Japanese Soc. Tribol. **62**, 228-235.
[11] J. Roberston, (2002) Diamond-Like Amorphous Carbon, Mater. Sci. Eng. R: Rep. **37**, 129-281.
[12] G. G. Stoney, (1909) the Tension of Metallic Films Deposited by Electrolysis", Proceedings of the Royal Society of London, Series A **82**, 172-175.
[13] M. Nakamura, K. Miura, T. Matsuoka and T. Hirayama, (2008) Effects of Deposition Conditions on Residual Stress in DLC films Prepared by UBM Sputtering, J. JSMS **57**, 488-494.
[14] H. Kobayashi, Fracture mechanics, 10th Press, Kyoritsu Publishing Co., Ltd., Japan, 2005, 61-63.
[15] S. Takezono, Foundation strength of materials, 19th Press, Asakusa Shoten, Japan, 2002, 134-135.

# Energy Saving Mechanism of Propeller with Endplates at Blade Tips

Yijie Liu, Decheng Wan*

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

**ABSTRACT**

Ship energy-saving has always been one of the performance improvements that researchers are trying to improve. It is an effective energy-saving way to maintain a certain chord length at the tip of the propeller blade and install an arc end plate at the tip of the blade. With the development of CFD(computational fluid dynamics), many researchers combine the design and research of energy-saving devices with the CFD methods. This paper calculated the hydrodynamic of propeller with endplates based on OpenFOAM, the flow field and models are processed by sliding grid technology. By comparing the open water characteristics and thrust performance of standard MP687 propeller and improved propeller, it is found that the propeller with endplates can maintain a larger circulation strength and prevent the transverse disturbance at the end of the blade; at the same time, it can reduce the cavitation zone at the blade of the propeller and play the role of noise reduction, erosion reduction and vibration reduction; besides, it can also eliminate tip vortices, achieve the effect of energy-saving.

**KEY WORDS**: endplates at blade tips; energy-saving; sliding mesh; OpenFOAM

## INTRODUCTION

For the high-energy-consuming shipping industry, researchers pay more and more attention to the "Green Ship" design, which can reduce energy consumption and cost. Among all the methods of energy saving and fuel consumption reduction, the improvement of propeller and propeller is the most effective. At present, most of the energy-saving propulsion technologies focus on improving the propeller inlet, improving the pressure on the blade, reducing the friction force on the blade surface[1, 2], so as to make more use of the energy that can not be used and improve the transmission shafting. Although many propulsion devices have been developed, there are still many problems affecting the efficiency of propulsion, but this shows that energy-saving propulsion devices still have great potential for development.

Early researchers mainly carried out research on energy-saving devices through ship model test or theoretical research. In recent years, the numerical simulation method for simulating fluid flow by means of computer has gradually emerged. It can greatly reduce the design time and cost, so it has attracted the attention of many scholars. Current numerical simulation methods mostly use viscous flow method. Viscous flow numerical simulation methods can be divided into three categories: DNS (direct turbulence numerical simulation), RANS (Reynolds average N-S equation) and LES (large eddy numerical simulation). DNS method is a direct discrete solution to the N-S model, which can accurately obtain the flow field information; RANS method can be solved by a smaller calculation process; LES method can save the calculation process by filtering small vortices and only calculating large vortices, and the results obtained are more accurate than RANS method. At present, many scholars [3] are continually studying and improving the viscous flow simulation methods.Chang [4] et al. simulated the cavitation flow around the propeller by RANS method, and analyzed the image results of cavitation on the suction surface of the propeller under the numerical simulation, which verified the feasibility of RANS simulation. Peters [5] et al. successfully simulated the cavitation flow around a marine propeller using a RANS-based solver and Volume Fluid (VOF) method.

Combining with the numerical simulation method, many energy-saving propellers have been designed and studied by researchers. Berger et al. [6] studied the optimal design process of hub cap fins by using CFD methods. Takafumi Kawamura et al. [7] used CFD methods to study the difference of energy-saving effect of hubcap fins on ship models and real-scale ships. Anirban et al. [8] discussed the influence of scale effect on hydrodynamic characteristics of ducted propeller based on CFD methods. Fahri [9] used commercial software to study the front catheter installed on chemical tankers. Sunho [10]et al. used commercial software STAR-CCM+ to study the wake field of the pre-rotating stator in the self-propelled sailing process of KVLCC2 in real ship scale. Based on OpenFOAM open source platform, the numerical simulation of MP687 propeller with end plates and standard MP687 propeller is carried out in open water. By comparing the force and vorticity changes of the propeller before and after adding endplates at blade tips, the energy-saving mechanism and energy-saving effect are analyzed.

## NUMERICAL METHOD

### Governing Equations

The governing equation of numerical simulation is incompressible fluid continuum equation. As for RANS turbulence model, the equations are presented below.

$$\frac{\partial \rho}{\partial t}+\frac{\partial\left(\rho u_i\right)}{\partial x_i}=0 \tag{1}$$

$$\frac{\partial\left(\rho u_i\right)}{\partial t}+\frac{\partial\left(\rho u_i u_t\right)}{\partial x_j}=-\frac{\partial p}{\partial x_i}+\frac{\partial}{\partial x_j}(\mu(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{x_i}-\frac{2}{3}\delta_{ij}\frac{\partial u_i}{\partial x_i})-\rho\overline{u_i^{'}u_j^{'}}) \tag{2}$$

In the above formulas, $u_i$ and $u_j$ are the $i$ and $j$ components of velocity, $P$ is the hydrostatic pressure, $\mu$ is the hydrodynamic viscous coefficient.

