ICCM Proceedings

# Proceedings
# of the International Conference
# on Computational Methods

## (Vol. 8, 2021)

12th ICCM, 4th-8th July 2021

Editors: **G. R. Liu, Nguyen-Xuan Hung**

# ICCM2021

Proceedings of the International Conference on Computational Methods (Vol. 8, 2021)

12th ICCM, 4$^{th}$-8$^{th}$ July 2021, Virtual Conference

Edited by

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

**Nguyen-Xuan Hung**
HUTECH University of Technology, Vietnam

Note: The papers/data included in this volume are directly from the authors. The editors are not responsible of the inaccuracy, error, etc. Please discuss with the authors directly, if you have any questions.

# WELCOME MESSAGE

**Dear Colleagues and Friends,**

We are delighted to once again welcome you to the 12th annual International Conference on Computational Methods (ICCM2021). Due to the prolonged COVID-19 pandemic, ICCM2021 will be hosted through Zoom from July 4-8, 2021 by Ho Chi Minh University of Technology (HUTECH), Vietnam. As the world is adjusting to the new normal, we, as a scientific community, will need to be at the forefront of this and make the most out of every opportunity presented. This year's conference will serve as another platform to reinforce our commitment to the continuous exchange of innovative ideas and production of high-quality research works.

Since its establishment, the ICCM has been an international forum for academic and industrial researchers in areas related to computational methods, numerical modelling & simulation, and machine learning techniques. It will offer presentations on a wide range of topics to facilitate the inter-disciplinary exchange of ideas in science, engineering and related disciplines, and foster various types of academic collaborations. Publications, which have been peer-reviewed and accepted, will be showcased in the form of oral presentations at the conference. All presentations, including abstracts and papers, will be published online at our website.

The ICCM conference series were originated in Singapore in 2004, followed by ICCM2007 in Hiroshima, Japan; ICCM2010 in Zhangjiajie, China; ICCM2012 in Gold Coast, Australia; ICCM2014 in Cambridge, England; ICCM2015 in Auckland, New Zealand; ICCM2016 in Berkeley, CA, USA; ICCM2017 in Guilin, China; ICCM2018 in Rome, Italy; ICCM2019 in Singapore, and the most recent ICCM2020 in Vietnam (online).

We would like to express our most sincere appreciation to all members of the Organizing Committee, the International Scientific Committee, and other members who have worked tirelessly to make this conference possible. To the international reviewers, we thank you for your endless support in reviewing submitted abstracts and papers.

As we come to conclude our remarks, we would like to convey our gratitude for your contributions to the ICCM2021 conference. Although it is virtually held, we are beyond ecstatic for your attendance and look forward to your continued engagement as well as support for future ICCM conferences.

**Professor Nguyen-Xuan Hung**
Conference Chairman
CIRTECH Institute, HUTECH University of Technology
President of Vietnam Association of Computational Mechanics
Vietnam

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

# ORGANIZATION COMMITTEES

**Conference Chairman**
Nguyen-Xuan Hung, Ho Chi Minh City University of Technology (HUTECH), Vietnam

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

**International Co-Chairs**
Magd Abdel-Wahab, Ghent University, Belgium
Stephane P.A. Bordas, Luxembourg University, Luxembourg
Tinh Quoc Bui, Tokyo Institute of Technology, Japan
Ha Bui, Monash University, Australia
Daining Fang, Beijng Institute of Technology, China
Jaehong Lee, Sejong University, South Korea
Hua Li, Nanyang Technological University, Singapore
Tuan Ngo, The University of Melbourne, Australia
Hiroshi Okada, Tokyo University of Science, Japan
Timon Rabczuk, Bauhaus University Weimar, Germany
Dia Zeidan, German Jordanian University, West Asia

**Local Co-Chairmen**
Canh Van Le, International University-VNU-HCMC
Hung Quoc Nguyen, Vietnam-German University, Vietnam
Kien Trung Nguyen, Ho Chi Minh City University of Technology and Education, Vietnam
Van Hai Luong, VNUHCM-University of Technology, Vietnam

**Secretary General**
Nhung Ngoc Hoang, Ho Chi Minh City University of Technology (HUTECH), Vietnam
Vuong Van Nguyen, Ho Chi Minh City University of Technology (HUTECH), Vietnam

**Local Organizing Committee**
Anh Ngoc Lai, Binh Anh Tran, Bang Quang Tao, Cuong Huu Ngo, Chien Hoang Thai, Bao Loi Dang, Hieu Van Nguyen, Long Minh Nguyen, Linh Ngoc Nguyen, Lieu Bich Nguyen, Phuc Hong Pham, Phuc Van Phung, Phuong Tran, Phuoc Trong Nguyen, Nam Van Hoang, Nghi Van Vu, Son Hoai Nguyen, Thanh Dinh Chau, Truong Van Vu, Viet Duc La, Binh Le, Hien Van Do, Tuan Ngoc Nguyen

**International Scientific Advisory Committee**
**(ordered by last name)**

Armas Rafael Montenegro (Spain)
Bui Ha (Australia)
Chen Bin (China)
Chen Jeng-Tzong (Taiwan)
Chen Jiye (UK)
Chen Shaohua (China)
Chen Songying (China)
Chen Weiqiu (China)
Chen Zhen (USA)
Cheng Yuan (Singapore)
Cheng Yumin (China)
Cui Fangsen (Singapore)
Dias-da-Costa Daniel (Australia)
Dong Leiting (China)
Fan Chia-Ming (Taiwan)
Fu Zhuojia (China)
Gan Yixiang (Australia)
Gao Wei (Australia)
Gravenkamp Hauke (Germany)
Gu Yuantong (Australia)
Gupta Murli (USA)
Hou Shujuan (China)
Jabareen Mahmood (Israel)
Jacobs Gustaaf (USA)
Jin Feng (China)
Jiang Chao (China)
Kanayama Hiroshi (Japan)
Kang, Zhan (China)
Kougioumtzoglou Ioannis (USA)
Lee Chin-Long (New Zealand)
Lee Ik-Jin (South Korea)
Lenci Stefano (Italy)

Li Chenfeng (UK)
Li Eric (UK)
Li Qing (Australia)
Li Yue-Ming (China)
Liu Yan (China)
Liu Yinghua (China)
Lombardi Domenico (UK)
Miller Karol (Australia)
Nguyen Anh Dong (Vietnam)
Nguyen Duc Dinh (Vietnam)
Nguyen Giang (Australia)
Nithiarasu Parumal (UK)
Niu Yang-Yao (Taiwan)
Ogino Masao (Japan)
Onishi Yuki (Japan)
Peng Qing (USA)
Picu Catalin (USA)
Quek Jerry Sinsin (Singapore)
Reali Alessandro (Italy)
Rebielak Janusz (Poland)
Reddy Daya (South Africa)
Sadowski Tomasz (Poland)
Saitoh Takahiro (Japan)
Shen Lian (USA)
Shen Luming (Australia)
Shioya Ryuji (Japan)
Son Gihun (South Korea)
Song Chongmin (Australia)
Stefanou George (Greece)
Su Cheng (China)
Tadano Yuichi (Japan)
Tian Zhao-Feng (Australia)
Trung Nguyen-Thoi (Vietnam)
Tsubota Ken-Ichi (Japan)

Wan Decheng (China)
Wang Dongdong (China)
Wang Hu (China)
Wang Jie (China)
Wang Lifeng (China)
Wang Yue-Sheng (China)
Wu Bin (Italy)
Wu Wei (Austria)
Xiang Zhihai (China)
Xiao Feng (Japan)
Xiao Jinyou (China)
Xu Chao (ZJU, China)
Xu Chao (NPU, China)
Xu Xiangguo George (Singapore)
Yang Judy (Taiwan)
Yang Qingsheng (China)
Yang Zhenjun (China)
Yao Jianyao (China)
Ye Hongling (China)
Ye Qi (China)
Yeo Jingjie (USA)
Zhang Chuanzeng (Germany)
Zhang Guiyong (China)
Zhang Jian (China)
Zhang Lihai (Australia)
Zhang Yixia Sarah (Australia)
Zhang Zhao (China)
Zhao Liguo (UK)
Zheng Hui (China)
Zhong Zheng (China)
Zhou Anna (Australia)
Zhou Kun (Singapore)
Zhuang Zhuo (China)

# TABLE OF CONTENTS

# Comparative analysis of some results of static calculations obtained in the two-stage method

**\*Janusz Rębielak**[1]

[1]Retired professor of the Faculty of Architecture,
Cracow University of Technology,
ul. Warszawska 24, 31-155 Kraków, Poland.

\*Presenting author: j.rebielak@wp.pl

**Abstract**

The paper presents comparative analysis of the calculation results obtained in the two-stage method applied for the static calculations of selected, simple forms of the statically indeterminate trusses. By usage of the two-stage method one can determine, in a easy and fast way, the approximate values of forces acting in members of such types trusses by usage of equally easy methods like for instance the Cremona's or Ritter's method, which are applied only for calculation of the statically determinate trusses. The two-stage method uses in each of its stages the theorems and features of the calculus of vectors as well as principle of the superposition method. In the two-stage method the values of forces acting in members of the statically indeterminate truss are resultants of forces calculated in each stage in the counterpart members of the statically determinate trusses. In the paper there are discussed some different ways of selection of forms of the statically determinate trusses applied in each stage of the two-stage method together with analysis of an impact of a selected truss form on the accuracy of the final values of the calculated forces.

**Keywords:** Truss, static analysis, statically indeterminate system, principle of superposition, approximate solution, Cremona's method

## Introduction

Due to the structural features of statically indeterminate trusses, appropriate calculation methods should be used in order to precisely determine the values of the forces acting in their bars. These methods must always take into account the essential feature of this type of truss that the distribution of forces transmitted in their nodes is significantly influenced by the stiffness ratios of the bars connecting in these nodes. There are many methods of calculating statically indeterminate trusses, which are presented, inter alia, in the papers [1-8]. The concepts of these methods, long used in engineering practice, have been adapted in recent decades to the requirements of modern computer techniques and they are nowadays applied in many types of specialized computer software.

The two-stage method for the approximate calculation of statically determinate trusses was developed in the process of the preliminary analysis of a tension-strut truss. In some areas of such structural systems certain overloaded component parts of such systems are excluded from the process of force transfer . The number of system elements not participating in this process is equal to the number of the static indeterminacy degree of the basic truss. This observed regularity was the reason for the development of this new method, an exemplary scheme of which is shown in Fig. 1. The essence of this method is to perform calculations in two stages using calculation procedures appropriate for statically determinate systems. The

truss, the scheme of which is shown in Fig. 1a, is the statically indeterminate truss. The force values in its bars can be calculated by determining in the first stage the force values in the truss bars of the diagram shown in Fig. 1b, and in the second stage by calculating the forces in the truss bars of the diagram shown in Fig. 1c. In each of these stages, a number of bars equal to the static indeterminacy of the basic truss is removed, the intermediate trusses have the same structural span and height, and are loaded at the appropriate nodes with forces corresponding to half the forces loading nodes of the basic truss. The final values of the forces in the basic truss members are calculated as the resultant values of the forces in the relevant truss members calculated in each of the two stages. It is assumed that forces with a value equal to zero act in the bars removed in each stage. The application of this procedure follows from the rules of the calculus of vectors and the rule of superposition.
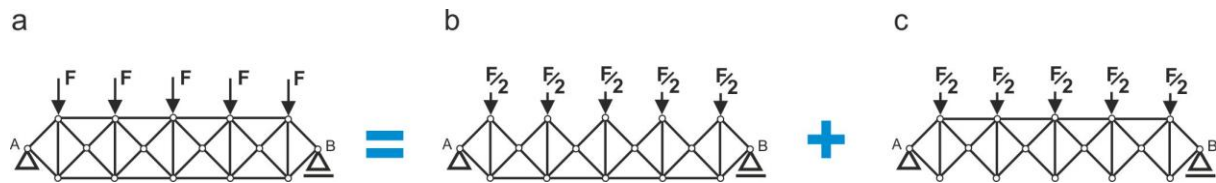


**Figure 1. Schemes of basic calculation procedure applied in the two-stage method**

The subject of considerations and analyzes carried out in the further part of this article are some possible variants of the application of the two-stage method for selected truss configurations with static schemes convergent or similar to the truss for the solution of which this new method was used for the first time [9,10].

**Initial comparative analysis of results obtained by application of the basic calculation procedure of the two-stage method**

It is assumed that number of nodes is defined by symbol "w", while symbol "p" defines number of members. Condition of the inner statically determinacy of plane truss is determined by the equation:

$$p = 2 \cdot w - 3 \tag{1}$$

The truss system presented in Fig.1a consists of number of nodes w = 16, what implies that the statically determinate truss created by means of this number of nodes has to be built by means of following number of members:

$$29 = 2 \cdot 16 - 3 \tag{2}$$

Truss of the scheme shown in Fig.1a is built by number of members p = 33 what indicates that the structure is the fourfold statically indeterminate system. It further implies that in each stage four members must be removed respectively from the top and then from the bottom chord. The basic form of the tested truss has a clear span between supports A and B equal to 5.0 meters, its construction depth equals 1.0 meter and it is loaded at each node of the upper chord with a concentrated force directed vertically downwards of 1.0 kN. Using the procedure of the two-stage method described above for the schemes shown in Fig. 1, the corresponding calculations were carried out, which were broadly and precisely presented in the paper [11], and the results are shown in Fig. 2a. For needs of the comparative analysis the forces in the members of the same truss were also calculated under the same conditions by application of Autodesk Robot Structural Analysis Professional 2016, which software takes into

consideration all requested mathematic tools necessary for precise computation of the force values in members of the statically indeterminate systems. It was assumed that the investigated truss is built of tubular members having diameter of 30.00 mm, thickness of section equal to 4.00 mm, while their steel material has the Young's modulus equal to 210 GPa. Results received in this way are shown in Fig. 2b.
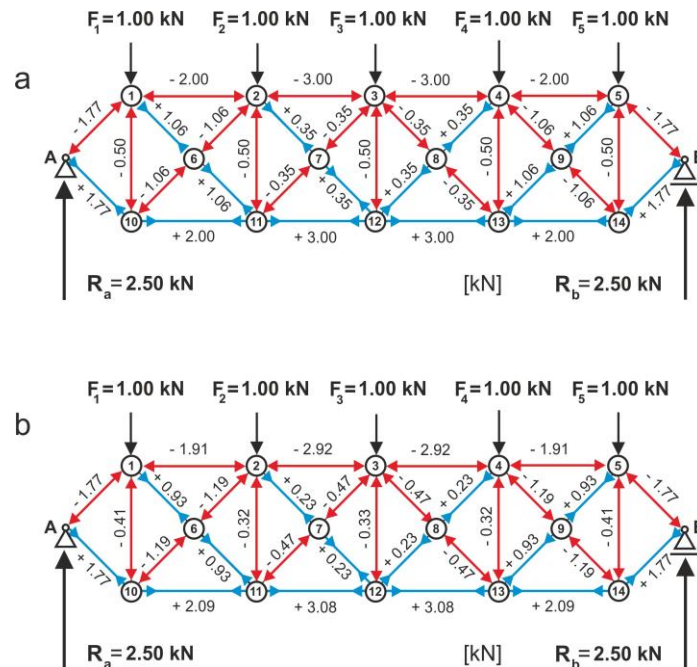


**Figure 2. Values of forces calculated in the same members of basic structure by application of, a) the two-stage method, b) a suitable computer software**

The comparative analysis of the data presented in Fig. 2 shows that in this case the results obtained by using the two-stage method are very similar to the results obtained by using the appropriate computer software. For example, the magnitude of the tensile force acting in the bar between nodes 11 and 12, determined using the two-stage method, is +3.00 kN, and calculated using a computer software is +3.08 kN. The relative difference of the obtained results is at the level of 2.66% in relation to the lower value and it can be considered negligible. Much greater differences one can be observe between the forces with the smallest absolute values. For example, the value of the force acting in the cross brace placed between the nodes 2 and 7 calculated in the two-stage method is equal to +0.35 kN, while using a computer software its value is defined as +0.23 kN. In this case the relative difference is as high as 52%. However such a large differentiation applies only to the magnitude of forces with absolute values close to zero, therefore it has a negligible impact on the final size of the bar cross-section, which in such cases results from the regulations contained in building standards.

**Comparative analysis of the calculated force values for the second case**

So far, the procedures of the two-stage method have been used, in which the number of bars located in the edge zones of the calculated trusses was reduced [12]. In the following parts of this article, cases of an appropriate reduction in the number of the cross braces of the trusses under study are considered and the impact of such a procedure on the entire calculation process is analyzed. The same basic truss, see Fig. 3a, is decided to be solved in two stages, in

which are solved the statically determinate trusses having the diagrams shown in Fig. 3b and Fig. 3c.
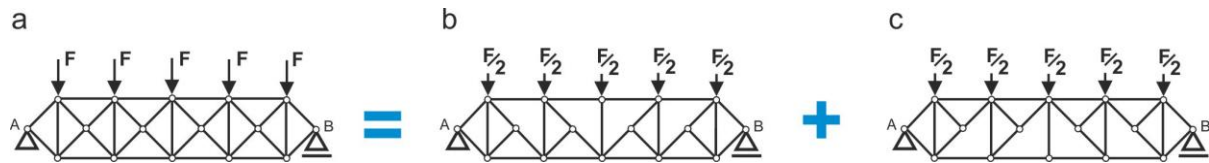


**Figure 3. Schemes of the second type of calculation procedure applied in the two-stage method**

Each of the both trusses is built by number of members p=29 and by number of nodes w=16. Thus the condition of the inner statically determinacy of plane truss is fulfilled:

$$29 = 2 \cdot 16 - 3 \qquad (3)$$

Although both of these trusses meet the condition of internal static determinability, but the direct application of, for example the Cremona's method to calculate the values of forces in their members, encounters certain difficulties. However, they are easily overcome thanks to one of the rules for determining the values of forces in the truss members without making a calculation. This useful rule states that if three bars join in an unloaded node, two of which are on one line, the third bar is a zero-force member. Taking into account this rule, the Cremona's method can be easily applied to determine the values of the forces acting in the truss members, calculated in each of the two stages. The results of these calculations together with the forms of the respective Cremona's polygons of forces are shown in Fig. 4 and in Fig. 5.
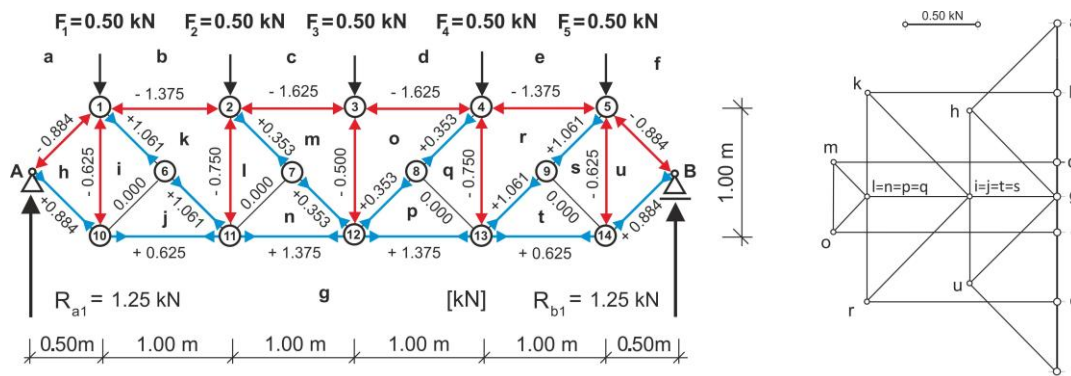


**Figure 4. Results of the first stage of calculation of the investigated truss system**
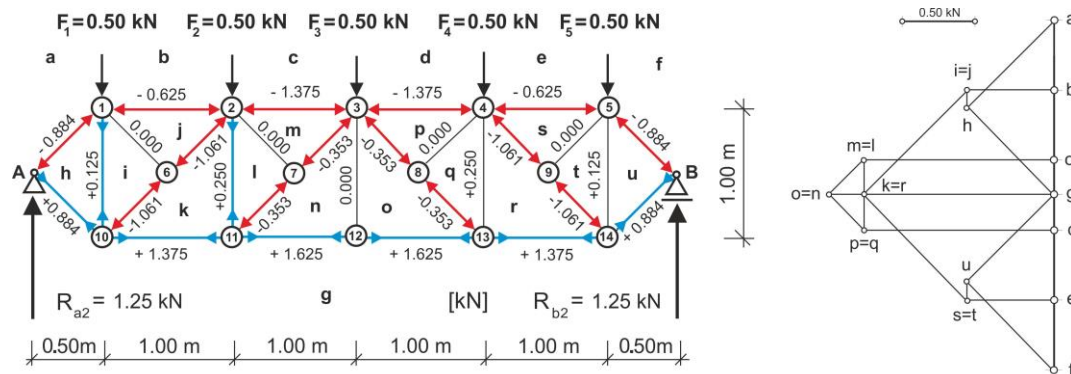


**Figure 5. Results of the second stage of calculation of the investigated truss system**

The final resultant values of the forces in the members of the basic truss, Fig. 3a, calculated in the two-step method using this procedure, are identical to the calculation results shown in Fig. 2a.

**Third comparative analysis of the calculated force values**

The analysis of the force system calculated in these two stages of the last procedure leads to an obvious conclusion that the basic statically indeterminate truss with the scheme shown in Fig. 6a can be calculated by the two-stage method using even simpler schemes of statically determinate trusses calculated in its two stages, see Fig. 6b and Fig. 6c.
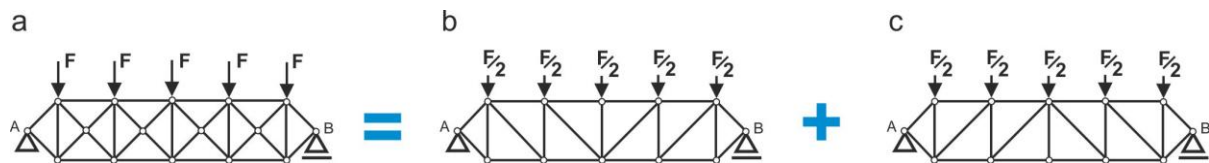


**Figure 6. Schemes of the third type of calculation procedure applied in the two-stage method**

Trusses having such static schemes were the subject of calculations at the appropriate stages of the two-stage method. The force values calculated in the first stage are shown in Fig. 7 together with the appropriate Cremona's polygon of forces. The results obtained in the second stage of calculations are shown in Fig. 8 also with the relevant Cremona's polygon of forces.
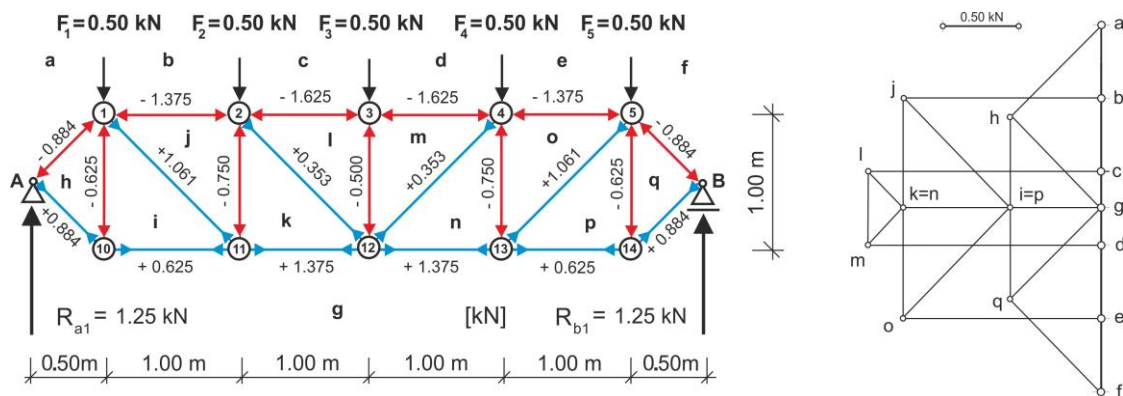


**Figure 7. Results of the first stage obtained in the third type of calculation procedure**
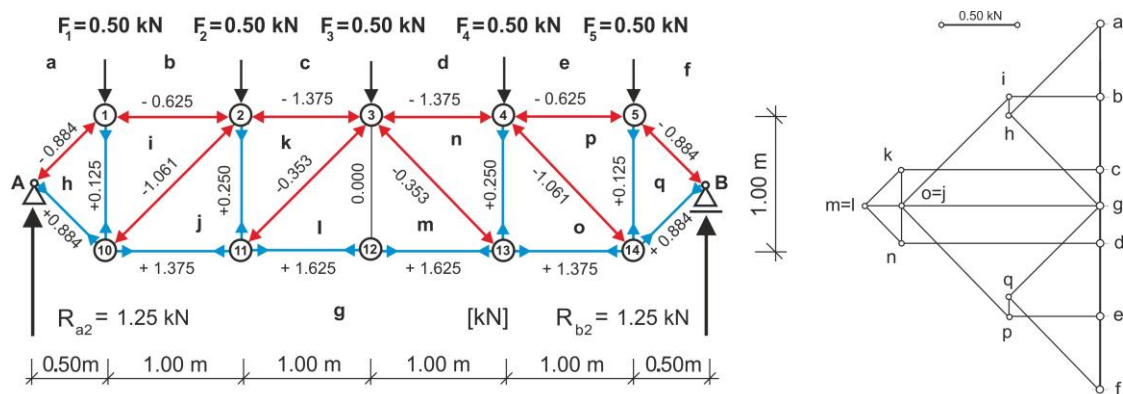


**Figure 8. Results of the second stage obtained in the third type of calculation procedure**

For easy recognition of the compared quantities in Fig. 7 and in Fig. 8 it was retained the numbering of nodes in the outer chords of the trusses used in the previous calculation procedures. In order to obtain the values of the forces acting in the basic truss, which is a statically indeterminate structure, it is necessary, in accordance with the principle of the two-stage method, to add together the values of forces calculated at each stage for members having the same positions in the area of the basic truss. As was the case for the second procedure, also the results of the calculations carried out according to the rule of the third procedure are identical to the results of the force values calculated in the course of the basic procedure, see Fig. 2a.

## Conclusions

The two-stage method makes possible the quick and easy calculation of the approximate values of forces acting in members of the statically indeterminate trusses, using in each of its two stages the procedures applied for calculating the statically determinate trusses. For a large part of the applied load cases and the geometric configuration of the trusses, the accuracy of the results obtained in this way can be considered sufficient. The application of the principles of symmetry and the superposition rule enables calculations to be carried out using the two-stage method according to various operating procedures. Some of them will be the subject of further research carried out in order to determine their accuracy and practical suitability for the design and analysis of the statically indeterminate trusses.

**References**

[1] Timoshenko, S.P. (1966) *History of strength of materials*, Arkady, Waszawa, - in Polish
[2] Makowski, Z.S. (1981) *Analysis, design and construction of double-layer grids*, Applied Science Publishers, London.
[3] Hibbeler, R.C.(1994). *Mechanics of materials*, Prentice Hall.
[4] Hibbeler, R.C. (1995). *Structural analysis*, Prentice Hall.
[5] Kolendowicz, T. (1993). *Theory of structures for architects*, Arkady, Warszawa, - in Polish.
[6] Przewłócki, J. and Górski, J. (2006). *Basis of theory of structures*, Arkady, Warszawa, - in Polish.
[7] Zienkiewicz, O.C. and Taylor R.L. (2000) *The finite element method*, Oxford Press, UK.
[8] Allen E., Zalewski W. and Boston Structures Group (2010) *Form and forces. Designing efficient, expressive structures*, John Wiley & Sons, Hoboken, New Jersey.
[9] Rębielak, J. (2014) A two-stage method for an approximate calculation of statically indeterminate trusses, *Journal of Civil Engineering and Architecture* **78**, 567-572.
[10] Rębielak, J. (2015) Examples of application of principle of superposition in the design of structural systems and in static analyses, *Journal of Mathematics and System Science* **5**, 150-155.
[11] Rębielak, J., (2018) Simple method of approximate calculation of statically indeterminate trusses, *International Journal of Computational Methods*, Vol. 15, No 1, DOI: 10.1142/S0219876218400261.
[12] Rębielak, J., (2021) Two-stage method applied for approximate calculations of selected types of statically indeterminate trusses, *International Journal of Computational Methods*, Vol. 18, No. 3, DOI: 10.1142/S0219876220410042.

# Blood Flow in Right Coronary Arteries with Multiple Stenoses

**Biyue Liu**

Department of Mathematics, Monmouth University, USA.

bliu@monmouth.edu

## Abstract

Atherosclerosis preferably occurs near the branches and bends of large and medium sized arteries. The presence of a moderate to severe stenosis significantly changes the downstream blood flow characteristics and might lead to the formation of the second stenosis. The fluid dynamic interaction of multiple stenoses is complex. The objective of this work is to study the blood flow in right coronary arteries with multiple stenoses. Various stenotic coronary models with different curvature of bend and different inter-stenotic distance are created. The computations are carried out using these models and the numerical results are compared with the blood flow simulation in a right coronary artery model reconstructed using the basic information from a coronary artery segment with two stenoses of a patient. Our results show that the curvature of bend significantly affects the wall shear stress (WSS) and pressure drop in curved artery with two moderate stenoses. The location of the distal stenosis strongly influences the flow pattern downstream, while the effect of the location of the proximal stenosis is insignificant.

**Keywords:** coronary artery, stenosis, blood flow

## 1. Introduction

A heart attack occurs when an area of heart muscle doesn't receive enough blood, usually by a clot in a coronary artery. The most frequent cause of the loss of blood supply to heart muscle is atherosclerosis. Atherosclerosis is a disease in which plaque builds up inside arteries, hardens and narrows the arteries (stenosis). The obstruction by a moderate to severe stenosis significantly affects the downstream blood flow pattern, such as flow shifting, pressure drop and the WSS. These flow characteristic changes may lead to the formation of a second stenosis in downstream. The fluid dynamic interaction of multiple stenoses in coronary arteries is complex.

Blood flows in arteries with multiple stenoses have been investigated by many researchers clinically and experimentally. De Bruyne *et al.* (2000) demonstrated that the hemodynamic assessment of a stenosis is significantly influenced by the presence of the second stenosis [1]. D'Souza *et al.* (2014) evaluated the diagnostic parameters, fractional flow reserve (FFR) and pressure drop coefficient (CDP), and observed the diagnostic uncertainties during assessment of serial coronary stenoses [2]. Computer simulations of blood flow in arteries with multiple stenoses have been conducted using two-dimensional models [3] or three dimensional models [4][5]. However, not much work have been reported on the systematic study of the blood flow in curved arteries with serial stenoses. The objective of this work is to study the blood flow in right coronary artery with multiple stenoses

## 2. Mathematical Models and Numerical Method

The blood flow in right coronary arteries with two stenoses is simulated using the time dependent three dimensional incompressible Navier-Stokes equations with the non-Newtonian viscosity of the Carreau model [6]. The blood density is assumed to be a constant at $1,050$ kg/m$^3$. The geometries of computational domain are shown in Fig. 1(a). Here P-patient model is reconstructed based on the inlet and outlet diameters of the IVUS slices, the locations of the proximal and distal stenoses, and the center-line curve of the angiographic image of the coronary segment of a patient [7]. The stenotic artery segment is 5.25 cm in length with a diameter of 0.3256 cm and 0.2954 cm at the inlet and the outlet, respectively. The inter-stenotic distance between the proximal and the distal stenoses is 2.69 cm (about 8.7 times the average diameter). Both stenoses have 75% of reduction in cross section area. To examine the effect of bend curvature on the blood flow through right coronary arteries with multiple stenoses, computer simulations are also performed on two more coronary models with the curvature of bend as 0.2 and 0.35 (labeled C-0.2 and C-0.35), respectively. These C-models have the same artery length, diameters of the inlet/outlet, and sizes of stenoses as those of the P-model.
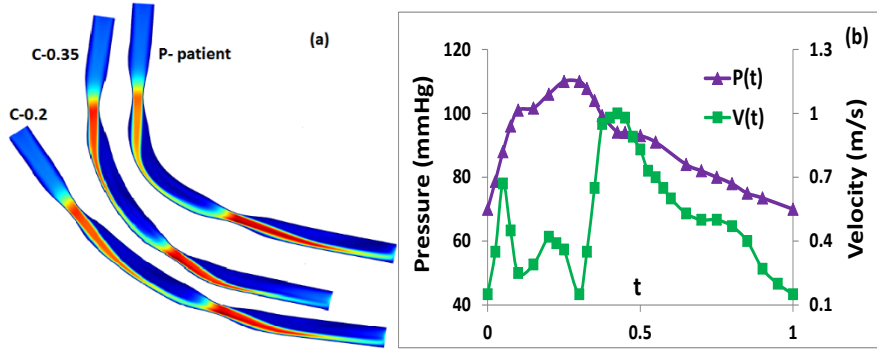


**Figure 1**. (a) Geometries of three coronary artery models with multiple stenoses; (b) Pressure and velocity waveforms at inlet/outlet

The boundary conditions are imposed as follows: A no-slip condition is applied to the lumen wall, treated to be inelastic and impermeable; A time dependent pressure with waveform P(t) (Fig. 1(b)) and a viscous-stress-free condition are applied at the inlet; A fully developed velocity profile with pulse waveform V(t) (Fig. 1(b)) is applied as the normal velocity at the outlet [4, 5]. Computations are performed using COMSOL5.3a. In computation each coronary artery segment is extended at the inlet and the outlet by 0.6 cm and 0.8 cm, respectively, to reduce the influence of the boundary conditions in the region of interest. Computations are repeated over four consecutive cardiac cycles to ensure a truly periodic flow.

## 3. Observation and Discussion

### 3.1 Effect of Curvature

The WSS plays an important role in the development of atherosclerosis. Intimal thickening likely occurs when the mean WSS is below 1 N/m$^2$, which presents an inverse hyperplasia with respect to the shear stress [8]. To examine how the curvature of bend affects the blood flow, computations are carried out on models with different curvature of bend. The curvature of bend for three models in Fig. 2 is in the order of C-0.2 < C-0.35 < P-patient. At each point on lumen wall, the time averaged WSS is calculated. In Fig. 2(a) the blue-colored areas

indicate where the temporal mean WSS are below 1 $N/m^2$. Forty-four cross section slices are evenly picked from the inlet to the outlet of the artery and a total of 128 points are picked approximately evenly spaced on the lumen boundary of each slice. The temporal mean pressure drop is calculated at each point and then averaged on 128 picked points of each slice. Fig. 2(b) plots the slice averaged temporal mean pressure drop along the axial length of the artery for each model. Here z = 0 and z = 5.25 cm correspond to the inlet and outlet, respectively. The extended inlet and outlet are not included in the plots of Fig. 2(b). Fig. 2 shows the effect of curvature on the WSS and pressure drop along the artery length. It is evident that as the curvature of bend increases the area with low WSS at the inner bend of the post-stenotic region for distal stenosis also increases significantly (Fig. 2(a)). The plots in Fig. 2(b) show that the pressure drop across the distal stenosis increases when the curvature of the coronary artery segment increases. Our observations from Fig. 2 indicate that the curvature is an important factor influencing the hemodynamics in coronary arteries with serial stenoses.
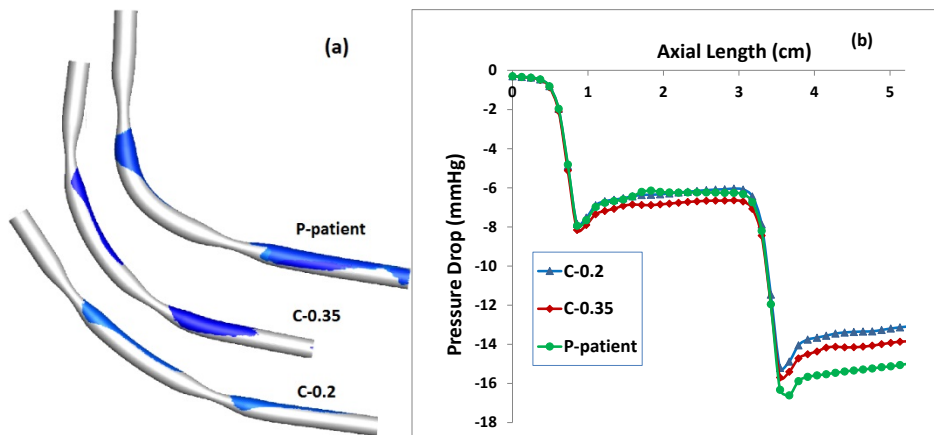


**Figure 2.** (a) Areas where the temporal mean WSS is below 1 $N/m^2$; (b) Slice averaged temporal mean pressure drop along the artery.

### 3.2 Effect of Stenosis Location

To examine the effect of the location of a stenosis on the blood flow in coronary arteries with multiple stenoses, computations are carried out on more models with various locations of either the proximal stenosis or the distal stenosis. Models of C-0.35P1, C-0.35P2 and C-0.35P3 are reconstructed by changing the location of the proximal stenosis of C-0.35 model. Thus inter-stenotic distance between the proximal and the distal stenoses changes from 2.69 cm to 2.32 cm, 1.96 cm and 1.59 cm, respectively. Models of C-0.35D1, C-0.35D2 and C-0.35D3 are reconstructed by changing the location of the distal stenosis of C-0.35 model. The inter-stenotic distance between the proximal and the distal stenoses is 1.59 cm, 1.96 cm and 2.32 cm, respectively.

The CDP is defined as the ratio of mean trans-stenotic pressure drop to proximal dynamic pressure. It is a non-dimensional flow resistance parameter for evaluating hemodynamic impact of coronary stenosis. Fig. 3 plots the slice averaged temporal mean CDP along the artery axial length for seven 0.35-curvature models with different location of (a) proximal or (b) distal stenosis. Fig. 3(a) shows that for the four models with the same location of the distal stenosis the varying location of the proximal stenosis has no significant effect on the CDP difference across the proximal stenosis. The small difference of the total CDP change at the outlet might be caused by the shorter inter-stenotic distance in the C-0.35P3 model. The

pressure drop of the flow from the proximal stenosis is not fully recovered yet before it moves into the distal stenosis. From Fig. 3(b) we can see that the CDP through the proximal stenosis is not influenced by the location of the distal stenosis, while the CDP in downstream significantly depends on the location of the distal stenosis. The change of CDP across the distal stenosis is larger when the distal stenosis locates further downstream. It is also evident from the plots in Fig. 3(b) that the change of CDP across the distal stenosis is larger than that across the proximal stenosis even though both stenoses have 75% reduction of cross-sectional area.
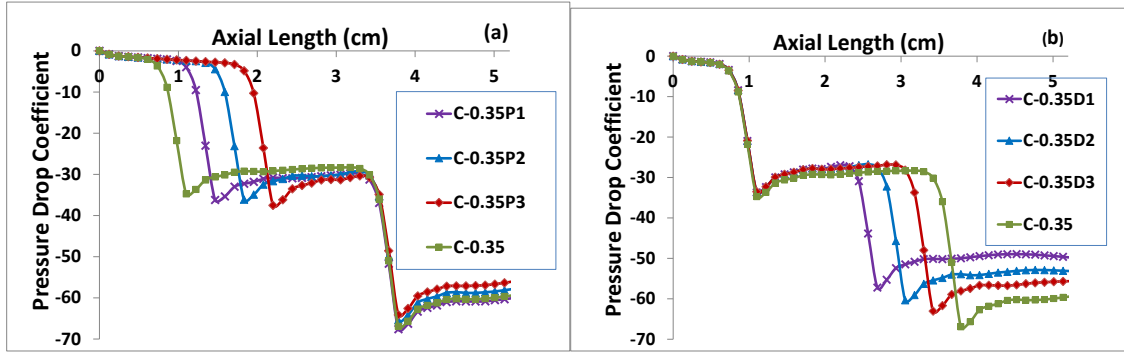


**Figure 3**. Slice averaged mean CDP along the axial length for 0.35-curvature models with (a) different location of proximal stenosis and (b) different location of distal stenosis

Fig. 4 plots the velocity magnitude along the center line for seven 0.35-curvature models with different location of (a) proximal or (b) distal stenosis. Fig. 4(a) demonstrates that four models with different location of the proximal stenosis have the same flow shifting magnitude through the proximal stenosis and have insignificant difference on the flow shifting through the distal stenosis. However, plots in Fig. 4(b) show that the location of the distal stenosis significantly affects the flow shifting through the distal stenosis while it has no influence on the flow shifting through the proximal stenosis.
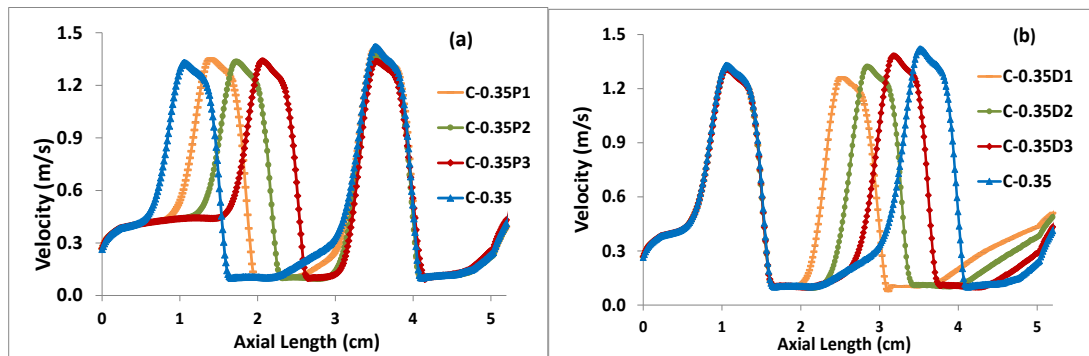


**Figure 4**. Plots of velocity magnitude along the center line for 0.35-curvature models with (a) different location of proximal stenosis and (b) different location of distal stenosis

## 4. Conclusion

Blood flows in coronary arteries with two moderate stenoses are simulated using various models with different curvature of bend and different location of either proximal or distal stenoses. Our observations show significant influences of the bend curvature and the distal stenosis location on the WSS, flow shifting and pressure drop of blood flow.

**References**

[1] De Bruyne, B. *et al.*. (2000) Pressure-derived fractional flow reserve to assess serial epicardial stenoses theoretical basis and animal validation. *Circulation* 101, 1840–184 .

[2] D'Souza, G. A., Peelukhana, S. V. and Banerjee, R. K.(2014) Diagnostic uncertainties during assessment of serial coronary stenoses: An *in vitro* study. *J. Biomech. Eng.* 136, 021026-1-11.

[3] Lee, T. S., Liao, W. and Low, H. T. (2003) Numerical simulation of turbulent flow through series stenosis. *Int. J. Numer. Meth. Fluids* 42, 717–740.

[4] Liu, B., Zheng, J., Bach, R. and Tang, D. (2017) Influences of flow parameters on pressure drop in a patient specific right coronary artery with two stenoses. In: O. Gervasi *et al.* (eds.) *Computational Science and Its Application — ICCSA 2017, LNCS 10404, Part I,* pp. 56–70. Springer.

[5] Liu, B., Tang, D. (2018) Influence of distal stenosis on blood flow through coronary serial stenoses: A numerical study. International Journal of Computational Methods doi.org/10.1142/S0219876218420033.

[6] Cho YI, Kensey KR (1991) Effects of the non-Newtonian viscosity of blood on flows in a diseased arterial vessel. Part 1: steady flows. Biorheology 28 pp. 241-262.

[7] Fan, R., *et al*.(2014) Human coronary plaque wall thickness correlated positively with flow shear stress and negatively with plaque wall stress: An IVUS-based fluid-structure interaction multi-patient study. *Biomed. Eng.Online* 13, 32.

[8] Ku, D.N. (1997) Blood flow in arteries, J. Ann. Rev. Fluid Mech, 29, pp.399-434.

# Structured grids based method with reformed boundary basis functions for solid and structure

[§][*]Yanan Liu[1], Fangxiong Tang[1] and Guansi Liu[1]

[1]China Special Equipment Inspection and Research Institute，Beijing 100029, China

*Presenting author: liuyanan@csei.org.cn

[§]Corresponding author: liuyanan@csei.org.cn

**Abstract:**

In this paper, a structured grid based basis functions are used for global approximation in solution domain. A boundary region is selected to create relevant boundary trial and test functions. The B-Spline basis functions are used to approximate the boundary curves (surfaces in 3D) and describe the given boundary region. In this boundary region, boundary basis functions are constructed based on the same B-Spline basis functions for representing the boundary region and weight functions are created for modification of the global basis functions and the boundary basis functions. The modified basis functions maintain high order continuity and can at least reconstruct linear polynomial. Based on the modified basis functions, the solution is created to make the essential boundary conditions automatically satisfied. Meanwhile, the accuracy of the solution near the essential boundaries can be guaranteed. Furthermore, the solution also can be constructed to improve the accuracy of solution in the local region in which complex deformation exists. The present method has been used to solve some 2-D elasticity problems. The numerical results are compared with analytical and finite element analysis solutions to show that the method is stable and effective.

**Keywords**: structured grids; boundary region; boundary basis functions; local refinement

## 1. Introduction

In finite element analysis, the domain of analysis is subdivided into elements and the resulting finite element mesh approximates the geometry and is used to construct the trial and test functions by piece-wise interpolation within each element. However, creating a satisfactory mesh that conforms to the boundaries of the domain is still one of the most burdensome tasks in finite element analysis. Mesh generation algorithms have been developed that work acceptably for most 2D problems but can be unreliable for some 3D geometries often resulting in poor or distorted elements in some regions that can lead to large error in the solution. A significant amount of user intervention is sometimes needed to correct such problems. An alternative to the conforming mesh approach has been to use non-conforming mesh, often a structured grid, for the analysis. A structured grid is much easier to generate than a finite element mesh and all the elements in the grid can have regular geometry (squares/rectangles/cubes).

In [1] the extended finite element method (XFEM) with a structured grid and implicit boundary representation is used for analysis. Approximate implicit function of the solid was

constructed using radial basis functions by fitting a set of sample points on the boundary. In [2] the structured grid based finite element approximation combining an implicit boundary method is used for elasticity analysis.

The shape functions constructed for the interpolation of the nodal variables in finite elements method only have $C^0$ continuity. The accuracy of solution with finite element method is not satisfactory in many problems. Many basis functions which have higher order continuity (at least $C^1$ continuity) and good approximation ability have been used in numerical methods for many years. Wavelet is a powerful mathematics tool in solving many problems in science and engineering. Wavelet scaling functions and wavelet functions are very popular basis functions for numerical analysis and simulation. Daubechies wavelet functions which are orthogonal have been widely used in various numerical methods [3-6]. The interpolation wavelet functions originated from Daubechies wavelets also perform well in function approximation [7]. Meanwhile, a variety of orthogonal and biorthogonal wavelet basis functions can be constructed by lifting scheme [8].

The meshless methods of analysis are based on a variety of basis functions which are constructed on discrete points. The most widely used basis (shape) functions for meshless methods are based on moving least squares (MLS) approximation which originated for the purpose of data fitting and surface construction. The MLS shape functions based meshless methods have been successfully used for solving many engineering problems [9]-[10]. Other meshless shape functions, such as the reproducing kernel particle (RKP) shape functions [11], the natural element interpolation shape functions [12] et al, also have good performance in numerical analysis and simulation.

The B-Spline functions are also important basis functions in numerical methods. In fact, a considerable body of literature now exists on the application of uniform and non-uniform B spline techniques to the solution of partial differential equations (PDEs) and mechanics problems. The recent studies of B-spline method can be found in some articles [13]-[14]. The isogeometric analysis method is a more general approach that uses B-Spline or non-uniform rational B-Spline (NURBS) for analysis [15]. In this method, B-Spline basis functions (or NURBS) are used for representing not only solution but also exact geometry of solution domain.

Whatever basis functions used for approximation, the traditional methods in FEM for applying essential boundary conditions cannot be used when a structured grid is used for analysis. The Lagrange multiplier methods [1][16], the penalty methods [17] and the modified variational principles [10] which are based on modified weak form can impose essential boundary conditions approximately. In the coupling with finite element method [18], essential boundary conditions can be applied to finite element nodes by standard methods. However, the solution around the boundary of the solution domain only has $C^0$ continuity in this approach.

The use of implicit equations for applying boundary conditions is another kind of method. In this approach, R-functions are used to construct implicit equations of the boundaries of the domain and solution structures are constructed to satisfy a variety of prescribed essential boundary conditions [2].

In the method presented here, a structured grid based basis functions are used for global approximation in problem domain. A boundary region is considered to create relevant

boundary trial and test functions. The B-Spline basis functions are used to approximate the boundaries curve (surface in 3D) and describe the given boundary region. In this boundary region, boundary basis functions are constructed based on the same B-Spline basis functions for representing the boundary region and a weight function is also created for modification of global basis functions and the boundary basis functions. The modified basis functions maintain high order continuity and can at least reconstruct linear polynomial. Based on the modified basis functions, the solution is created to make the essential boundary conditions automatically satisfied. Meanwhile, the solution is complete and the accuracy near the essential boundaries can be guaranteed even if complex characteristics exist. The main problem with structured grids is that the refinement for local regions is necessary because of the fixed global basis functions used for approximation. The solution also can be constructed for refinement in the local regions in which complex deformation exists. In this solution, the new B-spline basis functions are added into approximation. The local accuracy of solution is determined by the local meshes on which the new basis functions are constructed. The numerical examples of some 2-D elasticity problems are given to illustrate the stability and the effectiveness of the present method.

## 2 Approximation with a structure grid

A structure grid based approximation is very simple and easy to implement. Only one proper 1-D basic basis function is needed to represent the solution. For A proper 1-D basic basis function $\phi(x)$, it should at least possess following properties:

(1) Compact support

$$\text{Supp}\,\phi(x) = [0, m] \tag{1}$$

(2) Expressing polynomial up to $n(n \geq 1)$ order

$$x^{\alpha} = \sum_{i} a_{\alpha i} \phi(x-i) \qquad \alpha = 0, \cdots, n \tag{2}$$

where, $a_{\alpha i}$ are the coefficients.

(3) Enough Smoothness

$\phi(x)$ at least has $C^1$ continuity.

According to relevant theories and methods, some wavelet scaling functions, cardinal B-Spline functions and meshless shape functions constructed on nodes of regular distribution meet the above conditions.

In 1-D case, the approximation to solution $u$ defined on interval $[a, b]$ can be expressed as

$$u(x) = \sum_i c_i \phi(1/s * x - i) \tag{3}$$

where, $s$ is the size of 1-D grid.

The effective basis functions $\phi(1/s * x - i)$ which contribute to approximation can be determined by following rule

$$\text{Supp}\phi(1/s * x - i) \cap [a, b] \neq \varnothing \tag{4}$$

where,

$$\text{Supp}\phi(1/s * x - i) = [s * i, (m + i) * s] \tag{5}$$

In 2-D or 3-D case, the tensor product basis functions constructed by 1-D basis function can be used to express the solution $u$ defined in $\Omega$. The 3-D approximation can be written as

$$u(x, y, z) = \sum_{i_1} \sum_{i_2} \sum_{i_3} c_{i_1, i_2, i_3} \phi(1/s_1 * x - i_1) * \phi(1/s_2 * y - i_2) * \phi(1/s_3 * z - i_3) \tag{6}$$

where, $s_1, s_2, s_3$ are the size of 3-D grid in direction of $x, y, z$, respectively.

Similar to 1-D case, the effective 3-D tensor product basis functions which should be used for approximation can be determined by following rule

$$\text{Supp}\phi(1/s_1 * x - i_1) * \phi(1/s_2 * y - i_2) * \phi(1/s_3 * z - i_3) \cap \Omega \neq \varnothing \tag{7}$$

where, the support of 3-D tensor product basis functions can be easily determined by formula(5). For simplicity, the approximation to solution in high dimension can be written as

$$u(\mathbf{x}) = \sum_i c_i \phi_i(\mathbf{x}) \tag{8}$$

It should be noted that the multi-dimensional meshless basis functions for structure grid can be non-tensor product form. It can be constructed on regular nodes in multi-dimensional space.

The solution domain $\Omega$ will be covered by many small regular elements when a structure mesh is used for approximation. In practical computations, the integration should be evaluated in every active element which intersect domain $\Omega$. Figure 1 shows the active elements which consist of interior elements and boundary elements in 2-D case. The integration in interior elements is straightforward and relevant integral results in one interior element can be stored as parameters for all interior elements. The integration in boundary elements is more involved because only part of the element contributes to the solution domain $\Omega$. In this case, the boundary elements should be triangulated or quadrangulated and the integrations should be computed by performing quadrature on the triangles or quadrangles. This approach can lead

to relatively accurate integration computation in boundary elements. Some techniques have been developed to dispose of the boundary elements with complex boundary [1-2].
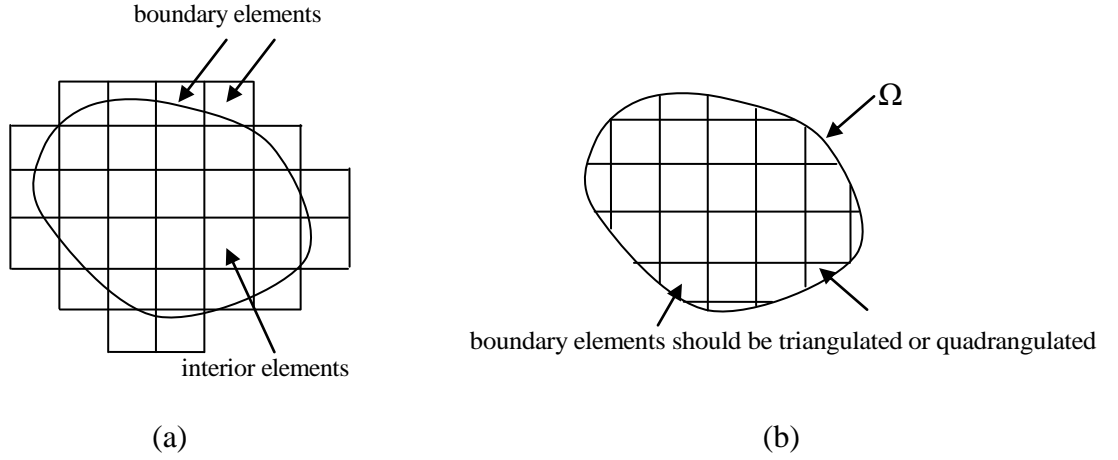


Figure 1. The active elements consist of interior elements and boundary elements

## 3 Basic method and formulation

### 3.1 B-Spline basis functions for describing geometry

B-spline functions are defined recursively starting with piecewise constants $(p = 0)$：

$$N_{i,0}(\xi) = \begin{cases} 1 & \text{if} \xi_i \le \xi < \xi_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

For $p = 1, 2, 3, \cdots$, they are defined by

$$N_{i,p}(\xi) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_{i,p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1,p-1}(\xi) \tag{10}$$

Important properties of B-spline basis functions are:

(1) They constitute a partition of unity, that is, $\forall \xi$

$$\sum_{i=1}^{n} N_{i,p}(\xi) = 1 \tag{11}$$

(2) The support of each $N_{i,p}$ is compact and contained in the interval $[\xi_i, \xi_{i+p+1}]$.

(3) Each basis function is non-negative, that is, $N_{i,p}(\xi) \ge 0, \forall \xi$.

B-spline curves in $\mathbb{R}^d$ are constructed by taking a linear combination of B-spline basis functions. The coefficients of the basis functions are referred to as control points. These are

somewhat analogous to nodal coordinates in finite element analysis. Piecewise linear interpolation of the control points gives the so-called control polygon. In general, control points are not interpolated by B-spline curves. given $n$ basis functions, $N_{i,p}$ $i = 1,2,3,\cdots,n$, and corresponding control points $B_i \in \mathrm{R}^d$ $i = 1,2,3,\cdots,n$, a piecewise-polynomial B-spline curve is given by

$$C(\xi) = \sum_{i=1}^{n} N_{i,p}(\xi)B_i \tag{12}$$

In 2-D space, the physical coordinate of the curve can be expressed as

$$x(\xi) = \sum_{i=1}^{n} N_{i,p}(\xi)x_i \tag{13}$$

$$y(\xi) = \sum_{i=1}^{n} N_{i,p}(\xi)y_i \tag{14}$$

where, $(x_i, y_i)$ is the coordinate of control points $B_i$.

Figure 2 shows quarter of a circle approximated by cubic B-spline functions. In this example, only five basis functions are used for approximation. It can be found that the error is negligible.
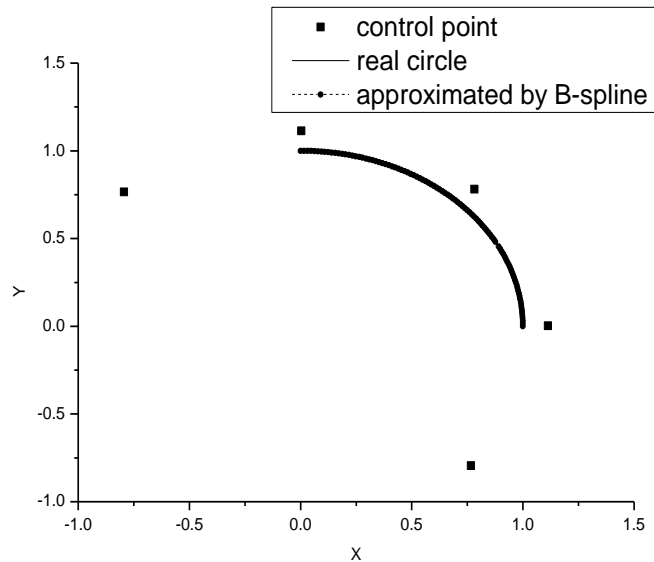


**Figure 2.   Quarter of a circle described by cubic B-spline**

Given a control net $B_{i,j} \in \mathrm{R}^d$, $i = 1,2,3,\cdots,n$, $j = 1,2,3,\cdots,m$, a tensor product B-spline surface is defined by

$$S(\xi,\eta) = \sum_{i=1}^{n}\sum_{j=1}^{m} N_{i,p}(\xi)M_{j,q}(\eta)B_{i,j} \tag{15}$$

where $N_{i,p}$ $M_{j,q}$ are the 2-D B-spline basis functions.

## 3.2 Boundary region and boundary basis functions

In 2-D problems, two curves can be used to describe the boundary region as shown in Figure 3. At first, the curves can be described by B-spline basis functions as mentioned above. For boundary $\Gamma_u$ and $\Gamma_{u'}$, the approximation can be respectively written as

$$\begin{cases} x_u(\xi) = \sum_{i=1}^{m} N_{i,p}(\xi)x_i^u \\ y_u(\xi) = \sum_{i=1}^{m} N_{i,p}(\xi)y_i^u \end{cases} \quad 0 \leq \xi \leq m \tag{16}$$

$$\begin{cases} x_{u'}(\xi) = \sum_{i=1}^{m} N_{i,p}(\xi)x_i^{u'} \\ y_{u'}(\xi) = \sum_{i=1}^{m} N_{i,p}(\xi)y_i^{u'} \end{cases} \quad 0 \leq \xi \leq m \tag{17}$$
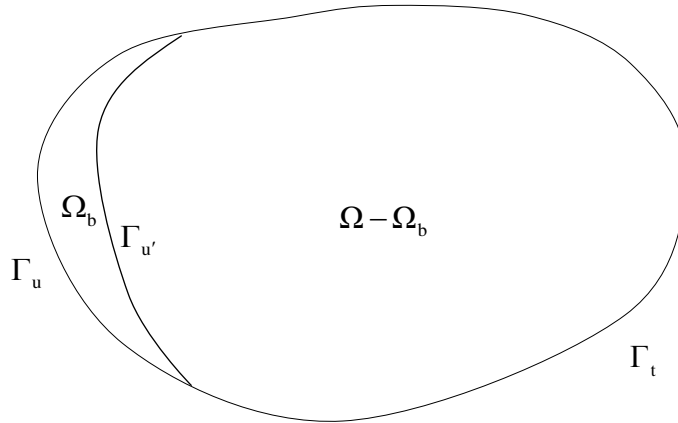


**Figure 3. The prescribed boundary region in 2-D domain**

Then, the boundary region $\Omega_b$ can be expressed as

$$\begin{cases} x(\xi,\eta) = \eta * \sum_{i=1}^{n} N_{i,p}(\xi)x_i^{u'} + (1-\eta)*\sum_{i=1}^{n} N_{i,p}(\xi)x_i^u \\ y(\xi,\eta) = \eta * \sum_{i=1}^{n} N_{i,p}(\xi)y_i^{u'} + (1-\eta)*\sum_{i=1}^{n} N_{i,p}(\xi)y_i^u \end{cases} \quad 0 \leq \eta \leq 1, 0 \leq \xi \leq n \tag{18}$$

In this boundary region, the boundary basis functions can be constructed as

$$\phi_{i,k}^{\text{b}}(\xi,\eta) = N_{i,p}(\xi) * L_k(\eta) \qquad 0 \le \eta \le 1, 0 \le \xi \le n \tag{19}$$

where, $N_{i,p}(\xi)$ are the same B-Spline basis functions in (16) and (17), and $L_k(\eta)$ can be the polynomial base functions.

$$L_k(\eta) = \eta^{k-1} \tag{20}$$

The B-Spline basis functions constructed in $[0,1]$ also can be chosen as $L_k(\eta)$ .

### 3.3 Solution for essential boundary condition

The governing equations for the elastostatics problem are expressed as follows

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{b} = 0 \quad \text{in} \quad \Omega \tag{21}$$

where

$$\boldsymbol{\sigma} = \mathbf{C} : \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} = \nabla_S \mathbf{u} \tag{22}$$

where, $\boldsymbol{\sigma}$ is the Cauchy stress, $\mathbf{b}$ is the body force, $\mathbf{C}$ is the Hooke's tensor, and $\boldsymbol{\varepsilon}$ is the strain, $\mathbf{u}$ is the displacement and $\nabla_S$ is the symmetric part of the gradient operator.

The essential and natural boundary conditions are, respectively

$$\mathbf{u} = \bar{\mathbf{u}} \qquad \text{on} \quad \Gamma_u$$
$$\mathbf{n} \cdot \boldsymbol{\sigma} = \bar{\mathbf{t}} \quad \text{on} \quad \Gamma_t \tag{23}$$

where $\Gamma_u$ and $\Gamma_t$ are the prescribed displacement and prescribed tension boundaries, respectively and $\mathbf{n}$ is the unit outward normal to domain $\Omega$ .

The weak form of the equations of elastostatics is

$$\int_\Omega \delta\boldsymbol{\varepsilon}^{\text{T}} \mathbf{C}\boldsymbol{\varepsilon}\,d\Omega = \int_{\Gamma_t} \delta\mathbf{u}^{\text{T}}\bar{\mathbf{t}}\,d\Gamma + \int_\Omega \delta\mathbf{u}^{\text{T}}\mathbf{b}\,d\Omega \tag{24}$$

where, $\delta\mathbf{u}$ is the virtual displacement and the corresponding virtual strain is $\delta\boldsymbol{\varepsilon}$ .

The space of trial functions $u$ and space of test functions $\delta u$ can be defined over $\Omega$ as

$$u = \{\mathbf{u} \in u \mid \mathbf{u} \text{ is smooth and } \mathbf{u}=\bar{\mathbf{u}} \text{ on } \Gamma_u\} \tag{25}$$

$$\delta u = \{\mathbf{u} \in \delta u \mid \mathbf{u} \text{ is smooth and } \mathbf{u}=\mathbf{0} \text{ on } \Gamma_u\} \tag{26}$$

For a structure mesh, the complete global approximation in $\Omega$ can be easily constructed on structure grid using wavelet scaling functions, meshless shape functions, and B-spline functions. We denote the global approximation as $\mathbf{u}^{\mathrm{g}} = \left\{ u_i^{\mathrm{g}}, i = 1, \cdots, n \right\}$ and $n$ is the dimension of problem.

The solution for admissible displacement can be constructed below

$$u_i = \begin{cases} u_i^{\mathrm{g}} & \text{in } \Omega - \Omega_{\mathrm{b}} \\ u_i^{\mathrm{b}} & \text{in } \Omega_{\mathrm{b}} \end{cases} \tag{27}$$

where, $u_i^{\mathrm{b}}$ denote the local approximation in the boundary region $\Omega_{\mathrm{b}}$. For the essential boundary condition $u_i = \overline{u}_i$ on $\Gamma_u$, $u_i^{\mathrm{b}}$ should be decomposed into two parts

$$u_i^{\mathrm{b}} = u_i^{\mathrm{bv}} + u_i^{\mathrm{bc}} \tag{28}$$

which can be respectively constructed in parameter space as

$$u_i^{\mathrm{bv}}(\xi, \zeta) = u_i^{\mathrm{g}}(\xi, \zeta) * (1 - w(\zeta)) + \sum_{i_1=1}^{m} \sum_{i_2=2}^{n} a_{i_1, i_2}^i \phi_{i_1, i_2}^{\mathrm{b}}(\xi, \zeta) * w(\zeta) \tag{29}$$

$$u_i^{\mathrm{bc}}(\xi, \zeta) = \sum_{i_1=1}^{m} a_{i_1}^{\overline{u}_i} \phi_{i_1, 1}^{\mathrm{b}}(\xi) * w(\zeta) \tag{30}$$

where, the $u_i^{\mathrm{bv}}$ is the variable that satisfy homogeneous boundary condition on $\Gamma_u$. The $u_i^{\mathrm{bc}}$ is related to the essential boundary condition and the control points $a_{i_1}^{\overline{u}_i}$ is constant which is used to construct following approximation

$$\overline{u}_i(\mathbf{x}(\xi)) =: \sum_{i_1=1}^{m} a_{i_1}^{\overline{u}_i} \phi_{i_1, 1}^{\mathrm{b}}(\xi) \quad \mathbf{x} \in \Gamma_u \tag{31}$$

$w(\zeta)$ is the basic weight function and can be chosen as

$$w(\zeta) = 1 - 6\zeta^2 + 8\zeta^3 - 3\zeta^4 \quad 0 \le \zeta \le 1.0 \tag{32}$$

$$w(\zeta) = \begin{cases} 1 - 2\zeta^3 & 0 \le \zeta \le 0.5 \\ 6\zeta - 12\zeta^2 + 6\zeta^3 & 0.5 \le \zeta \le 1.0 \end{cases} \tag{33}$$

$$w(\zeta) = \begin{cases} 1 - 6\zeta^2 + 6\zeta^3 & 0 \leq \zeta \leq 0.5 \\ 2 - 6\zeta + 6\zeta^2 - 2\zeta^3 & 0.5 \leq \zeta \leq 1.0 \end{cases} \tag{34}$$

It is obvious that the $u_i$ constructed above satisfies the essential boundary conditions $u_i = \bar{u}_i$ on

$\Gamma_u$. The high order continuity of modified global basis functions is also maintained.

Furthermore, the reconstruction of linear polynomial can be implemented using the modified basis functions.

The strains can be decomposed as

$$\boldsymbol{\varepsilon} = \begin{cases} \boldsymbol{\varepsilon}^g & \text{in } \Omega - \Omega_b \\ \boldsymbol{\varepsilon}^{bv} + \boldsymbol{\varepsilon}^{bc} & \text{in } \Omega_b \end{cases} \tag{35}$$

Substituting the formula (35) into weak form (24)

$$\int_{\Omega - \Omega_b} \delta\boldsymbol{\varepsilon}^T \mathbf{C} \boldsymbol{\varepsilon}^g d\Omega + \int_{\Omega_b} \delta\boldsymbol{\varepsilon}^T \mathbf{C} \boldsymbol{\varepsilon}^{bv} d\Omega = -\int_{\Omega_b} \delta\boldsymbol{\varepsilon}^T \mathbf{C} \boldsymbol{\varepsilon}^{bc} d\Omega + \int_{\Gamma_t} \delta\mathbf{u}^T \bar{\mathbf{t}} d\Gamma$$
$$+ \int_{\Omega - \Omega_b} \delta\mathbf{u}^T \mathbf{b} d\Omega + \int_{\Omega_b} \delta\mathbf{u}^T \mathbf{b} d\Omega \tag{36}$$

The virtual strain can be expressed as $\delta\boldsymbol{\varepsilon} = \mathbf{B}\delta\mathbf{a}$, where $\mathbf{B}$ is the strain-displacement matrix and $\mathbf{a}$ is the variable vector. From the solution structure, $\mathbf{B}$ and $\mathbf{a}$ can be decomposed into two parts respectively as

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_g & \mathbf{B}_b \end{pmatrix} \tag{37}$$

$$\mathbf{a} = \begin{pmatrix} \mathbf{a}_g \\ \mathbf{a}_b \end{pmatrix} \tag{38}$$

where, $\mathbf{B}_g$ and $\mathbf{a}_g$ are related to global basis functions, and $\mathbf{B}_b$ and $\mathbf{a}_b$ come from boundary

basis functions. $\boldsymbol{\varepsilon}^g$ and $\boldsymbol{\varepsilon}^{bc}$ can be expressed as

$$\boldsymbol{\varepsilon}^g = \mathbf{B}_g \mathbf{a}_g \qquad \text{in } \Omega - \Omega_b \tag{39}$$

$$\boldsymbol{\varepsilon}^{bc} = \begin{pmatrix} \mathbf{B}_g & \mathbf{B}_b \end{pmatrix} \begin{pmatrix} \mathbf{a}_g \\ \mathbf{a}_b \end{pmatrix} \qquad \text{in } \Omega_b \tag{40}$$

Then, the following linear algebraic equations can be obtained.

$$\mathbf{Ka} = \mathbf{f} \tag{41}$$

where, the stiffness matrix $\mathbf{K}$ can be decomposed into four parts.

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_g & \mathbf{K}_{gb} \\ \mathbf{K}_{gb}^T & \mathbf{K}_b \end{bmatrix} \tag{42}$$

and,

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_g \\ \mathbf{f}_b \end{bmatrix} \tag{43}$$

$$\mathbf{K}_g = \int_{\Omega-\Omega_b} \mathbf{B}_g^{\,T} \mathbf{C} \mathbf{B}_g d\Omega + \int_{\Omega_b} \mathbf{B}_g^{\,T} \mathbf{C} \mathbf{B}_g d\Omega \tag{44}$$

$$\mathbf{K}_b = \int_{\Omega_b} \mathbf{B}_b^{\,T} \mathbf{D} \mathbf{B}_b d\Omega \tag{45}$$

$$\mathbf{K}_{gb} = \int_{\Omega_b} \mathbf{B}_g^{\,T} \mathbf{D} \mathbf{B}_b d\Omega \tag{46}$$

$$\mathbf{f}_g^{\,T} = -\int_{\Omega_b} \mathbf{B}_g^{\,T} \mathbf{C} \boldsymbol{\varepsilon}^{bc} d\Omega + \int_{\Omega-\Omega_b} \boldsymbol{\phi}_g^{\,T} \mathbf{b} d\Omega + \int_{\Omega_b} \boldsymbol{\phi}_g^{\,T} \mathbf{b} d\Omega + \int_{\Gamma_t} \boldsymbol{\phi}_g^{\,T} \overline{\mathbf{t}} d\Gamma \tag{47}$$

$$\mathbf{f}_b^{\,T} = -\int_{\Omega_b} \mathbf{B}_b^{\,T} \mathbf{C} \boldsymbol{\varepsilon}^{bc} d\Omega + \int_{\Omega_b} \boldsymbol{\phi}_b^{\,T} \mathbf{b} d\Omega \tag{48}$$

where, $\boldsymbol{\phi}_g$ and $\boldsymbol{\phi}_b$ are the global basis functions vector and boundary basis functions vector, respectively. It should be noted that the relevant computation in boundary region $\Omega_b$ must be implemented in parameter space. The standard isoparametric transformation formulation should be used in computation.

**3.4 Solution for refinement in arbitrary boundary region**

The main problem with a structure grid is that the refinement for local region is necessary because of the fixed global basis functions used for approximation. When wavelet scaling functions are used as basis function for global approximation, the proper wavelet functions can be used for local refinement. In meshless analysis methods, local refinement can be implemented by constructing new basis functions in local region. For B-spline based refinement, using the hierarchical B-spline basis functions or T-spline method is an effective approach for local refinement.

In this paper, a general method is developed for local refinement. Similar to the method mentioned above, the approximation for solution can be constructed to improve the accuracy of solution in the local region where complex deformation exists. For simplicity, only boundary region in 2-D problems is considered for refinement in this paper. The present method can be easily extended to other complex cases.

The solution for displacement in arbitrary boundary region can be written as

$$u_i = \begin{cases} u_i^g & \text{in } \Omega - \Omega_b \\ u_i^b & \text{in } \Omega_b \end{cases} \tag{49}$$

where, $u_i^b$ denote the local approximation in the boundary region $\Omega_b$ and can be expressed in parameter space as

$$u_i^b(\boldsymbol{\xi}) = u_i^g(\boldsymbol{\xi}) * (1 - w_\alpha(\boldsymbol{\xi})) + \left( \sum_{i_1=1}^{m} \sum_{i_2=1}^{n} a_{i_1,i_2}^i \phi_{i_1,i_2}^b(\boldsymbol{\xi}) \right) * w_\alpha(\boldsymbol{\xi}) \tag{50}$$

where, $\boldsymbol{\xi} = (\xi, \zeta)$. In 2-D case, the weight function $w_\alpha(\boldsymbol{\xi})$ can be created as

$$w_\alpha(\xi, \zeta) = \begin{cases} 1 & 0 \le \zeta \le 1-\alpha \\ w((\zeta - 1 + \alpha)/\alpha) & 1-\alpha \le \zeta \le 1 \end{cases} \quad \alpha \le \xi \le m - \alpha \tag{51}$$

$$w_\alpha(\xi, \zeta) = \begin{cases} w((\alpha - \xi)/\alpha) & 0 \le \zeta \le 1-\alpha \\ w((\zeta - 1 + \alpha)/\alpha) * w((\alpha - \xi)/\alpha) & 1-\alpha \le \zeta \le 1 \end{cases} \quad 0 \le \xi \le \alpha \tag{52}$$

$$w_\alpha(\xi, \zeta) = \begin{cases} w((\zeta - m + \alpha)/\alpha) & 0 \le \zeta \le 1-\alpha \\ w((\zeta - 1 + \alpha)/\alpha) * w((\zeta - m + \alpha)/\alpha) & 1-\alpha \le \zeta \le 1 \end{cases} \quad m - \alpha \le \xi \le m \tag{53}$$

where, $\alpha$ is a parameter ($0 \le \alpha \le 1$). It is obvious that the hybrid basis functions in (50) are smooth and can at least reconstruct linear polynomial. Furthermore, the accuracy of refinement is determined by the local meshes on which the basis functions $\phi^b(\boldsymbol{\xi})$ are constructed

## 4 Numerical examples

In this part, numerical simulation of some elasticity problems is presented using the new method. The results are compared with those calculated by finite element method or analytical results to show the validity of the proposed method. For simplification, the units are omitted in this paper. In addition, B-Spline basis functions are used for global approximation in following examples.

### 4.1 Infinite plate with a central circular hole

An infinite plate with a central circular hole as shown in Figure 4 is considered. The plate is subjected to uniform unit traction along the $x$ axis at infinity. Symmetry of the problem is exploited to model only a quadrant of the analysis domain. The radius of the hole is chosen to be 0.25 and the length and width of the plate are chosen to be 1. The exact solution for this problem is used to compute the tractions on the traction boundaries (right and top boundaries) and symmetry boundary conditions are applied on the bottom and left edges while the inner

edge is traction free.
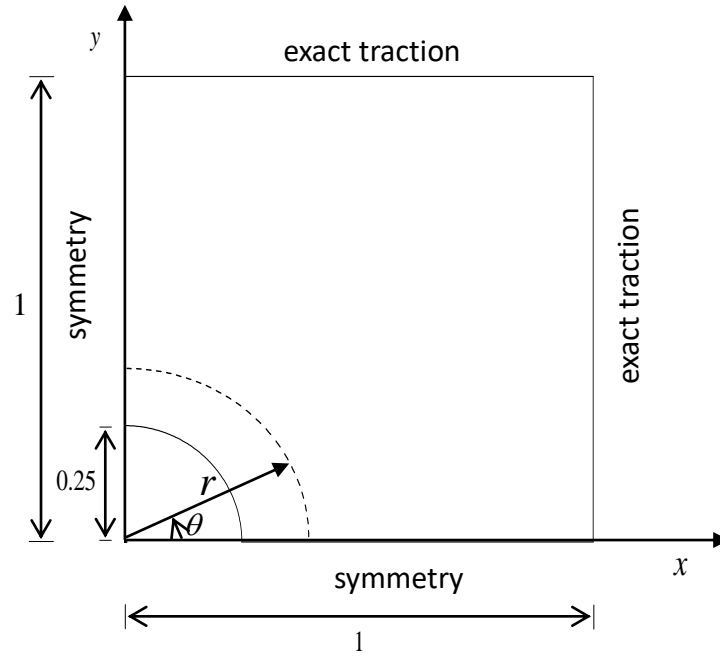


**Figure 4.    Quarter of a panel with a circle hole at its center**

The size of element is denoted by $w = w_x = w_y$. The selected boundary region is shown in

Figure 5 and its width is denoted by $t$. Cubic B-Spline basis functions are used for global approximation.
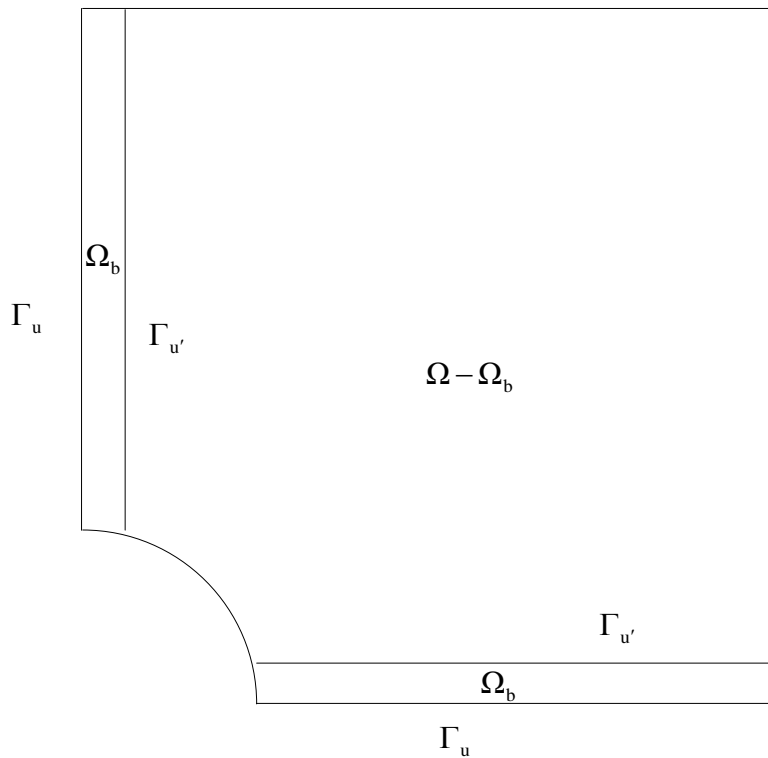


**Figure 5. The boundary regions in a quarter of a square plate with a circle hole**

Figure 6 shows the comparison of stress $(\sigma_x)$ value along the left edge with $w = 0.05$ and $t / w = 0.2$. It can be found that the results computed by the present method agree well with the analytical results even if stress concentration exists on the boundary.
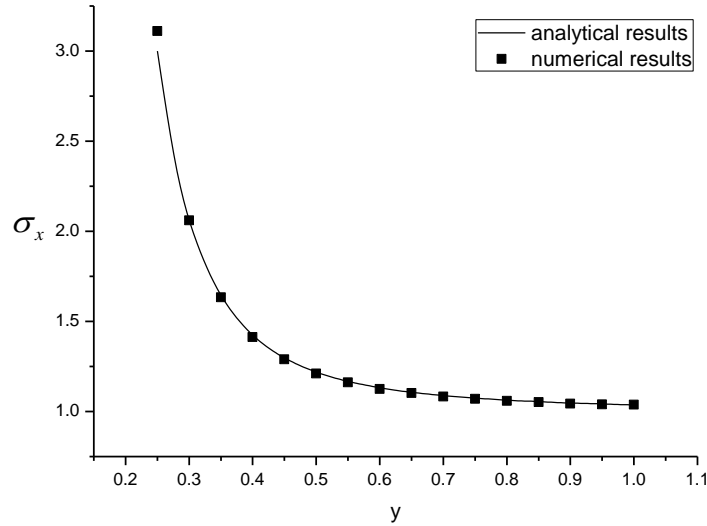


**Figure 6. The comparison of stress results along the left edge**
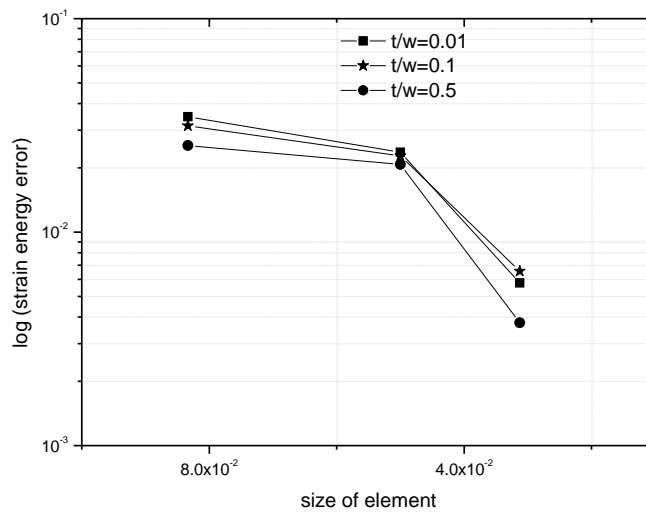


**Figure 7. Strain energy error for the problem of infinite plate with a circular hole**

The plot for strain energy error is shown in Figure 7. The x-axis represents the size of element. Different ratios of width of boundary region to size of grid are studied in error analysis for this problem. It can be found that the convergence rates under different ratios are almost same. However, the error is relatively small when large ratio is used. The results demonstrate that the accuracy of solution in selected boundary region can be improved using this method.

**4.2 Plate with a square hole**

Figure 8 depicts a quarter of a panel with a square hole of $a = 0.25$ at its center. The uniform

tension $p = 1$ in the horizontal direction is applied on the left and right edges of the panel. The boundary region and corresponding meshes for refinement are shown in Figure 9 The size of grid used for global approximation is $(0.05, 0.05)$.
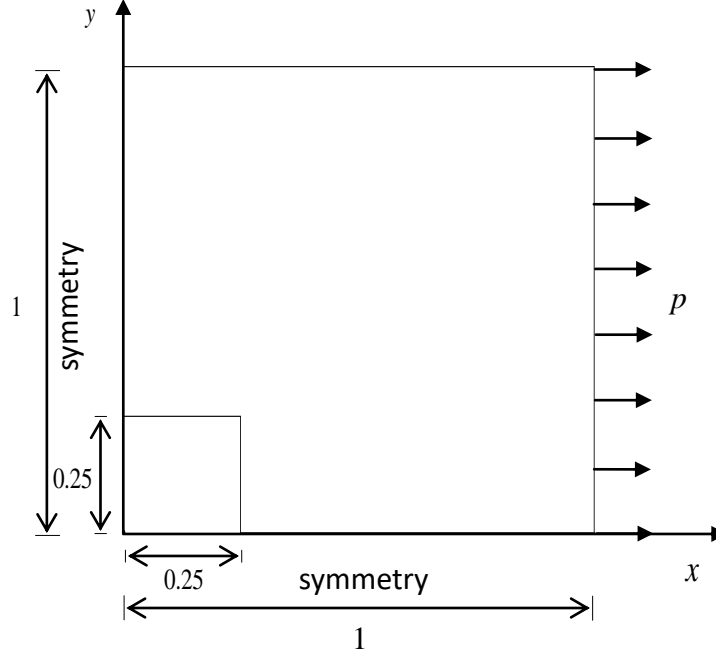


**Figure 8.   Quarter of a panel with a square hole at its center**

In this example, the parameter $\alpha$ for the weight function $w_\alpha(\xi)$ is determined as $\alpha = 1$. Figure 10 shows the stress results $(\sigma_x)$ along $y = 0.25$. Where, the results of FEM are computed using more than 22000 elements of plain eight nodes. It can be found that the results without refinement have unreasonable oscillations and the accuracy of solution near the singular point is very poor. As expected, these oscillations are disappeared in the results obtained by the present refinement method. Furthermore, the refined results can accurately catch the singular character even with less degree of freedoms added into approximation.
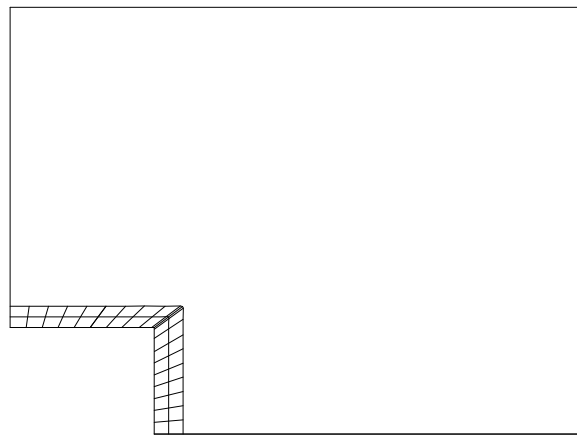


**Figure 9. The boundary region and corresponding meshes for refinement**
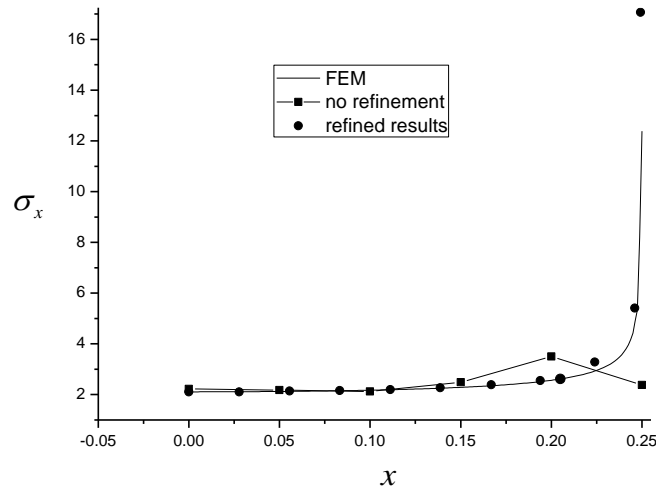
**Figure 10. The comparison of stress results along y=0.25**

The plot for strain energy error is shown in Figure 11. The x-axis represents the number of degrees of freedom. The strain energy obtained by FEM with more than 22000 elements of plain eight nodes is used for computing error. The x-axis represents the number of degrees of freedom. The local refinement is based on global approximation with the size of grid $(0.05, 0.05)$ and the region for refinement is shown in Figure 9. It is obvious that the convergence rate is remarkably improved when local refinement is performed by the present method.
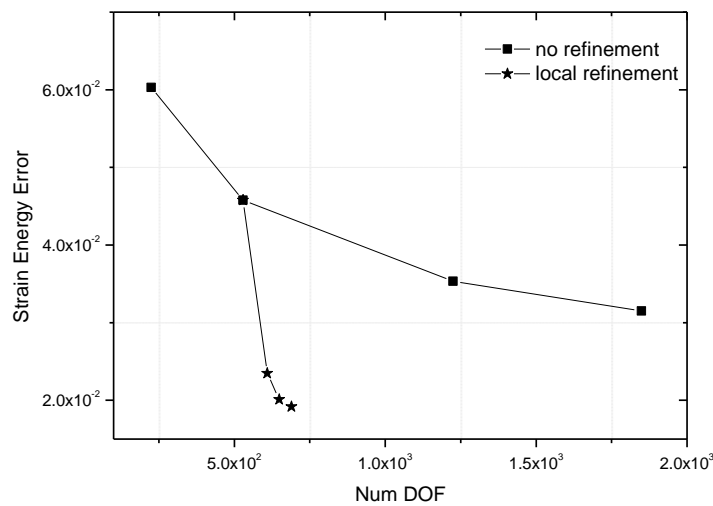


**Figure 11. Strain energy error for the problem of plate with a square hole**

## 5   Conclusion

In this paper, a structure grid is used for global approximation and a boundary region is selected to create relevant boundary trial and test functions. Based on this boundary region,

boundary basis functions and weight functions are constructed. The weight functions are used for modification of the global basis functions and the boundary basis functions. The modified basis functions maintain high order continuity and can at least reconstruct linear polynomial. Based on the modified basis functions, a highly accurate solution which automatically satisfies the essential boundary conditions can be created. Especially, the width of boundary region almost has no influence on the accuracy of solution.

The main problem with structured grids is that the refinement for local region is necessary because of the fixed global basis functions used for approximation. The solution also can be constructed for the local refinement by adding new basis functions constructed on local meshes into approximation. The numerical results in some elasticity problems show that the present method is stable and effective and it is potential for complex problems in sold and structure.

## Acknowledgements

## Reference

[1] T. Belytschko, C. Parimi, N. Moes, N. Sukumar, S. Usui, (2003) Structured extended finite element methods for solids defined by implicit surfaces, *International Journal for Numerical Methods in Engineering* 56, 609–635.

[2] A.V. Kumar, S. Padmanabhan, R.K. Burla, (2008) Implicit boundary method for finite element analysis using non-conforming mesh or grid, *International Journal for Numerical Methods in Engineering* 74, 1421–1447.

[3] K. Amaratunga, J.R. Williams, (1994) Wavelet-Galerkin solutions for one dimensional partial differential equation, *International Journal for Numerical Methods in Engineering* 37(16), 2703－2716.

[4] Y. Liu, Y. Liu, Z. Cen, (2008) Daubechies wavelet meshless method for 2-D elastic problems, *Tsinghua Science and Technology* 13, 605-608.

[5] Y. Liu, F. Qin, Y. Liu, Z. Cen, (2010) A Daubechies wavelet-based method for elastic problems, *Engineering Analysis with Boundary Elements* 34, 114-121.

[6] Y. Liu, F. Qin, Y. Liu, Z. Cen, (2010) The 2D large deformation analysis using Daubechies wavelet, *Computational Mechanics* 45, 179-187.

[7] S. Bertoluzza, G. Naldi, (1996) A wavelet collocation method for the numerical solution of partial differential equations, *Applied and Computational Harmonic Analysis* 3, 1–9.

[8] W. Sweldens, (1996) The lifting scheme: a custom-design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis* 3(2), 186–200.

[9] T. Belytschko, Y.Y. Lu, L. Gu, (1994) Element free Galerkin methods, *International Journal for Numerical Methods in Engineering* 37, 229-256

[10] Y.Y. Lu, T. Belytschko, (1994) A new implementation of the element free Galerkin method, *Computer Methods in Applied Mechanics Engineering* 113, 397-414

[11] W.K. Liu, S. Jun, Y.F. Zhang, (1995) Reproducing Kernel Particle methods, *International Journal for Numerical Methods in Fluids* 20, 1081-1106

[12] N. Sukumar, B. Moran, T. Bekytschko, (1998) The natural element method in solid mechanics,

*International Journal for Numerical Methods in Engineering* 43(5), 839–887.

[13] H. Naginoa, T. Mikamia, T. Mizusawa, (2008) Three-dimensional free vibration analysis of isotropic rectangular plates using the B-spline Ritz method, *Journal of Sound and Vibration* 317, 329–353.

[14] N. Caglar, H. Caglar, (2009) B-spline method for solving linear system of second-order boundary value problems, *Computers & Mathematics with Applications* 57, 757-762.

[15] T.J.R. Hughes, J.A. Cottrell, Y. Bazilevs, (2005) Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer Methods in Applied Mechanics Engineering* 194, 4135–4195.

[16] N. Moes, E. B échet, M.Tourbier, (2006) Imposing Dirichlet boundary conditions in the extended finite element method, *International Journal for Numerical Methods in Engineering* 67, 1641–1669.

[17] B.W. Clark, D.C. Anderson, (2003) The penalty boundary method, *Finite Elements in Analysis and Design* 39, 387–401.

[18] Y. Krongauz, T. Belytschko, (1996) Enforcement of essential boundary conditions in meshless approximations using finite elements, *Computer Methods in Applied Mechanics Engineering* 131, 133–145.

# Rapid seismic damage evaluation of subway stations using machine learning techniques

**\* Pengfei Huang [1], and † Zhiyi Chen [1,2]**

[1] Department of Geotechnical Engineering, Tongji University, Shanghai 200092, China.
[2] State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai

\* Presenting author: pfhuang@tongji.edu.cn
† Corresponding author: zhiyichen@tongji.edu.cn

**Abstract**

Rapid seismic damage evaluation of subway stations is critical for the efficient decision on the repair methods to damaged subway stations caused by earthquakes and rapid recovery of subway networks without much delay. However, the current methods to evaluate the damage state of a subway station after earthquakes are mainly field investigation by manual or computer vision, which is dangerous and time-consuming. Given this, a novel methodology that adopts machine learning technique as the classification model to rapidly and accurately evaluate the post-earthquake damage state of subway stations is proposed and the random forest (RF) is selected in the present paper. A two-story and three-span subway station in Shanghai is taken as an example to demonstrate the proposed methodology. The interrelated intensity measures of ground motions (IMs) and their uncorrelated principal components (PCs) are respectively taken as the input to investigate how the correlation between IMs affects the performance of the classification model. The results show that the accuracy of the model taking IMs as input and taking PCs as input is respectively 84.2% and 86.3%, revealing that taking PCs as input can improve the performance of RF.

**Keywords:** Rapid risk evaluation, Machine learning, Subway station, Random forest, Principal component analysis

## Introduction

To solve the conflict between limited land resources and the increasing demand for transportation capacity, recent years saw the rapid development of subway networks in China. However, researches have shown that severe damage to subway networks has occurred during earthquakes [1][2]. Specifically, the Daikai Subway Station was severely damaged during the 1995 Hyogoken-Namby earthquake in Japan. Considering the vulnerability of subway stations during earthquakes, it is of the essence to evaluate its post-earthquake damage state, which helps to estimate the economic losses and make decisions for the post-earthquake recovery process rapidly.

Normally, the damage state of structures can be evaluated through manual vision and fragility analysis [3]. For the former, the in-site inspection for the damaged subway station is very dangerous and time-consuming and the results can be subjectively affected according to the qualification of inspectors. The fragility analysis is a widely used method to quantitatively give the probability of a structure exceeding the predefined damage states. However, the application of fragility analysis to subway stations may be limited because only one intensity measure of ground motions (IM) and one demand measure (DM) of structures can be considered in the conventional fragility analysis [4]. Research from Liu et al. [5] reveals that

several DMs should be adopted to evaluate the post-earthquake damage state of subway stations. Additionally, taking only one IM to predicted the DMs of subway stations maybe not enough considering the random uncertainties in ground motions.

To avoid the above-mentioned issues, machine learning techniques such as Random Forest (RF) can be used as a prediction model to rapidly and accurately evaluate the post-earthquake damage state of subway stations. On the one hand, the machine-learning-based model can take multiple IMs as input to evaluate the post-earthquake damage state of subway stations. On the other hand, the output is the post-earthquake damage state, which means the model can predict the post-earthquake damage state by IMs directly. Compared with the fragility analysis, there is no intermediate process for predicting DMs and thus the limitations that only one DM can be considered can be avoided. Random Forest (RF), as a normally used machine learning technique, has been successfully applied to estimate the post-earthquake damage state of ground structures and bridges [3][6]-[8]. Multiple IMs are used as input and the relative importance of these IMs is well studied through the RF algorithm. However, the strong correlation the effects of the strong correlation between IMs on the accuracy of the RF model is not well investigated. Moreover, the relative importance of IMs for subway stations may be different from that for ground structures because they have different seismic response characteristics [9].

In the present study, the RF algorithm was adopted to develop a post-earthquake damage state prediction model. A two-story and three-span subway station in Shanghai was taken as an example to illustrate the proposed method. The Principal Component Analysis (PCA) [10] is used to transform the interrelated IMs into uncorrelated principal components (PCs). IMs and their PCs are respectively taken as the input of the prediction model, and thus two models are obtained. The performance of both models is evaluated to illustrate the effectiveness of the proposed method as well as to investigate how the correlation between IMs affects the performance of the prediction model.

**Overview of Random Forest**

*Decision Tree*

Decision tree (DT) is an independent classifier which can split a complex classification problem into a hierarchy of simple binary classification problem. The DT is comprised of a root node, interior nodes, and terminal nodes. The root node and each interior node can be regarded as a binary classifier. The predictor and the split point of the root node and each interior node are selected by minimizing the Gini index (GI). GI reflects the information entropy in a region, and a smaller GI means a better classifier. The GI is defined as:

$$GI = \sum_{m=1}^{M} p_{rm}(1 - p_{rm}),$$ (1)

where $M$ is the number of classes; $p_{rm}$ is the portion of samples in $r$ th region from $m$ th class. It is noteworthy that the performance of a DT on the training set can be improved by adding interior nodes. However, the number of interior nodes needs to be controlled to avoid overfitting.

**Fig. 1 Illustration of a decision tree**

Figure 1 gives the illustration of a decision tree. Starting from the root node, the DT recursively partitions the input space into two child nodes until the terminal nodes. Each terminal node represents a certain class. Assuming that there are a total of five predictor variables $(x_1, x_2, \cdots, x_5)$ and four classes (Class 1, Class 2, Class 3, and Class 4). The $x$ and $L$ in the root node and each interior node respectively denotes the predictor and the split point of the corresponding node. The tree initially judges the instance at the root node based on the $x_1$. If the $x_1$ is less than $L_1$, the instance will proceed to the Interior node 1, else the instance will proceed to the Interior node 2. Repeat this process until the terminal node is reached, and then the class of this instance can be determined.

*Random Forest*



**Fig. 2 Scheme to train a random forest**

Random forest (RF) is a commonly used machine-learning method for classification problems. As a revised version of bagging, the RF can be established by training a cluster of independent DTs (see Fig. 2). The training set for each DT is a subset of the original set with replacement (bootstrap sample). Assuming that there are $N$ samples in the original dataset. The probability of each sample not being selected $P$ can be calculated as:

$$P = \left(1 - \frac{1}{N}\right)^N. \tag{2}$$

Limit $N$ to positive infinity, and the $P$ will be equal to $1/e$, where $e \approx 2.718$. There are about $1/3$ of the original samples not selected, which is called out-of-bag (OOB). The OOB can be used as the validation samples to optimize the hyper-parameters of RF such as the number of interior nodes of each DT.

The RF can give the relative importance of each predictor variable by calculating the average changes in the impurity (represented by GI) of each node in all DTs caused by the predictor variable. The change in GI on the $m$ th node caused by the $j$ th predictor variable ($x_j$) is:

$$VIM_{jm}^{(i)} = GI_m - GI_l - GI_i, \tag{3}$$

where $VIM_{jm}^{(i)}$ is the variable importance measure of $x_j$ on $m$ th node for the $i$ th DT in RF; $GI_m$ is the GI on the $m$ th node before partition; $GI_l$ and $GI_r$ respectively denotes the GI on the left child node and right child node of the $m$ th node after partition. Assuming that there are $I$ decision trees in the RF and $M_i$ nodes where the predictor variable is $x_j$, the average changes in the GI caused by the $j$ th predictor variable ($VIM_j$) can be obtained as:

$$VIM_j = \sum_{i=1}^{I} \sum_{m=1}^{M_i} VIM_{jm}^{(i)}. \tag{4}$$

More detailed descriptions of DT and RF were provided by Mangalathu et al. [3].


**Case Study**

*Numerical Model*

A two-story three-span reinforced concrete subway station in Shanghai is selected to illustrate these machine-learning-based damage state prediction models. The subway station has a buried depth of 12.5 m and its typical cross-section is shown in Fig. 3 (a). The central columns have a width of 1 m and spacing of 8 m. The central columns are made of C45 concrete and other members are made of C35 concrete.

The numerical model of the soil-subway station dynamic system is established in the finite element software ABAQUS [11] as shown in Fig. 3 (b). The Mohr-Coulomb yield criterion is adopted to simulate the mechanical behavior of the surrounding soils, whose properties are shown in Table 1. The concrete damage plastic (CDP) model and idealized elastoplastic model are adopted to simulate the mechanical behavior of concrete and reinforcement, respectively. The CDP model adopting a degradation damage variable $d$ $(0 \leq d < 1)$ can simulate the degradation in the elastic modulus of concrete. The parameters of the CDP model

and the idealized elastoplastic model are listed in Table 2 – 4. The ABAQUS plane strain element CPE4 and CPE4R are respectively used to simulate the concrete and soils. The beam element B21 is adopted to simulate the reinforcement. The lateral boundary of surrounding soils is meshed by the CINPE4 element to eliminate the reflection of seismic waves at the boundary of the numerical model. NLTHA is conducted in two steps. Firstly, Gravity is applied to the numerical model whose bottom boundary is fixed in both horizontal and vertical directions (see Fig. 3 (b)), and thus the geostatic stress is obtained. Secondly, the ground motion is applied to the bottom boundary of the model after the bottom boundary is freed in the horizontal direction and then the seismic responses of the subway station can be obtained. More details of the numerical model were provided by Huang and Chen (2021) and such a method to establish the numerical model for subway stations has been validated by Chen and Liu [12].
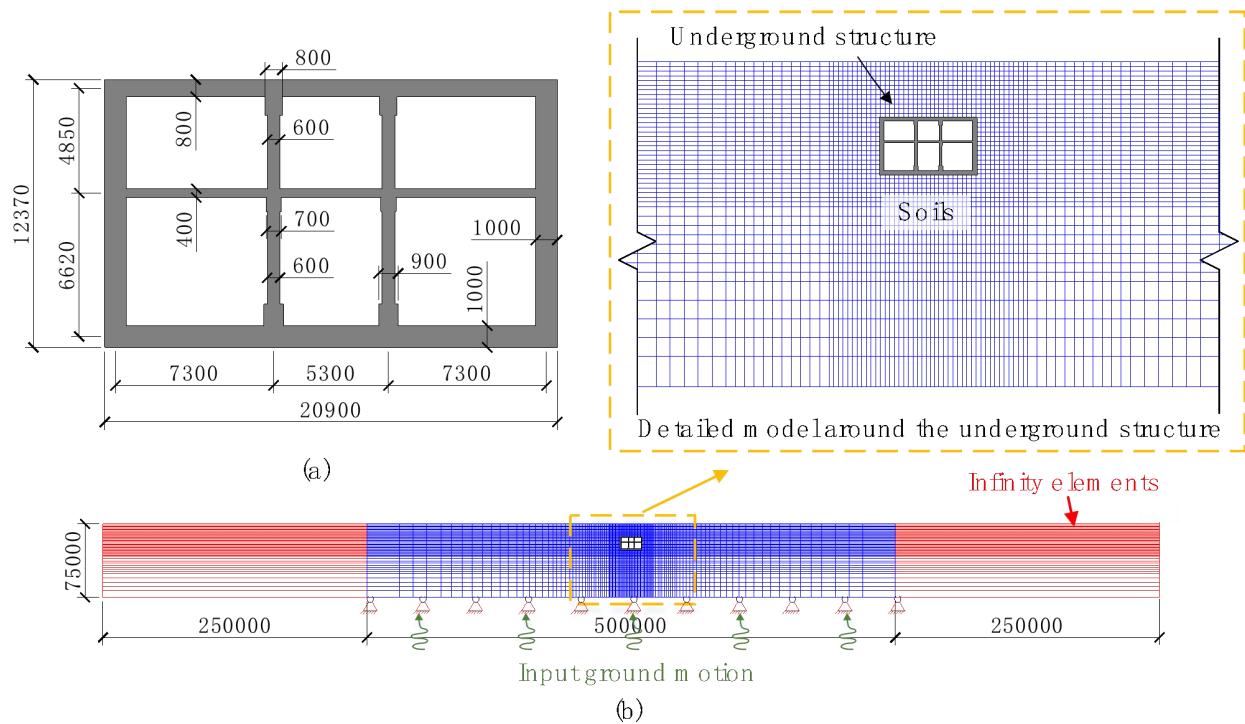


**Fig. 3 (a) Dimensions of the cross-section of the subway station and (b) numerical model for NLTHA (Unit: mm)**

**Table 1. Soil properties.**

| Soil layer | Thickness (m) | Density (kg/m³) | Elastic modulus (MPa) | Poisson Ratio | Cohesion (kPa) | Friction angle (º) |
|---|---|---|---|---|---|---|
| 1 | 1.30 | 1900 | 20.34 | 0.32 | 20.0 | 15.0 |
| 2 | 1.15 | 1920 | 20.34 | 0.32 | 9.5 | 31.3 |
| 3 | 1.11 | 1800 | 14.00 | 0.34 | 15.1 | 33.8 |
| 4 | 3.41 | 1740 | 10.85 | 0.38 | 5.3 | 28.3 |
| 5 | 8.13 | 1670 | 7.39 | 0.4 | 7.2 | 24.9 |
| 6 | 1.68 | 1740 | 11.55 | 0.35 | 10.0 | 29.7 |
| 7 | 4.39 | 1950 | 24.85 | 0.29 | 31.3 | 29.1 |
| 8 | 6.88 | 1820 | 32.2 | 0.29 | 2.0 | 31.1 |
| 9 | 15.05 | 1770 | 15.09 | 0.33 | 8.1 | 32.5 |
| 10 | 16.04 | 1840 | 28.7 | 0.32 | 8.0 | 28.1 |

**Table 2. Properties of C45 concrete.**

| Parameters | Value |
|---|---|
| Density (kg/m$^3$) | 2500 |
| Elastic modulus (MPa) | 31500 |
| Dilation angle (º) | 30 |
| Poisson ratio | 0.2 |
| Hardening and damage | (Table 3) |

**Table 3. Compression and tensile hardening and damage index versus plastic strain of C45 concrete.**

| Compression | | | Tensile | | |
|---|---|---|---|---|---|
| Inelastic strain (10$^{-3}$) | Stress (MPa) | Damage index | Cracking strain (10$^{-3}$) | Stress (MPa) | Damage index |
| 0 | 12.910 | 0 | 0 | 2.200 | 0 |
| 0.476 | 22.780 | 0.1664 | 0.083 | 1.785 | 0.1283 |
| 0.909 | 23.310 | 0.2709 | 0.160 | 1.244 | 0.2891 |
| 1.775 | 19.600 | 0.4631 | 0.295 | 0.777 | 0.545 |
| 3.5 | 12.370 | 0.7295 | 0.545 | 0.472 | 0.7846 |
| 4.69 | 9.388 | 0.8265 | 0.889 | 0.322 | 0.8971 |
| 6.26 | 6.977 | 0.8955 | 1.392 | 0.230 | 0.9502 |
| 7.79 | 5.522 | 0.9311 | 1.992 | 0.177 | 0.9726 |
| 9.31 | 4.560 | 0.9514 | 2.592 | 0.146 | 0.9824 |
| 10.82 | 3.880 | 0.964 | 2.991 | 0.133 | 0.9862 |
| 13.31 | 3.105 | 0.9764 | | | |

**Table 4. Reinforcement properties.**

| Parameters | Value |
|---|---|
| Density (kg/m$^3$) | 7800 |
| Elastic modulus (GPa) | 200 |
| Poisson ratio | 0.3 |
| Yield stress (MPa) | 400 |

*Intensity measures, demand measures, and damage states*

IMs are important indexes characterizing the intensity of ground motions, which can be used as the input of prediction models. One or a few IMs are unlikely to cover all the information in the time history of ground motions used in NLTHA. Therefore, the underground structures have different seismic responses when subjected to different ground motions even if these ground motions have the same IMs owing to the randomness of ground motions. According to Kiani et al. [13], 11 representative IMs are selected for the present study, as shown in Table 5.

**Table 5. Basic information of the selected 11 IMs.**

| No. | Name | Symbol | Definition |
|---|---|---|---|
| 1 | Peak acceleration | $P_a$ | $\max\left(\left\|a\left(t\right)\right\|\right)$ |

| | | | |
|---|---|---|---|
| 2 | Peak velocity | $P_v$ | $\max\left(\lvert v(t)\rvert\right)$ |
| 3 | Peak displacement | $P_d$ | $\max\left(\lvert d(t)\rvert\right)$ |
| 4 | Spectral acceleration | $S_a(T)$ | $S_a(t)\big\vert_{t=T}$ |
| 5 | Spectral velocity | $S_v(T)$ | $S_v(t)\big\vert_{t=T}$ |
| 6 | Spectral displacement | $S_d(T)$ | $S_d(t)\big\vert_{t=T}$ |
| 7 | Acceleration spectral intensity | $ASI$ | $ASI = \int_{0.1}^{0.5} S_a(T)dT$ |
| 8 | Spectral intensity | $SI$ | $SI = \int_{0.1}^{2.5} S_v(T)dT$ |
| 9 | Displacement spectral intensity | $DSI$ | $DSI = \int_{2}^{5} S_d(T)dT$ |
| 10 | Cumulative absolute energy | $CAV$ | $CAV = \int_{0}^{\infty} \lvert a(t)\rvert dt$ |
| 11 | Arias intensity | $AI$ | $AI = \dfrac{\pi}{2g}\int_{0}^{\infty} [a(t)]^2 dt$ |

DMs are the representative value of seismic responses of underground structures, based on which the damage state of underground structures can be determined. In the present study, the maximum story drift angle ( $\theta_l$ ), maximum column drift angle ( $\theta_c$ ), damage at the ends of central column ( $CCD$ ), and Energy ratio ( $R_E$ ) are selected as DM according to Liu et al. (2017). The four DMs are defined as Eq. (23) – (26)

$$\theta_l^{(i)} = \frac{d_{l,top}^{(i)} - d_{l,bottom}^{(i)}}{h_l^{(i)}}, \tag{5}$$

$$\theta_c^{(i)} = \frac{d_{c,top}^{(i)} - d_{c,bottom}^{(i)}}{h_c^{(i)}}, \tag{6}$$

$$CCD^{(j)} = \sum_{r \in R_j} \frac{E_d^{(r)}}{\sum_{r \in R_j} E_d^{(r)}} DC^{(r)}, \tag{7}$$

$$R_E = \frac{E_P + E_{DMD}}{E_P + E_{DMD} + E_{KE} + E_E}. \tag{8}$$

where $\theta_l^{(i)}$ and $\theta_c^{(i)}$ are respectively the layer drift angle and column drift angle of the $i$ th story ( $i = 1,\ 2$ in this paper); $d_{l,top}^{(i)}$ , $d_{l,bottom}^{(i)}$ , $d_{c,top}^{(i)}$ , and $d_{c,bottom}^{(i)}$ are respectively the displacements of the top plate, bottom plate, top end of the central column, and bottom end of the central column of the $i$ th story; $h_l^{(i)}$ and $h_c^{(i)}$ are respectively the height of the $i$ th story and that of the column at the $i$ th story; $CCD^{(j)}$ is the damage at the $j$ th end of the central column ( $j = 1, 2, \cdots, 8$ in this paper); $R_j$ is the set of elements in the region of the $j$ th end; $E_d^{(r)}$ is the dissipation energy of element $r$ ; $DC^{(r)}$ is the average compression damage index

of element $r$ ; $E_P$ , $E_{DMD}$ , $E_{KE}$ , and $E_E$ are respectively the plastic energy, damage energy, kinetic energy, and elastic energy.

Four limit states are defined in FEMA-273 (1997): operational (OP), slight damage (SD), life safety (LS), and collapse prevention (CP). The state of the subway station in the elastic stage is referred to as no damage (ND). The relationship between the column drift and the damage state was provided by Liu et al. [5] and shown in Table 6. As long as one of these DMs reaches the limit value defined in Table 6, the station will be regarded as entering the corresponding damage state.

**Table 6. Thresholds of each limit state of subway stations.**

| State | $\theta_l$ | $\theta_c$ | CCD | $R_E$ |
|-------|-----------|-----------|------|-------|
| ND | <0.0011 | <0.0018 | <0.07 | <0.25 |
| OP | 0.0011 | 0.0018 | 0.07 | 0.25 |
| SD | 0.0025 | 0.004 | 0.35 | 0.5 |
| LS | 0.0059 | 0.0083 | 0.8 | 0.75 |
| CP | 0.013 | 0.0167 | 0.95 | 0.9 |

*Ground motions and samples*

Enough ground motion records are required for generating sample sets and training machine-learning-based damage state prediction models. The two-story and three-span subway station is located in Shanghai, which covers deep soft soils. This kind of stratum has the characteristic of a long predominant period. Given that there are not enough natural ground motion records in Shanghai, artificial ground motions were generated according to the geological characteristics of Shanghai. The scheme of obtaining samples and training the RF is given in Fig. 4. A total of 730 Shanghai artificial ground motions are generated using the spectral representation method proposed by Liu et al. [14].
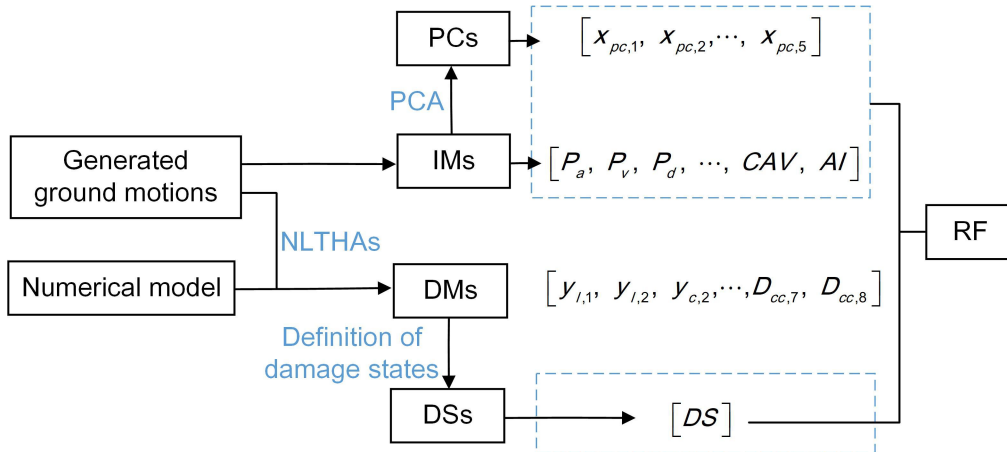


**Fig. 4 Scheme of obtaining samples and training the RF**

Intensity measures (IMs) of these artificial ground motions including $P_a$ , $P_v$ , $P_d$ , $S_a(T)$ , $S_v(T)$ , $S_d(T)$ , $ASI$ , $SI$ , $DSI$ , $CAV$ , and $AI$ can be calculated according to Table 5. It is noteworthy that the 11 IMs are highly correlated with each other. To investigate how the

correlation between IMs affect the performance of these damage state prediction model, principal component analysis (PCA) [10] is used to transform these interrelated IMs into uncorrelated variables and The first $p$ (In this paper $p=5$) principal components (PCs) ($x_{pc,1}$, $x_{pc,2}$, …, $x_{pc,p}$) with a cumulative contribution rate exceeding 99% were taken as the input variables. The IMs and their PCs are respectively taken as inputs to obtain the corresponding models. On the other hand, the DMs of the subway station including $\theta_l$, $\theta_c$, $CCD$ and $R_E$ can be calculated by conducting 730 NLTHA and the damage state of the subway station can be obtained following the definition in Table. 6. One-fifth of the total samples are taken as test samples and the other samples are taken as training samples. It is noteworthy that there are about one-third of the training samples not in the bootstrap sample and taken as validation samples as mentioned in section 2.2.

**Result and discussion**

The RF is adopted as a damage state prediction model for the subway station. The original IMs and their PCs are respectively taken as inputs to obtain the two models. The training set is used to optimize the parameters of the RF (predictor and the split point of each interior node of each DT), the validation set is used to optimize the hyper-parameters (the maximum tree depth, the trees numbers, etc. for RF) and the test set is used to evaluate the performance of both models.
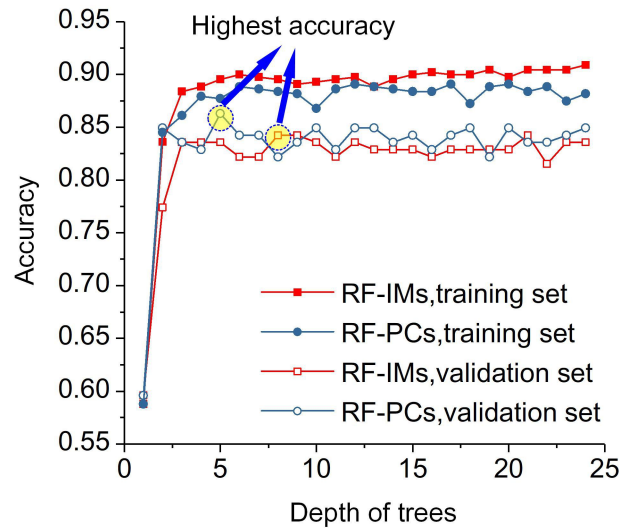


**Fig. 5 Training and Validation accuracy with the depth of trees**

Figure 5 gives the sensitivity analysis of the tree depth for RF. The RF taking PCs as inputs (PCs-RF) gets its best performance with a shallower tree depth than RF taking IMs as inputs (IMs-RF) does and the highest accuracy for the former is also higher than that for the latter. The reason may be that compared with taking IMs as input, taking PCs as input makes the selection of predictor and the split point of each interior node easier.

The performance of both prediction models can be further evaluated by a confusion matrix as shown in Fig. 6. The confusion matrix is a table of the actual damage state versus the

predicted damage state. The element $C_{ij}$ ( $i=1:5$ , $j=1:5$ ) in a confusion matrix means the number of the samples whose actual damage state is $DS_{i-1}$ but predicted damage state is $DS_{j-1}$ (Marks $DS_0$ , $DS_1$ , $DS_2$ , $DS_3$ , and $DS_4$ respectively denotes the five damage state i.e. no damage (ND), operational (OP), slight damage (SD), life safety (LS), and collapse prevention (CP)). For example, The element $C_{12}=2$ in Fig. 6 (a)'s matrix means that there are 2 samples whose actual damage state is $DS_0$ , but it is wrongly predicted as the $DS_1$ by the IMs-RF. Therefore, the diagonal elements in the confusion matrix indicate the number of samples whose damage state is correctly predicted by the machine-learning-based model, and for $C_{ij}$ the smaller the $|i-j|$ , the smaller the error of the predicted damage state. The confusion matrix for both models is the tridiagonal matrix, which means the predicted damage state is neighbor, if not equal to the actual damage state, showing a good performance. Additionally, there are three indexes namely recall ratio, precision ratio, and total accuracy can be used to evaluate the performance of the machine-learning-based models. The recall ratio given in the last column of each confusion matrix is defined as the percentage of the actual damage state that is correctly predicted by the machine-learning-based models; the precision ratio given in the last row of each confusion matrix is defined as the percentage of predicted damage states that are correctly predicted by the machine-learning-based models; the total accuracy given in the bottom right corner is defined as the ratio of the damage states that are correctly predicted to the total test samples.

(a)

| Actual state | Predicted state | | | | | Recall ratio (%) |
| --- | --- | --- | --- | --- | --- | --- |
| | $DS_0$ | $DS_1$ | $DS_2$ | $DS_3$ | $DS_4$ | |
| $DS_0$ | 20 | 2 | 0 | 0 | 0 | 90.1 |
| $DS_1$ | 0 | 17 | 4 | 0 | 0 | 81.0 |
| $DS_2$ | 0 | 3 | 22 | 1 | 0 | 84.6 |
| $DS_3$ | 0 | 0 | 2 | 38 | 8 | 79.2 |
| $DS_4$ | 0 | 0 | 0 | 3 | 26 | 89.7 |
| Precision ratio (%) | 100 | 77.3 | 78.6 | 90.5 | 76.5 | 84.2 |

IMs-RF

(b)

| Actual state | Predicted state | | | | | Recall ratio (%) |
| --- | --- | --- | --- | --- | --- | --- |
| | $DS_0$ | $DS_1$ | $DS_2$ | $DS_3$ | $DS_4$ | |
| $DS_0$ | 21 | 1 | 0 | 0 | 0 | 95.5 |
| $DS_1$ | 1 | 17 | 3 | 0 | 0 | 81.0 |
| $DS_2$ | 0 | 1 | 24 | 1 | 0 | 92.3 |
| $DS_3$ | 0 | 0 | 3 | 38 | 7 | 79.2 |
| $DS_4$ | 0 | 0 | 0 | 3 | 26 | 89.7 |
| Precision ratio (%) | 95.5 | 89.5 | 80.0 | 90.5 | 78.8 | 86.3 |

PCs-RF

**Fig. 6 Confusion matrix of test set for (a) IMs-RF, (h) PCs-RF.**

Figure 6 shows that the prediction accuracy of IMs-RF and PCs-RF is respectively 84.2% and 86.3%, revealing that the proposed method that using RF as a prediction model can evaluate the post-earthquake damage state accurately. The performance of these machine-learning-based models can be further evaluated by analyzing the recall ratio and precision ratio of these models. Figure 6 shows that the lowest recall ratio tends to appear at the $DS_3$ and the lowest precision ratio tends to appear at the $DS_4$ . The reason may be that the element $C_{45}$ in each confusion matrix is large, which means the samples whose actual damage state is $DS_3$ tend to be predicted as $DS_4$ and the samples whose predicted damage state is $DS_4$ actually belong to $DS_3$ . This finding reveals that the $DS_3$ and $DS_4$ cannot be well discriminated by the selected IMs.
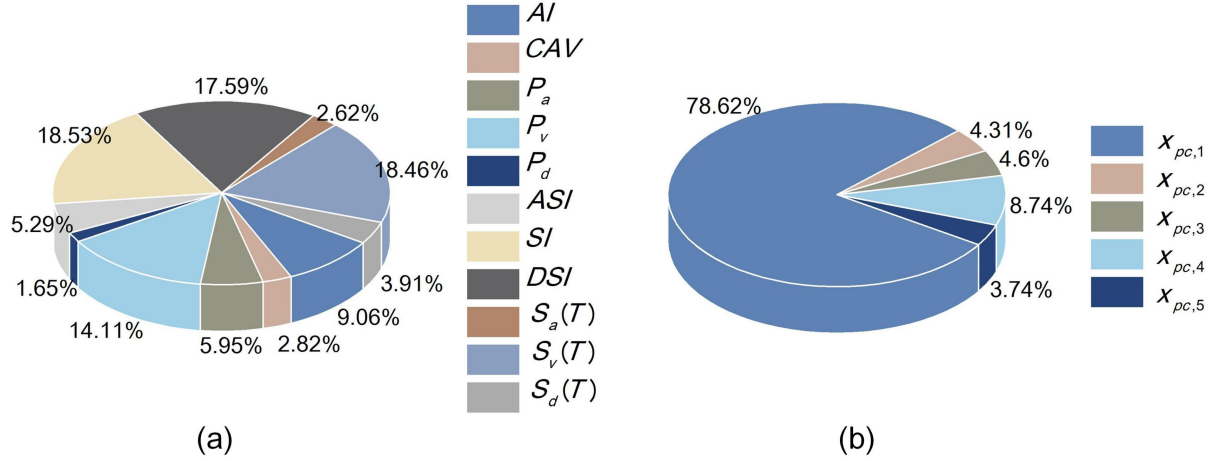
**Fig. 7 Relative importance for (a) IMs-RF and (b) PCs-RF.**

A comparison between the performance of IMs-RF and PCs-RF reveals that taking PCs as input effectively improves the prediction performance. The reason, as mentioned before, may be that taking PCs as input makes the selection of predictor and the split point of each interior node easier. The relative importance of inputs for IMs-RF and PCs-RF is given in Fig. 7 to further investigate which IM is most relevant to the post-damage state of the subway station. For IMs-RF, the five most important IMs are $SI$, $S_v(T)$, $DSI$, $P_v$, and $AI$ and their relative importance is respectively 18.53%, 18.46%, 17.59%, 14.11%, and 9.06 %. For the PCs-RF, the relative importance of $x_{pc,1}$, $x_{pc,2}$, …, $x_{pc,5}$ is respectively 78.62%, 4.31%, 4.6%, 8.74%, 3.74%. The PCA can be regarded as a pre-processing of input data and gives the most influential input, makes the election of predictor and the split point of each interior node more easy and effective. On the other hand, the five most important IMs reveals that the velocity-related parameters of ground motion (such as $SI$ and $P_v$) are closely related to the seismic response of underground structures, which is consistent with the previous researches [15].

**Conclusions**

In the present study, the RF is adopted as a classification model to rapidly and accurately evaluate the post-earthquake damage state of subway stations. The interrelated IMs and their uncorrelated PCs are respectively taken as the input to investigate how the correlation between IMs affects the performance of these models. The main conclusions are as follows:

The confusion matrix for both IMs-RF and PCs-PF is a tridiagonal matrix and the prediction accuracy of both models is more than 80%, revealing that the proposed methodology can rapidly and accurately evaluate the post-earthquake damage state of subway stations.

The PCs-PF has higher accuracy (86.3%) than IMs-RF (84.2%), which means taking PCs as input can improve the performance of RF because PCA can transform the interrelated IMs into uncorrelated variables, which makes the election of predictor and the split point of each interior node more easy and effective.

Among the selected 11 IMs, the five most important IMs are $SI$, $S_v(T)$, $DSI$, $P_v$, and $AI$ and their relative importance is respectively 18.53%, 18.46%, 17.59%, 14.11%, and 9.06%.

The results reveal that the velocity-related parameters of ground motion (such as $SI$ and $P_v$) is closely related to the seismic response of underground structures.

## Acknowledgments

**References**

[1] Iida H., Hiroto T., Yoshida N. and Iwafuji M. (1996) Damage to Daikai Subway Station, *Soils and Foundations* **36**, 283-300. https://doi.org/https://doi.org/10.3208/sandf.36.Special_283

[2] Huo H., Bobet A., Fernández G. and Ramírez J. (2005) Load Transfer Mechanisms between Underground Structure and Surrounding Ground: Evaluation of the Failure of the Daikai Station, *Journal of Geotechnical and Geoenvironmental Engineering* **131**, 1522-1533. https://doi.org/10.1061/(ASCE)1090-0241(2005)131:12(1522)

[3] Mangalathu S., Hwang S., Choi E. and Jeon J. (2019) Rapid seismic damage evaluation of bridge portfolios using machine learning techniques, *Engineering Structures* **201**, 109785. https://doi.org/10.1016/j.engstruct.2019.109785

[4] Huang P. and Chen Z. (2021) Fragility analysis for subway station using artificial neural network, Journal of Earthquake Engineering (submitted).

[5] Liu T., Chen Z., Yuan Y. and Shao X. (2017) Fragility analysis of a subway station structure by incremental dynamic analysis, *Advances in Structural Engineering* **20**, 1111-1124. https://doi.org/10.1177/1369433216671319

[6] Mangalathu S., Sun H., Nweke C. C., Yi Z. and Burton H. V. (2020) Classifying earthquake damage to buildings using machine learning, *Earthquake Spectra* **36**, 183-208. https://doi.org/10.1177/8755293019878137

[7] Xu Y., Lu X., Tian Y. and Huang Y. (2020) Real-time seismic damage prediction and comparison of various ground motion intensity measures based on machine learning, *Journal of Earthquake Engineering*, 1-21. https://doi.org/10.1080/13632469.2020.1826371

[8] Xie Y., Ebad Sichani M., Padgett J. E. and DesRoches R. (2020) The promise of implementing machine learning in earthquake engineering; a state-of-the-art review, *Earthquake Spectra* **36**, 1769-1801. https://doi.org/10.1177/8755293020919419

[9] Chen Z. and Huang P. (2021) Seismic response characteristics of multi-story subway station through shaking table test, Advances in Structural Engineering (submitted).

[10] Shang X., Li X., Morales-Esteban A. and Chen G. (2017) Improving microseismic event and quarry blast classification using Artificial Neural Networks based on Principal Component Analysis, *Soil Dynamics and Earthquake Engineering* **99**, 142-149. https://doi.org/https://doi.org/10.1016/j.soildyn.2017.05.008

[11] ABAQUS User Manual: Abaqus Analysis User's Guide Version 6.14. (2014) Dassault Systemes Simulia Corp., Providence, RI.

[12] Chen Z. Y. and Liu Z. Q. (2019) Effects of pulse-like earthquake motions on a typical subway station structure obtained in shaking-table tests, *Engineering Structures* **198**, 109557. https://doi.org/10.1016/j.engstruct.2019.109557

[13] Kiani J., Camp C. and Pezeshk S. (2019) On the application of machine learning techniques to derive seismic fragility curves, *Computers & Structures* **218**, 108-122. https://doi.org/10.1016/j.compstruc.2019.03.004

[14] Liu Z. J., Zeng B. and Wu L. Q. (2015) Simulation of non-stationary ground motion by spectral representation and random function, *Journal of Vibration Engineering* **3**, 411-417.

[15] Chen Z. and Wei J. (2013) Correlation between ground motion parameters and lining damage indices for mountain tunnels, *Natural Hazards* **65**, 1683-1702. https://doi.org/10.1007/s11069-012-0437-5

# Nonlinear analysis of steel frames under thermal loading

†*J. Petrolito[1], D. Ionescu[1], A. Sim[2] and T. Lougoon[3]

[1]Department of Engineering, La Trobe University, Australia.

[2]RMG, Australia.

[3]Jacobs, Australia.

*Presenting author: j.petrolito@latrobe.edu.au

†Corresponding author: j.petrolito@latrobe.edu.au

## Abstract

The analysis of building frames under thermal loading is an important task in structural engineering. Linear analysis is generally inadequate for large temperature changes, and nonlinear analysis is essential to correctly model the behaviour of the structure. Various nonlinear models are used in practice and these models usually involve some form of approximation of the nonlinear effects. In this case, it is not often clear what effects the approximations have on the analysis.

This paper discusses a general approach for the nonlinear analysis of steel frames under thermal loading. In particular, we consider the role of approximations in the governing equations of the structure and the effects of the changes in material properties at high temperatures. Several examples are given in the paper that demonstrate the various effects of the simplifying approximations on the results.

Keywords: Frame analysis, nonlinear analysis, thermal effects, fire conditions.

## Introduction

Thermal effects are generally required to be considered in the analysis of building frames. Two types of thermal effects need to be considered, namely:

1. Thermal effects under normal daily temperature ranges, and

2. Thermal effects under elevated temperatures, such as fire conditions.

In the first case, the material properties can be considered to be essentially unchanged due to the temperature changes. However, thermal effects may still adversely affect a structure due to the expansion and contraction of the members. An example of this is thermal buckling, such as in the case of continuously-welded railway tracks under high daily temperatures [1].

The second case is more complicated because the material properties can no longer be considered constant. In particular, steel loses both rigidity and strength at high temperatures, and any analysis under these conditions must take these effects into account [2].

Linear analysis is generally inadequate in both cases, and nonlinear analysis is essential to correctly model the behaviour of the structure. Various nonlinear models are used in practice and

these models usually involve some form of approximation of the nonlinear effects. Typically, this is done using either a traditional stiffness analysis combined with stability functions, or a finite element approach [3–11]. Both approaches can be combined with an iterative nodal updating scheme. However, it is not often clear what effects such approximations have on the analysis. Hence, it is important to study the effects of the approximations on the analysis, since significant differences in results are possible using different approximations [12].

This paper discusses a general approach for the nonlinear analysis of steel frames under thermal loading. In particular, we consider the role of approximations in the governing equations of the structure and the effects of the changes in material properties at high temperatures. We start with a fully nonlinear model for the member behaviour and then systematically consider the effects of simplifying approximations on the results of the analysis. The nonlinear and simplified models are combined with the changes in the material properties to explore the response of structures at high temperatures. Various examples are considered to demonstrate the effects of different assumptions and simplifications, and accurate solutions that can serve as benchmarks are given.

**Governing Equations**

*Nonlinear Beam Theory*

The behaviour of a beam is modelled using the extensible elastica theory [13], and shear deformation is ignored. The extension of the current formulation to include shear deformation is possible using the techniques discussed in our previous work [12].

Fig. 1 shows an element of a straight beam whose centroidal axis is initially along the $x$ axis, where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the $x$ and $y$ axes, respectively. After deformation, the axis deforms into a smooth curve $s(x)$ and the point $P$ is mapped to the point $P^*$. The beam displacements are $u(x)$ and $v(x)$, and the angle between $s$ and the $x$ axis at $P^*$ is $\phi(x)$.
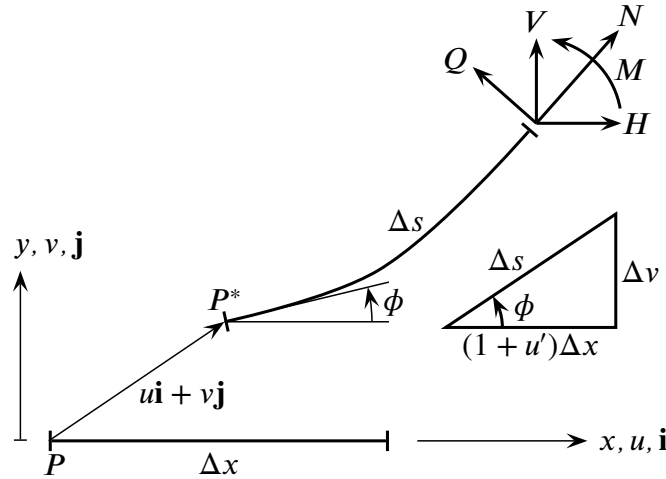


**Figure 1. Sign convention**

From the geometry of the deformation,

$$\cos \phi = (1 + u') \frac{dx}{ds} \tag{1a}$$

$$\sin \phi = \frac{dv}{ds} \tag{1b}$$

where a prime denotes differentiation with respect to $x$.

The strain measures for the deformation of the beam are the extensional strain, $e$, and the bending strain, $\kappa$, which are defined as

$$e = \frac{ds}{dx} - 1 = \sqrt{(1 + u')^2 + (v')^2} - 1 \tag{2a}$$

$$\kappa = \frac{d\phi}{dx} = \frac{(1 + u')v'' - v'u''}{(1 + u')^2 + (v')^2} \tag{2b}$$

The force resultants on the beam are the horizontal and vertical force components $H$ and $V$, and the bending moment $M$. The normal and shear forces are related to $H$ and $V$ by

$$N = H \cos \phi + V \sin \phi \tag{3a}$$

$$Q = -H \sin \phi + V \cos \phi \tag{3b}$$

The equilibrium equations are

$$H' + p_x = 0 \tag{4a}$$

$$V' + p_y = 0 \tag{4b}$$

$$M' - Hv' + V(1 + u') = 0 \tag{4c}$$

where $p_x$ and $p_y$ are the distributed loads acting on the beam in the $x$ and $y$ directions, respectively.

Assuming that the material is linear-elastic, the constitutive relationships are

$$N = EAe \tag{5a}$$

$$M = EI\kappa \tag{5b}$$

where $E$ is Young's modulus, $A$ is the cross-section area and $I$ is the second moment of area.

The governing equations are supplemented by appropriate boundary conditions for a particular problem. The theory results in a nonlinear system of ordinary differential equations of order six. Exact solutions are only possible in simple cases [13], and hence a simplified theory is often used.

The two commonly-used simplified nonlinear theories are the beam-column and beam-column with bowing theories [14]. These theories are based on ad hoc assumptions regarding the size of various nonlinear terms in the governing equations. If all the nonlinear effects are ignored, the theory reduces to the standard linear theory that is the basis of classical structural analysis. Table 1 compares the traditional simplified theories with the elastica theory.

*Thermal Effects*

We now consider the effects of temperature changes on the governing equations. For simplicity, we assume that any temperature changes are applied uniformly across the depth of the beam, and hence only additional axial strains occur. Hence, the axial thermal strain, $e_T$, is given by

$$e_T = \alpha \Delta T \tag{6}$$

**Table 1. Comparison of beam theories**

| Theory | Strain equations | Force equations | Equilibrium equations |
|---|---|---|---|
| Elastica | $e = \sqrt{(1+u')^2 + (v')^2} - 1$ <br> $\kappa = \dfrac{(1+u')v'' - v'u''}{(1+u')^2 + (v')^2}$ | $N = H\cos\phi + V\sin\phi$ <br> $Q = -H\sin\phi + V\cos\phi$ | $H' + p_x = 0$ <br> $V' + p_y = 0$ <br> $M' - Hv' + V(1+u') = 0$ |
| Beam-column with bowing | $e = u' + \frac{1}{2}(v')^2$ <br> $\kappa = v''$ | $N = H$ <br> $Q = -H\phi + V$ | $H' + p_x = 0$ <br> $V' + p_y = 0$ <br> $M' - Hv' + V = 0$ |
| Beam-column | $e = u'$ <br> $\kappa = v''$ | $N = H$ <br> $Q = -H\phi + V$ | $H' + p_x = 0$ <br> $V' + p_y = 0$ <br> $M' - Hv' + V = 0$ |
| Linear | $e = u'$ <br> $\kappa = v''$ | $N = H$ <br> $Q = V$ | $H' + p_x = 0$ <br> $V' + p_y = 0$ <br> $M' + V = 0$ |

where $\alpha$ is the coefficient of thermal expansion, which is assumed to be constant, and $\Delta T$ is the change in temperature from the reference temperature of the beam. If required, it is also possible to include the variation of $\alpha$ with temperature using an empirical formula [15].

The second effect is the change in the material properties at elevated temperatures. Several empirical formulas have been developed to account for this effect [15–17]. However, there are differences between the empirical formulas, and this may influence the results obtained from a given analysis depending on which formula is used. We adopt the empirical formulas given by Standards Australia [17] in this paper. Other empirical formulas can be readily incorporated into the general framework adopted in this paper if required.

According to [17], the change in Young's modulus with temperature can be approximated by

$$\frac{E(T)}{E(20)} = \begin{cases} 1 + \dfrac{T}{2000 \ln\left(\dfrac{T}{1100}\right)}, & \text{for } 0\,°\text{C} \leq T \leq 600\,°\text{C} \\[4mm] \dfrac{690\left(1 - \dfrac{T}{1000}\right)}{T - 53.5}, & \text{for } 600\,°\text{C} \leq T \leq 1000\,°\text{C} \end{cases} \tag{7}$$

where $T$ is the ambient temperature and $E(20)$ is the value of Young's modulus when $T = 20\,°\text{C}$.

The presence of the natural log function in Eq. (7) can potentially cause problems in a numerical solution if the temperature is close to zero or negative during an iteration. To avoid this issue, Eq. (7) was replaced by a fourth-order polynomial for the numerical results presented below. Accordingly, we approximate Eq. (7) with

$$\frac{E(T)}{E(20)} = \sum_{i=0}^{4} a_i T_s^i \tag{8}$$

where $T_s = T/1000$ and the $a_i$ coefficients are given in Table 2. The maximum difference between Eqs. (7) and (8) is approximately 3%, which is acceptable for numerical solutions. A higher-order polynomial function can be used if more accuracy is required.

**Table 2. Coefficients in Eq. (8)**

| | |
|---|---|
| $a_0$ | 1.0008372326824 |
| $a_1$ | -0.27234098307507 |
| $a_2$ | 0.87883333907696 |
| $a_3$ | -5.0888498704423 |
| $a_4$ | 3.5164356641168 |

Taking Eqs. (6) and (8) into account, the constitutive relationships from Eq. (5) are replaced with

$$N = E(T)A(e - e_T) \tag{9a}$$

$$M = E(T)I\kappa \tag{9b}$$

More complex constitutive relationships can be readily incorporated within the general framework of this paper using the techniques discussed in our previous work [18].

The last effect that needs to be considered at elevated temperatures is the change in the yield stress of the material, $f_y$. Again, we have used the empirical relationship from [17], namely

$$\frac{f_y(T)}{f_y(20)} = \begin{cases} 1, & \text{for } 0\,°C < T \le 215\,°C \\ \dfrac{905 - T}{690}, & \text{for } 215\,°C < T \le 905\,°C \end{cases} \tag{10}$$

where $f_y(20)$ is the value of the yield stress when $T = 20\,°C$.

**Formulation and Solution of Global Equations**

The formulation and solution of a typical problem follows the procedures used in our previous work [18]. This enables accurate results to be obtained under precisely defined conditions. In addition, the effects of different modelling assumptions can be easily assessed.

The solution procedure is briefly as follows.

1. Set up the structural data, including topology, member and joint characteristics, loads, supports and temperature changes. Initialise all unknown nodal displacements, which are the primary unknowns for the problem, to zero. Specify a tolerance for the analysis.

2. For each member, solve the system of differential equations and calculate the member end forces.

3. Establish the overall out of balance nodal forces for the structure from the applied loads and the member end forces.

4. Use an update strategy to adjust the nodal displacements based on the amount of out of balance forces.

5. Check for convergence using a suitable criterion relative to the specified tolerance. If convergence has been achieved, the analysis is complete. If not, go to step 2 and continue.

The solution procedure is augmented by a continuation technique, including an adaptive step algorithm [19], to efficiently and accurately generate the response of highly nonlinear structures.

Step 2 requires the use of a numerical boundary-value solver for a system of differential equations, and several reliable and robust methods are available [20]. We have used a collocation method in this paper [21]. This method incorporates an adaptive mesh method, and it can generate solutions with high accuracy if required. Moreover, it is not necessary to subdivide the members of the structure to achieve the specified accuracy, as this is done automatically by the software. Hence, this considerably reduces the data input for the structure.

Additional checks are required during the solution procedure if the aim is to establish the failure condition of the structure at elevated temperatures. In this case, we calculate the normal stress at the critical cross-sections of the structure. Assuming that the beam section is symmetrical about the axis of bending, the normal stress at any cross-section is

$$\sigma_n = \frac{|N|}{A} + \frac{|M|}{Z} \tag{11}$$

where $Z$ is the section modulus.

The structure is deemed to have failed when any such normal stress equals the current yield stress from Eq. (10). Hence, this is a first-yield check for failure and it results in a conservative estimate of the failure condition. If the full post-yield response is required, the current formulation can be extended using the methods discussed in our previous work [18].

**Examples**

Four examples are now considered, and accurate results for them are given to establish the effects of simplifying approximations and to enable the examples to be used as benchmarks. We define the examples in a precise way so that no uncertainties arise regarding either the problem or solution, following the guidelines proposed by us previously [22].

The examples are deliberately chosen as small structures to keep the data requirements simple. However, it is important to note that the numerical procedures used to solve the problems are not restricted in size and can solve structures of any practical size. Moreover, the examples illustrate all the nonlinear and simplifying effects we have discussed above, and hence serve our purposes.

The various results quoted below are identified by the subscripts *Lin* for linear theory, *BC* for beam-column theory, *Bow* for beam-column with bowing theory and *E* for extensible elastica theory.

Unless otherwise noted, the material properties are taken as

$$E(20) = 200\,\text{GPa}, \quad f_y(20) = 300\,\text{MPa}, \quad \alpha = 1.2 \times 10^{-5}/°\text{C} \tag{12}$$

and an initial temperature of $T_0 = 20\,°\text{C}$ is used.

