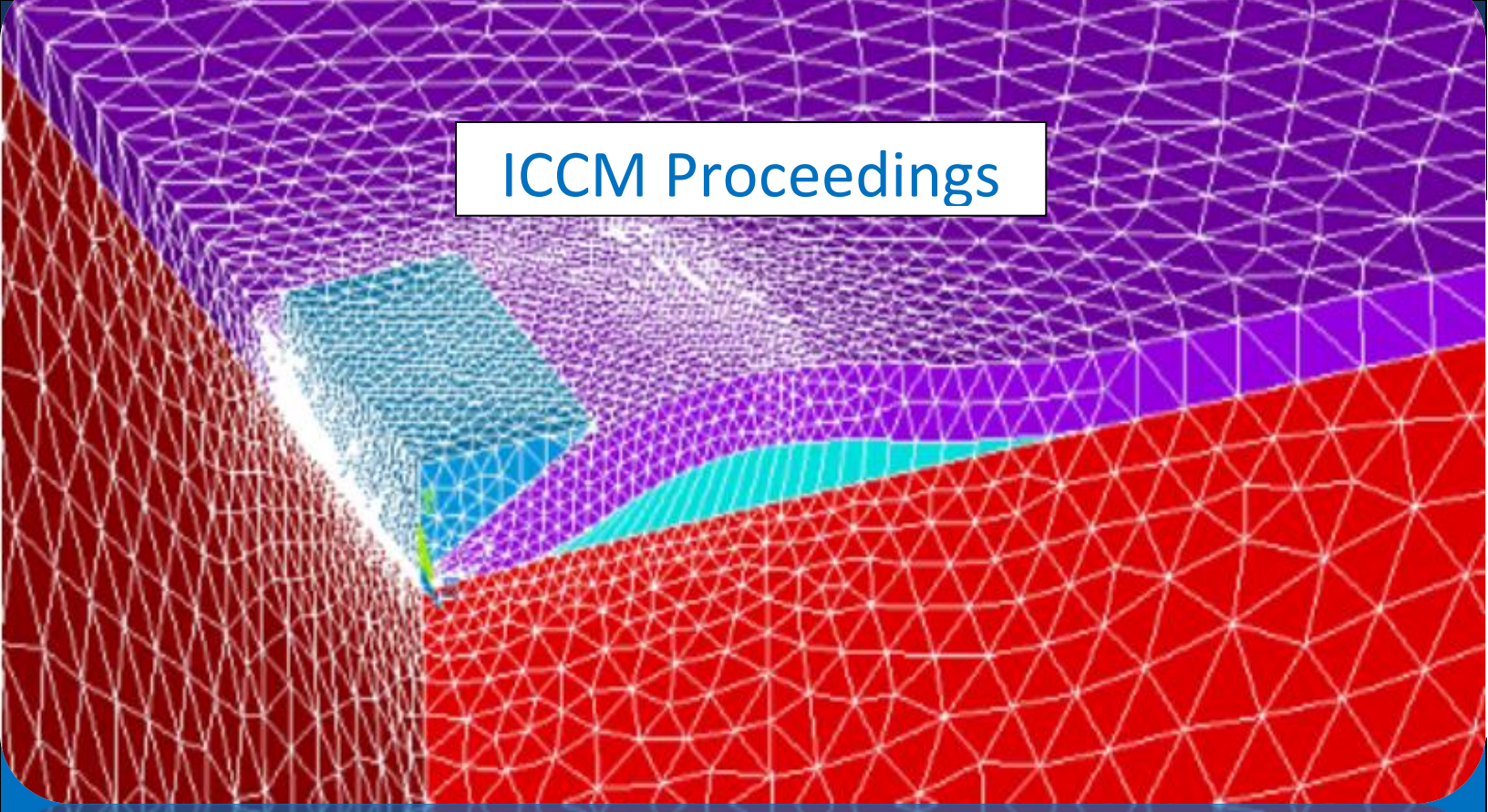


ICCM Proceedings

Proceedings of the International Conference on Computational Methods

(Vol.3, 2016)

7th ICCM2016, 1st-4th August 2016, Berkeley, CA, USA

Editors: **G. R. Liu and Shaofan Li**

ST ScienTech Publisher

ISSN 2374-3948 (online)

ICCM2016

Proceedings of the International Conference on Computational Methods (Vol.3, 2016)

7th ICCM2016, 1st-4th August 2016, Berkeley, CA, USA

Edited by

G. R. Liu
University of Cincinnati, USA

Shaofan Li
University of California at Berkeley, USA

Proceedings of the International Conference on Computational Methods, Vol.3, 2016

This volume contains full papers accepted by the 7th ICCM2016, 1st-4th August 2016, Berkeley, CA, USA

First Edition, August 2016

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

Papers in this Proceedings may be identically cited in the following manner: Author names, Paper title, *Proceedings at the 7th ICCM2016, 1st-4th August 2016, Berkeley, CA, USA*, Eds: G.R. Liu, Shaofan Li, ScienTech Publisher.

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

PREFACE

Dear Friends and Colleagues,

On behalf of the organizing committee and the co-chairs, we would like to welcome you to the 7th International Conference on Computational Methods (ICCM2016) at Berkeley, California, USA, between August 1st and 4th, 2016. The conference aims at to provide an international forum for scholars, researchers, industry practitioners, engineers, and graduate and undergraduate students to promote exchange and disseminate recent findings on both contemporary and traditional subjects in computational methods, numerical modeling and simulation, and their applications in science and engineering. It accommodates presentations on a wide range of topics to facilitate inter-disciplinary exchange of ideas in science, engineering and allied disciplines, and helps to foster collaborations.

Computational Modelling and Simulation are fundamental subjects in engineering and sciences. They can be applied to many of the primary engineering disciplines, including Aerospace, Bio-medical, Civil, Chemical, Mechanical, and Materials Engineering among others. Computational Modelling and Simulation cover a broad range of research areas, from conventional structural and mechanical designs, failure analysis, dynamic and vibration analysis, and fluid mechanics to cutting-edge computational mechanics, nano-micro mechanics, multiscale mechanics, coupled multi-physics problems and novel materials. This is reflected in the variety of fields featured in the conference topics.

The genesis of the ICCM series dates back to 2004, when the first ICCM2004 conference was held in Singapore founded and chaired by Professor Gui-Rong Liu, followed by ICCM2007 in Hiroshima, Japan, ICCM2010 in Zhangjiajie, China, ICCM2012 in Gold Coast, Australia, ICCM2014 in Cambridge, UK, and ICCM2015, Auckland, New Zealand. The present ICCM conference in Berkeley, USA encompasses over 330 oral presentations in 68 technical sessions, including 2 Plenary Talks, 6 Thematic Plenary Talks, and a number of Keynotes.

The ICCM conference is unique in the sense that it showcases the current developments and trends in the general topic of Computational Methods and their relationship to global priorities in science and engineering. The papers scheduled for presentation at ICCM address many urgent and grand challenges in modern engineering and sciences. All ICCM abstracts and full papers were peer-reviewed by independent reviewers. Selected papers may be invited to be developed into a full journal paper for publication in special issues of some international journals. These papers encompass a broad range of topics related to computational mechanics, including applied mechanics theory and formulation, computational methods and techniques, modelling techniques and procedures, nano and macro-mechanics of materials, dynamics, manufacturing, biomechanics, processing of advanced materials, welding and joining, surface engineering and other related processes.

We would like to express my gratitude for the contributions of all ICCM2016 participants and presenters at this international event. We gratefully acknowledge the contributions from the International Scientific Committee, Mini-Symposium Organizers, and the expert reviewers and volunteers for their efforts and assistance in the organization.

Finally, we would like to thank you for your contribution to the ICCM2016 conference. We are looking forward to your participation and continued engagement for the future ICCM conferences.

Professor Shaofan Li
Conference Chairman, ICCM2016
University of California at Berkeley, USA

Professor Gui-Rong Liu
Conference Chairman, ICCM2016
University of Cincinnati, USA

ORGANIZATION COMMITTEES

Conference Chairmen:
Shaofan Li (University of California, Berkeley, USA)
Gui-Rong Liu (University of Cincinnati, USA)

Co-Chairs:
Jeng-Tzong Chen (National Taiwan Ocean University, Taiwan)
Seiichi Koshizuka (University of Tokyo, Japan)
Raj Das (Auckland University, New Zealand)
Moubin Liu (Peking University, China)
Sau Cheong Fan (Nanyang Technological University, Singapore)
Paulo Pimenta (University of São Paulo, Brazil)
Yuantong Gu (Queensland University of Technology, Australia)
Jagdish Prakash (University of Botswana, Botswana)
Tarun Kant (Arid Forest Research Institute, India)
Xiao-Wei Gao (Dalian University of Technology, China)

Scientific Advisory Committee:

Umberto Alibrandi (Singapore)
Tinh-Quoc Bui (Viet Nam)
Song Cen (China)
Jeng-Tzong Chen (Taiwan)
Weiqiu Chen (China)
Zhen Chen (USA)
Raj Das (New Zealand)
Saucheong Fan (Singapore)
Xiqiao Feng (China)
Justin Fernandez (New Zealand)
Yuantong Gu (Australia)
Yu Huang (China)
Chao Jiang (China)
Hiroshi Kanayama (Japan)
Zhan Kang (China)
Tarun Kant (India)
Yoon-Young Kim (South Korea)
Adrian Koh (Singapore)
Seiichi Koshizuka (Japan)
Canh Le (Viet Nam)
Ik-Jin Lee (South Korea)
Quanbing Eric Li (China)
Wei Li (China)
Moubin Liu (China)
Ping Lu (USA)
Francesco Mammoliti (Italy)
Karol Miller (Australia)
Sundararajan Natarajan (India)
Francesco Noto (Italy)
Masao Ogino (Japan)
Marc Oudjene (France)
Joe Petrolito (Australia)
Paulo Pimenta (Brazil)
Jagdish Prakash (Botswana)
Ekkehard Ramm (Germany)
Alessandro Reali (Italy)
Daya Reddy (South Africa)
Erick Saavedra (Chile)
Lian Shen (USA)
Yuichi Tadano (Japan)
Zhaofeng Tian (Australia)
Cengiz Toklu (Turkey)
Patrizia Trovalusci (Italy)
Ken-ichi Tsubota (Japan)
Cheng Wang (China)
Yuesheng Wang (China)
Hengan Wu (China)
Yanling Wu (China)
Feng Xiao (Japan)
Jinyou Xiao (China)
Chao Xu (China)
Lixiang Yang (USA)
Jianyao Yao (China)
Hongling Ye (China)
Wenjing Ye (Hong Kong)
Jingjie Yeo (Singapore)
Sung-Kie Youn (South Korea)
Mengyan Zang (China)
Dia Zeidan (Jordan)
Chuanzeng Zhang (Germany)
Qing Zhang (China)
Xiong Zhang (China)
Kun Zhou (Singapore)

CONTENTS

Investigation of the Satellite Attitude Control System Performance Using as Actuator Reaction Wheels

Souza[1,2], L C G , Fonseca[2], J B S

1) Brasilia University – UnB
Área Especial de Indústria-A-UnB, 72444-240, Brasília - DF- Brasil.
E-mail: lcgs@unb.br
2) National Institute for Space Research- INPE
Avenida dos Astronautas, 1758, São Jose dos Campos – SP – Brasil
E-mail: luiz.souza@.inpe.br, jesusbravo85@yahoo.com.br

SUMMARY: Satellite Attitude Control System (SACS) pointing accuracy is dependent of its actuator and sensor performance and robustness, where the first design requirement can be associated with bandwidth while the second is related to the ability of SACS to keep performance in face of system parameters variation. One way to gain attitude control algorithms confidence is through the conjunction of computational methods and experimental design, which allows hardware and software interface test, besides decreasing the SACS design cost. As for maneuver pointing accuracy the reaction wheel (RW) is a key actuator, because its disturbance can influence the accuracy and stability of SACS. This paper studies how the dynamics and the control algorithm strategy of the reaction wheels with its respective DC motor can influence the performance and robustness of the SACS control in three axes. To do this one develops a 3D satellite simulator nonlinear model based on the State-Dependent Riccati Equation (SDRE) method taking into account the RW parameters. One compares the performance and robustness of the SACS where the RW is commanded by the SDRE control law with algorithm based on current and speed feedback compensation. Simulations of the computational methods developed have shown that the RW with speed feedback compensation has improved the SACS performance and robustness.

KEYWORDS: satellite attitude control, reaction wheel.

1. INTRODUCTION

The design of a SACS, that involves plant uncertainties and large angle maneuvers followed by stringent pointing control, may require new nonlinear attitude control techniques in order to have adequate stability, good performance and robustness. Experimental SACS design using nonlinear control techniques through prototypes is one way to increase confidence in the control algorithm. Experimental design has the important advantage of representing the satellite dynamics in a laboratory setting, from which it is possible to accomplish different simulations to evaluate the SACS [1]. However, the drawback of experimental testing is the difficulty of reproducing zero gravity and torque free space conditions. A Multi-objective approach [2] has been used to design a satellite controller with real codification. An investigated through experimental procedure has been used by Conti and Souza in [3] for simulator inertia parameters identification. An algorithm based on the least squares method to identify mass parameters of a rotating space vehicle during attitude maneuvers has been developed by Lee and Wertz in [4]. The H-infinity control technique was used in [5] to design robust control laws for a satellite composed of rigid and flexible panels. In the SDRE method, the nonlinear dynamics are brought to a time-invariant, linear-like structure containing state-dependent coefficients. Infinite-horizon LQR is then applied to the linear-like structure with the coefficient matrices being evaluated at the current operational point in the state space. The process is repeated in the next sampling periods therefore producing and controlling several state dependent linear models out of a non-linear one. The SDRE method was applied in [6] for controlling a nonlinear rotatory flexible beam system with two-degrees of freedom. However, it did not incorporate the SDRE filter (Kalman filter) as a state observer for the SDRE method, so that uncertainties could be accounted for in the filtering process. This paper studies how the dynamics and the control algorithm strategy of the reaction wheels with its respective DC motor can influence the performance and robustness of the SACS control in three axes. To do this one develops a 3D satellite simulator nonlinear model based on the State-Dependent Riccati Equation (SDRE) method taking into account the RW largest possible number of variables. One compares the performance and robustness of the SACS where the RW is commanded by the SDRE control law with algorithm based on current and speed feedback compensation. Simulations results have shown that the RW with speed feedback compensation has improved the SACS performance and robustness. As a result, the

simulations has shown the computational feasibility for real time implementation of the SDRE control method based on speed feedback algorithm in satellite´s onboard computer.

2. SDRE CONTROL METHODOLOGY

The Linear Quadratic Regulation (LQR) approach is well known and its theory has been extended for the synthesis of nonlinear control laws for nonlinear systems [7]. This is the case for satellite dynamics that are inherently nonlinear. A number of methodologies exist for the control design and synthesis of these highly nonlinear systems; these techniques include a large number of linear design methodologies such as Jacobean linearization and feedback linearization used in conjunction with gain scheduling [8]. Nonlinear design techniques have also been proposed including dynamic inversion and sliding mode control, recursive back stepping and adaptive control [9].
Compared to multi-objective optimization nonlinear control methods the SDRE method has the advantage of avoiding intensive interaction calculations, resulting in simpler control algorithms that are more appropriate for implementation on a satellite´s onboard computer.
The Nonlinear Regulator problem for a system represented in the State-Dependent Riccati Equation form with infinite horizon, can be formulated by minimizing the cost functional given by

$$J(x_0,u) = \frac{1}{2}\int_{t_0}^{\infty} (x^T Q(x)x + u^T R(x)u\ )dt \tag{1}$$

with the state $x \in \Re^n$ and control $u \in \Re^m$ subject to the nonlinear system constraints given by

$$\begin{aligned} \dot{x} &= f(x) + B(x)u \\ y &= C(x)x \\ x(0) &= x_0 \end{aligned} \tag{2}$$

where $B \in \Re^{nxm}$ and C are the system input and the output matrices, and $y \in \Re^s$ ($\Re^s$ is the dimension of the output vector of the system). The vector initial conditions is x(0), $Q(x) \in \Re^{nxn}$ and $R(x) \in \Re^{mxm}$ are the weight matrix semi defined positive and defined positive.
Applying a direct parameterization to transform the nonlinear system into State Dependent Coefficients (SDC) representation, the dynamic equations of the system with control can be write in the form

$$\dot{x} = A(x)x + B(x)u \tag{3}$$

with $f(x) = A(x)x$, where $A \in \Re^{nxn}$ is the state matrix. By and large $A(x)$ is not unique. In fact there are an infinite number of parameterizations for SDC representation. This is true provided there are at least two parameterizations for all $0 \le \alpha \le 1$ satisfying

$$\alpha A_1(x)x + (1-\alpha)A_2(x)x = \alpha f(x) + (1-\alpha)f(x) = f(x) \tag{4}$$

The choice of parameterizations to be made must be appropriate in accordance with the control system of interest. An important factor for this choice is not violating the controllability of the system, i.e., the matrix controllability state dependent $[B(x) + A(x)B(x) \dots A^{n-1}(x)B(x)]$ must be full rank.
The state-dependent algebraic Riccati equation (SDARE) can be obtained applying the conditions for optimality of the variational calculus. As a result, the Hamiltonian for the optimal control problem given by Equations (1) and (2) is given by

$$H(x,u,\lambda) = \frac{1}{2}(x^T Q(x)x + u^T R(x)u\ ) + \lambda^T (A(x)x + B(x)u) \tag{5}$$

where $\lambda \in \Re^n$ is the Lagrange multiplier.
Applying to the Eq.(5) the necessary conditions for the optimal control given by $\dot{x} = \frac{\partial H}{\partial \lambda}$, $\frac{\partial H}{\partial u} = 0$ and $\dot{\lambda} = -\frac{\partial H}{\partial x}$, one gets

$$\dot{\lambda} = -Q(x)x - \frac{1}{2}x^T \frac{\partial Q(x)}{\partial x}x - \frac{1}{2}u^T \frac{\partial R(x)}{\partial x}u - \left[\frac{\partial (A(x)x)}{\partial x}\right]^T \lambda - \left[\frac{\partial (B(x)u)}{\partial x}\right]^T \lambda \tag{6}$$

$$\dot{x} = A(x)x + B(x)u \tag{7}$$

$$0 = R(x)u + B(x)\lambda \tag{8}$$

Assuming the co-state in the form λ=P(x)x, which is dependent of the state, from Eq.(8) one obtains the feedback control law

$$u = -R^{-1}(x)B^T(x)P(x)x \tag{9}$$

Substituting this result into Eq. (7) one gets

$$\dot{x} = A(x)x - B(x)R^{-1}(x)B^T(x)P(x)x \tag{10}$$

To find the function P (x) one differentiates λ = P (x) with respect the time along the path from which one gets

$$\dot{\lambda} = \dot{P}(x)x + P(x)\dot{x} = \dot{P}(x)x + P(x)A(x)x - P(x)B(x)R^{-1}(x)B^T(x)P(x)x \tag{11}$$

Substituting Eq.(11) in the first necessary condition of optimal control (Eq.6) one obtains

$$\begin{aligned}&\dot{P}(x)x + P(x)A(x)x - P(x)B(x)R^{-1}(x)B^T(x)P(x)x \\ &= -Q(x)x - \frac{1}{2}x^T\frac{\partial Q(x)}{\partial x}x - \frac{1}{2}u^T\frac{\partial R(x)}{\partial x}u - \left[A(x) + \frac{\partial(A(x)x)}{\partial x}x\right]^T P(x)x \\ &- \left[\frac{\partial(B(x)u)}{\partial x}\right]^T P(x)x\end{aligned} \tag{12}$$

Arranging the terms more appropriately one has

$$\begin{aligned}&\dot{P}(x)x + \frac{1}{2}x^T\frac{\partial Q(x)}{\partial x}x + \frac{1}{2}u^T\frac{\partial R(x)}{\partial x}u + x^T\left[\frac{\partial\left(A(x)\right)}{\partial x}\right]^T P(x)x + \left[\frac{\partial(B(x)u)}{\partial x}\right]^T P(x) \\ &+ [P(x)A(x) + A^T(x)P(x) - P(x)B(x)R^{-1}(x)B^T(x)P(x) + Q(x)]x \\ &= 0\end{aligned} \tag{13}$$

In order to satisfy the equality of Eq.(13) one obtains two important relations. The first one is state-dependent algebraic Riccati equation (SDARE) which solution is P(x) given by

$$P(x)A(x) + A^T(x)P(x) - P(x)B(x)R^{-1}(x)B^T(x)P(x) + Q(x) = 0 \tag{14}$$

The second one is the necessary condition of optimality which must be satisfied and it is given by

$$\begin{aligned}&\dot{P}(x)x + \frac{1}{2}x^T\frac{\partial Q(x)}{\partial x}x + \frac{1}{2}u^T\frac{\partial R(x)}{\partial x}u + x^T\left[\frac{\partial\left(A(x)\right)}{\partial x}\right]^T P(x)x + \left[\frac{\partial(B(x)u)}{\partial x}\right]^T P(x)x \\ &= 0\end{aligned} \tag{15}$$

For the infinite time problem and considering the standard Linear Quadratic Regulator (LQR) problem, this is a condition that satisfies the optimality of the solution suboptimal control.
Finally, the nonlinear control law fed back by the states has the following form

$$u = -S(x)x\,,\ \ with\ \ S(x) = R^{-1}(x)B^T(x)P(x) \tag{16}$$

For some special cases, such as systems with little dependence on the state or with few state variables, Eq. (14) can be solved analytically. On the other hand, for more complex systems the numerical solution can be obtained using an adequate sampling rate. It is assumed that the parameterization of the coefficients dependent on the state is chosen so that the pair (A(x), B(x)) and (C(x) , A(x)) are in the linear sense for all x belonging to the neighborhood about the origin, point to point, stabilizable and detectable, respectively. Similar to the LQR method the SDRE nonlinear regulator need that all states are available to be feedback, otherwise one has to use the Kalman filter to estimates the data that is not measurable.

3. SIMULATOR MODEL

Figure 1 shows the INPE 3-D simulator which has a disk-shaped platform, supported on a plane with a spherical air bearing. Considering that the INPE 3-D simulator is not complete build, one assumes that there are three

reaction wheel configuration set capable to perform maneuver around the three axes and that there are three angular velocities sensor, like gyros. Apart from the difficulty of reproducing zero gravity and torque free condition, modeling a 3-D simulator, basically, follows the same step of modeling a rigid satellite with rotation in three axes free in space.

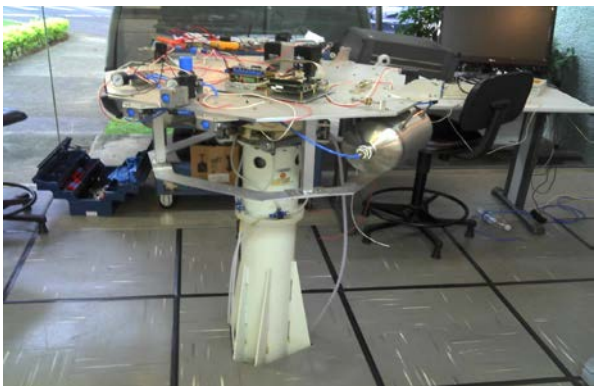


Figure 1- INPE 3-D simulator three reaction wheels.

The orientation of the platform is given by the body reference system F_b with respect to inertial reference system F_I considering the principal axes of inertia and using the Euler angles (θ_1, θ_2, θ_3) in the sequence 3-2-1, to guarantee that there is no singularity in the simulator attitude rotation. The equations of motions are obtained using Euler´s angular moment theorem given by

$$\dot{\vec{h}} = \vec{g} \tag{17}$$

where $\vec{g}$ and $\vec{h}$ are the torque and the angular moment of the system, which is given by

$$\vec{h} = I\vec{\omega} + I_w\left(\vec{\Omega} + \vec{\omega}\right) \tag{18}$$

where I = diag (I_{11}, I_{22}, I_{33}) is the system matrix inertia moment, $\vec{\omega}$ is the angular velocity of the platform, $\vec{I}_w$ = diag (I_{w1}, I_{w2}, I_{w3}) is the reaction wheel matrix inertia moment and Ω = (Ω_1, Ω_2, Ω_3) are the reaction wheel angular velocity.

Differentiating Eq. (18) and considering that the angular velocity of F_b is $\vec{\omega}$ and that the external torque is equal to zero, one has

$$\dot{\vec{h}} + \vec{\omega}^x\vec{h} = 0 \tag{19}$$

Substituting Eq.(18) into Eq.(19), the acceleration of the system is

$$\dot{\vec{\omega}} = \left(I + I_w\right)^{-1}\left[-\vec{\omega}^x\left(I + I_w\right)\vec{\omega} - \vec{\omega}^x I_w\vec{\Omega} - I_w\dot{\vec{\Omega}}\right] \tag{20}$$

The simulator attitude as function of the angular velocity is

$$\begin{pmatrix}\dot{\theta}_1\\ \dot{\theta}_2\\ \dot{\theta}_3\end{pmatrix} = \begin{pmatrix}0 & \sin\theta_3/\cos\theta_2 & \cos\theta_3/\cos\theta_2\\ 0 & \cos\theta_3 & -\sin\theta_3\\ 1 & \sin\theta_3\sin\theta_2/\cos\theta_2 & \cos\theta_3\sin\theta_2/\cos\theta_2\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\\ \omega_3\end{pmatrix} \tag{21}$$

Here one simulates the angular maneuver which represents the fine pointing mode control where the reaction wheel is the best actuator, so the state's x are $(\theta_1\ \theta_2\ \theta_3\ \omega_1\ \omega_2\ \omega_3)^T$ and the control are due to the reaction wheel velocities $(\dot{\Omega}_1\ \dot{\Omega}_2\ \dot{\Omega}_3)^T$ One knows that the reaction wheel generates internal torques and the attitude control is performed by exchange of angular moment between the reaction wheel and the satellite. From the union of the Equations (20) and (21) one obtains the matrices A(x), B(x) and C(x) in state space form, which represents the satellite simulator nonlinear plant (yellow block) as showed in Figure 5. It should be stressed, that a great advantage of the SDRE method is that it is not necessary to linearize the system. The SDRE method can deal with the nonlinearities of the system, which here come from the product of the angular velocities of the platform

and reaction wheel (Eq.(20)) and with the trigonometric function of Eq.(21) associated with the angular position that represent the attitude of the system.

4. REACTION WHEEL DYNAMICS

In the sequel one derives the reaction wheel dynamics which is triggered by a DC motor as show in Figure 2. For simplicity, here one ignores the losses due to the transformation of electrical energy into mechanical. Therefore, the electrical power is equal to the mechanical power given by

$$V(t)i(t) = T(t)w(t) \tag{22}$$

$$V(t) = Ri(t) + L\frac{di(t)}{dt} + e(t) \tag{23}$$

$$T(t) = Bw(t) + j\frac{dw(t)}{dt} \tag{24}$$

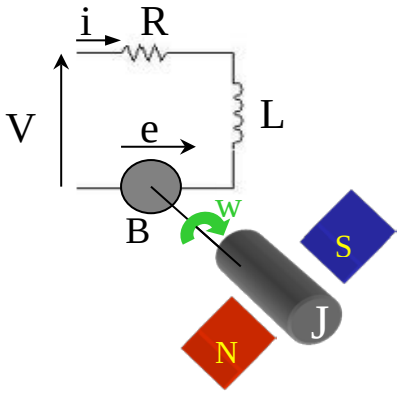


Figure 2- DC Motor dynamics representation.

Where R is the electrical resistance of the motor, L is the inductance of the motor, B is the viscous friction of the motor, J is the moment of inertia of the reaction wheel, w is the angular velocity of the wheel, i is the electric current of the motor, V is the electrical voltage at the motor terminals and e is the voltage generated due to movement of the motor rotor within a magnetic flux.
For a permanent magnet motor, the following relationship given below is valid

$$e(t) = K_e w(t) \tag{25}$$

where K_e is associated with the motor tension. One also knows that in an engine of this type the relationship between torque and current is given by

$$T(t) = K_t i(t) \tag{26}$$

where Kt is a constant associated with the motor torque. Substituting Eq. (25) into Eq. (23) one has

$$V(t) = Ri(t) + L\frac{di(t)}{dt} + k_e w(t) \tag{27}$$

Substituting Equation (26) into Equation (24) one has

$$K_t i(t) = Bw(t) + j\frac{dw(t)}{dt} \tag{28}$$

Arranging the Equations (27) and (28) with the first order terms in the left hand side and the zero order terms in the right hand one has

$$L\frac{di}{dt} = V - Ri - K_e w \tag{29}$$

$$j\frac{dw}{dt} = K_t\, i - Bw \tag{30}$$

Putting Equations (29) and (30) in the Matlab/Simulink form, one has the block diagram given by Figures 3 and 4, respectively.

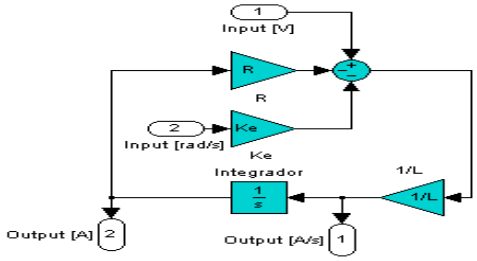


Figure 3 - block diagram of Eq. (29)

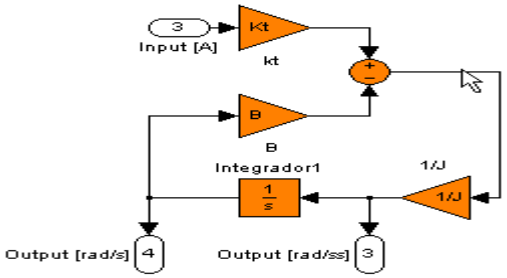


Figure 4 - block diagram of Eq. (30)

Joining the two block diagrams of the Figures 3 and 4 above, one gets the complete block diagram of the entire reaction wheel (blue bock) as showed in Figure 5.

5. SIMULATIONS RESULTS

Now one has the Simulink/Matlab model for the Satellite Simulator with Nonlinear Plant (yellow block), the control system using the SDRE Controller (green block) and the reaction wheel dynamics with velocity or current feedback (blue block), so grouping them one gets the Complete Simulator System , showed in Figure 5. In such system one has as input the reference angles to where the SDRE controller must maneuver the satellite and as output the angles and the angular velocity of the satellite. For simplicity the external torque is zero.

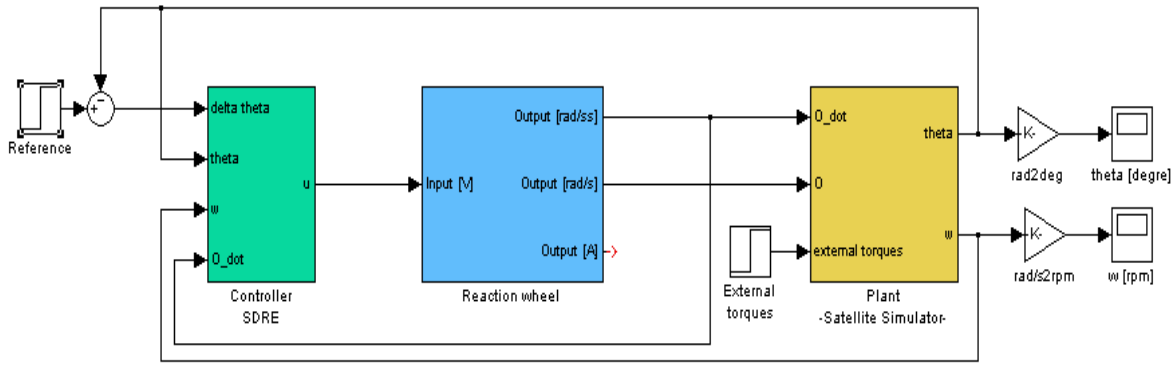


Figure 5 – Entire Simulator with plant of the satellite, SDRE Controller and the Reaction Wheel dynamics.

The satellite simulator model is inertia moment depend, so here one uses $I_{11} = I_{22}$=1185.0; I_{33}=1136.0 and for the DC motors parameters R = 7,3 , L = 2,5 , B = 0,00494, J = 2.0, Kt = 0,05, Ke = 0,05. The SDRE controller must maneuver the satellite from initial angles zeroes to final angles are Theta1 = 10°, Theta2 = 5°, Theta3 = - 5°. The control system has used three different reaction wheel configurations. In the first one the reaction wheel has no feedback, in the second and thirty configurations one employs velocity feedback and current feedback, as showed in Figure 6, in order to evaluate the reaction wheel performance for the three cases.

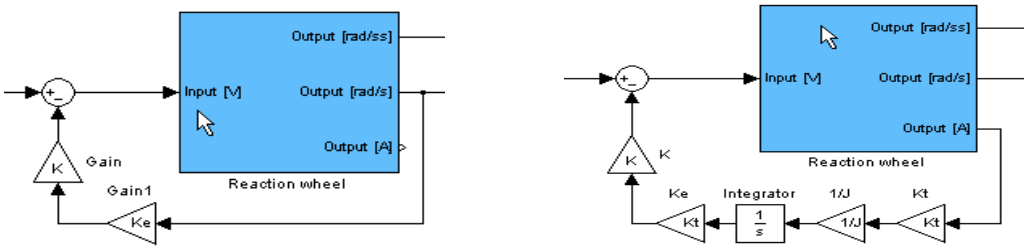


Figure 6 – reaction wheel block diagram with velocity and current feedback

The first simulation is the design of the SDRE controller where the reaction wheel loop has no feedback. The SDRE controller gain S(x) depend on matrices of the simulator model A(x), B(x) and C(x), see [16] for details, and of the tuning matrices Q and R which one assumes the values Q = diag(1, 1, 1,100, 100, 100) and R (0.001, 0.001,0.011). Once one has design the SDRE controller the next step is to design the reaction wheel control loop which can have velocity or current feedback. After some try and error one get the gain K = 50 to feedback with velocity or current the reaction wheel. The performance of the entire SACS for the previously angular maneuver is showed in Figures 7, 8 and 9 for each axis angles Theta1, 2 and 3, without feedback and with feedback of velocity and current, respectively

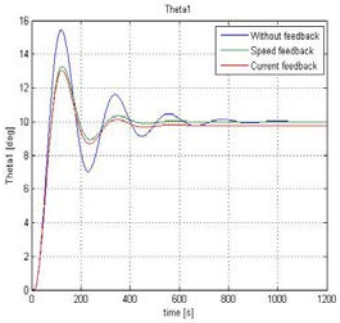


Figure 7 – Attitude angle Theta 1

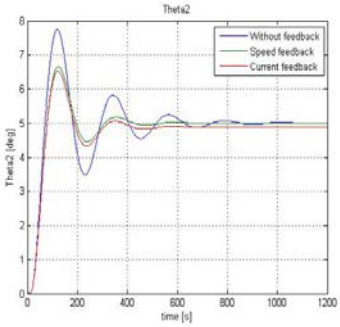


Figure 8 – Attitude angle Theta 2

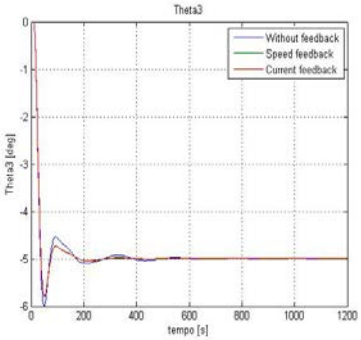


Figure 9 – Attitude angle Theta 3

In order to investigate the reaction wheel performance one increases its gain to K= 250 and perform the same previously angular maneuver. Figures 10, 11 and 12 show the SACS action for each angle Theta1, 2 and 3, without feedback and with feedback of velocity and current, respectively

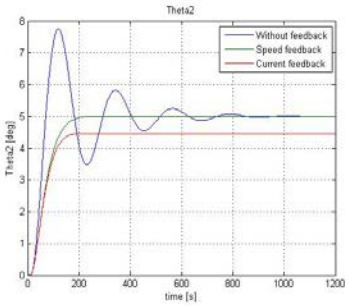


Figure 10 – Attitude angle Theta 1

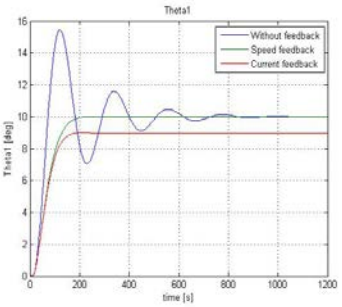


Figure 11 – Attitude angle Theta 2

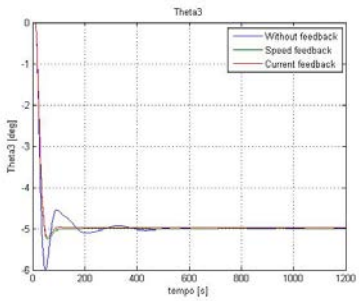


Figure 12 – Attitude angle Theta 3

As one observes the SACS performance has been improved when the reaction wheel gain increases, so one increases it a bit more to K= 500 and one performs the same angular maneuver. Figures 13, 14 and 15 show that the SACS performance to control the angles Theta1, 2 and 3 has been deteriorated both with velocity and current feedback.

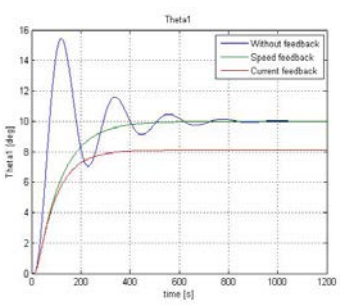


Figure 13 – Attitude angle Theta 1

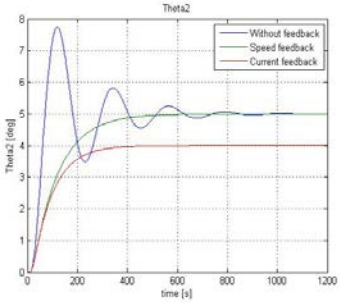


Figure 14 – Attitude angle Theta 2

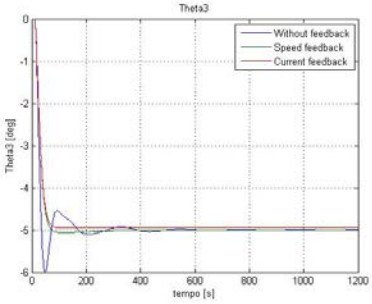


Figure 15 – Attitude angle Theta 3

4. CONCLUSIONS

From the first simulation one observes that the SACS with reaction wheel loop using the gain K=50 has better performance than the SACS with reaction wheel without both velocity or current feedback, since there is an improvement in the level of the overshoot and the maneuver has been done faster, although one observes that there is a stead state error when using the current feedback. So one can conclude that increasing the reaction wheel gain the velocity feedback has better performance that current feedback. In order to investigate this and to improve the maneuver one has increase the reaction wheel loop gain to K = 250, in that case one notices that stead state error introduced by the current feedback increase, although the overshoot has decreased. As a result, one could conclude that increasing the reaction wheel gain the SACS performance using the velocity feedback in the reaction wheel loop could be better than current. But this is not true since when one increase a bit more the gain to K = 500, the maneuver using the reaction wheel with velocity feedback has been performed in more time than the maneuver using K= 250. This just shows that there exists a limit value for the reaction wheel gain which possible is around 250. Besides, it is important to say that the reaction wheel gain is as function of its axis since the inertia moments are different for each axis. Finally, one observes that there are two ways to improve the SACS design, the first one could be using a kind of optimal control technique to obtain the reaction wheel gains, and the other one is including a Kalman filter to estimate the possible measurements that eventually are not available to be feedback, since here one has consider that all states are available to be feedback into the control loop.

5. REFERENCES

[1] Hall, C.D., Tsiotras and Shen, H. “Tracking Rigid Body Motion Using Thrusters and Momentum Wheels”. *Journal of the Astronautical Sciences.* (3), 2002, pp. 13-20.

[2] Mainenti-Lopes, I. , Souza, L. C. G., Sousa, F. L., Cuco, A. P. C. “ Multi-objective Generalized Extremal Optimization with real codification and its application in satellite attitude control”, Proceedings of 19th International Congress of Mechanical Engineering - COBEM, Gramado, 2009, Brasil.

[3] Conti, G T and Souza, L C G. “Satellite Attitude System Simulator”, *Journal of Sound and Vibration*,(15), 2008, pp. 392-395.

[4] Lee, A. Y. and Wertz, J. A. “In-flight estimation of the Cassini Spacecraft inertia tensor”. Journal Spacecraft, (39),1, 2002, pp.153-155.

[5] Pinheiro, E. R.; Souza, L. C. G.. Design of the Microsatellite Attitude Control System Using the Mixed H_2/H_∞ Method via LMI Optimization. Mathematical Problems in Engineering, v. 2013, p. 1-8, 2013.

[6] Bigot, P., Souza, L. C. (2014). Investigation of the State Dependent Riccati Equation (SDRE) adaptive control advantages for controlling non-linear systems as a flexible rotatory beam. *International journal of systems applications, engineering and development.* (Vol. 8, pp 92-99).

[7] Menon P.K. ; Lam T. ; Crawford L. S.; Cheng V. H L, “Real Time Computational Methods for SDRE Non-linear Control of Missiles”. *American Control Conference*, May 8-10, 2002, Anchorage, AK, USA.

[8] Shamma, J. S. and Athens, M. Analysis of gain scheduled control for nonlinear plants. IEEE Trans. on Auto. Control, 35(8):898{907, 1990.

[9] Slotine, J. J. E.. Applied nonlinear control. Prentice Hall, Englewood Clips, New Jersey, 1991.

Non Linear Strain Integral Damping (S.I.D.)

†U. Tornar[1]

[1]Department of Research and Development, PSA Peugeot Citroen, France.

†Presenting author: ugo.tornar@orange.fr

†Corresponding author: ugo.tornar@orange.fr

Abstract

Forces are generally defined in physics as functions of position (Newton: gravity) or velocity (Laplace: magnetic force on a moving electric charge). Damping forces are little known even today and represent one of the most intriguing subjects of physics. Maxwell elements and fractional derivatives are used to modelize time domain natural hysteretic damping. The resulting models are comparatively complicated and have a limited domain of validity especially when strong non-linearity is involved. The mathematical model we use is based on the introduction of a new state variable and is particularly suitable in the non-linear vibration case. S.I.D. (Strain Integral Damping: see ref. [2]) is a very suitable mean to modelize natural hysteretic damping in the time domain and for nonlinear rubber elements in particular. In the present paper the stress is on modelling of nonlinear elements. The effectiveness of SID is shown by an example concerning a strongly non-linear spring. A ‘’Scilab’’ script is provided to better explain.

Keywords: Natural hysteretic damping, Computational model, Vibration, Engine, Driveline, Startup

Introduction

Natural damping is only seldom viscous. Natural hysteretic damping is much more common and can be described as follows in the frequency domain.
If:
$M = Mass\ Matrix \quad K = Stiffness\ Matrix \quad f = force\ vector \quad \xi = displacement\ vector$
$\omega = angular\ frequency$

$$\left[-\omega^2 M + K \cdot \left(1 + j \cdot \mathrm{tg}(\varphi)\right)\right]\xi = f \qquad (1)$$

Where an imaginary part of the stiffness matrix is introduced (tg(φ)). We shall call this I.S.D. (Imaginary Stiffness Damping) in the following.
Such a formulation is much used in the frequency domain because it is simple and practical to use and not because there is a real physical theory behind it. S.I.D. (Strain Integral Damping) wants to be as simple and practical to use for hysteretic damping modeling in the time domain. The formulation of SID will be now briefly recalled. See references [2] and [3].