$-\rho\overline{u_i'u_j'}$ is the Reynolds stress term. Reynolds stress term ensures the closure of RANS equation by turbulence model.

### SST k-omega turbulence model

In SST k-omega model, k-omega model is used near the wall and k-epsilon model is used in the far field. At present, it is one of the models with high usability, which is mainly used to simulate Reynolds stress. The main equations for this model are as follows.

$$\frac{\partial \rho k}{\partial t}+\frac{\partial \rho u_i k}{\partial x_i}=\tilde{P_k}-\beta^* \rho k\omega+\frac{\partial}{x_i}\left[\left(\mu+\sigma_k\mu_t\right)\frac{\partial k}{\partial x_i}\right] \tag{3}$$

$$\begin{aligned}\frac{\partial \rho\omega}{\partial t}+\frac{\partial \rho u_i \omega}{\partial x_i}=\alpha\rho S^2 P_k-\beta\rho\omega^2+\frac{\partial}{x_i}\left[\left(\mu+\sigma_\omega\mu_t\right)\frac{\partial\omega}{x_i}\right]+\\ 2\left(1-F_1\right)\rho\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial k}{\partial x_i}\frac{\partial\omega}{\partial x_i}\end{aligned} \tag{4}$$

$$F_1=\tanh\left\{\left\{\min\left[\max\left(\frac{\sqrt{k}}{\beta^*\omega y},\frac{500\nu}{y^2\omega}\right),\frac{4\sigma_{\omega 2}k}{CD_{k\omega}y^2}\right]\right\}^4\right\} \tag{5}$$

$$CD_{k\omega} = \max\left(2\rho\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial k}{\partial x_i}\frac{\partial \omega}{\partial x_i}, 10^{-10}\right) \tag{6}$$

The decoupling of velocity and pressure adopts PIMPLE algorithm, which combines the advantages of PISO algorithm and SIMPLE algorithm. SIMPLE algorithm is used in the time step, and PISO algorithm is used in the time step, so that the time step can be increased properly on the premise of ensuring stability.

**Hydrodynamic Characteristics of Propellers**

The evaluation of propeller and improved propeller can be considered from thrust coefficient, torque coefficient and propulsion efficiency. The propeller's speed coefficient is $J$, thrust coefficient is $K_T$, torque coefficient is $K_Q$, and propulsion efficiency is $\eta_0$.

$$K_T = \frac{T_P}{\rho n^2 D^4} \tag{7}$$

$$K_Q = \frac{Q_P}{\rho n^2 D^5} \tag{8}$$

$$\eta_0 = \frac{K_T}{K_Q}\frac{J}{2\pi} \tag{9}$$

$\rho$ is density, $n$ is speed of propeller, $D$ is diameter of propeller.
Density is 998.2kg/m$^3$, and the kinetic viscosity coefficient of water is $v$=1.106×10$^{-6}$ m$^2$/s.

## CASE DESCRIPTION

**Computing Model**

The propeller used in this paper is MP687 propeller. The improved propeller is MP687 propeller with endplates at blade tips. The specific parameters of the propeller and the improved propeller are shown in Tables 1 and 2.

Table 1. MP687 propeller model parameters

| MP687 | | |
|---|---|---|
| Diameter (m) | D | 0.203 |
| Hub ratio | $r_h$/R | 0.180 |
| Pitch ratio | $P_{0.7}$/D | 0.750 |
| Disk ratio | Ae/Ao | 0.500 |
| Trim angle（deg） | θ | 5 |
| Number of propeller blades | Z | 5 |

Table 2. MP687 propeller model parameters

| endplates | | |
|---|---|---|
| Extend (m) | a | 0.09 |
| Intermediate (deg) | $\alpha$ | 87 |
| Height (m) | b | 0.03 |
| Number of propeller blades | Z | 5 |

The geometric model of MP687 propeller and MP687 propeller with endplates at blade tips is built by using

commercial software CATIA according to the spatial coordinates of propeller and hub cap fin, as shown in Figure 1 and Figure 2.

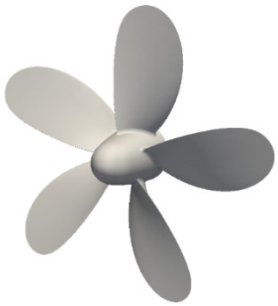


Fig. 1 MP687 propeller model.

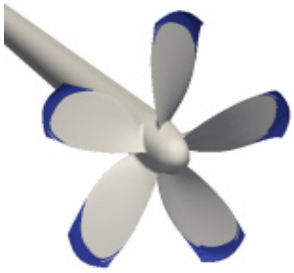


Fig. 2 MP687 propeller model with endplates at blade tips.

**Mesh Generation**

Arbitrary Mesh Interface (AMI) is a sliding mesh method, which can be used to solve unsteady flow problems. The computational domain is divided into two sub-regions. The propeller is completely enveloped in a cylindrical rotating sub-region, as shown in Fig. 3. Because the geometry of propeller blade is complex and has irregular surface, unstructured mesh is used in calculation. The mesh generation is shown in Fig. 4 and Fig. 5, and the total mesh is about 4 million.