*Simply supported beam*

Fig. 2 shows a simply supported beam under a small central load, $P = EI/10L^2$, and subject to a temperature change of $\Delta T$. The load simulates an imperfection of the beam from the ideal, straight-line configuration. The material properties are assumed to be constant with temperature for this example.
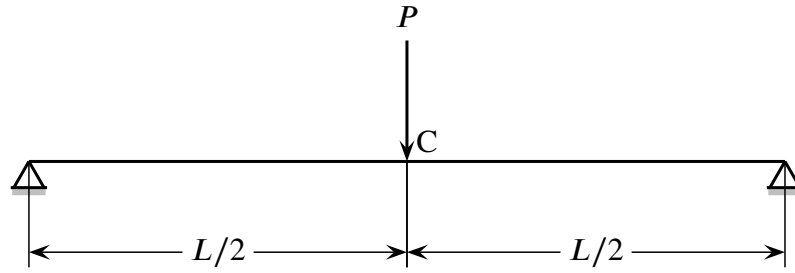


**Figure 2. Simply supported beam**

The beam buckles at a critical temperature change of [23]

$$\Delta T_b = \frac{\pi^2}{\alpha \lambda^2} \tag{13}$$

where $\lambda = L/r$ is the slenderness ratio of the beam and $r$ is the radius of gyration of the cross-section. Taking $\lambda = 150$, Eq. (13) gives $\Delta T_b \approx 36.55\,°\text{C}$. Values for the vertical displacement and bending moment at point C are given in Table 3. The results are quoted in terms of the non-dimensional quantities $\overline{v} = v/L$ and $\overline{M} = ML/EI$.

**Table 3. Vertical displacement and bending moment at point C for simply supported beam**

| $\Delta T$ (°C) | $1000\overline{v}_{Bc}$ | $1000\overline{v}_{Bow}$ | $1000\overline{v}_E$ | $100\overline{M}_{Bc}$ | $100\overline{M}_{Bow}$ | $100\overline{M}_E$ |
|---|---|---|---|---|---|---|
| 0 | 2.083 | 2.038 | 2.038 | 2.500 | 2.454 | 2.454 |
| 10 | 2.857 | 2.708 | 2.709 | 3.271 | 3.123 | 3.123 |
| 20 | 4.566 | 3.865 | 3.866 | 4.965 | 4.272 | 4.272 |
| 30 | 11.48 | 5.717 | 5.719 | 11.80 | 6.105 | 6.106 |
| 35 | 48.33 | 6.826 | 6.828 | 48.17 | 7.201 | 7.203 |

The linear theory results are

$$\overline{v}_{Lin} = \frac{1}{480} \approx 2.083 \times 10^{-3}, \quad \overline{M}_{Lin} = \frac{1}{40} = 2.5 \times 10^{-2} \tag{14}$$

and they are independent of $\Delta T$ for constant $E$. The results for the beam-column theory are the same as the linear theory results for $\Delta T = 0$, because the axial load is zero in this case. A non-zero axial load for the linear and beam-column theories requires $\Delta T > 0$. The beam-column

theory results are not the same as the linear theory results in this case, due to the coupling of the bending and axial modes of deformation. This coupling does not occur in the linear theory. However, the results for the beam-column theory become grossly inaccurate with increasing temperature change. In contrast, the beam-column with bowing and elastica theories correctly model the non-zero axial load due to the restrained horizontal displacement at the supports for all values of $\Delta T$. In addition, the results for the beam-column with bowing theory are acceptably close to the elastica theory results for the given range of $\Delta T$.

The temperature change versus displacement curves are shown in Fig. 3. The results for the beam-column theory asymptote to infinity as $\Delta T$ approaches $\Delta T_b$, and this theory cannot accurately represent the post-buckling response of the structure. Hence, the beam-column theory is only acceptable for modelling small temperature changes for this example. In contrast, the beam-column with bowing theory gives results that are acceptably close to the elastica results.
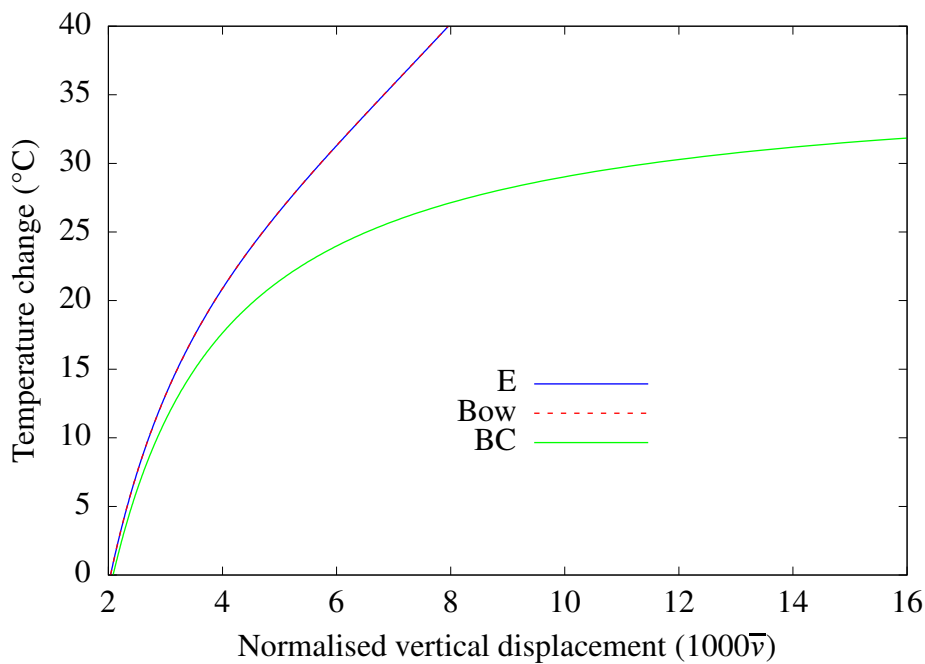


**Figure 3. Results for simply supported beam**

*Fixed-End Beam*

Fig. 4 shows a fixed-end beam subjected to a central load. A universal beam 150UB18 was chosen for the cross section. This selection is arbitrary, and any other section could have been used. The properties of the section are [24] $A = 2300 \, \text{mm}^2$, $I = 9.05 \times 10^6 \, \text{mm}^4$ and $Z = 117 \times 10^3 \, \text{mm}^3$.
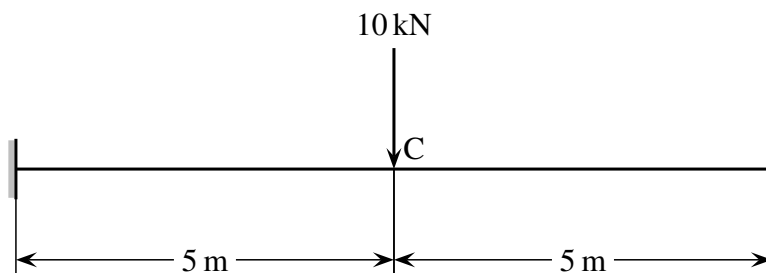


**Figure 4. Fixed-end beam**

Values for the vertical displacement and bending moment at point C are given in Table 4. The change in the vertical displacement for the linear theory is only due to the change in $E$ with increasing temperature. Hence, the linear theory becomes more inaccurate with increasing $\Delta T$, and underestimates the displacement by 35% at $\Delta T = 50\,°C$. The beam-column theory is more accurate, with a maximum error of 5.1%. The beam-column with bowing theory results are virtually identical to the elastica results, with a maximum error of 0.014%.

**Table 4. Vertical displacement (mm) and bending moment (kN m) at point C for fixed-end beam**

| $\Delta T$ (°C) | $v_{Lin}$ | $v_{Bc}$ | $v_{Bow}$ | $v_E$ | $M_{Lin}$ | $M_{Bc}$ | $M_{Bow}$ | $M_E$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 28.78 | 28.78 | 28.43 | 28.43 | 12.50 | 12.50 | 12.37 | 12.37 |
| 20 | 29.04 | 34.27 | 33.58 | 33.59 | 12.50 | 14.37 | 14.13 | 14.13 |
| 40 | 29.17 | 42.03 | 40.55 | 40.55 | 12.50 | 17.08 | 16.55 | 16.55 |
| 50 | 29.23 | 47.38 | 45.07 | 45.08 | 12.50 | 18.94 | 18.12 | 18.12 |

The results for the bending moment show a similar trend. For the linear theory, the central bending moment is constant for all $\Delta T$. The beam-column theory is more accurate, with a maximum error of 4.5%. Once again, the beam-column with bowing theory results are virtually identical to the elastica results, with a maximum error of 0.01%.

Table 5 gives the results for the failure condition of the beam. The linear theory results are grossly inaccurate, with an error of 28% for the failure temperature and a error of 42% for the failure displacement. The beam-column theory has an error of 6.5% for the failure temperature and an error of 0.17% for the failure displacement. The beam-column with bowing theory provides an accurate estimate of the failure conditions, with negligible errors.

**Table 5. Results at failure for fixed-end beam**

| Theory | Temperature (°C) | $v_C$ (mm) | $\sigma_C$ (MPa) |
|---|---|---|---|
| Linear | 102.4 | 29.45 | 300 |
| Beam-column | 74.7 | 50.40 | 300 |
| Beam-column with bowing | 79.9 | 50.49 | 300 |
| Elastica | 79.9 | 50.49 | 300 |

*Experimental Frame EGR*

Fig. 5 shows an experimental frame that was tested to failure [25]. This frame has also been used as a test problem for numerical approximation methods [3, 7, 10, 25]. All members had an IPE80 cross section, with properties $A = 764\,mm^2$, $I = 8.014 \times 10^5\,mm^4$ and $Z = 2.003 \times 10^4\,mm^3$. The material properties were $E = 210\,GPa$, $f_y = 382\,MPa$ and $\alpha = 1.4 \times 10^{-5}/°C$, and an initial temperature of $T_0 = 20\,°C$ was used.

Values for the horizontal displacement and bending moment at point A are given in Table 6. The nonlinear effects are significant for this structure for all values of $\Delta T$. The linear theory underestimates both the displacement and bending moment for all values of $\Delta T$, with a maximum error of 35% for the given range of $\Delta T$. The maximum errors for the beam column and
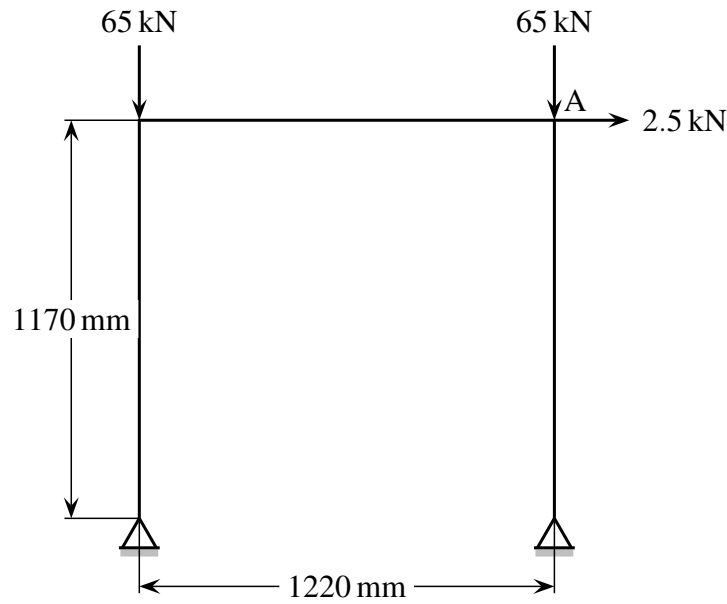
**Figure 5. Experimental frame EGR**

**Table 6. Horizontal displacement (mm) and bending moment (kN m) at point A for experimental frame**

| $\Delta T$ (°C) | $u_{Lin}$ | $u_{Bc}$ | $u_{Bow}$ | $u_E$ | $M_{Lin}$ | $M_{Bc}$ | $M_{Bow}$ | $M_E$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 6.071 | 8.588 | 8.588 | 8.579 | 1.463 | 2.020 | 2.020 | 2.018 |
| 100 | 7.095 | 9.805 | 9.823 | 9.844 | 1.583 | 2.164 | 2.165 | 2.168 |
| 200 | 8.183 | 11.17 | 11.21 | 11.26 | 1.693 | 2.309 | 2.311 | 2.318 |
| 300 | 9.509 | 13.09 | 13.15 | 13.25 | 1.785 | 2.471 | 2.475 | 2.486 |
| 400 | 11.26 | 16.17 | 16.28 | 16.45 | 1.843 | 2.675 | 2.683 | 2.699 |
| 450 | 12.39 | 18.57 | 18.73 | 18.96 | 1.856 | 2.817 | 2.827 | 2.848 |

beam column with bowing theories are 2.1% and 1.2%, respectively. Hence, these theories are sufficiently accurate for this range of $\Delta T$.

Table 7 gives the results for the failure condition of the frame. The linear theory overestimates the failure temperature, with an error of 21%. The errors for the beam column and beam column with bowing theories are 0.48% and 0.31%, respectively. The average experimental failure temperature for this frame was reported as 524 °C [25]. As expected, the elastica theory underestimates this temperature by 8.2%, since the present analysis stops at the first-yield point and hence provides a conservative estimate of the failure condition.

**Table 7. Results at failure for experimental frame**

| Theory | Temperature (°C) | $u_A$ (mm) | $\sigma_A$ (MPa) |
|---|---|---|---|
| Linear | 580.0 | 15.95 | 179.9 |
| Beam-column | 483.5 | 19.38 | 233.4 |
| Beam-column with bowing | 482.7 | 19.51 | 233.8 |
| Elastica | 481.2 | 19.66 | 234.6 |

The curves for temperature versus the horizontal displacement and normal stress at point A up to the failure point are shown in Figs. 6 and 7, and confirm the increasing error with temperature for the linear theory. Fig. 7 shows how the first-yield condition is reached when the normal stress equals the yield stress for the current temperature.
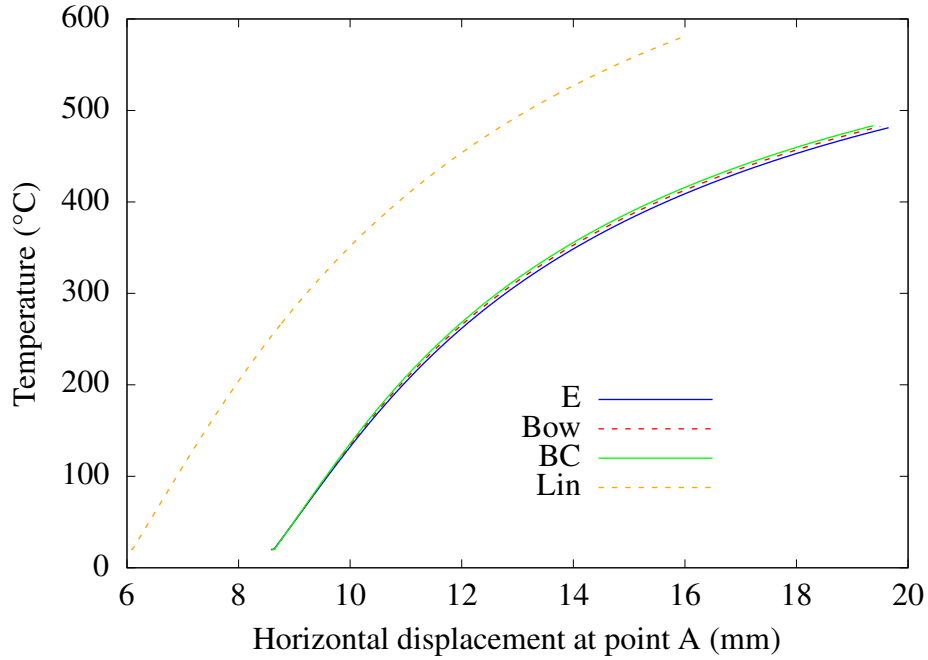
**Figure 6. Displacement results for experimental frame**
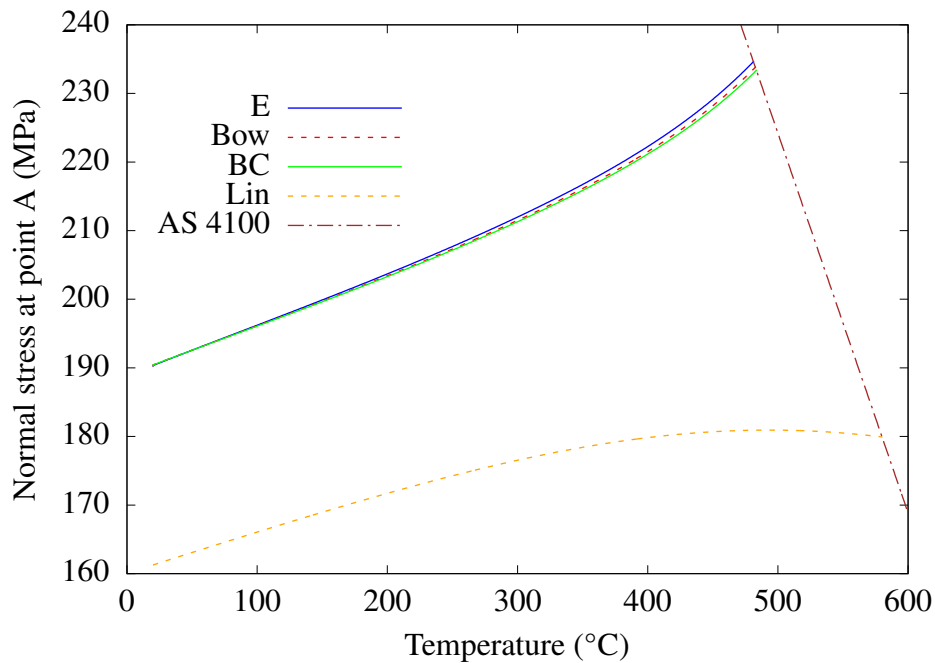
**Figure 7. Normal stress results for experimental frame**

*Portal Frame*

As a final example, we consider the portal frame shown in Fig. 8. This frame has been adapted from the design in [26], and models one bay of an industrial shed, similar to those that are

commonly constructed on rural properties.
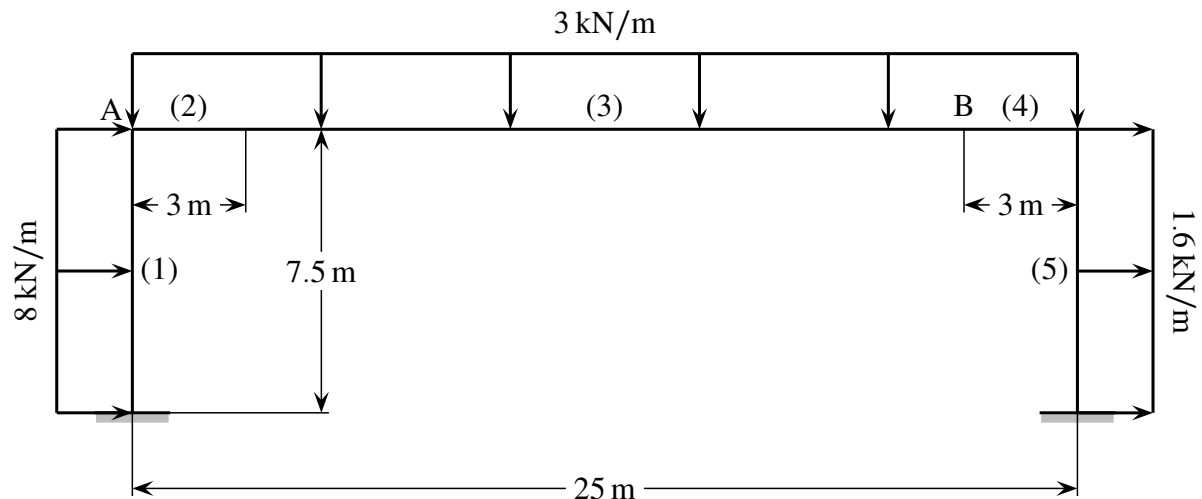


**Figure 8. Portal frame**

The loading on the frame represents an extreme combination of dead and wind loads, which attempts to simulate the loads that would occur on the structure during a catastrophic bushfire, as for example the Victorian (Australia) bushfires in 2009 [27]. The sections chosen for the design are given in Table 8. Although a practical design would have haunches (members 2 and 4) with variable properties, we have used constant-size haunches for simplicity.

**Table 8. Properties for portal frame**

| Member | Section | $A$ (mm$^2$) | $I$ ($10^6$ mm$^4$) | $Z$ ($10^3$ mm$^3$) |
|--------|---------|-------------|---------------------|---------------------|
| 1, 5 | 610UB125 | 16000 | 986 | 3230 |
| 2, 4 | 530UB92.4 | 11800 | 554 | 2080 |
| 3 | 360UB50.7 | 6470 | 142 | 798 |

Table 9 gives the results for the failure condition of the frame. The linear theory overestimates the failure temperature, with an error of 5.3%. In contrast, the errors for the beam column and beam column with bowing theories are less than 1%. The errors for the failure stress are low, with a maximum error of 2.9% for the linear theory and negligible errors for the beam column and beam column with bowing theories. The linear theory significantly underestimates the failure displacement, with an error of 10.7%. The errors for the beam column and beam column with bowing theories are more acceptable and are 2.4% and 1.1%, respectively. Hence, these theories provide an acceptable estimate of the failure conditions.

**Table 9. Results at failure for portal frame**

| Theory | Temperature (°C) | $u_A$ (mm) | $\sigma_B$ (MPa) |
|--------|------------------|-----------|------------------|
| Linear | 339.1 | 72.18 | 246.0 |
| Beam-column | 323.3 | 78.88 | 252.9 |
| Beam-column with bowing | 323.8 | 79.93 | 252.7 |
| Elastica | 322.1 | 80.84 | 253.4 |

## Conclusions

The paper has discussed the nonlinear analysis of steel frames under thermal loads, including fire conditions. Exact and approximate models for the members were discussed, together with their assumptions and limitations. The effects of different assumptions were demonstrated using several examples. Accurate results were presented for the examples, and these results can serve as benchmarks for this class of problems.

## References

[1] Li, D., Hyslip, J., Sussmann, T. and Chrismer, S. (2016) *Railway Geotechnics*, CRC Press, Boca Raton.

[2] Kodur, V. and Naser, M. (2020) *Structural Fire Engineering*, McGraw-Hill, New York.

[3] Kassimali, A. and Garcilazo, J. (2010) Geometrically nonlinear analysis of plane frames subjected to temperature changes, *Journal of Structural Engineering, ASCE*, **136**, 1342–1349.

[4] Lien, K., Chiou, Y., Wang, R. and Hsiao, P. (2010) Vector form intrinsic finite element analysis of nonlinear behavior of steel structures exposed to fire, *Engineering Structures*, **32**, 80–92.

[5] Jeffers, A. and Sotelino, E. (2012) Analysis of steel structures in fire with force-based frame elements, *Journal of Structural Fire Engineering*, **3**, 287–300.

[6] Luongo, A. and Contento, A. (2015) Nonlinear elastic analysis of steel planar frames under fire loads, *Computers and Structures*, **150**, 23–33.

[7] Iu, C. (2016) Nonlinear fire analysis of steel structure using equivalent thermal load procedure for thermal geometrical change, *Fire Safety Journal*, **86**, 106–119.

[8] Franssen, J. and Gernay, T. (2017) Modeling structures in fire with SAFIR (R): theoretical background and capabilities, *Journal of Structural Fire Engineering*, **8**, 300–323.

[9] Srivastava, G. and Prakash, P. (2017) An integrated framework for nonlinear analysis of plane frames exposed to fire using the direct stiffness method, *Computers and Structures*, **190**, 173–185.

[10] Prakash, P. and Srivastava, G. (2019) Distributed plasticity model for analysis of steel structures subjected to fire using the direct stiffness method, *Fire Safety Journal*, **105**, 169–187.

[11] Gernay, T. and Franssen, J. (2020) The introduction and the influence of semi-rigid connections in framed structures subjected to fire, *Fire Safety Journal*, **114**, 103007.

[12] Petrolito, J. and Legge, K. (2006) Nonlinear analysis of frames including shear deformation, *International Journal of Structural Stability and Dynamics*, **6**, 179–192.

[13] Antman, S. (2005) *Nonlinear Problems of Elasticity*, second edn., Springer-Verlag, New York.

[14] Meek, J. (1991) *Computer Methods in Structural Analysis*, Chapman Hall, London.

[15] CEN (2005) *EN 1993-1-2: Eurocode 3, Design Of Steel Structures, Part 1.2: Structural Fire Design*, European Committee for Standardization, Brussels.

[16] AISC (2016) *Specification for Structural Steel Buildings*, American Institute of Steel Construction, Chicago.

[17] Standards Australia (2020) *AS 4100: Steel Structures*, Standards Australia, Sydney.

[18] Petrolito, J. and Legge, K. (2012) Benchmarks for elasto-plastic analysis of steel frames, *Australian Journal of Structural Engineering*, **12**, 237–250.

[19] Rheinboldt, W. (1986) *Numerical Analysis of Parametrized Nonlinear Equations*, Wiley, New York.

[20] Petrolito, J. and Legge, K. (2005) Solution of boundary value problems in structural analysis, *Proceedings, Australian Structural Engineering Conference*, 1–9, Newcastle.

[21] Ascher, U., Mattheij, R. and Russell, R. (1995) *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Society for Industrial and Applied Mathematics, Philadelphia.

[22] Petrolito, J. and Legge, K. (2001) Formulating benchmarks for nonlinear analysis, *Proceedings, NAFEMS World Congress*, vol. 2, 993–1004, Lake Como.

[23] Timoshenko, S. and Gere, J. (1961) *Theory of Elastic Stability*, second edn., McGraw-Hill, New York.

[24] Onesteel (2014) *Hot Rolled and Structural Steel Products*, seventh edn., Onesteel Manufacturing, Melbourne.

[25] Rubert, A. and Schaumann, P. (1986) Structural steel and plane frame assemblies under fire action, *Fire Safety Journal*, **10**, 173–184.

[26] Woolcock, S., Kitipornchai, S. and Bradford, M. (1999) *Design of Portal Frame Buildings*, third edn., Australian Institute of Steel Construction, Sydney.

[27] Teague, B., McLeod, R. and Pascoe, S. (2010) *Final Report of 2009 Victorian Bushfires Royal Commission*, Government Printer for the State of Victoria, Melbourne.

# A method for Helmholtz operator and application to free vibration problem of thin plate on Pasternak foundation

**Jiarong Gan, †Hong Yuan, \*†Shanqing Li**

MOE Key Laboratory of Disaster Forecast and Control in Engineering, School of Mechanics and Construction
Engineering, Jinan University, Guangzhou 510632, China

\*Presenting author: lishanqing09@163.com

†Corresponding author: tyuanhong@jnu.edu.cn and lishanqing09@163.com

**Abstract**

A computing method is proposed to solve the two-dimensional Helmholtz operator, and the method is used to solve the vibration problem of thin plates on Pasternak foundation with complex boundary shape conditions. By using the shape function obtained by R-function, the Bessel function, the basic solution and boundary conditions of the two-dimensional Helmholtz operator, a quasi-Green function satisfying the homogeneous boundary conditions is constructed. The integral kernels are obtained by using the properties of Bessel function, and then the two-dimensional Helmholtz operator is transformed into integral equation by Green's formula. The appropriate boundary equation is selected by using R-function theory to eliminate the singularity, in which Bessel function is transformed into corresponding numerical expression. The method is applied to analyze the free vibration problem of thin plates on Pasternak foundation. The numerical examples show that the method is an effective numerical method.

**Keywords:** Two-dimensional Helmholtz operator; R-function; Bessel function; Quasi-Green function; Vibration problem

## Introduction

Helmholtz equation is widely used in many fields such as engineering technology, electromagnetic field theory, scattering theory, mechanics and so on. The study of its numerical solution has not only extensive practical significance but also important theoretical value. Traditional numerical methods, such as finite difference method, finite element method and boundary element method, need to establish network related to interpolation node topology.

Helmholtz operator and Pasternak foundation problem are extensively researched. Romero et al. [1] proposed a novel boundary element approach formulated on the Bézier-Bernstein basis to yield a geometry-independent field approximation. And the capability of the proposed method was demonstrated for solving the Helmholtz equation in longitudinally invariant problems. Jones et al. [2] proposed a novel dual-scale compact method for solving non paraxial Helmholtz equations at high wave numbers. Corbo et al. [3] proposed a hybridized

continuous/discontinuous Galerkin formulation for the Helmholtz problem that works with a continuous trace space. Li et al. [4] proposed the geometricmulti-grid solution to solve the grapes Helmholtz equation. Yu Du and Zhimin Zhang [5] researched super closeness of linear discontinuous Galerkinfinite element method and its super convergence based on the polynomial preserving recovery for Helmholtz equation. Galkowski et al. [6] considered solving the exterior Dirichlet problem for the Helmholtz equation with the h-version of the boundary element method using the standard second-kind combined-field integral equations. Hoshoudy et al. [7] investigated the combined effect of density, velocity and magnetic field gradients on the Kelvin–Helmholtz instability of two viscous fluid layers. Lin et al. [8] carried out the explorations for solving interface problems of the Helmholtz equation by immersed finite elements on interface independent meshes. Mandel et al. [9] proved an L p-version of the limiting absorption principle for a class of periodic elliptic differential operators of second order. And the result was applied to the construction of nontrivial solutions of nonlinear Helmholtz equations with periodic coefficient functions. Stein et al. [10] studied an acoustic model of a Helmholtz resonator under a grazing turbulent boundary layer. Shahbaztabar et al. [11] investigated the free vibration problem of functionally graded circular cylindrical shells resting on Pasternak elastic supports and fully or partially in contact with an ideal and irrotational fluid. Nebojša Radić [12] researched the buckling of porous double-layered functionally graded nanoplates in the Pasternak elastic foundation based on nonlocal strain gradient elasticity. Mohanty et al. [13] carried out the stability analysis of a two-layer elastic beam resting on a variable Pasternak foundation subjected to an axial pulsating load and thermal gradient. Wei et al. [14] carried out analysis of the impact on adjacent underground pipelines due to foundation pit excavation on the Pasternak foundation. Lin and Huang [15] studied the deflection of discontinuous underground pipelines excavated by shield tunnels based on Pasternak foundation. Ahmed et al. [16] studied quasi 3-D trigonometric plate theory for bending analysis of exponentially graded plates resting on Pasternak foundations. Chen et al. [17] applied energy method as solution for deformation of geosynthetic reinforced embankment on Pasternak foundation. Ghafarian et al. [18] carried out the free and forced vibration analysis of a Timoshenko beam on viscoelastic Pasternak foundation featuring coupling between flap wise bending and torsional vibrations. Zhang et al. [19] studied the sound-vibration behaviors of the thin orthotropic rectangular fluid-structure coupled system resting on varying elastic Winkler and Pasternak foundations. Mohanty et al. [20] investigated the free vibration characteristics of a two-layer elastic beam resting on a variable Pasternak foundation. Arani et al. [21] carried out the free vibration analysis of sandwich composite micro-beam subjected to multi physical fields using the sinusoidal shear deformation beam theory and modified strain gradient theory. Zhao et al. [22] carried out the inner-force calculation of bridge pile foundation in a high-steep slope based on the Pasternak double-parameter spring model. Chen et al. [23] solved two-dimensional Helmholtz equation indirectly by interpolated boundary element-free method. Dai et al. [24] solved the transmission eigenvalue problem of Helmholtz equation by spectral element method.

The R-function, the Bessel function, the basic solution and boundary condition are used to transform the two-dimensional Helmholtz operator into integral equation. The free vibration of an elastic arbitrary polygonal thin plate on a two-parameter foundation model with simply

supported periphery is analyzed, and the corresponding mode shapes are calculated by using the frequencies obtained. The developed method can be used Helmholtz operator with complicated shape.

## Solution of Two-Dimensional Helmholtz Operator

### *Basic equations of Two-Dimensional Helmholtz Operator*

Considered Two-dimensional Helmholtz operator's dirichlet problem

$$\frac{\partial^2 v}{\partial \mathrm{x}_1^2} + \frac{\partial^2 v}{\partial \mathrm{x}_2^2} - \lambda v = f_0 \tag{1}$$

Make its boundary condition become $\qquad v\mid_\Gamma = \varphi_0$ (2)

Where $\lambda$ is constant, $\Gamma = \partial\Omega$ is boundary of $\Omega$'s borders, extend the function $\varphi_0$ to $\Omega$, establish a function $\varphi$ that is smooth enough in $\Omega$, make $\varphi\mid_\Gamma = \varphi_0$. Therefore, using variable substitution $u = v - \varphi$, and substituting expressions Eq. (1) and Eq. (2), the original problem can be transformed into a boundary value problem with homogeneous boundary conditions as follow.

$$\nabla^2 u - \lambda u = f \tag{3}$$

$$u\mid_\Gamma = 0 \tag{4}$$

Where $f = f_0 - (\nabla^2 \varphi - \lambda \varphi)$.

### *Quasi-Green functions of two-dimensional Helmholtz operator*

Let $w = 0$ be the first-order normalized equation of boundary $\Gamma$ [25-26]. Satisfy the following expression.

$$\omega(\mathbf{x}) = 0 \quad |\nabla\omega| = 1 \quad \mathbf{x} \in \Gamma \tag{5}$$

$$\omega(\mathbf{x}) > 0 \qquad \mathbf{x} \in \Omega \tag{6}$$

Using the basic solution of the problem, the quasi-Green function is constructed as follow.

$$G(\mathbf{x}, \xi) = k_0(\sqrt{\lambda}r) - e(\mathbf{x}, \xi) \tag{7}$$

$$e(\mathbf{x}, \xi) = k_0(\sqrt{\lambda}R) \tag{8}$$

Where $k_0$ is the second kind of zero-order deformed Bessel function. And

$$r = \| \boldsymbol{\xi} - \mathbf{x} \| = \sqrt{(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2} \tag{9}$$

$$\vec{r} = (\xi_1 - x_1)\vec{i} + (\xi_2 - x_2)\vec{j} \tag{10}$$

$$R = \sqrt{r^2 + 4\omega(\boldsymbol{\xi})\omega(\mathbf{x})} \tag{11}$$

In the formula, $\vec{i}$, $\vec{j}$ represent the unit vectors in the direction $x_1$, $x_2$, respectively.

Where $\mathbf{x} = (x_1, x_2)$, $\boldsymbol{\xi} = (\xi_1, \xi_2)$. Obviously quasi-Green's function $G(\mathbf{x}, \boldsymbol{\xi})$ satisfies the condition

$$G(\mathbf{x}, \boldsymbol{\xi})|_{\xi \in \Gamma} = 0 \tag{12}$$

### *Integral equation*

The boundary value problem Eq. (3) and Eq. (4) are transformed into integral equation, and Green's formula of class $C^2(\Omega)$ function is applied. For all $U$, $V \in (\Omega \bigcup \Gamma)$, there are the following expression.

$$\int_{\Omega} [V(\nabla^2 U - \lambda U) - U(\nabla^2 V - \lambda V)] d_{\xi} \Omega = \int_{\partial \Omega} (V \frac{\partial U}{\partial n} - U \frac{\partial V}{\partial n}) d_{\xi} \Gamma \tag{13}$$

The $u$ and quasi-Green's function $G$ in formula (3) are used to replace $U$ and $V$ in formula (13), respectively. Noting that $-\frac{1}{2\pi} k_0(\sqrt{\lambda} r)$ is the fundamental solution of Helmholtz operator $\nabla^2 u - \lambda u$, using formulas (4) and (12), obtain the following expression

$$u(\mathbf{x}) = u_0(\mathbf{x}) + \int_{\Omega} u(\boldsymbol{\xi}) K(\mathbf{x}, \boldsymbol{\xi}) d_{\xi} \Omega \tag{14}$$

Where

$$u_0(\mathbf{x}) = -\frac{1}{2\pi} \int_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) f(\boldsymbol{\xi}) d_{\xi} \Omega \tag{15}$$

$$K(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{2\pi} (\nabla_{\xi}^2 - \lambda) e(\mathbf{x}, \boldsymbol{\xi}) \tag{16}$$

The expression (8) is substituted to the expression (16). By using the properties of Bessel function, we can get

$$K(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{2\pi} \sqrt{\lambda} k_1(\sqrt{\lambda} R)[-\nabla^2 R + \frac{1}{R}(\nabla R)^2]$$
$$-\frac{1}{2\pi} \lambda k_0(\sqrt{\lambda} R)[(\nabla R)^2 - 1] \tag{17}$$

According to the relationship between $R$ and $\omega$ of formula (10), formula (17) can eventually be transformed to the following form.

$$K(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{2\pi} \sqrt{\lambda} k_1(\sqrt{\lambda} R)\{-\frac{2}{R}[1 + \omega(\mathbf{x})\nabla^2 \omega] + \frac{2}{R^3}[\vec{r} + 2\omega(\mathbf{x})\nabla \omega]^2\}$$

$$-\frac{1}{2\pi}\lambda k_0(\sqrt{\lambda}R)\{\frac{1}{R^2}[\vec{r}+2\omega(\mathbf{x})\nabla\omega]^2-1\} \tag{18}$$

In which, $\omega=\omega(\xi)$, $\nabla=\nabla_\xi$.

***Elimination of kernel singularity***

When $R=0$, namely $\mathbf{x}=\xi$ and $\omega=0$, discontinuity may occur in $K(\mathbf{x},\xi)$ in expression

(18). In fact, when $\mathbf{x}=\xi$, equation (18) can be transformed to the following expression.

$$K(\mathbf{x},\xi)|_{\mathbf{x}=\xi}=-\frac{1}{2\pi}\sqrt{\lambda}k_1(2\omega\sqrt{\lambda})\frac{1}{\omega}[(\nabla\omega)^2-1-\omega\nabla^2\omega]$$

$$-\frac{1}{2\pi}\sqrt{\lambda}k_0(2\omega\sqrt{\lambda})[(\nabla\omega)^2-1] \tag{19a}$$

In order to make integral kernels $K(\mathbf{x},\xi)\in C(\Omega\cup\Gamma)$, hypothesis

$$\omega=\omega_0+\omega_0^2\phi \tag{19b}$$

The formula $\omega_0=0$ is the first order normalized equation of $\Gamma=\partial\Omega$, which satisfies Eq. (5)

and Eq. (6). Obviously, $\omega_0$ and $\omega$ are the first-order normalized equations of boundary $\Gamma$.

According to formulas (5) and (6), equation $\omega_0=0$, we can get

$$(\nabla\omega_0)^2=1+\omega_0\phi_0 \quad \phi_0\in C^0(\Omega\cup\partial\Omega)$$

It can be obtained from the upper formula.

$$\phi_0=[(\nabla\omega_0)^2-1]\omega_0^{-1} \tag{20}$$

By expanding the deformed Bessel function, we can get

$$k_1(2\omega\sqrt{\lambda})\approx\frac{1}{2\omega\sqrt{\lambda}} \quad (\omega\to 0)$$

$$k_0(2\omega\sqrt{\lambda})\approx\ln\frac{1}{\omega\sqrt{\lambda}} \tag{21}$$

Assuming the function $\omega$ of form (19b), we can obtain the power series of $\nabla\omega$, $(\nabla\omega)^2$ and

$\nabla^2\omega$ expressed by $\omega_0$. By simplifying the equation (20), we can get

$$\nabla\omega=\nabla\omega_0+2\omega_0\nabla\omega_0\phi+\omega_0^2\nabla\phi$$

$$(\nabla\omega)^2=1+\omega_0\phi_0+4\omega_0\phi+O(\omega_0^2)$$

$$\nabla^2 \omega = \nabla^2 \omega_0 + 2\phi + O(\omega_0) \tag{22}$$

Substituting formulas (20) and (21) into formula (22), In order to ensure the continuity of kernel function $K(\mathbf{x}, \xi)$ in the integral domain, only $\phi = \frac{1}{2}(\nabla^2 \omega_0 - \phi_0)$ is needed.

Substitute the expression of $\phi$ into formula (19b), we can get

$$\omega = \omega_0 + \frac{1}{2}\omega_0^2(\nabla^2\omega - \phi_0) = \omega_0 + \frac{1}{2}\omega_0^2[\nabla^2\omega_0 + \frac{1-(\nabla\omega_0)^2}{\omega_0}] \tag{23}$$

**Application**
*Basic equation*

The mode differential equation for free vibration of thin plates on Pasternak foundation is

$$D\nabla^4 W(\mathbf{x}) - G_p\nabla^2 W(\mathbf{x}) + kW(\mathbf{x}) = \overline{m}\overline{\omega}^2 W(\mathbf{x}) \quad \mathbf{x} \in \Omega \tag{24}$$

in which, $\nabla^4 = (\frac{\partial^2}{\partial x_1} + \frac{\partial^2}{\partial x_2})^2$ is the biharmonic operator, the bending stiffness is

$D = \frac{Eh^3}{12(1-v^2)}$, $h$ is the thickness of the plate, $E$ and $v$ are elastic modulus and Poisson's

ratio, respectively, $k$ is the coefficient of elastic foundation, $G_p$ is the shear modulus of

foundation, $\overline{\omega}$ is the natural frequency, $\overline{m}$ is the mass per unit area, $W(\mathbf{x})$ represents the

mode function, $\mathbf{x} = (x_1, x_2)$, $\Omega$ is the domain contained in the middle plane ( $x_1 - x_2$ plane)

of the plate in Cartesian coordinate system.

Introducing intermediate variables

$$M = \frac{1}{1+v}(M_1 + M_2) \tag{25}$$

In the formula, the expression of $M_1$ and $M_2$ is

$$M_1 = -D(\frac{\partial^2 W}{\partial x_1^2} + v\frac{\partial^2 W}{\partial x_2^2}) \quad M_2 = -D(\frac{\partial^2 W}{\partial x_2^2} + v\frac{\partial^2 W}{\partial x_1^2}) \tag{26}$$

By substituting equation (25) into differential equation (24), the mode control differential equation for free vibration of thin plates on Pasternak foundation can be decomposed into two coupled lower order differential equations.

$$\nabla^2 M - \frac{G_p}{D}M = -\lambda_1^2 W \quad \mathbf{x} \in \Omega \tag{27}$$

$$\nabla^2 W = -\frac{M}{D} \quad \mathbf{x} \in \Omega \tag{28}$$

In which, $\nabla^2 = \dfrac{\partial^2}{\partial x_1} + \dfrac{\partial^2}{\partial x_2}$ is a Laplace operator, $\lambda_1^2 = \overline{m\omega}^2 - k$, engineering

frequency $f = \dfrac{\overline{\omega}}{2\pi}$.

### Boundary conditions

The deflection and bending moment on the plate boundary should be zero under simply supported constraint boundary conditions.

$$W = 0 \quad \mathbf{x} \in \Gamma \tag{29}$$

$$M = 0 \quad \mathbf{x} \in \Gamma \tag{30}$$

### The Integral equations

Using the Solution of Two-Dimensional Helmholtz Operator, the boundary value problems Eq. (27) and Eq. (30) into integral equation. Substitution of $U$ and $V$ in Eq. (13) by $M$ and quasi-Green's function $G$, noting that $-\dfrac{1}{2\pi} k_0(\sqrt{\lambda}r)$ is the fundamental solution of Helmholtz operator. The following formulas can be obtained by using (12), (27) and (30).

$$M(\mathbf{x}) = M_0(\mathbf{x}) + \int_\Omega M(\xi) K_M(\mathbf{x},\xi) d_\xi \Omega \tag{31}$$

$$M_0(\mathbf{x}) = \frac{\lambda_1^2}{2\pi} \int_\Omega G_M(\mathbf{x},\xi) W(\xi) d_\xi \Omega \tag{32}$$

$$K_M(\mathbf{x},\xi) = -\frac{1}{2\pi}(\nabla_\xi^2 - \lambda) e_M(\mathbf{x},\xi) \tag{33}$$

Similarly, substitution of $U$ and $V$ in formula (34) by $W$ and quasi-Green's function $G$ of Laplace operator in references [27-31], noting that $\dfrac{1}{2\pi} \ln r$ is the fundamental solution of Laplace operator. The following formulas (35) can be obtained by using (34), (28) and (29).

$$\int_\Omega (V\nabla^2 U - U\nabla^2 V) d_\xi \Omega = \int_{\partial\Omega} (V\frac{\partial U}{\partial n} - U\frac{\partial V}{\partial n}) d_\xi \Gamma \tag{34}$$

$$W(\mathbf{x}) = W_0(\mathbf{x}) + \int_\Omega W(\xi) K_W(\mathbf{x},\xi) d_\xi \Omega \tag{35}$$

$$W_0(\mathbf{x}) = -\frac{1}{D} \int_\Omega G_W(\mathbf{x},\xi) M(\xi) d_\xi \Omega \tag{36}$$

$$K_W(\mathbf{x},\xi) = (\frac{\partial^2}{\partial\xi_1^2} + \frac{\partial^2}{\partial\xi_2^2}) e_W(\mathbf{x},\xi) \tag{37}$$

According to Eq. (18) and the corresponding formulas in reference [27-31], we can get

$$K_M(\mathbf{x},\xi) = -\frac{1}{2\pi}\sqrt{\lambda}k_1(\sqrt{\lambda}R)\{-\frac{2}{R}[1 + \omega(\mathbf{x})\nabla^2\omega] + \frac{2}{R^3}[\vec{r} + 2\omega(\mathbf{x})\nabla\omega]^2\}$$

$$-\frac{1}{2\pi}\lambda k_0(\sqrt{\lambda}R)\{\frac{1}{R^2}[\vec{r}+2\omega(\mathbf{x})\nabla\omega]^2-1\} \tag{38}$$

$$K_W(\mathbf{x},\xi)=\frac{R^2\omega(\mathbf{x})\nabla^2\omega+4\omega(\mathbf{x})\omega-4(\vec{r}\cdot\nabla\omega)\omega(\mathbf{x})-4\omega^2(\mathbf{x})(\nabla\omega)^2}{\pi R^4} \tag{39}$$

In the formulas, $k_1$ is the second kind of first-order deformed Bessel function, $\omega=\omega(\xi)$, $\nabla=\nabla_\xi$. The singularity of integral kernels is treated by references [27-31], which guarantees the continuity of integral kernels $K_M(\mathbf{x},\xi)$ and $K_W(\mathbf{x},\xi)$.

The integral equations (31) and (35) are discretized by the method of references [27-31]. The natural frequencies and corresponding modes can be obtained by programming of MATLAB.

## *Numerical examples*

**Example 1** Simply supported polygonal thin plate on Pasternak foundation is shown in Figure 1. We set $a=b=1$, Poisson ratio $v=0.3$, thickness $t=0.1$, Modulus of elasticity $E=3\times10^9$, the mass per unit area $\bar{m}=780$, Coefficient of elastic foundation $k=2\times10^7$, the shear modulus of foundation $G_P=2\times10^6$. According to R-function theory [25-26]

$$\omega_0=\omega_1+\omega_2-\sqrt{\omega_1^2+\omega_2^2}$$

in which $\omega_1=\frac{1}{2b}(b^2-y^2)$ and $\omega_2=\frac{1}{2a}(a^2-x^2)$. Then $\omega_0=0$ is the first-order boundary normalized equation of the square plate. $\omega_1=\frac{1}{2b}(b^2-y^2)$ and $\omega_2=\frac{1}{2a}(a^2-x^2)$ are the sides of a simply supported plate.



**Figure 1. Simply supported polygonal plate**

25(5×5), 49(7×7), 81(9×9) and 121(11×11) grid layout schemes are adopted for square plates, respectively. The results are shown in Table 1.

**Table 1．Natural frequencies *f* of square plate**

| Method | | Modal order | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | |
| | | Frequency | Error(%) | Frequency | Error(%) | Frequency | Error(%) |
| The method in this paper | 25 | 35.621 | 3.378 | 57.561 | 8.612 | 82.502 | 12.191 |
| | 49 | 34.997 | 1.567 | 55.378 | 4.493 | 78.772 | 7.119 |
| | 81 | 34.762 | 0.885 | 54.406 | 2.659 | 76.753 | 4.373 |
| | 121 | 34.651 | 0.563 | 53.916 | 1.73 | 75.670 | 2.901 |
| Analytic solutions | | 34.457 | --- | 52.997 | --- | 73.537 | --- |

The modes corresponding to each order of frequencies are shown in Figures 2-4.
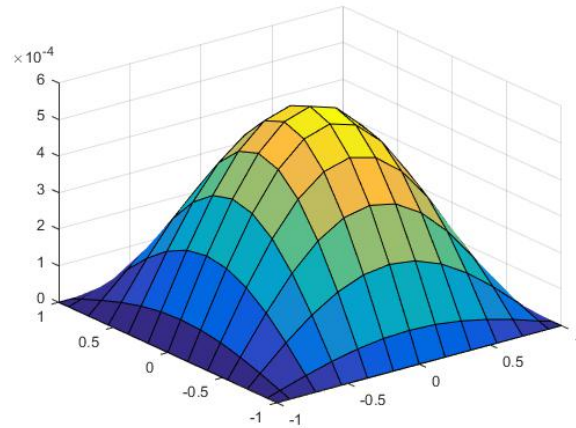


**Figure 2. First-order mode shape of square plate**
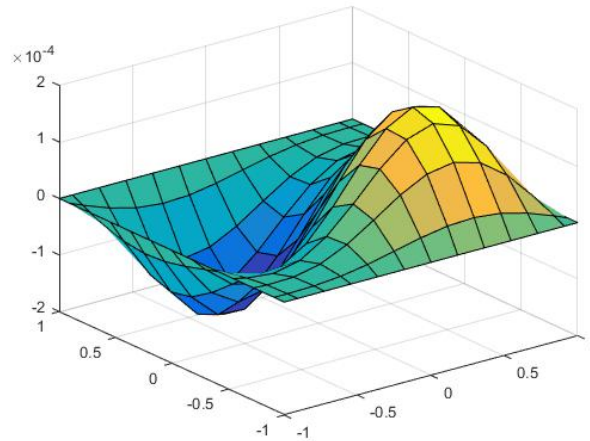


**Figure 3. Second-order mode shape of square plate**
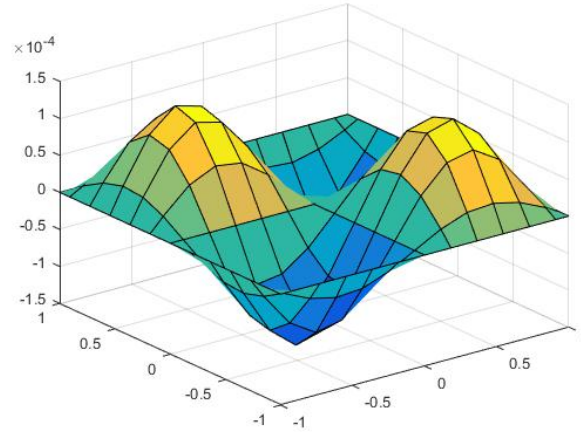
**Figure 4. Third-order mode shape of square plate**

**Example 2** Simply supported rectangular thin plates around Pasternak foundation are shown in Fig.1. We set *a*=1.25, *b*=1, other parameters applied are shown in example 1. According to R-function theory [25-26]

$$\omega_0 = \omega_1 + \omega_2 - \sqrt{\omega_1^2 + \omega_2^2}$$

in which $\omega_1 = \dfrac{1}{2b}(b^2 - y^2)$ and $\omega_2 = \dfrac{1}{2a}(a^2 - x^2)$. Then $\omega_0 = 0$ is the first-order boundary normalized equation of the rectangular plate. $\omega_1 = \dfrac{1}{2b}(b^2 - y^2)$ and $\omega_2 = \dfrac{1}{2a}(a^2 - x^2)$ are the sides of a simply supported plate.

25(5×5), 49(7×7), 81(9×9) and 121(11×11) grid layout schemes are adopted for rectangular plates, respectively. The results are shown in Table 2.

**Table 2. Natural frequencies *f* of rectangularplate**

| Method | | Modal order | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | |
| | | Frequency | Error(%) | Frequency | Error(%) | Frequency | Error(%) |
| The method in this paper | 25 | 33.528 | 3.058 | 46.685 | 6.867 | 55.503 | 9.644 |
| | 49 | 32.997 | 1.426 | 45.278 | 3.647 | 53.100 | 4.897 |
| | 81 | 32.796 | 0.808 | 44.629 | 2.161 | 52.083 | 2.888 |
| | 121 | 32.701 | 0.516 | 44.300 | 1.408 | 51.576 | 1.887 |
| Analytic solutions | | 32.533 | --- | 43.685 | --- | 50.621 | --- |

The modes corresponding to each order of frequencies are shown in Figures 5-7.
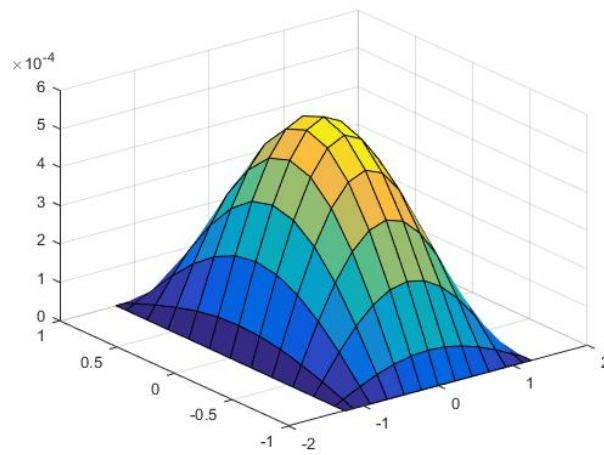
**Figure 5. First-order mode shape of rectangular plate**
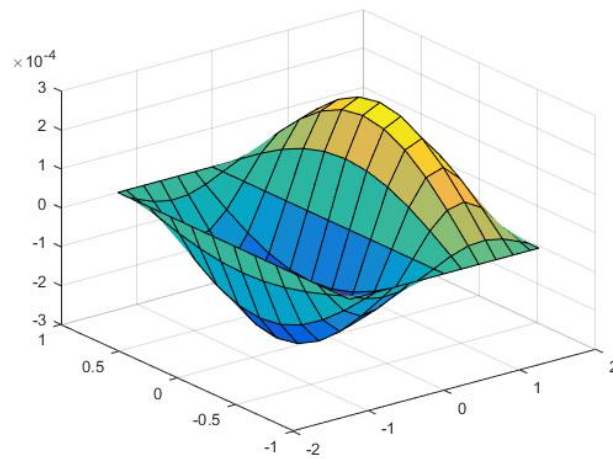


**Figure 6. Second-order mode shape of rectangular plate**
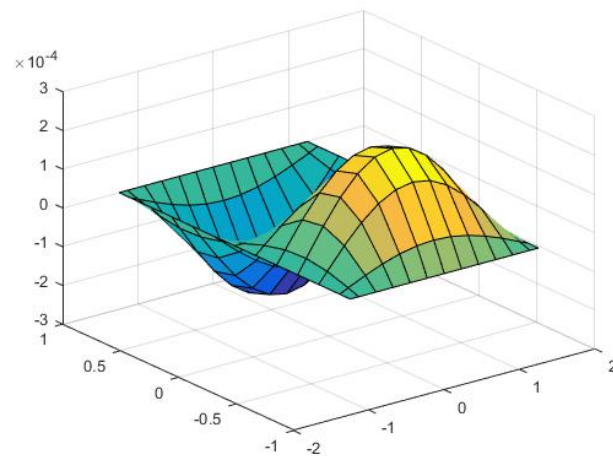


**Figure 7. Third-order mode shape of rectangular plate**

**Example 3** Simply supported trapezoidal thin plate on Pasternak foundation is shown in Fig.8. We set $a$=1.5, $b$=1 and $c$=2, other parameters applied are shown in example 1. According to R-function theory [25-26]

$$\omega_0 = \omega_1 + \omega_2 + \omega_3 - \sqrt{\omega_1^2 + \omega_2^2} - \sqrt{\omega_1^2 + \omega_3^2} - \sqrt{\omega_2^2 + \omega_3^2} + \sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2}$$

in which $\omega_1 = \dfrac{(c-y)y}{c}$, $\omega_2 = \dfrac{1}{\sqrt{1+(\dfrac{c}{a-b})^2}}(\dfrac{ac}{a-b}+\dfrac{c}{a-b}x-y)$,

and $\omega_3 = \dfrac{1}{\sqrt{1+(\dfrac{c}{a-b})^2}}(\dfrac{ac}{a-b}-\dfrac{c}{a-b}x-y)$. Then $\omega_0 = 0$ is the first-order boundary normalized

equation for the trapezoidal plate. $\omega_1 = 0$, $\omega_2 = 0$ and $\omega_3 = 0$ are each side of the simply
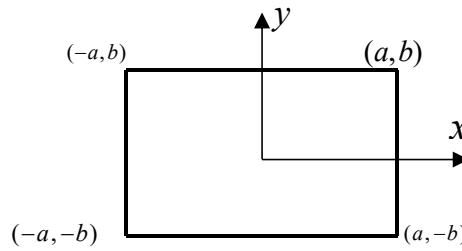
supported trapezoidal plate.
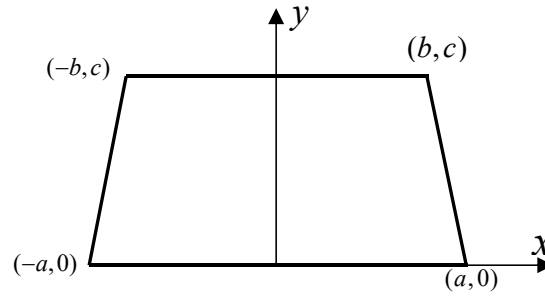


**Figure 8. Simply Supported Trapezoidal Plate**

25(5×5), 49(7×7), 81(9×9) and 121(11×11) grid layout schemes are adopted for rectangular plate, respectively. The results are shown in Table 3.

**Table 3. Natural frequencies $f$ of trapezoidal plate**

| Method | | Modal order | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | | frequency | frequency | frequency |
| The method in this paper | 25 | 33.573 | 46.928 | 55.514 |
| | 49 | 33.078 | 45.482 | 53.136 |
| | 81 | 32.892 | 44.862 | 52.147 |
| | 121 | 32.802 | 44.544 | 51.657 |

The modes corresponding to each order of frequencies are shown in Figures 9-11.
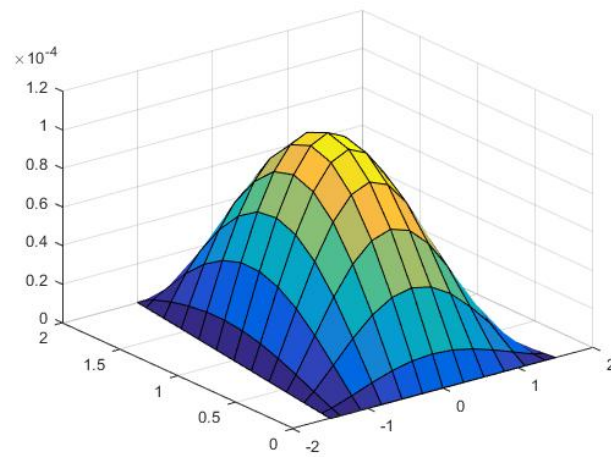
**Figure 9. First-order mode shape of trapezoidal plate**
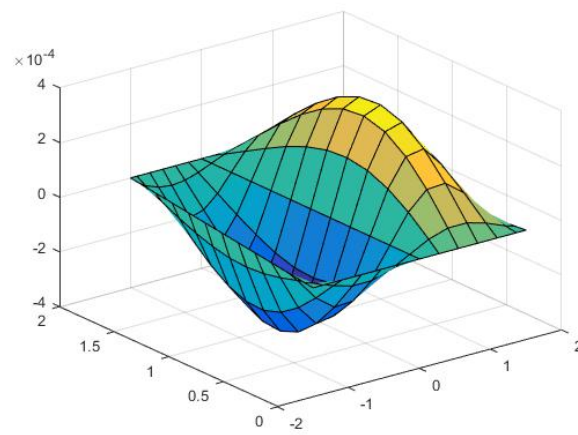


**Figure .10 Second-order mode of trapezoidal plate**
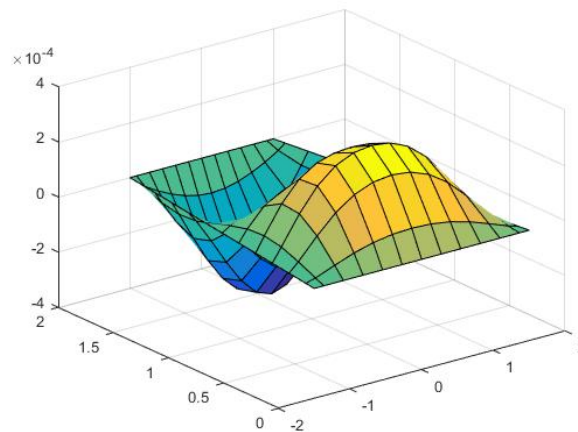


**Figure 11. Third-order mode of trapezoidal plate**

**Example 4** Simply supported polygonal thin plates on Pasternak foundation are shown in Fig.12 .We set $a$=1.5, $b$=2.0, c=0.75, d=1.0, other parameters applied are shown in example 1. According to R-function theory [25-26]

$$\omega_0 = \omega_5 + \omega_6 - \sqrt{\omega_5^2 + \omega_6^2}$$

in which $\omega_5 = \omega_1 + \omega_2 - \sqrt{\omega_1^2 + \omega_2^2}$ , $\omega_6 = \omega_3 + \omega_4 + \sqrt{\omega_3^2 + \omega_4^2}$ ,

and $\omega_1 = \dfrac{a^2 - x^2}{2a} \geq 0$ , $\omega_2 = \dfrac{b^2 - y^2}{2b} \geq 0$ , $\omega_3 = (c - x) \geq 0$ , $\omega_4 = (d - y) \geq 0$ . Then $\omega_0 = 0$ is the

first-order boundary normalized equation for L-shaped plate, $\omega_1 = 0$ , $\omega_2 = 0$ , $\omega_3 = 0$ ,

$\omega_4 = 0$ are the sides of the simply supported L-shaped plate.
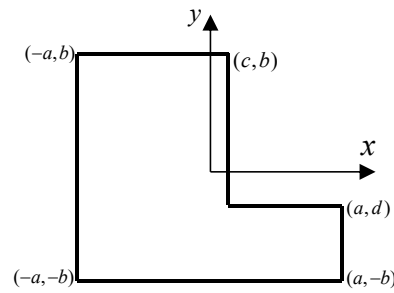


**Figure 12. Simply supported L-shaped plate**

L-shaped plate are arranged in 27(9×3), 75(25×3) and 147(49×3) grids respectively. The results are shown in Table 4.

**Table 4. Natural frequencies *f* of L-shaped plate**

| Method | | Modal order | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | | Frequency | Frequency | Frequency |
| | 27 | 29.146 | 33.701 | 38.056 |
| The method in this paper | 75 | 28.654 | 32.903 | 36.593 |
| | 147 | 28.486 | 32.566 | 36.005 |

The modes corresponding to each order of frequencies are shown in Figures 13-15.
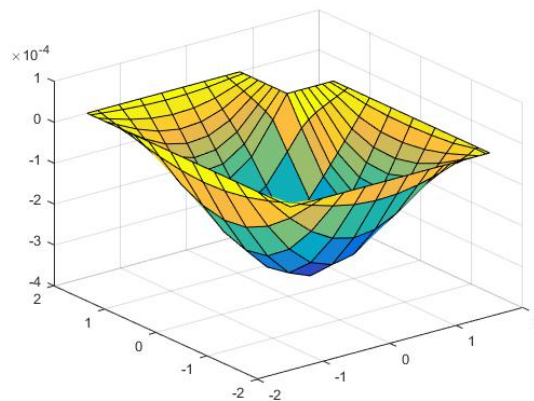


**Figure 13. First-order mode shape of L-shaped plate**
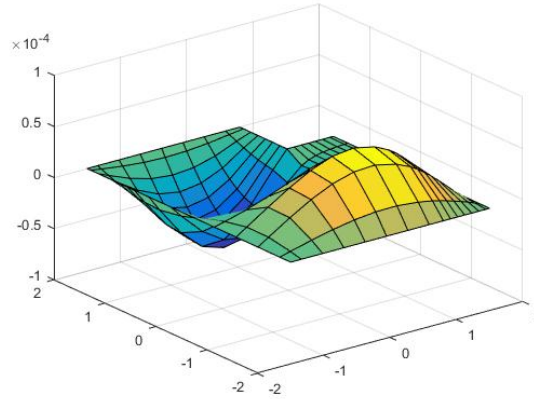
**Figure 14. Second-order mode L-shaped plate**
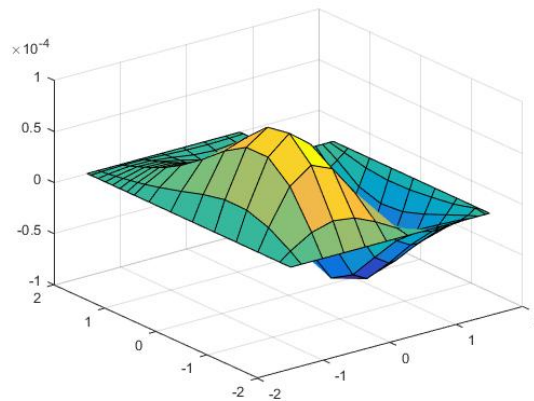


**Figure 15. Third-order modes of L-shaped plate**

**Example 5** Simply supported I-shaped thin plate on Pasternak foundation is shown in Fig.16. We set $a$=1.2, $b$=1.5, $c$=1, $d$=1, other parameters applied are shown in example 1. According to R-function theory [25-26]

$$\omega_0 = \omega_5 + \omega_6 - \sqrt{\omega_5^2 + \omega_6^2}$$

in which $\omega_5 = \omega_1 + \omega_2 - \sqrt{\omega_1^2 + \omega_2^2}$ , $\omega_6 = \omega_3 + \omega_4 + \sqrt{\omega_3^2 + \omega_4^2}$ and

$\omega_1 = \dfrac{a^2 - x^2}{2a} \geq 0$ , $\omega_2 = \dfrac{b^2 - y^2}{2b} \geq 0$ , $\omega_3 = \dfrac{c^2 - x^2}{2c} \geq 0$ , $\omega_4 = \dfrac{y^2 - d^2}{2d} \geq 0$ . Then $\omega_0 = 0$ is the

first-order boundary normalized equation of I-shaped plate. $\omega_1 = 0$ , $\omega_2 = 0$ , $\omega_3 = 0$ and

$\omega_4 = 0$ are each side of the I-shaped plate.
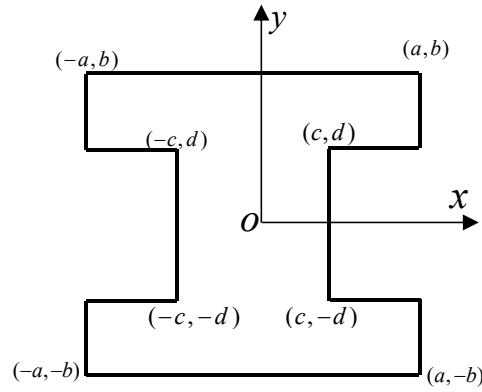
**Figure 16. Simply supported I-shaped plate**

The I-shaped plate are arranged in 27(9×3), 75(25×3) and 147(49×3) grids, respectively. The results are shown in Table 5.

**Table 5. Natural frequencies $f$ of simply supported I-shaped plate**

| Method | | Modal order | | |
| --- | --- | --- | --- | --- |
| | | 1 | 2 | 3 |
| | | Frequency | Frequency | Frequency |
| The method in this paper | 27 | 32.709 | 40.174 | 52.989 |
| | 75 | 32.008 | 39.520 | 52.159 |
| | 147 | 31.745 | 39.137 | 50.743 |

The modes corresponding to each order of frequencies are shown in Figures 17-19.
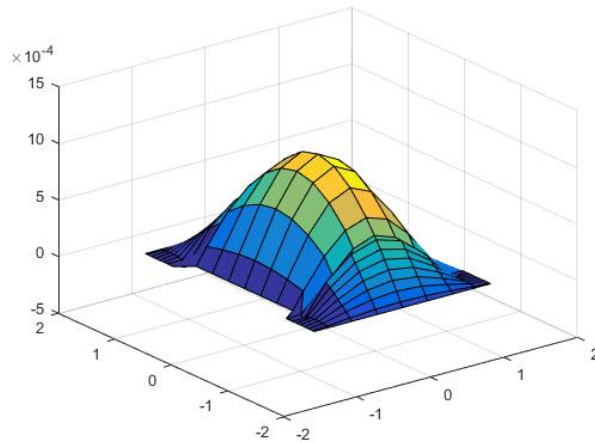


**Figure 17. First-order mode shape of I-shaped plate**
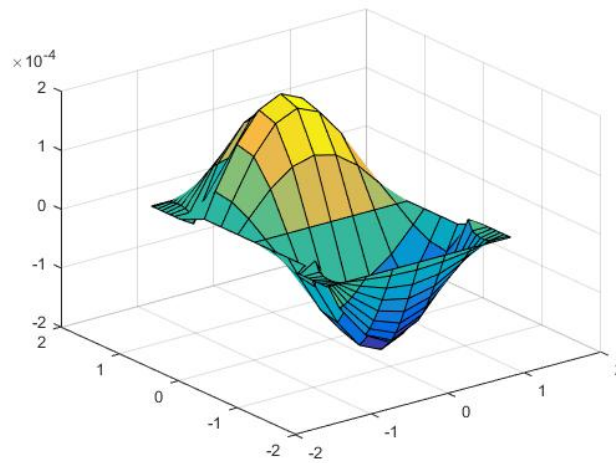
**Figure 18. Second-order mode shape of I-shaped plate**
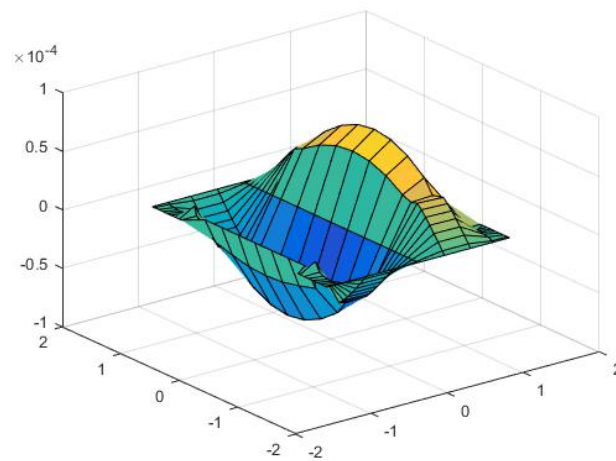


**Figure 19. Third-order mode shape of I-shaped plate**

**Conclusions**

The frequencies of square plate, rectangle plate, trapezoid plate, L-plate and I-plate shown in Table 1, Table 2, Table 3, Table 4 and Table 5 converge with the increase of grid, and the frequencies of square plate and rectangle plate shown in Table 1 and Table 2 are compared with the corresponding analytical solutions, indicating the convergence and high precision of the method. Moreover, the corresponding mode shapes are obtained by the calculated frequencies of each plate. It is proved that R-function theory is effective for the expression of complex shape. Compared with the analytical solution, the feasibility and effectiveness of this method are also proved. The computing method of two-dimensional Helmholtz operator introduced in this paper used to analyze the free vibration of simply supported thin plate with arbitrary shape is effective. The convergence speed of this proposed method is fast and it is very practical for solving boundary value problems. It can also further develop its application in interdisciplinary fields.

## Funding

## Acknowledgments

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

[1] Romero, A., Galvín, P., Mara-Molina, J. C., and Tadeu, A. (2019) On the formulation of a BEM in the Bézier-Bernstein space for the solution of Helmholtz equation, *Applied mathematical modelling* **74**, 301-319.

[2] Tiffany, N. and Jones, Q. S. (2019) Asymptotic stability of a dual-scale compact method for approximating highly oscillatory Helmholtz solutions, *Journal of computational physics* **392**, 403-418.

[3] Anna, R. C., Eduardo, G. D. C., Webe, J. M. and Katia, P. F. (2019) The size function for a HDG method applied to the Helmholtz problem, *Computational and Applied Mathematics* **38**.

[4] Li, J. M., Tian, Q., Zheng, F. Y., Wu, W. F. and Wang, J. X. (2018) Research on parallel solution of GRAPES Helmholtz equation, *Concurrency and Computation: Practice and Experience* **31**.

[5] Du, Y. and Zhang, Z. M. (2019) Supercloseness of Linear DG-FEM and Its Superconvergence Based on the Polynomial Preserving Recovery for Helmholtz Equation, *Journal of scientific computing* **79**, 1713-1736.

[6] Jeffrey, G., Eike, H. M. and Euan, A. (2019) Spence Wavenumber-explicit analysis for the Helmholtz h-BEM: error estimates and iteration counts for the Dirichletproblem, *Numerische Mathematik* **142**, 329-357.

[7] Hoshoudy, G. A., Cavus, H. and Mahdy, A. (2019) A viscous magnetohydrodynamic Kelvin-Helmholtz instability in the interface of two fluid layers: Part I. Basic mechanism, *Astrophysics and Space Science* **364**, 89.

[8] Lin, T., Lin, Y. P. and Zhuang, Q. (2019) Solving Interface Problems of the Helmholtz Equation by Immersed Finite Element Methods, *Communications on Applied Mathematics and Computation* **1**, 187-206.

[9] Mandel, R. (2019) The Limiting Absorption Principle for Periodic Differential Operators and Applications to Nonlinear Helmholtz Equations, *Communications in Mathematical Physics* **368**, 799-842.

[10] Stein, L. and Sesterhenn, J. (2019) An acoustic model of a Helmholtz resonator under a grazing turbulent boundary layer, *Acta Mechanica* **230**, 2013-2029.

[11] Shahbaztabar, A., Izadi, A., Sadeghian, M. and Kazemi, M. (2019) Free vibration analysis of FGM circular cylindrical shells resting on the Pasternak foundation and partially in contact with stationary fluid, *Applied Acoustics* **153**, 87-101.

[12] Radić, N. (2018) On buckling of porous double-layered FG nanoplates in the Pasternak elastic foundation based on nonlocal strain gradient elasticity, *Composites Part B: Engineering* **153**, 465-479.

[13] Mohanty, M., Pradhan, M. M. and Mishra, D. R. (2018) Stability analysis of a two-layer elastic beam

resting on a variable Pasternak foundation subjected to an axial pulsating load and thermal gradient[J]. *Materials Today-Proceedings* **5**, 23619-23628.

[14] Wei, W. (2019) Analytical analysis of the impact on adjacent underground pipelines due to foundation pit excavation on the Pasternak foundation, *Journal of Water Resources & Water Engineering* **30**, 212-216.

[15] Lin, C. G.and Huang, M. S. (2018) Deflections of discontinuous buried pipelines induced by shield tunnelling based on Pasternak foundation, *Chinese Journal of Geotechnical Engineering* **6**, 1-8.

[16] Ahmed, F. Ashraf, R. and Zenkour, M. (2018) Quasi 3-D trigonometric plate theory for bending analysis of EG plates resting on Pasternak foundations, *Curved and Layered Structures* **5**.

[17] Chen, F .Q., Lin, L. B. and Wang, J. J. (2018) Energy method as solution for deformation of geosynthetic-reinforced embankment on Pasternak foundation, *Applied Mathematical Modelling* **2018**.

[18] Ghafarian, M. and Ariaei, A. (2019) Forced vibration analysis of a Timoshenko beam featuring bending-torsion on Pasternak foundation, *Applied mathematical modelling* **2019**, 472-485.

[19] Zhang, H. Shi, D.Y., Zha, S. and Wang, Q. S. (2018) Sound-vibration behaviors of the thin orthotropic rectangular fluid-structure coupled system resting on varying elastic Winkler and Pasternak foundations, *Results in Physics* **2018**,188-200.

[20] Mohanty, M., Bhattacharya, S., Yadav, S. and Pradhan, M. (2018) Free vibration analysis of a two-layer elastic beam resting on a variable Pasternak foundation, *Materials Today: Proceedings* **5**.

[21] Arani, A. G., Hassan, B. Z., Pourmousa, P. and Eskandari, M. (2018) Investigation of free vibration response of smart sandwich micro-beam on Winkler-Pasternak substrate exposed to multi physical fields, *Microsystem Technologies* **24**, 3045-3060

[22] Zhao, M.H., Peng, W.Z., Yang, C.W. and Xiao, Y. (2018) Inner-force calculation of bridge pile foundation in a high-steep slope based on the Pasternak double-parameter spring model, *HYDROG ELOGY & ENGINE ERING GEOLOGY* **45**, 52-59.

[23] Chen L. C. and Li X. L. (2018) An interpolating boundary element-free method for 2D Helmhotz equation, *Applied Mathematics and Mechanics* **39**,470-484.

[24] Dai H. and Pan, W. F. (2018) Aspectral element method for transmission eigenvalue problems of the Helmholtz equation, *Applied Mathematics and Mechanics* **39**, 833-840.

[25] Kurpa, L., Shmatko, T. and Awrejcewicz, J. (2019) Vibration analysis of laminated functionally graded shallow shells with clamped cutout of the complex form by the Ritz method and the R-functions theory, *Latin American Journal of Solids and Structures* **16**, 1-16.

[26] Rvachev, V. L. and Kurpa, L.V. (1987) The R-functions for problems of plates theory, *Kiev, Naukova Dumka*, (in Russian).

[27] Li, S. Q. and Yuan, H. (2011) Quasi-Green's function method for free vibration of clamped thin plates on Winkler foundation, *Applied Mathematics and Mechanics (English Edtion)* **32**, 265-276.

[28] Li, S. Q. and Yuan, H. (2010) Quasi-Green's function method for free vibration of simply-supported trapezoidal shallow spherical shell, *Applied Mathematics and Mechanics (English Edtion)* **31**, 635-642.

[29] Li, S. Q. and Yuan, H. (2010) Green quasifunction method for free vibration of simply-supported trapezoidal shallow spherical shell on Winkler foundation, *Acta Mechanica Solida Sinica* **23**,370-376.

[30] Li, S. Q. and Yuan, H. (2012) Green quasifunction method for free vibration of clamped thin plates, *Acta Mechanica Solida Sinica* **25**, 37-45.

[31] Yuan, H., Li, S. Q. and Liu, R. H. (2007) Green quasifunction method for vibration of simply-supported thin polygonic plates on Pasternak foundation, *Applied Mathematics and Mechanics(English Edtion)* **28,** 847-853.

# Influence of random soil parameters on seismic reliability of underground structure

**†Yifan Fan[1], *Zhiyi Chen[1], and Zhiqian Liu[1]**

[1]Department of Geotechnical Engineering, Tongji University, China.

*Presenting author: 13754141231@163.com
†Corresponding author: zhiyichen@tongji.edu.cn

## Abstract

The underground structure in the soil is difficult to repair after the earthquake, and the rescue is also a difficult problem. Therefore, reasonable determination of seismic reliability of underground structure can ensure the seismic capacity of underground structure in seismic analysis and design. The response of underground structure in earthquake is mainly controlled by the deformation of surrounding soil. In addition to the uncertainty of ground motion and structural material and form, the uncertainty of soil parameters is the key factor affecting its seismic reliability. The seismic reliability analysis of underground structure is a stochastic dynamic problem. For the determined underground structure, in order to study the influence of the randomness of soil parameters, it is necessary to control the influence of the randomness of ground motion. Here, the method of generating the ground motion which belongs to the same probability complete set is used to control the variables. In order to improve the calculation effect and accuracy, an improved Monte Carlo simulation method is used to select the sample points, and the dynamic time history analysis is carried out with and without considering the randomness of soil parameters. The probability that the inter-storey drift ratio does not exceed the limit value given in the code is defined as the seismic reliability of subway station. The results show that ignoring the randomness of soil parameters will underestimate the potential earthquake disasters faced by underground structures, and reduce the requirements for the seismic capacity of underground structures. The influence of randomness of soil parameters on seismic reliability of underground structures should be included in the future seismic research.

**Keywords:** Subway station structure; Randomness of soil parameters; Seismic reliability; Monte Carlo simulation method; Uncertainty analysis

## Introduction

Subway station structure plays the role of urban lifeline in urban rail transit. Once it is damaged by earthquake, it is very difficult to repair the structure and rescue the personnel. Generally speaking, the randomness of structural seismic reliability mainly comes from two aspects. One is the randomness of ground motion, including frequency, intensity, propagation direction and other parameters [1]-[3]. On the other hand, it is the randomness of the structure itself, including the structural form and the material properties of the structure itself [4] [5].

Different from the above ground structure, the mechanical behavior of subway station during earthquake is mainly controlled by the deformation of soil, and the influence of soil parameter randomness on the seismic reliability of the structure can not be ignored.

It is a new trend to improve the seismic design method to predict the influence of soil conditions on seismic reliability of underground structures. In fact, the evolution of geological conditions occurs at all times, and even the soil properties of the same site have great variability [6]. For seismic analysis, in addition to the physical and mechanical parameters such as soil density and void ratio, it is worthier of attention to the shear wave velocity of soil. In the seismic code of China and the United States, shear wave velocity is used as the basis for site classification [7]-[9]. Some studies on the influence of shear wave velocity on seismic response have also been published [6] [10] [11]. In view of the uncertainty of shear wave velocity in soil, Toro et al. [12] established a shear wave velocity profile model based on the analysis of more than 500 shear wave velocity profiles. The model simplifies the analysis and takes into account the inter layer correlation of soil layers. This paper uses this method for reference, and establishes the shear wave velocity profile which is suitable for the model studied in this paper by Monte Carlo simulation method, and takes it as the input parameter of seismic analysis.

The purpose of this paper is to study the influence of the randomness of soil parameters on the seismic reliability of structures. Therefore, the randomness of ground motions should be eliminated or unified in the seismic reliability analysis of certain subway station structures. Liu et al. [13] used the idea of random function to reduce the number of random variables needed to generate ground motions to one. The ground motions generated by this method belong to the same probability complete set. For this kind of uncertainty problem, the same randomness of ground motions can be ensured statistically by generating ground motions according to the above method for dynamic time history analysis.

In conclusion, in order to explore the influence of soil parameters randomness on the seismic reliability of subway station structure, this paper uses Monte Carlo simulation method to generate random ground motion and shear wave velocity profiles, which are used as input parameters for dynamic time history analysis. For underground RC frame structures such as subway stations, the inter-storey drift ratio is an effective seismic performance evaluation index [14] [15]. The inter-storey drift ratio limit values of underground RC frame structure in elastic and elastoplastic state are also given in Chinese seismic codes [7] [8]. It is a reasonable method to define the probability of not exceeding the limit of inter-storey drift ratio given in the codes as the seismic reliability of subway station structure. The seismic reliability of subway station structure with and without considering the randomness of soil parameters is calculated respectively. The results show that the seismic damage risk of underground structure will increase without considering the randomness of soil parameters. Therefore, in the seismic analysis and design of underground structures, it is necessary to treat soil parameters as random variables for uncertainty analysis.

**Point selection strategy**

When sampling with Monte Carlo simulation method, the selection of sample points must be representative. The traditional Monte Carlo simulation method uses the method of generating random numbers to select sample points. When the number of sampling is less, the distribution of sample points will be uneven. In order to ensure the uniformity of sampling and reduce the cost of calculation, an improved Monte Carlo simulation method is adopted in this paper, that is, the Latin hypercube sampling method [16] is used to evenly divide the interval of generating random number, and then random points are selected in each new area, so as to ensure the uniformity of sample points. It is worth mentioning that, based on the randomly generated (0,1) interval obeying uniformly distributed sample points, the method of inverse CDF transform can be used to sample random variables of any distribution form [17].

Therefore, the point selection strategy of this paper is as follows: (1) according to the number of representative points n, the (0,1) interval is evenly divided, and n new areas are obtained; (2) Sampling points are randomly selected in each new area; (3) Through the inverse CDF transformation of the above sample points, random variable sampling under arbitrary distribution is completed.

**Finite element model**

The model studied in this paper is a typical two-story three span subway station structure in Shanghai. The height and width of the subway station structure are 12.37m and 20.9m respectively. The total height and width of the soil structure interaction model shown in Figure 1 are 60.5m and 1000m respectively. A 280m wide infinite element boundary is set on both sides of the site to eliminate the influence of seismic reflection.
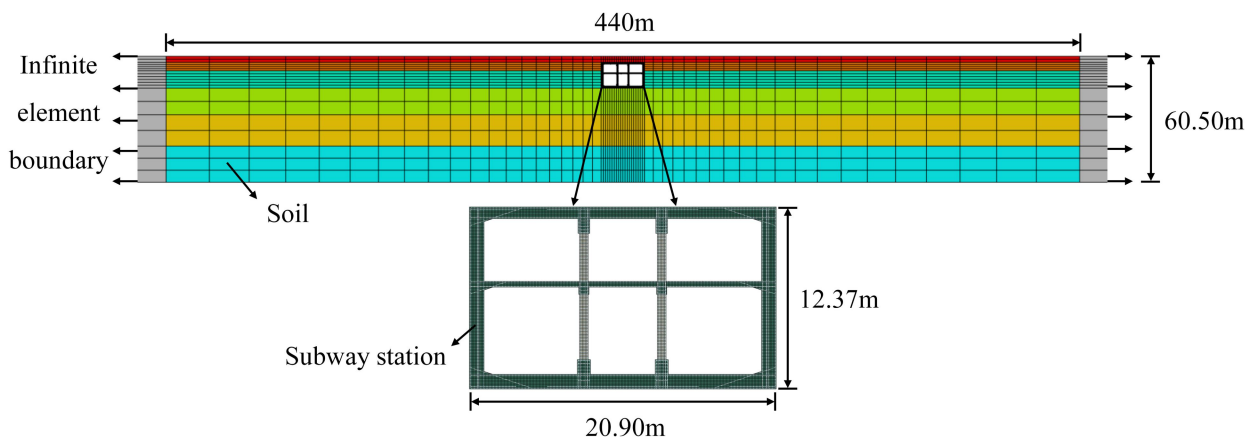


**Figure 1. Soil structure interaction model**

**Soil parameters and shear wave velocity profile**

*Soil parameters*

The soil parameters of the original site are shown in Table 1.

**Table 1. Soil layer information**

| Soil layer | Soil name | Mass density (kg•m⁻³) | Cohesion (kPa) | Angle of internal friction (°) | The mean of the shear wave velocity(m/s) | The standard deviation of the shear wave velocity(m/s) |
|---|---|---|---|---|---|---|
| S1 | Brown yellow silty clay | 1920 | 16.90 | 25.00 | 101 | 9.25 |
| S2 | Grey silty clay | 1670 | 11.00 | 11.20 | 130 | 14.24 |
| S3 | Dark green silty clay | 1950 | 44.0 | 16.40 | 292 | 44.38 |
| S4 | Grassy yellow, gray sandy silt | 1820 | 2.00 | 31.10 | 278 | 42.26 |
| S5 | Gray clay | 1770 | 17.00 | 15.20 | 231 | 43.66 |
| S6 | Gray silty clay interbedded with silt | 1840 | 8.00 | 28.10 | 279 | 24.55 |

*Shear wave velocity profile*

Toro [12] established a shear wave velocity profile model considering the correlation between soil layers, which assumes that the shear wave velocity obeys lognormal distribution at any depth. In order to facilitate the application of the model in different site conditions, Toro [12] analyzed more than 500 shear wave velocity profiles, divided the site into different categories according to the equivalent shear wave velocity within 30 meters depth, $V_{s30}$, and gave the model parameters with different correlations.

Assuming that $V_s$ obeys lognormal distribution at any depth, the shear wave velocity of layer i can be expressed as follows:

$$V_s(i) = \exp\left\{ \ln\left[ V_{s,m}(i) \right] + X_i \cdot \sigma_{\ln V_s} \right\}$$
(1)

Among them, $V_{s,m}(i)$ is the mean value of $V_s(i)$, $\sigma_{\ln V_s}$ is the natural logarithm of the standard deviation of $V_s(i)$, both are calculated according to the data in table 1; $X_i$ is a standard normal distribution random variable, which represents the standard deviation number between $\ln\left[ V_s(i) \right]$ and $\ln\left[ V_{s,m}(i) \right]$ in the logarithmic space.

The correlation of $X_i$ can be expressed by Eq. (2):

$$X_i = \begin{cases} \varepsilon_i & i = 1 \\ \rho_{IL} \cdot X_{i-1} + \varepsilon_i \cdot \sqrt{1 - \rho_{IL}^2} & i \geq 2 \end{cases} \tag{2}$$

$\varepsilon_i$ is a standard normal distribution random variable with a mean value of 0 and a standard deviation of 1; $\rho_{IL}$ is the correlation coefficient between layer *i* and layer *i-1*, and C can be expressed as a function of depth *d* and layer thickness *t*, as shown in Eq. (3):

$$\rho_{IL}(d,t) = \left[1 - \rho_d(d)\right]\rho_t(t) + \rho_d(d) \tag{3}$$

Where, $\rho_d(d)$ and $\rho_t(t)$ represent depth correlation coefficient and layer thickness correlation coefficient respectively, and their definitions are shown in Eq. (4) and Eq. (5) respectively:

$$\rho_d(d) = \begin{cases} \rho_{200}\left[\dfrac{d + d_0}{200 + d_0}\right]^b & d \leq 200m \\ \rho_{200} & d > 200m \end{cases} \tag{4}$$

$$\rho_t(t) = \rho_0 \exp\left(\frac{-t}{\Delta}\right) \tag{5}$$

Among them, $\rho_{200}$, $d_0$, $b$, $\rho_0$ and $\Delta$ are all related parameters of the model, which are determined according to the average shear wave velocity $V_{s30}$ in the field within 30 m depth. According to reference [12], the values of the above five parameters are $\rho_{200} = 0.98$, $d_0 = 0$, $b = 0.344$, $\rho_0 = 0.99$, $\Delta = 3.9$ respectively.

The random shear wave velocity profile generated in this paper is shown in Figure 2. In order to describe the variation range of shear wave velocity profile, the profiles at $\text{Exp}(\ln V_s \pm 2\sigma_{\ln V_s})$ are given. It can be seen from the figure that three randomly generated samples are roughly distributed in this range. In addition, it can also be noted that with the increase of depth, the overall trend of shear wave velocity increases, but there will be local reentry.
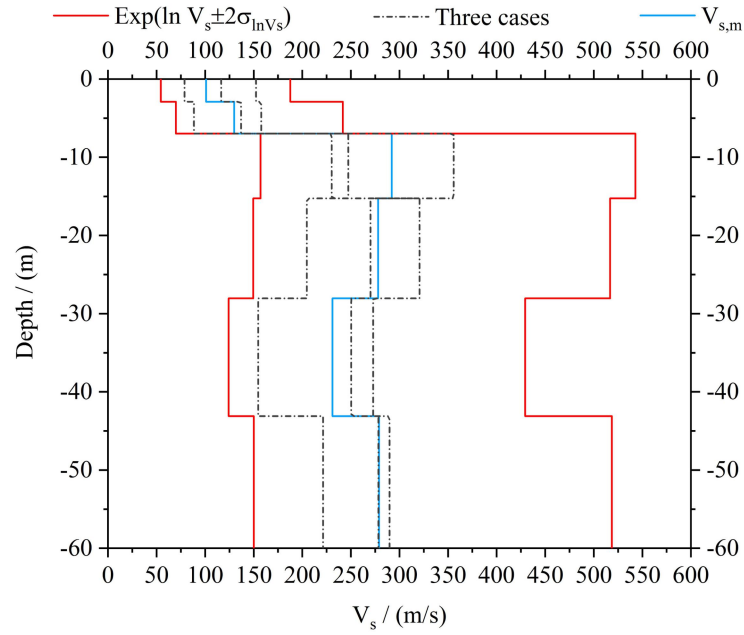
**Figure 2. Shear wave velocity profile**

**Influence of randomness of soil parameters on reliability of subway station**

In order to explore the influence of soil parameter randomness on the seismic reliability of subway station structure, the random shear wave velocity profile is used as the input parameter. According to the random ground motion generation method proposed by Liu et al. [13], the ground motions under the same probability complete set are obtained to control its influence on the seismic reliability of the structure. This method only needs to provide the phase angle θ, Figure 3 shows a typical time history acceleration-ground motion curve.
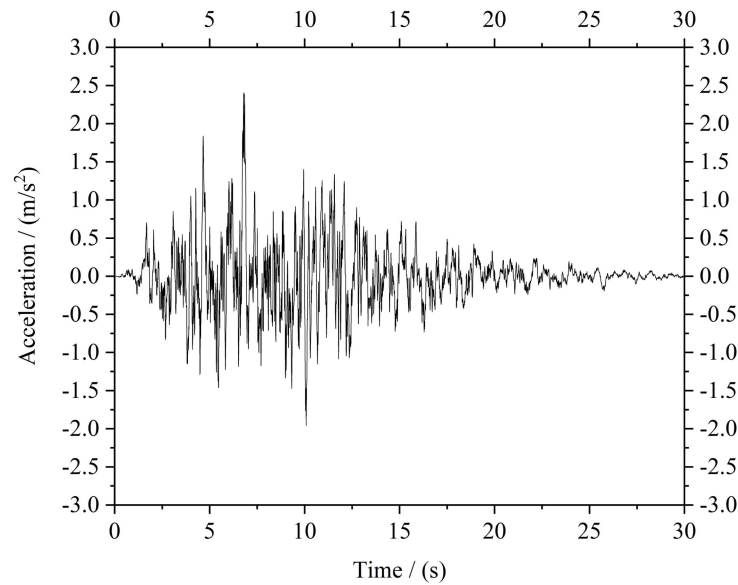


**Figure 3. Typical ground motion**

The inter-storey drift ratio is an effective index to evaluate the seismic performance of underground RC frame structure. Here, the reliability r of subway station structure is defined as the probability of not exceeding the specified inter-storey drift ratio limit. In China's seismic codes [7] [8], the limit of inter-storey drift ratio of underground RC frame structures under elastic and elastic-plastic state is given, as shown in Table 2.

**Table 2. Limit of the inter-storey drift ratio in Chinese codes**

| Chinese codes | Applicable objects | ES | EPS |
|---|---|---|---|
| Code for seismic design of subway structures (DG/TJ 08-2064-2009) | Concrete frame structure | 1/550 | - |
| | Reinforced concrete frame structure | - | 1/250 |
| Standard for seismic design of underground structures (GB/T 51336-2018) | Single-layer or double-layer rectangular section structure | 1/550 | 1/250 |
| | Three-storey and above rectangular section structure | 1/1000 | 1/250 |

Note: ES is the elastic state, EPS is the elastic-plastic state.

The seismic reliability analysis of the soil-structure interaction model studied in this paper involves a total of seven basic random variables, which are respectively the shear wave velocity of six soil layers and the phase angle of ground motion. According to the point selection strategy described in Chapter 2, 300 sample points, i.e., 300 cases of different ground motion and shear wave velocity profiles, were generated. The dynamic time history analysis was conducted for the above 300 working conditions (considering the randomness of soil parameters) and the original site condition using only the above 300 ground motions (without considering the randomness of soil parameters). The maximum interlayer displacement Angle in each working condition was extracted, and the inter-storey drift ratio corresponding to the ground motion samples was drawn in Fig. 4. Fig. 4 (a) is the distribution of inter-storey drift ratio without considering the randomness of soil parameters, and Fig. 4 (b) is the distribution of inter-storey drift ratio with the randomness of soil parameters considered.

Compared with figure 4 (a) and Figure 4 (b), it can be seen that under the premise of the same seismic samples, the inter-storey drift ratio distribution is more dispersed when considering the randomness of soil parameters, which well illustrates the influence of randomness of soil parameters on the seismic response of underground structures. Taking the maximum value as

an example, the maximum value of inter-storey drift ratio without considering the randomness of soil parameters is about 0.0027, and the minimum value is about 0.0008; Considering the randomness of soil parameters, the maximum value of inter-storey drift ratio is about 0.0053, and the minimum value is about 0.0006. The maximum difference between the two cases is as high as one time. In the seismic analysis and design, if the influence of random soil parameters on the seismic response of underground structures is not fully considered, the seismic risk of the structure will be underestimated.



(a)The randomness of soil parameters is not considered    (b)The randomness of soil parameters is considered
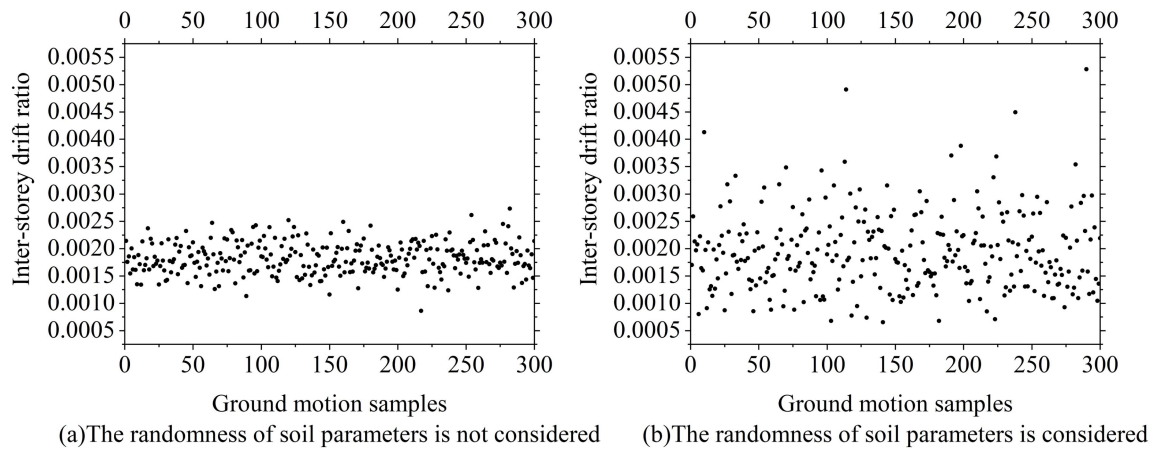
**Figure 4. Distribution of inter-storey drift ratio**

According to the code for the limit value of inter-storey drift ratio, the seismic reliability index of subway station structure is defined as the probability of not exceeding the limit value given in the code, which are the reliability $R_e$ in elastic state and the reliability $R_{ep}$ in elastic-plastic state. Table 3 lists the reliability of subway station structure with or without random soil parameters.

**Table 3. Comparison of reliability**

|  | The randomness of soil parameters is not considered | The randomness of soil parameters is considered |
|---|---|---|
| $R_e$ | 0.5300 | 0.4833 |
| $R_{ep}$ | 1 | 0.9867 |

Note: $R_e$ is the reliability of elastic state, $R_{ep}$ is the reliability of elastic plastic state.

It can be seen from Table 3 that the seismic reliability of subway station without considering the randomness of soil parameters is always greater than that of subway station with considering the randomness of soil parameters in both elastic and elastic-plastic states. In the elastic-plastic state, without considering the influence of soil parameter randomness on the seismic reliability of subway station, the reliability reaches 100%. Obviously, such evaluation will reduce the construction requirements of underground structures and bring potential risk factors. Therefore, reasonable consideration of the randomness of soil parameters is the key step to ensure the seismic capacity of underground structures.

## Conclusions

The seismic reliability of underground structures contains many uncertainties. Compared with the above ground structures, the most important source of uncertainty is the uncertainty of soil parameters. In this paper, the improved Monte Carlo simulation method is used to generate representative sample points. By controlling the ground motion variables under the same probability complete set, the seismic reliability of subway station structure with and without considering the randomness of soil parameters is studied. The results show that when the randomness of soil parameters is not considered, the seismic response distribution of the structure is more concentrated. And the maximum value of inter-storey drift ratio is only half of that considering the randomness of soil parameters, which will underestimate the potential seismic disaster of underground structures. According to the limit value of inter-storey drift ratio given by the code, the seismic reliability of subway station structure is defined. It is found that the corresponding seismic reliability will be overestimated without considering the influence of random soil parameters on the seismic response of subway station structure, which will reduce the seismic demand of underground structure and increase the risk of earthquake. Therefore, in the seismic analysis and design of underground structures, the uncertainty of soil parameters should be considered, and the seismic reliability of underground structures should be analyzed combined with the real site conditions to ensure the seismic capacity of underground structures.

## Acknowledgment

**References**

[1] Xiong, M., Huang, Y. and Zhao, Q. (2018) Effect of travelling waves on stochastic seismic response and dynamic reliability of a long-span bridge on soft soil. *Bulletin of Earthquake Engineering, 16*(9), 3721-3738.
[2] Sun, B. B., Deng, M. J., Zhang, S. R., Cui, W., Wang, C., Yu, L. C., and Cao, K. L. (2020) Inelastic dynamic analysis and damage assessment of a hydraulic arched tunnel under near-fault SV waves with arbitrary incoming angles. *Tunnelling and Underground Space Technology, 104.*
[3] Nejati, H. R., Ahmadi, M. and Hashemolhosseini, H. (2012) An investigation on the ground motion parameters and seismic response of underground structures. *Earthquake Science, 25*(3), 253-261.
[4] Lim, H. K., Kang, J. W., Pak, H., Chi, H S., Lee, Y. G. and Kim, J. (2018) Seismic Response of a Three-Dimensional Asymmetric Multi-Storey Reinforced Concrete Structure. *Applied Sciences, 8*(4).
[5] Wang, Z. Y., Zhang, N., Li, Q. and Chen, X. H. (2017) Dynamic response of bridge abutment to sand-rubber mixtures backfill under seismic loading conditions. *Journal of Vibroengineering, 19*(1), 434-446.
[6] Rathje, E. M., Kottke, A. R. and Trent, W. L. (2010) Influence of input motion and site property variabilities on seismic site response analysis. *Journal of Geotechnical and Geoenvironmental Engineering, 136(4)*, 607-619.
[7] Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). Standard for seismic design of underground structures (GBT51336-2018). Beijing: China Architecture & Building Press; 2010 (in Chinese).
[8] Shanghai Urban Constructions Communications (SUCC). Code for Seismic Design of Subway Structures (DG/TJ08-2064-2009). Shanghai: Shanghai Building Materials Industry Market Board; 2010 (in Chinese).
[9] SEAOC. Vision 2000: Performance based seismic engineering of buildings (Vols. I and II), Conceptual framework. Sacramento (CA): Structural Engineers Association of California; 1995.

[10] Mahsuli, M., Rahimi, H. and Bakhshi, A. (2018) Probabilistic seismic hazard analysis of Iran using reliability methods. *Bulletin of Earthquake Engineering, 17*(3), 1117-1143.

[11] Jakka, R. S., Roy, N. and Wason, H. R. (2014) Implications of surface wave data measurement uncertainty on seismic ground response analysis. *Soil Dynamics and Earthquake Engineering, 61-62*, 239-245.

[12] Toro, G. R. (1995) Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies. Technical Rep. No. 779574, Brookhaven National Laboratory, Upton, N.Y.

[13] Liu, Z. j., Liu, W. and Peng, Y. B. (2016) Random function based spectral representation of stationary and non-stationary stochastic processes. *Probabilistic Engineering Mechanics, 45*, 115-126.

[14] Chen, Z. Y., Fan, Y. F and Jia, P. (2021) Influence of buried depth on seismic capacity of underground subway stations through performance-based evaluation. *Structures, 32*, 194-203.

[15] Chen, Z. Y. and Liu, Z. Q. (2019) Stochastic seismic lateral deformation of a multi-story subway station structure based on the probability density evolution method. *Tunnelling and Underground Space Technology, 94*.

[16] Stein, M. (1987) Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics, 29*(2), 143-151.

[17] Hajian, M., Rosehart, W. D. and Zareipour, H. (2013) Probabilistic Power Flow by Monte Carlo Simulation With Latin Supercube Sampling. IEEE Transactions on Power Systems, 28(2), 1550-1559.

# Static equivalent modeling of dynamic seismic forces for realizing earthquake resistant spatial truss structures

**\*Zhiyuan Gao[1], †Koichiro Ishikawa[1]**

[1]Department of Architecture and Civil Engineering, University of Fukui, Japan.

\*Presenting author:  1340086909@qq.com
†Corresponding author: ishikawa@u-fukui.ac.jp

## Abstract

This study deals with spatial truss structures composed of a double layer spatial truss dome and two truss plate-type wall substructures subjected to horizontal earthquake motion. Comparative investigations of the structures are also carried out by considering the structural eccentricity varying the wall height. The equivalent static seismic force, which is hereinafter referred to as the static seismic force, is calculated by the response spectrum method, and the seismic force distribution and response deformation applied on the structure are analyzed and investigated. The elastic seismic response analysis using static seismic force and the influence of eccentricity on seismic response characteristics is carried out. In addition, a method to calculate the static seismic force of medium and medium-sized large-span double-layer truss dome is also proposed by considering the shape, height, span and support conditions of the structure. The purpose of this study is to investigate an effect of the substructure on the seismic static force because the roof truss structure with curved surface has higher stiffness. In addition, the seismic response characteristics of long-span structures are different from those of multi-story structures, which have the structural characteristics of its vertical response as well as horizontal response to horizontal earthquake motion. Based on the elastic seismic response analysis and the static analysis of the static seismic force calculated by the response spectrum method, the seismic force distribution and deformation characteristics of the analysis model with or without eccentricity are analyzed and discussed. It is seen that the static equivalent modeling with the maximum effective mass ratio can be used to obtain the seismic force distribution in this kind of long-span structure.

**Keywords:** Spatial truss structure, Truss wall, Truss dome, Horizontal earthquake motions, Equivalent static seismic force, Seismic force distribution, Structural eccentricity.

## Introduction

The in-plane stiffness of roof truss with curved surface is high, so the depth can be reduced relative to the span. In addition, the seismic response of long-span structure is different from that of multi-story structure, which has the structural characteristics of not only horizontal vibration but also vertical response. Considering the shape, height, span and support conditions, a method for calculating the static seismic force of medium and small-scale large-span double-layer solid vault is also proposed.

In this study, the equivalent static seismic force, which is hereinafter referred to as static seismic force, is calculated by the response spectrum method. The seismic force distribution and the response deformation of the double-layer truss roof supporting the three-dimensional truss walls are also carried out by the elastic dynamic analysis. In addition, the elastic static analysis using the static seismic force are compared with the dynamic analysis and the effect of the eccentricity on seismic response state is also investigated by the results of the both analyses.

**Analysis models**

The analysis models are composed of roofs and walls integrated double-layer three-dimensional truss structure as shown in Fig. 1. The one of the analysis models is generated by the structural eccentricity in the Z direction in Fig.1(a). The two roof structures in Fig.1(a) and (b) have the same structural characteristics as shown in Table 1. They have EP dome shape with the curvature radii $R_{x1}$, $R_{x2}$, $R_{z1}$ and $R_{z2}$ in Fig. 2. The half open angles of the dome have $\theta_x = \theta_z = 30°$ along two directions such as X and Z directions.

The two wall structures with the different height are also shown in Fig. 3. The wall model H is set to be the half height of wall model F.



(a) With eccentricity          (b) Without eccentricity
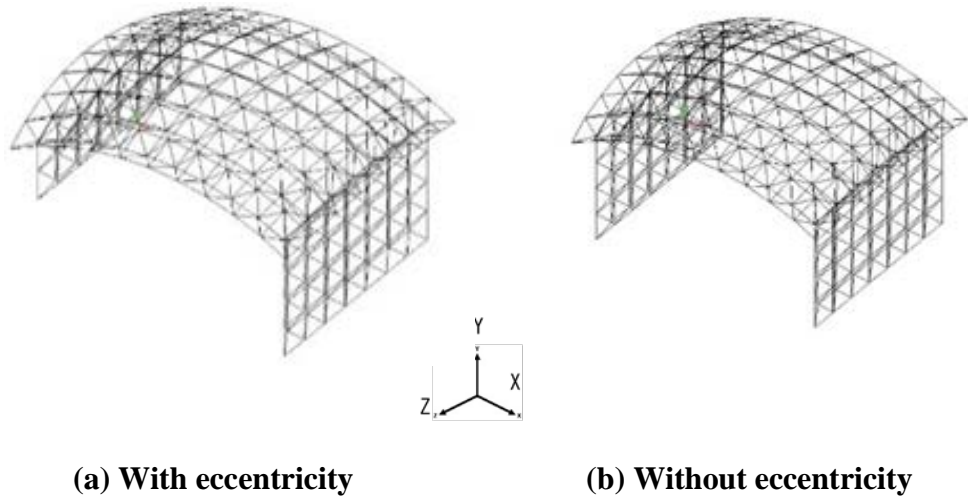
**Figure 1. Analysis model**
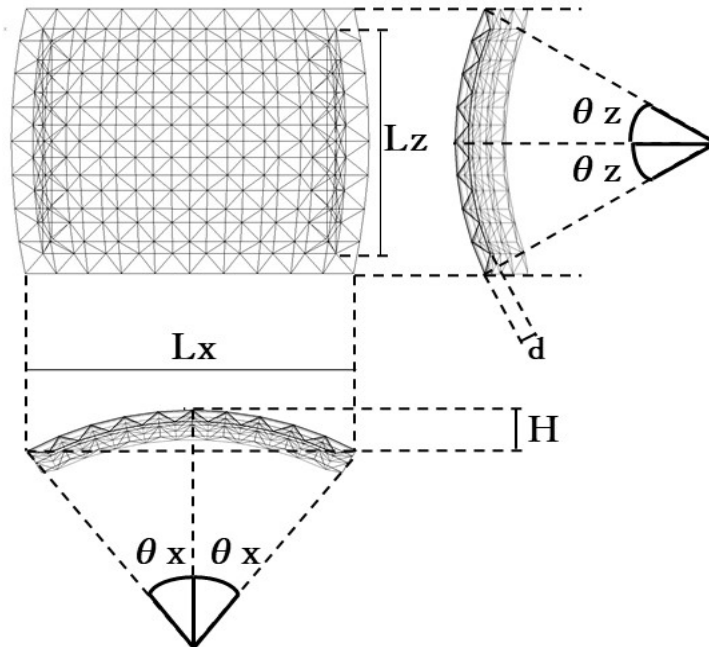


**Figure 2. Root type truss EP dome**

**Table 1. Structural form of analysis models**

| Structural classification | Ingredient direction | Analysis motel | | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| Roof structure | X direction | Half-open angle:$\theta_x$ | (deg) | 30 | 30 | 30 |
| | | Radius of curvature of the first chord: $R_{x1}$ | (m) | 17.882 | 12.534 | 10.234 |
| | | Radius of curvature of the first chord: $R_{x2}$ | (m) | 16.882 | 11.534 | 9.234 |
| | | Span: $L_x$ | (m) | 16.882 | 16.312 | 15.994 |
| | Z direction | Half-open angle: $\theta_z$ | (deg) | 30 | 45 | 50 |
| | | Radius of curvature of the first chord: $R_{zl}$ | (m) | 15.000 | 15.000 | 15.000 |
| | | Radius of curvature of the first chord: $R_{z2}$ | (m) | 14.000 | 14.000 | 14.000 |
| | | Rise: H | (m) | 3.923 | 5.019 | 6.085 |
| | | Layer spacing: d | (m) | 1.000 | 1.000 | 1.000 |
| Wall structure | Y direction | Height direction length: $L_y$ — With eccentricity | (m) | 4.000 | 4.000 | 4.000 |
| | | Height direction length: $L_y$ — Without eccentricity | (m) | 8.000 | 8.000 | 8.000 |
| | Z direction | Column direction span: $L_z$ | (m) | 14.000 | 14.000 | 14.000 |
| | | Layer spacing: d | (m) | 1.000 | 1.000 | 1.000 |
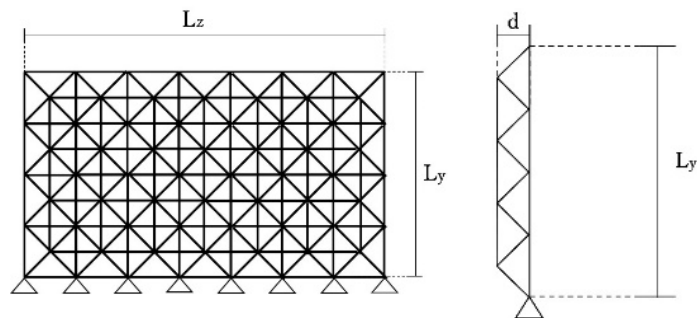


**(a) Wall model F**
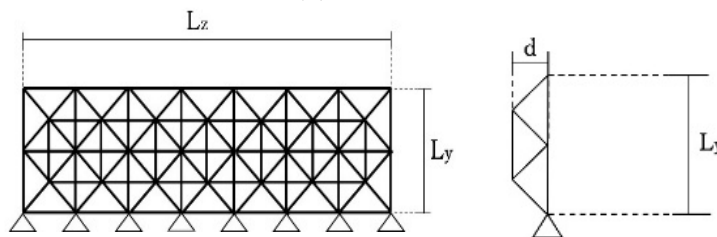


**(b) Wall model H**

**Figure 3. Wall type truss walls**

**Used members in the structures**

The used member characteristics such as the cross-sectional area and the slenderness ratio are determined by means of the allowable stress calculation conventional method. The snow load and dead load are used as the design loads of the analysis models.

The analysis models are assumed to be built in a heavy snow area. The deepest snow load of 2m with the unit volume weight of snow 30N / m$^2$ / cm is used as the snow load for the long-term structural design. The uniform distributed weight for seismic response analysis is taken to be 35% (0.7m) of the maximum depth (2m). The slenderness ratio of all the members such as the upper chord, lower chord, and web is taken to be 60 in the wall structures. The cross-sectional area of the members of the wall structures is determined by long-term allowable stress calculation, as shown in Table 2.

The member slenderness ratio of the dome structure is taken to be 80 at the upper and the lower chord members and 100 at the web members respectively. The member sectional area is shown in Fig. 4.
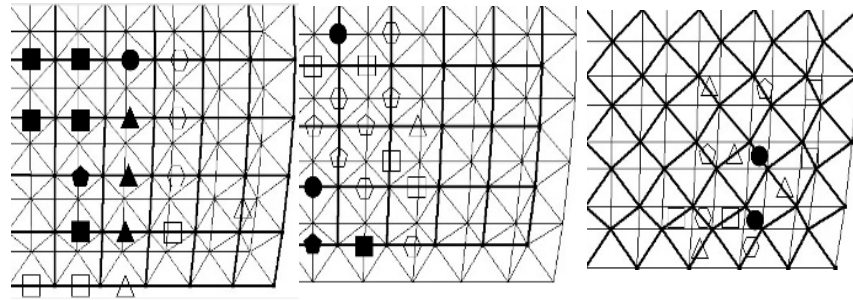


**(a) Upper chords members (b) Lower chord members (c) Web members**

| | | | |
|---|---|---|---|
| ———— | 700mm$^2$ | —●— | 2000mm$^2$ |
| —△— | 800mm$^2$ | —▲— | 2300mm$^2$ |
| —▱— | 1000mm$^2$ | —■— | 2700mm$^2$ |
| —⬠— | 1200mm$^2$ | —⬠— | 3000mm$^2$ |
| —⬡— | 1600mm$^2$ | —⬢— | 3200mm$^2$ |
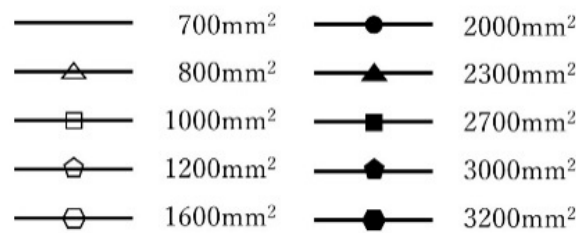
**(D) Legend of member cross-sectional area**

**Figure 4. Cross-sectional area of roof members**

**Table 2. Area of roof and wall structures, cane supported with roof and wall**

| | Semi-closed angle | Used members for wall structure | | | Cane members |
|---|---|---|---|---|---|
| | | Upper and lower chord members | Web members | Slenderness ratio 60 members | |
| Cross-sectional area of used members (mm$^2$) | 30 degrees | 1600 | 2700 | 3200 | 2700 |

## Comparison between elastic dynamic response analyses and response spectrum static analyses by using static seismic forces

The seismic force distribution and deformation properties of this truss structure will be analyzed and investigated based on the elastic seismic response analysis and the static analysis using the static seismic force calculated from the response spectrum method. The horizontal ground motion is acting for the longitudinal direction.

*Seismic force distribution and deformation property based on seismic force distribution of analysis model 1 without eccentricity*

Fig. 5 shows the static seismic force using the 6th-order mode (natural period $T_6$=0.146s) of the analysis model without eccentricity. For the acceleration response spectrum, the observed wave phase of El-Centro 1940 and the announced spectrum were set as the target spectra. The reason for adopting the 6th-order mode is that the effective mass ratio is 0.95, which is the largest value in the direction along the wall. The mode shape appears a shear deformation type of the wall.
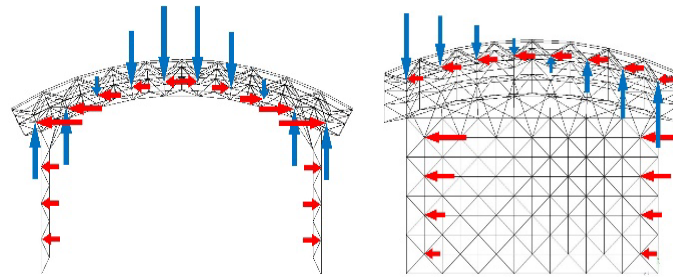
**(a) Seismic force distribution (x-y plane) (b) Seismic force distribution (y-z plane)**

**Figure 5. Distribution map of static seismic force**

The following seismic force distribution is recognized from this figure. That is, in the distribution (y-z plane), the roof response acceleration applies the seismic force in the vertical direction in a symmetrical distribution, and the walls response acceleration applies the seismic force in the out-of-plane direction due to the thrust from the roof. In the distribution (x-y plane), the roof receives a vertical seismic force in inverse symmetry, and the walls response acceleration applies a horizontal seismic force with an inverted triangular distribution. Regarding the seismic force acting on the roof cross section in the direction from the center point of the wall in the girder direction, the seismic response analysis and static seismic force are normalized and shown in Fig. 6 (a). The seismic force obtained by seismic response analysis was calculated as follows. That is, at the time when the maximum response acceleration in the girder direction of the center point of the upper surface of the dome became maximum, the response accelerations of all the nodes were obtained from the time history response analysis. Next, the seismic force is calculated by multiplying the mass of each node by the response acceleration. It can be seen that the seismic response analysis and static analysis have relatively close distributions from Fig. 6 (a).

Deformation properties in Fig. 6 (b) show the deformation diagram by the seismic response analysis, and Fig. 6 (c) shows a deformation diagram of static analysis by applying static seismic force. From this, the following can be confirmed for the roof structure. That is, in static analysis, it floats forward in the distribution (y-z plane). On the other hand, a deformation that sinks behind the girder appears. In the seismic response analysis, the same deformation is seen, but compared to the static analysis, the roof plate is smoothly deformed into a rigid body. The following was confirmed for the wall structure. That is, the outside of the wall surface is deformed due to the thrust from the roof structure in the girder direction. On the other hand, it was confirmed that the shear deformation property was shown in the wall direction.
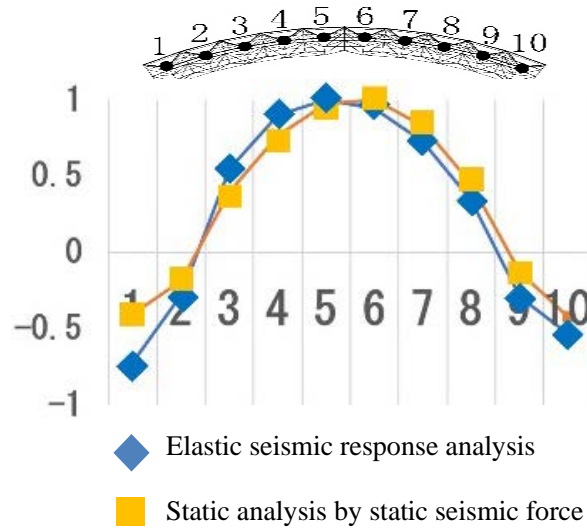
**(a) Normalized seismic force distribution**
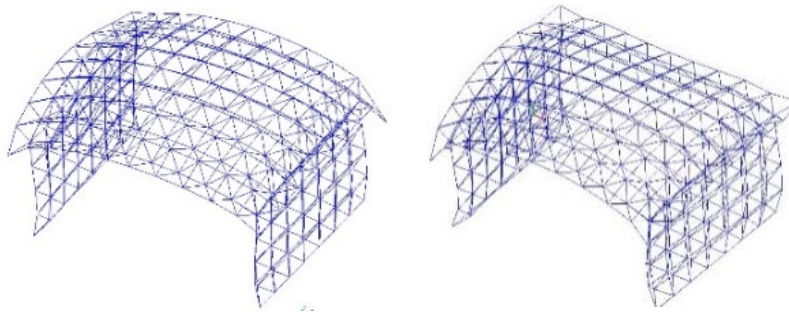


**(b) Seismic response analysis    (c) Static analysis by static seismic force**

**Figure 6. Comparison of seismic force distribution acting on the roof and deformation**

*Investigation of deformation properties of analysis model with and without eccentricity*

The deformation properties of the reverse object can be confirmed on the roof with eccentricity from the comparison of static analysis with and without eccentricity (Fig. 7).
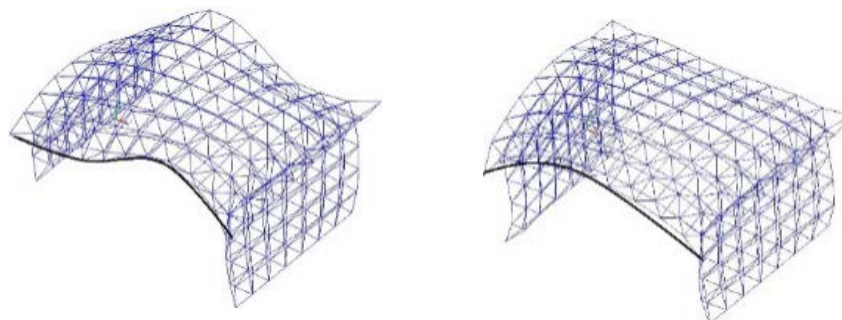


**(a) With eccentricity (b) Without eccentricity**

**Figure 7. Comparison of deformation properties with and without eccentricity**

## Conclusions

The equivalent static seismic forces and deformations are investigated by means of the static equivalent modeling of dynamic seismic forces for realizing earthquake resistant spatial truss structures. It has been confirmed that the proposed method shows a good agreement with the dynamic analysis of the spatial truss structures subjected to earthquake motion.

Whether there is eccentricity or not will affect the seismic response of the roof structure. The eccentricity analysis model can clarify the seismic response of the wave caused by the roof structure through the rigidity difference of the wall. In addition, no matter whether it is eccentric or not, the roof structure will induce the wall seismic force due to the vertical vibration. The wall structure also vibrates in the out of plane direction. The out of plane bending deformation has been occurred due to the force.

The seismic force distribution could be obtained using the mode with the maximum effective mass ratio in the case of this structure with a span of about 17 m. It was also confirmed that the out-of-plane response of the lower wall structure was affected by the thrust of the vertical vibration of the roof-type dome.

### References

[1] Ishikawa, K. (2009) Effects of resonance between spatial structures and ceiling systems on the seismic response", Proceedings of the International Association for Shell and Spatial Structures (IASS) Symposium 2009, Valencia Evolution and Trends in Design, Analysis and Construction of Shell and Spatial Structures Universidad Politecnica de Valencia, Spain Alberto DOMINGO and Carlos LAZARO (eds.).

[2] Ishikawa, K., Kato S. (1997) Elastic-plastic dynamic buckling analysis of reticular domes subjected to earthquake motion, *International Journal of Space Structures*, **12**, 205-215.

[3] Ishikawa K, Okubo S, Hiyama , Kato S. 2000. Evaluation method for predicting dynamic collapse of double layer lat-ticed space truss structures due to earthquake motion. In-ternational Journal of Space Structures, Vol.15, 249–257.

[4] Okubo S, Hiyama Y, Ishikawa K, Wendel R, Fischer L. (2001) Load capacity and plastic deformable ability of aluminum alloy double layer latticed wall subjected to plane load, IASS Symposium 2001, Nagoya, TP101.

[5] Ishikawa K, Kato S. (1997) Elastic-plastic buckling analysis of reticular dome subjected to earthquake motion. Interna-tional Journal of Space Structures, Vol.12, 205–215.

[6] Taniguchi Y, Gould P L, Kurano M. (2008) Earthquake input energy at dynamic collapse for double-layer cylindrical Lattice roofs. Journal of the IASS, Vol.49,No. 2.

[7] Fan F, Shen S Z, Parke G A R. (2004) Theoretical and experi-mental study of vibration reduction in braced domes using a viscous damper system. International Journal of Space Structures, Vol.19, No.4, 195-202.

[8] Midorikawa M. (2005) Performace-based seismic design provi-sions for buildings in Japan. Proceedings of the IASS 2005, Vol.I, 307-316.

[9] Giuliani G C. (2002) Overview on the dynamic control of struc-tures. Proceedings of the IASS 2002:561-567.

[10] Zeng Z P. 2007. Structural analysis and design of the latticed shell for Fujian Gymnasium. Journal of Spatial Structures, Vol.13, No.2, 44-48.

[11] Ilzarbe L. Álvarez, M.J. Viles E. Tanco M. (2008) Practical ap-plications of design of experiments in the field of engineer-ing. Abibliographical review, Qual. Reliab. Engng.Int., Vol.24, 417-428.

[12] European Committee for Standardization(CEN), Eurocode 8: Design of structures forearthquake resistance Part 1: Gen-eral rules. (2004) seismic actions and rules for buildings (EN 1998-1: 2004). Brussel.

[13] FEMA-356. (2000) NEHRP Guidelines for the Seismic rehabili-tation of buildings. Building seismic safety council, Wash-ington DC.

[14] Ishikawa, K. and Nagasaka, M. (2019) Evaluation method of performance to transmit horizontal seismic loads to substructures of double layer truss domes built in heavy snow region considering vertical load resistant capacity, *Journal of structural construction engineering, AIJ*, **84**, 1325-1335.

# An efficient reliability-based design optimization approach using PDF-based performance shift strategy

†*Z. Zhang[1], C. Jiang[1]

[1] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, College of Mechanical and Vehicle Engineering, Hunan University, Changsha, P. R. China, 410082

*Presenting author: zhangzhe0828@hnu.edu.cn
†Corresponding author: zhangzhe0828@hnu.edu.cn

**Abstract**

A probability density function (PDF)-based performance shift approach (PPSA) is proposed for sequential reliability-based design optimization (RBDO), which provides an effective tool for complex engineering design with consideration of uncertainties associated with design variables and simulation models. It converts the double-loop probabilistic optimization into a serial of cycles of reliability assessment and deterministic optimization through a performance shift strategy. The approach further proceeds by calculating the performance shift scalar and then solving a deterministic design optimization at each cycle, and eventually converges to the optimum design solution. The novelty of the proposed approach lies in two main aspects. First, the shift scalar is deducted for each probabilistic constraint in the response space, rather than in the design variable space, through which the violated constraints are moved towards the reliable region and the design solutions are improved progressively. Second, reliability analysis methods that calculate the moments or probability density functions, rather than most probable points or reliability indexes, are able to be integrated in the sequential RBDO framework, through which we could enhance the performance of existing sequential RBDO methods, especially the computational efficiency. Three numerical examples are investigated to demonstrate the effectiveness of the proposed approach, and the proposed approach is further utilized to perform crashworthiness design of a vehicle side impact problem.

**Keywords:** Reliability-based design optimization; Performance shift scalar; Probability density function; Dimension reduction method;

*Example - Crashworthiness design for vehicle side impact*
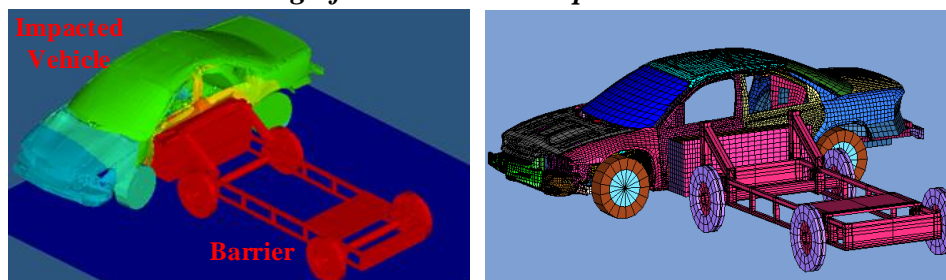


**Figure 1. A vehicle side impact problem**

Vehicle crashworthiness design is critical due to its extreme importance to guarantee the security of passengers. Among all the traffic accidents, vehicle side impact presently has been a principal factor leading to death of passengers, only second to frontal impact[1]-[3]. For a vehicle body, its two sides have a relatively weaker mechanical performance than the other parts. Vehicle side impact has a relatively small buffer area and hence its impact deformation is likely to cause severe injury of a passenger. Therefore, the crashworthiness design of a vehicle side impact problem is investigated, as shown in Fig. 1, in which the barrier impacts

the vehicle from the side with an initial velocity of 49.89kph. Generally, a crashworthiness design must satisfy internal and regulated side impact requirements specified by the vehicle market. In this study, the European Enhanced Vehicle-Safety Committee (EEVC) side impact criteria are used, as shown in Table 1, which include Head Injury Criterion (HIC), abdomen load, pubic symphysis force, VC's (viscous criteria), and rib deflections (upper, middle, and lower). Furthermore, the velocity of the B-pillar at the middle point and the velocity of the front door at the B-pillar should also meet regulated safety criteria in the vehicle side impact problems.

In the crashworthiness design of vehicle side impact, the optimization objective is to maintain or improve the vehicle safety performance as measured by the given safety rating criteria shown in Table 1, while minimizing the vehicle weight. Due to the existence of uncertainties in material properties, manufacturing processes and simulation models, etc., a reliability-based design optimization is required to obtain the reliable and robust crashworthiness design. The RBDO model is formulated as below[4]-[6]:

$$\min_{\boldsymbol{\mu_X}} \ \text{Weight}(\boldsymbol{\mu_X})$$

$$\text{s.t.} \ P\left(\text{Abdomen Load：} F_{\text{Abd}} \leq 1.0 \ \text{kN}\right) \geq R^t$$

$$P\left(\text{Upper/Middle/Lower Rib Deflection：} \text{Def}_{\text{rib\_l/rib\_m/rib\_u}} \leq 32 \ \text{mm}\right) \geq R^t$$

$$P\left(\text{VC}_{\text{upper/middle/lower}} \leq 0.32 \ \text{m/s}\right) \geq R^t$$

$$P\left(\text{Pubic Symphysis Force：} \text{Force}_{\text{pubic}} \leq 4.0 \ \text{kN}\right) \geq R^t \qquad (1)$$

$$P\left(\text{Velocity of B-pillar at Middle Point：} \text{Vel}_{\text{B-pillar}} \leq 9.9 \ \text{mm/ms}\right) \geq R^t$$

$$P\left(\text{Velocity of Front Door at B-pillar：} \text{Vel}_{\text{door}} \leq 15.7 \ \text{mm/ms}\right) \geq R^t$$

$$\boldsymbol{\mu_X^L} \leq \boldsymbol{\mu_X} \leq \boldsymbol{\mu_X^U}, \quad \boldsymbol{\mu_X} \in \mathbb{R}^9$$

In this model, ten probabilistic constraints are used to maintain the reliability of vehicle crashworthiness design. Nine random design variables are considered in the design optimization, which include thickness (X1–X7) and material properties (X8, X9) of critical vehicle parts, as shown in Table 2. $\boldsymbol{\mu_X}$ denotes the vector of mean values of the random design variables. Two random parameters are the barrier height and hitting position, which can vary from −30mm to 30mm according to the physical test.

A finite element analysis (FEA) model, as shown in Fig. 1, is established to compute the responses of the vehicle side impact problem. The model consists of 85941 shell elements and 96122 nodes. The computational time for one simulation of this FEA model using the RADIOSS software is approximately 20 hours on an SGI Origin 2000. Due to the high computational cost, the explicit response surface is generated for each vehicle impact response using the quadratic backward-Stepwise Regression method [4][7]:

$$\text{Weight} = 1.98 + 4.90x_1 + 6.67x_2 + 6.98x_3 + 4.01x_4 + 1.78x_5 + 2.73x_7$$

$$F_{\text{Abd}} = 1.16 - 0.3717x_2x_4 - 0.00931x_2x_{10} - 0.484x_3x_9 + 0.01343x_6x_{10}$$

$$\text{Def}_{\text{rib\_l}} = 46.36 - 9.9x_2 - 12.9x_1x_8 + 0.1107x_3x_{10}$$

$$\text{Def}_{\text{rib\_m}} = 33.86 + 2.95x_3 + 0.1792x_{10} + 5.057x_1x_2 - 11.0x_2x_8 - 0.0215x_5x_{10}$$
$$- 9.98x_7x_8 + 22.0x_8x_9$$

$$\text{Def}_{\text{rib\_u}} = 28.98 + 3.818x_3 - 4.2x_1x_2 + 0.0207x_5x_{10} + 6.63x_6x_9 - 7.7x_7x_8 + 0.32x_9x_{10}$$

$$\text{VC}_{\text{upper}} = 0.261 - 0.0159x_1x_2 - 0.188x_1x_8 - 0.019x_2x_7 + 0.0144x_3x_5$$
$$+ 0.0008757x_5x_{10} + 0.08045x_6x_9 + 0.00139x_8x_{11} + 0.00001575x_{10}x_{11}$$

$$\text{VC}_{\text{middle}} = 0.214 + 0.00817x_5 - 0.131x_1x_8 - 0.0704x_1x_9 + 0.03099x_2x_6 - 0.018x_2x_7$$
$$+ 0.0208x_3x_8 + 0.121x_3x_9 - 0.00364x_5x_6 + 0.0007715x_5x_{10}$$
$$- 0.0005354x_6x_{10} + 0.00121x_8x_{11} + 0.00184x_9x_{11} - 0.02x_2^2$$

$$\text{VC}_{\text{lower}} = 0.74 - 0.61x_2 - 0.163x_3x_6 + 0.001232x_3x_{10} - 0.166x_7x_9 + 0.227x_2^2$$

$$\text{Force}_{\text{pubic}} = 4.72 - 0.5x_4 - 0.19x_2x_3 - 0.0122x_4x_{10} + 0.009325x_6x_{10} + 0.000191x_{11}^2$$

$$\text{Vel}_{\text{B-pillar}} = 10.85 - 0.674x_1x_2 - 1.95x_2x_8 + 0.02054x_3x_{10} - 0.0198x_4x_{10} + 0.028x_6x_{10}$$

$$\text{Vel}_{\text{door}} = 16.45 - 0.489x_3x_7 - 0.843x_5x_6 + 0.0432x_9x_{10} - 0.0556x_9x_{11} - 0.000786x_{11}^2 \tag{2}$$

PPSA and SORA [8] are used to perform reliability-based crashworthiness design of the vehicle side impact problem, the results of which are presented in Table 1 and Table 2, respectively. It can be found that PPSA obtains the optimal design variables $\mu_X$ =(0.607, 1.329, 0.506, 1.5, 0.658, 1.5, 0.5, 0.345, 0.192), and at this case the vehicle weight is 25.90. On the contrary, SORA obtains the optimal design variables. $\mu_X$ = (0.5, 1.378, 0.5, 1.420, 0.691, 1.5, 0.5, 0.345, 0.192), and the vehicle weight is 25.40. In other words, PPSA and SORA obtains relatively close objective function values, e.g. vehicle weight, the deviation between which is only 1.93%. However, the computational cost of PPSA and SORA demonstrates a large deviation. PPSA requires only 2524 function evaluations to obtain the optimal design, while SORA requires 4913 function evaluations. The number of function evaluations of SORA is almost two times as much as that of PPSA. To make an intuitive comparison between PPSA and SORA, the convergence history of both methods is shown in Fig. 2. It is found that the convergence process of PPSA is much shorter than that of SORA, which also indicates the higher computational efficiency of PPSA. Therefore, for this vehicle side impact crashworthiness design, PPSA demonstrates a much higher computational efficiency than SORA, while it obtains a close vehicle weight as SORA.

**Table 1 EEVC regulations and requirements for vehicle side impact**

| Performance | | Unit | EEVC regulation | Given safety rating criteria | Initial design |
|---|---|---|---|---|---|
| Abdomen Load | | KN | 2.5 | $\leq 1.0$ | 0.663 |
| Rib Deflection | Upper | mm | 42 | $\leq 32$ | 28.5 |
| | Middle | | | | 39.0 |
| | Lower | | | | 34.0 |
| VC | Upper | m/s | 1.0 | $\leq 0.32$ | 0.22 |
| | Middle | | | | 0.21 |
| | Lower | | | | 0.31 |

| Pubic Symphysis Force | KN | 6 | $\leq 4.0$ | 4.067 |
| HIC | ~ | 1000 | $\leq 650$ | 229.4 |

**Table 2 Design variables and random parameters in the vehicle side impact model**

| | Symbol | Description | Distribution type | Standard deviation | Mean | Upper bound | Lower bound |
|---|---|---|---|---|---|---|---|
| Random design variable | $X_1$ | B-pillar inner thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_2$ | B-pillar thickness reinforce | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_3$ | Floor side inner thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_4$ | Cross member thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_5$ | Door beam thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_6$ | Door beltline thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_7$ | Roof rail thickness | Normal | 0.03 | 1 | 0.5 | 1.5 |
| | $X_8$ | B-pillar inner material | Normal | 0.006 | 0.3 | 0.192 | 0.345 |
| | $X_9$ | Floor side inner material | Normal | 0.006 | 0.3 | 0.192 | 0.345 |
| Random parameter | $X_{10}$ | Barrier height | Normal | 10 | 0 | | |
| | $X_{11}$ | Barrier hitting position | Normal | 10 | 0 | / | |

**Table 3 Convergence history of PPSA**

| Cycle | Design variable $\left(u_{X_1}, u_{X_2}, ..., u_{X_9}\right)$ | Objective function | Number of function evaluations |
|---|---|---|---|
| 0 | (1,1,1,1,1,1,1,0.3,0.3) | 29.05 | 0 |
| 1 | (0.680,1.415,0.5,1.456,0.762,1.5,0.5,0.345,0.192) | 26.80 | 706 |
| 2 | (0.690,1.291,0.5,1.497, 0.660,1.5,0.5,0.345,0.192) | 26.00 | 1312 |
| 3 | (0.609,1.327,0.504,1.5, 0.658,1.5,0.5,0.345,0.192) | 25.88 | 1918 |
| 4 | (0.607,1.329,0.506,1.5,0.658, 1.5,0.5,0.345,0.192) | 25.90 | 2524 |

**Table 4 Convergence history of SORA**

| Cycle | Design variable $\left(u_{X_1}, u_{X_2}, ..., u_{X_9}\right)$ | Objective function | Number of function evaluations |
|---|---|---|---|
| 0 | (1, 1, 1, 1, 1, 1, 1, 0.3, 0.3) | 29.05 | 0 |
| 1 | (0.5, 1.373, 0.5, 1.420, 0.683, 1.5, 0.5, 0.345, 0.192) | 25.35 | 1747 |
| 2 | (0.5, 1.378, 0.5, 1.420, 0.691, 1.5, 0.5, 0.345, 0.192) | 25.40 | 3381 |
| 3 | (0.5, 1.378, 0.5, 1.420, 0.691, 1.5, 0.5, 0.345, 0.192) | 25.40 | 4913 |
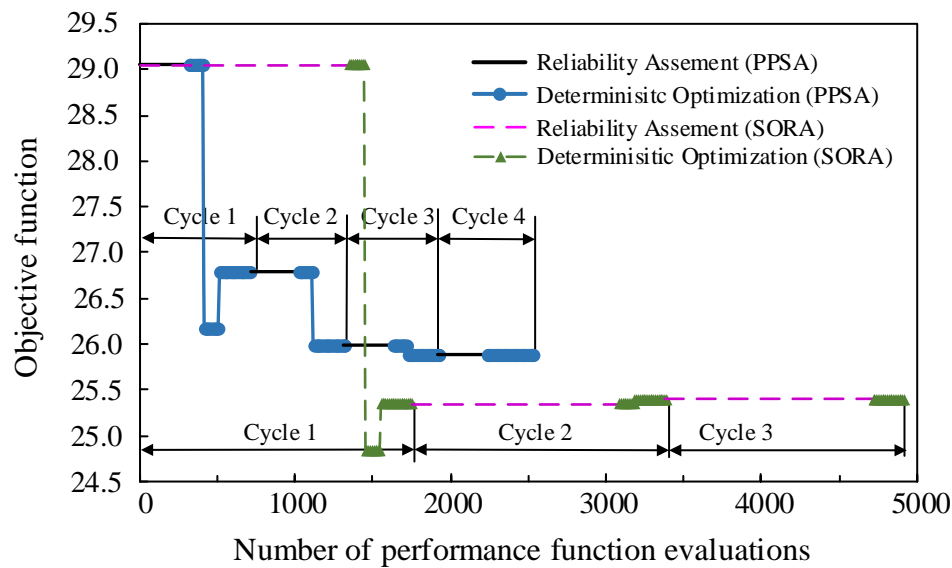


**Figure 2.  Convergence comparison between PPSA and SORA**

### References

[1] Gu L., Yang R-J., Tho C., et al (2001) Optimisation and robustness for crashworthiness of side impact. International *Journal of Vehicle Design*, 26 (1): 348-359, 2001.

[2] Youn B. D., Choi K. K. (2004a) A new response surface methodology for reliability-based design optimization. *Computers & Structures*, 82 (2-3): 241-256, 2004.

[3] Chakraborty S., Chatterjee T., Chowdhury R., et al (2017) Robust Design Optimization for Crashworthiness of Vehicle Side Impact. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 3 (3): 031002, 2017.

[4] Youn B. D., Choi K. K. (2004b) Yang R J, et al. Reliability-based design optimization for crashworthiness of vehicle side impact. *Structural and Multidisciplinary Optimization*, 26 (3-4): 272-283, 2004.

[5] Sinha K. (2007) Reliability-based multiobjective optimization for automotive crashworthiness and occupant safety. *Structural and Multidisciplinary Optimization*, 33 (3): 255-268, 2007.

[6] Goswami S., Chakraborty S., Chowdhury R., et al (2019) Threshold shift method for reliability-based design optimization. *Structural and Multidisciplinary Optimization*, 60 (5): 2053-2072, 2019.

[7] Yi P., Cheng G., Jiang L. (2008) A sequential approximate programming strategy for performance-measure-based probabilistic structural design optimization. *Structural Safety*, 30 (2): 91-109, 2008.