1 SID Formulation

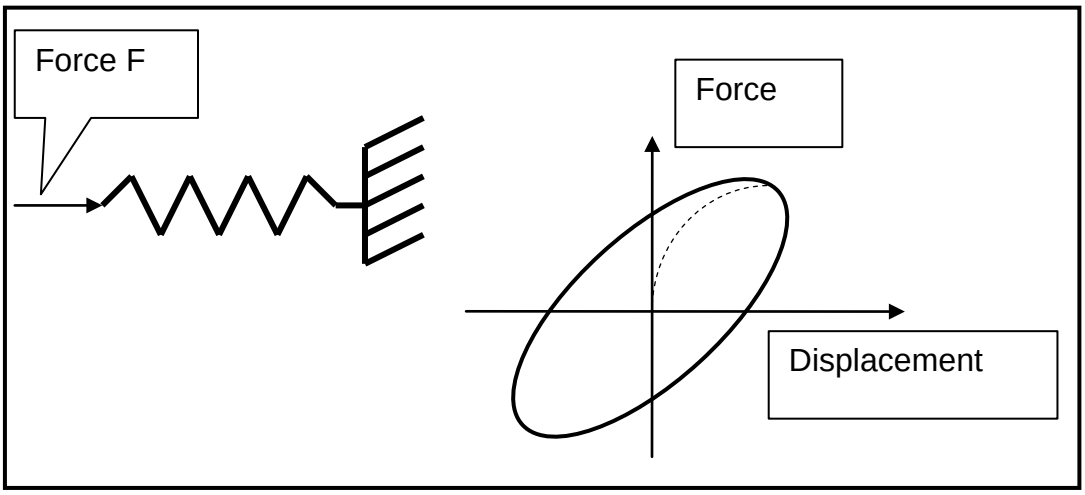


Figure 1. Hysteresis of a spring-damper

Let us consider the spring-damper of FIG. 1. If $x(t)$ is the displacement at time ‘’t’’ and we apply a sinusoidal force we shall obtain:

$$x(t) = X\ \cos(\omega t) \tag{2}$$

Velocity ‘’v’’ and acceleration ‘’a’’:

$$v(t) = \frac{dx}{dt} = -\omega\ X\ \sin(\omega t) \tag{3}$$

$$a(t) = \frac{d^2x}{dt^2} = -\omega^2\ X\ \cos(\omega t) \tag{4}$$

The applied force will be:

$$f(t) = F\ \cos(\omega t + \varphi) \tag{5}$$

‘’F’’ being the force amplitude. We can rewrite:

$$f(t) = F\cos(\omega t)\cos(\varphi) - F\ \sin(\omega t)\sin(\varphi) \tag{6}$$

Following equation (1) the springer-damper stiffness ‘’k’’ is defined by:

$$F\ \cos(\varphi) = kX \tag{7}$$

We can then write equation (6) in the form:

$$f(t) = kX\ \cos(\omega t) - k\ \mathrm{tg}(\varphi)\ X\ \sin(\omega t) \tag{8}$$

By substituting equations (2) and (3) in equation (8) we obtain:

$$f(t) = k\left(x(t) + \mathrm{tg}(\varphi)\frac{v(t)}{\omega}\right) \tag{9}$$

Where we have introduced the same $\mathrm{tg}(\varphi)$ factor of eq. (1).
Now we must express the ''ω'' of eq. (9) as a function of state variables only. We may think of expressing the ''1/ω'' factor of eq. (9) as the ratio:

$$\frac{1}{\omega} = \left|\frac{x(t)}{a(t)}\right|^{1/2} \tag{10}$$

But as it is shown in reference [1], forces cannot in general be expressed as functions of the accelerations and this leads us to define a new state variable which is the solution of the differential eq.:

$$\frac{dy}{dt} = -\omega_1 \cdot y + x(t) \tag{11}$$

The solution is:

$$y(t) = \int_0^t e^{-\omega_1(t-\tau)}x(\tau)d\tau \tag{12}$$

The constant ''ω_1'' is introduced to define as ''remote past'' all events for which:

$$(t-\tau) \gg 1/\omega_1 \tag{13}$$

Such events will have negligible effect on ''y'' (strain integral) and, as a consequence, on the damping force. We must remark that if ''ω_1'' is zero, ''y'' goes to infinity for all x(t) whose average is not zero (spring preloading). This of course wouldn't be physical. So ''ω_1'' can be seen as a high pass filter parameter: it has the same physical dimensions as a frequency and it must be set well lower than the frequencies of interest but it must not be negligible in comparison with the frequencies of interest to avoid ''y'' to go to infinity. We can better understand this by writing eq. (11) in the frequency domain:

$$\frac{Y}{X} = \frac{1}{\omega_1 + j\omega} \tag{14}$$

Where X and Y are the complex amplitudes of ''x'' and ''y'' respectively. We can see from this formula that if ω is an angular frequency of interest, it must be $\omega \gg \omega_1$ for ''y'' to be close to the integral of ''x''. $\omega_1 = 1\%(\omega)$ is a possible value.
We can then assume:

$$\frac{1}{\omega} \cong \left|\frac{y(t)}{v(t)}\right|^{1/2} \tag{15}$$

By substituting eq. (15) into eq. (9) we easily obtain:

$$f(t) = k\left(x + tg(\varphi)\ \mathrm{sign}\big(v(t)\big)\ |y(t)\ v(t)|^{1/2}\right) \tag{16}$$

We remark that the term $|y(t)\ v(t)|^{1/2}$ has the physical dimensions of a displacement but is ''phased'' like a velocity.

We assume as initial condition for ''y'':

$$(t = 0) \Rightarrow (y = 0) \tag{17}$$

We can easily see that, with this initial condition, the cycle starts at the origin like the dotted curve shown in Fig. 1.
Work experience has shown that the introduction of factor ''ω_1'' in eq. (11) is not enough to avoid that ''y'' goes to infinity. This problem of course can only exist in case of spring (engine mount) preloading. The problem is easily solved by the introduction of a moving average in equation (11):

$$\frac{dy}{dt} = -\omega_1 \cdot y + \left(x(t) - \overline{z(t)}\right) \tag{18}$$

The moving average $\overline{z(t)}$ is defined as the solution of the following differential equation:

$$\frac{dz}{dt} = -\omega_2 \cdot z + x \tag{19}$$

The solution is:

$$z(t) = \int_0^t e^{-\omega_2 (t-\tau)}\ x(\tau)\ d\tau \tag{20}$$

And the corresponding weighted moving average:

$$\overline{z(t)} = \int_0^t e^{-\omega_2 (t-\tau)}\ x(\tau)\ d\tau \cdot \omega_2/(1 - e^{-\omega_2 \mathrm{X} t}) \tag{21}$$

Where:

$$\omega_2/(1 - e^{-\omega_2 \mathrm{X} t}) \tag{22}$$

Is the normalization factor. We can see from eq. (21) that such an average has the important property that all events of the ''past'' that happened at time τ such that:

$$(t - \tau) \gg 1/\omega_2 \tag{23}$$

Are ''squeezed'' by the weighting factor:

$$e^{-\omega_2(t-\tau)} \ll 1 \tag{24}$$

So that only the most ‘’recent’’ events are really included in the average.
The fact that factor (22) goes to infinity when t=0 is generally avoided by the substitution:

$$\omega_2 \;/\; \max(0.0001, 1 - e^{-\omega_2 \times t}) \tag{25}$$

In practice we often assume:

$$\omega_2 = \omega_1 \tag{26}$$

But this is not a general rule: $\boldsymbol{\omega_2}$ depends on the speed by which the “quasi static” preloading varies and must be set accordingly. Sometimes static preloads are not really constant. For example the engine torque varies depending on how much the driver presses on the accelerator and ‘’static’’ loads on the mounts will vary accordingly. In the driveline model of section 4.2 for example we had:

$\omega_1 = 0.01$; $\omega_2 = 8$

Because of the quickly increasing engine torque due to quickly mounting RPM. The RPM rose quickly because of simulation of a steep sloping down startup of the vehicle. Consider for example the famous “Gross Glochner” very steep descent in Austria.

2 Nonlinearity

SID is most useful in nonlinear problems. To introduce non linearity we only need modifying eq. (16) as follows:

$$f(t) = \mathrm{spl}(x) + k_s \cdot \mathrm{tg}(\varphi)\ \mathrm{sign}\big(v(t)\big)\ |y(t)\ v(t)|^{1/2} \tag{27}$$

Where ‘’spl’’ is a spline representing the non-linear spring and k_S is the secant stiffness (very seldom the tangent stiffness as explained in reference [3]). In references [2] and [3] the user is provided with useful advice and cautions concerning the practical use of SID. For example the “boxcar effect” [4] needs sometimes being considered in analyzing results obtained by time step integration. It must be remarked that assuming the secant stiffness (load divided by displacement) to drive the damping phenomenon corresponds to assuming damping forces to be proportional to the loads acting on the nonlinear element. In the author’s experience such an assumption is often closer to reality than assuming damping forces to be proportional to the differential stiffness.

3 Frequency independence of SID nonlinear cycles (Numerical example)

We are now going to present with a numerical example concerning the property of a SID spring hysteresis cycle to remain the same whatever the frequency of the imposed displacement. Such a property is a feature of natural damping as it is observed in physical reality. SID has the remarkable power of insuring that such a property is verified also in the case of calculation of a strongly nonlinear spring. The Scilab script in the appendix was used to perform the calculations. By setting the imposed displacement frequency at 20, 40 and 60 Hz we are now going to see that the cycle doesn’t change. We can see that the cycle in figure (5) is practically identical to that in figure (3) although the frequency is twice and that the cycle in figure (7) is again practically identical to that in figure (3) although the frequency is 3 times higher.

3.1 Calculation at 20 Hertz.

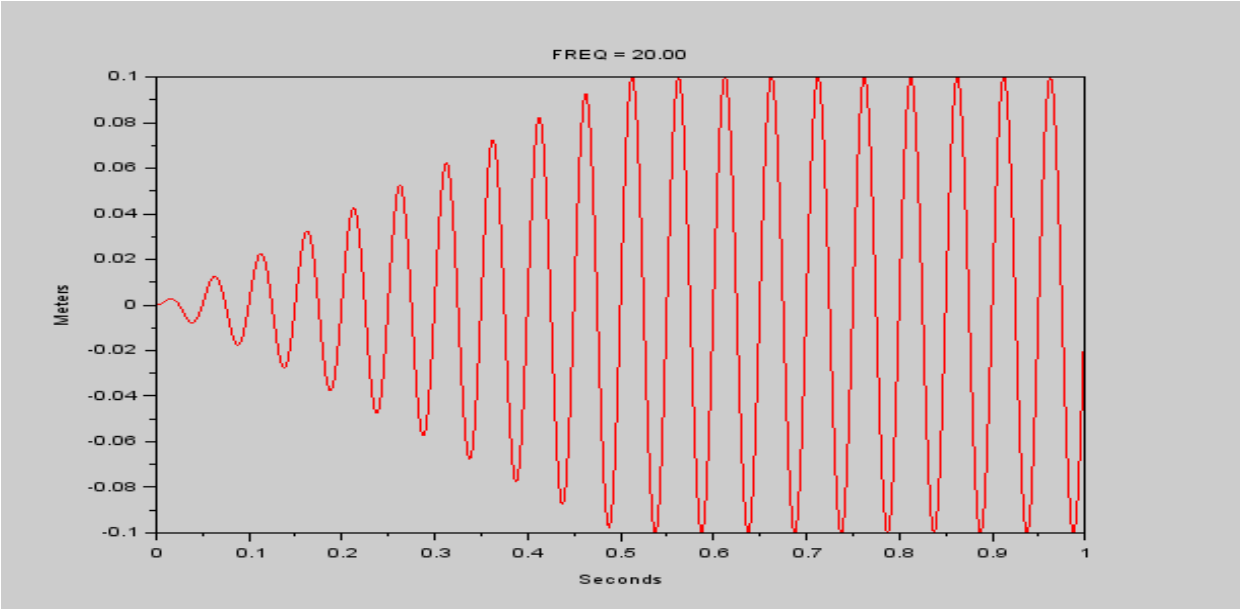


Figure 2. Displacement

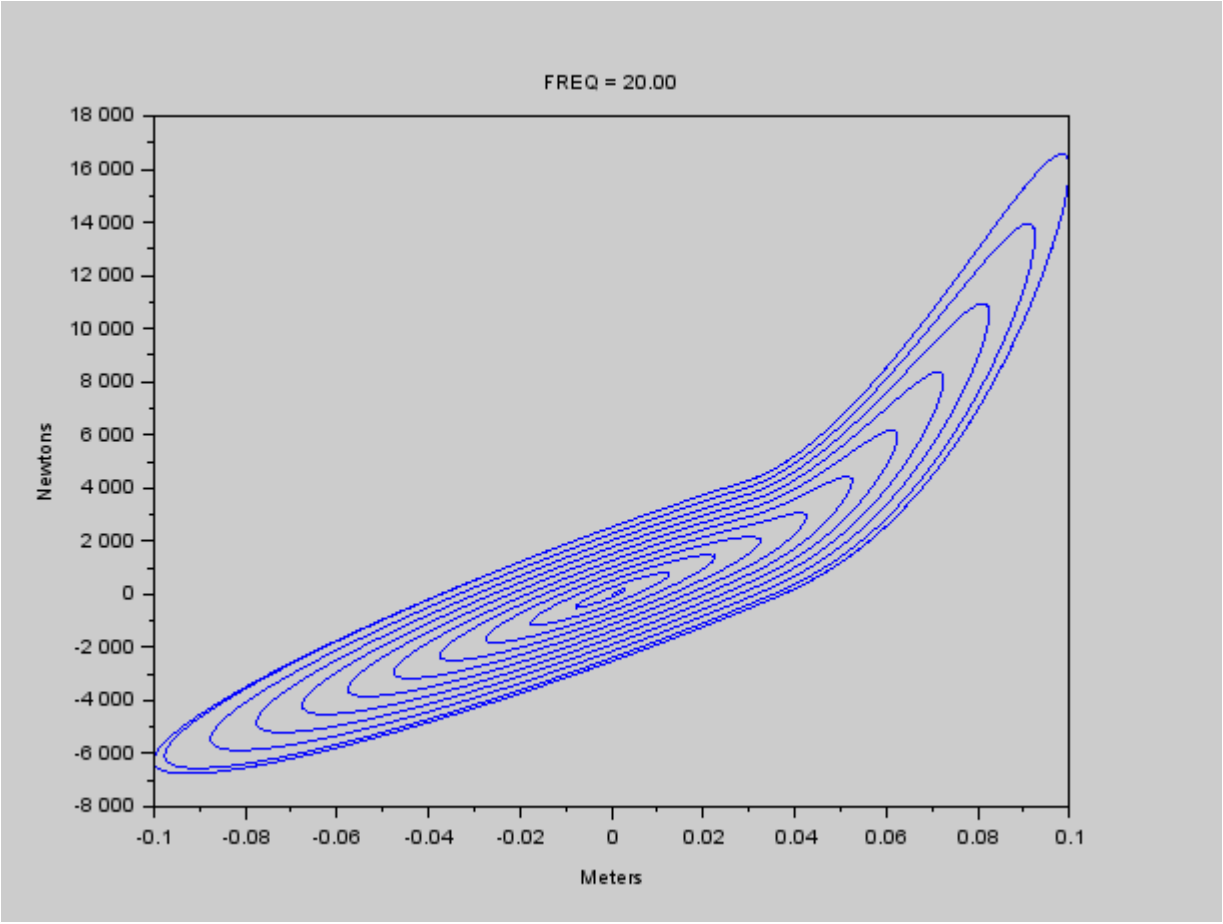


Figure 3. Cycle

3.2 Calculation at 40 Hertz.

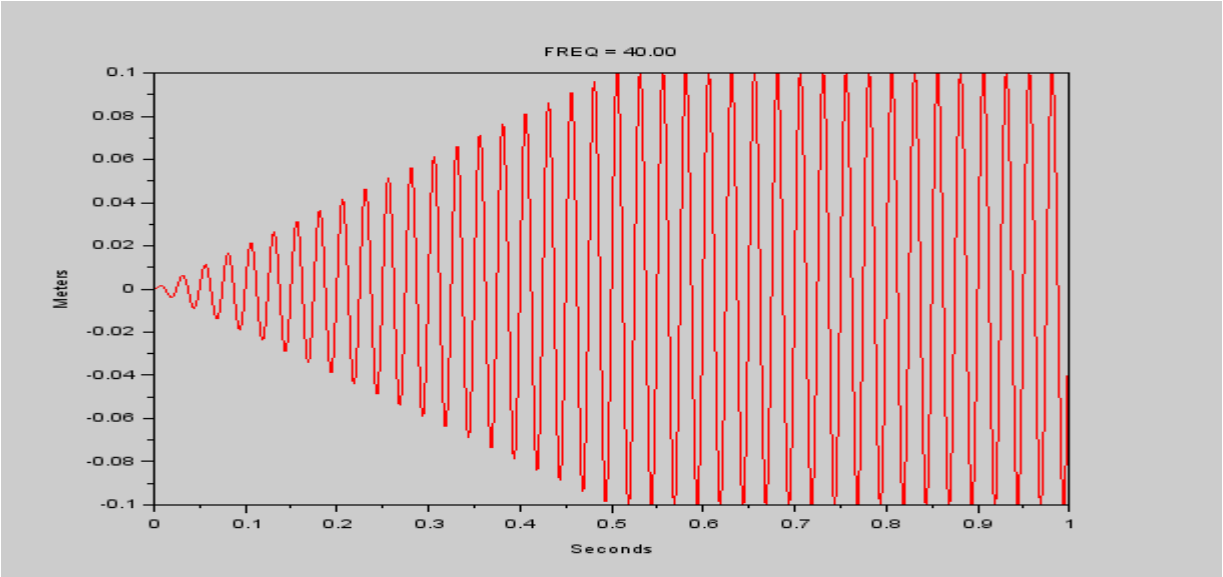


Figure 4. Displacement

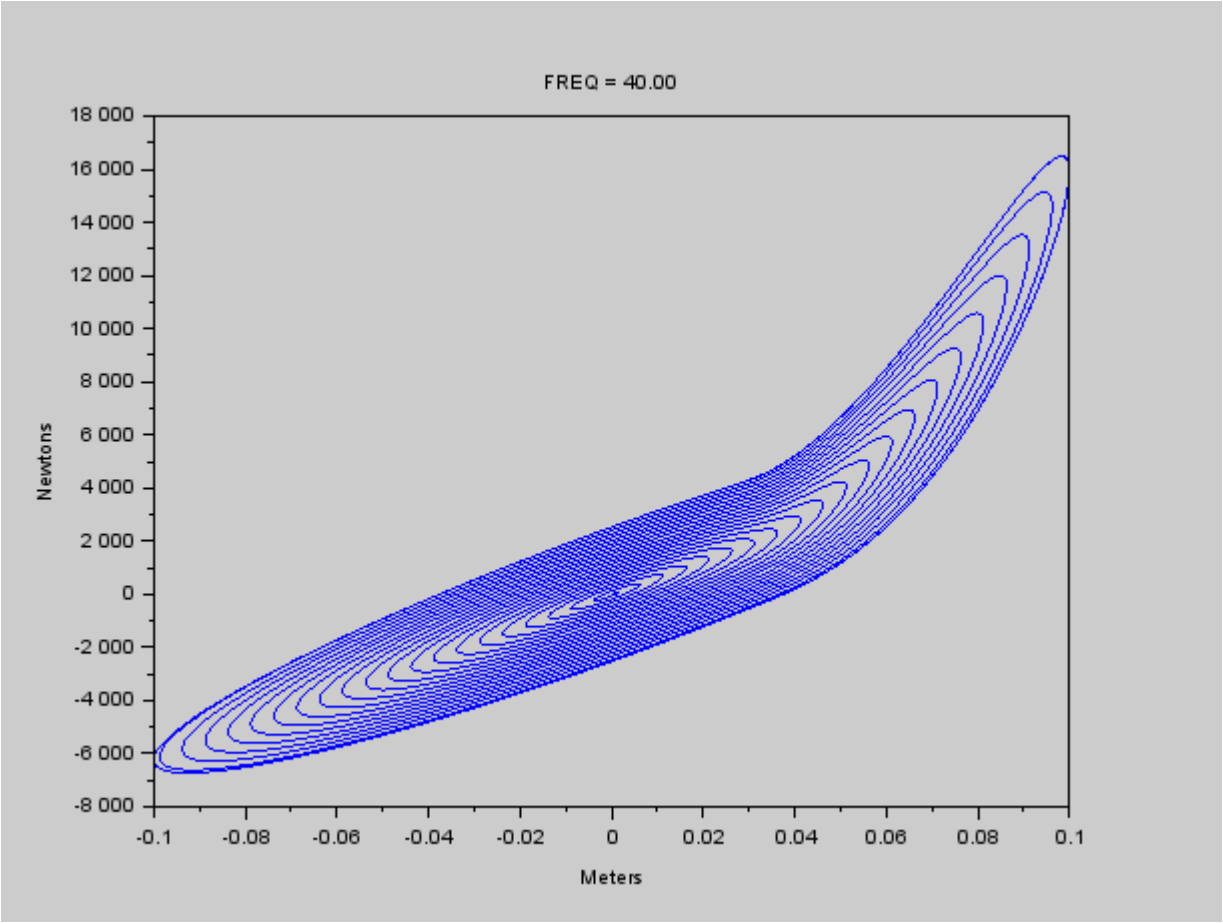


Figure 5. Cycle

3.3 Calculation at 60 Hertz.

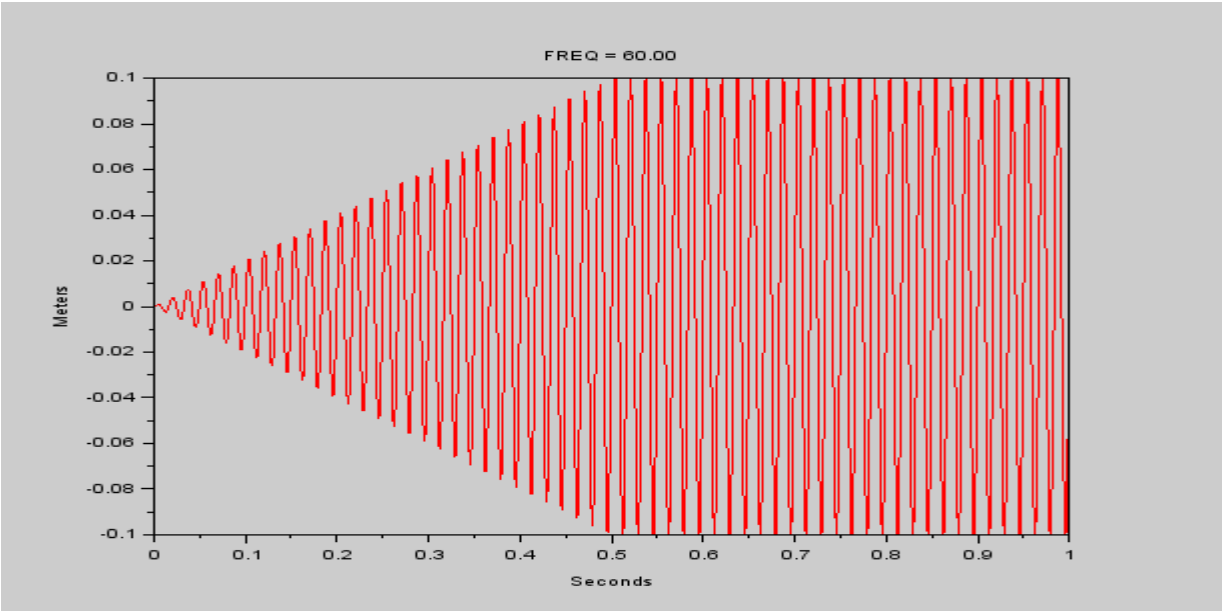


Figure 6. Displacement

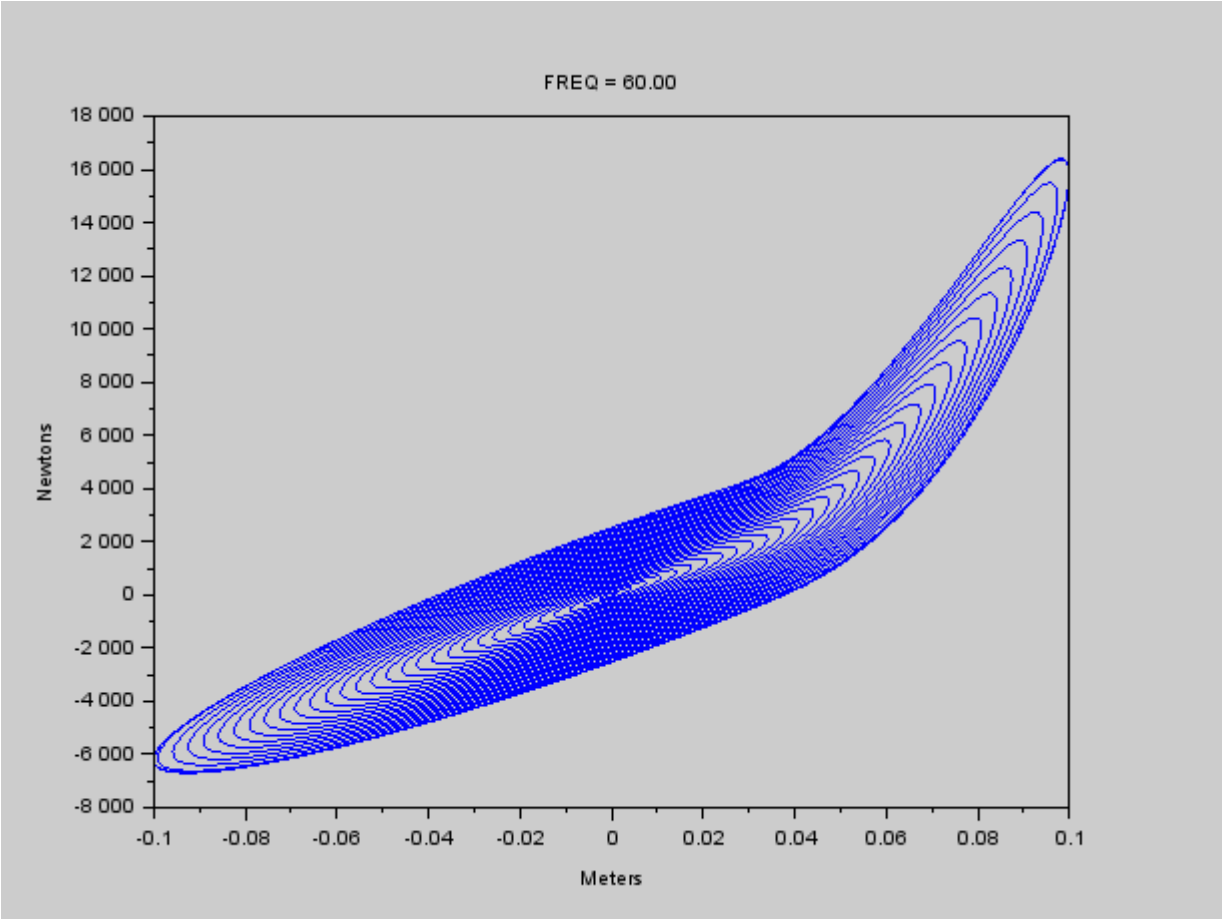


Figure 7. Cycle

4 The kind of models SID is used in

4.1 “Global car model”

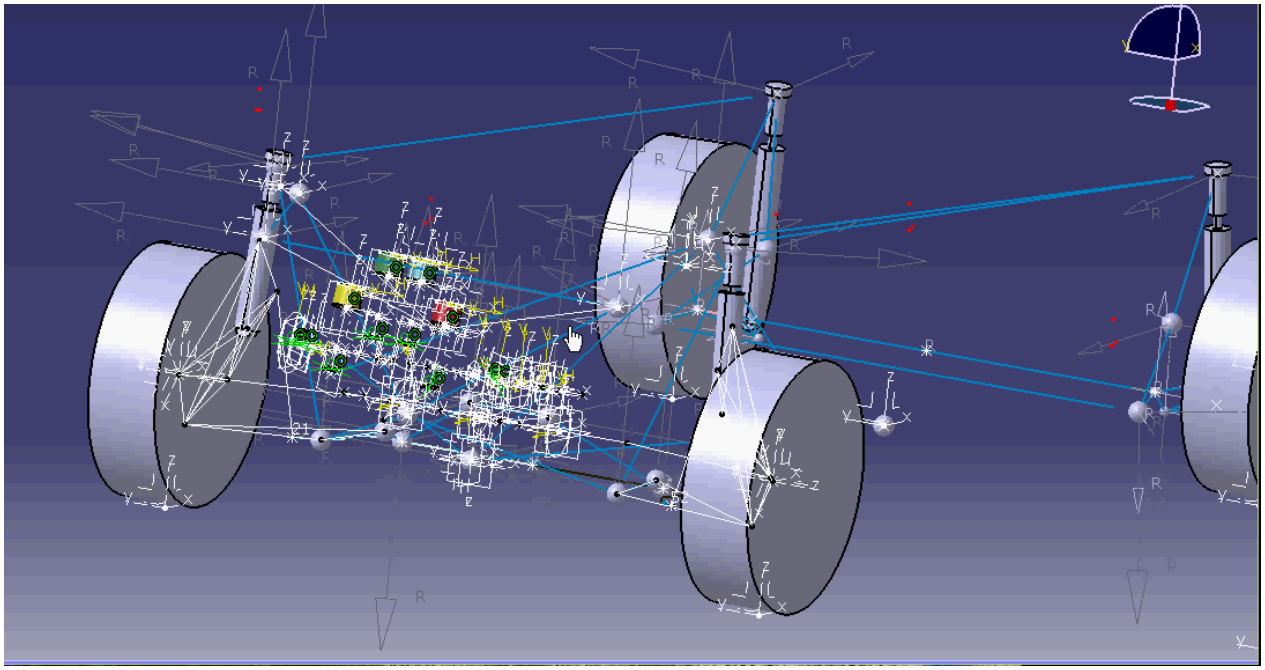


Figure 10. “Global car model”

Figure 10 shows a “VeLab” model inclusive of practically everything which is needed to predict a vehicle startup behavior. Models of the following subsystems are included:

- Starter
- Engine, pistons, crankshaft, links, engine mounts
- Clutch
- Gearbox, gears, differential, transmissions
- Suspensions, dampers, steering apparatus
- Wheels
- Tires
- Rigid or flexible car body and suspension frameworks
- Torsional dampers

4.1.1 Applicability

It is generally possible to devise and validate such subsystems separately and then assemble them in the global model. Such “global” models are seldom used except for special problems involving the whole of the vehicle like for example the study of vibration energy transmission from the engine through the suspensions and to the car body. Animation of this model helps understanding “global” problems sometimes.

4.2 “Driveline model”

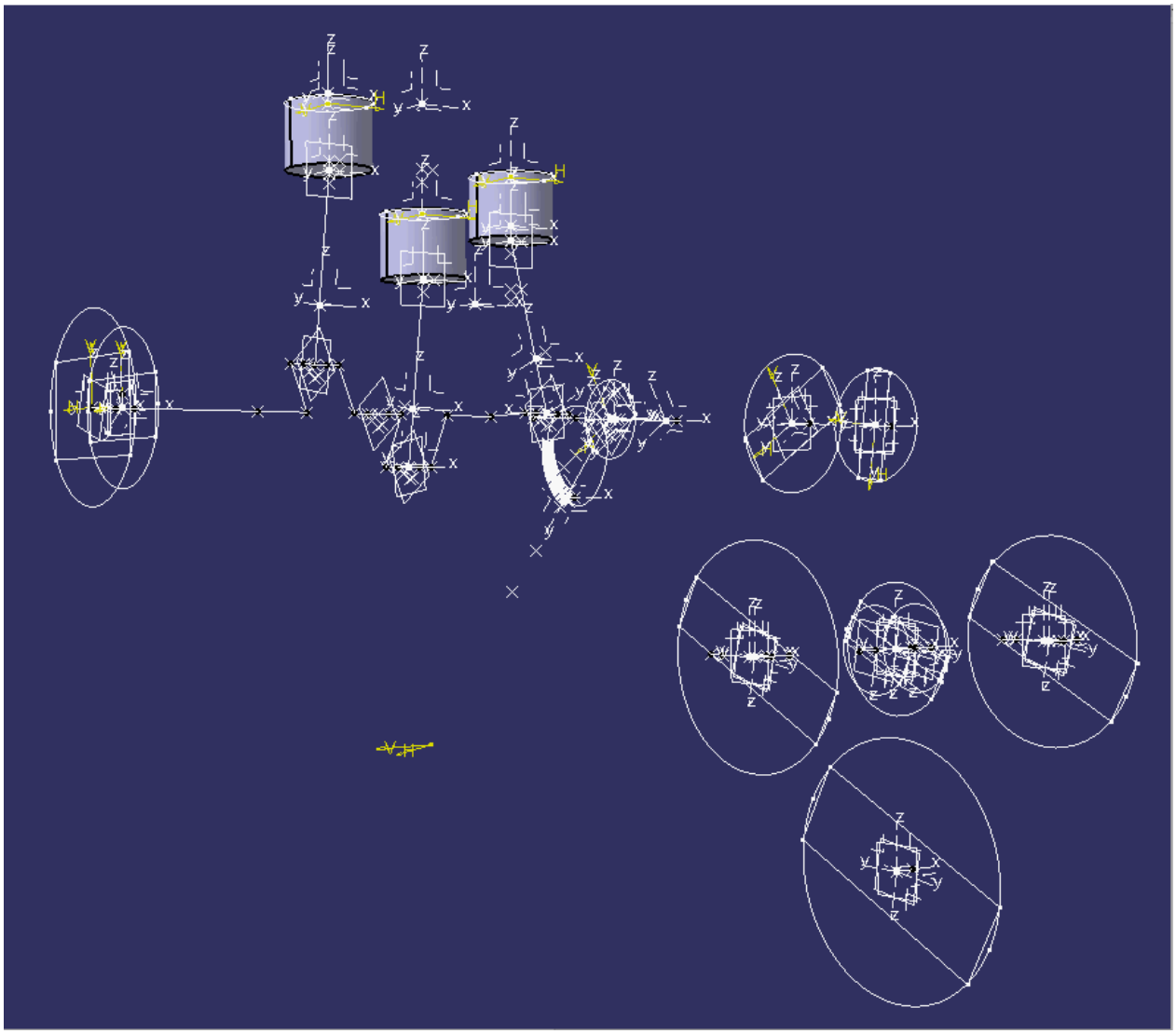


Figure 11. “Driveline model”

Figure 1 shows a driveline model that was used to study a pendulum damper dynamic behavior. The starter, gear, differential and wheels are modelized together with the vehicle which is represented by a big flywheel in such a model.

4.2.1 Applicability

Such models are more often used than the “global” one. The effects of the SHR (wheel longitudinal vibration) mode can be studied by such a model and SID is used to modelize practically everything flexible in the model. Only the tire model also includes viscous damping, tire longitudinal stiffness being concerned.

5 General remarks

SID is of great help in preparing such models because it provides the desired natural damping behavior. Using viscous damping for example would require adapting the damping to the new situation every time some eigenfrequencies change because of structural modification. Viscous damping cycles are in facts strongly frequency dependent. Once the 3 parameters governing SID are set, instead one can almost forget damping modelling and go on trying new solutions in a most

expedite way. SID also has the prize of simplicity: it would be very difficult obtaining the same result displayed in figures 3-5-7 by other methods and by 18 lines of code only (see the script in the appendix). One can quickly prepare macros that formulate SID for all elastic elements in a model. It is very important to remark that the phenomena dealt with by such models all start by low levels of vibration and then soar to higher vibration levels as the transient goes on: this is precisely the kind of phenomena SID was born to deal with. This is also the reason for the "rising amplitude imposed displacement" (figures 2-4-6) chosen for the examples of figures 3-5-7 and the corresponding SciLab script in the appendix.

Conclusions

The validity of a theory can only be proved by its agreement with reliable experimental results like the well-known result of eq. (5). In this sense SID has been shown to give the kind of results we expect (see figures 3-5-7). We don't know whether SID is a ''beable'' which is what the physicists call something that has a real link with physical reality: only the future can say. We can say however that it is a very practical and easy method that corresponds, in the time domain, to the imaginary stiffness damping of eq. (1) in the frequency domain: no physical base to it but everybody uses it because it is simple and practical (see the general remarks of paragraph 5). SID only needs three parameters to be defined. SID is a suitable mathematical description of hysteretic damping and gives fairly physical results when applied to non-linear problems (see figures 3-5-7).

References

[1] L.A. Pars, *A Treatise on Analytical Dynamics*, Ox Bow Press.
[2] U. Tornar, *Modelizing Structural Hysteretic Damping by Means of MDI ADAMS*, Proceedings of the ADAMS 1997 Users Conference, Marburg, German
[3] U. Tornar, *Application of Strain Integral Damping to nonlinear Engine Mounts*, ISMA 2014, Leuven, Belgium
[4] R.K. Otnes and L. Enochson, *Applied Time Series Analysis*, Wiley-Interscience

Appendix

In the following script the variables correspond to:

freqq = frequency

tt = time

dd = displacement

cs1 = natural hysteretic damping

k1 = linear stiffness

dt = time differential

zh1 = SID ω_2 parameter of eq. (19)

h1 = SID ω_1 parameter of eq. (11)

va = velocity

z1 = solution of eq. (19)

ss1 = solution of eq. (11)

z1av = moving average of eq. (21)

force1 = force of nonlinear spring: spline spl(x) of eq. (27)

secstiff = secant stiffness: (force/displacement) that is k_s of eq. (27)

force1 = after definition of secstiff, it is the total force including damping force

SCILAB SCRIPT

```
clear;
freqq = 20; fig1 = 1; fig2 = 2;
tt=(1:4096)/4096; ll = 2* %pi; dd = sin(ll*tt*freqq)/10;
for kk = 1 : 2048; dd(kk) = dd(kk) * tt(kk)/tt(2048); end;
figure(fig1); title('FREQ = ' + msprintf('%.2f',freqq)); plot(tt,dd,'r'); xlabel('Seconds'); ylabel('Meters');
csi1 = 0.4; m1 = 400; f1 = 2; k1 = (ll*f1)*(ll*f1)*m1; dt = 1/4096;
ss1 = 0; h1 = 0.2; z1 = 0; zh1 = 0.0001; force1 = 0.;
for kk = 1 : 4096 - 1;
va = (dd(kk+1) - dd(kk))/dt; z1 = z1 + (-z1 * zh1 + dd(kk)) * dt;
z1av = z1 * zh1 / max(0.0001,1-exp(-zh1*tt(kk)));
ss1 = ss1 + (-ss1 * h1 + dd(kk)-z1av) * dt;
force1 = (dd(kk))*k1 + 2.*((dd(kk)) > 0.03)*(dd(kk)-0.03)**2*1000000;
secstif = abs(force1/ (dd(kk)));
force1 = force1 + (-0.*0.15+sign(va))*(abs(ss1 * va))**0.5 * secstif * csi1;
force(kk) = force1;
end
dd = dd(1:length(dd)-1); figure(fig2); title('FREQ = ' + msprintf('%.2f',freqq)); plot((dd),force,'b');
xlabel('Meters'); ylabel('Newtons');
```

Analysis of the first modal shape using case studies

†*A.M. Wahrhaftig

[1]Department of Construction and Structures, Polytechnic School, Federal University of Bahia, Brazil

†*Presenting and Corresponding author: alixa@ufba.br

Abstract

Eigenvector analysis can be performed to determine the shapes and frequencies of the undampened free vibration modes of a system. These natural modes provide excellent insight into the behavior of a particular structure. Eigen vector analysis involves solving the generalized eigenvalue problem, which considers the stiffness and mass matrix of a structure. When a geometric nonlinear study must be performed, a situation that commonly occurs in the analysis of slender structures, nonlinear analysis or a more complete and rigorous evaluation that considers both parts of the total matrix is required. For instance, slender structures possess a small first frequency of vibration, less than 1 Hz, and can resonate due to wind excitation. The first frequency and shape of vibration are the most important parameters for calculating the response of a structure to environmental excitation. Therefore, when modal analysis depends on the stiffness of the structure, the effect of a reduction in stiffness on the modal shape of vibration must be determined. To this end, case studies were evaluated using the finite element method (FEM), considering and neglecting the geometric portion of the stiffness matrix. Mathematic functions were also applied for comparison.

Keywords: Modal Shape, Geometric Stiffness, Nonlinear Analysis, Computational Simulation, Mathematic functions, Case Studies

Introduction

For structures with a first natural frequency less than 1 Hz, the dynamic effects of wind are too important to be considered as pure static effort or deterministic in nature, which would only provide a rough approximation. Regarding the importance of the dynamic effects of wind, Durbey and Hansen (1996) suggested that flexible structures vibrate in different modes, frequencies and shapes when excited by the wind. Further, they stated that the dynamic effect of wind may allow slender structures to display resonance.

In many countries, models that consider the effects of wind in design structures are provided by governing codes. Many of these models consider that average wind speeds produce a static effect, whereas fluctuations or gusts of wind produce important oscillations, especially in tall constructions. When dealing with the dynamic response to the average wind speed, fluctuations are considered to occur in the band of the lower frequencies of the structure. This model of dynamic analysis was also considered by Simiu and Scalan (1996), who suggested that induced vibration analysis for floating loads was a necessary model component. Moreover, constructions with a basic period greater than 1 s and frequencies up to 1 Hz can undergo a floating response in the direction of the wind. Although the frequencies and vibration shapes of a structure should be considered, the most important parameter is the fundamental frequency.

Modal analysis and vibration shapes

A classical method for the dynamic analysis of a structure is modal analysis, in which sufficient information on the system or structure is obtained to reproduce their dynamics. Carrion et al. (2014) previously indicated that the natural frequencies (eigenvalues) and modes of vibration (eigenvectors) of the system are relevant information for classical modal analysis. Carrion further stated that a well-known concept used in the finite element method (FEM) is the stiffness matrix, which is used to relate the external forces applied at the nodes of the structural element to the nodal displacement.

Structural dynamics can be employed to obtain solutions to homogeneous differential equations, the shape of which represents vibration modes that exist in the coordinate system at the same frequency range and occur harmonically in time. The equation describes the vibration of the system according

to a normal mode of vibration and corresponds to the frequency. After deriving the solution twice with respect to time and canceling the harmonic function, the homogeneous algebraic equations shown in Eq. (1) were obtained. In the equation, ω^2 are the eigenvalues, and Φ are the eigenvectors in the FEM environment.

$$\left[\left[K\right]-\omega^2\left[M\right]\right]\ \left\{\Phi\right\}\ =\ 0 \tag{1}$$

[*K*] is the total stiffness matrix, which is composed of two parts, one being conventional, as shown in Eq. (2), the other being geometric, as shown in Eq. (3). [*M*] is the known mass matrix, pertaining to modal analysis with geometric nonlinear characteristics. When the mass matrix is a discrete mass distribution (lumped mass) of the structural system, a diagonal matrix containing the masses and moments of inertia for the nodal displacements is obtained.

$$\left[k_0\right]=E\begin{bmatrix} \frac{A}{L} & 0 & 0 & -\frac{A}{L} & 0 & 0 \\ & \frac{12I}{L^3} & \frac{6I}{L^2} & 0 & -\frac{12I}{L^3} & \frac{6I}{L^2} \\ & & \frac{4I}{L} & 0 & -\frac{6I}{L^2} & \frac{2I}{L} \\ & & & \frac{A}{L} & 0 & 0 \\ & \textit{symmetric} & & & \frac{12I}{L^3} & -\frac{6I}{L^2} \\ & & & & & \frac{4I}{L} \end{bmatrix} \tag{2}$$

$$\left[k_g\right]=\frac{F}{L}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ & \frac{6}{5} & \frac{L}{10} & 0 & -\frac{6}{5} & \frac{L}{10} \\ & & \frac{2L^2}{15} & 0 & -\frac{L}{10} & -\frac{L^2}{30} \\ & & & 0 & 0 & 0 \\ & \textit{symmetric} & & & \frac{6}{5} & -\frac{L}{10} \\ & & & & & \frac{2L^2}{15} \end{bmatrix}. \tag{3}$$

The mathematic solution to the dynamic problem is a polynomial equation of degree *n* that contains the variable ω^2 and is commonly known as the frequency equation. The *n* solutions for ω_I are real and positive and are considered the natural frequencies of the system. The smallest frequency is typically denoted as ω_1, while the largest frequency is denoted as ω_n. Thus, *n* modes of vibration can be determined and collected in a modal *n x n* matrix, which contains columns representing the *n* modes of undampened, normalized free vibration (Brazil, 2004). Each pair of eigenvalues and eigenvectors corresponds to a frequency and mode of vibration for the system. To consider values and characteristic vectors equal in number to the nodal displacements of the system, Venancio Filho (1975) suggested that Eq. (1) can be written as follows:

$$\left[\Phi\right]\left[\omega^2\right]=\left[K\right]\left[M\right]^{-1}\left[\Phi\right] \tag{4}$$

where $[\omega^2]$ is the diagonal matrix of order *n* and consists of the natural frequencies squared, and [Φ] is an *n x n* matrix and contains columns corresponding to the normal modes of vibration. The term $[K][M]^{-1}$ is a dynamic matrix, as previously mentioned by Blessmann (2005).

The formulation corresponding to the previous exposition of the FEM is a geometric nonlinear formulation and is based on the geometric stiffness matrix. Geometric stiffness has been introduced in several analyses of the FEM when nonlinear effects or geometric nonlinearity (GNL) are considered. The interpolation functions normally used in FEM formulations to determine the full stiffness matrix are third-degree polynomials, such as those evaluated by Filho (1975) and Wilson and Bathe (1976).

Computer models of actual structures were developed in the present study using a FEM-based computer modeling program, and modal analysis was performed linearly and nonlinearly to obtain the shape of the first mode of vibration. For comparative purposes, mathematic functions, such as the trigonometric function given in Eq. (5), the polynomial function given in Eq. (6), and the potential function given in Eq.(7). All of the functions were considered to be valid throughout the entire domain of the structure.

Trigonometric function

$$\phi(x) = 1 - \cos\left(\frac{\pi x}{2L}\right). \tag{5}$$

Polynomial function

$$\theta(x) = 3\frac{x^2}{L^2} - 2\frac{x^3}{L^3}. \tag{6}$$

Potential function

$$\psi(x) = \left(\frac{x}{L}\right)^{\gamma}. \tag{7}$$

The value of γ was determined in the present research.

Analysis of the first modal shape using case studies

Extremely slender structures possessing frequencies of the first vibration mode less than 1 Hz were selected. Modal analysis was achieved using finite element models, according to SAP2000 (integrated software for structural analysis and design, Analysis Reference Manual, Computer and Structures, Inc., Berkeley, California, USA), a commercial software package. Modal shapes for the structures were obtained linearly and nonlinearly. The procedure used to calculate the nonlinear modal shape considered geometric stiffness; therefore, the influence of axial loads was inserted in the stiffness matrix. The structures were modeled using bar elements with constant and variable cross sections, as appropriate.

Structure with a slenderness index of 310

The evaluated structure was 48 m high and possessed a hollow circular section with a variable external diameter (ϕ_{ext}) and thickness (t). The slenderness index of the pole was set to 310. The geometric details are shown in Figure 1(b), where t is the thickness of the wall of each segment of the structure. The metal pole was used to support the transmission system for mobile telephone signals. Table 1 lists the structural parameters and existing devices on the structure, and Table 2 specifies the structural properties and model discretization values.