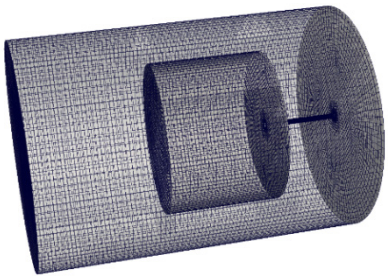


Fig. 3 Computational domain.

Fig. 4 Computational grid of MP687 propeller.

Fig. 5 Computational grid of MP687 propeller with endplates.

**Accuracy Verification of AMI Method**

In order to verify the accuracy of the AMI method in dealing with rotating grids, the open water performance of propeller is simulated and compared with the experimental data provided by NMRI. The calculation of open water propeller MP687 is performed by pimpleDyMFoam solver in OpenFOAM. During the calculation, the speed of the propeller is fixed and RPS=20 is maintained.

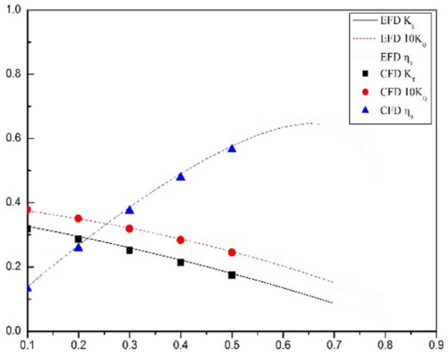



Fig. 6 Open Water Characteristic Calculating Curve of Propeller.

The numerical simulation is carried out under the condition of J=0.2, J=0.3, J=0.4, J=0.5. In Fig. 6, the error between the open water characteristics of propeller calculated by AMI method and the test results is within a small range, which verifies the accuracy and reliability of the AMI method in the calculation of rotating mesh.

## RESULT AND DISSCUSIONS

**Analysis and comparison of hydrodynamic characteristics**

The speed of the propeller remains unchanged at 20r/s, and J is adjusted by changing the flow velocity. The hydrodynamic performance of propeller with and without endplates under J=0.2 is compared in Table 3. The thrust of the propeller itself increases significantly, the torque decreases significantly, and the propulsion efficiency of the whole system improves significantly.

Table 3. comparison of hydrodynamic characteristics (J=0.2)

| | J | Thrust(N) | Torque(N*m) | KT | 10KQ | η0 |
|---|---|---|---|---|---|---|
| Without endplates | 0.2 | 195.380 | 4.848 | 0.2881 | 0.3522 | 0.2604 |
| With endplates | 0.2 | 198.831 | 4.763 | 0.2932 | 0.3460 | 0.2697 |

**Force Analysis of Propeller**

The surface pressure of two kinds of propellers are analyzed and the pressure distribution maps are drawn, as shown in Fig. 7 and Fig. 8. By comparing Fig. 7 with Fig. 8, it is not difficult to find that in 7, the high pressure area at the guide edge of the blade is obviously larger than that of the standard propeller due to the addition of endplates, which means that the pressure difference between the front and rear of the propeller will increase, thus leading to the increase of the propeller thrust. The reason for this phenomenon is related to the addition of endplates, which can increase the high pressure area and improve efficiency.

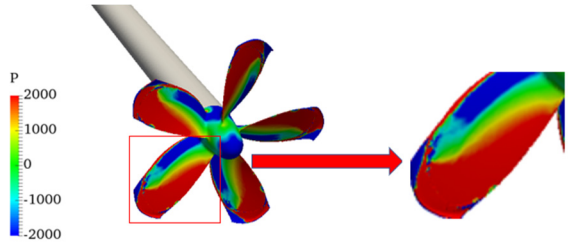



Fig. 7 Surface Pressure Diagram of MP687 Propeller with Endplates.

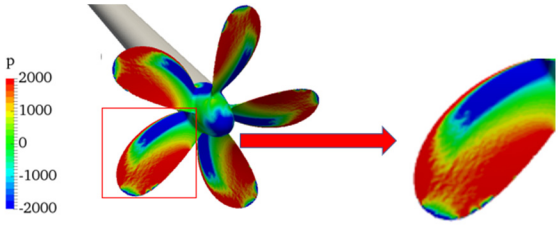



Fig. 8 Surface Pressure Diagram of MP687 Propeller.

**Contrast of Vortex Ejection**

The vortex spray charts of the two models are shown as fellows. It is easy to find that the improved propeller tail vorticity is more uniform, and because the end plate can eliminate the tip vorticity, thus eliminating unnecessary energy consumption, greatly improving the energy saving and propulsion capacity of the propeller.

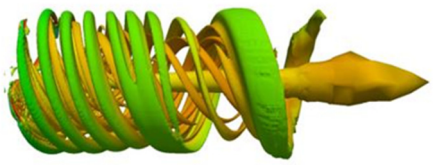


Fig. 9 Vortex Spray Chart of MP687 Propeller with Endplates.

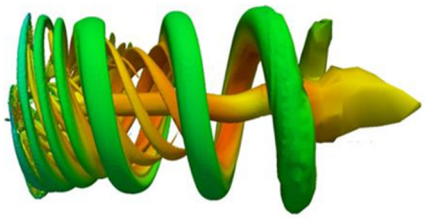


Fig. 10 Vortex Spray Chart of MP687 Propeller.