[8] Du X., Chen W. (2004) Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design. *Journal of Mechanical Design*, 126 (2): 225-233, 2004.

# Application of a 2D adaptive mesh refinement method to the flow over wall-mounted plate

## Zhenquan Li

School of Computing and Mathematics, Charles Sturt University, Australia.

Presenting author: jali@csu.edu.au
Corresponding author: jali@csu.edu.au

## Abstract

After verifying the accuracy of the 2D adaptive mesh refinement (AMR) method by the benchmarks of 2D lid-driven cavity flows and 2D backward facing step flows, this paper applied the AMR to simulate the flow over wall-mounted plate. 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 application of the AMR produces the estimated coordinates of the vortex center after the plate after the AMR applied once to the initial mesh. 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, flow over wall-mounted plate

## Introduction

When solving differential equations numerically, we find the values of the unknowns only at finite number of nodes due to limitation of computational capacity. Therefore, the selection of the finite number of nodes is one of the important issues for finding the accurate numerical solutions. The AMR adds more nodes in the domain where necessary based on prespecified criteria.

The AMR is a computational approach to increase the accuracy of numerical solutions of differential equations with low computational cost. There have been many publications on AMRs and their applications since the seminal work by Berger [4]-[5]. Some AMR methods take local truncation errors as a refinement criterion (e.g. [1][22][3][6]) and mass residuals as the criterion for mesh refinement (e.g. [7]). Henderson [9] introduced a method which refines everywhere that the solution gradients are large and two methods that use two different estimates of the approximation error as refinement criteria. Overall, these AMR methods based on numerical criteria aim to obtain a balance between the numerical accuracy and the computational cost in finding numerical solutions of differential equations.

Even though there might be some relations between the refined meshes obtained using numerical criteria and the velocity fields, no information may not be provided on the characteristics of the flow field for some cases such as locations of the centres of vortices [19]. The characteristics of velocity fields are hidden in the fields themselves. Therefore, if we want to include the characteristics of a velocity field in a mesh, we must generate the mesh based on the information from both the geometry of the domain and the velocity field.

We developed a theory for streamline tracking and AMR methods based on the qualitative theory of differential equations [12]. We have verified theory using the accuracy locations of singular points, asymptotic lines (planes for 3D) and drawing closed streamlines [11][13][14]. The AMR methods refine meshes based on information of calculated velocity fields. We have verified accuracy of these methods using 2D [16] and 3D analytical velocity fields [15], the benchmark of lid-driven cavity flow [10][17]-[19], the benchmark of backward-facing step flow [22], and 2D unsteady flow past a square cylinder [21]. We also investigated the computational complexity of the 2D AMR using lid-driven cavity flow and obtained positive outcomes [20].

Mesh refinement is necessary for obtaining accurate numerical velocity fields since different intensities of vortices requires different densities of mesh nodes [17]. The same conclusion was obtained by Armaly et al. [2]. We conducted a detailed study starting from relatively coarse initial meshes and demonstrating that the centers of vortices were captured within the refined cells of once refined meshes [18]. We applied the AMR method twice to the initial meshes and the twice refined meshes show that centers of the vortices are held within the twice refined cells [19].

This paper applied the 2D AMR we proposed to the flow over wall-mounted plate using open source software Navier2D for calculating the numerical velocity fields. We refine the initial mesh once and report the coordinates of centres of the vortices after the plate.

**Governing equations and domain**

This paper considers the 2D incompressible or steady flow over wall mounted plate. 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 [8], 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 channel is 3. The plate is a unit square. We choose the length of the upstream channel as $m$ 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 $n$.
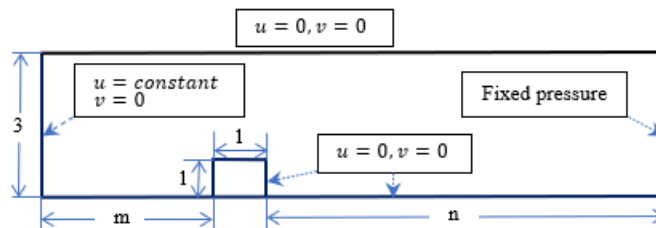


**Figure 1. Computational domain and boundary conditions**

The boundary conditions at the solid wall are $u = v = 0$, and fixed pressure at the outlet selected from Navier2D.

## Review of AMR Method

This section summarizes the 2D AMR method proposed by Li [16] based on the theory developed from qualitative theory of differential equations [11].

Assume that $\boldsymbol{V}_l = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ 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 $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{J}$ 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 [14]). The introduction of functions $f$ reduces the number of refined cells in refined meshes dramatically [12].

**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 [16]. 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 [19] and set $T = 1$, that is, we subdivide the cells on which one of the MC conditions is satisfied once.

**Results**

The residuals of the numerical velocity fields calculated in this study are less than $10^{-6}$ for both $u$ and $v$. Since the patterns of flows become more complex when Reynolds number is getting bigger, we choose the initial mesh with the same step size in both $x$ and $y$ directions for one case but different sizes for some different cases. The initial velocity field is $u = constant$ and $v = 0$. We show the profiles of $u$ of the calculated numerical velocity fields and the exact profiles in the same coordinate system. We obtain the information about the accuracies of the calculated fields from the comparisons of these profiles. The refined meshes show the accuracy of the AMR method. We record the coordinates of the center of the isolated refined cell as the estimate the locations of vortices after the plate.

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.

$Re = 10$

We chose the step size $1/6$ in the both directions and upstream length $m = 60$. Fig. 2 shows the comparison between the profile of the horizontal component $u$ of the exact analytical velocity and the calculated one on the initial mesh just before the plate. From this figure, we understand the flow has developed very well before the plate.



**Figure 2. Comparison of the profiles of $u$ before the plate on initial mesh for $Re = 10$ between the current study and the analytical solution**
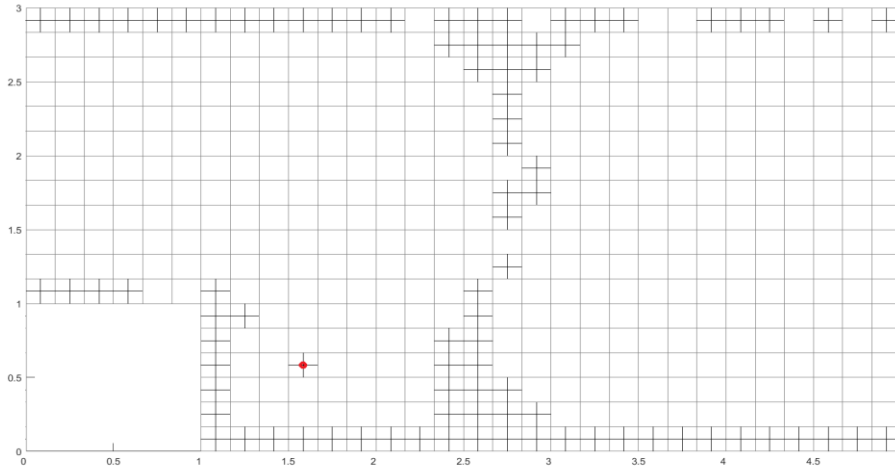


**Figure 3. Once refined mesh for $Re = 10$ with the estimated location of vortex center (red dot)**

Fig. 3 shows one isolated refined cell with a red dot in the middle after the plate. We take the coordinates of the red dot as the estimated location of the center of vortex after the plate.

$Re = 50$

We chose the same step size $1/6$ in the both directions but different upstream length $m = 300$ from the case of $Re = 10$. Fig. 4 shows the comparison between the profile of the horizontal

component $u$ of the exact analytical velocity and the calculated one on the initial mesh just before the plate. This figure demonstrates the flow has developed well before the plate.
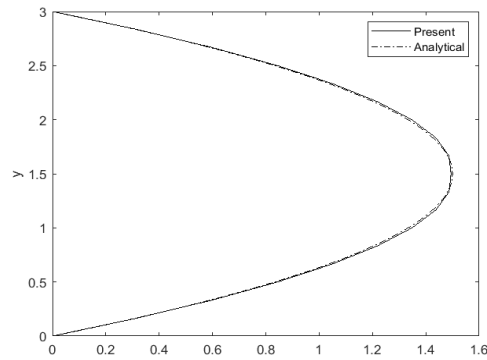


**Figure 4. Comparison of the profiles of $u$ before the plate on initial mesh for $Re = 50$ between the current study and the analytical solution**
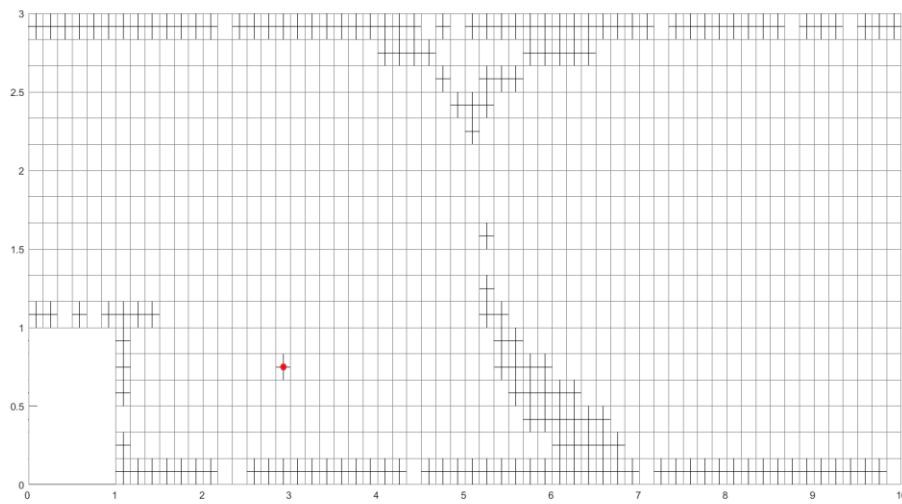


**Figure 5. Once refined mesh for $Re = 50$ with the estimated location of vortex center (red dot)**

Fig. 5 shows again one isolated refined cell with the red dot at the center after the plate. Again, we take the coordinates of the red dot as the estimated location of the center of vortex after the plate. Both horizontal and vertical coordinates of the red dot have increased.

$Re = 100$

We chose the step size $1/8$ in the both directions but different upstream length $m = 330$ from the cases of $Re = 10$ and $50$. Fig. 6 shows the comparison between the profile of the horizontal component $u$ of the exact analytical velocity and the calculated one on the initial mesh just before the plate. Even though we reduced the step size and increased the upstream length, the profile of calculated horizontal velocity component $u$ has not developed very well before the plate. There is a slight difference between the two profiles.
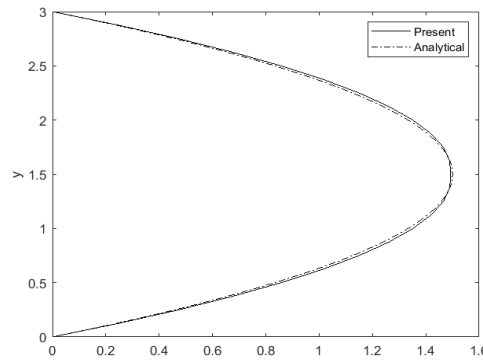
**Figure 6. Comparison of the profiles of $u$ before the plate on initial mesh for $Re = 100$ between the current study and the analytical solution**
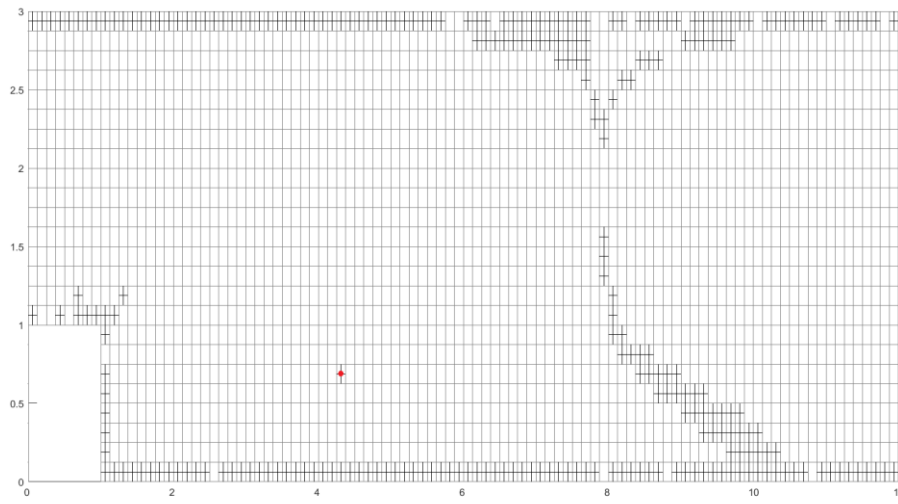


**Figure 7. Once refined mesh for $Re = 100$ with the estimated location of vortex center (red dot)**

There is one isolated refined cell with a red dot at the center after the plate as shown in Fig. 7. We take the coordinates of the red dot as the estimated location of the center of vortex after the plate. The horizontal coordinate increased more than one unit than that of the corresponding coordinate for $Re = 50$ but the vertical coordinate increased slightly.

$Re = 200$

We chose the same step size $1/8$ in the both directions as we did for $Re = 100$ but different upstream length $m = 570$ from all three cases above. Fig. 8 shows the comparison between the profile of the horizontal component $u$ of the exact analytical velocity and the calculated one on the initial mesh just before the plate. This figure demonstrates that the flow has developed worse than the one for $Re = 100$ before the plate even though we have increased the upstream length.
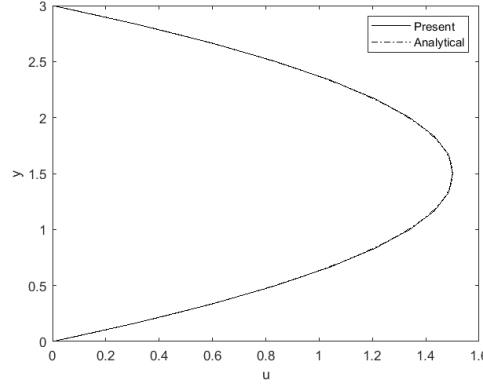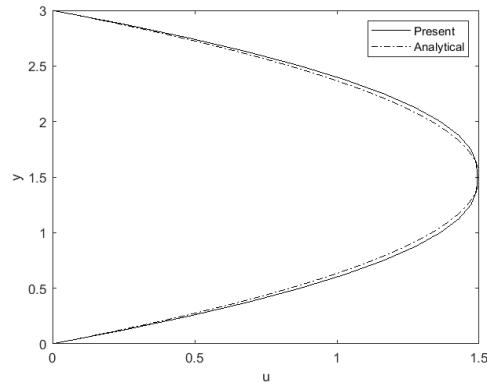
**Figure 8. Comparison of the profiles of $u$ before the plate on initial mesh for $Re = 200$ between the current study and the analytical solution**
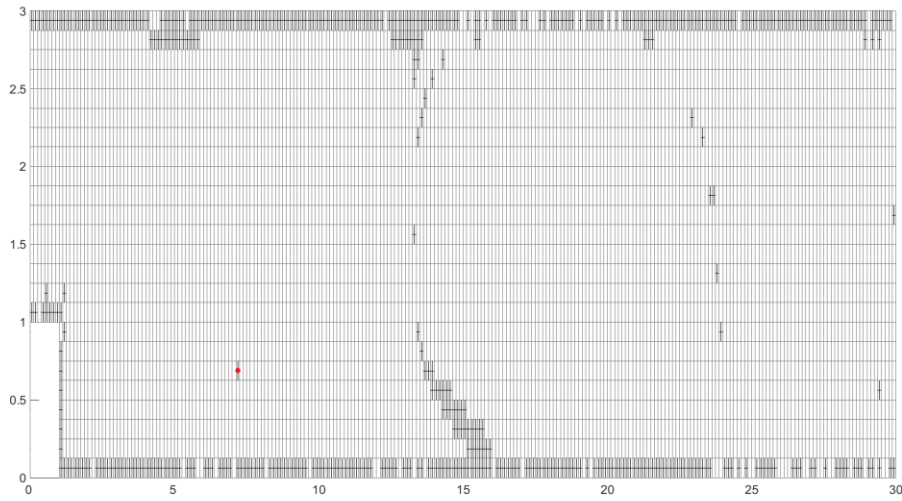


**Figure 9. Once refined mesh for $Re = 200$ with the estimated location of vortex center (red dot)**

Fig. 9 shows one isolated refined cell with red dot at the center again after the plate. We again take the coordinates of the red dot as the estimated location of the center of vortex after the plate. Comparing with the coordinates of the red dot for $Re = 100$, the horizontal coordinates increased dramatically but not the double as the Reynolds numbers. Since we chose the same step size, the vertical coordinate should increase slightly but this is not shown in Fig. 9.

**Estimated center locations**

Table 2 presents the estimated center locations of vortices (red dots) for $Re = 10, 50, 100,$ and 200.

**Table 2. Estimated locations of vortices centers**

| Reynolds number | Coordinates of estimated center |
|---|---|
| 10 | (1.58, 0.5833) |
| 50 | (2.90, 0.7500) |
| 100 | (4.30, 0.6875) |
| 200 | (7.20, 0.6875) |

## Discussions

This paper applied an AMR proposed based on qualitative theory of differential equations to the flow over wall-mounted plate. The refined meshes provided information about the vertex's centers after the plate and other information such as the attachment. Unlike the center of vortices, we cannot provide the estimated location of attachment from the once refined mesh. As we showed in lid-driven cavity flow [19], more refinements are required for more accurate estimated location of the centers of vortices. The other information such as the estimated locations of attachments may be clearly shown in the refined meshes. The software used in this research limited more accurate results to be produced at this stage.

## References

[1] 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.

[2] 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.

[3] 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.

[4] Berger, M. and Oliger, J. (1984) Adaptive mesh refinement for hyperbolic partial differential equations, *J. Comput. Phys.*, 53, 484-512.

[5] Berger, M. and Colella, P. (1989) Local adaptative mesh refinement for shock hydrodynamics, *J. Comput. Phys.*, 82, 64—85.

[6] 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.

[7] Dietrich, J.C., Kolar, R.L., and Dresback, K.M. (2008) Mass residuals as a criterion for mesh refinement in continuous Galerkin shallow water modles, *J. Hydraul. Eng.*, **134**, 520—532.

[8] Engwirda, D. (2006) Navier-Stokes Solver (Navier2d), MATLAB Central File Exchange. Last accessed May 2006.

[9] Henderson, R.D., Adaptive spectral element methods for turbulence and transition, *High-order methods for computational physics*, 1999, Barth, T.J., Deconick, H., Eds, 225—324, Springer, Berlin.

[10] 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.

[11] Li, Z. (2002) A mass conservative streamline tracking method for two-dimensional CFD velocity fields, *J. Flow Visual. Image Process.* **9**, 75--87.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[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., 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.

[20] Li, Z. (2017) Computational complexity of the algorithm for a 2D adaptive mesh refinement method using lid-driven cavity flows, *Comput. Therm. Sci*. **9**, 395--403.

[21] 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.

[22] Li, Z. and Li, M. (2021) Accuracy verification of a 2D adaptive mesh refinement method using backward-facing step flow of low Reynolds numbers, *Int. J. Comput. Methods.* **18**, https://doi.org/10.1142/S0219876220410121

# Two-dimensional meso-scale simulation of hydraulic fracture in concrete

## *Kelai Yu[1], †Zhenjun Yang[2], Xin Zhang[1] and Guohua Liu[1]

[1] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310012, China

[2] Hubei Key Laboratory of Geotechnical and Structural Safety, School of Civil Engineering, Wuhan University, Wuhan, 430072, China

*Presenting author: kelaiyu93@zju.edu.cn
†Corresponding author: zhjyang@whu.edu.cn

**Abstract**

A meso-scale finite element (FE) modelling approach is developed for meso-scale hydraulic fracture of concrete in this study. The FE models, consisting of aggregates, mortar and pores, are first generated by both random aggregate generation and packing and direct conversion of real micro-CT images. Pore pressure cohesive interface elements are then inserted into solid FE meshes by a developed Python code. The modeling approach was validated by hydro-fracturing experiments of concrete cubes. The results show that the developed models can reliably simulate complicated hydraulic fracture process, and the contents of aggregates and pores have significant effects on the hydraulic fracture resistance and the crack patterns.

**Key words**: Concrete; meso-scale finite element model; cohesive element; hydraulic fracture; X-ray Computed Tomography

**Introduction**

Concrete is widely used in hydraulic engineering structures, such as canal, gravity dams, conduits, tunnels, and port structures. As a typical quasi-brittle material, concrete is susceptible to tensile fracture, which can be exacerbated by high fluid pressure. Hydraulic fracture experiments of concrete specimens [1,2] are limited because the specimens must be tightly sealed to make sure the high fluid pressure is applied. Conducting such tests for large-sized structures such as dams is unrealistic. In addition, as a heterogeneous material, the concrete meso-structure, including aggregates, interfaces and pores which are randomly distributed in the mortar, is expected to interact with the fluid flowing and thus have significant effects on the hydraulic cracking behavior. Although numerical simulations, often based on the finite element (FE) method, are very promising and being increasingly used to investigate hydraulic fracturing in concrete, most of them have ignored the heterogeneous meso-structure of concrete [3-5].

In this study, 2D meso-scale FE models for hydraulic fracture of concrete are developed. Two approaches are used to generate the meso-scale concrete models. One is the random aggregate generation and packing approach, and another directly converts micro X-ray CT images into FE models. The zero-thickness cohesive elements coupling fluid pressure freedoms are pre-inserted within mortar and between the aggregate-mortar interfaces to model complicated hydraulic fracture processes, using an in-house Python code developed. Concrete cubes were simulated and the results were well compared with experiments available in the literature. The effects of the aggregate volume fraction and the porosity were analyzed by parametric simulations.

**Numerical method**

*Fluid flow within the crack*

The fluid flow inside a crack is idealized as the Newtonian fluid with no fluid-lag, and the Poiseuille flow equation is used

$$q = -k_t \nabla p_f \tag{1}$$

where $q$ is the volume flow rate, and $\nabla p_f$ is the fluid pressure gradient along the crack. $k_t$ is the tangential permeability calculated by the Reynolds equation for incompressible fluid

$$k_t = \frac{w_f^3}{12\mu} \tag{2}$$

where $w_f$ denotes the local crack width and $\mu$ is the viscosity of fluid.

The following equations are used to define the relationship between the fluid pressure and the flow rate in the direction normal to the crack surfaces

$$q_t = C_t \left( p_f - p_t \right)$$
$$q_b = C_b \left( p_f - p_b \right) \tag{3}$$

where $C_t$ and $C_b$ are fluid leak-off coefficients, $q_t$ and $q_b$ are the normal flow rate, and $p_t$ and $p_b$ are fluid pressures at top and bottom surfaces of the crack, respectively.

*Generation of FE models*

The mesoscale FE models are first generated with either random aggregates similar to [6] or from micro CT images similar to [7]. An algorithm similar to [8] is then devised to pre-insert cohesive interface elements with fluid pressure freedoms and implemented in a Python code. The building process of meso-scale models is illustrated in Fig. 1, where a typical node is connected with six quadratic-triangular elements in the initial mesh including an aggregate and mortar. This node is first replaced by five normal nodes (with displacement and pore pressure freedoms) and one special node with pore pressure freedom (shown by triangle in Fig. 1b) at the same position. Two COH2D4P elements with pore pressure freedoms are then generated along the element edges. COH2D4P elements are inserted both within the mortar and between the aggregate-mortar interfaces.
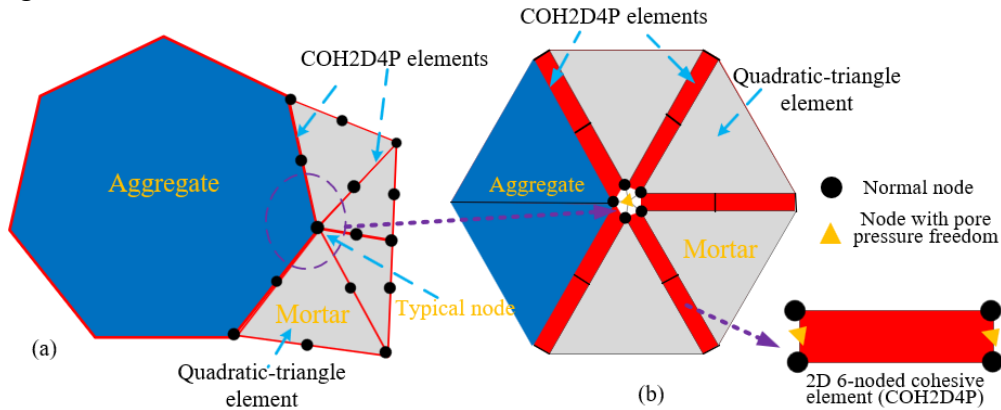


**Figure 1. Inserting process of pore pressure cohesive element**

**Simulation of hydro-fracturing experiments**

The cubic specimens of concrete (150mm size) with water injection in an initial central crack (as shown in Fig. 2a) were tested by Gan et al. [9] under different uniaxial loading ($P_u$). Fig. 2b presents a typical meso-scale FE model with a volume fraction of aggregates ($V_{agg}$) of 45% [9], built by the random generation and packing approach. The mesh consists of 12480 six-noded quadratic triangle elements with a maximum size of 2mm and 25706 COH2D4P elements. The aggregates and mortar are assumed to behave linear elastically. All the material properties are listed in Table 2.



| (a) A concrete cubic specimen (unit: mm) [9] | (b) 2D meso-scale model |
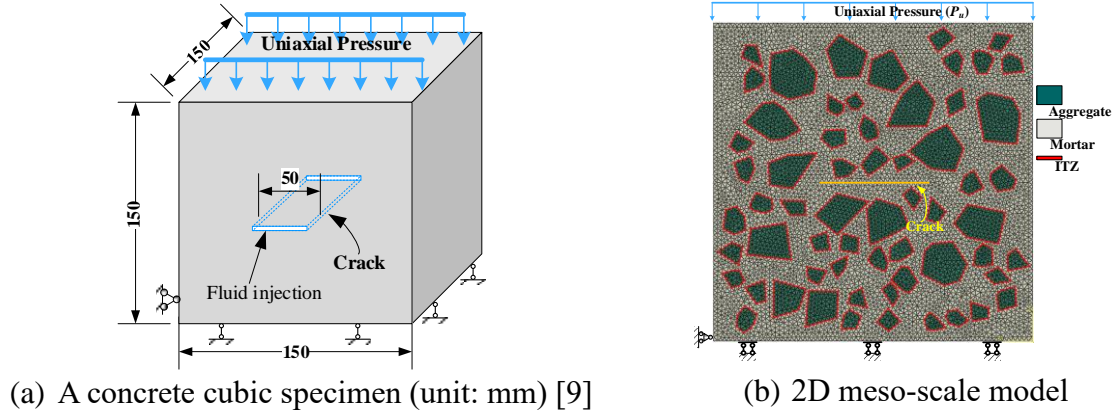
**Figure 2. Simulation of hydro-fracturing experiment: specimen and model**

**Table 2. Material properties for the meso-scale model**

| Properties | Mortar | Aggregate | Cohesive elements in mortar | Cohesive elements on interfaces | Fluid |
|---|---|---|---|---|---|
| Young's modulus $E$ (GPa) | 27.25Gpa | 70GPa | | | |
| Poisson's ratio $\nu$ | 0.2 | 0.2 | | | |
| Density $\rho$ ($10^{-9}$ tone/mm$^3$) | 2.5 | 2.7 | | | 1 |
| Fluid viscosity $\mu$ (MPa·s) | | | | | $10^{-9}$ |
| Injection rate $q$ (mm$^2$/s) | | | | | 0.2 |
| Fracture energy $G_n/G_s$ (N/mm) | | | 0.14 | 0.105 | |
| Tensile strength $f_t$ (MPa) | | | 2.01 | 1.51 | |
| Initial stiffness $K_n$ (MPa/mm) | | | $2.5\times10^4$ | $1.25\times10^4$ | |

Fig. 3 (a) shows the simulated crack path without loading on the boundary, indicating that the existence of aggregates affects the crack propagation considerably. The simulated peak fluid pressures under different uniaxial loadings are plotted in Fig. 3(b). The hydraulic fracture resistance increases almost linearly as the external compressive loading, which is in good agreement with the experimental results.
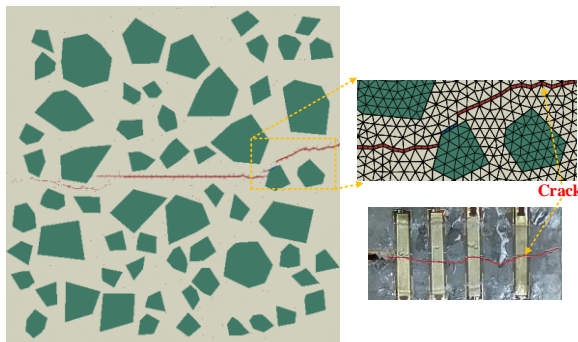


| (a) hydraulic fracturing process | (b) peak fluid pressure under different uniaxial loadings |
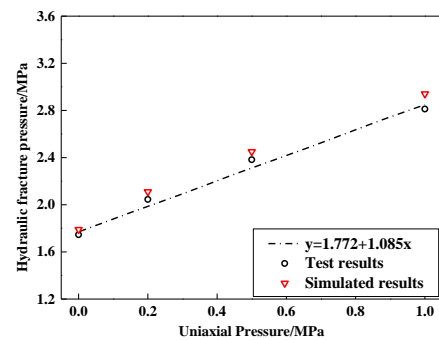
**Figure 3. The simulated crack pattern and peak fluid pressure**

Meso-models with $V_{agg}$=40%, 50% and 60% are generated and simulated to investigate its effects. Fig. 4a shows a model with element size of 2mm for $V_{agg}$=60%. The simulated crack pattern is presented in Fig. 4b. Figs 5a-c show the fluid pressure evolutions at the injection point for five random samples and the mean curve for the three $V_{agg}$, respectively. It can be seen that the peak fluid pressure increases by 22.5% from 1.78Pa for $V_{agg}$=40% to 2.18 MPa for $V_{agg}$=60%.



| (a) mesh | (b) crack pattern |
|---|---|
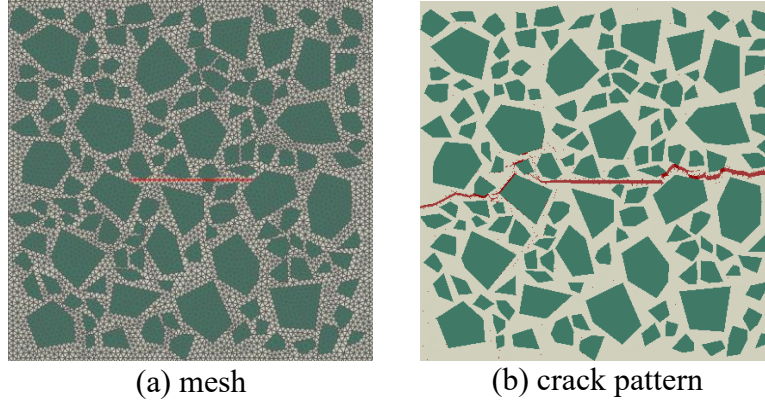
**Figure 4. Typical meso-scale model with aggregate volume fraction of 60%**



| (a) $V_{agg}$=40% | (b) $V_{agg}$=50% | (c) $V_{agg}$=60% |
|---|---|---|
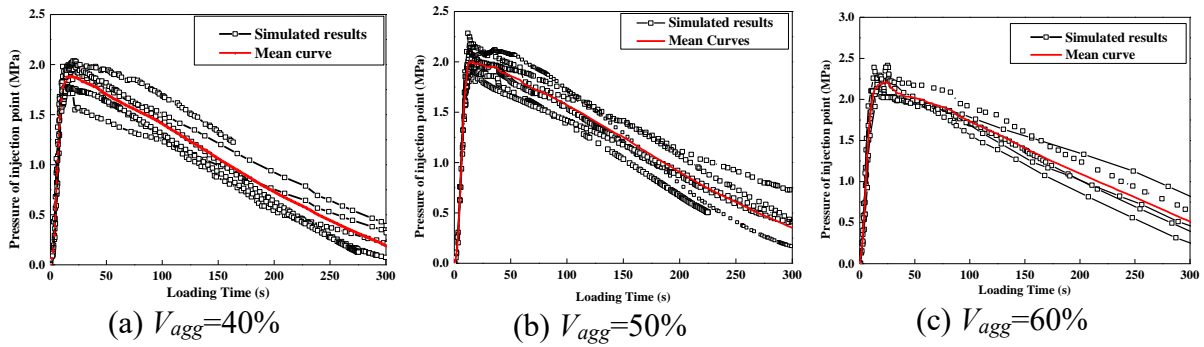
**Figure 5. Effects of aggregate volume fraction on the fluid pressure**

**X-ray CT image based meso-scale hydraulic fracturing modelling**

Fig. 6a shows a segmented X-ray tomography image of a 37.2mm cube tested by Yang et al [7,10]. One image slice with aggregates, mortar and pores (Fig. 6b) is modelled. The FE mesh in Fig. 6c has 23198 six-noded quadratic triangle elements and 33752 COH2D4P elements with 0.05-1mm size to accurately capture the geometry of aggregates and pores. In order to maintain the continuity of fluid flowing, the pores are modelled by solid elements with a high permeability coefficient of 0.5mm/s and a weak Young's modulus (10% of the mortar's). The same fluid properties listed in Table 2 and the same mechanical properties in [11] for aggregates and mortar are used. The Fluid is assumed to be injected into a central crack with an injection rate of 0.001 mm$^2$/s under the boundary conditions shown in Fig. 6c.
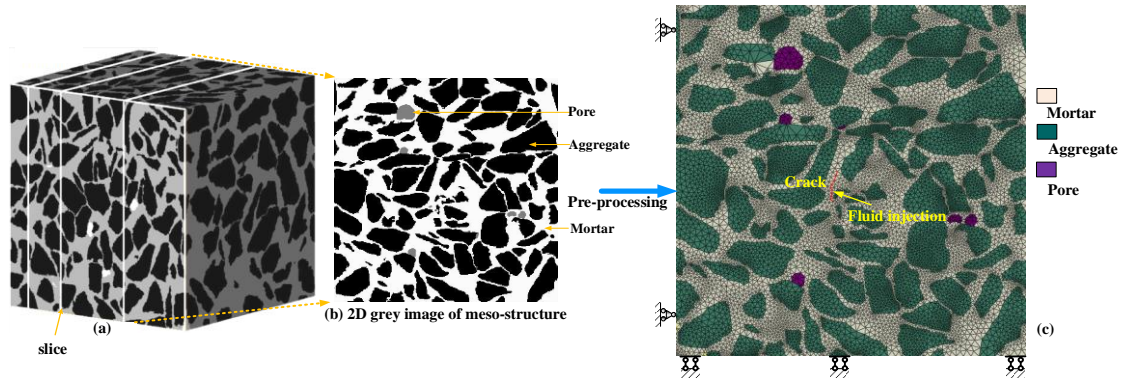
**Figure 6. The X-ray CT image-based model for an image slice**

Fig. 7a shows the simulated crack pattern at failure when a main crack reaches the top surface. To compare, the same image-based model without pores (with properties as the mortar) is also simulated with the crack pattern shown in Fig. 7b. It is clear that more crack bifurcations occur near the pores. The simulated fluid pressure evolutions are compared in Fig. 8, which indicates that the pores can considerably reduce the hydraulic fracture resistance.



(a) model with pores       (b) model without pores
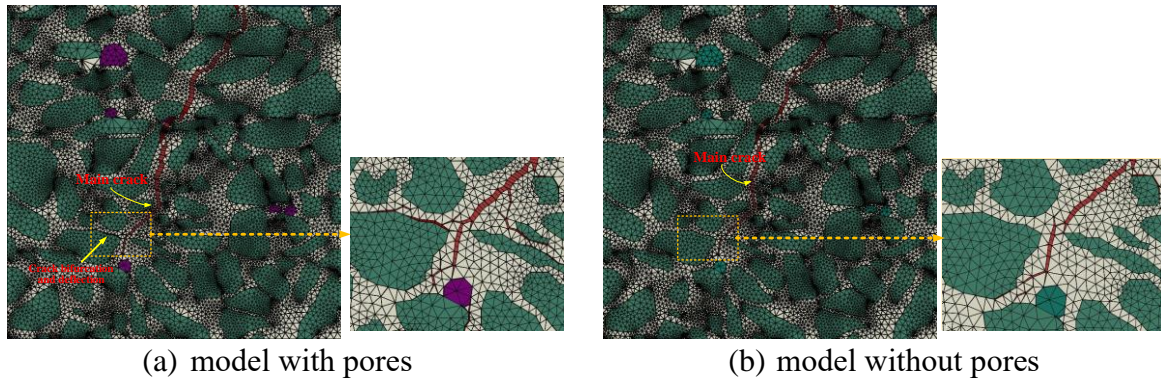
**Figure 7. The simulated results: hydraulic fracturing of X-ray CT image-based models**
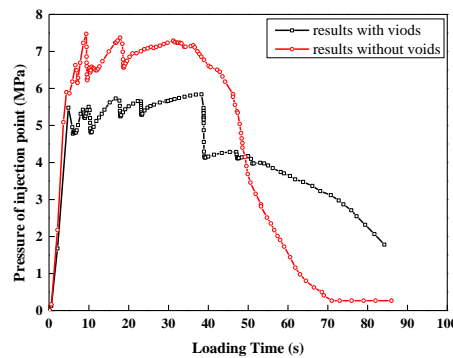


**Figure 8. Evolution of fluid pressure for X-ray CT image-based models**

**Conclusions**

2D meso-scale FE models were developed using the random aggregate generation and packing approach and from the micro CT images, to investigate the hydraulic fracture behavior of concrete. The effectiveness of pre-inserting cohesive elements (COH2D4P) with hydraulic freedoms has been demonstrated by good agreement of simulated results with experimental data. It is found that increasing the aggregate volume fraction and reducing the porosity can both improve the hydraulic fracture resistance of concrete.

**Acknowledgements**

**References**

[1] Slowik V, Saouma VE. (2000) Water pressure in propagating concrete cracks, *Journal of Structural Engineering* **126(2)**, 235–42.

[2] Bruhwiler E, Saouma VE. (1995) Water fracture interaction in concrete—part I: fracture properties, *Aci Material Journal* **92(3)**, 296–303.

[3] Elizaveta Gordeliy, Anthony Peirce. (2013) Coupling schemes for modeling hydraulic fracture propagation using the XFEM, *Computer Methods in Applied Mechanics and Engineering* **253**, 305–322.

[4] Yao F, Yang Z J, Hu Y J. (2018) A SBFEM-based for hydraulic fracturing in quasi-brittle materials, *Acta Mechanica Solida Sinica* **31(4)**, 416-432.

[5] Liu C, Shen Z Z, Gan L, et al. (2018) A Hybrid Finite Volume and Extended Finite Element Method for Hydraulic Fracturing with Cohesive Crack Propagation in Quasi-Brittle Materials, *Materials* **11**, 1921.

[6] Huang Y J., Yang Z J., Liu G H, et al. (2016) An efficient FE–SBFE coupled method for mesoscale cohesive fracture modelling of concrete, *Computational Mechanics* **58(4)**, 635-655.

[7] Ren W Y, Yang Z J, Sharma R, et al. (2015) Two-dimensional X-ray CT image based meso-scale fracture modelling of concrete, *Engineering Fracture Mechanics* **133**, 24-39.

[8] Yang Z J, Su X T, Chen J F, et al. (2009) Monte Carlo simulation of complex cohesive fracture in random heterogeneous quasi-brittle materials, *International Journal of Solids and Structures* **46(17)**, 3222-3234.

[9] Gan L, Shen X Z, Wang R, et al. (2017) Hydraulic fracturing test of concrete structures with single crack, *Advances in Science and Technology of Water Resources* **7(04)**, 30-35+86.

[10] Yang Z J, Ren W Y, Sharma R, et al. (2017) In-situ X-ray computed tomography characterisation of 3D fracture evolution and image-based numerical homogenisation of concrete, *Cement and Concrete Composites* **75**, 74-83.

[11] Yang Z J, Li B B, Wu J Y. (2019) X-ray computed tomography images based phase-field modeling of mesoscopic failure in concrete, *Engineering Fracture Mechanics* **208**, 151-170.

# The Competition between Entropy and Potential Energy in Contact Mechanics

**Zhaocheng Xuan**

Department of Computer Science, Tianjin University of Technology and Education, Tianjin 300222, China

Email: xuanzc@tute.edu.cn

## Abstract

In thermodynamics, equivalent to the growth of entropy in an adiabatic body is the decrease of the available free energy in a body whose boundary is kept as a constant temperature. In this study, we will show that there is also a competition between entropy and energy in the static elastic contact problems, which was found in the procedure of computing contact forces in the finite element model of contact problems. The potential contact nodes that are on the potential contact boundary are considered as a canonical ensemble of statistical physics, the nodal contact forces are normalized with a surrogate model of the primal model of contact problem, and a normalized nodal contact force is treated as the probability of a system occupying a microstate in statistical physics. Then analogous to the probability distribution that is obtained by maximizing the Gibbs entropy subject to the normalization condition of probability and the expectation value of system energy, the explicit formulation of normalized nodal contact forces is obtained by maximizing the entropy under the constraints of expectation of work done by the contact forces and nonpenetration conditions on the potential contact surface. An iterative algorithm for computing contact forces is constructed based on two principles - the principle of minimum potential energy and principle of maximum entropy. They are alternately applied in the iterative procedure, with the initial values of normalized nodal contact forces,  to find displacement with the principle of minimum potential energy, then to find an updated potential nodal contact force with the principle of maximum entropy using the displacement, keep the iteration until the termination condition is met. Examples show that the potential energy is increasing, and the entropy is decreasing throughout the iteration, this is contrary to the general understanding of entropy, that is, the potential energy should be decreased, and the entropy should be increased. This is because the "environment" of each step to calculate potential energy or entropy is different, even in an "environment" determined by nodal contact forces or displacements, the entropy reaches to the maximum or the potential energy reaches to the minimum, but in the whole process, the entropy is decreasing, and the potential energy is increasing, which shows a competitive relationship between the two concepts. (This research is supported by the National Natural Science Foundation of China under Grant No. 11772228.)

**Key words**: Entropy; Potential energy; Contact; Finite elements

**Appendix**: An illustration example.

An elastic T-shaped beam is supported by two elastic L-shaped beams, with the cross sections and dimensions of the structure (in centimeters) as shown in Figure 1, the numbers in brackets are rectangular coordinates. The Young's modulus of the elastic body is $E$=2900 $N$/cm$^2$, and the Poisson's ratio is $v$ =0.4. The pulling force $p$ =3 $N$/cm$^2$ acts on the bottom of the T-shaped beam. Since the T-shaped beam directly contacts the L-shaped beams, there is no initial gap between the elastic contact bodies. Considering the symmetry of the loading and geometry, only the left half of the structure is used to construct the finite element model as shown in

Figure 2, where 7 potential contact node pairs are arranged on the potential contact surfaces. Figure 3 shows the results of the normalized contact forces on the potential contact nodes in each iteration; the numbers on the lines indicate the iteration number, and the square marks show the reference solution produced by the commercial code. The initial value of normalized nodal contact force is set to 1/7, and 6 iterations are executed to calculate the solutions before the termination condition is met. Figure 4 shows the results of the potential energy and the entropy in each iteration, and it can again be seen that there is competition between the potential energy and the entropy. Figure 5 shows the von Misses stress distributions in the 6 iterations.
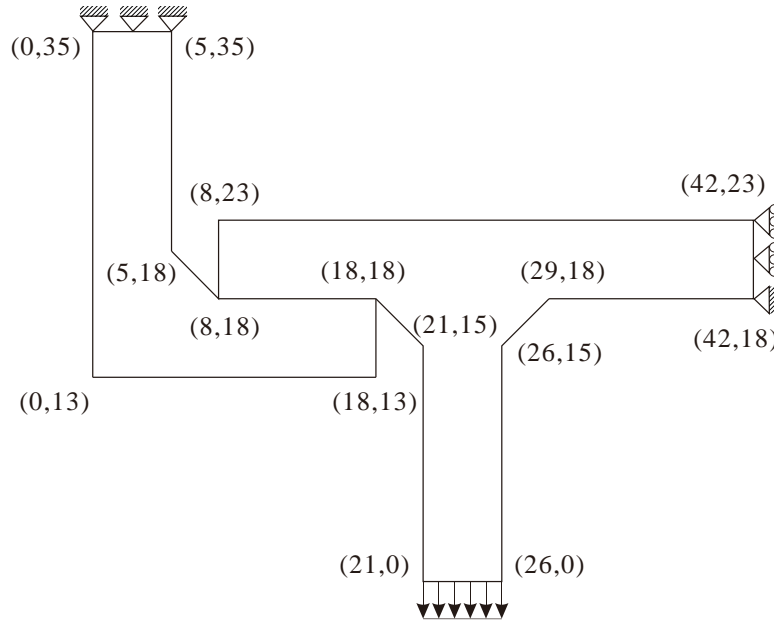


Figure 1: A T-shaped structure is in contact with two L-shaped structures.
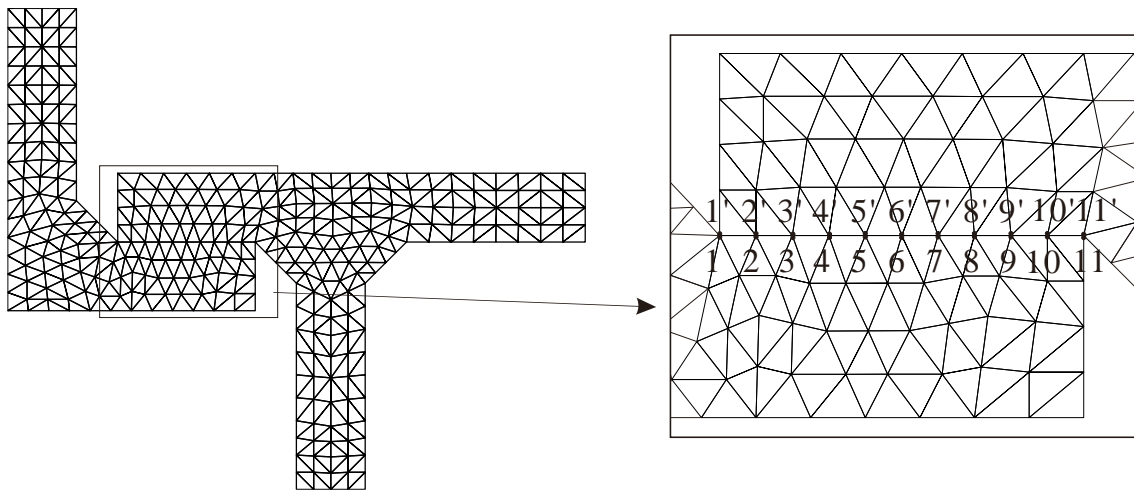


Figure 2: Finite element mesh and the potential contact node pairs (right).
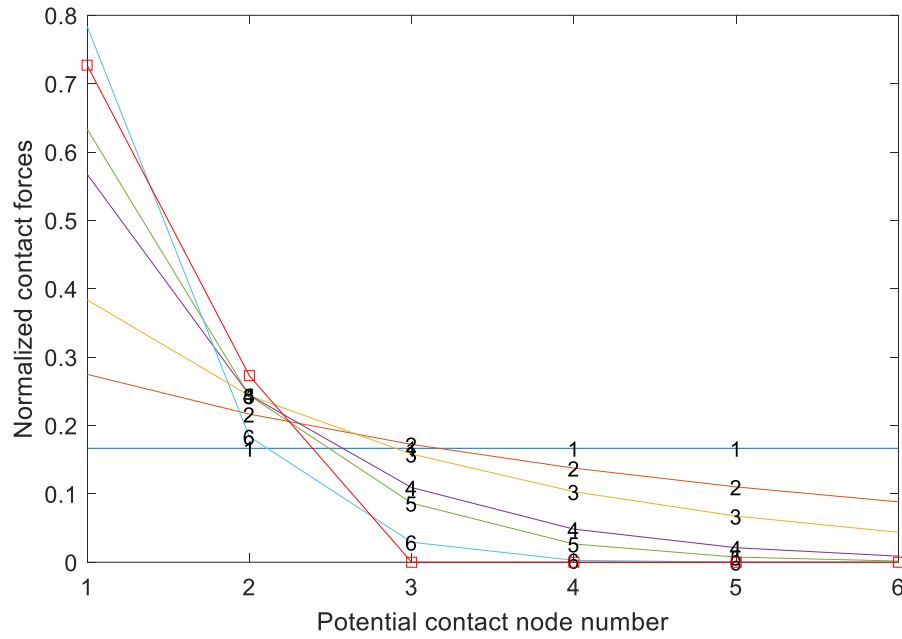
Figure 3: The results of normalized nodal contact forces in each iteration.
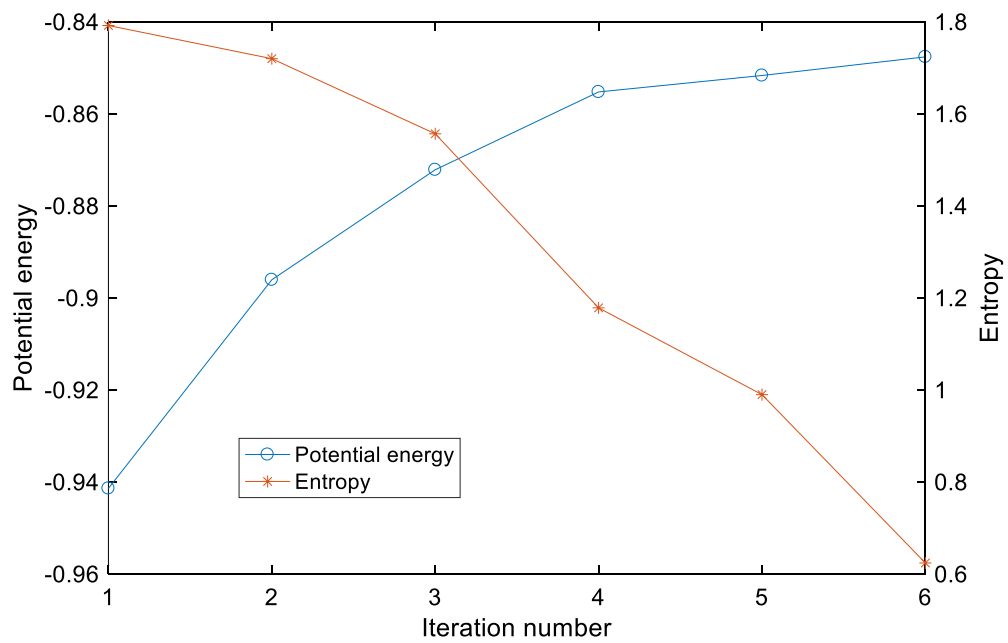


Figure 4: Potential energy vs. entropy in each iteration.
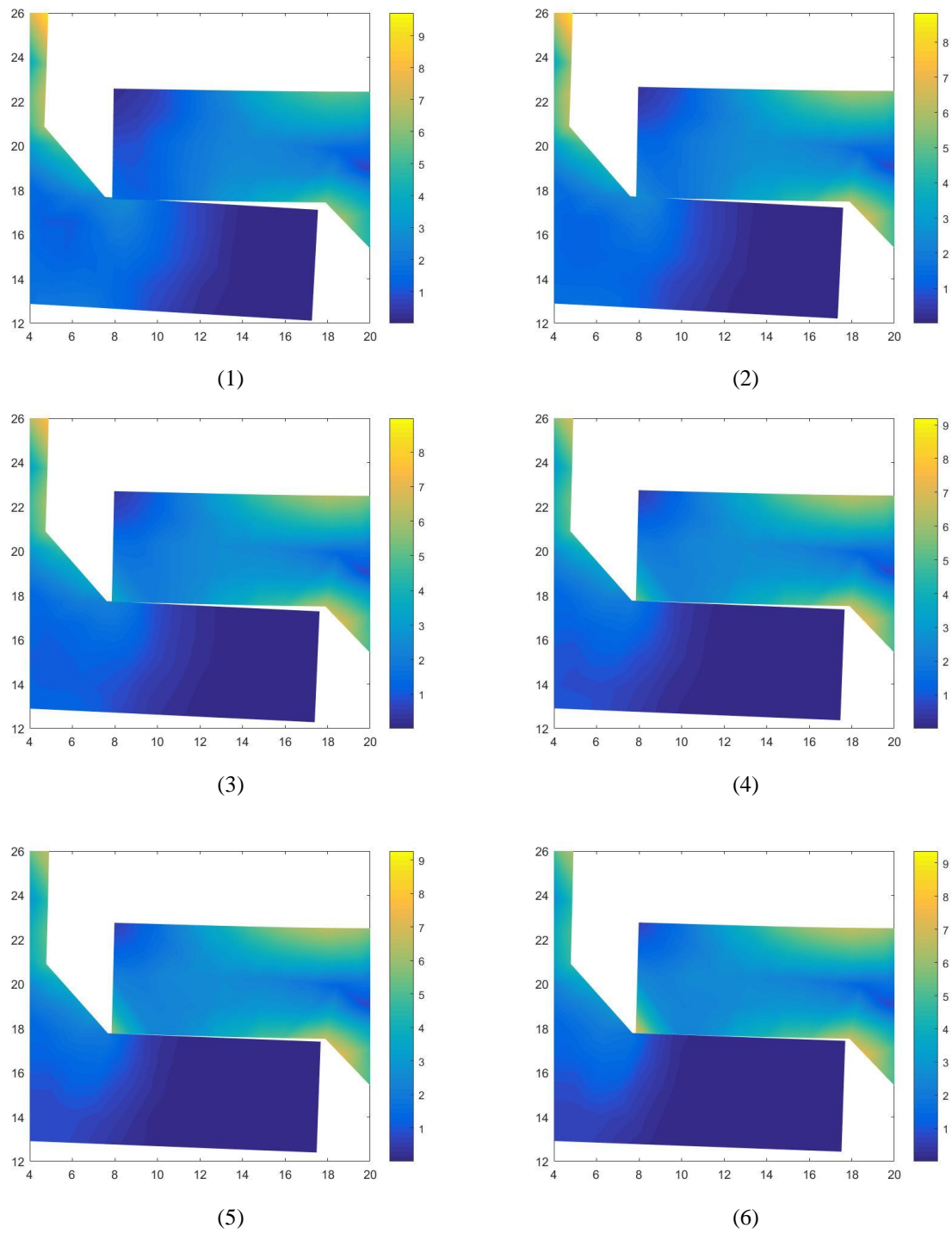
Figure 5: The von Misses stress distributions in the 6 iterations.

# BESO Approach for the Optimal Retrofitting Design of Steel Hollow-Section Columns Supporting Industry Cranes

## *Rut Su[1], Thu Huynh Van[1] and †Sawekchai Tangaramvong[1]

[1]Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Chulalongkorn University, Bangkok 10330, Thailand.

*Presenting author: rut-zu@hotmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

The local buckling phenomenon presents one of the main premature failures often prohibiting the steel hollow-section columns to attend the ultimate strength capacity. A strengthening method is then required to extend the service life of the member. This paper presents the optimal retrofitting design of standard steel hollow-section columns using external steel plates, such that the ultimate strength of the post-retrofitted column sufficiently resists the design load imposed by an industrial crane. The optimal design adopts a so-called bi-directional evolutionary structural optimization (BESO) algorithm that determines the cost-effective distribution of steel plate topology welded to the column. The proposed method realistically considers the influences of inelastic material properties and geometric nonlinearity, simultaneously. The BESO algorithm is encoded within the MATLAB modeling framework providing a direct application interface to ANSYS (a commercial-purposed finite element analysis software), which models the retrofitting joint between steel column and corbel using the comprehensive 3D finite elements. The robustness of the proposed scheme is illustrated through standard steel warehouse applications. The accuracy and integrity of the resulting design are validated by the full elastoplastic responses of the post-retrofitted column under applied forces.

**Keywords:** Local Buckling Failure; Topology Optimization; Evolutionary Structural Optimization; Column Retrofitting Design; Nonlinear Finite Element Analysis.

## 1. Introduction

Steel structural design is a combination process between architecture, safety and integrity. Good design complies with various performance criteria posed by all stakeholders at optimal resources. The design of hollow steel section (HSS) members has gained the popularity from designers. In views of strength consideration, the HSS provides the higher load capacity than the open sections [1]. Such physical property has made the HSS suitable for various structures and infrastructures, especially those with the special requirement on a long-span capability. One of the main drawbacks that unfortunately limits the general usage of steel hollow sections is the physically instabilizing local buckling failures under concentrated forces. A specific example is a steel HSS column with corbel heavily supporting industry crane loads of warehouses [2]-[4]. The HSS members has made itself prone to the premature failure caused by the eccentrically applied loads.

The intrinsic behaviors of HSS undergoing local buckling phenomena are rather sporadic. These have yet been precisely described by standard mathematical formulations [5]. The design guidelines were developed in the form of empirical formulations to predict the local buckling responses of steel open sections under compression and/or flexural forces, and the strengthening procedures were made available for the enhancement of the overall member

capacity. However, little is known about the behaviors and strengthening methods of the HSS applications.

The investigation of I-beam to box-column connection (IBBC) with external stiffener was reported in Ting [6] and Shanmugam [7]. The study of external T-stiffeners connected to IBBC was provided in [8]. In 2000, Hiroshi and Tanaka [9] conducted the experimental study of IBBC with external stiffener by wide spread flange. Furthermore, the numerical study of an external stiffener with IBBC was present in [10]. The work in [11] and [12] conducted experimental and numerical analyses of external diaphragm with IBBC in the concrete filled column applications. The aforementioned work adopted the strengthening techniques using external stiffening systems to facilitate the load transfer from beam to column members. The internal retrofits of HSS columns are generally inaccessible for practical constructions.
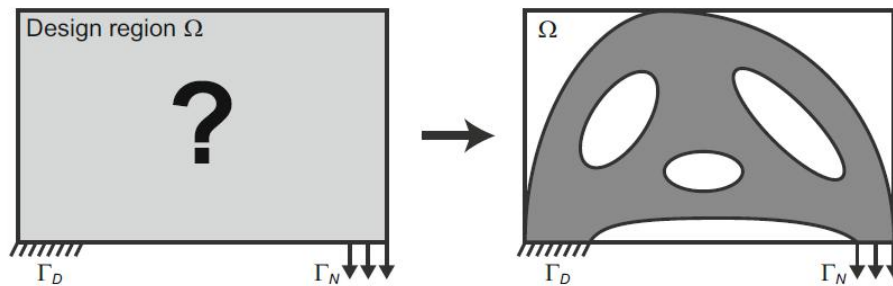


**Figure 1.** Illustration of typical structural topology optimization [13].

Topology Optimization (TO) determines the optimal shape as depicted in Fig. 1. The prior investigation of TO can be traced back to over a hundred years ago by Michell [14], who derived the optimality criteria of the least-weight truss layout. Mitchell's theory was extended 70 years later by Rozvany and his group [15]-[17] for the exact analytical optimal solutions of grid-type structures. In continuum mechanics, topology optimization can be formulated as a discrete problem or a binary design setting that the structure consists solely of either solids or voids [18]. However, the binary design for the structural compliance is ill-posed, where there exists non-convergent sequence of admissible designs with continuously refined geometrical details [19]-[22]. Bendsøe and Kikuchi [23] proposed the homogenization theory to relax the problem by assuming designable porous microstructures at a separated lower scale to claim this difficulty.

There has been the continuous development giving the emergence of various TO methods. One of the well-established techniques proposed by Xie and Steven [24]-[25] is called as an evolutionary structural optimization (ESO). Recently, the bi-evolutionary structural optimization (BESO) [27]-[29] has been developed. The method allows the recovery of the deleted elements close to the highly stressed areas. The convergent and mesh-independent BESO method [30] incorporates the sensitivity filter and stabilization schemes using some history information.

The present work proposes the optimal retrofit design of steel HSS columns in the IBBC region supporting heavily forces. The optimal topology of externally strengthening steel plates is appropriately designed using the BESO algorithm. The IBBC responses with and without optimal strengthening plates are mapped out by the nonlinear (incorporating inelastic materials and large deformations, simultaneously) finite element analysis method (FEM), modelled within the robust commercial-purposed (called ANSYS) analysis software. The applications of the proposed retrofit method are illustrated through the strengthening design of the HSS column connected to a corbel supporting industry crane loads.

## 2. Design guideline for column stiffening methods

*2.1 Crane Loads*

The crane load is determined according to [3] and [4]. The HSS columns are be designed to safely resist the applied loads and to prevent the local or global buckling failures. The detail of the top running cranes is depicted in Fig. 2.
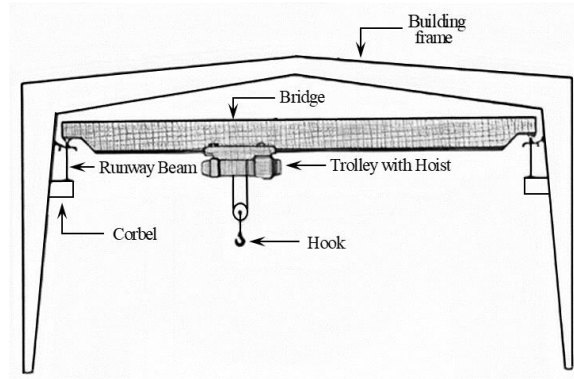


**Figure 2.** Top Running Bridge Crane with Suspended Trolley [3].

The model is constructed to predict the responses of the connection between HSS column and I-section corbel under the vertically applied forces. The vertical forces consider the wheel loads. The maximum magnitude occurs as when the crane is lifting its rated capacity load, and can be approximated as

$$WL = \frac{RC + HT + 0.5CW}{NW_b} \tag{1}$$

where $WL$ is the maximum wheel load, $RC$ is the rated crane capacity, $HT$ is the weight of hoist with trolley, $CW$ is the weight of the crane excluding the hoist with trolley. and $NW_b$ is the number of end truck wheels at one end of the bridge.

*2.2 External stiffener plates*

The retrofit of the beam-to-column connection focuses on the local failure behaviors (viz., chord deformations [31]) of the HSS columns. The external stiffening plates are designed for this purpose. The study of the connection between I-beam and Circular Hollow Section (CHS) column with external stiffeners is reported in [32] and [33]. The geometry of the HSS columns at IBBC zone with the proposed external ring plates are illustrated in Fig. 2,
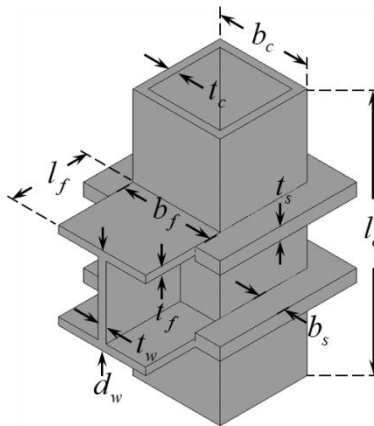


**Figure 2.** Geometry of IBBC with external ring plates.

where $b_c$ is the column width, $b_f$ and $d_w$ are the flange width and web depth of an I-beam, respectively, $t_c$ is the column thickness, $t_f$ and $t_w$ are the thickness of flange and web, respectively, $t_s$ is the thickness of external stiffener plates, $l_f$ is the beam length, and $l_c$ is the volume length.

*2.3 Topology optimization formulations*

The topology design of stiffener plates considers the minimization of the compliance function, namely $f_c = \dfrac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u}$, subject to the volume fraction constraint of the continuum model, and can be formulated by

$$
\begin{aligned}
\text{Minimization} \quad &: f_c(x, \mathbf{u}) \\
\text{subject to} \quad &: \mathbf{K}\mathbf{u} = \mathbf{F} \\
&: V(x) = \sum_{e=1}^{N_e} x_e v_e = V^*
\end{aligned}
\tag{2}
$$

where $\mathbf{K}$ and $\mathbf{u}$ are the global stiffness matrix and displacement vector at degrees of freedom, respectively, $v_e$ is the element volume, $V(x)$ and $V^*$ are the total and controlled material volumes (called volume fraction, i.e., ), respectively, $N_e$ and $x_e$ are the total number of finite elements and the material design variable, respectively.

The conventional ESO method known as a "hard-kill" method performs the complete elimination of some inefficient members within the design domain that can result in theoretical difficulties in some cases [34].

An alternative approach adopts the modulus reductions of the required elements to the very small values. This concept was applied to ESO by Hinton [35] and Rozvany [36]. Huang and Xie [37] replaced the virtual void elements with the soft members with very low Young's modulus values, and termed the method as a "soft-kill" BESO. This technique is an artificial material interpolation scheme with penalization that is similar to the treatment in the solid isotropic material with penalization (SIMP) model in steering the solution to the nearly solid-void design [38]-[40]. The elastic modulus of each intermediate material is interpolated as a function of the element density as

$$
E(x_e) = E^0 x_e^p, \text{ where } x_e = x_{\min} \text{ or } 1
\tag{3}
$$

$E_0$ denotes the elastic modulus of a solid material, and $p$ is the penalty exponent. The binary variable $x_e$ indicates the absence ($x_e = x_{\min}$)or presence ($x_e = 1$) of the element, where $x_{\min}$ is an artificially defined small parameter (e.g., 0.001). It is assumed that the Poisson's ratio is independent from the design variables, and the stiffness matrix $\mathbf{K}$ assemblies the products of elemental stiffness matrices $\mathbf{K}_e^0$ and design variables $x_e^p$ by

$$
\mathbf{K} = \sum_{e=1}^{N_e} x_e^p \mathbf{K}_e^0 .
\tag{4}
$$

At each design stage, the target volume in the current *l*-th design iteration $V^{(l)}$ is preset a priori. The required material volume can be greater or smaller than the volume of the initial guess design. Likewise, the target volume in each iteration may decrease or increase

progressively until the constraint volume is achieved. The evolution of the volume is expressed by

$$V^{(l)} = V^{(l-1)}\left(1 \pm c_{er}\right),$$ (5)

where the evolutionary ratio $c_{er}$ determines the percentage of material to be added or removed with reference to the design in the previous iteration. After the targeted material volume $V_{req}$ has been attained, the optimization alters only the topology whilst keeping the volume constant (up to a certain tolerance). The sensitivity of the structural compliance with regards to the change in the $e$-th element is evaluated by the adjoint function [41] as

$$\frac{\partial f_c}{\partial x_e} = -p x_e^{p-1}\left(\frac{1}{2}\mathbf{u}_e^T \mathbf{K}_e^0 \mathbf{u}_e\right),$$ (6)

where $\mathbf{u}_e$ denote elemental displacement vectors. The structure is optimized using the discrete $x_e$ design variables, and only two bound materials are allowed in the design [18]. Therefore, the sensitivity number used in the BESO method is defined by the relative ranking of the sensitivity associated with the individual element as

$$\alpha_e = -\frac{1}{p}\frac{\partial f_c}{\partial x_e} = x_e^{p-1}\left(\frac{1}{2}\mathbf{u}_e^T \mathbf{K}_e^0 \mathbf{u}_e\right).$$ (7)

In order to avoid checkerboard patterns and mesh-dependency, the sensitivity numbers are firstly smoothed by means of the filter scheme as

$$\alpha_e = \frac{\sum_{j=1}^{N_e} w_{ej}\alpha_j}{\sum_{j=1}^{N_e} w_{ej}},$$ (8)

where $w_{ej}$ is a linear weight factor computed by the prescribed filter radius $r_{min}$ and the elemental center-to-center distance $\Delta_{ej}$ between elements $\Omega_e$ and $\Omega_j$ as

$$w_{ej} = \max\left(0, r_{min} - \Delta_{ej}\right).$$ (9)

The filtered sensitivity numbers are averaged with the sensitivity numbers of the previous topology iteration to improve the convergence by

$$\alpha_e^{(l)} = \left(\alpha_e^{(l)} + \alpha_e^{(l-1)}\right)/2.$$ (10)

The BESO algorithm optimizes the structure by removing and adding the elements assigned in the ground domain. Two discrete values, namely $x_{min}$ for void elements and 1 for solid elements, are applied [42]. The sensitivity numbers for the solid and void elements are expressed by

$$\alpha_e = \begin{cases} \dfrac{1}{2}\mathbf{u}_e^T \mathbf{K}_e^0 \mathbf{u}_e & \text{when } x_e = 1 \\[2mm] x_{min}^{p-1}\left(\dfrac{1}{2}\mathbf{u}_e^T \mathbf{K}_e^0 \mathbf{u}_e\right) & \text{when } x_e = x_{min} \end{cases}.$$ (11)

The sensitivity numbers of solid elements are independent to the penalty exponent $p$, whilst those numbers of the soft elements depend on the $p$ value.

The recent BESO approach has been developed to the optimal topology for a wide range of structural design applications, involving multiple materials [37], multiple constraints [43], stiffness and frequency optimization [44] and nonlinear material and large deformation [45]-[48]. The present plate strengthening approach proposed adopts the BESO algorithm, and the procedures as how the optimal retrofit design is performed are described in Section 3.

## 3. Analysis and design of column stiffening

### 3.1 Traditional Retrofit design

The responses of steel HSS columns (over an IBBC area) under crane loads were mapped out using the nonlinear 3D finite element (FE) analyses. The SHS - 300 × 10 ($b_c \times t_c$) columns were considered as for illustrative purposes. The standard steel corbels H - 300 × 300 × 10 × 10 ($b_f \times d_w \times t_f \times t_w$) were employed, where the length of column $l_c$ and corbel $l_f$ equal to 1 m and 300 mm respectively. The analyses considered the responses of stiffened HSS columns for various plate thicknesses $t_s$, ranging from 3 to 10, 15 and 30 mm. The external ring stiffeners having a typical width of $b_s = 100$ mm.

The commercial purposed ANSYS Parametric Design Language (APDL) software modeled the structure as a number of (uniform 10 mm in size) eight-node solid (SOLID185) FEs (see Fig. 3). The HSS column was restrained in all directions at both ends, except that the deformation along a z-axis direction was released at the top end to permit its vertical translation under applied forces. The material properties employed were: the elastic modulus of 200,000 MPa, Poisson's ratio of 0.3 and yield stress of 235 MPa (for both column and beam) and 250 MPa (for the stiffening plate).



(a)                                                  (b)
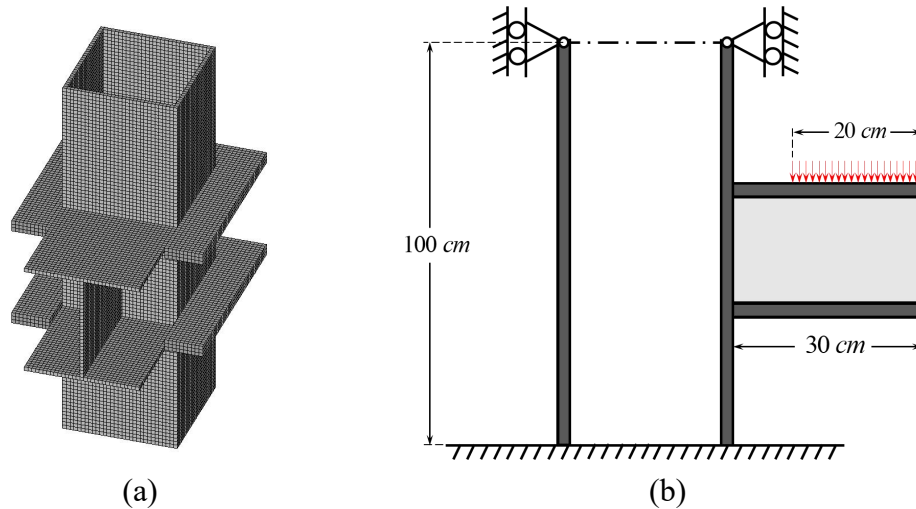
**Figure 3.** Structural discretization (a) FE model and (b) boundary conditions.

The FE analyses incorporated the influences of inelastic (elastic-perfectly plastic) materials and large (nonlinear geometry) deformations to realistically map out the full responses of the HSS column at an IBBC zone. The crane load was applied over a 20 cm width contact surface on an I-section corbel (see Fig. 3b).

*3.2 BESO algorithm procedure*

The topology of the two HSS column stiffening plates located at the top and bottom flanges of an I-beam were designed using the BESO algorithm [27]-[29]. The design procedures can be briefly summarized as follows:

Step 1: Discretize the design domain as in Fig. 3a with the sufficient number of eight-node solid (SOLID185) FEs. The domain initially sets the thickness of stiffener plates to $t_s = 30$ mm.

Step 2: Initialize the BESO parameters, including objective volume $V^*$, evolutionary ratio $c_{er}$, radius of filter $r_{min}$ and penalty exponent $p$.

Step 3: Perform the nonlinear FE analyses. Only the two stiffening plates are designed by the BESO schemes.

Step 4: Determine the targeted volume in Eq. (5) for the next iteration, if the current volume $V^{(l)}$ is larger than the objective volume $V^*$.

Step 6: Calculate the sensitivity functions in Eq. (8).

Step 7: Perform the elimination and addition process. The elemental density is switched from 1 to $x_{min}$ (i.e., a member addition) if $\alpha_e \leq \alpha_{th}$. In contrast, the elemental density for void element is switched from $x_{min}$ to 1 (member elimination) if $\alpha_e > \alpha_{th}$. The threshold $\alpha_{th}$ defines the threshold on the sensitivity number that is determined by the target material volume $V^{(l+1)}$ and the relative ranking of the sensitivity numbers, see [30].

Step 8: Repeat Steps 3 to 7. The algorithm terminates when the optimal topology of steel stiffening plates is converged.

It is noted that this work investigated the variation of the optimal plate topologies for different values of $V^*$, namely ranging from 0.5 down to 0.05 (e.g., indicating only 5 percent of initial design elements were remained in the final solutions). The responses of a HSS column with the resulting stiffening plates at an IBBC area were traced by the elastoplastic analyses in an ANSYS software to ensure the safety and integrity of the solutions.

## 4. Results and Discussions

The BESO method were successfully performed to achieve the optimal solution. The resulting optimal topologies of stiffening plates for difference volume fraction $V^*$ are depicted in Fig. 5. The full responses of the HSS columns with designed retrofitting steel plates at an IBBC zone were traced by performing the nonlinear material and geometry FE analyses. The corresponding maximum load capacities for all the designed cases are reported in Table 1. Clearly, the higher volume fractions $V^*$ yielded the stronger HSS columns retrofitted and hence the IBBC behaviours giving the capability to support the heavier crane loads applied on the I-section corbel.

The simple design was also performed for the ring plates in Fig. 2 with both top and bottom ring plates having the uniform thickness varying from $t_s = 3$ mm to 30 mm (without the BESO implementation). The maximum load capacities of the corresponding HSS columns at the IBBC zone captured by the comprehensive FE analyses are reported in Table 2 and are directly compared in Fig. 6 with those obtained after the retrofitting designs using the BESO method. The results obviously shown the volume fractions $V^*$ 0.1, 0.2, 0.3 and 0.5 have more

maximum load capacities than $t_s$ 3, 6, 9 and 15 mm around 33.8, 25.59, 18.55 and 5.92 %. It is evidenced that at the same volume of designed external plates the strength gained by the BESO process is more than that by the uniform plate thickness retrofits. The BESO provided significantly the effective plate strengthening designs of the HSS columns at the IBBC area with the high value of strength enhancement per unit volume of the plates employed.



(a) $V^* = 0.5$     (b) $V^* = 0.4$     (c) $V^* = 0.3$



(d) $V^* = 0.2$     (e) $V^* = 0.1$
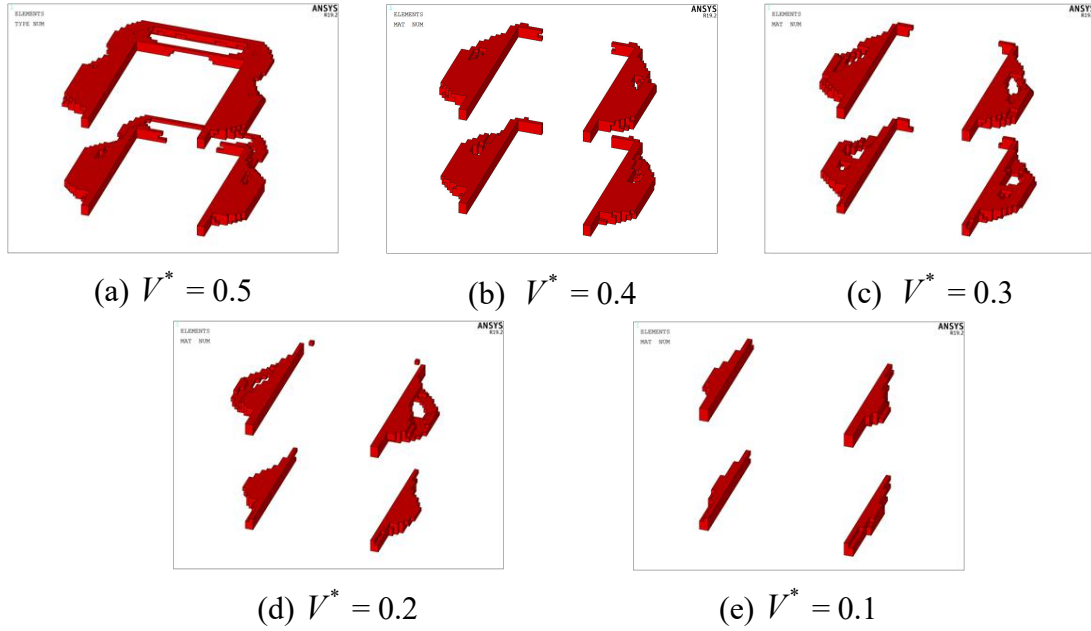
**Figure 5.** Optimal solution with different objective volume.

**Table 1.** Maximum load capacities for various design volumes by BESO.

| $V^*$ | Volume (cm³) | $P_{NL}$ (Tons) |
|---|---|---|
| 0.05 | 390 | 82.450 |
| 0.10 | 780 | 91.305 |
| 0.15 | 1,170 | 95.858 |
| 0.20 | 1,560 | 98.126 |
| 0.25 | 1,950 | 99.872 |
| 0.30 | 2,340 | 105.220 |
| 0.35 | 2,730 | 105.420 |
| 0.40 | 3,120 | 107.290 |
| 0.45 | 3,510 | 110.760 |
| 0.50 | 3,900 | 113.600 |

**Table 2.** Maximum load capacities for various design volumes by uniform-thickness plate design.

| $t_s$ (mm) | Volume (cm³) | $P_{NL}$ (Tons) |
|---|---|---|
| unstiffened | – | 58.487 |
| 3 | 780 | 68.146 |
| 4 | 1,040 | 71.508 |
| 5 | 1,300 | 74.907 |

| 6 | 1,560 | 78.319 |
| 7 | 1,820 | 81.749 |
| 8 | 2,080 | 85.220 |
| 9 | 2,340 | 88.753 |
| 10 | 2,600 | 91.888 |
| 15 | 3,900 | 107.890 |
| 30 | 7,800 | 126.690 |
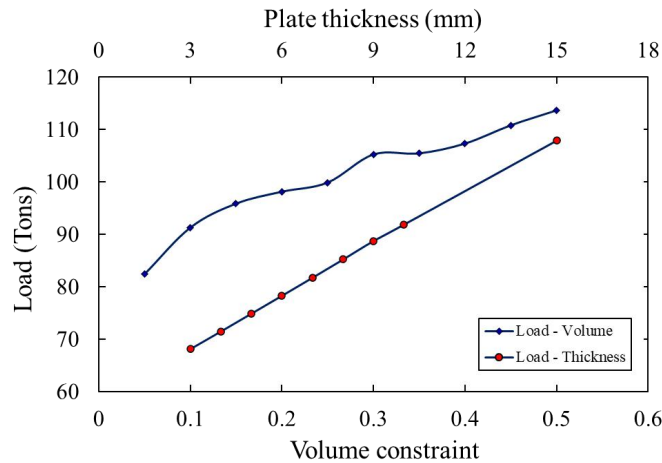


**Figure 6.** Maximum load capacities with different techniques.

## 5. Concluding remarks

This paper presents the steel plate topology optimization for the strengthening design of HSS columns at an IBBC zone subjected to the industry crane loads. The initial design domains are the thick ring-shape steel plates enveloping the HSS column at the contacts on top and bottom flanges of the I-section corbel. The BESO method has been developed for the determination of optimal material distributions through the process of a soft-kill optimization, permitting not only elimination but also addition of steel masses for various targeted volume fractions. The influences of inelastic material as well as local steel buckling failures are incorporated using the comprehensive nonlinear material and geometry FE analyses.

The optimal topology of the retrofitting plates computed provides the maximum strength enhancement to the HSS columns at the IBBC area, whilst the associated design volumes are maintained the indicated fractions. The efficiency of the BESO method has been evidenced, where the maximum load capacities of the HSS columns retrofitted by the steel plates with the optimal layouts in the IBBC can be achieved at the far lower plate volumes, as compared to those with the simple uniform-thickness plate designs. Extensions to the present work focus on the applications of steel open-section columns having different slenderness practically employed in engineering structures.

## 6. Acknowledgements

**References**

[1] Kurobane, Y., Packer, J.A., Wardenier, J. (2004) Design guide for structural hollow section column connections. CIDECT Series "Construction with hollow steel sections" No. 9, TÜV-Verlag, Köln, Germany, ISBN 3- 8249-0802-6.

[2] Maccrimmon, R. (2009) Crane-Supporting Steel Structure. 2nd Edition. Canadian Institute of Steel Construction, Niagara Falls, Ontario, Canada.

[3] MBMA (2012) Metal Building Systems Manual, based on the 2012 IBC and ASCE 7-10, Cleveland, Ohio.

[4] NBCC (2015) National Building Code of Canada, Ottawa: Associate Committee on the National Building Code, National Research Council.

[5] Lu, L. H. (1997). The static strength of I-beam to rectangular hollow section column connections. Ph.D. Dissertation, Delft University, Netherlands.

[6] Ting, L. C., Shanmugam, N.E., Lee, S.L. (1991). Box-column to I-beam connections with external stiffeners. *Journal of Constructional Steel Research* **18,** (3): 209-226.

[7] Shanmugam, N.E., Ting, L. C., Lee, S.L. (1993) Static behaviour of I-beam to box-column connections with external stiffeners. *The Structural Engineer* **71,** (15): 269–275.

[8] Ting, L. C., Shanmugam, N.E., Lee, S.L. (1993). Use of external T-stiffeners in box-column to I-beam connections. *Journal of Constructional Steel Research* **26,** (2-3): 77–98.

[9] Hiroshi, M., and Tanaka, A. (2000). Statical characteristics of the improved beam-to-column connections of steel structure. *Proceedings of 12th World Conference on Earthquake Engineering*, Auckland, New Zealand.

[10] Goswami, R., Murty, C. V. R. (2010). Externally Reinforced Welded I-Beam-to-Box-Column Seismic Connection. *Journal of Engineering Mechanics* **136**, (1): 23–30.

[11] Vulcu, C., Stratan, A., Ciutina, A., and Dubina, D. (2017a). Beam-to-CFT High-Strength Joints with External Diaphragm. I: Design and Experimental Validation. *Journal of Structural Engineering* **143**, (5).

[12] Vulcu, C., Stratan, A., Ciutina, A., and Dubina, D. (2017b). Beam-to-CFT High-Strength Joints with External Diaphragm. II Numerical Simulation of Joint Behavior. *Journal of Structural Engineering* **143,** (5).

[13] Xia, L., & Breitkopf, P. (2016). Recent Advances on Topology Optimization of Multiscale Nonlinear Structures. *Archives of Computational Methods in Engineering* **24**, (2): 227–249.

[14] Michell, A. G. M. (1904) Lviii. the limits of economy of material in frame-structures. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **8**, (47): 589–597.

[15] Prager, W., Rozvany, G. (1977) Optimal layout of grillages. *Journal of Structural Mechanics* **5**, (1): 1–18.

[16] Rozvany, G. (1972) Grillages of maximum strength and maximum stiffness. *International Journal of Mechanical Sciences* **14**, (10): 651–666.

[17] Rozvany, G. (1972) Optimal load transmission by flexure. *Computer Methods in Applied Mechanics and Engineering* **1**, (3): 253–263.

[18] Xia, L., Xia, Q., Huang, X., and Xie, Y. M. (2016). Bi-directional Evolutionary Structural Optimization on Advanced Structures and Materials: A Comprehensive Review. *Archives of Computational Methods in Engineering* **25**, (2): 437–478.

[19] Cheng, K., Olhoff, N. (1981) An investigation concerning optimal design of solid elastic plates. *International Journal of Solids and Structures* **17**, (3): 305–323.

[20] Kohn, R., Strang, G. (1986a) Optimal design and relaxation of variational problems (part i). *Commun Pure Applied Math* **39**, (1): 113–137.

[21] Kohn, R., Strang, G. (1986b) Optimal design and relaxation of variational problems (part ii). *Commun Pure Applied Math* **39**, (2): 139–182.

[22] Kohn, R., Strang, G. (1986c) Optimal design and relaxation of variational problems (part iii). *Commun Pure Applied Math* **39**, (3): 353–377.

[23] Bendsøe, M. P., & Kikuchi, N. (1988). Generating optimal topologies in structural design using a homogenization method. *Computer Methods in Applied Mechanics and Engineering*, **71,** (2): 197–224.

[24] Xie, Y. M., Steven, G.P. (1992) Shape and layout optimization via an evolutionary procedure. *In: Proceedings of the International Conference on Computational Engineering Science*, Hongkong, China.

[25] Xie, Y. M., Steven, G.P. (1993) A simple evolutionary procedure for structural optimization. *Computers and Structures* **49**, (5): 885-896.

[26] Xie, Y. M., Steven, G.P. (1997) Evolutionary Structural Optimization. Springer-Verlag, London.

[27] Querin, O. M., Steven, G. P., and Xie, Y. M. (1998). Evolutionary structural optimisation (ESO) using a bidirectional algorithm. *Engineering Computations* **15**, (8): 1031–1048.

[28] Querin, O. M., Steven, G. P., and Xie, Y. M. (2000) Evolutionary structural optimization using an additive algorithm. *Finite Elements in Analysis and Design* **34**, (3): 291–308.

[29] Querin, O. M., Young, V., Steven, G. P., and Xie, Y. M. (2000). Computational efficiency and validation of bi-directional evolutionary structural optimisation. *Computer Methods in Applied Mechanics and Engineering* **189**, (2): 559–573.

[30] Huang, X., and Xie, Y. M. (2007). Convergent and mesh-independent solutions for the bi-directional evolutionary structural optimization method. *Finite Elements in Analysis and Design* **43**, (14): 1039–1049.

[31] Lee, M. M., and Llewelyn-Parry, A. (1999). Strength of ring-stiffened tubular T-joints in offshore structures. *Journal of Constructional Steel Research* **51**, (3): 239–264.

[32] Bagheri Sabbagh, A., Chan, T. M., and Mottram, J. T. (2013). Detailing of I-beam-to-CHS column joints with external diaphragm plates for seismic actions. *Journal of Constructional Steel Research* **88**: 21–33.

[33] Tritep, U. and Kitjapat, P. (2016). Strength Enhancement of Connections between Steel I-Beam and Circular Hollow Steel Column by Nonlinear Finite Element Analysis. *National Convention on Civil Engineering (NCCE21)*, Songkla, Thailand, 28-30 June 2016, pp.35-43.

[34] Zhou, M., and Rozvany, G. I. N. (2001). On the validity of ESO type methods in topology optimization. *Structural and Multidisciplinary Optimization* **21**, (1): 80–83.

[35] Hinton, E., and Sienz, J. (1995). Fully stressed topological design of structures using an evolutionary procedure. *Engineering Computations* **12**, (3): 229–244.

[36] Rozvany, G. I. N., and Querin, O. M. (2002). Combining ESO with rigorous optimality criteria. *International Journal of Vehicle Design* **28,** (4): 294.

[37] Huang, X., and Xie, Y. M. (2008). Bi-directional evolutionary topology optimization of continuum structures with one or multiple materials. *Computational Mechanics* **43**, (3): 393–401.

[38] Bendsøe, M. P. (1989). Optimal shape design as a material distribution problem. *Structural Optimization* **1**, (4): 193–202.

[39] Bendsøe, M. P., and Sigmund, O. (1999). Material interpolation schemes in topology optimization. *Archive of Applied Mechanics (Ingenieur Archiv)* **69**, (9-10): 635–654.

[40] Rietz, A. (2001). Sufficiency of a finite exponent in SIMP (power law) methods. *Structural and Multidisciplinary Optimization* **21**, (2): 159–163.

[41] Bendsøe, M. P. and Sigmund, O. (2003) Topology optimization: theory, methods and applications. Springer-Verlag, Berlin.

[42] Huang, X., and Xie, Y. M. (2007). Convergent and mesh-independent solutions for the bi-directional evolutionary structural optimization method. *Finite Elements in Analysis and Design* **43**, (14): 1039–1049.

[43] Huang, X., and Xie, Y. M. (2009). Evolutionary topology optimization of continuum structures with an additional displacement constraint. *Structural and Multidisciplinary Optimization* **40**, (1-6): 409–416.

[44] Huang, X., Zuo, Z. H., and Xie, Y. M. (2010). Evolutionary topological optimization of vibrating continuum structures for natural frequencies. *Computers and Structures* **88**, (5-6): 357–364.

[45] Huang, X. H., and Xie, Y. (2007). Bidirectional Evolutionary Topology Optimization for Structures with Geometrical and Material Nonlinearities. *AIAA Journal* **45**, (1): 308–313.

[46] Huang, X., and Xie, Y. M. (2008). Topology optimization of nonlinear structures under displacement loading. *Engineering Structures* **30**, (7): 2057–2068.

[47] Huang, X., and Xie, Y.-M. (2010). A further review of ESO type methods for topology optimization. *Structural and Multidisciplinary Optimization* **41**, (5): 671–683.

[48] Xia, L. (2016) Multiscale structural topology optimization. Elsevier-ISTE, London.

# Optimal Inspection Period for Structures Subjected to Fatigue

**†Feng Li, *Qianhui Xu, Liming Zhou, and Heng Zhao**

College of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022

*Presenting author: xuqh19@mails.jlu.edu.cn

†Corresponding author: fengli@jlu.edu.cn

**Abstract**

From the view of reliability, an optimal inspection model for three-dimensional structures under random loading is presented. Neuber method and finite element analysis as well as experiential stress intensity factor(**SIF**) were applied in the analysis of fatigue life. Based on the series system theory, the reliability function for the structural fatigue life is given. Treat the crack initiation life (**CIL**)as service life and the crack propagation life(**CPL**) as inspection period, the optimal inspection period was obtained under a given reliability. The proposed method is proved to be feasible by a numerical example, which involves a bracket under random loading.

**Key words:** Inspection period; Fatigue; Reliability; Stochastic finite element method; Neuber method

## 1 Introduction

The resistance of structures will deteriorate during service life. One of the important mechanisms of deterioration is the effect of fatigue on structures subjected to repeated or cyclic loading patterns. To maintain acceptable levels of safety, it needs to perform frequent periodic inspections in case of sudden failure. Performing these inspections is often time consuming and costly. However, inadequate inspection may result in disasters which cause heavy casualties. Therefore, in order to reduce inspection costs and ensure structural safety, an optimal inspection strategy is very important.

Because of the significant uncertainties inherent in the degradation process of the structural fatigue life, most inspection models[1-7] were based on the reliability method. Using cost-benefit criteria, Rackwitz[1] developed a maintenance interventions model for deteriorating structures, but only S-N curves was considered in this paper and the crack propagation stage was neglected. Zou [2] gave a minimum cost preventive maintenance schedule for deteriorating structure. He suggested that intervals between inspections were equal. However, with the fatigue damage growing, fatigue crack growth rate is gradually accelerated. Intervals between inspections should be gradually shortened to ensure the reliability by repeated inspections. Based on linear elastic fracture mechanics (**LEFM**), many authors [3-5] proposed some inspection strategies. Karan Doshi[3] proposed an inspection planning for ship structural details using fracture mechanics. Based on dynamic Bayesian networks, Yang and Frangopol [4] presented a framework for optimizing inspection methods of fatigue-critical details in order to minimize the posterior expected life-cycle cost .Eltaief [5] proposed a dynamic inspection time optimization method for scheduling preventive

maintenance of mechanical components subjected fatigue.

As seen from the above literature review, most of the research on inspection planning for structures is based on the damage tolerance approach. This approach assumes that initial cracks were existed for structures. Obviously, these methods are not suitable for uncracked components with relatively long crack initiation life(**CIL**). An optimal inspection model was established by Gao[6]. Meng et al. [7] later extended this work by finite element analysis. However, Meng's model is restricted to two-dimensional structures and the complicate calculation.

Based on the work of Meng et al. [7], this paper proposes an optimal inspection period model for three-dimensional structures under random loading. Neuber method and Finite Element Analysis as well as experiential Stress Intensity Factor(**SIF**) were applied in the analysis of fatigue life. The proposed method reduces the calculation time and is more suitable for common engineering structures.

## 2 Optimised inspection planning methodology

Usually, the fatigue life is separated into a **CIL** and a crack propagation life (**CPL**). Denote the **CIL** and the **CPL** as $T_p$ and $T_p^*$ respectively, the total fatigue life $T$ can be expressed as

$$T = T_p + T_p^* \tag{1}$$

Because of uncertainties existed in the process of fatigue, it is difficult for a deterministic methodology to assess the fatigue life of the structure. Therefore, reliability analyses [8-10] now are widely applied in fatigue design to ensure that the structure fulfills its design purpose for some specified design life.

If the structure has $m$ key components , for the $i$th component, denote the reliability function $p_i$ for the **CIL** $T_{pi}$ and the reliability function $p_i^*$ for the **CPL** $T_{pi}^*$ as follows

$$p_i = f_i(T_{pi}) \tag{2}$$

$$p_i^* = f_i^{'}(T_{pi}^*) \tag{3}$$

Because the fatigue failure is a series system, the failure probability is [6]

$$F_i = (1 - p_i)(1 - p_i^*) \tag{4}$$

The reliability is

$$R_i = 1 - F_i = 1 - (1 - p_i)(1 - p_i^*) \tag{5}$$

The reliability function for the whole structure will be

$$R = \prod_{i=1}^{m}(1 - (1 - p_i)(1 - p_i^*)) = \prod_{i=1}^{m}(f_i(T_{pi}) + f_i^{'}(T_{pi}^*) - f_i(T_{pi})f_i^{'}(T_{pi}^*)) \tag{6}$$

Treat the **CIL** as service life and **CPL** as inspection period. Under a given reliability, when the **CIL** is equal to the **CPL**, the first optimal inspection period $T1$ can be obtained. If there is no crack observed in the first inspection, the next inspection period $T2$ can be obtained as follows

$$R = \prod_{i=1}^{m}(f_i(T1 + T2) + f_i^{'}(T2) - f_i(T1 + T2)f_i^{'}(T2)) \tag{7}$$

Similarly, other inspection periods can also be got until a crack is observed.

Optimised inspection planning can be illustrated graphically as shown in Fig.1.
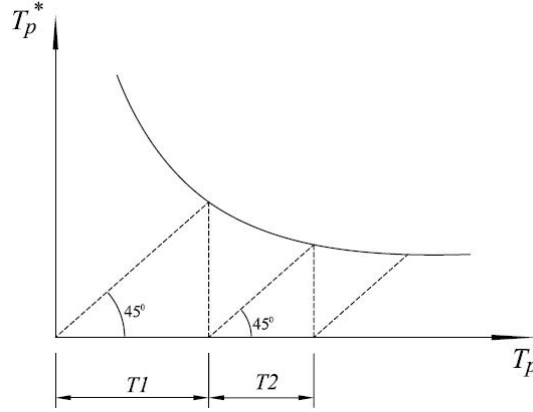
Fig.1    Sketch of optimal inspection period

## 3    The fatigue life of structures

In practice, it is necessary to reduce the spectrum of varying loading into a set of simple constant amplitude loading by rainflow cycle counting method. The relationship between the total fatigue life under the spectrum of varying loading and fatigue lives under constant amplitude loadings can be quantified in terms of the Miner's rule[11]. According to this rule, the damage $D$ caused by variable amplitude loadings can be defined as

$$D = \sum_{j=1}^{k} \frac{n_j}{N_j} \qquad (8)$$

where $n_j$ is the number of cycles at the $j$th loading level; $N_j$ is the number of cycles to failure under the $j$th loading level .

When the damage   reaches 1, it is supposed that failure will occur. So, the structural fatigue life $T$ under the spectrum of varying loading can be obtained by

$$T = \frac{1}{D} = 1 \bigg/ \sum_{j=1}^{k} \frac{n_j}{N_j} \qquad (9)$$

It should be noted that $N_j = N_{pj}$ and $T = T_p$ for the **CIL** as well as $N_j = N_{pj}^*$ and $T = T_p^*$ for the **CPL**.

### 3.1 The crack initiation life

For components in the **CIL**, the local strains become the parameter which can govern fatigue behavior. The stress-stain response can be obtained by Elasto-Plastic Finite Element Analysis (**EPFMA**) [7]. However, this method means a large amount of calculation and it is time-consuming. So, Neuber method [12] is adopted to simulate the local nonlinear stress-strain behavior. Neuber method is expressed as follows

$$K_t^2 = K_\sigma K_\varepsilon = \frac{\sigma}{S} \frac{\varepsilon}{e} \qquad (10)$$

where $K_t$ is the elastic stress concentration factor , $K_\sigma$ is the true stress concentration factors, $K_\varepsilon$ is the strain concentration factors, $\sigma$ and $S$   are the true and nominal stress   respectively, $\varepsilon$   and $e$ are the true and nominal strain respectively.

For the cyclic loading, Topper *et al.*[13] suggested that

$$K_f^2 = \frac{\Delta\sigma}{\Delta S} \frac{\Delta\varepsilon}{\Delta e} \qquad (11)$$

where $K_f$ is the fatigue strength reduction factor.

Given the further linear relationship of nominal elasticity

$$\Delta S = E \Delta e \tag{12}$$

where $E$ is the Yong's modulus.

Rearrangement of Eq (11) leads to:

$$\Delta\sigma\Delta\varepsilon = \frac{K_f^2 (\Delta S)^2}{E} \tag{13}$$

Meanwhile, the expression for the cyclic stress-strain curve is

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\sigma}{2E} + \left(\frac{\Delta\sigma}{2K'}\right)^{\frac{1}{n'}} \tag{14}$$

in which $K'$ is the cyclic strength coefficient, $n'$ is the cyclic-hardening exponent.

The true strain amplitude $\Delta\varepsilon$ is composed of the elastic and plastic components

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p \tag{15}$$

Basquin [14] gave the expression for the elastic strain amplitude-life curve:

$$\frac{\Delta\varepsilon_e}{2} = \frac{\sigma'_f}{E}\left(2N_p\right)^b \tag{16}$$

where $\sigma'_f$ and $b$ are the fatigue strength coefficient and exponent respectively.

Both Manson[15] and Coffin[16] simultaneously given the expression for the plastic strain amplitude- life curve:

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f \left(2N_p\right)^c \tag{17}$$

where $\varepsilon'_f$ and $c$ are the fatigue ductility coefficient and exponent respectively.

From Eqs. (16) and (17), the expression of strain–life curve is established as

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}\left(2N_p\right)^b + \varepsilon'_f \left(2N_p\right)^c \tag{18}$$

When referring the mean stress $\sigma_m$, Morrow [17] suggested that the elastic part of the strain-life curve could be modified as follows

$$\varepsilon_a = \frac{\sigma'_f - \sigma_m}{E}\left(2N_p\right)^b + \varepsilon'_f \left(2N_p\right)^c \tag{19}$$

if $\sigma_m < 0$, then specify $\sigma_m = 0$.

The local stress and strain at time $j$ can be obtained by the following expression

$$\sigma_j = \sigma_{j-1} + sign(\Delta S'_{j,j-1})\Delta\sigma_{j,j-1} = \sum_{k=1}^{j} sign(\Delta S'_{k,k-1})\Delta\sigma_{k,k-1} \tag{20}$$

$$\varepsilon_j = \varepsilon_{j-1} + sign(\Delta S'_{j,j-1})\Delta\varepsilon_{j,j-1} = \sum_{k=1}^{j} sign(\Delta S'_{k,k-1})\Delta\varepsilon_{k,k-1} \tag{21}$$

where $\Delta S'_{k,k-1}$ is the nominal stress increment from time k-1 and k, if $\Delta S'_{k,k-1} = S_k - S_{k-1} > 0$, $sign(\Delta S'_{k,k-1}) = 1$; while $\Delta S'_{k,k-1} < 0$, $sign(\Delta S'_{k,k-1}) = -1$.

## 3.2 The crack propagation life

For the **CPL** $N_p^*$, fatigue crack growth laws were based on **LEFM**. The crack growth rate is presented by a general form

$$\frac{da}{dN_p^*} = f\left(\Delta K, R, K_{IC} \cdots\right) \tag{22}$$

where $\Delta K$ is the **SIF** range, $R$ is the stress ratio, $K_{Ic}$ is the fracture toughness.

Paris and Erdogan [18], who postulated that the rate of fatigue crack growth depends on the **SIF** range $\Delta K$, proposed the basic model of the fatigue crack growth rate

$$\frac{da}{dN_p^*} = C\left(\Delta K_{eq}\right)^m \tag{23}$$

where $C$ and $m$ are material parameters; $\Delta K_{eq}$ is the equivalent stress intensity range of mode I, which is

$$\Delta K_{eq} = K_{eq}^{\max} - K_{eq}^{\min} \tag{24}$$

where $K_{eq}^{\max}$ is the value of the maximum **SIF** and $K_{eq}^{\min}$ is the minimum **SIF**.

According to the theory of maximum circumferential stress[19], the equivalent **SIF** of mode I can be written as

$$K_{eq} = \left[K\right]^T \left[\Psi\right] \tag{25}$$

in which

$$\left[K\right] = \begin{bmatrix} K_I \\ K_{II} \end{bmatrix} \,,\quad \left[\Psi\right] = \cos^2\frac{\theta}{2}\begin{bmatrix} \cos\dfrac{\theta}{2} \\ -3\sin\dfrac{\theta}{2} \end{bmatrix} \tag{26}$$

$$\theta = 2\,arc\tan\frac{1}{4}\left(\frac{K_I}{K_{II}} \pm \sqrt{\left(\frac{K_I}{K_{II}}\right)^2 + 8}\right) \tag{27}$$

where $\theta$ is the crack propagation angle and its sign is the opposite of the $K_{II}$.

From Eq. (23), the **CPL** $N_p^*$ can be got

$$N_p^* = \int_{a_0}^{a_c} \frac{da}{C\left(\Delta K_{eq}\right)^m} \tag{28}$$

where the initial crack length $a_0$ are determined by the threshold **SIF** $K_{th}$ and the final crack length $a_c$ are determined by the fracture toughness $K_{IC}$[20].

To account for the effect of loading ratio $R$ and the fracture toughness $K_{IC}$, Forman[20] suggested that

$$N_p^* = \int_{a_0}^{a_c} \frac{(1-R)K_{IC} - \Delta K_{eq}}{C\left(\Delta K_{eq}\right)^m}\,da \tag{29}$$

For two-dimensional structure, Meng et al. [7] proposed triangular quarter-point elements around the crack tip to calculate the **SIF**. However, the complexity of crack mesh makes it difficult to apply finite element method to the crack propagation analysis of three-dimensional structure. According to **LEFM**, the stress intensity which depends on the loading, crack shape, mode of crack displacement can be expressed as

$$K_i = Y(a)S\sqrt{\pi a} \quad (i = I, II) \tag{30}$$

where $Y(a)$ is the geometry correction factor.

For a semi-elliptical surface crack(in Fig.2) which is the most common problem in engineering, the geometry correction factor [21] is

$$Y(a) = \left[ \frac{\left( \sin^2 \phi + \left( \dfrac{a}{c} \right)^2 \cos^2 \phi \right)^{\frac{1}{4}}}{\psi(k)} \right] M_f \tag{31}$$
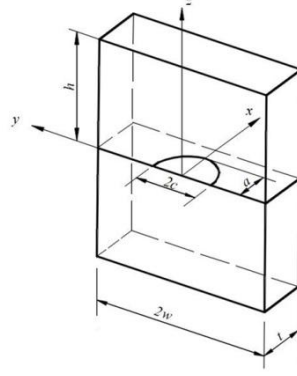


Fig. 2 The semi-elliptical surface crack

where $a$ is the depth, $c$ is the half-length, $\phi$ is the angle in the parametric equation of ellipse, $\psi(k)$ is the complete integral of the second kind that can be approximated as

$$\psi(k) = \left( 1 + 1.464 \left( \frac{a}{c} \right)^{1.65} \right)^{1/2} \tag{32}$$

$M_f$ is the correction factor, for finite width $W$ and finite thickness $t$, it can be calculated as

$$M_f = \left( M_1 + M_2 \left( \frac{a}{t} \right)^2 + M_3 \left( \frac{a}{t} \right)^4 \right) g f_w \tag{33}$$

where

$$M_1 = 1.13 - 0.09 \frac{a}{c}$$

$$M_2 = -0.54 - \frac{0.89}{0.2 + \dfrac{a}{c}}$$

$$M_3 = 0.5 - \frac{1}{0.6 + \dfrac{a}{c}} + 14 \left( 1 - \frac{a}{c} \right)^{24}$$

$$g = 1 + \left( 0.1 + 0.35 \left( \frac{a}{t} \right)^2 \right) (1 - \sin \phi)^2$$

$$f_w = \sec \left( \frac{\pi c}{2W} \sqrt{\frac{a}{t}} \right)$$

## 4   Reliability analysis of fatigue life

Note random variables as $X = (x_1, x_2, \cdots, x_n)$ and define the performance function of fatigue life as

$$g(X) = \ln T - \ln T_D = -\ln\left(\sum_{j=1}^{k}\frac{n_j}{N_j}\right) - \ln T_D \tag{34}$$

in which $T_D$ is the design life.

The reliability of the structure is as follows

$$R = \int_{g(X)>0} f(X)dX \tag{35}$$

where $f(X)$ is the joint probability density function(**PDF**).

In general, Eq.(35) cannot be solved analytically. The most common approach to compute the reliability is the first order reliability method (**FORM**). In the **FORM**,the failure surface is represented by a hyperplane at the most probable failure point(**MPP**), which is the closed point to the origin in the standard normal space. So, the reliability analysis of fatigue life can be treated as a constrained optimization problem as follows

$$\text{find:} \quad Y$$
$$\text{which minimizes:} \quad \beta = \sqrt{Y^T Y}$$
$$\text{subject to:} \quad H(Y) = 0 \tag{36}$$

where $\beta$ is the reliability index, $Y = [y_1, y_2, \ldots, y_n]$ denote the standard Gaussian variables, $H(Y)$ in the standard normalized space is transformed by the performance function $G(X)$ in the physical space.

The structure reliability can also be expressed as

$$R = \Phi(\beta) \tag{37}$$

where $\Phi$ is the standard normal cumulative distribution function(**CDF**).

Assumed the basic variables X are stochastically independent, a transformation from the basic x-space to the standard normal y-space is defined by

$$F_{x_i}(x_i) = \Phi(y_i) \tag{38}$$

where $F_{x_i}(x_i)$ is the **CDF** in terms of the basic variable $x_i$.

The transformed standard normal variable is then given by

$$y_i = \Phi^{-1}\left[F_{x_i}(x_i)\right] \tag{39}$$

Taking the derivative of both sides of Eq.(38) yields

$$f_{x_i}(x_i)dx_i = \phi(y_i)dy_i \tag{40}$$

$$\frac{dx_i}{dy_i} = \frac{\partial x_i}{\partial y_i} = \frac{\phi(y_i)}{f_{x_i}(x_i)} \tag{41}$$

where $f_{x_i}(x_i)$ is the basic **PDF**, $\phi(\cdot)$ is the standard normal **PDF**.

Since it is impossible to find the **MPP** in Eq. (36) analytically, the iterative expression for the **MPP** search is given by[22]

$$\beta^{k+1} = \beta^k + \frac{g(Y^k)}{\left\|\nabla g(Y^k)\right\|} \tag{42}$$

$$Y^{k+1} = -\frac{\nabla g(Y^k)}{\left\|\nabla g(Y^k)\right\|}\left\{\beta^k + \frac{g(Y^k)}{\left\|\nabla g(Y^k)\right\|}\right\} \tag{43}$$

where $\nabla g(Y) = \left(\dfrac{\partial g(Y)}{\partial y_1}, \dfrac{\partial g(Y)}{\partial y_2}, \ldots, \dfrac{\partial g(Y)}{\partial y_n}\right)$

For the *j*th load level, if the random variable $x_i$ is independent to the structural displacement,

the partial derivative of *g* function with respect to $y_i$ can be theoretically calculated according to the chain rule of the composite function

$$\frac{\partial g}{\partial y_i} = \frac{1}{T}\frac{\partial T}{\partial D}\frac{\partial D}{\partial N_j}\frac{\partial N_j}{\partial x_i}\frac{\partial x_i}{\partial y_i} = \frac{1}{T}\frac{1}{D^2}\sum_{j=1}^{k}\frac{n_j}{N_j^2}\frac{\partial N_j}{\partial x_i}\frac{\partial x_i}{\partial y_i} \tag{44}$$

If the random variable $x_i$ is relevant to the structural displacement , there are

$$\frac{\partial g}{\partial y_i} = \frac{1}{T}\frac{1}{D^2}\sum_{j=1}^{k}\frac{n_j}{N_j^2}\left(\frac{\partial N_{pj}}{\partial \Delta\varepsilon_j}\frac{\partial \Delta\varepsilon_j}{\partial x_i} + \frac{\partial N_{pj}}{\partial \sigma_{m,j}}\frac{\partial \sigma_{m,j}}{\partial x_i}\right)\frac{\partial x_i}{\partial y_i} \tag{45}$$

$$\frac{\partial g}{\partial y_i} = \frac{1}{T}\frac{1}{D^2}\sum_{j=1}^{k}\frac{n_j}{N_j^2}\frac{\partial N_{pj}^*}{\partial \Delta K_{eq,j}}\frac{\partial \Delta K_{eq,j}}{\partial x_i}\frac{\partial x_i}{\partial y_i} \tag{46}$$

where

$$\frac{\partial \Delta\varepsilon_j}{\partial x_i} = \frac{1}{2}\left(\frac{\partial \varepsilon_{j,\max}}{\partial x_i} - \frac{\partial \varepsilon_{j,\min}}{\partial x_i}\right) \tag{47}$$

$$\frac{\partial \sigma_{m,j}}{\partial x_i} = \frac{1}{2}\left(\frac{\partial \sigma_{j,\max}}{\partial x_i} + \frac{\partial \sigma_{j,\min}}{\partial x_i}\right) \tag{48}$$

$$\frac{\partial \Delta K_{eq,j}}{\partial x_i} = \frac{\partial K_{eq,j}^{\max}}{\partial x_i} - \frac{\partial K_{eq,j}^{\min}}{\partial x_i} \tag{49}$$

From Eqs. (20) and (21) ,we have

$$\frac{\partial \sigma_j}{\partial x_i} = \sum_{k=1}^{j} sign(\Delta S_{k,k-1}^{'})\frac{\partial \Delta\sigma_{k,k-1}}{\partial x_i} \tag{50}$$

$$\frac{\partial \varepsilon_j}{\partial x_i} = \sum_{k=1}^{j} sign(\Delta S_{k,k-1}^{'})\frac{\partial \Delta\varepsilon_{k,k-1}}{\partial x_i} \tag{51}$$

Also from Eqs.(13) and (14), we obtain

$$\frac{\partial \Delta\sigma_{k,k-1}}{\partial x_i} = \frac{2K_f^2\Delta S_{k,k-1}}{G}\frac{\partial \Delta S_{k,k-1}}{\partial x_i} \tag{52}$$

$$G = E\Delta\varepsilon_{k,k-1} + \Delta\sigma_{k,k-1}W \tag{53}$$

$$W = 1 + \frac{E}{n'K'}\left(\frac{\Delta\sigma_{k,k-1}}{2K'}\right)^{\frac{1-n'}{n'}} \tag{54}$$

$$\frac{\partial \Delta\varepsilon_{k,k-1}}{\partial x_i} = \frac{W}{E}\frac{\partial \Delta\sigma_{k,k-1}}{\partial x_i} \tag{55}$$

$$\frac{\partial \Delta S_{k,k-1}}{\partial x_i} = \frac{\partial S_k}{\partial x_i} - \frac{\partial \Delta S_{k-1}}{\partial x_i} \tag{56}$$

And $\dfrac{\partial S_k}{\partial x_i}$ can be obtained by From the view of reliability (**SFEM**).

## 5. Taylor expansion stochastic finite element method

Due to the available technological advances in computing power, SFEM has acquired quantity attention over the past three decades and has become an excellent tool to solve problems with uncertain. There are mainly three types of **SFEM** :Taylor expansion **SFEM** [23], perturbation

**SFEM** [24] and Neumann expansion **SFEM[** 25]. Due to the effectiveness and easy coding, Taylor expansion **SFEM** is chose here to obtain the partial derivative $\dfrac{\partial S_k}{\partial x_i}$.

The finite element equation is as follows

$$KU = F \tag{57}$$

in which $U$ is the displacement vector, $K$ is the global stiffness matrix, $F$ is the global load vector. Both $K$ and $F$ are functions of random variables.

By using the first-order Taylor expansion method, there are

$$E[U] \approx \overline{U} = \overline{K}^{-1}\overline{F} \tag{58}$$

$$\frac{\partial U}{\partial x_i} = K^{-1}\left(\frac{\partial F}{\partial x_i} - \frac{\partial K}{\partial x_i}U\right) \tag{59}$$

in which $\overline{U}$, $\overline{K}$ and $\overline{F}$ are the mean values and $E[\cdot]$ is the expectation operator.

According to the stress-displacement relationship, there is

$$S = DBU \tag{60}$$

in which $S$ is the stress vector, $D$ is the elasticity matrix, $B$ is the displacement shape function matrix.

From Eq. (60), the mean value of stress vector can be written as

$$E(S) = E(D)E(B)E(U) \tag{61}$$

The partial derivative $\dfrac{\partial S}{\partial x_i}$ can be got by Eqs. (60)

$$\frac{\partial S}{\partial x_i} = \frac{\partial D}{\partial x_i}BU + D\frac{\partial B}{\partial x_i}U + DB\frac{\partial U}{\partial x_i} \tag{62}$$

## 6 Example

As shown in Fig.3, a bracket, whose material is alloy 16Mn[26], is fixed at its left surface. The parameters are listed as follows: the Young's modulus $E$ is $71.02\,GPa$, the cyclic strength coefficient $K'$ is 645.79, the cyclic-hardening exponent $n'$ is 0.0069, the threshold stress intensity $K_{th}$ is $16.8\text{MPa}\sqrt{m}$, the fracture toughness $K_{IC}$ is $68.6\text{MPa}\sqrt{m}$. The loading history can be seen in Fig.4 and the coefficient of variation(**CV**) for each nominal load is 0.1. The other random variables are given in Table 1.
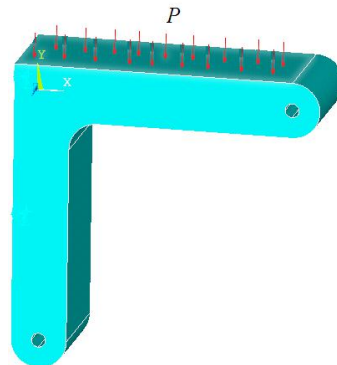


Fig.3    The bracket under random loading
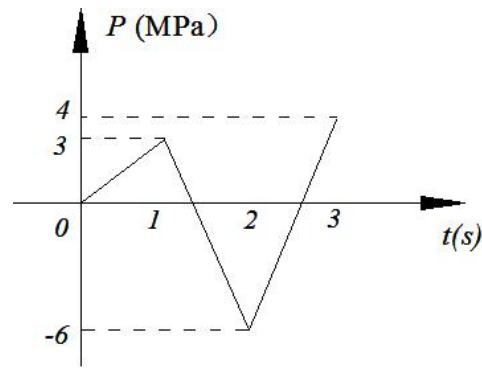
Fig.4 The loading history

Tab. 1　　Random Variables

|  | Mean | CV |
|---|---|---|
| $\sigma'_f$ | 847MPa | 0.03 |
| $b$ | -0.0943 | 0.03 |
| $\varepsilon'_f$ | 0.4644 | 0.04 |
| $c$ | -0.5396 | 0.04 |
| $C$ | $1.062 \times 10^{-13}$ | 0.07 |
| $m$ | 4.443 | 0.06 |



Fig.5　　Finite element mesh of the bracket

Finite element analysis during the **CIL** was made using commercial software Ansys. Finite element mesh of the bracket is presented in Fig.5. The relationship between the life $T_p$ and probability $p$ is shown in Fig 6. The result of this paper　is also compared with the 100,000 Monte-Carlo simulations. The expression of reliability $p$ with the **CIL** $T_p$ is as follows

$$p = -9.20396270 \times 10^{-10} \cdot T_p^2 + 8.2537296 \times 10^{-6} \cdot T_p + 0.984009 \qquad （63）$$
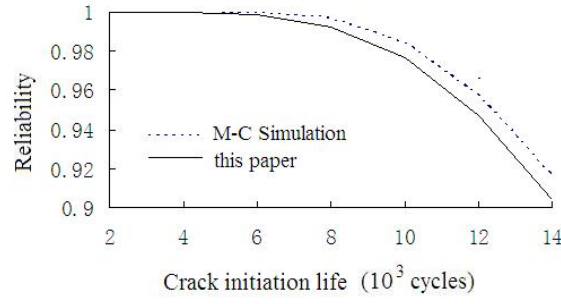
Fig.6    Reliability for **CIL**
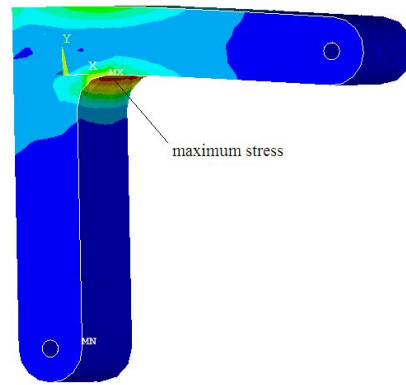


Fig. 7 Von Mises Stress distribution of the bracket in the **CIL**
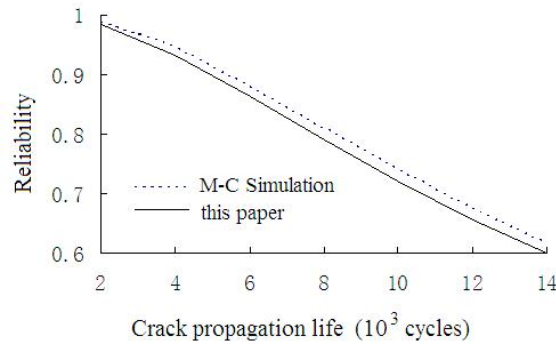


Fig.8    Reliability for **CPL**

From the finite element analysis as shown in Fig. 7, the maximum stress occurs in the fillet at the corner.

Assuming the variable ratio a/c is 0.5, the relationship between $T_p^*$ and $p*$ during the **CPL** is shown in Fig.8. The expression of reliability $p*$ with the propagation life $T_p^*$ is as follows

$$p^* = -1.61785714 \times 10^{-9} \cdot T_p^{*2} - 1.75335714 \times 10^{-5} T_p^* + 1.02763142 \qquad (64)$$

Substitute Eq. (63) and (64) into (6), then given $T_p = T_p^* = T1$, $R = 0.9999$ and the calculated first inspection period $T1$ is 6419 cycles and the 10000 Monte-Carlo simulation result is 6858 cycles.

Fig. 9 presents the relationship between inspection period and variable ratio a/c. It can be seen that the inspection period grows with the increasing of the variable ratio a/c. For safety's sake, the value of a/c can be taken as 0.1.
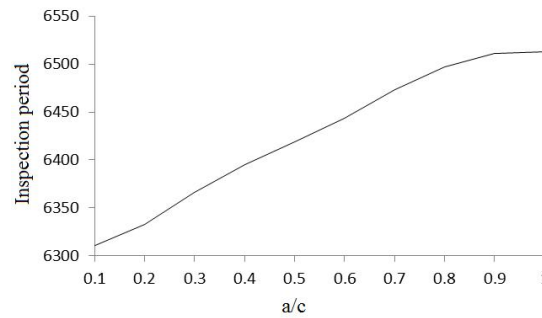
Fig. 9    Relationship between inspection period and a/c

## 7    Conclusion

From the view of probability theory and mathematical statistics, an optimal inspection period model is presented for structures, based on **SFEM** and the reliability method**.**

This method can ensure the safety and reliability of the structure and make full use of it by prediction its service life. Since the structural fatigue life includes the **CIL** and the **CPL**, the inspection policy that takes the **CIL** as service life and the **CPL** as inspection period could ensure the safe operation of the structure, even if there is no crack.

With the finite element analysis and Neuber method in **CIL**, experiential **SIF** in **CPL** analysis , the proposed method reduces the calculation time. The result of this paper shows good accuracy when it is compared with Monte Carlo Simulation.

## Reference

1    R. Rackwitz.Optimization- the basis of code-making and reliability verification. Structural Safety, 2000, 22:27-60.

2    Zou XiaoLi. An optimal preventive maintenance policy for structures[J]. Acta    Aeronautica ET Astronautica Sinica , 1997 , 18 (3) : 363 -366.

3    Karan Doshi,Tuhin Roy,Yogendra Singh Parihar. Reliability based inspection planning using fracture mechanics based fatigue evaluations for ship structural details[J].Marine Structures,2017,54:1-22.

4    David Y. Yang,Dan M. Frangopol. Probabilistic optimization framework for inspection/repair planning of fatigue-critical details using dynamic Bayesian networks[J].Computers and Structures, 2018, 198:40–50.

5    M. Eltaief, A. Chateauneuf, Ch. Bouraoui, T. Hassine, Dynamic approach for optimal inspection planning of fatigue cracked components. Journal of Constructional Steel Research,2015, 115:263–275.

6    Fu Huimin, Gao Zhentong. A model for reliability assessment of structures. Acta Aeronautica et Astronautica Sinca, 1991; 12(5): A264-A271.(in Chinese)

7    Guangwei Meng，Feng Li，Lirong Sha，Zhenppin Zhou. Prediction of economic inspection time on structural fatigue life . International Journal of Fatigue, 2007, 29:1516-1522.

8    Jinting Guo, Ge Wang, Anastassios N. Perakis, Lyuben Ivanov. A study on reliability-based inspection planning – Application to deck plate thickness measurement of aging tankers. Marine Structures, 2012,25(1):85-106.

9    John Leander,Mohammad Al-Emrani.Reliability-based fatigue assessment of steel bridges using LEFM – A sensitivity analysis, International Journal of Fatigue, 2016,93:82–91.

10   Ming Li, Li-Yang Xie, Hai-Yang Li，Jun-Gang Ren. Life distribution transformation model of planetary gear system. Chinese Journal of Mechanical Engineering, 2018, 31:24.

11    Miner MA. Cumulative damage in fatigue. Journal of Applied Mechanics,1945,67,A159-A164.

12    Neuber H. Theory of stress concentration for shear-strained prismatic bodies with arbitrary nonlinear stress–strain law. Journal of Applied Mechanics, ASME 1961,28:544–550.

13    Topper, T. H., Wetzel, R. M., and Morrow, JoDean, Neuber's    Rule Applied to Fatigue of Notched Specimens, Journal of Materials, ASTM, 1969,4(1), 200-209 .

14    Basquin OH. The exponential law of endurance tests. Proc ASTM, ASTEA 1910,10:625–630.

15    Coffin LF. A study of the effects of cyclic thermal stresses on a ductile metal. Trans ASME 1954; 76:931–50.

16    Manson SS. Behavior of materials under conditions of thermal stress. Heat Transfer Symposium, University of Michigan Engineering Research Institute, 1953, p. 9–75.

17    Morrow JD. Cyclic plastic strain energy and fatigue of metals. Internal friction, damping, and cyclic plasticity. ASTM STP 378. Philadelphia (PA): American Society for Testing and Materials,1964, p. 45–87.

18    Paris PC, Erdogan F. A critical analysis of crack propagation laws. Journal of Basic Engineering, 1963, 85: 528-534.

19    Erdogan F, Sih GC. On the crack extension in plates under plane loading and transverse shear. Journal of Basic   Engineering , 1963, 85:519–27.

20    Wang Ziqiang, Chen Shaohua. Advanced fracture mechanics. Beijing:Science Press, 2009 (in Chinese) .

21    Yibing Xiang, Zizi Lu, Yongming Liu. Crack growth-based fatigue life prediction using an equivalent initial model. Part I: Uniaxial loading. International Journal of Fatigue, 2010, 32: 341–349.

22    Xiaoping Du, Wei Chen.A Most Probable Point-Based Method for Efficient Uncertainty Analysis. Journal of Design and Manufacturing Automation, 2001,1(1):47–66

23    Marcin Kamiński.On stochastic finite element method for linear elastostatics by the Taylor expansion. Structural and multidisciplinary optimization, 2008, 35 (3) : 213-223.

24    Baizhan Xia, Dejie Yu, Jian Liu.Transformed perturbation stochastic finite element method for static response analysis of stochastic structures.Finite Elements in Analysis and Design, 2014,79:9-21.

25    Zhao Lei,Chen Qiu.Neumann dynamic stochastic finite element method of vibration for structures with stochastic parameters to random excitation. Computers & Structures, 2000, 77(6):651-657.

26    Gao Zhentong, Jiang Xintong, Xiong Junjiang,etc. Design of fatigue performance test and data processing. Pekin: Pekin University of Aeronautic and Astronautic Press, 1999.(in Chinese)

# A parameter-free methodology for boundary identification of topology optimization results

**\*†Yang Liu[1], Yoshinori Nishio[2], and Hiroki Asayama[3]**

[1]Department of Mechanical Engineering, Sojo University, Japan.
[2,3]Graduate School of Mechanical Engineering, Sojo University, Japan.

\*Presenting author: liuyang60212@gmail.com
†Corresponding author: liu@mec.sojo-u.ac.jp

## Abstract

Topology optimization is an advanced design method for generating lightweight and high-performance structures by determination of the material distribution. One of important drawbacks of the topology optimization, especially performed by the density-based approach, is that distinct and smooth boundaries cannot be directly obtained owing to checkerboard patterns, grayscales and irregular shapes. This drawback make it difficult to manufacture the results of topology optimization. In this paper, a novel methodology is proposed to automatically obtain optimal smooth boundaries of topology optimization results using an efficient boundary smoothing technique and $H^1$ gradient method, which is a node-based parameter-free optimization method. With this methodology, distinct and smooth optimal boundaries can be determined without any shape design parameterization. Moreover, re-mesh is not necessary in the shape updating process and the process is fully automatic. The validity and practical utility of this method is verified through a numerical example with respect to a mean compliance minimization problem.

**Keywords:** Topology optimization, Density approach, Boundary smoothing, Shape optimization, $H^1$ gradient method

## Introduction

Topology optimization is an advanced design method that is generally performed in the initial phase of the design to predict the optimal material distribution within a given initial design space of a structure. Compared to size and shape optimization, topology optimization is the most comprehensive and attractive type of structural optimization. Both shape and configuration can be varied in the topology optimization to obtain lightweight and high-performance structures that are difficult to obtain with conventional design ideas.

Since the "homogenization approach to topology optimization"[1] is introduced in the seminal paper by Bendsøe and Kikuchi in 1988, topology optimization has been extensively studied and achieved a tremendous development. Other than the homogenization approach, a number of topology optimization approaches have been proposed, including: density based approach [2]; level set method [3]; evolutionary approaches [4]; topological derivative [5]; phase field [6], and several others [7]. Among these approaches, the density approach is the most popular technology due to its efficiency in computation. Moreover, this method is relatively easy to implement numerically and has been employed in many commercial software. In addition, it can be easily applied to physical problems other than structural mechanics. Since 2000, the density approach has been applied to various physical problems

such as heat diffusions, fluids, acoustics, electromagnetics, optics, etc. Unfortunately, the topology optimization performed by the density-based approach tends to suffer from numerical instability problems, including mesh dependency, checkerboard patterns and grayscales. Distinct and smooth boundaries cannot be directly. This drawback makes it difficult to directly manufacture the results of topology optimization, and makes these results difficult to use in subsequent studies.

To obtain distinct and smooth boundaries from topology results, many approaches have been proposed. For example, the density contour approach is widely used to identify geometric boundaries by selecting a value of density threshold [8]. Basic shape templates [9], such as parametric spheres, cylinders, and rectangles, are applied to determine the size and location of holes inside topology results. Interpolation functions [10], such as B-spline curves, bi-quartic surface splines, and T-spline curves, are employed to describe complicated boundary shapes. More recently, Yi and Kim proposed a method to identify the boundaries of topology optimization results using basic parametric features, such as lines, arcs, circles, and fillets [11]. This method can deal with complicated boundary shapes with a relatively modest number of fitting variables. S. Liu et al. presented a method for interpreting topology results into smooth, parametric CAD models by using an adaptive B-spline fitting method after several preprocessing steps, and stereo lithography (STL) models for additive manufacturing [12]. Nevertheless, most of previously proposed methods are parametric methods. These methods are effective to reduce the design variables, but require the time-consuming shape parameterization preprocess in advance, especially for complicated structure. Moreover, the obtained shapes are strongly dominated by the parameterization process.

In this paper, a parameter-free methodology is proposed for boundary identification of topology optimization results. Firstly, the topology optimization is carried out by the density approach without any sensitivity filtering. Secondly, checkerboard patterns are filled with material by an automatic algorithm that accompany the element removal procedure. At this point, a boundary smoothing technique is applied to regulate the zig-zag elements at the design boundary to be distinct and smooth enough for the subsequent parameter-free shape optimization. Finally, the $H^1$ gradient method [13] is employed to the newly established boundaries to obtain optimal smooth shapes.

## Formulation of the topology optimization problem

Consider a structure occupying a domain $\Omega^{mat}$, which is part of a larger reference domain $\Omega$ in $\mathbf{R}^d$ (d=2 or 3). The reference domain $\Omega$ is chosen so as to allow for a definition of the applied loads and boundary conditions and the reference domain is sometimes called the ground structure. Referring to the reference domain $\Omega$, the optimal design problem can be defined as the problem of finding the optimal choice of stiffness tensor $E_{ijkl}(x)$, which is a variable over the domain [14]. Here, let's consider the mean compliance minimization (global stiffness maximization) problem, which is widely used for testing various new methods in the literature. The formulation of this optimization problem can be written as follows:

$$\min_{\boldsymbol{u} \in U} \quad l(\boldsymbol{u}) \tag{1}$$

$$\text{Subject to} \quad a(\boldsymbol{u}, \boldsymbol{w}) = l(\boldsymbol{w}), \text{ for all } \boldsymbol{w} \in U \tag{2}$$

$$V = \int_{\Omega} 1 d\Omega \leq \widehat{V} \tag{3}$$

Here the equilibrium equation expressed in Eq. (2) is written in its weak, variational form, with $U$ denoting the space of kinematically admissible displacement fields. Moreover, $\boldsymbol{w} = \{w_1, w_2, w_3\}$ expresses adjoint variables of the variational equilibrium equation. $\hat{V}$ denotes constraint value of the volume $V$. In addition, the bilinear forms $a(\boldsymbol{u}, \boldsymbol{w})$ and the linear form $l(\boldsymbol{w})$ are defined in Eqns. (4) and (5), respectively, using tensor representation.

$$a(\boldsymbol{u}, \boldsymbol{w}) = \int_\Omega E_{ijkl} u_{k,l} w_{i,j} d\Omega \tag{4}$$

$$l(\boldsymbol{w}) = \int_\Omega \boldsymbol{f} \boldsymbol{w} d\Omega + \int_\Gamma \boldsymbol{t} \boldsymbol{u} ds \tag{5}$$

where $E_{ijkl}$ is the elasticity tensor and $E_{ijkl} = E_{klij} = E_{jikl} = E_{ijlk}$. $\boldsymbol{f}$ are the body forces and $\boldsymbol{t}$ indicate the boundary tractions on the traction part $\Gamma_t \subset \Gamma \equiv \partial\Omega$ of the boundary.

An extremely efficient and commonly used method to solve the formulated topology optimization problem, is the SIMP (Solid Isotropic Material with Penalization) method. The basic concept of the SIMP method is to use the "density" $\rho(x)$ to describe that a point $x$ is filled with a given isotropic material ($\rho$=1) or void ($\rho$=0). Then, the volume of the structure can be evaluated as $\int_\Omega \rho(x) d\Omega$. In other words, this topology optimization problem can be mathematically formulated as a distributed, discrete valued design problem (a 0-1 problem) as follows:

$$\rho(x) = \begin{cases} 1 & \text{if } x \in \Omega^{\text{mat}} \\ 0 & \text{if } x \in \Omega \setminus \Omega^{\text{mat}} \end{cases} \tag{6}$$

In the SIMP, another point is to define the stiffness tensor $E_{ijkl}$ of the point $x$ as follows, in order to solve the formulated 0-1 problem.

$$E_{ijkl}(x) = \rho(x)^p E_{ijkl}^0, \quad p > 1 \tag{7}$$

where, $E_{ijkl}^0$ represents the material properties of a given isotropic material. $p$ indicates the penalization factor, and $p \geq 3$ is usually required to ensure a good convergence. It means that if a final design has density 0 or 1 in all points, this design is a black-and-white design, the performance has been evaluated with a correct physical model. Unfortunately, the topology optimization performed by the density-based approach tends to suffer from numerical instability problems, including mesh dependency, checkerboard patterns and grayscales. These drawbacks result in the topology results cannot be directly transformed to models for manufacturing.

**Parameter-free methodology for boundary identification**
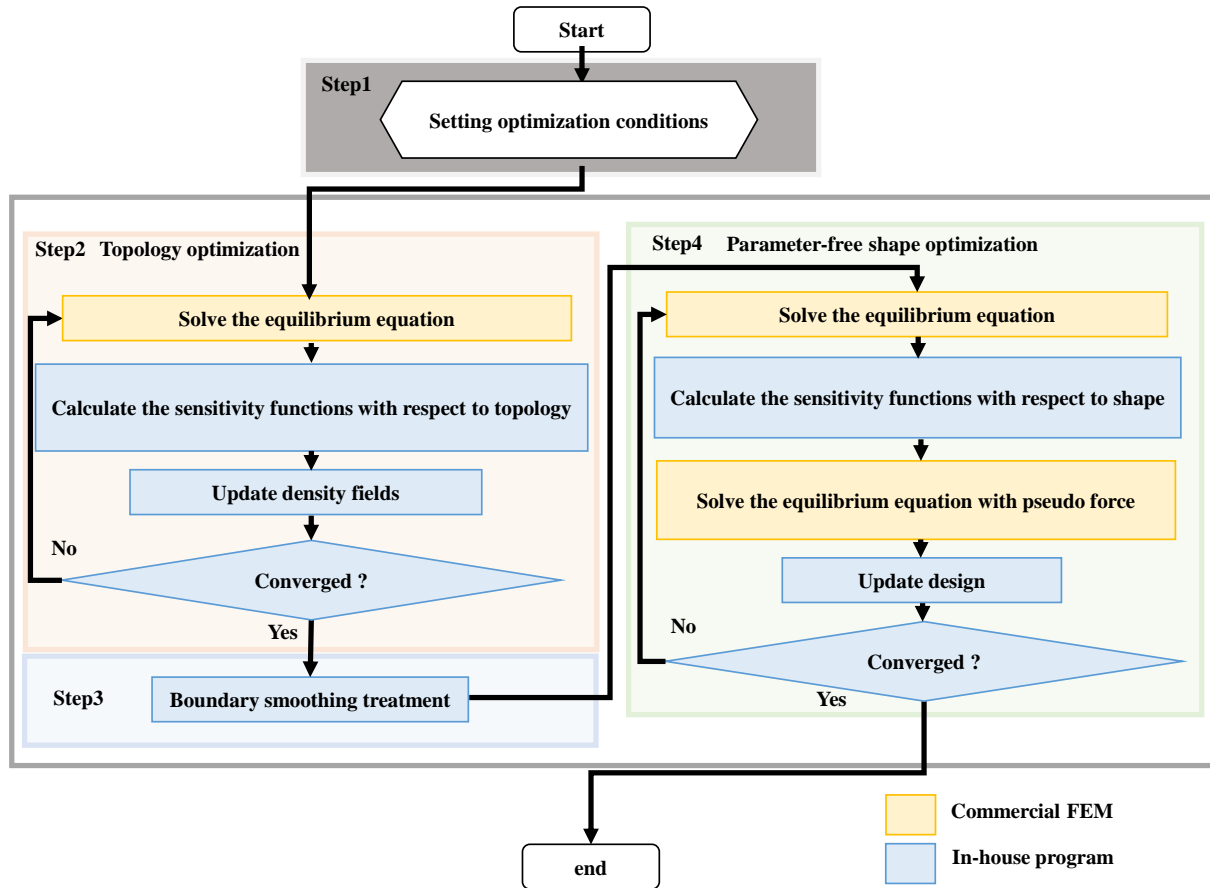
Fig.1 Flowchart diagram of the parameter-free methodology for boundary identification

This section illustrates a summary of the main steps involved in our methodology for the boundary identification. Fig.1 shows the flowchart diagram of the boundary identification system, which is composed of in-house C-codes and a commercial FEM code. First, the topology optimization is carried out by the density approach without any sensitivity filtering. Then, the boundary smoothing technique is initiated once the topology optimization has reached its termination criterion. In this step, checkerboard patterns are filled with material by an automatic algorithm that accompany the element removal procedure. Furthermore, a boundary smoothing technique is applied to regulate the zig-zag elements of the boundary to be distinct and smooth enough for the subsequent parameter-free shape optimization.

The boundary smoothing is performed with geometric data only, which may result in increase of the objective functional. Thus, shape optimization method should be employed to the newly established boundaries to minimize the objective functional, and to obtain optimal smooth shape. The parameter-free shape optimization method proposed here for the boundary identification of topology optimization results is based on the H1 gradient method, which is also called the traction method and is a type of gradient method in a Hilbert space. It is a node-based shape optimization method that can treat all nodes as design variables and does not require any shape design parametrization. The original H1 gradient method was proposed by Azegami in 1994 [13], and was modified by Shimoda and Liu to perform the free-form shell optimization [15], and the shape optimum design of stiffeners on thin-walled structures [16].

In this work, we apply the $H^1$ gradient method to obtain the optimal boundary shape variation of topology optimization results. Here, the Dirichlet conditions are defined for a pseudo-elastic body. A distributed force proportional to the shape gradient function $-Gn$ is applied in

the normal direction of boundary. The analysis for the optimal boundary shape variation is called the "velocity analysis". The shape gradient function is not applied directly to the shape variation but rather is replaced by a force, to vary the shape of boundary. This makes it possible both to reduce the objective functional and to maintain the smoothness, i.e., mesh regularity.

**Numerical example**

The proposed methodology was applied to a numerical example in order to confirm its validity and practical utility for the boundary identification of topology optimization results. Fig. 2(a) shows the fixed design domain and the boundary conditions of a cantilever model, which occupies a rectangle domain of size 500mm×400mm and is discretized by 50×40 elements. The left-hand edge of the model is fully constrained and the middle portion of the right-hand edge is subject to a concentrated load acting in the negative vertical-direction. In this example, the model is made of an isotropic linear elastic material with Young's modulus $E= 210$GPa, Poisson's ratio $v = 0.31$.

First of all, the topology optimization is carried out to minimize the mean compliance by the density approach without any sensitivity filtering. The results is shown in Fig. 2(b), where the black elements represent the portions filled with the isotropic material and the white elements represent the void (no material). Obviously, grey density elements with density range $0<\rho<1$ and the checkerboard pattern exist in the topology optimization results. These grey elements has no physical meaning and is difficult to be dealt with in the manufacture process.

Here, we apply the proposed parameter-free methodology to identify the distinct and smooth boundaries of the topology optimization results. Following the above procedures of the proposed methodology, the boundary identification of the cantilever model is conducted, and the details of each step are shown in Fig. 3, where the color contour plots represent the distribution of the topological or shape sensitivity and are scaled to the same range of values. Fig. 3(a) to (d) provide the details of analysis results of the initial shape, the results after the density based topology optimization, the obtained shape by the zigzagging boundaries regulation and the smoothing treatment and the identified optimal shape via the $H^1$ gradient method, respectively. Compared with the results shown in Fig. 3, distinct and smooth boundaries are identified by the proposed method and are more suitable for manufacturing.



(a) Design domain        (b) Topology optimization result
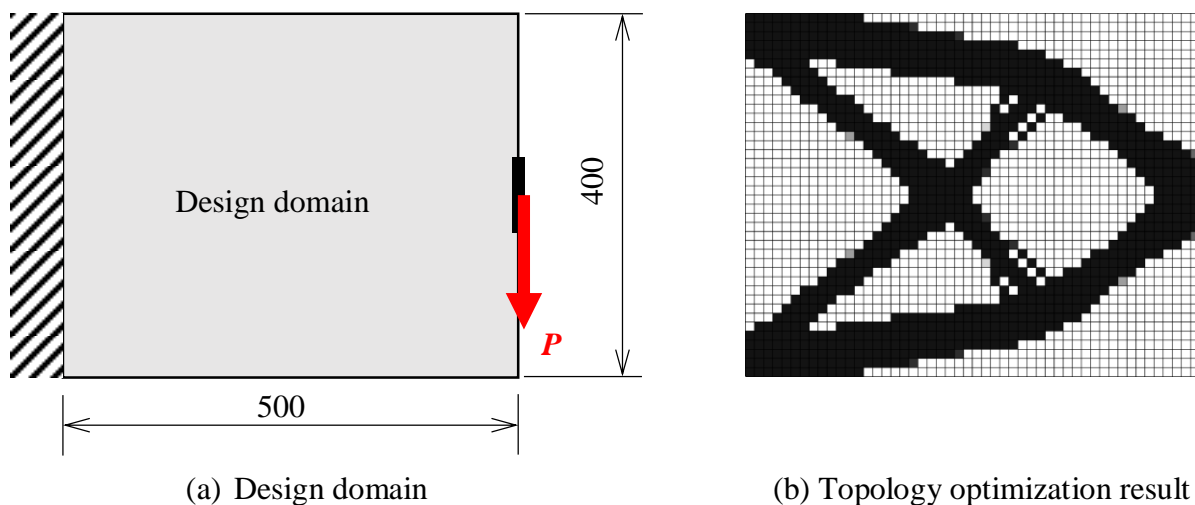
Fig. 2 Design domain of a cantilever model and its topology optimization result.