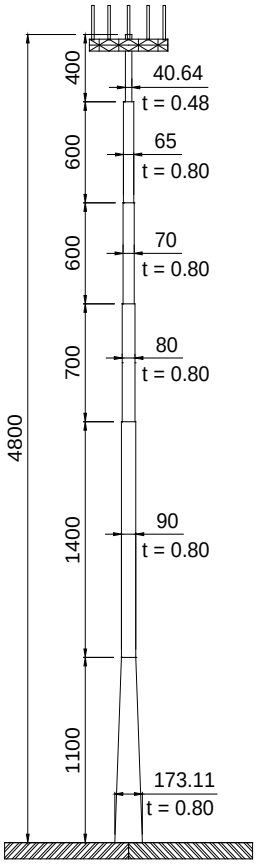
Table 1. Devices and weights on the structure

Device	Height	Weight and distributed weight
Pole	from 0 to 48 m	7850 kN m^{-3}
Ladder	from 0 to 48 m	0.15 kN m^{-1}
Cables	from 0 to 48 m	0.25 kN m^{-1}
Antenna and supports	48 m	3.36 kN

Table 2. Structural properties and discretization of the FEM model

Height (m)	ϕ_{ext} (cm)	t (cm)	Height (m)	ϕ_{ext} (cm)	t (cm)	Height (m)	ϕ_{ext} (cm)	t (cm)	Height (m)	ϕ_{ext} (cm)	t (cm)
48.00	40.64	0.48	30.00	80.00	0.80	20.00	90.00	0.80	10.00	97.56	0.80
46.00	40.64	0.48	29.00	80.00	0.80	19.00	90.00	0.80	9.00	105.11	0.80
44.00	40.64	0.48	28.00	80.00	0.80	18.00	90.00	0.80	8.00	112.67	0.80
42.00	65.00	0.80	27.00	80.00	0.80	17.00	90.00	0.80	7.00	120.22	0.80
40.00	65.00	0.80	26.00	80.00	0.80	16.00	90.00	0.80	6.00	127.78	0.80
38.00	65.00	0.80	25.00	80.00	0.80	15.00	90.00	0.80	5.00	135.33	0.80
36.00	70.00	0.80	24.00	90.00	0.80	14.00	90.00	0.80	4.00	142.89	0.80
34.00	70.00	0.80	23.00	90.00	0.80	13.00	90.00	0.80	3.00	150.44	0.80
32.00	70.00	0.80	22.00	90.00	0.80	12.00	90.00	0.80	2.00	158.00	0.80
31.00	80.00	0.80	21.00	90.00	0.80	11.00	90.00	0.80	1.00	165.56	0.80
									0.00	173.11	0.80

(a) Slender metallic pole



(b) Geometric details

Figure 1. Slender metallic pole and its geometric details

The modal shapes obtained by FEM and the aforementioned mathematic functions are provided in the graph shown in Figure 3. The exponent of the potential function that best fit the curve was equal to 1.965.

Structure with a slenderness index of 256

This investigated structure is a truncated cone metallic pole with 52 cm and 82 cm top and bottom diameters respectively. It is intended for the sustaining of the mobile phone broadcasting system. It is 30 meters high, hollow section. The external diameter (ϕ_{ext}) and thickness (t) vary along of the height. The assessed slenderness of the structure is 256.

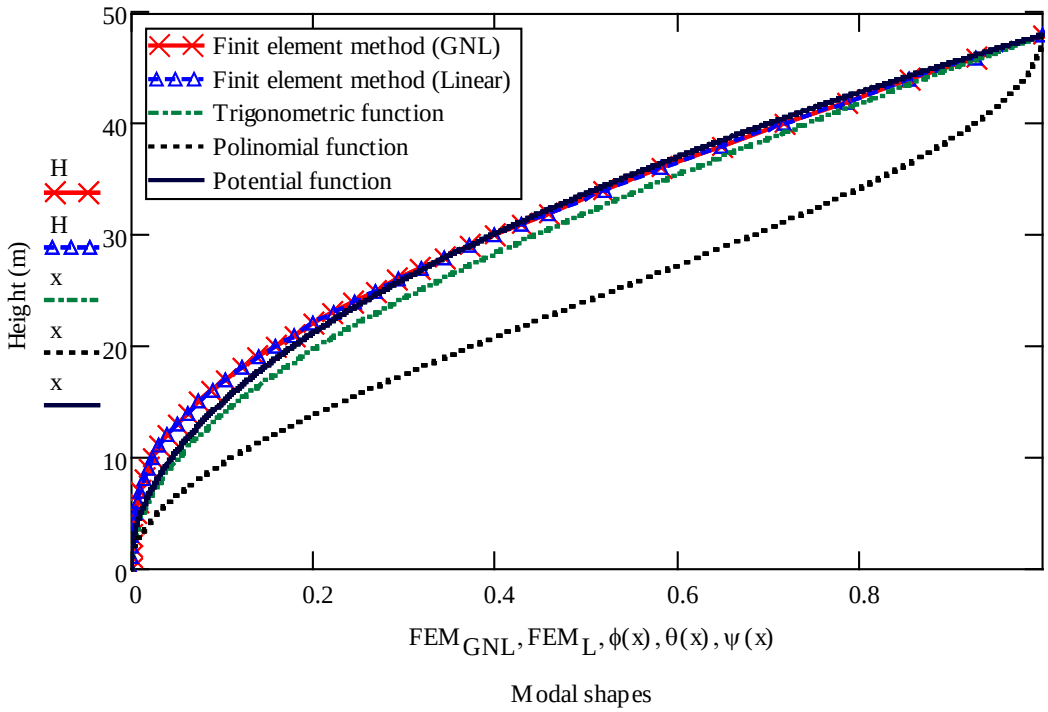


Figure 2. Modal shapes of the structure with slenderness 310

The structure data were acquired in the field. The diameters were measured with a metallic tape measure and the thickness with ultrasound equipment. For a given vertical line, several thickness measurements were carried out to obtain a relative average of the band. The union of the pole segments is formed by successive fittings, by placing and screw-fastening the metallic parts. Each superpositioning band has 20 cm length. In these joint areas, the thickness of the transverse section corresponds to the sum of the measures of the superpositioning bands, conform is indicated in Figure 3. In Table 3 it can be found the properties and the discretization used to model the structure.

Table 3: Structural properties and discretization of the FEM model.

Height (m)	ϕ_{ext} (cm)	t (cm)	Height (m)	ϕ_{ext} (cm)	t (cm)	Height (m)	ϕ_{ext} (cm)	t (cm)
30.00	52.00	0.60	20.00	62.00	0.60	10.00	72.00	0.76
29.00	53.00	0.60	19.00	63.00	0,60	9.00	73.00	0.76
28.00	54.00	0.60	18.10	63.90	0.60	8.00	74.00	0.76
27.00	55.00	0.60	17.90	64.10	0.60	7.00	75.00	0.76
26.00	56.00	0,60	17.00	65.00	0.60	6.10	75.90	0.76
25.00	57.00	0.60	16.00	66.00	0.60	5.90	76.10	0.76
24.10	57.90	0.60	15.00	67.00	0.60	5.00	77.00	0.76
23.90	58.10	0.60	14.00	68.00	0.60	4.00	78.00	0.76
23.00	59.00	0.60	13.00	69.00	0.60	3.00	79.00	0.76
22.00	60.00	0.60	12.10	69.90	0.60	2.00	80.00	0.76
21.00	61.00	0.60	11.90	70.10	0.76	1.00	81.00	0.76
						0.00	82.00	0.76

The metallic pole sustains two working platforms, one situated at 20 m height and the other at the superior extremity. There is still a set of antennas located at 27 m from the base and attached to the body of the pole through metallic devices. The platforms and the supporting devices follow the composition presented in Table 4 where ϕ designate the diameter of the platform. The local assessment revealed the presence of microwave (MW) antennas and of radio frequency (RF), which are listed with the rest of the structure accessories in

Table 5. The data related to the antennas were obtained from the catalogue of the manufacturer. All the aforementioned devices represent additional masses and concentrated forces on the structure, as shown in
Table 6, which presents the structural parameters and the parameters of the existing devices, the specific weight adopted for the material of the structure, the localized and distributed axial load. The geometry of the structure and the existing devices are schematically represented in Figure 3. In Figure 4 they are presented photographic images of the pole.

Table 4. Composition of the platform and support

Platform ϕ = 2.5 m	Mass (kg)
Floor sheet	116
Lateral floor sheet	46
Channel (U) 150 × 12.2 mm – Banister	96
Angle (L) 102 × 76 × 6.4 mm – Banister	68
Angle (L) 102 × 76 × 6.4 mm – Banister	77
Angle (L) 102 × 76 × 6.4 mm – Floor support	43
Platform lower ring	14
Joints	3
Banister bolts	5
Angles (L) 152 × 102 × 9.5 mm – Platform lower support	33
Total =	500
Support set for antenna	**Mass (kg)**
Pipe ϕ = 1´ (25.4 mm)	6
Angle (L) 203 × 152 × 19 mm	50
Staples U (ϕ = 1´ = 25.4 mm)	1
Top plate	1
Total =	58

Table 5. Composition of the localized nodal masses

Device	Mass (kg/unit)	1st Plat (20 m) Quant.	(kg)	Support (27 m) Quant.	(kg)	2nd Plat (30 m) Quant.	(kg)
Antenna RF 2.6 m	19	2	37	3	56	1	19
Antenna RF 1.23 m	4	1	4	0	0	1	4
Antenna MW	19	2	38	0	0	0	0
Platform	500	1	500	0	0	1	500
Support for antennas	58	6	345	3	173	6	345
Pipe ϕ = 1´ (25.4 mm) (Guide)	6	0	0	0	0	1	6
Pipe ϕ = 3/4´ (19 mm) (LC)	6	0	0	0	0	1	6
Total (kg) =			924		228		880

(LC = Lightning conductor, MW = Microwave, RF = Radio frequency, Plat = Platform)

Table 6. Localized axial load and characteristics of the devices

Device	Frontal area	Height	Weight, distributed weight
Pole	Variable	0-30 m	77 kNm^{-3}
Ladder	0.05 m^2/m	0-30 m	0.15 kN m^{-3}
Cables	0.15 m^2/m	0-30 m	0.25 kN m^{-3}
1st Platform Antenna of the 1st platform	2.60 m^2 1.99 m^2	20 m	9.06 kN
Intermediate antennas Intermediate supports	2.11 m^2 0.56 m^2	27 m	2.24 kN
2nd Platform Antennas of the 2nd platform	2.36 m^2 0.90 m^2	30 m	8.63 kN

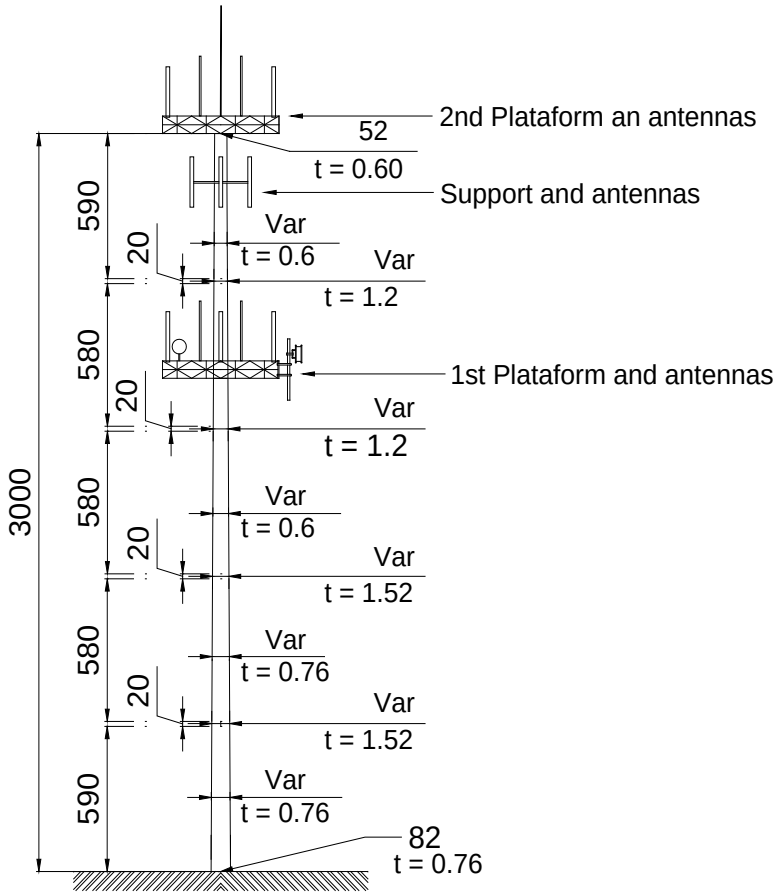


Figure 3. Geometry – Measures in centimeters

Figure 4. General photographic views

The modal shapes obtained by FEM and by the mathematic functions can seem in graph of Figure 5. The exponent of the potential function which best adjusts the curve is 1.85.

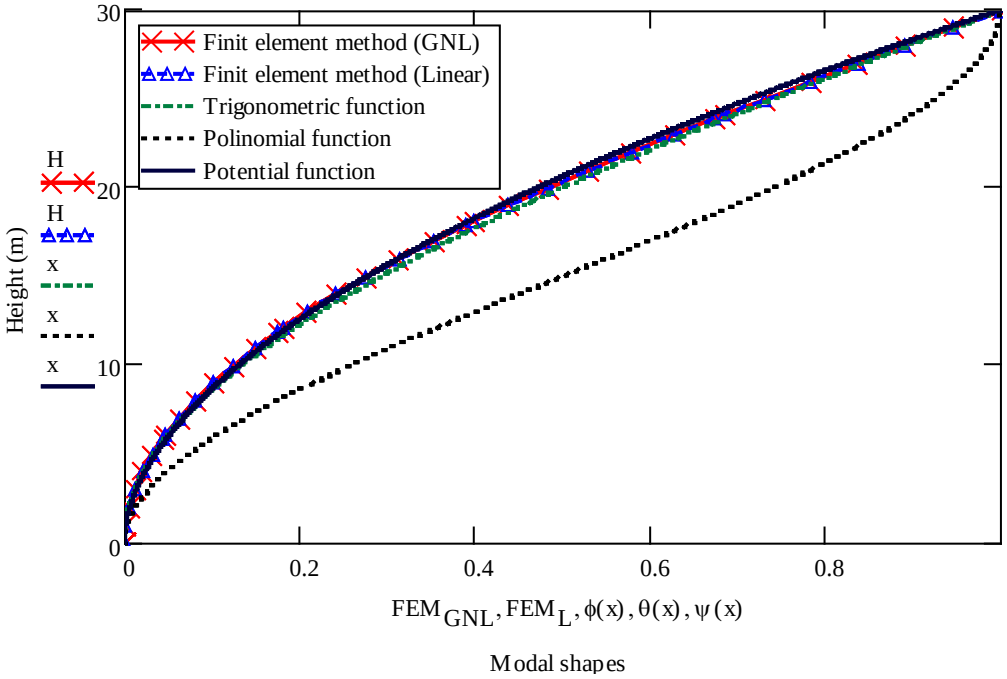


Figure 5. Modal shapes of structure with slenderness 256

Conclusions

In the present study, the shape of the first mode of vibration was investigated using case studies. Analysis by finite element method (FEM) was performed using two different procedures, including a linear procedure, where the geometric stiffness was not considered, and a nonlinear procedure, called the geometric nonlinear formulation (GNL), which considered the geometric stiffness. For comparison, several mathematic functions were studied, and all of the functions were valid throughout the entire domain of the structure.

For the studied cases, geometric stiffness did not have a significant effect on the shape of the first mode of vibration, and the trigonometric function was shown to be a good approximation for the nonlinear vibration shape. The mathematic potential function also represented the first shape of the vibration. For the structure with a slenderness index of 310, the exponent of the function was equal to 1.965, while the structure with a slenderness index of 256 corresponded to an exponent of 1.865. With this information, the weight-averaged rate of slenderness (r_s) was determined to be $r_s = 0.006812$. Thus, an adequate exponent could be obtained by multiplying the slenderness index by r_s. For example, for a structure with a slenderness of 200, the exponent is equal to 1.36 (200 times 0.006812).

Finally, the polynomial function did not provide an accurate representation of the vibration shape of the first mode.

References

Dyrbye C., Hansen S.O. (1996) Wind Loads on Structures. John Wiley & Sons, England.

Brasil, M. L. R. F. R (2004) Dinâmica das Estruturas (Dynamics of Structures). Notes. São Paulo. (in Portuguese).

Isyumov, N. (2012) Alan G. Davenport's mark on wind engineering, *J. Wind Eng. Ind. Aerodyn.* 104–106, 12–24. doi:10.1016/j.jweia.2012.02.007.

Simiu, E. and Scanlan, R. (1996) Wind effects on Structures. John Wiley & Sons, New York.

Carrion, R., Mesquita, E., Ansoni, J.L. (2014) Dynamic response of a frame-foundation-soil system: a coupled BEM–FEM procedure and a GPU implementation, *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 1–9. doi:10.1007/s40430-014-0230-3.

Filho, F.V. (1975) Matrix Analysis of Structures (Static Stability, Dynamics), Almeida Neves – Editors, Technological Institute of Aeronautics, Rio de Janeiro.

Wilson, E.L. and Bathe, K.J. (1976) Numerical Methods in Finite Element Analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Shape identification of steady-state viscous flow fields to prescribe flow velocity distribution

∗†**E. Katamine**[1] **and R. Kanai**[2]

[1,2]Department of Mechanical Engineering, Gifu National College of Technology, Motosu-city, Gifu, Japan
∗Presenting author: katamine@gifu-nct.ac.jp
†Corresponding author: katamine@gifu-nct.ac.jp

Abstract

This paper presents a numerical solution to shape identification problem of steady-state viscous flow fields. In this study, a shape identification problem is formulated for flow velocity distribution prescribed problem, while the total dissipated energy is constrained to less than a desired value, in the viscous flow field. The square error integral between the actual flow velocity distributions and the prescribed flow velocity distributions in the prescribed sub-domains is used as the objective functional. Shape gradient of the shape identification problem is derived theoretically using the Lagrange multiplier method, adjoint variable method, and the formulae of the material derivative. Reshaping is carried out by the traction method proposed as an approach to solving shape optimization problems. The validity of proposed method is confirmed by results of 2D numerical analysis.

Keywords: Inverse problem, Shape identification, Optimum design, Flow control, Traction method

Introduction

Shape optimization problems of viscous flow fields for improving performance are important in mechanical engineering fields. The theory of shape optimization for incompressible viscous flow fields was initiated by Pironneau [Pironneau(1973; 1974; 1984)], who formulated a shape optimization problem for an isolated body located in a uniform viscous flow field to minimize the drag power on this body. The distributed shape sensitivity, which is called the shape gradient, was derived with respect to the domain variation by means of an adjoint variable method based on optimal control theory. The adjoint variable method introduces adjoint variables into variational forms of the governing equations as variational variables; it also determines the adjoint variables using adjoint equations derived from criteria defining an optimality condition with respect to the domain variation.

The present authors have proposed an approach for the shape optimization of such channels or bodies based on a gradient method using the distributed shape sensitivity. In previous studies, the present authors presented a numerical method for the minimization of the dissipation energy of steady-state viscous flow fields [Katamine and Azegami(1995); Katamine et al.(2005)] and extended this method to 3D problems [Katamine et al.(2009)]. Also, the present authors applied this method to the shape optimization solution for the drag minimization and lift maximization of an isolated body located in a uniform viscous flow field [Katamine and Matsui(2012)].

The present study describes the extension of this method for solving a shape identification problem of flow velocity distribution prescribed problem in sub-domains of steady-state viscous flow fields. Reshaping is accomplished using the traction method [Azegami el al.(1995; 1997); Azegami(2000)], which was proposed as a means of solving boundary shape optimization problems of domains. In the traction method, domain variations that minimize the objective

functional are obtained as solutions of pseudo-linear elastic problems for continua defined in the design domain. These continua are loaded with pseudo-distributed traction in proportion to the shape gradient in the design domain.

In this study, the shape identification problem is formulated for flow velocity distribution prescribed problem, while the total dissipated energy is constrained to less than a desired value, in the viscous flow field. The square error integral between the actual flow velocity distributions and the prescribed flow velocity distributions in the prescribed sub-domains is used as the objective functional. Shape gradient of the shape identification problem is derived theoretically using the Lagrange multiplier method, adjoint variable method, and the formulae of the material derivative. The validity of proposed method is confirmed by results of 2D numerical analysis.

Flow velocity distribution prescribed problem

Let Ω be a viscous flow fields in a steady state. The fluid flows in from sub-boundaries Γ_0 and flows out from sub-boundaries Γ_1, where we write velocity vector $u = \{u_i\}_{i=1}^n$ and pressure p. A domain variation problem where the flow velocity distribution u is specified with u_D in sub-domains $\Omega_D \subset \Omega$ can be regarded as a shape optimization problem. For simplicity, we assume that the sub-domains Ω_D, sub-boundaries Γ_0 and Γ_1 are invariables. The flow velocity distribution prescribed problem considering constraint for dissipation energy is formulated as

$$
\begin{aligned}
\text{Given} \quad & \Omega && && (1)\\
\text{find} \quad & \Omega_s && && (2)\\
\text{that minimizes} \quad & E(u-u_D,\ u-u_D) && && (3)\\
\text{subject to} \quad & a^V(u,w)+b(u,u,w)+c(w,p)=l(w) && \forall w \in W && (4)\\
& c(u,q)=0 \quad \forall q \in Q && && (5)\\
& a^V(u,u) \le a_M^V && && (6)
\end{aligned}
$$

where Eqs.(4) and (5) are variational forms, or weak forms, using adjoint velocity $w = \{w_i\}_{i=1}^n$ and adjoint pressure q a for the state equations. Eq.(6) is the constraint with respect to the dissipation energy, and a_M^V is the limit of dissipation energy. The flow velocity square error integral $E(u-u_D,\ u-u_D)$ and the terms such as the $a^V(u,w)$ are defined as

$$
\begin{aligned}
& E(u-u_D,\ u-u_D) = \int_{\Omega_D} (u_i-u_{Di})\cdot(u_i-u_{Di})\ dx,\\
& a^V(u,w) = \frac{2}{Re}\int_\Omega \varepsilon_{ij}(u)\varepsilon_{ij}(w)\,dx = \frac{1}{Re}\int_\Omega w_{i,j}(u_{i,j}+u_{j,i})\,\mathrm{d}x,\\
& b(v,u,w) = \int_\Omega w_i v_j u_{i,j}\,\mathrm{d}x, \quad c(w,p) = -\int_\Omega w_{i,i}p\,\mathrm{d}x, \quad l(w) = \int_{\Gamma_1} w_i\hat{\sigma}_i\,\mathrm{d}\Gamma
\end{aligned}
$$

where $\varepsilon_{ij}(u) = \frac{1}{2}(u_{i,j}+u_{j,i})$, Reynolds number Re and the traction $\hat{\sigma}_i$ are given as known values or functions.

Applying the concept of the Lagrange multiplier method and the adjoint variable method, this problem can be rendered as a stationary problem for the Lagrange functional $L(u,p,w,q,\Lambda)$:

$$
\begin{aligned}
L = \quad & E(u-u_D,\ u-u_D)\\
& -a^V(u,w)-b(u,u,w)-c(w,p)+l(w)-c(u,q)+\Lambda(a^V(u,u)-a_M^V) \qquad (7)
\end{aligned}
$$

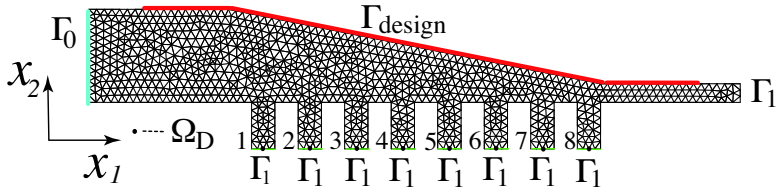


Figure 1: 2D numerical analysis model

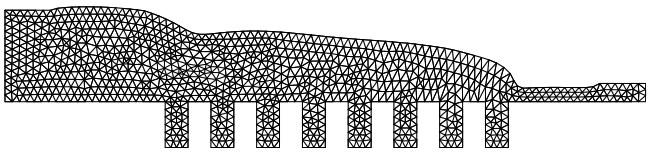


Figure 2: Identified shape

where Λ is the Lagrange multiplier with respect to the dissipation energy constraint. The derivative $\dot{L}$ with respect to domain variation for shape optimization is calculated. Letting this $\dot{L} = 0$, the Kuhn-Tucker conditions with respect to u, p, w, q, Λ are obtained by

$$a^V(u, w') + b(u, u, w') + c(w', p) = l(w') \quad \forall w' \in W \tag{8}$$
$$c(u, q') = 0 \quad \forall q' \in Q \tag{9}$$
$$a^V(u', w) + b(u', u, w) + b(u, u', w) + c(u', q) = 2E(u - u_D,\ u') + 2\Lambda a^V(u, u') \quad \forall u' \in W \tag{10}$$
$$c(w, p') = 0 \quad \forall p' \in Q \tag{11}$$
$$\Lambda \geq 0, \quad a^V(u, u) \leq a_M^V, \quad \Lambda(a^V(u, u) - a_M^V) = 0 \tag{12}$$

that indicate the variational forms of the original state equations for u and p, the variational forms of the adjoint equations for w and q which we call adjoint equations, respectively. Where $(\,\cdot\,)'$ is the shape derivative for domain variation of the distributed function fixed in spatial coordinates. Under the condition satisfying Eqs.(8)- (12), the derivative $\dot{L}$ agrees with the linear form $< G\nu, V >$ with respect to the velocity function V of domain variation:

$$\dot{L}|_{u,p,w,q,\Lambda} =< G\nu, V >= \int_\Gamma G\nu_i V_i \, d\Gamma, \tag{13}$$
$$G = -\frac{1}{Re} w_{i,j}(u_{i,j} + u_{j,i}) + \Lambda \frac{1}{Re} u_{i,j}(u_{i,j} + u_{j,i}) \tag{14}$$

where ν is an outward unit normal vector on the boundary.
The coefficient vector function $G\nu$ in Eq.(13) has the meaning of a sensitivity function relative to domain variation and is so-called the shape gradient function. The scalar function G is called the shape gradient density function. Since the shape gradient function is obtained, the traction method[Azegami el al.(1995; 1997); Azegami(2000)] can be applied to this shape identification problem.

Numerical results

We present the results of a numerical analysis for a 2D shape identification problem using the traction method and the shape gradient derived as described in the above sections.

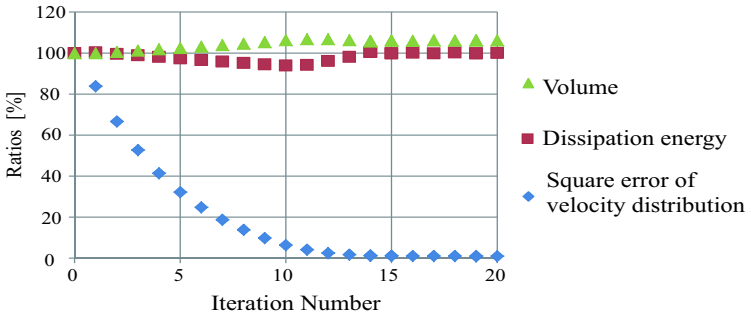


Figure 3: Iterative histories

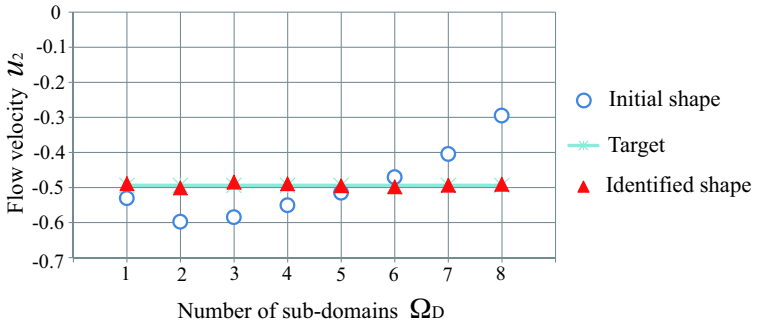


Figure 4: Flow velocity distribution on 8 lower-side sub-domaines Ω_D

We analyzed the 2D problem as one fundamental problem, as shown in Figure 1 The fluid flows in from left-side sub-boundary Γ_0 and flows out from a right-side and 8 lower-side sub-boundaries Γ_1. The sub-domain Ω_D to prescribe the flow velocity distribution was set as 8 lower-side sub-domains. The purpose of this analysis is to determine the shape for which the flow velocity distribution in the 8 lower-side sub-domains becomes as uniform as possible.

In this numerical analysis of the flow field, we used the Hood-Taylor type finite element. That is, the complete polynomial series of the second-order terms was used to provide the interpolation functions for u and w, while the linear polynomial series was used to provide the interpolation functions for p and q. Further, finite elements with six nodes for u and w and three nodes for p and q were also used. The total numbers of nodes and elements were 3,902 and 1,803, respectively. For the analyses of the domain variation V, we used the finite element method with second-order finite elements. The Reynolds number is 100. The dissipation energy is less than the initial shape measure.

The numerical results for the shape identification are shown in Figures 2, 3 and 4. Figures 2 shows the obtained identified shape. Figure 3 shows the iterative history ratios of the square error of velocity distribution $E(u - u_D,\ u - u_D)$, the dissipation energy, and the volume normalized by their respective initial values. Figure 4 shows the flow velocity distribution in the 8 lower-side sub-boundaries Γ_1 for the target, the initial shape, and the identified shape. These results confirm that the flow velocity distribution of the identified shape analyzed by the proposed method approached the target uniform distribution and that the value for the objective functional became zero. The validity of the present method was confirmed based on the numerical results obtained for the basic problems described above.

Conclusions

In the present study, we formulated a shape identification problem in which the square error integral between the actual flow velocity distributions and the prescribed distributions in the

prescribed sub-domains on viscous flow fields was used as the objective functional. The shape gradient of the shape identification problem was derived theoretically. The validity of the proposed method was confirmed based on the results of a 2D numerical analysis. The present study was supported in part by JSPS KAKENHI Grant Numbers 26420161.

References

[1] Pironneau, O.(1973) On Optimum Profiles in Stokes Flow, *J. Fluid Mechanics* **59**, Part 1, 117-128.

[2] Pironneau, O.(1974) On Optimum Design in Fluid Mechanics, *J. Fluid Mechanics* **64**, Part 1, 97-110.

[3] Pironneau, O.(1984) *Optimal Shape Design for Elliptic Systems*, Springer-Verlag.

[4] Katamine, E. and Azegami, H.(1995) Domain Optimization Analyses of Flow Fields, *Computational Mechanics'95*, S. N. Atluri, G. Yagawa, and T. A. Cruse eds., Springer,**Vol. 1**, 229-234.

[5] Katamine, E., Azegami, H., Tsubata, T., and Itoh, S.(2005) Solution to Shape Optimization Problems of Viscous Flow Fields, *International Journal of Computational Fluid Dynamics* **19**, 45-51.

[6] Katamine, E., Nagatomo, Y., and Azegami, H.(2009) Shape optimization of 3D viscous flow fields, *Inverse Problems in Science and Engineering* **17**, No.1, 105-114.

[7] Katamine, E. and Matsui Y.(2012) Multi-objective shape optimization for drag minimization and lift maximization in low Reynolds number flows, *Theoretical and Applied Mechanics Japan* **61**, 83-92.

[8] Azegami, H., Shimoda, M., Katamine, E., and Wu, Z. C.(1995) A Domain Optimization Technique for Elliptic Boundary Value Problems, *Computer Aided Optimum Design of Structures IV*, Hernandez S. and Brebbia C.A. eds., Computational Mechanics Publications, 51-58.

[9] Azegami, H., Kaizu, S., Shimoda, M., and Katamine, E.(1997) Irregularity of Shape Optimization Problems and an Improvement Technique, *Computer Aided Optimum Design of Structures V*, Hernandez S. and Brebbia C. A. eds., Computational Mechanics Publications, 309-326.

[10] Azegami, H.(2000) Solution to Boundary Shape Identification Problems in Elliptic Boundary Value Problems using Shape Derivatives, *Inverse Problems in Engineering Mechanics II*, Tanaka, M. and Dulikravich, G. S. eds., Elsevier, 277-284.

Effectiveness of Load Balancing in a Distributed Web Caching System

Brandon Plumley, Richard Hurley

Department of Computing and Information Systems

Trent University

Peterborough, ON Canada

bplumley@trentu.ca, rhurley@trentu.ca

Abstract

In this paper, we investigate the effects of load balancing in a distributed Web caching system. Our investigation is focused specifically on adaptive load sharing: an approach that reacts to the current state of the system. Load balancing has been shown to improve system performance in other applications and in this paper, we investigate it in a distributed Web caching environment using both a unified and partitioned approach. The goal of this work is threefold: (1) to determine the conditions under which load balancing can be beneficial in a distributed Web caching system, (2) to compare load balancing in a unified and partitioned Web caching system, and (3) to determine how much state information is required to achieve any benefit. Discrete-event simulation is used as the tool to generate results for these different environments.

Keywords: Web Caching, Load balancing, Performance Evaluation, Simulation, Computer Modelling

1. Introduction

Web caching is a technique that is heavily utilized on the Internet and has been shown to be highly effective in improving network performance by reducing bandwidth and latency [1][2][3]. The premise of Web caching is to store frequently-accessed pages from an originating server *closer* to the clients to reduce bandwidth and workload on the originating server[4]. This can result in a reduction in the time to deliver a page from the server to the client [5].

One of the more common approaches is to implement multiple Web caches in a distributed system where additional Web caches are considered peers with each cache being contained within the same level (similar ”distance” from the client). This arrangement allows the peer caches to be relatively *close* to one another. Distributive Web caching allows for better load sharing when compared to other approaches [6].

Traditionally Web caches hold both large and small pages together, where one large page would replace multiple small pages or a single large page. This storage model is referred to as Unified caching. However, partitioned Web caching, where large and small pages are stored in separate

areas, has been shown in previous work to result in increased performance [7]. This approach ensures that large pages will not replace many small pages in the cache.

Since there are multiple caches working together in a distributed environment, a key mechanism to fully harness the potential of the system is load balancing. There have been many load balancing algorithms proposed in the past that have been applied to diverse applications such as telecommunications, processing process on a computer and network traffic [8][9]. In the last decade there has been an increase into research for applying load balancing to a distributed Web caching system [10][11]. In a system without load balancing, requests are typically assigned randomly to the distributed Web caches. With no direction as to which assignment of requests, the issue that arises is that one cache could be congested while other caches are underutilized; this uneven utilization can degrade performance [12].

Typically, there are two common transfer policies used in adaptive load sharing: sender-initiated and receiver-initiated [13]. A sender-initiated policy attempts to balance the workload in the system at the point in time when a Web cache receives an incoming request. A threshold value, which is based on the number of requests in the local queue, is used to determine whether the system needs to transfer the incoming request to another peer cache. This approach will only transfer newly arriving requests with them being placed at the end of the selected Web cache queue. A receiver-initiated policy, on the other hand, attempts to load balance as requests are serviced (not when they arrive). If the Web cache queue falls below a given threshold, the system attempts to find additional work from a peer cache with queue length above the given threshold. If such a cache can be found, a request from its tail will be transferred to the tail of the Web cache that initiated the transfer. In this paper, we focus on sender-initiated policies but more information on the performance of receiver-initiated policies can be found in [14].

For this paper, we used a discrete-event simulation model to investigate the performance of load balancing in a distributed unified and partitioned Web caching system. We present the models and assumptions for our distributed Web caching systems for both unified and partitioned storage in Section 2, while in Section 3 the input parameters are discussed. Section 4 presents the simulation results derived from the models and finally, Section 5 summarizes our findings.

2. Performance Models

Our system model is divided into two parts: a Web reference model and a Web cache model. By varying the architecture in the Web cache model, we produce two distinct systems: unified and partitioned. We can simplify our system models since we are concerned with the relative performance achieved by each load balancing algorithm relative to the distributed Web caching system without load balancing (i.e. we are not concerned with the absolute performance of the system).

2.1. Web Reference Model

The pages stored in a Web cache and their request probabilities vary over time. Pages such as news articles, viral videos, course assignments and memes become popular for periods of time and then eventually the frequency of access decreases. To represent this behavior, we use a dynamic page reference model (shown in Figure 1) as described in [15].

2.1.1. Page Popularity

The request probabilities are shown in Equation (1) and defined by: $p_i(t)$, the probability of requesting page i at time t is $i = 1,2,...,M$, where M is the number of Web pages. From Figure 1, we can see that there are two states for the probability of requesting a page: normal and popular. Pages in the popular state have a higher request probability than that of the pages in the normal state, where v represents the ratio of the rate of requests in the popular to normal state.

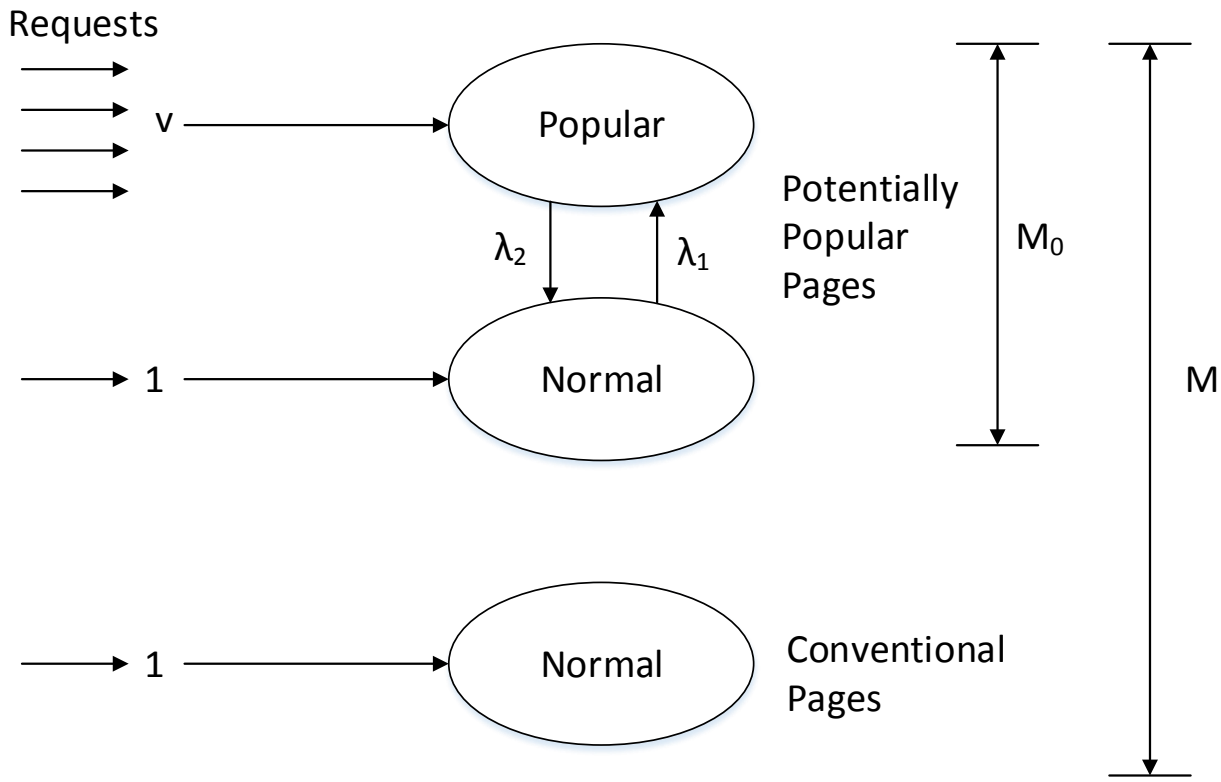


Figure 1: Dynamic Page Reference Model

The model also assumes that there are two types of pages: conventional (M) and potentially popular (M_0). Conventional pages remain in the normal state while potentially popular pages shift between the normal and popular sate based on a continuous-time Markov chain. The rate at which a page transitions from a normal to popular state is λ_1 and from popular to normal is λ_2 (the time spent in either state is assumed to be exponentially distributed). We let $M_0 < M$ denote the number of potentially popular pages and thus $M_p(t) < M_0$ represents the total number of pages in the popular state at time t. The time-dependent request probability for page i is defined as:

$$p_i(t) = \begin{cases} \frac{v}{vM_p(t)+(M-M_p(t))} & \text{popular state} \\ \frac{1}{vM_p(t)+(M-M_p(t))} & \text{normal state} \end{cases} \quad (1)$$

2.1.2. Page Size

To simplify our model we assume that a page is either large or small, we assume that a large page is k times larger than a small page. Small pages have a service time that is assumed to be exponentially distributed with a mean rate of μ^{-1}, while large pages have a exponentially distributed service time of $k\mu^{-1}$.

It has been shown that the majority (ninety percent) of Web pages are in the range of 100 bytes to 100 KB, with less than ten percent being greater than 100 KB [16]. As a result, we assume that the probability of requesting a large page would be $1-s$, where s is the probability of requesting a small page. Since 90% of pages requested are small, we set s to 0.9, which based on previous observations is reasonable.

2.2. Web Cache Model

Our Web cache model is comprised of a page replacement model, an architectural model, and a storage mode.

2.2.1. Page Replacement Model

One of the most critical components of a good Web caching system is the page replacement algorithm. The page replacement algorithm is responsible for storing or discarding pages in the Web cache once it becomes full. Without this component, once the cache is full, no new pages would be stored and the cache would become stale. Although there are many different page replacement algorithms, our model uses The Least Recently Used (LRU) [17].

As the name implies, the LRU algorithm selects the least recently used page (determined from the last accessed timestamp) to be removed from the cache. We have chosen to implement the LRU since it is one of the most widely-used cache replacement algorithms for Web pages [18]. One of the main advantages of the LRU is that it is straightforward to incorporate in the system model, while being highly efficient. Some of the determents to the algorithm are that it excludes certain state information such as the and latency of a page. However, since we are considering only relative performance, these effects will be negligible.

2.2.2. Architectural Model

Our work expands on a Web cache model that was first introduced by [19], and is shown in Figure 2. The distributed system contains D peer (or co-operative) caches which are assumed to exist as the same level. When a Web cache receives a page request, the cache fist checks if there is a copy currently stored in its own cache and if the page is found, it is returned to the client. However, if no copy of the requested page can be found at the current Web cache, the request is forwarded randomly to one of the peer caches. If the page cannot be found at the new Web cache, the request is again transfered to another peer Web cache, until the page is found. If all D peer caches are exhausted the request will be forwarded to the originating server, a copy is made at the original cache and the page is returned to the client.

It is assumed that if the request is satisfied by the first cache in D peer cache, than the processing time is considered to be T_0 (this includes the service time and propagation delay). If the first cache

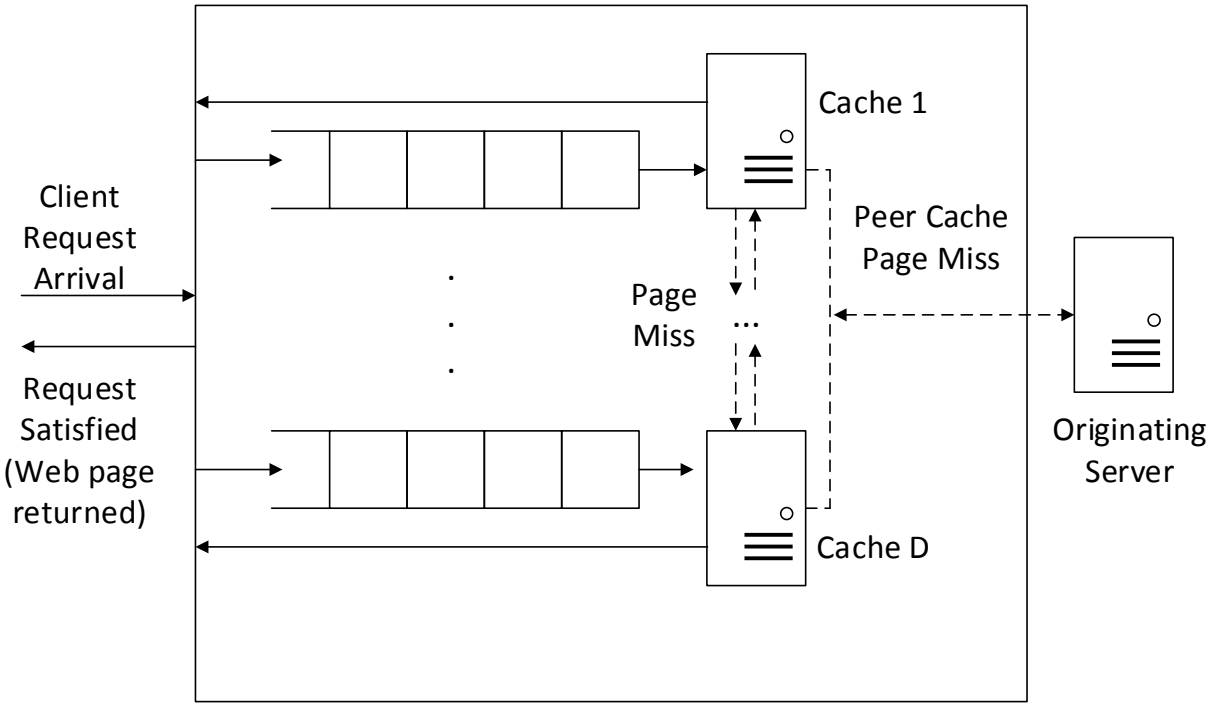


Figure 2: Web Caching Architecture Model

can not satisfy the request (a {textitmiss) but can still be satisfied within the D peer cache, then the service time is assumed to be $T_1 + T_0$. While exact for $D = 2$ caches, this value is an approximation for larger values of D as it would be a factor of the number of cache misses. If the request can not be satisfied within our distributed Web caching system and therefore must be completed by the originating server, then the processing time is considered to be T_2.

2.2.3. Storage Model

We also investigate two variations of the cache storage model: a unified cache and a partitioned cache. A unified cache is simply a single cache that treats both large and small pages the same (they are stored together). If the cache was full and needed to make room for a incoming large page, the cache would have to discard one large page or k small pages. A partitioned cache on the other hand treats large and small pages differently. The cache is split into two separate ares: one for large pages and one for small pages. This approach ensures that large pages will not replace k small pages and that k small pages will not replace a single large page. It is assumed that the ratio of space reserved for large pages is (P_L).