**CONCLUSION AND PROSPECT**

(1) Using solver under OpenFOAM platform, based on RANS equation, adding A turbulence model and

combining with the AMI meshing method is an effective numerical simulation approach for propeller, which can achieve good accuracy. An effective way to predict propeller performance and design energy-saving device is discussed for researcher to carry out numerical simulation of propeller in this paper.

(2) The improvement of propeller propulsion efficiency by endplates is mainly based on two aspects: on the one hand, endplates can increase the high pressure area and increase the pressure difference between front and rear blades. On the other hand, the endplates can reduce the tip vorticity and eliminate the tip vorticity, so as to improve the propulsion efficiency of the propeller.

(3) In the follow-up work, the influence of endplates on wake field of propeller can be further studied.

## ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (51879159, 51490675, 11432009, 51579145), Chang Jiang Scholars Program (T2014099), Shanghai Excellent Academic Leaders Program (17XD1402300), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23/09) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

## REFERENCES

[1] TAKEO N, NORIO I, HISASHI K. Energy saving technology of pbcf (propeller boss cap fins) and its evolution[C]. Okayama, Japan, 2010.

[2] OUCHI K, OGURA M, KONO Y, et al. A research and development of PBCF(propeller boss cap fins)-improvement of flow from propeller boss[J]. Journal of Society of Naval Architects of Japan. 1988, 163: 66-78.

[3] Kim S E. A numerical study of unsteady cavitation on a hydrofoil. Proceedings of the 7th international symposium on cavitation, [J]., 2009.

[4] Chang Y C , Hu C N , Tu J C , et al. Experimental investigation and numerical prediction of cavitation incurred on propeller surfaces [J]. Journal of Hydrodynamics, Ser. B, 2010, 22(5-supp-S1):764-769.

[5] Peters A , Lantermann U , Moctar O E . Numerical Prediction of Cavitation Erosion on a Ship Propeller in Model-andFull-Scale[J]. Wear, 2018.

[6] Berger S , Druckenbrod M , Pergande M , et al. A Two-Stage Optimisation Method for Full-Scale Marine Propellers Working Behind a Ship[J]. Ship Technology Research, 2014, 61(2):64-79.

[7] Kawamura T, Ouchi K, Nojiri T. Model and full scale CFD analysis of propeller boss cap fins (PBCF)[J]. Journal of Marine Science and Technology, 2012, 17(4):469-480.

[8] Bhattacharyya A,Krasilnikov V,Steen S.A CFD-based scaling approach for ducted propellers[J]. Ocean Engineering,2016,(7):116-130.

[9] FAHRI C. A numerical study for effectiveness of a preduct[J]. Ocean Engineering, 2007, 34: 2138-2145.

[10] SUNHO P, GWANGHO O, SHIN H R, et al. Full scale wake prediction of an energy saving device by using computational fluid dynamics[J]. Ocean Engineering, 2015, 101: 254-263.

# An Efficient Adaptive Strategy for Acoustic Problems with the Edge-based Smoothed Point Interpolation Method (ES-PIM)

***†Q. Tang[1,2], K.X.Wei[1,2], X.G.Sun[1,2]**

[1]Department of Mechanical Engineering, Hunan Institute of Engineering, Xiangtan, 411101，China

[2]Hunan Provincial Key Laboratory of Vehicle Power and Transmission System, Xiangtan, 411101, China

*Presenting author: tq1618@163.com

†Corresponding author: tq1618@163.com

## Abstract

In this paper we carry out an H-adaptive strategy for acoustic problems with the Edge-based smoothed point interpolation method. The key features of the adaptive procedure are an error indicator, H-type refinement strategy, local critical values, Delaunay mesh generation and ES-PIM analysis. The computations are performed on meshes with three-node triangles are adapted to the solution by locally changing element sizes, taking advantage of the background mesh which is convenient to discrete and conducts numerical simulation for any complicated model. The error indicator has been performed considering the maximum values of velocity difference among the vertexes in each cell. The adaptive meshes are then obtained through global mesh regeneration using a Delaunay mesh generator. The adaptive analysis is applied to 2D acoustic frequency response analysis, especially for expansion chamber. Numerical examples are shown to illustrate the properties of the error indicator technique and the procedure of the proposed adaptive strategy. The results highlight the efficiency of adaptive analysis, which reduces computation consumption significantly, and the results also have shown the validity and efficiency of the proposed error indicator.

**Keywords:** Adaptive analysis; Error indicator; H-refinement; Point interpolation method; Acoustic

## Introduction

Numerical computation has been widely applied for scientific research and solving practical engineering problems in many fields. A large number of research results show that the adaptive analysis is an effective way to improve the efficiency and precision of the numerical calculation[1]-[7]. Through a lot of theoretical and numerical analysis, the ES-PIM can reduce the softening effects and give a quite close-to-exact stiffness by using the edge-based strain smoothing operation [8]-[12],it is found that ES-PIM is more suitable for solving acoustic problems, and a series of innovative research results have been obtained [13]-[14]. He and Liu applied the generalized gradient smoothing technique to the field of acoustic numerical computation [15]-[16].

Adaptive analysis is a reliable way to improve the accuracy and efficiency of acoustic problems, and error estimation with high reliability is also a very important factor in the study of acoustic adaptive analysis.In summary, ES-PIM have many aforementioned features which make it become an ideal candidate for adaptive analyses.