(a) Initial shape



(b) topology optimization



(c) Boundary smoothing



(d) Shape optimization
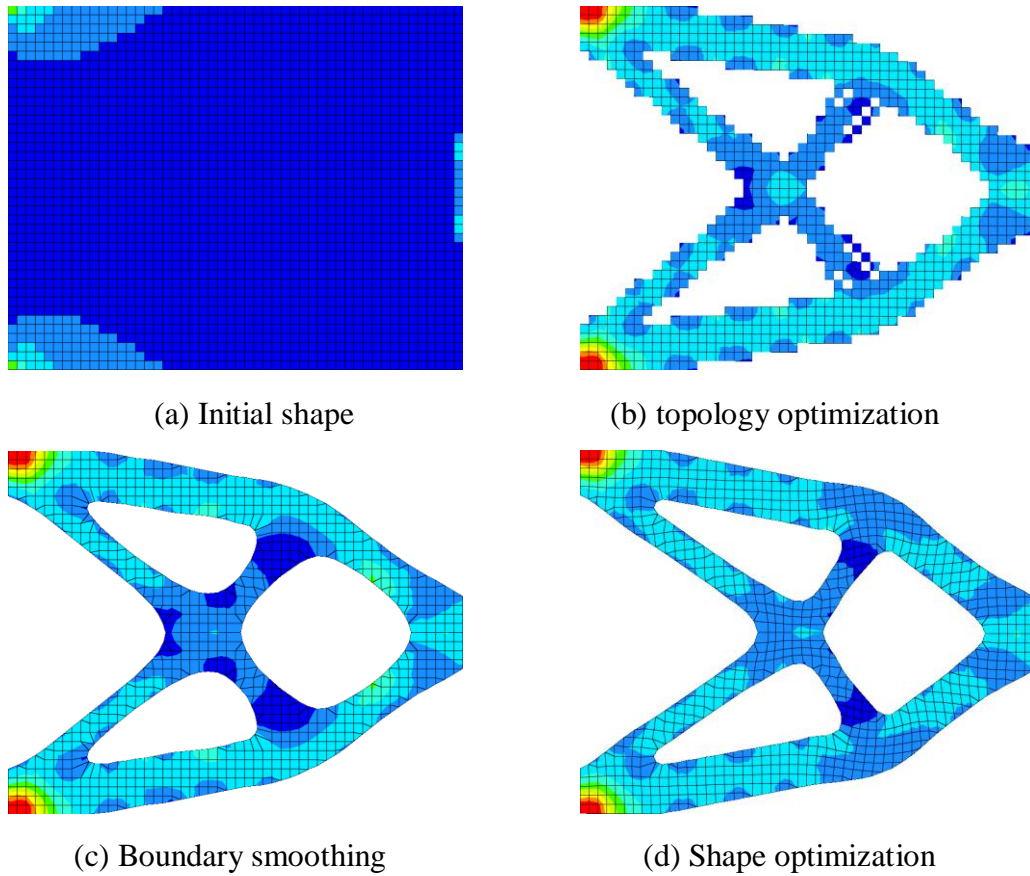
Fig. 3 Distribution of the topological/shape sensitivity of configuration obtained in each step.



(a) 50×40 mesh



(b) 100×80 mesh



(c) 200×160 mesh
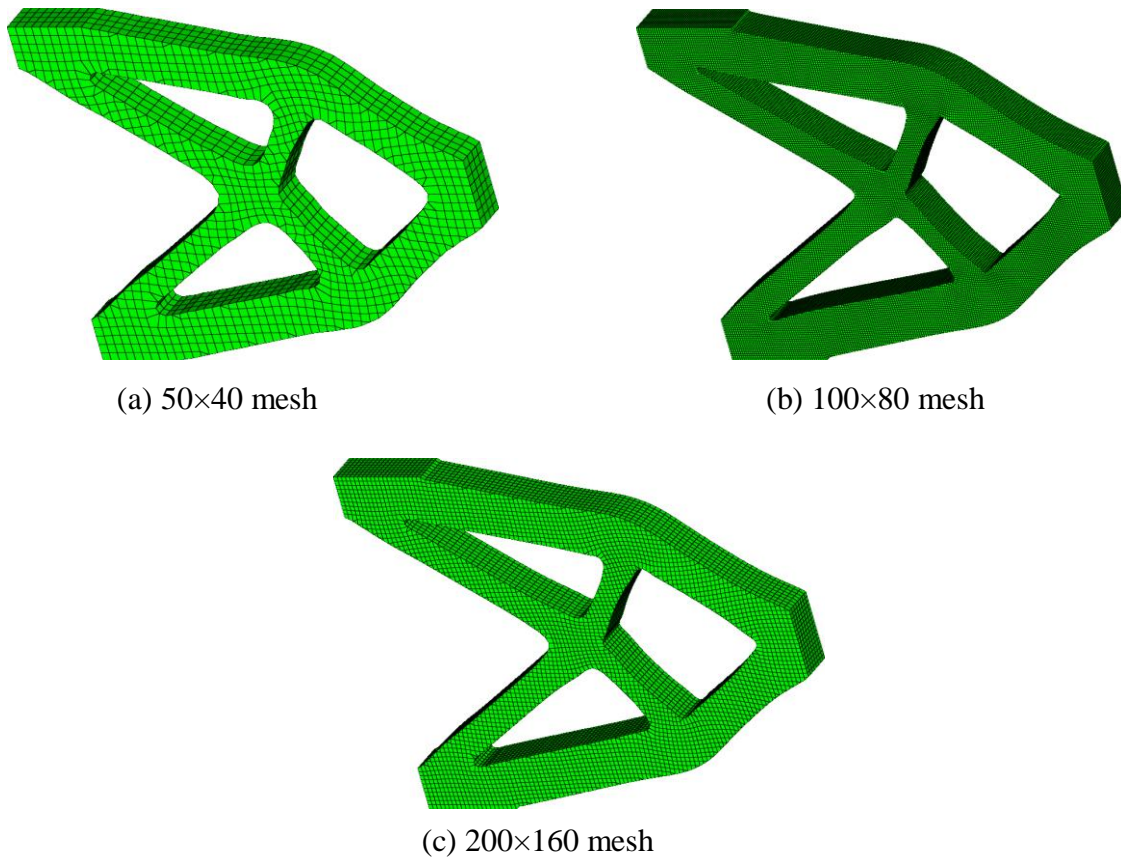
Fig. 4 Optimal configurations using different meshes

As shown in Fig.4, using the same model, the effect of the mesh size upon the resulting optimal configurations is investigated in three cases, in which degree of discretization is subject to the following mesh parameters: 50×40, 100×80 and 200×160. The comparison of optimized configurations obtained in the three cases is shown in Fig 6, where all of the obtained optimal configurations are distinct, smooth, and practically identical. It is confirmed that dependency with regard to the mesh size is extremely small, and an appropriate optimal configuration can be obtained even if the model is discretized by coarse meshes.

## Conclusions

This paper proposed a parameter-free methodology to identify the boundaries of topology optimization results. The proposed methodology based on the boundary smoothing technique and the $H^1$ gradient method, which is a node-based, parameter-free shape optimization method. The validity and practical utility of this method were verified through a design example. The main advantages of this method are summarized as follows:

・Optimal smooth boundaries can be determined without any shape design parameterization.

・Re-mesh is not necessary in the shape updating process, because mesh smoothing is simultaneously implemented by the $H^1$ gradient method.

・It can be easily implemented in combination with a commercial FEA code and the process is fully automatic, which means it has generality and practical utility for actual design work.

## Acknowledgment

## References

[1] Bendsøe, M.P. and Kikuchi, N. (1988) Generating optimal topologies in structural design using a homogenization method, *Computer Methods in Applied Mechanics and Engineering* **71**, 197–224.

[2] Bendsøe, M.P. and Sigmund, O. (1999) Material interpolation schemes in topology optimization, *Archive of Applied Mechanics* **69**,635–654.

[3] Allaire, G., Jouve, F. and Toader, A.M (2002) A level-set method for shape optimization, *Comptes Rendus Mathematique* **334**, 1125–1130.

[4] Xie, Y.M. and Steven, G.P. (1993) A simple evolutionary procedure for structural optimization, *Computers and Structures* **49**, 885–896.

[5] Sokołowski, J. and Zochowski, A. (1999) On the topological derivative in shape optimization, *SIAM Journal on Control and Optimization* **37**, 1251–1272.

[6] Bourdin, B. and Chambolle, A. (2003) Design-dependent loads in topology optimization, *ESAIM: Control, Optimisation and Calculus of Variations* **9**,19–48.

[7] Sigmund, O. and Maute, K. (2013) Topology optimization approaches, *Structural and Multidisciplinary Optimization* **48**, 1031–1055.

[8] Hsu, Y.L., Hsu, M.S. and Chen, C.T. (2001) Interpreting results from topology optimization using density contours, *Computers and Structures* **79**, 1049–1058.

[9] Larsen, S. and Jensen, C.G. (2009) Converting topology optimization results into parametric CAD models, *Computer-Aided Design and Applications* **6**, 407–418.

[10] Koguchi, A. and Kikuchi, N. (2006) A surface reconstruction algorithm for topology optimization. *Engineering with Computers* **22**, 1–10.

[11] Yi, G. and Kim, NH. (2017) Identifying boundaries of topology optimization results using basic parametric features. Structural and Multidisciplinary Optimization 55, 1641–1654.

[12] Liu, S.T., Li, Q.H., Liu, J.H., Cheng, W.J. and Zhang, Y.C. (2018) A realization method for transforming a topology optimization design into additive manufacturing structures, *Engineering* **4**, 277–285.

[13] Azegami, H. (1994) Solution to domain optimization problems, *Transactions of the Japan Society of Mechanical Engineers*, *Series A* **60**, 1479-1486. (in Japanese)

[14] Bendsøe, M.P. and Sigmund, O. (2013) Topology optimization: theory, methods, and applications.

[15] Shimoda, M. and Liu, Y. (2014) A non-parametric free-form optimization method for shell structures, *Structural and Multidisciplinary Optimization* **50**, 409-423.

[16] Liu, Y. and Shimoda, M. (2015) Non-parametric shape optimization method for natural vibration design of stiffened shells, *Computers and Structures* **146**, 20-31.

# Analysis of failure process of bonded pipe joints under tension loads

## Hong YUAN¹, Jun HAN¹, †*Lan ZENG¹, Ziyong MO¹

¹MOE Key Laboratory of Disaster Forecast and Control in Engineering, School of Mechanics and Construction Engineering, Jinan University, Guangzhou 510632, China

*Presenting author: zenglan@jnu.edu.cn
†Corresponding author: zenglan@jnu.edu.cn

**Abstract**

In order to numerically describe the whole failure process of bonded pipe joints under tension loads which is common in mechanical engineering, a cohesive interface modeling approach about debonding analysis of adhesively bonded interface of pipe joints is presented in this paper. A rigid-softening cohesive model was adopted to characterize the deformation of the interface. The whole failure process divided into different sections was discussed in detail and closed-form solutions of bond slip and shear stress were obtained. Finite element method (FEM) model for the bonded pipe joints under certain situations was established using commercial software ABAQUS, and the analytical and FEM results were compared. The stress transfer mechanism, the interface crack propagation and the ductility behavior of the joints under tension loads could be explained. These outcomes do help for the design and application of bonded pipe joints and can be extended to other orthotropic materials.

**Keywords:** Adhesively bonded; Cohesive law; Pipe joints; Failure process; Tension loads

## 1. Introduction

Pipe structure is a very important structural form for energy and construction industries. With the development of material science and manufacturing, the mechanical properties of pipe have been dramatically improved. However, the limitations of the overall system performance usually come from the capacity of pipe joints. Therefore, the pipe joints play the most important role in the overall integrity of most piping systems [1][2].

Among all the possible loading configurations, tension loading is one of the fundamental loading types. Because of the difficulties in the analysis of interfacial behavior, few theoretical studies are available in the literature. Moreover, all the existing analytical studies were focused on the elastic region of interfacial behavior. Lubkin and Reissner [5] used the ordinary thin-shell theory to study the adhesive shear and normal stresses and Poisson's effect was taken into consideration. Based on the Lubkin and Reissner model, an explicit closed-form solution was obtained by means of the Laplace transform [6][7]. By means of the principle of minimum complementary energy, closed-form solutions are obtained by Shi and Cheng [8]. Nemes et al. [9][10] introduced $\sigma_{rr}$ and $\sigma_{zz}$ into the potential energy formulation to predict the intensity and the distributions of stresses. In order to understand the mechanical behavior, Yang [11] developed an analytical model based on the first-order laminated anisotropic plate theory. Adhesive peel stress and shear stress distributions are obtained. Pugno and Carpinteri [12] focused on both static and dynamic behavior of the joint. A fracture energy criterion to predict brittle crack propagation for conventional and optimized joint was presented.

Finite element methods have been used increasingly since the 1970s to analyse lap joints. Finite element and experimental results are good supplements to analytical research. The analysis of adhesively bonded joints has been based, generally, on the assumption that the adhesive behaves as a linearly elastic material. Many adhesives, however, exhibit nonlinear stress-strain behaviour, particularly near failure. Nagaraja and Alwar [13] reported an attempt on the analysis of an adhesive tubular lap joint with the adhesive obeying a nonlinear stress-strain law. The cross section was modelled with two dimensional finite element mesh by Hossein and Ochsner [14]. Based on Tsai-Wu coupled stress criterion, three-dimensional stress analyses had been carried out [15]-[17]. By using a linear elastic fracture mechanics methodology, Reedy and Guess [18] studied the axial strength and fatigue resistance of tubular lap joints. Cognard et al. [19] analysed stress concentrations for the first time, starting from refined finite element computations under a linear elastic assumption. The main parameter of failure process is interface slip which is difficult to be measured during experiment. To the authors' acknowledgement, very little experimental data can be found in the existing references. An experimental study was concerned with the monotonic and cyclic damage behavior of adhesively bonded glass-fiber reinforced epoxy polymer tubes [20]. The results revealed characteristic damage modes for the pipe body and the joint area.

Cohesive zone model has evolved as a preferred method to analyse fracture problems in composite material systems because it avoids the singularity and can be easily implemented in a numerical method of analysis such as in finite element modeling [3][4]. It is generally accepted that cohesive zone models can be described by two or three independent parameters [21]. These parameters may be the fracture toughness, the cohesive strength $\sigma_f$ (or $\tau_f$) and the shape of the cohesive law. Based on the existing cohesive zone models, some analytical solutions have also been developed to solve interface debonding problem for the adhesively bonded joints [22]-[27].

In the present study, failure process of bonded interface based on rigid-softening cohesive zone model is investigated. The expressions for the interface slip and shear stress are derived for the different loading stages. And the maximum load capacity is derived. The simple expressions of analytical study can be directly used for practical design.

## 2. Interface model of pipe joints

### 2.1. Interface model and assumptions

The inner and outer pipes are bonded together by a thin and soft adhesive layer shown in Fig. 1. Here the inner and outer pipes are defined as pipe 1 and 2, respectively. Due to symmetry, only the right half of the pipe joints is considered. The distance between the left end of pipe 1 and the right end of pipe 2 is expressed by $L$.



(a) A cross-sectional view          (b) A side view
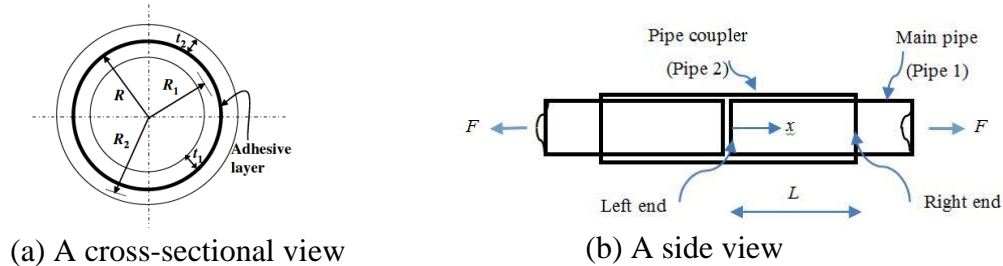
**Figure 1. Adhesively bonded pipe joints.**

Before starting the derivations, the basic assumptions adopted in the present study are summarized as follows:

(1) The adherents are homogeneous and linear elastic;
(2) The adhesive is only exposed to shear forces;
(3) Radial displacements produced by the axial stress because of Poisson's effect are neglected.
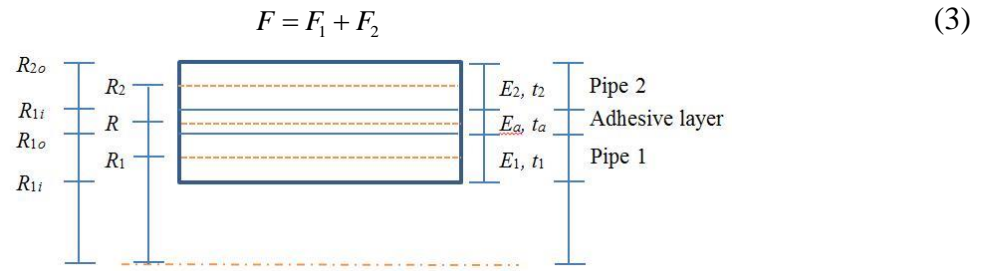
*2.2. Governing equation*

If at the given cross-section as illustrated in Fig.2, the slips of pipe 1 and 2 are different from each other, a relative slip occurs accompanied by a longitudinal relative displacement at the bond layer. Considering the elastic constitutive law of two tubes, their axial force $F_1$ and $F_2$ are written as:

$$F_1 = E_1 A_1 \frac{du_1}{dx} \tag{1}$$

$$F_2 = E_2 A_2 \frac{du_2}{dx} \tag{2}$$

where $E_i$, $A_i$, $u_i$ are Young's modulus, cross section and axial displacement of pipe $i=1$ and 2, respectively. $R_i$, $R_{ii}$, $R_{io}$, $t_i$ are the average radius, inner radius, outer radius, thickness of pipe $i=1$ and 2, respectively. $E_a$ and $t_a$ are Young's modulus and thickness of the adhesive layer. According to the above assumptions, the tension load carried by the soft and thin adhesive layer is ignored. Thus, the equilibrium between external and internal tension load in the pipe joints requires:

$$F = F_1 + F_2 \tag{3}$$



(a) Schematic for the right half of pipe joints
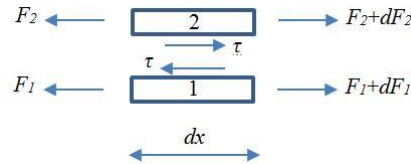
(b) Infinitesimal isolated body

**Figure 2. Deformation and equilibrium in the bonded joints.**

The interface slip $\delta$ is defined as the relative displacement of two bonded pipes:

$$\delta = u_1 - u_2 \tag{4}$$

In this model the study starts from the axial equilibrium, considering pipe 1 in Fig. 2.

$$\tau = \frac{1}{2\pi R} \frac{dF_1}{dx} \tag{5}$$

where $\tau$ is the interfacial shear stress along the axial direction and $R$ is the distance between the center of the pipe and mid-height of the adhesive layer. Substituting Eqs.(1)-(4) into Eq.(5) yields, by introducing interfacial fracture energy $G_f$.

$$\frac{d^2\delta}{dx^2} - \frac{2G_f}{\tau_f^2} \lambda^2 \tau = 0 \tag{6}$$

$$F_1 = \frac{\tau_f^2}{2G_f} \frac{2\pi R}{\lambda^2} \left( \frac{d\delta}{dx} + \frac{F}{E_2 A_2} \right) \tag{7}$$

where

$$\lambda^2 = \frac{\tau_f^2}{2G_f} 2\pi R \left( \frac{1}{E_1 A_1} + \frac{1}{E_2 A_2} \right) \tag{8}$$

Eq.(6) is the governing differential equation of the adhesive bonded joints in Fig. 1. When the local bond-slip law $\tau = \tau(\delta)$ is found, this equation can be solved.

*2.3. Bond-slip law*
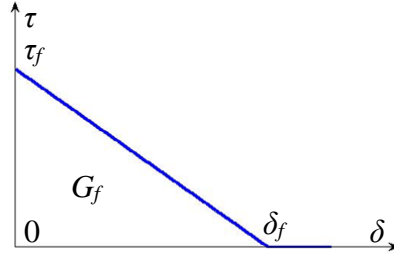


**Figure 3. Bond-slip model.**

As shown in Fig. 3, the rigid-softening model has only a softening branch and can be obtained from the bi-linear model by omitting the linear ascending branch (i.e. $\delta_1 = 0$). The elastic deformation $\delta_1$ at the peak bond stress $\tau_f$ is much smaller than the ultimate slip $\delta_f$ when the shear stress reduces to zero, which signifies the shear fracture (or debonding or macro-cracking) of a local bond element. Therefore, this model can be treated as simplifications of the bi-linear bond-slip law. The rigid-softening model is described by the following equation:

$$\tau = f(\delta) = \begin{cases} 0 & (\delta = 0) \\ \dfrac{\tau_f}{\delta_f}(\delta_f - \delta) & (0 < \delta \le \delta_f) \\ 0 & (\delta > \delta_f) \end{cases} \tag{9}$$

When $\delta = 0$, it is called a rigid stage which is simplified as R. When $0 < \delta \le \delta_f$, it is called a softening stage which is simplified as S. When $\delta > \delta_f$, it is called a debonding stage which is simplified as D. Once the bond-slip law is defined, the governing equation (i.e. Eq. (6)) can be solved to find the shear stress distribution along the interface and the load-displacement response of the bonded joints. The states of the interface and the possible debonding processes are discussed below.

Substituting Eq. (9) into Eq. (6) givens the following differential equation for different stages:

$$\frac{d^2\delta}{dx^2} = 0 \quad \text{for rigid stage} \tag{10}$$

$$\frac{d^2\delta}{dx^2} - \lambda_1^2 \left[ \delta_f - \delta(x) \right] = 0 \quad \text{for softening stage} \tag{11}$$

where

$$\lambda_1^2 = \frac{2G_f}{\tau_f^2} \frac{\tau_f}{\delta_f} \lambda^2 = \frac{\tau_f}{\delta_f} 2\pi R \left( \frac{1}{E_1 A_1} + \frac{1}{E_2 A_2} \right) \tag{12}$$

**3. Analysis of the failure process**

*3.1. SRS stage*

At small load, the whole length of the interface is in a softening-rigid-softening state (SRS stage) and experiences no debonding. The lengths of softening regions at the left and right sides are denoted by $a_L$ and $a_R$, respectively. The differential equation for this stage is Eq. (11) with boundary conditions:

$$F_1(0) = 0 \tag{13}$$

$$F_1(L) = F \tag{14}$$

$$\delta(a_L) = 0 \tag{15}$$

$$\delta(L - a_R) = 0 \tag{16}$$

In the whole bond length, the distribution of axial normal stress is continuous. Therefore, it can be known from Eqs. (1) and (7) that $\delta'(x)$ is related to $F_1$. Thus $\delta'(x)$ is continuous as expressed below:

$$\delta'(x) \text{ is continuous at } x = a_L \tag{17}$$

$$\delta'(x) \text{ is continuous at } x = L - a_R \tag{18}$$

Based on the conditions (15) and (17), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $0 \le x \le a_L$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \cos\left[\lambda_1 (x - a_L)\right] \tag{19}$$

$$\tau(x) = \tau_f \cos\left[\lambda_1 (x - a_L)\right] \tag{20}$$

The expressions of shear displacement and stress for rigid part of the bond length in the middle can be easily obtained based on the bond-slip law. Therefore, the expressions for the rigid region are not given. In addition, based on the conditions (16) and (18), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $L - a_R \le x \le L$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \cos\left[\lambda_1 (L - a_R - x)\right] \tag{21}$$

$$\tau(x) = \tau_f \cos\left[\lambda_1 (L - a_R - x)\right] \tag{22}$$

The interface slips at the two ends are defined as $\Delta_L$ and $\Delta_R$, respectively. According to this definition, $\Delta_L$ and $\Delta_R$ can be obtained from Eqs. (19) and (21):

$$\Delta_L = \delta_f - \delta_f \cos(\lambda_1 a_L) \tag{23}$$

$$\Delta_R = \delta_f - \delta_f \cos(\lambda_1 a_R) \tag{24}$$

Based on the conditions (13)-(14), the expressions of load can be obtained:

$$\delta_f \lambda_1 \sin(\lambda_1 a_L) = \frac{F}{E_2 A_2} \tag{25}$$

$$\delta_f \lambda_1 \sin(\lambda_1 a_R) = \frac{F}{E_1 A_1} \tag{26}$$

By comparing the two above equations, the relationship of the length of two softening regions is given in the form:

$$\sin(\lambda_1 a_R) = \rho \sin(\lambda_1 a_L) \tag{27}$$

where

$$\rho = \frac{E_2 A_2}{E_1 A_1} \tag{28}$$

It can be known from Eqs. (23)-(24) and (27) that the relationship of interface slips at the two ends is related to the non-dimensional parameter $\rho$. If $\rho = 1$, the interface slips at the two ends are the same and the two ends enter into debonding at the same time. If $\rho$ is not equal to 1, the interface slips at the two ends are different and the sequence of entering into debonding should

be considered. The shear stress at the right end reaches $\delta_f$ first when $\rho>1$ since slip at the right end is larger than that at the left end. When $\rho<1$, the result is the opposite. Therefore, two cases of $\rho=1$ and $\rho>1$ are considered without loss of generality. The load-displacement curve could be obtained from Eqs. (23)-(26).

### 3.2. Case ρ=1

As load increases, shear stresses at the two ends reach $\tau_f$ and the interface enters into DSRSD stage. If the bond length is not long enough, the whole interface enters into softening region while the shear stresses at the two ends are still less than $\tau_f$. Therefore, there exists a critical bond length to distinguish the following failure process. By letting $\Delta_L=\Delta_R=\delta_f$ and $a_L+a_R=L$, the critical bond length can be easily obtained as:

$$L_{cr1} = \frac{\pi}{\lambda_1} \tag{29}$$

If $L>L_{cr1}$, the interface enters into DSRSD stage from SRS stage. But if $L<L_{cr1}$, the interface enters into S stage from SRS stage.

### 3.2.1. L<Lcr1

### 3.2.1.1. Critical state

At the end of SRS stage when $a_L=a_R=L/2$, the interface enters into S stage. The expression of ultimate load can be obtained from Eqs. (25)-(26):

$$F_u = E_1 A_1 \delta_f \lambda_1 \sin\left(\frac{1}{2}\lambda_1 L\right) = E_2 A_2 \delta_f \lambda_1 \sin\left(\frac{1}{2}\lambda_1 L\right) \tag{30}$$

### 3.2.1.2. S stage

The differential equation of S stage is Eq. (11) with boundary conditions (13)-(14). The expressions of displacement and shear stress can be written as the follows:

$$\delta = \delta_f - \frac{F}{E_1 A_1}\frac{1}{\lambda_1}\frac{\cos(\lambda_1 x)}{\sin(\lambda_1 L)} - \frac{F}{E_2 A_2}\frac{1}{\lambda_1}\frac{\cos\left[\lambda_1(L-x)\right]}{\sin(\lambda_1 L)} \tag{31}$$

$$\tau_f = \frac{\tau_f}{\delta_f}\frac{F}{E_1 A_1}\frac{1}{\lambda_1}\frac{\cos(\lambda_1 x)}{\sin(\lambda_1 L)} + \frac{\tau_f}{\delta_f}\frac{F}{E_2 A_2}\frac{1}{\lambda_1}\frac{\cos\left[\lambda_1(L-x)\right]}{\sin(\lambda_1 L)} \tag{32}$$

In this stage, the interface slips at the left and right ends can be obtained as:

$$\Delta_L = \delta_f - \frac{F}{E_1 A_1}\frac{1}{\lambda_1}\frac{1}{\sin(\lambda_1 L)} - \frac{F}{E_2 A_2}\frac{1}{\lambda_1}\cot(\lambda_1 L) \tag{33}$$

$$\Delta_R = \delta_f - \frac{F}{E_1 A_1}\frac{1}{\lambda_1}\cot(\lambda_1 L) - \frac{F}{E_2 A_2}\frac{1}{\lambda_1}\frac{1}{\sin(\lambda_1 L)} \tag{34}$$

The load-displacement curve could be obtained from Eqs. (33) and (34).

### 3.2.2. L>Lcr1

### 3.2.2.1. Critical state

At the end of SRS stage, the two softening lengths can be derived as $a_L=a_R=\pi/(2\lambda_1)$. When $\Delta_L=\Delta_R=\delta_f$, the interface enters into DSRSD stage. The expression of ultimate load can be obtained from Eqs. (25)-(26):

$$F_u = E_1 A_1 \delta_f \lambda_1 = E_2 A_2 \delta_f \lambda_1 \tag{35}$$

### 3.2.2.2. DSRSD stage

During DSRSD stage, debonding (or macro-cracking or fracture) commences and propagates along the interface. We assume that the debonding lengths of the interface starting at the left and

right ends are $d_L$ and $d_R$, respectively. The differential equations of this stage are Eqs. (10)-(11) with boundary conditions (13)-(14) and:

$$\delta(d_L) = \delta_f \tag{36}$$

$$\delta(L - d_R) = \delta_f \tag{37}$$

$$\delta(d_L + a_L) = 0 \tag{38}$$

$$\delta(L - d_R - a_R) = 0 \tag{39}$$

and continuous conditions are given as follows:

$$\delta'(x) \text{ is continuous at } x = d_L \tag{40}$$

$$\delta'(x) \text{ is continuous at } x = L - d_R \tag{41}$$

$$\delta'(x) \text{ is continuous at } x = d_L + a_L \tag{42}$$

$$\delta'(x) \text{ is continuous at } x = L - d_R - a_R \tag{43}$$

Based on the conditions (36) and (40), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $0 \leq x \leq d_L$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \lambda_1 \sin(\lambda_1 a_L)(x - d_L) \tag{44}$$

$$\tau(x) = 0 \tag{45}$$

Based on the conditions (38) and (42), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $d_L \leq x \leq d_L + a_L$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \cos\left[\lambda_1(x - d_L - a_L)\right] \tag{46}$$

$$\tau(x) = \tau_f \cos\left[\lambda_1(x - d_L - a_L)\right] \tag{47}$$

Based on the conditions (39) and (43), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $L - d_R - a_R \leq x \leq L - d_R$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \cos\left[\lambda_1(L - d_R - a_R - x)\right] \tag{48}$$

$$\tau(x) = \tau_f \cos\left[\lambda_1(L - d_R - a_R - x)\right] \tag{49}$$

In addition, based on the conditions (37) and (41), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $L - d_R \leq x \leq L$ can be written in the form:

$$\delta(x) = \delta_f - \delta_f \lambda_1 \sin(\lambda_1 a_R)(L - d_R - x) \tag{50}$$

$$\tau(x) = 0 \tag{51}$$

Based on the conditions (13)-(14), the expressions of load can be obtained:

$$F = E_2 A_2 \delta_f \lambda_1 \sin(\lambda_1 a_L) \tag{52}$$

$$F = E_1 A_1 \delta_f \lambda_1 \sin(\lambda_1 a_R) \tag{53}$$

Substituting conditions (36)-(37) into Eqs. (46) and (48), respectively, expression of the lengths of softening regions is derived as the following equation:

$$a_L = a_R = \frac{\pi}{2\lambda_1} \tag{54}$$

Therefore, the debonding lengths increase while the softening lengths stay the same during this stage. Thus the two softening regions move to the middle. In this stage, the interface slips can be obtained as:

$$\Delta_L = \delta_f + \delta_f \lambda_1 \sin(\lambda_1 a_L) d_L \tag{55}$$

$$\Delta_R = \delta_f + \delta_f \lambda_1 \sin(\lambda_1 a_R) d_R \tag{56}$$

The load-displacement curve could be obtained from (52)-(56). When $a_L + a_R + d_L + d_R = L$, the interface enters into DSD stage. At the end of this stage, the length of softening zone is defined as $a_u$. Based on this definition, the expression of $a_u$ is given as:

$$a_u = \frac{\pi}{\lambda_1} \tag{57}$$

### 3.2.2.3. DSD stage

When the stress peaks reach together, the interface enters into DSD stage. This stage is governed by (10)-(11) with boundary conditions (13)-(14), (36)-(37) and continuous conditions (40)-(41). The length of softening region in this stage is defined as *a*. Based on the conditions (13) and (36), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $0 \leq x \leq d_L$ can be written in the form:

$$\delta(x) = \delta_f - \frac{F}{E_2 A_2}(x - d_L) \tag{58}$$

$$\tau(x) = 0 \tag{59}$$

Based on the conditions (36) and (40), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $d_L \leq x \leq L - d_R$ can be written in the form:

$$\delta(x) = \delta_f - \frac{F}{E_2 A_2} \frac{1}{\lambda_1} \sin\left[\lambda_1(x - d_L)\right] \tag{60}$$

$$\tau(x) = \frac{\tau_f}{\delta_f} \frac{F}{E_2 A_2} \frac{1}{\lambda_1} \sin\left[\lambda_1(x - d_L)\right] \tag{61}$$

In addition, based on the conditions (14) and (37), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $L - d_R \leq x \leq L$ can be written in the form:

$$\delta(x) = \delta_f - \frac{F}{E_1 A_1}(L - d_R - x) \tag{62}$$

$$\tau(x) = 0 \tag{63}$$

Substituting condition (37) into Eq. (60), the expression of softening length can be obtained:

$$a = L - d_L - d_R = \frac{\pi}{\lambda_1} = a_u \tag{64}$$

The above equation indicates the softening length remains constant during this stage. The interface slips at the two ends can be obtained from Eqs. (58) and (62) giving the following load-displacement relationships:

$$\Delta_L = \delta_f + \frac{F}{E_2 A_2} d_L \tag{65}$$

$$\Delta_R = \delta_f + \frac{F}{E_1 A_1} d_R \tag{66}$$

From the two above equations, it can be concluded that the displacement reduces linearly with the load.

### 3.3. Case $\rho > 1$

If $\rho > 1$, the interface slip at the right end is larger than left. As load increases, shear stress at the right end reaches $\tau_f$ and the interface enters into SRSD stage. But if the bond length is not long enough, the whole interface enters into softening region while the shear stress at the right end is still less than $\tau_f$. Therefore, there exists another critical bond length. By letting $\Delta_R = \delta_f$ and $a_L + a_R = L$, the critical bond length can be easily obtained as:

$$L_{cr2} = \frac{1}{\lambda_1} \arccos\left(-\frac{1}{\rho}\right) \tag{67}$$

If $L > L_{cr2}$, the interface enters into SRSD stage from SRS stage. But if $L < L_{cr2}$, the interface enters into S stage from SRS stage.

### 3.3.1. L<Lcr2

### 3.3.1.1. Critical state

At the end of SRS stage when $a_L + a_R = L$, the interface enters into S stage. The expression of

ultimate load can be obtained from Eqs. (25) and (26):

$$F_u = E_1 A_1 \delta_f \lambda_1 \frac{\sin(\lambda_1 L)}{\sqrt{\left[\sin(\lambda_1 L)\right]^2 + \left[1/\rho + \cos(\lambda_1 L)\right]^2}} \tag{68}$$

### 3.3.1.2. S stage

This stage is the same as that in 3.2.1.2. Thus the expressions of the interface slip, the interface shear stress, as well as the relationship of the load-displacement are also the same.

### 3.3.2. L>Lcr2

### 3.3.2.1. Critical state

When $\Delta_R = \delta_f$, the interface enters into SRSD stage. At the end of SRS stage, the two softening lengths can be derived as $a_L = (1/\lambda_1)\arcsin(1/\rho)$ and $a_R = \pi/(2\lambda_1)$. The expression of ultimate load can be obtained from Eq. (26):

$$F_u = E_1 A_1 \delta_f \lambda_1 \tag{69}$$

### 3.3.2.2. SRSD stage

The differential equations of this stage are Eqs. (10)-(11) with conditions (13)-(15), (17), (37), (39), (41) and (43). Based on the conditions (15) and (17), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $0 \le x \le a_L$ can be written the same as (19)-(20). Based on the conditions (39) and (43), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $L - d_R - a_R \le x \le L - d_R$ can be written the same as (48)-(49). In addition, based on the conditions (37) and (41), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $L - d_R \le x \le L$ can be written the same as (50)-(51). Based on the conditions (13)-(14), the expressions of load can be obtained the same as (25)-(26). The lengths of softening regions stay the same which means that the two softening regions move to the middle. In this stage, the left and right slips can be obtained the same as (23) and (56). Thus the load-displacement curve could be obtained from (26) and (56). When $a_L + a_R + d_R = L$, the interface enters into SD stage. At the end of this stage, the length of softening zone is defined as $a_u$. Based on this definition, the expression of $a_u$ is given as:

$$a_u = \frac{1}{\lambda_1} \arccos\left(-\frac{1}{\rho}\right) \tag{70}$$

It can be obtained from Eq. (70) that if $\rho = 1$ the result is the same as Eq. (57).

### 3.3.2.3. SD stage

When the stress peaks move together, the interface enters into SD stage. This stage is governed by (10)-(11) with boundary conditions (13)-(14) and (37) and continuous conditions (41). The length of softening region in this stage is defined as $a$. Based on the conditions (13) and (37), the solutions of Eq. (11) for the relative shear displacement and the shear stress of $0 \le x \le a$ can be written in the form:

$$\delta(x) = \delta_f + \frac{F}{E_2 A_2} \frac{1}{\lambda_1} \frac{\sin\left[\lambda_1(L - d_R - x)\right]}{\cos\left[\lambda_1(L - d_R)\right]} \tag{71}$$

$$\tau(x) = -\frac{\tau_f}{\delta_f} \frac{F}{E_2 A_2} \frac{1}{\lambda_1} \frac{\sin\left[\lambda_1(L - d_R - x)\right]}{\cos\left[\lambda_1(L - d_R)\right]} \tag{72}$$

In addition, based on the conditions (37) and (41), the solutions of Eq. (10) for the relative shear displacement and the shear stress of $a \le x \le L$ can be written in the form:

$$\delta(x) = \delta_f + \frac{F}{E_2 A_2} \frac{1}{\cos\left[\lambda_1(L-d_R)\right]}(L-d_R-x) \tag{73}$$

$$\tau(x) = 0 \tag{74}$$

Substituting condition (14) into Eq. (73), the expression of softening length can be obtained:

$$a = L - d_R = \frac{1}{\lambda_1}\arccos\left(-\frac{1}{\rho}\right) = a_u \tag{75}$$

The above equation indicates the softening length remains constant during this stage. In this stage, the interface slips at two ends can be obtained as:

$$\Delta_L = \delta_f + \frac{F}{E_2 A_2}\frac{1}{\lambda_1}\tan\left[\lambda_1(L-d_R)\right] \tag{76}$$

$$\Delta_R = \delta_f - \frac{F}{E_2 A_2}\frac{1}{\cos\left[\lambda_1(L-d_R)\right]}d_R \tag{77}$$

From the above equation, it can be concluded that the displacement reduces linearly with the load.

## 4. Numerical simulations

In this section, numerical examples and parametric study are conducted for the bonded pipe joints. The failure processes identified above are analyzed. The material properties and geometry parameters in the numerical analysis are selected as follows: $t_1$=5 mm, $R_1$=147.5 mm, $t_2$=5 mm and $R_2$=153 mm. For $\rho$=1, Young's modulus for two pipes are given as: $E_1$=124.47 GPa and $E_2$=120 GPa. For $\rho$>1, Young's modulus for two pipes are given as: $E_1$=120 GPa and $E_2$=200 GPa. And the interfacial characteristic parameters are selected as: $\tau_f$=7.2 MPa, $\delta_f$=0.16 mm and $G_f$=0.58 N/mm. Two critical bond lengths for $\rho$=1 and $\rho$>1 are then calculated as 259 mm and 199 mm, respectively. According to the above analysis, the failure process for different bond lengths could be discussed.

The results of FEM using ABAQUS modeling are also given. Considering the assumptions given above, the pipe joints are only subjected to tension load in the present study neglecting bending and Poisson's effect. The FE model is implemented in the software package ABAQUS. In the model, the two pipes are idealised as truss elements (T2D2). As the bondline between the two pipes is assumed to be under pure shear deformation, it is modelled using the 4-node, two-dimensional interfacial cohesive element COH2D4. The cohesive elements share common nodes with the truss elements.

### 4.1. Load-displacement curves

For both case $\rho$=1 and $\rho$>1, take $L$=500 mm when the bond length is larger than the critical bond length and $L$=100 mm when the bond length is shorter than the critical bond length. The corresponding external loading vs. right bond slip curves are exhibited in the following.

### 4.1.1. Case $\rho$=1

The load-displacement curve for $L$=100 mm is shown as in Fig. 4(a). OA is the SRS stage and AB is the S stage. The load-displacement curve for $L$=500 mm is shown as in Fig. 4(b). OA is the SRS stage, AB is the DSRSD stage, and BC is the DSD stage.
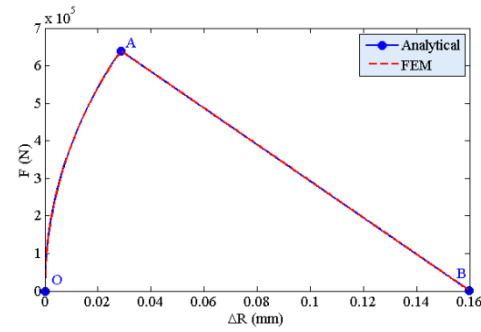
(a) L=100 mm
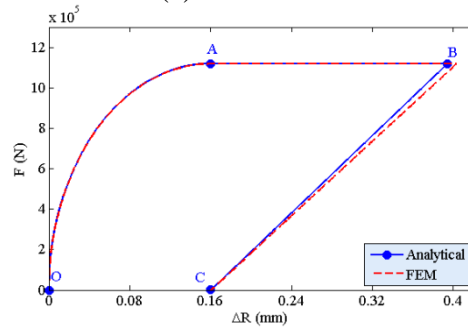


(b) L=500 mm

**Figure 4. Load-displacement curves for case ρ=1.**

*4.1.2. Case ρ>1*

The load-displacement curve for *L*=100 mm is shown as in Fig. 5(a). OA is the SRS stage and AB is the S stage. The load-displacement curve for *L*=500 mm is shown as in Fig. 5(b). OA is the SRS stage, AB is the SRSD stage, and BC is the SD stage.
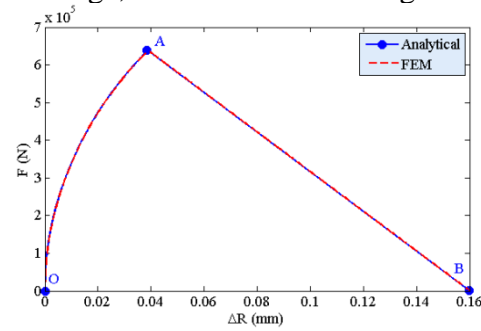


(a) L=100 mm
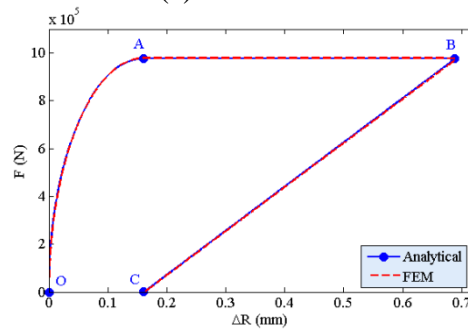


(b) L=500 mm

**Figure 5. Load-displacement curves for case ρ>1.**

4.2. Shear stress distribution

*4.2.1. Case ρ=1*

The shear stress distribution for case $\rho=1$ is shown as in Fig. 6. When the load is small, the interface slips at the two ends are less than $\delta_f$ and the interface is in SRS stage. As load increases, the peak stresses move towards the middle and the lengths of softening zones increases. For $L=100$, the interface enters into S stage when the rigid region disappears while the interface slips at the two ends are still less than $\delta_f$ which means no debonding occurs. For $L=500$, the interface enters into DSRSD stage when the interface slips at the two ends reach $\delta_f$. The debonding and softening regions move towards the middle and the lengths of softening regions stay the same during this stage. When the rigid region disappears and stress peaks reach together, the interface enters into DSD stage.



(a) L=100 mm          (b) L=500 mm
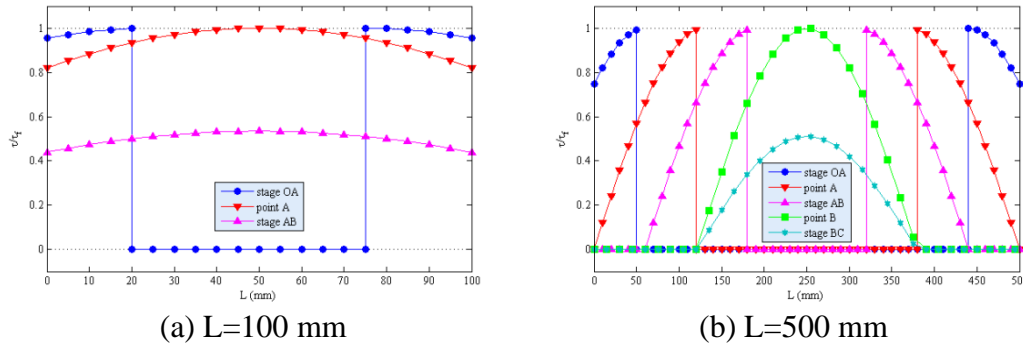
**Figure 6. Interfacial shear stress distribution for case ρ=1.**

*4.2.2. Case ρ>1*

The shear stress distribution for case $\rho>1$ is shown as in Fig. 7. When the load is small, the interface slip at the right end is less than $\delta_f$ and the interface is in SRS stage. As load increases, the peak stresses move towards the middle and the lengths of softening zones increases. For $L=100$, the interface enters into S stage when the rigid region disappears while the interface slip at the right end is still less than $\delta_f$ which means no debonding occurs. For $L=500$, the interface enters into SRSD stage when the interface slip at the right end reaches $\delta_f$. The debonding and right softening regions move towards the left end and the lengths of softening regions stay the same during this stage. When the rigid region disappears and stress peaks reach together, the interface enters into SD stage.



(a) L=100 mm          (b) L=500 mm
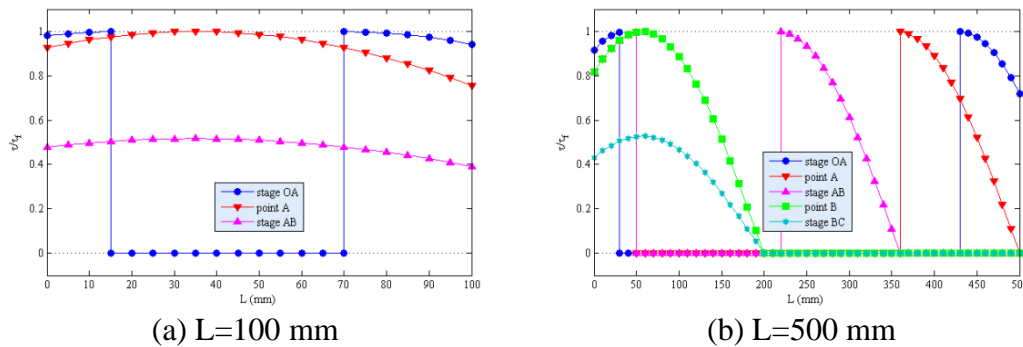
**Figure 7. Interfacial shear stress distribution for case ρ>1.**

*4.3. Parametric study*

Figs. 8(a) and (b) are the influence for different bond lengths of $\rho=1$ and $\rho>1$ on the load-displacement curves, respectively. From the figures, the significant influence for bond

lengths on the curves could also be observed. In the range of the effective bond length, as the bond length increases, not only the interface failure processes change but also the ultimate load and interface slip. Specifically, the increase of the bond length can increase damage ductility. However, when the bond length reaches a certain length (effective bond length), the ultimate load stays the same.



(a) Case ρ=1           (b) Case ρ>1
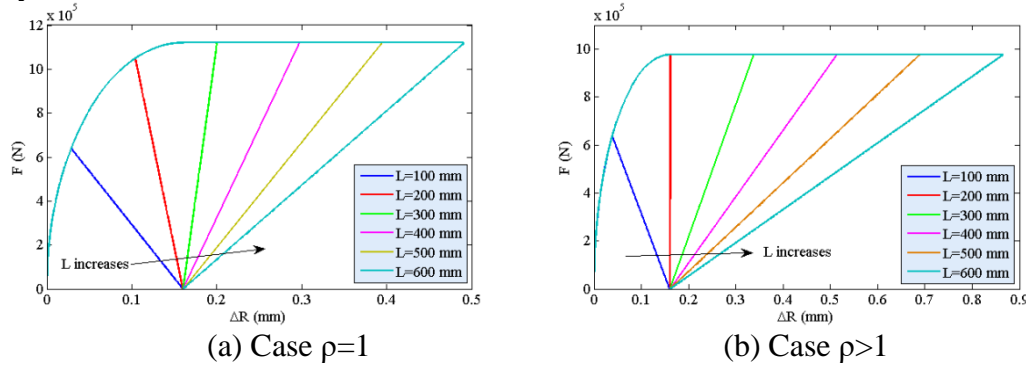
**Figure 8. Load-displacement curves for different bond lengths.**

Fig. 9 shows the relationship of ultimate load and $\rho$. The Young's modulus of pipe 1 and the bond length are taken as 128 GPa and 1000 mm, respectively. From the figure it can be seen that the ultimate load reduces with $\rho$.
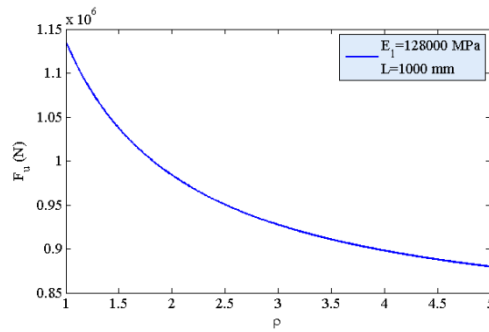


**Figure 9. Effect of ρ on the ultimate load.**

Through the numerical computation, the ultimate load for different bond lengths could be obtained. Fig. 10 shows the ultimate load for different bond lengths. From the figure, when the bond length is short, the ultimate load increases significantly with the bond length. When the bond length is large, the ultimate load stays unchanged. The ultimate load of $\rho=1$ is larger than $\rho>1$ for the reason that when $\rho=1$ both sides enter into debonding zone together, thus can resist greater external force.
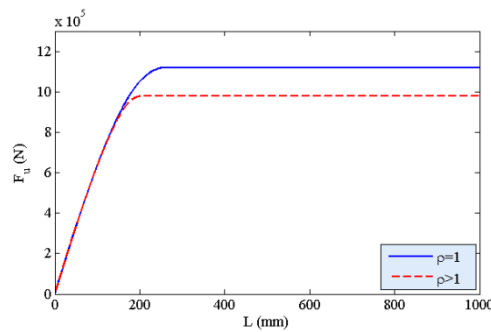


**Figure 10. Effect of bond length on the ultimate load.**

## 5. Conclusions

On the basis of the mechanical behavior of interface, this paper gives a further understanding of the key factors of interfacial debonding through a nonlinear fracture mechanics approach (a cohesive zone model). A rigid-softening cohesive zone model is thus adopted and then used to simulate the initiation, growth and failure of interfacial debonding. Through the nonlinear fracture mechanics, the analytical expressions of the interfacial shear stress and the load-displacement relationship of pipe joints under tension loads could be obtained. Thus the shear stress propagation and the failure progress of the whole interface for different bond lengths could be predicted. The influences of different bond lengths on the load-displacement curve and the ultimate load are studied through the analytical solutions. The stress transfer mechanism, the interface crack propagation and the ductility behavior of the joints could be explained. The present models may be further extended to orthotropic materials, such as fiber-reinforced composite pipe joints.

## Acknowledgments

### References

[1] Fame, C. M., Correia, J. R., Ghafoori, E., Wu, C. (2021) Damage tolerance of adhesively bonded pultruded GFRP double-strap joints, *Composite Structures* **263**.

[2] Bregar, T., An, D., Gharavian, S., Burda, M., Durazo-Cardenas, I., Thakur, V. K. et al. (2020) Carbon nanotube embedded adhesives for real-time monitoring of adhesion failure in high performance adhesively bonded joints, *Scientific Reports* **10**.

[3] Ouyang, Z. Y., Li, G. Q. (2009) Interfacial debonding of pipe joints under torsion loads: a model for arbitrary nonlinear cohesive laws, *International Journal Fractures* **155**, 19-31.

[4] Ouyang, Z.Y., Li, G. Q. (2009) Cohesive zone model based analytical solutions for adhesively bonded pipe joints under torsional loading, *International Journal of Solids and Structures* **46**,1205-1217.

[5] Lubkin, J. L., Reissner, E. (1956) Stress distribution and design data for adhesive lap joints between circular tubes. *Transactions of the ASME*, 1213-1221.

[6] Dragoni, E., Goglio, L. (2013) Adhesive stresses in axially-loaded tubular bonded joints-Part I: critical review and finite element assessment of published models. *International Journal of Adhesion & Adhesives* **47**, 35-45.

[7] Goglio, L., Paolino, D.S. (2014) Adhesive stresses in axially-loaded tubular bonded joints-Part II: development of an explicit closed-form solution for the Lubkin and Reissner model, *International Journal of Adhesion & Adhesives* **48**, 35-42.

[8] Shi, Y. P., Cheng, S. (1993) Analysis of adhesive-bonded cylindrical lap joints subjected to axial load, *Journal of Engineering Mechanics* **119**,584-602.

[9] Nemes, O., Lachaud, F., Mojtabi, A. (2006) Contribution to the study of cylindrical adhesive joining, *International Journal of Adhesion & Adhesives* **26**, 474-480.

[10] Nemes, O., Lachaud, F. (2009) Modeling of cylindrical adhesively bonded joints, *Journal of Adhesion Science and Technology* **23**, 1383-1393.

[11] Yang, C. (2000) Design and analysis of composite pipe joints under tensile loading, *Journal of Composite Materials* **34**, 332-349.

[12] Pugno, N., Carpinteri, A. (2003) Tubular adhesive joints under axial load. *Transactions of the ASME* **70**, 832-839.

[13] Nagaraja, Y. R., Alwar, R.S. (1979) Nonlinear stress analysis of an adhesive tubular lap joint, *The Journal of Adhesion* **10**, 97-106.

[14] Hossein, S. M., Ochsner, A. (2009) A comparative numerical study of the stress states in flat and cylindrical lap joints, *Journal of Adhesion Science and Technology* **23**, 1369-1382.

[15] Das, R. R., Pradhan, B. (2010) Adhesion failure analyses of bonded tubular single lap joints in laminated fiber reinforced plastic composites, *International Journal of Adhesion & Adhesives* **30**,425-438.

[16] Das, R. R., Pradhan, B. (2011) Finite element based design and adhesion failure analysis of bonded tubular socket joints made with laminated FRP composites, *Journal of Adhesion Science and Technology* **25**, 41-67.

[17] Das, R. R., Pradhan, B. (2014) Delamination damage analysis of laminated bonded tubular single lap joint made of fiber-reinforced polymer composite, *International Journal of Damage Mechanics* **23**, 772-790.

[18] Reedy, E. D., Guess, T. R. (1993) Composite-to-metal tubular lap joints: strength and fatigue resistance, *International Journal of Fracture* **63**, 351-367.

[19] Cognard, J. Y., Devaux, H., Sohier, L. (2010) Numerical analysis and optimisation of cylindrical adhesive joints under tensile loads, *International Journal of Adhesion & Adhesives* **30**, 706-719.

[20] Mertiny, P., Ursinus, K. (2007) Damage behavior of joined fiber-reinforced polymer pipe under monotonic and cyclic loading, *ASME Pressure Vessels and Piping Division Conference*.

[21] Hutchinson, J. W., Evans, A. G. (2000) Mechanics of materials: topdown approaches to fracture, *Acta Materialia* **48**, 125-135.

[22] Wu, Z. S., Yuan, H., Niu, H. D. (2002) Stress transfer and fracture propagation in different kinds of adhesive joints, *Journal Engineering Mechanics* **128**, 562-573.

[23] Yuan, H., Teng, J. G., Seracino, R., Wu, Z.S., Yao, J. (2004) Full-range behavior of FRP-to-concrete bonded joints, *Engineering Structure* **26**, 553-565.

[24] Yuan, H., Lu, X.S., Hui, D., Feo, L. (2012) Studies on FRP-concrete interface with hardening and softening bond-slip law. *Composite Structures* **94**, 3781-3792.

[25] Chen, F. L., Qiao, P. Z. (2009) Debonding analysis of FRP-concrete interface between two balanced adjacent flexural cracks in plated beams, *International Journal of Solids and Structures* **46**, 2618-2628.

[26] Lorenzis, L.D., Zavarise, G. (2009) Cohesive zone modeling of interfacial stresses in plated beams, *International Journal of Solids and Structures* **46**, 4181-4191.

[27] Wang J. L. (2007) Cohesive zone model of FRP-concrete interface debonding under mixed-mode loading, *International Journal of Solids and Structures* **44**, 6551-6568.

# Phase transition in polymer derived ceramics (PDCs) and its effect on mechanical response

## Chi Ma and Yan Li[*]

Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, U.S.A

\* Presenting author: yan.li@dartmouth.edu
\* Corresponding author: yan.li@dartmouth.edu

**Abstract**

Polymer-derived ceramics (PDCs) which are fabricated through pyrolysis of preceramic polymers have attracted increasing attention due to their versatility in structure architecture design and property tailoring. Shaping at the polymer state using 3D printing allows the final ceramic products to exhibit arbitrary shapes and complex architectures that are otherwise impossible to achieve through traditional processing routes. The polymer-to-ceramic phase transition also provides additional space for mechanical property tailoring. A multiscale computational model is developed to explore the phase transition mechanisms and their correlations with processing parameters and mechanical response. Calculations in this work concern PMHS/DVB. Molecular dynamics simulations are carried out first to track the chemical reaction mechanisms and atomic structure evolution. The density of generated gas during pyrolysis is transferred to the finite element model (FEM) for coupled heat transfer and phase transition analysis. FEM calculations reveal the effect of pyrolysis temperature and heating rate on structure-level phase composition and elastic modulus. It is found that there is a threshold of pyrolysis temperature above which full ceramic phase is formed. Higher heating rate promotes ceramization and leads to higher elastic modulus. In addition, volume shrinkage is found to accelerate ceramic formation which slightly enhances material strength.

**Keywords:** polymer derived ceramics; phase transition; molecular dynamics simulation; finite element thermal simulation.

## 1. Introduction

Advanced ceramics represent a key enabling technology in aerospace, defense, power generation, and healthcare industries due to their superior properties, such as lightweight [1], high strength [2], excellent thermal stability [3] and high corrosion resistance [4]. Traditional ceramic processing technique has very little control over material geometry and does not provide enough room for property tailoring [5]. The discovery of polymer derived ceramics (PDCs) in 1960 has enabled significant technological breakthroughs in ceramic science and technology [6]. This fabrication approach, which converts preceramic polymers to ceramics through heat treatment under an inert or reacting atmosphere, opens up new opportunities for property tailoring through phase transition control [7-9]. Recently, additive manufacturing technology has enabled fabrication of preceramic polymers with complex shapes and architectures [10, 11]. Shaping at the polymer state not only avoids problems related to tool wear and brittle fracture upon finishing the ceramic component, but also provides new opportunities for geometric design which is of great importance in applications, such as customized biomedical implants, body armor, and energy storage devices, etc. Understanding the effect of key processing parameters on mechanical properties of PDCs requires in-depth

understanding of the phase transition process. Experimental characterizations, e.g. thermogravimetric analysis (TGA) [12] and infrared spectroscopy [7], can track the mass loss associated with preceramic polymer decomposition during pyrolysis. However, these approaches alone cannot directly reveal the molecular structure evolution which is an important aspect of phase transition. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) [13], which can provide detailed nano/micro structure characterization, are only available after sample pyrolysis. Computational models can address some of the underlying physics that cannot be directly captured during experiment. Molecular dynamics (MD) models have been employed to simulate the chemical reaction mechanisms and atomic structure change during pyrolysis [14, 15]. However, conclusions from MD simulations cannot be directly employed to guide the manufacturing process for tailored mechanical properties due to the large time and length scale gaps. Bernard et al. [16] proposed a diffusion-controlled kinetic model which predicts polymer-to-ceramic phase transition at the structure scale. Their prediction of polymer-to-ceramic conversion did not account for the temperature field evolution or the change of heat transfer behavior during the dynamic phase transition process. In fact, the current state phase composition and distribution will largely affect the heat transfer behavior and temperature field evolution that will ultimately determine the subsequent polymer decomposition and phase redistribution. This is because the thermal conductivity of ceramics is about ten times higher than that of polymers. The thermal conductivity of the entire material tends to increase when the polymer phase is gradually converted to the ceramic phase, leading to more intensified subsequent polymer decomposition. A computational model which finds the missing link between the atomic level structure evolution and macroscale phase composition map will promote in-depth understanding of the process physics and its relationship with material response.

In this paper, a multiscale computational model is developed to study the effect of phase transition on mechanical properties of pyrolyzed PMHS/DVB by accounting for a set of systematically varied pyrolysis parameters. Continuum-scale ceramic phase formation is predicted based on the competition between gas generation and gas diffusion in Section 3.1. The effect of heating rate and pyrolysis temperature on elastic modulus is presented in Section 3.2. The effect of volume shrinkage on ceramization is discussed in Section 3.3. The developed model, which correlates key processing parameters with mechanical properties, will reduce the time and cost in developing future PDCs with tailored mechanical properties.

## 2. Model description

### 2.1 Multiscale modeling of phase transition

#### 2.1.1 Molecular dynamics simulation based on reaction force field

A preceramic polymer system, in which polymethylhydrosiloxane (PMHS) is crosslinked by divinylbenzene (DVB), is modeled in this work. The polymer chains with the molar mass of 1500 g/mol are first constructed and randomly packed in the simulation box while DVB molecules are bonded to different polymer chains to create the network structure. The network structure is imported to large-scale atomic/molecular massively parallel simulator (LAMMPS) to obtain system equilibrium [17, 18]. The parameters of reaction force field are selected based on the work of Kulkarni et al. [19]. Constant temperature and pressure ensemble (NPT) are utilized with a time step of 0.1 fs. The pre-pyrolysis system is equilibrated at 300 K as shown in Fig. 1(a). The pyrolysis process is simulated in MD by considering a range of pyrolysis temperatures from 873 K to 5000 K with constant heating rate of 0.1 K/fs and time step of 0.2 fs. The top surface of the simulation box is set to move freely along the vertical direction. At the height of three times of the initial box length, the temperature is set to 0.1 K

by Berendsen thermostat in order to trap the diffused gas molecules. Periodic boundary conditions are applied on the rest of the surfaces.
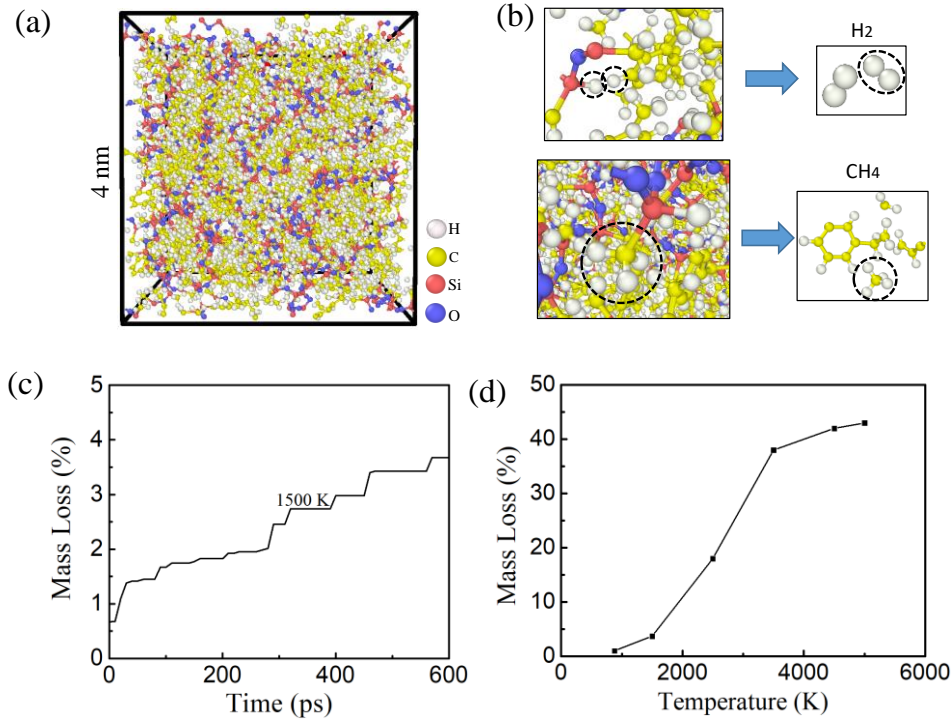


**Fig. 1 (a) Equilibrated PMHS/DVB systems before pyrolysis; (b) Atomic debonding and rebonding process at 1500 K during pyrolysis; (c) Temporal evolution of mass loss at 1500 K; (d) Mass loss percentage as a function of pyrolysis temperature.**

Chemical reactions during pyrolysis are elucidated in Fig. 1(b). At a pyrolysis temperature of 1500 K, bond breakage occurs to form free radicals (e.g. -CH$_3$) and atoms (e.g. H atoms). Gaseous products, such as H$_2$ and CH$_4$ are generated due to the new bond formation. Mass loss occurs as the gaseous products are gradually released out of the system during pyrolysis. As indicated in Fig. 1(c), the system possesses a low degree of mass loss ($< 5\%$) when the pyrolysis temperature is 1500 K. It is noted from Fig. 1(d) that there is a steep increase of mass loss when the sample is pyrolyzed between 1500 K to 3000 K. No obvious mass loss change was observed when the pyrolysis temperature is outside this range. This conclusion agrees with the thermogravimetric analysis (TGA) [20, 21]. It should be noted that the mass loss prediction from MD cannot directly represent the experiment result at the same pyrolysis temperature. A much higher temperature is required in MD to reproduce the mass loss value from TGA [14, 22]. This is primarily due to the limited time and length scales that can be considered in current MD simulations. But the trend of mass loss as indicated in Fig. 1(d) agrees with the experimental observations. In this work, the temperature range is calibrated according to the experimental data from Li et al. [23]. The generated gas density is calculated as

$$\psi\left(T\right) = m_{\text{gas}}\left(T\right)\rho_{\text{initial}} / m_{\text{initial}}, \tag{1}$$

where $\rho_{\text{initial}} = 1.21\,\text{g/cc}$ and $m_{\text{gas}}\left(T\right)$ is the mass loss data from MD results. $\psi\left(T\right)$, which is the amount of generated gas per unit volume, is implemented in phase transition and heat transfer analysis in Section 2.1.2.

*2.1.2 Finite element simulation of phase transition*

Gaseous products, which are generated during pyrolysis, need to release out of the system so that the ceramic structure can be formed. Therefore, phase transition requires in-depth understanding of the interplay between gas generation and gas diffusion. At the structure level, a PDC sample during pyrolysis may include three phases: polymer phase (phase 1), ceramic phase (phase 2) and intermediate phase with partially decomposed polymers. Due to the huge discrepancy of thermal conductivity in each phase, non-uniform temperature field is expected when different phases coexist. Gas diffusion is triggered as a result of the gas density gradient. Gas diffusion rate $\mathrm{d}\psi / \mathrm{d}t$, is calculated as

$$\mathrm{d}\psi / \mathrm{d}t = D\sum_{i=1}^{3}\mathrm{d}\psi^2 / \mathrm{d}^2 x_i , \tag{2}$$

where $D$ is the diffusion coefficient from the work of Merkel et al. [24]. According to the gas diffusion rate, ceramic fraction can be predicted based on the competition between gas generation and gas diffusion. At a given moment during pyrolysis, a selected volume of the sample is either under gas gain or gas loss. Ceramic formation requires $\mathrm{d}\psi / \mathrm{d}t < 0$ when gas loss is activated. The ceramic fraction $\varphi$ is defined as

$$\varphi = \frac{\psi_{\mathrm{release}}}{\psi_{\mathrm{release}}^{\mathrm{max}}} , \tag{3}$$

where $\psi_{\mathrm{release}}$ is the current gas release density. $\psi_{\mathrm{release}}^{\mathrm{max}}$ is the maximum gas density that can be generated in a given unit volume. Calculation of $\varphi$ is carried out through a user subroutine UMATHT in ABAQUS. The finite element model in Fig. 2(a), which simulates macroscale phase transition process by accounting for both heat transfer and gas diffusion kinetics, can explicitly resolve the real-time phase composition map.
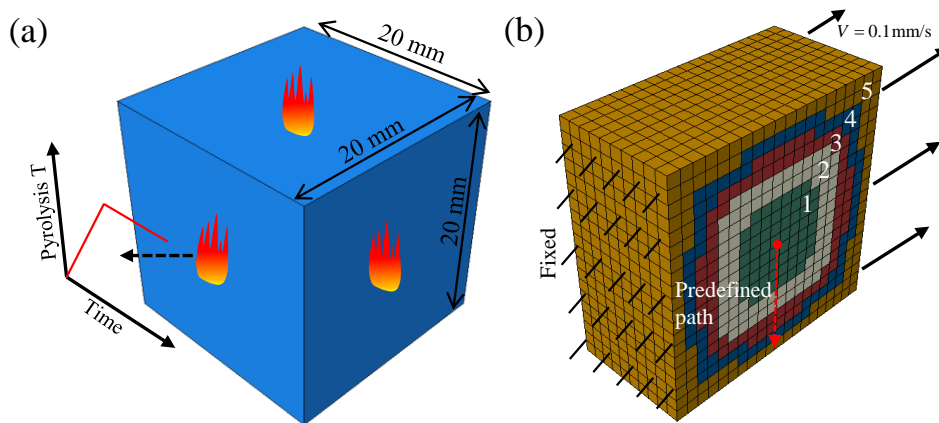


**Fig. 2 (a) Scheme of the finite element model with temperature boundary conditions; (b) Schematic illustration of the phase composition map at 1600 s with pyrolysis temperature of 1273 K and heating rate of 0.63 K/s. The effective elastic modulus is extracted through the simulation of simple tension based on the given boundary and loading conditions.**

*2.2 Prediction of effective elastic modulus*

In this study, five material regions are created based on the range of ceramic fraction $\varphi$ as listed in Table 1. Fig. 2(b) illustrates the region distribution at 1600 s when the sample is pyrolyzed at 1273 K with a heating rate of 0.63 K/s. The elastic modulus at $\varphi=0$ (pure polymer phase) and $\varphi=1$ (pure ceramic phase) is taken as 3.7 MPa [25] and 106.8 GPa [9], respectively. It is assumed that all the regions follow isotropic linear elastic constitutive relations. The equivalent elastic modulus of each region is determined by averaging the elastic modulus of all the including phases. Simple tension simulation is carried out to extract the effective elastic modulus. The simulation prediction is compared with the analytical solution based on the Mori-Tanaka (MT) method according to Lee [26], Fisher et al. [27], Thorvaldsen [28] and Li et al. [29]. In the MT model, the effective stiffness tensor $C$ is formulated as

$$C = A_0 \left[ (1 - \sum_{i=1}^{N=5} f_i)C_{\text{polymer}} + \sum_{i=1}^{N=5} f_i C_i A_i^{\text{dil}} \right] , \tag{4}$$

where $C_{\text{polymer}}$ is polymer stiffness tensor; $f_i$ and $C_i$ are the volume fraction and stiffness tensor of region $i$; $A_0$ and $A_i^{\text{dil}}$ are the strain concentration factors following the following expressions as,

$$A_0 = \left[ (1 - \sum_{i=1}^{5} f_i)I + \sum_{i=1}^{N=5} f_i A_i^{dil} \right]^{-1} \text{ and} \tag{5}$$

$$A_i^{dil} = \left[ I + S_i C_{\text{polymer}}^{-1}(C_i - C_{\text{polymer}}) \right]^{-1}. \tag{6}$$

Here, $I$ and $S_i$ are the identity tensor and the Eshelby tensor of region $i$, respectively. Under the uniaxial tension condition in this study, only the diagonal elements of Eshelby tensors are considered. The diagonal elements can be calculated as

$$S_{diagonal} = \frac{7 - 5\nu_0}{15(1 - \nu_0)}, \tag{7}$$

with $\nu_0 = 0.48$ as Poisson's ratio of the polymer phase. The diagonal value along the $C$ tensor is extracted as the effective elastic modulus. This analytical solution will be compared with the prediction from the simple tension simulation in Section 3.2.

**Table 1. Region division criterion**

| Region number | Ceramic fraction $\varphi$ |
|---|---|
| Region 1 | $\varphi \leq 20\%$ |
| Region 2 | $20\% < \varphi \leq 40\%$ |
| Region 3 | $40\% < \varphi \leq 60\%$ |
| Region 4 | $60\% < \varphi \leq 80\%$ |
| Region 5 | $80\% < \varphi \leq 100\%$ |

## 3. Results and discussion

*3.1 Phase distribution under different pyrolysis conditions*

Macroscale phase transition simulations concern seven samples with the identical geometry as illustrated in Fig. 2(a). The pyrolysis parameters associated with each sample are listed in Table 2. In the first set of calculations, four heating rates of 0.27 K/s, 0.38 K/s, 0.63 K/s and 1.9 K/s are considered while the pyrolysis temperature is kept at 1273 K. Fig. 3 illustrates the phase distribution under each heating rate at 720 s. Fully converted ceramic phase is found at the outer layer of the sample when the heating rate is 1.9 K/s. No more fully converted ceramic phase is observed at the outer surface when the heating rate reduces to 0.63 K/s and below. Lower heating rate leads to larger undecomposed region. In the second set of calculations, heating rate is kept at 0.63 K/s, while the pyrolysis temperature of 673 K, 873 K, 1073 K and 1273 K are applied, respectively. As shown in Fig. 4, fully converted ceramic phase is observed at 1273 K when the sample is being pyrolyzed for 1600 s. In the other three samples under lower pyrolysis temperatures, no fully converted ceramic phase is observed. According to the heating history, it only takes 1227 s to reach the surface temperature of 1073 K. Further pyrolysis to 1600 s cannot help further ceramization. It can be concluded that there is a threshold of pyrolysis temperature above which full ceramic phase is formed.
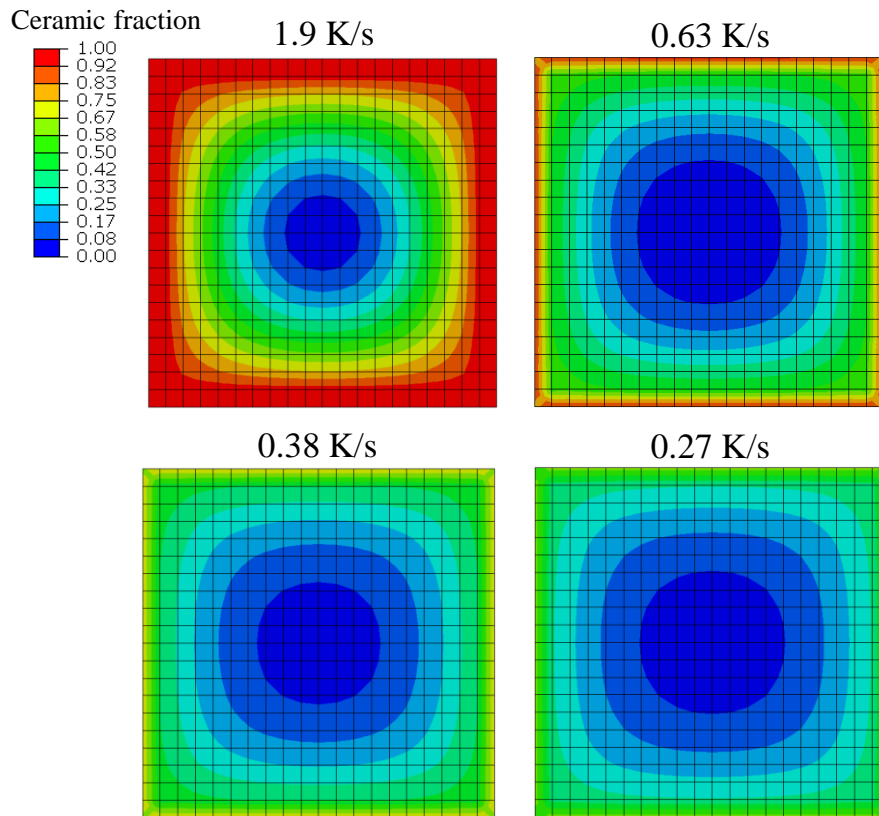


**Fig. 3 Phase distribution under different heating rates at 720 s with final pyrolysis temperature of 1273 K.**
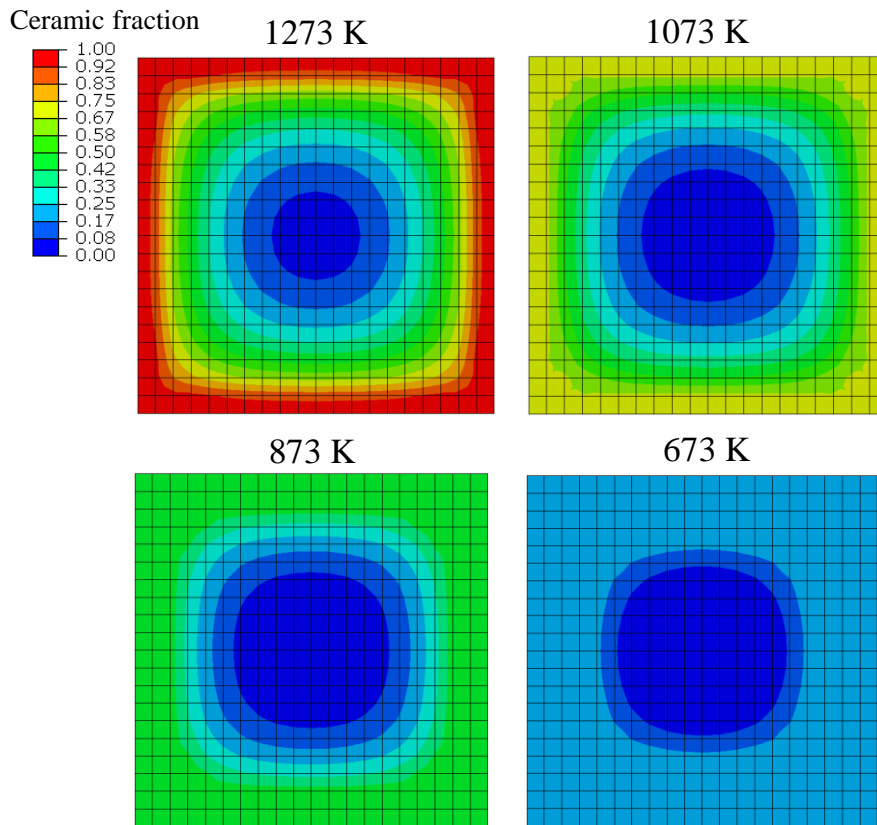
**Fig. 4 Phase distribution under different pyrolysis temperature with heating rate of 0.63 K/s at 1600 s.**

**Table 2. Pyrolysis parameters**

| Sample Number | Heating rate (K/s) | Pyrolysis Temperature (K) |
|---|---|---|
| Sample 1 | 0.27 | 1273 |
| Sample 2 | 0.38 | 1273 |
| Sample 3 | 0.63 | 1273 |
| Sample 4 | 1.9 | 1273 |
| Sample 5 | 0.63 | 1073 |
| Sample 6 | 0.63 | 873 |
| Sample 7 | 0.63 | 673 |

*3.2 Effect of pyrolysis parameters on elastic modulus*

Based on the phase composition map from Fig. 3 and Fig. 4, the corresponding effective elastic modulus of each sample is predicted by both simple tension simulation and MT method as discussed in Section 2.2. As indicated in Fig. 5(b), predictions from both approaches are very close when the heating rate is below 0.63 K/s. As the heating rates increases, the discrepancy becomes larger. Overall, the MT predictions are relatively lower than the FEM predictions from the simple tension test. This trend is observed by Mortazavi et al. [30] as well. It is noticed from Fig. 5 (b) that a 41% increase of heating rate from 0.27 K/s to 0.38 K/s results in a 12% increase of elastic modulus, while a 202% increase of heating rates from 0.63 K/s to 1.9 K/s only leads to a 29% increase of elastic modulus. It can be argued that there exists an upper bond of heating rate beyond which no obvious increase of effective elastic modulus is observed at the given pyrolysis temperature. Based on the given

pyrolysis temperature range as shown in Fig. 6(b), the elastic modulus linearly increases with the pyrolysis temperature.



(a)  (b)
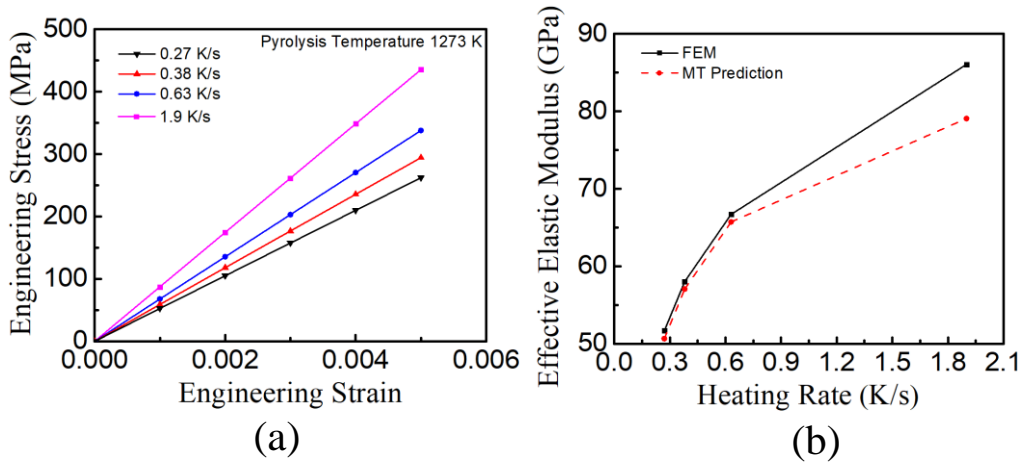
**Fig. 5 (a) Effect of heating rate on engineering stress-strain behavior from simple tension simulation; (b) Comparison of effective elastic modulus as predicted from simple tension simulation and MT method at different heating rates, respectively.**



(a)  (b)
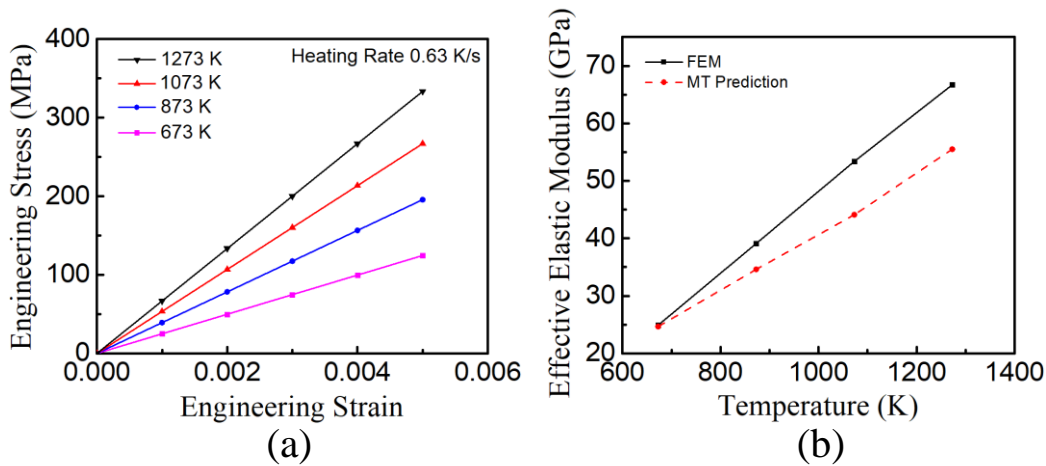
**Fig. 6 (a) Effect of pyrolysis temperature on engineering stress-strain behavior from simple tension simulation; (b) Comparison of effective elastic modulus as predicted from simple tension simulation and MT method at different pyrolysis temperatures, respectively.**

*3.3 Effect of volume shrinkage on phase distribution and elastic modulus*

Gas release during the phase transition not only leads to mass loss, but also results in volume shrinkage of the entire sample. In the following studies, we model the volume shrinkage effect by applying an equivalent hydrostatic pressure $P$ on the sample surfaces. $P$ is calculated according to

$$P = \frac{\Delta V}{V} \times \frac{\overline{E}}{3(1 - 2\overline{v})} \quad , \tag{8}$$

where $\overline{E}$ and $\overline{v}$ are the effective elastic modulus and Poisson's ratio, respectively. $\Delta V$ and $V$ are the change of volume and initial volume of the sample, respectively.

169

Fig. 7 shows the effect of volume shrinkage on spatial distribution of temperature, released gas density and ceramic fraction along the pre-defined path as illustrated in Fig. 2(b). Calculations concern heating rate of 0.63 K/s and pyrolysis temperature of 1273 K. It can be inferred from Fig. 7(a) that volume shrinkage promotes heat transfer as higher temperatures are predicted along the pre-defined path. This is especially obvious towards the sample center where the normalized distance is 0. Higher temperature promotes polymer decomposition and leads to a greater amount of gas generation. On the other hand, gas diffusion which depends on the spatial gas density gradient according to eqn. (2) is also affected by the volume shrinkage. As indicated in Fig. 7(b), volume shrinkage essentially increases the gas density gradient due to the decreased spatial distance. As a result, higher ceramic fraction is expected as shown in Fig. 7(c). The effective elastic modulus predicted from the simple tension model would increase from 66.7 GPa to 72.3 GPa when the volume shrinkage effect is considered.
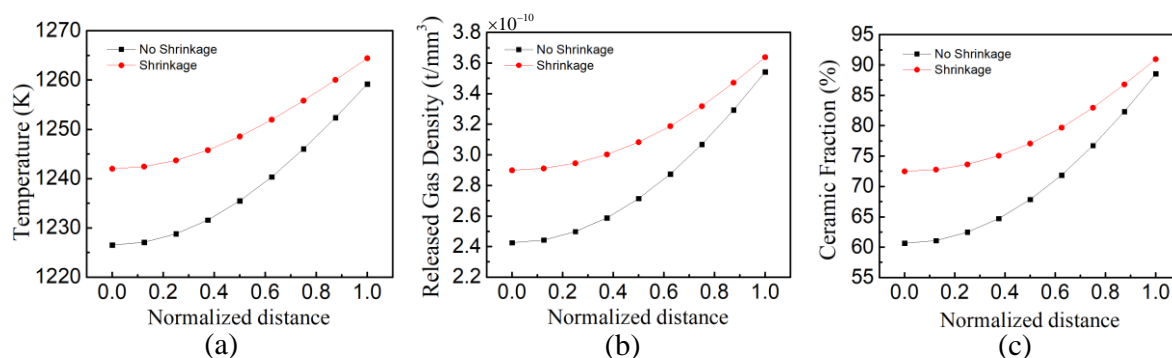


**Fig. 7 Effect of volume shrinkage on spatial distribution of (a) temperature, (b) released gas density and (c) ceramic fraction at the pyrolysis temperature of 1273 K with heating rate of 0.63 K/s.**

### 4. Summary

A multiscale computational model is developed to find the relationship among pyrolysis condition, phase transition and mechanical response. The macroscale phase distribution is determined from the interplay between gas generation and gas diffusion. Gas generation, which is associated with polymer decomposition, is calculated from the MD simulation and calibrated with the experiment data. Gas diffusion, which occurs due to the inhomogeneous temperature distribution induced gas density gradient, is analyzed through coupled heat transfer-phase transition analysis. The phase composition map, which corresponds to a given processing condition, can be explicitly extracted. The effective elastic modulus of a sample is predicted from the simple tension simulation and MT method based on its phase composition map. It is found that predictions from both approaches agree well at low heating rate and pyrolysis temperature. Volume shrinkage promotes heat transfer and ceramic formation, leading to higher prediction of effective elastic modulus. The model developed in this work will be further validated with future experiment.

### Acknowledgement

# References

[1] M. Alizadeh-Osgouei, Y. Li, C. Wen, (2018), A comprehensive review of biodegradable synthetic polymer-ceramic composites and their manufacture for biomedical applications, Bioact Mater 4(1) 22-36.

[2] Y. Waku, N. Nakagawa, T. Wakamoto, H. Ohtsubo, K. Shimizu, Y. Kohtoku, (1997), A ductile ceramic eutectic composite with high strength at 1,873 K, Nature 389(6646) 49-52.

[3] J.F. Justin, A. Jankowiak, (2011), Ultra High Temperature Ceramics: Densification, Properties and Thermal Stability, AerospaceLab (3) p-1.

[4] T. Wei, F. Yan, J. Tian, (2005), Characterization and wear-and corrosion-resistance of microarc oxidation ceramic coatings on aluminum alloy, Journal of Alloys and Compounds 389(1-2) 169-176.

[5] J. Gonzalez Gutierrez, S. Cano, S. Schuschnigg, C. Kukla, J. Sapkota, C. Holzer, (2018), Additive manufacturing of metallic and ceramic components by the material extrusion of highly-filled polymers: A review and future perspectives, Materials 11(5) 840.

[6] P. Colombo, R. Riedel, G.D. Sorarù, H.K. Kleebe (2009), Historical review of the development of polymer derived ceramics (PDCs), Polymer Dervied Ceramics: From Nanostructure to Apllications, DEStech Publictions Inc., pp. 1-12.

[7] L. Gottardo, S. Bernard, C. Gervais, M. Weinmann, P. Miele, (2012), Study of the intermediate pyrolysis steps and mechanism identification of polymer-derived SiBCN ceramics, Journal of Materials Chemistry 22(34) 17923-17933.

[8] J. Wan, M.J. Gasch, A.K. Mukherjee, (2004), InSitu Densification Behavior in the Pyrolysis Consolidation of Amorphous Si-N-C Bulk Ceramics from Polymer Precursors, Journal of the American Ceramic Society 84(10) 2165-2169.

[9] G.D. Sorarù, L. Kundanati, B. Santhosh, N. Pugno, (2019), Influence of free carbon on the Young's modulus and hardness of polymer-derived silicon oxycarbide glasses, Journal of the American Ceramic Society 102(3) 907-913.

[10] Z.C. Eckel, C. Zhou, J.H. Martin, A.J. Jacobsen, W.B. Carter, T.A. Schaedler, (2016), Additive manufacturing of polymer-derived ceramics, Science 351(6268) 58-62.

[11] G. Konstantinou, E. Kakkava, L. Hagelüken, P.V.W. Sasikumar, J. Wang, M.G. Makowska, G. Blugan, N. Nianias, F. Marone, H. Van Swygenhoven, (2020), Additive micro-manufacturing of crack-free PDCs by two-photon polymerization of a single, low-shrinkage preceramic resin, Additive Manufacturing 35 101343.

[12] S. Venkatachalam, D. Hourlier, (2019), Heat treatment of commercial Polydimethylsiloxane PDMS precursors: Part I. Towards conversion of patternable soft gels into hard ceramics, Ceramics International 45(5) 6255-6262.

[13] S. Vry, M. Roumanie, R. Laucournet, G. Bernard-Granger, (2020), Transmission Electron Microscopy Investigations on a Polysiloxane Preceramic Polymer Pyrolyzed at High Temperature in Argon, Ceramics 3(4) 421-427.

[14] I. Ponomarev, A.C.T. Van Duin, P. Kroll, (2019), Reactive Force Field for Simulations of the Pyrolysis of Polysiloxanes into Silicon Oxycarbide Ceramics, Journal of Physical Chemistry C 123(27) 16804-16812.

[15] X. Lu, X. Wang, Q. Li, X. Huang, S. Han, G. Wang, (2015), A ReaxFF-based molecular dynamics study of the pyrolysis mechanism of polyimide, Polymer Degradation and Stability 114 72-80.

[16] S. Bernard, K. Fiaty, D. Cornu, P. Miele, P. Laurent, (2006), Kinetic modeling of the polymer-derived ceramics route: Investigation of the thermal decomposition kinetics of poly[B-(methylamino)borazine] precursors into boron nitride, Journal of Physical Chemistry B 110(18) 9048-9060.

[17] S. Plimpton, (1995), Fast parallel algorithms for short-range molecular dynamics, Journal of computational physics 117(1) 1-19.

[18] C. Ma, T. Ji, C.G. Robertson, R. Rajeshbabu, J. Zhu, Y. Dong, (2017), Molecular insight into the Mullins effect: irreversible disentanglement of polymer chains revealed by molecular dynamics simulations, Physical Chemistry Chemical Physics 19(29) 19468-19477.

[19] A.D. Kulkarni, D.G. Truhlar, S. Goverapet Srinivasan, A.C.T. Van Duin, P. Norman, T.E. Schwartzentruber, (2013), Oxygen interactions with silica surfaces: Coupled cluster and density functional investigation and the development of a new ReaxFF potential, Journal of Physical Chemistry C 117(1) 258-269.

[20] X. You, Q. Hu, X. Hu, H. Chen, W. Yang, X. Zhang, (2019), An Effective, Economical and Ultra-Fast Method for Hydrophobic Modification of NCC Using Poly (Methylhydrogen) Siloxane, Polymers 11(6) 963.

[21] R. Li, P. Hou, N. Xie, Z. Ye, X. Cheng, S.P. Shah, (2018), Design of SiO2/PMHS hybrid nanocomposite for surface treatment of cement-based materials, Cement and Concrete Composites 87 89-97.

[22] H. Gao, H. Wang, Z. Zhao, M. Niu, L. Su, Y. Wei, (2018), Reactive Dynamics Simulation Study on the Pyrolysis of Polymer Precursors to Generate Amorphous Silicon Oxycarbide Structures, Journal of Physical Chemistry C 122(10) 5767-5773.

[23] Y.J. Li, C.Y. Fan, J.P. Zhang, X.L. Wu, (2018), A promising PMHS/PEO blend polymer electrolyte for all-solid-state lithium ion batteries, Dalton transactions 47(42) 14932-14937.

[24] T.C. Merkel, V.I. Bondar, K. Nagai, B.D. Freeman, I. Pinnau, (2000), Gas sorption, diffusion, and permeation in poly (dimethylsiloxane), Journal of Polymer Science Part B: Polymer Physics 38(3) 415-434.

[25] Z. Wang, A.A. Volinsky, N.D. Gallant, (2014), Crosslinking effect on polydimethylsiloxane elastic modulus measured by custom-built compression instrument, Journal of Applied Polymer Science 131(22).

[26] D. Lee, (2018), Local anisotropy analysis based on the Mori-Tanaka model for multiphase composites with fiber length and orientation distributions, Composites Part B: Engineering 148 227-234.

[27] F.T. Fisher, L.C. Brinson, (2006), Nanomechanics of nanoreinforced polymers, Handbook of theoretical and computational nanoscience, American Scientific Publishers, pp. 253-360.

[28] T. Thorvaldsen, (2015), Modelling the elastic stiffness of nanocomposites using the Mori-Tanaka method.

[29] Y. Li, M. Zhou, (2013), Prediction of fracture toughness of ceramic composites as function of microstructure: I. Numerical simulations, Journal of the Mechanics and Physics of Solids 61(2) 472-488.

[30] B. Mortazavi, M. Baniassadi, J. Bardon, S. Ahzi, (2013), Modeling of two-phase random composite materials by finite element, Mori–Tanaka and strong contrast methods, Composites Part B: Engineering 45(1) 1117-1125.

# Dynamic fracture modelling of impact test specimens by the polygon scaled boundary finite element method

**\*Xinxin Jiang, †Hong Zhong, Deyu Li**

State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

*Presenting author: Jiangxx_dlut@163.com

†Corresponding author: zhonghong@iwhr.com

**Abstract**

The polygon scaled boundary finite element method(SBFEM) poses the advantages of flexible meshing and semi-analytical solution, which is capable of various domains and allows the stress intensity factors to be directly extracted from the definition. Hence no local mesh refinement around the crack tip is required. These salient features are used to develop an efficient method to model the impact test specimens, one of which occurs without crack propagation, the other with crack propagation. The linear dynamic equations of polygon SBFEM are derived and then time history of the nodal displacement, velocity and acceleration are obtained using the Newmark direct integration method in time domain. The dynamic stress intensity factors are then computed directly from the instantaneous displacement field and crack velocity. A simple and efficient local remeshing algorithm is used to simulate the crack propagation in polygon element, which greatly simplifies the modelling of crack propagation. Numerical results of the two specimens correspond well with the experimental data and other numerical results reported in the literature. The effects of time step, mesh density and damping coefficient on the resultant accuracy are extensively discussed. Furthermore, the displacement contours are extracted from the numerical simulation to show the impact process, which is useful for the interpretation of the experiment observations.

**Keywords:** Impact test, The polygon SBFEM, Dynamic stress intensity factors, Crack propagation, Time domain

## 1. Introduction

The initiation and propagation of cracks may lead to sudden failure of structures, and prevention of such cases have always been a major concern in engineering. For the dynamic fracture analysis of cracks, the dynamic stress intensity factors (DSIFs) are very important parameters to describe and predict the complex process of crack initiation and growth. Obviously, the DSIFs are time dependent and to extract them is more complex than that from the case subjected to static load. Moreover, it is not easy to deal with the moving boundary with time changed when cracks propagate [1].

Progress toward understanding the fracture process is driven by the researchers all the time. Analytical solutions of the DSIFs are very complex due to the time dependency of the dynamic fracture process and are mainly limited to the infinite plate with semi-infinite crack. Experimental tests are extremely important for the discovery and verification of fracture behavior and several experimental techniques have been developed to investigate the singular stress field around the crack tip. For example, the method of caustics is applied to investigate the dynamic behavior of notched bend specimens, and a series of experimental results can be

referred to the literature [2,3]. In terms of the experimental test, the main difficulties are that the parameters of interest must be measured, and the boundary conditions must be enforced with time changed. Comparing with the analytical solution and experimental test, the numerical methods are available to more complex problems for the relaxation of geometric models and boundary conditions. The FDM is the first to be used in dynamic fractures and successfully solved some problems. Chen analyzes a centrally cracked rectangular plate in time domain using the FDM [4]. Stöckl H and Böhme W simplify the impact tests specimen into a numerical model of the FDM and reproduce the impacting process very well with the experimental observations [5]. However, the difficulty in modelling irregular geometries limit its applications. The FEM has the merit on meshing complex geometries. The main challenge for modelling the crack using the conventional FEM is that enough fine mesh around the crack tip is required to obtain accurate result. Various techniques are developed in modelling the dynamic stress singularity around the crack tip and extend to simulate crack propagation. For example, the dynamic J-integral method [6], the quarter point element (QPE) [7], the extended finite element method (XFEM) [8,9,10,11]. The merit of the BEM over the FEM is that only boundary of the structure is discretized, which can great simplify the modelling of crack and crack propagation. The dual boundary element method has shown its accuracy and efficiency in linear elastic crack problems [12,13,14]. The demerit of the BEM is the requirement of the convoluted fundamental solution. Numerical methods include but are not limited to the above methods. Although many methods are available in crack modelling, both accurate and efficient methods are still necessary due to the complexity of dynamic fractures, especially for problems of crack propagation.

A recent development in structural analysis is the application of arbitrary n-sided polygons [15,16] in the numerical model. The polygon elements have more nodes, which lead to higher order shape functions and more accurate results. In the SBFEM, the polygons can be regarded as sub-domains, which are convenient for the construction of the polygons. The SBFEM is a semi-analytical method with only the boundaries of sub-domains are discretized yet no requirement for fundamental solutions of sub-domains [17]. This appealing feature makes it be a very competitive method for dynamic fractures. Song develops a super-element for the dynamic analysis of crack problems and use its property of semi-analytical solution to extract the DSIFs directly from the transient stress field [18]. Yang computes the frequency-response functions and extracts the DSIFs from the response function in the frequency domain [19,20]. To obtain the time history of the DSIFs, a fast Fourier transform (FFT) is conducted. Later E.T. and Yang use the matrix [M] of the elements and then evaluate the transient response directly using the time-integration method in time domain [21]. Song extends to the dynamic analysis of interface cracks in an-isotropic bi-materials [22].

In terms of dynamic fracture mechanics, experiment tests and numerical methods are mutually reinforcing and developing. On the one hand, the experimental results become the standard to verify the numerical methods; On the other hand, numerical methods can be applied to more complex problems and can obtain physical quantities that are difficult to obtain in experimental tests. And for the numerical methods, not only the accuracy of the results should be validated by the experimental data, but also the computational efficiency is the focus due to the complexity of the problem.

This study aims to develop an accurate and efficient method based on the polygon SBFEM to

model the impact test specimens. The shape functions of polygon elements using SBFEM are obtained and the governing elastic-dynamic equations of the polygon SBFEM in time domain are derived in section 2. The solution to governing elastic-dynamic equations and the extraction of dynamic stress intensity factors are presented in section 3. The local remeshing algorithm is introduced in section 4. Two impact test specimens are modelled using the presented method and the results are compared with the published results in detail. Parametric studies of time step, mesh density and damping coefficient are conducted to verify the numerical accuracy and stability.

## 2. Elastic-dynamic equation of the polygon SBFEM

2.1 Shape functions of polygons using the SBFEM

Fig.1 shows a domain meshed by three n-sided polygons, where n can be greater than 4. Each polygon is regarded as a SBFEM sub-domain with the only requirement that any point on the polygonal boundary should be directly visible from the scaling center. For the first and third polygon without crack, the scaling center can usually be chosen on the centroid. But for the second polygon in Fig.1 with an edge crack, if the scaling center locates on the crack tip, then the crack faces are represented by the edges that connect the crack opening nodes and the scaling center. They can also be called side-faces in SBFEM [23].
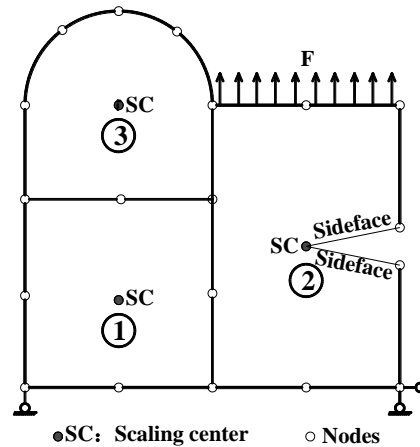


**Fig.1. A domain containing a crack meshed by 3 polygons: the dark solid circles and SC represent the scaling centers, the hollow circles denote the nodes.**
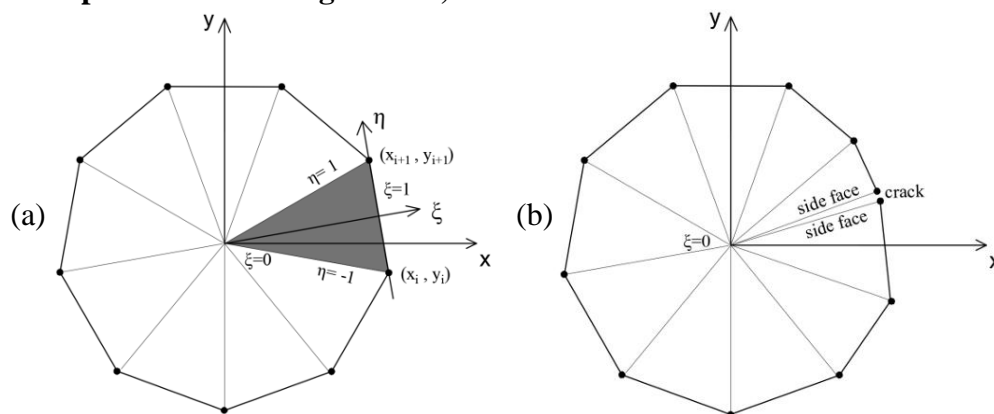


**Fig.2. Examples of polygons using 2-node linear elements.(a) a polygon enclosed by a closed loop; (b)a cracked polygon with side-faces.**

Fig.2 shows polygons modelled by the 2-node linear elements in SBFEM. Each edge of the polygon is discretised by a 2-node linear element. For each linear element, the radial direction

from the scaling centre to the boundary is represented by a dimensionless radial coordinate, $\xi$, with $\xi = 0$ at the scaling centre and $\xi = 1$ at the boundary. The circumferential direction passing through the nodes at the boundary in a counterclockwise direction is represented by a dimensionless circumferential coordinate, $\eta$, with $\eta = -1$ at the first node and $\eta = 1$ at the second node. No special treatment is required for the cracked polygon. The origin of Cartesian coordinates can be moved to the scaling centre, thus (x,y) in a polygon are represented by ($\xi, \eta$) as

$$x = \xi[N(\eta)]\{x\} \tag{1}$$

$$y = \xi[N(\eta)]\{y\} \tag{2}$$

$$[N(\eta)] = [N_1(\eta) \ N_2(\eta)] = [(1\text{-}\eta)/2 \ \ (1+\eta)/2] \tag{3}$$

Here, $\xi, \eta$ is the dimensionless radial and circumferential coordinate, respectively. $\{x\}, \{y\}$ is the nodal coordinates column vector of x direction and y direction, respectively. $N(\eta)$ denotes the shape function of a 2-node linear element.

The transformation relation between the scaled boundary coordinate system with respect to ($\xi, \eta$) and the Cartesian coordinate system with respect to (x,y) is expressed as

$$\begin{Bmatrix} \dfrac{\partial}{\partial \xi} \\ \dfrac{\partial}{\partial \eta} \end{Bmatrix} = [J] \begin{Bmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \end{Bmatrix} \tag{4}$$

with the Jacobian matrix defined as

$$[J] = \begin{bmatrix} x_{,\xi} & y_{,\xi} \\ x_{,\eta} & y_{,\eta} \end{bmatrix} \tag{5}$$

Take the partial derivative of Eqs.(1) and (2) and then substituting the partial derivatives into Eq.(5) results in

$$[J] = \begin{bmatrix} 1 & 0 \\ 0 & \xi \end{bmatrix} \begin{bmatrix} x_b & y_b \\ x_{b,\eta} & y_{b,\eta} \end{bmatrix} \tag{6}$$

with [$J_b$] defined as

$$[J_b] = \begin{bmatrix} x_b & y_b \\ x_{b,\eta} & y_{b,\eta} \end{bmatrix} \tag{7}$$

Based on the transformation relation from the Cartesian coordinate system to the Scaled boundary coordinate system, the linear operator [L] which relates the strains and the displacements can be obtained by the inverse of Eq.(4) and grouping the terms of scaled boundary coordinates leads to

$$[L] = \begin{bmatrix} \dfrac{\partial}{\partial x} & 0 \\ 0 & \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \end{bmatrix} = [b_1]\dfrac{\partial}{\partial \xi} + \dfrac{1}{\xi}[b_2]\dfrac{\partial}{\partial \eta} \tag{8}$$

where the matrices are expressed as

$$[b_1] = \frac{1}{|J_b|} \begin{bmatrix} y_{b,\eta} & 0 \\ 0 & -x_{b,\eta} \\ -x_{b,\eta} & y_{b,\eta} \end{bmatrix}, \quad [b_2] = \frac{1}{|J_b|} \begin{bmatrix} -y_b & 0 \\ 0 & x_b \\ x_b & -y_b \end{bmatrix} \tag{9}$$

The nodal displacement functions $\{u(\xi)\}$ and nodal force functions $\{q(\xi)\}$ along the radial line passing through the scaling center and the modes are introduced. They are the basic unknowns and depend on only the radial coordinate, $\xi$. The displacement field $\{u\}$ of any point inside a sector is related to the nodal displacement functions $\{u(\xi)\}$ as

$$\{u\} = \{u(\xi,\eta)\} = [N_u]\{u(\xi)\} \tag{10}$$

$$[N_u] = \begin{bmatrix} N_1 & 0 & N_2 & 0 \\ 0 & N_1 & 0 & N_2 \end{bmatrix} \tag{11}$$

For the case of vanishing body loads and side-face loads, the relations of the nodal force functions $\{q(\xi)\}$ and the displacement functions $\{u(\xi)\}$ [17,18] can be expressed as

$$\{q(\xi)\} = [E_0]\xi\{u(\xi)\}_{,\xi} + [E_1]^T\{u(\xi)\} \tag{12}$$

$$\xi\{q(\xi)\}_{,\xi} = [E_1]\xi\{u(\xi)\}_{,\xi} + [E_2]\{u(\xi)\} \tag{13}$$

The nodal force functions and the nodal displacement functions depend on only $\xi$. The coefficient matrices depend on only $\eta$

$$[E_0] = \sum_e \int_{-1}^{+1} [B_1]^T [D][B_1]|J_b|\,\mathrm{d}\eta \tag{14}$$

$$[E_1] = \sum_e \int_{-1}^{+1} [B_2]^T [D][B_1]|J_b|\,\mathrm{d}\eta \tag{15}$$

$$[E_2] = \sum_e \int_{-1}^{+1} [B_2]^T [D][B_2]|J_b|\,\mathrm{d}\eta \tag{16}$$

with

$$[B_1] = [b_1][N_u], \quad [B_2] = [b_2][N_u]_{,\eta} \tag{17}$$

Eqs.(12) and (13) contain only two unknowns for $\{u(\xi)\}$ and $\{q(\xi)\}$. Introducing $\{X(\xi)\}$ which consists of $\{u(\xi)\}$ and $\{q(\xi)\}$ as

$$\{X(\xi)\} = \begin{Bmatrix} \{u(\xi)\} \\ \{q(\xi)\} \end{Bmatrix} \tag{18}$$

Using $\{X(\xi)\}$ to rewritten Eqs. (12) and (13) as

$$\xi\{X(\xi)\}_{,\xi} = [Z_p]\{X(\xi)\} \tag{19}$$

with the coefficient matrix $[Z_p]$ defined as

$$[Z_p] = \begin{bmatrix} -[E_0]^{-1}[E_1]^T & [E_0]^{-1} \\ [E_2]-[E_1][E_0]^{-1}[E_1]^T & [E_1][E_0]^{-1} \end{bmatrix} \tag{20}$$

A standard procedure of the eigenvalue decomposition is applied to obtain the solution of nodal displacement and force functions

$$\{u(\xi)\} = [\Phi_b^{(u)}]\xi^{\langle\lambda_b\rangle}\{c\} \tag{21}$$

$$\{q(\xi)\} = \left[\Phi_b^{(q)}\right]\xi^{\langle\lambda_b\rangle}\{c\} \tag{22}$$

where $\lambda_b$ is the positive eigenvalues of $\left[Z_p\right]$ by eigenvalue decomposition and sorted as an ascending order; $\Phi$ is the eigenvector matrix corresponding to the positive eigenvalues. $\{c\}$ are the integration constants vector.

At the boundary ($\xi = 1$) of a polygon, the nodal forces at the boundary are denoted as $\{F\}=\{q(\xi=1)\}$ and the nodal displacements at the boundary are denoted as $\{d\}=\{u(\xi=1)\}$, respectively. Using the boundary condition results in

$$\{d\} = \left[\Phi_b^{(u)}\right]\{c\} \tag{23}$$

$$\{F\} = \left[\Phi_b^{(q)}\right]\{c\} \tag{24}$$

The integration constants vector $\{c\}$ can be obtained by the reverse of Eq.(23) and then substituting it to Eq.(21) leads to

$$\{c\} = \left[\Phi_b^{(u)}\right]^{-1}\{d\} \tag{25}$$

$$\{u\} = \left[N_p\right]\{d\} \tag{26}$$

with the shape function of polygons expressed as

$$\left[N_p\right] = [N_u]\left[\Phi_b^{(u)}\right]\xi^{\langle\lambda_b\rangle}\left[\Phi_b^{(u)}\right]^{-1} \tag{27}$$

Here $\{u\}$ denotes the displacement of any point in the polygon element, $\{d\}$ is the nodal displacements of the polygon element.

2.2 Linear dynamic equations of the Polygon SBFEM

Using the Galerkin approach, the equivalent integral form of the dynamic equilibrium equation and the boundary conditions within the domain V can be expressed as

$$\int_V \delta u_i\left(\sigma_{ij,j} + f_i - \rho u_{i,u} - c u_{i,t}\right)\mathrm{d}V - \int_{S_\sigma} \delta u_i\left(\sigma_{ij}n_j - \bar{T}_i\right)\mathrm{d}s = 0 \tag{28}$$

Applying integration by parts to the first part $\int_V \delta u_i \sigma_{ij,j}\,\mathrm{d}V$ of Eq. (28) results in

$$\int_V \delta u_i\sigma_{ij,j}\mathrm{d}V = \int_{S_\sigma} \delta u_i\sigma_{ij}n_j\mathrm{d}s - \int_V \delta\varepsilon_{ij}\sigma_{ij}\mathrm{d}V \tag{29}$$

Substituting Eq. (29) into (28), Eq. (28) becomes

$$\int_V\left(\delta\varepsilon_{ij}D_{ijkl}\varepsilon_{kl} + \delta u_i\rho u_{i,tt} + \delta u_i c u_{i,t}\right)\mathrm{d}V = \int_V \delta u_i f_i\mathrm{d}V + \int_{S_\sigma} \delta u_i\bar{T}_i\mathrm{d}s \tag{30}$$

The relationship of the strain and the displacement within a polygon can be expressed as

$$\{\varepsilon\} = [L]\{u\} \tag{31}$$

where [L] is the linear operator in Eq.(8).

Substituting Eqs. (28) and (31) into (30), Eq. (30) becomes

$$\{\delta d\}^T[K]\{d\} + \{\delta d\}^T[M]\{\ddot{d}\} + \{\delta d\}^T[C]\{\dot{d}\} = \{\delta d\}^T\{F\} \tag{32}$$

The Eq. (32) should be satisfied for any choice of $\{\delta d\}$ leads to

$$[K]\{d\}+[M]\{\ddot{d}\}+[C]\{\dot{d}\}=\{F\} \tag{33}$$

with the stiffness matrix, the mass matrix, the damping matrix and load vector expand as follows

$$[K]=\sum_e\left[K^e\right]=\sum_e\int_{V^e}\left[N_p\right]^T[L]^T[D][L]\left[N_p\right]dV \tag{34}$$

$$[M]=\sum_e\left[M^e\right]=\sum_e\int_{V^e}\left[N_p\right]^T\rho\left[N_p\right]dV \tag{35}$$

$$[C]=\sum_e\left[C^e\right]=\sum_e\int_{V^e}\left[N_p\right]^Tc\left[N_p\right]dV \tag{36}$$

$$[F]=\sum_e\left[F^e\right]=\int_{V_e}[N_p]^T\{f\}dV+\int_{S_\sigma^e}[N_p]^T\{\tau\}ds \tag{37}$$

Here D is the material constitutive matrix, $\rho$ is the material density.

It can be seen that the dynamic equations of the polygon SBFEM and that of the conventional finite element have the same form. So similar methods can be used to obtain the displacement field as in the finite element method. The Newmark direct integration method in time domain is applied to solve the Eq.(34) and the dynamic response history of the domain is obtained.

2.3 Formulation of [M], [K], [C]

The mass matrix of an S-element is determined by substituting [$N_p$] of Eq. (28) into Eq. (35) and is expressed as

$$[M]=\left[\Phi_b^{(u)}\right]^{-T}\int_0^1\xi^{\langle\lambda_b\rangle}\sum_e\left[\Phi_b^{(u)}\right]^T\int_{-1}^{+1}[N_u]^T\rho[N_u]\xi|J_b|\times d\eta\left[\Phi_b^{(u)}\right]\xi^{\langle\lambda_b\rangle}d\xi\left[\Phi_b^{(u)}\right]^{-1}$$
$$=\left[\Phi_b^{(u)}\right]^{-T}\int_0^1\xi^{\langle\lambda_b\rangle}\left[\Phi_b^{(u)}\right]^T[M_0]\left[\Phi_b^{(u)}\right]\xi^{\langle\lambda_b\rangle}\xi d\xi\left[\Phi_b^{(u)}\right]^{-1} \tag{38}$$

with the coefficient matrix $[M_0]$ is expressed as

$$[M_0]=\sum_e\int_{-1}^{+1}[N_u]^T\rho[N_u]|J_b|d\eta \tag{39}$$

The coefficient matrix $[M_0]$ of a polygon is obtained by assembling the element coefficient matrices of each 2-node linear element according to the element connectivity of a polygon.

The stiffness matrix can also be obtained by substituting the shape function of polygons to Eq. (34). However, a much simpler way is to utilize Eqs.(23) and (24) with the boundary conditions. Eliminating $\{c\}$ and enforcing $\xi=1$ at the boundary in Eqs.(23) and (24) leads to

$$\{F\}=[K]\{d\} \tag{40}$$

with the stiffness matrix defined as

$$[K]=\left[\Phi_b^{(q)}\right]\left[\Phi_b^{(u)}\right]^{-1} \tag{41}$$

Although the expression of the damping matrix can be obtained by using Eq.(36), the Rayleigh damping is more commonly used in practical applications. The damping matrix of a polygon is proportional to the combination of the mass matrix and the stiffness matrix

$$[C]=\alpha[M]+\beta[K] \tag{42}$$

## 3. Dynamic stress intensity factors

The dynamic stress intensity factors of a semi-infinite crack in an infinite plane are related

with the crack velocity[1], which is expressed as

$$\begin{Bmatrix} K_I^{dyn}(\theta,t) \\ K_{II}^{dyn}(\theta,t) \end{Bmatrix} = \begin{bmatrix} k_I(v) & k_{II}(v) \end{bmatrix} \begin{Bmatrix} K_I^{eq}(\theta,t) \\ K_{II}^{eq}(\theta,t) \end{Bmatrix} \tag{43}$$

Here the superscript 'eq' denotes the equivalent stress intensity factors computed from the instantaneous singular stress field. The coefficients $k_I(v), k_{II}(v)$ relate to the crack velocity as

$$k_I(v) \approx \frac{1-v/c_R}{\sqrt{1-v/c_d}}, \quad k_{II}(v) \approx \frac{1-v/c_R}{\sqrt{1-v/c_s}} \tag{44}$$

where $c_d$, $c_s$ and $c_R$ denote the dilatational wave speed, shear wave speed and the Rayleigh wave speed of the material, respectively.

The equivalent stress intensity factors [ $K_I^{eq}(\theta,t)$, $K_{II}^{eq}(\theta,t)$ ]$^T$ are obtained similar to the definition of the static stress intensity factors but using the instantaneous singular stress field.

Different from the eigenvalue decomposition of the coefficient matrix $\begin{bmatrix} Z_p \end{bmatrix}$ in Eq.(20) mentioned above, the generalized stress intensity factors are obtained by the Schur decomposition of $\begin{bmatrix} Z_p \end{bmatrix}$. In a polygon with a crack, the scaled boundary coordinates will be transformed to the polar coordinates r and $\theta$ to seek the maximum singular stress terms around the crack tip. In addition, a characteristic length L is introduced to make the generalized SIFs applicable to the more cases such as the bi-material interface. L is suggested to be the length of the crack in a homogeneous body and twice the length of the crack in other cases. The generalized stress intensity factors are used here, a detailed discussion can be refered to [24]. The equivalent generalized stress intensity factors are expressed as

$$\begin{Bmatrix} K_I^{eq}(\theta) \\ K_{II}^{eq}(\theta) \end{Bmatrix} = \sqrt{2\pi L} \begin{bmatrix} \Psi_L^{(s)}(\theta) \end{bmatrix} \begin{Bmatrix} c^s \end{Bmatrix} \tag{45}$$

$$\begin{bmatrix} \Psi_L^{(s)}(\theta) \end{bmatrix} = \begin{bmatrix} \Psi_\sigma^s(\eta(\theta)) \end{bmatrix} \left( \frac{L}{r_b(\theta)} \right)^{[S^s]-[I]} \tag{46}$$

where $[S^s]$ is the sub-matrix of values between 0 and 1, which causes singularity in the stress field; the $\{c\}$ corresponding to the $[S^s]$ is denoted as $\{c^s\}$.

## 4. Crack propagation simulation procedure

### 4.1 Crack propagation criterion

The initiation and propagation of the crack subjected to dynamic loads are determined by the dynamic stress intensity factors. It is considered that the crack will grow in the direction perpendicular to the largest tension at the crack tip and the detailed introduction about the propagation criteria can refer to [25]. The direction of the crack propagation has a relationship with the stress intensity factors, which is expressed as

$$K_I \sin\theta_c + K_{II}(3\cos\theta_c - 1) = 0 \tag{47}$$

where $\theta_c$ denotes the angle between the direction of the crack propagation and the radial direction of the crack.

### 4.2 Local remeshing

The notched specimens are modelled by polygons. The polygons have more nodes to represent the geometry of the domain, so large size of polygonal element are used to mesh the

domain. The crack propagation length is determined by the instantaneous crack velocity and time step. Since the time step is small for impact testing, the crack propagation length of each time step will also be small. The crack propagate in the polygon element is classified into two categories:

(a) For the crack that grows along the initial direction, a very simple operation is conducted as shown in Fig.3(b). The scaling center of the polygon moves to a new position according to the crack propagation length. This usually occurs in the pure fracture mode *I*.

(b) For the mixed mode loading, the crack will deviate from its original direction. The direction can be determined using Eq.(49). Fig.3(c-d) shows the crack propagation process. The scaling center of the polygon is moved to the location of the new crack tip, then two nodes are added at the crack tip and the polygon is divided into three sub-polygons, which are shown in Fig.3(c). Fig.3(d) descibes next crack propagation step. The previous polygon is not changed, the edges of the divided sub-polygons move with the new crack propagation length.
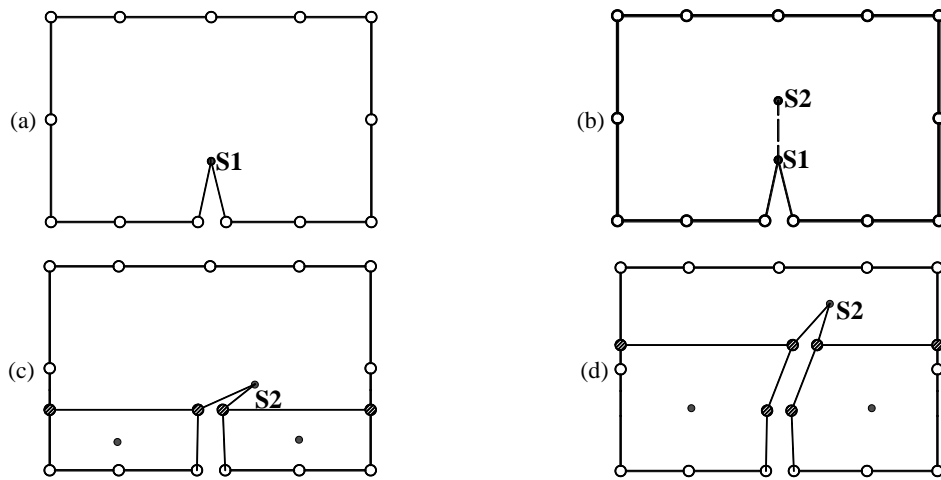


**Fig.3. The local remeshing algorithm:(a) initial mesh of polygon; (b) moving the scaling center to the new crack tip; (c) the polygon is divided into three sub-polygons; (d) the polygon is divided into three new sub-polygons with new edges changed; the solid black circle represent the scaling center; the hollow white circle represent the nodes;the slash circle represent new added nodes.**

## 5. Numerical examples

Two notched bend specimens in impact testing are modelled to validate the presented method. The impact test specimens meshed by polygons are briefly introduced before the numerical simulation. The impact tests are very complicated with many factors, and the numerical models are simplified. The thickness of the samples are very small compared with their plane dimensions, so they are simplified to plane stress problems. The instrumented tup is regarded as a particle and the contact area of the impact specimen is also considered as a point. The impact load during the impact process is recorded and used for the input load of the numerical model. The contact areas between the supports and the sample are also simplified to points. The supports are assumed to be rigid and there is no friction between the sample and the supports. In this way, the numerical models are simplified into the simply supported plates under impact at the midpoint of the top plane. The domains are discretized with polygon elements, and the notches are seen as cracks, which include in the polygon with no special treatment.

5.1 Impact testing of a steel sample without crack propagation

The first example is a notched steel sample impacted by a tup. Fig.4 shows the geometry, dimensions, material properties, boundary and loading conditions. The crack length is a=19.9mm. The measured impact load is used as load input, which is shown in Fig.5. Experimental data and numerical results by FDM are reported in [5].
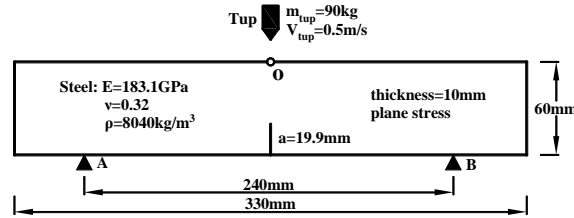


**Fig.4. Impact on a steel sample: the hollow circle denotes the impacting node.**



**Fig.5. The measured impact load.**



(a)Coarse mesh with 3 polygons and 47 nodes



(b)Intermediate mesh with 3 polygons and 63 nodes



(c) Fine mesh with 3 polygons and 127 nodes
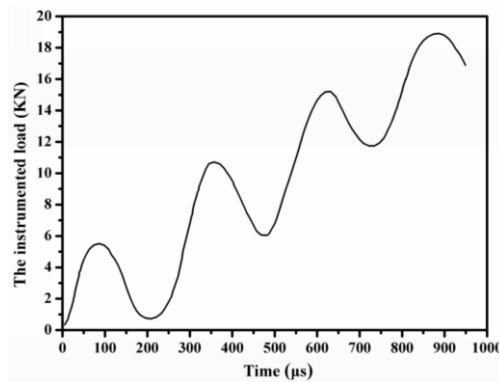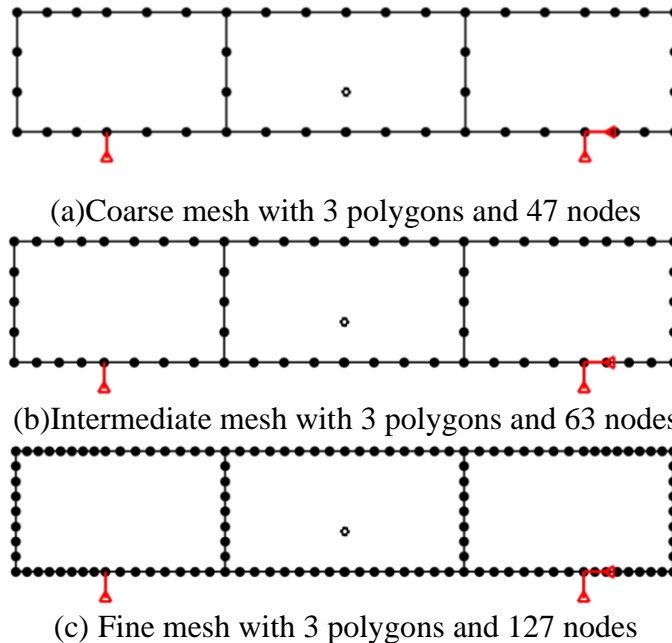
**Fig.6. The initial model and the three meshes of the numerical model: the black solid circles denote the nodes and the black hollow circles denote the crack tip; the crack tip is used as the scaling center of the polygon.**

Fig.6(a-c) shows the three meshes with 47 nodes, 63 nodes and 127 nodes, respectively. Note that two independent nodes with a tiny space are used to represent the crack mouth and the

side-faces passing through the scaling center are not discretized. The dynamic response is solved by the Newmark direct integral method. The simulation is terminated at 960μs. The time step is an important parameter, which not only affects the stability of the algorithm but also the accuracy of the results. Generally speaking, the smaller time step leads to more accurate result. Fig.7 shows time history of the $K_I^{dyna}$ using different time steps with the coarse mesh. It can be seen that the results of time steps from 20μs to 0.4μs are very close with each other, while the $K_I^{dyna}$ using time steps at 40μs shows somewhat different. It indicates the range of the time step is relatively large, which proves the good stability of the method.
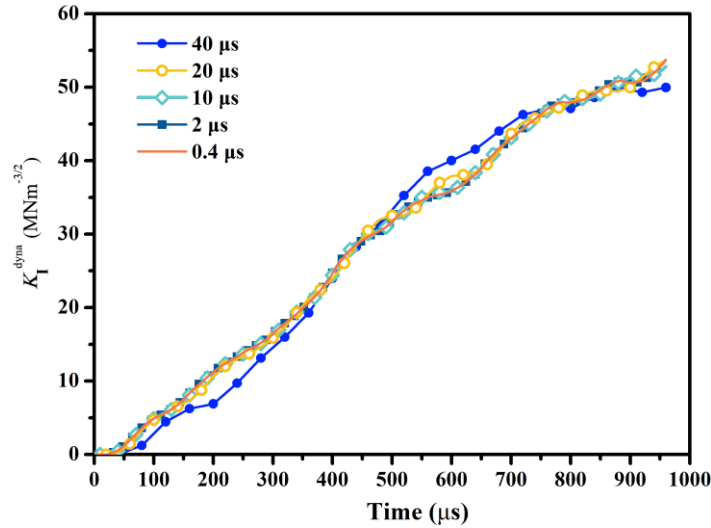


**Fig.7. Effect of time step on $K_I^{dyna}$ using the coarse mesh.**

The CPU time on a Core(TM) 3.20GHz for the three different meshes are listed in Table 1. It is shown that the CPU time gradually increases with the increase of the number of nodes. It only takes 1.2463s, 1.7813s and 3.8831s to obtain time history of the $K_I^{dyna}$ for the three meshes, which demonstrates high computational efficiency of the presented method. The comparison of the CPU time with the FDM is not conducted here since the CPU time is not listed in the literature.

**Table 1 The CPU time for computation of $K_I^{dyna}$ with the three meshes**

| Total Time | Time step | Total steps | CPU time(s) | | |
|---|---|---|---|---|---|
| | | | Coarse mesh | Intermediate mesh | Fine mesh |
| 960μs | 2μs | 480 | 1.2463 | 1.7813 | 3.8831 |
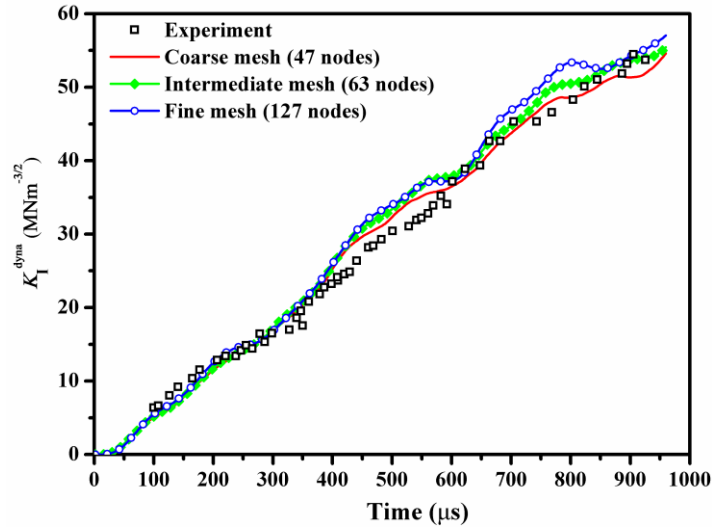
**Fig.8. Effect of mesh density on** $K_I^{dyna}$ **: damping coefficient** $\beta{=}0$ **.**
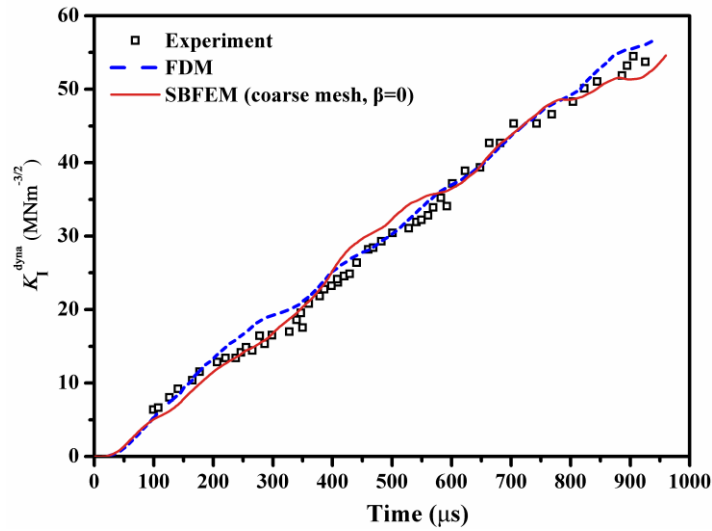


**Fig.9. Dynamic stress intensity factor** $K_I^{dyna}$ **: comparison with experiment and FDM;**

$\beta$ **-damping coefficient.**

Fig.8 shows the computed $K_I^{dyna}$ without damping using coarse mesh, intermediate mesh and fine mesh, respectively, and compared with the experimental data. The overall results of $K_I^{dyna}$ using the three meshes agree very well with the experimental data, which indicates that the mesh density has little effect on the result. Fig.9. shows the computed $K_I^{dyna}$ by the developed method with damping coefficient $\beta{=}0$ using the coarse mesh, compared with the results from experimental data and FDM[5]. It is shown that the numerical results of $K_I^{dyna}$ computed by the SBFEM and FDM correspond very well with the experimental result, which demonstrates good accuracy of the presented methodology.
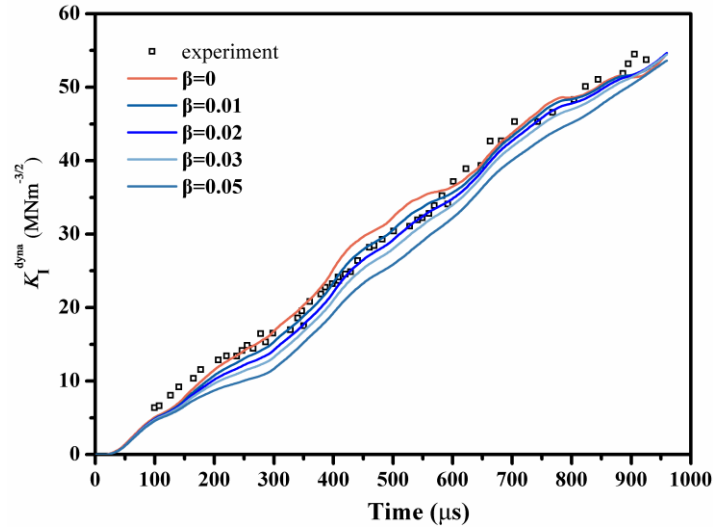
**Fig.10. Effect of material damping coefficient β(0-0.05) on** $K_I^{dyna}$ **using the coarse mesh.**

Fig.10 compares the time history of $K_I^{dyna}$ using different damping coefficient β. Good agreement of the $K_I^{dyna}$ with the experimental result can be observed for β=0 and β=0.01, which indicates that a reasonable damping coefficient is favorable to the results but not necessary for impact testing.
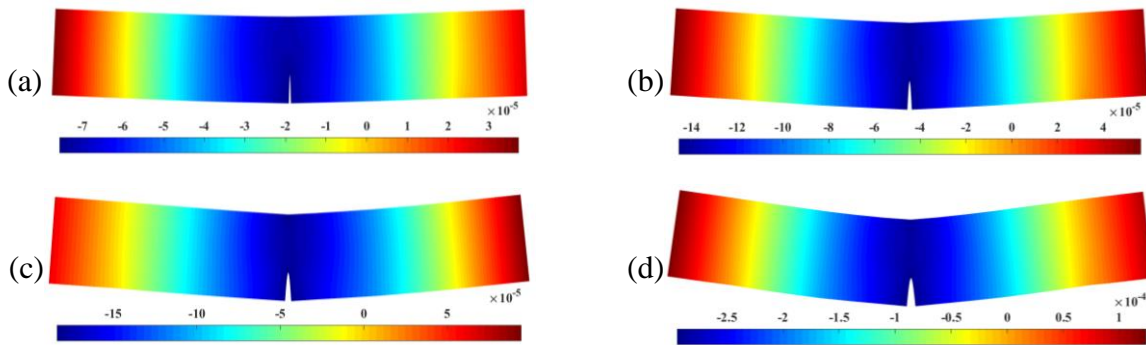


**Fig.11. The vertical displacement contours (m): (a) 200μs; (b)400μs; (c)600μs; (d)800μs .**

Fig.11 demonstrates the impacting process along with the vertical displacement contours at 200μs, 400μs, 600μs and 800μs, respectively. It is shown that the vertical downward displacements and the crack opening displacements increase with the time, which is consistant with the experimental data.

5.2 Impact testing of a specimen with crack propagation
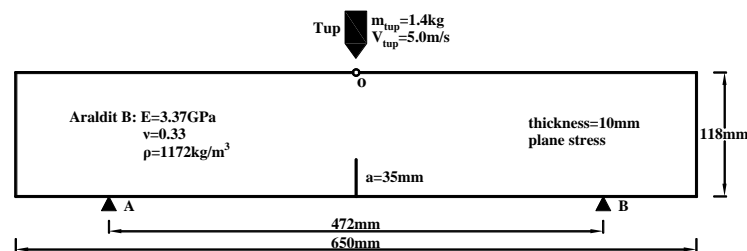


**Fig.12. Impact on a sample made from the epoxy resin Araldite B : the hollow circle denotes the impacting node.**

185

The second example is a sample made from the epoxy resin Araldite B with a crack impacted by a tup on its top midpoint. Fig.12 shows the geometry, dimension, boundary and loading conditions. The crack length is a=35mm. The crack propagates upward under the impact load, and the crack propagation length at each time step depends on the crack velocity and the specified time step. For the impacting test, the crack velocity is related to the $K_I^{dyna}$. The input load and the relation of $V - K_I^{dyna}$ are shown in Fig 13. Experimental data and numerical results by FDM and FEM are reported in[5]. It is noted that the experimental data is recorded only before 232μs.
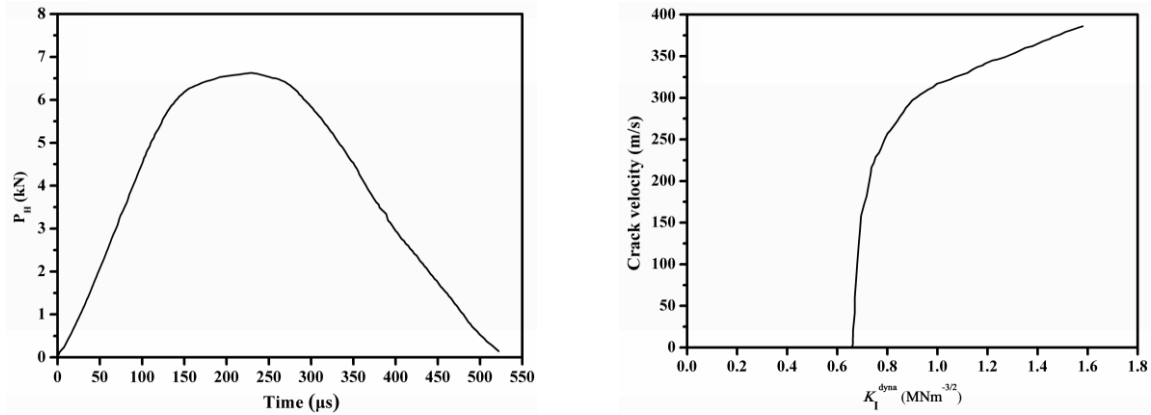


**Fig.13. Input load history(left) and the relation of $V - K_I^{dyna}$ (right).**



**(a) Coarse mesh with 3 polygons and 31 nodes**



**(b) Intermediate mesh with 3 polygons and 47 nodes**



**(c) Fine mesh with 3 polygons and 63 nodes**
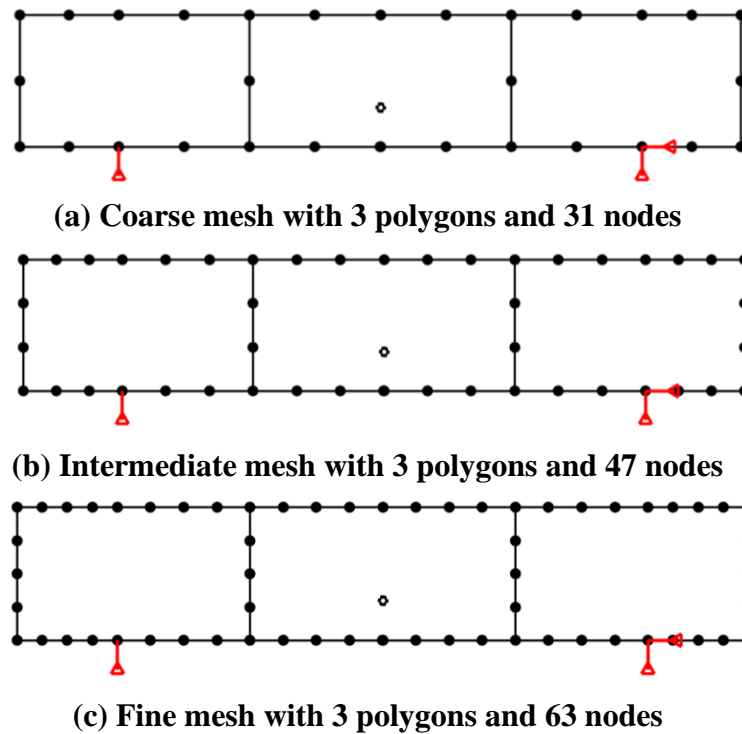
**Fig.14. The initial model and the three meshes of the numerical model: the black solid circles denote the nodes and the black hollow circles denote the crack tip; the crack tip is used as the scaling center of the polygon.**

Fig.14(a),(b) and (c) shows the three meshes with 31 nodes, 47 nodes and 63 nodes, respectively. The crack mouth is denoted by two nodes with a tiny space(1e-12m). The dynamic response is solved by the Newmark direct integral method and the simulation is

186

terminated at 520μs. In each time step, the $K_I^{dyna}$ should be calculated to judge whether the crack propagates or not. If $K_I^{dyna}$ exceeds the critical value, the propagation algorithm is used to determine the crack velocity and the crack length. This process is repeated until the specimen fails. To study the influence of the specified time step, five different time steps are used.
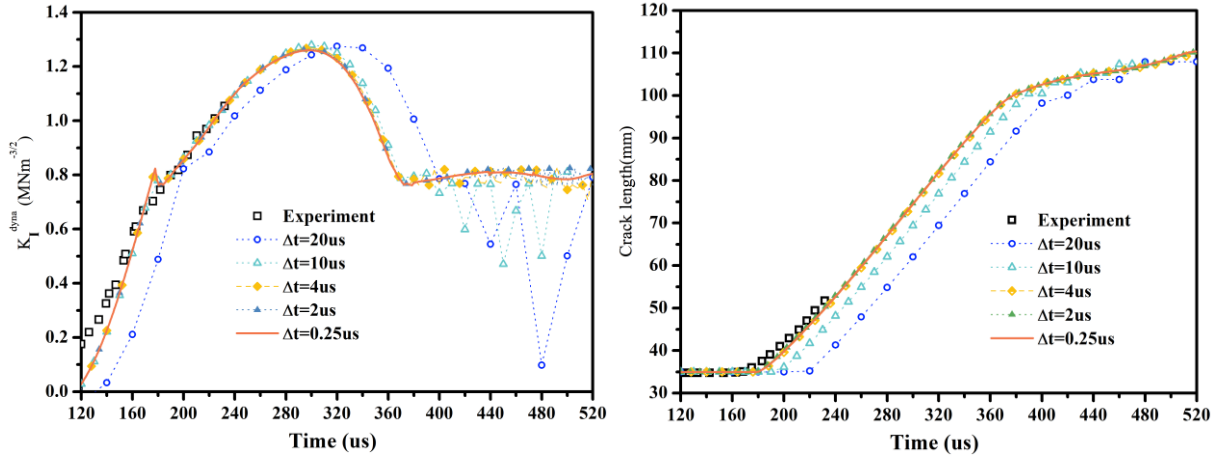


**Fig.15. Effect of time step on the** $K_I^{dyna}$ **and crack length using the coarse mesh.**

Fig.15 shows time history of the $K_I^{dyna}$ and crack length using five different time steps, and compared with the experimental data. It can be seen that results of the $K_I^{dyna}$ and crack length using time steps of 4μs, 2μs and 0.25μs lead to identical results and are close with the experimental data. Results using time steps of 10μs and 20μs show somewhat different. It indicates that smaller time steps are required for the simulation of crack propagation. This is because the time step also influences the crack velocity at each moment, and the two mutually determine the crack propagation length, which is critical for the fracture analysis. It is also found that time steps of 4μs and 2μs lead to slightly osillations between 380μs and 520μs. Therefore, a time step of 0.25μs is used to calculate the results in this example.
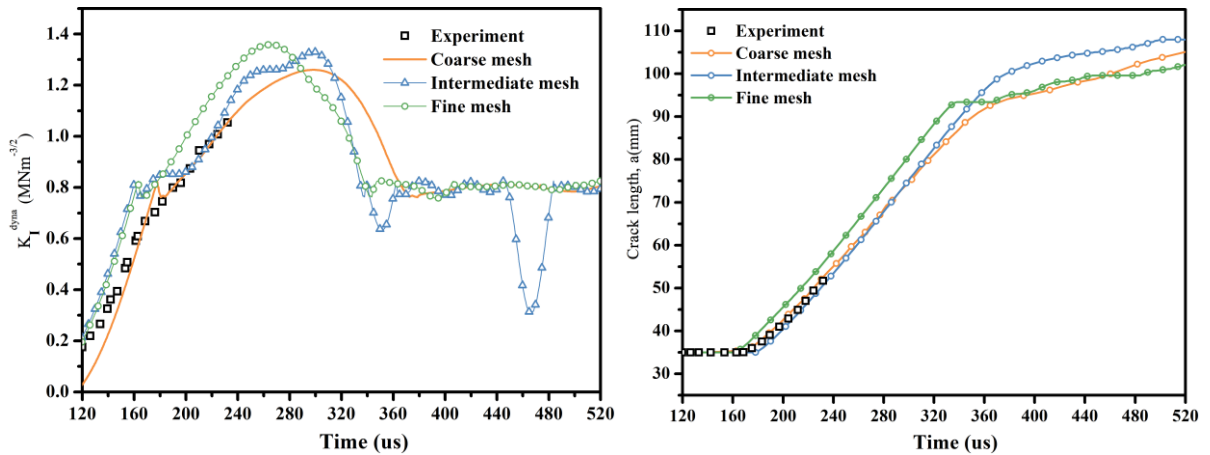


**Fig.16. Effect of mesh density on the** $K_I^{dyna}$ **and crack length using the three meshes.**

Fig.16 shows time history of the computed $K_I^{dyna}$ and crack length without damping using coarse mesh, intermediate mesh and fine mesh, respectively, and compared with the experimental data. The overall agreement is general good with experimental data. An oscillation is observed for the intermediate mesh between 450μs and 500μs for the $K_I^{dyna}$,

which provides no useful information.