2.3. Load Balancing Algorithms

Our investigation considers two variants of a sender-initiated load balancing algorithm:

- *Short-Sender (SS).* Once a threshold value (Θ) is reached, the algorithm looks for the Web cache with the shortest queue (including itself).

- *Random-Sender (RS).* Once a threshold value (Θ) is reached, the algorithm randomly selects a Web cache (excluding itself).

3. Input Parameters

In order to simplify our investigation, the following model parameters are fixed for all simulations: $M = 1000$, $M_0 = 100$, $\mu = 1$, $P_L = 0.4$, $T_0 = 0.5$, $T_1 = 0.1$, $T_2 = 1$, $s = 0.9$, $z = 100$ and $k = 10$. It is assumed that our system has a finite population of N client workstations, with D peer caches. Each Web cache is assumed to have a size of C bytes, which is defined to be the percentage of total bytes available for storage within the entire system, initially we set C to 0.05 [19].

4. Performance Results

The main objective of this investigation is to evaluate the relative performance of our load balancing algorithms in a distributed Web caching system using both unified and partitioned storage against the same system without load balancing. That is, the chief concern is whether load balancing will be effective in a distributed Web caching system. We are not concerned with the absolute performance of our system but that said, it would be beneficial to also be able to compare the results from the simluation models with experimental data from an implemented system but at this point in time, none was available. This is an area underwhich current work is being applied. Our performance measure of interest in our simulation models is mean response time (the time from when a request is generated until the web page has been returned to the client). The complexity of the system and the number of possible parameters is such that an analytic solution is not tractable thus results are gathered using a discrete event simulation written in C++. For more information on the acutal simluation program, please see [14].

4.1. Threshold Limit

We begin by examining threshold limit (Θ) for a sender-initiated approach in a unified and partitioned storage environment. We simulate the system under a high system load ($\rho_{N_U} = 0.85$) for $D = 2$ and 10 peer Web caches (Figures 3 and 4). The results indicate that all four systems (RS_U - Random-Sender-Unified, RS_P - Random-Sender-Partitioned, SS_U - Short-Sender-Unified, SS_P - Short-Sender-Partitioned) preform at least as well as to that of the distributed Web caching system without load balancing (N_U - No LB-Unified, $N_P - NoLB - Partitioned$). In some cases, response time is decreased by as much as 60.0%.

One of the more prominent trends is that as we increase Θ, the mean response time also increases: this is as a result of the fact that less load balancing is occurring up until the point where no pages are being transfered. However, with the RS_U algorithm we observe a small dip: the valley of the dip tends to be achieved with a threshold value (Θ) just greater than 0 (1 or 2). This can be explained by the fact that when the threshold is set to 0, the system will transfer the work randomly even if the local Web cache queue is the shortest. As the threshold is increased, the probability that the arriving Web cache is the shortest in the system decreases. As the threshold is increased to the point where the algorithm stops initiating transfers, there appears to be an *optimal* value that would be dependent on factors such as system load and number of caches. Going forward, we will be using a threshold value (Θ) of 3, as this is a reasonable choice given that the optimal value can not be directly determined.

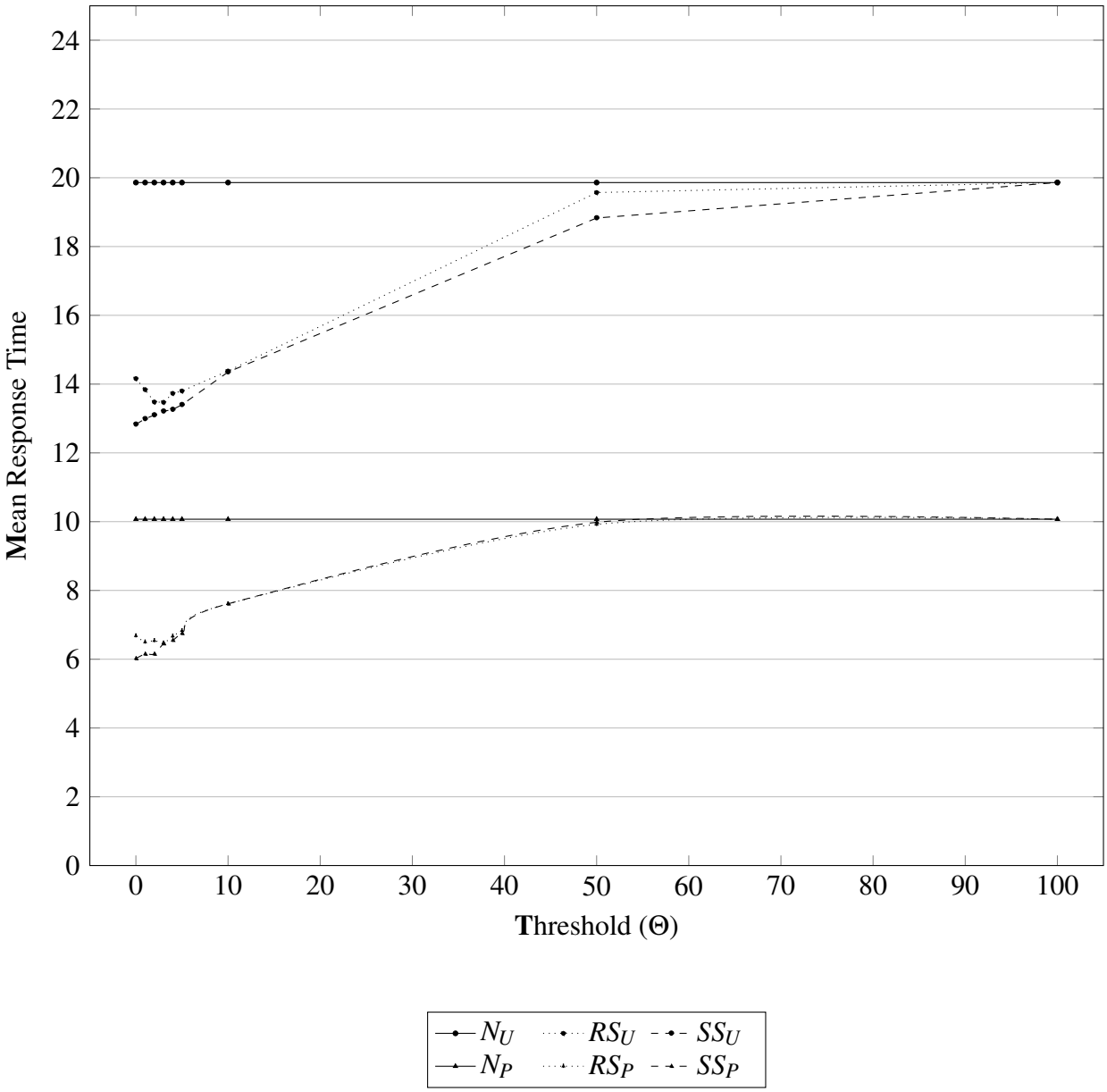


Figure 3: The Effect of Threshold (Θ) on the Mean Response Time for Unified and Partitioned Web Caching: $D = 2$, $\rho_{N_U} \approx 0.85$

4.2. System Workload

We next examine the effects of system workload for $D = 2$ and 10 peer Web caches (Figures 5 and 6). We observe that the system is relatively underutilized (workloads less than 20%), there is little difference between the load balancing algorithms and the respective systems without load balancing. As the utilization increases, we start to see a dramatic improvement (with respect to response time) with our load balancing algorithms relative to the systems without load balancing: this trend becomes more noticeable as the number of peer caches increase. Specifically from Figure 6, the Short (SS_U) algorithm has a decrease in response time of 37.7% over N_U ($\rho_{N_U} \approx 90\%$), while RS_U has a decrease in mean response time of 30.6% over N_U ($\rho_{N_U} \approx 90\%$). The Short

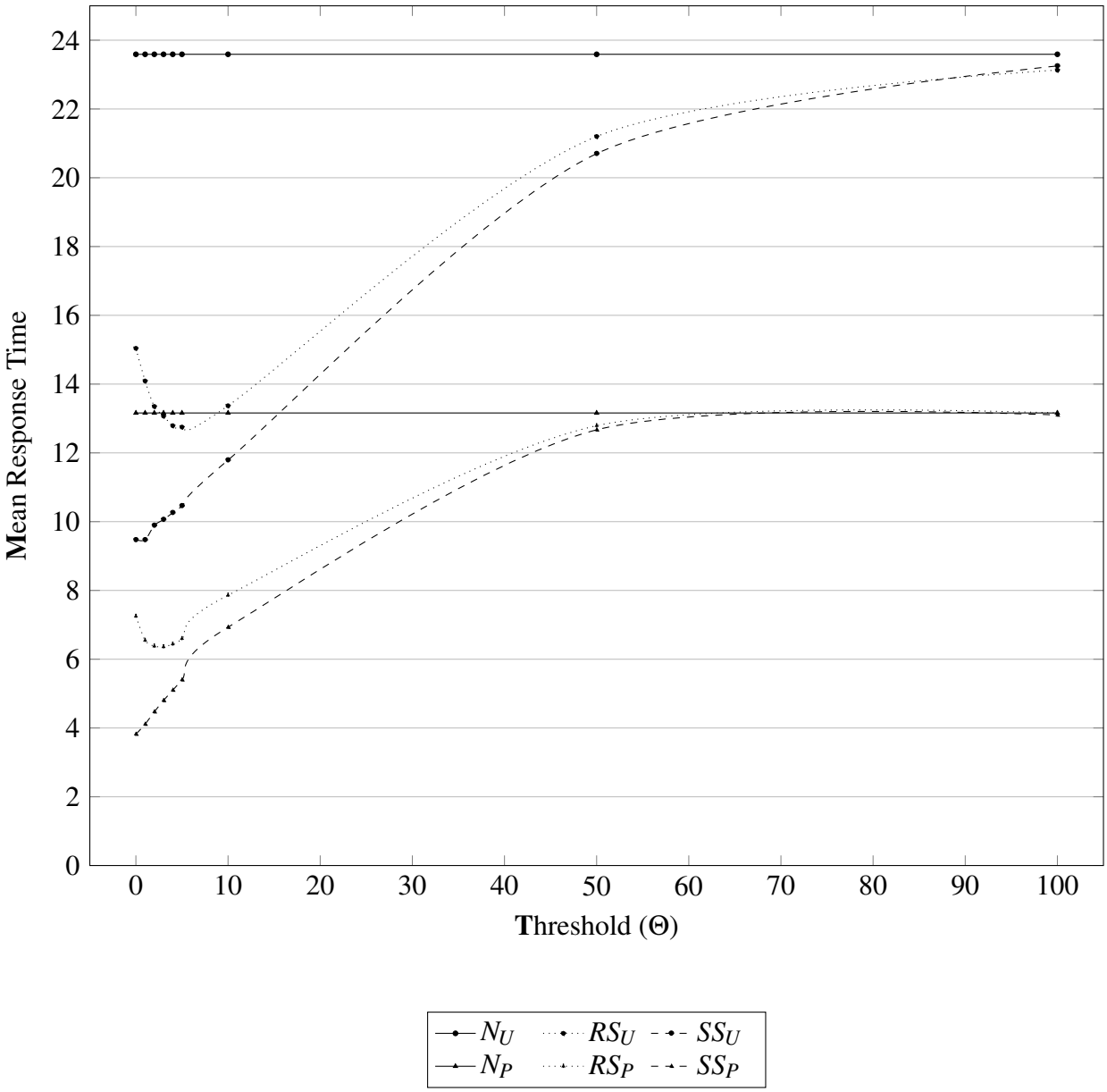


Figure 4: The Effect of Threshold (Θ) on the Mean Response Time for Unified and Partitioned Web Caching: $D = 10$, $\rho_{N_U} \approx 0.85$

(SS_U) algorithm seems to outperform the Random (RS_U) algorithm by 7.1% ($\rho_{N_U} \approx 90\%$). The results also indicate that partitioned load balancing systems follow the same trends as their unified counterparts with partitioning tending to perform better overall.

Additional workload seems to provide more opportunity for the load balancing algorithms to reduce the mean response time and so we can conclude that the higher the system load, the more potential the load balancing algorithms have to make a positive impact on the performance of the distributed Web caching system in both a unified and partitioned storage model.

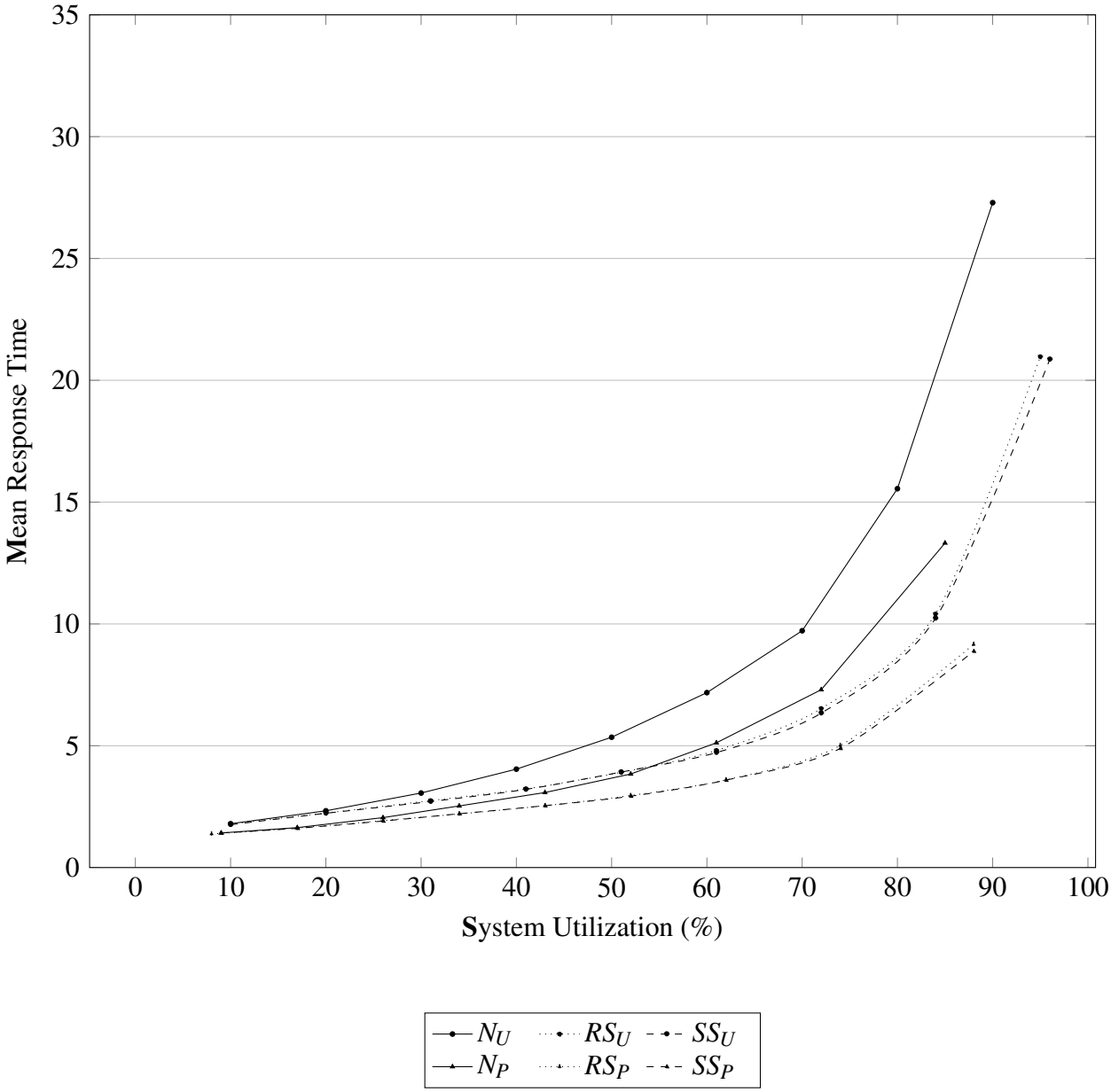


Figure 5: The Effect of System Workload on the Mean Response Time for Unified and Partitioned Web Caching
$\Theta = 3$, $D = 2$

4.3. Number of Peer Caches

We examine both Web caching systems under a medium (Figure 7) and a high system load (Figure 8). As additional peer Web caches are added, the systems without load balancing (N_U, N_P) have mean response times which tend to increase marginally. From Figure 8, the increase in mean response time from 2 to 10 peer Web caches for the systems without load balancing (N_U and N_P) N_U is 12.0% and N_P respectively. Each additional Web cache added to the distributed Web caching system tend to increase the probability that one of the Web caches will become overloaded, leading

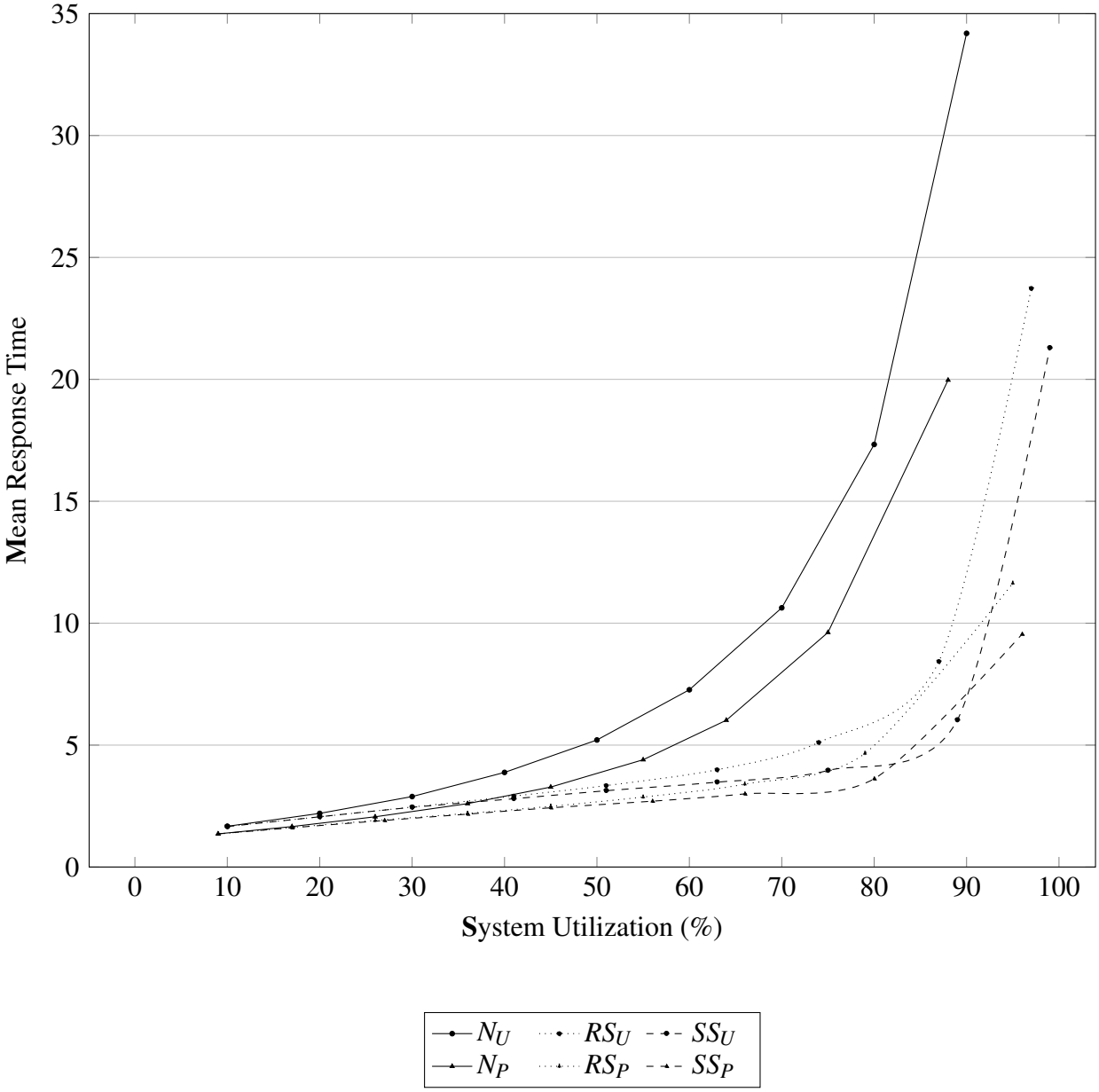


Figure 6: The Effect of System Workload on the Mean Response Time for Unified and Partitioned Web Caching
$\Theta = 3$, $D = 10$

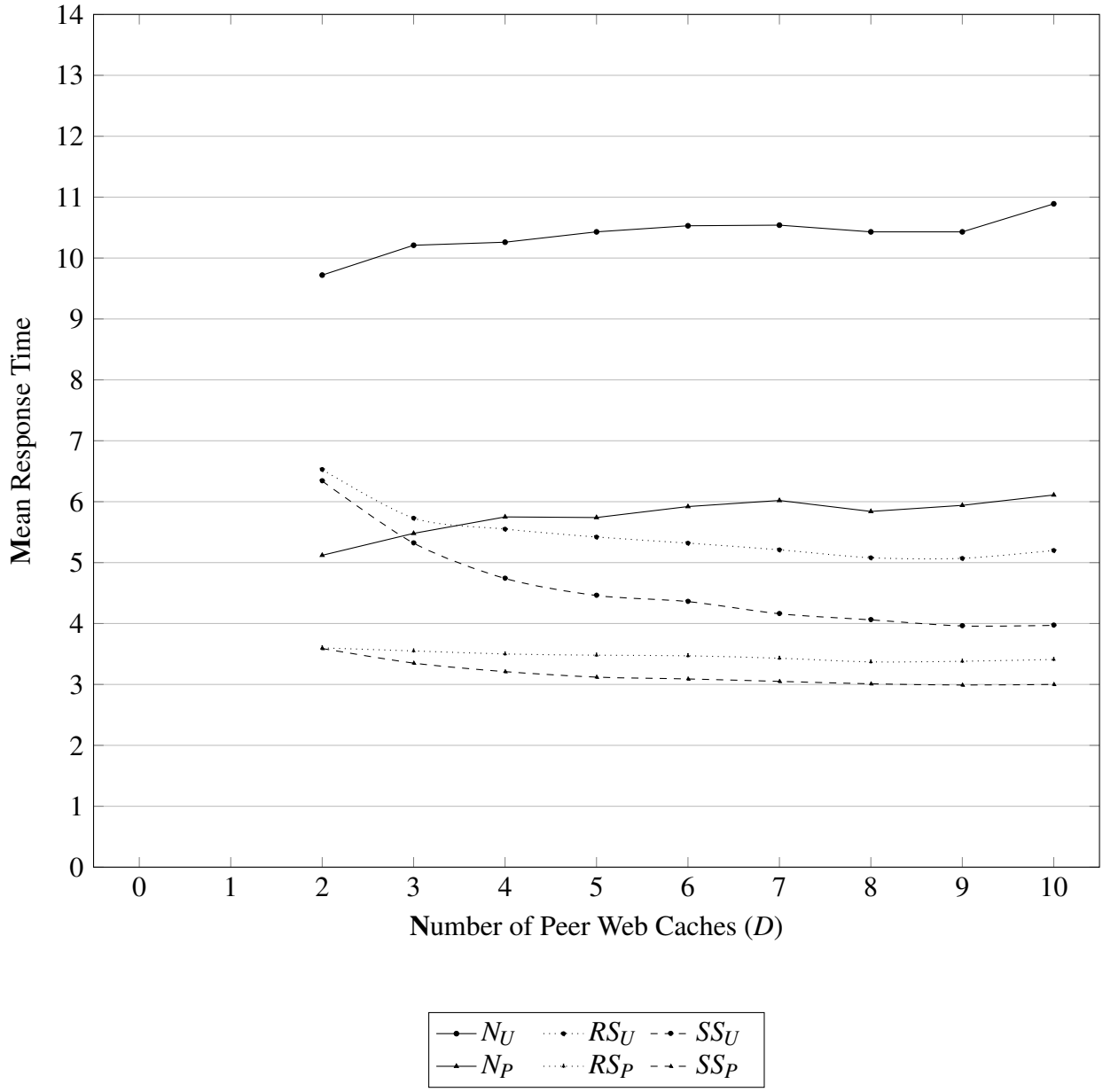


Figure 7: The Effect of the Number of Peer Caches (D) on the Mean Response Time for Unified and Partitioned Web Caching: $\Theta = 3$, $\rho_{N_U} \approx 0.70$

to higher response times.

As additional Web caches are introduced in our load balancing environments, the Short (SS_U, SS_P) and Random (RS_U, RS_P) algorithms tend to lead to a decrease in mean response time. For each additional cache added to the system, the system is provided with more opportunities to attempt to balance the workload in the system, leading to a decrease in mean response time. Again from Figure 8, when D = 10, SS_U and RS_U have a decrease in mean response time of 57.1% and 44.3% respectively with regards to the system without load balancing (N_U). The algorithms seem to follow the same pattern with the Short algorithm outperforming the Random algorithm by 30.0%

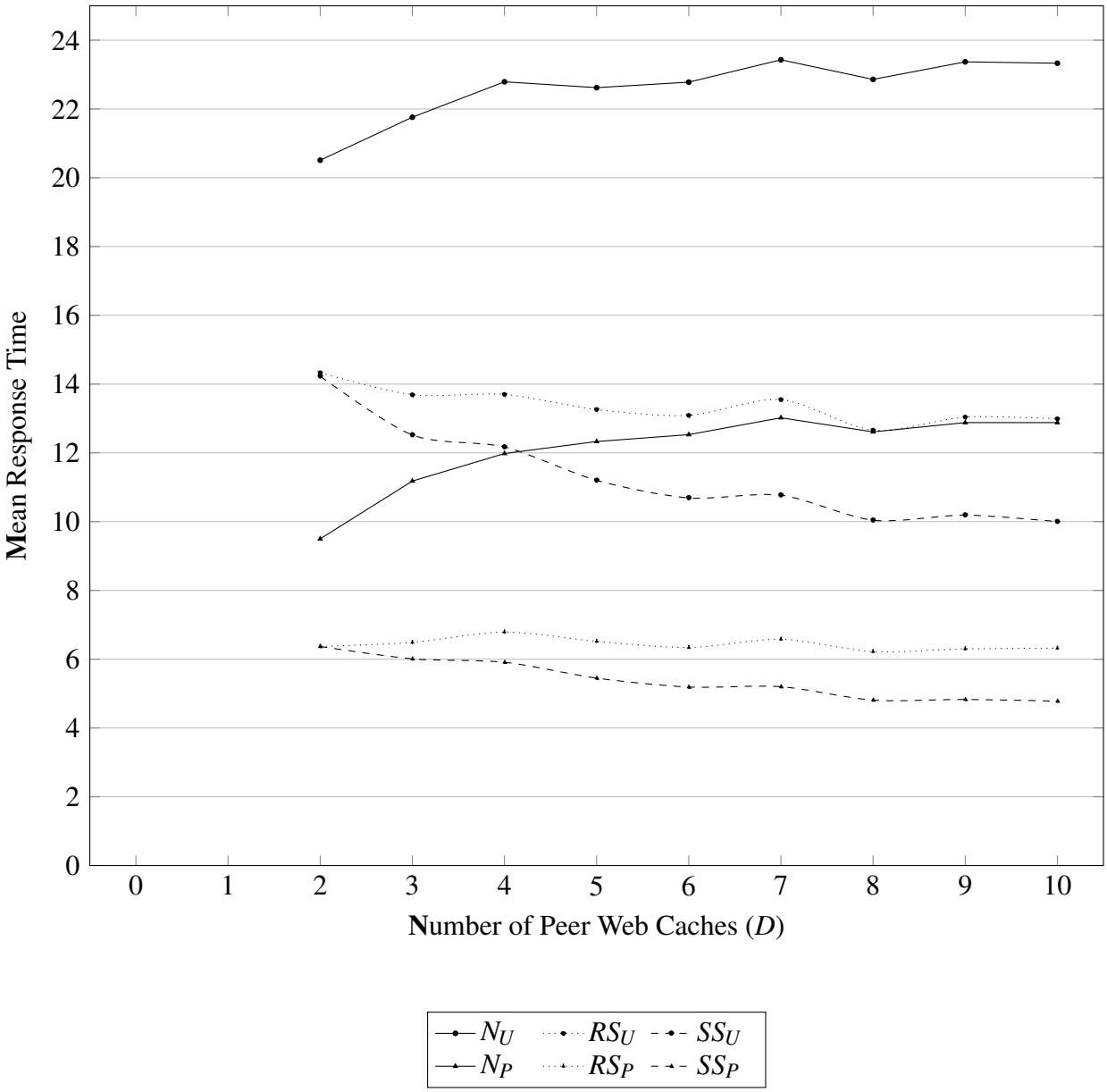


Figure 8: The Effect of the Number of Peer Caches (D) on the Mean Response Time for Unified and Partitioned Web Caching: $\Theta = 3$, $\rho_{N_U} \approx 0.85$

with respect to response time. However, it is important to note that Short algorithm would incur more overhead than Random algorithm due to the need to collect queue lengths from peer caches.

Partitioned Web caching system tends to again outperform a unified Web caching system. As we observe from Figure 8, there is a 52.2% performance difference between SS_U and SS_P and a 51.3% performance difference between RS_U and RS_P when $D = 10$. From these results, we conclude that a Web caching system with load balancing tends to scale gracefully relative to a Web caching system without load balancing as the number of peer Web caches (D) increases.

4.4. Cache Size

In Figure 9, we investigate the effects of the cache size (C) on the mean response time for our unified and partitioned Web caching systems. As expected, we observe that there is a dramatic decrease in the mean response time when the Web cache size is greater than 0. As soon as Web caching is introduced, there is an immediate performance benefit that can be observed, N_U has a performance increase of 62.9% when $C = 5\%$ when compared to when $C = 0\%$. As the Web cache size increases (C), the mean response time tends to decrease. However, we observe that after the initial dramatic decrease in mean response time, the system does not see the same large performance benefit as the cache size continues to increase. The performance improvements over time tends to decrease until mean response time plateaus. This occurs when the cache size is large enough to store most of the pages from the originating servers (an unlikely event but does provide a lower bound for the mean response time). Both algorithms follow the same trend and tend to outperform the systems without load balancing. For example SS_U and SS_P have a performance increase of 35.8% over N_U and N_P, while RS_U and RS_P have an increase of 33.7% over N_U and N_P when $C = 100$.

We also observe that as we increase the cache size, the partitioned system collapses into a unified system. With ample cache space, both storage models achieve the same level of performance. We find that smaller values of cache size (as long as it is greater than 0), tends to benefit partitioned storage over unified storage(i.e. when $C = 5\%$, the mean response time for N_P is 53.7% lower than that of N_U). It is again important to observe that irrespective of the value of the cache size, load balancing tends to improve the performance of the system with respect to the mean response time.

5. Conclusion

The results from this study have shown that the load balancing algorithms in a distributed Web caching system can be effective from a performance standpoint. In fact, any of the algorithms we examined achieved a level of performance equal to or better than a system without load balancing. We also determined that both unified and partitioned systems scale well with respect to additional peer Web caches, with the performance gains actually increasing as additional caches are introduced (unlike the system without load balancing (N_U) which degrades with additional Web caches). The use of the partitioned storage system has also been shown to increase the performance benefits of the load balancing algorithms in the Web caching environment. Performance benefits are seen even if a simple algorithm such as Random is incorporated. The benefits tend to increase with the use of state information (such as that seen with Short versus Random algorithms). In all of our cases, the use of load balancing in a distributed Web caching system tends to be much more desirable relative to a Web caching system without load balancing.

The research from this investigation has opened the door to a variety of potential extensions. A natural extension would be to utilize more state information from the requests; specifically, what page is being requested. For example, it may be beneficial to transfer a request (or multiple requests) to a Web cache that contains the requested page so as not to have to retrieve the page from the originating server. This will result in a cache hit for the local Web cache, thus increasing the hit rate of the cache at the same time as reducing the mean response time. As well, our system model was based on a distributed Web caching system, it may be possible to adapt our sender-initiated load balancing algorithms to a hierarchal Web caching system where caches are assumed to reside at various levels (i.e."distances") from the client (a receiver-initiated would not be appropriate for this

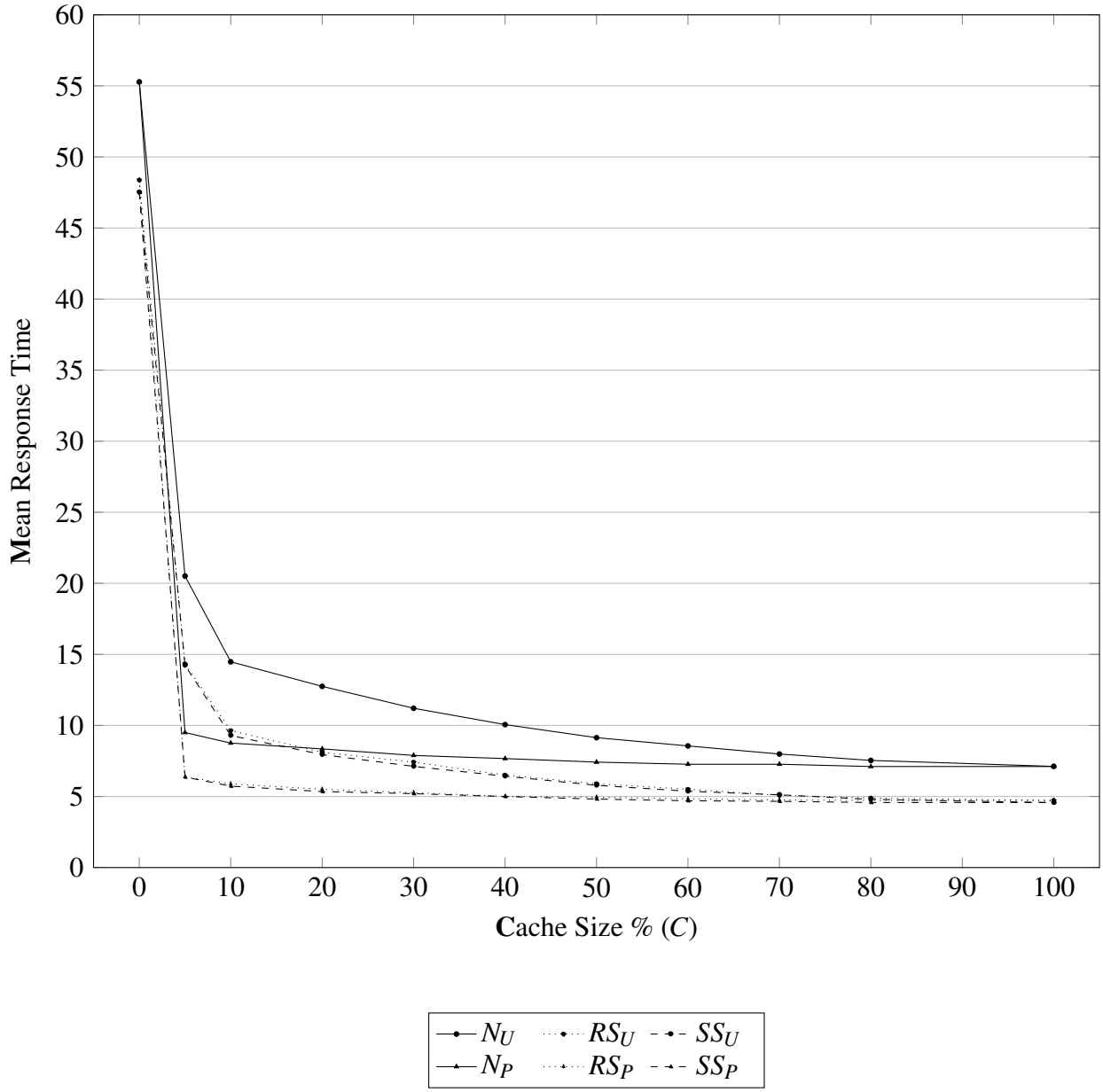


Figure 9: The Effect of Threshold Cache Size (C) on the Mean Response Time for Unified and Partitioned Web Caching: $\Theta = 3$, $D = 2$, $\rho_{N_U} \approx 0.85$

environment). Finally, our system model did not directly model the effects of overhead, such as the cost of transferring a request or the cost of collecting state information. It would be interesting to examine the effects of these overhead costs as they would likely impact some of the load balancing algorithms differently.

References

[1] W. Ali, S. M. Shamsuddin, and A. S. Ismail. A survey of web caching and prefetching. *Int. J. Advance. Soft Comput. Appl*, 3(1):18–44, 2011.

[2] Jay Chen and Lakshmi Subramanian. Interactive web caching for slow or intermittent networks. In *ACM DEV-4 '13: Proceedings of the 4th Annual Symposium on Computing for Development*, pages 1–10. ACM, 2013.

[3] Ali Raza amd Yasir Zaki, Thomas Pötsch, Jay Chen, and Lakshmi Subramanian. Extreme web caching for faster web browsing. In *SIGCOMM '15: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, pages 111–112. ACM, 2015.

[4] Sunghwan Ihm and Vivek S. Pai. Towards understanding modern web traffic. In *ACM SIGMETRICS Performance Evaluation Review*, volume 39, pages 335–336. ACM, 2011.

[5] W. Feng, S. Man, and G. Hu. Markov tree prediction on web cache prefetching. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pages 105–120. Springer, 2009.

[6] R. T. Hurley, W. Feng, and B. Y. Li. Partitioning in distributed and hierarchical web-caching architectures: A performance comparison. *Proc. of the16th International Conference on Computer Applications in Industry and Engineering, Las Vegas, Nevada, USA*, pages 11–13, Nov 2003.

[7] W. Feng, R. T. Hurley, and Z. Tan. Increasing web cache hit rate by dynamic load partitioning. *Proceedings of the 7th Joint Conference on Information Sciences, Cary, North Carolina, USA*, pages 405–409, September 2003.

[8] D. Ferrari and S. Zhou. An empirical investigation of load indices for load balancing applications. Technical report, DTIC Document, 1987.

[9] S. Hofmeyr, C. Iancu, and F. Blagojević. Load balancing on speed. In *ACM Sigplan Notices*, volume 45, pages 147–158. ACM, 2010.

[10] G. Barish and K. Obraczke. World wide web caching: Trends and techniques. *IEEE Communications magazine*, 38(5):178–184, 2000.

[11] Rohan Gandhi, Hongqiang Harry Liu, Y. Charlie Hu, Guohan Lu, Jitendra Padhye, Lihua Yuan, and Ming Zhang. Duet: cloud scale load balancing with hardware and software. In *SIGCOMM '14: Proceedings of the 2014 ACM conference on SIGCOMM*, pages 27–38. ACM, 2014.

[12] M. E. Soklic. Simulation of load balancing algorithms. *ACM - SIGCSE Bulletin*, December 2002.

[13] D. L. Eager, E. D. Lazowsk, and J. Zahorjan. A comparison of receiver-initiated and sender-initiated adaptive load sharing. *ACM 0-89791-169-5/85/007/0001*, 1985.

[14] B. Plumley. An investigation of load balancing in a distributed web caching systems. Master's thesis, Trent University, 2015.

[15] R. T. Hurley and B. F. Hircock. Benefits of vertical file migration in a horizontal file migration system. *IASTED International Conference on Parallel and Distributed Computing Systems, MIT, Boston, MA, USA*, pages 365–370, Nov. 3-6 1999.

[16] M. F. Arlitt and C. L. Williamson. Web server workload characterization: The search for invariants. In *ACM SIGMETRICS Performance Evaluation Review*, volume 24, pages 126–137. ACM, 1996.

[17] H. Bahn, H. Lee, and S. H. Noh. Replica-aware caching for web proxies. *Computer Communications Journal, Elsevier Science*, 2001.

[18] R. T. Hurley and B. Y. Li. Effects of dynamic content on web caching. *ISCA PDCCS*, 2008.

[19] B. Y. Li. An investigation of partitioned caching in the world wide web. Master's thesis, Trent University, 2002.

A case study of time step validation strategy and convergence method for oscillation numerical simulation in a heat transfer process

J. Zhu, †X.H.Zhang

School of Energy, Soochow University, China.
†Corresponding author: xhzhang@suda.edu.cn

Abstract A convergence identification method for oscillation numerical simulation is proposed, the numerical solutions can converge at the inflection point with respect to the time steps. In this way, it is possible to determine which time step is the appropriate convergence solutions, it can be ensured to obtain the accurate solution as much as possible, the results of the numerical experiments are presented and they confirm analytical predicts. In addition, an algorithm to verify the appropriate time step is suggested also, first use one time step to compute a case until it reaches a stable periodic solution; then sequentially reducing time step to check its convergence. The feasibility of the proposed method is further verified via its applications to the case study of the combined natural and MHD convection in a Joule-heated cavity using the finite volume methods. It is found that the two approaches have the same results and can judge the validity of the time step in computation, this might accurately predict the fluid flow and heat transfer.

Keywords : oscillation numerical simulation, time step, convergence, algorithm

Nomenclature

A	amplitude
g	gravitational acceleration [m/s^2]
Ha	Hartmann number
L	enclosure height [m]
Pr	Prandtl number
Ra	Rayleigh number
T	temperature [K]； period
u	x-velocity component [m/s]
U	dimensionless x-velocity component
v	y-velocity component [m/s]
V	dimensionless y-velocity component
W	enclosure width [m]
x	x coordinate [m]
X	dimensionless x coordinate
y	y coordinate [m]
Y	dimensionless y coordinate

Greek symbols

θ	dimensionless temperature
σ	electrical conductivity [ms/s]
τ	dimensionless time
φ	potential difference [V]

1. Introduction

The most common approach for approximating the derivatives is the finite difference methods due to their accuracy, stability, and easy of implementation. Different types and orders of finite difference methods are available to model the diffusions and the convection derivatives, and this method is widely used in the fluid flow and heat transfer field. The improvement in computer capabilities, especially in memory and speed, has made an accurate numerical predictions of the complex fluid flow and heat transfer cases.

However in the scientific computing, there are many sources of uncertainty including the model inputs, the form of the model, and poorly characterized numerical approximation errors [1]. In fact, all of these sources of uncertainty can give false results.

Therefore, several lines of researches have been proposed in the literature to solve these serious problems. One of them is for the scheme and algorithm, for example, a scheme called SGSD (Stability Guaranteed Second Order Difference Scheme) is proposed [2] which is absolutely stable and possesses at least second-order accuracy. A new weighted essentially non-oscillatory (WENO) procedure for solving hyperbolic conservation laws is proposed on uniform meshes [3]. An algorithm called IDEAL algorithm was conducted by Sun et al. [4] [5] in the IDEAL algorithm where the inner doubly iterative processes for the pressure equation are used to almost completely overcome the two approximations in the SIMPLE algorithm. Furthermore , a general method to remove the numerical instability of partial differential equations was presented by [6].

The previous studies on the computation of the discretization equation mainly focused on the finite difference method, the issue of consistency still remains several problems far from totally solved in the actual numerical computation, most transient simulations consist of a considerable number of time steps, therefore, the choice of the time step size is critical for the efficiency of the transient simulations. An alternative approach is to focus on the numerical solution and computer round-off errors. It is well known that Von-Neumann established that discretized algebraic equations must be consistent with the differential equations, and must be stable in order to obtain a convergent numerical solutions for the given differential equations. Eça and Hoekstra [7] offered a procedure for the estimation of the numerical uncertainty of any integral or local flow quantity as a result of a fluid flow computation. Teixeira *et al*. [8] explored the time step sensitivity of non-linear atmospheric models and illustrated how solutions with small but different time steps will decoupled from each other after a certain finite amount of the simulation time. Li [9] carried out systematic investigations on the sensitivity of the numerical solutions of non-linear ordinary differential equations (ODEs). A review on the computational uncertainty principle could be seen in Li and Wang [10]. Wang *et al.* [11] developed a high-performance parallel Taylor solver to do the Lorenz equations computation.

Depending on the study and analysis of those representative works mentioned above, the present paper finds that most of them are concerned to the Lorenz system, namely the ordinary differential equations. We know that the governing equations on the fluid flow and heat transfer problems are usually partial differential equations (PDEs). It can be proved mathematically that linear differential equations should have unique solutions, the situation is more complex for non-linear PDE's, and ,in some cases the numerical solutions are not chaotic but are still spurious and time periodic, making it difficult for the researchers to determine if the solution is representative of the true physics of the problem or not? Explicit methods have been coupled with spatial variable and time step for a particular problem to obtain simulations with a low computational cost, efforts have been made to identify the correct time step from the physical viewpoint, the time step size is restricted by stability reasons to fulfill the Courant–Friedrichs–Lewy (CFL) condition, while, few attentions on the time step with fully implicit scheme which is unconditionally stable in the non-steady

computation and few time step with fully implicit scheme validations are studied but on the grid independency, meanwhile, there is not a suitable convergence method for the oscillation simulation.

So, this is the motivation of our work, where a suitable convergence method for the oscillation simulation and an algorithm were established to overcome previous convergence method shortcoming, extensive calculations were performed and examined to a Joule heating flow in order to confirm the two independent methods.

2. Convergence method and algorithm

The rigorous convergent criterion has only been established for the equilibrium solution: the difference between two consecutive iterations is less than a predetermined value is considered to be convergence, the iteration process convergence to one steady-state solution. This is only applicable for the system which has the static values as time approaches to infinity. Therefore, it is no appropriate to use convergent criterion aforementioned above in the oscillation numerical simulation cases.

A convergence method in the numerical simulation is addressed here which states that if the system is a stable oscillation system, as the time step decreases, the calculated values (including velocity and temperature) should be monotonous, theoretical speaking, at the same point in the same moment time, the reason is that the even smaller truncation error can be achieved because of decreasing time step size for the fixed grid spacing. It is desirable to use the smallest time step possible throughout the computation, the difference of the computation values with different two time steps at the same space point in the same moment time is less than a predetermined value is considered to be the convergence solution. But in practical simulation, the computer is finite precision, so as the time step decreases more, the round-off error is primary. Consequently, the smallest time step cannot be viewed as the solution approached to the correct one, the solution properties at the same point in the same moment time as the time step is refined is non-monotonic. Therefore, the numerical solutions can converge at the inflection point with respect to the time step, in this way, it is possible to determine which time step is the appropriate convergence solutions, and it can be ensured to obtain the accurate solution as much as possible. This is the convergence concept for the stable oscillation case.