In the following section, we describe the implementation of the adaptive analysis based on ES-PIM in acoustic problems. Our adaptive strategy contains two main issues:error indicator and refinement strategy.The proposed adaptive procedure construct an error indicator combining refinement strategy and remeshing technique with the available open source

packages.

**Adaptive scheme**

1. Error indicator

Error indicator plays a crucial role in adaptive procedure.It is able to accurately detect the regions for mesh refinement. According to the characteristics of acoustic problems, a new error indicator has been designed considering the maximum values of velocity difference among the vertexes in each cell.Generally, the high error region is consistent with the area of steep gradient velocity.

For each three-node triangular element, we can obtain the difference value of velocity components between two different nodes, namely node 1 and node 2 by the following equation.

$$\begin{aligned} \Delta v_{xx}^{(12)} &= \left| v_{xx}^{(1)} - v_{xx}^{(2)} \right| \\ \Delta v_{yy}^{(12)} &= \left| v_{yy}^{(1)} - v_{yy}^{(2)} \right| \end{aligned} \tag{1}$$

Eq(2) means the absolute value of the difference of velocity components between any two nodes in the same element is modulo.

$$E = \sqrt{\Delta v_{xx}^2 + \Delta v_{yy}^2} \tag{2}$$

Substituting (1) into (2), we can obtain the following Eq(3).

$$E_{12} = \sqrt{\left(\Delta v_{xx}^{(12)}\right)^2 + \left(\Delta v_{yy}^{(12)}\right)^2} = \sqrt{\left| v_{xx}^{(1)} - v_{xx}^{(2)} \right|^2 + \left| v_{yy}^{(1)} - v_{yy}^{(2)} \right|^2} \tag{3}$$

In the same way, we can calculate $E_{13}$ and $E_{23}$ for every two nodes in a cell. Finally for every background cell, we use the maximum modulo values of all nodes as the error indicator of the three-node triangular cell. Thus the error indicator $E_i$ for the ith cell can be obtained as following equation.

$$\left\| E_i \right\| = Max\left( E_{12}, E_{23}, E_{13} \right) \tag{4}$$

2. Refinement strategy

In this section, we will describe a very simple and rapid method to implement refinement. As shown in Fig 1, white node denotes initial node, black node denotes new node which is inserted into the high error area, three additional black nodes will be inserted at the midpoints of the three edges and the original cell will be further divided into four triangles.

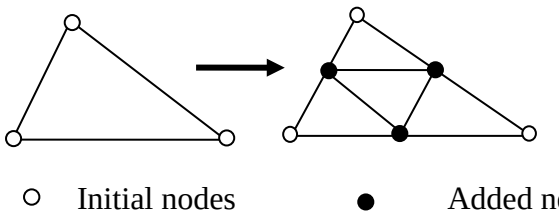



**Figure 1. Illustration of the h-refinement strategy for three-node triangle**

Local critical values , refinement rate and Delaunay mesh generator used in this part are similar to the counter part of adaptive ES-PIM on solid mechanics,details can be found in the previous work[2]-[5].

**Numerical examples**

**（1）Case1: Two dimensional cavity**

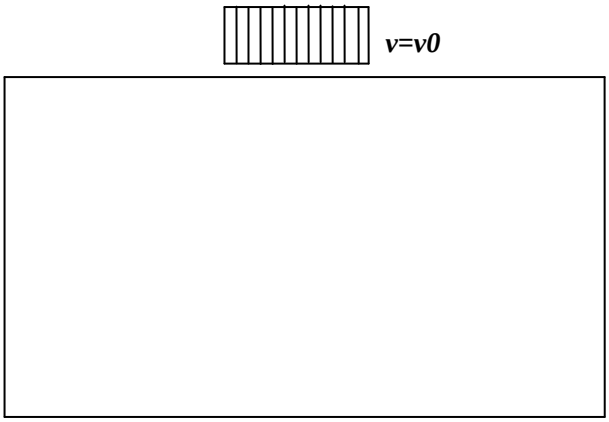



**Figure 2. Geometry,boundary conditions of the two dimensional cavity**

The first example is a room of length L=5 meters and width W=3 meters, which is considered as a two dimensional cavity.As shown in Fig 2, the acoustic excitation is at the top end of the room by a vibrating panel with velocity $\upsilon_0 = 1m/s$.On the other boundaries, the normal velocity is set to be zero. Fluid density in cavity is $\rho_0 = 1.225kg/m^3$, the sound travels speed in this medium is $c_0 = 340m/s$.

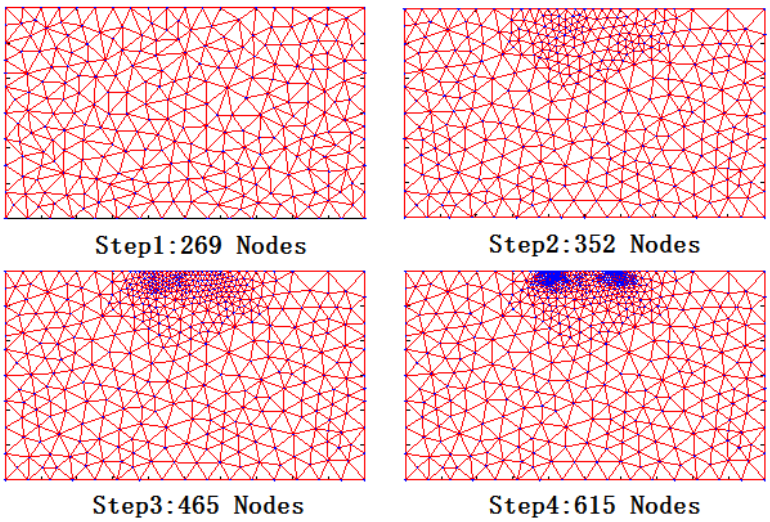



**Figure 3. Nodes distribution and corresponding meshes of each adaptive step for two dimensional cavity at 40Hz**

We studied this problem using 40Hz frequency, the ES-PIM adaptive analysis started from quite coarse mesh of 269 nodes and is performed for 4 steps with the refinement rate