The CPU time on a Core(TM) 3.20GHz for the three different meshes are listed in Table 2. It is observed that the CPU time with 1.842s, 2.804s and 4.602s increase to 15.770s, 23.365s and 37.965s when the specified time step changes from 2μs to 0.25μs. They are in an approximate linear relationship, which indicates that the time occupied by local remeshing is very small. The results demonstrate the efficiency of the local remeshing algorithm and the presented method.

**Table 2 The CPU time for computation of $K_I^{dyna}$ with the three meshes**

| Total Time | Time step | Total steps | CPU time(s) | | |
|---|---|---|---|---|---|
| | | | Coarse mesh | Intermediate mesh | Fine mesh |
| 520μs | 2μs | 260 | 1.842 | 2.804 | 4.602 |
| 520μs | 0.25μs | 2080 | 15.770 | 23.365 | 37.965 |
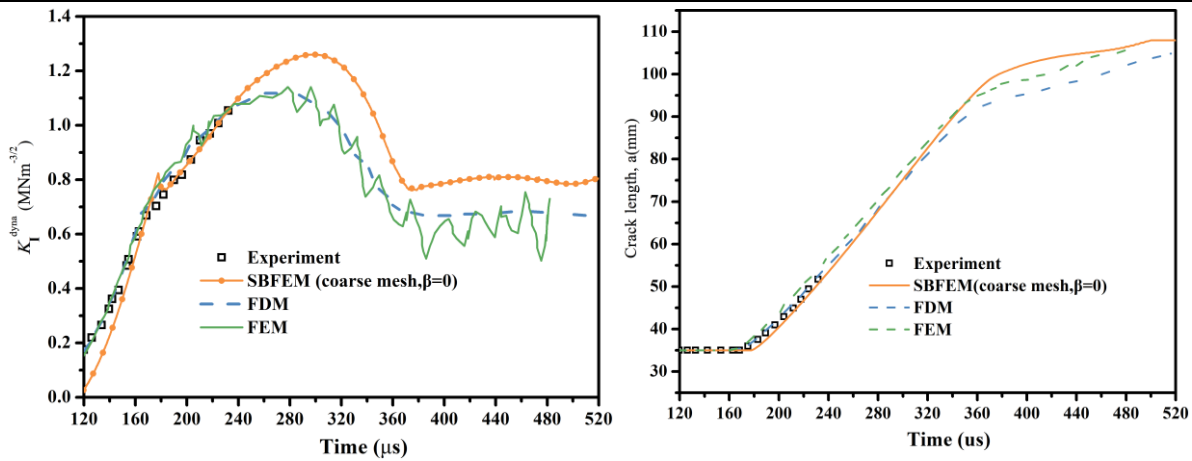


**Fig.17. The $K_I^{dyna}$ and crack length: comparison with experiment, FEM and FDM.**

Fig.17 compare the results of $K_I^{dyna}$, crack length computed by the presented method with the experimental data, FDM and FEM reported in the literature [5]. In terms of $K_I^{dyna}$ and crack length before 232μs, good agreement is observed between the numerical methods and the experimental data. After 232μs, the $K_I^{dyna}$ of the SBFEM reaches a maximum of 1.257MNm$^{-3/2}$ at 298μs, and then gradually decreases to about 0.77MNm$^{-3/2}$ at 298μs and stays around 0.80MNm$^{-3/2}$ until the end. The $K_I^{dyna}$ of the FDM is similar but less than the SBFEM. The $K_I^{dyna}$ of the FEM oscillates around the results of FDM. The predicted crack lengths of the three methods show a good agreement with the experimental data. It can be seen that the crack accelerates rapidly and propagates toward the impact point at nearly the same speed after initiation.
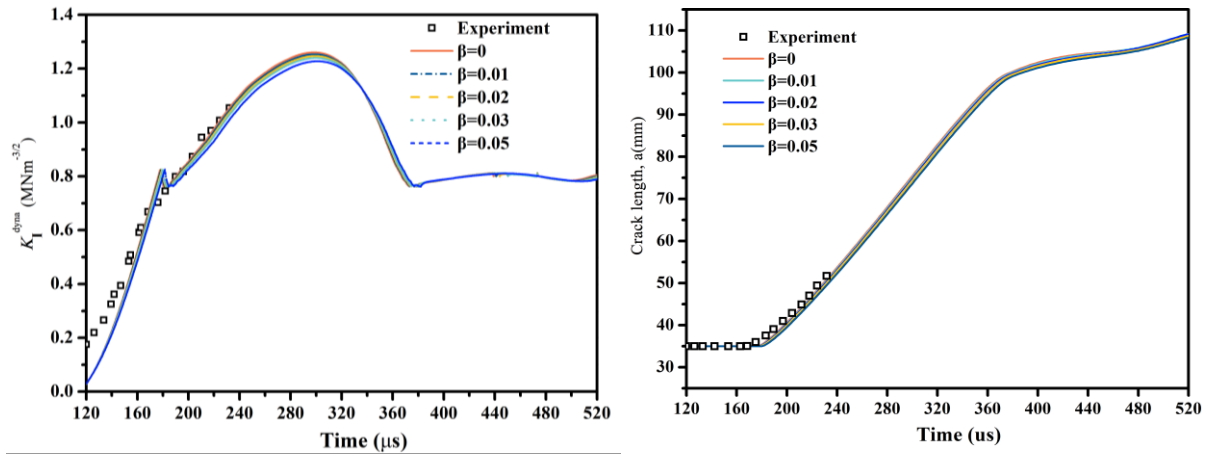
**Fig.18. Effect of material damping coefficient β on the $K_I^{dyna}$ and crack length.**

Fig.18 shows the effect of material damping coefficient β on $K_I^{dyna}$ using the coarse mesh. It is shown that the results of $K_I^{dyna}$ have subtle difference with the increase of damping coefficient, which indicates that a zero damping coefficient is reasonable.
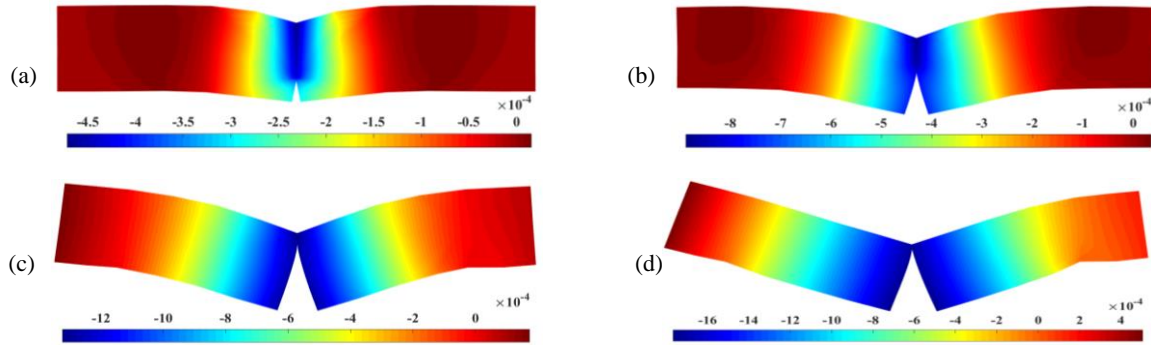


**Fig.19. The vertical displacement contours (m) : (a) 200μs; (b)300μs; (c)400μs; (d)500μs.**
Fig.19 shows the vertical displacement contours to demonstrate the crack propagation progress. It is seen that the crack propagates vertically towards the midpoint on the top. The vertical downward displacements and the crack opening displacements increase with the time, which is consistant with the experimental measurments.

**6. Conclusion**

The impact test specimens have been successfully simulated by the presented method based on the polygon SBFEM. In the method, the time histories of DSIFs can be directly extracted from the semi-analytical solution at each time step and the crack propagation path is predicted using local remeshing algorithm. Numerical results show that the method can compute the DSIFs accurately and efficiently whether the crack propagates or not. For the case with crack propagation, the crack propagation path agrees well with the experimental data. It is observed that the specimens are only meshed by three polygons in the presented method, which leads to high computational efficiency. Compared with FDM and FEM reported in the literature, it requires less number of degrees of freedom to obtain accurate results.

Effect of time step, mesh density and damping coefficient on the DSIFs and crack length are extensively studied. The time step can be selected in a relatively large range. It is shown that the numerical model is not sensitive to the mesh density. And it is also found that the appropriate damping coefficient for impact testing is very small, so it is reasonable not to

consider the damping coefficient here. The vertical displacement contours are also extracted to show the impacting process. Therefore, the developed method is accurate and efficient in simulating the impact test specimens, which makes it to be an attractive and competitive alternative for impact testing simulation.

More work can be done to increase the functionality of the presented method. For example, the mass and velocity of the tup are used to predict the impacting load. In this way, the whole process of numerical simulation of the impact specimens can be realized without relying on the instrumented load from the experiment. It is anticipated that the method can be used as a guide to predict the dynamic behavior of the impact tests. In addition, a very interesting and meaningful research is the extension to the simulation of 3D crack propagation. The polygon elements are then extended to polyhedra elements, and also the definition of DSIFs in 2D calculation method should be replaced by DSIFs in 3D. It is a worthwile job despite the challenges.

## References

[1]   Freund LB. Dynamic Fracture Mechanics. Cambridge University Press: Berlin, 1998.

[2]   Böhme, W., & Kalthoff, J. (1982). The behavior of notched bend specimens in impact testing. International Journal of Fracture, 20, R139-R143.

[3]   Kalthoff, J. (2004). Characterization of the dynamic failure behaviour of a glass-fiber/vinyl-ester at different temperatures by means of instrumented Charpy impact testing. Composites Part B-engineering,35, 657-663.

[4]   Chen, Y.M. (1975). Numerical computation of dynamic stress intensity factors by a Lagrangian finite-difference method (the HEMP code). Engineering Fracture Mechanics, 7, 653-660.

[5]   Stöckl H and Böhme W. Numerische Simulation von Kerbschlagbiege-versuchen an Proben aus Araldit B und Stahl，Berichtw 4/83 des Fraunhofer-Institut fur Werkstoffmechanik(IWM), Freiburg, 1983.

[6]   Enderlein, M., Ricoeur, A., & Kuna, M. (2003). Comparison of finite element techniques for 2D and 3D crack analysis under impact loading. International Journal of Solids and Structures, 40, 3425-3437.

[7]   Murti, V., & Valliappan, S. (1986). The use of quarter point element in dynamic crack analysis. Engineering Fracture Mechanics, 23, 585-614.

[8]   Richardson, C., Hegemann, J., Sifakis, E., Hellrung, J., & Teran, J. (2011). An XFEM method for modeling geometrically elaborate crack propagation in brittle materials. International Journal for Numerical Methods in Engineering, 88, 1042-1065.

[9]   Hojjati-Talemi, R., Cooreman, S., & Hoecke, D.V. (2018). Finite element simulation of dynamic brittle fracture in pipeline steel: A XFEM-based cohesive zone approach. Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications, 232, 357 - 370.

[10] Asareh, I., Yoon, Y., & Song, J. (2018). A numerical method for dynamic fracture using the extended finite element method with non-nodal enrichment parameters. *International Journal of Impact Engineering, 121*, 63-76.

[11] Yan, Z., Feng, W., Zhang, C., & Liu, J. (2019). The extended finite element method with novel crack-tip enrichment functions for dynamic fracture analysis of interfacial cracks in piezoelectric–piezomagnetic bi-layered structures. *Computational Mechanics*, 1-17.

[12] Fedeliński, P. (2004). Boundary element method in dynamic analysis of structures with cracks. Engineering Analysis With Boundary Elements, 28, 1135-1147.

[13] Portela, A., Aliabadi, M., & Rooke, D. (1992). The dual boundary element method: Effective implementation for crack problems. International Journal for Numerical Methods in Engineering, 33, 1269-1287.

[14] Xiao, S., Yue, Z., & Xiao, H. (2019). Dual boundary element method for analyzing three-dimensional cracks in layered and graded halfspaces. *Engineering Analysis With Boundary Elements, 104*, 135-147.

[15] Dai, K.Y., Liu, G., & Nguyen, T. (2007). An n-sided polygonal smoothed finite element method (nSFEM) for solid mechanics. Finite Elements in Analysis and Design, 43, 847-860.

[16] Guo, H., Ooi, E., Saputra, A., Yang, Z., Natarajan, S., Ooi, E., & Song, C. (2019). A quadtree-polygon-based scaled boundary finite element method for image-based mesoscale fracture modelling in concrete. Engineering Fracture Mechanics, 211, 420-441.

[17] Song, C., & Wolf, J. (2002). Semi-analytical representation of stress singularities as occurring in cracks in anisotropic multi-materials with the scaled boundary finite-element method. Computers & Structures, 80, 183-197.

[18] Song, C. (2004). A super-element for crack analysis in the time domain. International Journal for Numerical Methods in Engineering, 61, 1332-1357.

[19] Yang, Z. (2006). Fully automatic modelling of mixed-mode crack propagation using scaled boundary finite element method. Engineering Fracture Mechanics, 73, 1711-1731.

[20] Yang, Z., Deeks, A., & Hao, H. (2007). Transient dynamic fracture analysis using scaled boundary finite element method: a frequency-domain approach. Engineering Fracture Mechanics, 74, 669-687.

[21] Ooi ET, Yang Z. Modelling dynamic crack propagation using the scaled boundary finite element method. International Journal for Numerical Methods in Engineering. 2011;88:329-49.

[22] Song, C., Tin-Loi, F., & Gao, W. (2010). Transient dynamic analysis of interface cracks in anisotropic bimaterials by the scaled boundary finite-element method. International Journal of Solids and Structures, 47, 978-989.

[23] Song, C., & Wolf, J. (1997). The scaled boundary finite-element method—alias consistent infinitesimal finite-element cell method—for elastodynamics. Computer Methods in Applied Mechanics and Engineering, 147, 329-355.

[24] Song C, Tin-Loi F, Gao W. (2010).A definition and evaluation procedure of generalised stress intensity factors at cracks and multi-material wedges. Engineering Fracture Mechanics, 77:2316–36.

[25] Erdogan, F., & Sih, G. (1963). On the Crack Extension in Plates Under Plane Loading and Transverse Shear. Journal of Basic Engineering, 85, 519-525.

# Modeling analysis and mechanical properties research of three-dimensional thin shell composite materials in multi-physical fields

**†\*Wenxing Chen, Shuyang Dai, Hao Lin and Shujun Wang**

[1]School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China
\*Presenting author: wenxingchen@whu.edu.cn
†Corresponding author: wenxingchen@whu.edu.cn

## Abstract

Composites material combined the performance advantages of various materials, which can form some good comprehensive properties. For example, thermal conductivity, electrical conductivity and flexibility. Composites material have been widely used in many fields such as daily necessities, electronic products and medical devices. In this paper, the coupling problem of temperature field and mechanical field of Cu-Zn thin shell material is studied. The material characteristics of the simulation experiment are: cylindrical metal heating vessel, considering the initial temperature conditions and mechanical boundary conditions, when the local heat source temperature will rise rapidly, the thermal stress and deformation will also have a great influence on the material itself. Therefore, the establishment of a mathematical model of thermal field and mechanical field coupling is of great significance to the study of the mechanical behavior of composite cylindrical vessels under heating conditions. The mesh characteristics of this model are the combination of CST element and Q4 element, which can improve the overall solution efficiency and mesh quality. The FEM weak form of the coupling equation is given. Finally, according to the established multi-physical field model, the thermal stress cloud map and temperature diffusion process map of thin shell materials are calculated and output. Our work enriches the multi-physical field theory, which is a process of combining the practical application of multilayer composites with numerical simulation. Numerical calculation plays a good role in assisting and promoting industrial applications.
**Keywords**: multi-physical field, composite material, thin shell structure, thermal stress, thermal strain, temperature field nephogram.

## Introduction

Composite materials are widely used in various engineering structures, medical prosthetics, electronic circuit boards, sports equipment, etc. The subject of these materials is an interdisciplinary field in which chemists, materials scientists, chemical, mechanical and structural engineers contribute to the overall product. Simple multi-laminated composite materials such as phenolic resin laminated materials with low price, heat resistance, flame retardant, insulation, radiation protection and other excellent properties [1,2,3]. Laminated materials are usually added to the motor for flame retardant protection. Such laminated materials are generally divided into two types: hot-pressed semiconductor plate and vulcanized fiber laminated materials [4]. In addition, laminated materials are often used in the design of circuit boards, circuit boards contain PCB laminated materials, the current use of oil heating type or electric heating type vacuum laminating machine for multi-laminated plate

pressing. According to the thickness of resin, buffer film, laminate and so on.

Fiber composite materials are more and more widely used in manufacturing industry. Compared with traditional metal engineering materials, fiber composite materials are lighter and more resistant to corrosion. The properties such as strength, stiffness and toughness can usually be customized according to specific applications. The fiber composite is composed of carrying fibers embedded in polymer resin. Composite materials are usually laminated plates formed by multi-layer superposition, in which the fibers in each layer are unidirectional [5]. In this model, we conduct stress analysis on laminated composite cylinders. It is unrealistic to model the single fiber in each layer of the laminated plate. Some researchers use the simplified micromechanical model of single carbon fiber in epoxy resin to calculate the elastic properties of single layer. Then these properties are used for the homogeneous model of laminated composite cylinders. Here, two methods are used to model the laminated plate, namely the layered (LW) theory and the equivalent monolayer (ESL) theory.

Layered materials can be formed in various ways, and graphene films can be combined into composites by CVD growth [6,7]. Folding method can also be used to fold single-layer graphene into macro-composites, and then form a 3D solid structure from 2D geometric lamellar structure. Folding method can be extended / extended to other 2D nanomaterials to design and manufacture macro-laminated composites with enhanced mechanical properties. Folding leads to additional hardening of composites relative to simple stacked counterparts, and the elastic modulus of newly formed composites will also be significantly increased. The research on the impact resistance and damage performance of laminated materials is also a very complex process[8]. For example, the main factors affecting the drilling of CFRP laminates include delamination, wear, thermal degradation and surface roughness. The mechanical analysis of composite laminates includes many aspects, such as free vibration and dynamic analysis, buckling and post-buckling analysis, geometric nonlinearity and large deformation analysis, and fatigue damage. There are many numerical methods for solving the plates and shells , such as, particle dynamics, molecular dynamics [9,10].

In recent years, the application of multi-physical fields in material mechanics has attracted much attention, that is, the mathematical model that combines the differential equations of multiple physical fields to solve. Simple problems include two-field coupling, thermal field-electromagnetic field coupling, thermal field-force field, light field-thermal field and so on. Multi-field coupling problems include the fracture of thermoelectric materials, thermal-electric-magnetic interaction. In real life, many physical phenomena are multi-physical field coupling problems, and the simplified single-field model is easy to ignore the interaction between fields. In this paper, the application of thermal field-force field coupling in cylindrical heating vessels with composite thin shells is mainly studied. The thermal stress, thermal strain, temperature and so on. They are generated under the action of multi-physical fields. In this paper, the finite element method is used to solve the problem. The innovation of this paper is that the triangular element and quadrilateral element are combined to improve the mesh quality and computational efficiency.

**Temperature-force field coupling model**

**Space term discrete**

In the real three-dimensional world, heat transfer is an objective physical phenomenon, and heat transfer is also accompanied by energy conversion. The heat transfer of objects belongs to three-dimensional transient heat. Temperature variable $T(x,y,z,t)$ is a multivariate function of coordinates and time. When $T(x,y,z,t)$, it indicates that Q does not change with time, and it indicates that the temperature of the thermally conductive object does not change with time after heat exchange. This process is called the steady-state temperature field. When $\frac{\partial T}{\partial t} = 0$ is the transient temperature field, the difference between the transient temperature field and the steady-state temperature field is time variable $t$ [11,12,13]. According to the Fourier heat transfer law and the energy conservation law, the energy balance differential equation in the rectangular coordinate system satisfies the following relationship:

$$\rho c \frac{\partial T}{\partial t} - \frac{\partial}{\partial x}(k_x \frac{\partial T}{\partial x}) - \frac{\partial}{\partial y}(k_y \frac{\partial T}{\partial y}) - \frac{\partial}{\partial z}(k_z \frac{\partial T}{\partial z}) - \rho Q = 0 \qquad (1)$$

For Equation (1), $\rho$ is the material density of $kg/m^3$, $c$ is the specific heat capacity of the material $J/(kg \cdot k)$, plate requires energy when heating up, and $Q$ is the heat generated by the internal heat source, $k_x, k_y, k_z$ is thermal conductivity, they go in three directions $x,y,z$, $W/(m \cdot K)$ is thermal conductivity. $\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}$ respectively represents the heat that flows in the x,y,z direction per unit of time, balancing the heat of the incoming object with the outgoing object when heating up. In addition, the temperature field distribution in the solution domain $\Omega$ needs to meet certain boundary conditions.

(1) Class I boundary conditions : The solid surface temperature is a known function of the time $t$.

$$T_1(x,y,z,t) = \overline{T}(x,y,z,t) \qquad\qquad T_1(x,y,z,t) \in \Gamma. \qquad (2)$$

(2) Class II boundary conditions: The thermal flow density of the solid surface is equal to the change value of the temperature T in the direction of each component.

$$k_x \frac{\partial T_2}{\partial x} n_x + k_y \frac{\partial T_2}{\partial y} n_y + k_z \frac{\partial T_2}{\partial z} n_z = q \qquad T_2 \in \Gamma_2. \qquad (3)$$

(3) Class III boundary conditions: the difference between the heat flow density of the solid surface is proportional to the surface temperature $T$ and the fluid surface temperature $T_c$.

$$k_x \frac{\partial T_3}{\partial x} n_x + k_y \frac{\partial T_3}{\partial y} n_y + k_z \frac{\partial T_3}{\partial z} n_z = h(T_a - T_3) \qquad T_3 \in \Gamma_3. \qquad (4)$$

$n_x, n_y, n_z$ is the direction cosine of the normal line outside the boundary, $\overline{T}(x,y,z,t)$ is a given temperature, $q(x,y,z,t)$ is the heat flow density $W/m^2$ on the boundary $\Gamma_2$, $h$ is the thermal conductivity coefficient $W/(m^2 \cdot K)$ on the boundary, $T_a$ is the insulating temperature of the boundary layer under natural convection conditions, and a combination of all boundaries can be expressed as $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3$.

The heat conduction equation also establishes the finite element format for discretization, which needs to be considered in two parts, namely, spatial discretization and time discretization. If the problem is a regular region, the spatial region is discretized by Q4 finite element, and the solution effect will be better. Namely, triangular elements are more flexible for discrete irregular regions. The total solution domain can be assembled by n small elements, and its functional form can be written as Eq.(5):

$$I = \sum I^e \tag{5}$$

According to the variational principle, there exists $T^* \in H_0^1$, integrals on the unit e can be represented as:

$$I^e = \iiint\limits_{\Omega} \left( \frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right) \right)T^* + \frac{\partial T}{\partial t}T^* dxdydz - \iint\limits_{S} h(T_a - T)T^* ds \tag{6}$$

If the temperature $T$ has equal thermal conductivity along each components, there are $k_x = k_y = k_z = a$

$$I^e = \iiint\limits_{\Omega} a\left( \frac{\partial T^*}{\partial x}\frac{\partial T}{\partial x} + \frac{\partial T^*}{\partial y}\frac{\partial T}{\partial y} + \frac{\partial T^*}{\partial z}\frac{\partial T}{\partial z} \right) + \frac{\partial T}{\partial t}T^* dxdydz - \iint\limits_{S} h(T_a - T)T^* ds \tag{7}$$

Then for the above formula deviation guide, you can get:

$$\begin{aligned}
\frac{\partial I^e}{\partial T_i} &= \iiint\limits_{\Omega} a\left( \frac{\partial T}{\partial x}\frac{\partial}{\partial T_i}\frac{\partial T}{\partial x} + \frac{\partial T}{\partial y}\frac{\partial}{\partial T_i}\frac{\partial T}{\partial y} + \frac{\partial T}{\partial z}\frac{\partial}{\partial T_i}\frac{\partial T}{\partial z} \right) + \frac{\partial T}{\partial T_i}\frac{\partial T}{\partial t} dxdydz \\
&\quad - \iint\limits_{S} h(T_a - T)\frac{\partial T}{\partial T_i} ds
\end{aligned} \tag{8}$$

Then, it is necessary to find the extreme value of the functional form of Eq. (7), and there is

$$\sum \frac{\partial I^e}{\partial T_i} = 0 \tag{9}$$

Let the node temperature of the elemnt e is:

$$\boldsymbol{T}^e = [T_1(t), T_2(t),..., T_n(t)]^{\mathrm{T}} \tag{10}$$

The 3 D interpolation base function is selected as follows:

(1)Linear element (8-junction point)

$$N_i = \frac{1}{8}(1+\xi\xi_i)(1+\eta\eta_i)(1+\varsigma\varsigma_i) \qquad (i=1,2,...,8) \tag{11}$$

(2) Secondary element (20 junction point)

Basis function of cube angular vertices:

$$N_i = \frac{1}{8}(1+\xi\xi_i)(1+\eta\eta_i)(1+\varsigma\varsigma_i)(\xi\xi_i+\eta\eta_i+\varsigma\varsigma_i-2) \qquad i=1,2,..,8 \tag{12}$$

Basis function of side midpoint:

$$N_i = \frac{1}{4}(1+\xi\xi_i)(1-\eta^2)(1+\varsigma\varsigma_i) \qquad (i=9,11,13,15) \tag{13}$$

$$N_i = \frac{1}{4}(1+\xi\xi_i)(1+\eta\eta_i)(1+\varsigma\varsigma_i) \qquad (i=10,12,14,16) \tag{14}$$

$$N_i = \frac{1}{8}(1+\xi\xi_i)(1+\eta\eta_i)(1-\varsigma^2) \qquad (i=17,18,19,20) \tag{15}$$

Temperature interpolation function in a element at any point can be expressed as $T(x,y,z)=\boldsymbol{N}\boldsymbol{T}^e$

$$T(x,y,z,t) = [N_1(x,y,z),N_2(x,y,z),...,N_n(x,y,z)][T_1(t),T_2(t),...,T_n(t)]^T \tag{16}$$

This numerical experiment adopts the specific linear combination of temperature function $T(x,y,z,t)$:

$$\frac{\partial T}{\partial x} = \frac{\partial N_1}{\partial x}T_1 + \frac{\partial N_2}{\partial x}T_2 + ... + \frac{\partial N_n}{\partial x}T_n \tag{17}$$

$$\frac{\partial}{\partial T_i}(\frac{\partial T}{\partial x}) = \frac{\partial N_i}{\partial x} \qquad \frac{\partial T}{\partial T_i} = N_i \tag{18}$$

$$\frac{\partial T}{\partial t} = N_1\frac{\partial T_1}{\partial t} + N_2\frac{\partial T_2}{\partial t} + ... + N_n\frac{\partial T_n}{\partial t} = \boldsymbol{N}\frac{\partial \boldsymbol{T}^e}{\partial t} = \boldsymbol{N}\dot{\boldsymbol{T}}^e \tag{19}$$

Bring the upper expression (17) -(19) into (7), we can gain the equation available:

$$\frac{\partial I^e}{\partial T_i} = \sum k_{ij}^e T_j + \sum c_{ij}^e \frac{\partial T_j^e}{\partial t} + \sum q_{ij}^e T_j - g_i^e T_c \tag{20}$$

$$k_{ij}^e = \iiint_\Omega a(\frac{\partial N_i}{\partial x}\frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y}\frac{\partial N_j}{\partial y} + \frac{\partial N_i}{\partial z}\frac{\partial N_j}{\partial z})dxdydz \tag{21}$$

$$c_{ij}^e = \iiint_\Omega N_i N_j dxdydz \qquad q_{ij}^e = \iint_S h N_i N_j ds \qquad g_i^e = \iint_S h\overline{N}_i ds \tag{22}$$

Among them, $q_{ij}^e = \iint_S h N_i N_j ds$ and $g_i^e = \iint_S h\overline{N}_i ds$ is mainly integral on the surface boundary.

After obtaining the integral on each element, and then assemble the total stiffness matrix. The matrix expression of the finite element discrete equation of the 3 D heat conduction equation is:

$$\boldsymbol{KT} + \boldsymbol{C\dot{T}}^e + \boldsymbol{QT} - \boldsymbol{g}T_c = 0 \tag{23}$$

The element $\boldsymbol{Q}$ is not zero when the $i,j$ is all on the boundary , Let $\boldsymbol{H} = \boldsymbol{K} + \boldsymbol{Q}$ , we can get

$$\boldsymbol{HT} + \boldsymbol{C\dot{T}} - \boldsymbol{P} = 0 \tag{24}$$

Among them, $\boldsymbol{P} = \boldsymbol{g}T_c$ .

**Time term discrete**

The recursive process can be realized by combining the finite difference method with the weighted residual method [14]. In the time period $\Delta t = t_{n+1} - t_n$ , the shape function of each node temperature is as follows :

$$\boldsymbol{T}(t) = N_n(t)\boldsymbol{T}_n + N_{n+1}(t)\boldsymbol{T}_{n+1} \tag{25}$$

where, the $N_n(t) = 1 - \xi, N_{n+1}(t) = \xi$ , reguides the base function $N_n$ about the $t$

$$\dot{N}_n = \frac{dN_n}{dt} = -\frac{1}{\Delta t} \qquad \dot{N}_{n+1} = \frac{dN_{n+1}}{dt} = \frac{1}{\Delta t} \tag{26}$$

Derivative derivative derivative by (25):

$$\dot{\boldsymbol{T}} = \dot{N}_n \boldsymbol{T}_n + \dot{N}_{n+1}\boldsymbol{T}_{n+1} = -\frac{1}{\Delta t}\boldsymbol{T}_n + \frac{1}{\Delta t}\boldsymbol{T}_{n+1} \tag{27}$$

Bring (25) into (24) for the remainder item:

$$\boldsymbol{R} = \boldsymbol{H}(N_n \boldsymbol{T}_n + N_{n+1}\boldsymbol{T}_{n+1}) + \boldsymbol{C}\dot{N}_n \boldsymbol{T}_n + \dot{N}_{n+1}\boldsymbol{T}_{n+1} - \boldsymbol{P} \tag{28}$$

Let $W_j$ be the weight function, according to the weighted margin method

$$\int_0^1 W_j \boldsymbol{R} d\xi = 0 \tag{29}$$

Bring the upper formula (28) into (29), available

$$(\boldsymbol{H}\int_0^1 W_j\, \xi d\xi + \frac{1}{\Delta t}\boldsymbol{C}\int_0^1 W_j\, \xi d\xi)\boldsymbol{T}_{n+1} + (\boldsymbol{H}\int_0^1 W_j(1-\xi)d\xi$$

$$-\int_0^1 \frac{1}{\Delta t}\boldsymbol{C}\int_0^1 W_j\, \xi d\xi)\boldsymbol{T}_n - \int_0^1 W_j\, \boldsymbol{P} d\xi = 0 \tag{30}$$

After finishing, both sides divided by $\int_0^1 W_j\, d\xi$ , we can simplify (30) to equation (31):

$$(s\,\boldsymbol{H} + \frac{1}{\Delta t}\boldsymbol{C})\boldsymbol{T}_{n+1} + (1-s)(\boldsymbol{H} - \frac{1}{\Delta t}\boldsymbol{C})\boldsymbol{T}_n - \boldsymbol{P} = 0 \tag{31}$$

Denote as: $s = \int_0^1 W_j\,\xi d\xi \,/\, \int_0^1 W_j\,d\xi$

$$T_c = N_n(t)T_c^n + N_{n+1}(t)T_c^{n+1} = (1-\xi)T_c^n + \xi T_c^{n+1} \tag{32}$$

$$\boldsymbol{P} = [sT_c^{n+1} + (1-s)T_c^n]\boldsymbol{g} \tag{33}$$

With $T_c^n, T_c^{n+1}, \boldsymbol{T}^n, \boldsymbol{P}$, we can get (33). and then, $\boldsymbol{T}^{n+1}$ can be obtained by iteration. What's

more, according to the three-dimensional stress $(\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{yz}, \tau_{zx})$ and strain

$(\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{yz}, \gamma_{zx})$ in the generalized Hook 's law, assuming that the thermal strain of

$\alpha T$ can be compensated in all directions from the temperature $T$ of the initial stress-free state,
the coordination equation (Bettelm formula ) is derived [15,16]:

$$\begin{cases} (1+u)\nabla^2\sigma_x + \dfrac{\partial^2 S}{\partial x^2} + \dfrac{1+2u}{1-2u}E\alpha\nabla^2 T + E\alpha\dfrac{\partial^2 T}{\partial x^2} = 0 \\[2mm] (1+u)\nabla^2\sigma_y + \dfrac{\partial^2 S}{\partial y^2} + \dfrac{1+2u}{1-2u}E\alpha\nabla^2 T + E\alpha\dfrac{\partial^2 T}{\partial y^2} = 0 \\[2mm] (1+u)\nabla^2\sigma_z + \dfrac{\partial^2 S}{\partial z^2} + \dfrac{1+2u}{1-2u}E\alpha\nabla^2 T + E\alpha\dfrac{\partial^2 T}{\partial z^2} = 0 \\[2mm] (1+u)\nabla^2\tau_{yz} + \dfrac{\partial^2 S}{\partial y\partial z} + E\alpha\dfrac{\partial^2 T}{\partial y\partial z} = 0 \\[2mm] (1+u)\nabla^2\tau_{zx} + \dfrac{\partial^2 S}{\partial z\partial x} + E\alpha\dfrac{\partial^2 T}{\partial z\partial x} = 0 \\[2mm] (1+u)\nabla^2\tau_{xy} + \dfrac{\partial^2 S}{\partial x\partial y} + E\alpha\dfrac{\partial^2 T}{\partial x\partial y} = 0 \end{cases} \tag{34}$$

The formular $S = \sigma_x + \sigma_y + \sigma_z$ and $\nabla^2$ of the above equations are second-order differential

operators, and $E$ is elastic modulus. $\mu$ is Poisson ' s ratio. $\alpha$ is the expansion coefficient. The

above equations are also the coupled equations of temperature field and force field, which can
not only reflect the mutual influence between temperature field and stress, but also obtain the

strain $(\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{yz}, \gamma_{zx})$. Naturally, the change of these values is very helpful for the

study of material properties in high temperature environment. Next, numerical calculation is
needed, to solve the finite element equations, the variational principle is needed to simplify
and deform each equation, and finally a new algebraic equation group is formed.

$$\iiint\limits_\Omega \left\{ \left[ (1+\mu)\nabla^2\sigma_x + \frac{\partial^2 S}{\partial x^2} \right]\sigma_x^* + \left[ \frac{1+2\mu}{1-2\mu}E\alpha\nabla^2 T + E\alpha\frac{\partial^2 S}{\partial x^2} \right]T^* \right\} dxdydz = 0 \tag{35}$$

Using partial integral and Green formula to obtain:

$$\iiint_{\Omega} \{[(1+\mu)(\frac{\partial \sigma_x}{\partial x}\frac{\partial \sigma_x^*}{\partial x}+\frac{\partial \sigma_x}{\partial y}\frac{\partial \sigma_x^*}{\partial y}+\frac{\partial \sigma_x}{\partial z}\frac{\partial \sigma_x^*}{\partial z})+\frac{\partial \sigma_x^*}{\partial x}\frac{\partial S}{\partial x}]dV -$$

$$\iint_{\Gamma} \sigma_x^*[(1+\mu)\frac{\partial \sigma_x}{\partial n}+\frac{\partial S}{\partial x}cos(n,x)]dS+\iiint_{\Omega}\frac{1+2\mu}{1-2\mu}E\alpha(\frac{\partial T}{\partial x}\frac{\partial T^*}{\partial x}+\frac{\partial T}{\partial y}\frac{\partial T^*}{\partial y}+\frac{\partial T}{\partial z}\frac{\partial T^*}{\partial z})+$$

$$+E\alpha\frac{\partial T}{\partial x}\frac{\partial T^*}{\partial x}]dV-\iint_{\Gamma}T^*[\frac{1+2\mu}{1-2\mu}E\alpha\frac{\partial T}{\partial x}+\frac{\partial T}{\partial x}cos(n,x)]ds = 0 \qquad (36)$$

As for $\forall \sigma_x^*,T^* \in H_0^1, t > 0$ , among them, $T^* \in H_0^1$ is the Sobolev space and meet the boundary conditions, bilinear formula as following:

$$B(\sigma_x,\sigma_x^*)-(\frac{\partial^2 S}{\partial x^2},\sigma_x^*)= -B(T,T^*)+(\frac{\partial^2 T}{\partial x^2},T^*) \qquad (37)$$

After finishing, we can get:

$$J(\sigma_x,T)=\frac{1}{2}[B(\sigma_x,\sigma_x)+B(T,T)]-(\frac{\partial^2 S}{\partial x},\sigma_x)-(\frac{\partial^2 T}{\partial x^2},T) \qquad (38)$$

The original problem was transformed into:

$$J(\sigma_x^*,T)= min\, J(\sigma_x,T) \qquad \sigma_x,T \in H^1(\Omega).$$

Similarly, the bilinear discrete formulas of the other formulas are:

$$B(\sigma_y,\sigma_y^*)-(\frac{\partial^2 S}{\partial y^2},\sigma_y^*)= -B(T,T^*)+(\frac{\partial^2 T}{\partial y^2},T^*) \qquad (39)$$

$$B(\sigma_z,\sigma_z^*)-(\frac{\partial^2 S}{\partial z^2},\sigma_z^*)= -B(T,T^*)+(\frac{\partial^2 T}{\partial z^2},T^*) \qquad (40)$$

$$B(\tau_{yz},\tau_{yz}^*)-(\frac{\partial^2 S}{\partial y\partial z},\tau_{yz}^*)= (\frac{\partial^2 T}{\partial y\partial z},T^*) \qquad (41)$$

$$B(\tau_{zx},\tau_{zx}^*)-(\frac{\partial^2 S}{\partial z\partial x},\tau_{zx}^*)= (\frac{\partial^2 T}{\partial z\partial x},T^*) \qquad (42)$$

$$B(\tau_{xy},\tau_{xy}^*)-(\frac{\partial^2 S}{\partial x\partial y},\tau_{xy}^*)= (\frac{\partial^2 T}{\partial x\partial y},T^*) \qquad (43)$$

The distribution function of the object temperature difference is $\Delta T(x,y,z)$ , due to the principle of thermal expansion and cold contraction, if the temperature increase will cause the expansion of the material itself, the thermal strain will occur on each component of the right-Angle coordinate system. The physical equation of thermal strain using temperature and

thermal stress is as follows:

$$
\begin{cases}
\varepsilon_x = \dfrac{1}{E}[\sigma_x - \mu(\sigma_y + \sigma_z) + \alpha\Delta T] \\[2mm]
\varepsilon_y = \dfrac{1}{E}[\sigma_y - \mu(\sigma_x + \sigma_z) + \alpha\Delta T] \\[2mm]
\varepsilon_z = \dfrac{1}{E}[\sigma_z - \mu(\sigma_x + \sigma_y) + \alpha\Delta T] \\[2mm]
\gamma_{xy} = \dfrac{1}{G}\tau_{xy} \qquad \gamma_{yz} = \dfrac{1}{G}\tau_{yz} \qquad \gamma_{zx} = \dfrac{1}{G}\tau_{zx}
\end{cases}
\tag{44}
$$

**Numerical example:**

**Material parameter setting:** This numerical experiment mainly studies the numerical simulation experiment of a container with heating function in a multi-physical field. The coupling simulation experiment of the material in a multi-physical field. The single-layer composite material mainly selects the metal material of Cu-Zn combination. The geometric shape of the object is the shape of a metal cup. The bottom radius of the cylindrical container is $R = 2m$, the height is $h = 6m$, and the thickness is $d = 0.05m$. The bottom of the material will provide a continuous heat source for heating $Q_b = 50w/m^2$, and the surrounding will convection with the air. The temperature of the air selects the normal outdoor temperature $T_0 = 293k$. For the thin shell, the steady-state heat conduction equation is:

$$
\rho C_\rho \boldsymbol{u} \cdot \nabla T + \nabla \cdot \boldsymbol{q} = Q + Q_{ted}
\tag{45}
$$

$$
\boldsymbol{q} = -k\nabla T
\tag{46}
$$

Transient thermal conduction equation can also be written as:

$$
\rho C_\rho \frac{\partial T}{\partial t} + \rho C_\rho \boldsymbol{u} \cdot \nabla T + \nabla \cdot \boldsymbol{q} = Q
\tag{47}
$$

$$
\boldsymbol{q} = -k\nabla T
\tag{48}
$$

The $\rho$ of the upper formula is called material density, $C_\rho$ is constant pressure heat capacity, $\boldsymbol{u} = (u,v,w)$ is displacement field vector, $q$ is heat flux, $Q$ is heat source function, $k$ heat conduction coefficient. Heat flow exchange on the boundary, can also be expressed as:

$$
q_0 = h(T_{ext} - T) \qquad (x_i, y_i, z_i, t_i) \in \Gamma
\tag{49}
$$

Among them, the external temperature $T_{ext} = 273k$, the heat transfer coefficient is $h = 750kW/(m^2 \cdot k)$, the top cover of the container will be subjected to downward gravity load $F_{zg} = -1000N$.

Transient structural mechanical equation:

$$\rho(\frac{\partial^2 u}{\partial t^2} + z\frac{\partial^2 ar}{\partial t^2}) = \nabla \cdot S + F_v + 6(M_v \times n)\frac{z}{d} \tag{50}$$

$$S = \sigma_x + \sigma_y + \sigma_z, \quad \sigma_z = 0, \quad -1 \leq z \leq 1. \tag{51}$$

Among them, $u = (u,v,w)$ is the shift field vector of material, $a_r$ is normal displacement, $F_v$ is external load, $M_v$ is moment, $n$ is normal vector, $d$ is the thickness of material. Mechanical equation of steady state structure:

$$0 = \nabla \cdot S + Fv + 6(M_v \times n)\frac{z}{d} \tag{52}$$

$$\sigma_z = 0, \quad -1 \leq z \leq 1 \tag{53}$$

Table 1. The basic parameters of Cu-Zn mixed materials

| Name | Sign | Value | Unit |
|---|---|---|---|
| Thickness | $d$ | 0.01[m] | $m$ |
| Young's modulus | $E$ | 90000 | Pa |
| Heat capacity at Constant pressure | $C_\rho$ | 420 | J/(kg·K) |
| Density | $\rho$ | 8800 | kg/m³ |
| Rotation | $w$ | 0.0 | rad |
| Thermal conductivity factor | $k$ | 52 | W/(m·K) |
| Poisson ratio | $v$ | 0.35 | 1 |

**Construction of geometrical shape and mesh subdivision**
In this model, two kinds of grids are used to divide, one is CST element and Q4 element to divide the region, that is, the quadrilateral element is used around the cylinder container, and the triangle element is selected for the bottom circle.Generally speaking, for regular graphics, we use rectangular elements, and for irregular graphics, triangles can also be used in combination.However, it is necessary to pay attention to the vertex alignment (common nodes) on the interface of different elements.The purpose is that the data between the elements can be correctly transmitted after the discretization of different regions, and the calculation accuracy of the model can be improved.The number of vertices in the cylindrical entity grid is 312. The bottom area $\Gamma_c$ contains 118 elements, the minimum mesh mass $Min_q = 0.6975$,

$Max_q = 0.8689$, the maximum element size length $l = 0.6$, the curvature factor $\rho_\varsigma = 0.6$,

and the resolution $R_{ratio} = 0.5$ of the narrow area. The side area $\Gamma_R$ of the cylinder has 240

rectangular sides, and the average mesh quality is $\overline{M} = 1.0$.



(a)Triangle grid unit division of the      (b) Rectangle unit division on the side of the cylinder
bottom circle area
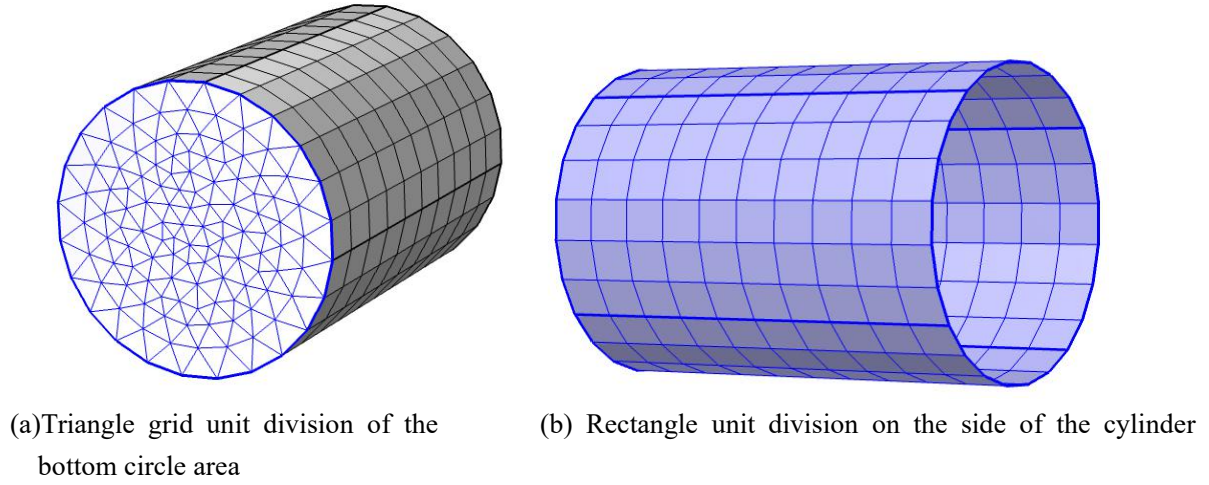
Fig1. Heating container is grided with the CST and Q4 element.

**Mesh quality Assessment**

For triangular mesh, the shape coefficient is also defined to evaluate the quality of the mesh, where $f_s$ is the shape coefficient and $A_e$ is the mesh area. Any triangular element can draw an circumcircle, and this circumcircle can draw an equilateral triangle, whose area is the optimal mesh area of $A_{e,\ opt}$.

$$f_s = \frac{A_e}{A_{e,opt}} \tag{54}$$

In addition, there are seven indicators of the quality of the mesh quality are:
   (1)  Aspect Ratio, Horizontal ratio, the software default value is 15.
   (2)  Skew Angle, Software default value 45.
   (3)  Warpage, different software definitions vary.
   (4)  Taper, default value for the software is 0.25-0.5.
   (5)  Jacobian Ratio, default value for the software is 0.7.
   (6)  Twist Angle, default value for the software is 30.
   (7)  Element length, default minimum length of 0.1mm, maximum of 100mm.

The following is the visualization of the grid quality evaluation results. The green represents the optimal grid quality close to $M_q = 1$. Red represents the worst grid quality $M_q = 0$,

From the division effect, the grid quality of this physical container division is good, the part is very uniform, and there is no area of poor grid quality.

| (a) Cylindrar mesh division | (b) Triangle grid offset histogram |
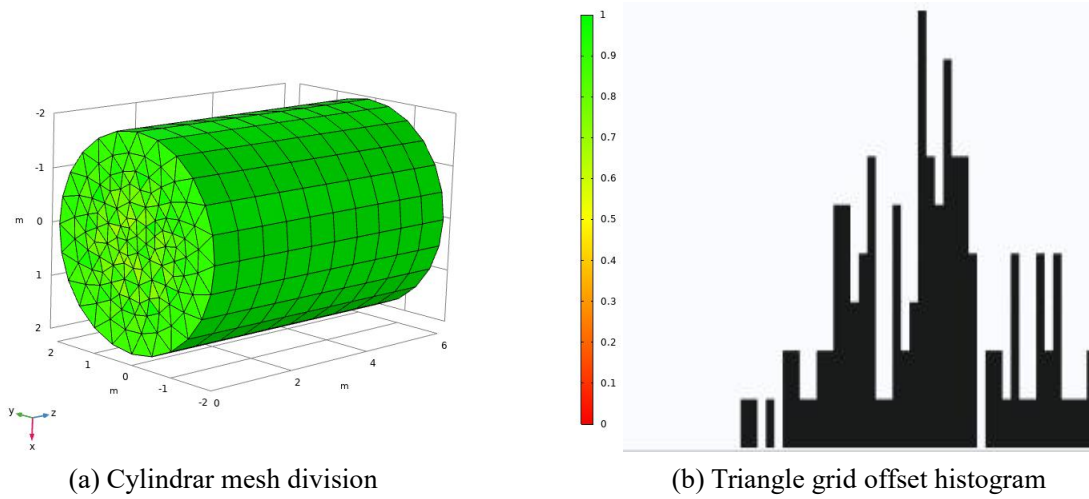
Fig2. Mesh quality test result diagram

**Multi-physical field boundary conditions**

The physical model is mainly the coupling of temperature field and structural mechanics field. This model selects the steady-state change, which is independent of time change[17,18]. The purpose of coupling is to observe the influence of temperature on mechanical properties of Cu-Zn thin shell material at high temperature, and then draw the stress nephogram by visualization technology. The coupling equation of the two physical fields is as follows:

$$\rho C_\rho \boldsymbol{u} \cdot \nabla T + \nabla \cdot \boldsymbol{q} = Q + Q_{ted} \tag{55}$$

$$\boldsymbol{q} = -k\nabla T \tag{56}$$

$$0 = \nabla \cdot S + Fv + 6(\boldsymbol{M}_v \times \boldsymbol{n})\frac{z}{d} \tag{57}$$

$$\sigma_z = 0, \quad -1 \le z \le 1 \tag{58}$$

**Temperature initial condition**

the boundary information of temperature field is as follows : the cylinder keeps convection with air around, that is, the initial temperature of surface keeps at :

$$T_{surf} = 293k, \quad (x, y, z) \in \Gamma_R \tag{59}$$

Bottom circle provides continuous heat source, temperature retention: $Q_b = 50w/m^2, (x, y, z)$

$\in \Gamma_c$. In addition, the cylinder side provides a thermal flux of: $q_0 = 15w/m^2$, $(x, y, z) \in \Gamma_{surf}$.

The convective heat flux at the top of a cylindrical container is $q_0 = h(T_{ext} - T)$. Among them,

the heat transfer coefficient is $h = 750w/m^2 \cdot k$. External temperature is $T_{ext} = 273.15k$.

**Displacement and velocity conditions**

Meanwhile, the physical model coupled with 2 fields also needs to set mechanical boundary conditions, fixed for the bottom circle boundary $\Gamma_{fix}$, the boundary displacement vector.

$u(x_i, y_i, z_i) = 0$    $\Gamma_{fix} = \{(x, y, z) \mid x^2 + y^2 = r^2, z = 0, r = 2)\}$ . And the speed field is  $\dfrac{\partial \boldsymbol{u}}{\partial t} = 0$ .

Normal displacement of the shell is  $\boldsymbol{a}_r = 0$ ,  and normal velocity of the shell $\dfrac{\partial \boldsymbol{a}_r}{\partial t} = 0$ .

**Side load**: the upper bottom circle cover, due to gravity factors, will produce vertical and downward force, that is

$$F(x_i, y_i, z_i) = -1000N \mid_{\Gamma_G} \qquad \Gamma_G = \{(x, y, z) \mid x^2 + y^2 = r^2, z = 6, r = 2)\} . \qquad (60)$$

**Face load:** If the external side of the vessel receives a squeeze pressure $\overline{F} = 200N$ , it can be

represented as:

$$F_{n\times 3} = (F_i^x, F_i^y, F_i^z) = (100, 0, 0) \qquad \Gamma_G = \{(x, y, z) \mid x^2 + y^2 = r^2, 0 \le z \le 6)\} \qquad (61)$$
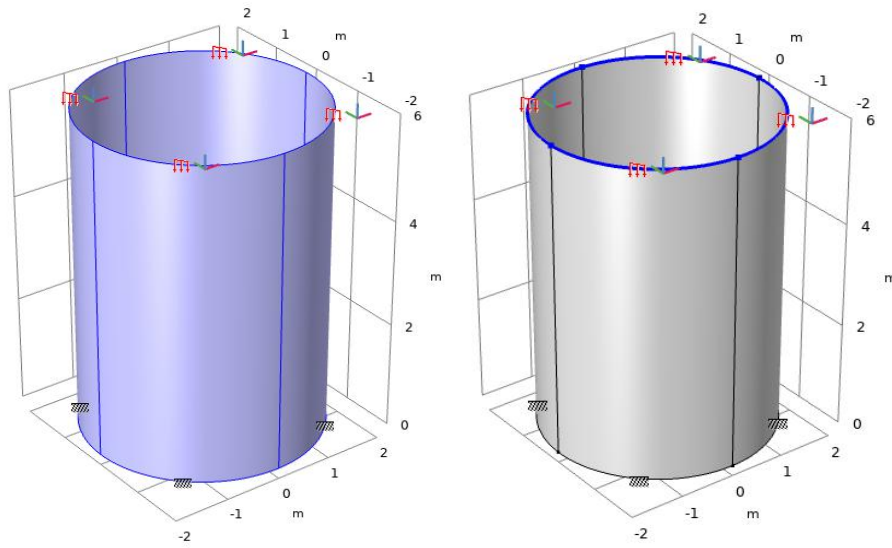


(a) Face load force area calibration        (b) top boundary force calibration

Fig3. Application of boundary load in different regions

**Numerical results**

Define the problem, and establish a suitable physical model, using a simple mathematical differential equation language to describe, through the establishment of the above multi-physical field theory and the boundary information given, the original problem needs to be pre-processed, the partial differential equation is transformed into the corresponding variational principle to obtain the corresponding Galerkin weak situation, and then calculate the stiffness matrix one by one, and finally the single stiffness matrix and into the total stiffness matrix.The total stiffness matrix is generally with weak singularity, because the corresponding boundary conditions are also added, so that the problem has a unique solution.There are many methods to add the boundary, such as replacement method, large number method, additional method and so on.Then the corresponding algebraic equation group is obtained, the corresponding algebraic equation group is solved, the coupling algebraic equation group solution method about temperature and force field is used, the LU decomposition embedded in COMSOL, and the multigrid algorithm is used.

Then, the temperature and stress information are solved, and the corresponding cloud images

are output. It can be seen from Fig.4 that Fig.4.(a) is the inner and outer layers of the container, the inner layer is blue, and the outer layer is red. Fig.4.(b) is mainly the local vector field of the cylindrical surface, and its effect can clearly distinguish the direction of the force when the load is loaded on the boundary. Fig.4.(c) is the three-dimensional steady-state temperature function cloud image solved in the coupling model, which can clearly observe that the high temperature region is concentrated in the middle part because of the heat source at the bottom and the heat flux at the top. Fig.4.(d) is the stress cloud image, and the source of the stress is two parts, one is the thermal stress caused by the heat source, and the other is the boundary load added to the mechanical mechanics. In addition, this map is also the deformation and stress color. On the same graph, the force on both sides of the symmetrical side at the bottom is the largest, and the force on the other positions is uniform. Due to the obvious deformation at the top of the extrusion container, the Fig4 simulation results can excavate different physical information from the same object, and the multi- physical field is more accurate than the single, considering the interaction between the fields.
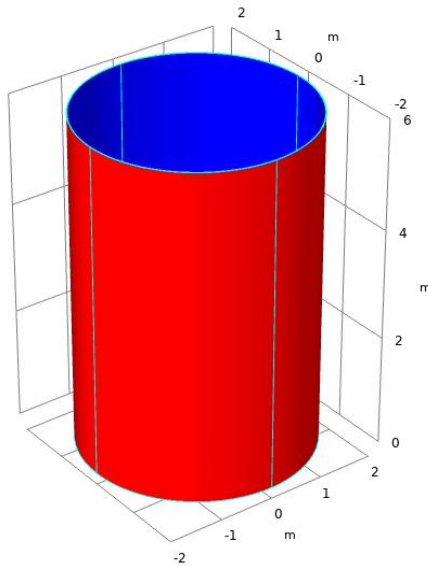


(a)Inner and outer layer geometry
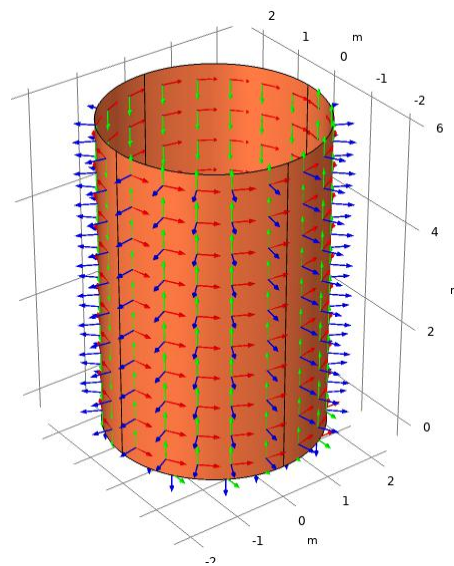


(b) Material thickness and local direction vector field of the container
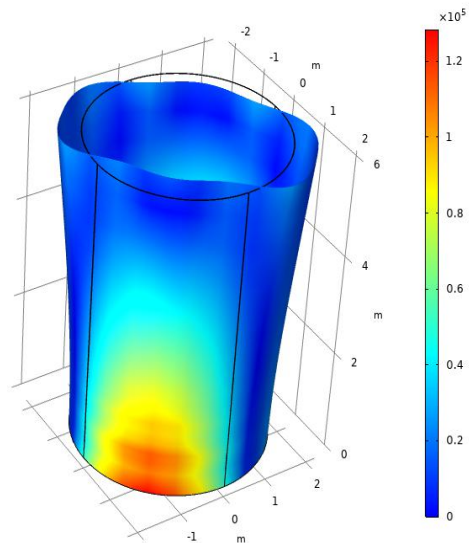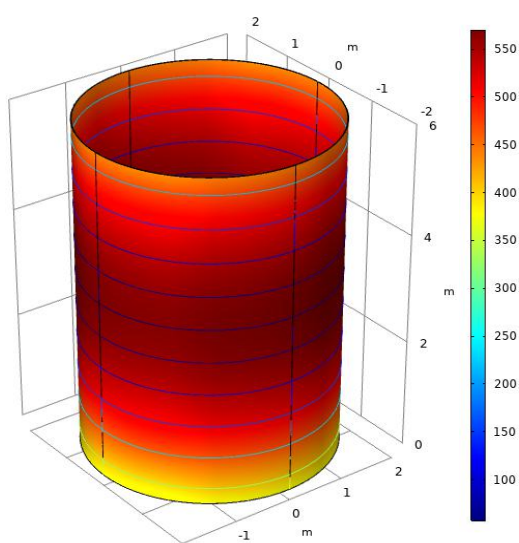
(c) 3D temperature field cloud map          (d) 3D thermal stress cloud map

Fig4.Numerical results visualization of temperature field and force field coupling

## Conclusion

This paper mainly studied the mechanical application of Cu-Zn composite materials in the couple physical field, involving the temperature field and structural mechanics field. In this work, the differential equation of transient two-field coupling and the steady-state two-field coupling equation are given. The numerical example is a three-dimensional steady-state two-field coupling physical model, which needs to be through multiple steps: setting geometric parameters, constructing geometric region, meshing, adding boundary conditions, simplifying the algebraic equation of weak form, and finally outputting the results. This example can well reflect the temperature information, stress, deformation and other information. The conclusion of this paper is that the quality of the mesh needs to be improved in meshing, which is helpful to the calculation speed and accuracy of the model. In addition, a physical model can be composed of multiple elements, such as CST, LST, Q4 and Q8. However, the common nodes on the interface of different elements must be reunited to ensure the correct transmission of physical data and computational information, and reduce the error of mesh construction. The selection of multiple physical fields must conform to the actual and objective development law. The small number of coupled fields cannot truly reflect the original problem, which is too difficult to solve. At the same time, the interference term is also added, the boundary conditions must be given when coupling, because the lack of boundary information may lead to unsolvable results or excessive errors. The iterative optimization method can be added to improve the iterative speed. In short, this paper is a simple two-physics coupling model, which provides the theoretical basis for thermal-mechanical multi-physics field analysis, gives numerical experiments, and uses COMSOL software to simulate the original problem. In the future, we will continue to explore the relevant applications of new functional materials in multi-physical fields.

## References

[1] Konovalov D A, Veretennikova I A , Bykova T M , et al. Development of an Approach to Det- ermining the Representative Volume Element of the Al/SiC Metal Matrix Composite Material Fabricated by Squeeze Casting. *Russian Metallurgy (Metally)*, 2020, **7**:738-745.

[2] Shen L, Zhao S, Bai Y, et al. An experimental study of spallation criterion in phenolic-resin based woven roving glass fibre reinforced composite material. *Acta Mechanica Sinica*, 1985, **1**(1):81-93.

[3] Zhao X, Ouyang J, Tan Q, et al. Interfacial characteristics between mineral fillers and phenolic resin in friction materials . *Materials Express*, 2020, **10**(1):70-80.

[4] None. Impact damage resistance of several laminated material systems. *Aiaa Journal*, 2015, **33**(6):1106-1113.

[5] Wisnom M R . Size effects in the testing of fibre-composite materials. *Composites Science & Technology*, 1999, **59**(13):1937-1957.

[6] Suk J W ,   Kitt A ,   Magnuson C W , et al. Transfer of CVD-grown monolayer graphene onto arbitrary substrates. *Acs Nano*, 2011, **5**(9):6916-6924.

[7] Bai H ,   Li C ,   Shi G . Functional composite materials based on chemically converted graphene. *Advanced Materials*, 2015, **23**(9):1088-1088.

[8] Zhang Y X , Yang C H . Recent developments in finite element analysis for laminated composite plates. *Composite Structures*, 2009, **88**(1):147-157.

[9] Liu, G. R. and Liu, M. B. (2003) *Smoothed Particle Hydrodynamics* — A Meshfree Particle Method, Chapter 6, World Scientific, Singapore.

[10] Liu, G. R. and Gu, Y. T. (2001) A point interpolation method for two-dimensional solids, *International Journal for Numerical Methods in Engineering* **50**, 937–951.

[11] Luo Q, Tong L. Multi-physics field models of photostrictive unimorphs and heterogeneous bimorphs subjected to light illumination and mechanical loading. *International Journal of Solids & Structures*, 2010, **47**(16):2006-2016.

[12] Du A K, Yang G, Lu D, et al. Densification of MnZn Ferrite Sintered under Multi-Physical Field Coupling. *Applied Mechanics and Materials*, 2013, 271-272.

[13] Li Z, Chen Q, Wang Q. Analysis of Multi-Physics Coupling Field of Multi-Degree-of-Freedom Permanent Magnet Spherical Motor. *IEEE Transactions on Magnetics*, 2019:1-5.

[14] Saadatfar M . Multiphysical time-dependent creep response of FGMEE hollow cylinder in thermal and humid environment. *Mechanics of Time-Dependent Materials*, 2019:1-23.

[15] Wan Q M . Research Progress in Physical Properties and Structural Design of Metamaterials. *Materials Science Forum*, 2021, **1027**:22-32.

[16] Li W, Wu Z, Tang H, et al. Research on Multi-Physical Fields of High-Power PMSM/G Used for FESS During the Process of Controllable Charging and Uncontrollable Discharging. *IEEE Transactions on Energy Conversion*, 2020, **35**(1):454-461.

[17] Wan Q M. Research Progress in Physical Properties and Structural Design of Metamaterials. *Materials Science Forum*, 2021, **1027**:22-32.

[18] Li Z, Wang Q, Qin B, et al. Vibration and acoustic radiation of magneto-electro-thermo-elastic functionally graded porous plates in the multi-physics fields. *International Journal of Mechanical Sciences*, 2020, **185**(1):105850.

# The finite element method for inverse wave scattering in rods

**\*Bang X. Trinh[1], †Cuong T. Nguyen[2]**

[1]Vietnamese-German University, Vietnam.
[2]School of Science and Technology, RMIT University, Vietnam.
\*Presenting author: trinhxuanbang103@gmail.com
†Corresponding author: cuong.nguyentan@rmit.edu.vn

## Abstract

This paper is concerned with the inverse problem for reconstructing the mass density of thin rods, directly in the time domain, based on the complete waveform response at one end. Mathematically, the inverse problem can be cast as a partial-differential-equation-constrained optimization problem which is regularized by the time-dependent Tikhonov regularization. Specifically, we seek to minimize a cost functional of the misfit between measured and computed responses. Numerical results are presented to show the accuracy and effectiveness of the proposed scheme in reconstructing sharp profiles of mass density even in the presence of noise in the data.

**Keywords:** wave motion, inverse problems, finite elements, Tikhonov regularization

## 1  Introduction

The reconstruction of the material profile based on collected measurements is central to many problems in engineering [1]. In this paper, we present a study in solving the inverse problem of recovering the mass density distribution along a thin rod. Specifically, to tackle this problem we utilize Tikhonov regularization for stabilizing the obtained ill-posed inverse problem. This technique was first introduced by Tadi [3][4], the basic idea is to treat the unknown parameters as time dependent and then the obtained equations are solved explicitly by the finite difference method. In this study, we generalize the governing equations in the inverse problem for thin rods with general values of the elastic modulus and the cross-sectional area. The main difference from [3][4] is that we solve the obtained strong-form equations using the finite element framework.

The problem setting is described as follows: consider a one-dimensional prismatic thin rod (Figure 1) with the cross-section area $A$; the left end is fixed, the right end is free and subjected to a force which is a prescribed excitation $f(t)$. We assume that the elastic modulus $E(x)$ is given whereas the mass density $\rho(x)$ is an unknown variable, which can be varied along the rod and need to be found. By using Lagrange multiplier to minimize the misfit functional under constraints, we end up with a partial-differential-equation-constrained optimization problem.
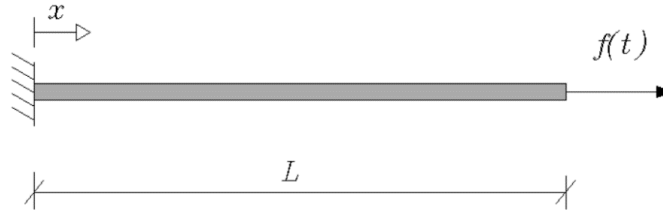


**Figure 1.** A thin rod subject to a pulse excitation $f(t)$ at the right end.

## 2 Governing equations

Consider a thin, straight, and prismatic rod, the longitudinal wave equation [2] can be written as

$$\frac{\partial}{\partial x}\left[E(x)\frac{\partial u(x,t)}{\partial x}\right]+q(x,t)=\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2},\tag{1}$$

where $u(x,t)$ is the axial displacement. If the elastic modulus is constant $E(x)=E$, we obtain

$$E\frac{\partial^2 u(x,t)}{\partial x^2}+q(x,t)=\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2}.\tag{2}$$

The effect of body force is neglected, i.e. $q(x,t)=0$ then Eq. (2) can be reduced to

$$\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2}=E\frac{\partial^2 u(x,t)}{\partial x^2},\tag{3}$$

where $x\in[0,L]$ and $t\in[0,T]$.

Next, the initial and boundary conditions can be described as follows:

$$\begin{cases}u(x,0)=u_t(x,0)=u(0,t)=0,\\ EAu_x(L,t)=f(t),\end{cases}\tag{4}$$

where $f(t)$ denotes the prescribed excitation at the right end of the rod. We assume that the displacement at the right end, $u(L,t)=u_m(t)$ can be measured. The inverse problem aims to recover the distribution of mass density $\rho(x)$ based on the observed displacement $u_m(t)$. In order to regularize the inversion, the unknow parameter is treated as time dependent [3], i.e. $\rho(x)\to \rho(x,t)$. We spatially approximate the mass density by using the linear shape functions of the finite element method. In particular, $\rho(x,t)$ can be approximated as

$$\rho(x,t)=\sum_{i=1}^{n}N_i(x)p_i(t)=\mathbf{N}\cdot\mathbf{p},\tag{5}$$

where $N_i(x)$ 's are the linear shape functions. In the master element, they are defined as:

$$N_1(\xi)=\frac{1}{2}(1-\xi);N_2(\xi)=\frac{1}{2}(1+\xi).$$

Mathematically speaking, one way to check how close of the real mass density and the guess one is to consider how small the misfit functional (least square) $J$ is, where $J$ is defined as:

$$J=\int_0^T\frac{1}{2}\left[u_m(t)-u(L,t)\right]^2 dt.\tag{6}$$

The computed response from Eq. (3) is denoted by $u(L,t)$ which is associated with a guess of $\rho(x)$. It is well-known that minimizing the misfit functional $J$ leads to an ill-posed problem. In

order to go around this issue, one should add a quadratic term to the misfit functional to make it more like a parabola, and thus making the optimization problem easier. This is the basic idea behind the Tikhonov regularization. Specifically, we shall try to minimize the *nearby* misfit functional defined as:

$$J = \int_0^T \left\{ \frac{1}{2} \left[ u_m(t) - u(L,t) \right]^2 + \frac{1}{2} \alpha \dot{\mathbf{p}} \cdot \dot{\mathbf{p}} \right\} dt. \tag{7}$$

where $\dot{\mathbf{p}}$ is the time rate of $\mathbf{p}$. The unconstrained minimization problem can be obtained by applying Lagrange multipliers to the wave equation (Eq. (3)), the misfit functional is modified as:

$$J = \int_0^T \left\{ \frac{1}{2} \left[ u_m(t) - u(L,t) \right]^2 + \frac{1}{2} \alpha \dot{\mathbf{p}} \cdot \mathbf{p} + \int_0^L \lambda(x,t) \left[ \sum_{i=1}^n N(x) p_i(t) u_{tt} - E u_{xx} \right] dx \right\} dt, \tag{8}$$

in which $\lambda(x,t)$ is the Lagrange multiplier.

Next, in order to find the minimum of $J$, we consider the first variation of $J$ as follows

$$\delta J = \int_0^T \left[ -\hat{u} \delta u(L,t) + \alpha \dot{\mathbf{p}} \cdot \delta \dot{\mathbf{p}} + \int_0^L \delta \lambda (\rho u_{tt} - E u_{xx}) dx + \right.$$

$$\left. \int_0^L \lambda(x,t) \left( \rho \delta u_{tt} + u_{tt} \sum_{i=1}^n N_i(x) \delta p_i - \delta u_{xx} \right) dx \right] dt = 0, \tag{9}$$

where

$$\hat{u}(t) = u_m(t) - u(L,t). \tag{10}$$

Integrating the term $\dot{\mathbf{p}} \cdot \delta \dot{\mathbf{p}}$ by parts leads to

$$\int_0^T \dot{\mathbf{p}} \cdot \delta \dot{\mathbf{p}} dt = \dot{\mathbf{p}} \cdot \delta \mathbf{p} \big|_0^T - \int_0^T \ddot{\mathbf{p}} \cdot \delta \mathbf{p} dt. \tag{11}$$

We choose $\dot{p}(T) = 0$ and $\delta p(0) = 0$ as in [3], two following terms in Eq. (9) can be expanded as

$$\int_0^T \lambda \rho \delta u_{tt} dt = \int_0^T (\lambda \rho)_{tt} \delta u dt + (\lambda \rho) \delta u_t \big|_o^T - (\lambda \rho)_t \delta u \big|_o^T, \tag{12}$$

$$\int_0^L \lambda \delta u_{xx} dx = \lambda \delta u_x \big|_0^L - \lambda_x \delta u \big|_0^L + \int_0^L \lambda_{xx} \delta u dx.$$

The left end is fixed, so its displacement is zero. At the right end, the external load is prescribed. In other words, $\lambda \delta u_x \big|_L = \delta u \big|_0 = 0$, substituting the above expansions into Eq. (9) results in

$$\delta J = \int_0^T \left[ \left( \lambda_x(L,t) - \hat{u}(t) \right) \delta u(L,t) + \int_0^L \delta\lambda (\rho u_{tt} - Eu_{xx}) dx + \int_0^L \left( (\lambda\rho)_{tt} - \lambda_{xx} \right) \delta u dx + \right.$$

$$\left. \sum_{i=1}^n \left( -\alpha\ddot{p}_i + \int_0^L \lambda u_{tt} N_i(x) dx \right) \delta p_i \right] dt = 0. \tag{13}$$

The terms $\delta u, \delta\lambda, p$ are arbitrary, in order to obtain the minimization of $J$, the following terms must be set to be zero

$$\begin{cases} \lambda_x(L,t) - \hat{u}(t) = 0, \\ \rho(x)u_{tt} - Eu_{xx} = 0, \\ (\lambda\rho(x))_{tt} - \lambda_{xx} = 0, \\ \alpha\ddot{p}_i = \int_0^L \lambda(x,t_i)u_{tt}(x,t_i)N_i(x)dx. \end{cases} \tag{14}$$

The boundary conditions for Lagrange multiplier are described as follows:

$$\lambda(0,t) = \lambda(x,T) = \lambda_t(x,T) = 0. \tag{15}$$

## 3   The Finite Element Method
### 3.1. The forward equation

We utilize the finite-element method to solve the wave equation of the rod with the associated Dirichlet and Neumann boundary conditions. The strong form can be stated as

$$\begin{cases} \rho u_{tt} - Eu_{xx} = 0 \\ u(0,t) = 0, EAu_x(L,t) = f(t). \\ u(x,0) = u_t(x,0) = 0 \end{cases} \tag{16}$$

The domain is discretized into $n$ equal elements. By using the linear shape functions, the displacement field can be approximated as

$$u(\xi) \approx \tilde{u}(\xi) = N_1(\xi)u^{e1} + N_2(\xi)u^{e2}. \tag{17}$$

The stiffness and mass matrices are given by [5]:

$$\mathbf{K}_e = \frac{EA}{L}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \mathbf{M}_e = \frac{\rho AL}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \tag{18}$$

Finally, by assembling all elements we obtain the following final system of equations:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}(t), \mathbf{u}(0) = \mathbf{0}, \dot{\mathbf{u}}(0) = \mathbf{0}. \tag{19}$$

Then, Newmark method with the average acceleration algorithm is applied to solve this second-order differential equation in time.

### 3.2. The adjoint equation

The strong form of adjoint problem is described by Eqs. $(14)_1$ & $(14)_3$ and (15). It is worth to note that Eq. $(14)_1$ plays the role of representing for the external load prescribed on the boundary

$$\begin{cases} \rho\lambda_{tt} - \lambda_{xx} = 0, \\ \lambda(0,t) = 0, \lambda_x(L,t) = \hat{u}(t), \\ \lambda(x,T) = \lambda_t(x,T) = 0. \end{cases} \tag{20}$$

In order to obtain $\lambda(x, t)$, we realize the similarity in the pattern of two systems of equations (16) and (20). Specifically, we proceed to solve for $\lambda(x, t)$ by the same manner as used to solve for $u(x, t)$, the only difference is that we need to go backward from $t = T$ to $t = 0$.

### 3.3. The updating equation

Eq. $(14)_4$ can be simplified to

$$\ddot{p}_i(t_j) = \frac{1}{\alpha} \int_0^L \lambda(x,t_j) \ddot{u}(x,t_j) N_i(x) dx. \tag{21}$$

With the stored values of $u(x, t), \lambda(x, t)$ and the piecewise *linear* shape functions $N_i(x)$, the definite integral on the right-hand side of Eq. (21) can be evaluated at node $i$ and time step $j$ by using the *trapezoidal* rule as follows

$$\ddot{p}_i^j := \ddot{p}_i(t_j) = \frac{1}{\alpha} \int_0^L b_i^j N_i dx$$
$$\Rightarrow \ddot{p}_i^j = \frac{1}{\alpha} \left[ \frac{1}{6} h(b_{i-1}^j + 2b_i^j) + \frac{1}{6} h(2b_i^j + b_{i+1}^j) \right] = \frac{h}{6\alpha} (b_{i-1}^j + 4b_i^j + b_{i+1}^j), \tag{22}$$

where

$$b_i^j = \lambda(x_i,t_j) u_{tt}(x_i,t_j). \tag{23}$$

At node 2 and time step $j$, since we do not need to consider node 1,

$$\ddot{p}_2^j = \frac{h}{6\alpha} \left( 4b_2^j + b_3^j \right). \tag{24}$$

At node $n$ and time step $j$, we have another approximation:

$$\ddot{p}_n^j = \frac{h}{6\alpha} \left( b_{n-1}^j + 2b_n^j \right). \tag{25}$$

In summary, we can compute the "acceleration" $\ddot{p}_i^j$ as follows:

$$\ddot{p}_i^j = B_i^j = \frac{h}{6\alpha}\begin{cases} 4b_i^j + b_{i+1}^j, i = 2 \\ b_{i-1}^j + 4b_i^j + b_{i+1}^j, i = 3,\ldots n-1. \\ b_{i-1}^j + 2b_i^j, i = n \end{cases} \tag{26}$$

Next, the values of $p_i^j$ can be determined by mean of forward Euler method. In particular, we approximate the acceleration as

$$\ddot{p}_i^j = \frac{\dot{p}_i^{j+1} - \dot{p}_i^j}{\Delta t} = B_i^j, \tag{27}$$

hence

$$\dot{p}_i^{j+1} = B_i^j \Delta t + \dot{p}_i^j, \tag{28}$$

In addition, the first derivatives of $p$ at initial and final time steps are chosen to be zero. Once the value of time rate of $\dot{p}$ is computed, we can also track for $p$ by reusing forward Euler method

$$p_i^{j+1} = \dot{p}_i^j \Delta t + p_i^j \tag{29}$$

### 3.4. Algorithm

We follow a procedure of three steps to recover the distribution of the mass density:

- Solving the forward equation: apply the finite element method for spatial discretization and Newmark method for temporal discretization, solve the wave equation with a guess for mass density, and then the obtained displacements at all nodes are stored as $u(x,t)$.
- Solving the adjoint equation: keep the mass density is constant and obtain $\lambda(x,t)$.
- Solving the updating equation: with the values of $u(x,t)$ and $\lambda(x,t)$ are stored, forward Euler method is applied via two steps to update $\rho$. The updated guess for mass density is obtained from the value of $\rho$ at the final time step $T$. We then check the error between the measured displacement and the computed displacement which is correspondent to the new guess mass density by considering the norm of misfit functional. If the error is still larger than a prescribed tolerance, we repeat the above procedure. The value of $\rho(x,T)$ then is used to update the new guess for density.

---

**Algorithm: Inversion of the mass density**

---

0: Input $E, A, L, n, f(t), \mathbf{u}_m(t)$

1: Choose $\alpha$

2: Set $k = 0$

3: Set initial guess for $\rho_0 \rightarrow \mathbf{p}(x, 0)$

4: Set convergence tolerance TOL

5: **while** (misfit > TOL) **do**

6:      Solve the forward problem to obtain $\mathbf{u}(t)$

7:      Solve the adjoint problem to obtain $\lambda(t)$

8:      Compute $\ddot{p}_i(t_j)$ using

$$\ddot{p}_i^j = B_i^j = \frac{h}{6\alpha} \begin{cases} 4b_i^j + b_{i+1}^j, i = 2 \\ b_{i-1}^j + 4b_i^j + b_{i+1}^j, i = 3, \ldots n-1. \\ b_{i-1}^j + 2b_i^j, i = n \end{cases}$$

9:      Update the parameter using

$$\dot{p}_i^{j+1} = B_i^j \Delta t + \dot{p}_i^j, \quad p_i^{j+1} = \dot{p}_i^j \Delta t + p_i^j$$

10: $k = k + 1$

11: **end while**

12: Save final estimates $\rho(x) = \rho(x, T)$.

---

## 4   Numerical examples

In this section, we consider three scenarios in which the mass density is constant, the mass density varies along the rod and the noise appears in the measurement data.

### 4.1. *Constant mass density*

In this case, the rod is discretized into 100 elements with equal lengths, the observation time $T = 2$ s and the time step is chosen as $\Delta T = 0.001$ s. The right end is subject to a unit-step excitation $f(t) = H(t)$, which is the Heaviside step function. Parameters are chosen as $TOL = 10^{-5}, \alpha = 10^{-6}$. In choosing the initial guess, we try three different values $\rho_0 = 0.3; 0.5; 0.7$, which are smaller than the real mass density ($\rho_{\text{target}} = 1$). Figure 2 shows the history of searching the mass density using the presented algorithm. It can be seen from this Figure, all three initial guesses lead to the correct convergence result. The bigger difference between the target value and the initial guess requires a higher number of iterations toward the convergence. In Figure 3, we test various values of the initial guess on another side of the real mass density. The convergence can also be obtained accurately by using the algorithm we present.
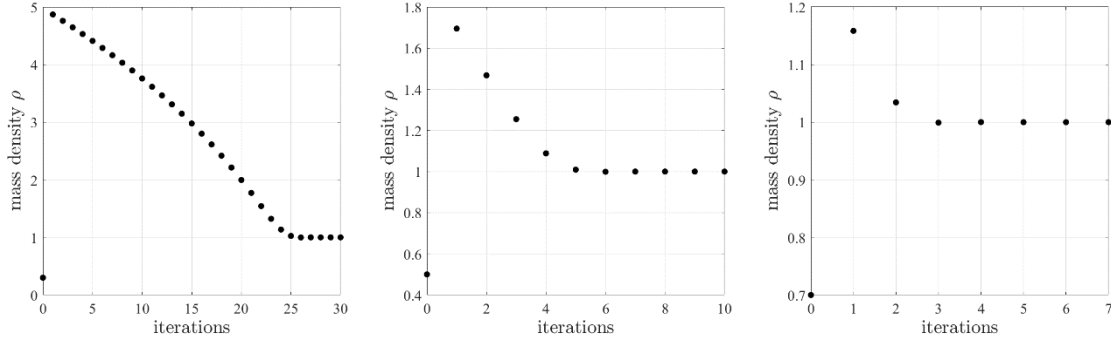
**Figure 2.** Initial guess $\rho_0 = 0.3$ (left), $\rho_0 = -0.5$ (center), $\rho_0 = 0.7$ (right).
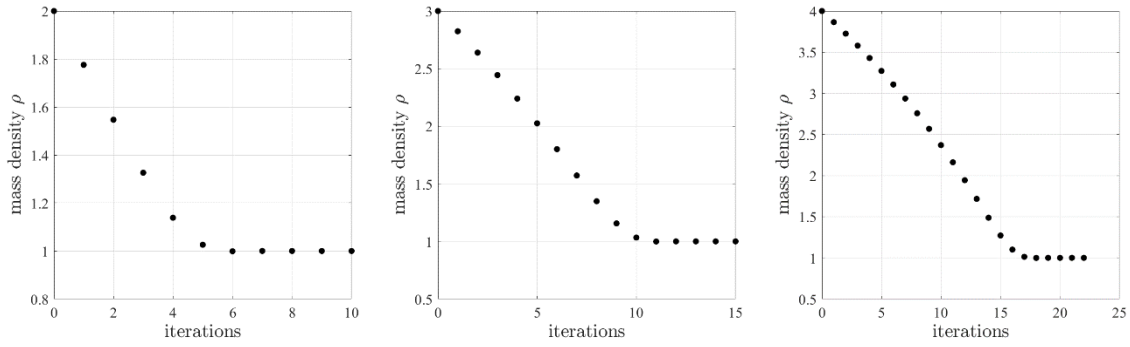


**Figure 3.** Initial guess $\rho_0 = 2$ (left), $\rho_0 = 3$ (center), $\rho_0 = 4$ (right).

### 4.2. The mass density varies along the rod

We reconsider four numerical examples presented in section 4 of [3]. In the first three numerical experiments, the mass density profile is described by smoothly exponential functions. Meanwhile, the profile is given by a sharp function in Example 4. The applied force for all cases can be expressed as

$$f(t) = \exp\left[-\frac{(t-0.1)^2}{0.0025}\right].$$

The rod is divided into 240 elements with equal lengths. The observation time is $T = 3$ s while time step is $\Delta t = 0.0015$ s. The time range is chosen such the wave starting from the free-end then reflects from the fixed-end and goes back to the free-end. One important parameter of this regularization scheme is the parameter $\alpha$, the smaller value of $\alpha$ causes faster convergence. However, the inverse problem is ill-posed when $\alpha$ goes to zero. In this study, we start to consider the value of $\alpha$ at 0.0001 and gradually reduce to $\alpha = 10^{-8}$.

*Example 1:* The real density of material is given by the following expression:

$$\rho(x) = 1 + \exp\left[-\frac{(x-0.5)^2}{0.04}\right].$$
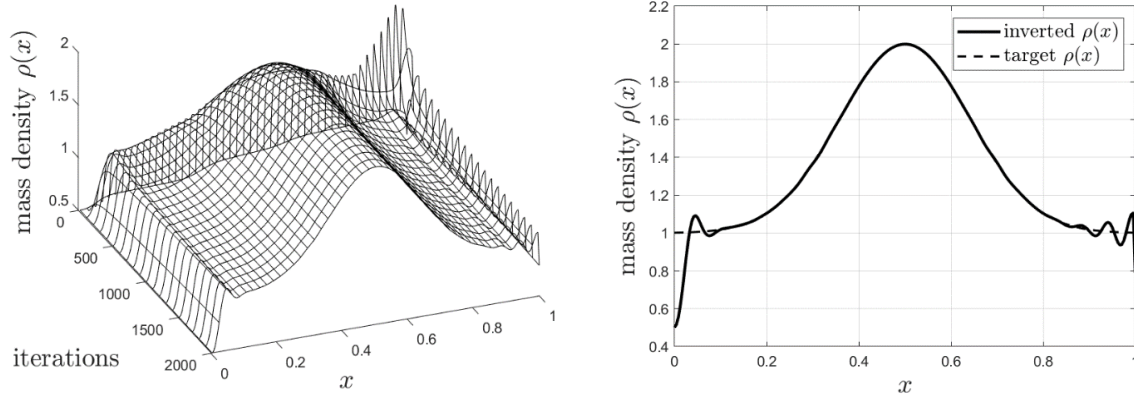
**Figure 4.** Example 1: Initial guess $\rho_0(x) = 0.5$.
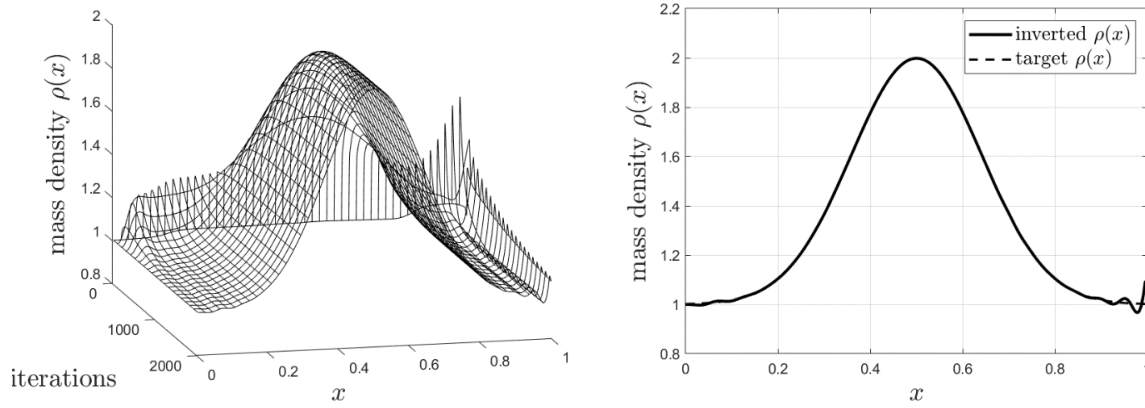


**Figure 5.** Example 1: Initial guess $\rho_0(x) = 1$.

Figure 4 (right) shows the inverted mass density at the 2000$^{th}$ iteration when $\rho_0(x) = 0.5$, the error-norm of misfit function is 0.008078. The displacement at the fixed end is not evaluated, therefore the inverted density at this node remains as the initial guess and it causes the recovered density equals 0.5 as shown in the Figure. The oscillation at both ends reduces when initial guess chosen closer to the actual value of density (Figure 5 (right)).

*Example 2:* We now consider the recovery problem in which the actual density is given as:

$$\rho(x) = 1 + \exp\left[-\frac{(x-0.26)^2}{0.02}\right] + \exp\left[-\frac{(x-0.74)^2}{0.02}\right].$$
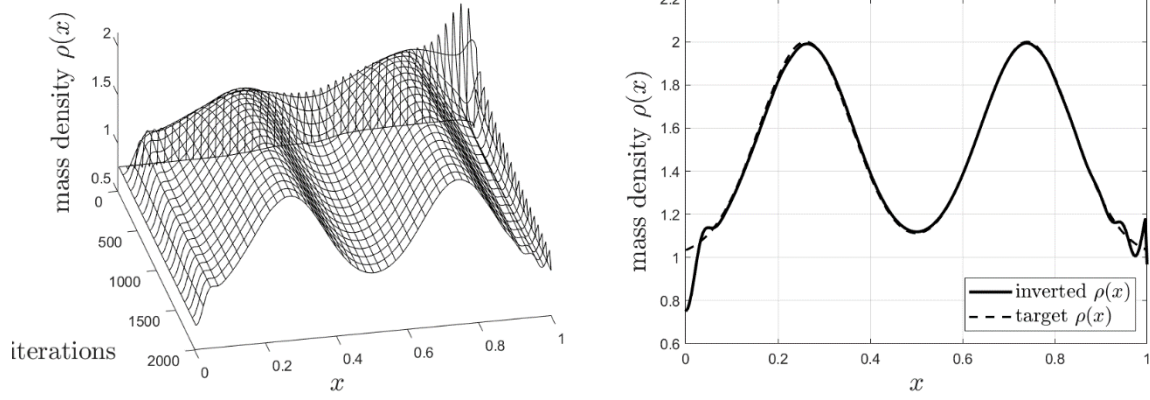
**Figure 6.** Example 2: Initial guess $\rho_0(x) = 0.75$.
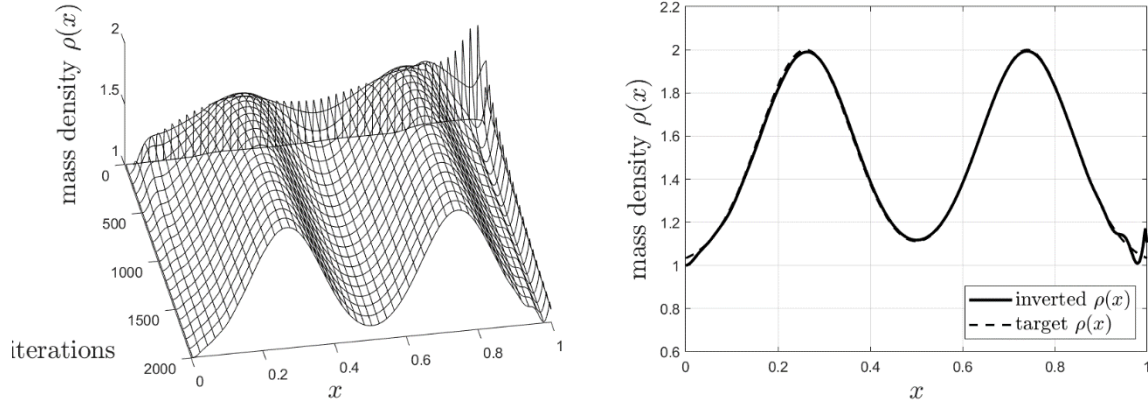


**Figure 7.** Example 2: Initial guess $\rho_0(x) = 1$.

It is clear that the real mass density has two waveforms. In both cases of the initial guess $\rho_0(x) = 0.75$ and $\rho_0(x) = 1$, the inverted results match the target values.

*Example 3:* We next consider another case in which the mass density rapidly changes in space. It is possible to check the density at the free-end and propose a better initial guess. The density is given by:

$$\rho(x) = 1 + 0.8\left( \exp\left[ -\frac{(x-0.5)^2}{0.005} \right] + \exp\left[ -\frac{(x-0.6)^2}{0.005} \right] \right).$$
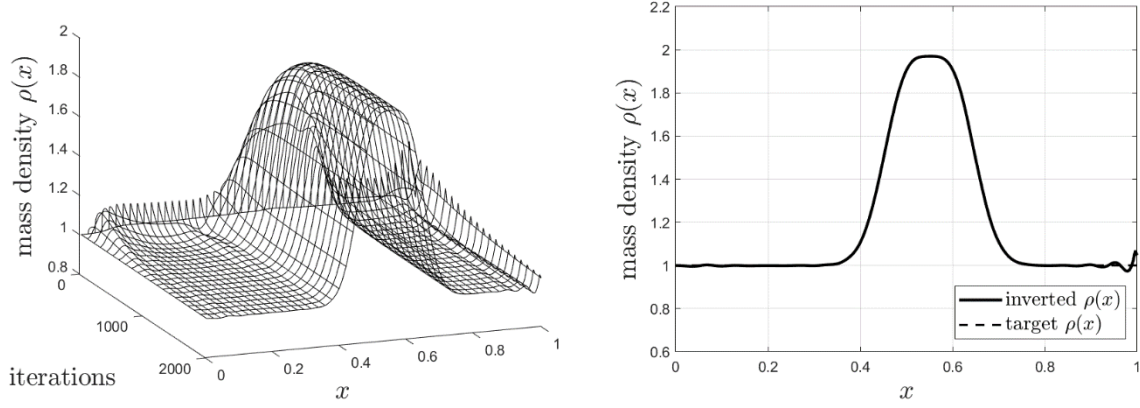
**Figure 8.** Example 3: Initial guess $\rho_0(x) = 1$.

*Example 4:* Finally, we consider the inverse problem in which the density is not a continuous function. The mass density is described is by

$$\rho(x) = 1.5H(x) + 0.5H(x - 0.35) - H(x - 0.7) - H(x - 1).$$

We applied the same numerical parameters as in *Example 3*, Figure 9 demonstrates that the regularization scheme successfully recovers the density profile.
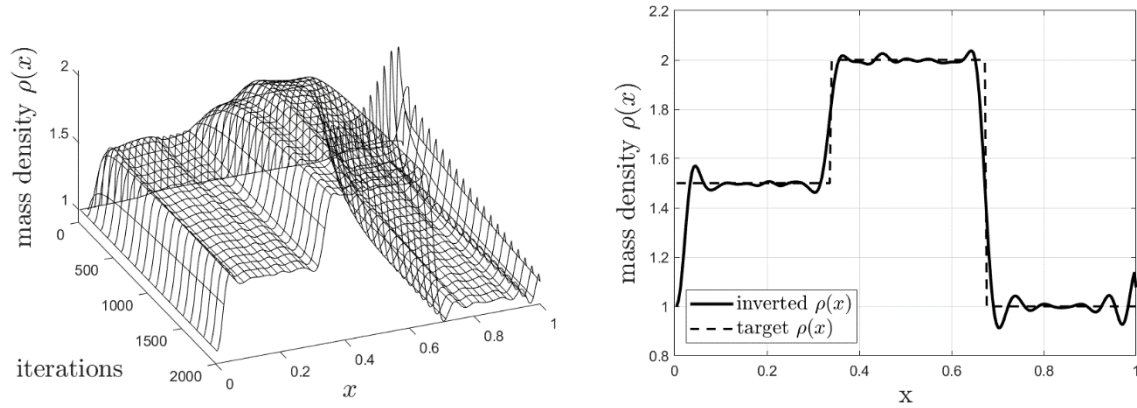


**Figure 9.** Example 4: Initial guess $\rho_0(x) = 1$.

### 4.3. Noisy measurements

In practice, we are typically not able to observe $u_m(t)$ directly but its noise-corrupted value $u^{obs}(t_j) = u_m(t_j) + e_j$. The noise can be considered as the inaccuracy in observation devices. In order to test the algorithmic performance in such a case, we again consider four numerical experiments in section 4.2, but the obtained measurements now are noisy. Figure 10 - Figure 13 show the inverted results for three profiles of Gaussian noises in which the mean value of $\{e_j\}$ is zero. As can be seen, the presented algorithm works well even in the presence of noise.
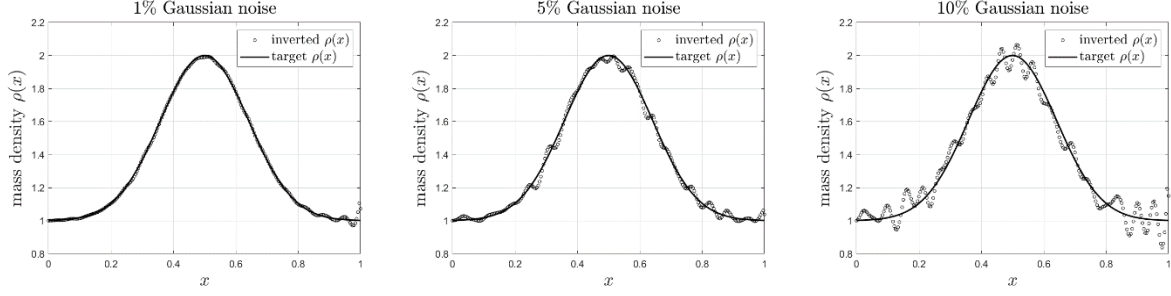
**Figure 10.** Example 1: Initial guess $\rho_0(x) = 1$ with Gaussian noises in $u_m(t)$.
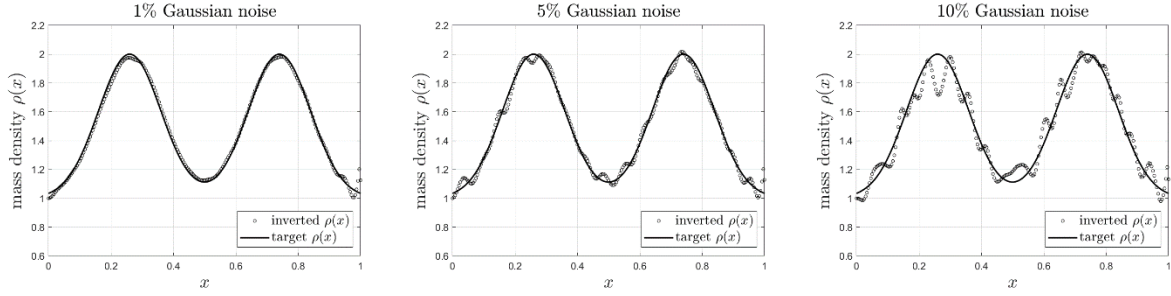


**Figure 11.** Example 2: Initial guess $\rho_0(x) = 1$ with Gaussian noises in $u_m(t)$.
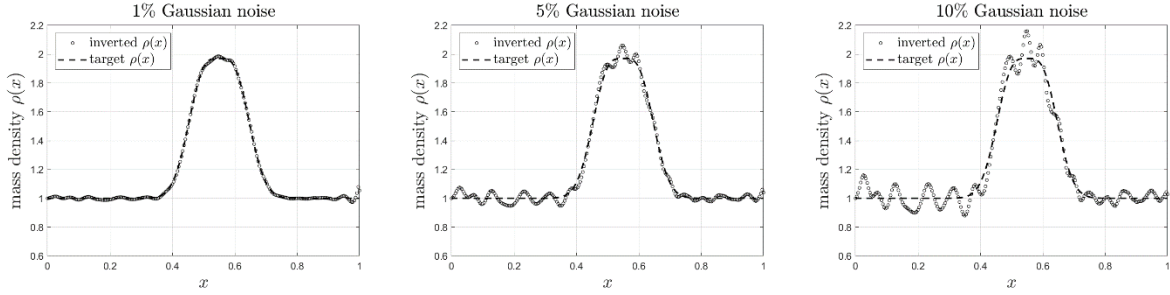


**Figure 12.** Example 3: Initial guess $\rho_0(x) = 1$ with Gaussian noises in $u_m(t)$.
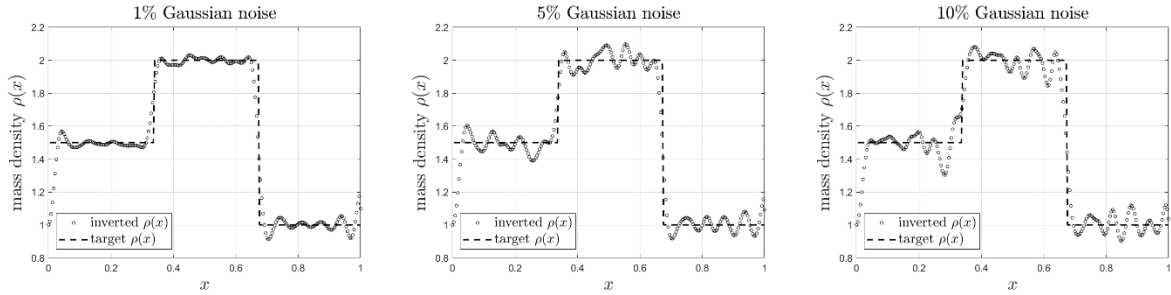


**Figure 13.** Example 4: Initial guess $\rho_0(x) = 1$ with Gaussian noise in $u_m(t)$.

## 5   Conclusions

Based on the numerical results presented in section 4, we conclude that our approach to the regularization scheme successfully recovers the distribution of mass density which varies along the domain. For the problem in which the mass density is a discontinuous function (layered media), our approach can also provide acceptable results. It was also shown that the algorithm can be used to predict accurately the system parameters even when noise appears in the measurement data.

## References

[1] Marc Bonnet, A. Constantinescu (2005), Inverse problems in elasticity, *Inverse Problems,* 21, R1-R50.
[2] Karl F. Graff (1975), Wave motion in elastic solids, *Oxford University Press.*
[3] M. Tadi (1997), Explicit method for inverse wave scattering in solids, *Inverse Problems* **13**, 509-521.
[4] M. Tadi (1998), Evaluation of the elastic property based on boundary measurement, *Acta Mechanica* **129**, 231-241.
[5] TJR Hughes (2012), The finite element method: linear static and dynamic finite element analysis, *Courier Corporation.*

# Transformed Newton's method with a fixed-point iteration for highly nonlinear problems in structural mechanics

**\*Ngoc M. La[1], †Cuong T. Nguyen[2], †Minh N. Dinh[2]**

[1]Vietnamese German University, Vietnam.
[2]School of Science and Technology, RMIT University, Vietnam.

\*Presenting author: lmngocdhsp@gmail.com
†Corresponding authors: cuong.nguyentan@rmit.edu.vn, minh.dinh4@rmit.edu.vn

**Abstract**

In this paper, a nonlinear solver combining fixed-point iteration and transformed Newton's method is first proposed. The transformed Newton's method was recently introduced to decrease the degree of nonlinearity of problems in solid mechanics. The key contribution behind this work is to modify the starting point of each iteration of the transformed method. Specifically, the transformed method gets started with the previous converged solution while the proposed solver starts at an initial guess theoretically proved to be close to the converged root of the current step. The advantage of the proposed nonlinear solver lies in the simple implementation and the significant reduction in number of iterations compared with the purely transformed Newton's method. Numerical results are presented to show the accuracy and efficiency of the proposed solver in dealing with highly nonlinear problems in structural mechanics.

**Keywords:** Nonlinear elasticity, Newton's method, fixed-point iteration.

## 1    Introduction

Nonlinear behavior of solids takes two typical forms: material and geometric nonlinearity. The former occurs when the stress is not linearly proportional to the strain, whereas the latter is important when changes in geometry, however large or small, have a significant effect on the response of structures [1]. Modelling and simulating nonlinear structures require robust solvers for solving the obtained nonlinear system of equations. A well-known and widely used iterative algorithm is Newton's method. The basic idea of Newton's method is to linearize the governing equations for obtaining the solution. Specifically, in each force increment step a successive equation is established to compute iteratively the next estimate closer to the real root. Highly nonlinear problems can lead to slow convergence which results in expensive computational cost.

In biomechanics, the common characteristic of many of the proposed constitutive laws is the exponential relation between stress and strain [2]. The exponential feature drastically increases the level of nonlinearity of problems. Recently, Yue Mei et al. [3] proposed a transform technique toward reducing the nonlinear degree to improve the performance of Newton's method. In particular, Newton's method is modified by applying a transformation before linearization. The transformed problem possesses significantly reduced nonlinearity, and thus convergence properties can be improved.

It is also well-known that the convergence properties of Newton's method depend heavily on the initial guess, poor choices often lead to slow convergence or divergence and it is quite time consuming. The main idea of this study is to develop a technique that leads to a reduction in

nonlinearity and alleviating the limitation of Newton's method in choosing the initial guess. The fixed-point iteration and its corresponding natural Newton's method [4] have been used to improve the initial guess and increase the order of convergence. The mathematical formulation of the proposed approach is introduced in sections 2 and 3. We then illustrate the application of our proposed combination to numerical examples about 1DOF and 2DOF truss systems in section 4. We proceed to assess the performance of the new formulation and show the improved convergence property compared with Newton's method and transformed Newton's method.

## 2 Fixed point iteration

### 2.1 Newton's method

Suppose that the nonlinear function $f(x)$ is continuous and there exists its first derivative $f'(x) \neq 0$ for all $x \in (x^* - \delta, x^* + \delta)$, where $f(x^*) = 0$.

Let $x_0 \in (x^* - \delta, x^* + \delta)$ be an initial guess and $|x^* - x_0|$ is sufficiently "small". The function $f(x)$ is evaluated at $x^*$ by expanding in Taylor's series about a point up to second order as follows

$$f(x^*) = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^* - x_0)^2}{2} f''(\xi(x^*)), \tag{1}$$

where $\xi(x^*)$ lies between $x^*$ and $x_0$. Since $f(x^*) = 0$, Eq. (1) leads to

$$0 = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^* - x_0)^2}{2} f''(\xi(x^*)). \tag{2}$$

Newton's method is derived by assuming that $|x^* - x_0|$ is small, thus the term involving $(x^* - x_0)^2$ is much smaller, so

$$0 \approx f(x_0) + (x^* - x_0)f'(x_0).$$

Slightly rearranging this equation leads to $x^* \approx x_0 - \dfrac{f(x_0)}{f'(x_0)} \equiv x_1$. By applying this process iteratively, we obtain a sequence $\{x_n\}_{n=0}^{\infty}$ which is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{3}$$

in each iteration.

### 2.2 Iteration function

We would like to find $x = x^*$ such that $f(x^*) = 0$ for a given differentiable function $f : K \to K \ (K = \mathbb{R} \text{ or } \mathbb{C})$. In order to solve this problem, starting with an initial guess to the solution $x_n = x_0$ the guess is iteratively updated using $x_{n+1} = \Phi(x_n), \ n = 0, 1, 2, \ldots$ where the

iteration function $\Phi(x)$ depends on $f(x)$. It is required that $x_n \to x^*$ as $n \to \infty$ for the numerical scheme to be converged to the true solution.

## *2.3 Choosing the initial guess*

Following theorem 2.1 in [4], let $\Phi(x)$ is an iteration function of $f(x)$, $\Phi^{(1)}(x)$ denotes its first derivative and $x_0$ is an initial guess. If $\left\| \Phi^{(1)}(x^*) \right\| < 1$, then there exists a neighborhood of $x^*$ such that for any $x_0$ in that neighborhood the sequence converges to $x^*$.

*Convergence analysis:*
From the proof for theorem 2.1 in [4], by continuity, there is an interval $I_\rho(x^*) = (x^* - \rho, x^* + \rho)$ such that $\left\| \Phi^{(1)}(x) \right\| \leq \dfrac{1 + \left\| \Phi^{(1)}(x^*) \right\|}{2} = L < 1$. Then, if $x_n \in I_\rho(x^*)$, we have

$$\left\| x_{n+1} - x^* \right\| \leq L \left\| x_n - x^* \right\| \leq \left\| x_n - x^* \right\| \leq \rho \tag{4}$$

and $x_{n+1} \in I_\rho(x^*)$. Moreover,

$$\left\| x_n - x^* \right\| \leq L^n \left\| x_0 - x^* \right\|, \tag{5}$$

and the sequence $x_{n+1} = \Phi(x_n)$ converges to $x^*$ as $n \to \infty$ because $0 \leq L < 1$.

*Remark:* We assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ and define the spectral radius of $\mathbf{A}$ by
$$\rho(\mathbf{A}) := \max \left\{ |\lambda| : \lambda \in \mathbb{C} \text{ is an eigenvalue of } \mathbf{A} \right\}.$$
Then every subordinate matrix norm on $\mathbb{R}^{n \times n}$ satisfies the inequality $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$. That leads to $\|\mathbf{A}\| < 1$ if and only if $\rho(\mathbf{A}) < 1$. Thus, the main condition of the previous theorem could also be equivalently reformulated as $\rho(\Phi^{(1)}) < 1$. When we extend to *vector* nonlinear equations, $\Phi^{(1)}(\mathbf{x})$ is the Jacobian matrix of $\Phi(x)$ at $x$,. The norms are defined as

$$\|.\|_\infty = \text{row sum norm} \qquad \|\mathbf{M}\|_\infty := \max_{i=1,\ldots,m} \sum_{j=1}^{n} |M_{ij}| \tag{6}$$

$$\|.\|_1 = \text{column sum norm} \qquad \|\mathbf{M}\|_1 := \max_{j=1,\ldots,n} \sum_{i=1}^{m} |M_{ij}| \tag{7}$$

*Application:*
Newton's method for finding roots of a given function $f(x)$ is Eq. (3)

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

By choosing $\Phi(x_n) = x_n - \dfrac{f(x_n)}{f'(x_n)}$ is an iteration function, we can rewrite the Newton iteration as the fixed-point iteration: $x_{n+1} = \Phi(x_n)$, then

$$\Phi^{(1)}(x_n) = \frac{f(x_n)f''(x_n)}{f'^2(x_n)}. \tag{8}$$

## 3    Transformed Newton with fixed-point iteration method

### 3.1 Degree of nonlinearity

In this section, we summarize the basic idea behind the transformed Newton method presented in [3]. The degree of nonlinearity at the current point $x_n$ is defined as $N(\zeta, x_n) := \sup\limits_{\zeta \in (x_n, x^*)} \left| \dfrac{f''(\zeta)}{2f'(x_n)} \right|$ for $\zeta \in [x_n, x^*]$. We start with a simple exponential function $f(x): Ae^{Bx} = H$. The standard formulation reads:

$$N_{\text{standard}} = \sup_{\zeta \in (x_n, x)} \left| Be^{B(\zeta - x_n)} \right| = Be^{B\Delta x}. \tag{9}$$

It shows clearly that the error in the Newton's method will rise as when increasing $\Delta x$. A more practical model of force–displacement relation is $f(x): A(e^{Bx} - 1) = H$, using the transformation $\log\left[ A(e^{Bx} - 1) \right] = \log[H]$, we obtain

$$N_{\text{transform}} = \sup_{\zeta \in (x_n, x)} \left| \frac{-\left( \dfrac{1}{f(\zeta)} ABe^{B\zeta} \right)^2 + \dfrac{1}{f(\zeta)} AB^2 e^{B\zeta}}{\dfrac{1}{f(x_n)} ABe^{Bx_n}} \right| = \frac{B}{e^{Bx_n} - 1}. \tag{10}$$

It is worth to note that the nonlinearity in the log formulation is non-zero but does not depend on $\Delta x$. It only depends on the initial guess $x_n$ and reduces when increasing $x_n$. Then, the ideal scenario to apply the transformed method occurs when its degree of nonlinearity is smaller than the one produced by the standard method, i.e.

$$N_{\text{transform}} < N_{\text{standard}} \Rightarrow \frac{B}{e^{Bx_n} - 1} < \frac{Be^{Bx}}{e^{Bx_n}}. \tag{11}$$

As can be seen from Eq. (11), when $Bx_n$ is extremely small the standard method is expected to be better than the transform method.

### 3.2 Transformed Newton's method

In this section, we briefly go through the important steps in the procedure using the transformed technique proposed by Yue Mei et. al. in [3]. Equilibrium equations are established with respect

to the current position by assembling the typical internal forces $T_i$ and the external forces $F_i$ at all nodes $(i = 1, 2, ..., N)$, the residual or out of balance nodal force $R_i$ as the balance between the internal and the external forces as

$$R_i(x) = T_i(x) - F_i(x) = 0 \qquad \forall i = 1, ..., N. \tag{12}$$

We utilize a transformed equation to solve the standard equation Eq. (12)

$$\mathcal{T}(T_i(x)) = \mathcal{T}(F_i(x)) \qquad \forall i = 1, ..., N \tag{13}$$

for a pre-determined bijective transformation $\mathcal{T} : \mathbb{R} \to \mathbb{R}$.

Our aim is to determine a simple transformation that reduces the nonlinearity. Many of the constitutive models include an exponential function, which is the primary source of nonlinearity. Thus, we suppose a highly-simplified form $\sigma \sim \exp(\lambda)$ which could be reduced the nonlinearity by taking a logarithm, i.e., $\mathcal{T} \equiv \log$. Using finite element discretization, we write the linearized system of equations as

$$\sum_j K_{ij} \Delta x_j = \bar{R}_i(x_n), \tag{14}$$

where the modified residual

$$\bar{R}_i(x_n) = \begin{cases} T_i(x_n) \log\left(\dfrac{F_i(x_n)}{T_i(x_n)}\right) & \text{if condition (16) is satisfied} \\ F_i(x_n) - T_i(x_n) & \text{otherwise} \end{cases} \tag{15}$$

and with tolerance TOL, the transformed condition is:

$$|F_i(x_n)| > TOL, |T_i(x_n)| > TOL, \text{ and } \frac{F_i(x_n)}{T_i(x_n)} > 0. \tag{16}$$

### 3.3 Transformed Newton method enhanced by fixed-point iteration

In this section, we present algorithms in detail for applying transformed Newton method approach with fixed-point iteration. For problems of multi degree of freedom (MDOF), the fixed-point iteration requires computing the Jacobian matrix so its algorithm need to be presented separately.

*3.3.1  Scalar nonlinear equations for problems of single degree of freedom (SDOF)*

**Algorithm: Transformed Newton method enhanced by fixed-point iteration for SDOF problems**

Input: geometry, material properties, and solution parameters, $x$: displacement vector, and $f(x)$: internal force vector.

1: Input   A: the initial shear modulus
$\qquad$   B: dimensionless parameter
$\qquad$   L: initial length truss
$\qquad$   nincr: number of load increment
$\qquad$   fincr: force increment
$\qquad$   area: initial area
$\qquad$   E: Young modulus
$\qquad$   maxiteration: maximum number of fixed-point iteration

2: $F \leftarrow 0, \; R \leftarrow 0$

3: $TOL \leftarrow$ convergence tolerance

4: $x_0 \leftarrow$ initial guess

5: Compute the tangent stiffness $K_t(x)$

6: **for** k = 1: nincr **do**

7: $\qquad F \leftarrow F + fincr$

8: $\qquad$ **for** m = 1:maxiteration **do**

9: $\qquad\qquad r(x) \leftarrow f(x) - F$  (residual)

10: $\qquad\qquad N \leftarrow x - \dfrac{r(x)}{r'(x)}$  (establish the iteration function)

11: $\qquad\qquad dN \leftarrow \dfrac{r(x)r''(x)}{\left[r'(x)\right]^2}$  (typically Eq.(8))

12: $\qquad\qquad$ **if** $\left|dN(x = x_0)\right| < 1$

13: $\qquad\qquad$ break

14: $\qquad\qquad$ **end if**

15: $\qquad\qquad x_0 \leftarrow N(x_0)$

16: $\qquad$ **end for**

17: $\qquad x \leftarrow x_0$

18: $\qquad K \leftarrow K_t(x_0)$

19: $\qquad R \leftarrow R - fincr$

20: $\qquad$ **while** $\left(\|R\| / \|F\| > tolerance\right)$ **do**

21: $\qquad\qquad$ Solve $Ku = R$

22: $\qquad\qquad x \leftarrow x + u$

23: $\qquad\qquad T \leftarrow f(x), \; K \leftarrow K_t(x)$

24: $\qquad\qquad$ **if** $|F| > TOL, |T| > TOL,$ and $F/T > 0$ **do**

| 25: | | $R \leftarrow T*\log(F/T)$ |
| 26: | **else** | |
| 27: | | $R \leftarrow F-T$ |
| 28: | **end if** | |
| 29: | **end while** | |
| 30: | $x_0 \leftarrow x$ | |
| 31: **end for** | | |

*3.3.2    Vector nonlinear equations for problems of multi degree of freedom (MDOF)*

**Algorithm 2: Transformed Newton method enhanced by fixed-point iteration for MDOF problems**

Input: geometry, material properties, and solution parameters, $u = [u_1; u_2]$: displacement vector, and $f(u_1, u_2)$: internal force vector.

1: Input    A: the initial shear modulus
               B: dimensionless parameter
               L: initial length truss
               nincr: number of load increment
               fincr: force increment
               area: initial area
               E: Young modulus
               maxiteration: maximum number of fixed-point iteration

2: $F \leftarrow [0;0]$, $R \leftarrow [0;0]$

3: $TOL \leftarrow$ convergence tolerance

4: $u_0 \leftarrow$ initial guess

5: Compute the tangent stiffness matrix: $K_t(u_1, u_2)$

6: **for** k = 1: nincr **do**

7:       $F \leftarrow F + fincr$

8:       $r(u_1, u_2) \leftarrow f(u_1, u_2) - F$ (residual)

9:       $N(u_1, u_2) \leftarrow [u_1; u_2] - \dfrac{r(u_1, u_2)}{K_t(u_1, u_2)}$

10:      $DN = jacobian(N, [u_1; u_2])$

11:      **for** m = 1: maxiteration **do**

12:             $A \leftarrow DN(u_0)$

13:             $spectral \leftarrow \max(abs(eig(A)))$

14:             **if** $(spectral) < 1$

15:             **break**

16:             **end if**

17:             $u_0 \leftarrow N(u_0)$

18:      **end for**

| | |
|---|---|
| 19: | $u \leftarrow u_0$ |
| 20; | $K \leftarrow Kt(u_0)$ |
| 21: | $R \leftarrow R - fincr$ |
| 22: | **while** $(\|R\|/\|F\| > tolerance)$ **do** |
| 23: | $\quad$ Solve $Kx = R$ |
| 24: | $\quad u \leftarrow u + x$ |
| 25: | $\quad T \leftarrow f(u),\ K \leftarrow Kt(u)$ |
| 26: | $\quad$ **for** $i = 1:2$ **do** |
| 27: | $\quad\quad$ **if** $\|F_i\| > TOL, \|T_i\| > TOL,$ an*d* $F_i/T_i > 0$ **do** |
| 28: | $\quad\quad\quad R_i \leftarrow T_i * \log(F_i/T_i)$ |
| 29: | $\quad\quad$ **else** |
| 30: | $\quad\quad\quad R_i \leftarrow F_i - T_i$ |
| 31: | $\quad\quad$ **end if** |
| 32: | $\quad$ **end for** |
| 33: | **end while** |
| 34: | $u_0 \leftarrow u$ |
| 35: | **end for** |

## 4   Numerical examples

In this section, we consider two truss systems as depicted in Figure 1. The bar elements in these trusses are modelled by human tracheal cartilage with the elastic modulus $E = 25\,\text{MPa}$ [5], the cross-sectional $A = 1\,\text{cm}^2$. For these bar elements, the difference between the cross-section area in the current configuration, $A$, and that in the original configuration, $A_0$, is negligible. The highly nonlinear component comes from the spring attached at node 2. In particular, the spring represents for a bar made by aortic tissue, which is modelled using an isotropic Veronda-Westmann constitutive law [5][6], with the initial shear modulus $A = 0.5$ kPa. The aortic tissue bar is in uniaxial stretch and under incompressibility constraint, the first Piola-Kirchhoff stress along stretch direction can be expressed as

$$P(\lambda) = 2A\left(\lambda - \frac{1}{\lambda^2}\right)e^{B\left(\lambda^2 + \frac{2}{\lambda} - 3\right)} - A\left(1 - \frac{1}{\lambda^3}\right),$$

where the stretch $\lambda = l/L$, the initial length is $L = 25\,\text{cm}$.
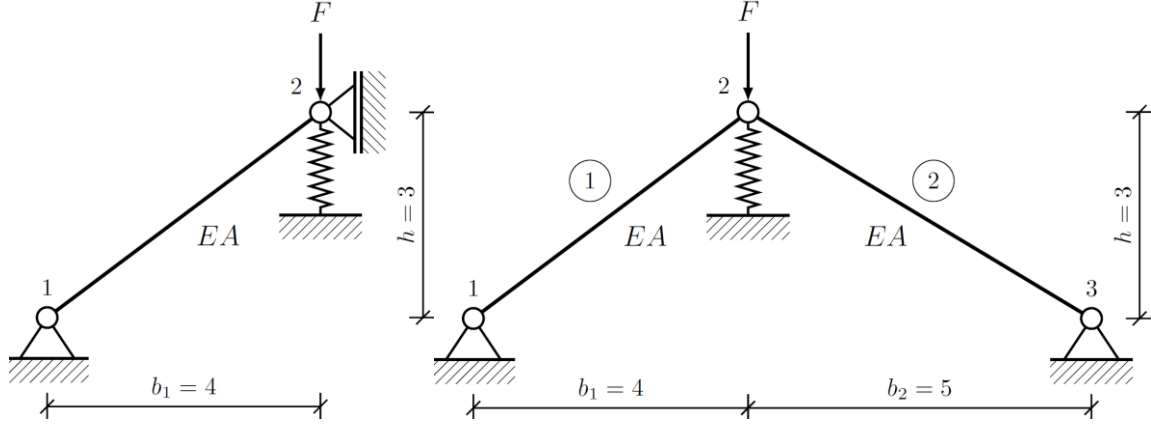
**Figure 1.** Truss systems: 1DOF (left) and 2DOF (right).

### 4.1 One degree of freedom problem

In case of 1DOF system as shown in Figure 1, applying the Crisfield truss-element method [1], we first get the reference configuration

$$\mathbf{X}^e = \begin{pmatrix} \mathbf{X}^{e1} \\ \mathbf{X}^{e2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 4 & 3 \end{pmatrix}^T,$$

and the element displacement vector is expressed as: $\boldsymbol{u}^e = \begin{pmatrix} u^{e1} \\ u^{e2} \end{pmatrix} = \boldsymbol{x}^e - \mathbf{X}^e = \begin{pmatrix} 0 & 0 & 0 & v \end{pmatrix}^T$. The

undeformed and deformed lengths are computed as

$$L^2 = \mathbf{X}^e \cdot \mathbf{A} \cdot \mathbf{X}^e = 4^2 + 3^2 = 25,$$

$$l^2 = \boldsymbol{x}^e \cdot \mathbf{A} \cdot \boldsymbol{x}^e = (4+0)^2 + (3+v+0)^2 = 25 + 6v + v^2,$$

in which $v$ is the vertical displacement at node 2. We proceed to compute the internal force vector

$$\boldsymbol{F}_{int}^e(\boldsymbol{u}^e) = \frac{EA}{2L^3}(l^2 - L^2) \begin{pmatrix} a \\ b \\ -a \\ -b \end{pmatrix}, \text{ where } \begin{cases} a = X_1^{e1} + u_1^{e1} - X_1^{e2} - u_1^{e2} = -4 \\ b = X_2^{e1} + u_2^{e1} - X_2^{e2} - u_2^{e2} = -(3+v) \end{cases}.$$

The internal force therefore is reduced to $F_{22}^e = (6v + v^2)(3+v) = 18v + 9v^2 + v^3$. The stiffness matrix comprises of two parts, geometric and material ones as follows

$$\mathbf{K}^e = \mathbf{K}_{geo}^e + \mathbf{K}_{mat}^e = \frac{EA}{L^3} \left( \frac{(l^2 - L^2)}{2} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} + \begin{pmatrix} a^2 & ab & -a^2 & -ab \\ . & b^2 & -ab & -b^2 \\ . & . & a^2 & ab \\ sym. & . & . & b^2 \end{pmatrix} \right),$$

then $K_{22}^e = K_{geo_{22}}^e + K_{mat_{22}}^e = 18 + 18v + 3v^2$. The current force acting on the aortic tissue bar is calculated as $F_s = P(\lambda)S_0$, where the initial cross-section area $S_0 = 1\,\text{cm}^2$.
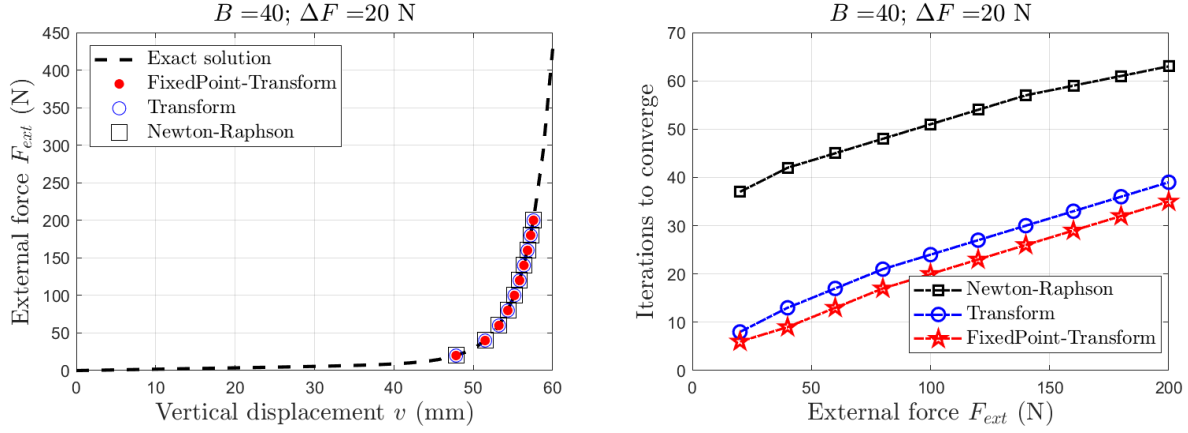
**Figure 2.** 1DOF ($B = 40, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
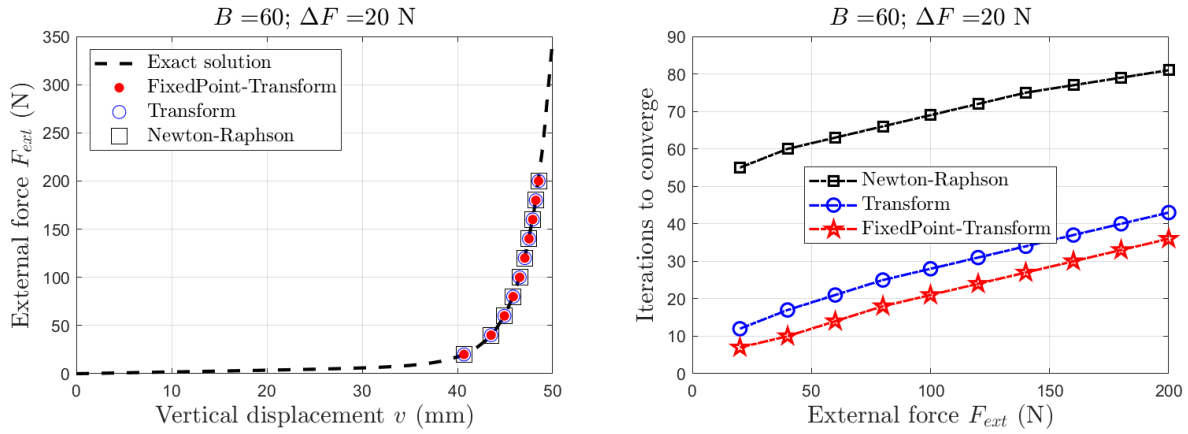


**Figure 3.** 1DOF ($B = 60, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
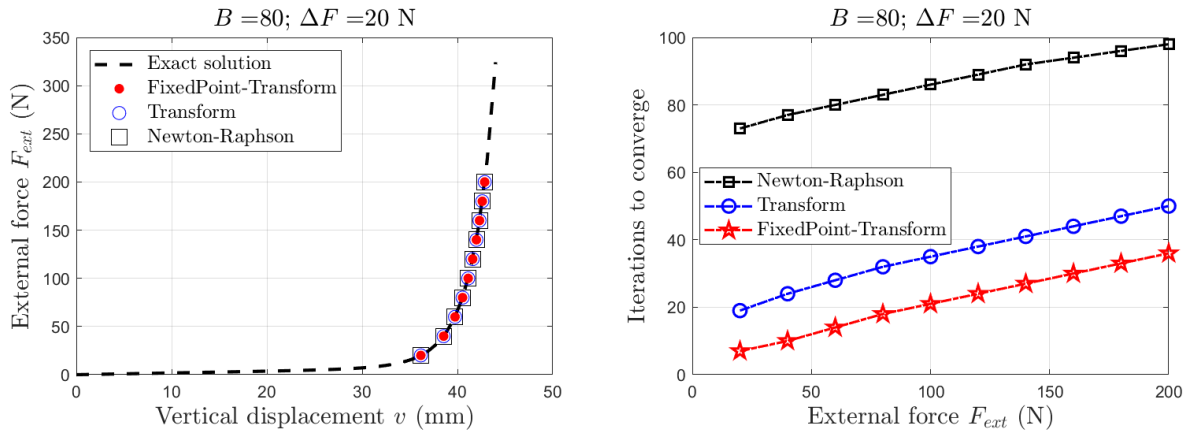


**Figure 4.** 1DOF ($B = 80, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
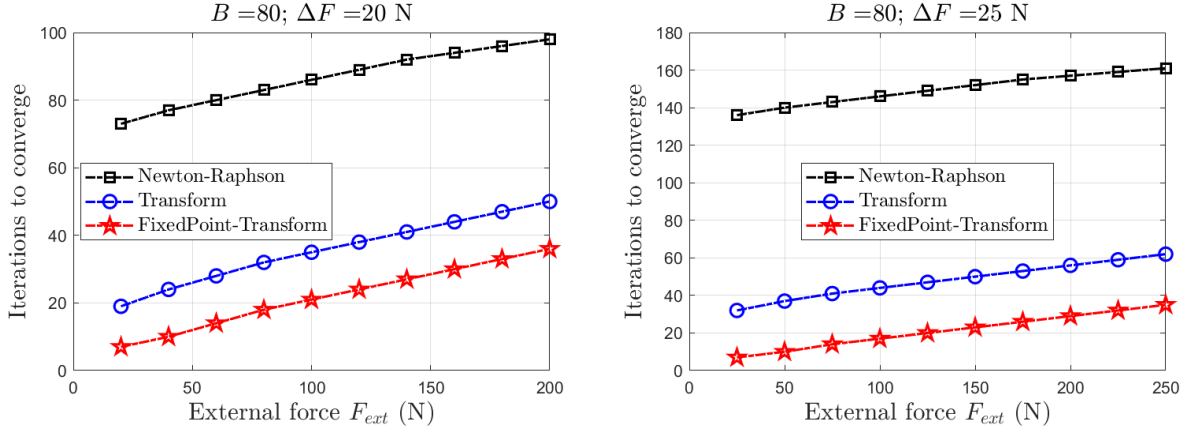
**Figure 5.** 1DOF ($B = 80$): No. of iterations when $\Delta F = 20 \, N$ (left) and $\Delta F = 25 \, N$ (right).
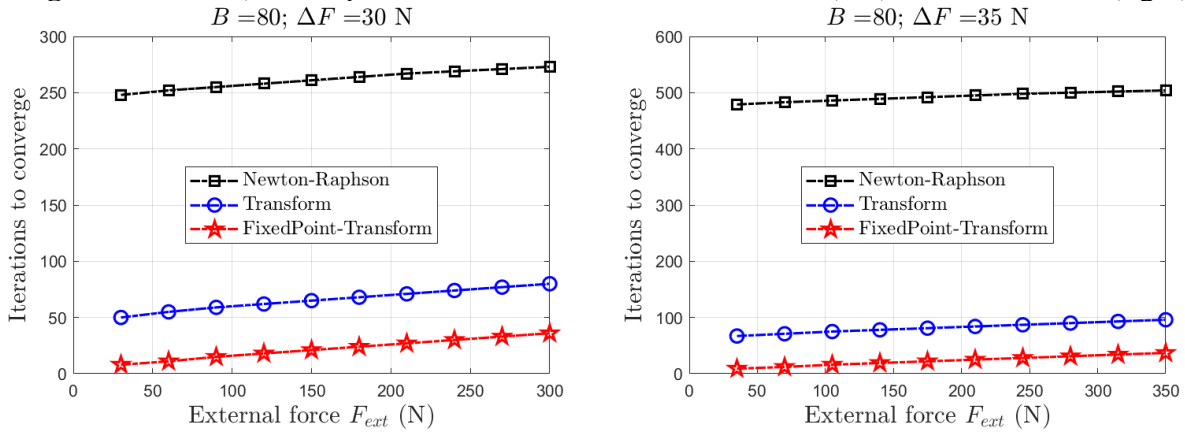


**Figure 6.** 1DOF ($B = 80$): No. of iterations when $\Delta F = 30 \, N$ (left) and $\Delta F = 35 \, N$ (right).

In order to study the performance of transformed Newton's enhanced by fixed-point iteration when applied to highly nonlinear problems, we gradually increase the parameter $B$, the numerical results are shown in Figure 2-Figure 6. As can be seen from these graphs, all three methods lead to numerical results perfectly matching with the exact solutions. The implementation of Fixed-point Transform method to our scalar nonlinear equation brings us incredible improvements. Notably, the improvement of decreasing the nonlinearity using Fixed-point Transform method is even greater than the Transform method. It can also be noted that when $B = 40$ the Newton-Raphson method needs a number of iterations that is significantly higher than for Transform and Fixed-point Transform. If we rise the degree of nonlinearity to B = 60 or B = 80, the numbers of iterations using the Newton's and Transform method increase accordingly, whereas the convergence of Fixed-point Transform remains nearly constant (around 35 or 36 iterations). On the other hand, another significant advantage of using the Fixed-point Transform method is that the number of iterations to converge is deeply smaller than for Newton's and better than Transform method when we increase the force increment $\Delta F$ from 20 to 35. In a nutshell, for the scalar nonlinear equation under consideration the enhancement of Transform method by Fixed-point iteration does not increase the computational cost in each force increment step.