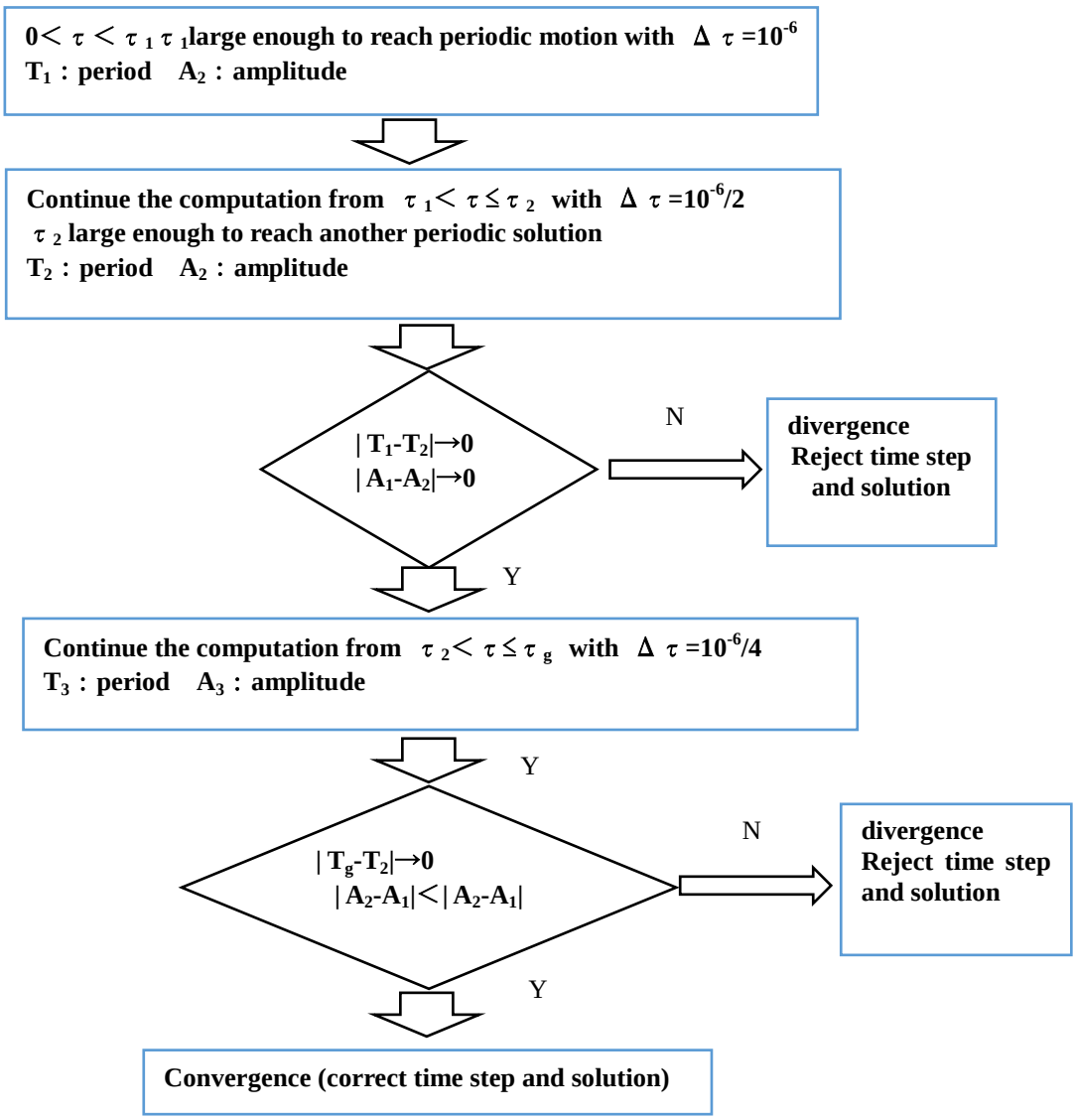


Figure 1 flow chart of time step identification

A practical algorithm of judging the accuracy of the above analysis for oscillations results is suggested below (see Figure 1 for more details), first we use one time step to compute a case until it reaches a stable periodic solution; then sequentially reducing time step to check its convergence, for example, the time step equals to $\Delta\tau = 10^{-6}$:

Step 1 From $0 < \tau \le \tau_1$, choose of τ_1 is large enough for the computational result reached a periodic motion whose period is T1 and the amplitude is A1. The purpose of this period is to lock the numerical solution into a special mode, we hope that the truncation error is sufficient to alter the initial condition and leads to a special solution among many possibility.

Step 2 Continue the computation from $\tau_1 < \tau \le \tau_2$ with $\Delta\tau = 10^{-6}/2$. τ_2 is large enough for the computational results to reach another periodic solution, its period is T2, and the amplitude is A2. If (T2=T1), and A2 is close to A1, then the solution may have some meaning.

Step 3 Continue the computation from $\tau_2 < \tau \le \tau_3$ with $\Delta\tau = 10^{-6}/4$. If (T3 = T2) and A3-A2 is smaller than A2-A1, then the results have chance to converge. Then , return to the other time step, repeat the above steps until time step corresponding the convergence of the

solution is reached. The alternative convergence method and choosing the correct time step size algorithm for the solution of the oscillation numerical simulation are more accurate than the previous convergent method, and this is more general approach. In the next section, the method presented above will be validated and analyzed by the numerical simulation test.

3. Numerical experiments

In the previous section, the convergence approach and algorithm of indentifying adequate time step were discussed. In this section, we investigate the convergence approach using an example of case study.

3.1 Physical model and the problem formulation

The problem under consideration is the combined natural and MHD convection, as demonstrated in Zhang [12], the system considered is shown in Figure 2. The fluid contained in the rectangular pool is heated by a pair of vertical electrodes, which are assumed to be isopotential surfaces with an externally applied potential difference of ϕ_0 across them. The bottom boundary is assumed to be electrically insulated. In the present study, low frequency alternating current sources are considered for Joule heating. All the boundaries of the cavity are solid–fluid interfaces, which can be treated as no-slip and no-penetration boundaries. The upper boundary of the liquid cavity is an isothermal surface at $T = T_0$, while the rest of the boundaries are assumed to be thermally insulated. The aspect ratio of the pool is set to be W:L=2:1.

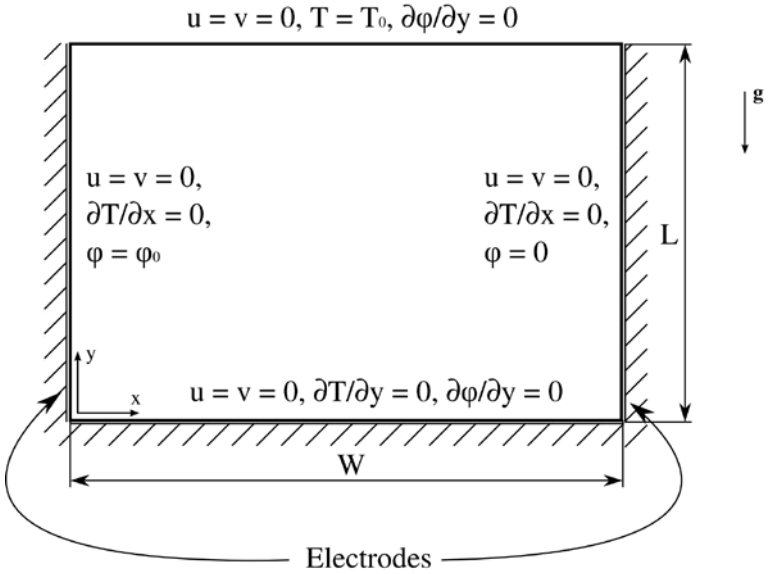


Figure 2 Schematic of the system under consideration

In the present model, flow is simulated as a two dimensional phenomenon with the following assumptions or simplifications: a) the fluid is Newtonian, incompressible and the flow is laminar; b) the effect of temperature on fluid density is expressed adequately by the Boussinesq approximation; c) the local electrical conductivity is independent of the thermal field.

The governing equations presented in Zhang [12] will not be repeated here just for the brevity. In order to guarantee both the numerical stability and solution accuracy, the SGSD scheme [2] is employed for the discretization of the convection terms, which is absolutely stable and adaptive. The SGSD scheme can automatically choose a different difference scheme according to the available local field information in difference space or time. The diffusion terms are discretized by the central difference scheme. The IDEAL [4] [5] algorithm is

adopted which exists inner doubly iterative processes for the pressure equation. The coupling between the velocity and pressure is fully guaranteed, greatly enhancing the convergence rate and the stability of the iteration process. While dealing with the time-dependent physics problem for the un-steady state governing equations. It has been theoretical analyzed that the fully implicit scheme is unconditionally stable for SGSD scheme in un-steady convection diffusion equation, it is not repeated here for simplicity.

It must be noted that, the Rayleigh number and the Hartman number which are investigated here are smaller than the critical Rayleigh number and the critical Hartman number respectively. The zero initial conditions are set for velocity and temperature fields.

Grid sensitivity analysis is performed and the accuracy of the numerical procedure is further validated by comparing predicted results with the solutions obtained by Sugilal [13] on the same test case, the present procedure adequately predicts the flow and heat transfer inside the system considered.

3.2 Numerical Results

The main goal of the present study is not only to obtain the accurate solution but also to investigate its stability. The computational efficiency (low demand on CPU time) of the present study is not considered here.

3.2.1 Time step validation for Pr=1 ,Ra=15000 and Ha=0

We perform the numerical simulations for four values of the time step ($\Delta\tau$) ranging from $\Delta\tau=10^{-3}$ to $\Delta\tau=10^{-6}$, while keeping the other relevant parameters fixed (i.e., Ra =15000, Pr =1 and Ha =0) . This approach is aimed to evaluate the sensitivity of the time step. All the computations start from a zero field initialization and are stopped at τ = 4. Throughout the simulations, the time histories of the dimensionless temperature and velocity components are recorded at a monitoring point (X,Y) = (0.25,0.483) inside of the cavity. All the simulation results exhibit a common behavior as depicted in Fig. 3, where the dimensionless temperature reaches a steady state of the solution as the time increases, and it has a similar behavior for the velocity components. The solution for a particular time step is considered converged when the iteration makes no change to the solution in any of the variables *U*, *V* or θ**.** This convergence method is not necessarily the best, but it is a commonly used.

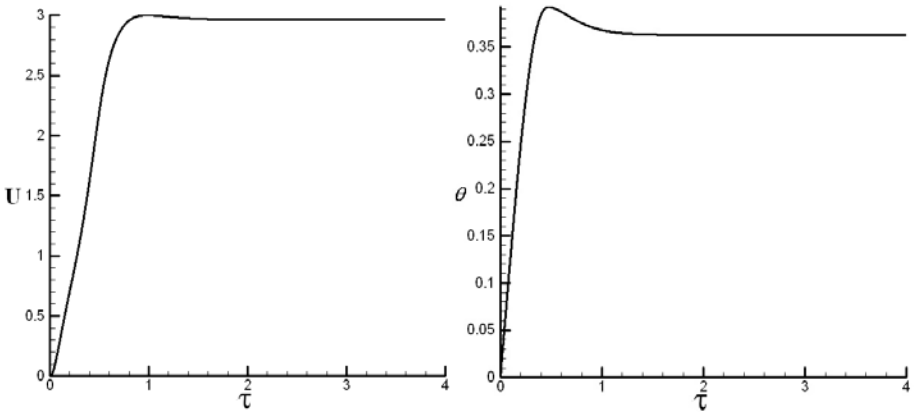


Fig. 3 Evolution of *U*-velocity (left) and temperature (right) in monitoring point (X,Y=0.25,0.483) of the cavity for $\Delta\tau$=0.0001

The only difference in Table 1 is the momentum residual ,we find that as the $\Delta\tau$ decrease from 0.001 to 0.0001, the momentum residual decreases. While when $\Delta\tau$ decreases more the momentum residual increases, this can be explained that the truncation error is smaller

when $\Delta\tau$ decreases，while when $\Delta\tau$ decreases more the round-off error is bigger and the more accurate time step is 10^{-4}.

Table 1. Residuals , dimensionless temperature at τ = 4 at a monitoring point (*X*,*Y*) = (0.25,0.483)

Case	Time step	Mass residual	Momentum	Residual
a	**0.001**	**1.2822E-09**	**1.7986E-02**	**8.5379E-03**
b	**0.0001**	**3.3605E-13**	**4.6960E-06**	**2.9421E-06**
c	**0.00001**	**2.6585E-13**	**2.2607E-05**	**1.5130E-05**
d	**0.000001**	**3.3216E-13**	**4.2578E-04**	**2.6565E-04**

3.2.2 Time step validation for Pr=0.01, Ra =15000 and Ha=0

We perform the numerical simulations for four values of the time step ($\Delta\tau$) ranging from $\Delta\tau$= 10^{-4} to $\Delta\tau$=10^{-7}, while keeping the other relevant parameters fixed (i.e., Ra =15000, Pr =1 and Ha =0) . All the computations start from a zero-field initialization and are stopped at τ=1. Throughout the simulations, the time histories of the dimensionless temperature and velocity components are recorded at a monitoring point (*X*,*Y*) = (0.25,0.483) as shown in Fig.4.

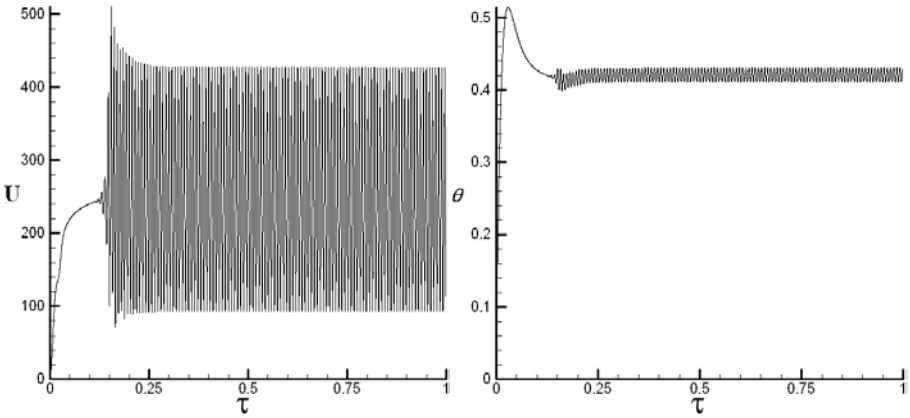


Fig. 4 Evolution of *U*-velocity (left) and temperature (right) in a monitoring point (X, Y) =(0.25 ,0.483) of the cavity for $\Delta\tau$=0.0001

The time history of the dimensionless temperature(θ) and the time history of the dimensionless *x*-velocity component (*U*) exhibit a common behavior in different time steps for all the cases examined. It is worthwhile to note that the sensitivity to the initial conditions associated with a set of non-linear differential equations is a reflection of a characteristic of a non-linear physical system, to pursue this property more fully. It can be verified by a non-zero field in procedure at τ=0 whose components take random values from -1 to 1 generated by the computer. The results keep the same as those of zero initial conditions. It should be noted that the computation for Rayleigh number (Ra=15000) is less than the critical Rayleight number, verifies the system is to make stable oscillation.

The question is which time step corresponds to the accurate solution and how to identify the convergence, while the method of considering convergence when the monitoring value makes a small change cannot be applied in this case, as the θ and *U* are oscillated with the time. These results suggest that there is no apparent convergence of comparing the numerical

values during the iterations. It can be verified with proposed method in section 2 by the numerical simulation results below. Fig.5 shows that the *V*-velocity and temperature are monotonically decrease as the time step decreases. The truncation errors become the primary, on the contrary when $\Delta\tau$ is 10^{-6}, as the time step decreases, the *V*-velocity monotonically increases. This is because the round-off errors become the primary errors. In order to get more accurate results, the correct time step should be 10^{-6}, where in this case the residuals are relatively smaller (see Table. 2), so the more accurate solutions can be obtained. From the experiment we validate the convergence analysis method.

Table 2. Comparisons of the mass and momentum residuals

Case	Time step	Mass residual	Momentum residual	
A	**0.0001**	**6.4119E-04**	**1.0649E-02**	**1.8723E-02**
B	**0.00001**	**8.6406E-05**	**6.9381E-03**	**9.3104E-03**
C	**0.000001**	**2.5270E-06**	**2.1513E-02**	**2.0302E-02**
D	**0.0000001**	**2.3201E-08**	**7.4870E-03**	**9.3305E-03**

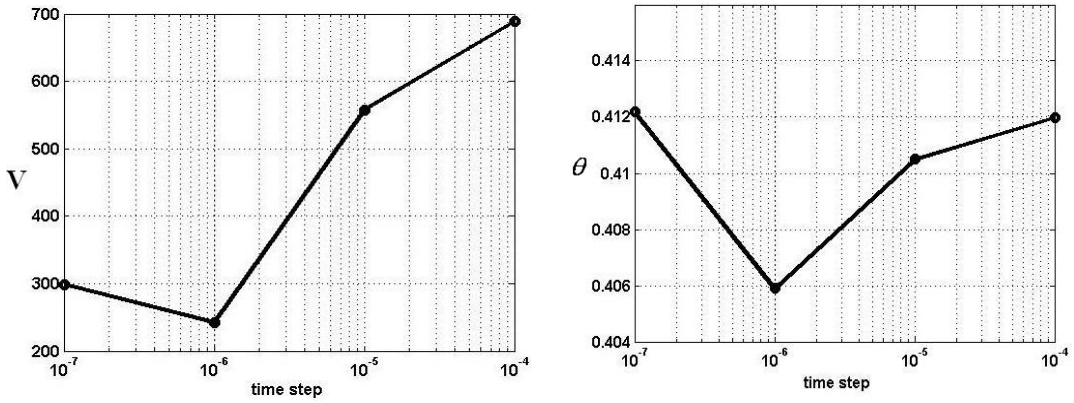


Fig. 5. Comparison of *V*-velocity and temperature calculated by different time steps at the same moment time (τ=1) in a monitoring point (X,Y=0.25,0.483) of the cavity

A practical algorithm of judging the accuracy for oscillations results in section 2 is implemented, the experiment results for different time steps are listed in Table 3 which confirm our analysis, and the correct time step should be 10^{-6}.

Table 3. Periods and amplitudes of periodic oscillation for each $\Delta\tau$

time step/ $\Delta\tau$	periods of the periodic oscillations/T	amplitudes of the periodic oscillations/A
10^{-4}	**0.00765**	**0.0179**
10^{-4}/2	**0.00487**	**0.01424**
10^{-4}/4	**0.003437**	**0.010512**
10^{-5}	**0.002563**	**0.00823**
10^{-5}/2	**0.002287**	**0.00728**
10^{-6}	**0.002055**	**0.008**
10^{-6}/2	**0.002007**	**0.0081**
10^{-6}/4	**0.002114**	**0.00814**

3.2.3 Time step validation for Ha = 7000 and Ra = 0

The numerical simulations for four values of the time step are performed where , $\Delta\tau$, ranging from $\Delta\tau = 10^{-4}$ to $\Delta\tau=10^{-7}$, while keeping the other relevant parameters fixed (i.e., Ha = 7000, Pr =0.01 and Ra =0) .All the computations start from a zero-field initialization and are stopped at τ=0.2. Throughout the simulations, the time histories of the dimensionless temperature and velocity components are recorded at a monitoring point (*X*,*Y*)=(0.25,0.483) as shown in Fig. 6. The computed *U* results at a monitoring point (X=0.25,Y=0.483) take the oscillation in the average of 400 , 460 and 100 for the three different time steps respectively. It can be seen that, the solutions are apparently quite close to each other for the different time steps except $\Delta\tau$=0.0000001.

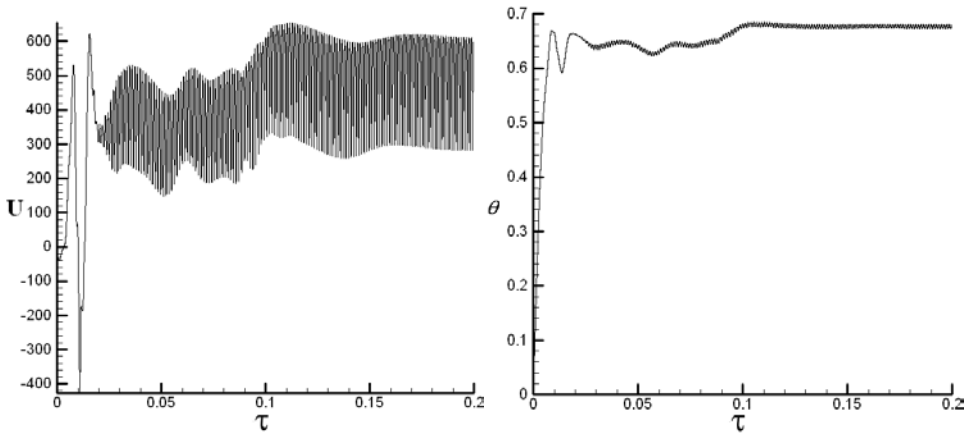


Fig. 6.Evolution of *U*-velocity (left) and temperature (right) in a monitoring point (X,Y=0.25,0.483) of the cavity for $\Delta\tau$=0.00001

The non- zero field in procedure at τ = 0 whose components take random values from -1 to 1 which are generated by the computer is implemented, where the results keep the same as those of the zero initial condition. This verifies the system is not non-linear at present computation conditions. It can be seen from Fig. 7 , that the moment time records increase monotonically with decreasing time step to $\Delta\tau = 10^{-6}$, then it decreases with decreasing time step furthermore. The optimal time step should be 10^{-6}, and the residuals are relatively small one (Table 4) in this case. Similarly, the method stated in section 2 for the selection time step is utilized again with sequentially reducing $\Delta\tau$ by factor two and comparison of the results. It can be got clearly that the correct time step should be 10^{-6}.

Table 4. Comparisons of the mass and momentum residuals

Case	Time step	Mass residual	Momentum residual	
A	**0.0001**	**1.0177E-03**	**1.9489E-02**	**1.1435E-02**
B	**0.00001**	**1.1858E-04**	**5.5880E-03**	**3.5529E-03**
C	**0.000001**	**2.2849E-06**	**4.1785E-03**	**4.8898E-03**
D	**0.0000001**	**3.2633E-08**	**4.1209E-03**	**4.6373E-03**

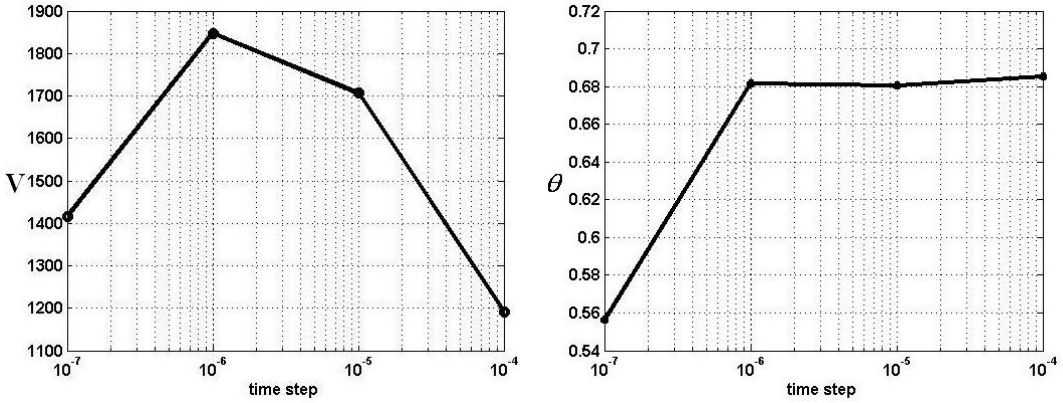


Fig. 7. Comparison of *V*-velocity and temperature calculated by different time steps at the same moment time (τ=0.2) in a monitoring point (X,Y=0.25,0.483) of the cavity

3.2.4 Time step validation for Pr = 0.01, Ha=7000 and Ra=15000

The time-periodic solutions are predicted shown in Fig. 8 which reports the time dependent behavior of the dimensionless velocity and temperature at the monitoring point (X,Y=0.25,0.483) of the cavity. Fig.9 shows that the oscillations start at $\tau \sim 0.08$ and the computed *U* at a monitoring point takes the oscillatory center value of 230.

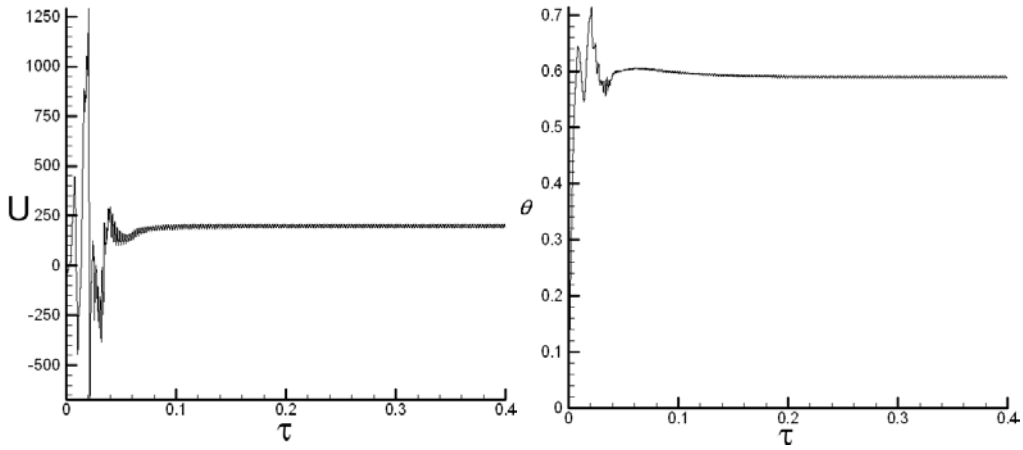


Fig. 8.Evolution of *U*-velocity (left) and temperature (right) in a monitoring point (X,Y=0.25,0.483) of the cavity for $\Delta\tau$ =0.00001

We find that the results are different in different time steps as shown in Table 5. For cases A and B, the time step width is of the order of 10^{-3} and 10^{-4}, residuals for momentum equation and mass equation are of the order of 10^{-4}. The time step width is of the order of 10^{-5} for case C, and the residuals are of the order of 10^{-5}.For cases D and E, the considered smaller time steps are, $10^{-6}$and 10^{-7} respectively, the residuals of the order of 10^{-2}.Such small time step width gives much larger residuals, the different truncation errors associated with different time-steps, in effect, lead to a series of residuals. A non-zero field in procedure at τ=0 whose components take random values from -1 to 1 which are generated by the computer is implemented and the experiment results are the same as the zero initial condition. Therefore, this confirms the system is not a non-linear system.

The convergence of the solution properties as the time step refined is no monotonically at the same zero initial condition, this can be seen from Fig. 9, where the moment time records increase monotonically with decreasing time step to $\Delta\tau = 10^{-5}$, then it decreases with decreasing time step. The correct time step should be 10^{-5}. In this case the residuals (see Table

5) are the smallest one, accuracy of the solution can be obtained, and the total errors keep in an admissible bound. Consequently, we can also check these time steps as the step stated in section 2 by sequentially reducing ($\Delta\tau$) by factor two. It is found that the results obtained are in excellent agreement with the analytical numerical results, and it is confirmed that the optimal time step should be 10^{-5} .

TABLE 5. Comparisons of the mass and momentum residuals

Case	Time step	Mass residual	Momentum residual	
A	**0.001**	**2.9206E-4**	**4.7253E-4**	**7.6886E-4**
B	**0.0001**	**3.1185E-5**	**1.6895E-4**	**1.5762E-4**
C	**0.00001**	**2.5432E-6**	**7.2870E-5**	**5.7573E-5**
D	**0.0000001**	**2.4611E-6**	**1.8177E-2**	**1.4063E-2**
E	**0.00000001**	**3.6968E-8**	**2.9424E-2**	**2.0477E-2**

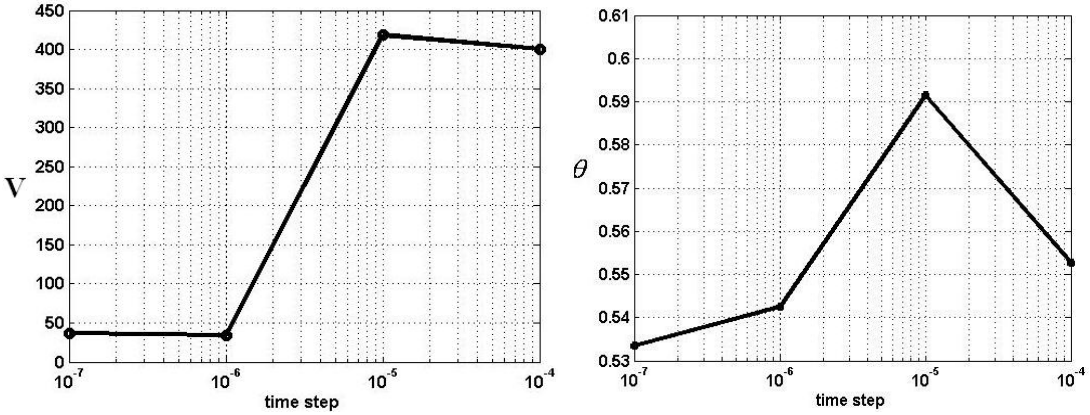


Fig. 9. Comparison of *V*-velocity (left) and temperature (right) calculated by different time steps at the same moment time ($\Delta\tau$=0.4) in a monitoring point (X, Y=0.25,0.483) of the cavity

4. Conclusions

The convergence method in the numerical simulation provided that the system is stable oscillation is present in the present paper , where the solution properties at the same point in the same moment time with refined time steps are non-monotonic for the stable oscillation model. So, the numerical solutions can converge at the inflection point with respect to the time step, therefore in this way it is possible to determine which time step is the appropriate convergence solution. In order to obtain the accurate solution as much as possible, the results of the numerical experiments are presented and they confirm our theoretical predictions. Therefore, an algorithm to verify the appropriate time step is suggested. First use one time step to compute a case until it reaches a stable periodic solution; then sequentially reducing time step to check its convergence. The numerical accuracy of the proposed method has also been demonstrated via its application to more complex two-dimensional Joule heating flow problem. The feasibility of the proposed method is further verified. It is found that the results obtained in all the test cases with the suggested algorithm are in excellent agreement with the analytical as well as the established numerical results, underlining the high validity of the method. The new methods are somewhat more complex and the accuracy of the results is greatly improved. Meanwhile, the proposed methods are considered universal and can be applied to other unsteady computation engineering calculations.

Acknowledgments

This work is supported by the National Science Foundations of China. (51176132).

References

[1] Roy, C.J. and Oberkampf, W.L. (2011), A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, Comput. *Methods Appl. Mech. Engrg.,* Vol.200, pp.2131–2144.

[2] Li, Z.Y. and Tao, W.Q. (2002) A new stability-guaranteed second-order difference scheme, *Numer. Heat Transfer.* Part B. Vol.42 ,pp.349–365.

[3] Abedian,R., Adibi,H. and Dehghan. M. (2014)A high-order symmetrical weighted hybrid ENO-flux limiter scheme for hyperbolic conservation laws, *Computer Physics Communications,* Vol.185,pp.106–127.

[4] Sun, D. L. Qu.Z.G., He, Y. L. and Tao. W. Q., (2008a) An Efficient Segregated Algorithm for Incompressible Fluid Flow and Heat Transfer Problems - IDEAL (Inner Doubly Iterative Efficient Algorithm for Linked Equations) Part I: Mathematical Formulation and Solution Procedure Numerical Heat Transfer, Part B, Vol.53 ,pp.1-17.

[5] Sun, D. L. Qu.Z.G., He, Y. L. and Tao. W. Q., (2008b)An Efficient Segregated Algorithm for Incompressible Fluid Flow and Heat Transfer Problems - IDEAL (Inner Doubly Iterative Efficient Algorithm for Linked Equations) II: Application Examples, *Numer. Heat Transfer Part B,* Vol.53 ,pp.18-38.

[6] Duchemin,L and Eggers,J., (2014) The Explicit–Implicit–Null method: Removing the numerical instability of PDEs, *Journal of Computational Physics,* Vol.263, No.15, pp.37–52.

[7] Eça, L. and Hoekstra. M., (2014)A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies. *Journal of Computational Physics,* Vol.262,No.1, pp.104–130.

[8] Teixeira, J.,Reynolds, C.A., and Judd, K., (2007)Time step sensitivity of nonlinear atmospheric models: numerical convergence, truncation error growth, and ensemble design, *J. Atmos. Sci.*, Vol.64 ,pp.175–189.

[9] Li. J., (2000)Computational uncertainty principle: meaning and implication. *Bull. Chin. Acad. Sci.* Vol.15, pp.428–430.

[10] Li, J. and Wang. S., (2008) Some mathematical and numerical issues in geophysical fluid dynamics and climate dynamics. *Commun. Comput. Phys.* Vol.3 ,pp.759–793.

[11] Wang, P., Li, J.,and Li. Q., (2012)Computational uncertainty and the application of a high-performance multiple precision scheme to obtaining the correct reference solution of Lorenz equations. *Numer Algor,* Vol.59, pp.147-159.

[12] Zhang, X.H., (2013) Unsteady numerical simulation of flow and heat transfer in a uniformly Joule-heated rectangular pool. *Nuclear Engineering and Design,* 255,pp.280– 286.

[13] Sugilal, G., Wattal, P.K., Iyer,K., (2005)Convective behaviour of a uniformly Joule-heated liquid pool in a rectangular cavity. *International Journal of Thermal Sciences 44* ,pp.915–925.

Hydromagnetic Nanofluids flow through porous media with thermal radiation, chemical reaction and viscous dissipation using spectral relaxation method

Nageeb A.H Haroun[1], S. Mondal[1,a)] and P. Sibanda[1]

[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal
Private Bag X01 Scottsville 3209, Pietermaritzburg, South Africa

[a)]Corresponding author and Presenting author: sabya.mondal.2007@gmail.com

Abstract.

We investigate the convective heat and mass transfer in magnetohydrodynamic a nanofluid through a porous medium over a stretching sheet subject to a magnetic field, heat generation, thermal radiation, viscous dissipation and chemical reaction effects. The effects of porosity, heat generation, thermal radiation, magnetic field, viscous dissipation and chemical reaction to the flow field are thoroughly explained for various values of the governing parameters. We have further assumed that the nanoparticle volume fraction at the wall may be actively controlled. Two types of nanofluids, namely Cu-water and Al_2O_3-water are studied. The physical problem is modeled using systems of nonlinear differential equations which have been solved numerically using the spectral relaxation method. Comparing the results with those previously published results in the literature shows excellent agreement.

Keywords: MHD Nanofluids flow; Porous media; Thermal radiation; Spectral relaxation method.

Introduction

Nanofluids are suspensions of metallic, non-metallic or polymeric nano-sized powders in a base liquid which are employed to increase the heat transfer rate in various applications. In recent years, the concept of nanofluid has been proposed as a route for increasing the performance of heat transfer liquids. Due to the increasing importance of nanofluids, there is an enormous amount of literature on convective transport of nanofluids and problems linked to a stretching surface. Today nanofluid are sought to have more range of applications in power generation in nuclear reactors, medical application, biomedical industry, detergency, and more specifically in any heat removal involved industrial applications. The ongoing work ever since then has extended to utilization of nanofluids in microelectronics, fuel cells, pharmaceutical processes, vehicle thermal management, domestic refrigerator, chillers, heat exchanger, nuclear reactor coolant, grinding, machining, space technology, defence and ships, and boiler flue gas temperature reduction. The majority of the previous studies have been restricted to boundary layer flow and heat transfer in nanofluids. Following the early work by Crane [1], Khan and Pop [2] were the first to work on nanofluid flow due to stretching sheet. A mathematical analysis of momentum and heat transfer characteristics of the boundary layer flow of an incompressible and electrically conducting viscoelastic fluid over a linear stretching sheet was carried out by Abd El-Aziz [3]. In addition, radiation effects on the viscous flow of a nanofluid and heat transfer over a nonlinearly stretching sheet were studied by Hady et al. [4]. Theoretical studies include, for example, modelling unsteady boundary layer flow of a nanofluid over a permeable stretching/shrinking sheet by Bachok et al. [5]. Rohni et al. [6] developed a numerical solution for the unsteady flow over a continuously shrinking surface with wall mass suction using the nanofluid model proposed by Buongiorno [7]. The effect of an applied magnetic field on nanofluids has substantial applications in chemistry, physics and engineering. These include cooling of continuous filaments, in the process of drawing, annealing and thinning of copper wire. Drawing such strips through an electrically conducting fluid subject to a magnetic field can control the rate of cooling and stretching, thereby furthering the desired characteristics of the final product. In other work, Jafar et al. [8] studied the effects of magnetohydrodynamic(MHD) flow and heat transfer due to a stretching/shrinking sheet with an external magnetic field, viscous dissipation and joule effects. Murthy and Singh [9] studied viscous dissipation on non-Darcy natural convection regime in porous media saturated with Newtonian fluid. In the past few years, convective heat and mass transfer in nanofluids has become a topic of major contemporary interest. In this paper we examine the study analyzed of magneto-hydrodynamics (MHD), heat and mass transfer in nanofluid flow over a stretching sheet subject to Porous media, hydromagnetic, heat generation, thermal

radiation, viscous dissipation, chemical re- action and Soret effects. The spectral relaxation method (SRM) was proposed by Motsa [10]. It is used to solve the governing partial differential equations numerically. This spectral relaxation method has been successfully applied to other problems of fluid mechanics and heat transfer. In this paper we discuss the fluid flow and heat transfer as well as highlight the strengths of the solution method.

Governing Equations

Consider the two-dimensional steady boundary layer flow of an incompressible heat and mass transfer nanofluid past a stretching sheet. The origin of the system is located at the slit from which the sheet is drawn. In this coordinate frame the x-axis is taken along the direction of the continuous stretching surface. The y-axis is measured normal to the surface of the sheet. It is assumed that the induced magnetic field is negligible in comparison to the applied magnetic field. It is assumed that the induced magnetic field, the external electric field and the electric field due to the polarization of charges are negligible in comparison to the applied magnetic field. In addition to these, the effects of chemical heating, agglomeration and sedimentation of nanoparticles are not included in the work.

The fluid is a water based nanofluid containing two different types of nanoparticles; Copper (Cu) and Alumina (Al_2O_3) nanoparticles. It is assumed that the base fluid and the nanoparticles are in thermal equilibrium and no slip occurs between them. The thermophysical properties of the nanofluid are given in Table 1.

With the above assumptions, the governing boundary layer equations of the nanofluid flow, the continuity, momentum, energy and the concentration fields with diffusion with radiation, heat generation, viscous dissipation and chemical reaction effects can be written in dimensional form as proposed by Tiwari and Das [11]

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

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \frac{\mu_{nf}}{\rho_{nf}}\frac{\partial^2 u}{\partial y^2} - \left\{\frac{\mu_{nf}}{\rho_{nf}}\frac{1}{K} + \frac{\sigma B_0^2}{\rho_{nf}}\right\}u, \tag{2}$$

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \alpha_{nf}\frac{\partial^2 T}{\partial y^2} + \frac{Q}{(\rho c_p)_{nf}}(T - T_\infty) + \frac{1}{(\rho c_p)_{nf}}\frac{16\sigma^* T_\infty^3}{3K^*}\frac{\partial^2 T}{\partial^2 y} + \frac{\mu_{nf}}{((\rho c_p)_{nf}}\left(\frac{\partial u}{\partial y}\right)^2, \tag{3}$$

$$u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = D_B\frac{\partial^2 C}{\partial y^2} + \frac{D_T}{T_\infty}\frac{\partial^2 T}{\partial y^2} - K_0(C - C_\infty), \tag{4}$$

Here q_r is the radiation heat flux given by

$$q_r = -\frac{4\sigma^*}{3K^*}\frac{\partial T^4}{\partial y} \tag{5}$$

where σ^* is the Stefen-Boltzmann constant and K^* is the Rosseland mean absorption coefficient. The temperature variation T^4 is expanded in a Taylor series expansion form. Neglecting higher order terms and expanding T^4 about T_∞ we obtain,$T^4 \cong 4T_\infty^3 T - 3T_\infty^4$. where u and v are the fluid velocity and normal velocity components along $x-$ and $y-$directions, respectively, $\mu_{nf}, \rho_{nf}, \alpha_{nf}$ are the effective dynamic viscosity of the nanofluid, nanofluid density and the thermal diffusivity of the nanofluid respectively. The boundary conditions for equations (1) - (4) are as follows

$$\begin{aligned} &u = ax,\ v = 0,\ T = T_{\mathrm{w}}(x) = T_\infty + H\left(\tfrac{x}{\omega}\right)^2, \\ &C = C_{\mathrm{w}}(x) = C_\infty + Q\left(\tfrac{x}{\omega}\right)^2 \quad \text{at} \quad y = 0, \\ &u \to 0,\ T \to T_\infty,\ C \to C_\infty \quad \text{as} \quad y \to \infty, \end{aligned} \tag{6}$$

where Q, H and a are constants, $a > 0$ and ω is the characteristic length. The effective dynamic viscosity of the nanofluid was given by Brinkman [14] as

$$\mu_{nf} = \frac{\mu_f}{(1-\phi)^{2.5}}, \tag{7}$$

where ϕ and μ_f are the solid volume fraction of nanoparticles and the dynamic viscosity of the base fluid. In equations (1) to (4) the heat capacitance of the nanofluid and the thermal conductivity of nanofluids restricted to spherical nanoparticles is approximated by the Maxwell-Garnett model (see Maxwell Garnett [15]).

$$\begin{aligned} (\rho c_p)_{nf} &= (1-\phi)(\rho c_p)_f + \phi(\rho c_p)_s, \\ \rho_{nf} &= (1-\phi)\rho_f + \phi\rho_s, \nu_{nf} = \tfrac{\mu_{nf}}{\rho_{nf}}, \\ \alpha_{nf} &= \tfrac{k_{nf}}{(\rho c_p)_{nf}}, k_{nf} = k_f\left[\tfrac{(k_s+k_f)-2\phi(k_f-k_s)}{(k_s+k_f)+\phi(k_f-k_s)}\right], \end{aligned} \tag{8}$$

where $\nu_{nf}, \rho_{nf}, (\rho c_p)_{nf}, k_{nf}, k_f, k_s, \rho_s, (\rho c_p)_f, (\rho c_p)_s$ are the nanofluid kinematic viscosity, the electrical conductivity, the nanofluid heat capacitance, thermal conductivity of the nanofluid, thermal conductivity of the fluid, the thermal conductivity of the solid fractions, the density of the solid fractions, the heat capacity of base fluid, the effective heat capacity of nanoparticles, respectively, (see Abu-Nada [16] and Kameswaran et al. [18]).
The continuity equation (1) is satisfied by introducing a stream function $\psi(x, y)$ such that

$$u = \frac{\partial \psi}{\partial y}, v = -\frac{\partial \psi}{\partial x}. \tag{9}$$

Introducing the following non-dimensional variables,

$$\psi = \left[a\nu_f\right]^{\frac{1}{2}} \; x f(\eta), u = axf'(\eta), v = -\left(a\nu_f\right) f(\eta), \tag{10}$$

$$\theta(\eta) = \frac{T - T_\infty}{T_\mathrm{w} - T_\infty}, \varphi(\eta) = \frac{C - C_\infty}{C_\mathrm{w} - C_\infty}, \eta = \left[\frac{a}{\nu_f}\right]^{\frac{1}{2}} \; y \tag{11}$$

where η, is the similarity variable, $f(\eta)$ is the dimensionless stream function, $\theta(\eta)$ is the dimensionless temperature and $\varphi(\eta)$ is the dimensionless concentration. By using (7), (8) and (11) the governing equations (2), (4) and (3) along with the boundary conditions (6) are reduced to the following two-point boundary value problem:

$$f''' + \phi_1 \left[ff'' - f'^2 - \frac{1}{\phi_2} M f'\right] - K_1 f' = 0, \tag{12}$$

$$\left(1 + \frac{4R}{3}\right)\theta'' + Pr\frac{k_f}{k_{nf}}\phi_3 \left[f\theta' - 2f'\theta + \delta\theta + \frac{E_c}{\phi_4} f''^2\right] = 0, \tag{13}$$

$$\varphi'' + Sc\,(f\varphi' - 2f'\varphi + \gamma\varphi) + Sr\theta'' = 0, \tag{14}$$

subject to the boundary conditions

$$f(0) = 0, f'(0) = 1, \theta(0) = 1, \varphi(0) = 1, \eta = 0, \tag{15}$$

$$f'(\infty) \to 0, \theta(\infty) \to 0, \varphi(\infty) \to 0, \eta \to \infty, \tag{16}$$

Where primes denote differentiation with respect to η, $\alpha_f = k_f/(\rho c_p)_f$ and $\nu_f = \mu_f/\rho_f$ are the thermal diffusivity and kinetic viscosity of the base fluid, respectively. Other non-dimensional parameters appearing in equations (12) to (14) are M, K_1, R, Pr, δ, E_c, Sc, γ and Sr denote the magnetic parameter, porous medium parameter, thermal radiation parameter, Prandtl number, heat generation parameter, Eckert number, Schmidt number, scaled chemical reaction parameter and Soret number. These parameters are defined mathematically as

$$M = \frac{\sigma B_0^2}{a\rho_f}, K_1 = \frac{\nu_f}{ak}, R = \frac{4\sigma^* T_\infty^3}{k^* k_{nf}}, Sc = \frac{\nu_f}{D}, \tag{17}$$

$$Pr = \frac{\nu_f (\rho c_p)_f}{k_f}, \delta = \frac{Q}{a(\rho c_p)_{nf}}, \gamma = \frac{K_0}{a}, \tag{18}$$

$$E_c = \frac{u_w^2}{(T_w - T_\infty)(c_p)_f}, Sr = \frac{D_1(T_w - T_\infty)}{D(C_w - C_\infty)}. \tag{19}$$

The nanoparticle volume fraction ϕ_1 and ϕ_2 are defined as

$$\phi_1 = (1-\phi)^{2.5}\left[1 - \phi + \phi\left(\frac{\rho_s}{\rho_f}\right)\right], \phi_2 = 1 - \phi + \phi\frac{(\rho_s)}{(\rho_f)},$$
$$\phi_3 = 1 - \phi + \phi\frac{(\rho c_p)_s}{(\rho c_p)_f}, \phi_4 = (1-\phi)^{2.5}\left[1 - \phi + \phi\frac{(\rho c_p)_s}{(\rho c_p)_f}\right]. \tag{20}$$

Skin friction, heat and mass transfer coefficients

The quantities of engineering interest are the skin friction coefficient C_f, the local Nusselt number Nu_x and the local Sherwood number Sh_x characterize the surface drag, wall heat and mass transfer rates respectively. The shearing stress at the surface of the wall τ_w is defined as

$$\tau_w = -\mu_{nf}\left(\frac{\partial u}{\partial y}\right)_{y=0} = -\frac{1}{(1-\phi)^{2.5}}\rho_f \sqrt{\nu_f a^3}\; x\, f''(0), \tag{21}$$

where μ_{nf} is the coefficient of viscosity. The skin friction coefficient is obtained as

$$C_{fx} = \frac{2\tau_w}{\rho_f U_w^2}, \tag{22}$$

and using equation (21) in (22) we obtained

$$\frac{1}{2}(1-\phi)^{2.5} \quad C_{fx} = -\,Re_x^{-\frac{1}{2}} f''(0). \tag{23}$$

The heat transfer rate at the surface flux at the wall is defined as

$$q_w = -k_{nf}\left(\frac{\partial T}{\partial y}\right)_{y=0} = -k_{nf}\frac{(T_w - T_\infty)}{x}\sqrt{\frac{U_w x}{\nu_f}}\theta'(0), \tag{24}$$

where k_{nf} is the thermal conductivity of the nanofluid. The local Nusselt number is defined as

$$Nu_x = \frac{x q_w}{k_f (T_w - T_\infty)}. \tag{25}$$

Using equation (24) in equation (25), the dimensionless wall heat transfer rate is obtained as

$$\left(\frac{k_f}{k_{nf}}\right) Nu_x = -\,Re_x^{\frac{1}{2}}\,\theta'(0). \tag{26}$$

The mass flux at the wall surface is defined as

$$q_m = -D\left(\frac{\partial C}{\partial y}\right)_{y=0} = -DQ\left(\frac{x}{\omega}\right)^2\sqrt{\frac{a}{\nu_f}}\;\varphi'(0), \tag{27}$$

and the local Sherwood number is obtained as

$$Sh_x = \frac{x q_m}{D(C_w - C_\infty)}. \tag{28}$$

The dimensionless wall mass transfer rate is obtained as

$$Sh_x = -\,Re_x^{\frac{1}{2}}\,\varphi'(0), \tag{29}$$

where Re_x represents the local Reynolds number and is defined as

$$Re_x = \frac{x u_w}{\nu_f}. \tag{30}$$

Method of Solution

The equations (12) to (14) are highly non-linear, it is difficult to find the closed form solutions. Thus, the solutions of these equations with the boundary conditions 15 and 16 were solved numerically using the SRM, Motsa [10].
The SRM is an iterative procedure that employs the Gauss-Seidel type of relaxation approach to linearise and decouple the system of differential equations. The linear terms in each equation is evaluated at the current iteration level (denoted by $r+1$) and non-linear terms are assumed to be known from the previous iteration level (denoted by r). The linearised form of (12) to (14) is

$$f'''_{r+1} + a_{1,r} f''_{r+1} - a_{2,r} f'_{r+1} = R_{1,r}, \tag{31}$$

$$(1 + \tfrac{4R}{3})\theta''_{r+1} + b_{1,r}\theta'_{r+1} + b_{2,r}\theta_{r+1} = R_{2,r},, \tag{32}$$

$$\varphi''_{r+1} + c_{1,r}\varphi'_{r+1} + c_{2,r}\varphi_{r+1} = R_{3,r}, \tag{33}$$

Results and Discussion

The nonlinear boundary value problem 12 to 14 subject to the boundary conditions 15 and 16 connot be solved in closed form, so these equations are solved numerically using the spectral relaxation method (SRM) for Cu-water and Al_2O_3-water nanofluids with water as the base fluid (i.e. with a constant Prandtl number $Pr = 6.7850$). The thermophysical properties of the nanofluids used in the numerical simulations are given in Table 1. Extensive calculations have been performed to obtain the velocity, temperature, concentration profiles as well as skin friction, local Nusselt number and local Sherwood

number for various values of physical parameters such as ϕ, M, K_1, R, Pr, Sc, δ, E_c, γ and Sr. To determine the accuracy of our numerical results, the skin friction and the heat transfer coefficient are compared with the published results of Hamad [17], Kameswaran et al. [18] and Grubka and Bobba [19]in Tables 2 and 3. Here we have varied the M with ϕ while keeping other physical parameters fixed for Cu-water and Al_2O_3-water in Table 2. It is observed that increasing the values of M results in an increase in the skin friction coefficient. The calculated values show good agreements with Hamad [17] and Kameswaran et al. [18].