$\eta = 0.1$.Four uniform refinement models with 269, 820, 2001 and 2910 nodes are studied respectively. The nodal distributions and corresponding meshes for 4 adaptive steps are shown in Fig. 3. The figure illustrates that our error indicator can accurately catch the steep gradient of velocity, the dense nodes are inserted near the top end of the room where the velocity gradient is high.

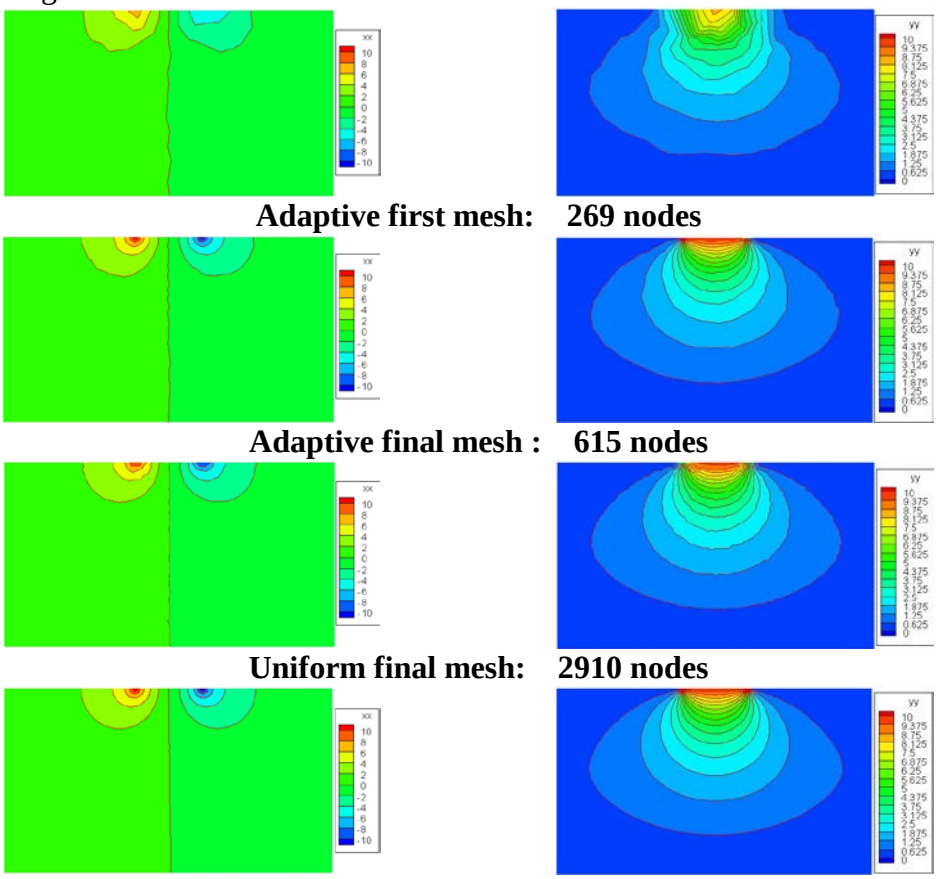



**Figure 4. Comparison of velocity distributions at the first and final stage for the problem of two dimensional cavity at 40Hz**

Fig. 4 compares the contours of velocity components at the first and final adaptive mesh with the final uniform mesh and FEM reference solution mesh results of 13123 uniformly distributed nodes. It clearly indicates that the velocity contours at the final stage with adaptive mesh(only 615 nodes) and uniform mesh(2910 nodes) are in good agreement with the reference solution results obtained using a very fine mesh.