### *4.2 Two degree of freedom problem*

Again, following the Crisfield truss-element method, we proceed the same procedure as used in subsection 4.1 to obtain the stiffness matrix

$$\mathbf{K}^e = \mathbf{K}^e_{geo} + \mathbf{K}^e_{mat} = \begin{bmatrix} 5.1u_1 + 4.89u_1^2 + 9.78u_2 + 1.63u_2^2 + 63.5 & 9.78u_1 + 3.26u_1u_2 + 1.7u_2 + 5.1 \\ 9.78u_1 + 3.26u_1u_2 + 1.7u_2 + 5.1 & 1.7u_1 + 1.63u_1^2 + 29.34u_2 + 4.89u_2^2 + 29.34 \end{bmatrix}.$$

Combining the internal forces inside of two bars and the soft tissue spring leads to the total internal force as follows

$$\mathbf{F}_{int} = \begin{bmatrix} 63.5u_1 + 2.55u_1^2 + 1.63u_1^3 + 9.78u_1u_2 + 1.63u_1u_2^2 + 5.1u_2 + 0.85u_2^2 \\ 5.1u_1 + 4.98u_1^2 + 1.7u_1u_2 + 1.63u_1^2u_2 + 29.34u_2 + 14.67u_2^2 + 1.63u_2^3 + F_S \end{bmatrix}.$$
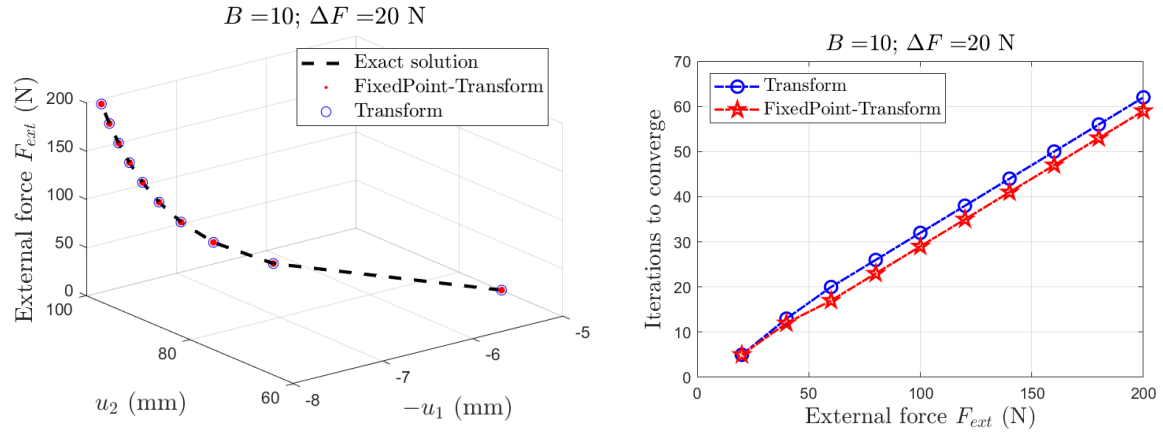


**Figure 7.** 2DOF ($B = 10, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
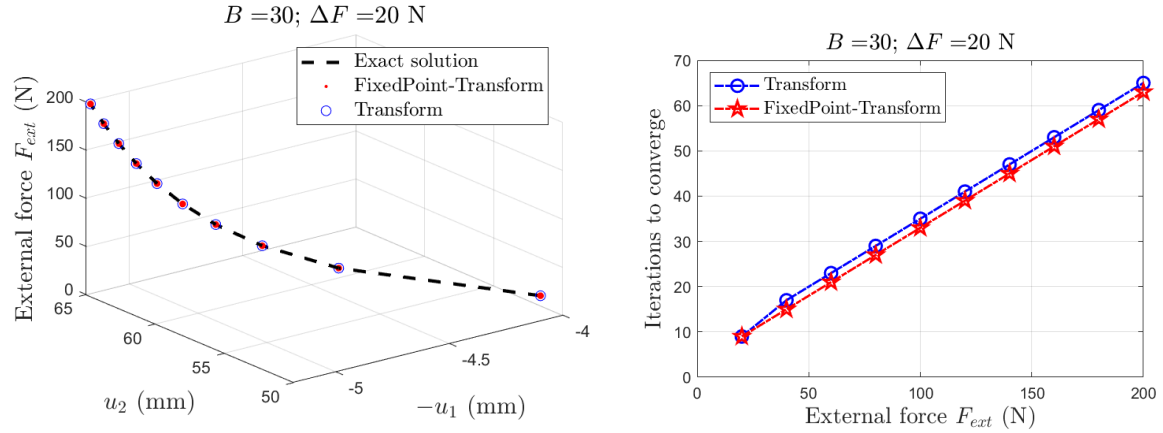


**Figure 8.** 2DOF ($B = 30, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
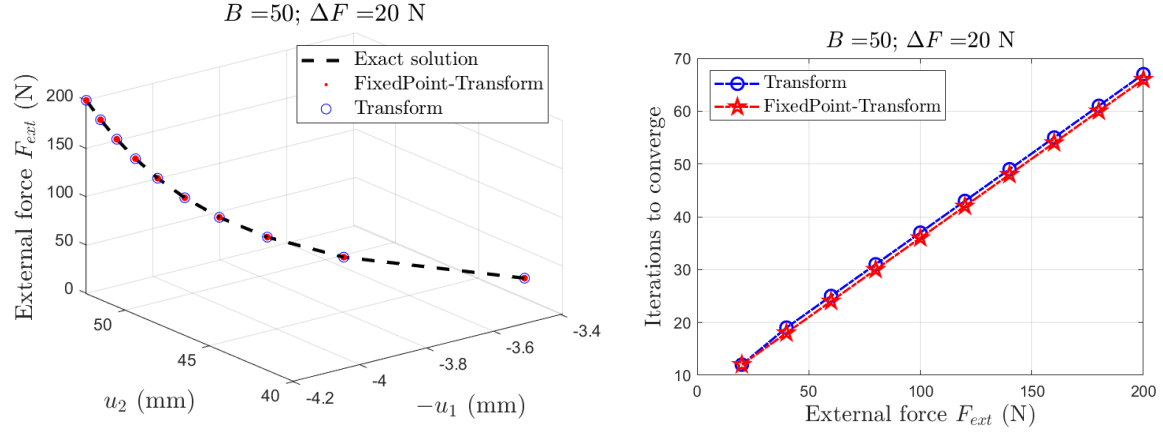
**Figure 9.** 2DOF ($B = 50, \Delta F = 20\ N$): Equilibrium path (left) and No. of iterations (right).
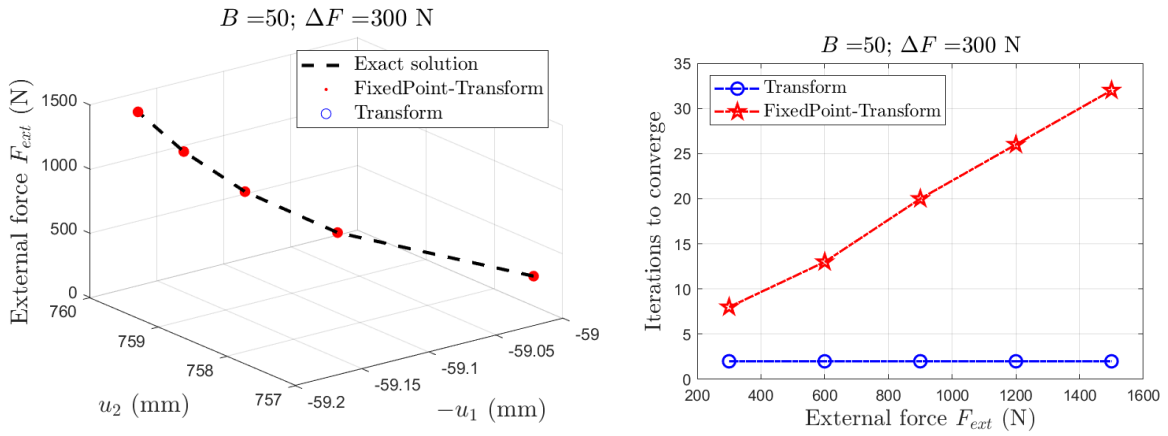


**Figure 10.** 2DOF ($B = 50, \Delta F = 300\ N$):
Equilibrium path (left) and No. of iterations (right).
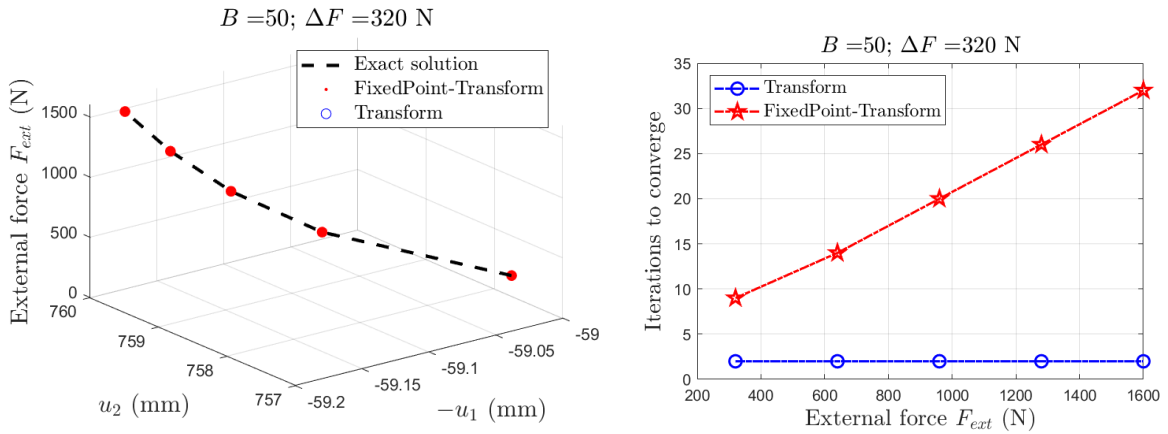


**Figure 11.** 2DOF ($B = 50, \Delta F = 320\ N$):
Equilibrium path (left) and No. of iterations (right).
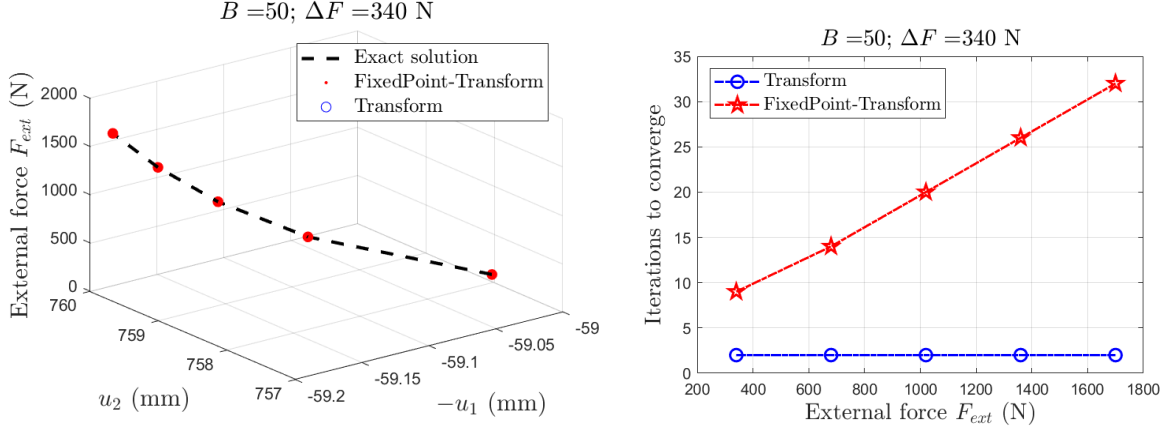
**Figure 12.** 2DOF ($B = 50, \Delta F = 340\ N$):
Equilibrium path (left) and No. of iterations (right).

Both Transform and Fixed-point Transform methods work perfectly as shown in Figure 7-Figure 9, the results fit the exact equilibrium paths, the number of iterations to converge of Fixed-point Transform method is lower than those of Transform method. The extension of Fixed-point Transform method to vector nonlinear equations shows that the convergence improved significantly when increasing force increment to a certain high value, e.g. $\Delta F = 300$ N, whereas Transform method diverges (noted that: the horizontal line describes the number of iterations corresponding to external force remain constant which means that it *diverges*). In other words, the Fixed-point Transform method allows us to increase the permissible load step size (Figure 10-Figure 12). One drawback of Fixed-point Transform method is that it slightly increases the computational cost of each iteration since it needs the second derivative. However, the development of newer methods based on symbolic computation and automatic differentiation, this limitation is becoming less important.

## 5    Conclusion

Nonlinear solvers based on the classical Newton's method to find roots of equations are at the heart of computational science. The combination of fixed-point iteration and the transform Newton's method presented in this work, which results in improving the quality of the initial guess and decreasing the nonlinearity, provides an efficient technique to deal with highly nonlinear problems. This approach is simple in implementation and can be easily integrated into any nonlinear finite element solvers. Hence, it has the potential of attracting the interest of the community of scientific computing to fully explore its capacities in solving highly nonlinear problems with various materials and structures.

# References

[1] Javier Bonet, Richard D. Wood (2008), *Nonlinear continuum mechanics for finite element analysis*, Second Edition, Cambridge University Press, New York.

[2] Ankush Aggarwal (2017), An improved parameter estimation and comparison for soft tissue constitutive models containing an exponential function, *Biomech Model Mechanobiol.*

[3] Yue Mei, Daniel E. Hurtado, Sanjay Pant, Ankush Aggarwal (2018), On improving the numerical convergence of highly nonlinear elasticity problems, *Computer Methods in Applied Mechanics and Engineering* **337**,110–127.

[4] Francois Dubeau, Calvin Gnang (2014), Fixed Point and Newton's Methods for Solving a Nonlinear Equation: From Linear to High-Order Convergence, *SIAM REVIEW Society for Industrial and Applied Mathematics*, Vol. 56, No. 4, 691–70.

[5] F. Safshekan, M. Tafazzoli-Shadpour, M. Abdouss, M.B. Shadmehr (2016), Mechanical characterization and constitutive modeling of human trachea: age and gender dependency*, Materials* 9 (6) 456.

[6] M.J. Girard, D. Tan, M. Ang, J.S. Mehta, L. Zhang, C.W. Chung, B. Mani, T.A. Tun, T. Aung (2015), An engineering-based methodology to characterize the in vivo nonlinear biomechanical properties of the cornea with application to glaucoma subjects, *Invest. Ophthalmol.* Vis. Sci. 56 (7), 1099–1099.

# Comparative study to analyze deformation/fracture behavior under high-velocity impact using different dynamic transient operators

**Yogeshwar Jasra[1], Pardeep Kumar[2], Nikesh K Ojha[3] and \*†Ravinder K. Saxena[4]**

[1,3,4]Department of Mechanical Engineering, Sant Longowal Institute of Engineering and Technology, India.
[2]R V Industries, India

\*Presenting author: rks@sliet.ac.in
†Corresponding author: rks@sliet.ac.in

## Abstract

Impact and dynamic fracture behavior studies in the structures always hold importance in many areas of science and everyday phenomenon. A process of fracture phenomenon and mechanics of blunt-shaped projectile moving at a high impact velocity will result in fragmentation due to dynamic stress loading also known as the classical Taylor rod impact problem. The present investigation demonstrates the deformation behavior of the flat-faced Taylor rod using different "Direct Integration Schemes" viz. Single/Multi step Houbolt, Newmark beta, Generalized Alpha, etc. under implicit transient dynamic operator and central difference under explicit operator. The finite element analysis is performed using MSC Marc Mentat™. It is further added the plasticity model is incorporated as a Fortran routine in MSC Marc. It is found that instability in the algorithm and irregular deformation can occur using the Newmark Beta scheme. It is also found that the results can be overpredicted using a multi-step Houbolt operator due to high numerical dissipation and the Generalized-alpha method has been presented as an unconditionally stable that allows controllable numerical dissipation and the deformation consistent with experimental results.

**Keywords:** Computational Mechanics, Finite Element Analysis, Dynamic transient operators, Taylor Rod, Newmark Beta, Houbolt, Generalized Alpha.

## Introduction

In the Taylor Rod impact problem, a flat deformable projectile is fired against a fixed rigid target. The experiment is generally used to determine the dynamic yield stresses in the material. It is also performed to validate the constitutive model by comparing the numerical model with the experiments. Taylor [1] was the first researcher who analytically determined the dynamic yield stress by firing the blunt-nosed deformable projectile against the rigid wall. In the study, the mechanical response was determined for the problem involving large plastic deformations, high strain rate, and elevated temperatures. The mathematical model is generally formulated using equations of motion and these equations are solved with the finite element technique (FE) to obtain the solution. The FE formulation of the equation of motion for a dynamic system will result in the ordinary differential equations. A few numerical procedures are available in the literature to solve such coupled differential equations [2]-[7]. The solution is obtained by discretization of the time variable in time intervals. The various functions are assumed to predict the variation in displacement, velocity, and acceleration. The dynamic equilibrium is obtained in the discrete-time increment. So, the numerical methods generally used for solving governing equations of a dynamic system are known as direct integration numerical procedures. Direct integration can be performed in two different ways (Explicit and Implicit). In the explicit integration scheme, the current configuration (t) of the body is known and used to predict the deformed configuration at a time ($t + \triangle t$), In the

implicit integration scheme, the current configuration and the deformed configuration or the dynamic states at both time intervals t and $t + \triangle t$ are required to obtain the solution. In the finite element simulations, the dynamic state at deformed configuration is obtained by the iterative scheme. This procedure is for both linear and nonlinear problems and includes geometric, material, and boundary nonlinearities. The stability and accuracy of the direct integration schemes are determined by the quality of the solutions like in a conditionally stable method the error remains within the prescribed tolerance. The direct integration schemes are in general imprecise. Each integration scheme has various problems like the high-frequency numerical dissipation improves the overall convergence but an uncontrollable numerical dissipation in high-frequency modes can damp out the response in low-frequency mode, but it overall reduces the quality by incorporating the excessive artificial numerical damping in the system.

Woodward et al. [8] performed the two-dimensional axisymmetric finite element simulation and predicted the failure mechanics using an explicit time integration scheme. Worwick et al. [9] analyzed the initiation and propagation of voids in brass. The 2D axis-symmetric analysis was performed using an explicit time integration scheme. Addessio et al. [10] performed the Taylor rod impact simulation using an explicit finite element solver and analyzed the evolution of confined fracture during an impact test. Celentano et al. [11] analyzed the coupled thermomechanical analysis of the Taylor rod. The integration of the time derivative terms was performed using Hilber- Hughes-Taylor (HHT) method. The parameters were chosen using the midpoint rule algorithm. The results were compared with the experimental literature available. Teng et al. [12] determined the fracture patterns in the Taylor rod at the velocity of 600 m/sec. An explicit integration scheme was used to determine the dynamic behavior and fracture pattern. Bao and Wierzbicki [13] analyzed the failure of the target material impacted by a blunt-shaped projectile. The dynamic equations are solved using an explicit time integration scheme. The analysis also suggested that the damage increment is computed only when the triaxiality is greater than -1/3 or, the material is under the tensile loading. This phenomenon has also been introduced in the present study in both implicit and explicit integration schemes. Gautam et al. [14]-[16] predicted the ductile fracture in the Taylor rod impact problem. In the study, Newmark's algorithm is implemented for the implicit integration scheme and algorithmic damping was introduced so to maintain the stability of the solution and to improve the overall response of the system. Rathore et al. [17] analyzed the impact phenomenon by comparing the effect of contact constraints on the overall failure of the target body. In the study, two contact constraints were compared. The dynamic equations were solved using the Newmark integration scheme. Xiao et al. [18] determined the effect of lode parameters in the fracture pattern of the Aluminum alloy Taylor rods. In the analysis, an explicit time integration scheme was implemented to determine the effect of lode angle in the Taylor rod.

It is known that the deformation in the Taylor rod will happen if it impacts the rigid surface at a high velocity. The deformation behavior will depend upon the formulated mathematical model as well as the solution techniques used to solve the model. It will also depend upon the assumed displacement field and the choice of the direct integration schemes in the simulation. Studies are available in the literature which analyze the deformation and fracture behavior of the Taylor rod. Most of the studies used explicit time integration schemes to solve the response of the dynamic system [10] [12] 13] [18]. A very few studies were found in the literature which used the implicit time integration schemes [11] [14-16]. Although, the number of schemes is available and used by the researchers to predict the response of the dynamic system, the effect of different numerical integration schemes on the deformation,

mushrooming and stress distribution in the Taylor rod test appears to be lacking in the literature. So, the objective of the present study is to formulate the finite element model and analyze the deformation in the Taylor rod impact problem using different dynamic transient operators. The investigation will demonstrate the deformation behavior of the flat-faced Taylor rod using different direct integration schemes (Single/Multi step Houbolt, Newmark beta, Generalized Alpha) under an implicit transient dynamic operator. The effect of the Explicit time integration schemes on the deformation behavior has also been analyzed in the present study.

**Finite element modeling**

*Computational model*

In the classical Taylor rod impact problem, the flat-nosed projectile is impacted onto a rigid surface. The cylindrical projectile geometric dimensions are presented in Fig. 1. The material parameters are presented in Table 1. The friction coefficient is assumed to be 0.1. Different implicit and explicit direct integration schemes are used to simulate the deformations and fracture pattern of the Taylor rod. The contact between two surfaces is implemented using node to contact segment [17]. The iterative penetration algorithm available MSC Marc Mentat™ [20] is applied to avoid penetration during the impact process. The one-fourth finite element model is simulated using eight noded, brick elements with reduced integration due to symmetry.
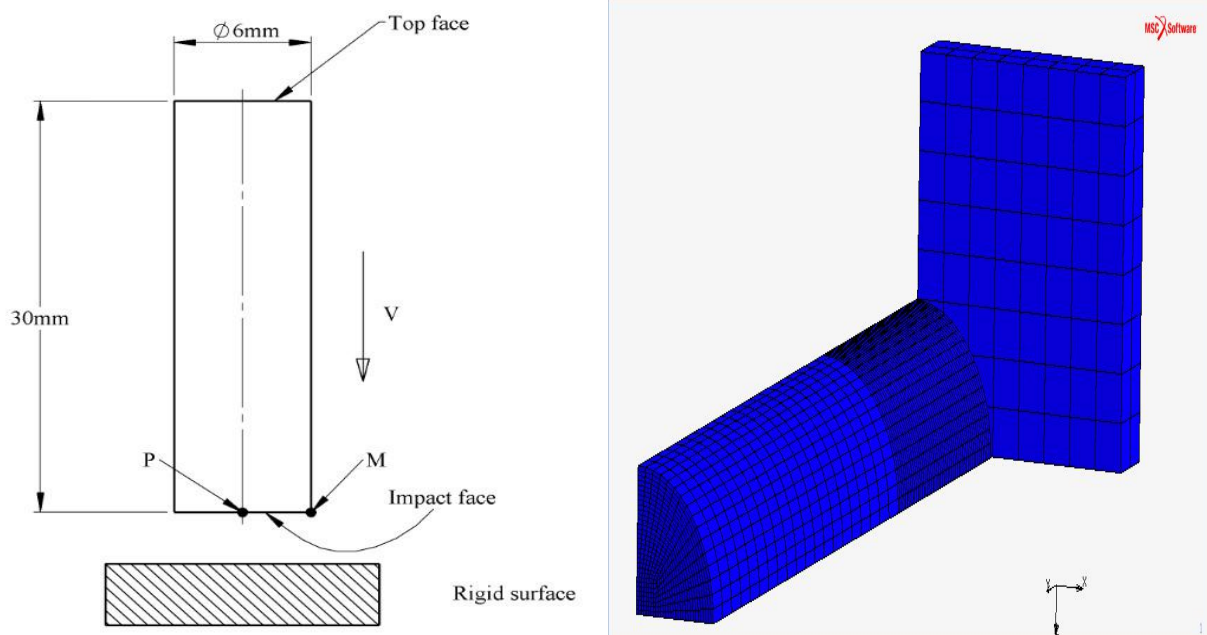


Fig.1 (a). Schematic of Taylor rod impact problem. (b). Finite Element Model of the Taylor rod impact problem.

A fine mesh is modeled in the front part as it involves high deformation and non-linear behavior. It is also expected that the failure will initiate from the front portion of the Taylor rod, therefore, number of elements in the front portion is more as compared to the rear portion of the Taylor rod.

The projectile velocity is taken as 400 m/sec in the simulation to capture the fracture deformation of the Taylor rod. The target is assumed to be much stiffer as compared to the

| Property | Notations | Value |
|---|---|---|
| Young's modulus | $E$ (N/m$^2$) | $203 \times 10^9$ |
| Poisson's ratio | $\upsilon$ | 0.33 |
| Density | $\rho$ | 7850 |
| Initial yield | $A$ (N/m$^2$) | $304.330 \times 10^6$ |
| Strain hardening constant | $B$ (N/m$^2$) | $422.007 \times 10^6$ |
| Strain hardening exponent | $n$ | 0.345 |
| Viscous effect | $C$ | 0.0156 |
| Thermal softening constant | $m$ | 0.87 |
| Reference strain rate | $\varepsilon_0$ | $0.0001 \text{s}^{-1}$ |
| Melting temperature | $\theta_{melt}$ (K) | 1800 |
| Transition temperature | $\theta_{transition}$ (K) | 293 |

Table 1. Material parameters for the mild steel [19]

projectile The projectile rebounds back from the rigid surface after the completion of the impact process.

*Material Constitutive Model*

The Johnson-Cook (JC) material model is implemented in the MSC Marc Mentat™ in all the direct integration schemes. In the model, the dynamic equivalent stress depends upon the equivalent plastic strain, equivalent plastic strain rate, and temperature. The material is assumed to be elastic up to the initial yield of the material then behaves according to the following relation of the JC plasticity model [21].

$$\sigma = \left[ A + B \left( \varepsilon_{eq}^{p} \right)^n \right] \left[ 1 + C \ln \left( \frac{\dot{\varepsilon}}{\dot{\varepsilon}_0} \right) \right] \left[ 1 - \left( T^* \right)^m \right] \tag{1}$$

where $\sigma$ is the equivalent stress, $\varepsilon_{eq}^{p}$ is the equivalent plastic strain, A, B, n, C and m are the constants, $\varepsilon_o$ is the reference strain rate and $T^*$ is expressed as

$$T^* = \frac{T - T_r}{T_m - T_r}$$

in which $T_r$ and $T_m$ are reference temperature and melting point temperatures. The first bracket represents the quasi-static stress-strain relationship at room temperature, the second term represents the strain-rate hardening, the third term represents the effect of temperature on the plastic behavior of the material. The JC plasticity parameters for the mild steel are listed in Table 1.

*Formulation of Direct Integration Schemes*

The dynamic equation of motion for the structural analysis is

$$Ma + Cv + Ku = F \tag{2}$$

where M represents the mass matrix, C is the damping matrix and K is the stiffness matrix. The acceleration is represented by "a", "v" is the velocity, "u" is the displacement and F is the external force vector. There are different integration schemes available in the literature to integrate the equation of motions and to obtain the dynamic response of the system.

**Newmark-beta operator**

The general form of Newmark beta is represented as [20]

$$u^{n+1} = u^n + \Delta t v^n + \left(\frac{1}{2} - \beta\right)\Delta t^2 a^n + \beta\Delta t^2 a^{n+1} \tag{3}$$

The superscript n represents the nth finite time increment. u is the displacement; v is velocity and a is acceleration.

$$v^{n+1} = v^n + (1 - Y)\Delta t a^n + Y\Delta t a^{n+1} \tag{4}$$

If $\gamma = \frac{1}{2}$, $\beta = \frac{1}{4}$

The equations become

$$\left(\frac{4}{\Delta t^2}M + \frac{2}{\Delta t}C + K\right)\Delta u = F^{n+1} - R^n + M\left(a^n + \frac{4}{\Delta t}v^n\right) + Cv^n \tag{5}$$

where R is the internal force.

$$R = \int_v \beta^T \sigma dv$$

After solving the equations implicitly, the solution of the dynamic system becomes
$$u^{n+1} = u^n + \Delta u \tag{6}$$

**Houbolt operator**
The velocity and acceleration functions of the Houbolt operator are detailed. The Houbolt operator [20] uses the last three values and by cubic fitting obtain the following expression:

For velocity
$$v^{n+1} = \left(\frac{11}{6}u^{n+1} - 3u^n + \frac{3}{2}u^{n-1} - \frac{1}{3}u^{n-2}\right)/\Delta t \tag{7}$$

For acceleration
$$a^{n+1} = (2u^{n+1} - 5u^n + 4u^{n-1} - u^{n-2})/\Delta t^2 \tag{8}$$

By substituting this expression in the equations of motion

$$\left(\frac{2}{\Delta t^2}M + \frac{11}{6\Delta t}C + K\right)\Delta u = F^{n+1} - R^n + \frac{1}{\Delta t^2}(3u^n - 4u^{n-1} + u^{n-2})M + \frac{1}{\Delta t}\left(\frac{7}{6}u^n - \frac{3}{2}u^{n-1} + \frac{1}{3}u^{n-2}\right)C \tag{9}$$

By solving the equation implicitly and $\Delta u$ is obtained by using the value from the previous two increments.

**Single-Step Houbolt operator**

In comparison to the standard Houbolt operator, the single-step Houbolt operator contain few terms related to the beginning of the increment. [20]

$$\alpha^{m1}Ma^{n+1} + \alpha^{c1}Cv^{n+1} + \alpha^{k1}Ku^{n+1} + \alpha^{m}Ma^{n} + \alpha^{c}Cv^{n} + \alpha^{k}Ku^{n} = \alpha^{f1}F^{n+1} + a^{f}F^{n}$$

$$u^{n+1} = u^{n} + \Delta tv^{n} + \beta\Delta t^{2}a^{n} + \beta^{1}\Delta t^{2}a^{n+1}$$

$$v^{n+1} = v^{n} + \Upsilon\Delta ta^{n} + \Upsilon^{1}\Delta ta^{n+1} \tag{10}$$

To solve the equations few assumptions were made to reduce all the unknown parameters into two.

$$\alpha^{k} = 0, \beta = \gamma, \beta^{'} = \gamma + \gamma^{'}, \alpha^{m} = -1/2, \alpha^{k1} = 1/2\beta^{l}$$

$$\alpha^{c} = -(2\beta + \beta')\big/4(\beta')^{2}, \alpha^{c1} = (2\beta + 3\beta')\big/4(\beta^{1})^{2}, \alpha^{f} = \alpha^{k}, \alpha^{f1} = \alpha^{k1} \tag{11}$$

Later based on the Taylor series expansion of the displacement function.

$$\beta + \beta' = 1/2$$

$$\gamma = 1/2(1/2 - \gamma')$$

By substituting the velocity and acceleration functions into the equilibrium equations will result in the following simplified form

$$\left\{\frac{1}{\beta^{1}\Delta t^{2}\alpha^{k1}}M + \frac{\alpha^{c1}\Upsilon^{1}}{\beta^{1}\Delta t\alpha^{k1}}C + K\right\}\Delta u = F^{n+1} - Ku^{n} + \frac{1}{\beta^{1}\Delta t^{2}\alpha^{k1}}M\{\Delta tv^{n} + \beta\Delta t^{2}a^{n}\} - \frac{\alpha^{m}}{\alpha^{k1}}Ma^{n} -$$

$$\frac{\alpha_{c1}}{\alpha^{k1}}C\left\{v^{n} + \Upsilon\Delta ta^{n} - \frac{\Upsilon^{1}}{\beta^{1}\Delta t}\{\Delta tv_{n} + \beta\Delta t^{2}a_{n}\}\right\} - \frac{\alpha^{c}}{\alpha^{k1}}Cv^{n} \tag{12}$$

After solving the equations implicitly, the solution of the dynamic system becomes
$$u^{n+1} = u^{n} + \Delta u$$

## Generalized Alpha Operator

From the equations of motion, the equilibrium equations for the generalized alpha can be expressed in the form [20]

$$M_{a}^{n+1+\alpha_{m}} + C_{v}^{n+1+\alpha_{f}} + K_{u}^{n+1+\alpha_{f}} = F^{n+1+\alpha_{f}}$$

where

$$u^{n+1+\alpha_{f}} = (1 + \alpha_{f})u^{n+1} - \alpha_{f}u^{n}$$

$$v^{n+1+\alpha_{f}} = (1 + \alpha_{f})v^{n+1} - \alpha_{f}v^{n}$$

$$a^{n+1+\alpha_m} = \left(1 + \alpha_m\right)a^{n+1} - \alpha_m a^n \tag{13}$$

The displacement and velocity functions are similar to the Newmark's functions:

$$u^{n+1} = u^n + \Delta t v^n + \left(\frac{1}{2} - \beta\right)\Delta t^2 a^n + \beta\Delta t^2 a^{n+1}$$
$$v^{n+1} = v^n + (1 - \Upsilon)\Delta t a^n + \Upsilon\Delta t a^{n+1} \tag{14}$$

As shown in the literature, the optimized values for the parameters $\beta$ and $\gamma$ are related by the following expressions.

$$\beta = \frac{1}{4}(1 + \alpha_m - \alpha_f)^2$$

$$\Upsilon = \frac{1}{2} + \alpha_m - \alpha_f \tag{15}$$

If the values for the $\alpha_f$ and $\alpha_m$ are set to zero, then the Generalized alpha equations become equivalent to the Newmark-beta scheme. If the values are varied from $\alpha_m = 0$ and $-0.33 \leq \alpha_f \leq .0$ then the equations become equivalent to HHT method and if the values for the parameters varied from $\alpha_f = 0$ and $0 \leq \alpha_m \leq .1$, it gives rise to the WBZ method. The values of $\alpha_f$ and $\alpha_m$ are related to the behavior of the dynamic system. These parameters are also used to restrict numerical dissipation. Numerical dissipation is related inversely to the spectral radius. The values of $\alpha_f$ and $\alpha_m$ in the MSC Marc Mentat™ FE package can be varied in terms of spectral radius. The relations between the spectral radius and $\alpha_f$ and $\alpha_m$ are presented as:

$$\alpha_f = -\frac{S}{1+S}$$

$$\alpha_m = \frac{1-2S}{1+S} \tag{16}$$

## Central Difference Operator

In the central difference operator, the displacement is assumed to vary quadratically over the finite time interval. The following equations are used to describe the variation of displacement over time.

$$a^n = (v^{n+\frac{1}{2}} - v^{n-\frac{1}{2}})/(\Delta t)$$

$$v^n = (u^{n+\frac{1}{2}} - u^{n-\frac{1}{2}})/(\Delta t)$$

$$a^n = (\Delta u^{n+1} - \Delta u^n)/(\Delta t^2) \tag{17}$$

where the incremental displacement is,

$$\Delta u^n = u^n - u^{n-1}$$

$$\frac{M}{\Delta t^2}\Delta u^{n+1} = F^n - R^n + \frac{M}{\Delta t^2}\Delta u^n \tag{18}$$

In the explicit central difference operator, the inverse operation of the matrix is not involved in the explicit analysis.

## Results and Discussions

The 3D, high-velocity large deformation analysis of the Taylor test is performed. The constant material properties are assumed in the analysis. The finite element formulation is validated with experimental and numerical results of Celentano et al [11]. It is observed that after impact, the stress wave travel from the outer edge of the projectile towards the central axis of the projectile. Initial elastic wave followed by the plastic wave evolved during the impact in the initial stages using all the direct integration schemes. Initially, i.e., before 2 μs of the impact, mushroom-like deformation of the projectile is observed in the cylinder. Later, the material at the center of the impacting cylinder rebounds in-between 2 to 5μs while the outer edge of the cylinder remains in contact with the target in all the different integration schemes. The deformation process using different integration schemes has been presented in Fig. 3 at different time intervals. This observation of deformation is consistent with Celentano [11]. The stress at the outer surface grows at a higher rate as compared to the central axis of the cylinder due to the large deformation at the outer edge. The von Mises stress contours at different time intervals using different integration schemes have been presented in Figs. 3, 4, and 5. Initially, the standard values of the direct integration schemes have been chosen while analyzing the effect of different integration schemes on the deformation process. The standard values have been taken from the literature. Then these constant values have been varied within well-defined ranges reported in the literature. The overall effect of the variation of the constant parameters on the deformation process of the Taylor rod has also been presented in the following sections. Six different types of dynamic operators (Explicit, Newmark-Beta operator, Wood-Bossak-Zienkiewicz (WBZ)-α modification, Generalized Alpha, multistep Houbolt operator, and the single-step Houbolt operator) are used to analyze the effect of different dynamic operators on the deformation process of the Taylor rod impact problem. Different constant values for these dynamic operators are used in the present work as recommended in the literature. The different values used are presented in Table 2. There are no parameters in the Explicit and Multistep Houbolt operators. The response of the solution depends upon the time step in the case of the multistep Houbolt operator and independent in the Explicit operator.

*Validations*

The impact test of the flat-ended cylindrical projectile against the rigid wall is performed and validated with the literature available [11]. Figure 1 represents the initial geometry of the projectile. The material properties are taken from the literature [11] for validation. Due to

|   | Dynamic Operator | Constant Parameter 1 | Constant Parameter2 |
|---|---|---|---|
| 1 | Newmark-Beta | $\beta = 0.5$ | $\gamma = 0.9142$ |
| 2 | WBZ | $\alpha_f = 0$ | $0 \leq \alpha_m \leq .1$ $\alpha_m = 0.5$ |
| 3 | Genralized aplha | Spectral Radius 0, $\alpha_f = 0$ | $0 \leq \alpha_m \leq .1$, $\alpha_m = 1$ |
| 4 | Multistep Houbolt | Time step 5e-9 | nil |
| 5 | Single Step Houbolt | $\gamma' = 1.5$ | $\gamma = -0.5$ |
| 6 | Explicit | Time step 4e-9 | nil |

Table 2 The constant parameters in different dynamic transient operators
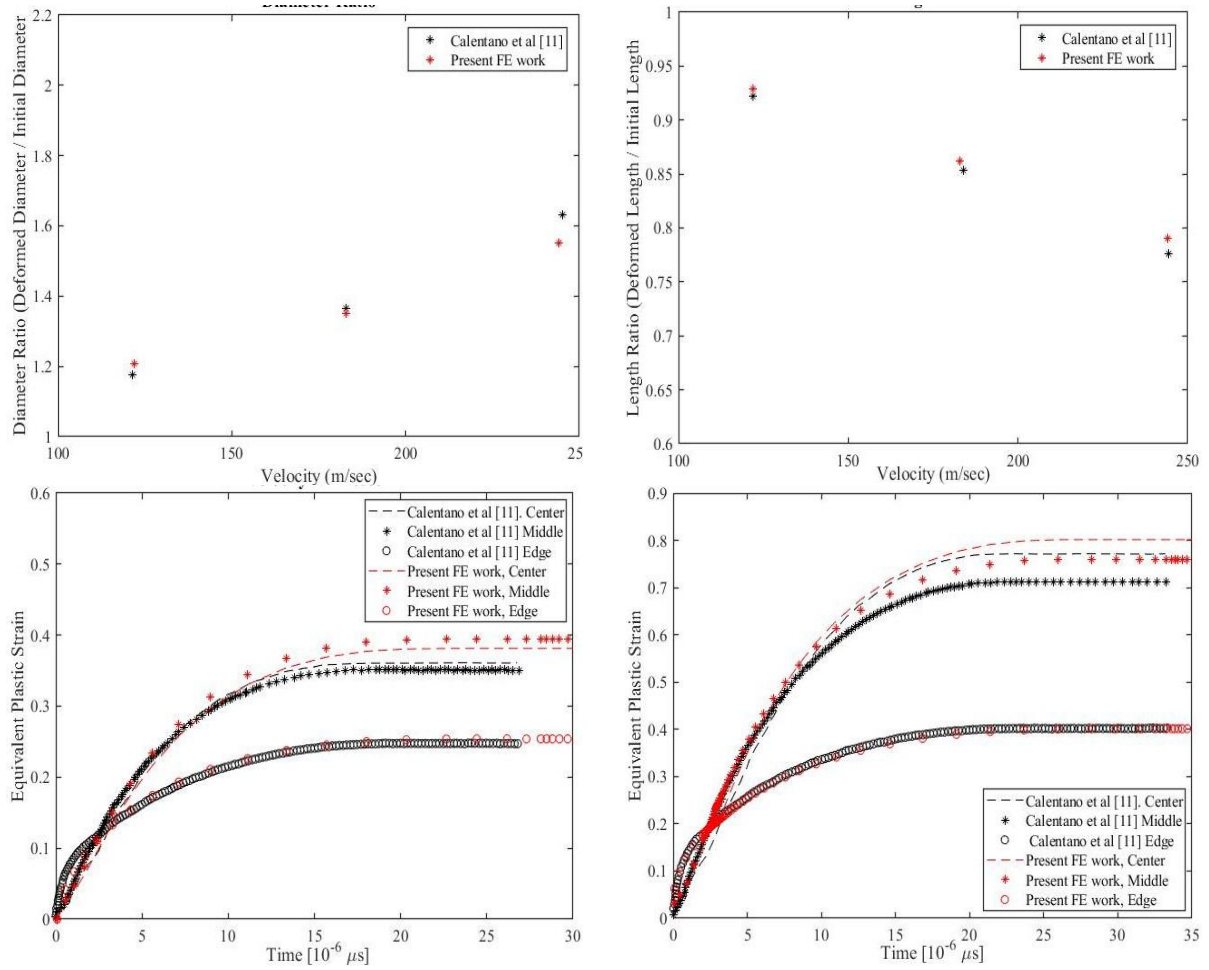
Figure 2. The variation of different characteristics with velocity.

symmetry of the problem, only one-fourth model is considered for the analysis. The heat transfer to the surrounding environment is assumed to be negligible because of the short duration of impact. The predicted result in Fig. 2 is consistent with the literature [11].

*Deformation pattern in Taylor rod*

The von-Mises stress distribution at three different time intervals has been presented in the section. The first-time interval represents the stress distribution during the mushrooming process just after the initiation of the impact process and at the starting time of the lifting of the central axis reported in the study [16]. The second time step is chosen when the central axis again comes in contact with the target i.e., around 5 to 7 μs. The overall effect of the different integration schemes on the jump of the central point has also been presented in the present study.

(a) Explicit Dynamic       (b) Newmark-Beta.       (c) Generalized alpha

(d) Multi-step Houbolt       (e) Single step Houbolt
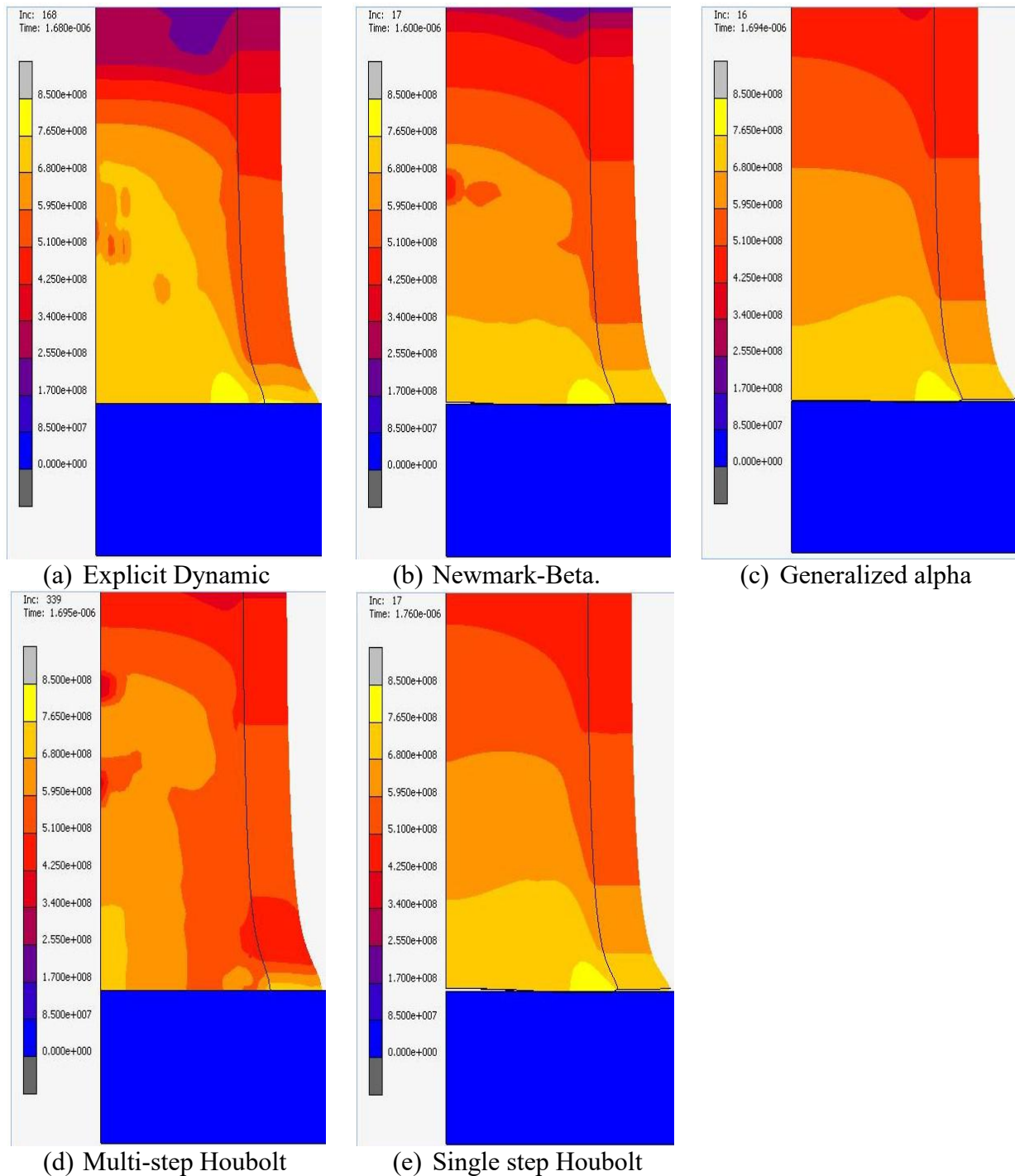
Fig 3. The von-mises stress contours at the initial time step of 1.6 μs.

The third time step is chosen at the end of the analysis where the analysis failed to converge to the required tolerance.

**At the initial time interval (The starting time of the jump at the central axis)**

In Fig 3. (a), the stress contours are obtained by solving the dynamic equations explicitly. The constant time steps are used in the explicit dynamic operator as there is no need for an adaptive time step during the analysis. The initial time step is 1.6 μs when the central point starts to rebound. In the explicit dynamic operator, the central point of the projectile does not rebound back as compared to other implicit schemes in which the central point starts lifting at a time

(a) Explicit Dynamic     (b) Newmark-Beta.     (c) Generalized alpha

(d) Multi-step Houbolt     (e) Single step Houbolt
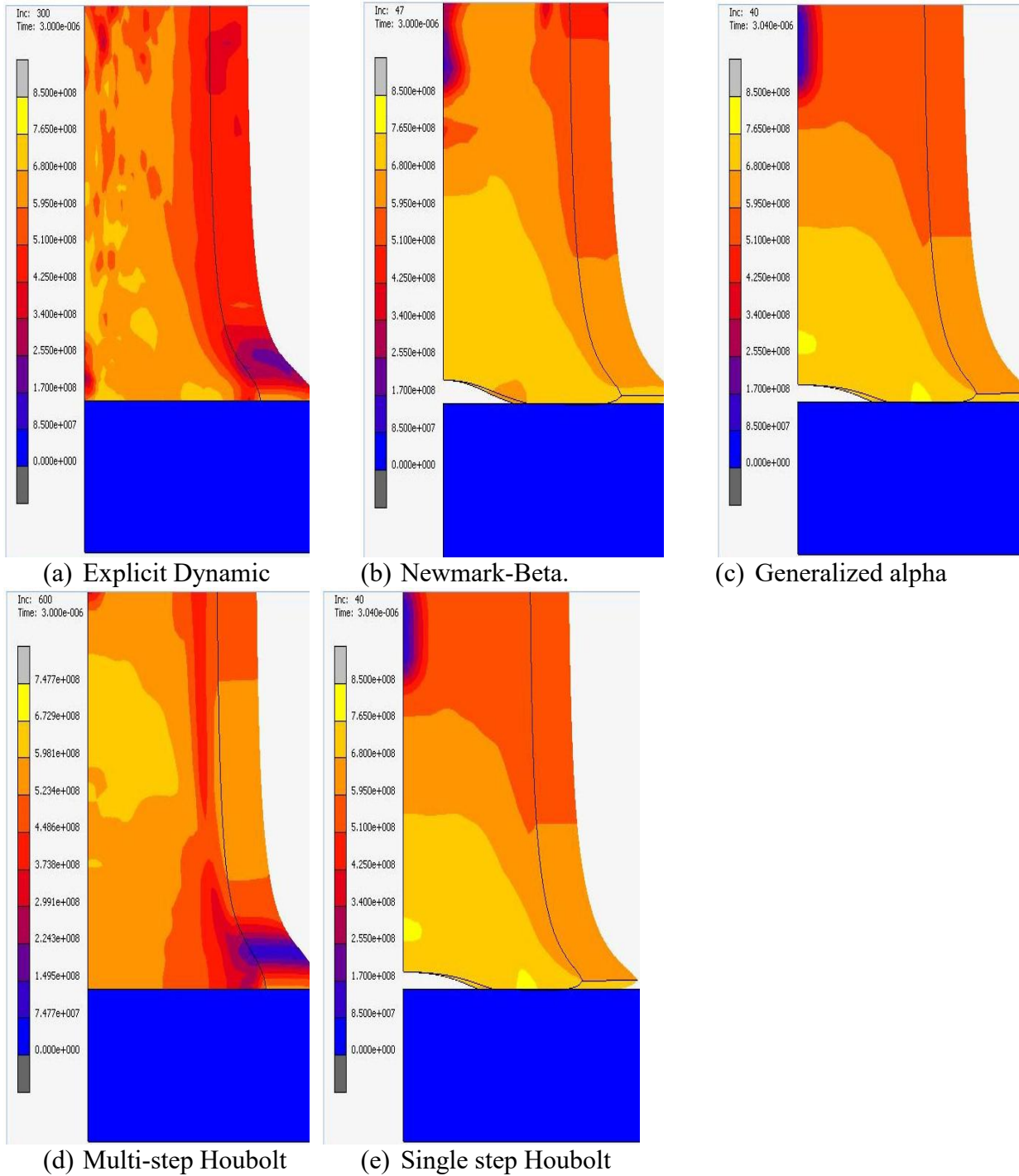
Fig 4. The von-mises stress contours at the initial time step of 3.0 μs.

i.e., around 1.68 μs. The stress in the case of explicit dynamics is around 827 MPa. In Fig. 3(b), the solution is achieved by solving the Newmark-beta integration scheme. The stress distribution generated during the analysis is higher as compared to other solutions at the outer periphery of the rod. It happens because the outer edge bulges out after the initial stages of the impact. The lifting of the central point of the rod initiates at 1.60 μs. The Newmark dynamic operator at this initial stage is unconditionally stable with no numerical damping. In Fig 3(c), the stress distribution is obtained by solving equations using the second-order algorithm generalized alpha method. The maximum stress obtained by generalized alpha is around 850 MPa, which is consistent with the previous two distributions. The jump occurs at the

(a) Explicit Dynamic    (b) Newmark-Beta.    (c) Generalized alpha



(d) Multi-step Houbolt    (e) Single step Houbolt
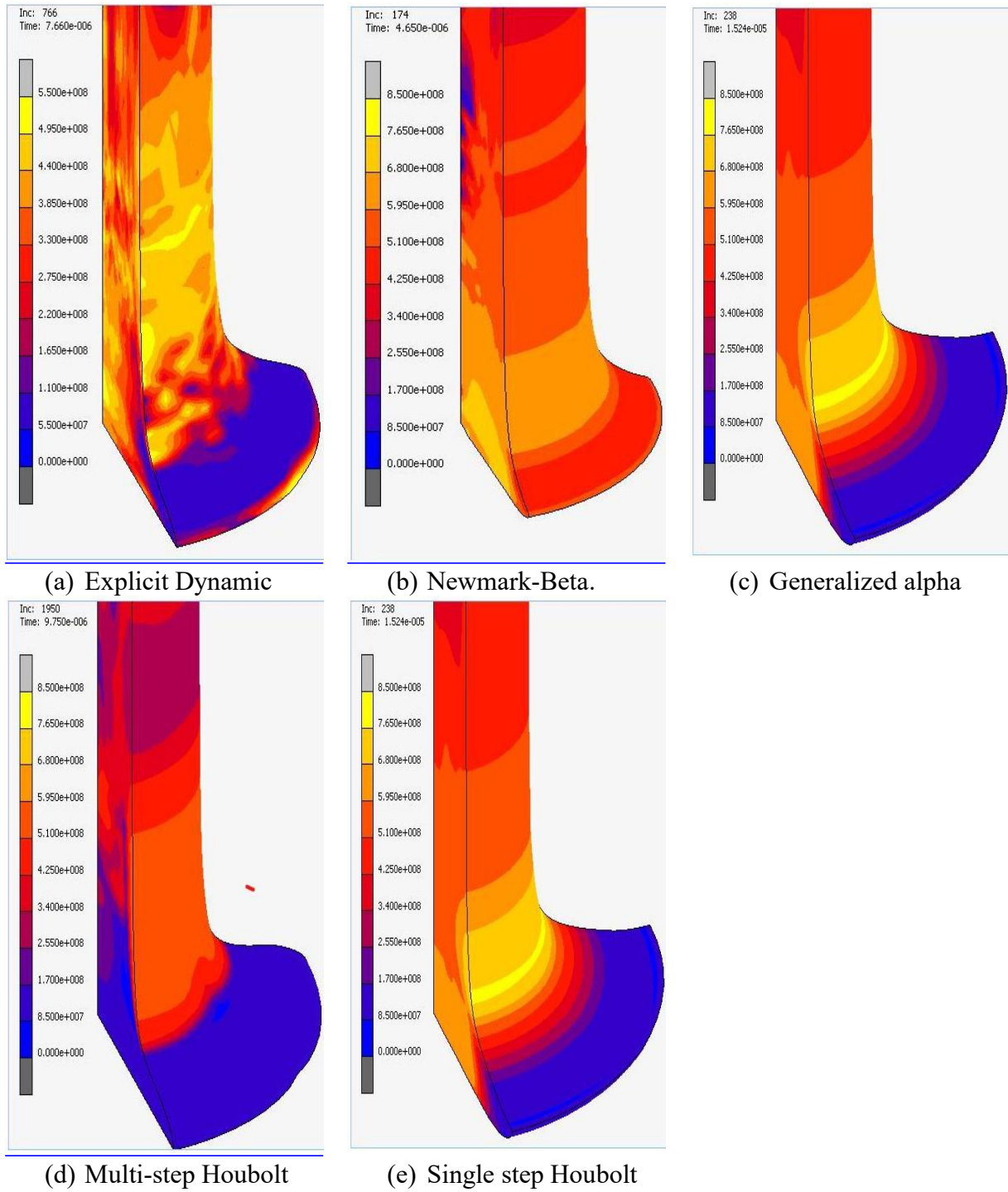
Fig 5. The von-mises stress contours at the initial time step of 3.0 μs.

central axis at around 1.6 μs. The single-step Houbolt algorithm can be treated as special cases of generalized-alpha algorithm. The stress distribution obtained in Fig. 3(e) by the single-step Houbolt algorithm is exactly similar due to the generalized-alpha distribution. The single-step generalized alphanumerical dissipation is controlled by two constant parameters $\alpha_f$ and $\alpha_m$ or by spectral radius. The generalized-alpha algorithm constant parameters comprise a number of other time integration algorithm. The equations which are solved using the generalized-alpha method $\alpha_f = 0$ and $\alpha_m = 1$ or spectral radius 0 are exactly similar to the stress distribution obtained by the single-step Houbolt algorithm. In Fig. 3(d) the stress distribution is obtained by solving time integrals by multistep Houbolt algorithm. The maximum stress

obtained is 850 MPa at the periphery of the outer edge of the Taylor rod while in other cases the maximum stress generated a little closer than the boundary of the Taylor rod. The jump at the central axis has not been observed in this case.

## At the intermediate time interval (When the jump achieves the maximum distance at the central axis)

To analyze the effect of different time integration schemes on the stress distribution and the deformation behavior of the impact problem, it has become necessary to analyze the deformation/stress contour at different time intervals. The intermediate time interval in the present study is taken when the jump of the central point of the projectile achieves the maximum value. In the three integration schemes, the central point achieves maximum height at the time step of around 3.0 µs. As discussed in the previous section, in the explicit time integrations scheme there is no jump at the central axis of the projectile. The projectile remains in contact with the rigid target at the central point throughout the analysis as shown in Fig. 4(a). The outer edge of the projectile also remains in contact with the target. The deformation in the outer mushroomed part of the projectile is not similar to the deformation shown by Teng et al. [12]. The stress distribution is consistent with the literature, but the deformation pattern is not consistent. In the Newmark-beta algorithm the center portion lifts to 0.3 mm at 3.0 µs. The jump of the central portion is maximum as compared to other integration schemes. This represents the larger deformation of the projectile than the usual as shown in Fig. 4(b). For the Newmark integration operator, this will generate the requirement for the numerical damping to exclude the unnecessary numerical dissipation during the analysis.

The jump in the generalized-alpha scheme as well as in the single-step Houbolt operator as discussed earlier exhibits the same behavior as 0.2 mm while the jump in the multistep Houbolt operator has not been observed. The jump is considered to be an important factor in determining the stress and fracture pattern in the Taylor rod. When the jump-starts the triaxiality in the rod become positive which will in return increases the damage at the central axis of the rod. But the larger value of the jump in the case of the Newmark integration scheme is due to the uncontrolled numerical dissipation. The analysis will fail to achieve the required convergence in the later stages if the numerical dissipation is not controlled at the initial stages.

## At the last time interval (Timestep where analysis failed to achieve the convergence)

In the last time step, the analysis is failed to achieve the required convergence. The maximum value of the stress in the explicit analysis is 550 MPa around the periphery of the projectile as shown in Fig. 5(a). The outer edge of the projectile lifts and this portion provide less resistance to the deformation as compared to the portion at the central axis or along the length of the rod. The stress distribution is nonlinear, generating the non-uniform stress distribution around the periphery of the rod as shown in Fig. 5(a). The analysis failed to achieve the required convergence at 7.0 µs. In the explicit dynamic integration, due to high-frequency dissipation, the excessive numerical damping has been introduced by the algorithm. This is the most likely reason for the response and non-occurrence of the jump in the rod using the dynamic operator. In Fig 5(b), the analysis failed to achieve the required convergence at 4.0 µs. The default parameters for the Newmark beta operators are $\beta = 0.25$ and $\gamma = 0.5$ as recommended by the literature failed to achieve the balance within 1 µs. The Newmark beta scheme implemented in the MSC Marc Mentat™ is not consistent with the default parameters. In the second study, the default parameters have been varied to $\beta = 0.5$ and $\gamma = 0.9142$. Using the Newmark integration scheme, the maximum time achieved by the analysis is 4.0 µs. The

stress contours are consistent with other integration schemes, but the analysis failed to achieve the prescribed tolerance. But the jump of the central point is also overpredicted, which represents the excessive numerical dissipation. In the study [16], the Newmark time integration scheme has been implemented to simulate the behavior of the Taylor rod impact problem. The algorithmic damping has been introduced in the study to analyze the Taylor rod impact problem. In the Newmark integration scheme, the incremental time is inversely proportional to the displacement and acceleration distribution. So, the dependence on the incremental time step leads to the instability of the structure. In Fig. 5(c), the analysis failed to converge up to the required tolerance at the time interval of 15 µs. The stress distribution is consistent with the literature. The maximum stress of 700 MPa is generated at the outer periphery of the projectile above the petal as shown in Fig. 5(c). The jump at the central point is also validated with the literature. The maximum von Mises stress is achieved at the central axis above the central point of the projectile. In Fig. 5(d), the stress distribution is similar to the Explicit dynamic analysis. The central axis as well as the outer periphery of the projectile remain in contact with the target throughout the analysis. No jump of the central axis has been observed using the multistep step Houbolt operator. In the multi-step Houbolt operator, the incremental time step is fixed during the analysis. The maximum stress in generated at the central axis as well at the outer periphery of the projectile. The stress is relieved as the outer edge of the rod lifts at the end of the analysis. As compared to the single-step Houbolt operator the multistep step Houbolt cannot be used with the adaptive time steeping which is necessary to capture the phenomenon of the impact process.

**Conclusions**

The numerical study of the Taylor rod has been performed to analyze the effect of different time integration schemes on the overall deformation behavior of the Taylor rod. The finite element method using MSC Marc Mentat™ is employed for this purpose. The triaxiality effect, Johnson-Cook plasticity, and stiffness reduction had been introduced in the study through an external Fortran-based user subroutine. The numerical results are validated with the experimental results. The effect of different integration schemes on the von-Mises stress distribution had particularly been discussed in the present work. The following conclusions have been drawn:

- The Explicit and Multi-step Houbolt operators coded in FE code MSC Marc Mentat™ are not able to capture the exact deformation in the Taylor rod impact test. The jump of the central point in the Taylor rod has not been observed in the aforementioned operators.
- In Explicit and Multistep Houbolt operators the central axis, as well as the outer edge of the rod remains in contact with the rigid wall throughout the analysis. The
- The Newmark-Beta operator in FE code MSC Marc Mentat™ overpredicts the deformation in the rod.
- The Newmark-beta operator exhibits excessive numerical dissipation. This results in the loss of accuracy, and this is the most likely reason for the overprediction of the deformation in Newmark integration.
- The WBZ time integration scheme can be treated as a special case of the single-step Houbolt operator and with parameters $\alpha_f = 0$ and $\alpha_m = 0$ the deformation is exactly similar to the Newmark-Beta integration scheme. If the parameters are fixed at $\alpha_f = 0$ and $\alpha_m = 0$, the deformation is exactly similar with the single-step Houbolt scheme.

- The Generalized-alpha method is the most suitable operator time integration operator to capture the exact deformation and stress distribution in the Taylor rod impact problem.

## Acknowledgments

### References

[1] Taylor, G. (1948). The Use of Flat-ended Projectiles for Determining Dynamic Yield Stress, I: Theoretical Considerations, Proceedings of the Royal Society of London Series A. Mathematical and Physical Sciences, 194: 289–299.

[2] Chung, J. and Hulbert, G.M., "A family of single-step Houbolt time integration algorithms for structural dynamics", Comp. Meth. in App. Mech. Engg., 118, 1994.

[3] Chung, J. and Hulbert, G.M., "A time integration algorithm for structural dynamics with improved numerical dissipation: The Generalized-Method", Journal of Applied Mechanics, Vol. 60, pp. 371 - 375, June 1993.

[4] Hilber, H.M., Hughes, T.J.R., and Taylor, R.L., "Improved Numerical Dissipation for Time Integration Algorithms in Structural Dynamics", Earthquake Engineering and Structural Dynamics, Vol. 5, pp. 283-292, 1977.

[5] Newmark, N. M. "A Method of Computation for Structural Dynamics." ASCE of Eng. Mech. 85, 67-94, 1959.

[6] Wood, W.L., Bossak, M., and Zienkiewicz, O.C., "An Alpha Modification of Newmark's Method", International Journal for Numerical Methods in Engineering, Vol., 15, pp. 1562-1566, 1981.

[7] Chung, J. and Hulbert, G.M., "A time integration algorithm for structural dynamics with improved numerical

[8] Woodward, R.L., O'donnell, R.G. and Flockhart, C.J. (1992). Failure Mechanisms in Impacting Penetrators, Journal of Material Science, 27: 6411–6416.

[9] Worswick, M.J. and Pick, R.J. (1995). Void Growth and Coalescence During High Velocity Impact, Mechanics of Materials, 19: 293–309.

[10] Addessio, F.L., Johnson, J.N. and Maudlin, P.J. (1993). The Effect of Void Growth on Taylor Cylinder Impact Experiments, Journal of Applied Physics, 73: 7288–7297.

[11] Celentano, D.J. (2002). Thermomechanical Analysis of the Taylor Impact Test, Journal of Applied Physics, 91: 3675–3686.

[12] Teng, X., Wierzbicki, T., Hiermaier, S. and Rohr, I. (2005). Numerical Prediction of Fracture in the Taylor Test, International Journal of Solids and Structures, 42: 2929–2948.

[13] Bao, Y., Lee, Y., Wierzbicki, T. Evaluation and calibration of seven fracture models. International Journal of Mechanical Sciences, submitted.

[14] Gautam, S.S. and Saxena, R.K., 2012. A numerical study on effect of strain rate and temperature in the Taylor rod impact problem. The International Journal of Structural Changes in Solids, 4, pp.1-11.

[15] Gautam, S. and Saxena, R., 2015. A finite element study on effect of frictional heating in the Taylor rod impact problem. World Journal of Engineering.

[16] Gautam, S.S., Babu, R. and Dixit, P.M., 2011. Ductile fracture simulation in the Taylor rod impact test using continuum damage mechanics. International Journal of Damage Mechanics, 20(3), pp.347-369.

[17] Rathore, K.K., Jasra, Y. and Saxena, R.K., 2020. Numerical simulation of fracture behavior under high-velocity impact for Aluminium alloy 6060 target plate. Materials Today: Proceedings, 28, pp.1809-1815.

[18] Xiao, X., Mu, Z., Pan, H. and Lou, Y., 2018. Effect of the Lode parameter in predicting shear cracking of 2024-T351 aluminum alloy Taylor rods. International Journal of Impact Engineering, 120, pp.185-201.

[19] Iqbal, M.A., Senthil, K., Bhargava, P. and Gupta, N.K., 2015. The characterization and ballistic evaluation of mild steel. International Journal of Impact Engineering, 78, pp.98-113.

[20] MSC Software 2013 Volume A: Theory and user information (MSC. Software Corp.)

[21] Yadav, S., Singhal, S., Jasra, Y. and Saxena, R.K., 2020. Determination of Johnson-Cook material model for weldment of mild steel. Materials Today: Proceedings, 28, pp.1801-1808.

# Box test and numerical simulation on settlement and energy dissipation characteristics of ballast under cyclic load

**\*†Tianqi Zhang, Yanyun Luo**

Institute of Rail Transit, Tongji University, China

**\***Presenting author, †Corresponding author: ztqljs1994@hotmail.com

## Abstract

In order to analyze the settlement and energy dissipation characteristics of ballast under cyclic load, the box test of ballast under cyclic loading was carried out by using hydraulic servo fatigue machine, and the settlement law of ballast was obtained. In addition, the discrete element method simulation is used to simulate the test of ballast box. On this basis, the settlement law of ballast under different loading conditions is analyzed, and the change law of ballast energy in the process of cyclic loading and unloading is also analyzed. The results show that the settlement of sleeper under cyclic load can be divided into two stages, the first is the rapid settlement stage, after a period of circulation into the stable settlement stage. The discrete element simulation method can be used to simulate the box test, and this numerical simulation method can also be used to analyze the mesoscopic characteristics of ballast. Both friction coefficient and damping ratio have effects on the settlement and energy dissipation of the ballasted track. The larger the coefficient of friction, the smaller the settlement of the sleeper, which is due to the larger the coefficient of friction, the more likely the interlocking between the particles is to occur. But for the damping ratio, there is a different rule. Although the greater the damping ratio, the greater the sleeper settlement, but when the damping ratio is 0.2 rather than 0, the settlement is minimum. When the damping is large, the damping consumes more energy. However, the friction coefficient also has an effect on the damping energy consumption. Under the same damping ratio, the smaller the friction coefficient is, the greater the damping energy consumption will be. The damping energy consumption increases linearly with the increase of loading times. For the total friction energy consumption, it does not mean that the greater the friction coefficient, the greater the energy consumption. This may be because a higher coefficient of friction may make it more difficult for particles to slide relatively to each other and between particles and walls.

## Key words

Ballasted track, box test, discrete element method, settlement, energy dissipation

## Introduction

Ballasted track is a very common kind of track form in the world at present. The track bed part is composed of a certain gradation of gravel ballast. It has a strong dispersive nature, so it is difficult to use the traditional finite element method to analyze its mechanical characteristics. At present, many scholars use experiment and discrete element method to study ballast track.

In the aspect of the research on the modeling of the micro scale discrete element of ballast particles. Professor Powrie's research group uses the approximate polyhedral particle modeling method. Ahmed et al. [1] calibrated the contact parameters of ballast through the indoor triaxial static compression test, and analyzed the effects of strain rate, damping coefficient, contact stiffness and contact friction coefficient on the simulation results. On this basis, Harkness et al. [2] improved the Hertz contact model for solving the contact force between ballast particles, proposed a cone-shaped damage model considering the contact wear of particle surface to simulate the change of particle contact stiffness, and adopted the friction coefficient related to the contact normal force. The research team of Professor McDowell at the University of Nottingham in the UK has also constructed cluster particle simulation ballast. Lu and McDowell [3], [4] used spherical particles and cluster particles composed of eight spherical particles to simulate their compression test of a ballast box.

In addition, many researchers are concerned about energy dissipation mechanism of particle damping and granular matter. Cempel and Lotz [5] investigated the damping performance of particle dampers and found that the energy dissipation is related to the impacts between particles or particles and the wall of the container and the friction between them. Mao et al. [6] utilized a 3D discrete element method to simulate the particle dampers and showed that the particle damping is a combination of the impact and the friction damping. Xu et al. [7] and Lu et al. [8], [9] found that energy dissipation mechanisms of particle damping are primarily related to friction and impact phenomena.

Based on the mechanics of granular materials and the thermodynamics of continuous media, Kong Liang [10] studied the macro and micro research ideas of the constitutive model of rock and soil, and drew the conclusion that the energy dissipation is caused by particle breakage, particle rotation, rolling and sliding. Yang Yugui et al. [11] analyzed the energy characteristics of the typical frozen silt through the stress-strain curve of loading and unloading in the triaxial test, but in the indoor test, the method of loading and unloading cannot achieve the fine statistics of energy in the meso level.

To sum up, the current research on ballast track mainly adopts experiments and discrete element method to study the mechanical characteristics, but the research on the settlement characteristics and energy dissipation characteristics of ballast under cyclic loading is relatively limited, and the energy dissipation law of ballast as a kind of discrete material under load is worth studying. Therefore, this paper is based on the ballast box test under cyclic load, using the method of combining test and numerical simulation, analysis of the settlement and energy dissipation law of your majesty ballast under cyclic load.

## 1 Box test

In the box test, one 650 * 500 * 550mm ballast box is used to observe the mechanical characteristics of ballast particles under cyclic load. The blue rectangle in the Fig. 1 is the range of the study area, which is located directly under the rail.
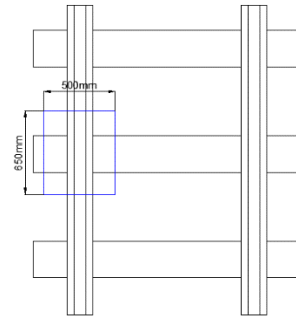
**Figure 1. Study area**

The box test is shown in Fig. 2. The ballasts used in the box test are in accordance with the industry standard TB/T2140-2008 "railway ballast" in the provisions of the super-grade ballast requirements, as shown in Table 1.

**Table 1. Particle grading**

| Side length of square sieve (mm) | | 22.5 | 31.5 | 40.0 | 50.0 | 63.0 |
|---|---|---|---|---|---|---|
| Percentage passing mass (%) | Standard specification | 0~3 | 1~25 | 30~65 | 70~90 | 100 |
| | Box test | 1 | 20 | 50 | 80 | 100 |

The cyclic load is with the frequency of 5Hz and the amplitude of 70kN and loading for a total of 1200 cycles. The ballast particles will be constantly squeezed, collided and rubbed, and the lateral box will move accordingly, due to the effect of external load. The signal of sleeper vertical displacement is collected to the data acquisition system during this process.



**Figure 2. The box test**

## 2 DEM simulation of box test

The settlement characteristics of ballast under cyclic load can be obtained through box test, but it is difficult to obtain some meso-parameters directly, so discrete element software is used to simulate it. According to the previous studies [12], the boundary effect of the container could be ignored if the size of the container is 8 times greater than the average particle size. The minimum length of the container is more than times the average particle size in this paper, which is less than 50mm, so the influence of the boundary effect could be ignored.

*2.1 The model parameters*

Table 2 shows the parameters used in the numerical simulation in the paper. After several trial calculations in the early stage, the simulation results obtained by using the parameters shown in Table 2 are in accordance with the sleeper settlement law in the test. As shown in Fig. 3, 24 kinds of ballast particles are simulated in accordance with the real ballast shapes used in the box test, and the gradation in Table 1 is adopted. A force loading board is used to simulate the sleeper. The left, right and bottom wall are used to simulate the box, and the wall is extended to avoid the ballast running out from the intersection.

**Table 2. Parameters of DEM modelling**

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| Normal contact stiffness of ballast($N \cdot m^{-1}$) | $7.5 \times 10^8$ | Normal contact stiffness of wall($N \cdot m^{-1}$) | $1.2 \times 10^9$ |
| Tangential contact stiffness of ballast($N \cdot m^{-1}$) | $6.5 \times 10^8$ | Tangential contact stiffness of wall($N \cdot m^{-1}$) | $1.2 \times 10^9$ |
| Loading particle density($kg \cdot m^{-3}$) | $2.8 \times 10^3$ | Ballast particle density ($kg \cdot m^{-3}$) | $2.7 \times 10^3$ |
| Friction coefficient | 0.5 | Local damping | 0.05 |
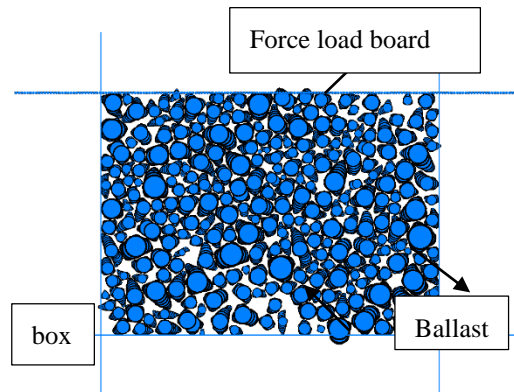


**Figure 3. Simulation model**

*2.2 Model validation*
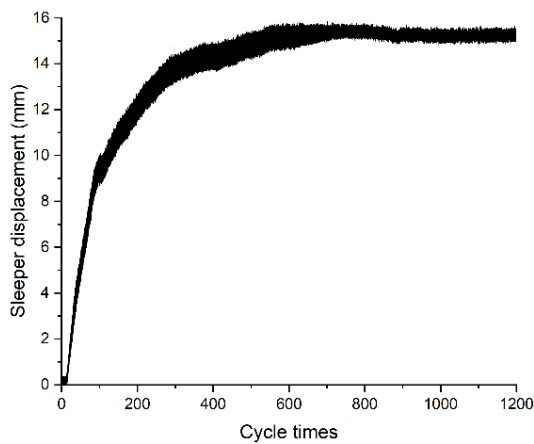


| a) Displacement of sleeper(test) | b) Displacement of sleeper(simulation) |
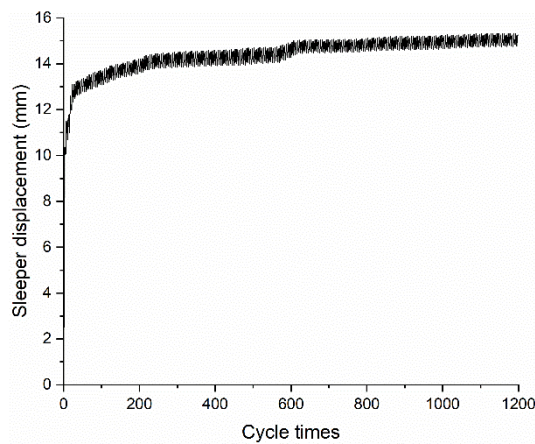
**Figure 4. Contrast between simulation results and test results**

Fig. 4 is the comparison between the simulation and test result. According to the comparison, the displacement trend of the sleeper under the cyclic load of the simulation result is consistent with the measured result, indicating that the DEM established in this paper could be used to simulate the box test of ballast under the cyclic load.

*2.3 The law of settlement*

Fig. 5 shows the settlement of sleeper with different contact friction coefficient (fric) and normal damping ratio (dp), as well as tangential damping ratio (ds). As can be seen from Fig. 6, there are two main stages of sleeper settlement under cyclic load. At the beginning of the cycle, the settlement changes rapidly when the bed is subjected to cyclic load, but the settlement trend gradually slows down. After a certain number of cycles, the settlement of the bed tends to be stable.
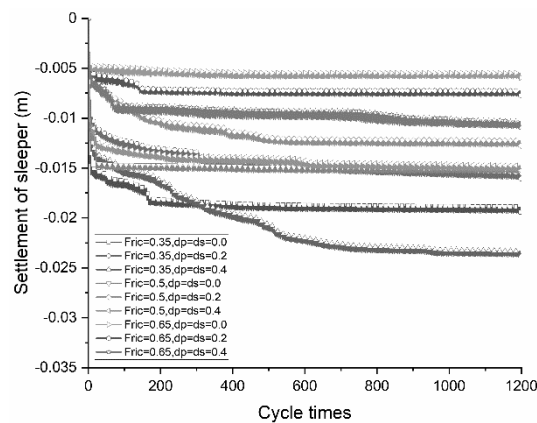


**Figure 5. The settlement of sleeper**

In addition, it can also be seen that the settle of sleeper is smaller with larger friction coefficient because the larger the friction coefficient is, the more likely the interlocking between particles will occur. But for the damping ratio, there is a different rule. Although the bigger the damping ratio is, the more settlement of sleeper is, the settlement is smallest while the damping ratio is 0.2 rather than 0.

*2.4 The law of energy evolution*

The influence of different friction coefficients and contact damping ratio of ballast under cyclic load on the total energy consumed by friction (Eslip) and energy consumed by particle contact damping (Edashpot) are mainly studied tin this paper.
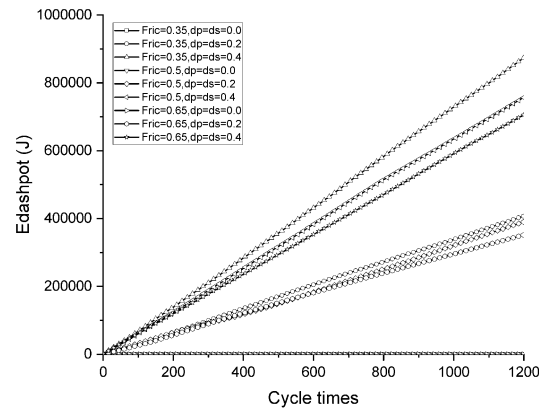
**Figure 6. The dissipation energy of contact damping**

Fig. 6 shows the dissipation energy of contact damping among ballasts. It can be seen that, the Edashpot is 0 while there is no damping among ballasts, which is obvious. When the damping is large, the damping consumes more energy. But the friction coefficient also has an effect on the damping energy dissipation. Specifically, in the case of the same damping ratio, the smaller the friction coefficient, the greater the damping energy consumption. This may lie in the smaller the friction coefficient, the ballasts are more likely to produce relative sliding under cyclic load, and leading to the greater the damping energy consumption. In addition, the damping energy consumption basically increases linearly with the increase of loading times, indicating that in the cyclic loading process, the damping energy consumption is basically the same in each cycle.
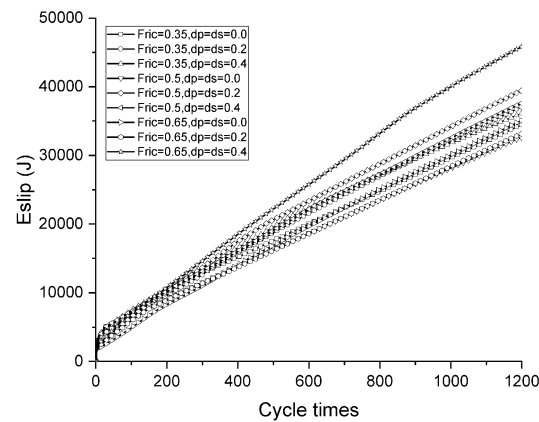


**Figure 7. The dissipation energy of contact friction**

Fig. 7 shows the dissipation energy of contact friction coefficient. It can be seen that the dissipation energy of contact friction experiences a very obvious increase stage at first 20 cycles, and then the increase rate slowed down and showed an almost linear trend.

For the total friction energy consumption, when the friction coefficient is 0.65 and the contact damping ratio is 0.4, the total friction energy consumption is the maximum; when the friction coefficient is 0.5 and the contact damping ratio is 0, and when the friction coefficient is 0.65 and the contact damping ratio is 0.2, the total friction energy consumption is the minimum. In other words, it's not that the greater the friction coefficient, the greater the energy consumption. This is because a higher coefficient of friction may make it more difficult for the particles to

slide relatively to each other and between the particles and the wall.

## Conclusion

Based on the ballast box test under cyclic load and discrete element method simulation, the settlement law of sleeper and the change of ballast energy in the process of cyclic loading and unloading are analyzed, and the following conclusions are obtained :

(1) Under cyclic loading and unloading conditions, sleeper settlement will occur because of the interaction between ballast. The settlement is mainly divided into two stages, the first is the rapid settlement stage, after a period of circulation, it goes into the stable settlement stage.

(2) The discrete element method simulation can be used to simulate the box test, and the numerical simulation method can be used to analyze the mesoscopic characteristics of ballast. Both friction coefficient and damping ratio have effects on the settlement and energy dissipation of the ballasted track. The larger the coefficient of friction, the smaller the settling of the sleeper. It is due to the larger the coefficient of friction is, the more likely the interlocking between the particles is to occur.

(3) But for the damping ratio, there are different laws. Although the greater the damping ratio, the greater the sleeper settlement, but while the damping ratio is 0.2 rather than 0, the settlement is minimum. When the damping is large, the damping consumes more energy. However, the friction coefficient also has an effect on the damping energy consumption. Under the same damping ratio, the smaller the friction coefficient is, the greater the damping energy consumption will be. The damping energy consumption increases linearly with the increase of loading times.

(4) For the total friction energy consumption, it does not mean that the greater the friction coefficient, the greater the energy consumption. This may be because a higher coefficient of friction may make it more difficult for particles to slide relatively to each other and between particles and walls.

**References**

[1]   Ahmed S, Harkness J, Pen L L, et al. Numerical modelling of railway ballast at the particle scale [J]. International Journal for Numerical & Analytical Methods in Geomechanics, 2015, 40(5):713-737.

[2]   Harkness J, Zervos A, Pen L L, et al. Discrete element simulation of railway ballast: modelling cell pressure effects in triaxial tests[J]. Granular Matter, 2016, 18(3):65.

[3]   Lu M, Mcdowell G R. The importance of modelling ballast particle shape in the discrete element method[J]. Granular Matter, 2007, 1(1):69-80.

[4]   Ferellec J F, Mcdowell G R. A method to model realistic particle shape and inertia in DEM[J]. Granular Matter, 2010, 12(5):459-467.

[5]   C. Cempel, G. Lotz, Efficiency of vibrational energy dissipation by moving shot, Journal of Structural Engineering, 1993, 119(9), 2642-2652.

[6]   Xu, Z., Wang, M. Y., and Chen, T. An Experimental Study of Particle Damping for Beams and Plates. Journal of Vibration and Acoustics-Transactions of the Asme. January 2004; 126(1): 141–148.

[7]   Xu, Z. W., Wang, M. Y., and Chen, T. N., 2002, "Particle Damping for Passive Vibration Suppression: Numerical Modeling," Journal of Sound and Vibration, 2005, 279(3–5), 1097.

[8]   Z. Lu, X. L. Lu, S. F. Masri, Studies of the performance of particle dampers under dynamic loads , Journal of Sound and Vibration, 2010, 329(26), 5415.

[9]   Z. Lu, X. Lu, W. Lu, S. F. Masri, Experimental studies of the effects of buffered particle dampers attached to a multi-degree-of-freedom system under dynamic loads, Journal of Sound and Vibration, 2012, 331(9), 2007.

[10] KONG Liang. Constitutive modeling for soils based on granular matter mechanics and continuous media thermomechanics [J]. Rock and Soil Mechanics, 2010, 31(Supp. 2): 1－6.

[11] YANG Yu-gui, LAI Yuan-ming, LI Shuang-yang, et al. Experimental study of deformation failure and energy properties of frozen silt under triaxial compression [J]. Rock and Soil Mechanics, 2010, 31(11): 3505－3510.

[12] INDRARATNA B. Behaviour of Fresh and Fouled Railway Ballast Subjected to Directed Shear Testing-Discrete Element Simulation [J]. International Journal of Geomechanics, 2012, 14(10): 3-4.

# On the spatial symmetry of high-resolution simulation
# for compressible flow

**†*Hiro Wakimura[1], Shinichi Takagi[1], and †Feng Xiao[1]**

[1]Department of Mechanical Engineering, Tokyo Institute of Technology, Japan.

*Presenting author: wakimura.h.aa@m.titech.ac.jp
†Corresponding author: ijsp090724@gmail.com (Hiro Wakimura)
†Corresponding author: xiao.f.aa@m.titech.ac.jp (Feng Xiao)

**Abstract**

Compressible flow has complex structures of wide-range scales such as vortices, turbulence, shock waves and contact discontinuities. In order to accurately capture the small fluid structures and discontinuous solutions, low-dissipation schemes have been developed so far. Meanwhile, the improvement of computational power enables simulations with high grid resolution. As a result, numerical simulations with extremely low dissipation can be obtained. However, in such high-resolution numerical simulations, errors of floating-point arithmetic may be amplified and result in asymmetric solutions even the physical configurations are symmetric. In this study, we have thoroughly examined the causes of symmetry-breaking and proposed numerical techniques to completely eliminate the numerical errors for asymmetric solutions.

**Keywords:** Computational Fluid Dynamics; Compressible Flow; High-resolution Simulation; Low-dissipation Scheme; Spatial Symmetry Preserving

## Introduction

In the compressible flow, a wide-range of scale fluid structures such as vortices, turbulence, shock waves and contact discontinuities are generated simultaneously. Solving all of these structures with high accuracy is a challenging problem, and the high-resolution schemes has been developed for decades. Methods based on the Godunov-type finite volume method [1] have been widely studied since they satisfy the conservation law in principle. In finite volume method, the interpolation of the distribution of physical quantities, called spatial reconstruction, is an important factor in terms of the calculation accuracy. TVD (Total Variation Diminishing) schemes [2] including MUSCL (Monotone Upstream-centered Schemes for Conservation Law) scheme [3] preserves monotonicity and has a 2nd-order accuracy for smooth solutions by performing linear interpolation with a slope limiter. WENO (Weighted Essentially Non-Oscillatory) schemes [4][5] achieve high-order accuracy for smooth solutions and avoid numerical oscillations near discontinuous solutions by weighting the smooth ones among multiple sub-stencils. While these schemes are currently the mainstream computational methods for compressible flows, they have large numerical dissipation error and are unable to accurately capture discontinuous solutions and small fluid structures. The low-dissipation property of the original linear scheme is lost due to the non-linear limiter, which is introduced to suppress the numerical oscillation in the vicinity of discontinuous solutions.

BVD (Boundary Variation Diminishing) schemes [6]-[9] have been developed based on BVD principle of reducing the numerical dissipation term in the equation of approximate Riemann

solver. The first BVD scheme, BVD-WENO-THINC scheme [6], switches WENO scheme for smooth solutions and THINC (Tangent of Hyperbola for INterface Capturing) scheme for discontinuous solutions. THINC schemes [10]-[12] were originally developed for capturing moving interfaces and suitable for representing discontinuous physical distribution. The selection of the appropriate interpolation function is performed according to the BVD algorithm [6]. As a result, BVD-WENO-THINC scheme can capture the discontinuous solutions more accurately than WENO scheme. Furthermore, $P_n T_m$-BVD (polynomial of $n$-degree and THINC function of $m$-level reconstruction based on BVD algorithm) scheme [8][9] achieved arbitrarily high order accuracy (($n + 1$)th-order) for smooth solutions, suppression of the numerical oscillations, and accurate capture of the discontinuous solutions without non-linear limiter. This scheme uses a $n$th-degree linear polynomial scheme and THINC schemes with $m$ different steepness parameters as the admissible interpolation functions and applies the BVD algorithm in $m$-stage. The numerical results of the $P_n T_m$-BVD scheme can resolve both smooth and discontinuous solutions with superior fidelity compared with other existing schemes.

In practical applications, a computational scheme with sufficient resolution is required. In recent years, as the computational power is improved, it has become possible to perform simulations with high grid resolution. As a result, extremely high-resolution numerical results can be obtained by combining low-dissipation schemes with fine grids. On the other hand, in such a high-resolution simulation, rounding errors due to floating-point arithmetic are amplified and affect the calculation results. Particularly, it is known that the spatial symmetry of the flow structure is lost even though the physical configuration is set symmetrically. An asymmetric result is non-physical solution and should be avoided. Remacle et al. [13] initially identified that the cause of asymmetry is the rounding error and it grows with time evolution. Fleischmann et al. [14] pointed out that the lack of associativity in the summation or multiplication of more than two components generates unexpected asymmetry results, and completely eliminated the symmetry error by carefully examining the calculation procedure of the WENO-like schemes. To our knowledge, [14] is the only research that directly approaches the sources of the symmetry error and preserves the exact symmetry despite the extremely high resolution.

In this study, symmetry-preserving $P_4 T_2$-BVD scheme is proposed. Since the original THINC scheme does not satisfy the symmetry, a new formulation of the THINC scheme is developed. Moreover, we thoroughly analyze the entire calculation procedure including the characteristic decomposition and the approximate Riemann solver. Based on this analysis, we present new symmetry-preserving techniques, which have not been specified in [14] or other papers. These techniques are efficient not only for the BVD scheme but also for all Godunov-type schemes. As the proposed symmetry-preserving techniques are modifications in the level of the rounding errors, they do not affect the original properties of the schemes.

**Governing equations**

We solve 2D Euler equations as the governing equations,

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{U})}{\partial y} = 0, \tag{1}$$

where $\mathbf{U}$ is the conservative variables, and $\mathbf{F}(\mathbf{U})$ and $\mathbf{G}(\mathbf{U})$ are the flux functions in $x$- and $y$-direction respectively.

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix}, \qquad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E + p)u \end{pmatrix}, \qquad \mathbf{G}(\mathbf{U}) = \begin{pmatrix} \rho v \\ \rho vu \\ \rho v^2 + p \\ (E + p)v \end{pmatrix}, \tag{2}$$

where $\rho$ is the density, $u$ and $v$ are the $x$- and $y$-components of the velocity, $E$ is the total energy per unit volume, and $p$ the static pressure.

To close the system, the ideal gas law is added,

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho(u^2 + v^2)\right), \tag{3}$$

where $\gamma$ is the specific heat ratio.

The system of the Euler equations (1) can be diagonalized in $x$- and $y$-direction respectively as below,

$$\frac{\partial \mathbf{W}_x}{\partial t} + \mathbf{\Lambda}_x \frac{\partial \mathbf{W}_x}{\partial x} = 0, \tag{4}$$

$$\frac{\partial \mathbf{W}_y}{\partial t} + \mathbf{\Lambda}_y \frac{\partial \mathbf{W}_y}{\partial y} = 0, \tag{5}$$

where $\mathbf{\Lambda}_x = \text{diag}(u - c, u, u + c, u)$, $\mathbf{\Lambda}_y = \text{diag}(v - c, v, v + c, v)$, are the diagonal matrices and $\mathbf{W}$ is the characteristic variables. The sound speed $c$ is obtained by $c = \sqrt{\gamma p / \rho}$. Each component of $\mathbf{W}_x$ and $\mathbf{W}_y$ is depending on the eigenvalue of the system (1) and expressed as following in this paper,

$$\mathbf{W}_x = \begin{pmatrix} w^{(u-c)} \\ w^{(u)} \\ w^{(u+c)} \\ w^{(u\perp)} \end{pmatrix}, \qquad \mathbf{W}_y = \begin{pmatrix} w^{(v-c)} \\ w^{(v)} \\ w^{(v+c)} \\ w^{(v\perp)} \end{pmatrix}. \tag{6}$$

In Eq. (6), the symbol "$\perp$" is introduced in order to distinguish two same eigenvalues. For the same reason, the eigenvectors are expressed as,

$$\mathbf{L}_x = \begin{pmatrix} \vec{l}^{(u-c)} \\ \vec{l}^{(u)} \\ \vec{l}^{(u+c)} \\ \vec{l}^{(u\perp)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(b_1 + \frac{u}{c}\right) & -\frac{1}{2}\left(\frac{1}{c} + b_2 u\right) & -\frac{1}{2}b_2 v & \frac{1}{2}b_2 \\ 1 - b_1 & b_2 u & b_2 v & -b_2 \\ \frac{1}{2}\left(b_1 - \frac{u}{c}\right) & -\frac{1}{2}\left(\frac{1}{c} - b_2 u\right) & -\frac{1}{2}b_2 v & \frac{1}{2}b_2 \\ -v & 0 & 1 & 0 \end{pmatrix} = (\vec{l}_{x1} \ \vec{l}_{x2} \ \vec{l}_{x3} \ \vec{l}_{x4}), \tag{7}$$

$$\mathbf{L}_y = \begin{pmatrix} \vec{l}^{(v-c)} \\ \vec{l}^{(v)} \\ \vec{l}^{(v+c)} \\ \vec{l}^{(v\perp)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(b_1 + \frac{v}{c}\right) & -\frac{1}{2}b_2 u & -\frac{1}{2}\left(\frac{1}{c} + b_2 v\right) & \frac{1}{2}b_2 \\ 1 - b_1 & b_2 u & b_2 v & -b_2 \\ \frac{1}{2}\left(b_1 - \frac{v}{c}\right) & -\frac{1}{2}b_2 u & -\frac{1}{2}\left(\frac{1}{c} - b_2 v\right) & \frac{1}{2}b_2 \\ -u & 1 & 0 & 0 \end{pmatrix} = (\vec{l}_{y1} \ \vec{l}_{y2} \ \vec{l}_{y3} \ \vec{l}_{y4}), \tag{8}$$

$$\mathbf{R}_x = (\vec{r}^{(u-c)} \ \vec{r}^{(u)} \ \vec{r}^{(u+c)} \ \vec{r}^{(u\perp)}) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ u - c & u & u + c & 0 \\ v & v & v & 1 \\ H - uc & \frac{u^2 + v^2}{2} & H + uc & v \end{pmatrix}, \tag{9}$$

$$\mathbf{R}_y = (\vec{r}^{(v-c)} \ \vec{r}^{(v)} \ \vec{r}^{(v+c)} \ \vec{r}^{(v\perp)}) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ u & u & u & 1 \\ v - c & v & v + c & 0 \\ H - vc & \frac{u^2 + v^2}{2} & H + vc & u \end{pmatrix}, \tag{10}$$

where $b_1 = \frac{u^2 + v^2}{2}\frac{\gamma - 1}{c^2}$, $b_2 = \frac{\gamma - 1}{c^2}$, and the enthalpy $H = \frac{E + p}{\rho}$.

**Finite volume method**

We review the numerical framework of the finite volume method (FVM). In order to explain the numerical methods briefly, we here use 1D scalar conservation law as shown below,

$$\frac{\partial q}{\partial t} + \frac{\partial f(q)}{\partial x} = 0, \tag{11}$$

where $q$ is the conservative variable and $f(q)$ is the flux function. The computational domain is divided into $N$ non-overlapping cells, $\Omega_i = [x_{i-1/2}, x_{i+1/2}]$, $i = 1, 2, \dots, N$. The cell size $\Delta x = x_{i+1/2} - x_{i-1/2}$ is assumed to be constant over the computational domain for brevity.

In a standard FVM, the spatial distributions of the physical quantities are discretized by volume integrated average (VIA). Hence, the spatial discrete value in the cell $\Omega_i$ is defined as,

$$\bar{q}_i(t) \equiv \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} q(x, t) dx. \tag{12}$$

Integrating Eq. (11) and using Eq. (12), the semi-discrete version of Eq. (11) can be derived,

$$\frac{d\bar{q}_i}{dt} = -\frac{1}{\Delta x}\left(\hat{f}_{i+\frac{1}{2}} - \hat{f}_{i-\frac{1}{2}}\right), \tag{13}$$

where $\hat{f}_{i+1/2}$ is the numerical flux at the cell boundary $x_{i+1/2}$. Since Eq. (13) is now an ODE (ordinary differential equation), the solution $\bar{q}_i(t)$ can be updated by an ODE solver such as Runge-Kutta method. The physical flux passing through the cell interface is calculated by the approximate Riemann solver. While there are some types of the Riemann solver, they are given the following canonical form,

$$\hat{f}_{i+\frac{1}{2}} = \frac{1}{2}\left(f\left(q^L_{i+\frac{1}{2}}\right) + f\left(q^R_{i+\frac{1}{2}}\right)\right) - \frac{1}{2}\left|a_{i+\frac{1}{2}}\right|\left(q^R_{i+\frac{1}{2}} - q^L_{i+\frac{1}{2}}\right), \tag{14}$$

where $a$ is the characteristic speed and $q^{L/R}$ is the left/right cell boundary value. The boundary value is obtained by the reconstruction procedure. In this paper, as mentioned above, we use the $P_4T_2$-BVD reconstruction [8], which can resolve both smooth and discontinuous solutions with low dissipation compared to existing shock-capturing schemes. The calculation procedure of the $P_4T_2$-BVD scheme including the symmetry-preserving techniques is explained from the next subsection.

**The mechanisms of the symmetry error**

Fleischmann et al. [14] revealed that the symmetry errors are caused from a lack of the associativity of summation and multiplication in the floating-point arithmetic, shown as

$$(a + b) + c \neq a + (b + c), \tag{15}$$

$$(a \times b) \times c \neq a \times (b \times c). \tag{16}$$

In other words, the result of summation and multiplication depends on the order of calculation when there are three or more components. The lack of the associativity is the main cause of the asymmetry and we thoroughly examine the calculation procedure in terms of the order of summation or multiplication.

**Symmetry-preserving P₄T₂-BVD reconstruction**

In this section, we show the calculation procedure of the symmetry-preserving $P_4T_2$-BVD scheme, obtained by introducing some techniques into the original scheme. The $P_4T_2$-BVD reconstruction has 4th-degree polynomial function and THINC functions with two different steepness parameters as the candidate interpolants, called admissible reconstruction functions. The selection from the admissible reconstruction function is conducted by 2-stage BVD

algorithm. In this subsection, we explain the calculation procedures of each admissible reconstruction first, and specify the 2-stage BVD algorithm later.

*Candidate interpolant 1: 4th-degree polynomial scheme*

The reconstruction function of the 4th-degree polynomial scheme ($P_4$) for target cell $\Omega_i$ is given as,

$$Q_i^{P_4}(x) = \sum_{k=0}^{4} a_k (x - x_i)^k, \tag{17}$$

where $a_k$ ($k = 0,1,2,3,4$) is unknown coefficients, obtained from the following constraint conditions,

$$\frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} Q_i^{P_4}(x)dx = \bar{q}_j, \qquad (j = i, i \pm 1, i \pm 2). \tag{18}$$

Solving $a_k$ ($k = 0,1,2,3,4$) from Eq. (18) and substituting them into the reconstruction function (17), the cell boundary values can be obtained as,

$$\begin{aligned}
q_{i+\frac{1}{2}}^{L,P_4} &= Q_i^{P_4}\left(x_{i+\frac{1}{2}}\right) = \frac{1}{60}\left((2\bar{q}_{i-2} - 3\bar{q}_{i+2}) + (-13\bar{q}_{i-1} + 27\bar{q}_{i+1}) + 47\bar{q}_i\right), \\
q_{i-\frac{1}{2}}^{R,P_4} &= Q_i^{P_4}\left(x_{i-\frac{1}{2}}\right) = \frac{1}{60}\left((-3\bar{q}_{i-2} + 2\bar{q}_{i+2}) + (27\bar{q}_{i-1} - 13\bar{q}_{i+1}) + 47\bar{q}_i\right).
\end{aligned} \tag{19}$$

Note that the order of the summation of $\bar{q}_j$ ($j = i, i \pm 1, i \pm 2$) in Eq. (19) is adjusted so that the symmetry property of the cell boundary value is preserved.

*Candidate interpolant 2: THINC scheme*

THINC scheme [10]-[12] was originally developed for capturing moving interfaces. THINC function is good at representing the discontinuous solution and contributes significantly to the low dissipation of the BVD scheme.

In order to preserve the symmetry property of the THINC function, we propose a new formulation of the THINC function for target cell $\Omega_i$, written as

$$Q_i^T(x) = q_a + q_d \tanh\big(\beta(X_i - d_i)\big), \tag{20}$$

where

$$q_a = \frac{\bar{q}_{i+1} + \bar{q}_{i-1}}{2}, \qquad q_d = \frac{\bar{q}_{i+1} - \bar{q}_{i-1}}{2}, \qquad X_i = \frac{x - (x_{i+1/2} + x_{i-1/2})/2}{x_{i+1/2} - x_{i-1/2}}. \tag{21}$$

Except for parameter $\beta$, the only unknown parameter in the reconstruction function (20) is $d_i$, which indicates the jump location of the sigmoid function. $d_i$ is determined by the following condition,

$$\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} Q_i^T(x)dx = \bar{q}_i. \tag{22}$$

By solving $d_i$ from Eq. (22) and substituting it into the reconstruction function (20), the symmetry-preserving cell boundary values can be obtained as,

$$\begin{aligned}
q_{i+\frac{1}{2}}^{L,T} &= Q_i^T\left(x_{i+\frac{1}{2}}\right) = q_a + q_d \frac{T_1 + T_2/T_1}{1 + T_2}, \\
q_{i-\frac{1}{2}}^{R,T} &= Q_i^T\left(x_{i-\frac{1}{2}}\right) = q_a - q_d \frac{T_1 - T_2/T_1}{1 - T_2},
\end{aligned} \tag{23}$$

where

$$T_1 = \tanh\left(\frac{\beta}{2}\right), \qquad T_2 = \tanh\left(\frac{\alpha_i \beta}{2}\right), \qquad \alpha_i = \frac{\bar{q}_i - q_a}{q_d}. \tag{24}$$

The steepness parameter $\beta$ determine the gradient of the THINC function and the characteristic of the THINC scheme. THINC scheme with $\beta = 1.6$ can resolve the discontinuous solution within only four cells [11]. According to ADR (approximate dispersion relation) analysis [15][7], THINC scheme with $\beta = 1.1$ has similar characteristic to MUSCL scheme with van Leer limiter and has ability to suppress numerical oscillations near the discontinuities. We denote THINC with $\beta = 1.1$ by THINC($\beta_s$) and that with $\beta = 1.6$ by THINC($\beta_l$). In the P$_4$T$_2$-BVD reconstruction, three functions, $P_4$, THINC($\beta_s$) and THINC($\beta_l$) are used as the admissible reconstruction functions.

*2-stage BVD algorithm*

The BVD principle [6] suggests that the jump of the left and right boundary values, called boundary variation ($BV$),

$$BV_{i+\frac{1}{2}} = \left| q^L_{i+\frac{1}{2}} - q^R_{i+\frac{1}{2}} \right|. \tag{25}$$

By diminishing the value of $BV$, the numerical dissipation term in the Riemann solver equation (14) becomes small and the dissipation error in numerical results considered to be suppressed.

The selection from the candidate reconstruction functions is performed in the following procedure called 2-stage BVD algorithm:

I. The 1st-satge

   I.I. The reconstruction function selected in the 1st-stage $Q_i^{1st}$ is initially set to the 4th-degree polynomial,

$$Q_i^{1st} = Q_i^{P_4}. \tag{26}$$

   I.II. Calculate the values of $TBV$ (Total $BV$) for $P_4$ and THINC($\beta_s$) as following,

$$TBV_i^{P_4} = \left| q^{L,P_4}_{i-\frac{1}{2}} - q^{R,P_4}_{i-\frac{1}{2}} \right| + \left| q^{L,P_4}_{i+\frac{1}{2}} - q^{R,P_4}_{i+\frac{1}{2}} \right|,$$
$$TBV_i^{T_s} = \left| q^{L,T_s}_{i-\frac{1}{2}} - q^{R,T_s}_{i-\frac{1}{2}} \right| + \left| q^{L,T_s}_{i+\frac{1}{2}} - q^{R,T_s}_{i+\frac{1}{2}} \right|. \tag{27}$$

   I.III. Change the reconstruction function to THINC($\beta_s$) for the cells $\Omega_{i-1}, \Omega_i, \Omega_{i+1}$ using following BVD algorithm,

$$Q_j^{1st} = Q_j^{T_s} \quad (j = i, i \pm 1), \qquad \text{if } TBV_i^{T_s} < TBV_i^{P_4}. \tag{28}$$

II. The 2nd-stage

   II.I. Calculate the values of $TBV$ (Total Boundary Variation) for function selected in the 1st-stage and THINC($\beta_l$) as,

$$TBV_i^{1st} = \left| q^{L,1st}_{i-\frac{1}{2}} - q^{R,1st}_{i-\frac{1}{2}} \right| + \left| q^{L,1st}_{i+\frac{1}{2}} - q^{R,1st}_{i+\frac{1}{2}} \right|,$$
$$TBV_i^{T_l} = \left| q^{L,T_l}_{i-\frac{1}{2}} - q^{R,T_l}_{i-\frac{1}{2}} \right| + \left| q^{L,T_l}_{i+\frac{1}{2}} - q^{R,T_l}_{i+\frac{1}{2}} \right|. \tag{29}$$

   II.II. Determine the final reconstruction function for the target cell $\Omega_i$ using BVD algorithm as

$$Q_i = \begin{cases} Q_i^{T_l}, & \text{if } TBV_i^{T_l} < TBV_i^{1st}, \\ Q_i^{1st}, & \text{otherwise.} \end{cases} \tag{30}$$

In the 1st-stage, the smooth solutions are accurately calculated by the 4th-degree polynomial scheme with 5th-order convergence rate and the numerical oscillations near the discontinuities

are suppressed by the THINC($\beta_s$) scheme. In the 2nd-stage, the discontinuous solutions are captured with high resolution by the THINC($\beta_l$) scheme.

## Symmetry-preserving characteristic decomposition

Solving 2D Euler equations (1), the P$_4$T$_2$-BVD reconstruction is performed in terms of the characteristic variables $\mathbf{W}$ shown in Eq. (6), in order to avoid numerical oscillations [16]. For this reason, the transformation between the conservative variables $\mathbf{U}$ and the characteristic variables $\mathbf{W}$ is performed before and after the reconstruction.

The transformation from the solution $\overline{\mathbf{U}}$ to the characteristic variables $\overline{\mathbf{W}}$ is given as

$$\overline{\mathbf{W}}_x = \mathbf{L}_x \cdot \overline{\mathbf{U}}, \qquad \overline{\mathbf{W}}_y = \mathbf{L}_y \cdot \overline{\mathbf{U}}. \tag{31}$$

As Fleischmann et al. [14] have already proposed, from the point of view of the symmetry preserving, the product of the left-eigenvectors and the conservative variable vector in Eq. (31) should be calculated in the following order,

$$\overline{\mathbf{W}}_x = \vec{l}_{x1}\bar{\rho} + \left(\vec{l}_{x2}\overline{\rho u} + \vec{l}_{x3}\overline{\rho v}\right) + \vec{l}_{x4}\bar{E},$$
$$\overline{\mathbf{W}}_y = \vec{l}_{y1}\bar{\rho} + \left(\vec{l}_{y2}\overline{\rho u} + \vec{l}_{y3}\overline{\rho v}\right) + \vec{l}_{y4}\bar{E}. \tag{32}$$

In other words, the terms corresponding to $\overline{\rho u}$ and $\overline{\rho v}$ should be added first. This is because the terms $\overline{\rho u}$ and $\overline{\rho v}$ have different values in the case of diagonal symmetry, and the order of the summation in Eq. (32) will be different if added in the order $\bar{\rho}$, $\overline{\rho u}$, $\overline{\rho v}$, and $\bar{E}$.

The transformation from the reconstructed characteristic variables $\mathbf{W}$ to the reconstructed conservative variables $\mathbf{U}$ is given as

$$\mathbf{U}_x = \mathbf{R}_x \cdot \mathbf{W}_x, \qquad \mathbf{U}_y = \mathbf{R}_y \cdot \mathbf{W}_y. \tag{33}$$

In order to avoid the symmetry errors, the product of the right-eigenvectors and the characteristic variable vector in Eq. (33) should be calculated in the following order,

$$\mathbf{U}_x = \left(\vec{r}^{(u-c)}w^{(u-c)} + \vec{r}^{(u+c)}w^{(u+c)}\right) + \vec{r}^{(u)}w^{(u)} + \vec{r}^{(u\perp)}w^{(u\perp)},$$
$$\mathbf{U}_y = \left(\vec{r}^{(v-c)}w^{(v-c)} + \vec{r}^{(v+c)}w^{(v+c)}\right) + \vec{r}^{(v)}w^{(v)} + \vec{r}^{(v\perp)}w^{(v\perp)}. \tag{34}$$

The first equation of Eq. (34) means that the terms corresponding to the eigenvalues $(u - c)$ and $(u + c)$ should be summed up first. This is because the values of the characteristic variables $w^{(u-c)}$ and $w^{(u+c)}$ are interchanged and have different values in the case of $y$-axis symmetry. The meaning of the second equation of Eq. (34) is same as the first one in $y$-direction.

## Symmetry-preserving HLLC Riemann solver

In this paper, we take HLLC Riemann solver [17] as an example and propose techniques to preserve symmetry. The numerical flux of the HLLC solver for $x$-direction is expressed as

$$\hat{\mathbf{F}}^{HLLC} = \frac{1 + \text{sgn}(s^*)}{2}\left(\mathbf{F}^L + s^-(\mathbf{U}^{*L} - \mathbf{U}^L)\right) + \frac{1 - \text{sgn}(s^*)}{2}\left(\mathbf{F}^R + s^+(\mathbf{U}^{*R} - \mathbf{U}^R)\right), \tag{35}$$

where $s^- = \min\left(s^L, 0\right)$ and $s^+ = \max\left(s^R, 0\right)$. $s^L$ and $s^R$ are the left and right wave speeds respectively, and they are estimated by PVRS method [18] in this study. The asterisk mark ($*$) on the physical quantity indicates that it is located in the region between the left and right waves. The conservative variables in the intermediate region $\mathbf{U}^{*L}$ and $\mathbf{U}^{*R}$ can be derived from the Rankine-Hugoniot condition, as following,

$$\mathbf{U}^{*K} = \frac{s^K - u^K}{s^K - s^*}\begin{pmatrix} \rho^K \\ \rho^K s^* \\ \rho^K v^K \\ E^K + (s^* - u^K)\left(\rho^K s^* + \frac{p^K}{s^K - u^K}\right) \end{pmatrix}, \tag{36}$$

where $K = L$ or $R$. Fleischmann et al. [14] proposed the intermediate wave speed $s^*$ with symmetry-preserving treatment,

$$s^* = \frac{p^R - p^L + \{\rho^L u^L(s^L - u^L) - \rho^R u^R(s^R - u^R)\}}{\rho^L(s^L - u^L) - \rho^R(s^R - u^R)}. \tag{37}$$

Note that Eq. (37) is simply the original one with a bracket added.

It is noted that the flux function should be coded with reference to Eq. (2) to avoid the symmetry errors. Especially, the 3rd component of **F** ($F_3$) and the 2nd component of **G** ($G_2$) are products of three numbers, and the order of the calculation should be checked carefully. If $F_3$ and $G_2$ are set as follows,

$$(F_3, G_2) = (\rho uv, \rho vu) \text{ or } (\rho vu, \rho uv), \tag{38}$$

the lack of associativity (16) can be avoided in diagonal symmetry case.

**Numerical results**

In order to verify the symmetry-preserving techniques proposed above, some of the numerical tests are solved. The reconstruction is performed by P4T2-BVD scheme [8] in terms of the characteristic variables. The numerical flux is calculated by HLLC Riemann solver [17]. The time evolution is performed by 3rd-order Runge-Kutta method [19]. The CFL number is set to 0.6 for all benchmark tests. The calculations have been done on Intel Xeon CPU E5-2687W0 @ 3.10 GHz in a multi-threaded fashion on CentOS 6.10 operating system. GCC compiler version 4.4.6 is used without any special optimization options.

*Rayleigh-Taylor instability*

The Rayleigh-Taylor instability (RTI) is a typical benchmark test [20] that has a symmetric flow structure. In this test, many of the high-resolution schemes result in asymmetric numerical results [21][22][13].

The initial condition which Fleischmann et al. [14] proposed in terms of symmetry preserving is set as

$$(\rho_0, u_0, v_0, p_0) = \begin{cases} (2, 0, v_0(x), 2y + 1), & \text{if } y < 0.5, \\ (1, 0, v_0(x), y + 1.5), & \text{otherwise,} \end{cases} \tag{39}$$

where

$$v_0(x) = \begin{cases} -0.025c \cos(8\pi x), & \text{if } x < 0.125, \\ -0.025c \cos(0.25 - 8\pi x), & \text{otherwise.} \end{cases} \tag{40}$$

The boundary condition is reflective at the left and right boundaries and fixed values at the top and bottom boundaries as $(\rho, u, v, p)_{top} = (1, 0, 0, 2.5)$ and $(\rho, u, v, p)_{bottom} = (2, 0, 0, 1)$. The computational domain is $[0, 0.25] \times [0, 1]$. The specific heat ratio is $\gamma = 5/3$. The numerical results with the mesh resolution of $1024 \times 4096$ at time $t = 1.95$ are shown in Fig. 1.

Fig. 1 shows that the numerical result of the original P4T2-BVD scheme (left) generates $y$-axis asymmetric flow structure, and the proposed scheme (right) can completely preserve the $y$-axis symmetry property despite the extremely low-dissipation condition.

*Implosion test*

This test demonstrates an implosion phenomenon where a diamond-shaped low-pressure fluid is crushed by surrounding high-pressure [23]. Preserving symmetry in this test is challenging problem because $x$-axis, $y$-axis, and diagonal symmetry are existing simultaneously, and trajectories of diagonal jets are sensitive to the symmetry errors.
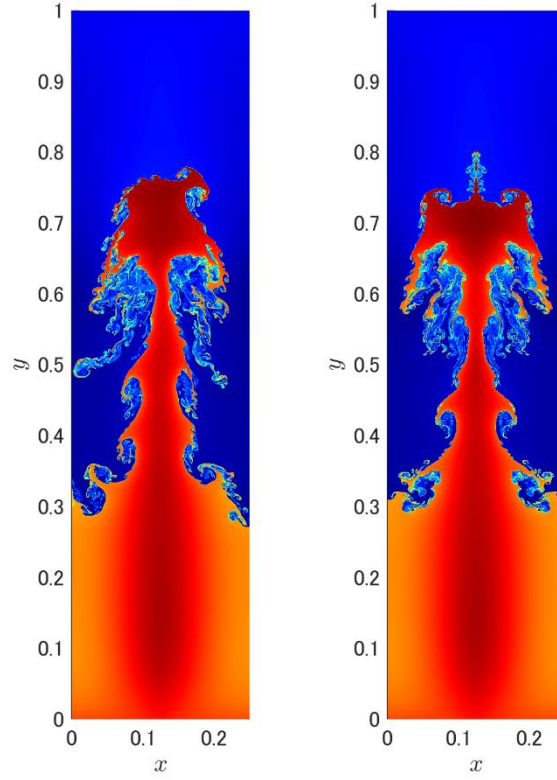
**Figure 1.  Numerical solutions of density (blue=0.85 to red=2.25) for RTI test.**

The initial condition is set as

$$(\rho_0, u_0, v_0, p_0) = \begin{cases} (0.125, 0, 0, 0.14), & |y + x| < 0.15 + \epsilon \text{ and } |y - x| < 0.15 + \epsilon, \\ (1, 0, 0, 1), & \text{otherwise,} \end{cases} \quad (41)$$

where $\epsilon = 10^{-10}$ is introduced in order to avoid asymmetric initial condition. The boundary condition is reflective at all boundaries. The computational domain is $[-0.3, 0.3] \times [-0.3, 0.3]$. The specific heat ratio is $\gamma = 1.4$. The numerical results with the mesh resolution of $1600 \times 1600$ at time $t = 2.5$ are shown in Fig. 2.
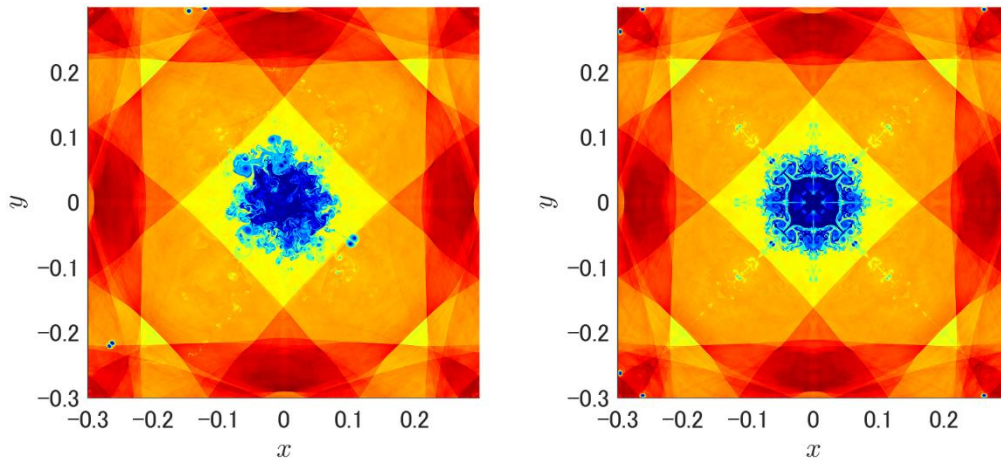


**Figure 2.  Numerical solutions of density (blue=0.41 to red=1.07) for implosion test.**

It is observed from Fig. 2 that the trajectories of the diagonal jets are distorted by the original $P_4T_2$-BVD scheme (left). By adopting the symmetry-preserving techniques discussed above, both axis and diagonal symmetry can be completely preserved (right).

## Conclusions

In this paper, the numerical method of the finite volume method is thoroughly examined and the countermeasures are proposed. Mainly in the three calculation procedures, i.e., $P_4T_2$-BVD reconstruction, characteristic decomposition, and HLLC Riemann solver, some formulas that cause lacks of the associativity of summation or multiplication are improved with symmetry-preserving techniques. The numerical results completely eliminate the symmetry errors even though the numerical dissipation is extremely low.

**References**

[1] S. K. Godunov, A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics, *Math. Sb.* **47** (1959) 271–306.

[2] A. Harten, High Resolution Schemes for Hyperbolic Conservation Laws, *International Journal for Numerical Methods in Fluids* **23** (1983) 309–323.

[3] B. Van Leer, Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, *Journal of Computational Physics* **23** (1977) 276–299.

[4] X.-D. Liu, S. Osher, T. Chan, Weighted Essentially Non-oscillatory Schemes, *Journal of Computational Physics* **115** (1994) 200–212.

[5] G.-S. Jiang, C.-W. Shu, Efficient implementation of weighted ENO schemes, *Journal of Computational Physics* **126** (1996) 202–228.

[6] Z. Sun, S. Inaba, F. Xiao, Boundary Variation Diminishing (BVD) reconstruction: A new approach to improve Godunov schemes, *Journal of Computational Physics* **322** (2016) 309–325.

[7] X. Deng, B. Xie, R. Loubère, Y. Shimizu, F. Xiao, Limiter-free discontinuity-capturing scheme for compressible gas dynamics with reactive fronts, *Computers and Fluids* **171** (2018) 1–14.

[8] X. Deng, Y. Shimizu, F. Xiao, A fifth-order shock capturing scheme with two-stage boundary variation diminishing algorithm, *Journal of Computational Physics* **386** (2019) 323–349.

[9] X. Deng, Y. Shimizu, B. Xie, F. Xiao, Constructing higher order discontinuity-capturing schemes with upwind-biased interpolations and boundary variation diminishing algorithm, *Computers and Fluids* **200** (2020) 104433.

[10] F. Xiao, Y. Honma, T. Kono, A simple algebraic interface capturing scheme using hyperbolic tangent function, *International Journal for Numerical Methods in Fluids* **48** (2005) 1023–1040.

[11] F. Xiao, S. Ii, C. Chen, Revisit to the THINC scheme: A simple algebraic VOF algorithm, *Journal of Computational Physics* **230** (2011) 7086–7092.

[12] B. Xie, F. Xiao, Toward efficient and accurate interface capturing on arbitrary hybrid unstructured grids: The THINC method with quadratic surface representation and Gaussian quadrature, *Journal of Computational Physics* **349** (2017) 415–440.

[13] J. F. Remacle, J. E. Flaherty, M. S. Shephard, An adaptive discontinuous Galerkin technique with an orthogonal basis applied to compressible flow problems, *SIAM Review* **45** (2003) 53–72.

[14] N. Fleischmann, S. Adami, N. A. Adams, Numerical symmetry-preserving techniques for low-dissipation shock-capturing schemes, *Computers and Fluids* **189** (2019) 94–107.

[15] S. Pirozzoli, On the spectral properties of shock-capturing schemes, *Journal of Computational Physics* **219** (2006) 489–497.

[16] J. Qiu, C. Shu, On the construction, comparison, and local characteristic decomposition for high-order central WENO schemes, *Journal of Computational Physics* **183** (2002) 187-209.

[17] E. F. Toro, M. Spruce, W. Speares, Restoration of the contact surface in the HLL-Riemann solver, *Shock Waves* **4** (1994) 25–34.

[18] E. F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, *Elsevier B.V.*, 2009.

[19] S. Gottlieb, On High Order Strong Stability Preserving Runge-Kutta and Multi Step Time Discretizations, *Journal of Scientific Computing* **25** (2005) 105–128.

[20] Z. Xu, C. W. Shu, Anti-diffusive flux corrections for high order finite difference WENO schemes, *Journal of Computational Physics* **205** (2005) 458–485.

[21] J. Shi, Y. T. Zhang, C. W. Shu, Resolution of high order WENO schemes for complicated flow structures, *Journal of Computational Physics* **186** (2003) 690–696.

[22] L. Fu, X. Y. Hu, N. A. Adams, A family of high-order targeted ENO schemes for compressible-fluid simulations, *Journal of Computational Physics* **305** (2016) 333–359.

[23] R. Liska, B. Wendroff, Comparison of several difference schemes on 1D and 2D test problems for the Euler equations, *SIAM Journal on Scientific Computing* **25** (2003) 995–1017.

# Extremely accurate solutions using block decomposition and extended precision for solving very ill-conditioned equations

**†Kansa, E.J.[1], Skala V.[2], and Holoborodko, P.[3]**

**[1]**Convergent Solutions, Livermore, CA 94550 USA
**[2]**Computer Science & Engineering Dept., Faculty of Applied Sciences, University of West Bohemia, University 8, CZ 301 00 Plzen, Czech Republic
**[3]**Advanpix LLC, Maison Takashima Bldg. 2F, Daimachi 10-15-201, Yokohama, Japan 221-0834
†Corresponding author: edwardjkansa@gmail.com

## Abstract

Many authors have complained about the ill-conditioning associated the numerical solution of partial differential equations (PDEs) and integral equations (IEs) using as the continuously differentiable Gaussian and multiquadric continuously differentiable ($C^\infty$) radial basis functions (RBFs).   Unlike finite elements, finite difference, or finite volume methods that lave compact local support that give rise to sparse equations, the $C^\infty$-RBFs with simple collocation methods give rise to full, asymmetric systems of equations. Since $C^\infty$ RBFs have adjustable constent or variable shape parameters, the resulting systems of equations that are solve on single or double precision computers can suffer from "ill-conditioning". Condition numbers can be either the absolute or relative condition number, but in the context of linear equations, the absolute condition number will be understood.   Results will be presented that demonstrates the combination of Block Gaussian elimination, arbitrary arithmetic precision, and iterative refinement can give remarkably accurate numerical salutations to large Hilbert and van der Monde equation systems.

## 1. Introduction

An accurate definition of the condition number, $\kappa$, of the matrix, **A**, is the ratio of the largest to smallest absolute value of the singular values, $\{\sigma_i\}$, obtained from the singular value decomposition (SVD) method, see [1]:

$$\kappa(\mathbf{A}) \quad = \max_j |\sigma_j| / \min_j |\sigma_j| | \qquad (1)$$

This definition of condition number will be used in this study.

Whenever the absolute condition number is comparable to the inverse of the machine epsilon, see [2, 3,4], the numerical results become unreliable.   In some instances, the absolute condition number may be exceptionally large, but the relative is small.

The important cause of ill-conditioning is the number of bits assigned to a computer word. The Institute of Electrical and Electronics Engineers (IEEE) defined a single precision word to have 8 bits and double precision word to have 16 bits, and (the double precision maximum condition number is O(1e16). Since scientific computing comprises a minuscule fraction of the user market, extended memory chips are unlikely, hence software methods are needed to obtain extended precision.

$C^\infty$ RBFs have the advantage of being deponently convergent and converges faster as the dimensions increases.   Socially important problems such as controlled fusion [3], designing new medical drugs, option markets, etc. requires numerical solutions of higher dimensional

PDEs and IEs made more difficult because operator splitting is not a viable approach. Because of the curse of dimensionality, a f the fewer the fewer the number of points per dimension the better.

The approach that we are advocating is using $C^\infty$ RBFs, shape parameters that make the RBFs flatter, minimizing the number of data centers, using extended arithmetic precision and employing block decomposition to obtain many smaller ranked block matrices, each of which is better conditioned

Assume the original matrix of rank N is subdivided into K blocks, each of which contains P points. Thus, the matrix is subdivided as:

| $\mathbf{A}_{1,1}$ | $A_{1,2}$ | $\mathbf{A}_{1,3}$ | $\mathbf{A}_{1,4}$ | $\cdots$ | $\mathbf{A}_{1,K}$ |
|---|---|---|---|---|---|
| $\mathbf{A}_{2,1}$ | $A_{2,2}$ | $A_{2,3}$ | $A_{2,4}$ | $\cdots$ | $\cdots$ |
| $\mathbf{A}_{3,1}$ | $A_{3,2}$ | $A_{3,3}$ | $A_{3,4}$ | $\cdots$ | $\cdots$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $\mathbf{A}_{K,1}$ | $\mathbf{A}_{K,2}$ | $\mathbf{A}_{K,3}$ | $\mathbf{A}_{K,4}$ | $\cdots$ | $\mathbf{A}_{K,K}$ |

**Figure 1. The full matrix A is partitioned to K blocks each of which contains P points.**

Each block, $\mathbf{A_{j,k}}$ is a square k×k, matrix that is a submatrix of the original N × N matrix, **A**.

There are various possible block decomposition methods available: block Gaussian decomposition, block singular value (SVD) decomposition, and block quotient- is a remainder (QR) decomposition. The simplest block decomposition method is the block Gaussian elimination method (BGEM) analog without pivoting. The BGEM is combined with extended arithmetic precision, and iterative refinement. The resulting block operations transform the original fully populated block matrix into identity block diagonal matrices and zero block matrices on the off-diagonal matrices. In additions, all operations are vectorized for maximum computational efficiency.

## 2. Example Test Problem of ill-conditioned linear equations

The test problem is the notoriously ill-conditioned Hilbert matrix whose elements are:

$$\mathbf{H}(i,j) = 1/(i+j-1), \tag{2}$$

where i is i-th row index and j is the column index. The Hilbert matrix is invertible to all orders, and the inverse Hilbert matrix exists. However, on a double precision computer, the condition number is O(1e16). A heuristic rule for ill-conditioning is that the condition number increase with rank of the matrix.

If the unknown vector, **x**. is specified as $\mathbf{x} = [1,1\ldots,1]^T$, the right vector, **b**, is found by multiplying

$$\mathbf{b} = \mathbf{Hx}. \tag{3}$$

271

The problem that is considered is to assume the right-hand-side, **b**, is known, and to find x

A more accurate definition of the condition number of the matrix, **A**, is the ratio of the largest to smallest absolute value of the singular values, $\{\sigma i\}$, obtained from the singular value decomposition (SVD) method:

$$\kappa(A) \quad = \max_j |\sigma_j|/\min_j |\sigma_j,| \tag{4}$$

This definition of condition number will be used.

## 3. Example calculations

The first set of numerical experiments examines the root mean square errors RMS)

$$\text{RMS error} = [\ \Sigma_i(x^{exact}_j - x^{numerical}_j)^2/N\ ]^{\frac{1}{2}} \tag{5}$$

The first set of examples is the block partitioning a 300×300 Hilbert matrix with 48 and 200 digits of extended precision with various sized blocks.
Table 1. Comparison of RMS errors of a 300×300 Hilbert matrix using different block sizes digits of precision.

**Table 1. RMS errors for block decomposition of a (300×300) Hilbert matrix with 48 and 200 digits of precision.**

| 48 digits | Block size | RMS error | 200 digits | Block size | RMS error |
|---|---|---|---|---|---|
| | 10 (30×30) | 0.73 | | 10 (30×30) | 7.1e-527 |
| | 15(20×20) | 2,72e-15 | | 15(20×20) | 3.9e-527 |
| | 20(15×15) | 2.19e-16 | | 20(15×15) | 4.3e-527 |
| | 30(10×10) | 1.95e-20 | | 30(10×10) | 6.5e-537 |
| | 60 (5×5) | 1.26e-24 | | 60 (5×5) | 4.9e-537 |

One last example was a 1000 × 1000 Hilbert matrix, with the right-hand vector being generated by the solution, x= [1,1….,1]$^T$ in which 50 (20×20) blocks in which 200 digits of precision were used.   Such a problem required so much memory that the run time required over 42hrs. The RMS error was 2.72e-126.
Because of the extremely slow execution speed observed with the 1000×1000 Hilbert matrix, no additional tests were conducted.
I some examples of the 300×300 Hilbert matrix, partitioning the original matrix into smaller blocks. The full the 300×300 Hilbert matrix, condition number is 1.1e20. For the following block sizes, the maximum block condition numbers are:

**Table 2: Block size and maximin condition number of a 300×300 Hilbert matrix**

| Block size | Maximum condition number |
|---|---|
| (20 ×20) | 2.23e+18 |
| ((15   × 15) | 2.59e+17 |
| (10 × 10) | 1.60e+13 |

An alternative method to the block Gaussian elimination method involving P blocks containing a uniform number of k elements per block, it the decomposition of a larger matrix into many $2 \times 2$ block matrices, see [10].

The next examples are the van der Monde matrices defined as a matrix of vectors raised to a power, commonly given by:

$$A(i,j) = v(j)^{(N-j)} \tag{6}$$

We considered the vector composed of N elements starting with 1 and increasing by increments of 0.5. Many other variations can be considered.

We considered a vector of 100 elements ranging from 1.0 to 50.5, incremented by 0.5. The van der Monde matrix of rank 100 of such a vector is an estimated condition number of 8e+202. With 360 digits of precision, we considered 25 (4x4) blocks with a RMS error of 4e-44, and with 50 (2x2) blocks, the RMS error is 4e-95. So even with a horribly conditioned van der Monde matrix, we are still able to obtain very accurate RMS errors.

## 4. Discussion

Traditional domain decomposition methods are intrinsically iterative in nature whether they are overlapping or non-overlapping methods. A large domain over a large PDE or IE problems is subdivided into many smaller sub-domains. Artificial boundaries are constructed on which artificial boundary conditions are imposed, and the solution is obtained iteratively. For elliptic or time dependent problems for diffusion or viscosity dominates, the iterative matching of the function and normal and tangential derivatives can achieve a sufficient degree of convergence. The basic problem is that there is not a sufficient number of equations for the number of unknowns to enforce all the derivative continuity conditions.

For advective dominated problems, convergence is more of a problem unless one is willing to sacrifice physics to additional numerical viscosity by way of up wind spatial differencing. Domain decomposition methods embody large potential for a parallelization of the finite difference, finite element, or finite volume methods for distributed, parallel computations.

A non-iterative domain decomposition method was introduced in which large submatrices were solved individually and potentially in parallel with linear equation matching at surfaces, lines, and points, see ]11]. Because of the different sizes of each type of sub-domain, parallelization is hampered.

Additional acceleration can be obtained by classical iterative refinement, and the geometric projective algebra, see [12].

These examples illustrate the well-known fact that the condition number depends upon the rank of the system of equations, and the number of digits available. However, as the numbers in the set of equations approaches the ideal Platonic limit of infinity, the time required also approaches infinity. The computation of important applications such as plasma fusion, designer medicine based upon the first principles of quantum mechanics, etc. will involve multi-dimensional calculations and will require us to expand our vision beyond horse blinders.

Literature Cited

**1.** Cline, A.K.; Miler, C.B.; Stewart; G.W.; Wilkinson., J.H, "An Estimate for the Condition Number of a Matrix",, *SIAM J. Numer. Anal.*, V. **16**(2), pp. 368–375, 1977.

2. Kansa, E.J.; Holoborodko," On the ill-conditioned nature of C$^\infty$ **RBF** strong collocation", E*ng. Anal. Bound. Elem*. vol. **78,** pp. 26–30, 2017.

3. https://www.advanpix.com.

4. Kansa, E.J.; Holoborodko: "Fully and sparsely supported radial basis functions", *Int. J. Comput. Meth. Exper, Measur.,* V. **8**(3)m pp208-219 ,2020.

5. Chen. F.E., *Introduction to Plasma Physics and controlled Fusion,* 3$^{rd}$ edition, ISSN 978-3-319-79391-7, Springer, Heidelberg, 2015.

6. Kirk, D, B, Hwu, W-M.W., *Programming Massively Parallel Computers, Elsevier*, Amsterdam ISBN 978-0- 12-811986-0, 2017.

7. Hilbert, D.: Ein Beitrag zur Theorie des Legendre'schen Polynoms", *Acta Mathematica*,  V. **18**: pp.155–159, 1893.

8. Choi, M-D, *"*Tricks or Treats with the Hilbert Matrix*". The American Mathematical Monthly*., V.**90** (5): pp. 301–312, 1983.

9. Salam.: Conditionality of Linear Systems of Equations and Matrices Using Projective Geometric Algebra, ICCSA 2020 proceedings, Part II, LNCS 12250, pp. 3-17, DOI: 10.1007/978-3-030-58802-1_1, Springer, 2020).

10. Lu, T-T; Shiou, S-H, "Inverses of 2x2 block matrices", Comput. *Math. Applic*., V**43**, pp. 119-129, 2002.

11. Kansa E, Hon Y "Circumventing the ill-conditioning problem with multiquadric radial basis functions: Applications to elliptic partial differential equations" *Computers & Mathematics with Applications*, V.. **39**, (7-8) pp. 123-137, 2000.'

12. kala, V., "Conditionality of linear systems of equations and matrices using projective geometric algebra", ICCSA 2020 Proc., Part II LNCS 12250, pp. 3-17, DOI: 10.1007/987-4-030-53603-2,2, Springer, 2020.

# Application of boundary face method in cathodic protection problems

**†Sanshan Tu[1,2], Yanjie Wei[1], Shengzhong Feng[2], and Zhongyang Dai[2]**

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
[2]National Supercomputing Center in Shenzhen, China
†Corresponding author: tss71618@163.com

**Abstract**

The cathodic protection problems are mostly simulated by boundary element method. However, boundary element method consumes a large amount of computer memory and is very inefficient when it is used to simulate large-scale models. Thus, we developed the boundary face method to simulate the cathodic protection problems to improve the efficiency and accuracy. The cathodic protection problems of a subsea pipeline and a tension-leg platform (TLP) are simulated. The results show that BFM is more accurate than BEM and costs less time when the same precision is achieved in most cases.

**Keywords:** Cathodic Protection; Numerical Simulation; Boundary Element Method; Boundary Face Method;

**Introduction**

Cathodic protection (CP) plays an important role in protecting offshore structures from corrosion. More and more numerical simulations are employed in designing and evaluating cathodic protection systems with the development of computers and numerical methods. The CP problem is governed by Laplace's equation, and mostly simulated by boundary element method (BEM) [1-3]. However, BEM consumes a large amount of computer memory and is very inefficient when it is used to simulate large-scale models.

The meshless methods have been used in many engineering problems such as fluid mechanics problems [4], solid mechanics problems [5, 6], elasticity problems [7,18], viscoelasticity problems [8], heat conduction problems [9, 23] and elastodynamic problems [17, 19].

Boundary integral equation (BIE) based meshless methods are an important part of meshless methods. The boundary node method (BNM) is firstly proposed by Mukherjee et al. [10-13] based on the moving least square (MLS) method [14] and BIE. Unlike many other 'domain' type meshless methods, the BNM only requires scattered nodes on the 1-D bounding curve of a 2-D area or the 2-D bounding surface of a 3D body and a simple boundary cell structure for numerical integration. The idea of BNM has been developed by many other researchers such as Zhu et al. [15,16], Liew et al. [17-19], Ren et al. [20, 21] and Li et al. [22] with different approximating methods.

Zhang et al. developed the Boundary Face Method (BFM) [24-26] based on BNM to solve more complex geometrical problems. It uses MLS to approximate boundary field variables with scattered points on the boundary and uses background grids for numerical integration. However, it is different from BNM in specific implementation and is the inheritance and development of BNM. BFM is based on CAD geometric modeling data, and can be seamlessly integrated with the CAD system, so that it can handle more complex physical problems.

Approximating methods play an important role in meshless methods. The moving Kriging interpolation (MKI) method [20], also known as the radial point interpolation method (RPIM) [2], is an important approximating method to construct shape functions in meshless methods. MKI has partition of unity property, consistency property and high approximation precision. Besides, the shape functions constructed by MKI possess Kronecker delta property. Then, the essential boundary condition can be imposed directly and easily. Li et al. [19] have proposed the moving Kriging interpolation-based boundary node method (MKIBNM) by combing moving Kriging interpolation (MKI) with BIE for potential problems. MKIBNM has a high precision and can directly impose boundary conditions. Therefore, we chose MKI to construct shape functions in BFM.

The BFM is developed to simulate the CP problem. The CP problems of a subsea pipeline and a tension-leg platform (TLP) will be simulated to show that BFM can be applied in cathodic protection problems to improve the efficiency and accuracy.

## Boundary face method for cathodic protection problems

*Governing equations*

The equation governing the potential distribution and the current flow in the electrolyte can be derived from charge conservation. The continuity equation requires that the current per unit volume, $J$, relates to the charge, $q$, by

$$\nabla J = \frac{\partial q}{\partial t} \tag{1}$$

Taking into account the relationship of electric field intensity, $E$,

$$E = -\nabla \phi \tag{2}$$

and Ohms law,

$$I = \sigma E \tag{3}$$

where $\sigma$ is the conductivity of the electrolyte, the continuity equation transforms to

$$\nabla(\sigma \nabla \phi) = -\nabla(\frac{\partial q}{\partial t}) \tag{4}$$

Galvanic corrosion is a very slow process, thus we can make the following assumptions:
1. The electrolyte solution is well mixed that the conductivity is isotropic, $\sigma$ is a constant.

2. The solution is electro-neutral, $\frac{\partial q}{\partial t} = 0$.

With the above assumptions, Eq. (4) can be simplified as,

$$\nabla^2 \phi = 0 \tag{5}$$

Therefore, for a uniform, isotropic electrolyte, the potential obeys the Laplace equation.

*Boundary condition*

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = 0$$

Electrolyte

$$\nabla^2 \phi = 0$$

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = 0 \qquad\qquad \frac{\partial \phi}{\partial \boldsymbol{n}} = 0$$

Cathode          Anode

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = -\frac{f_c(\phi)}{\sigma} \qquad \frac{\partial \phi}{\partial \boldsymbol{n}} = -\frac{f_a(\phi)}{\sigma}$$
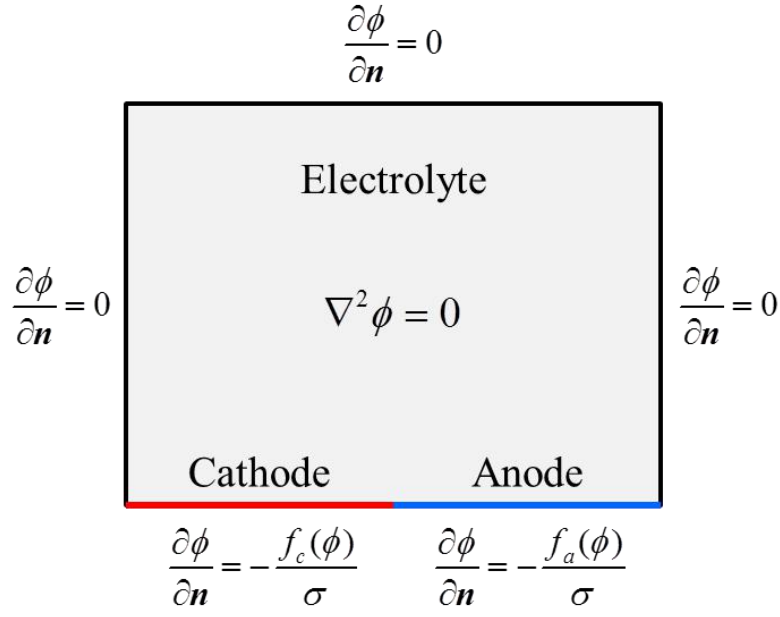
**Figure 1. Schematic of the mathematical model of cathodic protection problems**

Eq. (5) can be solved with the boundary conditions shown schematically in Figure 1. The boundary conditions at the anode and the cathode surfaces are vital to predict the state of the cathodic protection. The boundary condition applied at the anode surface $\Gamma_a$ is:

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = -\frac{j}{\sigma} = -\frac{f_a(\phi)}{\sigma} \qquad\qquad (6)$$

Where, $\sigma$ is the electrical conductivity of the electrolyte solution and $f_a(\phi)$ is the current density determined by anodic species. $f_a(\phi)$ is a piecewise linear interpolation function which is obtained from the polarization curve of the anodic species. Thus, we use a piecewise linear interpolation approach to handle non-linear boundary conditions.

Similarly, the boundary condition applied at the cathode surface $\Gamma_c$ is

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = -\frac{f_c(\phi)}{\sigma} \qquad\qquad (7)$$

The boundary condition applied at the insulation surface and the electrolyte-air interface $\Gamma_{ins}$ is

$$\frac{\partial \phi}{\partial \boldsymbol{n}} = -\frac{j}{\sigma} = 0 \qquad\qquad (8)$$

*BFM formulation*

In summary, the governing equation and the boundary conditions are

$$\begin{cases} \Delta\phi = \dfrac{\partial^2\phi}{\partial x^2} + \dfrac{\partial^2\phi}{\partial y^2} = 0 & \text{in } \Omega \\[2mm] \dfrac{\partial\phi}{\partial \boldsymbol{n}} = -\dfrac{f_a(\phi)}{\sigma} & \text{on } \Gamma_a \\[2mm] \dfrac{\partial\phi}{\partial \boldsymbol{n}} = -\dfrac{f_c(\phi)}{\sigma} & \text{on } \Gamma_c \\[2mm] \dfrac{\partial\phi}{\partial \boldsymbol{n}} = 0 & \text{on } \Gamma_{ins} \end{cases} \tag{9}$$

where $\phi$ is the potential, $j = -\sigma\dfrac{\partial u}{\partial \boldsymbol{n}}$ is the current density along the normal of the boundary, the whole boundary is $\Gamma = \partial\Omega = \Gamma_a + \Gamma_c + \Gamma_{ins}$. The integral representation of the solution for Eq. (9) is

$$c(\boldsymbol{\xi})\phi(\boldsymbol{\xi}) + \int_\Gamma \frac{\partial\phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})}{\partial \boldsymbol{n}}\phi(\boldsymbol{x})\mathrm{d}\Gamma = \frac{-1}{\sigma}\int_\Gamma \phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})j(\boldsymbol{x})\mathrm{d}\Gamma \tag{10}$$

where $\boldsymbol{\xi}$ denotes the given source point on the boundary,

$\boldsymbol{x}$ denotes a filed point on the boundary,

$\phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi}) = \dfrac{1}{4\pi|\boldsymbol{x}-\boldsymbol{\xi}|}$ is the fundamental solution of Laplace's equation,

$\dfrac{\partial\phi^*}{\partial \boldsymbol{n}}$ is the normal derivative of $\phi^*$ on the boundary,

$c(\boldsymbol{\xi})$ is a coefficient related to the boundary smoothness.

The boundary $\Gamma$ is divided into background cells $\Gamma_i \ (i = 1,2,3,\mathsf{L}\ ,m)$ for numerical integration, and the discrete form of Eq. (10) is

$$c(\boldsymbol{\xi})u(\boldsymbol{\xi}) + \sum_{i=1}^{m}\int_{\Gamma_i} \frac{\partial\phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})}{\partial \boldsymbol{n}}\phi(\boldsymbol{x})\mathrm{d}\Gamma = \frac{-1}{\sigma}\sum_{i=1}^{m}\int_{\Gamma i} \phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})j(\boldsymbol{x})\mathrm{d}\Gamma \tag{11}$$

$u(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are approximated by MKI.

$$\begin{cases} u = \boldsymbol{N}(s,t)\boldsymbol{\Phi} = \displaystyle\sum_{k=1}^{n}\varphi_k(s,t)\phi_k \\[3mm] j = \boldsymbol{N}(s,t)\boldsymbol{Q} = \displaystyle\sum_{k=1}^{n}\varphi_k(s,t)j_k \end{cases} \tag{12}$$

Substituting Eq. (12) into Eq. (11) yields

$$c(\boldsymbol{\xi})\phi(\boldsymbol{\xi}) + \sum_{i=1}^{m}\int_{\Gamma_i} \frac{\partial\phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})}{\partial \boldsymbol{n}}\boldsymbol{N}(s,t)\boldsymbol{\Phi}\mathrm{d}\Gamma = \frac{-1}{\sigma}\sum_{i=1}^{m}\int_{\Gamma_i} \phi^*(\boldsymbol{x}\text{-}\boldsymbol{\xi})\boldsymbol{N}(s,t)\boldsymbol{J}\mathrm{d}\Gamma \tag{13}$$

Employing numerical methods for the integrals in Eq. (13), at every nodes, we can obtain the linear algebraic equations.

$$C\boldsymbol{\Phi} + \hat{H}\boldsymbol{\Phi} = GJ \tag{14}$$

where

$$C = diag(c(\boldsymbol{\xi}_1), c(\boldsymbol{\xi}_2), \mathsf{L}, c(\boldsymbol{\xi}_n)) \tag{15}$$

$$\hat{\boldsymbol{H}}_i = \sum_{k=1}^{m} \int_{\Gamma_k} \frac{\partial \phi^*(\boldsymbol{x} - \boldsymbol{\xi}_i)}{\partial \boldsymbol{n}} \boldsymbol{N}(s, t) \mathrm{d}\Gamma \tag{16}$$

$$\boldsymbol{G}_i = \frac{-1}{\sigma} \sum_{k=1}^{m} \int_{\Gamma_k} \phi^*(\boldsymbol{x} - \boldsymbol{\xi}_i) \boldsymbol{N}(s, t) \mathrm{d}\Gamma \tag{17}$$

Let

$$\boldsymbol{H} = \boldsymbol{C} + \hat{\boldsymbol{H}} \tag{18}$$

Eq. (14) can then be rewritten as

$$\boldsymbol{H\Phi} = \boldsymbol{GQ} \tag{19}$$

Normally, we do not directly compute $\boldsymbol{C}$, because the diagonal elements of $\boldsymbol{H}$ can be computed by constant potential method.

$$H_{ii} = -\sum_{k=1, k \neq i}^{n} H_{ik} \tag{20}$$

Finally, we can solve Eq. (19) with the boundary conditions in Eq. (9) and obtain the nodal values of potential and current density on the boundary.