In Table 3 gives a comparison of the values of wall temperature gradient $-\theta'(0)$ results with those obtained by Kameswaran et al. [18] and Grubka and Bobba [19] when $M = E_c = K_1 = \delta = R = \phi = 0$, $Sc = 1$, $Sr = 0.2$ and $\gamma = 0.08$ for different values of Prandtl number Pr. As it is shown in the table thewall temperature gradient $-\theta'(0)$ increases with an increase of Prandtl number. This is fact because the definition of Prandtl number is the ratio of kinematic viscosity to thermal diffusivity. An increase in the values of Prandtl number implies that momentum diffusivity dominates thermal diffusivity. Hence, the rate of heat transfer at the surface increases with increasing values of Pr. It is observed that the present results are in good agreement with results in the literature by Kameswaran et al. [18] and Grubka and Bobba [19]. In Table 4 approximate solutions of the skin friction coefficient, surface heat transfer and the surface mass transfer rates at different values of flow parameters are presented. All SRM results were generated using $L = 30$, $Nx = 60$ and $Nt = 1000$, these values were found to give accurate solutions after a numerical experimentation. The L and Nt in the tables represent the maximum Lth and Ntth iteration required to produce converging results. It is observed that increasing the values of Sr increase Sherwood numbers in case of Cu-water but the opposite trend is observed with Al_2O_3-water. Also increasing the values of Sc increase the Sherwood numbers for both cases of nanofluids while increasing in heat generation parameter δ is tend to decrease the heat transfer rate for both nanofluids. The table also shows that surface mass transfer rates increase with increasing in the values of the chemical reaction parameter γ as can be seen from the table.

The effects of physical parameters on various fluid dynamic quantities are show in Figures 1 - 13. Figures 1 - 4 illustrate the effect of the nanoparticle volume fraction ϕ on the velocity, temperature and concentration profiles, respectively, in the case of a Cu-water nanofluid and Al_2O_3-water nanofluid. It is clear that as the nanoparticle volume fraction increases, the Cu-water nanofluid velocity decreases while the Al_2O_3-water nanofluid velocity increases. As it is shown in Figure 1 while the temperature profile increases with increase in the values of nanoparticle volume fraction this is clear from Figure 2. increasing the volume fraction of nanoparticles increases the thermal conductivity of nanofluid and in turn results a thickening of the thermal boundary layer. It is also observed that the temperature distribution in a Cu-water nanofluid is higher than that of Al_2O_3-water nanofluid; this is an anticipated results because Cu-water is good conductor of heat and electricity. The Al_2O_3-water nanofluid concentration profile decreases as the nanoparticle volume fraction increases but reveres it true to that of Cu-water nanofluid as shown in Figure 3.

Figure 4 shows the effect of the porous medium parameter K_1 on the velocity in case of a cu-water and Al_2O_3-water nanofluids. increasing the porous medium parameter K_1 decreases the velocity profiles of both nanofluids. We observed from the Figure, the velocity profile of Al_2O_3-water nanofluid is higher than that of Cu-water nanofluid. Figures 5 and 6 show the effect of porous medium parameter K_1 on the temperature and solutal concentration profiles respectively, in the case of Cu-water and Al_2O_3-water nanofluids. It is clear that as the porous medium parameter K_1 increases the temperature and solutal concentration profiles increase.It is observed that the temperature and concentration profiles increment of Al_2O_3-water nanofluid is less than that of Cu-water nanofluid. Figure 7 illustrates the influence of heat generation parameter δ on the temperature profile in the case of Cu-water and Al_2O_3-water nanofluids. We observed that the temperature profile increases for both cases of nanofluids with increasing in the values of heat generation parameter δ. It found that the temperature in case of Cu-water is more than that of Al_2O_3-water nanofluids. Increasing the values of heat generation parameter δ increases the thermal conductivity of nanofluid and the thickening of the thermal boundary layer. Figure 8 shows the influence of the magnetic parameter M on nanofluid velocity profile in the case of Cu-water and Al_2O_3-water nanofluids. When the magnetic parameter M increases, the nanofluid velocity profile of Cu-water and Al_2O_3-water decrease. This is because of the application of the transverse magnetic field in an electrically conducting fluid produces a ratarding lorenz force slows down the fluid motion in the boundary layer and hence decreases the velocity at the expense of increasing it is temperature and the solutal concentration. But we observed the opposite for solutal concentration of Al_2O_3-water nanofluid is against this fact as illustrates in Figure 4. The velocity profile of the Al_2O_3-water nanofluid is higher than that of the Cu-water nanofluid as it shown in the Figure.

Figure 9 shows the effect of the viscous dissipation parameter E_c on the temperature profile in the case of Cu-water and Al_2O_3-water nanofluids. It is observed that the temperature profile increases of both nanofluids with increasing in the values of E_c; we notice that the influence of an increment in E_c is to increase the temperature distribution. This is due

to the fact that the energy is stored in the fluid region as a consequence of dissipation because the viscosity and elastic deformation. It is observed that the temperature profile in the case of Cu-water nanofluid is higher than that of Al_2O_3-water nanofluid. Figure 10 shows the effect of the thermal radiation parameter R on the temperature profile in the case of both nanofluids. Increasing the thermal radiation Parameter R increases the temperature profile of Cu-water and Al_2O_3-water nanofluids. We observed that the temperature increases of Cu-water is higher than that of Al_2O_3-water nanofluids. The thermal radiation parameter R is responsible to thickening of thermal boundary layer. This enables the nanofluids to release the heat energy from the flow region and cases the system to be cool. This is true because of increasing the Rosseland approximation results in an increase in the temperature profile. Figure 11 illustrates the effect of the Schmidt number Sc on the solutal concentration profile in the case of Cu-water and Al_2O_3-water nanofluids. Increasing the values of Sc decreases the solutal concentration profile of both case of nanofluids. It is observed that the concentration profile of Cu-water nanofluid increases more than that of AI_2O_3-water nanofluid. Figures 12 and 13 show the effect of two parameters namely by chemical reaction parameter γ and the Soret number Sr on the concentration profiles in the case of Cu-water and Al_2O_3-water nanofluids in Figure 12 and 13 respectively. We observed that the concentration profiles decreases with an increase in the values of the scale chemical reaction parameter γ whereas the chemical reaction parameter γ effect shows no substation changes on the nanofluid velocity and temperature profile in the two case of the nanofluids. It is clear that the solutal concentration profiles in case of Al_2O_3-water nanofluid is relatively less than that of Cu-water nanofluid in Figure 12. While the Figure 13as the Soret number Sr increases, the solutal concentration boundary layer thickness of both case of nanofluids also increase. We found that the solutal concentration profiles increment of Al_2O_3-water nanofluid exhibits less than that of Cu-water nanofluid.

Conclusions

We have investigated the heat and mass transfer in steady MHD boundary layer flow in nanofluids through a porous due to an stretching surface subjected to a magnetic field, heat generation, chemical reaction, viscous dissipation and thermal radiation effects. From the numerical simulations, some results can be drawn as follow:
[i] The velocity profile of Cu-water nanofluid decreases with increasing in the nanoparticle volume fraction whereas the velocity profile of Al_2O_3-water nanofluids increases with increasing in the nanoparticle volume fraction while the velocity profile of both nanofluids decrease with an increase in magnetic and porous medium parameters.
[ii] The temperature profile of both nanofluids increase with increasing in the values of the nanoparticle volume fraction while the concentration of Al_2O_3-water nanofluids decreases with increasing in the values of the nanoparticle volume fraction and the opposite trend is observed for the concentration of Cu-water nanofluids with increasing in the values of the nanoparticle volume fraction.
[iii] The temperature profile of both nanofluids increase with increase in the values of the Viscous dissipation, heat generation and thermal radiation parameters.
[iv] The concentration profile of both nanofluids decreases with increase in the values of chemical reaction parameter and Schmidt number while the opposite trend is observed for the increasing values of the Soret number in the both case of nanofluids.
[v] The rate of thermal boundary layer thickness of both nanofluids decreases with the presence of nanoparticle volume fraction, thermal radiation, porous media and viscous dissipation in the flow field.
[vi]In general, the Al_2O_3-water nanofluid shows thicker velocity layer at the plate than a Cu-water nanofluids; Al_2O_3-water nanofluid exhibits thicker thermal and concentration boundary layer than that of a Cu-water nanofluid.

Acknowledgment:

The authors are grateful to the University of KwaZulu-Natal, South Africa for the necessary support. Also, this work is based on the research supported by the National Research Foundation, South Africa.

References

[1] Crane, L.J. (1970) Flow past a stretching plate, *ZAMP. Angew Math. Phys.* 21, 645 – 647.
[2] Khan, W.A. and Pop, I. (2010) Boundary layer flow of a nanofluid past a stretching sheet, *International Journal of Heat and Mass Transfer* 53, 2477 – 2483.
[3] Abd El-Aziz, M. (2007) Thermal-diffusion and diffusion-thermo effects on combined heat and mass transfer by hydromagnetic three-dimensional free convection over apermeable stretching surface with radiation, *Physics Letters* 372, 3, 263 – 272.

[4] Hady, F.M., Ibrahim, F.S., Abdel-Gaied, S.M. and Mohamed Eid, R. (2012) Radiation effect on viscous flow of a nanofluid and heat transfer over a nonlinearly stretching sheet, *Nanoscale Res. Lett* 7, 229.

[5] Bachok, N., Ishak, A. and Pop, I. (2012) Unsteady boundary-layer flow and heat transfer of a nanofluid over a permeable stretching/shrinking sheet, *International Journal of Heat and Mass Transfer* 55, 2102 – 2109.

[6] Rohni, A.M., Ahmad, A.S. and Ismail, Md I. and Pop, I. (2013) Flow and heat transfer over an unsteady shrinking sheet with suction in a nanofluid using Buongiorno's model, *International Communications in Heat and Mass Transfer* 43, 75 – 80.

[7] Buongiorno, J., (2006) Convective transport in nanofluids, *ASME Journal of Heat Transfer* 128, 240 – 250.

[8] Jafar, K., Nazar, R., Ishak, A., Pop, I. (2011) MHD Flow and Heat Transfer Over stretching/ shrinking sheets with external magnetic field, viscous dissipation and Joule Effects, *Canadian Journal on Chemical Engineering* 99, 1 – 11.

[9] Murthy, P.V. and Singh, P. (1997) Effect of viscous dissipation on a non-Darcy natural convection regime, *International Journal of Heat and Mass Transfer* 40, 1251 – 1260.

[10] Motsa , S.S.(2013) A New spectral relaxation method for similarity variable nonlinear boundary layer flow systems, *Chemical Engineering Communications* 16, 23 – 57.

[11] Tiwari, R.K. and Das, M.N. (2007) Heat tranfer augmentation in a two - sided lid-driven diffrentially heated square cavity utilizing nanofluids, *International Journal of Heat and Mass Transfer* 50, 2002 – 2018.

[12] Sheikholeslami, M., Bandpy, M.G., Ganji, D.D., Soleimani, S. and Seyyedi, S.M. (2012) Natural convection of nanofluids in an enclosure between a circular and a sinusoidal cylinder in the presence of magnetic field, *International Communications in Heat and Mass Transfer* 39, 1435 – 1443.

[13] Oztop, H.F. and Abu-Nada, E. (2008) Numerical study of natural convection in partially heated rectangular enclosures filled with nanofluids, *International Journal of Heat and Fluid Flow* 29, 1326 – 1336.

[14] Brinkman , H.C.(1952) The viscosity of concentrated suspensions and solution, *Journal of Chemical Physics* 20, 571 – 581.

[15] J.C., Maxwell Garnett (1904) Colours in metal glasses and in metallic films, *Philosophical Transactions of the Royal Society of London* 203, 385 – 420.

[16] Abu-Nada , E.(2008) Application of nanofluids for heat transfer enhancement of separated flows encountered in a backward facing step, *International Journal of Heat and Fluid Flow* 29, 242 – 249.

[17] Hamad, M.A.A. (2011) Analytical solution of natural convection flow of a nanofluid over a linearly stretching sheet in the presence of magnetic field, *International Communications in Heat and Mass Transfer* 38, 487 – 492.

[18] Kameswaran, P.K., Narayana, M., Sibanda, P. and Murthy, P.V. (2012) Hydromagnetic nanofluid flow due to a stretching or shrinking sheet with viscous dissipation and chemical reaction effects, *International Journal of Heat and Mass Transfer* 55, 7587 – 7595.

[19] Grubka, L.G. and Bobba, K.M. (1985) Heat transfer characteristics of a continuous stretching surface with variable temperature, *The ASME Journal of Heat Transfer* 107, 248 – 250.

Table 1. Thermophysical properties of the water and copper and alumina nanoparticles, (see Sheikholeslami et al. [12] and Oztop and Abu-Nada [13])

Physical properties	Base fluid (Water)	Copper (Cu)	Alumina (Al_2O_3)
$C_p(J/kgK)$	4179	385	765
$\rho(Kg/m^3)$	997.1	8933	3970
$k(W/mK)$	0.613	401	40
$\alpha \times 10^7(m^2/s)$	1.47	1163.1	131.7
$\beta \times 10^5(K^{-1})$	21	1.67	0.85

Table 2. Comparison of $-f''(0)$ for various values of M and ϕ when $Pr = 6.2$, $Sc = 1$, $Sr = 0.2$, $E_c = 0$, $K_1 = 0.0$, $R = 0$, $\delta = 0.02$, $\gamma = 0.08$

		Hamad[17]		Kameswaran et al.[18]		Present results	
M	ϕ	Cu-water	Al_2O_3	Cu-water	Al_2O_3	Cu-water	Al_2O_3
0	0.05	1.10892	1.00538	1.108919904	——-	1.108920	1.005385
	0.1	1.17475	0.99877	1.174746021	——-	1.174746	0.998781
	0.15	1.20886	0.98185	1.208862320	——-	1.208862	0.981854
	0.2	1.21804	0.95592	1.218043809	——-	1.218043	0.955931
0.5	0.05	1.29210	1.20441	1.292101949	——-	1.292102	1.204412
	0.1	1.32825	1.17548	1.328248829	——-	1.328249	1.175484
	0.15	1.33955	1.13889	1.339553714	——-	1.339554	1.138892
	0.2	1.33036	1.09544	1.330356126	——-	1.330356	1.095444
1	0.05	1.45236	1.37493	1.452360679	——-	1.452361	1.374930
	0.1	1.46576	1.32890	1.465763175	——-	1.465763	1.328901
	0.15	1.45858	1.27677	1.458581570	——-	1.458582	1.276766
	0.2	1.43390	1.21910	1.433898227	——-	1.433898	1.219104
2	0.05	1.72887	1.66436	1.728872387	——-	1.728872	1.664356
	0.1	1.70789	1.59198	1.707892022	——-	1.707892	1.591984
	0.15	1.67140	1.51534	1.671398302	——-	1.671398	1.515336
	0.2	1.62126	1.43480	1.621264175	——-	1.621264	1.434799

Table 3. Comparison of the values of wall temperature gradient $-\theta'(0)$ from currents with Kameswaran et al. [18] and Grubka and Bobba [19] for different values of Prandtl numbers Pr when $M = E_c = K_1 = \delta = R = 0$, $Sc = 1$, $Sr = 0.2$, $\gamma = 0.08$ and $\phi = 0$.

Pr	0.72	1	3	10	100
Kameswaran et al. [18]	1.08852	1.33333	2.50973	4.79687	15.71163
Grubka and Bobba [19]	1.0885	1.3333	2.5097	4.7969	15.7120
Present result (SRM)	1.088524	1.333333	2.509725	4.796873	15.711967

Table 4. Comparison of the SRM solutions for $f''(0)$, $-\theta'(0)$, and $-\varphi'(0)$ for different values of Sr, Sc, δ and γ. $\phi = 0.1, E_c = 1$, $M = 0.5$, $Sc = 1$, $\delta = 0.01$, $Pr = 6.2$, $K_1 = 1$, $\gamma = 0.08$, $Sr = 0.2$.

	$Cu - water$					$Al_2O_3 - water$
	$\phi = 0.1, E_c = 1$		$R = 2, Pr = 6.2$		$K_1 = 1, M = 0.5$	
Sr	$f''(0)$	$-\theta'(0)$	$-\varphi'(0)$	$f''(0)$	$-\theta'(0)$	$-\varphi'(0)$
0.0	1.662602	0.262150	1.202677	1.543296	0.387825	1.231631
0.1	1.662602	0.262150	1.203733	1.543296	0.387825	1.223237
0.3	1.662602	0.262150	1.205845	1.543296	0.387825	1.206449
0.4	1.662602	0.262150	1.206901	1.543296	0.387825	1.198055
Sc						
0.6	1.662602	0.262150	1.204789	1.543296	0.387825	1.214843
0.7	1.662602	0.262150	1.574980	1.543296	0.387825	1.587530
0.8	1.662602	0.262150	1.887968	1.543296	0.387825	1.901663
0.9	1.662602	0.262150	2.163376	1.543296	0.387825	2.177705
δ						
0.6	1.662602	2.350214	0.854174	1.543296	2.408433	0.877492
0.7	1.662602	2.305154	0.861916	1.543296	2.365163	0.884957
0.8	1.662602	2.258469	0.869887	1.543296	2.320555	0.892611
0.9	1.662602	2.044322	0.905376	1.543296	2.122996	0.925819
γ						
0.6	1.662602	0.262150	1.140069	1.543296	0.387825	1.155418
0.7	1.662602	0.262150	1.219003	1.543296	0.387825	1.228159
0.8	1.662602	0.262150	1.438686	1.543296	0.387825	1.439333
0.9	1.662602	0.262150	1.642761	1.543296	0.387825	1.639504

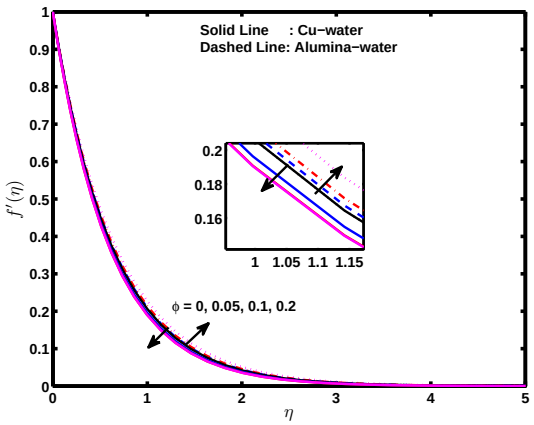


Figure 1. Effect of various nanoparticle values fraction ϕ on velocity profile for $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.4$.

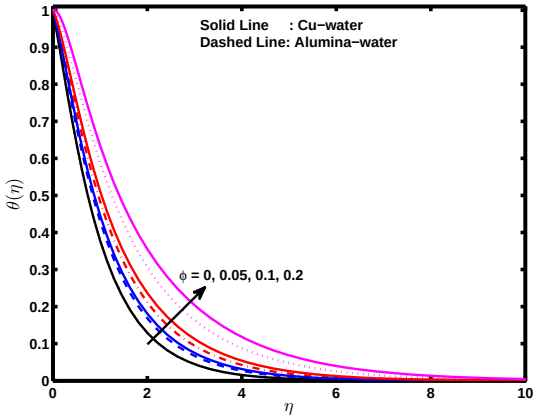


Figure 2. Effect of various nanoparticle values fraction ϕ on temperature profile for $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.4$.

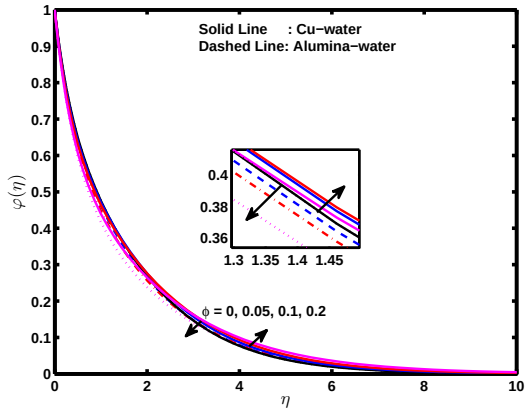


Figure 3. Effect of various nanoparticle values fraction ϕ on the concentration profile for $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.4$.

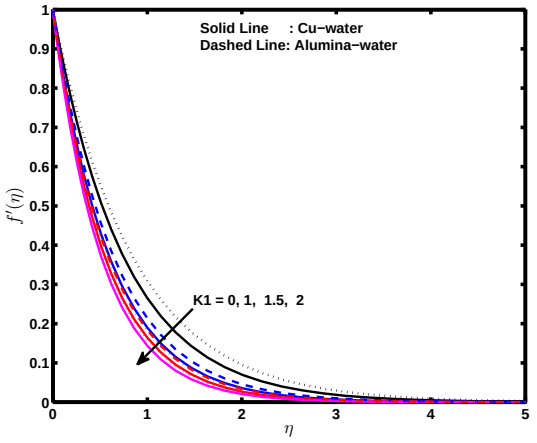


Figure 4. Effect of various nanoparticle values fraction ϕ on the velocity profile for $\phi = 0.2$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

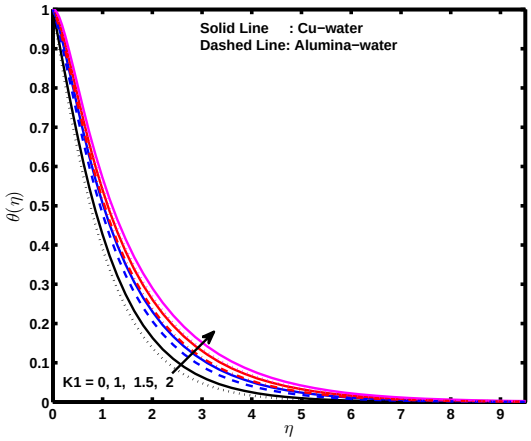


Figure 5. Effect of the porous medium parameter K_1 on temperature profile for $\phi = 0.2$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

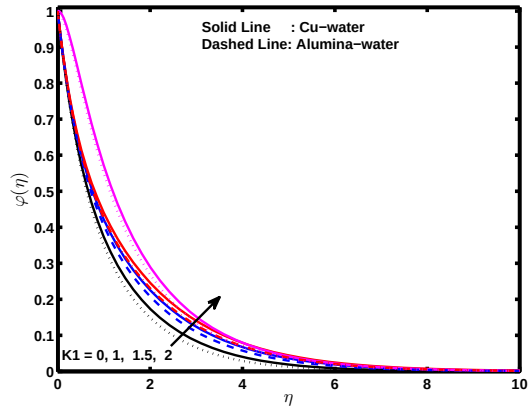


Figure 6. Effect of the porous medium parameter K_1 on concentration profile for $\phi = 0.2$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

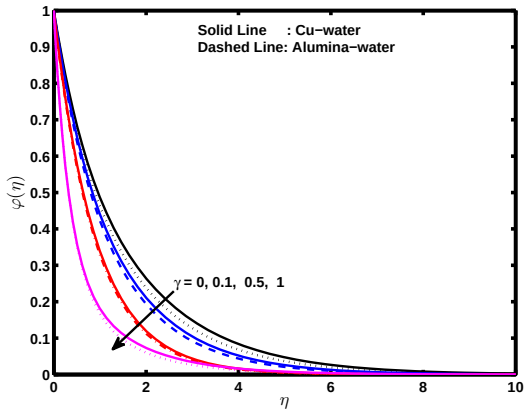


Figure 7. Effect of heat generation parameter δ on the temperature profile for $\phi = 0.2$, $M = 0.5$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $K_1 = 1.0$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

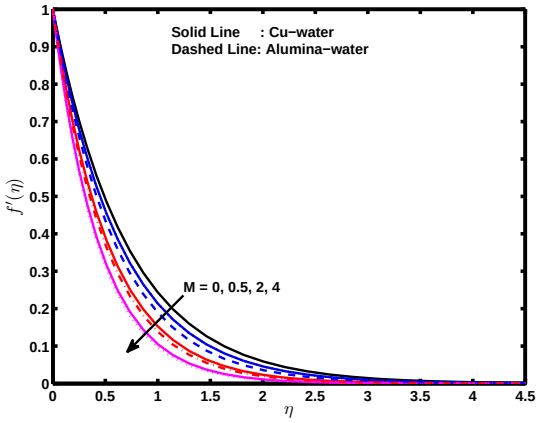


Figure 8. Effect of magnetic parameter M on the velocity profile for $\phi = 0.1$, $K_1 = 1.0$, $E_c = 1.0$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

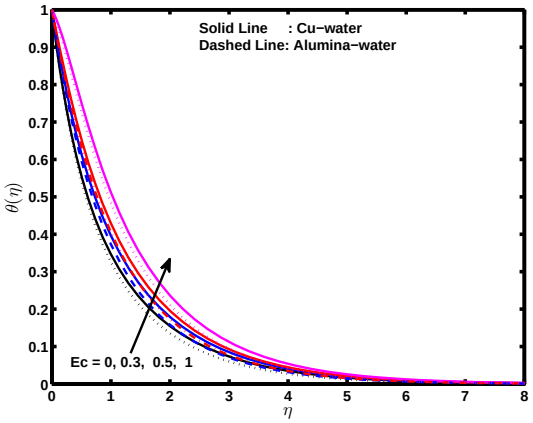


Figure 9. Effect of viscous dissipation parameter E_c on the temperature profile for $\phi = 0.1$, $K_1 = 1.0$, $M = 0.5$, $R = 2.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

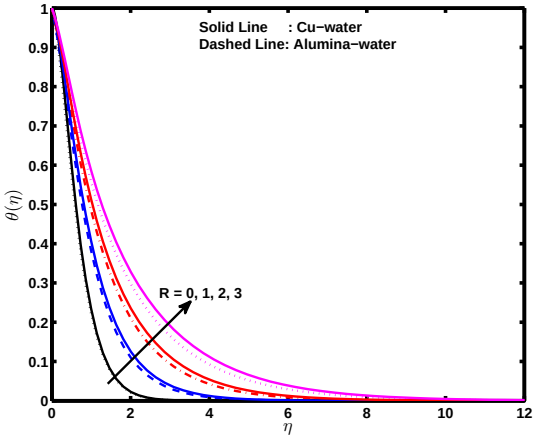


Figure 10. Effect of thermal radiation parameter R on the temperature profile for $\phi = 0.1$, $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $Sc = 1$ and $Sr = 0.2$.

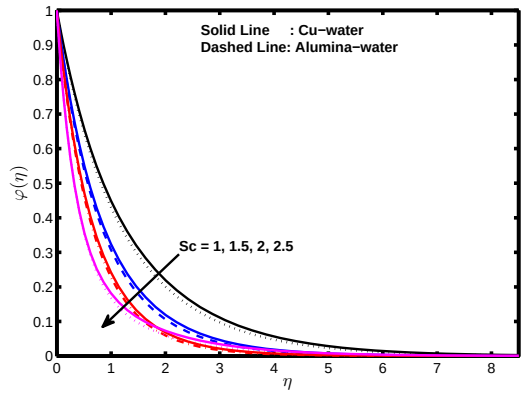


Figure 11. Effect of the Schmidt number Sc on concentration profile for $\phi = 0.1$, $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $Pr = 6.2$, $\delta = 0.01$, $\gamma = 0.08$, $R = 2$ and $Sr = 0.2$.

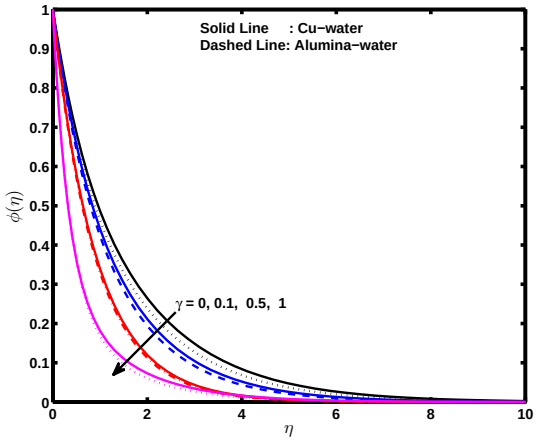


Figure 12. Effect of the chemical reaction parameter γ and Soret number Sr on concentration profiles for $\phi = 0.1$, $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $Pr = 6.2$, $\delta = 0.01$, $Sc = 1$ and $R = 2$.

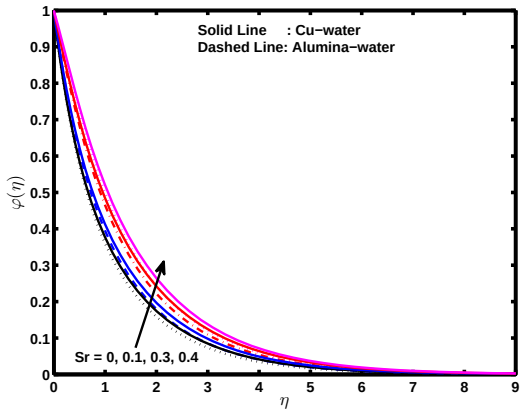


Figure 13. Effect of the chemical reaction parameter γ and Soret number Sr on concentration profiles for $\phi = 0.1$, $K_1 = 1.0$, $M = 0.5$, $E_c = 1.0$, $Pr = 6.2$, $\delta = 0.01$, $Sc = 1$ and $R = 2$.

A unified computational method of differential analysis for solving the Navier-Stokes equations.

Mike Mikalajunas

CIME, 38 Neuville, Montreal, Canada J7V 8L1

michelmikalajunas@bellnet.ca
jpnelson_mfc@yahoo.ca

Abstract

Certain traditional methods of Calculus for solving DEs and systems of DEs in engineering analysis depend in one form or another on the use of some general initially assumed analytical representation of the intended solution. Unfortunately this often leads to defining one or several integrals that cannot always be resolved exactly. In order to avoid this complication we propose that the complete "*differential*" of a general initially assumed analytical representation of the intended solution with unknown coefficients to solve for be used instead as a means of solving for DEs and systems of DEs. Such a novel method of differential analysis has led to the development of what appears to be some form of a unified theory of integration. This would represent the greatest opportunity by which the complete Navier-Stokes equations for incompressible flow in the presence of any external forces may be investigated for the existence of any "*generalized*" analytical solutions under the three most commonly used coordinate systems.

Keywords: Universal Polynomial Transform, ODEs, PDEs, Multinomial Expansion Theorem**,** Quantum Physics, Quantum computers, Navier-Stokes equations, Theory of everything.

Introduction

Such a non-traditional method of using this unique form of differential analysis in Calculus would have the real potential of defining integrals that can be completely resolved because a certain number of these initially assumed "*differentials*" are expected to become "*exact*" from the application of a well defined computational process. This would represent a very significant departure from current traditional methods of engineering analysis favoring a purely "*numerical*" method of integration in cases by which no real analytical solution to many fundamental DEs and systems of DEs in engineering science is possible. The greatest advantage of performing such a type of analysis strictly at the differential level has led to the development of some type of a unified theory of integration that can be applied for finding approximate or in some cases exact analytical solutions to "*all types*" of DEs and systems of DEs encountered in engineering analysis. The entire process of analytical integration now becomes a matter of pure computational analysis just for identifying those differentials that are exact and thus completely integrable. Such a very unique method of differential analysis will be applied for the complete *analytical* solution of a number of randomly selected DEs that would include a first and second order ODE as well as a second order PDE. The outcome of having performed such a detailed differential analysis on these very simple DEs may provide us in the long term with some basic fundamental tools of analysis by which a generalized theory of the Navier-Stokes equations may be possible in the foreseeable future. Not surprisingly since such a novel method of differential analysis has led to the development of a *computational based* unified *analytical* theory of integration. Beyond the Navier-Stokes equations are other equations of significant importance to the physical sciences that would include Maxwell's equations, Einstein's field equations, the Schrödinger equation just to name a few. Each of these fundamental equations of science would define their own very unique ideology all of which may one day be consolidated into one gigantic universal theory of everything.

1. Universal differential form representation of all mathematical equations

For solving a DE or a system of DEs, an alternative representation in *complete differential form* for a generally assumed system of "k" number of implicitly defined *multivariate* mathematical equations in the form of "$f_k(z_m, x_n) = 0$" that consist of "m" number of dependent variables and "n" number of independent variables [Mikalajunas (2015)] may be *completely* defined as :

(1). *Primary Expansion:*

$$F_i(W_1, W_2, \dots, W_{p+q}) \;=\; 0 \;=\; \sum_{t}^{r} a_{i,t} \left(\prod_{j}^{p+q} W_j^{E_{i,kj}} \right) \qquad (1 \le i \le k) \tag{1}$$

where "W_j" for $1 \le j \le p$ are arbitrarily defined auxiliary variables that take part in representing the complete initially assumed analytical solution of a DE or a system of DEs. For any number of basis functions that are present in a DE or a system of DEs we would have to define an additional "q" number of known *supplemental* auxiliary variables for including each of their differential expansion as part of the complete overall expansion for representing the system of "k" number of implicitly defined multivariate equations. In such cases, the total number of auxiliary variables would grow from "p" to "$p + q$" when such basis functions are present in these types of DEs. Each of the "p" number of arbitrarily defined auxiliary variables are always initially assumed as raised to some floating point number and finally, "r" refers to the total number of multivariate polynomial terms that are present in each of the "k" number of implicitly defined multivariate polynomial equations.

(2). *Secondary Expansion:*

$$dz_i \;=\; dW_i \qquad (1 \le i \le m) \tag{2}$$

$$dx_i \;=\; dW_{m+i} \qquad (1 \le i \le n) \tag{3}$$

$$\sum_{t=1}^{m} N_{i(m+n+1)-m-n-1+t} dz_t \;+\; \sum_{t=1}^{n} N_{i(m+n+1)-n-1+t} dx_t \;=$$

$$=\; N_{i(m+n+1)} dW_j \qquad [1 \le i \le p+q-m-n] \;[m+n+1 \le j \le p+q] \tag{4}$$

As in the case of the *Primary Expansion*, each of the expressions for "N_u" in equation (4) is also defined as a multivariate polynomial with unknown coefficients and floating point exponent values to solve for.

And finally we have,

$$\sum_{t=1}^{m} T_{i(m+n+1)-m-n-1+t} dz_t \;+\; \sum_{t=1}^{n} T_{i(m+n+1)-n-1+t} dx_t \;=$$

$$=\; T_{i(m+n+1)} dW_j \qquad [1 \le i \le q] \;[p \le j \le p+q] \tag{5}$$

where each of the expression for "T_u" in equation (5) is also a multivariate polynomial but this time containing only *known* coefficient and exponent values that are reserved exclusively for defining each of the basis functions that would be present inside a DE or a system of DEs.

At the present time there is no other known *universal* representation of *all* mathematical equations consisting only of algebraic and elementary basis functions other than the one suggested above.

In complete expanded form we would write this as follow:

(1). *Primary Expansion:*

$$F_1 = 0 = a_{1,1}W_1^{m_{11}}W_2^{m_{12}}\cdots W_{p+q}^{m_{1,p+q}} + a_{1,2}W_1^{m_{1,p+q+1}}W_2^{m_{1,p+q+2}}\cdots W_{p+q}^{m_{1,2(p+q)}} + \ldots + a_{1,r}W_1^{m_{1,(p+q)(r-1)+1}}W_2^{m_{1,(p+q)(r-1)+2}}\cdots W_{p+q}^{m_{1,r(p+q)}} \quad (6)$$

$$F_2 = 0 = a_{2,1}W_1^{m_{21}}W_2^{m_{22}}\cdots W_{p+q}^{m_{2,p+q}} + a_{2,2}W_1^{m_{2,p+q+1}}W_2^{m_{2,p+q+2}}\cdots W_{p+q}^{m_{2,2(p+q)}} + \ldots + a_{2,r}W_1^{m_{2,(p+q)(r-1)+1}}W_2^{m_{2,(p+q)(r-1)+2}}\cdots W_{p+q}^{m_{2,r(p+q)}} \quad (7)$$

. . .

. . .

. . .

$$F_k = 0 = a_{k,1}W_1^{m_{k1}}W_2^{m_{k2}}\cdots W_{p+q}^{m_{k,p+q}} + a_{k,2}W_1^{m_{k,p+q+1}}W_2^{m_{k,p+q+2}}\cdots W_{p+q}^{m_{k,2(p+q)}} + \ldots + a_{k,r}W_1^{m_{k,(p+q)(r-1)+1}}W_2^{m_{k,(p+q)(r-1)+2}}\cdots W_{p+q}^{m_{k,r(p+q)}} \quad (8)$$

(2). *Secondary Expansion:*

$$dz_i = dW_i \qquad (1 \le i \le m) \quad (9)$$

$$dx_i = dW_{m+i} \qquad (1 \le i \le n) \quad (10)$$

$$[N_1dz_1 + N_2dz_2 + \ldots + N_mdz_m] + [N_{m+1}dx_1 + N_{m+2}dx_2 + \ldots + N_{m+n}dx_n] = N_{m+n+1}dW_{m+n+1} \quad (11)$$

$$[N_{m+n+2}dz_1 + N_{m+n+3}dz_2 + \ldots + N_{2m+n+1}dz_m] + [N_{2m+n+2}dx_1 + N_{2m+n+3}dx_2 + \ldots + N_{2(m+n+1)-1}dx_n] = N_{2(m+n+1)}dW_{m+n+2} \quad (12)$$

$$\begin{array}{ccc} . & . & . \\ . & . & . \\ . & . & . \end{array}$$

$$\left[N_{(p+q-1)(m+n+1)+1} dz_1 + N_{(p+q-1)(m+n+1)+2} dz_2 + \ldots + N_{(p+q-1)(m+n+1)+m} dz_m \right] +$$

$$+ \left[N_{(p+q-1)(m+n+1)+m+1} dx_1 + N_{(p+q-1)(m+n+1)+m+2} dx_2 + \ldots + N_{(p+q)(m+n+1)-1} dx_n \right] =$$

$$= N_{(p+q)(m+n+1)} dW_{p+q} \qquad (13)$$

The actual process of transforming a complete mathematical equation or a system of mathematical equations in terms of the above universal differential form representation is referred to as taking its *Multivariate Polynomial Transform.* The complete *reverse* process of going from a differential form representation back to the original complete mathematical equation or system of mathematical equations would be referred to as taking the *inverse* of a *Multivariate Polynomial Transform*. This would involve following a very unique integration process in the *Secondary Differential Expansion* for determining the complete analytical expression corresponding to each auxiliary variable. They in turn would each be substituting back into the *Primary Expansion* for arriving at the complete original expression in the form of "$f_k(z_m, x_n) = 0$".

Appendix A provides a list of the *Multivariate Polynomial Transform* corresponding to a variety of univariate and multivariate mathematical equations. For simplicity, both the Sine and Cosine function have been expressed as a rational combination of the Tangent function using the following basic trigonometric identity:

$$Sin(x) = \frac{2Tan(x/2)}{1 + Tan^2(x/2)} \qquad (14)$$

$$Cos(x) = \frac{1 - Tan^2(x/2)}{1 + Tan^2(x/2)} \qquad (15)$$

Just by increasing the total number of dependent and independent variables, the concept of a *Multivariate Polynomial Transform* is still applicable for including all *systems* of mathematical equations as well. However, space limitation prevents the inclusion of these types of mathematical equations as good illustrative examples.

2. Unique template for investigating the probable existence of complete *"general"* analytical solutions to DEs and systems of DEs by using a method of conjecture

A necessary condition for defining a complete unified analytical theory of integration is by substituting an initially assumed version with unknown coefficients to solve for of the universal differential form representation of all mathematical equations as described by equations (1) through (5) into *any type* of DEs and systems of DEs. This would always result into defining a very unique type of system of nonlinear simultaneous equations to solve for. The exact *numerical* solution sets obtained would then be used as a means of inverting the corresponding initially assumed differential expansions for arriving at an exact or approximate analytical solution that would be expressible only in terms of the algebraic and elementary basis functions.

Such an initially assumed differential expansion form would possess all the characteristics of a complete mathematical transform so we would refer to it as an *initially assumed Multivariate Polynomial Transform* or in short IAMPT.

The entire process of using an IAMPT for solving DEs and systems of DEs can be divided into two fundamental stages. The first, is the computational stage by which the corresponding nonlinear simultaneous equations of a DE or a system of DEs are numerically derived and completely solved for. The second, is the analytical stage by which every numerical solution set obtained is converted to pure analytical form. This would involve the process of identifying and solving for those exact integrals that are present in the *Secondary Expansion* which have successfully pass the complete test for exactness. From this exact integration process, the complete expression for each initially assumed set of auxiliary variables are obtained and substituted into the *Primary Expansion* for arriving at the complete analytical solution of the DE or system of DEs.