**（2）Case2: Expansion chamber**

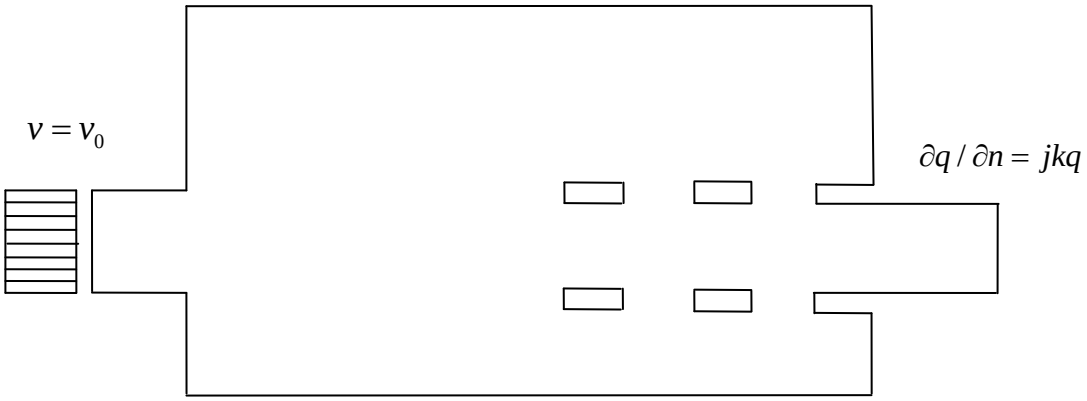



**Figure 5. Geometry,boundary conditions of expansion chamber**

The second problem is a 2D expansion chamber,whose geometry and boundary conditions are shown in Fig.5. Acoustic excitation is presented at the left side by a vibrating panel with velocity $\upsilon_0 = 0.1m/s$. In order to better simulate the real situation,the boundary condition of the equation $\partial q / \partial n = jkq$ is presented at the right side, meanwhile, at the other side the normal boundary velocity is set to be zero. Fluid density in this model is $\rho_0 = 1.225kg/m^3$,the sound travels speed in this medium is $c_0 = 340m/s$.

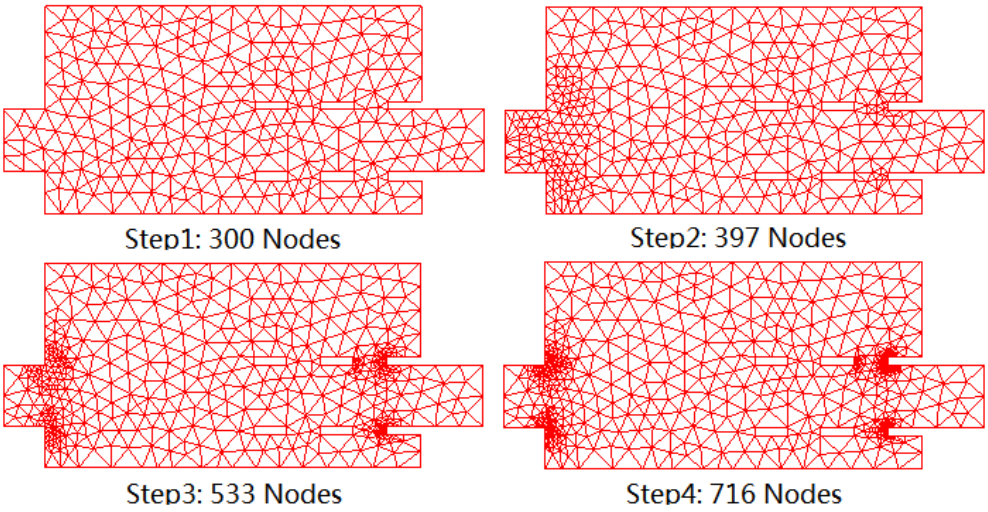
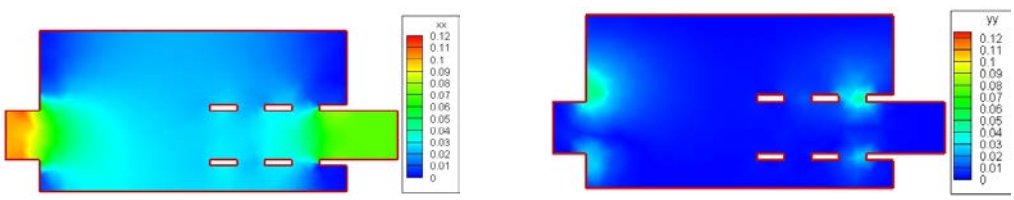



**Figure 6. Nodes distributions at each adaptive step for the problem of expansion chamber**

In this example, we studied expansion chamber using frequency of excitation is 200Hz. The adaptive procedure starts from an initial mesh of 300 nodes, and is performed for 4 steps with $\eta = 0.1$. The adaptive meshes for each step are shown in Fig. 6. One can notice that the proposed error indicator effectively detects all the regions for high velocity gradient and implements the refinement of nodes. For comparison, four models of uniformly refined models with 300, 1177, 2658 and 4076 nodes are also adopted to study this problem.



**Adaptive first mesh: 300 nodes**

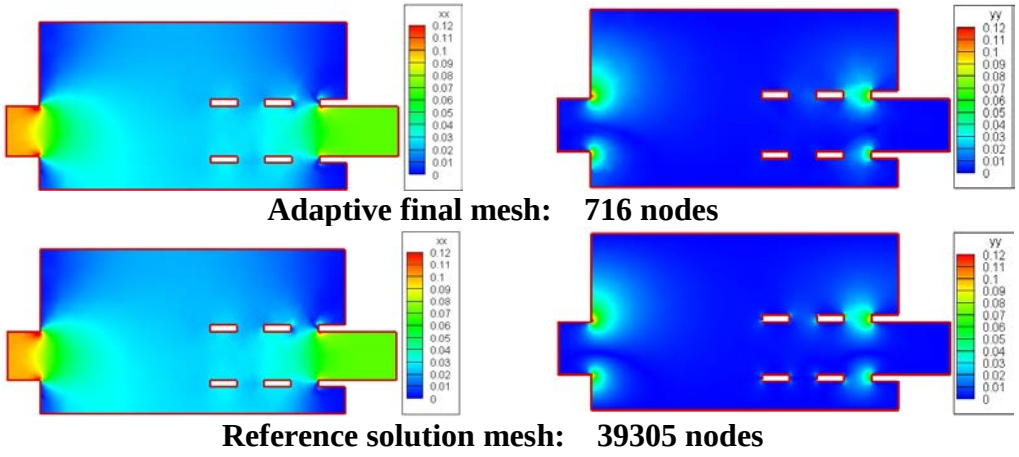



**Figure 7. Comparison of velocity distributions at the first and final stage for the problem of expansion chamber at 200Hz**

In Fig. 7 we compare the contours of velocity components at the first and final adaptive mesh with the FEM reference solution mesh results obtained using a very fine mesh. It can be seen that the velocity contours at the final stage with adaptive mesh (only contains 716 nodes) is in good agreement with the reference solution mesh of 39305 uniformly distributed nodes.