**Case study**

In the following, we will use two numerical simulation examples of cathodic protection of marine structures to study the calculation accuracy and efficiency of the method. The polarization curve is an important boundary condition in the numerical simulation of cathodic protection. The empirical polarization curves of the steel structure, zinc sacrificial anode and auxiliary anode used in simulation are as follows ( $q_d$ is the demand current density of the protected structure in Table 1):

*Galvanic anode cathodic protection for submarine pipelines*

This example is the cathodic protection problem of a submarine pipeline (See Figure 2). The demand current density is $20\mathrm{mA/m}^2$. At the end of the protection, the anode is exhausted, which is the most dangerous working condition. Therefore, the protection state at the end is simulated, and the thickness of the anode is approximately considered to be 0 during the simulation.

At first, we use BEM and BFM with 380 nodes to simulate the problem respectively. The protection potential distribution is shown in Figure 4. The potential range of the submarine pipeline simulated by BEM is -991~-800mV, and the protection potential of the submarine pipeline simulated by BFM is -990. ~-794mV, the accurate value is -980~-789mV (from BEM model with 81122 nodes), the calculation accuracy of BFM is higher than that of BEM, and the potential distribution obtained by BFM simulation is smoother and more consistent with the accurate solution.

Then, we change the mesh size to compare the calculation accuracy and efficiency of the two methods. The results under different number of nodes are shown in Table 4. Since the cathodic protection problem generally has no analytical solution, it is impossible to directly

calculate the solution error. Here, we compare the calculation accuracy and calculation efficiency by comparing the number of nodes and calculation time used by the two methods to obtain the convergence results (in the bold part of the table). It can be seen from Table 4 that the calculation accuracy and convergence of BFM are better than that of BEM. When the same calculation accuracy is reached, the calculation efficiency of BFM is higher than that of BEM, and the number of nodes used is less than that of BEM (that is, the memory consumption of BFM is less than that of BEM).
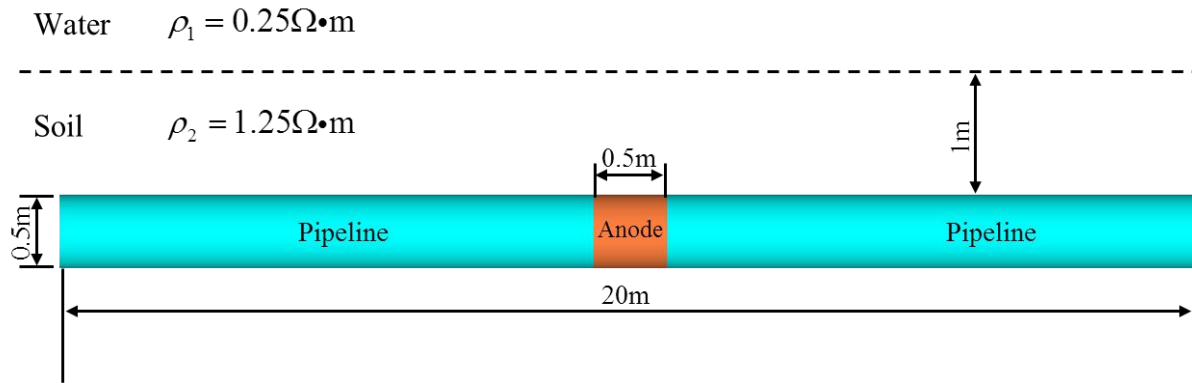


**Figure 2. Subsea pipeline**
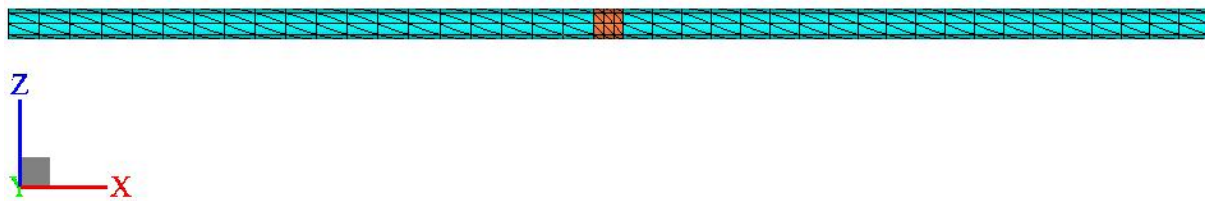


**Figure 3. Mesh of the subsea pipeline**
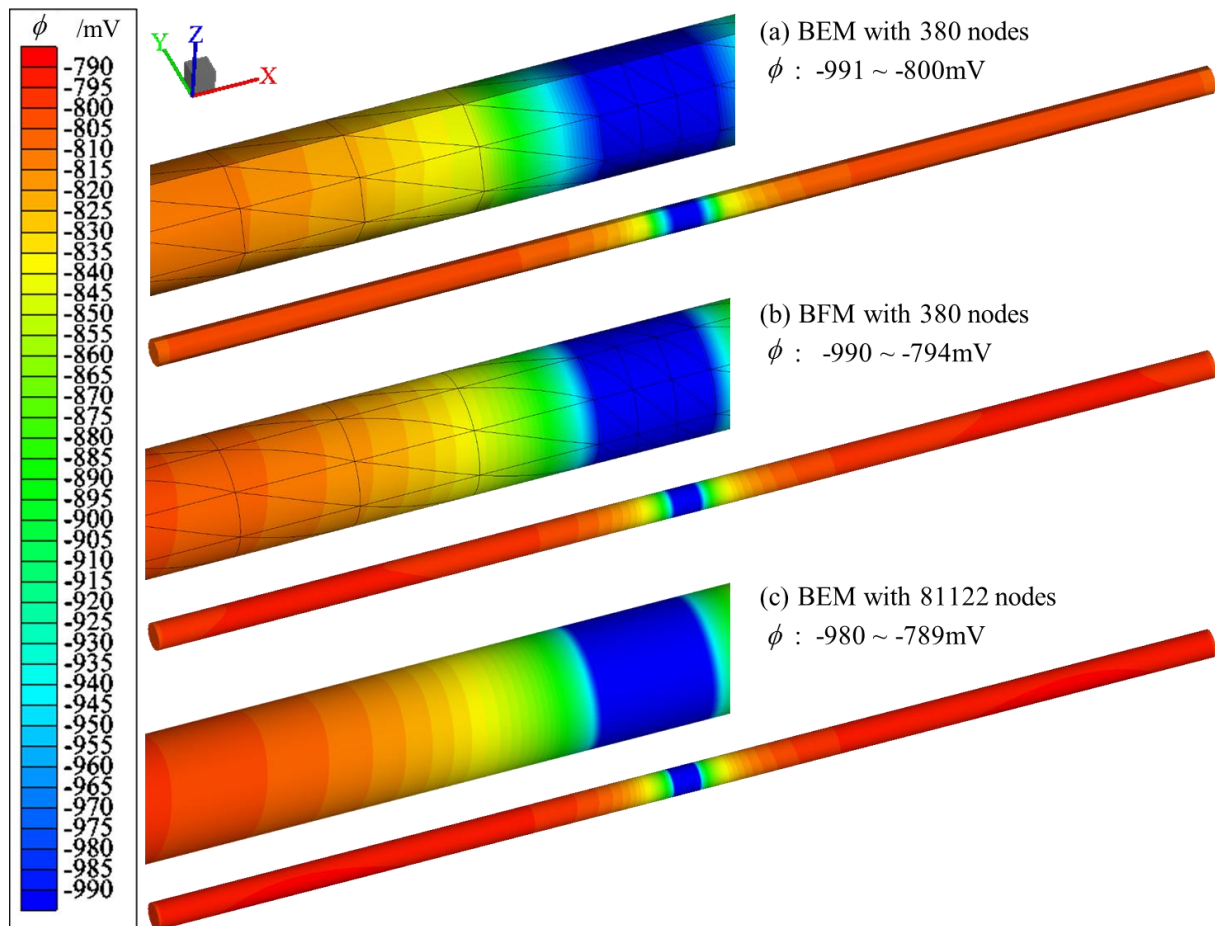
**Figure 4. Potential distribution on the surface of the subsea pipeline: (a) BEM with 380 nodes, (b) BFM with 380 nodes and (c) BEM with 81122 nodes**

**Table 4. Simulation results of the subsea pipeline with different mesh sizes**

| Nodes | BEM | | BFM | |
|---|---|---|---|---|
| | $\phi$ | CPU time | $\phi$ | CPU time |
| | mV | s | mV | s |
| 156 | -995~-814 | 0.045 | -992~-795 | 2.6 |
| 380 | -991~-800 | 0.18 | -990~-794 | 11 |
| 2720 | -986~-793 | 1.8 | -984~-791 | 99 |
| 7011 | -985~-791 | 52 | **-980~-789** | **658** |
| 13058 | -983~-790 | 279 | -980~-789 | 1698 |
| 20403 | -982~-790 | 1251 | | |
| 81122 | **-980~-789** | **41631** | | |

*Impressed current cathodic protection for tension leg platform*

This example is the cathodic protection of the underwater part of the tension leg platform (See Figure 5). The surface is covered by anticorrosive coating, and the design life of the anticorrosion is 20 years. The terminal protection status was simulated, and the demand current density of the terminal protection is $25.22 \text{mA/m}^2$. A total of twelve ICCP anodes were arranged, the radius of the anode screen was 3.5m (the pink area in Figure 5), the ICCP anodes are in the center of the anode screen, and the output current of the anodes are 400A.

The mesh of the model is shown in Figure 5(c), with a total of 1756 nodes. The results are shown in Figure 6. The protection potential range simulated by BEM is -1107~-833mV, the

protection potential simulated by BFM is -1115~-830mV, and the accurate value is -1115~-830mV. (From BEM model with 66812 nodes)

The results under different number of nodes are shown in Table 5. It can be seen that the calculation accuracy and convergence of BFM are better than BEM. When the same calculation accuracy is reached, the number of nodes used by BFM is less than that of BEM, and the calculation efficiency of BFM is better.



(a)     (b)     (c)
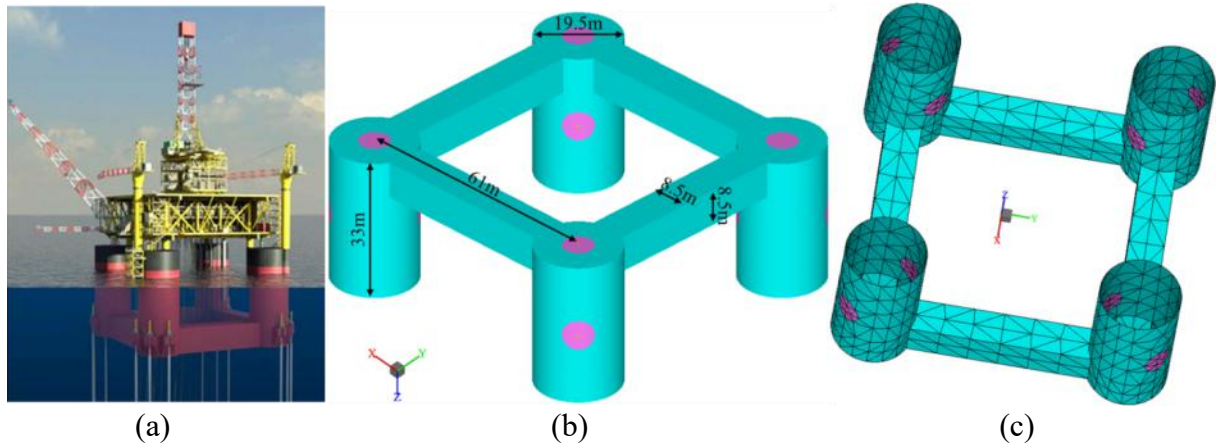
**Figure 5. Model of the TLP and arrangement of the ICCP anodes: (a) Concept model, (b) Size of the underwater part and (c) Mesh model**



(a)     (b)     (c)
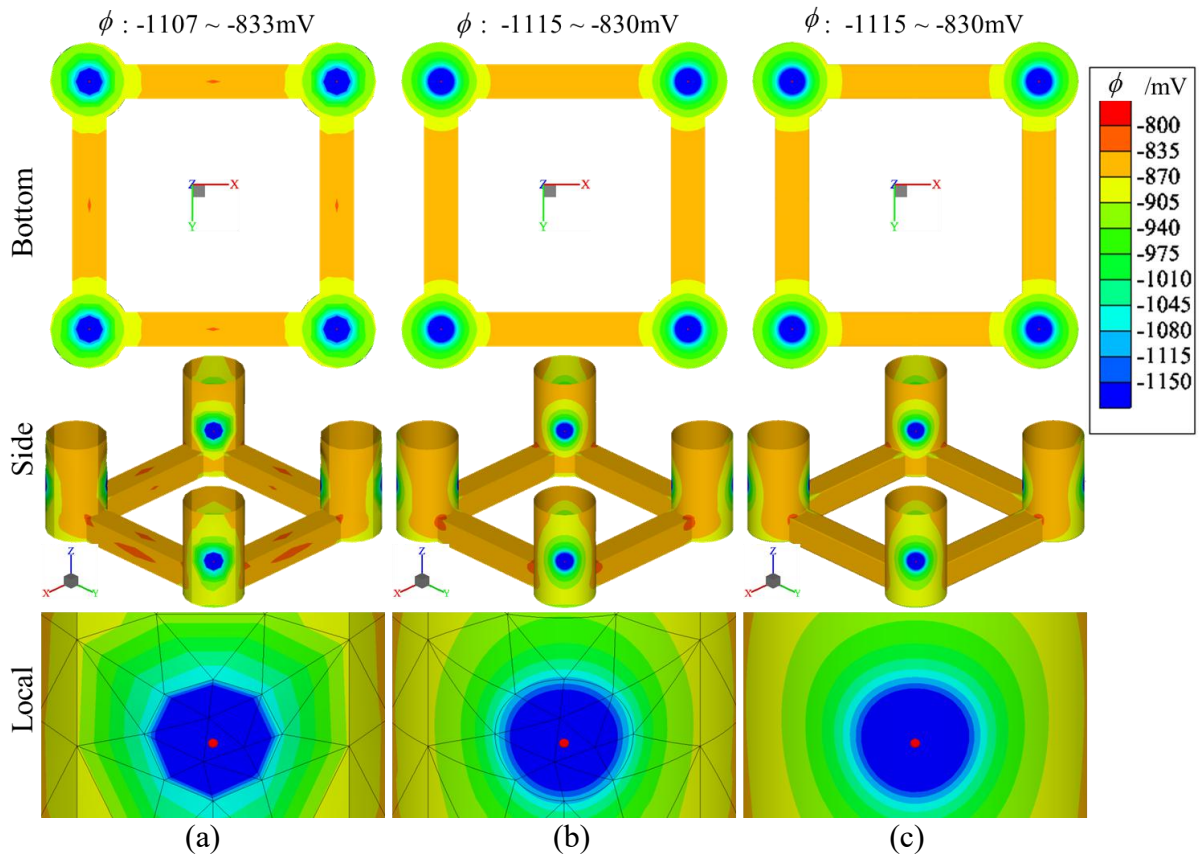
**Figure 6. Simulation results of the TLP model: (a) BEM with 1756 nodes, (b) BFM with 1756 nodes and (c) BEM with 66812 nodes**

**Table 4. Simulation results of the subsea pipeline with different mesh sizes**

| Nodes | BEM | | BFM | |
|---|---|---|---|---|
| | $\phi$ | CPU time | $\phi$ | CPU time |
| | mV | s | mV | s |
| 1756 | -1107~-833 | 2.4 | -1115~-830 | 24 |
| 3233 | -1113~-826 | 11.0 | -1114~-830 | 92 |
| 5840 | -1113~-830 | 38.0 | **-1115~-830** | **291** |
| 18579 | -1114~-829 | 569.0 | -1115~-830 | 2968 |
| 66812 | **-1115~-830** | **17888** | | |

## Conclusions

The BFM is developed to simulate the CP problem. The CP problems of a subsea pipeline and a tension-leg platform (TLP) are simulated. The results show that BFM is more accurate than BEM and costs less time when the same precision is achieved in most cases. Therefore, BFM can be applied in cathodic protection problems to improve the efficiency and accuracy.

**References**

[1] DeGiorgi V. G.. Evaluation of perfect paint assumptions in modeling of cathodic protection systems[J]. Engineering Analysis with Boundary Elements, 2002, 26(5):435-445.
[2] DeGiorgi V. G., Wimmer S. A., Hogan E., et al. Modeling the experimental environment for shipboard ICCP systems[J]. WIT Transactions on Modelling and Simulation, 2002, 32:439-447.
[3] DeGiorgi V. G., Wimmer S. A. Geometric details and modeling accuracy requirements for shipboard impressed current cathodic protection system modeling[J]. Engineering Analysis with Boundary Elements, 2005, 29(1):15-28.
[4] Liu, G. R. and Liu, M. B. (2003) Smoothed Particle Hydrodynamics — A Meshfree Particle Method, World Scientific, Singapore.
[5] Cui, X. Y., Feng, H., Li, G. Y. and Feng, S. Z. (2015) A cell-based smoothed radial point interpolation method (CS-RPIM) for three-dimensional solids, Engineering Analysis with Boundary Elements 50, 474-485
[6] Zheng, B. J. and Dai, B. D. (2011) A meshless local moving Kriging method for two-dimensional solids, Applied Mathematics and Computation 218, 563-573
[7] Peng, M. J., Liu, P. and Cheng, Y. M. (2009) The Complex Variable Element-Free Galerkin (Cvefg) Method for Two-Dimensional Elasticity Problems, International Journal of Applied Mechanics 1, 367-385
[8] Cheng, Y. M., Li, R. X. and Peng, M. J. (2012) Complex variable element-free Galerkin method for viscoelasticity problems, Chinese Physics B 21, 090205
[9] Deng, Y. J., Liu, C., Peng, M. J. and Cheng, .Y. M. (2015) The Interpolating Complex Variable Element-Free Galerkin Method for Temperature Field Problems, International Journal of Applied Mechanics 7, 1550017
[10] Mukherjee, Y. X. and Mukherjee, S. (1997) The boundary node method for potential problems, International Journal for Numerical Methods in Engineering 40, 797-815
[11] Chati, M. K., Mukherjee, S. and Mukherjee, Y. X. (1999) The boundary node method for three-dimensional linear elasticity, International Journal for Numerical Methods in Engineering 46, 1163-1184
[12] Kothnur, V. S., Mukherjee, S. and Mukherjee, Y. X. (1999) Two-dimensional linear elasticity by the boundary node method, International Journal of Solids and Structures 36, 1129-1147
[13] Chati, M. K. and Mukherjee, S. (2000) The boundary node method for three-dimensional problems in potential theory, International Journal for Numerical Methods in Engineering 47, 1523-1547
[14] Lancaster, P. and Salkauskas, K. (1981) Surfaces generated by moving least squares methods, Mathematics of Computation 37, 141-158
[15] Zhu, T., Zhang, J. and Atluri, S. N. (1998) A meshless local boundary integral equation (LBIE) method for solving nonlinear problems, Computational Mechanics 22, 174-186
[16] Zhu, T., Zhang, J. D. and Atluri, S. N. (1998) A local boundary integral equation (LBIE) method in computational mechanics, and a meshless discretization approach, Computational Mechanics 21, 223-235
[17] Liew, K. M., Cheng, Y. M. and Kitipornchai, S. (2005) Boundary element-free method (BEFM) for two-dimensional elastodynamic analysis using Laplace transform, International Journal for Numerical Methods in Engineering 64, 1610-1627

[18] Liew, K. M., Cheng, Y. M. and Kitipornchai, S. (2006) Boundary element-free method (BEFM) and its application to two-dimensional elasticity problems, International Journal for Numerical Methods in Engineering 65, 1310-1332

[19] Liew, K. M. and Cheng, Y. M. (2009) Complex variable boundary element-free method for two-dimensional elastodynamic problems, Computer Methods in Applied Mechanics and Engineering 198, 3925-3933

[20] Ren, H. P., Cheng, Y. M. and Zhang, W. (2009) An improved boundary element-free method (IBEFM) for two-dimensional potential problems, Chinese Physics B 18, 4065-4073

[21] Ren, H. P., Cheng, Y. M. and Zhang, W. (2010) An interpolating boundary element-free method (IBEFM) for elasticity problems, Science China-Physics Mechanics & Astronomy 53, 758-766

[22] Li, X. G., Dai, B. D. and Wang, L. H. (2010) A moving Kriging interpolation-based boundary node method for two-dimensional potential problems, Chinese Physics B 19, 120202

[23] Gu, L. (2003) Moving Kriging interpolation and element-free Galerkin method, International Journal for Numerical Methods in Engineering 56, 1-11

[24] Zhang J M, Qin X Y, Han X, et al. A boundary face method for potential problems in three dimensions[J]. Int J Numer Methods Eng, 2009, 80(3): 320-337.

[25] Qin X Y, Zhang J M, Li G Y, et al. An element implementation of the boundary face method for 3D potential problems[J]. Eng Anal Bound Elem, 2010, 34(11): 934-943.

[26] Zheng B J, Dai B D. A meshless local moving Kriging method for two-dimensional solids[J]. Appl Math Comput, 2011, 218(2): 563-573.

# Economic Parametric Optimization and Uncertainty Analysis in Ship Design using Monte Carlo Simulations

**†*Yuanhang Hou[1], Chong Fu[2], and Yeping Xiong[3]**

[1]Department of Naval architecture and Ocean Engineering, Dalian Maritime University, China.

[2]Department of Naval architecture and Ocean Engineering, Dalian University of Technology, China

[3]Faculty of Engineering and Physical Sciences, University of Southampton, UK

*Presenting author: houyuanhang@dlmu.edu.cn

†Corresponding author: houyuanhang@dlmu.edu.cn

## Abstract

Economic efficiency must be considered in ship concept design. There are many uncertain internal and external factors in the ship design process. This paper concentrates on the optimization of a ship's economic performance while considering the influence of uncontrollable factors on the output response. Firstly, the economic-objective function and the mathematical optimization model of a bulk carrier are established, and design space and constraints are proposed. Secondly, two algorithms are adopted to perform deterministic multi-objectives optimization. Thirdly, sensitivity analysis of the design parameters is conducted as well as the output response uncertainty analysis based on Monte Carlo simulations. The results reveal that, when random variables obey a specific distribution, the corresponding distribution of uncertainty effects will also exist in the output response. Therefore, the necessity of uncertainty analysis in parametric ship concept design is verified.

**Keywords:** Ship design, economic, Optimization design, Uncertainty analysis, Monte Carlo simulation

## 1 Introduction

The economy of a ship, which is one of its most important properties, is usually set as a design objective in the concept optimization design. Nowadays, research of ship concept design and hull form optimization has accumulated many achievements involving various design objectives, design variables, and optimization systems [1-3], which embody the development ideas of 'from simple to complex' and 'from coarse to fine'.

There are various internal and external parameters that cannot be precisely described or obtained in the process of ship optimization design. These parameters usually participate in calculations as constants. Inevitably, these parameters fluctuate all the time according to probability distributions. This fluctuation makes the output response uncertain. Since this uncertainty would be magnified by continuous iterative optimization, the influence of these parameters has a practical significance on ship optimization design. In recent years, Diez introduced uncertainty optimization design to ship hull design systems, and a series of studies were conducted [4-6]. Diez [7] considered the uncertainty of the economic parameters of bulk carriers, and a robust optimization study was carried out. However, the uncertain parameters were only expressed in interval form; probability distributions and responses to the output have not yet been studied.

In this paper, the bulk carrier conceptual design tool by Sen and Yang [8] is referenced and redefined. An economic objective function and its mathematical model of the ship are established, and design space and constraint conditions are defined. Two optimization algorithms are adopted to conduct the economic multi-objectives optimization calculation. Pointing to those internal parameters with random characteristics, sensitivity analysis and

uncertainty analysis based on Monte Carlo simulations are carried out. As a result, the response relationship between system output and random variables is obtained, which can be used to guide future ship optimization design.

## 2 Establishment of optimization model

### 2.1 Optimization objective and derivation

The optimization function in this study can be divided into two parts: (a) a mathematical model of hull cost based on the ship's principal dimensions and form coefficient and (b) the economic model of overall cargo shipping considering the other factors in operation.

Hull cost can be calculated based on steel weight, outfitting weight, and main power:

$$C_s = 1.3(2000W_h^{0.85} + 3500W_f + 2400P^{0.8}) \tag{1}$$

In this equation, $C_s$ is the hull cost (pounds); $W_h$ and $W_f$ are the steel weight and outfitting weight (t), respectively; and $P$ is the main power of the ship (kW), which is calculated as follows:

$$\begin{cases} P = (\Delta^{(2/3)}V_k^3)/(a + b \cdot Fn) \\ a = 4977.06C_b^2 - 8105.61C_b + 4456.51 \\ b = -10847.2C_b^2 + 12817C_b - 6960.32 \end{cases} \tag{2}$$

Here, $\Delta$ is displacement (t), $V_k$ is speed (kn), $Fn$ is the Froude number, and $C_b$ is the block coefficient.

The weight of each part is calculated as follows:

$$\begin{cases} W_h = 0.034 \cdot L^{1.7}B^{0.7}D^{0.4}C_b^{0.5} \\ W_f = L^{0.8}B^{0.6}D^{0.3}C_b^{0.1} \\ W_m = 0.17P^{0.9} \end{cases} \tag{3}$$

In these equations, $L$, $B$, and $D$ are the length between perpendiculars, breadth, and depth (m), respectively, and $W_m$ is the mechanical and electrical equipment weight (t).

The main evaluation indexes are annual shipping cost, annual freight volume, and unit shipping cost:

$$C_{apt} = C_a / D_a \tag{4}$$

In this equation, $C_{apt}$ is the unit transportation cost (pounds/t), $D_a$ is the annual freight volume (t), and $C_a$ is the annual shipping cost (pounds), which consists of three parts: shipping cost ($C_c$), operation cost ($C_r$), and voyage cost ($C_v$). These are calculated as follows:

$$\begin{cases} C_c = 0.2C_s, \quad C_r = 40000DW^{0.3} \\ C_v = (C_f + C_{po})RTPA, \quad C_{po} = 6.3DW^{0.8}, \quad D_a = DW \cdot RTPA \\ C_f = 1.05\, C_d \cdot d_s \cdot P_f, \quad C_d = 0.19 \times P \times 24/1000 + 0.2 \\ d_s = RTM/(24V_k), \quad RTPA = 350/(d_s + d_P), \quad d_P = 2(D_c/R_h + 0.5) \end{cases} \tag{5}$$

Here, $C_f$ and $C_{po}$ are fuel cost and port cost (pounds), respectively; $RTPA$ is the number of round-trips a ship travels in one year; $C_d$ is the daily consumption of oil (t); $d_s$ is the number of shipping days; $P_f$ is the fuel price (pounds/t), where the default is 100; $RTM$ is the ship's

single-trip mileage (n miles), where the default is 5,000; $d_p$ is the ship's days in anchorage; $R_h$ is the cargo handling efficiency (t/day), where the default is 8,000; $D_c$ is the cargo dead weight (t); and $DW$ is the dead weight of the ship (t). The latter is obtained as follows:

$$DW = \Delta - LW \tag{6}$$

Accordingly, the design variables of the optimization model in this study can be identified as: length ($L$), breadth ($B$), depth ($D$), draft ($T$), speed ($V_k$), and the block coefficient ($C_b$).

## 2.2 Constraints

While evaluating and optimizing the economy of a ship, the technical performance of the design should also be taken into account. Thus, it is necessary to propose constraints in the optimization model, including dimension ratio, manoeuvrability, stability, and so on. These constraints are defined as follows:

$$\begin{cases} T_1 = 5.5 - L/B \geq 0, \quad T_2 = 18 - L/D \geq 0 \\ T_3 = 20 - L/T \geq 0, \quad T_4 = 0.45 \cdot DW^{0.31} - 0.6 \cdot T \geq 0 \\ T_5 = 0.85D - T \geq 0, \quad T_{61} = DW - 25000 \geq 0, \quad T_{62} = 500000 - DW \geq 0 \\ T_7 = 0.32 - Fn \geq 0, \quad T_8 = 0.15B - GM \geq 0 \end{cases} \tag{7}$$

In these equations, $T_1$, $T_2$, $T_3$, and $T_5$ are the dimension ratio constraints to ensure feasibility, manoeuvrability, and stability; $T_4$, $T_{61}$, and $T_{62}$ are the constraints for ship displacement; $T_7$ is the constraint for ship speed; and $T_8$ is the constraint for ship stability and seakeeping.

## 3 Deterministic parametric optimization design

In order to obtain satisfactory designs with low unit-transportation cost and high freight capacity, optimization objectives are set for the minimum unit-transportation cost ($C_{apt}$) and the maximum annual freight volume ($D_a$). Two heuristic algorithms are adopted here to achieve multiobjective optimization: adaptive simulated annealing (ASA) [9] and the multi-island genetic algorithm (MIGA) [10]. These algorithms have superior performance for nonlinear optimization problems. The parameters for the optimization model are shown in Table 1.

**Table 1. Description of deterministic optimization model**

**Objective：**
- ▶ Minimum $C_{apt}$ and maximum $D_a$

**Design variables：**
- ▶ $L$: [100, 400], initial value 217 (*m*) ▶ $B$: [10, 45], initial value 32.3 (*m*)
- ▶ $T$: [10, 15], initial value 12.5 (*m*) ▶ $D$: [10, 25], initial value 19.7 (*m*)
- ▶ $V_k$: [14, 18], initial value 14.5 (*m*) ▶ $C_b$: [0.63, 0.83], initial value 0.82 (*m*)

**Constraints：**
- ▶ $T_1$–$T_8 \geqslant 0$

**Constants：**
- ▶ $P_f$: 100 (*pound/t*) ▶ RTM: 5000 (*n mile*) ▶ $R_h$: 8000 (*t/day*)

**Optimization techniques：**
- ▶ ASA ▶ MIGA

The main internal parameters of ASA and MIGA are set as follows: the maximum number of generated designs is 500, the relative rate of parameter annealing is 1.0, the convergence epsilon is 1e-8, the subpopulation size is 20, the number of islands is 10, and the number of generations is 10. The optimization results are shown in Figure 1.



(a) ASA          (b) MIGA
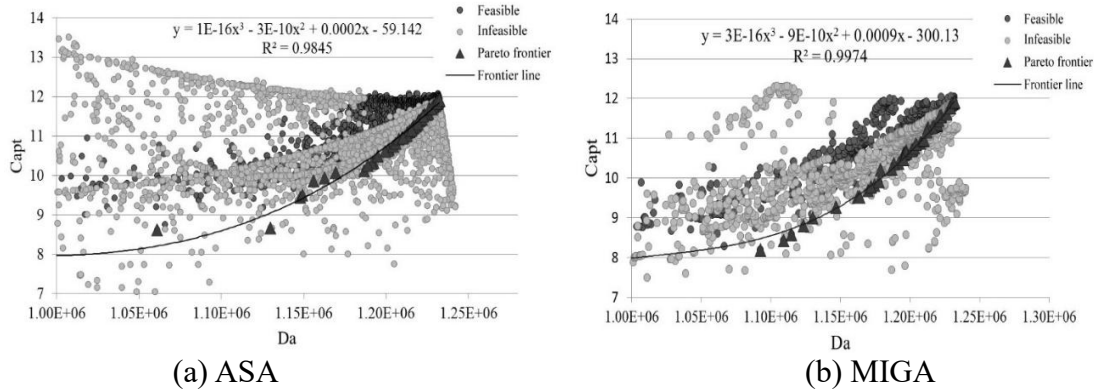
**Figure 1. Multiobjective optimization results graphs**

Generally, the feasible solution is centralized, and the frontiers are very concentrated with a clear Pareto frontier. The point distribution of the MIGA in the optimization process is relatively more concentrated and uniform in its concentrated area. By comparison, ASA is more uneven in the optimization process: in addition to the concentration of several lines, the focus is almost exclusively on the range of 10~12 on the $C_{pat}$ axis and 1,200,000~1,250,000 on the $D_a$ axis. With this kind of optimal result, designer can select some excellent plans based on the frontier curve and carry out detailed design for the next step.

## 4 Uncertainty analysis

### 4.1 Sensitivity analysis

In order to analyse the influence of variable changes on the outputs, it's necessary to perform a sensitivity analysis of the design variables ($L$, $B$, $D$, $T$, $V_k$, and $C_b$) and the important constants ($P_f$, $RTM$, and $R_h$) toward the optimization objective.

One experiment's design technique is adopted here: the Latin hypercube design, in which the engineer has total freedom in selecting the number of designs to run. A total of 1,200 points are generated for the Latin hypercube. The main effect and Pareto contributions of $C_{apt}$ and $D_a$ are shown in Figures 2~5, which reflect the degree each input has an effect on each output.
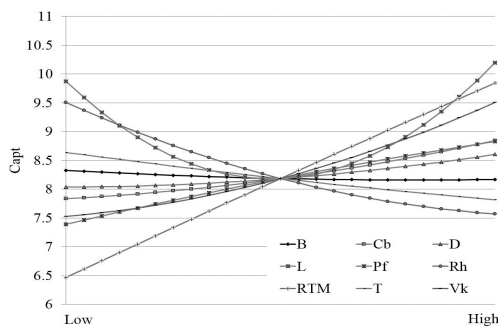


**Figure 2. $C_{apt}$ main effect graph**          **Figure 3. $D_a$ main effect graph**
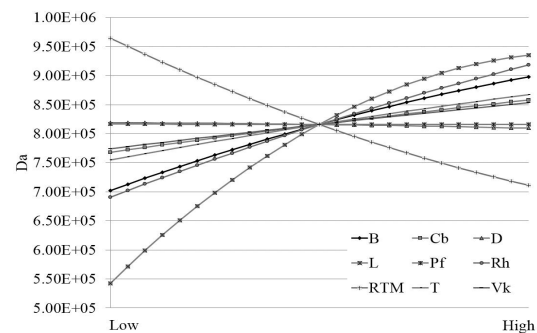
288
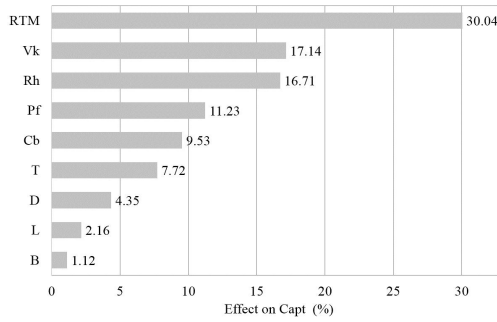
**Figure 4. $C_{apt}$ Pareto graph**      **Figure 5. $D_a$ Pareto graph**
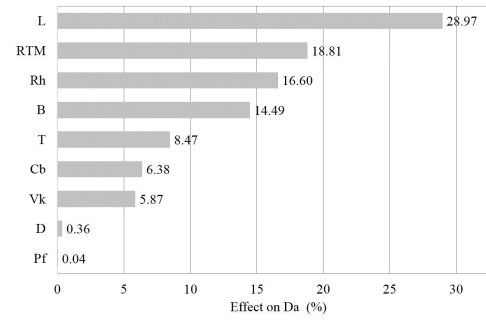
According to Figures 2 and 4, the parameters that influence $C_{apt}$ the most are $RTM$, $V_k$, $R_h$, and $P_f$, and all of their contributions exceed 10%. According to Figures 3 and 5, the parameters that influence $D_a$ the most are $L$, $RTM$, $R_h$, and $B$, and their contributions also exceed 10%. Therefore, uncertainty in the important constants, such as $P_f$, $RTM$, and $R_h$, would lead to uncertainty in the output and affect the whole optimization design.

### 4.2 Uncertainty analysis

To analyse the uncertainty influence, a Monte Carlo simulation (MCS) was adopted [11]. In an MCS, the probability distribution of random variables is known. Through random sampling, the probability distribution of a system's response can be estimated, and the contribution of each random variable to the response results can be obtained.

There are two sampling techniques in an MCS: simple random sampling and descriptive sampling. Compared to the former, descriptive sampling reduces the variance of the statistical estimates derived from the population data. Descriptive sampling also ensures the quality of statistical analysis with less sampling and simulation time, so it becomes a more representative method and so is used in this study. The uncertainty analysis model is shown in Table 2.

**Table 2. Uncertainty analysis model**

| |
|---|
| **Objective：** |
|   ▶ Uncertainty influence on the optimization object |
| **Design variables：** |
|   ▶ $L$: 217 (m);  ▶ $B$: 32.26 (m);  ▶ $D$: 19.7 (m); |
|   ▶ $T$: 12.5 (m);  ▶ $C_b$: 0.82;  ▶ $V_k$: 14.5 (kn); |
| **Uncertainty factors:** |
|   ▶ $P_f$, Normal, $\mu$=100 (Pound/t), $\sigma$=1%*$\mu$ |
|   ▶ $RTM$, Normal, $\mu$=5000 (n mile), $\sigma$=1%*$\mu$ |
|   ▶ $R_h$, Normal, $\mu$=8000 (t/day), $\sigma$=1%*$\mu$ |
| **Constraints：** |
|   ▶ $T_1$–$T_8 \geqslant 0$ |
| **Analysis Method** |
|   ▶ Monte Carlo simulation: descriptive sampling |

The maximum number of simulations is set at 10,000, and then a normal distribution simulation of three uncertain parameters ($P_f$, $RTM$, and $R_h$) is conducted. Their effects on the result optimization object ($C_{apt}$, as a more important factor to be considered,) are calculated independently. A histogram is then drawn of the frequency distribution and the frequency fitting curve according to system response parameters, a normal distribution hypothesis test is conducted, and the normal distribution curve is drawn. The influence of different parameters

on the response uncertainty can then be analysed comparatively. The results of the uncertainty analysis by MCS are shown in Figure 6.
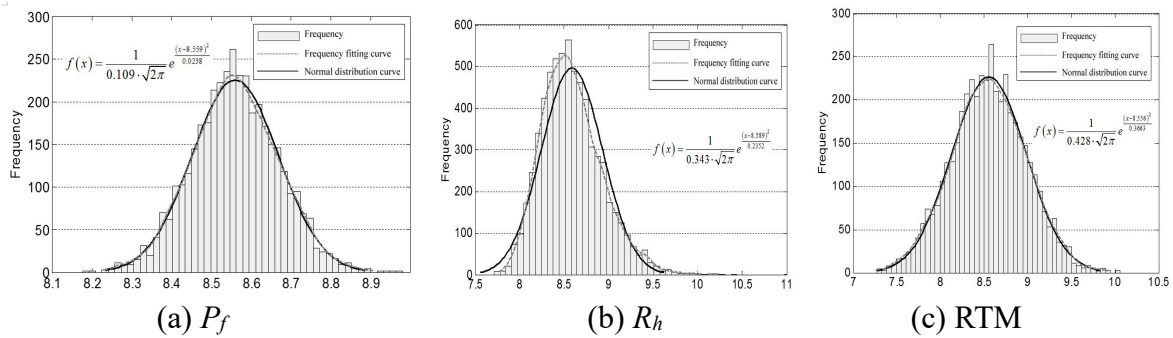


**Figure 6. Distribution of $C_{apt}$ when a single parameter obeys normal distribution**

A normal hypothesis test is then done. The related results and parameters are shown in Table 3. The statistical results of the system response, $C_{apt}$, are shown in Table 4.

**Table 3. Results of the normal hypothesis test**

| Input parameters | $R_h$ obeys normal distribution separately | RTM obeys normal distribution separately | $P_f$ obeys normal distribution separately |
|---|---|---|---|
| Test statistics | 920.9634 | 1.1631 | 1.9229 |
| Critical value | 5.7458 | 6.1611 | 5.7143 |
| H | 1 | 0 | 0 |
| $C_{apt}$ obeys normal distribution? | No | Yes | Yes |
| P | 0 | 0.5570 | 0.3790 |

**Table 4. Statistical results of system response, $C_{apt}$**

| Statistical indicators | $R_h$ obeys normal distribution separately | RTM obeys normal distribution separately | $P_f$ obeys normal distribution separately |
|---|---|---|---|
| Expectation (E) | 8.589 | 8.556 | 8.559 |
| Standard deviation (S.D) | 0.343 | 0.428 | 0.109 |
| S.D/E | 0.0399 | 0.0500 | 0.0127 |
| Skewness | 2.87E-02 | 2.51E-03 | 4.48E-05 |
| Kurtosis | 5.79E-02 | 9.87E-02 | 4.32E-04 |

The distribution of the system response parameters can be compared directly through the ratio of standard deviation to expectation. The ratio for $R_h$, RTM, and $P_f$ is 0.040, 0.050, and 0.013, respectively, showing that the fluctuation of the system response parameter, $C_{apt}$, is more obvious with the random variable RTM.

As is shown, when random variables $P_f$ and RTM obey normal distribution, the distribution of $C_{apt}$ also strictly obeys normal distribution. When $R_h$ alone obeys normal distribution, $C_{apt}$ can also be approximated as a normal distribution. However, by comparing the standard-deviation-to-expectation ratio of the three groups of data, it can be seen that the fluctuation of the system response caused by RTM is more obvious.

Therefore, when the uncertainty of parameters is considered in ship design, different

uncertainty parameters have different effects. The more-obviously-effect factors should be set in a more clearly pattern (probability distribution or interval, with accurate description), while the remainder can be set in an approximate range.

## 5 Conclusion

This research focuses on economic ship optimization design and its uncertainty analysis due to the fluctuation of internal parameters. Through the above simulation, calculation, and analysis, the response relationship between system output and random variable input is obtained, which can be used to guide the optimal design of actual ship optimization. The following conclusions are drawn:

(1) The uncertainty analysis based on MCS with descriptive sampling can clearly depict the impact of uncertain parameters on output response, thus the research conducted could hopefully promote the development of uncertainty optimization ship design.

(2) Different uncertainty parameters have different effects on the output response, thus they should be considered separately in ship design.

(3) Further studies can be conducted to investigate the properties of uncertain parameters and the applicability of uncertainty optimization algorithms with uncertainty.

**References**

[1] Papanikolaou A . Holistic ship design optimization[J]. Computer-Aided Design, 2010, 42(11):1028-1044.

[2] Yuan hang H, You Y, Xiao L. Minimum Resistance Ship Hull Uncertainty Optimization Design Based on Simulation-Based Design Method[J]. Journal of Shanghai Jiaotong University, 2017, 22(6):1-7.

[3] Chen J, Wei J, Jiang W. Optimization of a twin-skeg container ship by parametric design and CFD simulations[J]. International Journal of Naval Architecture & Ocean Engineering, 2016, 8(5):466-474.

[4] Wei H, Diez M., Zou Z J, etc. URANS study of Delft catamaran total/added resistance, motions and slamming loads in head sea including irregular wave and uncertainty quantification for variable regular wave and geometry[J]. Ocean Engineering, 2013, 74: 189-217

[5] Diez, M., Wei H, Campana E F. Uncertainty quantification of Delft catamaran resistance, sinkage and trim for variable Froude number and geometry using meta-models, quadrature and Karhunen–Loeve expansion[J]. J Mar Sci Technol, 2014,19: 143-169

[6] Leotardi C, Campana E F, Diez M. On the use of uncertainty quantification in variable-accuracy simulation-based multidisciplinary optimisation[C]. In: 3rd Workshop on Uncertainty Quantification in Computational Fluid Dynamics, 2015

[7] Diez M, Peri D. Robust optimization for ship conceptual design[J]. Ocean Engineering, 2010, 37(11):966-977.

[8] Sen P, Yang J B. Multiple Criteria Decision Support in Engineering Design[C]. Engineering in a Changing Economy, 1998.

[9] Chen S, Luk B L. Adaptive simulated annealing for optimization in signal processing applications[J]. Signal Processing, 1999, 79(1):117-128.

[10] Mu S, Du C, Mu F, et al. Topology Optimization of a Continuum Structure Using Multi-Island Genetic Algorithm (MIGA)[J]. Mechanical Science & Technology for Aerospace Engineering, 2009, 28(10):1316-1320.

[11] Binder K, Heermann D W. Monte Carlo Simulation in Statistical Physics[M]. World Book Inc, 2014.

# A kinematic yield design of materials using computational homogenization analysis and stabilized radial point interpolation method

**\*Phuc L. H. Ho[1], †Canh V. Le[1] and Phuong H. Nguyen[1]**

[1]Department of Civil Engineering, International University - VNU-HCMC, Vietnam

\*Presenting author: hlhphuc@hcmiu.edu.vn
†Corresponding author: lvcanh@hcmiu.edu.vn

## Abstract

This study presents a novel numerical procedure for kinematic yield design of periodic heterogeneous microstructures. The fluctuating part of displacement field is approximated using stabilized radial point interpolation method (RPIM). Taking advantage of truly mesh-free approximation and second order cone programming (SOCP), the resulting optimization problem formulated on a representative volume element (RVE) can be efficiently solved. The macroscopic strength domains at limit state and failure mechanisms associated with different load cases are obtained. Several numerical examples are investigated to perform the computational aspect of proposed method.

**Keywords:** Computational homogenization analysis, limit analysis, radial point interpolation method, second order cone programming.

## 1 Introduction

Composite and heterogeneous materials are increasingly used in practical engineering, and the prediction of their effective mechanical response plays a crucial role in safety assessment as well as structural design. Various mathematical theories of homogenization have been developed to obtained the mechanical behaviors of heterogeneous media, involving the multiple scale nature. Well-known as the most suitable treatment for composites with periodic microstructures, the computational homogenization approach [3], based on the solution of two boundary value problems in macroscopic and microscopic scales, has been widely exploited in recent years. In framework of limit analysis for periodic microstructures, the first formulation has been proposed in [4]. Then, the theory was extended to fiber-reinforced composite using Drucker-Prager, Mohr-Coulomb or von Mises yield criterion by [5, 6]. The first numerical implementation was developed in [7] with the use of finite element method and linear mathematical programming. A static finite element formulation for analysis of isotropic microstructures in three-dimensions was developed by [8, 9]. Similarly, a pseudo-lower bound formulation for periodic composite and heterogeneous materials using the nonlinear programming was presented in [10]. The kinematic approaches in combination with nonlinear algorithms, for which both of isotropic and anisotropic materials obeying the von Mises or elliptic yield criterion were considered, have been reported in [12, 13, 14, 15, 16]. In order to improve the computational aspect, a kinematic formulation combined with second order cone programming for periodic materials were proposed in [17, 18].

In this study, a novel computational homogenization approach for upper bound limit analysis of microstructures using stabilized radial point interpolation method is developed. The stability conforming nodal integration (SCNI) technique proposed in [20] is employed to improve the computational effect of the numerical formulation. In addition, the plastic dissipation will be transformed into the form of a sum of norms and the resulting optimization are then formulated

as conic programming. The benchmark numerical examples will be investigated to test the performance of proposed method.

## 2 Kinematic limit formulations for computational homogenization analysis of materials

Consider a heterogeneous representative volume element (RVE) $\Omega \in \mathbb{R}^2$ subjected to the body force $\mathbf{f}$, the surface load $\mathbf{t}$ on the static boundary $\Gamma_t$ and fixed by the displacement field $\mathbf{u}$ on the kinematic boundary $\Gamma_u$. The local fields for microscopic structures are decomposed into a mean part and a fluctuation part as follows

$$\mathbf{u} = \mathbf{E} \cdot \mathbf{X} + \tilde{\mathbf{u}} \tag{1a}$$

$$\boldsymbol{\epsilon} = \mathbf{E} + \tilde{\boldsymbol{\epsilon}} \tag{1b}$$

$$\boldsymbol{\sigma} = \boldsymbol{\Sigma} + \tilde{\boldsymbol{\sigma}} \tag{1c}$$

where $\mathbf{u}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\sigma}$ denote displacement, strain and stress fields; $\boldsymbol{\Sigma}$ and $\mathbf{E}$ are the overall stress and strain; $\tilde{\mathbf{u}}$, $\tilde{\boldsymbol{\epsilon}}$ and $\tilde{\boldsymbol{\sigma}}$ are the fluctuation parts of displacement, strain and stress rate; matrix $\mathbf{X}$ consists of the position components of each material point in the computational domain.

Note that, at micro-scale, the fluctuate displacement and stress fields must satisfy the periodic and anti-periodic conditions on boundaries

$$\tilde{\mathbf{u}}(\mathbf{x}^+) = \tilde{\mathbf{u}}(\mathbf{x}^-), \qquad \text{on } \Gamma_u \tag{2a}$$

$$\mathbf{t}(\mathbf{x}^+) = -\mathbf{t}(\mathbf{x}^-), \quad \text{on } \Gamma_t \tag{2b}$$

where $\tilde{\mathbf{u}}(\mathbf{x}^+)$ and $\tilde{\mathbf{u}}(\mathbf{x}^-)$ are the fluctuation displacement field, $\mathbf{t}(\mathbf{x}^+)$ and $\mathbf{t}(\mathbf{x}^-)$ are the traction field of positive and negative boundaries, respectively. As a result, the average of $\tilde{\boldsymbol{\epsilon}}$ and $\tilde{\boldsymbol{\sigma}}$ over the RVE should vanish, meaning that

$$\langle \tilde{\boldsymbol{\epsilon}} \rangle = 0; \quad \langle \tilde{\boldsymbol{\sigma}} \rangle = 0 \tag{3}$$

where the operation $\langle \cdot \rangle$ stands the volume average of fields over the RVE.

Denoting $|\Omega|$ for the area of RVE, the overall quantities can be calculated via the average relations

$$\mathbf{E} \equiv \langle \boldsymbol{\epsilon} \rangle = \frac{1}{|\Omega|} \int_\Omega \boldsymbol{\epsilon} \, \mathrm{d}\Omega; \quad \boldsymbol{\Sigma} \equiv \langle \boldsymbol{\sigma} \rangle = \frac{1}{|\Omega|} \int_\Omega \boldsymbol{\sigma} \, \mathrm{d}\Omega \tag{4}$$

For any admissible velocity and stress field satisfying the periodic and anti-periodic conditions, the principle of macroscopic virtual work can be expressed as

$$\langle \boldsymbol{\sigma} : \boldsymbol{\epsilon} \rangle = \boldsymbol{\Sigma} : \mathbf{E} \tag{5}$$

Denoting $\mathscr{X}$ and $\mathscr{Y}$ for the appropriate spaces of statically admissible stress state and kinematically admissible velocity state, respectively. Solving one of following optimization problems, the exact collapse multiplier will be obtained

$$\lambda_{exact} = \max\{\lambda \mid \exists \boldsymbol{\sigma} \in \mathscr{B} : a(\boldsymbol{\sigma}, \mathbf{u}) = \lambda F(\mathbf{u}), \forall \mathbf{u} \in \mathscr{Y}\} \tag{6a}$$

$$= \max_{\boldsymbol{\sigma} \in \mathscr{B}} \min_{\mathbf{u} \in \mathscr{C}} a(\boldsymbol{\sigma}, \mathbf{u}) \tag{6b}$$

$$= \min_{\mathbf{u} \in \mathscr{C}} \max_{\boldsymbol{\sigma} \in \mathscr{B}} a(\boldsymbol{\sigma}, \mathbf{u}) \tag{6c}$$

$$= \min_{\mathbf{u} \in \mathscr{C}} D(\mathbf{u}) \tag{6d}$$

where $\mathscr{C} = \{\mathbf{u} \in \mathscr{Y} \mid F(\mathbf{u}) = 1\}$; $\mathscr{B} = \{\boldsymbol{\sigma} \in \mathscr{X} \mid \psi(\boldsymbol{\sigma}) \leq 0\}$; the plastic dissipation rate $D(\mathbf{u})$ and the external work $F(\mathbf{u})$ can be expressed in terms of $\boldsymbol{\sigma}$ and $\mathbf{u}$ as follow

$$D(\mathbf{u}) = \max_{\boldsymbol{\sigma} \in \mathscr{B}} a(\boldsymbol{\sigma}, \mathbf{u}) \tag{7a}$$

$$F(\mathbf{u}) = \int_{\Omega} \mathbf{f}^T \mathbf{u} \mathrm{d}\Omega + \int_{\Gamma_t} \mathbf{t}^T \mathbf{u} \mathrm{d}\Gamma \tag{7b}$$

It is worth noting that most of yield criterion can be expressed in the following form

$$\psi(\boldsymbol{\sigma}) = \sqrt{\boldsymbol{\sigma}^T \mathbf{P} \boldsymbol{\sigma}} - 1 \tag{8}$$

where $\mathbf{P}$ is the coefficient matrix consisting of strength properties of materials. Using von Mises criterion, which is frequently applied for isotropic materials, matrix $\mathbf{P}$ for plane stress problem can be expressed as

$$\mathbf{P} = \frac{1}{\sigma_p} \begin{bmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \tag{9}$$

where $\sigma_p$ is the yield stress of material.

In framework of limit analysis, the power of dissipation can be formulated in terms of strain rates assumed to obey the normality rule as

$$D(\boldsymbol{\epsilon}) = \int_{\Omega} \sqrt{\boldsymbol{\epsilon}^T \boldsymbol{\Theta} \boldsymbol{\epsilon}} \mathrm{d}\Omega = \int_{\Omega} \sqrt{(\mathbf{E} + \tilde{\boldsymbol{\epsilon}})^T \boldsymbol{\Theta} (\mathbf{E} + \tilde{\boldsymbol{\epsilon}})} \mathrm{d}\Omega \tag{10}$$

where $\boldsymbol{\Theta} = \mathbf{P}^{-1}$.

In the absence of body force $\mathbf{f}$, the external work can be obtained by applying the principle of virtual work as

$$F(\mathbf{u}) = \int_{\Gamma_t} \mathbf{t}^T \mathbf{u} \mathrm{d}\Gamma = \boldsymbol{\Sigma}^T \mathbf{E} = 1 \tag{11}$$

The kinematic approach in framework of limit analysis for computation homogenization described in [11, 12, 13, 15, 16, 18] will be taken into account in this paper. The upper bound limit formulation for a periodic microstructure can be expressed as

$$\lambda^+ = \min \int_{\Omega} \sqrt{\boldsymbol{\epsilon}^T \boldsymbol{\Theta} \boldsymbol{\epsilon}} \mathrm{d}\Omega \tag{12a}$$

$$\text{s.t} \begin{cases} \boldsymbol{\Sigma}^T \mathbf{E} = 1 \\ \tilde{\mathbf{u}} \text{ periodic on } \Gamma_u \end{cases} \tag{12b}$$

The upper bound of macroscopic limit strength $\lambda^+ \boldsymbol{\Sigma}$ will be determined by solving the nonlinear problem described by Eq. (12). It should be note that only continuous velocity fields are considered in this study. For the cases when the velocity fields are assumed to be discontinuous, the dissipated power generated by discontinuities must be taken into account.

## 3 Stabilized radial point interpolation method

Consider a set of scattered nodes $\mathbf{x}_Q^T = [x_1, x_2, ..., x_{\mathscr{N}}]$ within a closed domain of are $\Omega$, the approximate function obtained by interpolating pass-through nodal value is expressed as

$$u^h(\mathbf{x}) = \mathbf{R}(\mathbf{x})\mathbf{a} + \mathbf{p}(\mathbf{x})\mathbf{b} \tag{13}$$

where $\mathbf{a}^T = \{a_1, a_2, ..., a_{\mathscr{N}}\}$ and $\mathbf{b}^T = \{b_1, b_2, ..., b_M\}$ are the coefficient vectors associating with radial basis function (RBF) $\mathbf{R}(\mathbf{x})$ and polynomial basis function (PBF) $\mathbf{p}(\mathbf{x})$; $\mathscr{N}$ and $M$ are number of terms in $\mathbf{R}(\mathbf{x})$ and $\mathbf{p}(\mathbf{x})$, respectively.

Enforcing $u^h(\mathbf{x})$ at scattered nodes in problem domain, the matrix form of Eq. (13) can be expressed as follows

$$\mathbf{U} = \mathbf{R}_Q \mathbf{a} + \mathbf{P}_M \mathbf{b} \tag{14}$$

where $\mathbf{R}_Q$ is given by

$$\mathbf{R}_Q = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ R_I(r_1) & R_I(r_2) & \cdots & R_I(r_{\mathscr{N}}) \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}_{\mathscr{N} \times \mathscr{N}} \tag{15}$$

with $r_k = ||\mathbf{x}_k - \mathbf{x}_I||$ is the distance from node $I^{th}$ to node $\mathbf{x}_k \in \mathbf{x}_Q$. The best ranked RBF function in terms of accuracy so-called multi-quadric (MQ) is employed in this study

$$R_I(r_k) = \sqrt{r_k^2 + (\alpha_s d_I)^2} \tag{16}$$

where $\alpha_s > 0$ is the shape parameter, $d_I$ is the minimal distance from node $I^{th}$ to its neighbors within the influence domain limited by the circle of radius $\beta_s d_I$, in which $\beta_s$ denotes the influence parameter.

To guarantee the unique property of approximate function, the PBF term must satisfy following condition

$$\mathbf{P}_M^T \mathbf{a} = 0 \tag{17}$$

Combining Eqs. (14) and (17), we obtains

$$\mathbf{G} \begin{Bmatrix} \mathbf{a} \\ \mathbf{b} \end{Bmatrix} = \begin{Bmatrix} \mathbf{U} \\ 0 \end{Bmatrix} \tag{18}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{R}_Q & \mathbf{P}_M \\ \mathbf{P}_M^T & 0 \end{bmatrix} \tag{19}$$

The coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ can be determined by an efficient manner presented in [19] as

$$\mathbf{a} = \chi_a \mathbf{U}; \quad \mathbf{b} = \chi_b \mathbf{U} \tag{20}$$

where

$$\chi_a = \mathbf{R}_Q^{-1}[1 - \mathbf{P}_M \chi_b] = \mathbf{R}_Q^{-1} - \mathbf{R}_Q^{-1}\mathbf{P}_M \chi_b \tag{21a}$$

$$\chi_b = [\mathbf{P}_M^T \mathbf{R}_Q^{-1} \mathbf{P}_M]^{-1} \mathbf{P}_M^T \mathbf{R}_Q^{-1} \tag{21b}$$

The approximate function in Eq. (13) can be now rewritten as

$$u^h(\mathbf{x}) = [\mathbf{R}(\mathbf{x})\chi_a + \mathbf{p}(\mathbf{x})\chi_b]\mathbf{U} = \sum_{I=1}^{\mathscr{N}} \Phi_I(\mathbf{x})u_I \tag{22}$$

where $\Phi_I$ denotes RPIM shape function at node $I^{th}$, for which the related values of shape

function and its partial derivatives for node $\mathbf{x}_k$ can be calculated as follows

$$\Phi_k = \sum_{I=1}^{\mathscr{N}} R_I \chi_{Ik}^a + \sum_{J=1}^{M} p_J \chi_{Jk}^b; \tag{23a}$$

$$\frac{\partial \Phi_k}{\partial x} = \sum_{I=1}^{\mathscr{N}} \frac{\partial R_I}{\partial x} \chi_{Ik}^a + \sum_{J=1}^{M} \frac{\partial p_J}{\partial x} \chi_{Jk}^b; \quad \frac{\partial \Phi_k}{\partial y} = \sum_{I=1}^{\mathscr{N}} \frac{\partial R_I}{\partial y} \chi_{Ik}^a + \sum_{J=1}^{M} \frac{\partial p_J}{\partial y} \chi_{Jk}^b \tag{23b}$$

It is interesting to note that RPIM shape function satisfies Kronecker delta property; consequently, the essential boundary condition in problems can be easily imposed in the similar way to finite element method.

To improve the computational effect and ensure the truly mesh-free features of RPIM approximation, the stabilized conforming nodal integration (SCNI) scheme introduced in [20] is employed in this study. The idea of SCNI is that strains will be smoothed over the nodal representative domain, detail calculation can be found in [20, 21]. The smoothed derivatives of shape function can be computed by

$$\tilde{\Phi}_{I,\alpha}(\mathbf{x}_J) = \frac{1}{a_J} \oint_{\Gamma_J} \Phi_I(\mathbf{x}_J) n_\alpha(\mathbf{x}) \mathrm{d}\Gamma = \frac{1}{2a_J} \sum_{k=1}^{ns} \left( n_\alpha^k \, L^k + n_\alpha^{k+1} \, L^{k+1} \right) \Phi_I \left( \mathbf{x}_J^{k+1} \right) \tag{24}$$

where $\tilde{\Phi}$ is the smoothed version of $\Phi$; $a_J$ is the area of representative domain $\Omega_J$; $n_s$ is number of edges bounded $\Omega_J$; $\mathbf{x}_J^k$ and $\mathbf{x}_J^{k+1}$ are the coordinates of the two end points of boundary segment $\Gamma_J^k$ having length $L^k$ and outward surface normal $n^k$.

## 4 Stabilized RPIM-based discrete formulation

In kinematic formulation, the fluctuation terms of displacement field $\tilde{u}^h(\mathbf{x})$ and fluctuation strain $\tilde{\epsilon}(\mathbf{x})$ can be approximated using RPIM method as

$$\tilde{u}^h(\mathbf{x}) = \sum_{i=1}^{\mathscr{N}} \Phi_i(\mathbf{x}) \tilde{u}_i = \mathbf{Nd} \tag{25a}$$

$$\tilde{\epsilon}(\mathbf{x}) = \sum_{i=1}^{\mathscr{N}} \tilde{\Phi}_{i,\alpha}(\mathbf{x}) \tilde{u}_i = \mathbf{Bd} \tag{25b}$$

where $\mathbf{N}$ denotes the RPIM shape function; $\mathbf{B}$ is the strain matrix consisting of the smoothed shape function derivatives; and $\mathbf{d}$ is the nodal displacement vector.

The plastic dissipation work can be computed using nodal integration as

$$D_p(\epsilon) = \sum_{i=1}^{\mathscr{N}} \sigma_p A_i \sqrt{(\mathbf{E} + \mathbf{Bd})^T \Theta (\mathbf{E} + \mathbf{Bd})} \tag{26}$$

where $\sigma_p$ is the yield stress of material; $A_i$ is the smoothed area of node $i^{th}$.

Introducing the additional variable $\rho$ such that

$$\boldsymbol{\rho}_i = \mathbf{Q}^T (\mathbf{E} + \mathbf{Bd}), \quad i = 1, 2, \ldots, \mathscr{N} \tag{27}$$

where $\mathbf{Q}$ is the Cholesky factor of $\Theta$, the internal dissipation power can be expressed in form

of sum of norms as

$$D_p(\boldsymbol{\epsilon}) = \sum_{i=1}^{\mathcal{N}} \sigma_p A_i \|\boldsymbol{\rho}_i\| \tag{28}$$

where $\|.\|$ denotes the Euclidean norms.

Denoting $\Gamma^+$ and $\Gamma^-$ for the positive and negative boundaries of RVE, the periodic feature of fluctuation field is enforced as follows

$$\tilde{\mathbf{u}}(\mathbf{x}^+) - \tilde{\mathbf{u}}(\mathbf{x}^-) = 0 \tag{29}$$

where $\{\mathbf{x}^+, \mathbf{x}^-\}$ denotes a pair of nodes on opposite boundaries. The periodic condition can be then rewritten in matrix form as

$$\mathbf{Cd} = 0 \tag{30}$$

Finally, introducing the auxiliary variables $(t_1, t_2, \ldots, t_{\mathcal{N}})$, the optimization problem is formulated as conic programming as follows

$$\lambda^+ = \min \sum_{i=1}^{\mathcal{N}} \sigma_p A_i \|\boldsymbol{\rho}_i\| \tag{31a}$$

$$\text{s.t} \begin{cases} \boldsymbol{\Sigma}^T \mathbf{E} = 1 \\ \mathbf{Cd} = 0 \\ \|\boldsymbol{\rho}_i\| \leq t_i, \quad i = 1, 2, \ldots, \mathcal{N} \end{cases} \tag{31b}$$

## 5 Numerical examples

In this section, several microscopic structures, for which the numerical solutions are available, are investigated to test the performance of proposed method. In all problems, the RVE is considered with the size of $a \times a = 1 \times 1$ mm, and subjected to the overall in-plane stresses $(\Sigma_{11}, \Sigma_{22})$ forming with $x_1$-axis an angle $\varphi$. The materials are assumed to be isotropic in plane stress state and governed by von Mises yield criterion. The resulting optimization problems are solved using Mosek software package on a 2.8 GHz Intel Core i7 PC running Window 10.

### 5.1 Perforated material with a central circular hole

In this example, a perforated material with a central circular void, as shown in Fig. 1(a), is considered first. The matrix material is mild steel ($St3s$) with yield stress $\sigma_p = 273$ MPa. The computational domain and nodal discretization are presented in Fig. 1(b).

Various ratios of $R/a$ are investigated to exam the effect of hole size on macroscopic strength. For the load case of $\Sigma_{11} = 1$ and $\Sigma_{22} = 0$, the numerical solutions associating with inclined angles $\varphi = 0°$ and $\varphi = 45°$ are respectively summarized in Table 1 and plotted in Fig. 2. The yield surfaces related to different values of ratio $R/a$ and loading angle $\varphi$ are also performed in Fig. 3. Obviously, the macroscopic strength decrease when increasing volume of the void. The problem has been also numerically investigated in [12, 17] using kinematic finite element method, and in [18] using kinematic stabilized iRBF mesh-free formulation. Generally, present solutions are slightly lower than those in [12] and in good agreement compared with those in [17, 18].

The distributions of dissipation power for RVE with circular holes subjected to uniaxial load $(\Sigma_{11} = 1, \Sigma_{22} = 0)$ are plotted in Fig. 4.

(a) Geometry and loading       (b) Nodal discretization
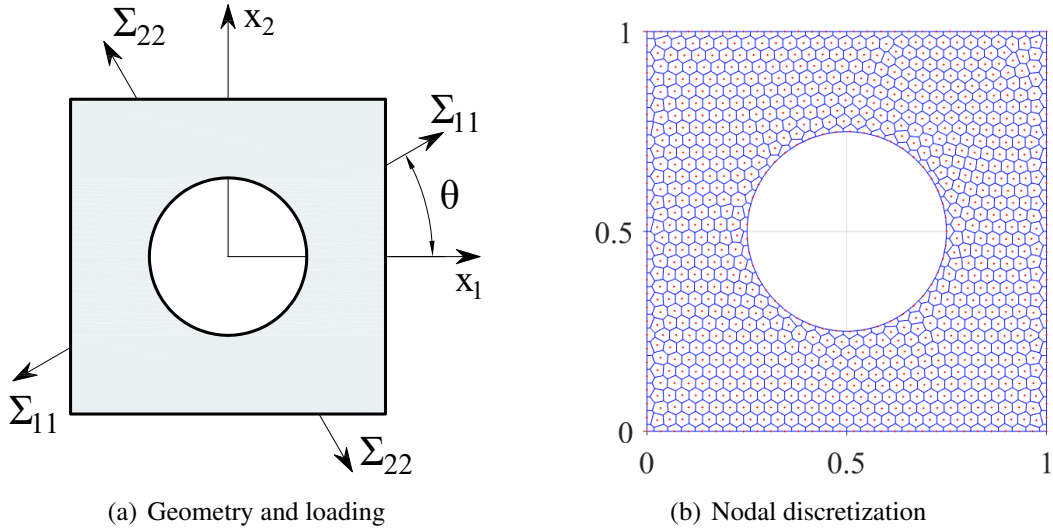
**Figure 1: Perforated RVE with circular hole.**

**Table 1: Perforated RVE with circular hole: uniaxial strength $\Sigma_{11}/\sigma_p$**

| $R/a$ | $V_f$ (%) | $\varphi = 0°$ | | $\varphi = 45°$ | |
|---|---|---|---|---|---|
| | | RPIM | iRBF in [18] | RPIM | iRBF in [18] |
| 0.1 | 3.14 | 0.8592 | 0.8441 | 0.8045 | 0.8029 |
| 0.2 | 12.57 | 0.6782 | 0.6780 | 0.5847 | 0.5829 |
| 0.3 | 28.27 | 0.4403 | 0.4402 | 0.3750 | 0.3706 |
| 0.4 | 50.27 | 0.2122 | 0.2118 | 0.1703 | 0.1691 |

$V_f$: hole volume fractions.



(a) $\varphi = 0°$       (b) $\varphi = 45°$
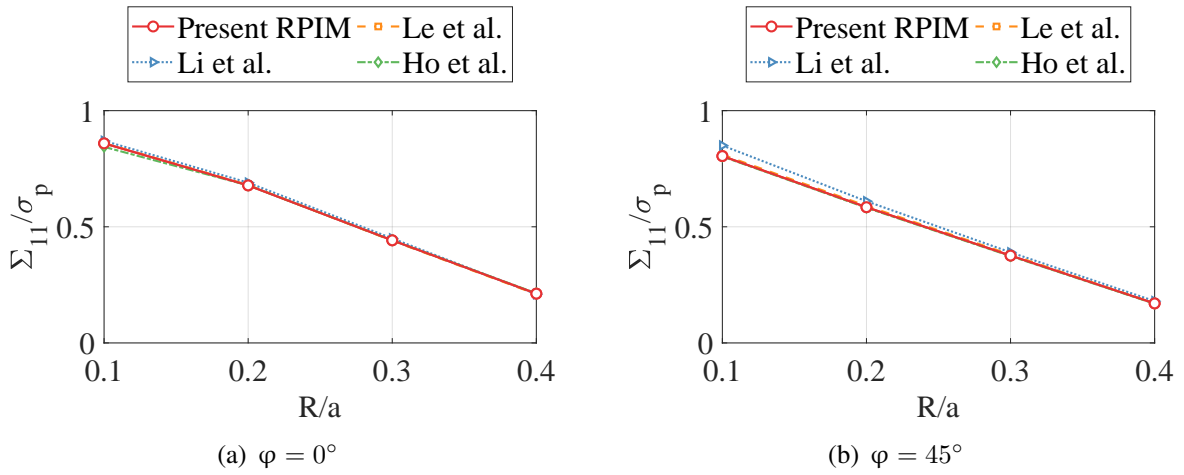
**Figure 2: Perforated RVE with circular hole: uniaxial strength for various values of $(R/a, \varphi)$.**

## 5.2 Fiber-reinforced composite material

In this example, a fiber-reinforced metal matrix composites, as shown in Fig. 5, is considered. Denoting $\sigma_p$ for yield strength of the matrix, the yield stress of fiber is given to be $20\sigma_p$; and a perfect bond between matrix and fiber is assumed.

Under biaxial in-plane loads $(\Sigma_{11}, \Sigma_{22})$, the macroscopic strength domain and distribution of

(a) $\varphi = 0°$

(b) $\varphi = 45°$
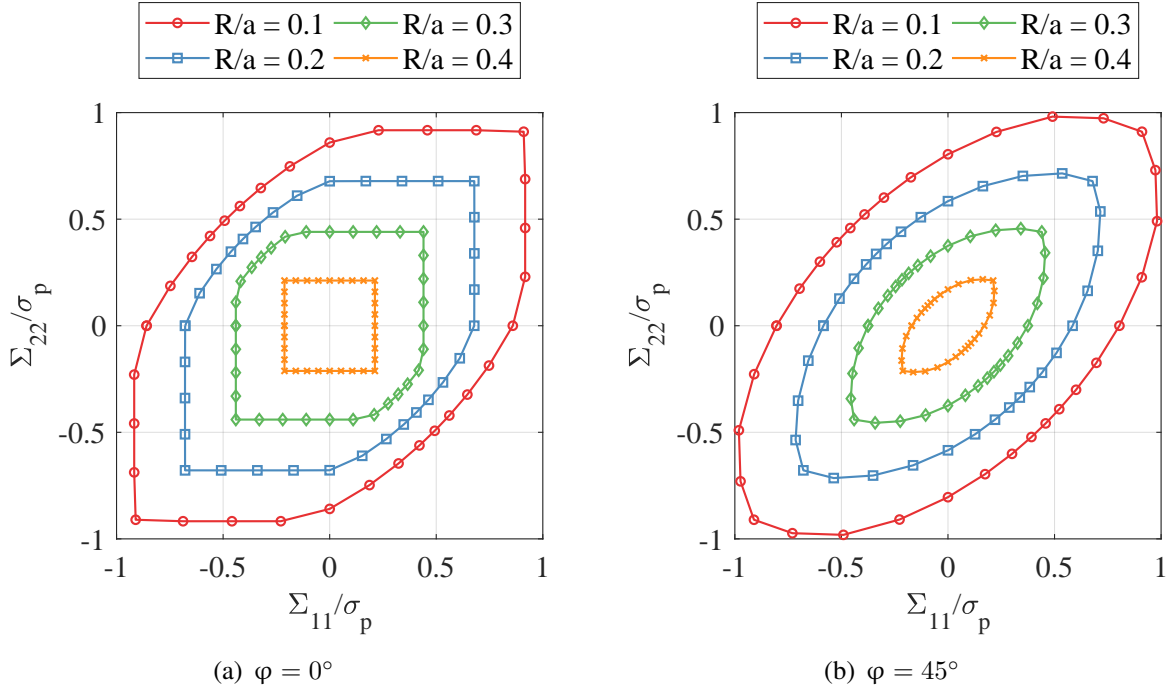
**Figure 3: Perforated RVE with circular hole: macroscopic strength domain with different values of of $(R/a, \varphi)$.**



(a) $\varphi = 0°$

(b) $\varphi = 45°$
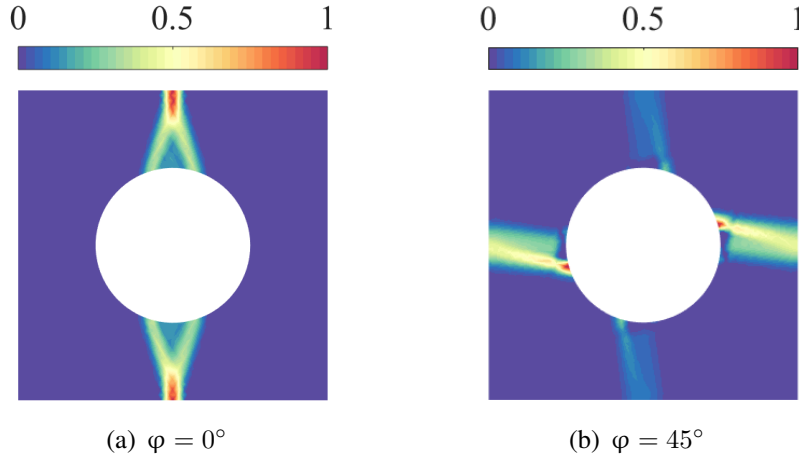
**Figure 4: Perforated RVE with regular hole arrangements ($\Sigma_{11} = 1, \Sigma_{22} = 0$): distributions of plastic dissipation at limit state.**

dissipation power for the case of $V_f = 20\%$ and $\varphi = 0°$ are plotted in Fig. 6. From the solutions seen in Fig. 6(a), it can be observed that present results are in good agreement in comparison with the numerical those reported by [14, 17]. The plastic zone occurs within the weak phase, i.e. the matrix ones, and concentrates at the interface between two materials.

## 6  Conclusions

A kinematic mesh-free formulation based on computational homogenization theory for yield design of microstructures has been presented in this study. The fluctuation displacement field is approximated using stabilized radial point interpolation method. By means of SCNI and SOCP, the resulting optimization problems are formulated in minimum size, leading to the

(a) Geometry and loading

(b) Nodal discretization
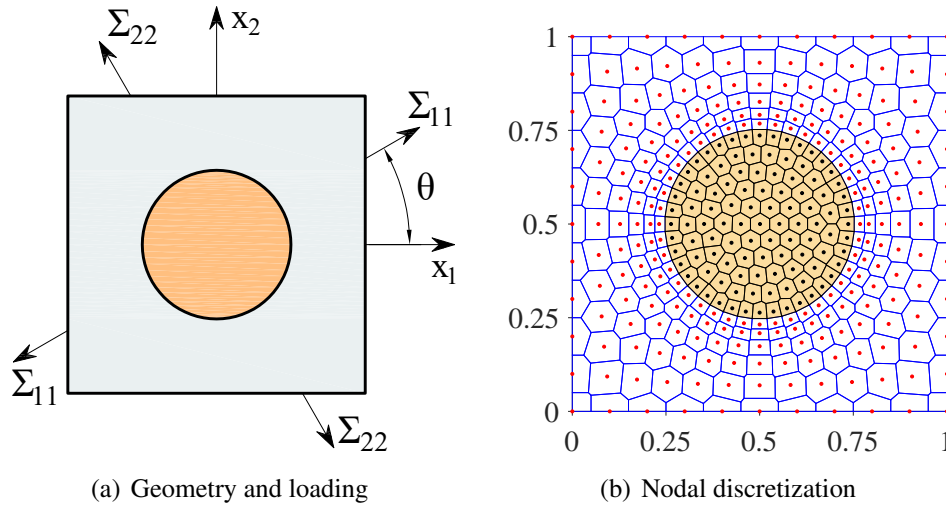
**Figure 5: Fiber-reinforced composite material.**



(a) Macroscopic strength domain

(b) Distribution of dissipation power
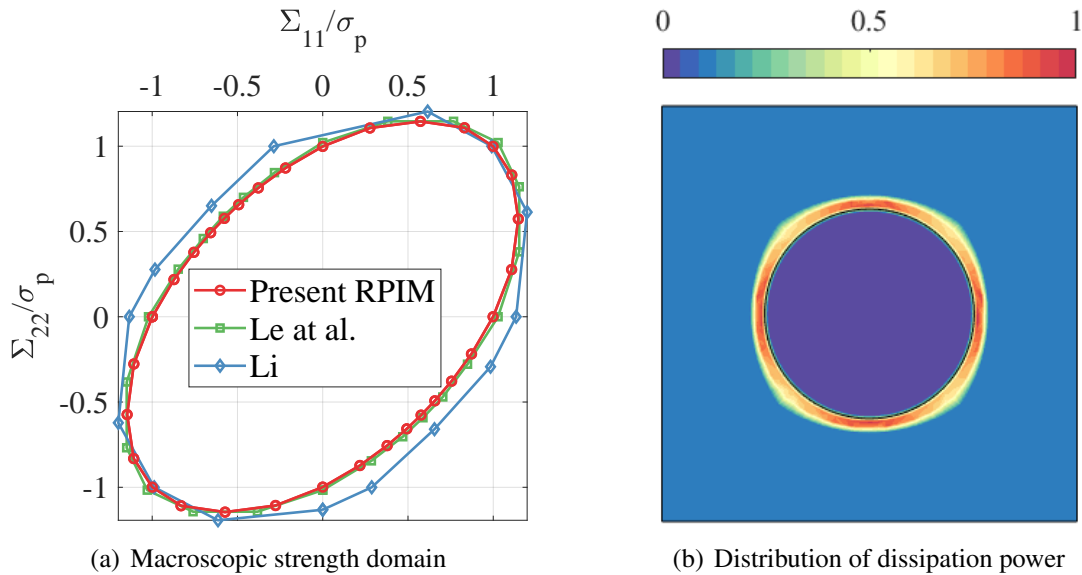
**Figure 6: Fiber-reinforced composite ($V_f = 20\%$, $\varphi = 0°$).**

reduction of computational effort. The periodic boundary conditions of RVE are easily imposed owing to the Kronecker delta property of RPIM shape function. Several numerical examples have been investigated; and the good agreement in comparison with other studies demonstrates the computational efficiency of proposed method. This study focuses on problems in plane stress state only; the two- and three-dimension RVEs in plane strain state with more complicate constitution and material distribution are extended in future works.

## Acknowledgements

## References

[1] Save, M. A., Massonnet, C. E., & Massonnet, C. C. (1972). *Plastic analysis and design of plates, shells and disks* (Vol. 15). North-Holland.

[2] Salençon, J. (2013). *Yield design*. John Wiley & Sons.

[3] Squet, P. M. (1985). Local and global aspects in the mathematical theory of plasticity. *Plasticity today*, 279-309.

[4] Buhan, D. E. (1991). A homogenization approach to the yield strength of composite materials. *European Journal of Mechanics, A/Solids*, 10(2), 129-154.

[5] Taliercio, A. (1992). Lower and upper bounds to the macroscopic strength domain of a fiber-reinforced composite material. *International journal of plasticity*, 8(6), 741-762.

[6] Taliercio, A., & Sagramoso, P. (1995). Uniaxial strength of polymeric-matrix fibrous composites predicted through a homogenization approach. *International Journal of Solids and Structures*, 32(14), 2095-2123.

[7] Francescato, P., & Pastor, J. (1997). Lower and upper numerical bounds to the off-axis strength of unidirectional fiber-reinforced composites by limit analysis methods. *European journal of mechanics. A. Solids*, 16(2), 213-234.

[8] Weichert, D., Hachemi, A., & Schwabe, F. (1999). Application of shakedown analysis to the plastic design of composites. *Archive of Applied Mechanics*, 69(9), 623-633.

[9] Weichert, D., Hachemi, A., & Schwabe, F. (1999). Shakedown analysis of composites. *Mech. Res. Commun.*, 26, 309-18.

[10] Zhang, H., Liu, Y., & Xu, B. (2009). Plastic limit analysis of ductile composite structures from micro-to macro-mechanical analysis. *Acta Mechanica Solida Sinica*, 22(1), 73-84.

[11] Carvelli, V., Maier, G., & Taliercio, A. (2000). Kinematic limit analysis of periodic heterogeneous media.

[12] Li, H., Liu, Y., Feng, X., & Cen, Z. (2003). Limit analysis of ductile composites based on homogenization theory. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2031), 659-675.

[13] Li, H. X., & Yu, H. S. (2006). Limit analysis of composite materials based on an ellipsoid yield criterion. *International journal of plasticity*, 22(10), 1962-1987.

[14] Li, H. X. (2011). Limit analysis of composite materials with anisotropic microstructures: A homogenization approach. *Mechanics of materials*, 43(10), 574-585.

[15] Li, H. X. (2013). Microscopic limit analysis of cohesive-frictional composites with non-associated plastic flow. *European Journal of Mechanics-A/Solids*, 37, 281-293.

[16] Li, H. X. (2013). A microscopic nonlinear programming approach to shakedown analysis of cohesive–frictional composites. *Composites Part B: Engineering*, 50, 32-43.

[17] Le, C. V., Nguyen, P. H., Askes, H., & Pham, D. C. (2017). A computational homogenization approach for limit analysis of heterogeneous materials. *International Journal for Numerical Methods in Engineering*, 112(10), 1381-1401.

[18] Ho, P. L., Le, C. V., & Nguyen, P. H. (2021). Kinematic yield design computational homogenization of micro-structures using the stabilized iRBF mesh-free method. *Applied Mathematical Modelling*, 91, 322-334.

[19] Liu, G. R., & Karamanlidis, D. (2003). Mesh free methods: moving beyond the finite element method. *Appl. Mech. Rev.*, 56(2), B17-B18.

[20] Chen, J. S., Wu, C. T., Yoon, S., & You, Y. (2001). A stabilized conforming nodal integration for Galerkin mesh-free methods. *International journal for numerical methods in engineering*, 50(2), 435-466.

[21] Le, C. V., Gilbert, M., & Askes, H. (2009). Limit analysis of plates using the EFG method and second-order cone programming. *International Journal for Numerical Methods in Engineering*, 78(13), 1532-1552.

# Numerical modelling investigation of wave overtopping on stepped-type seawall structures

**Bao-Loi Dang[1,2], Viet Quoc Dang[2], Magd Abdel Wahab[1], H. Nguyen-Xuan[3]**

[1]Department of Electrical energy, metal, mechanical construction and systems, Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium.
[2]Department of Civil Engineering, Mientrung University of Civil Engineering, Tuy Hoa city, Vietnam.
[3]CIRTech Institute, Ho Chi Minh University of Technology (HUTECH), Ho Chi Minh City, Vietnam.

*Presenting author: baoloi.dang@ugent.be
†Corresponding author: baoloi.dang@ugent.be

## Abstract

This paper presents a large-scale 2-D numerical simulation of wave-structure interactions using IH2VOF which is one of the most advanced RANS models. A non-conventional seawall structure was deployed on Vietnam's coastline as a case study. The value of mean wave overtopping discharge is used to evaluate the hydraulic performance of such construction. It is one of the most significant factors to consider while assessing the feasibility of proposed designs. For novel structure design with complex shape, it is inconvenient to adopt the traditional methods like empirical formulae, or the current approaches using machine leaning algorithms due to the geometrical difference. Therefore, numerical modelling is considered for analyzing such types of structure as a supplement to the experimental study. The reliability of numerical model initially is validated by experimental data, and theoretical results as well. Thereafter, the large-scale numerical model of wave-structure interaction is conducted with various wave conditions. The good agreement is found when comparing the numerical results to empirical solution and prediction methods using neural networks. Hence, the numerical set-up model and calibration for this problem are proposed.

**Keywords:** coastal structure, wave overtopping, RANS models, VOF method.

## Introduction

Seawall is important coastal protection construction that prevent coastal damage from tides and sea waves attacks. Overtopping has been identified as one of the main causes of coastal failure in a number of studies, making overtopping research important for coastal design. Saville conducted a number of physical model experiments on regular wave run-up and overtopping on sloping revetment in the early 1950s [1]. Following that, Owen and Van der Meer conducted more comprehensive model testing on a variety of overtopping influencing variables, and their empirical equations have become important standards and references for coastal engineering[2-4]. However, these empirical formulations are becoming increasingly difficult to adapt to novel design conditions due to complicated coastal structure profiles and extreme wave attacks. As a result, specific physical model tests, which are undoubtedly costly and time-consuming, must be carried out for such scenario.

More numerical models are being used to study wave-structure interactions, including wave overtopping, thanks to advancements in wave theory and computation technology. Kobayashi et al. computed wave overtopping against a sloping dike using the nonlinear shallow water equation and compared their results to Saville's experimental data [5]. Hu et al. also employed

the AMAZON model, which is based on the improvement of shallow water equation [6]. In addition, Soliman and Garcia et al. investigated overtopping using the Reynolds averaged Navier-Stokes equation and the Volume of Fluid approach [7, 8]. In recent years, Smoothed Particle Hydrodynamics, a revolutionary meshfree Lagrange treatment, has been widely used in numerical modeling of free surface fluid, particularly in single phase and multiphase hydrodynamic problems with substantial deformation, complicated boundary, and material interface [9, 10]. In a variety of situations, including coastal structures, vertical breakwaters, and other ocean structures, the IH2VOF hydrodynamic model has been used to predict wave run-up and overtopping. Because of their ease of use and low cost, empirical formulas based on physical model testing are commonly used to forecast wave run-up and mean overtopping discharge [11-13].

## Numerical model

The IH2-VOF model, which is a two-dimensional model that describes water particle movement by solving Navier-Stokes-type equations on a structured grid, is utilized in this study. It tracks the free surface using the Volume of Fluid approach [14]. The IH2VOF user manual and Lara et al. provide more details on the governing equations and numerical techniques [15, 16]. In this section, the model set-up process and calibration are presented. It should be noted that the numerical model in this study was developed at the prototype scale.

### Boundary condition

Solitary, irregular, and regular waves may be generated at the boundary with IH2VOF. We may import the wavemaker movement, the time series of the free surface elevation or specify intended wave characteristics, and the model regenerates the time series of velocity and free surface depending on the input. The first approach was utilized in model validation to mimic the exact identical wave signals utilized in both the physical (see Figure 1&2). The movement of wave paddle in experiment is assigned into numerical paddle in order to rebuild the wave series for the boundary. For large-scale case study in this investigation, the dynamic wave-paddle is utilized by specify the intended wave characteristics as listed in Table 1. In addition, the wave paddle is also defined as an active wave absorb system to reduce the effect of reflected waves.
The real scale two-dimensional geometry of structure is extracted from planning drawings of project provided by BUSADCO [17] and defined in IH2VOF. To ensure the accuracy of wave propagation, the distance from wave paddle to structure is about 2.5-3 times of wavelength. The detail of model set-up as well as location of wave gauge are shown in Figure 7.

### Meshing generation

In numerical model, meshing is an important aspect to guarantee the accuracy the convergence of model. The ideas and conclusions of prior modeling research with IH2VOF, such as Lara and Maza is considered while developing the model in this work [18]. For instance, at least 10 cells per wave height and 100 cells per wavelength of the incident waves are required for mesh generation. The final grid was determined based on a sensitivity analysis of the grid resolution in order to achieve the highest level of accuracy in the shortest possible time. The detail of meshing will be described for each problem in the following sections.

*Model validation and application*

The reliability of numerical model can be found in many studies in the literature. In this work, numerical model is again validated by comparing with experimental results. The experiment conducted in Universitat Politecnica de Catalunya, investigates the wave propagating and pressure acting on a slope with vertical wall [19]. Experimental set-up such as water depth, wave gauges position, etc. can be seen in Figure 1. The wave paddle movement in experiment (see Figure 2) is assigned to numerical wave paddle for modelling. A uniform mesh is generated based on the model domain which is dx=0.01m and dy=0.02m corresponding to horizontal and vertical directions. The wave gauge WG#2 and WG#3 with 2.83m and 4.12m far from wave paddle is chosen for analyzing the surface elevation as shown in Figure 3. The water surface elevation performed by numerical model seems slightly smaller than the experimental one, but a good agreement can be found visibly in term of wave propagation. It is confident to adopting this numerical model in next steps.
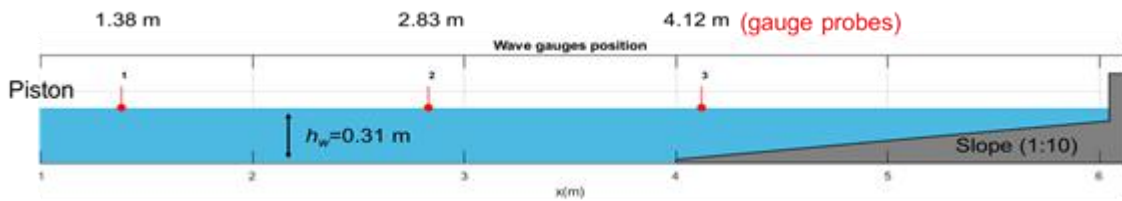


Figure 1. Set-up of numerical for validation
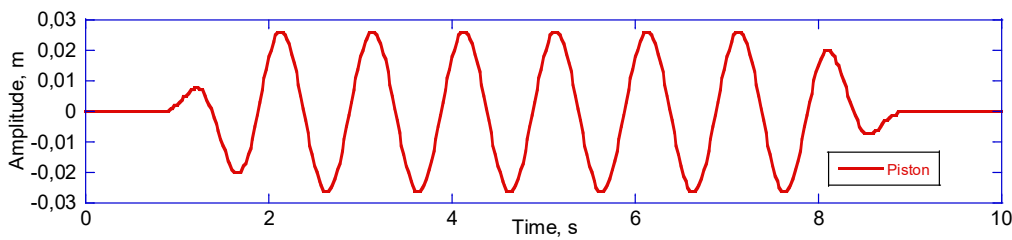


Figure 2. Wave paddle movement used in experiment.
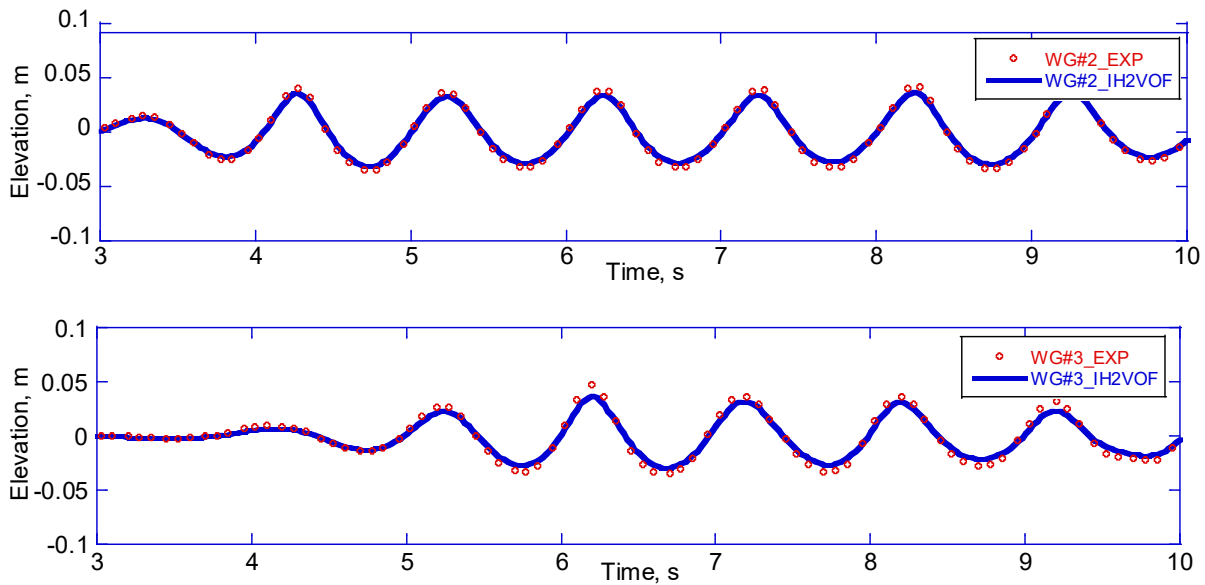


Figure 3. Comparison of free surface elevation between experimental and numerical model at wave gauge #2&#3.

In this step, the performance of the model in predicting mean overtopping discharge on a new seawall structure is investigated under regular in irregular wave attacks as listed in Table 1. The total numerical domain is 100m of length and 15m of heigh. The distance from wave generation boundary to structure is about 85m which is in range of 2.5-3 times of wavelength as shown in Figure 7. Wave gauges are located at distances 20m, 40m, 60m, and 82m from wave paddle for monitoring the water elevation. A gauge at distance 92.1m is used for obtaining wave overtopping discharge. For stable and accurate modelling, a uniform mesh is performed which follows the recommendations in previous studies. In particular, dx=0.1m and dy=0.2m, resulting 500 cells and 150 cells corresponding to x-direction and y-direction, respectively.

Firstly, the large-scale numerical model is validated against theoretical one in term of wave propagation. For instance, a numerical wave tank without structure is modelled as shown in Figure 4. The left boundary is defined as wave generation (both regular and irregular wave) by specifying wave characteristics as well as is used as active wave absorption system to absorb the reflected waves. The right boundary is defined as passive wave absorption system to damp the incident waves. The total time of simulation is 200 seconds. As seen in the Figure 5, the result of water elevation at the wave gauge WG#3 is compared with the theoretical regular waves. The well-agreement can be found visibly. For irregular wave, it is quite difficult to reproduce the irregular wave exactly comparing to theoretical one, however the shape and trend of numerical irregular wave well follows the theoretical one as shown in Figure 6. Therefore, it is acceptable to procedure the analysis of wave-structure interaction using this model.
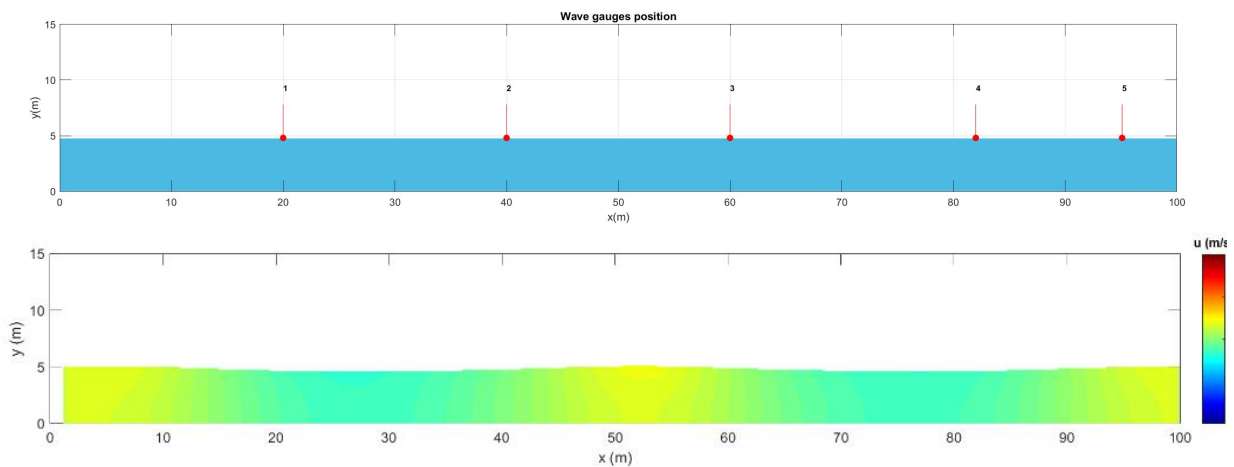
Figure 4. Numerical wave tank for large scale model verifecation.
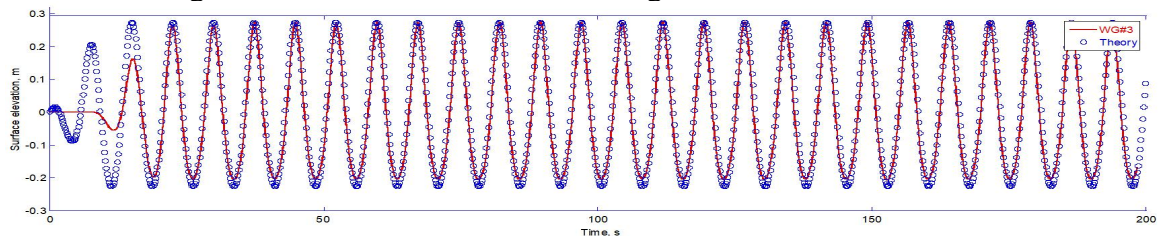
Figure 5. Comparison of free surface elevation between theoretical and numerical model at wave gauge #3 for regular wave.
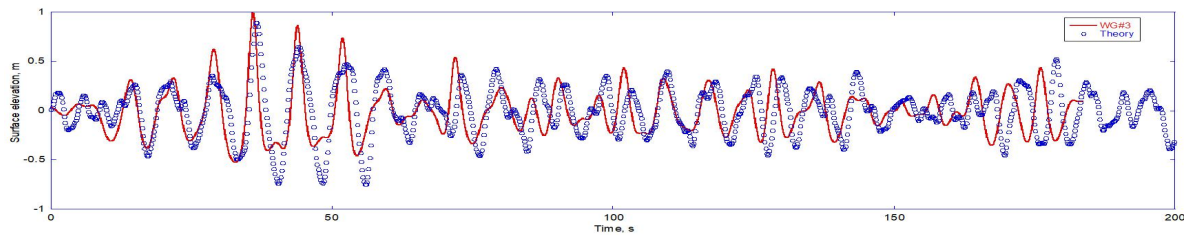
Figure 6. Comparison of free surface elevation between theoretical and numerical model at wave gauge #3 for irregular wave.
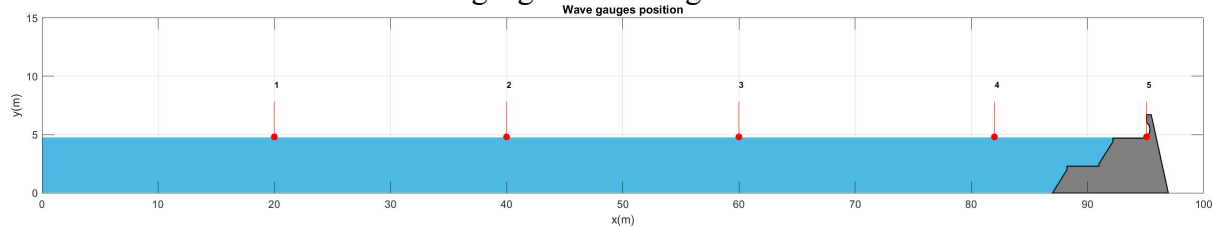


Figure 7. Large-scale numerical model of case study in Vietnam
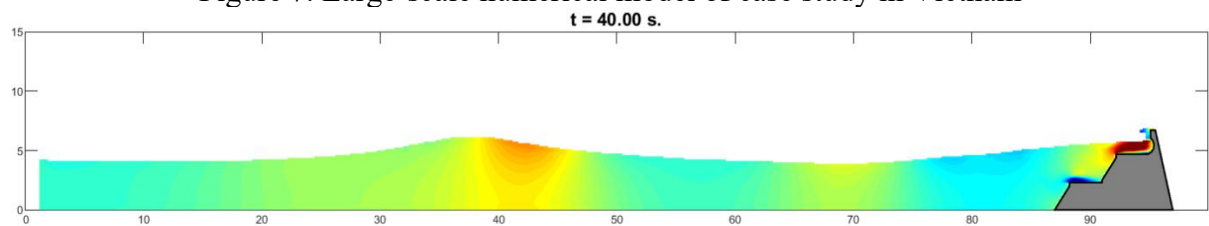


Figure 8. Snapshot of wave surface elevation at 40 second of simulation time.

Table 1. Wave condition used in modelling of case study.

| Regular waves conditions | | | | |
|---|---|---|---|---|
| Parameter | RE#1 | RE#2 | RE#3 | RE#4 |
| $H_m$ (m) | 1.35 | 1.50 | 1.95 | 2.10 |
| $T_m$ (s) | 8.67 | 7.64 | 8.35 | 9.20 |
| Irregular waves conditions | | | | |
| Parameters | IR#1 | IR#2 | IR#3 | IR#4 |
| $H_s$ (m) | 1.68 | 1.78 | 2.07 | 2.36 |
| $T_s$ (s) | 8.99 | 9.31 | 9.32 | 9.24 |

## Result and discussion

The performance of the model in predicting mean overtopping discharge on a new seawall structure is investigated under regular and irregular wave attacks. Figure 9 shows the example of numerical result of water elevation captured at wave gauge #2 and wave gauge #3 for regular wave cases. It is seen that numerical model maintain the regular of predefined waves.
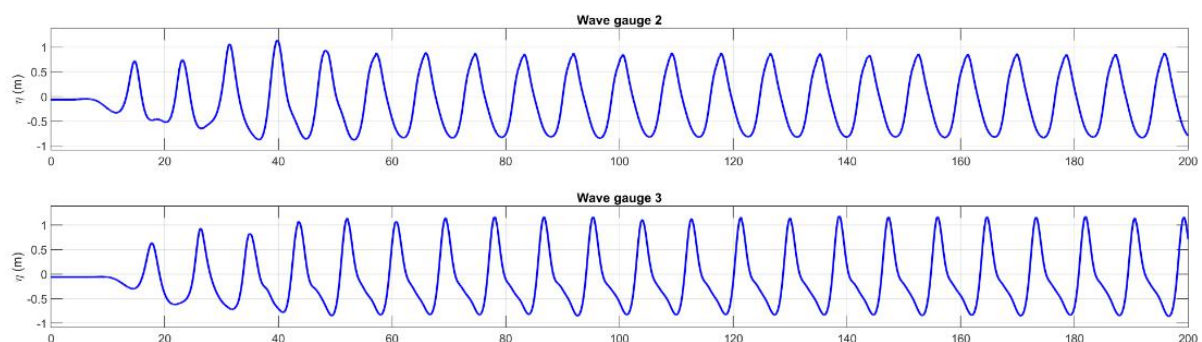


Figure 9. Numerical results of free surface elevation at wave gauges #2 and #3 in case study.

For comparison purpose, the empirical formula (EurOtop, 2018) for calculating the mean overtopping wave discharge is used. The detail of formula and calculation can be found in the EurOtop manual [20]. In addition, a recent approach which is based on machine learning algorithms is also adopted [21]. Figure 10 and Figure 11 show the comparison of the prediction of mean overtopping discharge for different methods under regular and irregular wave, respectively. In general, the numerical results are well-agreed with empirical and neural network ones. For small value of relative crest board ($R_c/H_s$) which is $H_s$ is high, the numerical results are quite not accurate. The reason may be that with the wave height being too large, wave propagation is not able to take place smoothly in the low water depth flume. In practice, a large wave height will correspond to a large water depth or a particular extreme situation. The detail result of mean overtopping discharge predicted by various methods is presented in Table 2.



(a)  (b)
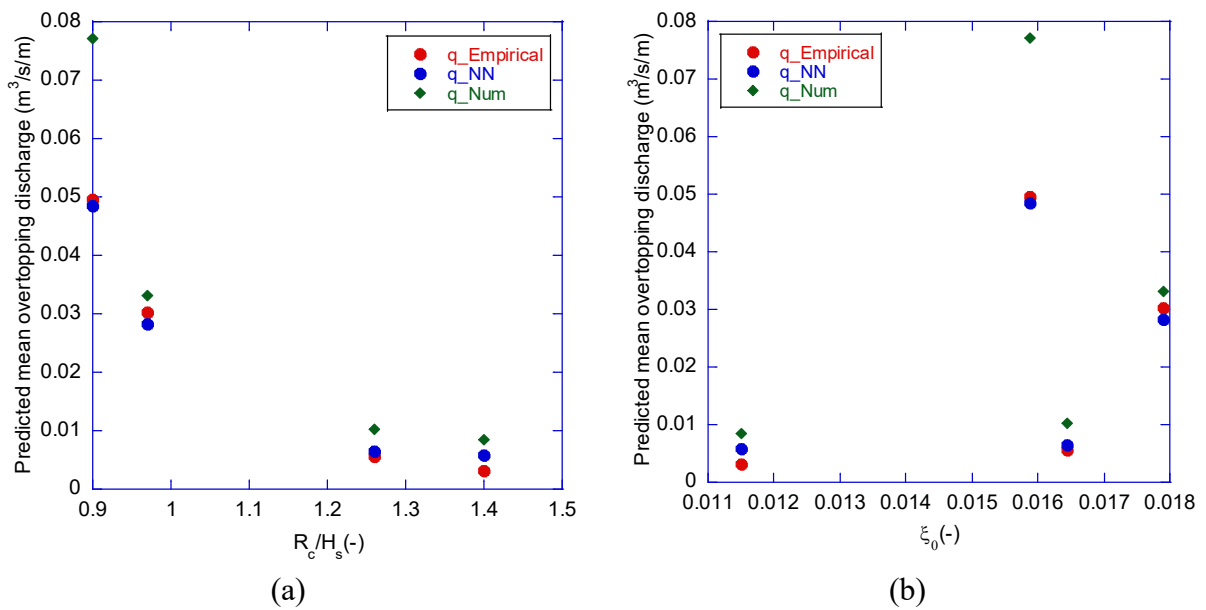
Figure 10. Comparison of mean wave overtopping discharge under regular wave conditions for different methods in term of relative crest board ratio (a), and wave steepness (b).



(a)  (b)
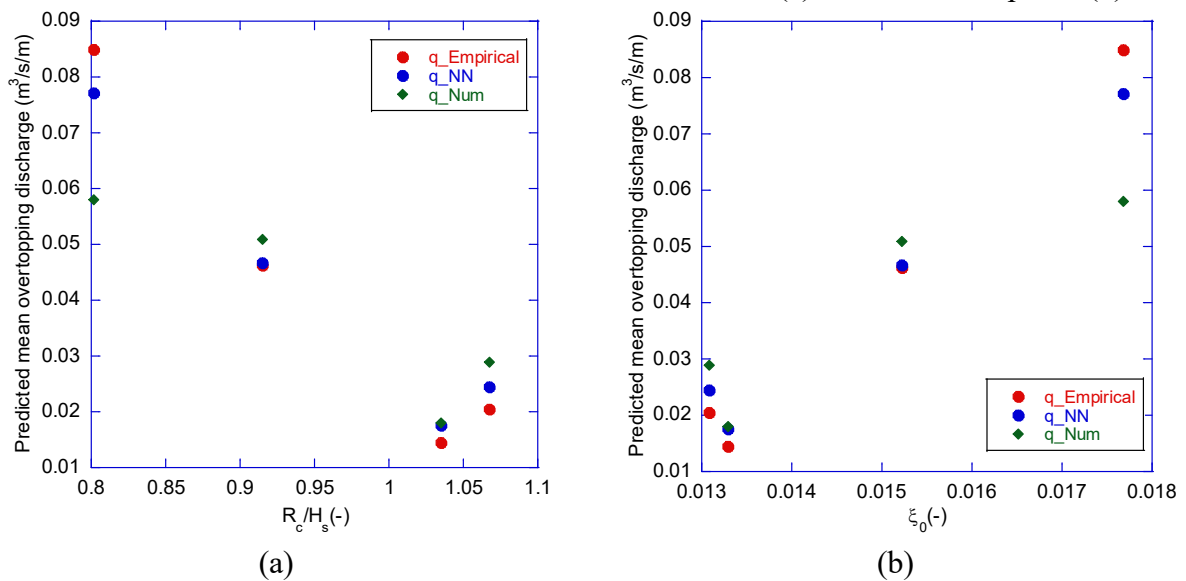
Figure 11. Comparison of mean wave overtopping discharge under irregular wave conditions for different methods in term of relative crest board ratio (a), and wave steepness (b).

Table 2. Comparison of the mean overtopping discharge for different approaches.

| Test case | Label | $H_s$ | $T_s$ | Empirical formula | Neural network prediction | Numerical model (present study) |
|---|---|---|---|---|---|---|
| | | *m* | *s* | *m³/s/s* | *m³/s/s* | *m³/s/s* |
| Regular waves | RE#1 | 1.35 | 8.67 | 0.00321 | 0.00586 | 0.00849 |
| | RE#2 | 1.50 | 7.64 | 0.00562 | 0.00656 | 0.01021 |
| | RE#3 | 1.90 | 9.20 | 0.03034 | 0.02823 | 0.03307 |
| | RE#4 | 2.10 | 9.24 | 0.04950 | 0.04852 | 0.07706 |
| Irregular waves | IR#1 | 1.68 | 8.99 | 0.01436 | 0.01760 | 0.01796 |
| | IR#2 | 1.78 | 9.31 | 0.02044 | 0.02435 | 0.02895 |
| | IR#3 | 2.07 | 9.32 | 0.04631 | 0.04671 | 0.05080 |
| | IR#4 | 2.36 | 9.24 | 0.08494 | 0.07714 | 0.05799 |

## Conclusion

A numerical mesh-based method for modelling wave overtopping discharge on novel seawall structure has been presented in this work. The model was firstly verified by using experimental and theoretical results, then adopted to investigate the wave overtopping discharge on a novel seawall structure.

The numerical results show that well-agreement can be found when comparing with other methods, i.e., empirical and neural network-based solutions. It is noticed that for small ratio of relative crest board ($R_c/H_s$) the numerical results are quite different from other approaches, but the overall performance of model is acceptable. This brings out the promising applications for such similar problems in further.

## Acknowledgment

### References

1. Saville, T. *Laboratory data on wave run-up and overtopping on shore structures*. 1955.
2. Owen, M.W. and A. Steele. *Effectiveness of recurved wave return walls*. 1993.
3. van der Meer, J. and H. Janssen, *Wave run-up and wave overtopping at dikes*. 1995.
4. van der Meer, J., *Technical Report Wave Run-up and Wave Overtopping at Dikes*. 2002, Advisory Committee on Flood Defence, Delft.
5. Kobayashi, N. and A. Wurjanto, *Wave Overtopping on Coastal Structures*. Journal of Waterway, Port, Coastal, and Ocean Engineering, 1989. **115**(2): p. 235-251.
6. Hu, K., C.G. Mingham, and D.M. Causon, *Numerical simulation of wave overtopping of coastal structures using the non-linear shallow water equations*. Coastal Engineering, 2000. **41**(4): p. 433-465.
7. Soliman, A.S.M., *Numerical study of irregular wave overtopping and overflow*. 2003.

8.  Garcia, N., J.L. Lara, and I.J. Losada, *2-D numerical analysis of near-field flow at low-crested permeable breakwaters.* Coastal Engineering, 2004. **51**(10): p. 991-1020.

9.  Gingold, R.A. and J.J. Monaghan, *Smoothed particle hydrodynamics: theory and application to non-spherical stars.* Monthly Notices of the Royal Astronomical Society, 1977. **181**: p. 375.

10. Monaghan, J.J., *Smoothed particle hydrodynamics.* Reports on Progress in Physics, 2005. **68**(8): p. 1703-1759.

11. Neves, M., et al., *Wave Overtopping of Póvoa de Varzim Breakwater: Physical and Numerical Simulations.* Journal of Waterway Port Coastal and Ocean Engineering-asce - J WATERW PORT COAST OC-ASCE, 2008. **134**.

12. Di Lauro, E., et al., *Advantages of an innovative vertical breakwater with an overtopping wave energy converter.* Coastal Engineering, 2020. **159**: p. 103713.

13. Inverno, J., et al., *Numerical simulation of wave interacting with a submerged cylinder using a 2D RANS model.* Journal of Hydro-environment Research, 2016. **12**: p. 1-15.

14. Hirt, C.W. and B.D. Nichols, *Volume of fluid (VOF) method for the dynamics of free boundaries.* Journal of Computational Physics, 1981. **39**(1): p. 201-225.

15. Lara, J.L., N. Garcia, and I.J. Losada, *RANS modelling applied to random wave interaction with submerged permeable structures.* Coastal Engineering, 2006. **53**(5): p. 395-417.

16. Torres-Freyermuth, A., J.L. Lara, and I.J. Losada, *Numerical modelling of short- and long-wave transformation on a barred beach.* Coastal Engineering, 2010. **57**(3): p. 317-330.

17. *Ba Ria – Vung Tau Urban Sewerage and Development One Member Limited Company (BUSADCO)*. Available from: www.busadco.com.vn.

18. Lara, J.L., et al., *Large-scale 3-D experiments of wave and current interaction with real vegetation. Part 1: Guidelines for physical modeling.* Coastal Engineering, 2016. **107**: p. 70-83.

19. Aller, A.B. *Smoothed Particle Hydrodynamics model for civil and coastal engineering applications*. 2015.

20. van der Meer, J., et al., *EurOtop: Manual on wave overtopping of sea defences and related sturctures - An overtopping manual largely based on European research, but for worlwide application (2nd edition)*. 2016.

21. van Gent, M.R.A., et al., *Neural network modelling of wave overtopping at coastal structures.* Coastal Engineering, 2007. **54**(8): p. 586-593.

# Optimization analysis of stiffened composite plate using iJaya algorithm

**\*Lam Phat Thuan[1], †Nguyen Hoai Son[1], Nguyen Nhut Phi Long[2], Nguyen Quan[3], Doan Dinh Thien Vuong[1]**

[1]Faculty of Civil Engineering, HCMC University of Technology and Education (HCMUTE), Viet Nam
[2] Faculty of Mechanical Engineering, HCMC University of Technology and Education (HCMUTE), Vietnam
[3]Faculty of Engineering – Technology, Pham Van Dong University (PDU),

\* Presenting author: thuanlp@hcmute.edu.vn,
†Corresponding author: sonnh@hcmute.edu.vn,

## Abstract

In this paper, an improved version of Jaya algorithm, called iJaya, is proposed to optimize the fiber orientation of the stiffened composite plate. The improvement has been executed in the selection step of the original algorithm, in which one-third of the next population is chosen by the original procedure, while two-thirds are selected by the elitist technique. This way of modification aims at effectively balancing the exploration and exploitation capabilities of the algorithm and lead to the improvement in convergence rate. The improved algorithm is also proven to effectively handle optimization problems with both continuous and discontinuous variables.

The iJaya algorithm is then applied to solve the optimization problem of the stiffened composite plate with the design variables are the fiber angles and the objective function is the strain energy of the plate. The Numerical results obtained from solving optimization problems show that iJaya outperforms some other methods and prove the effectiveness and the efficiency of the proposed method.

**Key words**: Stiffened composite plate, improved Jaya algorithm, optimization analysis, composite structure

## 1. Introduction

Optimization is a very potential field and has attracted the interest of many scientists. Over the past decades, many optimal methods have been researched, developed, improved and applied in many fields. The algorithms can be classified into two main groups: gradient-based and popular-based approach. The gradient-based method group has the advantage of finding the optimal solution quickly, however, the drawback is that the solution is easy to be trapped in the local extrema. Moreover, they are limited to problems with continuous constraint and objective functions. Some of the prominent algorithms of gradient-based groups include: sequential linear programming (SLP) [1,2], sequential quadratic programming (SQP) [3,4], steepest descent method, and Newton's method. Population based methods are capable of handling both contiguous and discrete variables and are easy to be implemented. Moreover, it is able to avoid local extrema and reach the global optimization. Some of famous population-based methods are: genetic algorithm (GA) [5], particle swarm optimization (PSO) [6], differential evolution (DE) [7], artificial bee colony (ABC) [8] and Jaya algorithm, etc.

Among them, the Jaya algorithm recently proposed by Rao [9] and its variants have appeared as one of the most efficient and effective algorithm. The Jaya algorithm has been applied in many engineering optimization problems and has obtained very good computational efficiency [10,11,12,13,14]. However, selecting the population of the next generation based on the paired comparison as in the original Jaya algorithm may lead to the possibility that a

good individual might be eliminated when paired with a stronger one and hence decrease the convergence speed. On the other hand, selecting the next generation population based only on the best individuals in the entire population would reduce the exploration of the algorithm and therefore may results in a premature convergence.

In order to select good individuals while ensuring a balance between exploration and exploitation, a new improved version of Jaya algorithm, called iJaya, is proposed in this paper. In the iJaya algorithm, the population for the next generation is selected from two different group with different selection techniques. This improvement of selection step aims at effectively balancing between the exploration and exploitation capabilities of the algorithm, and hence the convergence rate is improved, while the quality of the optimal solution is still guaranteed

In another aspect, composite materials are increasingly used in many engineering fields because of their superior properties compared to other materials [15,16]. Among many composite structures used, the stiffened composite plate structure is one of the most prominent and increasingly popular structure [17,18,19]. By combining the advantages of the composite plate with the reinforcing beam, the stiffened composite panel is both high bending strength and lightweight. Thanks to its high applicability in the fields of aerospace, construction, engineering, etc, it is essential to optimize the design of the stiffened composite plate structure to save costs and increase efficiency. However, due to the difficulty of computing the behavior of such a complex structure, finding the optimal design also requires an efficient and highly accurate algorithm.

In this paper, the iJaya algorithm will be applied to solve a number of unconstrained and constrained optimization problems of stiffened composite plate structures including: fiber angle and thickness optimization problems. The solutions obtained from the iJaya algorithm will be compared with the reference solutions to prove the efficiency and accuracy of the proposed algorithm.

## 2. Brief introduction on stiffened composite plate



**Figure 1. Model of a stiffened composite plate**

- Stiffened composite plate are created by combining a composite panel with one or more composite beams as in Figure 1.

- The composite beam used in this structure is a Timoshenko beam and serves as a stiffener.

- The neutral plane of the beam and plate is spaced by a distance *e*

- The beam-plate structure is modeled by FEM based on a set of nodes. Each node of the plate element has five degrees of freedom (DOF) and is represented by the vector $\mathbf{d} = [u, v, w, \beta_x, \beta_y]^T$, where $u, v, w$ and $\beta_x, \beta_y$ are the displacements at the middle of the plate and the rotations around the *y*-axis and *x*-axis, respectively. Each node of the beam

also has the DOF $\mathbf{d}_{st} = [u_r, u_s, u_z, \beta_r, \beta_s]^T$, where, $u_r, u_s, u_z$ are centroid displacements of beam and $\beta_r, \beta_s$ are the rotations of the beam around r-axis and s-axis.

- For static analysis, the global equations governs the behavior of the stiffened composite plate is $[\mathbf{K}]\{\mathbf{d}\} = \{\mathbf{F}\}$ and can be found in [20]

## 3. Improvement of Jaya algorithm

### 3.1 The original Jaya algorithm

Jaya is a simple yet powerful population-based optimization algorithm, first time proposed by Rao [9] in 2016, for solving constrained and unconstrained optimization problems. This algorithm is based on the concept that the optimal solution for a given problem should move closer to the best solution and should avoid the worst solution. The Jaya algorithm requires only the common control parameters and does not require any algorithm-specific control parameters. Jaya algorithm includes in three simple steps: initialization, modification and selection. The Figure 2 shows flowchart of the original Jaya algorithm.
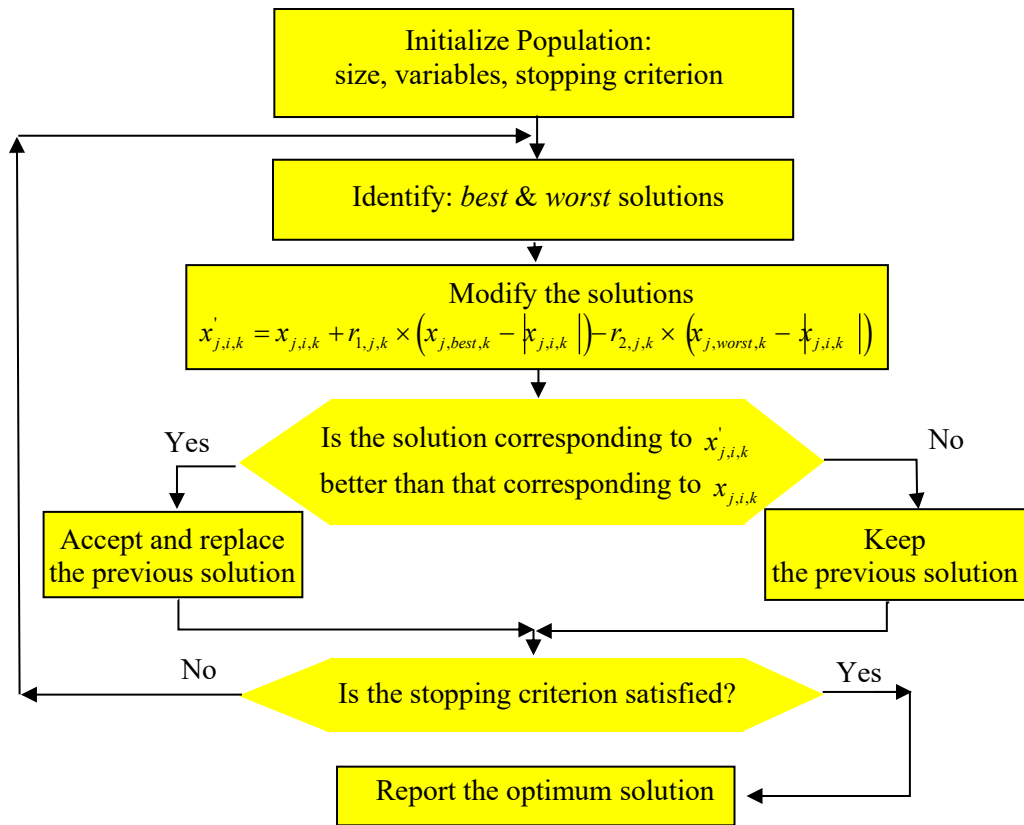


**Figure 2. Flowchart of the original Jaya algorithm**

First, a population of $N_p$ individuals is randomly initialized in the search space. Each candidate is a vector consisting of n design variables $x = (x_1, x_2, ..., x_n)$ with values in the upper and lower bounds:

$$x_{j,i} = x_{j,i}^l + rand[0,1] \times \left( x_{j,i}^u - x_{j,i}^l \right), j = (1, 2, ..., n), i = (1, 2, ..., N_p) \tag{1}$$

where $x_{j,i}^l$ and $x_{j,i}^u$ are the upper and lower bounds of the design variable $x_j$; $rand[0,1]$ generates a random number within the interval $[0,1]$.

Let *f(x)* be the objective function of the optimization problem. The fitness of each candidate in the population is then evaluated via the values of *f(x)*. The best ( $x_{best}$ ) and the worst ( $x_{worst}$ ) candidates denote the individuals with the best and worst fitness function values in the population entire, respectively. If $x_{j,i,k}$ is the value of the *j*th variable of *i*th candidate in the *k*th iteration, then a new vector $x'_{j,i,k}$ will be generated by stochastically modifying $x_{j,i,k}$ as follows:

$$x'_{j,i,k} = x_{j,i,k} + r_{1,j,k} \times \left(x_{j,best,k} - \left|x_{j,i,k}\right|\right) - r_{2,j,k} \times \left(x_{j,worst,k} - \left|x_{j,i,k}\right|\right) \tag{2}$$

where $x_{j,best,k}$ and $x_{j,worst,k}$ are the values of *j*th variable corresponding to the best ( ) and the worst candidate in the entire population at *k*th iteration. $r_{1,j,k}$ and $r_{2,j,k}$ are the random numbers in the range of [0,1]. the term $r_{1,j,k} \times \left(x_{j,best,k} - \left|x_{j,i,k}\right|\right)$ indicates the tendency of the design variable moving closer to the best solution and the term $r_{2,j,k} \times \left(x_{j,worst,k} - \left|x_{j,i,k}\right|\right)$ shows the tendency of the solution moving away from the worst one.

At the final step, the objective function values calculated from $x_{j,i,k}$ and $x'_{j,i,k}$ will be the criterion for choosing which candidate will be held for the next generation. All accepted candidates at the end of iteration are maintained and become the input to the next iteration

$$x_{i,k+1} = \begin{cases} x'_{i,k} & \text{if } f(x'_{i,k}) \leq f(x_{i,k}) \\ x_{i,k} & \text{otherwise} \end{cases} \tag{3}$$

### 3.2 The improvement of Jaya algorithm

In the original Jaya algorithm, the population of the next generation is selected by comparing individual pairs (paired comparison) based on their fitness function values. This results in a good individual being disqualified when compared to a stronger individual in the population. This means that an individual that loses in one pair can outperform the winner in another.

On the other hand, selecting the next generation population based only on the best individuals in the entire population would reduce the exploration of the algorithm and therefore may results in a premature convergence. In order to select good individuals while ensuring a balance between exploration and exploitation, a new 2-step modification of optimization procedure is proposed.

In that, the population of the next generation is firstly separated into two different groups. Specifically, the first group holds one-third of the population size (N/3) and the rest of them (2N/3) belongs to the second group. Secondly, for each group of population, different selection techniques are applied to choose the suitable candidates for the next generation. When the algorithm reaches the selection step, one-third of the individuals of the next population set are selected in the first group using a paired comparison technique as in the original Jaya algorithm. Meanwhile, the rest of the next population are selected from the second group using the elitist procedure [21].

Particularly, two third of the parent population and offspring population are combined and haft best individuals of them are selected for the next generation based on their fitness values. The modification of selection step can be illustrated as in Figure 3. This algorithm can ensure the balance between the exploration and exploitation capabilities effectively and effectively

help increase the convergence speed of the algorithm.

| Group 1 | | $X_i$ | |
|---|---|---|---|
| 7 | 5 | 7 | |
| 1 | 2 | 2 | |
| 5 | 9 | 9 | |
| 4 | 3 | 4 | |

| Group 2 | | $X_i$ |
|---|---|---|
| 1 | 2 | 9 |
| 2 | 5 | 8 |
| 6 | 9 | 8 |
| 7 | 8 | 7 |
| 8 | 4 | 7 |
| 3 | 7 | 6 |

Pair Comparison · Elitist Selection Technique

**Figure 3. Improvement of selection phase of the original Jaya algorithm**

## 4. Numerical results

In this section, an improved version of Jaya algorithm (iJaya) will be applied to solve two problems of stiffened composite plate to verify the efficiency and accuracy of the algorithm. In the first problem, iJaya algorithm is applied to find the optimal fiber angle of the reinforced composite plate to achieve the minimum objective function of strain energy. In the second problem, the objective function to be optimized is the weight of the stiffened composite plate with the design variable being the thickness of the composite plate and reinforcement beam. The obtained results are compared with the reference solutions to analyze the efficiency, accuracy and the convergence rate of the algorithm.

To facilitate the investigation of example problems, models and parameters of stiffened composite plate will be briefly presented as below:

- Two types of plates studied in this section are square and rectangular plates reinforced in the X, Y or X-Y directions. All the plate models are under simply-supported condition.

- Figure 4 illustrates a model of an X-direction reinforced composite plate with basic dimensional parameters. Composite plate has the thickness t and the lengths in two directions x, y are $l_x$ and $l_y$, respectively.
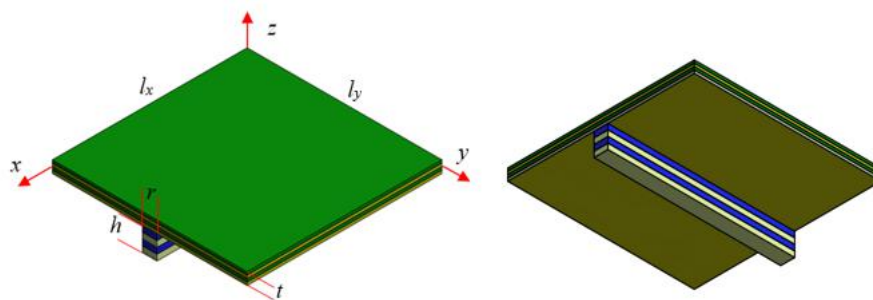


**Figure 4. Parameters of a stiffened composite plate**

- The composite beam has cross-sectional dimension $h$ x $r$, where h is the thickness and r is the beam width. The length of the beams in optimization problems of this section is $l_x$ and $l_y$ corresponding to the cases of reinforcement in the X and Y directions.

### 4.1 Fiber orientation optimization of stiffened composite plate using iJaya

In this section, the fiber orientation optimization of two models of square and rectangular stiffened composite plate are studied to verify the accuracy and the effectiveness of iJaya algorithm. In both cases, the stiffeners are set in X-direction as in Figure 5. The objective

function of the problem is chosen to be the strain energy of the plate.



a) square plate     –     b) rectangular plate
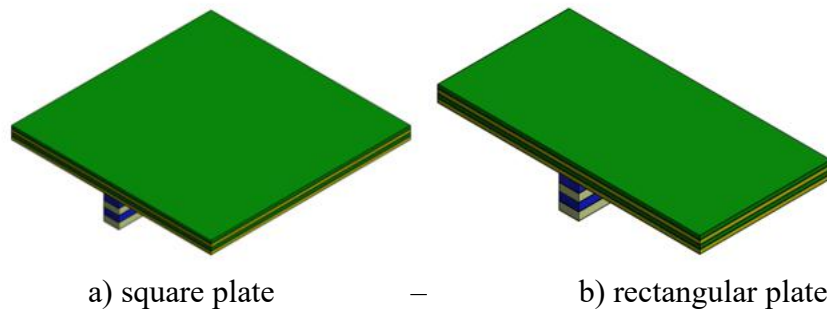
**Figure 5. Models of square and rectangular stiffened composite plates**

The optimization problem can be described as follows:

$$\begin{cases} \min_{\theta} \quad \mathbf{U} = \dfrac{1}{2}\mathbf{d}^T\mathbf{Kd} \\ \text{subject to} \quad 0 \le \theta_i \le 180, i = 1,...,4 \end{cases} \qquad (4)$$

**U** is strain energy

$\theta_i$ is fiber orientation of $i$th layer

- Optimal results of the fiber direction for both cases are presented in Table 1. The design variables used in the problem include both continuous and discrete variables (integers).

- The results reveal that the solutions obtained by the iJaya agree well with those by the SQP [4], DE [22] and Jaya. The strain energy obtained from the iJaya in two cases is better than that of the SQP in the paper of T. Nguyen-Thoi [2], 6183.1 to 6363.8 for the square plate and 30366 to 31471 for the rectangular plate.

- In terms of computational time, the iJaya method also outperforms the DE and Jaya methods especially for the case of rectangular plate. In this case, the computational cost of iJaya is just 4462 seconds in comparison with 6217 and 7762 of Jaya and DE, respectively.

- Moreover, when comparing the results obtained by using continuous and discrete variables, it is clear that optimization solutions using discrete variables has much better computational time than that of continuous variables. In the case of rectangular plate, the computational cost decreased about 35% from 4462 seconds (continuous variables) to 2896 seconds (discrete variables). This strongly proved the effectiveness and the accuracy of the iJaya method.

**Table 1. The optimal results of two problems with continuous and discrete variable**

| Type | Methods | Optimal Angle [$^o$ Degree] | | | | Strain energy | CPU time (s) |
|---|---|---|---|---|---|---|---|
| | | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | | |
| Square (a=b=254 mm) | SQP [4] | 134.5 | 47.4 | 0 | 179.8 | 6363.8 | 129 |
| | DE [22] | 134.5 | 46.6 | 0.25 | 179.5 | 6364.0 | 2626 |
| | Jaya | 135.0 | 47.9 | 0 | 180 | 6183.1 | 1564 |
| | **iJaya** | **134.9** | **47.8** | **0** | **180** | **6183.1** | **1386** |
| | iJaya (D) | 135 | 48 | 0 | 180 | 6183 | 1274 |
| Rectangular (a=254 mm, | SQP [4] | 160.3 | 35.6 | 0 | 179.8 | 31471 | 154 |
| | DE [22] | 159.9 | 37.1 | 0 | 0 | 30366 | 7762 |

| b=508) | Jaya | 159.2 | 37.0 | 0 | 180 | 30300 | 6217 |
| | **iJaya** | **159.9** | **37.1** | **0** | **180** | **30366** | **4462** |
| | iJaya (D) | 160 | 37 | 0 | 180 | 30365 | 2896 |
| * iJaya (D): *iJaya using discrete variables* | | | | | | | |

### 4.2 Thickness optimization of stiffened composite plate

Consider the problem of optimizing composite plate reinforced by composite beams as shown in Figure 5. 7 under a simply-supported condition. The parameters of the problem are given as follows: the plate has the length $a = 254$ mm and the thickness $t_p$, cross section of reinforced beam has the width of $c_x = 6.35$ mm, and the height is $t_b$. The optimal analysis is carried out with two cases of square plate (b = 254 mm) and rectangular plate (b = 508 mm).

Both plate and beam have a symmetrical four-layer structure. The fiber direction of each layer on the plate is respectively [90 45 45 90], and the fiber layer of the beam is [180 0 0 180].

Beams and plate are made with the same materials with parameters $E_1 = 144.8\,\text{GPa}$, $E_2 = E_3 = 9.65\,\text{GPa}$, $G_{12} = G_{13} = 4.14\,\text{GPa}$, $G_{23} = 3.45\,\text{GPa}$, $\upsilon_{12} = \upsilon_{13} = \upsilon_{23} = 0.3$. The plate is loaded evenly with the value of $f = 0.6895$ (N/mm$^2$).

The problem can be represented as follows:

$$\begin{cases} \min_{t_p, t_b} & Weight(h, d_x) \\ \text{subject to} & Disp \le 1 \\ & \sigma_{Tsai-wu} \le 1 \end{cases} \tag{5}$$

In particular, the objective function is the stiffened composite plate mass and subjected to two constraints including the displacement of the stiffened plate less than 1 and the Tsai-Wu stress is also less than 1.

The thickness optimization problems are solved for three different cases: rectangular plate stiffened in X-direction (R-X), square plate stiffened in X-direction (S-X) and square plate with two sitffeners in X-Y direction (S-XY). The optimal results are presented in Table 2. It is shown from the results that, in the case of square plate, the optimal results obtained from iJaya algorithm agreed well with those of DE, Jaya for all 3 cases S-X, R-X and S-XY. This proves the accuracy of the iJaya algorithm.

In terms of computational costs, the iJaya algorithm proved to be superior to the two algorithms DE, Jaya, especially in the case of a square plate. Specifically, in the case of S-X, the computation time of iJaya was reduced by nearly half compared to DE and the original Jaya algorithm. Meanwhile, in the case of R-X, the computation time is also reduced by up to 13% using the iJaya algorithm.

The results also reveal that the objective function reaches the lowest value for the case of square plate with 2 stiffened beams S-XY. The reason is that two reinforcement beams support the structure better, thereby reducing the size of the plate thickness and contributing to the reduction of the structure weight. Although the difference from the S-X case is only about 10%, the 2-beam reinforcement model gives a better balance to the structure

In the case of R-X, since the beam is reinforced in the X direction (the direction of the smaller dimension), the support capacity of the stiffener to the plate is reduced. This lead to the incresement in the thickness of the plate to ensure the whole structure under the loading. Therefore, the total weight of the structure has also increased significantly.

From the above results, one observation is that the optimal option is usually achieved with the smallest plate thickness combined with the more number of stiffened beams.
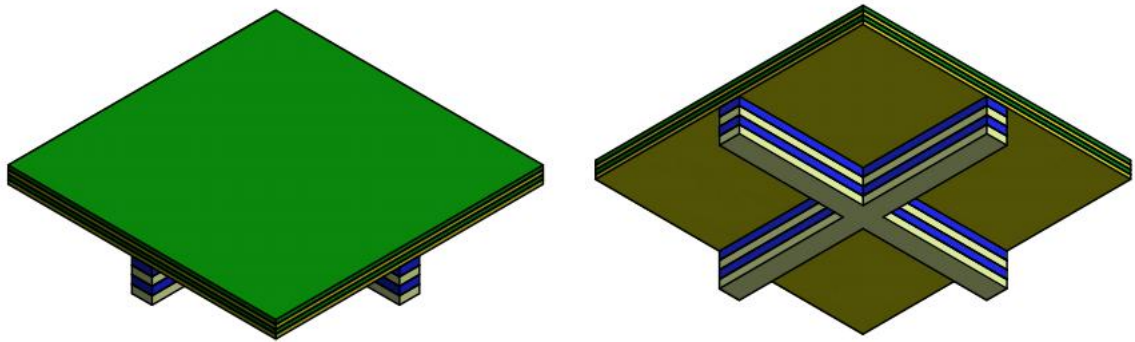


**Figure 6. Square composite plate stiffened by two beams**

**Table 2. Thickness optimization of square and rectangular stiffened composite plates**

| Type | Method | Thickness | | Weight | CPU's time |
| | | $t_p$ | $t_b$ | | |
|---|---|---|---|---|---|
| S-X (a=b=254mm) | DE [22] | 13 | 83 | 1.351135507 | 2545 |
| | Jaya | 13 | 83 | 1.351135507 | 2684 |
| | **iJaya** | **13** | **83** | **1.351135507** | **1469** |
| | | | | | |
| R-X (a=254 mm b=508 mm) | DE [22] | 18 | 20 | 3.271406038 | 2984 |
| | Jaya | 18 | 20 | 3.271406038 | 2903 |
| | **iJaya** | **18** | **20** | **3.271406038** | **2565** |
| | | | | | |
| S-XY (a=b=254mm) | DE | 10 | 66 | 1.192046584 | 1803 |
| | Jaya | 10 | 65 | 1.191838584 | 7496 |
| | **iJaya** | **10** | **66** | **1.192046584** | **1393** |

### 4.3 Convergence of the iJaya algorithm

Based on the results obtained in the Table II, in terms of computational efficiency as well as accuracy, iJaya algorithm is completely superior to the two methods DE and Jaya. This can also be seen through the convergence curves shown in Figure 7 for the case of S-X. As seen, the iJaya algorithm almost reaches the optimal target function after just over 200 loops, while Jaya and DE algorithm takes more than 300 structural analyzes to reach this value.

**Figure 7. Convergence curves for the case of square plate reinforced in X-direction**

## 5. Conclusion

In this paper, an improved version of the Jaya algorithm (iJaya) has been introduced. The iJaya algorithm is developed by modifying the selection 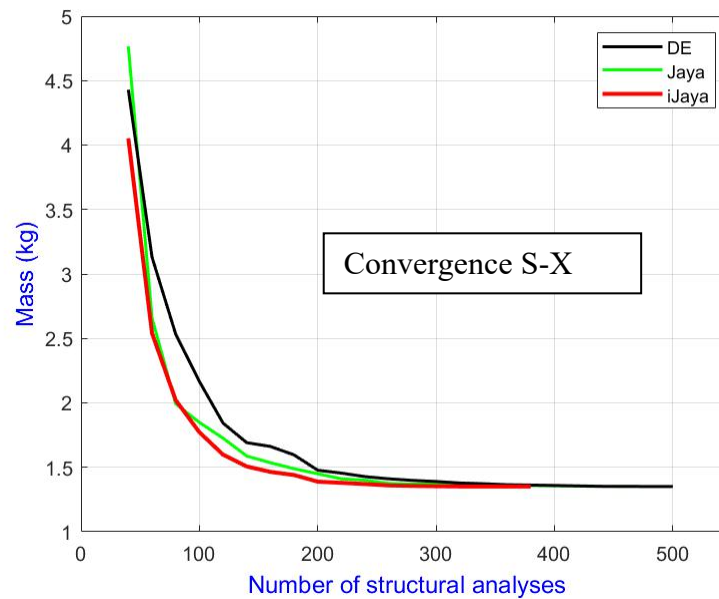step of the original algorithm. By dividing the population into two groups and selecting it by two different techniques: paired comparison and elitist selection technique, the iJaya algorithm ensures a balance between exploitation and exploration.

The iJaya algorithm is then applied to solve two optimization problems of reinforced composite plates, which are the optimization of the fiber direction angle and the optimization of the thickness of the structure. The results achieved showed that iJaya algorithm has very high accuracy and efficiency.

**References**

[1] L, Lamberti; C, Pappalettere;, "Comparison of the numerical effi ciency of different sequential linear programming based algorithms for structural optimisation problems," *Comput,* p. 713 – 728, 2000.

[2] L, Lamberti; C, Pappalettere;, "Improved sequential linear programming formulation for," *Comput Methods Appl Mech Eng,* vol. 193, p. 3493 –3521, 2004.

[3] S. R, "Benchmark case studies in structural design optimization using the force method," *Int J Solids Struct,* vol. 42, p. 5848 – 5871, 2005.

[4] T. Nguyen-Thoi, T. Ho-Huu, T. Dang-Trung, T. Bui-Xuan and T. Lam-Phat, "Optimization analysis of stiffened composite plate by sequential quadratic programming," *J Sci Techno l,* vol. 51, p. 156 – 165, 2013.

[5] J. R. Koza, "Genetic programming: On the programming of computers by means of natural selection," *Cambridge:MA: MIT Press,* 1992.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of the IEEE Int. Conf. on Neural Networks*, 1995.

[7] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim,* vol. 11, no. 4, p. 341–359, 1997.

[8] B. Basturk and D. Karaboga, "An artificial bee colony (ABC) algorithm for numeric function optimization," in *Proceedings of the IEEE Swarm Intelligence Symposium*, 2006.

[9] R. V. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *Int. J. Ind. Eng. Comput.,* vol. 7, p. 19–34, 2016.

[10] Dinh-Cong, D.; Ho-Huu, V.; Vo-Duy, T.; Ngo-Thi Hong, Q.; Nguyen Thoi, T.;, "Efciency of Jaya algorithm for solving the opti-mization-based structural damage identifcation problem based on a hybrid

objective function," *Eng Optim,* vol. 50, no. 8, p. 1233–1251, 2017.

[11] Dinh-Cong, D.; Vo-Duy, T.; Ho-Huu, T.; Nguyen-Thoi, T.;, "Damage assessment in plate-like structures using a two-stage method based on modal strain energy change and Jaya algorithm," *Inverse Probl Sci Eng,* vol. 27, no. 2, p. 166–189, 2018.

[12] R. V. Rao and K. More, "Design optimization and analysis of selected thermal devices using self-adaptive Jaya algorithm," *Energy Convers Manag,* vol. 14, p. 24–35, 2017.

[13] R. V. Rao and K. More , "Optimal design and analysis of mechan-ical draft cooling tower using improved Jaya algorithm," *Int J Refrig,* vol. 82, p. 312–324, 2017.