When selecting a suitable IAMPT for solving a particular DE or a system of DEs, the total number of unknown coefficients and floating point exponent values to solve for becomes purely arbitrary and should be as high as possible. This is necessary as a means of capturing those "*exact*" analytical solutions that can successfully resolve a DE or a system of DEs uniquely in terms of some combination of algebraic and elementary basis functions. The limitations on the total number of unknown coefficients and exponent values to solve for as defined from an IAMPT is generally set by the capacity of a computer system to handle extremely large numbers of very complex nonlinear simultaneous equations to solve for.

The resultant system of nonlinear simultaneous equations to solve for will always consist of an *infinite number* of exact numerical solutions sets provided that the IAMPT has been chosen large enough to contain the exact solution of the DE or system of DEs that is being solved for.

Some of the reasons that would account for the existence of such an infinite number of numerical solution sets are:

- The ability for an exact solution to a DE or a system of DEs to satisfy an infinite number of initial conditions.

- The permutation of each auxiliary variable present in both the *Primary* and *Secondary Expansion* for representing the same identical exact analytical solution of the DE or system of DEs.

- As a result of the natural computational process involved in solving for a very large number of complex nonlinear simultaneous equations, many numerical solutions sets obtained are expected to define numerous types of <u>*trivial*</u> algebraic identities from the process of inverting the corresponding IAMPT. Such type of identities will always be present in one form or another in the final representation of the analytical solution. A good example is the "$Sin^2(x) + Cos^2(x) = 1$" or any other algebraic variations of this trigonometric identify that would also include other types of basis functions as well.

- The presence of singular solutions.

- As a result of the natural computational process involved in solving for a very large number of complex nonlinear simultaneous equations, many numerical solutions sets obtained will naturally lead to the formation of one or several expressions in the *Secondary Expansion* that would be represented as a ratio of two exactly identical multivariate polynomials. These types of ratios would be considered as <u>*trivial ratios*</u> that would have to be all completely eliminated before any attempts is made for inverting a *Secondary Expansion.*

For every numerical solution set obtained as a result of solving for these nonlinear simultaneous equations there will always be a corresponding exact analytical solution satisfying a "<u>*unique*</u>" set of initial conditions. We would refer to the existence of such a type of exact analytical solution as an "*instance solution*". As there are an infinite number of possible numerical solution sets of the nonlinear simultaneous equations this will give rise to an infinite number of such *instance solutions*.

By consolidating a sufficient number of such instance solutions we can by using a method of conjecture potentially uncover more complete "*generalized*" versions of analytical solutions satisfying a *general* DE or a system of DEs. It therefore becomes quite imperative that as a result of solving for the nonlinear simultaneous equations we always continuously keep track of all instance analytical solutions obtained in the form of a table that we would like to refer as a *"numerically controlled system of analytics table"* or in short an (NCSA) table.

The following general system of PDEs of any order can be used for describing the most general case of an NCSA table:

$$G_k = G_k\left(z_1, z_2, \dots, z_m, x_1, x_2, \dots, x_n, \frac{\partial z_1}{\partial x_1}, \dots, \frac{\partial z_1}{\partial x_n}, \frac{\partial z_2}{\partial x_1}, \dots, \frac{\partial z_2}{\partial x_n}, \dots, \frac{\partial z_m}{\partial x_1}, \dots, \frac{\partial z_m}{\partial x_n}, \dots, \right.$$
$$\dots, \quad \frac{\partial^2 z_m}{\partial x_1 \partial x_1}, \dots, \quad \frac{\partial^2 z_m}{\partial x_1 \partial x_n}, \dots, \quad \frac{\partial^2 z_m}{\partial x_2 \partial x_1}, \dots, \frac{\partial^2 z_m}{\partial x_2 \partial x_n}, \dots,$$
$$\left. \dots, \frac{\partial^2 z_m}{\partial x_n^2}, \dots \dots, \quad \frac{\partial^r z_m}{\partial x_n^r} \right) = 0 \qquad (16)$$

In this case, the NCSA table would be represented as follow:

$G_k = 0$		
Initial Conditions	*Coefficient values present in the DE or system of DEs*	*Exact analytical solution obtained using the Multivariate Polynomial Transform method*
$z_{01}, z_{02}, \dots, z_{0m}, x_{01}, \dots, x_{0n}$ ...	$a_0, b_0, c_0, \dots$	$U_1 = 0$
$z_{11}, z_{02}, \dots, z_{0m}, x_{01}, \dots, x_{0n}$ ...	$a_1, b_0, c_0, \dots$	$U_2 = 0$
$z_{01}, z_{12}, \dots, z_{0m}, x_{01}, \dots, x_{0n}$ ...	$a_0, b_1, c_0, \dots$	$U_3 = 0$
. . .	. . .	. . .
. . .	. . .	. . .
. . .	. . .	. . .

Table 2.1

where "$U_i = 0$" would then be referred to as an *instance solution* satisfying the unique set of parameters contained in this table.

Example (2.1). For the simple two dimensional case that can be represented by the following *general* first order ODE,

$$x\frac{dy}{dx} + ay + bx^n y^2 = 0 \tag{17}$$

the corresponding NCSA table may be constructed in the following manner:

$x\frac{dy}{dx} + ay + bx^n y^2 = 0$		
Initial Conditions	*Coefficient Values*	*Exact analytical solution obtained using the Multivariate Polynomial Transform method*
$x_0 = 1$ $y_0 = 1$	$a = 1.0$ $b = 1.0$ $n = -1.0$	$(-3x + x^{-1})y + 2 = 0$
$x_0 = 1$ $y_0 = 2$	$a = 1.2$ $b = -1.0$ $n = 2.0$	$(1.4x^{1.2} - x^2)y - 0.80 = 0$
$x_0 = 1$ $y_0 = -1$	$a = 1.2$ $b = 1.5$ $n = -2.0$	$(1.7x^{1.2} + 1.5^{-2})y + 3.2 = 0$
$x_0 = 1$ $y_0 = 2$	$a = 2.0$ $b = -1.0$ $n = 2.0$	$x^2 y(0.5 - \ln(x)) - 1 = 0$
$x_0 = 1$ $y_0 = -2$	$a = 1.5$ $b = 2.0$ $n = 3.0$	$(-2.75x^{1.5} + 2x^3)y - 1.5 = 0$
$x_0 = 1$ $y_0 = 1$	$a = 1.0$ $b = 1.0$ $n = 1.0$	$xy(1 + \ln(x)) - 1.0 = 0$
$x_0 = 1$ $y_0 = -1$	$a = -1.0$ $b = 1.5$ $n = -1.0$	$x^{-1}y(-1 + 1.5\ln(x)) - 1.0 = 0$

Table 2.2

The evidence gathered from each of the above *instance solutions* allows us to conclude by conjecture that:

$$f_1(x,y) \;=\; 0 \;=\; (Ax^B \;+\; Cx^D)y \;+\; E \tag{18}$$

and:

$$f_2(x,y) \;=\; 0 \;=\; x^A y(B \;+\; C\ln(x)) \;+\; D \tag{19}$$

both appear to be perfect candidates for the general exact analytical solution of the ODE where the coefficients "A", "B", "C", "D" and "E" are to be expressed in terms of the coefficients "a", "b", "n" and the initial conditions of the ODE.

By substituting any one of these generally assumed analytical solution into the ODE and equating like terms to zero, we can derive a complete relationship that can exist between the known and the unknown coefficients.

The general formula used for determining the first derivative of "y" is:

$$\frac{dy}{dx} \;=\; -\frac{\partial f}{\partial x}\Big/\frac{\partial f}{\partial y} \tag{20}$$

In our first assumption that "$f_1(x,y) = 0$" and upon equating like terms to zero in the ODE, this would define the following system of equations to solve for:

$$A(a \;-\; B) \;=\; 0 \tag{21}$$

$$C(a \;-\; n) - \; bE \;=\; 0 \tag{22}$$

$$(Ax_0^a \;+\; Cx_0^n)y_0 \;+\; E \;=\; 0 \tag{23}$$

with exact solution [Mikalajunas 2015]:

$$A \;\neq\; 0 \tag{24}$$

$$B \;=\; a \tag{25}$$

$$C \;=\; \frac{-Abx_0^a y_0}{a \;+\; bx_0^n y_0 - \; n} \tag{26}$$

$$E \;=\; \frac{(a \;-\; n)C}{b} \qquad (a \;\neq n) \tag{27}$$

Following the same type of logic for our second assumption that "$f_2(x,y) = 0$", this would define the following system of nonlinear equations to solve for:

$$B(a \;-\; n) - \; C \;-\; bD \;=\; 0 \tag{28}$$

$$C(a \;-\; n) \;=\; 0 \tag{29}$$

$$x_0^n y_0(B \;+\; C\ln(x_0)) \;+\; D \;=\; 0 \tag{30}$$

with exact solution [Mikalajunas 2015]:

$$D \neq 0 \tag{31}$$

$$C = -bD \tag{32}$$

$$B = \frac{-D}{x_0^n y_0} - C\ln(x_0) = \frac{-D - Cx_0^n y_0 \ln(x_0)}{x_0^n y_0} \tag{33}$$

Without having constructed the NCSA table it would have been very difficult to have correctly arrived at the complete "*general analytical solution*" of this first order ODE that would satisfy all initial conditions as well. There are currently no known traditional method of integration capable of deriving complete "*general*" closed form solutions to "*any type*" of DEs and systems of DEs that would be entirely based on the use of a well defined "*exact*" method of computational analysis such as the one being proposed in this paper.

The very unique mathematical properties of an IAMPT when substituted into a DE or a system of DEs allows for all initial conditions to be fully accounted for. This is because the exact integration process that is performed in the *Secondary Expansion* for determining an exact expression for each auxiliary variable must always include the constant of integration which in turn would automatically define each of their initial values. For every *instance solution* obtained, the overall contribution of each of these initial values for the auxiliary variables can easily succeed in completely matching the initial conditions of a DE or a system of DEs. This becomes very obvious by noticing that the *Primary Expansion* of an IAMPT is always expressed as some algebraic combination of initially assumed auxiliary variables as well as known auxiliary variables. Its the initial values of each of these auxiliary variables that can easily be adjusted numerically for satisfying the overall initial conditions of a DE or a system of DEs by solving for the type of system of nonlinear equations in which there will always be more unknowns than available equations to solve them.

Based on our previous example for the general first order ODE, we notice that every *instance solution* obtained would potentially lead towards defining a more generalized version of the exact analytical solution. It is only through the painstaking gathering of this type of information in the form of a large distribution sample of *instance solution* sets can we succeed in determining only by the method of conjecture complete general closed form solutions of a DE or a system of DEs.

The complete consolidation of a large number of these generalized exact analytical solutions which would be the result of having solved for a large number of very distinct classes of DEs and systems of DEs can potentially lead to defining some very fundamental theorems. Case in point is the *superposition theorem* being the result of having solved mostly by *trial and error* a very distinct class of linear second order ODEs.

By consolidating each of these fundamental mathematical theorems into one gigantic universal theory might represent our most *realistic* hope yet of ever arriving at some *unified theory of everything.*

3. The *theory of everything* not just about modern physics anymore

To this day, the most accepted definition of the *theory of everything* is that it must remain an integral part of modern physics on the principle of defining a unique Space-Time model that would explain all the basic laws of this universe.

However, what appears to be clearly lacking in our attempt to create such a *grandiose physical theory* for explaining everything about this universe is an equivalent *grandiose mathematical theory* that would have to succeed in explaining everything about the complete analytical integration of all types of DEs as well as all types of systems of DEs.

Because DEs are completely universal and not linked to any specific area of the physical sciences, there is really no evidence to support that modern physics is the only real subject by which a complete theory of everything may be entirely constructed from.

Rather, it would have to be through the application of some unified theory of analytical integration that a theory of everything would be achievable. This would be result of consolidating each fundamental theorem associated with a single *Unified Physical System* at a time into one gigantic theory capable of explaining everything about this physical universe.

The following block diagram suggests such a scenario by which DEs and systems of DEs would play a central role for establishing such a theory of everything where each *Unified Physical System* would have its own very unique story to tell us that in the end we would need to know about:

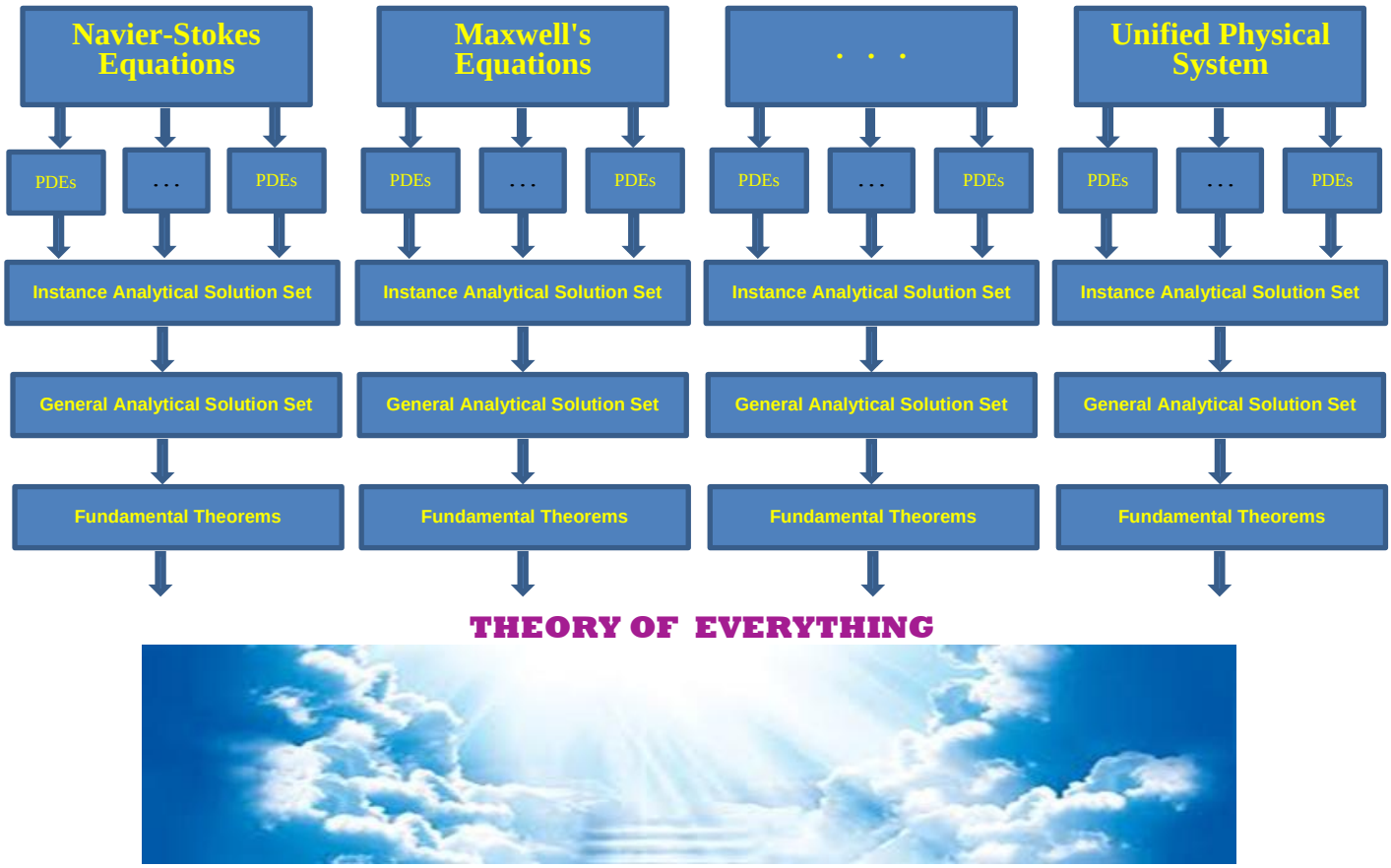


Figure 3.1

The very mathematical nature of our proposed unified theory of analytical integration is built on the principle that "***analytical solutions***" to DEs and systems of DEs *must* be constructed entirely on pure computational analysis.

In the absence of a unified theory of analytical integration, our understanding of the physical sciences cannot be complete as our method of analysis becomes reduced to a process that is mostly governed by unpredictable events. Because Calculus is so deeply embedded into all of the physical sciences, how can we expect to devise a *theory of everything* without the use of some form of a *unified analytical theory of integration that would be entirely driven by some well defined method of exact computational analysis* ?

4. Complete numerical example for a second order ODE

In our first example for the general first order ODE, we highlighted the importance of creating a special type of table called the *NCSA table* for providing much greater visibility towards the acquisition of *general* closed form solutions. Such a table would be constructed on the principle of creating a special type of database that would consist of a large number of instance solutions each satisfying a predetermined number of control parameters that would include initial conditions and all the variable coefficients that take part in defining a DE or a system of DEs.

Corresponding to a unique set of control parameters would define a unique instance solution that would be obtained as a result of substituting an IAMPT into a DE or a system of DEs and numerically solving for the resultant system of nonlinear simultaneous equations. This would be followed by the complete transformation of the resultant IAMPT into a unique instance solution.

As the number of instance solutions grows, this would allow for much greater insight in determining by method of conjecture if a more *general* analytical solution actually exists. These types of closed form solutions have a far greater capacity towards a much better understanding on the very long term behavior of a physical system. By consolidating each and every *general* analytical solutions obtained over a large class of DEs and systems of DEs into basic fundamental theorems, an even far much better understanding of the same physical system is possible. Only as we progress further in the complete formulation of a large number of such specialized fundamental theorems can we expect to move closer towards the complete development of some form of a *theory of everything*.

In the following example, we have randomly selected a second order ODE and provided a complete step by step process for arriving at its complete exact analytical solution satisfying all initial conditions.

Example (4.1). Starting with the following *second* order ODE:

$$y\frac{d^2y}{dx^2} \; - \; \left(\frac{dy}{dx}\right)^2 \left\{1 \; - \; \frac{dy}{dx}Sin(y) \; - \; y\frac{dy}{dx}Cos(y)\right\} \; = \; 0 \tag{34}$$

there are two *external* inputs that are defined in terms of the Sine and Cosine function.

For the sake of simplicity in our analysis, we can use the following identities for expressing each of the two trigonometric functions as a *rational* combination of the half angle tangent function:

$$Sin(u) = \frac{2Tan(u/2)}{1 + Tan^2(u/2)} \tag{35}$$

$$Cos(u) = \frac{1 - Tan^2(u/2)}{1 + Tan^2(u/2)} \tag{36}$$

Based on the use of this half angle formula for the Tangent function, we begin by selecting a much simpler alternative representation for the Sine and Cosine function by defining:

$$H \; = \; Tan(y/2) \; = \; W_{p+1} \tag{37}$$

where "p" is the total number of arbitrarily defined auxiliary variables from the IAMPT that will be selected for solving this second order ODE.

For this choice of auxiliary variable the corresponding *Multivariate Polynomial Transform* would be defined as follow:

(1). *Primary Expansion:*

$$H(W_{p+1}) = W_{p+1} \tag{38}$$

(2). *Secondary Expansion:*

$$dy = dW_2 \tag{39}$$

$$0 \cdot dx + (1 + W_{p+1}^2)dy = 2dW_{p+1} \tag{40}$$

We can arbitrarily select our IAMPT as consisting of a maximum of *five* arbitrarily defined auxiliary variables so that "$p = 5$". There will be a total number of *six* terms in the *Primary Expansion* so that "$u_P = 6$" and a total number of *four* terms in the *Secondary Expansion* so that "$u_S = 4$". Because there is only one external input in the form of the Tangent function for representing both the Sine and Cosine function, "$q = 1$" thereby bringing the total number of auxiliary variables in the entire initially assumed expansion to *six*.

For this selection of parameters, the corresponding IAMPT for solving this second order ODE can be expanded as:

(1). *Primary Expansion:*

$$F = 0 = a_1 W_1^{m_1} W_2^{m_2} \cdots W_6^{m_6} + a_2 W_1^{m_7} W_2^{m_8} \cdots W_6^{m_{12}} + \ldots + + \ldots + a_6 W_1^{m_{31}} W_2^{m_{32}} \cdots W_6^{m_{36}} \tag{41}$$

(2). *Secondary Expansion:*

$$dx = dW_1 \tag{42}$$

$$dy = dW_2 \tag{43}$$

$$N_1 dx + N_2 dy = N_3 dW_3 \tag{44}$$

$$N_4 dx + N_5 dy = N_6 dW_4 \tag{45}$$

$$N_7 dx + N_8 dy = N_9 dW_5 \tag{46}$$

$$N_{10} dx + N_{11} dy = N_{12} dW_6 \tag{47}$$

where:

$$N_1 = b_1 W_1^{m_1} W_2^{m_2} \cdots W_6^{m_6} + \ldots + b_4 W_1^{m_{19}} W_2^{m_{20}} \cdots W_6^{m_{24}} \tag{48}$$

$$N_2 = b_5 W_1^{m_{25}} W_2^{m_{26}} \cdots W_6^{m_{30}} + \ldots + b_8 W_1^{m_{45}} W_2^{m_{46}} \cdots W_6^{m_{48}} \tag{49}$$

. . .

. . .

. . .

$$N_9 = b_{33} W_1^{m_{193}} W_2^{m_{194}} \cdots W_6^{m_{198}} + \ldots + b_{36} W_1^{m_{211}} W_2^{m_{212}} \cdots W_6^{m_{216}} \quad (50)$$

To account for the presence of both the Sine and Cosine function inside the ODE we must define the following three multivariate polynomials with *known* coefficient values:

$$N_{10} = 0 \quad (51)$$

$$N_{11} = 1 + W_{p+1}^2 = 1 + W_6^2 \quad (52)$$

and :

$$N_{12} = 2 \quad (53)$$

We can compute the total number of unknowns to solve for in our IAMPT using the following general formula with "$p = 5$", "$u_P = 6$", "$u_s = 4$" and "$q = 1$" :

$$N_{Total} = N_{Primary} + N_{Secondary} \quad (54)$$

$$= u_P(p + q + 1) + 3\,u_S(p + q + 1)(p - 2) \quad (55)$$

$$= 6(5 + 1 + 1) + 3(4)(5 + 1 + 1)(5 - 2) \quad (56)$$

$$= 6(7) + 12(7)(3) = 42 + 252 = 294 \quad (57)$$

We can express the entire ODE in terms of the following single large multivariate polynomial by taking its complete *Multivariate Polynomial Transform* using equation (35), (36) and (37):

$$W_2 \frac{d^2Y}{dX^2} - \left(\frac{dY}{dX}\right)^2 \left\{1 - \frac{dY}{dX}\left(\frac{2W_{p+1}}{1 + W_{p+1}^2}\right) - W_2 \frac{dY}{dX}\left(\frac{1 - W_{p+1}^2}{1 + W_{p+1}^2}\right)\right\} = 0 \quad (58)$$

where we have selected:

$$W_1 = X \quad (59)$$

$$W_2 = Y \quad (60)$$

and where capital letters are used to indicate that a transformation from rectangular to complete multivariate polynomial form has taken place.

A very general formula for calculating the first derivative of a general IAMPT may be defined as:

$$\frac{dY}{dX} = \frac{P_1}{Q_1} = \frac{-\frac{\partial F}{\partial W_1}\prod_{k=1}^{p+q-2} N_{3k} - \sum_{j=3}^{p+q}\left\{N_{3j-8}\frac{\partial F}{\partial W_j}\prod_{\substack{k=1\\k\neq j-2}}^{p+q-2} N_{3k}\right\}}{\frac{\partial F}{\partial W_2}\prod_{k=1}^{p+q-2} N_{3k} + \sum_{j=3}^{p+q}\left\{N_{3j-7}\frac{\partial F}{\partial W_j}\prod_{\substack{k=1\\k\neq j-2}}^{p+q-2} N_{3k}\right\}} \tag{61}$$

where both P_1 and Q_1 are each defined as a multivariate polynomial.

By expressing this equation in the following form:

$$\frac{dY}{dX}Q_1 - P_1 = 0 \tag{62}$$

we can *numerically* determine the second and higher derivatives of the dependent variable by successively differentiating both sides of this equation using the product rule and the general formula provided in equation (61).

Section 6 describes an *exact computational method* for calculating the various derivative of a product of two or more expressions using the *Multinomial Expansion Theorem* without resorting to any type of symbolic algebraic manipulation.

Our system of nonlinear simultaneous equations of interest to solve for is obtained by first taking the various derivatives of equation (58) that represents the ODE in complete multivariate polynomial form. This would include the various derivatives of each auxiliary variable that define the *Multivariate Polynomial Transform* of the single external input as provided in equations (37) through (40) which are "W_2" and "W_{p+1}".

Next, we replace the various derivatives of the dependent variable in equation (58) with the computed values obtained from the various derivatives of our IAMPT using equations (61) and (62).

The resultant nonlinear simultaneous equations can then be numerically solved for using various optimization technics where our objective function to be minimized would be represented as the sum of the squares of each of the various derivatives of equation (58):

$$G_n = \frac{d^n}{dx^n}\left[W_2\frac{d^2Y}{dX^2} - \left(\frac{dY}{dX}\right)^2\left\{1 - \frac{dY}{dX}\left(\frac{2W_{p+1}}{1+W_{p+1}^2}\right) - W_2\frac{dY}{dX}\left(\frac{1-W_{p+1}^2}{1+W_{p+1}^2}\right)\right\}\right] \tag{63}$$

Our main objective function to minimize would thus be represented as:

$$F = \sum_n G_n^2 \tag{64}$$

By succeeding in completely minimizing the above objective function to zero, the corresponding inverse *Multivariate Polynomial Transform* would define an *exact* analytical solution of the ODE that would satisfy a completely random set of initial conditions. Such a type of analytical solution obtained was earlier described as an *instance solution*. Any numerical solution set that would depart from this minima would represent only an approximation of the actual *exact* analytical solution of the ODE. The further away we are from this minima, the greater will be the error of approximation between the exact analytical solution and the one arrived at.

As we are only interested in obtaining as many exact instance solutions as possible each satisfying their own very unique initial conditions when solving for these nonlinear simultaneous equations, we must treat all initial values of the auxiliary variables as unknown coefficients to solve for in order to achieve the highest numerical solution set rate possible. It is the initial values of each auxiliary variable defined from the exact integration of a *Secondary Expansion* that when substituted into the *Primary Expansion* would completely define the initial conditions of a DE or a system of DEs. Keep in mind that our primary objective in this type of analysis is to acquire as many instance solutions as possible so that by applying a unique method of conjecture, we would be able to arrive at a more *generalized* version of the closed form solution satisfying a DE or a system of DEs.

For solving these nonlinear simultaneous equations using an optimization technic, all gradient calculations can become fairly complex quite often leading to very unpredictable results. A preferred method of optimization that generally does not require any type of gradient calculations is the *pattern search method* as described in the book by [Adby and Dempster 1974].

All calculations involving very high order partial derivatives of an IAMPT require a great deal amount of precision and thus not recommended to be performed on a regular PC. Instead, the entire computational process would become more manageable if it were conducted on a very *advanced super computer system*.

Future generations of computer hardware may begin to take full advantage of the multistate quantum bit (or Qubit) technology originating from the principles of quantum physics as they are expected to become much more powerful than the conventional types that operate only on the principle of two states being a 0 or 1. Over time the semi conductor industry that currently powers our conventional computers will eventually reach its own physical limitations in terms of its ability for designing super fast switching devices. Some estimate that because of the multi state capability of a Qubit, it would succeed in outperforming even the most powerful conventional super computer of our time in the *billion-fold* under the most demanding condition of computational requirements.

Upon the gathering of as many numerical solution sets of the nonlinear simultaneous equations as possible, the next step to follow afterwards is in the complete construction of an NCSA table that would be very specific to the particular DE or system of DEs being solved for.

For solving our second order ODE, we were able to acquire a large number of instance solutions each satisfying its own very unique set of initial conditions that would also become the initial conditions of the ODE as well. The greater the number of instance solutions that can be gathered and fully documented accordingly, the greater is the amount of information that can be made available for facilitating the entire process of deducing by *conjecture* the complete general exact analytical solution of the second order ODE.

The NCSA table for our example of a second order ODE would therefore appear as follow:

$y\frac{d^2y}{dx^2} - \left(\frac{dy}{dx}\right)^2\left\{1 - \frac{dy}{dx}Sin(y) - y\frac{dy}{dx}Cos(y)\right\} = 0$		
Initial Conditions	*Coefficient Values*	*Exact analytical solution obtained using the Multivariate Polynomial Transform method*
$x_0 = -1.28$ $y_0 = 1.591$	N/A	$Cos(y) + x + 1.662 - 0.778\,ln(y)$
$x_0 = 0.2473$ $y_0 = 0.76$	N/A	$Cos(y) + x - 0.111 + 3.138\,ln(y)$
$x_0 = -3.2542$ $y_0 = 1.442$	N/A	$Cos(y) + x + 2.662 + 1.267\,ln(y)$
$x_0 = 1.2223$ $y_0 = 3.865$	N/A	$Cos(y) + x + 0.579 - 0.778\,ln(y)$
$x_0 = -0.837$ $y_0 = 2.691$	N/A	$Cos(y) + x - 1.051 + 2.817\,ln(y)$
$x_0 = -1.668$ $y_0 = 1.877$	N/A	$Cos(y) + x - 0.871 + 4.511\,ln(y)$

Table 4.1

Based entirely on the information provided in this table and following the same basic procedure as was done in our first example for a first order ODE, a plausible conjecture for the exact analytical solution of this second order ODE satisfying all initial conditions would be:

$$f(x,y) = 0 = Cos(y) + x + A_1 + A_2\,ln(y) \tag{65}$$

where "A_1" and "A_1" are each defined as a constant of integration.

5. Complete numerical example for a second order PDE

For PDEs and for systems of ODEs as well as for system of PDEs, the NCSA table is always constructed in pretty much the same way as we did for the first order ODE described in the first example. In all cases involved, we always allow for the initial conditions of a DEs or a system of DEs to become part of the unknown coefficients to solve for as originally defined from within an IAMPT.

In the following example, we have randomly selected a second order PDE and provided a complete step by step process for arriving at its complete exact analytical solution satisfying all initial conditions.

Example (5.1). For the following second order PDE :

$$x_2\left(\frac{\partial^2 z}{\partial x_1 \partial x_2}\right) - \frac{\partial z}{\partial x_1} - x_1 x_2^2 Sin(x_1 x_2) = 0 \tag{66}$$

there is only one external input that is defined in terms of the Sine function.

As we did in our previous example for a second order ODE, we can use the following trigonometric identity for expressing the Sine function as a *rational* combination of the tangent function:

$$f(x_1, x_2) = Sin(x_1 x_2) = \frac{2Tan(x_1 x_2/2)}{1 + Tan^2(x_1 x_2/2)} \tag{67}$$

Based on the use of this half angle formula for the Tangent function, we begin by selecting a much simpler alternative representation for the Sine function by defining:

$$H(x_1, x_2) = W_{p+1} = Tan(x_1 x_2/2) = Tan(W_2 W_3/2) \tag{68}$$

where "p" is the total number of arbitrarily defined auxiliary variables from the IAMPT that will be selected for solving this second order PDE.

For this choice of auxiliary variable the corresponding *Multivariate Polynomial Transform* would be defined as follow:

(1). <u>*Primary Expansion:*</u>

$$H(W_{p+1}) = W_{p+1} \tag{69}$$

(2). <u>*Secondary Expansion:*</u>

$$0 \cdot dz + \left(1 + W_{p+1}^2\right) W_3 dx_1 + \left(1 + W_{p+1}^2\right) W_2 dx_2 = 2dW_{p+1} \tag{70}$$

where we have selected:

$$W_1 = z \tag{71}$$

$$W_2 = x_1 \tag{72}$$

and:

$$W_3 = x_2 \tag{73}$$

We can arbitrarily select our IAMPT as consisting of a maximum of *eight* arbitrarily defined auxiliary variables so that "$p = 8$". There will be a total number of *eight* terms in the *Primary Expansion* so that "$u_P = 8$" and a total number of *four* terms in the *Secondary Expansion* so that "$u_S = 4$". Because there is only one external input in the form of the Tangent function for representing only the Sine function, "$q = 1$" thereby bringing the total number of auxiliary variables in the entire initially assumed expansion to *nine*.

For this selection of parameters, the corresponding IAMPT for solving this second order PDE can be expanded as:

(1). *Primary Expansion:*

$$F = 0 = a_1 W_1^{m_1} W_2^{m_2} \cdots W_9^{m_9} + a_2 W_1^{m_{10}} W_2^{m_{11}} \cdots W_9^{m_{18}} + \ldots + \\ + \ldots + a_8 W_1^{m_{64}} W_2^{m_{65}} \cdots W_9^{m_{72}} \quad (74)$$

(2). *Secondary Expansion:*

$$dz = dW_1 \quad (75)$$

$$dx_1 = dW_2 \quad (76)$$

$$dx_2 = dW_3 \quad (77)$$

$$N_1 dz + N_2 dx_1 + N_3 dx_2 = N_4 dW_4 \quad (78)$$

$$N_5 dz + N_6 dx_1 + N_7 dx_2 = N_8 dW_5 \quad (79)$$

$$N_9 dz + N_{10} dx_1 + N_{11} dx_2 = N_{12} dW_6 \quad (80)$$

$$N_{13} dz + N_{14} dx_1 + N_{15} dx_2 = N_{16} dW_7 \quad (81)$$

$$N_{17} dz + N_{18} dx_1 + N_{19} dx_2 = N_{20} dW_8 \quad (82)$$

$$N_{21} dz + N_{22} dx_1 + N_{23} dx_2 = N_{24} dW_9 \quad (83)$$

where :

$$N_1 = b_1 W_1^{m_1} W_2^{m_2} \cdots W_9^{m_9} + \ldots + b_4 W_1^{m_{28}} W_2^{m_{29}} \cdots W_9^{m_{36}} \quad (84)$$

$$N_2 = b_5 W_1^{m_{37}} W_2^{m_{38}} \cdots W_9^{m_{45}} + \ldots + b_8 W_1^{m_{64}} W_2^{m_{65}} \cdots W_9^{m_{72}} \quad (85)$$

. . .

. . .

. . .

$$N_{20} = b_{77} W_1^{m_{685}} W_2^{m_{686}} \cdots W_9^{m_{693}} + \ldots + b_{80} W_1^{m_{712}} W_2^{m_{713}} \cdots W_9^{m_{720}} \quad (86)$$

To account for the presence of the Sine function inside the PDE we must define the following three multivariate polynomials with *known* coefficient values:

$$N_{21} = 0 \tag{87}$$

$$N_{22} = (1 + W_{p+1}^2)W_3 = (1 + W_9^2)W_3 \tag{88}$$

$$N_{23} = (1 + W_{p+1}^2)W_2 = (1 + W_9^2)W_2 \tag{89}$$

and :

$$N_{24} = 2 \tag{90}$$

We can compute the total number of unknowns to solve for in our IAMPT using the following general formula with "$n = 2$", "$p = 8$", "$u_P = 8$", "$u_s = 4$" and "$q = 1$" :

$$N_{Total} = N_{Primary} + N_{Secondary} \tag{91}$$

$$= u_P(p + q + 1) + u_S(p + q + 1)(n + 2)(p - n - 1) \tag{92}$$

$$= 8(8 + 1 + 1) + 4(8 + 1 + 1)(2 + 2)(8 - 2 - 1) \tag{93}$$

$$= 8(10) + 4(10)(4)(5) = 80 + 800 = 880 \tag{94}$$

As in the case for the second order ODE, the entire PDE may be expressed in terms of the following single large multivariate polynomial by taking its complete *Multivariate Polynomial Transform* using equations (68) through (73):

$$W_3\left(\frac{\partial^2 Z}{\partial W_2 \partial W_3}\right) - \frac{\partial Z}{\partial W_2} - 2W_2W_3^2\left(\frac{W_{p+1}}{1 + W_{p+1}^2}\right) = 0 \tag{95}$$

where we have selected:

$$W_1 = z \tag{96}$$

$$W_2 = x_1 \tag{97}$$

$$W_3 = x_2 \tag{98}$$

and where capital letters are used to indicate that a transformation to complete multivariate polynomial form has taken place.

A very general formula for calculating the first partial derivative of our IAMPT that is based on the use of the product rule and the *Multinomial Expansion Theorem* can also be derived in a very similar manner as was done in our last example of a second order ODE which was provided in equation (61).

Our system of nonlinear simultaneous equations of interest to solve for is obtained by first taking the various partial derivatives of equation (95) that represents the PDE in complete multivariate polynomial form. This would include the various partial derivatives of each auxiliary variable that define the *Multivariate Polynomial Transform* of the single external input as provided in equations (68) through (73) which are "W_2", "W_3" and "W_{p+1}".

Next, we replace the various partial derivatives of the dependent variable in equation (95) with the computed values obtained from the various partial derivatives of our IAMPT.

The resultant nonlinear simultaneous equations can then be numerically solved for using various optimization technics where our objective function to be minimized would be represented as the sum of the squares of each of the various partial derivatives of equation (95):

$$G_i = \frac{\partial^{m_1}}{\partial W_2^{m_1}} \frac{\partial^{m_2}}{\partial W_3^{m_2}} \frac{\partial^{m_3}}{\partial W_4^{m_3}} \cdots \frac{\partial^{m_k}}{\partial W_{p+1}^{m_k}} \cdots \left\{ W_3 \left(\frac{\partial^2 Z}{\partial W_2 \partial W_3} \right) - \frac{\partial Z}{\partial W_2} - 2W_2 W_3^2 \left(\frac{W_{p+1}}{1 + W_{p+1}^2} \right) \right\} = 0 \qquad (99)$$

Our main objective function to minimize would therefore be represented as:

$$F = \sum_i G_i^2 \qquad (100)$$

By succeeding in completely minimizing the above objective function to zero, the corresponding inverse *Multivariate Polynomial Transform* would define an *exact* analytical solution of the PDE that would satisfy a completely random set of initial conditions. Such a type of analytical solution obtained was earlier described as an *instance solution*. Any numerical solution set that would depart from this minima would represent only an approximation of the actual *exact* analytical solution of the PDE. The further away we are from this minima, the greater will be the error of approximation between the exact analytical solution and the one arrived at.

As we are only interested in obtaining as many exact instance solutions as possible each satisfying their own very unique initial conditions when solving for these nonlinear simultaneous equations, we must treat all initial values of the auxiliary variables as unknown coefficients to solve for in order to achieve the highest numerical solution set rate possible. It is the initial values of each auxiliary variable defined from the exact integration of a *Secondary Expansion* that when substituted into the *Primary Expansion* would completely define the initial conditions of a DE or a system of DEs. Keep in mind that our primary objective in this type of analysis is to acquire as many instance solutions as possible so that by applying a unique method of conjecture, we would be able to arrive at a more *generalized* version of the closed form solution satisfying a DE or a system of DEs.

For solving these nonlinear simultaneous equations using an optimization technic, all gradient calculations can become fairly complex quite often leading to very unpredictable results. A preferred method of optimization that generally does not require any type of gradient calculations is the *pattern search method* as described in the book by [Adby and Dempster 1974].

All calculations involving very high order partial derivatives of an IAMPT require a great deal amount of precision and thus not recommended to be performed on a regular PC. Instead, the entire computational process would become more manageable if it were conducted on a very *advanced super computer system*.

Upon the gathering of as many numerical solution sets of the nonlinear simultaneous equations as possible, the next step to follow afterwards is in the complete construction of an NCSA table that would be very specific to the particular DE or system of DEs being solved for.

For solving our second order PDE, we were able to acquire a large number of instance solutions each satisfying its own very unique set of initial conditions that would also become the initial conditions of the PDE as well. The greater the number of instance solutions that can be gathered and fully documented accordingly, the greater is the amount of information that can be made available for facilitating the entire process of deducing by *conjecture* the complete general exact analytical solution of the second order PDE.

The NCSA table for our example of a second order PDE would therefore appear as follow:

$$x_2\left(\frac{\partial^2 z}{\partial x_1 \partial x_2}\right) - \frac{\partial z}{\partial x_1} - x_1 x_2^2 Sin(x_1 x_2) = 0$$

Initial Conditions	*Coefficient Values*	*Exact analytical solution obtained using the Multivariate Polynomial Transform method*
$x_{01} = 3.61$ $x_{02} = 1.771$	N/A	$2x_2 x_1^{1.68} + Sin(\ln[x_2^{-1.6}] + x_2^{0.78}) - Sin(x_1 x_2) - z = 0$
$x_{01} = 1.29$ $x_{02} = -1.88$	N/A	$x_2 \sqrt[6]{x_1^{0.23} + 1.78} + 1.22\ln\left(\sqrt{x_2^2 + 1} + 3.5\right) - Sin(x_1 x_2) - z = 0$
$x_{01} = 3.555$ $x_{02} = 2.76$	N/A	$0.56x_2 e^{x_1^{-0.46}} - 4.6Tan(x_2^{1.86} + \sqrt[4]{x_2^{1.1} - 6.1}) - Sin(x_1 x_2) - z = 0$
$x_{01} = -0.723$ $x_{02} = 1.58$	N/A	$3.06x_2 Sinh(x_1^2) - 2.45x_2^{1.46\sqrt{x_2^{3.1} - 2.3}} - Sin(x_1 x_2) - z = 0$

Table 5.1

Based entirely on the information provided in the above table, there appears to be no obvious patterns by which a plausible conjecture for the exact analytical solution of this second order PDE satisfying all initial conditions can be made.

The main reason for this is that the exact analytical solution consists of a number of expressions that are completely arbitrarily defined. This would call for the development of a very sophisticated method of *comparison analysis* just for identifying those arbitrary expressions that are present in all of the instance solutions obtained. Some of these arbitrarily defined expressions may be easier to detect than others for establishing a plausible conjecture by which a complete analytical solution of the PDE satisfying all initial conditions may be arrived at.

In the final analysis, all results would be pointing towards the following expression as representing the complete exact analytical solution of the PDE satisfying all initial conditions:

$$f(z, x_1, x_2) = 0 = x_2\varphi_1(x_1) + \varphi_2(x_2) - Sin(x_1 x_2) - z \qquad (101)$$

where upon conducting such a type of special method of *comparison analysis*, each of the expression for "$\varphi_1(x_1)$" and "$\varphi_2(x_2)$" would eventually have been singled out in the end as completely arbitrarily defined.

Once again it is very important to mention that without having constructed the NCSA table it would have been virtually impossible to have correctly arrived at the complete *general* analytical solution of this second order PDE satisfying all initial conditions.

6. Exact computational method for calculating the various derivatives and partial derivatives of an initially assumed Multivariate Polynomial Transform (IAMPT)

The method of substituting an IAMPT into a DE or a system of DEs for defining a valid system of nonlinear simultaneous equations to solve for requires that the numerical values of each of the various derivatives of the DE or system of DEs become equal to that of an IAMPT. An alternative method is to substitute an IAMPT into a DE or a system of DEs and afterwards equating like multivariate polynomial terms to zero. However, this would result into defining a completely invalid system of nonlinear simultaneous equations to solve for as it would automatically impose a major restriction on each auxiliary variable for becoming totally independent from one another. The evidence is clearly provided in Appendix A where as you will notice that for the vast majority of the cases involved, it is always necessary to maintain a certain degree of dependency among auxiliary variables especially when very complex mathematical equations are involved.

The actual process of computing the *exact* values for the various derivatives and partial derivatives of an IAMPT to any desirable order of differentiation without any loss of accuracy whatsoever can always be reduced at a *computational level*. The reason for this is that we take full advantage of a well known fact in numerical analysis that taking the various derivatives of a product of several expressions is very much similar to algebraically expanding to some exponent value the sum of several terms. The only major difference between the two is that in the case of differentiation, exponentiation becomes treated purely as an order of differentiation while all the remaining algebraic operations remain completely identical.

For the simple case of differentiating a product involving only two expressions, this would require the use of the *Binomial Expansion Theorem* which is defined by:

$$\frac{d^n}{dx^n} fg \;=\; \sum_{k=0}^{n} \binom{n}{k} f^{(k)} g^{(n-k)} \tag{102}$$

where:

$$\binom{n}{k} \;=\; B_{n,k} \;=\; \frac{n!}{k!\,(n-k)!} \tag{103}$$

are the binomial coefficients and where it is to be clearly understood that all exponent values are to be treated purely as order of differentiation.

In complete expanded form, the various derivatives of a product consisting of *two* expressions can be *symbolically* defined as :

$$[f+g]^{(n)} \;=\; f^{(0)}g^{(n)} \;+\; B_{n-1,1} f^{(1)} g^{(n-1)} \;+\; B_{n-2,2} f^{(2)} g^{(n-2)} \;+\; \ldots \;+\; f^{(n)} g^{(0)} \tag{104}$$

where the product is being substituted by the sum inside a square bracket and "n" is the order of differentiation.

When a product always involves more than two expressions, we can instead replace the *Binomial Expansion Theorem* with the following *Multinomial Expansion Theorem*:

$$(a_1 \;+\; a_2 \;+\cdots+\; a_k)^n \;=\; \sum_{\substack{n_1,n_2,\ldots,n_k \geq 0 \\ n_1+n_2+\cdots+n_k=n}} \frac{n!}{n_1!\,n_2!\cdots n_k!}\; a_1^{(n_1)}\, a_2^{(n_2)} \;\cdots\; a_k^{(n_k)} \tag{105}$$

where $n = n_1 \;+\; n_2 \;+\; \ldots \;+\; n_k$

For determining the various derivatives of a product involving any number of expressions and in accordance to our previously defined notation we can define:

$$\frac{d^n}{dx^n}(f_1 f_2 \cdots f_k) \;=\; [f_1 \;+\; f_2 \;+\cdots+\; f_k]^{(n)} \tag{106}$$

$$= \sum_{\substack{n_1,n_2,\ldots,n_k \geq 0 \\ n_1+n_2+\cdots+n_k=n}} \frac{n!}{n_1!\, n_2! \cdots n_k!}\; f_1^{(n_1)}\, f_2^{(n_2)} \;\ldots\; f_k^{(n_k)} \tag{107}$$

where the square bracket is used to symbolize differentiation with all exponents treated as order of differentiation.