## Conclusions

In this paper, an adaptive procedure using ES-PIM is proposed for acoustic problem. Our adaptive strategy uses initial meshes and remeshing which are implemented by the open source package TRIANGLE, FEM reference solution results using the software SYSNOISE. Numerical problems demonstrated that the proposed error indicator is able to detect the location steep gradient of velocity, h-refinement is performed by adding in new nodes. According to the practise in the research about 2D problems, which provides experience for extending to 3D problems in future.

## Acknowledgment

This work is funded by Natural Science Foundation of Hunan Province of China (2017JJ2059), Scientific Research Fund of Hunan Provincial Education Department (18B384,14B042).

## References

[1] Nguyen, X. H., Liu, G. R., Bordas, S., Natarjan, S. and Rabczuk, T. (2013) An adaptive singular ES-FEM for mechanics problems with singular fifield of arbitrary order, *Computer Methods in Applied Mechanics and Engineering* **253**, 252–273.

[2] Tang, Q., Zhong, Z. H., Zhang, G. Y. and Xu, X. (2011a) An effiffifficient adaptive analysis procedure for node-based smoothed point interpolation method (NS-PIM). *Applied Mathematics and Computation* **217**, 8387–8402.

[3] Tang, Q., Zhang, G. Y., Liu, G. R., Zhong, Z. H. and He, Z. C. (2011b) A three-dimensional adaptive analysis using the meshfree node-based smoothed point interpolation method (NS-PIM). *Engineering Analysis with Boundary Elements* **35**, 1123–1135.

[4] Tang, Q.,Wei, K. X., Zhang, G. Y., Sun, X. G. (2018) A fully automatic h-adaptive analysis procedure using the edge-based smoothed point interpolation method. *International Journal of Computational Methods*, 15(7): 1845001

[5] Tang, Q., Zhang, G. Y., Liu, G. R., Zhong, Z. H. and He, Z. C. (2012) An efficient adaptive analysis procedure using the edge-based smoothed point interpolation method (ES-PIM) for 2D and 3D problems. *Engineering Analysis with Boundary Elements* **36**, 1424–1443.

[6] Nguyen, X. H., Liu, G. R., Bordas, S., Natarjan, S. and Rabczuk, T. (2013) An adaptive singular ES-FEM for

mechanics problems with singular field of arbitrary order. *Computer Methods in Applied Mechanics and Engineering* **253**, 252–273

[7] Li, Y., Liu, G. R. and Zhang, G. Y. (2011) An adaptive NS/ES-FEM approach for 2D contact problems using triangular elements, *Finite Elements in Analysis and Design* **47**, 256–275.

[8] Liu, G. R. and Zhang, G. Y. (2008a) Edge-based smoothed point interpolation methods, *International Journal of Computational Methods* **5**, 621–645.

[9] Liu, G. R., Nguyen, T. T. and Lam, K. Y.(2009a) An edge-based smoothed fifinite element method (ES-FEM) for static, free and forced vibration analyses in solids, *Journal of Sound and Vibration* **320**, 1100–1130.

[10] Liu, G. R. and Zhang, G. Y. (2009) A normed G space and weakened weak (W2) formulation of a cell-based smoothed point interpolation method, *International Journal of Computational Methods* **6**, 147–179.

[11] Li, E., He, Z. C., Xu, X. and Liu, G. R. (2015) Hybrid smoothed fifinite element method for acoustic problems,e singular ES-FEM for mechanics problems with singular fifield of arbitrary order. *Computer Methods in Applied Mechanics and Engineering* **283**, 664–688.

[12] Li, E., He, Z. C. and Xu, X.(2013) A novel edge-based smoothed tetrahedron finite element method (ES-T-FEM) for thermomechanical problems, *International Journal of Heat and Mass Transfer* **166**, 723–732.

[13] Li, E., He, Z. C. Jiang, Y. (2016) 3D mass-redistributed finite element method in structural–acoustic interaction problems. *Engineering Analysis with Boundary Element* **227,**857-879

[14] Li, E., He, Z. C., Hu, J. Y. and Long, X. Y. (2017) Volumetric locking issue with uncertainty in the design of locally resonant acoustic metamaterials, *Computer Methods in Applied Mechanics and Engineering* **324**, 128–148.

[15] He, Z. C., Liu, G. R., Zhong, Z. H., et al.(2009) An edge-based smoothed finite element method (ES-FEM) for analyzing three-dimensional acoustic problems. *Computer Methods in Applied Mechanics and Engineering*, **199**, 20-33

[16] He, Z. C., Liu, G. R., Zhang, G. Y., et al. (2011) Dispersion error reduction for acoustic problems using the edge-based smoothed finite element method (ES-FEM). *International Journal for Numerical Methods in Engineering*, **86**,1322–1338

# Authors Index