[14] Rao , R. V.; Rai , DP.;, "Optimisation of welding processes using quasi-oppositional-based Jaya algorithm," *J Exp Theor Artif Intell ,* vol. 29, no. 5, p. 1099–1117, 2017.

[15] J. Reddy, Mechanics of laminated composite plates and shells: theory and analysis, CRC Press, 2003.

[16] Ghasemi, H.; Park, HS; Rabczuk , T;, "A multi-material level set-based topology optimization of fexoelectric composites," *Comput Methods Appl Mech Eng,* vol. 332, p. 47–62, 2017.

[17] Nguyen-Thoi, T; Ho-Huu, V; Dang-Trung, H; Bui-Xuan, T; Lam-Phat, T;, "Optimization analysis of stiffened composite plate by sequential quadratic programming," *Journal of Science and Technology,* vol. 51, no. 4B, pp. 156-165, 2013.

[18] Marin, L; Trias, D; Badallo, P; Rus, G; Mayugo, J. A.;, "Optimization of composite stiffened panels under mechanical and hygrothermal loads using neural networks and genetic algorithms," *Composite Structures,* vol. 94, pp. 3321-3326, 2012.

[19] B. G. Falzon and A. Faggiani, "The use of a genetic algorithm to improve the postbuckling strength of stiffened composite panels susceptible to secondary instabilities," *Composite Structures,* vol. 94, pp. 883-895, 2012.

[20] M. Kolli and K. Chandrashekhara , "Finite element analysis of stiffened laminated plates under transverse loading," *Compos Sci Technol,* vol. 56, p. 1355–61, 1996.

[21] Padhye , N; Bhardawaj , P; Deb , K;, "Improving differential evolution through a unified approach," *Glob Optim,* vol. 55, p. 771–99, 2013.

[22] Lam-Phat, T; Nguyen-Hoai, S; Ho-Huu, V; Nguyen-Thoi, T;, "Optimization of stiffened composite plate using adjusted different evolution algorithm," in *Proceeding of the international conference on computational methods*, Berkeley, CA, USA, 2016.

# Estimation of the welding current in Gas Tungsten Arc welded (GTAW) process with a specified width penetration

**\*Nguyen Quan[1], †Nguyen Hoai Son[2], Nguyen Nhut Phi Long[3], and Lam Phat Thuan[4]**

[1]Faculty of Engineering – Technology, Pham Van Dong University (PDU), Quang Ngai, Vietnam
[2, 4]Faculty of Civil Engineering, HCMC University of Technology and Education (HCMUTE), Viet Nam
[3]Faculty of Mechanical Engineering, HCMC University of Technology and Education (HCMUTE), Vietnam

\*Presenting author: nquan@pdu.edu.vn
†Corresponding author: sonnh@hcmute.edu.vn

**Abstract**

This paper presents the welding current estimation in the Gas Tungsten Arc welded (GTAW) process with a specified width penetration. The welding current determination of the GTAW process is performed by using the inverse method. In this work, the inverse method utilized is the Levenberg – Marquardt method. The advantage of the method is that the algorithm is simple, easy to apply, and has a high convergence rate. Two examples are considered to demonstrate the proposed method. The results show that the proposed method efficiently estimates the welding current for a specified width penetration in the GTAW process.

**Keywords:** Gas Tungsten Arc Welding (GTAW), Welding current, Levenberg – Marquardt method (LMM)

**Introduction**

Many studies demonstrated that the weld size, especially the weld depth, is one of the essential parameters of the weld, which significantly affects the quality of the weld. Karadeniz et al. [1] presented that the welding current greatly influences the weld size during the gas metal arc welding. Therefore, the determination of an suitable welding current to achieve the desired weld size in advance is essential for good quality welds, where the welding current is the input parameter, and the weld size is the output parameter, which results from the welding process. In the case of the problem when knowing the input parameters of the GTAW process: welding current, welding rod size, and welding speed, the determination of weld parameters are very easy through solving the direct problem by Computer-Aided Engineering (CAE) software, such as COMSOL, ABAQUS, ANSYS,... However, the estimation of the welding current value for the desired weld parameters (in this study, the width penetration of the weld) is a problem that needs to be proposed. This is a problem of the input parameter estimation to achieve the preset output parameter. Thus, the level of difficulty more increases which reflected in the problem of the welding current determination to achieve the desired width penetration in the GTAW weld process in this paper. So far, the solution of selecting/adjusting welding current almost through some trial and error method is common. The basic disadvantage of this solution is that it depends on the welder's experience and is difficult to achieve high accuracy. Proposing a method to determine the welding current for a preset weld depth is significant both scientifically and practically.

Therefore, this paper proposes a method based on inverse technical to determinate welding current with the desired width penetration in the GTAW process. The study focuses on building the steps involved in using inverse technical to solve the above problem. The article proposes using the Levenberg-Marquardt method (LMM) to determine the appropriate welding current to attain the specified width penetration of the weld. The Levenberg-Marquardt method (LMM) has simplicity, stability, and fast convergence speed, especially for solving high-order non-linear inverse problems. It has been successfully applied in several studies using the inverse technical for optimization problems [2] and some studies of other authors [3-5]. The implementation results will be presented in detail in the following sections.
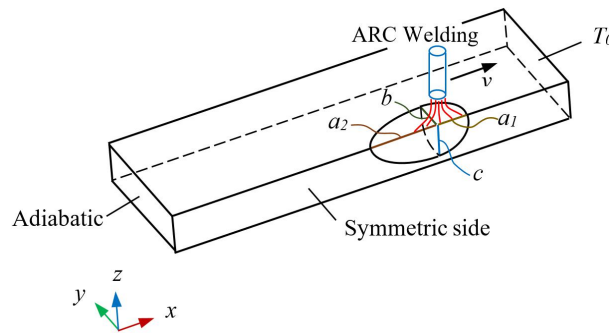
**Problem**



**Figure 1. Model of butt joint welding in the GTAW**

Consider the process of GTAW welding with butt type on the steel plates. Due to symmetry, only half of the models were examined (see Fig. 1). Based on the principle of energy conservation, the equation describing the temperature change when the welding rod moves with speed v in the x-direction is shown as follows [6]:

$$\frac{\partial}{\partial \xi}\left(k\frac{\partial T}{\partial \xi}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+\dot{q}_v = \rho C_p v \frac{\partial T}{\partial \xi} \tag{1}$$

with $\xi = x - vt$; and k, ρ, C is respectively the coefficient of thermal conductivity, the density, and the specific heat of the weld sample (note that these depend on the temperature)

The volumetric heat source $\dot{q}_v$ of the GTAW process defined as follows [7]:

$$\dot{q}_v(\xi, y, z) = \frac{6\sqrt{3}f_{1,2}\eta IU}{a_{1,2}bc\pi\sqrt{\pi}} \times \exp\left(-3\frac{\xi^2}{a_{1,2}}-3\frac{y^2}{b^2}-3\frac{z^2}{c^2}\right) \tag{2}$$

where $\eta$: the heat generation efficiency of electric ar; $I$: the welding amperage; $U$: the voltage; $f_1$ and $f_2$: the front and behind elliptical heat sources' weights ( $f_1 \approx 0.6$ and $f_2 \approx 0.4$ ); $a_1$, $a_2$, $b$, and $c$: parameters of the double ellipse.

The boundary conditions are also shown in Fig. 1. Due to symmetry, the longitudinal side of the weld line is established with thermal insulation boundary conditions. The back side is set to adiabatic boundary condition, and the front side is set to isothermal boundary condition. The remaining sides of the welded plate are cooled through natural radiation and heat exchange. The following equation expresses this cooling process:

$$-k\frac{\partial T}{\partial n} = h(T-T_0)+\sigma\varepsilon(T^4-T_0^4) \tag{3}$$

where *k* and *h:* the coefficients of heat transfer and heat exchange, $\sigma$: the Stefan – Boltzmann constant, $\varepsilon$: the emissivity coefficient

When the welding input parameters are predefined, the temperature distribution and the weld zone size (including the width penetration of the weld) can be determined from Eq. (1) above. In this paper, the finite element method is used to solve this equation [8].

## Method

### *The Levenberg – Marquardt method (LMM)*

This section will show how to determine the welding current to achieve the width penetration of the weld using the Levenberg – Marquardt method. First, the study investigates the formation of the weld's width penetration with different welding currents.
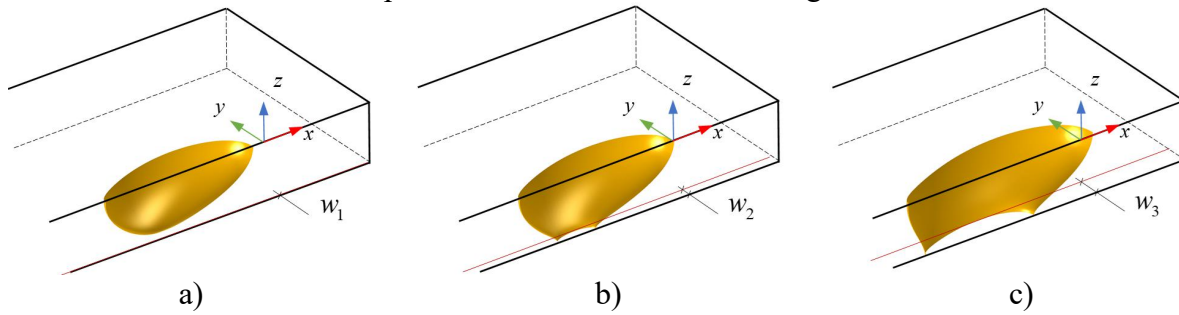


**Figure 2. The width penetration of the weld with different welding currents**

Fig. 2 shows the width penetration of the weld $w_1$, $w_2$ and $w_3$ with amperage $I_1$, $I_2$, and $I_3$, where $I_1 < I_2 < I_3$. In case 1, the current is too small to form the weld's width penetration, and the welding process has not fully penetration; case 2, the welding process has fully penetrated, and the weld's width penetration is achieved; while, in case 3, the weld's width penetration is too large because the current $I_3$ is the largest. In case 1, the weld quality will not be better than in cases 2 and 3 because the penetration has not fully. Moreover, compared to case 2, the width penetration of the weld in case 3 is too large, which will be bad for the welded structure and waste energy. Therefore, it is necessary to determine a reasonable amperage to achieve the appropriate width penetration of the weld. This helps to achieve the best weld quality as well as save on welding costs.

Since then, this study proposes to use LMM to implement the above problem. The proposed method is based on the inverse problem-solving principle. Accordingly, the LM method will find out a reasonable welding current by comparing the calculated temperature for any curent with the melting temperature at the location where we want to achieve the desired width penetration of the weld. Mathematically, setting the current for the desired weld's width penetration using the LM method is the minimization of the following quadratic function [9]:

$$S(\mathbf{x}) = \sum_{i}^{N}\left(\Phi_c^i - \Phi_m^i\right)^2 \tag{4}$$

where *N*: the number of nonlinear equations, **x**: the variable vector to find

To minimize Eq. (4), the values of the variables in the variable vector **x** must satisfy the following condition:

$$\sum_{i=1}^{N}\frac{\partial\Phi}{\partial\mathbf{x}}(\Phi_c^i - \Phi_m^i) = 0 \tag{5}$$

Eq. (5) is a system of nonlinear equations. The LM method is used to solve this equation. In the LM method, a "damping" factor $\mu$ controls the convergence of the nonlinear problem. The advantage of the LM method is that it is both simple and can solve problems with high nonlinearity. The values of the variables in vector $\mathbf{x}$ solved by the LM method are described in the following loop:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{D}^T\mathbf{D} + \mu\mathbf{\Omega})^{-1}\mathbf{D}^T(\mathbf{\Phi}_c - \mathbf{\Phi}_m) \tag{6}$$

where $\mathbf{D} = \dfrac{\partial\mathbf{\Phi}}{\partial\mathbf{x}}$ : sensitivity matrix, $\mathbf{\Omega}$ : diagonal matrix with diagonal values taken from the matrix $\mathbf{D}$, $k$: the loop index.
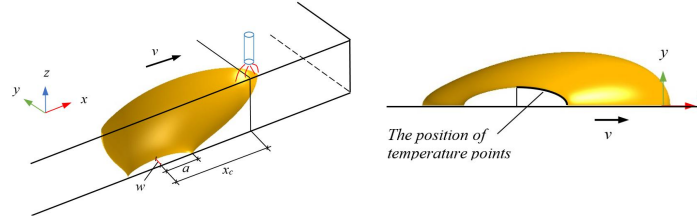


**Figure 3. Parameters $w$, $a$, $x_c$ and a quarter of the ellipse used to extract the temperature**

From Eq. (6), it can be inferred that the solution can be obtained when the values $\mathbf{\Phi}_c$ and $\mathbf{\Phi}_m$ are known. However, this determination of welding amperage is based on the inverse method. To ensure stability in solving this problem, the number of equations must be greater than the number of variables. In this study, the number of temperature extraction points to compare with the melting temperature was taken per quarter of the elliptic profile of the weld forming process (some studies have shown that, for the welding process with the tip welding moves at a fixed speed, the shape of the weld at the bottom surface is elliptical [10, 11]). Furthermore, to increase the stability and convergence speed in finding the welding amperage, two auxiliary variables are included in the LM algorithm: the length variable $a$ and the ellipse center position $x_c$, as shown in Fig. 3.

*The stopping criteria for the LM method*

The LM method is used to determine the amperage and two auxiliary variables in the vector $\mathbf{x}$. The solving process is done by loop solving in Eq. (6) to determine the variable $\mathbf{x}_k$ value. This loop will stop when one of the stopping criteria is satisfied. Two stopping criteria proposed by Frank and Wolfe [12] are used in this LM method as follows:

$$\|\mathbf{x}_k - \mathbf{x}_{k-1}\| / \|\mathbf{x}_k\| \leq \delta \tag{7}$$

$$\|S(\mathbf{x}_k) - S(\mathbf{x}_{k-1})\| / \|S(\mathbf{x}_k)\| \leq \varepsilon \tag{8}$$

with

$$\|S(\mathbf{x}_k)\| = \sum_{i=1}^{N}\left[\Phi_c^i - \Phi_m^i\right]^2 \tag{9}$$

where $\delta$ and $\varepsilon$: the convergence tolerances are small positive value.
Or the width penetration of the weld at the $k_{th}$ loop ($w_k$) satisfying the following condition:

$$\|\mathbf{w}_k - \mathbf{w}_0\| \leq \delta_0 \tag{10}$$

where $w_0$: the desired width penetration

*Algorithm*

The algorithm of the LM method is described as follows:

Given the stopping criteria value ($\delta$, $\varepsilon$, $\delta_0$) the coefficient $\mu$ and the initial amperage value $I_0$, solve Eq. (1) to define the initial variable $\mathbf{x}_0 = \{I_0, a_0, x_{c,0}\}$. The value of variable $x_k$ at the $k_{th}$ loop is determined as follows:

    Step 1: Solve Eq. (1) and calculate the temperature vector $\Phi_c^i$

    Step 2: Put the value $\Phi_c^i$ in Eq. (4).

    Step 3: Calculate $\mathbf{x}_k$ from Eq. (6) and calculate the value $S(\mathbf{x}_k)$ from Eq. (4).

    Step 4: If $S(\mathbf{x}_k) \geq S(\mathbf{x}_{k-1})$, increase the damping value $\mu = 10\mu$, return to Step 3. Otherwise, accept the value of $\mathbf{x}_k$ and decrease the damping value $\mu = 0.1\mu$.

    Step 5: Stop the loop if the condition stops in Eq. (7) – (10) satisfied. Otherwise, increase $k$ = $k$+1 and go back to Step 2.

## Results and Discussion

In the present study, the Gas Tungsten Arc Welding (GTAW) process is considered. The diameter of the welding rod is $d$ = 1.6 mm with the heat generation efficiency of the electric arc $\eta$ = 0.6 and the DC voltage is $U$ = 20 V (these parameters refer to the studies [13] , 14]) welding speed is $v$ = 15 mm/s. The welding process is simulated on a sheet of size 70 × 30 × 0.8 mm. The material of the welded plate is commercial stainless steel AISI304, which has temperature-dependent thermal properties obtained from the study of Sabarikanth [15].
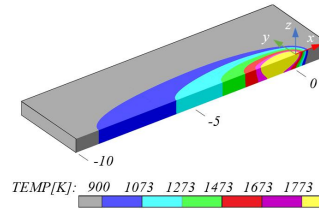


**Figure 4. The heat distribution in the case $U$ = 20 V, $I$ = 32.5 A, $\eta$ = 0.6 and $v$ = 15 mm/s**

When the welding configuration is known in advance (that is, the input parameters including amperage $I$ (A), voltage $U$ (V), arc heat generation efficiency hàn and welding speed $v$ (mm/s) are known in advance), heat distribution, weld size and the width penetration of the weld can determine the solution of the forward problem. Fig. 4 shows the temperature distribution and weld size in the case of $U$ = 20 V, $I$ = 32.5 A, $\eta$ = 0.6 and $v$ = 15 mm/s. In this case, the width penetration of the weld is $w$ = 0.57 mm.

In the case where the weld's width penetration is known, we need to find the welding amperage to achieve the width penetration of the weld. At this point, the LM method is used to solve this problem. Two cases of finding a reasonable welding amperage with two different width penetration of the weld will be performed to demonstrate the proposed method.

*Case 1*

In case 1, the desired weld depth width is $w_0$ = 0.1 (mm). As in the LM algorithm section, the initial value for the welding amperage and the stopping values. Here, choose $I_0$ = 32.5 A, and the stopping parameters are 10e-4. The results performed by the LM method are shown in Table 1 and Table 2

**Table 1. The function value of *S(x)* and the stopping criterion in case 1**

| Loop | $S(x)$ | $\dfrac{\|S(\mathbf{x}_k)-S(\mathbf{x}_{k-1})\|}{\|S(\mathbf{x}_k)\|}$ | $\dfrac{\|\mathbf{x}_k-\mathbf{x}_{k-1}\|}{\|\mathbf{x}_k\|}$ | $\dfrac{\|\mathbf{w}_k-\mathbf{w}_0\|}{\|\mathbf{w}_0\|}$ |
|------|--------|------|------|------|
| 1 | 2.07E+06 | 4.26E-01 | 3.85E-06 | 4.74E+00 |
| 2 | 2.95E+06 | 4.23E-01 | 6.00E-02 | 4.00E+00 |
| 3 | 1.71E+06 | 9.77E-01 | 1.64E-01 | 1.36E+00 |
| 4 | 3.88E+04 | 9.73E-01 | 4.21E-02 | 3.66E-01 |
| 5 | 1.05E+03 | 9.99E-01 | 1.13E-02 | 9.50E-03 |
| 6 | 8.07E-01 | 9.93E-01 | 2.64E-04 | 2.30E-03 |
| 7 | 5.56E-03 | 1.67E+00 | 7.79E-05 | 2.00E-03 |

**Table 2. Value the welding amperage *I*, the power *P*, and the width penetration of the weld *w* after loops of LM algorithm in case 1**

| Loop | $I$ (A) | $P = I.U$ (W) | $w$ (mm) |
|------|---------|---------------|----------|
| 1 | 32.500 | 650.000 | 5.749E-01 |
| 2 | 32.500 | 650.003 | 5.002E-01 |
| 3 | 30.550 | 611.007 | 2.364E-01 |
| 4 | 25.550 | 510.995 | 1.366E-01 |
| 5 | 24.475 | 489.493 | 1.134E-01 |
| 6 | 24.210 | 484.205 | 1.052E-01 |
| 7 | 24.205 | 484.101 | 1.002E-01 |

The results in Table 2 show that the welding current result is $I_7 = 24.205$ A, corresponding to the power $P_7 = 484.101$ W, after 7 loops. The value of the weld's width penetration after 7 loops is respectively $w_7 = 0.1002$ mm. The relative error between the desired width penetration of the weld value and the estimated width penetration of the weld by the LM method is 0.002 (0.2%), as shown in Table 1. This error is very small. This proves that the LM method is very effective in determining the welding amperage to achieve the desired width penetration of the weld.

*Case 2*

In case 2, the desired weld depth width is $w_0 = 7.500$E-2 mm. In this case, the initial amperage value is $I_0 = 25$ A. Welding input values are the same as in case 1. The results in case 2 are similar to those in case 1 in Tables 3 and 4.

**Table 3. The function value of S(x) and the stopping criterion in case 2**

| Loop | $S(x)$ | $\dfrac{\|S(\mathbf{x}_k)-S(\mathbf{x}_{k-1})\|}{\|S(\mathbf{x}_k)\|}$ | $\dfrac{\|\mathbf{x}_k-\mathbf{x}_{k-1}\|}{\|\mathbf{x}_k\|}$ | $\dfrac{\|\mathbf{w}_k-\mathbf{w}_0\|}{\|\mathbf{w}_0\|}$ |
|------|--------|------|------|------|
| 1 | 1.83E+04 | 9.20E-01 | 2.59E-02 | 3.82E-01 |
| 2 | 1.46E+03 | 9.94E-01 | 1.14E-02 | 4.46E-02 |
| 3 | 8.08E+00 | 9.79E-01 | 7.30E-04 | 1.45E-02 |
| 4 | 1.73E-01 | 6.63E-02 | 2.21E-04 | 1.14E-02 |
| 5 | 1.61E-01 | 5.69E-01 | 1.67E-04 | 6.53E-03 |
| 6 | 2.96E-05 | 9.84E-01 | 1.83E-05 | 8.15E-04 |

**Table 4. Value the welding amperage *I*, the power *P*, and the width penetration of the weld *w* after loops of LM algorithm in case 2**

| Loop | $I$ (A) | $P = I.U$ (W) | $w$ (mm) |
|------|---------|---------------|----------|
| 1 | 25.000 | 500.000 | 1.215E-01 |
| 2 | 24.353 | 487.059 | 7.851E-02 |
| 3 | 24.075 | 481.507 | 7.392E-02 |
| 4 | 24.058 | 481.155 | 7.587E-02 |
| 5 | 24.059 | 481.172 | 7.549E-02 |
| 6 | 23.962 | 479.242 | 7.494E-02 |

From Table 4, to achieve the width penetration of the weld $w_0$ = 7.5E-2 mm, the loop process in the LM method has found the welding amperage of $I_6$ = 23.96 A, corresponding to the power is $P_6$ = 479.242 W, after 6 loops with the initial amperage value set to $I_0$ = 25 A. The value of the width penetration of the weld corresponding to the current $I_6$ = 23.96 A is $w_6$ = 7.494E-02 mm. The relative error between the desired weld depth width value and the estimated weld depth width value by the LM method is 8.153E-04 (0.08%), as shown in Table 3. This error is also very small, like case 1.

Besides, in both cases, the number of iterations to achieve the result is very small (namely, case 1 is 7 loops and case 2 is 6 loops). This means that the LM method gives a very high convergence rate.

From the above two cases, it can be seen that the LM method is very effective in determining the welding current to achieve a given weld depth width. In both cases, the number of iterations required to obtain the result is very little, and the relative error between the estimated weld depth width by the LM method and the desired weld depth width value is very small. This shows that the LM method is very effective in solving this problem.

## Conclusions

This paper has performed the estimating of welding current to attain the desired width penetration in the Gas Tungsten Arc Welding (GTAW) by using the Levenberg – Marquardt method (LMM). To do this, the paper describes the computational model in the GTAW process, the method of solving the welding problem, the study of the Levenberg – Marquardt method, and the building of relationships, constraints, conditions, stopping criteria, and algorithms to solve the proposed problem.

Two cases of estimating the welding amperage with the desired weld depth width of GTAW welding were performed. The results show that the welding current is achieved after only 7 loops (case 1) and 6 loops (case 2) in the LM algorithm. The relative error between the desired width penetration of the weld value and the estimated width penetration of the weld by the LM method is very small (0.2% for case 1 and 0.08% for case 2). It shows that the LM method is very effective in finding the welding current to achieve the desired width penetration in GTAW. This setting helps to choose reasonable welding parameters to achieve quality welding and save energy in the welding process.

## References

[1] E. Karadeniz, U. Ozsarac, C. Yildiz, The effect of process parameters on penetration in gas metal arc welding processes, *Materials & Design*, **28**(2) (2007) 649-656.

[2] Q. Nguyen, S. Nguyen-Hoai, T. Chuong-Thiet, T. Lam-Phat, Optimization of the Longitudinal Cooling Fin by Levenberg–Marquardt Method, in: *International Conference on Advances in Computational Mechanics*, Springer, 2017, pp. 217-227.

[3] S. Rouquette, J. Guo, P. Le Masson, Estimation of the parameters of a Gaussian heat source by the Levenberg–Marquardt method: Application to the electron beam welding, *International Journal of Thermal Sciences*, **46**(2) (2007) 128-138.

[4] J. Fan, J. Zeng, A Levenberg–Marquardt algorithm with correction for singular system of nonlinear equations, *Applied Mathematics and Computation*, **219**(17) (2013) 9438-9446.

[5] A. Kleefeld, M. Reißel, The Levenberg–Marquardt method applied to a parameter estimation problem arising from electrical resistivity tomography, *Applied Mathematics and Computation*, **217**(9) (2011) 4490-4501.

[6] T.L. Bergman, F.P. Incropera, D.P. DeWitt, A.S. Lavine, *Fundamentals of heat and mass transfer*, John Wiley & Sons, 2011.

[7] D. Gery, H. Long, P. Maropoulos, Effects of welding speed, energy input and heat source distribution on temperature variations in butt joint welding, *Journal of Materials Processing Technology*, **167**(2–3) (2005) 393-401.

[8] BMR Frewin, D.A. Scott, *Finite Element Model of Pulsed Laser Welding*, Welding Research Supplement, 78 (1999) 15s-22s.

[9] M.N. Ozisik, *Inverse heat transfer: Fundamentals and applications*, (2000)

[10] D. Rosenthal, Mathematical theory of heat distribution during welding and cutting, *Welding journal*, **20**(5) (1941) 220s-234s.

[11] D. Rosenthal, *The theory of moving sources of heat and its application to metal treatments*, in, *ASME*, 1946.

[12] M. Frank, P. Wolfe, An algorithm for quadratic programming, *Naval research logistics quarterly*, **3**(1-2) (1956) 95-110.

[13] S. Bag, A. De, Error analysis of forward and reverse heat conduction and convection calculations considering uncertainties in welding, *Sci Technol Weld Joi*, **14**(7) (2009) 662-668.

[14] C.V. Gonçalves, L.O. Vilarinho, A. Scotti, G. Guimarães, Estimation of heat source and thermal efficiency in GTAW process by using inverse techniques, *Journal of Materials Processing Technology*, **172**(1) (2006) 42-51.

[15] R. Sabarikanth, K. Sankaranarayanasamy, N. Siva Shanmugam, G. Buvanashekaran, A study of laser welding modes with varying beam energy levels, Proceedings of the Institution of Mechanical Engineers, *Part C: Journal of Mechanical Engineering Science*, **223**(5) (2009) 1141-1156.

# An improvement of linear polyhedral finite element method using consecutive interpolation scheme in transient analysis

**\*Hau Nguyen-Ngoc[1,2], Hung Nguyen-Xuan[3], Magd Abdel-Wahab[1]**

[1] Department of Electrical energy, metals, mechanical constructions and systems, Faculty of Engineering and Architecture, Ghent University, Belgium.
[2] University of Science and Technology, the University of Danang, Vietnam.
[3] CIRTech Institute, Ho Chi Minh City University of Technology (HUTECH), Vietnam.

\* Corresponding author: Hau.NguyenNgoc@ugent.be; NNHau@dut.udn.vn.

**Abstract**

In finite element analysis, piecewise linear shape functions are often powerful and quite simple to construct and implement, which are very popular in the finite method using conventionally tetrahedral and hexahedral elements. Recently, the piecewise linear shape functions have been presented for arbitrary polyhedral elements (PFEM). The new interpolation scheme is relied on the well-known hierarchical shape functions of tetrahedral element. The proposed method shows its flexibility, simplicity and accuracy in the solid mechanics problems. However, due to the nature of linear interpolants, the obtained strains are constant within sub-domains of the polyhedral element and across the elements. Therefore, the consecutive interpolation (CI) scheme has been recently applied to further improve the performance and accuracy of PFEM. In this study, the CI scheme is implemented for 3D polyhedral elements with arbitrary number of vertices, edges and the facets can be non-planar. The scheme is constructed based on the fact of the continuity of strain across the discretized nodes of the problem domains. The consecutive shape functions are formulated by using average derivatives of the nodal displacements and the piecewise linear shape functions of PFEM. As a result, the newly formulated shape functions are high-order polynomials over linear polyhedral elements. The application of this method in the analysis of transient responses of 3D solid structure is carried out to prove the rationality, feasibility, accuracy and performance of the method.

**Keywords:** PFEM, 3D polyhedral elements, consecutive interpolation scheme, transient analysis.

## Introduction

In three-dimensional solid mechanics, the structures with complex geometry can be meshed effectively with 4-node tetrahedral elements, while the hexahedral elements are sensitive to the shape of geometry. However, the accuracy of analysis using tetrahedral elements is usually moderate, element distortion sensitivity and so on. On the other hand, arbitrary polyhedral elements not only can be implemented to complex geometry structures but also usually produce better engineering solutions compared to conventionally tetrahedral interpolation. Therefore, the polyhedral element method (pFEM) has been developed significantly recent years.

Within the framework of pFEM, there are various techniques that have been proposed to formulate and/or solve the engineering problems [1]. For instance, Voronoi cell finite element method [2] and Hybrid polygonal element method [3] were developed to analyze microstructures. Various shape functions were proposed based on conforming polygonal finite element method, which was effective to interpolate over arbitrary polygonal/polyhedral [4]–[6]. The polygonal smoothed finite element method can increase the solutions accuracy and smooth the results within the analyzed domain [7]. The computational time can be reduced significantly compared to conventional pFEM by coupling virtual element method and polygonal/polyhedral smoothed finite element method [8]. Whereas, the polyhedral finite element method using piecewise linear shape functions (PFEM) shows its flexibility, simplicity and accuracy in the analysis of solid mechanics problems [6][9]. Nevertheless, the approximated strains are constant across the elements. Therefore, the PFEM was improved by using the consecutive interpolation scheme (CIPFEM) [10], which shows great advantages compared to the conventional linear FEM such as higher accuracy, continuous stress and strain fields without any requirement of smoothing operators, higher convergence rate with the same mesh size, high-order polynomial shape functions processing Kronecker-delta property and mesh distortion insensibility.

In this article, the CIPFEM will be implemented in the transient analysis of 3D solid structure so that the feasibility of the method will be further studied.

**Formulations of CIPFEM for transient analysis**

Let us consider a problem domain $\Omega$ with boundary $\Gamma$ in the 3D space $\mathbb{R}^3$, which is discretized into $N_e$ finitely polyhedral elements $\Omega_e$. The procedure of solving Galerkin weak form using CIPFEM is similar to conventional FEM. Whereas, the nodal unknowns of solid structures are approximated using the discretized equivalent equation which can be written for CIPFEM as follows [11]

$$\widetilde{\mathbf{M}}\ddot{\mathbf{d}}(t) + \widetilde{\mathbf{C}}\dot{\mathbf{d}}(t) + \widetilde{\mathbf{K}}\mathbf{d}(t) = \mathbf{F}(t) \tag{1}$$

where $\widetilde{\mathbf{M}}, \widetilde{\mathbf{C}}$ and $\widetilde{\mathbf{K}}$ are respectively the mass, damping and stiffness matrices using CI shape functions, and $\mathbf{F}$ is the vector of time-dependent external forces.

The stiffness matrix $\widetilde{\mathbf{K}}$ and mass matrix $\widetilde{\mathbf{M}}$ are formulated as follows:

$$\widetilde{\mathbf{K}}_e = \int_{\Omega_e} \mathbf{B}_e^{\mathrm{T}}(\mathbf{x}_j)\mathbf{D}\mathbf{B}_e(\mathbf{x}_j)\mathrm{d}\Omega = \sum_{i=1}^{n_\tau}\sum_{j=1}^{n_{GP}} \mathbf{B}_s^{\mathrm{T}}(\mathbf{x}_j)\mathbf{D}\mathbf{B}_s(\mathbf{x}_j)|\mathbf{J}_j^i|W_j \tag{2}$$

$$\widetilde{\mathbf{M}}_e = \int_{\Omega_e} \mathbf{N}_e^{\mathrm{T}}(\mathbf{x}_j)\rho\mathbf{N}_e(\mathbf{x}_j)\mathrm{d}\Omega = \sum_{i=1}^{n_\tau}\sum_{j=1}^{n_{GP}} \mathbf{N}_s^{\mathrm{T}}(\mathbf{x}_j)\rho\mathbf{N}_s(\mathbf{x}_j)|\mathbf{J}_j^i|W_j \tag{3}$$

where $n_\tau$ is the number of sub-tetrahedra in the polyhedral element $e$; $n_{GP}$ is the number of Gauss points in each sub-domain; $\Omega_e$ is the volume of element $e$; $W_j$ is the weight of the $j^{\text{th}}$ Gauss point $\mathbf{x}_{GPj}$ in sub-domain $i^{\text{th}}$; $\mathbf{D}$ is the matrix of material and $|\mathbf{J}_j^i|$ is the determinant of Jacobian matrix of 4-node tetrahedron formulated. $\mathbf{B}_s$ is the displacement transformation matrix at $\mathbf{x}_j$ in sub-

domain $\Omega_s$. Unlike the matrix of PFEM that includes $n_e$ nodes of the element $e$, the CIPFEM's matrix contains all the nodes ($n_s$) in the support domain $\Omega_{sup}$ as follows [10]:

$$[\mathbf{B}_s]^\mathrm{T} = \begin{bmatrix} \frac{\partial \tilde{N}_l}{\partial x} & 0 & 0 & \frac{\partial \tilde{N}_l}{\partial y} & 0 & \frac{\partial \tilde{N}_l}{\partial z} \\ 0 & \frac{\partial \tilde{N}_l}{\partial y} & 0 & \frac{\partial \tilde{N}_l}{\partial x} & \frac{\partial \tilde{N}_l}{\partial z} & 0 \\ 0 & 0 & \frac{\partial \tilde{N}_l}{\partial z} & 0 & \frac{\partial \tilde{N}_l}{\partial y} & \frac{\partial \tilde{N}_l}{\partial x} \end{bmatrix}, \quad l = 1 \div n_s \tag{4}$$

Where $\tilde{N}_l$ is the consecutive shape functions over arbitrary polyhedral element given as

$$\tilde{N}_l(\mathbf{x}) = \sum_{i=1}^{n_e}\left(\phi_i(\mathbf{x})N_l^i + \phi_{ix}(\mathbf{x})\bar{\bar{N}}_{l,x}^i + \phi_{iy}(\mathbf{x})\bar{\bar{N}}_{l,y}^i + \phi_{iz}(\mathbf{x})\bar{\bar{N}}_{l,z}^i\right) \tag{5}$$

with $\phi_i, \phi_{ix}, \phi_{iy}, \phi_{iz}$ are the auxiliary functions as follows:

$$\phi_i = L_i + L_i^2(\Sigma_1 - L_i) - L_i(\Sigma_2 - L_i^2); \tag{6}$$

$$L_i = N_i^{PFEM}; \quad \Sigma_1 = \sum_{j=1}^{n_e} L_j; \quad \Sigma_2 = \sum_{j=1}^{n_e} L_j^2; \quad \Sigma_3 = \sum_{j=1}^{n_e} L_j(\nabla L_j) \tag{7}$$

where $N_i^{PFEM}$ is the piecewise shape functions of PFEM.

and

$$\phi_{ix} = \sum_{j=1, j \neq i}^{n_e}(x_j - x_i)\left[L_j L_i^2 + 0.5 L_i L_j(\Sigma_1 - L_i - L_j)\right] \tag{8}$$

The formulations of $\phi_{iy}, \phi_{iz}$ are similar to that of $\phi_{ix}$, in which the $x$-coordinates are replaced respectively by $y$- and $z$-coordinates.

Finally, the damping matrix $\mathbf{C}$ is given as:

$$\tilde{\mathbf{C}} = c_M \tilde{\mathbf{M}} + c_K \tilde{\mathbf{K}} \tag{9}$$

**Numerical example**

The implementation of CIPFEM in the forced vibration analysis of solid structure is carried out for the concrete corbel shown in Fig. 1a. The sine-type impulsive load is implemented, $\{f^s(t) = f_0 \sin(\pi t/t_1) [H(t) - H(t - t_1)]\}$. Whereas, $f_0$ is the magnitude of the load, $t_1$ is the pulse's duration and H($t$) is the Heaviside function given as: H($t$) = 1 with $t \geq 0$ and H($t$) = 0 with $t < 0$. The load duration is $t_1 = 0.01$ s, and the time step, $\Delta t = 5 \times 10^{-5}$ s. The transient vibration responses of the structure, which are numerically approximated by CIPFEM and CT4, are investigated. The very coarse meshes of unstructured polyhedral elements ($h = 97.0$ mm) and tetrahedral elements ($h = 99.7$ mm) are used for CIPFEM and CT4, respectively (Fig. 1b, c).

The intensive of vibration force is assigned the value of $f_0 = 16$ kN. The structure is fixed at one end and subjected to transient load at the other end, while the other faces are traction free. The concrete material has Young's modulus, $E = 30 \times 10^3$ MPa, weight density, $\rho = 2300$ kg/m$^3$ and Poisson's ratio, $v = 0.25$. The analysis formulations using Newmark integration method are shown in Reference [12]. The initial conditions and the damping coefficients of all numerical cases are assumed to be zero. In this study, the solutions of ANSYS are also used as reference to evaluate the CIPFEM.

Fig. 2 to Fig. 4 show the comparisons of transient responses obtained from CIPFEM and CT4. In detail, Fig. 2 illustrates the time history of vertical displacement along *y* axis. Fig. 3 illustrates the time history of longitudinal displacement along *x*-direction, while Fig. 4 illustrates the time history of maximum von-Mises tress of the edge CD. All three figures show that the transient solutions predicted by CIPFEM agree very well with reference results compared to CT4, especially, for the cases of maximum von-Mises stress.
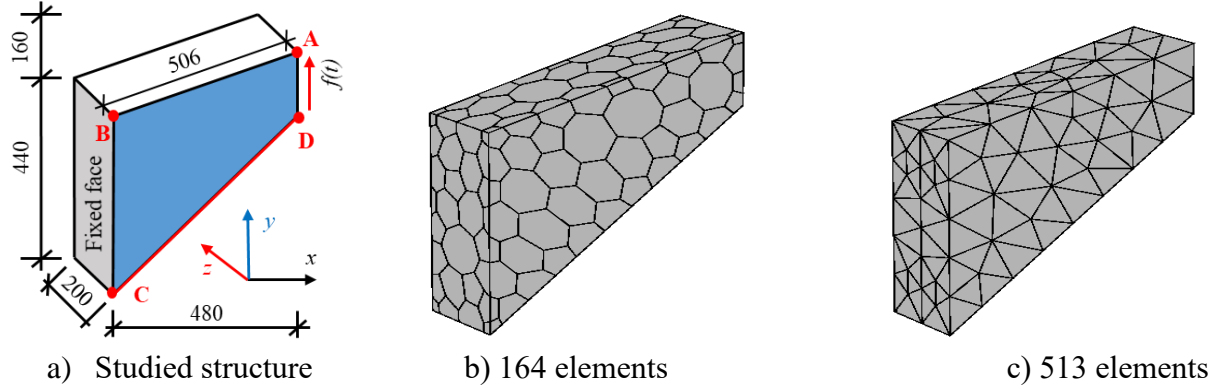


a) Studied structure      b) 164 elements      c) 513 elements

**Figure 1: Coarse meshes are used for analyses using consecutive interpolation scheme.**
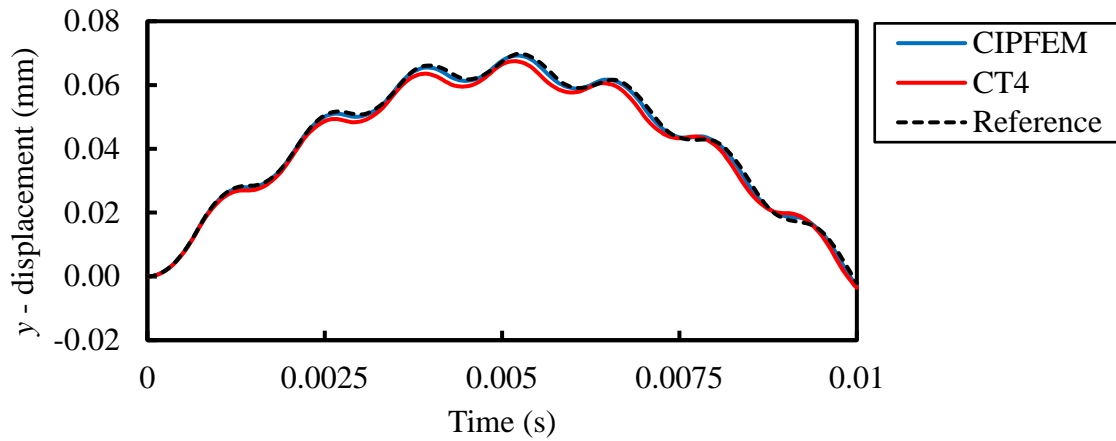


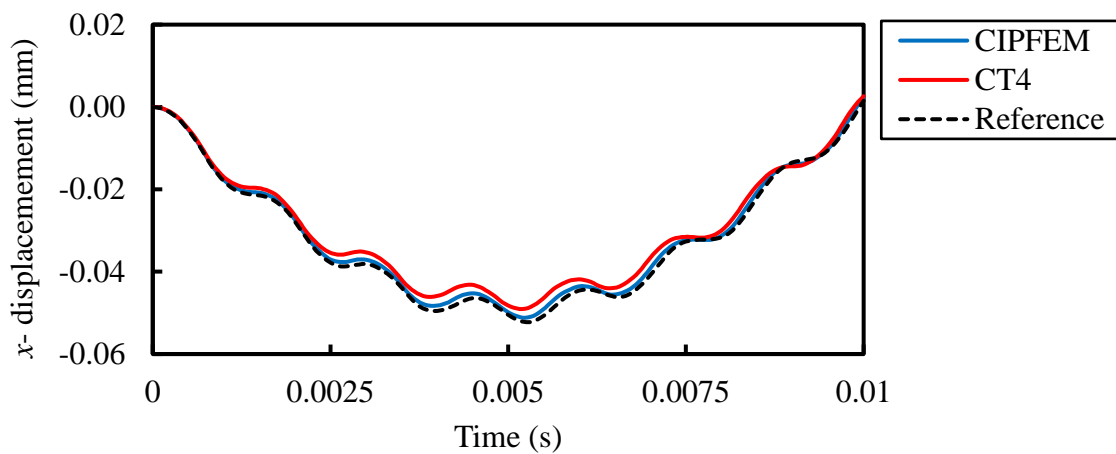**Figure 2: Time history of vertical (*y*-) displacement at node A.**



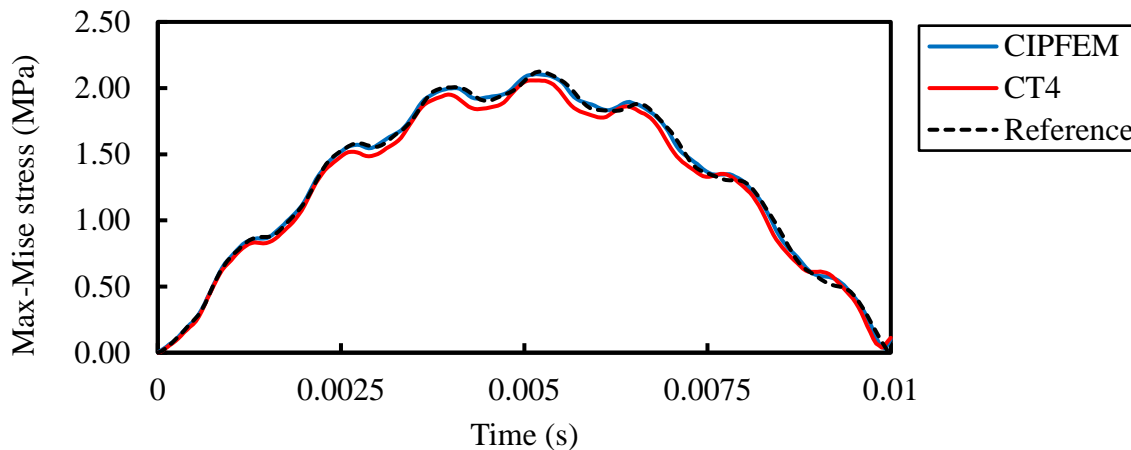**Figure 3: Time history of longitudinal (*x*-) displacement at node A.**

**Figure 4: Time history of maximum von-Mises tress along edge CD.**

## Conclusion

This study presents the implementation of CIPFEM in the analysis of three-dimensional concrete corbel. The method is formulated based on the consecutive interpolation scheme over arbitrary star-convex polyhedral elements. The PFEM shape functions are used to construct the high-order shape functions of CIPFEM because of its flexibility in interpolation over various type of elements. The performance of the method in the numerical analyses of dynamic problem is considered. All the numerical cases are analyzed on the MATLAB program. Some conclusions are derived as following:

- With the same mesh size, the CIPFEM gives better results than CT4.
- The CIPFEM presents very high accurate solutions of transient problems, even though the structure is mesh into very coarse elements.
- The CIPFEM can inherit the advantages of PFEM in the analysis of arbitray star-convex polyhedral elements.

### References

[1] Perumal, L. (2018) A Brief Review on Polygonal/Polyhedral Finite Element Methods, *Mathematical Problems in Engineering* **2018**, 1–22.

[2] Eder, P.M., Giuliani, J.E. and Ghosh, S. (2005) Multilevel Parallel Programming for Three-Dimensional Voronoi Cell Finite Element Modeling of Heterogeneous Materials, *The International Journal of High Performance Computing Applications* **19**, 29–45.

[3] Zhang, J. and Dong, P. (1998) A hybrid polygonal element method for fracture mechanics analysis of resistance spot welds containing porosity, *Engineering Fracture Mechanics* **59**, 815–825.

[4] Floater, M.S., Hormann, K. and Kós, G. (2006) A general construction of barycentric coordinates over convex polygons, *Advances in Computational Mathematics* **24**, 311–331.

[5] Bishop, J.E. (2014) A displacement-based finite element formulation for general polyhedra using harmonic shape functions, *International Journal for Numerical Methods in Engineering* **97**, 1–31.

[6]    Nguyen-Ngoc, H., Phung-Van, P., Dang, B.-L., Nguyen-Xuan, H. and Abdel Wahab, M. (2019) Static and dynamic analyses of three-dimensional hollow concrete block revetments using polyhedral finite element method, *Applied Ocean Research* **88**, 15–28.

[7]    Dai, K.Y., Liu, G.R. and Nguyen, T.T. (2007) An n-sided polygonal smoothed finite element method (nSFEM) for solid mechanics, *Finite Elements in Analysis and Design* 43, 847–860.

[8]    Natarajan, S., Bordas, S.P. and Ooi, E.T. (2015) Virtual and smoothed finite elements: A connection and its application to polygonal/polyhedral finite element methods, *International Journal for Numerical Methods in Engineering* **104**, 1173–1199.

[9]    Nguyen-Ngoc, H, Dang, B.L., Nguyen-Xuan, H., Thao, H.D. and Abdel Wahab, M., Three-dimensional analysis of an innovative hollow concrete block of interlocking revetment, *Lecture Notes in Civil Engineering*, Springer, 2018.

[10]   Nguyen-Ngoc, H., Chau, K.N., Duy, K.N., Nguyen-Xuan, H. and Abdel-Wahab, M. (2021) A Consecutive-interpolation polyhedral finite element method for solid structures, *International Journal for Numerical Methods in Engineering* (Accepted, unedited).

[11]   Öchsner, A., Computational Statics and Dynamics, Springer Singapore, Singapore, 2020.

[12]   Nguyen-Ngoc, H., Nguyen-Xuan, H. and Abdel-Wahab, M. (2020) A numerical investigation on the use of pervious concrete for seawall structures, Ocean Engineering **198**, 106954.

# Combined Gaussian process regression model and comprehensive learning particle swarm optimizer in reliability-based structural optimization

**\*Thu Huynh Van[1], Bach Do[2], Suchart Limkatanyu[3] and †Sawekchai Tangaramvong[1]**

[1] Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Chulalongkorn University, 10330 Bangkok, Thailand

[2] Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto-Daigaku Katsura, Nishikyo, Kyoto 615-8540, Japan

[3] Department of Civil Engineering, Faculty of Engineering, Prince of Songkla University, Songkla 90110, Thailand

*Presenting author: thuxd11@gmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

Reliability-based design optimization (RBDO) addresses the cost-effective integrity design of structures in the presence of inherent uncertain parameters. Processing this class of problem is challenging from the computational burden to determine the failure probability of structures violating the limit-state function. This paper proposes an efficient decoupling RBDO method that advantageously couples a comprehensive learning particle swarm optimization (CLPSO) algorithm with a Gaussian process regression (GPR) model, termed as GPR-CLPSO approach. In essence, the proposed method iteratively performs the CLPSO with deterministic parameters based on the most probable point underpinning the limit-state function iterative updated by the reliability evaluation process. The GPR model approximates from the design data given by CLPSO the spectrum of the limit-state function under uncertain parameters, and hence enables a significant reduction of Monte-Carlo simulations for estimation of failure probability. What is more, is that a so-called expected feasibility function is maximized to systematically refine the GPR model by locating new sampling points in the region with high-reliability sensitivity leading to the more accurate prediction of failure probability. The RBDO terminates when the resulting failure probability reaches the allowable threshold. The CLPSO is primarily adopted in the optimization process for the GPR hyperparameters and the expected feasibility function. A numerical example is provided to illustrate the applications and robustness of the proposed schemes in solving the RBDO problems.

**Keywords:** Gaussian process regression; Comprehensive learning; Particle swarm optimization; Expected feasibility function; Reliability-based design optimization.

## Introduction

Deterministic optimization has been extensively applied in engineering structures to improve the design performance with minimum resources. The design solution computed by the deterministic optimization becomes unreliable in some cases, especially when the influences of uncertainties inheriting structural dimensions, material properties, loading and operating conditions are significant and cannot be eliminated. By addressing the performance and reliability of the structure together, the structural reliability-based design optimization (RBDO) has been considered as the alternative approach in recent years. More explicitly, the RBDO problem minimizes the cost function, denoted as $C$, and satisfies the certain deterministic and probabilistic constraints, as state by the following generic mathematical formulations [1]:

$$\min C(\mathbf{s})$$
$$\text{s.t.} \quad \mathsf{P}\big[G(\mathbf{s},\mathbf{x}) - \overline{z} \leq 0\big] - P_a \leq 0, \tag{1}$$
$$\mathbf{s} \in \big[\mathbf{s}_L, \mathbf{s}_U\big]$$

where $\mathbf{s}$ and $\mathbf{x}$ are the vectors of deterministic design variables and random parameters, respectively. The vector $\mathbf{x}$ is characterized by the joint probability density function (PDF) $f(\mathbf{x})$ in the space $\Omega$. The two $\mathbf{s}_L$ and $\mathbf{s}_U$ denote the lower and upper bound on the variables $\mathbf{s}$, respectively. The functions $G(\mathbf{s}, \mathbf{x})$ and $G(\mathbf{s},\mathbf{x}) - \overline{z}$ are respectively the performance and limit-state design expressions considered, where $\overline{z}$ is a constant threshold of $G(\mathbf{s},\mathbf{x})$. The failure domain specified for the design variables $\mathbf{s}$ (i.e., $\mathsf{F} = \big\{\mathbf{x} \in \Omega \,|\, G(\mathbf{s},\mathbf{x}) - \overline{z} \leq 0\big\}$) reads the probability of failure, namely $P_f = \mathsf{P}[G(\mathbf{s}, \mathbf{x}) - \overline{z} \leq 0] = \int \ldots \int_{\mathsf{F}} f(\mathbf{x})\, \mathrm{d}\mathbf{x}$, where $P_a$ is the allowable threshold of $P_f$.

Various optimization approaches have been developed to solve the problem stated in Eq. (1), mainly categorized by the double-loop, single-loop and decoupling approaches [2, 3]. Despite of some gain in efficiency, solving the RBDO problem in large-scale applications is very challenging. For example, some of the widely-employed conventional methods, such as Monte-Carlo simulation (MCS) [4] and first-order reliability method (FORM) [5], are endowed by the drawbacks related to the expensive computational cost and/or the result error. The quest for the efficient optimization method is essential and ongoing in the research field. Recently, the surrogate models [6-8], such as artificial neural networks, radial basic function, support vector machine and Gaussian process regression (GPR), have been adopted to alleviate the computational burden and the solution inaccuracy that would be possibly experienced by standard techniques.

From the above comments, this paper proposes the novel combined GPR-CLPSO approach that efficiently process the accurate solution of the decoupling form of the problem in Eq. (1), where CLPSO is an acronym for comprehensive learning particle swarm optimization [9]. At the beginning, the CLPSO solves the counterpart to Eq. (1) formulated using the initial vector of random parameters, called the most probable point (MPP). From the optimal result determined by the CLPSO, the GPR is constructed as a surrogate model to approximate the spectrum of the limit state functions under uncertainty parameters, thereby enabling the low-cost MCSs to estimate the failure probability associated with the optimal design. To enhance the accuracy of the failure probability estimate, the GPR model is strategically refined by adding new sampling points to the region with high-reliability sensitivity, which is further transferred to maximize a so-called expected feasibility function (EFF). The CLPSO is then adopted to optimize both the GPR hyperparameters and the learning function EFF. For each decoupling iteration, the MPP is redefined in the CLPSO algorithm to search for the new optimal design solution. The optimization terminates as when the failure probability associated with the design by the deterministic procedures reaches the allowable thresholds. Finally, the robustness and accuracy of the proposed method is illustrated through the benchmark on the problem of a ten-bar truss.

**CLPSO Algorithm**

Whilst the general PSO algorithm provides the good convergence rate to the optimal solutions, the method is often trapped into the local optima. To enable the diversity of design particles and overcome the premature convergence, Liang and Huang [9] developed the CLPSO algorithm using the comprehensive learning (CL) strategy. The velocity and position functions of CLPSO in the next time step $t + 1$ are written as follows:

$$\mathbf{V}_i(t+1) = w\mathbf{V}_i(t) + c_1 r_1 \left( \mathbf{pbest}_{fi(D)} - \mathbf{X}_i(t) \right) + c_2 r_2 \left( \mathbf{gbest} - \mathbf{X}_i(t) \right) \tag{2}$$

$$\mathbf{X}_i(t+1) = \mathbf{X}_i(t) + \mathbf{V}_i(t+1) \tag{3}$$

where $fi(D) = [fi(1), fi(2),\ldots, fi(N_d)]$ indicates if the $i$-th particle follows its own or some other best (called **pbest**$_i$) position for each dimension $D \in \{1,\ldots,N_d\}$; $N_d$ defines the total number of searching dimensions; $\mathbf{X}_i(t) = [X_i^1(t),\ldots, X_i^D(t)]$ and $\mathbf{V}_i(t) = [V_i^1(t),\ldots,V_i^D(t)]$ denote the position and velocity of the $i$-th particle for $i \in \{1,\ldots,N_p\}$ ($N_p$ is the total number of particle populations) at the $t$-th time, respectively; $w$ is an inertial weight; $c_1$ and $c_2$ are the two accelerating coefficients; $r_1$ and $r_2$ are the two independent random numbers lying uniformly within the [0,1] interval; $\mathbf{pbest}_{fi(D)}^D$ is the exemplar of the $i$-th particle on the $D$-th dimension, as well as those with reference to the best position found by its own **pbest**; and **gbest** is the best position of the whole swarm

The CL strategy can be briefly described in the following three-step implementations [10].

*Step i*. For a generic $i$-th particle, the position $\mathbf{X}_i(t)$ is tested for its solution convergence and optimality. If the deterministic objective functions $C(\mathbf{s} = \mathbf{X}_i(t))$ over $m$ consecutive time steps ($m$ is defined as a refreshing gap) all are larger than the recorded best result $C(\mathbf{s} = \mathbf{pbest}_i)$, perform the CL searching on the $i$-th particle for the best new best position $\mathbf{pbest}_i$ in *Step ii*.

*Step ii*. For each of the $D$-th dimensions of the $i$-th best particle **pbest**$_i$, update based on the learning probability $Pc_i$ in Eq. (4) an exemplar $\mathbf{pbest}_{fi(D)}^D$ from the two random particles $\mathbf{pbest}_{fi(D)}$ within the whole **pbest** populations that yields the lower objective function $C(\mathbf{s} = \mathbf{pbest}_{fi(D)})$.

$$Pc_i = 0.05 + 0.45 \left( \exp\left( \frac{10(i-1)}{N_p - 1} \right) - 1 \right) / \left( \exp(10) - 1 \right) \tag{4}$$

*Step iii*. Repeat *Step ii* for all dimensions $D \in \{1,\ldots,N_d\}$ to generate the new best position $\mathbf{pbest}_i = \mathbf{pbest}_{fi(D)}$ of the $i$-th particle. In the case when some exemplar $\mathbf{pbest}_{fi(D)}^D$ returns the current $\mathbf{pbest}_i^D$, it randomly learns from another $i$-th best particle of the same $D$-th dimension.

**GPR Model**

Consider the training set $\Delta = (\mathbf{X}, \mathbf{y}) = \{(\mathbf{x}_i, y_i) \mid i = 1,\ldots, N\}$ of $N$ observations, where $\mathbf{x}_i \in \mathbb{R}^D$ are the input variables and $y_i \in \mathbb{R}$ is the corresponding output variable (e.g., structural response in this case). The goal is to construct an input-output mapping $y = f(\mathbf{x}) + \epsilon: \mathbb{R}^D \to \mathbb{R}$, where $f(\mathbf{x})$ is the unknown regression function and the noise $\epsilon \sim N(0, \sigma^2_N)$.

The GPR is defined as the distribution over the set of functions $f(\mathbf{x})$ [11, 12]. The distribution is characterized by the mean value function $m(\mathbf{x})$ to represent a prior knowledge about the regression function and the kernel function $\kappa(\mathbf{x}, \mathbf{x}')$ to control the smoothness of the function, such that

$$f(\mathbf{x}) \sim \mathrm{GP}\left(m(\mathbf{x}), \kappa(\mathbf{x}, \mathbf{x}')\right) \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^D. \tag{5}$$

Commonly, a zero-mean function (i.e., $m(\mathbf{x}) = 0$) and the following squared exponential kernel are used in Eq. (5) [11, 12].

$$\kappa(\mathbf{x}, \mathbf{x}') = \alpha_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d - x_d')^2}{l_d^2}\right), \tag{6}$$

where the hyperparameters $\theta = \{\alpha_f, l_d\}$ consist of signal variance $\alpha_f$ and length-scale $l_d$.

Suppose one already defines $\Delta$ and aims to predict the responses $\mathbf{y}^* \in \mathbb{R}^{N^* \times 1}$ for the new set of input variables $\mathbf{X}^* \in \mathbb{R}^{N^* \times D}$. As the GPR nature, the jointed PDF $p(\mathbf{y}, \mathbf{y}^*; \mathbf{X}, \mathbf{X}^*)$ of $\mathbf{y}$ and $\mathbf{y}^*$ placing the condition on $\mathbf{X}^*$ and $\mathbf{X}$ is described by the Gaussian [13]

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_N^2\mathbf{I} & \mathbf{K}(\mathbf{X},\mathbf{X}^*) \\ \mathbf{K}^{\mathrm{T}}(\mathbf{X},\mathbf{X}^*) & \mathbf{K}(\mathbf{X}^*,\mathbf{X}^*) \end{bmatrix}\right) \tag{7}$$

where $\mathbf{K}(\mathbf{X},\mathbf{X}) \in \mathsf{R}^{N \times N} = \{\kappa(\mathbf{x}_i, \mathbf{x}_j) \mid \forall i, \forall i \in (1, \ldots, N)\}$ is the covariance matrix of all input variables (viz., similarly for $\mathbf{K}(\mathbf{X},\mathbf{X}^*) \in \mathsf{R}^{N \times N^*}$ and $\mathbf{K}(\mathbf{X}^*,\mathbf{X}^*) \in \mathsf{R}^{N^* \times N^*}$), and $\mathbf{I} \in \mathsf{R}^{N \times N}$ is an identity matrix.

By applying the conditional distribution to the PDF in Eq. (7), the predictive posterior PDF of the GPR model can be established. Both the posterior mean vector $\boldsymbol{\mu}^*$ and covariance matrix $\boldsymbol{\Sigma}^*$ are analytically expressed as

$$\begin{aligned} \boldsymbol{\mu}^* &= \mathbf{K}^{\mathrm{T}}(\mathbf{X},\mathbf{X}^*)\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_N^2\mathbf{I}\right]^{-1}\mathbf{y} \\ \boldsymbol{\Sigma}^* &= \mathbf{K}(\mathbf{X}^*,\mathbf{X}^*) - \mathbf{K}^{\mathrm{T}}(\mathbf{X},\mathbf{X}^*)\left[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_N^2\mathbf{I}\right]^{-1}\mathbf{K}(\mathbf{X},\mathbf{X}^*) \end{aligned}. \tag{8}$$

Determining the hyperparameters $\theta$ that best fits the data set $\Delta$ is essential for the construction of the GPR model. Based on the statistical regression, the optimal parameter $\hat{\theta}$ reads the maximizer of the marginal likelihood $\mathsf{L} = p(\mathbf{y}; \mathbf{X}, \theta)$, and hence the minimize of the (negative) log-marginal likelihood $\mathsf{L}$ of the training set $\Delta$. As is clear,

$$\hat{\theta} = \arg\min_{\theta}\left(-\log(\mathsf{L})\right) \tag{9}$$

where

$$-\log(\mathsf{L}) = \frac{N}{2}\log(2\pi) + \frac{1}{2}\log\left|\mathbf{K}(\mathbf{X},\mathbf{X};\theta) + \sigma_N^2\mathbf{I}\right| + \frac{1}{2}\mathbf{y}^T\left|\mathbf{K}(\mathbf{X},\mathbf{X};\theta) + \sigma_N^2\mathbf{I}\right|^{-1}\mathbf{y} \tag{10}$$

In this study, the problem in Eq. (9) is processed by the CLPSO schemes.

**Failure Probability Estimation Using MCS and Learning Function EFF**

Let $\hat{G}(\mathbf{x}) \sim N\left(\mu_{\hat{G}}(\mathbf{x}), \sigma_{\hat{G}}^2(\mathbf{x})\right)$ denote the GPR model of the performance function $G(\mathbf{s}, \mathbf{x})$ evaluated at the particular design variables $\mathbf{s}$ (or $G(\mathbf{x})$ for simplification), where $\mu_{\hat{G}}(\mathbf{x})$ and $\sigma_{\hat{G}}^2(\mathbf{x})$ are provided in Eq. (8). The MCSs can be performed to approximate the mean function $\mu_{\hat{G}}(\mathbf{x})$ and thus the failure probability in Eq. (1) by

$$P_f = \int_{\Omega} f(\mathbf{x}) \quad [\mu_{\hat{G}}(\mathbf{x}) - \overline{z}] \mathrm{d}\mathbf{x} = \frac{1}{N_{MCS}} \sum_{k=1}^{N_{MCS}} [\mu_{\hat{G}}(\mathbf{x}_k) - \overline{z}] \tag{11}$$

where $N_{MCS}$ is the number of random samples generated within the space $\Omega$, and $[.]$ denotes the indicator function, namely

$$[\mu_{\hat{G}}(\mathbf{x}_k) - \overline{z}] = \begin{cases} 1 & \text{if} \quad \mu_{\hat{G}}(\mathbf{x}_k) - \overline{z} \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

The prediction of $P_f$ in Eq. (11) considers the uncertainty, but $\hat{G}(\mathbf{x})$ may not be able to capture the failure domain of the performance function, especially at the border between safe and unsafe ones. For this reason, the present approach adopts the learning function EFF [14] for the generation of the new training point added to the training set $\Delta$ giving the systematic refinement of the GPR model $\hat{G}(\mathbf{x})$. In essence, the learning function EFF describes the expectation of the true value of responses that can satisfy the equality constraint $G(\mathbf{x}) - \overline{z} = 0$ based on $\mu_{\hat{G}}(\mathbf{x})$ and $\sigma_{\hat{G}}(\mathbf{x})$ in the current GPR model $\hat{G}(\mathbf{x})$. This expectation enables the searches over the vicinity of the response threshold, where the estimation of $P_f$ is the most sensitive. The analytical expressions of EFF can be found in [14]. Its determination processes the CLPSO schemes.

**Combined GPR-CLPSO Algorithm**

This section summarizes the proposed GPR-CLPSO algorithm as the following step-by-step procedures:

*Step 1*: Initialize the random variables in the original RBDO Eq. (1).
*Step 2*: Perform the deterministic optimization using the CLPSO algorithm to obtain the optimal design solutions.
*Step 3*: Employ the Latin hypercube sampling (LHS) method to generate random samples for the design obtained in *Step 2*. Calculate the actual responses associated with the samples using the limit state functions to create the training set.
*Step 4*: Construct the GPR model for the limit state functions based on the training set obtained in *Step 3*. The optimal hyperparameters for the GPR model are determined by the CLPSO algorithm.
*Step 5*: Compute the maximizer of the learning function EFF, and evaluate the associated actual response functions using the CLPSO schemes.
*Step 6*: Add the point obtained in *Step 5* and the associated response to the training data set. Reconstruct the GPR model, and re-iterate *Step 5* until the number of added points reaches the specified value $n$. The recommended number of added points is $n \in [5, 10]$.

*Step 7*: Based on the GPR model in *Step 6*, perform the MCSs with $N_{MCS}$ random samples to estimate the failure probability $P_f$. Then, update the new MPP for the next deterministic optimization.

*Step 8*: Check the convergent criteria on the failure probability. If the estimated failure probability $P_f$ converges, terminate. Otherwise, repeat *Steps 2 to 8*.
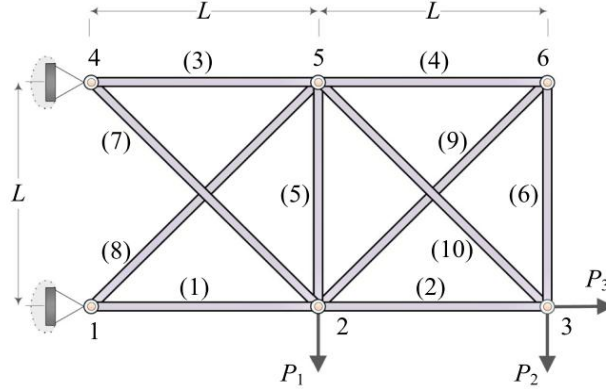
## Illustrative Example



**Figure 2.** Ten-bar truss.

A 10-bar truss in Fig. 2 was considered, and its RBDO problem was formulated in Eq. 1 [15]. The cross-sectional areas were employed the design variables, namely $\mathbf{s} = [s_1, \ldots, s_{10}]^T$, where each variable is independently bounded within the intervals, i.e., $s_i \in [1, 20] \times 10^{-4}$ m2. The intrinsic (indeterministic) uncertainties inherited the external loads ($P_1$, $P_2$ and $P_3$), Young's modulus ($E$) and dimension length ($L$), simultaneously. The random variables were $\mathbf{x} = [P_1, P_2, P_3, E, L]^T$ with the probabilistic properties listed in Table 1.

The vertical displacement at node 3, denoted as $\Delta_3$, was considered as the response performance of interest ($y = -\Delta_3$), whose probability of exceeding the allowable value of $\bar{z} = 4 \times 10^{-3}$ m was less than or equal to $P_a = 6.21 \times 10^{-3}$. Thus, the specific RBDO Eq. (1) was written as follows:

$$\min C(\mathbf{s}) = \sum_{i=1}^{10} s_i$$

$$\text{s.t.} \quad \mathrm{P}\left[G(\mathbf{s},\mathbf{x}) + 4\times10^{-4} \leq 0\right] - 6.21\times10^{-3} \leq 0 .$$

$$s_i \in [1, 20]\times10^{-4}, \ i = 1,\ldots,10 \quad (13)$$

$$G(\mathbf{s},\mathbf{x}) = -\Delta_3$$

In the first instance, the initial random variables were assigned to take the mean values given in Table 1. The deterministic counterpart of the problem in Eq. (13) was solved using the CLPSO algorithm. The parameters adopted were: the total number of particle populations of $N_p = 20$; $c_1 = c_2 = 2$; the inertial weight of $w$ linearly declining from 0.9 to 0.2; and the maximum number of iterations of 500 as per each particle set.

The optimal deterministic results $\mathbf{s}$ computed were adopted as the initial inputs for the construction of the GPR model that approximated the limit state functions in Eq. (13), where the LHS employed 75 training points and each decoupling iteration adopted $n = 5$ added points. The failure probability of the design problem was estimated using the CLPSO

algorithm, and the MPP was updated. The MCSs were performed over the constructed GPR model without the need of any further finite element analysis (FEA) solves giving the computational advantages in approximating the failure probability $P_f$ in Eq. (11). The coupling CLPSO and GPR procedures were iterated until the estimated probability of failure (viz., $P_f = 5.43 \times 10^{-3}$) was converged and well complied with the limit of $P_a = 6.21 \times 10^{-3}$. The proposed GPR-CLPSO method was encoded in Python, and run using the computer hardware with Intel Core i5-9400 CPU @ 2.9 GHz and 16 GB RAM.

The optimal solutions successfully designed by the present GPR-CLPSO method are reported in Table 2 that also provides the direct comparisons to some other benchmarks [1, 15, 16]. It presents the good agreement between the proposed approach and various design techniques in determining the optimal solutions to the RBDO problem in Eq. (13). More explicitly, the proposed GPR-CLPSO computed the optimal design of $C(\mathbf{s}) = 61.293 \times 10^{-4}$ m$^2$ that is not only close but also lies (most minimum) lower than all reported values, namely $62.367 \times 10^{-4}$ m$^2$ in FEA-FORM [15], $63.649 \times 10^{-4}$ m$^2$ in RSM2-FORM-MCS [15], $62.347 \times 10^{-4}$ m$^2$ in PSA-ISAP [16] and $61.482 \times 10^{-4}$ m$^2$ in MGP-SA [1].

The computional efforts (i.e., taking 2,213 s in CPU times) were modest, and the scheme only required 225 FEA iterates with 15 adddd points in the learning function EFF. The corresponding failure probability of $P_f = 5.43 \times 10^{-3}$ complied with the limit of $P_a = 6.21 \times 10^{-3}$, and was close to the reference value of $P_f = 5.21 \times 10^{-3}$ given by standard MCSs exhaustively generating $5 \times 10^5$ random samples on the designed structure. Finally, the plot of the resulting optimal design layout is depicted in Fig. 3, where the sizes of the lines indicate the relative normalized member areas.

**Table 1.** Probabilistic properties of random parameters.

| Variable | Distribution | Mean value | COV |
|---|---|---|---|
| $P_1$ [kN] | Normal | 60 | 0.20 |
| $P_2$ [kN] | Normal | 40 | 0.20 |
| $P_3$ [kN] | Normal | 10 | 0.20 |
| $E$ [GPa] | Normal | 200 | 0.10 |
| $L$ [m] | Normal | 1 | 0.05 |

**Table 2.** Comparisons of optimization results for various analysis approaches.

| Design variables | FEA - FORM [15] | RSM2- FORM- MCS [15] | PSA-ISAP [16] | MGP-SA [1] | GRP-CLPSO (This Study) |
|---|---|---|---|---|---|
| $s_1$ | 10.493 | 10.705 | 10.482 | 10.333 | 10.635 |
| $s_2$ | 5.772 | 5.914 | 4.421 | 5.371 | 5.589 |
| $s_3$ | 14.098 | 14.424 | 15.685 | 13.579 | 13.481 |
| $s_4$ | 1.000 | 1.000 | 1.089 | 1.000 | 1.000 |
| $s_5$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $s_6$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $s_7$ | 5.460 | 5.531 | 7.851 | 6.418 | 5.883 |
| $s_8$ | 11.586 | 11.853 | 10.048 | 11.273 | 11.149 |
| $s_9$ | 1.000 | 1.000 | 1.121 | 1.000 | 1.000 |
| $s_{10}$ | 10.958 | 11.223 | 9.650 | 10.508 | 10.555 |
| $C(\mathbf{s})$ [ $\times 10^{-4}$ m$^2$] | **62.367** | **63.649** | **62.347** | **61.482** | **61.293** |
| No. of FEA iterates | **2,240** | **1,904** | **524** | **20,000** | **225 + 15** |
| Computational time [s] | - | - | - | **3,208** | **2,213** |

| | | | | | |
|---|---|---|---|---|---|
| $[G(\mathbf{s}, \mathbf{x}) \leq 0]$ | $8.51 \times 10^{-3}$ | $6.19 \times 10^{-3}$ | $6.19 \times 10^{-3}$ | $4.34 \times 10^{-3}$ | $5.43 \times 10^{-3}$ |
| $P_f$(MCS) (with $5 \times 10^5$ samples) | $4.22 \times 10^{-3}$ | $2.95 \times 10^{-3}$ | $6.15 \times 10^{-3}$ | $5.64 \times 10^{-3}$ | $5.21 \times 10^{-3}$ |

Note that: +15 is the total number of added points from the learning function EFF.
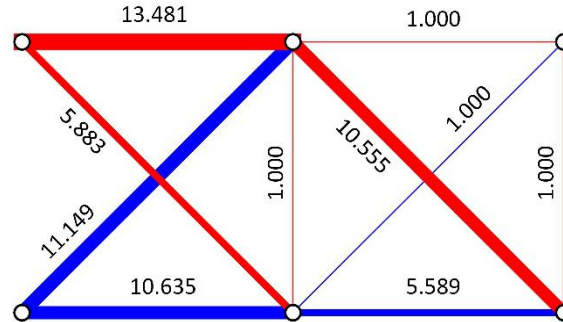


**Figure 3.** Normalized cross-sectional areas of optimally designed members, where blue and red colors indicate compression and tension, respectively.

## Concluding remarks

The paper has presented the computationally efficient GPR-CLPSO method that couples the CLPSO algorithm with the GPR model in processing the decoupling RBDO problem. The CLPSO schemes in particular initialize the nominal design points (assuming the deterministic optimization), which are subsequently adopted to estimate the failure probability and the updated MPP underpinning the limit state functions. The GPR model is constructed such that the reliability responses of limit state functions can be mapped out using only few training points (enhanced by the use of a so-called learning function EFF). What is interesting is that the method advantageously by-passes the finite element analysis (time consuming) computations as when the GPR model has been estimated. The MCSs can be fully enumerated on the GPR model at modest computing efforts even for a large number of samples.

A number of RBRO problems and benchmarks (one of which has been provided herein) have been successfully performed by the GPR-CLPSO approach. These illustrate the accuracy and robustness of the proposed method.

## Acknowledgements

### References

[1] Do B, Ohsaki M, Yamakawa M. (2021) Sequential mixture of Gaussian processes and saddlepoint approximation for reliability-based design optimization of structures. *Structural & Multidisciplinary Optimization*, 1-24.
[2] Aoues Y, Chateauneuf A. (2010) Benchmark study of numerical methods for reliability-based design optimization. *Structural & Multidisciplinary Optimization* **41**, 277-94.
[3] Moustapha M, Sudret B. (2019) Surrogate-assisted reliability-based design optimization: a survey and a unified modular framework. *Structural & Multidisciplinary Optimization*, 1-20.
[4] Papadrakakis M, Lagaros ND. (2002) Reliability-based structural optimization using neural networks and Monte Carlo simulation. *Computer Methods in Applied Mechanics and Engineering* **191**, 3491-507.

[5] Zhao Y-G, Ono T. (1999) A general procedure for first/second-order reliability method (FORM/SORM). *Structural Safety* **21**, 95-112.

[6] Cheng K, Lu Z. (2020) Structural reliability analysis based on ensemble learning of surrogate models. *Structural Safety* **83**, 101905.

[7] Lehký D, Slowik O, Novák D. (2018) Reliability-based design: Artificial neural networks and double-loop reliability-based optimization approaches. *Advances in Engineering Software* **117**, 123-35.

[8] Jing Z, Chen J, Li X. RBF-GA: (2019) An adaptive radial basis function metamodeling with genetic algorithm for structural reliability analysis. *Reliability Engineering & System Safety* **189**, 42-57.

[9] Liang JJ, Qin AK, Suganthan PN, Baskar S. (2006) Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE Transactions on Evolutionary Computation* **10**, 281-95.

[10] Zhang K, Huang Q, Zhang Y. (2019) Enhancing comprehensive learning particle swarm optimization with local optima topology. *Information Sciences* **471**, 1-18.

[11] Murphy KP. (2012) *Machine learning: a probabilistic perspective*: MIT press.

[12] Rasmussen CE, Williams C. (2006) *Gaussian processes for machine learning*. MIT press Cambridge MA.

[13] Schulz E, Speekenbrink M, Krause A. (2018) A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology* **85**, 1-16.

[14] Bichon BJ, McFarland JM, Mahadevan S. (2011) Efficient surrogate models for reliability analysis of systems with multiple failure modes. *Reliability Engineering & System Safety* **96**, 1386-95.

[15] Zhao W, Qiu Z. (2013) An efficient response surface method and its application to structural reliability and reliability-based optimization. *Finite Elements in Analysis Design* **67**, 34-42.

[16] Ghasemi MR, Camp CV, Dizangian B. (2019) Novel decoupled framework for reliability-based design optimization of structures using a robust shifting technique. *Frontiers of Structural and Civil Engineering* **13**, 800-20.

# Phasor particle swarm optimization of dome structures under limited natural frequency conditions

**\*Ei Cho Pyone[1], Thu Huynh Van[1], Toan Minh Le[1], Linh Van Hong Bui[2] and †Sawekchai Tangaramvong[1]**

[1]Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Chulalongkorn University, Bangkok 10330, Thailand.
[2]Faculty of Civil Engineering, Ho Chi Minh City Open University, Ho Chi Minh City 700000, Vietnam.

\*Presenting author: eicho.ecpyone@gmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

The paper proposes the novel numerical approach, called phasor particle swarm optimization (PPSO), for the optimal design of structures under natural frequency constraints. The proposed scheme develops the phasor theory in mathematics directly within the PSO algorithm. In essence, a phase angle incorporating the periodic sine and cosine functions is applied during the optimization process to model particle control parameters. This technique enables the fast-learning strategy of the particle velocity and is thus able to capture the optimal sizing distributions of structural members under some specified conditions at modest computing efforts. The application of the proposed method is illustrated for the optimal design of dome structures subjected to the responded natural frequency limits. To demonstrate the effectiveness and robustness of the proposed PPSO algorithm, a simple 3-D dome structure is successfully tested, and then the obtained results are compared with those of the other meta-heuristic algorithms in the literature.

**Keywords:** Phasor Theory, Particle Swarm Optimization, Dome Structures, Natural Frequency Conditions, Meta-Heuristic Algorithms.

## Introduction

The natural frequency is one of the major parameters that indicates the dynamic performance of the structural system. Imposing some appropriate constraints on the frequency avoids undesirable vibrations and resonance under external excitations [1, 2]. The natural frequency constraints generate complexity in the optimal design of structures. The sizing optimization of the dome structure under natural frequency constraints for instance is regarded as the highly implicit nonlinear and/or non-convex problem [1, 3]. To address this problem, the efficient and robust optimization techniques are developed to provide both the most minimum of design structural weight and reasonable computing efforts.

Many metaheuristic algorithms have been employed to systematically capture the weight minima of practical structures without the need of mathematical programming implementations. A wide class of these methods has been studied in the structural optimization with natural frequency conditions. Some of which are genetic algorithm (GA) [4], particle swarm optimization (PSO) [5], democratic PSO (DPSO) [6], colliding-bodies optimization (CBO) [7], harmony search-based mechanism into PSO with aging leader and challengers (HALC-PSO) [8], modified sub-population TLBO (MS-TLBO) [9], colliding-bodies optimization (CBO) [7], vibrating particles system (VPS) [10], improved symbiotic organism search (ISOS) [11], improved differential evolution (IDE) [12] and adaptive hybrid evolutionary firefly (AHEFA) [13].

This paper proposes the sizing optimization method, based on a so-called phasor PSO (PPSO) [14], of dome-like truss structures under the constraints on the associated natural frequency. The implementation of PPSO is simple and adaptive, as it entails only the phase angle incorporating the periodic sine and cosine functions to model particle control parameters. The technique enables the fast-learning strategy of the particle velocity, and quickly captures the structural members under some specified natural frequency conditions.

**Sizing optimization problem and formulations**

The optimization problem minimizes the total weight ($W$) of the dome structure, where the design variables read member cross-sectional areas, namely $A_d$ for each $d$-th member. The constraints consider the intrinsic structural responses and the required natural frequency. This can be mathematically described as follows:

$$
\begin{aligned}
&\text{Find} && A_d \ \text{ for } \ \forall d \in \{1, \ldots, n_d\} \\
&\text{Minimize} && W = \sum_{d=1}^{n_d} \rho_d A_d L_d \\
&\text{Subjected to} && \omega_j \leq \omega_j^*, \quad \omega_k \leq \omega_k^* \\
&&& A_{\min} \leq A_d \leq A_{\max}
\end{aligned}
\tag{1}
$$

where $n_d$ is the total number of (pin-jointed) truss members, $\rho_d$ the material density, $L_d$ the length of a generic $d$-th member, $\omega_j$ and $\omega_k$ the response natural frequencies (i.e., the $j$-th and $k$-th modes, respectively), $\omega_i^*$ and $\omega_j^*$ the natural frequency limits, $A_{min}$ and $A_{max}$ the lower and upper limits on the available sectional areas, respectively.

The problem in Eq. (1) is further reformulated by applying the penalty function $f$ to the objective function (total design weight) $W$:

$$
\left.
\begin{aligned}
f &= W(1 + \varepsilon_1.C)^{\varepsilon_2} \\
C &= c_\omega^j + c_\omega^k
\end{aligned}
\right\},
\tag{2}
$$

$$
\text{where } c_\omega^j = \begin{cases} \dfrac{|\omega_j|}{|\omega_j^*|} - 1, & \text{if } \ \omega_j > \omega_j^* \\[2mm] 0, & \text{if } \ \omega_j \leq \omega_j^* \end{cases}, \quad c_\omega^k = \begin{cases} \dfrac{|\omega_k|}{|\omega_k^*|} - 1, & \text{if } \ \omega_k > \omega_k^* \\[2mm] 0, & \text{if } \ \omega_k \leq \omega_k^* \end{cases},
$$

$C$ is the penalty factor associated with the violation of natural frequency constraints, $c_\omega^j$ and $c_\omega^k$ the parameters indicating the satisfaction or violation of the natural frequency conditions. The parameters $\varepsilon_1$ and $\varepsilon_2$ are set to 1 and 1.5 (subsequently increase to 2), respectively.

**Phasor Particle Swarm Optimization Algorithm**

The PPSO method, also known as the nonparametric variant PSO algorithm, is based on the phasor theory in mathematics. The algorithmic procedures are summarized in Fig. 1. The algorithm searches for the global best position of each particle within the design domains. All particles cross over the neighbor particles within $d$ dimensional search spaces using some social cooperation and swarm around the position to optimize (minimize) the objective function. In essence, each generic $i$-th particle in the swarm randomly generates the information, including the current positions $\mathbf{X}_i \in \mathfrak{R}^{n_d} = [X_{i,1}, \ldots, X_{i,n_d}]$, the velocities

$\mathbf{V}_i \in \mathfrak{R}^{n_d} = [V_{i,1}, \ldots, V_{i,n_d}]$ , the distances between the particle best ($\mathbf{pbest}_i \in \mathfrak{R}^{n_d} = [pbest_{i,1}, \ldots, pbest_{i,n_d}]$) and current positions $\mathbf{X}_i$, and the distances between the global best positions $\mathbf{gbest} \in \mathfrak{R}^{n_d} = [gbest_1, \ldots, gbest_{n_d}]$ and $\mathbf{X}_i$. These parameters are iteratively updated to keep the particle converging toward the optimal solutions [14].

In the PPSO algorithm, the phase angle $\theta$ lying within a range between 0 and $2\pi$ radians is introduced as the particle control parameter that incorporates the periodic sine (i.e., in an interval of [−1, 1]) and cosine (in [0, 1]) functions. The variations of these periodic functions with phase angles are depicted in Fig. 2. Both functions $\left|\cos\theta_i^t\right|^{2\sin\theta_i^t}$ and $\left|\sin\theta_i^t\right|^{2\cos\theta_i^t}$ construct the adaptive searching characters through the phase angels of individual particles. The velocities $\mathbf{V}_i$ and positions $\mathbf{X}_i$ associated with the $i$-th particle are updated at the $(t + 1)$-th iteration as follows:

$$X_{i,d}^{t+1} = X_{i,d}^t + V_{i,d}^{t+1} , \tag{3}$$

$$V_{i,d}^{t+1} = \left(\frac{\left|\cos\theta_i^t\right|^{2\sin\theta_i^t}}{npop}V_{i,d}^t\right) + \left|\cos\theta_i^t\right|^{2\sin\theta_i^t}(\mathbf{pbest}_{i,d}^t - X_{i,d}^t) + \left|\sin\theta_i^t\right|^{2\cos\theta_i^t}(\mathbf{gbest}_d^t - X_{i,d}^t) , \tag{4}$$

where $\left|\cos\theta_i^t\right|^{2\sin\theta_i^t} / npop$ is an inertia weight factor controlling the balance between the global and local searching abilities, and $\left|\cos\theta_i^t\right|^{2\sin\theta_i^t}$ and $\left|\sin\theta_i^t\right|^{2\cos\theta_i^t}$ the two acceleration control periodic functions. All particles learn from their $\mathbf{pbest}_i$ and $\mathbf{gbest}$ underpinning the swarm to update the corresponding velocities and positions:

$$\mathbf{pbest}_i^{t+1} = \begin{cases} \mathbf{pbest}_{i,d}^t, & \text{if } f(\mathbf{pbest}_{i,d}^t) \le f(X_{i,d}^{t+1}) \\ X_{i,d}^{t+1}, & \text{otherwise} \end{cases} \quad \text{for } d \in \{1, \ldots, n_d\} , \tag{5}$$

$$\mathbf{gbest}^{t+1} = \begin{cases} \mathbf{pbest}_{i,d}^{t+1}, & \text{if } f(\mathbf{pbest}_{i,d}^{t+1}) \le f(\mathbf{gbest}_d^t) \\ \mathbf{gbest}_d^t, & \text{otherwise} \end{cases} \quad \text{for } d \in \{1, \ldots, n_d\} . \tag{6}$$

The phase angle $\theta_i$ and maximum velocity $\mathbf{V}_{(\max)i}$ of the $i$-th particle are updated in the $(t + 1)$-th iteration by [15].

$$\theta_i^{t+1} = \theta_i^t + \left|\cos\theta_i^t + \sin\theta_i^t\right| 2\pi , \tag{7}$$

$$V_{(\max)i,d}^{t+1} = \left|\cos\theta_i^{t+1}\right|^2 (X_{(\max)d} - X_{(\min)d}) . \tag{8}$$

This technique enables the particles to overcome the searches in the local optimal region as well as premature solution convergence.
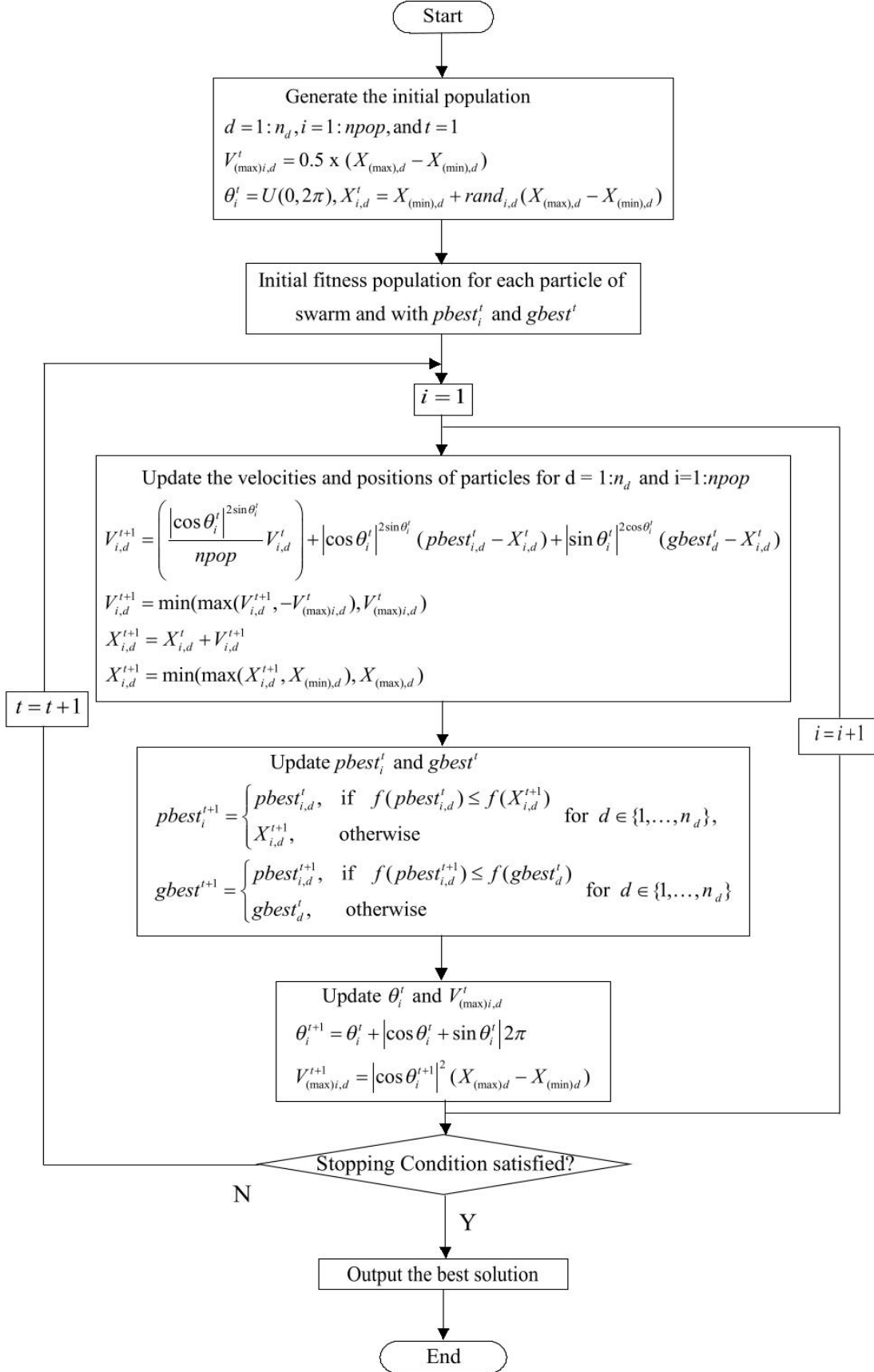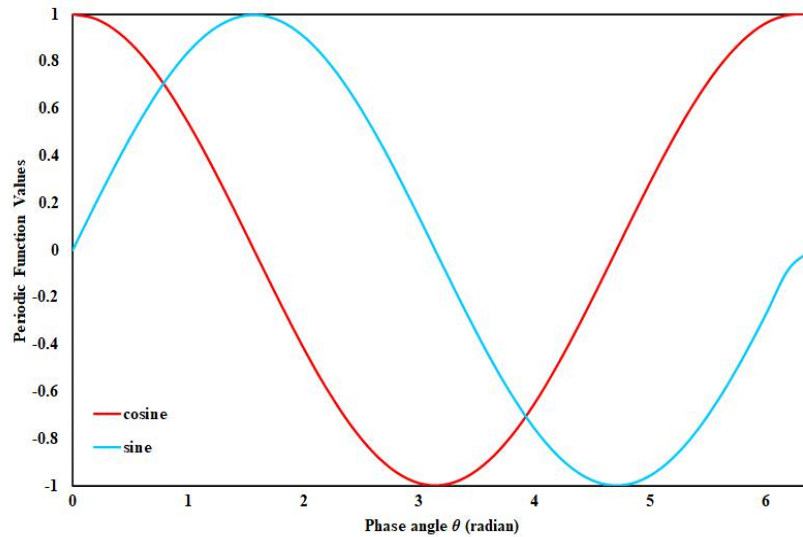
**Fig. 1. PPSO procedures.**

**Fig. 2. Variation of periodic functions for different phase angles.**

**Illustrative Example**

The 120-member dome-like truss structure in Fig. 3 was designed for the cost minimization of all member sizes under the limited natural frequency conditions. The lumped mass of $m_1$ was assigned at node 1, $m_2$ at nodes 2 to 13, and $m_3$ for the remaining nodes. All design member areas were categorized into 7 design groups. The material properties and frequency constraints employed are listed in Table 1.

The optimal sizing design of the dome was successfully performed by the proposed PPSO method with the total of 20 independent runs and the population of 30 particles. The optimal solutions, see Fig. 4, converged at the early number of analysis iterations for all the repeating PPSO solves.

The resulting member sizes of all 7 design groups and the total weight of $W = 8736.24$ kg are reported in Table 2, where the solutions from various analysis methods are also compared. More explicitly, the designed weight values referred are 8896.74 kg by vibrating particles system (VPS) [10], 8890.48 kg by democratic particle swarm optimization (DPSO) [6], 8889.13 kg by colliding-bodies optimization (CBO) [7], 8889.96 kg by harmony search-based mechanism into the particle swarm optimization with an Aging Leader and Challengers (HALC-PSO) [8], 8710.06 kg by improved symbiotic organisms search (ISOS) [11], 8708.73 kg by modified sub-population teaching-learning-based optimization (MS-TLBO) [9], 8707.29 kg by improved differential evolution (IDE) [12], and 8707.26 kg by adaptive hybrid evolutionary firefly (AHEFA) [13]. The optimal design result computed by the present method agrees very well with all benchmarks with the comparable numerical efforts, where the response natural frequencies comply with the imposed limits.

**Table 1. Material properties and design parameters.**

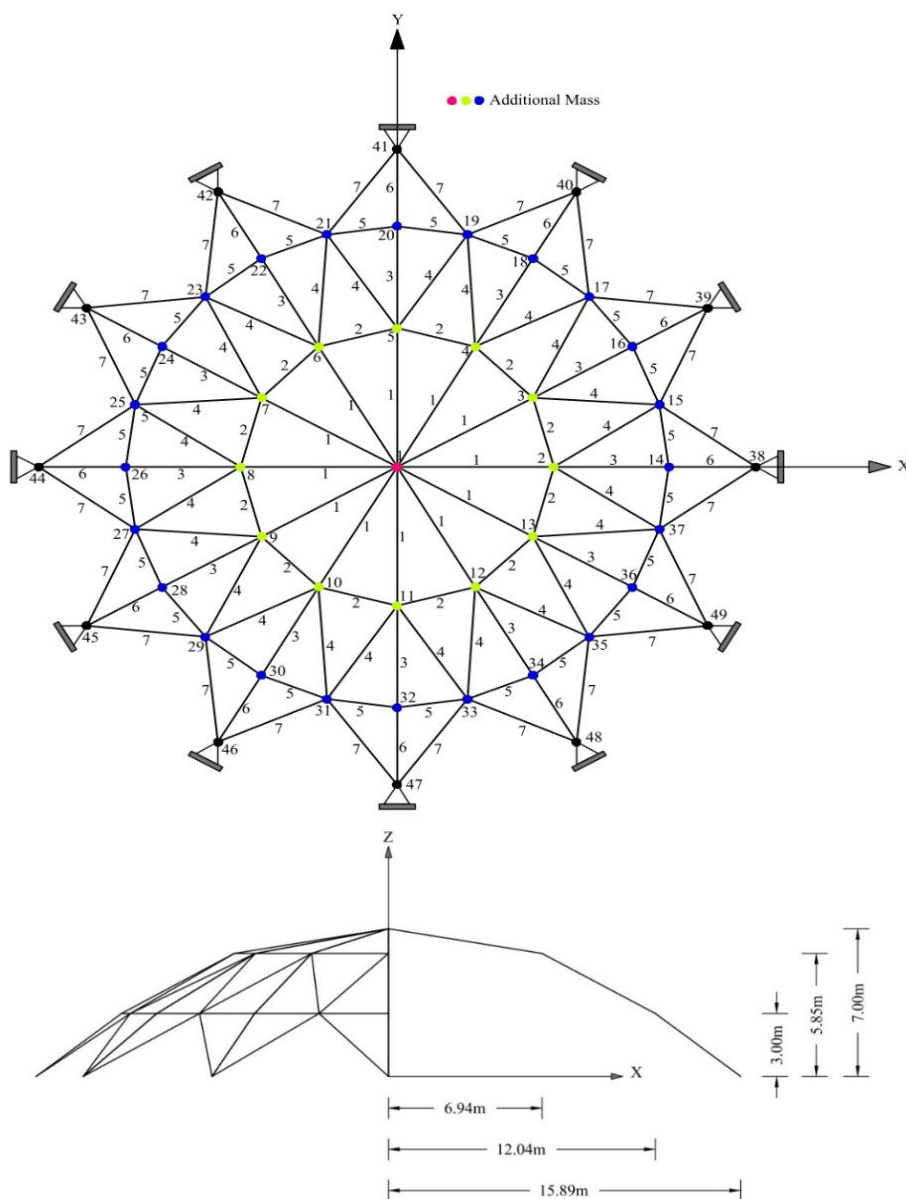| Parameters | Value |
|---|---|
| Modulus of elasticity $E$ (N/m$^2$) | $2.1 \times 10^{11}$ |
| Material density $\rho$ (kg/m$^3$) | 7971.81 |
| Additional mass (kg) | $m_1 = 3000$; $m_2 = 500$; $m_3 = 100$ |
| Allowable range of cross-section (cm$^2$) | $1 \leq A \leq 129.3$ |
| Constraints on the first two frequencies (Hz) | $\omega_1 \geq 9$; $\omega_2 \geq 11$ |



**Fig. 3.   120-bar dome truss geometry.**

**Table 2. Optimal design solutions by variations analysis methods.**

| Design Variables (Areas) cm$^2$ | Kaveh and Ghazaan | Kaveh, A. and A. Zolghadr | Kaveh, A. and V. Mahdavi Dahoei | A. Kaveh, M. Ilchi Ghazaan | Tejani et al. | Tejani et al. | Ho-Huu et al. | Lieu et al. | Present |
|---|---|---|---|---|---|---|---|---|---|
| | [10] | [6] | [7] | [8] | [11] | [9] | [12] | [13] | |
| | VPS | DPSO | CBO | HALC-PSO | ISOS | MS-TLBO | IDE | AHEFA | PPSO |
| A$_1$ | 19.6836 | 19.607 | 19.6917 | 19.8905 | 19.6662 | 19.4886 | 19.4670 | 19.5094 | 19.4084 |
| A$_2$ | 40.9581 | 41.290 | 41.1421 | 40.4045 | 39.8539 | 40.3949 | 40.5004 | 40.3867 | 40.1976 |
| A$_3$ | 11.3325 | 11.136 | 11.1550 | 11.2057 | 10.6127 | 10.6921 | 10.6136 | 10.6033 | 10.7976 |
| A$_4$ | 21.5387 | 21.025 | 21.3207 | 21.3768 | 21.2901 | 21.3139 | 21.1073 | 21.1168 | 21.0787 |
| A$_5$ | 9.8867 | 10.060 | 9.8330 | 9.8669 | 9.7911 | 9.8943 | 9.8417 | 9.8221 | 9.8050 |
| A$_6$ | 12.7116 | 12.758 | 12.8520 | 12.7200 | 11.7899 | 11.7810 | 11.7735 | 11.7735 | 12.1316 |
| A$_7$ | 14.9330 | 15.414 | 15.1602 | 15.2236 | 14.7437 | 14.5979 | 14.8269 | 14.8405 | 14.9168 |
| Best Weight (kg) | 8888.74 | 8890.48 | 8889.1303 | 8889.96 | 8710.062 | 8708.729 | 8707.2898 | 8707.2559 | 8736.242 |
| Number of analyses | 30000 | 6000 | 6000 | 17000 | 4000 | 4000 | 4060 | 3560 | 30000 |
| Mean Weight(kg) | 8896.04 | 8895.99 | 8891.254 | 8900.39 | 8728.5951 | 8734.7450 | 8707.8147 | 8707.5580 | 8737.6056 |
| f$_1$ (Hz) | 9.000 | 9.0001 | 9.000 | 9.000 | 9.001 | 9.0002 | 9.000 | 9.000 | 9.000 |
| f$_2$ (Hz) | 11.000 | 11.0007 | 11.0000 | 11.0000 | 10.998 | 11.0000 | 11.0000 | 11.0000 | 11.0000 |
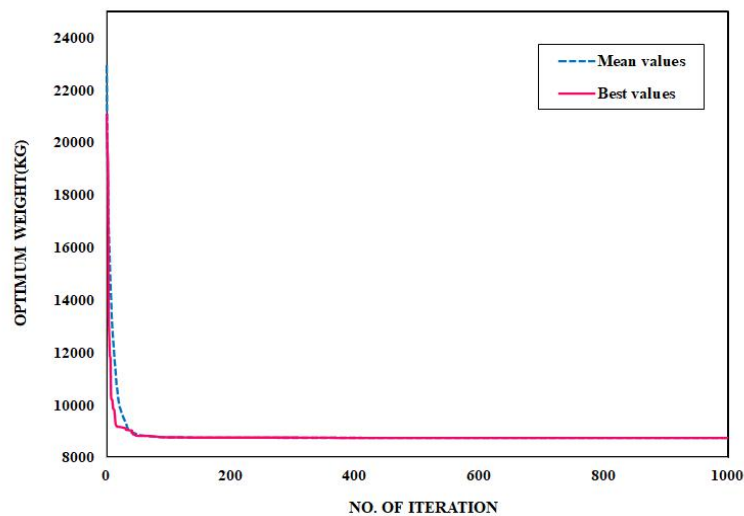


**Fig. 4. Solution convergence by PPSO method.**

**Concluding remarks**

The paper presents the novel PPSO method for the sizing optimization of dome-like truss structure subjected to the constraints on its natural frequencies. The approach mathematically adopts the phasor theory (namely the periodic sine and cosine functions) directly to the PSO that enhances its fast solution searching schemes and more importantly overcomes the likelihood of local optimal (premature) convergence as would be expected in standard PSO techniques. Its applications have been illustrated through the optimal sizing design of the modest-size dome truss structure, where its accuracy and robustness are evidenced by the

good comparisons with various available benchmarks. An ongoing extension of the present work is the development of an effective PPSO algorithm with the capability to incorporate the time-dependent dynamic (seismic) constraints.

## Acknowledgements

## References

[1]     Grandhi, R., Structural optimization with frequency constraints-a review. AIAA journal, 1993. 31(12): p. 2296-2303.
[2]     Bellagamba, L. and T. Yang, Minimum-mass truss structures with constraints on fundamental natural frequency. AIAA Journal, 1981. 19(11): p. 1452-1458.
[3]     Sedaghati, R., A. Suleman, and B. Tabarrok, Structural optimization with frequency constraints using the finite element force method. AIAA journal, 2002. 40(2): p. 382-388.
[4]     Lingyun, W., et al., Truss optimization on shape and sizing with frequency constraints based on genetic algorithm. Computational Mechanics, 2005. 35(5): p. 361-368.
[5]     Gomes, H.M., Truss optimization with dynamic constraints using a particle swarm algorithm. 2011. 38(1 %J Expert Syst. Appl.): p. 957–968.
[6]     Kaveh, A. and A. Zolghadr, Democratic PSO for truss layout and size optimization with frequency constraints. Computers & Structures, 2014. 130: p. 10-21.
[7]     Kaveh, A. and V. Mahdavi Dahoei, Colliding-Bodies Optimization for Truss Optimization with Multiple Frequency Constraints. Journal of Computing in Civil Engineering, 2015. 29.
[8]     Kaveh, A. and M. Ilchi Ghazaan, Hybridized optimization algorithms for design of trusses with multiple natural frequency constraints. Advances in Engineering Software, 2015. 79: p. 137-147.
[9]     Tejani, G.G., V.J. Savsani, and V.K. Patel, Modified sub-population teaching-learning-based optimization for design of truss structures with natural frequency constraints. Mechanics Based Design of Structures and Machines, 2016. 44(4): p. 495-513.
[10]    Kaveh, A. and M. Ilchi Ghazaan, Vibrating particles system algorithm for truss optimization with multiple natural frequency constraints. Acta Mechanica, 2017. 228(1): p. 307-322.
[11]    Tejani, G.G., et al., Truss optimization with natural frequency bounds using improved symbiotic organisms search. Knowledge-Based Systems, 2018. 143: p. 162-178.
[12]    Ho-Huu, V., et al., Optimal design of truss structures with frequency constraints using improved differential evolution algorithm based on an adaptive mutation scheme. Automation in Construction, 2016. 68: p. 81-94.
[13]    Lieu, Q.X., D.T.T. Do, and J. Lee, An adaptive hybrid evolutionary firefly algorithm for shape and size optimization of truss structures with frequency constraints. Computers & Structures, 2018. 195: p. 99-112.
[14]    Camp, C.V., Design of space trusses using big bang-big crunch optimization. 2007. 133(7): p. 999-1008.
[15]    Ghasemi, M., et al., Phasor particle swarm optimization: a simple and efficient variant of PSO. Soft Computing, 2019. 23(19): p. 9701-9718.

# A Turbulent Flow of Water-Based Optimization for the Optimal Sizing Design of Steel Trusses

**\*Saw Thiri Khaing[1], Thu Huynh Van[1], and †Sawekchai Tangaramvong[1]**

1 Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Chulalongkorn University, 10330 Bangkok, Thailand

\*Presenting author: sawthirikhaing@gmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

The paper presents the novel meta-heuristic, called turbulent flow of water-based optimization, algorithm to determine the optimal distribution of steel member sizes allocated to the truss structure that can safely sustain the specified design forces. The problem states the minimization of the cost function, described by the total weight of the designed structure, complying with the limit state specifications. The TFWO method performs the random searches among various whirlpool sets, where the best particle position of each group is pulled down by the centripetal force to the cavity in the center of a whirlpool. A centrifugal force acting in an opposite direction to the centripetal force randomly transfers the particle to the new position. The interaction between different whirlpools applies the individual centripetal forces to iteratively unify those surrounding whirlpools into ones with stronger tractions, and subsequently converges the optimal design solution. The accuracy of the TFWO scheme is illustrated through comparisons with some benchmarks processed by various recent optimization algorithms. These examples present the robustness of the proposed approach in the optimal design of steel structures at modest computing resources.

**Keywords:** Turbulent Flow of Water-based Optimization, Structural Optimization, Meta-Heuristic Algorithm, Steel Structures, Optimal Sizing Design**.**

## Introduction

The structural optimization determines the optimal distribution of members and sizes assigned to the structure under the required strength and serviceability performance criteria. The problem is typically written in the mathematical formulations aiming to computing the objective function (typically cost minimization) subjected to the constraints intrinsically describing the targeted design responses. The fast growing of recent computing technologies has encouraged the development of meta-heuristic methods that systematically perform the iterative-type design procedures to converge the optimal solutions. The methods are generally inspired by the concepts observing the nature-like collective birds and animal behaviors, e.g., genetic algorithm [1], particle swarm optimization (PSO) [2], artificial bee colony (ABC) [3], flower and big bang-big crunch [4] etc. On the other hand, one of the major drawbacks underlying is the return of local optimum leading to the premature solution convergence. The ability in obtaining the accurate optimal designs is largely problem dependent. The exploration of new and suitable methods is thus necessary for the specific structural design problems considered.

This study proposes the development of a so-called turbulent flow of water-based optimization (TFWO) to process the sizing design of steel trusses under the required forces[5]. The TFWO is inspired by the random behaviors in nature established for examples in rivers, seas and oceans. It provides the optimal solutions of various complex problems with real-parameter

benchmark functions for different dimensions. The specific problem considers the cost (total weight) minimization as the objective function subject to the constraints on the limited strength and serviceability responses of the design structures.

## Optimal Sizing Design Formulations

The minimum weight design of the pin-connected steel truss structure can be mathematically described as follows:

$$\text{Minimize} \quad W = \sum_{i=1}^{n} \rho_i A_i L_i \tag{1}$$

$$\text{subject to} \quad \sigma_i \leq \sigma_{all}, \, i = 1,2, \text{------},n$$
$$\delta_i \leq \delta_{all}, \, i = 1,2, \text{------},n \tag{2}$$

where $W$ is the total weight of the designed structure; $n$ is the number of all truss members; $\rho_i$ is the material density of the $i$-th member for $\forall i \in \{1,\dots,n\}$; $A_i$ is the member cross-sectional area (defined as the design variables); $L_i$ is the member length; $\sigma_i$ is the member stress; $\sigma_{all}$ is the allowable stress; $\delta_j$ is the nodal displacement for $\forall j \in \{1,\dots,d\}$; and $\delta_{all}$ is the limited displacement at some $j$-th specified degree of freedom.

## Turbulent Flow of Water-based Optimization

The TFWO is based on the whirlpool behaviors in developing the robust grouping optimization. The algorithm defines $X_i$ as the position of the object and $Wh_j$ the position of each whirlpool (i.e., being the best position occurring in the object). In the beginning, the method divides the population into various whirlpool sets, where the best position in each group generates the traction strength. Each whirlpool is unifying the object positions to the whirlpool center by applying the centripetal force (i.e., $X_i = Wh_j$). Other whirlpools lead to some deviations resulting in the new position of the object described by

$$\Delta X_i = \left( \cos(\delta_i^{new}) * rand\,(1,D) * \left(Wh_f - X_i\right) - \sin(\delta_i^{new}) * rand\,(1,D) * \left(Wh_w - X_i\right) \right) *$$
$$(1 + |\cos(\delta_i^{new}) * -\sin(\delta_i^{new})|) \tag{3}$$

$$X_i^{new} = Wh_j - \Delta X_i \tag{4}$$

At variance with the centripetal force attracting the moving object toward its whirlpool, the centrifugal force pushes the object away the center. In the instance when the centrifugal force overcomes the centripetal counterpart as defined in Eq. (5), the object position transfers to the new position. The centrifugal force $FE_i$ is described in Eq. (6) if it is greater than the random values:

$$FE_i = ((\cos(\delta_i^{new}))^2 * (\sin(\delta_i^{new}))^2)^2 \tag{5}$$

$$X_{i,p} = X_p^{min} + rand * (X_p^{max} - X_p^{min}) \tag{6}$$

Moreover, the position of whirlpools can be influenced by the other whirlpool. The whirlpool positions is updated as follows:

$$Wh_j^{new} = Wh_f - \Delta Wh_j \tag{7}$$

$$\Delta Wh_j = rand\,(1, D) * \left| cos(\delta_j^{new}) + sin(\delta_j^{new}) * (Wh_f - Wh_j) \right| \tag{8}$$

In the case where the best object among all members in the set is stronger than the whirlpool itself, the new whirlpool is updated by this best object for the consequent iteration.

**Illustrative Example**

The applications of the proposed TFWO method have been illustrated through the sizing optimization of 3D steel trusses under strength and serviceability constraints. A number of design benchmarks and examples have been successfully solved by the present scheme. One of which, namely the design of 25-bar space tress in Fig. 1 [6] is considered in this work. The design forces applied were 0.5 kip at node 3 in the positive x-direction, 0.6 kip at node 6 in both negative y and z directions and 10 kips at nodes 1 and 2 in both negative y and z directions.
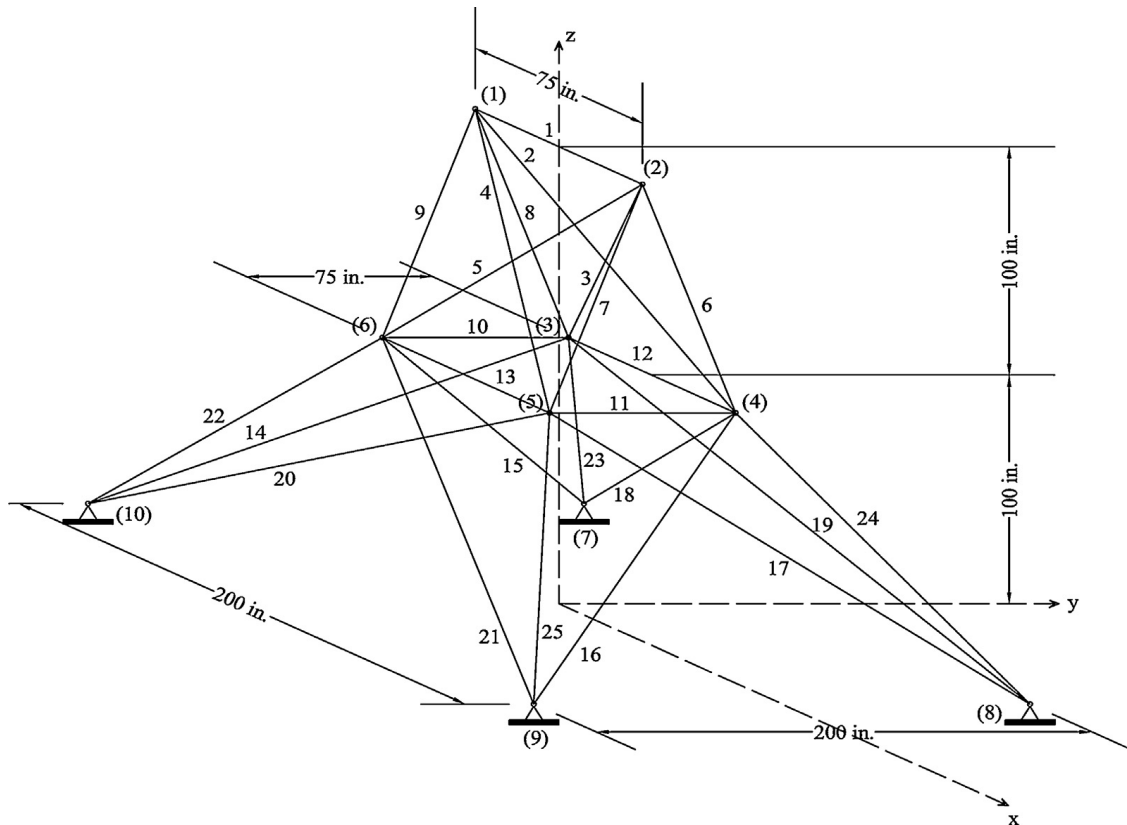


**Fig. 1. 25-bar space truss structure**

The design variables defined the member areas categorized into 8 different groups as listed in Table 1. The material properties employed were elastic modulus of 10,000 ksi (68,950 MPa) and uniform material density of 0.1 lb.in$^{-3}$ (2767.99 kg.m$^{-3}$). The cross-sectional areas were selected within the range between 0.01 in$^2$ and 3.4 in$^2$. The allowable displacements of each node were limited to the variation of 0.35 in at x- and y-directions. The maximum stress limits in all compression and tension members are listed in Table 1.

**Table 1 Member group and stress limits.**

| Group | Members | Compression stress limit (ksi) | Tension stress limit (ksi) |
|---|---|---|---|
| 1 | $A_1$ | 35.092 | 40 |
| 2 | $A_2 - A_5$ | 11.59 | 40 |
| 3 | $A_6 - A_9$ | 17.305 | 40 |
| 4 | $A_{10}, A_{11}$ | 35.092 | 40 |
| 5 | $A_{12}, A_{13}$ | 35.092 | 40 |
| 6 | $A_{14} - A_{17}$ | 6.759 | 40 |
| 7 | $A_{18} - A_{21}$ | 6.959 | 40 |
| 8 | $A_{22} - A_{25}$ | 11.082 | 40 |

The optimal design of the steel space truss was successfully performed by the proposed TFWO method within 50 analysis iterations. The solution (total weight) convergence with the number of analysis (up to 400) iterations is clearly depicted in Fig. 2. More explicitly, the minimum weight of 482.026 lbs was computed at the 42-th iteration and took only 23 seconds. The optimal results, including the total weight and designed member areas, are reported in Table 2, and agree well with those from benchmarks [6], [7], [8], [9], [10], and [4]. In essence, the present TFWO approach provides the most minimum weight solution with the satisfaction of all constraints.
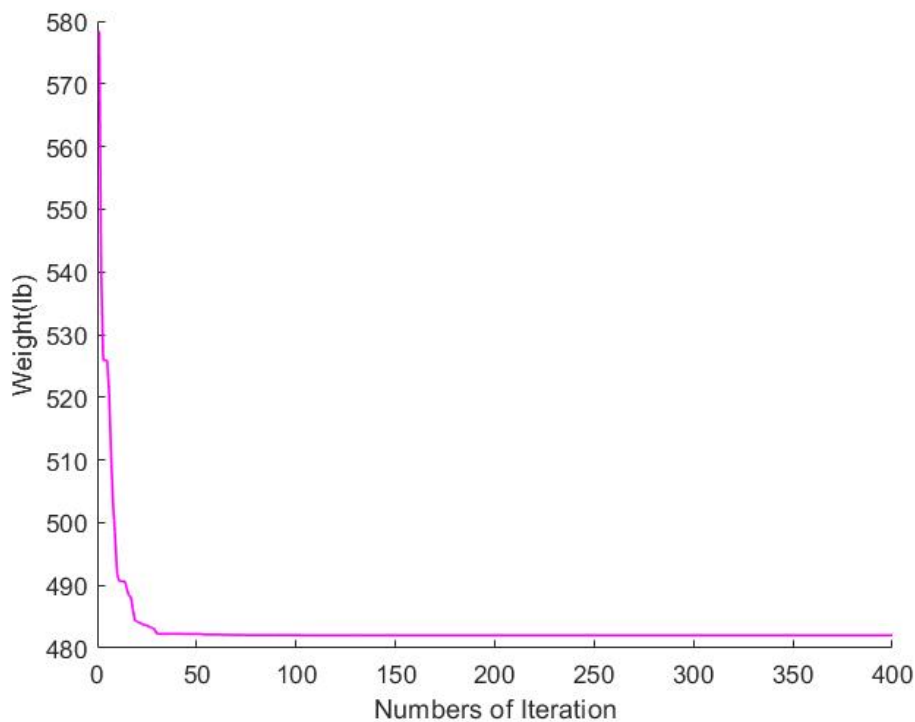


**Fig. 2. Convergence History of 25-bar space truss**

**Table 2 Optimum results for various design methods.**

| Design Variables | Cao et al. GA [8] | Li et al. HPSO [9] | TLBO [10] | Camp [4] | ACO [7] | FPA [6] | TFWO (Present) |
|---|---|---|---|---|---|---|---|
| $A_1$ | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 |
| $A_2$ | 2.0119 | 1.9700 | 1.9878 | 2.0920 | 2.0000 | 1.8300 | 0.4231 |
| $A_3$ | 2.9493 | 3.0160 | 2.9914 | 2.9640 | 2.9660 | 3.1834 | 3.4000 |
| $A_4$ | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 |
| $A_5$ | 0.0295 | 0.0100 | 0.0100 | 0.0100 | 0.0120 | 0.0100 | 1.9170 |
| $A_6$ | 0.6838 | 0.6940 | 0.6828 | 0.6890 | 0.6890 | 0.7017 | 0.9653 |
| $A_7$ | 1.6798 | 1.6810 | 1.6764 | 1.6010 | 1.6790 | 1.7266 | 0.4728 |
| $A_8$ | 2.6759 | 2.6430 | 2.6656 | 2.6860 | 2.6680 | 2.5713 | 3.4000 |
| Weight (lb) | 545.8000 | 545.1900 | 545.1750 | 545.3800 | 545.5300 | 545.1590 | 482.0268 |

## Concluding Remarks

The novel TFWO method has been presented for the optimal sizing design of steel space truss structures under applied forces. Both the material capacities and displacement limits are imposed directly as the design constraints. The TFWO determines the minimum of the total weight (cost) of the design structure with the strict satisfaction of all constraints. A number of design examples have been successfully proceeded by the proposed scheme at modest computing efforts. The accuracy of the optimal designed variables as compared to the benchmarks can be achieved. In essence, the more minimum weight of all members employed as illustrated in this paper is evidenced.

## Acknowledgments

## References

[1] Coello, C.C., M. Rudnick, and A.D. Christiansen. Using genetic algorithms for optimal design of trusses. in Proceedings Sixth International Conference on Tools with Artificial Intelligence. TAI 94. 1994. IEEE.
[2] Camp, C.V., B.J. Meyer, and P.J. Palazolo, Particle swarm optimization for the design of trusses, in Structures 2004: Building on the Past, Securing the Future. 2004. p. 1-10.
[3] Sonmez, M.J.A.S.C., Artificial Bee Colony algorithm for optimization of truss structures. 2011. 11(2): p. 2406-2418.
[4] Camp, C.V., Design of space trusses using Big Bang–Big Crunch optimization. Journal of Structural Engineering, 2007. 133(7): p. 999-1008.
[5] Ghasemi, M., et al., A novel and effective optimization algorithm for global optimization and its engineering applications: Turbulent Flow of Water-based Optimization (TFWO). 2020. 92: p. 103666.

[6] Bekdaş, G., S.M. Nigdeli, and X.-S.J.A.S.C. Yang, Sizing optimization of truss structures using flower pollination algorithm. 2015. 37: p. 322-331.

[7] Camp, C.V. and B.J. Bichon, Design of space trusses using ant colony optimization. Journal of structural engineering, 2004. 130(5): p. 741-751.

[8] Cao, G., Optimized design of framed structures using a genetic algorithm. 1997.

[9] Li, L., et al., A heuristic particle swarm optimizer for optimization of pin connected structures. 2007. 85(7-8): p. 340-349.

[10] Camp, C.V. and M. Farshchin, Design of space trusses using modified teaching–learning based optimization. Engineering Structures, 2014. 62: p. 87-97.

# Simultaneous Size and Shape Structural Optimization using Enhanced Comprehensive Learning Particle Swarm Optimization

## *Soviphou Muong[1], Thu Huynh Van[1], Chung Nguyen Van[1, 2], and †Sawekchai Tangaramvong[1]

[1] Applied Mechanics and Structures Research Unit, Department of Civil Engineering, Chulalongkorn University, Bangkok 10330, Thailand.
[2] Faculty of Civil Engineering, HCMC University of Technology and Education, Ho Chi Minh 721400, Vietnam.

*Presenting author: soviphou@gmail.com
†Corresponding author: sawekchai.t@chula.ac.th

## Abstract

The paper proposes an enhanced version of the comprehensive learning particle swarm optimization (CLPSO) method for the simultaneous optimal size and shape design of steel truss structures under applied forces. The CLPSO approach incorporates the two novel enhancing techniques, namely perturbation-based exploitation and adaptive learning probabilities, in addition to its distinctive diversity of particles preventing the premature local optimum solutions. In essence, the perturbation enables the robust exploitation of the updating velocity of particles, whilst the learning probabilities are dynamically adjusted by the ranking information of personal best particles. A combination of these techniques results in the fast convergence and likelihood of the global optimum solution. Applications of the enhanced CLPSO method are illustrated through a number of successfully solved truss design examples. The robustness and accuracy of the proposed scheme are evidenced by the comparisons with available benchmarks processed by some other metaheuristic algorithms in obtaining the optimal size and shape distributions of steel trusses complying with limit state specifications.

**Keywords:** Comprehensive Learning; Particle Swarm Optimization; Perturbation-Based Exploitation; Adaptive Learning Probabilities; Size and Shape Optimization

## Introduction

Particle swarm optimization (PSO) first introduced by Kennedy and Eberhart [1] in 1995 emulates the movement or social behavior of a bird flock. An PSO is a popular swarm-intelligence-based algorithm which is used in many real-world optimization problems. However, the solutions of complex problems can often be trapped in the local optima. Many approaches have been studied to improve the performance of PSO methods. In 2014, Xiang Yu and Xueqing Zhang [3] proposed an enhanced comprehensive learning PSO (ECLPSO) based on the original concept of comprehensive learning PSO (CLPSO) [2], where two enhancing techniques construct the perturbation-based exploitation together with adaptive learning probabilities.

This paper presents the novel ECLPSO method in the simultaneous size and shape optimization of planar truss structures. By considering both size and shape variables, the optimization provides the more economical material design than the size optimization alone. The applications of the proposed ECLPSO and its accuracy in obtaining the optimal solution are illustrated through the comparisons with some available benchmarks.

**State Optimization Problem**

The weight (cost) minimization problem of the truss structure consisting of $n$ pin-connected members can be mathematically formulated in two design size (namely $\mathbf{X}_A \in \mathfrak{R}^n = \{A_1,\ldots,A_n\}$) and shape ($\mathbf{X}_G \in \mathfrak{R}^{ng} = \{G_1,\ldots,G_{ng}\}$) variables as follows:

$$
\left.
\begin{aligned}
&\text{find} &&\mathbf{X} \in \mathfrak{R}^{n+ng} = \{\mathbf{X}_A, \mathbf{X}_G\} \\
&\text{minimize} &&W(\mathbf{X}_A, \mathbf{X}_G) = \sum_{i=1}^{n} \rho_i L_i A_i \\
&\text{subject to} &&\sigma_i^c \leq \sigma_i \leq \sigma_i^t &&\text{for} \quad \forall i \in \{1,\ldots,n\} \\
& &&\delta_{\min} \leq \delta_j \leq \delta_{\max} &&\text{for} \quad \forall j \in \{1,\ldots,m\} \\
& &&A_{\min} \leq A_i \leq A_{\max} &&\text{for} \quad \forall i \in \{1,\ldots,n\}
\end{aligned}
\right\},
\tag{1}
$$

where $W$ is the total weight of the design structure defined as the function of member density $\rho_i$, physical length $L_i$ and cross-sectional area $A_i$, $m$ the total number of degrees of freedom, $ng$ the total number of geometry variables, $\delta_j$ the displacement at the $j$-th degree of freedom, and $\sigma_i$ the member stress. The optimization problem in Eq. (1) minimizes the total weight $W$ of the structure under the bounds on permissible compression $\sigma_i^c$ and tension $\sigma_i^t$ stresses, minimum $\delta_{\min}$ and maximum $\delta_{\max}$ displacements, and minimum $A_{\min}$ and maximum $A_{\max}$ areas.

The penalty method reformulates the constrained design Eq. (1) to an unconstrained optimization problem [7]-[9]. The function is defined below:

$$
W' = W(\mathbf{X}_A, \mathbf{X}_G)\varphi_p K,
\tag{2}
$$

$$
\varphi_p = (1+C)^\varepsilon,
\tag{3}
$$

where $K$ and $\varepsilon$ are the penalty constant and penalty exponent (viz., $\varepsilon = 1$ in this study), $C$ is the parameter measuring the violation of penalty constraints:

$$
C = \sum_{j=1}^{m} C_\delta^j + \sum_{i=1}^{n} C_\sigma^i
\tag{4}
$$

$C_\delta^j$ and $C_\sigma^i$ are the displacement and stress constraints, respectively:

$$
\left.
\begin{aligned}
C_\delta^j &= \left| \frac{\delta_j - \delta_{\min}}{\delta_{\min}} \right| &&\text{if} \quad \delta_j < \delta_{\min} \\
C_\delta^j &= \left| \frac{\delta_j - \delta_{\max}}{\delta_{\max}} \right| &&\text{if} \quad \delta_j > \delta_{\max} \\
C_\delta^j &= 0 &&\text{if} \quad \delta_{\min} < \delta_j < \delta_{\max}
\end{aligned}
\right\},
\tag{5}
$$

$$
\left.\begin{array}{ll}
C_\sigma^i = \left| \dfrac{\sigma_i - \sigma_i^t}{\sigma_i^t} \right| & \text{if } \sigma_i > \sigma_i^t \\[3mm]
C_\sigma^i = \left| \dfrac{\sigma_i - \sigma_i^c}{\sigma_i^c} \right| & \text{if } \sigma_i < \sigma_i^c \\[3mm]
C_\sigma^i = 0 & \text{if } \sigma_i^c \le \sigma_i \le \sigma_i^t
\end{array}\right\} . \tag{6}
$$

## Comprehensive Learning Particle Swarm Optimization

The CLPSO approach is a variant of the PSO algorithm pioneered by Liang and Qin [2] in 2006. The method employs the strategy that updates the particles' velocity by learning from all other particles' best information to prevent the solution from premature convergence. The velocity and position of a generic particle are determined by:

$$
V_{i,d} = wV_{i,d} + cr_d (pbest_{f_i,d} - X_{i,d}) \tag{7}
$$

$$
X_{i,d} = X_{i,d} + V_{i,d} \tag{8}
$$

where $i$ and $d$ are the array indices of the particles (population) and dimensions (design variables), respectively; $X_{i,d}$ and $V_{i,d}$ are the position and velocity of the $i$-th particle at the $d$-th dimension, respectively; $w$ is the inertia weight; $c$ is the acceleration coefficient being equal to 1.5; $r_d$ is a random number in the range of $[0,1]$; $f_i$ is the exemplar index the $i$-th particle follows; $pbest_{f_i,d}$ is the best position of the $f_i$-th particle for the $d$-th dimension.

$$
L_i = L_{\min} + (L_{\max} - L_{\min}) \frac{\exp\left(\dfrac{10(i-1)}{N-1}\right) - 1}{\exp(10) - 1} . \tag{9}
$$

The selection of the exemplar index is based on the learning probability $L_i$. More explicitly, if the random number in a range of $[0,1]$ is greater than $L_i$ then the exemplar index $f_i$ reads the $i$-th particle's index. Otherwise, the index $f_i$ takes the location of particle associated with the best of the two fitness values randomly selected from the populations of the considered $d$-th dimension.

The inertia weight is defined by:

$$
w = w_{\max} - \frac{k}{k_{\max}} (w_{\max} - w_{\min}) . \tag{10}
$$

The maximum ($w_{\max}$) and minimum ($w_{\min}$) inertia weights are respectively equal to 0.9 and 0.4 for most cases, and $k_{\max}$ is the maximum generation number.

## Enhanced Comprehensive Learning Particle Swarm Optimization (ECLPSO)

Two techniques, namely perturbation-based exploitation and adaptive learning probabilities, are developed within the ECLPSO such that the optimal sizes and shapes of structures are designed. These concepts are described as follows.

*Perturbation-based Exploitation*

The perturbation-based exploitation term mainly improves the exploitation and accuracy of CLPSO algorithm. The scheme is considered if Eq. (11) and Eq. (12) is true. Then, the perturbation term is added to the velocity equation in Eq. (7).

$$\overline{P}_d - \underline{P}_d \leq \alpha\left(\overline{X}_d - \underline{X}_d\right) \tag{11}$$

$$\overline{P}_d - \underline{P}_d \leq \beta \tag{12}$$

$$V_{i,d} = w_{PbE}V_{i,d} + cr_d\left(pbest_{f_i,d} + \eta\left(\frac{\overline{P}_d - \underline{P}_d}{2} - pbest_{f_i,d}\right) - X_{i,d}\right) \tag{13}$$

where $\overline{P}_d$ and $\underline{P}_d$ are the upper and lower bounds of the personal best position on the *d*-th dimension, respectively; $\overline{X}_d$ and $\underline{X}_d$ are respectively the upper and lower bounds of the search space; $\alpha$ is a relative ratio equal to 0.01; $\beta$ is the small absolute bound with value equals to 2; $w_{PbE}$ is the inertia weight for the approach exploitation ($w_{PbE} = 0.5$ in this work); $\eta$ is the perturbation coefficient assisting the supports optimization process to capture the global optimum.

*Adaptive Learning Probabilities*

A new adaptive learning probabilities strategy is introduced to replace the original function in Eq. (9) as follows:

$$L_i = L_{\min} + (L_{\max} - L_{\min})\frac{\exp\left(\frac{10(K_i - 1)}{N - 1}\right) - 1}{\exp(10) - 1}, \tag{14}$$

$$L_{\max} = L_{\min} + 0.25 + 0.45\log_{(D+1)}(M_k - 1). \tag{15}$$

The value $K_i$ is a rank of personal best fitness value for the *i*-th particle, defined in an ascending order of the personal best fitness values. Moreover, $L_{\min}$ is fixed at 0.05, and $M_k$ is the number of dimensions as when both Eqs. (11) and (12) are complied.

**Illustrative Example**

The example considers the 15-bar planar truss structure in Fig. 1 subjected to the vertical load of $F = 10$ kips. The material properties employed were the modulus of elasticity of $10^4$ ksi, the material density of $0.1$ lb.in$^{-3}$, and the permissible tensile and compressive stresses of $\sigma_i^t = \sigma_c^t = 25$ ksi for all members $i \in \{1,\ldots,15\}$.

The sizing and shape optimization problem in Eq. (1) defined the design variables, namely the unknown member areas of $W_A = [A_1,\ldots,A_{15}]$ and some coordinates variations in *x*- and *y*-axes of $W_G = [x_2, x_3, x_6, x_7, y_2, y_3, y_4, y_6, y_7, y_8]$, where the geometry constraints of $x_2 = x_6$ and $x_3 = x_7$ were imposed. Moreover, the limits on both geometry variables are listed in Table 1.

The area variables are selected from the set available sections consisting of the discrete areas of A = {0.111, 0.141, 0.174, 0.220, 0.270, 0.287, 0.347, 0.440, 0.539, 0.954, 1.081, 1.174, 1.333, 1.488, 1.764, 2.142, 2.697, 2.800, 3.131, 3.565, 3.813, 4.805, 5.952, 6.572, 7.192, 8.525, 9.300, 10.850, 13.330, 14.290, 17.170, 19.180} in$^2$

**Table 1. Limit on geometry variables**

| Geometry variables: | $X_G,\min$ (in) | $X_G,\max$ (in) |
|---|---|---|
| $x_2$ | 100 | 140 |
| $x_3$ | 220 | 260 |
| $y_2$ | 100 | 140 |
| $y_3$ | 100 | 140 |
| $y_4$ | 50 | 90 |
| $y_6$ | -20 | 20 |
| $y_7$ | -20 | 20 |
| $y_8$ | 20 | 60 |
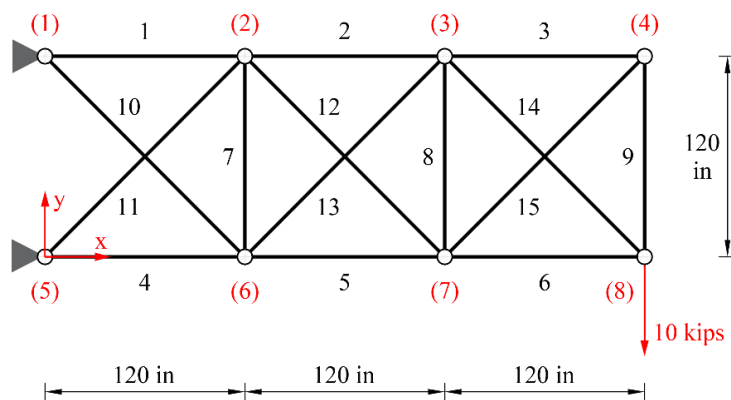


**Figure 1. Schematic of the planar 15-bar truss structure.**
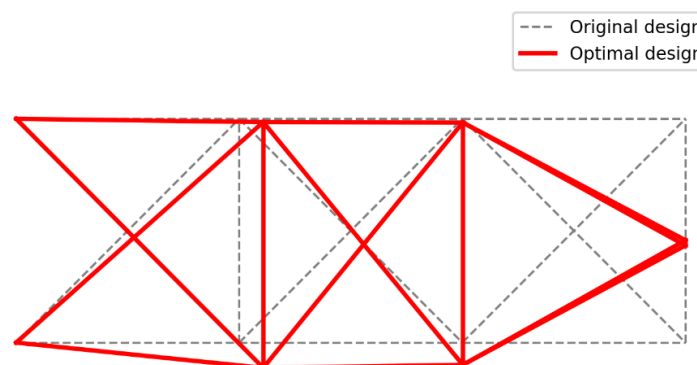


**Figure 2. Optimum layout of the 15-bar truss structure.**

The ECLPSO method adopted 30 particles with the maximum number of 800 iterations. All imposed constraints were fully complied. The resulting optimal member areas and shapes of the planar truss are depicted in Table 2, where those from various design methods are also summarized. It is evidenced that the optimal design weight value of 75.1552 lb given by the present ECLPSO achieved the most minimum (viz., the economical design with some 8% lighter than that from standard PSO scheme) as compared to all other benchmarks. The optimal

design shape is depicted in Fig. 2, and the plot of solution convergence in Fig. 3 presents monotonic variations of the design weights decreasing to the optimum over the increasing number of iterations.
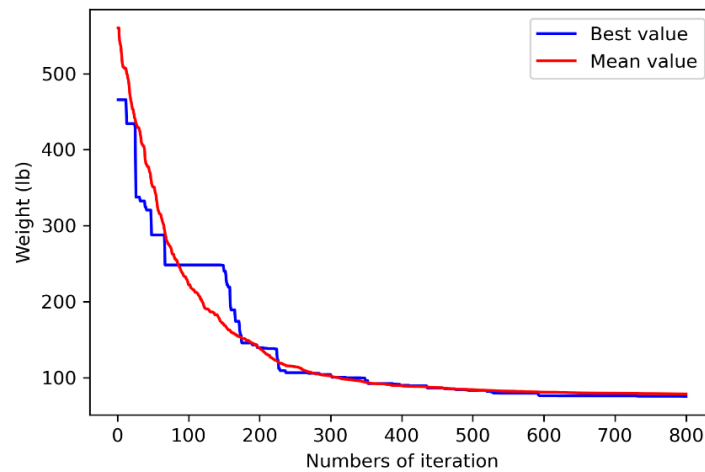


**Figure 3. Solution convergence in the ECLPSO process.**

**Table 2. Optimal area and shape solutions computed various design methods.**

| Design variable | Rahami et al. [4] | Tang et al. [5] | Gholizadh [6] | | ECLPSO (this study) | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | PSO | CPSO | | Stress (ksi) |
| A1 | 1.081 | 1.081 | 0.954 | 1.174 | 0.954 | 24.9120 |
| A2 | 0.539 | 0.539 | 1.081 | 0.539 | 0.539 | 23.8234 |
| A3 | 0.287 | 0.287 | 0.270 | 0.347 | 0.174 | 24.8164 |
| A4 | 0.954 | 0.954 | 1.081 | 0.954 | 0.954 | -23.4293 |
| A5 | 0.539 | 0.954 | 0.539 | 0.954 | 0.539 | -24.8863 |
| A6 | 0.141 | 0.220 | 0.287 | 0.141 | 0.287 | -24.2336 |
| A7 | 0.111 | 0.111 | 0.141 | 0.141 | 0.111 | -2.0217 |
| A8 | 0.111 | 0.111 | 0.111 | 0.111 | 0.141 | -4.3943 |
| A9 | 0.539 | 0.287 | 0.347 | 1.174 | 0.539 | 7.0051 |
| A10 | 0.440 | 0.220 | 0.440 | 0.141 | 0.347 | 24.9239 |
| A11 | 0.539 | 0.440 | 0.270 | 0.440 | 0.440 | -24.2596 |
| A12 | 0.270 | 0.440 | 0.111 | 0.440 | 0.270 | 20.0072 |
| A13 | 0.220 | 0.111 | 0.347 | 0.141 | 0.174 | -24.9816 |
| A14 | 0.141 | 0.220 | 0.440 | 0.141 | 0.287 | 24.9373 |
| A15 | 0.287 | 0.347 | 0.220 | 0.347 | 0.174 | -24.3701 |
| X2 | 101.5775 | 133.612 | 106.0521 | 102.2873 | 132.8913 | |
| X3 | 227.9112 | 234.752 | 239.0245 | 240.5050 | 240.2414 | |
| Y2 | 134.7986 | 100.449 | 130.3556 | 112.5840 | 118.2489 | |
| Y3 | 128.2206 | 104.738 | 114.2730 | 108.0428 | 118.0375 | |
| Y4 | 54.8630 | 73.762 | 51.9866 | 57.7952 | 52.0899 | |
| Y6 | -16.4484 | -10.067 | 1.8135 | 6.4299 | -13.5400 | |
| Y7 | -16.4484 | -1.339 | 9.1827 | 1.8006 | -11.9345 | |
| Y8 | 54.8572 | 50.402 | 46.9087 | 57.7987 | 54.9136 | |
| Weight (lb) | 76.6854 | 79.820 | 82.2344 | 77.6153 | 75.1552 | |

## Concluding Remarks

The paper has presented the simultaneous sizing and shape optimization of the in-plane truss structures under limited stress and serviceability constraints. The ECLPSO shows its efficient and robust optimizer that incorporates the two enhancing techniques, called the perturbation-based exploitation and the adaptive learning probabilities. The scheme by adjusting the ranking of personal best information advantageously overcomes the burdens associated with the premature solution convergence as would be experienced in standard PSO methods. The applications of the method have been illustrated through the design of modest-scale sizing and shape truss designs given in some available benchmarks. The optimal design solution can be achieved with the fast convergence to the optimum by processing the ECLPSO approach. The accuracy of the results computed has been well compared with those reported in the literatures.

## Acknowledgements

### References

[1] Kennedy, J. and R. Eberhart (1995). Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks.
[2] Liang, J. J., A. K. Qin, P. N. Suganthan and S. Baskar (2006). "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions." IEEE Transactions on Evolutionary Computation 10(3): 281-295.
[3] Yu, X. and X. Zhang (2014). "Enhanced comprehensive learning particle swarm optimization." Applied Mathematics and Computation 242: 265-276.
[4] Rahami, H., A. Kaveh and Y. Gholipour (2008). "Sizing, geometry and topology optimization of trusses via force method and genetic algorithm." Engineering Structures 30(9): 2360-2369.
[5] Tang, W., L. Tong and Y. Gu (2005). "Improved genetic algorithm for design optimization of truss structures with sizing, shape and topology variables." International Journal for Numerical Methods in Engineering 62(13): 1737-1762.
[6] Gholizadeh, S. (2013). "Layout optimization of truss structures by hybridizing cellular automata and particle swarm optimization." Computers & Structures 125: 86-99.
[7] Rajeev, S. and C. S. Krishnamoorthy (1992). "Discrete Optimization of Structures Using Genetic Algorithms." Journal of Structural Engineering 118(5): 1233-1250.
[8] Farshchin, M., C. V. Camp and M. Maniat (2016). "Optimal design of truss structures for size and shape with frequency constraints using a collaborative optimization strategy." Expert Systems with Applications 66: 203-218.
[9] Jawad, F. K. J., C. Ozturk, W. Dansheng, M. Mahmood, O. Al-Azzawi and A. Al-Jemely (2021). "Sizing and layout optimization of truss structures with artificial bee colony algorithm." Structures 30: 546-559.

# AUTHOR INDEX