Example (6.1). To test the validity of our symbolic notation, let us consider the simple two dimensional case for calculating the various derivatives up to the 5th order at "x = 2" for the following equation:

$$y \;=\; e^{2x} \;=\; e^{-x} e^{0.5x} e^{2.5x} \tag{108}$$

Here we can start by letting:

$$\mathrm{f_1} = \mathrm{e^{-x}},\quad \mathrm{f_2} = \mathrm{e^{0.5x}} \text{ and } \quad \mathrm{f_3} = \mathrm{e^{2.5x}} \tag{109}$$

so that each of their various derivatives up to 5 may be defined as:

$$f_1^{(0)} = \mathrm{e^{-x}},\; f_2^{(0)} = \mathrm{e^{0.5x}} \text{ and } \; f_3^{(0)} = \mathrm{e^{2.5x}} \tag{110}$$

$$f_1^{(1)} = -\mathrm{e^{-x}},\; f_2^{(1)} = 0.5\mathrm{e^{0.5x}} \text{ and } \; f_3^{(1)} = 2.5\mathrm{e^{2.5x}} \tag{111}$$

$$f_1^{(2)} = \mathrm{e^{-x}},\; f_2^{(2)} = 0.25\mathrm{e^{0.5x}} \text{ and } \; f_3^{(2)} = 6.25\mathrm{e^{2.5x}} \tag{112}$$

$$f_1^{(3)} = -\mathrm{e^{-x}},\; f_2^{(3)} = 0.125\mathrm{e^{0.5x}} \text{ and } \; f_3^{(3)} = 15.625\mathrm{e^{2.5x}} \tag{113}$$

$$f_1^{(4)} = \mathrm{e^{-x}},\; f_2^{(4)} = 0.0625\mathrm{e^{0.5x}} \text{ and } \; f_3^{(4)} = 39.0625\mathrm{e^{2.5x}} \tag{114}$$

$$f_1^{(5)} = -\mathrm{e^{-x}},\; f_2^{(5)} = 0.03125\mathrm{e^{0.5x}} \text{ and } \; f_3^{(5)} = 97.65625\mathrm{e^{2.5x}} \tag{115}$$

At "x = 0.5" we thus have:

$$f_1^{(0)} = \mathrm{e^{-0.5}} = 0.607,\; f_2^{(0)} = \mathrm{e^{0.25}} = 1.284 \text{ and } \; f_3^{(0)} = \mathrm{e^{1.25}} = 3.490 \tag{116}$$

$$f_1^{(1)} = -\mathrm{e^{-0.5}} = -0.607,\; f_2^{(1)} = 0.5\mathrm{e^{0.25}} = 0.642 \text{ and } \; f_3^{(1)} = 2.5\mathrm{e^{1.25}} = 8.726 \tag{117}$$

$$f_1^{(2)} = \mathrm{e^{-0.5}} = 0.607,\; f_2^{(2)} = 0.25\mathrm{e^{0.25}} = 0.321 \text{ and } \; f_3^{(2)} = 6.25\mathrm{e^{1.25}} = 21.815 \tag{118}$$

$$f_1^{(3)} = -\mathrm{e^{-0.5}} = -0.607,\; f_2^{(3)} = 0.125\mathrm{e^{0.25}} = 0.161 \text{ and } \; f_3^{(3)} = 15.625\mathrm{e^{1.25}} = 54.537 \tag{119}$$

$$f_1^{(4)} = \mathrm{e^{-0.5}} = 0.607,\; f_2^{(4)} = 0.0625\mathrm{e^{0.25}} = 0.080 \text{ and } \; f_3^{(4)} = 39.0625\mathrm{e^{1.25}} = 136.342 \tag{120}$$

$$f_1^{(5)} = -\mathrm{e^{-0.5}} = -0.607,\; f_2^{(5)} = 0.03125\mathrm{e^{0.25}} = 0.040 \text{ and } \; f_3^{(5)} = 97.65625\mathrm{e^{1.25}} = 340.854 \tag{121}$$

Applying the *Multinomial Expansion Theorem* on these three individual components, we arrive at:

$$\frac{d^5y}{dx^5} = [f_1 + f_2 + f_3]^{(5)} = \sum_{\substack{n_1,n_2,n_3\geq 0 \\ n_1+n_2+n_3=5}} \frac{n!}{n_1!\,n_2!\,n_3!}\, f_1^{(n_1)}\, f_2^{(n_2)}\, f_3^{(n_3)} \tag{122}$$

$$\begin{aligned} = \; & (\mathbf{1})(-0.607)(1.284)(3.490) + (\mathbf{5})(0.607)(0.642)(3.490) + (\mathbf{10})(-0.607)(0.321)(3.490) + (\mathbf{10})(0.607)(0.161)(3.490) + \\ & (\mathbf{5})(-0.607)(0.080)(3.490) + (\mathbf{1})(0.607)(0.040)(3.490) + (\mathbf{5})(0.607)(1.284)(8.726) + (\mathbf{20})(-0.607)(0.642)(8.726) + \\ & (\mathbf{30})(0.607)(0.321)(8.726) + (\mathbf{20})(-0.607)(0.161)(8.726) + (\mathbf{5})(0.607)(0.080)(8.726) + (\mathbf{10})(-0.607)(1.284)(21.815) + \\ & (\mathbf{30})(0.607)(0.642)(21.815) + (\mathbf{30})(-0.607)(0.321)(21.815) + (\mathbf{10})(0.607)(0.161)(21.815) + (\mathbf{10})(0.607)(1.284)(54.537) + \\ & (\mathbf{20})(-0.607)(0.642)(54.537) + (\mathbf{10})(0.607)(0.321)(54.537) + (\mathbf{5})(-0.607)(1.284)(136.342) + (\mathbf{5})(0.607)(0.642)(136.342) + \\ & (\mathbf{1})(0.607)(1.284)(340.854) \end{aligned} \tag{123}$$

where there are a total number of 21 terms satisfying the criteria that "$n_1, n_2, n_3 \geq 0$" and "$n_1 + n_2 + n_3 = 5$".

We can define the *multinomial coefficient vector* has having a total number of 21 elements and these are:

$$C_M = [\,1, 5, 10, 10, 5, 1, 5, 20, 30, 20, 5, 10, 30, 30, 10, 10, 20, 10, 5, 5, 1\,] \tag{124}$$

We can also define the *multinomial exponent vector* as also consisting of 21 elements and they are:

$$E_M = [500, 410, 320, 230, 140, 050, 401, 311, 221, 131, 041, 302, 212, 122, 032, 203, 113, 023, 104, 014, 005] \tag{125}$$

By writing a short computer program for performing the arithmetical operation in equation (123) using equation (122) but with higher precision, the value obtained based on the *Multinomial Expansion Theorem* was determined as "$\mathbf{86.985019}$".

The 5th derivative of "e^{2x}" is "$2^5 e^{2x}$" so that at "x = 0.5" this value becomes $32e^{2(0.5)} = 32e = \mathbf{86.9850\,1851}$ which is roughly the same value as the one computed using the *Multinomial Expansion Theorem* in equation (123).

For calculating the various *partial derivatives* with respect to any number of independent variables involving any number of products of multivariate expressions, the *Multinomial Expansion Theorem* is still applicable but with some minor modifications of the general formula that was derived for the two dimensional case.

The various partial derivatives of a product of several multivariate expressions may be written in a more general form as:

$$\frac{\partial^{m_1}}{\partial x_1^{m_1}} \frac{\partial^{m_2}}{\partial x_2^{m_2}} \frac{\partial^{m_3}}{\partial x_3^{m_3}} \cdots \frac{\partial^{m_k}}{\partial x_j^{m_k}} \left[f_1(x_1, x_2, \ldots, x_j) \cdot f_2(x_1, x_2, \ldots, x_j) \cdots f_i(x_1, x_2, \ldots, x_j)\right] \tag{126}$$

which can symbolically be expanded as:

$$\left[f_1^{(0)} + f_2^{(0)} + \ldots + f_i^{(0)} \right]_{1(m_1)}^{m_1} \;\Delta\; \left[f_1^{(0)} + f_2^{(0)} + \ldots + f_i^{(0)} \right]_{2(m_2)}^{m_2} \;\Delta\; \cdots \;\Delta$$

$$\Delta \;\cdots\; \Delta \quad \left[f_1^{(0)} + f_2^{(0)} + \ldots + f_i^{(0)} \right]_{j(m_k)}^{m_k} \tag{127}$$

where "Δ" is a special operator that is used to mimic the process of algebraically expanding *term by term* the product of two or more expressions with the only exception that all exponents are to be treated as order of differentiation.

In complete notational form using the *Multinomial Expansion Theorem* this may be rewritten as:

$$\left[\sum_{\substack{n_1,n_2,\ldots,n_i \geq 0 \\ n_1+n_2+\cdots+n_i=m_1}} \frac{n!}{n_1!\, n_2! \cdots n_k!}\; f_{1,1(n_1)}^{(n_1)} f_{2,1(n_2)}^{(n_2)} \;\cdots\; f_{i,1(n_i)}^{(n_i)} \right] \Delta$$

$$\left[\sum_{\substack{n_1,n_2,\ldots,n_i \geq 0 \\ n_1+n_2+\cdots+n_i=m_2}} \frac{n!}{n_1!\, n_2! \cdots n_k!}\; f_{1,2(n_1)}^{(n_1)} f_{2,2(n_2)}^{(n_2)} \;\cdots\; f_{i,2(n_i)}^{(n_i)} \right] \quad \Delta \;\cdots\; \Delta$$

$$\left[\sum_{\substack{n_1,n_2,\ldots,n_i \geq 0 \\ n_1+n_2+\cdots+n_i=m_k}} \frac{n!}{n_1!\, n_2! \cdots n_k!}\; f_{1,j(n_1)}^{(n_1)} f_{2,j(n_2)}^{(n_2)} \;\cdots\; f_{i,j(n_i)}^{(n_i)} \right] \tag{128}$$

When expanding the various partial derivatives of a product of several multivariate expressions using the above notational form, it is very important to insure that "*all*" the multivariate expressions present in "*each product*" are also "*all*" present in "*each term*" of the resultant expansion.

Example (6.2). Based entirely on our standard notation for representing the various partial derivatives of a product of several multivariate expressions, we will determine "$\frac{\partial f_1 f_2}{\partial x_1 \partial x_2^2}$" where "$f_1$" and "$f_2$" are each defined as arbitrary multivariate function.

$$\frac{\partial^3 f_1 f_2}{\partial x_1 \partial x_2^2} = [f_1 + f_2]_{1(1)}^{(1)} \;\Delta\; [f_1 + f_2]_{2(2)}^{(2)} \tag{129}$$

$$= \left[f_{1,1(1)}^{(1)} + f_{2,1(1)}^{(1)} \right] \Delta \left[f_{1,2(2)}^{(2)} + 2 f_{1,2(1)}^{(1)} f_{2,2(1)}^{(1)} + f_{2.2(2)}^{(2)} \right] \tag{130}$$

Algebraically performing a term by term symbolic multiplication by treating all exponent values as order of differentiation, we obtain:

$$= f^{(1)}_{1,1(1)} f^{(2)}_{1,2(2)} + 2 f^{(1)}_{1,1(1)} f^{(1)}_{1,2(1)} f^{(1)}_{2,2(1)} + f^{(1)}_{1,1(1)} f^{(2)}_{2,2(2)} + \\ + f^{(1)}_{2,1(1)} f^{(2)}_{1,2(2)} + 2 f^{(1)}_{2,1(1)} f^{(1)}_{1,2(1)} f^{(1)}_{2,2(1)} + f^{(1)}_{2,1(1)} f^{(2)}_{2,2(2)} \quad (131)$$

which in the conventional symbolic form may be translated as:

$$= \frac{\partial^3 f_1}{\partial x_1 \partial x_2^2} + 2\frac{\partial^2 f_1}{\partial x_1 \partial x_2}\frac{\partial f_2}{\partial x_2} + \frac{\partial f_1}{\partial x_1}\frac{\partial^2 f_2}{\partial x_2^2} + \frac{\partial^2 f_1}{\partial x_2^2}\frac{\partial f_2}{\partial x_1} + 2\frac{\partial f_1}{\partial x_2}\frac{\partial^2 f_2}{\partial x_1 \partial x_2} + \frac{\partial^3 f_2}{\partial x_1 \partial x_2^2} \quad (132)$$

To insure that every term in the above expansion always contains the two functions that is being differentiated, we must include "f_2" and "f_1" in the first and last term of the expansion respectively.

The final results are:

$$= \frac{\partial^3 f_1}{\partial x_1 \partial x_2^2} f_2 + 2\frac{\partial^2 f_1}{\partial x_1 \partial x_2}\frac{\partial f_2}{\partial x_2} + \frac{\partial f_1}{\partial x_1}\frac{\partial^2 f_2}{\partial x_2^2} + \frac{\partial^2 f_1}{\partial x_2^2}\frac{\partial f_2}{\partial x_1} + 2\frac{\partial f_1}{\partial x_2}\frac{\partial^2 f_2}{\partial x_1 \partial x_2} + f_1\frac{\partial^3 f_2}{\partial x_1 \partial x_2^2} \quad (133)$$

We can validate the use of our symbolic notations by performing the same operation manually and compare the results with the one obtained in the above equation:

$$\frac{\partial^2 f_1 f_2}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f_1}{\partial x_2} f_2 + f_1\frac{\partial f_2}{\partial x_2}\right) = \frac{\partial^2 f_1}{\partial x_2^2} f_2 + 2\frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_2} + f_1\frac{\partial^2 f_2}{\partial x_2^2} \quad (134)$$

$$\frac{\partial^3 f_1 f_2}{\partial x_1 \partial x_2^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial^2 f_1 f_2}{\partial x_2^2}\right) = \frac{\partial}{\partial x_1}\left(\frac{\partial^2 f_1}{\partial x_2^2} f_2 + 2\frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_2} + f_1\frac{\partial^2 f_2}{\partial x_2^2}\right) \quad (135)$$

$$= \frac{\partial^3 f_1}{\partial x_1 \partial x_2^2} f_2 + \frac{\partial^2 f_1}{\partial x_2^2}\frac{\partial f_2}{\partial x_1} + 2\frac{\partial^2 f_1}{\partial x_1 \partial x_2}\frac{\partial f_2}{\partial x_2} + 2\frac{\partial f_1}{\partial x_2}\frac{\partial^2 f_2}{\partial x_1 \partial x_2} + \frac{\partial f_1}{\partial x_1}\frac{\partial^2 f_2}{\partial x_2^2} + f_1\frac{\partial^3 f_2}{\partial x_1 \partial x_2^2} \quad (136)$$

As can be verified, the above expansion is exactly identical to the one in equation (133) thereby completely validating our standard use of special notations for taking the various partial derivatives of a product of several multivariate expressions.

The greatest advantage for using this notational convention is that it can reduce the entire process of determining the various partial derivatives of a product consisting of any number of expressions entirely on a "*computational level*".

In general, an IAMPT will always consist of multivariate polynomials as well as the differential of multivariate polynomials where each multivariate polynomial term will always be expressible as a product of several auxiliary variables. For calculating the various derivatives and partial derivatives of an IAMPT would require that each of the products of several auxiliary variables be differentiated under the product rule. So its therefore quite easy to visualize how the use of the *Multinomial Expansion Theorem* would become a very valuable tool for computing the various derivatives and partial derivatives of an IAMPT to any desirable degree of accuracy.

The complete development of all the formulas related to the calculations of the various derivatives and partial derivatives of an IAMPT for solving all types of DEs and systems of DEs is of course much beyond the scope of this paper. However, this can always be made available to anyone by special request provided you contact me at either one of the following email addresses michelmikalajunas@bellnet.ca or at jpnelson_mfc@yahoo.ca .

7. *General* closed form solutions of the Navier-Stokes equations by method of conjecture involving the use of computational differential analysis

The Navier-Stokes equations is the direct application of Newton's second law of motion for the complete analysis of both compressible and incompressible fluids.

For the case of incompressible flow and assuming constant viscosity, the equations may be described as follow:

$$\begin{matrix} & \textbf{\textit{Inertia}} & = & \textbf{\textit{Pressure gradient}} & + & \textbf{\textit{Viscosity}} & + & \textbf{\textit{Other forces}} \\ \rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) & & = & -\nabla P & + & \mu\nabla^2\mathbf{v} & + & F \end{matrix} \tag{137}$$

along with the mass continuity equation which states that:

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \tag{138}$$

Since we will restrict our analysis to *incompressible* flow only, the density is always assumed constant so that the above equation may be rewritten as:

$$\nabla\cdot\mathbf{v} = 0 \tag{139}$$

By assuming that gravitational forces are the only external forces present, the vector equations in ***Cartesian*** coordinates expand as follow:

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = -\frac{\partial P}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + \rho g_x \tag{140}$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) = -\frac{\partial P}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) + \rho g_y \tag{141}$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) = -\frac{\partial P}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) + \rho g_z \tag{142}$$

along with the mass continuity equation defined as:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{143}$$

We would construct the NCSA table by defining the *variable* coefficients as the fluid density "ρ", the fluid dynamic viscosity "μ" and the gravitational force components in the x, y and z direction. Since no external inputs are present in these equations other then the external forces due to gravity then we can set "q = 0" in the IAMPT that will be selected for solving these vector equations.

In the *Secondary Expansion* of our IAMPT, the first set of auxiliary variables will be used for representing the dependent and independent variables in that order. This will be followed by the remaining initially assumed auxiliary variables used for representing all basis functions in complete differential form that will be present in the exact analytical solution of the system of PDEs.

Our IAMPT will be selected on the basis of solving the above system of PDEs in terms of a system of *implicitly* defined equations that would consist only of the algebraic and elementary basis functions. The various initial conditions possible for this type of generalized flow are of course expected to be infinite. So in order to maximize our numerical solution rate of the corresponding nonlinear simultaneous equations, we can set all the coefficients defining the initial conditions in our IAMPT as part of the unknowns to solve for that would be represented by the initial values of each initially assumed auxiliary variable. Other unknowns to solve for are the variable coefficients defined in our NCSA table as well as those present in both the *Primary* and *Secondary Expansion* of our IAMPT.

Over time, the NCSA table should eventually succeed in capturing from the numerical solution set of the nonlinear simultaneous equations all those *exact instance analytical solutions* that would conform with experimental results obtained under controlled laboratory conditions.

It is only through the gathering of this type of information over a span of say many years or even many decades that a large number of *generalized* analytical solutions may potentially be uncovered. This would in the very long term enable us to acquire a far better understanding of general fluid behavior than having to depend entirely on the use of laboratory experiments as a result of the non-integrability of many integrals that would have originated from the use of conventional methods of pure mathematical analysis.

In terms of ***Cylindrical*** coordinates this would be written as:

$$\rho\left(\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_r}{\partial \theta} + u_z\frac{\partial u_r}{\partial z} - \frac{u_\theta^2}{r}\right) = -\frac{\partial P}{\partial r} + \mu\left[\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_r}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u_r}{\partial \theta^2} + \frac{\partial^2 u_r}{\partial z^2} - \frac{u_r}{r^2} - \frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta}\right)\right] + \rho g_r \quad (144)$$

$$\rho\left(\frac{\partial u_\theta}{\partial t} + u_r\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_\theta}{\partial \theta} + u_z\frac{\partial u_\theta}{\partial z} + \frac{u_r u_\theta}{r}\right) = -\frac{1}{r}\frac{\partial P}{\partial \theta} + \mu\left[\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_\theta}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u_\theta}{\partial \theta^2} + \frac{\partial^2 u_\theta}{\partial z^2} - \frac{u_\theta}{r^2} + \frac{2}{r^2}\frac{\partial u_r}{\partial \theta}\right)\right] + \rho g_\theta \quad (145)$$

$$\rho\left(\frac{\partial u_z}{\partial t} + u_r\frac{\partial u_z}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_z}{\partial \theta} + u_z\frac{\partial u_z}{\partial z}\right) = -\frac{\partial P}{\partial z} + \mu\left[\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u_z}{\partial \theta^2} + \frac{\partial^2 u_z}{\partial z^2}\right)\right] + \rho g_z \quad (146)$$

along with the mass continuity equation defined as:

$$\frac{1}{r}\frac{\partial}{\partial r}(r u_r) + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} = 0 \quad (147)$$

Such a coordinate system may in some cases prove to be easier for the analysis of certain types of fluid motion that would mainly involve symmetry thereby allowing for the elimination of a velocity component.

A very common case is axisymmetric flow where there is no tangential velocity ($u_\theta = 0$) and the remaining quantities are independent of θ:

$$\rho\left(\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + u_z\frac{\partial u_r}{\partial z}\right) = -\frac{\partial P}{\partial r} + \mu\left[\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_r}{\partial r}\right) + \frac{\partial^2 u_r}{\partial z^2} - \frac{u_r}{r^2}\right)\right] + \rho g_r \tag{148}$$

$$\rho\left(\frac{\partial u_z}{\partial t} + u_r\frac{\partial u_z}{\partial r} + u_z\frac{\partial u_z}{\partial z}\right) = -\frac{\partial P}{\partial z} + \mu\left[\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_z}{\partial r}\right) + \frac{\partial^2 u_z}{\partial z^2}\right)\right] + \rho g_z \tag{149}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(ru_r) + \frac{\partial u_z}{\partial z} = 0 \tag{150}$$

For this type of coordinate system we would proceed in constructing the NCSA table in exactly the same manner as for the *Cartesian* coordinate system where in both cases there are no external inputs so that "q = 0". This would also include managing in exactly the same manner all initial conditions and the variable coefficients defined by the fluid density "ρ", the fluid dynamic viscosity "μ" and the gravitational components in the x, y and z direction.

In terms of ***Spherical*** coordinates this would be written as:

$$\rho\left(\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + \frac{u_\theta}{rSin(\emptyset)}\frac{\partial u_r}{\partial \theta} + \frac{u_\emptyset}{r}\frac{\partial u_r}{\partial \emptyset} - \frac{u_\theta^2+u_\emptyset^2}{r}\right) = \frac{\partial P}{\partial r} + \rho g_r +$$
$$+ \mu\left\{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u_r}{\partial r}\right) + \frac{1}{r^2Sin(\emptyset)^2}\frac{\partial^2 u_r}{\partial \theta^2} + \frac{1}{r^2Sin(\emptyset)}\frac{\partial}{\partial \emptyset}\left(Sin(\emptyset)\frac{\partial u_r}{\partial \emptyset}\right) - 2\left(\frac{u_r + \frac{\partial u_\emptyset}{\partial \emptyset} + u_\emptyset Cot(\emptyset)}{r^2}\right) + \frac{2}{r^2Sin(\emptyset)}\frac{\partial u_\theta}{\partial \theta}\right\} \tag{151}$$

$$\rho\left\{\frac{\partial u_\theta}{\partial t} + u_r\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{rSin(\emptyset)}\frac{\partial u_\theta}{\partial \theta} + \frac{u_\emptyset}{r}\frac{\partial u_\theta}{\partial \emptyset} + \left(\frac{u_r u_\theta + u_\theta u_\emptyset Cot(\emptyset)}{r}\right)\right\} = -\frac{1}{rSin(\emptyset)}\frac{\partial P}{\partial \theta} + \rho g_\theta +$$
$$+ \mu\left\{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u_\theta}{\partial r}\right) + \frac{1}{r^2Sin(\emptyset)^2}\frac{\partial^2 u_\theta}{\partial \theta^2} + \frac{1}{r^2Sin(\emptyset)}\frac{\partial}{\partial \emptyset}\left(Sin(\emptyset)\frac{\partial u_\theta}{\partial \emptyset}\right) + \left(\frac{2\frac{\partial u_r}{\partial \theta} + 2Cos(\emptyset)\frac{\partial u_\theta}{\partial \theta} - u_\theta}{r^2Sin(\emptyset)^2}\right)\right\} \tag{152}$$

$$\rho\left\{\frac{\partial u_\emptyset}{\partial t} + u_r\frac{\partial u_\emptyset}{\partial r} + \frac{u_\theta}{rSin(\emptyset)}\frac{\partial u_\emptyset}{\partial \theta} + \frac{u_\emptyset}{r}\frac{\partial u_\emptyset}{\partial \emptyset} + \left(\frac{u_r u_\emptyset - u_\theta^2 Cot(\emptyset)}{r}\right)\right\} = -\frac{1}{r}\frac{\partial P}{\partial \emptyset} + \rho g_\emptyset +$$
$$+ \mu\left\{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u_\emptyset}{\partial r}\right) + \frac{1}{r^2Sin(\emptyset)^2}\frac{\partial^2 u_\emptyset}{\partial \theta^2} + \frac{1}{r^2Sin(\emptyset)}\frac{\partial}{\partial \emptyset}\left(Sin(\emptyset)\frac{\partial u_\emptyset}{\partial \emptyset}\right) + \frac{2}{r^2}\frac{\partial u_r}{\partial \emptyset} - \left(\frac{u_\emptyset + 2Cos(\emptyset)\frac{\partial u_\theta}{\partial \theta}}{r^2Sin(\emptyset)^2}\right)\right\} \tag{153}$$

along with the mass continuity equation defined as:

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2u_r) + \frac{1}{rSin(\emptyset)}\frac{\partial u_\theta}{\partial \theta} + \frac{1}{rSin(\emptyset)}\frac{\partial}{\partial \emptyset}(Sin(\emptyset)u_\emptyset) = 0 \tag{154}$$

In this coordinate system, there are two external inputs in the form of the Sine and Cosine function which according to equation (35) and (36) can each be expressed in terms of the Tangent half angle formula so that we can set "q = 1" in our IAMPT. All initial conditions and variable coefficients are handled in exactly the same manner as with the *Cartesian* and *Cylindrical* coordinate system.

Because of the universality of the new method of analytical integration we can extend this analysis to cover all possible cases for both compressible and incompressible flow where the concept of an NCSA table would still be applicable throughout.

8. The development of a universal software for the analytical solutions of all DEs and systems of DEs under a single unified theory of analytical integration

The highly computational nature of the universal differential expansion described by equations (1) through (5) for representing all mathematical equations makes it very difficult for conducting any real meaningful numerical experimentations even for solving the simplest type of DE. For solving the vast majority of DEs and systems of DEs of greatest importance to the physical sciences, super computers are by far more suitable for this type of high level and very advanced form of computational analysis.

The advent of Quantum computers in the near future could significantly improve the performance of handling even the most complex systems of PDEs. They would by far exceed the capabilities of even our most powerful super computer of our time because they would operate entirely on the fundamental principles of Quantum theory which is based on the study of energy at the atomic and subatomic level. Such advanced computer technology would allow for the capability of performing multiple tasks in parallel thereby resulting in a significant increase in the billion-fold when compared to conventional computer systems.

Among the many possible states of operation is the *binary* state of a Quantum bit or Qubit that would either be defined as spin-down or spin-up with each mode entirely controlled by a pulse of energy originating from a laser beam. Major centers of research in Quantum computing are currently in operation that would include MIT, IBM, Oxford, Harvard, Stanford and the Los Alamos National Laboratory.

The greatest advantage for having arrived at a unified theory of analytical integration is that it can be converted into a *single major universal software* by which all DEs and systems of DEs may be resolved under a single *common mathematical ideology*. Such a universal software development would be referred to as a "*Numerical Control Analytics Software*" or NCAS. It would operate on the principle of determining the existence of *general analytical solutions* to DEs and systems of DEs through the application of a very unique method of conjecture that would be driven entirely by computational analysis. This would represent a far better alternative than having to maintain a large number of highly *dispersed* mathematical theories all of which could never be consolidated in terms of a single universal software development package such as the one proposed here.

If such a Numerical Control Analytics Software would be applied only to Physics, it would certainly qualify as being "the complete unified theory of physics" but only in its most "raw state". Human intervention would then only be necessary for complete translation of all computer results that would appear in the form of exact numerical computations into practical decipherable mathematical equations.

If such a Numerical Control Analytics Software would be applied only into Engineering Science, it would become the standard method of all engineering analysis by which the concept of an IAMPT would be applied very rigorously for resolving all relevant DEs and systems of DEs in the form of general closed form solutions only. This would set the stage for the complete formulation of many fundamental key theorems similar to what the famous Superposition Theorem has succeeded in accomplishing in the general theory of linear physical systems.

9. Conclusions

The problem of integration has always presented itself as a real challenge when attempting to find closed formed solutions for the vast majorities of DEs and systems of DEs. The main reason for this is the frequent occurrences of integrals from which the vast majority of them cannot always be resolved exactly under any existing methods of mathematical analysis. This complication can be completely avoided altogether if rather than proceeding with some *initially assumed closed form solution* for attempting to solve a DE or a system of DEs, we instead work only with the complete *differential* representation of the same *initially assumed closed form solution*. The greatest advantage for proceeding in that fashion is the highest expectation that many of the assumed differentials will in the end appear *exact* and thus always completely integrable in the end. This in fact is quite achievable because every *differentiable* mathematical equation can always be converted in complete differential form by following the same basic unique mathematical structure as the one

introduced by equations (1) through (5). Such a unique differential expansion form is so universal to all mathematical equations that it would certainly qualify by all mathematical standards as being a *complete unified analytical theory of integration* for resolving all types of DEs and systems of DEs in terms of closed form solutions. Many key mathematical properties of this unified analytical theory of integration have been quite extensively investigated in the past mainly by myself. But the one that stands out the most is the ability for resolving "*all types*" of DEs and systems of DEs uniquely in terms of "*general closed form solutions*" by utilizing a method of conjecture that would be driven entirely on computational analysis alone. We use the Navier-Stokes equations as a perfect model for illustrating this very unique approach of working with initially assumed differentials. In our example, we explore the various types of systems of PDEs that were developed in the past under the three most popular set of coordinate systems. In the final analysis, we were able to establish that independent of the type of flow whether compressible or incompressible, the boundary conditions and various external forces present can always be completely accounted for during the process of working with these types of initially assumed differential forms. From the very unique properties of such a proposed unified differential method of analysis, it is expected that many cases of the Navier-Stokes equations will always be completely integrable in terms of such "*general*" closed form solutions by following a very unique method of conjecture. From the Navier-Stokes equations we can apply the same type of universal differential analysis for investigating other types of *fundamental equations* that would include Maxwell's equations, Einstein's field equations, the Schrödinger equation just to name a few. Figure 3.1 provides a direct relationship between the method of universal differential analysis and the elusive "*theory of everything*". From this table, one is very tempted to conclude that for arriving at such a gigantic theory for explaining everything about our universe may no longer be just a matter for modern physics to resolve over time. Rather, it is expected that such a theory of everything may only be achievable in the end from the complete consolidation of every single theory describing its own unique physical system under one big gigantic universal theory that in the end will succeed in explaining everything about our universe.

10. Appendix A

$$(\mathbf{1.1}) \quad \mathbf{f(x,y)} = \mathbf{0} = a_1x^2 + a_2y^2 + a_3xy + a_4$$

$W_1 = x$
$W_2 = y$

(1). *Primary Expansion:*

$$F(W_1, W_2) = 0 = a_1W_1^2 + a_2W_2^2 + a_3W_1W_2 + a_4$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$

$$(\mathbf{1.2}) \quad \mathbf{f(x,y)} = \mathbf{0} = a_1y + a_2e^{a_3x}Sin(a_4x)$$

$W_1 = x$
$W_2 = y$
$W_3 = e^{a_3x}$
$W_4 = Tan(a_4x/2)$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, W_4) = 0 = a_1W_2(1 + W_4^2) + 2a_2W_3W_4$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$
$$a_3 W_3 dx + 0 \cdot dy = dW_3$$
$$a_4(1 + W_4^2)dx + 0 \cdot dy = 2dW_4$$

$(\mathbf{1.3})\quad \mathbf{f(x,y) = 0 = x^2 + y^2\sqrt{(x-y)} + 3e^{3x}}$

$$W_1 = x$$
$$W_2 = y$$
$$W_3^2 = x - y = W_1 - W_2$$
$$W_4 = e^x = e^{W_1}$$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, W_4) = 0 = W_1^2 + W_2^2 W_3 + 3W_4^3$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$
$$dx - dy = 2W_3\, dW_3$$
$$3W_4 dx + 0 \cdot dy = dW_4$$

$(\mathbf{1.4})\quad \mathbf{f(x,y) = 0 = x\sqrt{x^2 + y^2} + y\sqrt{x^2 - y^2}}$

$$W_1 = x$$
$$W_2 = y$$
$$W_3^2 = W_1^2 + W_2^2$$
$$W_4^2 = W_1^2 - W_2^2$$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, W_4) = 0 = W_1 W_3 + W_2 W_4$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$
$$W_1 dx + W_2 dy = W_3 dW_3$$
$$W_1 dx - W_2 dy = W_4 dW_4$$

$(\mathbf{1.5})\quad \mathbf{f(x,y) = 0 = \ln\left(1 + \sqrt[3]{x+1}\right) - \sqrt[6]{y+1} - 1}$

$$W_1 = x$$
$$W_2 = y$$
$$W_3^3 = x + 1 = W_1 + 1$$
$$W_4 = \ln\left(1 + \sqrt[3]{x+1}\right) = \ln(1 + W_3)$$
$$W_5^6 = y + 1 = W_2 + 1$$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, W_4, W_5) = 0 = W_4 - W_5 - 1$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$

$$dx + 0 \cdot dy = 3W_3^2 dW_3$$
$$dx + 0 \cdot dy = 3W_3^2(1 + W_3)dW_4$$
$$0 \cdot dx + dy = 6W_5^5 dW_5$$

$$\mathbf{(1.6)} \quad \mathbf{f(x,y) = 0 = 3Sin(x+y) - ln\left(e^x + \sqrt{Cos(x)}\right) + ln\left(\frac{x}{y}\right) + \sqrt{ArcTan(2x)}}$$

$$W_1 = x$$
$$W_2 = y$$
$$W_3 = Tan\left(\frac{x+y}{2}\right)$$
$$W_4 = e^x$$
$$W_5 = Tan(\frac{x}{2})$$
$$W_6^2 = Cos(x) = \frac{1 - W_5^2}{1 + W_5^2}$$
$$W_7 = ln\left(e^x + \sqrt{Cos(x)}\right) = ln(W_4 + W_6)$$
$$W_8 = ln(x)$$
$$W_9 = ln(y)$$
$$W_{10}^2 = ArcTan(2x)$$

(1). *Primary Expansion:*

$$F(W_1, W_2, \dots, W_{10}) = 0 = \frac{6W_3}{1 + W_3^2} - W_7 + W_8 - W_9 + W_{10}$$

(2). *Secondary Expansion:*

$$dx = dW_1$$
$$dy = dW_2$$

$$(1 + W_3^2)dx + (1 + W_3^2)dy = 2dW_3$$
$$W_4 dx + 0 \cdot dy = dW_4$$
$$(1 + W_5^2)dx + 0 \cdot dy = 2dW_5$$
$$-W_5 dx + 0 \cdot dy = W_6(1 + W_5^2)dW_6$$
$$\{W_4 W_6(1 + W_5^2) - W_5\}dx + 0 \cdot dy = W_6(1 + W_5^2)(W_4 + W_6)dW_7$$
$$dx + 0 \cdot dy = W_1 dW_8$$
$$0 \cdot dx + dy = W_2 dW_9$$
$$dx + 0 \cdot dy = (1 + 4W_1^2)W_{10} dW_{10}$$

$$(\mathbf{2.1}) \quad \mathbf{f}(\mathbf{z}, \boldsymbol{x_1}, \boldsymbol{x_2}) = \mathbf{0} = \mathbf{z} + \boldsymbol{z^3 x_1 x_2} - \boldsymbol{x_2} + \mathbf{1}$$

$W_1 = z$
$W_2 = x_1$
$W_3 = x_2$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3) = 0 = W_1 + W_1^3 W_2 W_3 - W_3 + 1$$

(2). *Secondary Expansion:*

$$\begin{aligned} dz + 0 \cdot dx_1 + 0 \cdot dx_2 &= dW_1 \\ 0 \cdot dz + dx_1 + 0 \cdot dx_2 &= dW_2 \\ 0 \cdot dz + 0 \cdot dx_1 + dx_2 &= dW_3 \end{aligned}$$

$$(\mathbf{2.2}) \quad \mathbf{f}(\mathbf{z}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3},) = \mathbf{0} = \mathbf{5}\boldsymbol{x_2 x_3 Sin}(\boldsymbol{z x_1 x_2}) + (\boldsymbol{x_1} + \boldsymbol{x_2})\boldsymbol{Cos}(\boldsymbol{z} + \mathbf{3}\boldsymbol{x_2} + \mathbf{2}\boldsymbol{x_3}) + \mathbf{3}$$

$W_1 = z$

$W_2 = x_1$

$W_3 = x_2$

$W_4 = x_3$

$W_5 = Tan(z x_1 x_2 / 2)$

$W_6 = Tan\left\{\frac{z + 3x_2 + 2x_3}{2}\right\}$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, W_4, W_5, W_6) = 0 = 5W_3 W_4 \left[\frac{2W_5}{1 + W_5^2}\right] + (W_2 + W_3)\left[\frac{1 - W_6^2}{1 + W_6^2}\right] + 3$$

(2). *Secondary Expansion:*

$$\begin{aligned} dz + 0 \cdot dx_1 + 0 \cdot dx_2 + 0 \cdot dx_3 &= dW_1 \\ 0 \cdot dz + dx_1 + 0 \cdot dx_2 + 0 \cdot dx_3 &= dW_2 \\ 0 \cdot dz + 0 \cdot dx_1 + dx_2 + 0 \cdot dx_3 &= dW_3 \\ 0 \cdot dz + 0 \cdot dx_1 + 0 \cdot dx_2 + dx_3 &= dW_4 \end{aligned}$$

$$(1 + W_5^2) W_2 W_3 dz + (1 + W_5^2) W_1 W_3 dx_1 + (1 + W_5^2) W_1 W_2 dx_2 + 0 \cdot dx_3 = 2dW_5$$

$$(1 + W_6^2) dz + 0 \cdot dx_1 + 3(1 + W_6^2) dx_2 + 2(1 + W_6^2) dx_3 = 2dW_6$$

$$(\mathbf{2.3}) \quad \mathbf{f(x,y)} = \mathbf{0} = \mathbf{3\ln}\left(\sqrt[3]{\boldsymbol{z}+\boldsymbol{x_1^2}+\boldsymbol{x_2^2}} - \mathbf{25}\boldsymbol{e^{2zx_1x_3}}\right) + \\ + \sqrt[5]{\boldsymbol{x_1^2}+\boldsymbol{x_2^2}+\boldsymbol{x_3^2}} - \mathbf{4}\boldsymbol{z^3} + \mathbf{1}$$

$W_1 = z$

$W_2 = x_1$

$W_3 = x_2$

$W_4 = x_3$

$W_5^3 = z + x_1^2 + x_2^2 = W_1 + W_2^2 + W_3^2$

$W_6 = e^{2zx_1x_3} = e^{2W_1W_2W_4}$

$W_7 = \ln\left(\sqrt[3]{z + x_1^2 + x_2^2} - 25e^{2zx_1x_3}\right) - \ln(W_5 - 25W_6)$

$W_8^5 = x_1^2 + x_2^2 + x_3^2 = W_2^2 + W_3^2 + W_4^2$

(1). *Primary Expansion:*

$$F(W_1, W_2, W_3, \dots, W_8) = 0 = 3W_7 + W_8 - 4W_1^3 + 1$$

(2). *Secondary Expansion:*

$$dz + 0\cdot dx_1 + 0\cdot dx_2 + 0\cdot dx_3 = dW_1$$
$$0\cdot dz + dx_1 + 0\cdot dx_2 + 0\cdot dx_3 = dW_2$$
$$0\cdot dz + 0\cdot dx_1 + dx_2 + 0\cdot dx_3 = dW_3$$
$$0\cdot dz + 0\cdot dx_1 + 0\cdot dx_2 + dx_3 = dW_4$$

$$dz + 2W_2dx_1 + 2W_3dx_2 + 0\cdot dx_3 = 3W_5dW_5^2$$

$$2W_2W_4W_6dz + 2W_1W_4W_6dx_1 + 0\cdot dx_2 + 2W_1W_2W_6dx_3 = dW_6$$

$$(1 - 150W_2W_4W_5^2W_6)dz + (2W_2 - 150W_1W_4W_5^2W_6)dx_1 + 2W_3dx_2 - \\ 150W_1W_2W_5^2W_6dx_3 = 3W_5^2(W_5 - 25W_6)dW_7$$

$$0\cdot dz + W_2dx_1 + W_3dx_2 + W_4dx_3 = 2.5W_8^4dW_8$$

11. References

Adby and Dempster (1974) Introduction to Optimization Methods, John Wiley & Sons, NY.

Anderson H. I. (1995) An exact solution of the Navier-Stokes equations for magneto hydrodynamic flow Acta Mechanica, Springer Journal, Volume 113, Issue 1, pp 241-244

Bilson-Thompson, Sundance O.; Markopoulou, Fotini; Smolin, Lee (2007). "Quantum gravity and the standard model".

Brady J.F. and. Acrivos A. (1981) Steady flow in a channel or tube with an accelerating surface velocity. An exact solution to the Navier Stokes equations with reverse flow, Journal of Fluid Mechanics / Volume 112 / pp 127- 150 Cambridge University Press

Davis H. T. (1962) Introduction to nonlinear differential and integral equations, Dover publications.

Einstein, A. (1974) The meaning of Relativity, Princeton Paperback, Fifth Edition.

Feynman, Richard (1982). "Simulating physics with computers". International Journal of Theoretical Physics 21 (6–7):

Guillen, M. (1995) Five equations that changed the world, Hyperion, NY.

Hazewinkel, M ed. (2001), "Multinomial coefficient", Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4

Jaeger, Gregg (2006). Quantum Information: An Overview. Berlin: Springer. ISBN 0-387-35725-4. OCLC 255569451.

Mikalajunas, M. (1981) New algorithm for integrating DEs, SIAM 1981 Fall meeting, Netherland Hilton Hotel Cincinnati. OH.

Mikalajunas, M. (1983) Representing a PDE in terms of an infinite variety of integrable and non-integrable systems of ODEs, Abstracts of papers presented to the AMS Vol. 4, Number 5, 87th summer meeting, Albany, NY.

Mikalajunas, M. (1983) On the use of Multivariate Polynomials for integrating ODEs, NY State Mathematical Assoc. of Two Year Colleges, Seaway Section MAA, Spring 1983 meeting, Utica NY.

Mikalajunas, M. (2015) A better way for managing all of the physical sciences under a single unified theory of analytical integration, *Proceedings of the 6th International Conference on Computational Methods*, 14th – 17th July 2015, Auckland, G.R. Liu and Raj Das, ID 845-3475-1-PB, ScienTech Publisher.

Murphy G.M. (1960) Ordinary differential equations and their solutions, D. Van Nostrand Company.

Newton, Sir Isaac (1729). The Mathematical Principles of Natural Philosophy II. p. 255.

Nielsen, Michael; Chuang, Isaac (2000). Quantum Computation and Quantum Information. Cambridge: Cambridge University Press. ISBN 0-521-63503-9. OCLC 174527496

Simon, Daniel R. (1994). "On the Power of Quantum Computation". Institute of Electrical and Electronic Engineers Computer Society Press.

Simon, D.R. (1994). "On the power of quantum computation". Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on: 116–123. doi:10.1109/SFCS.1994.365701. ISBN 0-8186-6580-7.

Singer, Stephanie Frank (2005). Linearity, Symmetry, and Prediction in the Hydrogen Atom. New York: Springer. ISBN 0-387-24637-1. OCLC 253709076.

Wang C. Y. (2003) Exact Solutions of the Steady-State Navier-Stokes Equations Annual Review of Fluid Mechanics Vol. 23: 159-177

Weinberg, Steven (1993) Dreams of a Final Theory: The Search for the Fundamental Laws of Nature, Hutchinson Radius, London, ISBN 0-09-177395-4

Reliability Analysis of Slope Stability using Monte Carlo Simulation and Comparison with Deterministic Analysis

R. K. Sharma

Professor, Department of Civil Engineering,
National Institute of Technology Hamirpur (H.P.) – 177005 (India)

Presenting author & corresponding author: rksnithp61l@gmail.com

Abstract

Traditional slope stability analysis is limited to the use of single valued parameters to analyze a slope's characteristics. Consequently, traditional analysis methods yield single valued estimates for factor of safety of a slope's stability. However, the inherent variability of the soil characteristics which affect slope stability indicates that the stability of a slope is a probabilistic rather than a deterministic situation. In other words, the stability of a slope is a random process which is dependent on the relative distribution of controlling soil parameters. For a natural slope, the stability deciding parameters vary considerably throughout the extent of slope. In this paper, the variability of soil properties and their effect on stability of a natural slope has been studied incorporating the probabilistic analysis using Monte Carlo simulation and deterministic analysis using Geo-Studio and PLAXIS. The factors of safety have been determined using the two approaches and effect of dynamic loading input on slope stability has been studied.

Keywords: Slope stability, deterministic approach, probabilistic analysis Monte Carlo method.

Introduction

Slope instability is responsible for damage to public and private property every year. Slope failures can be manifested as landslides or by other slowly occurring processes such as soil seriously damaged or destroyed. Slope instability is a complex phenomenon that can occur at many scales and for many reasons. Slope stability analyses and stabilization require an understanding and evaluation of the processes that govern the behavior of slopes.

Real life failures in naturally deposited mixed soils are not necessarily circular, but prior to computers, it was far easier to analyze such a simplified geometry. Nevertheless, failures in 'pure' clay can be quite close to circular. Such slips often occur after a period of heavy rain, when the pore water pressure at the slip surface increases, reducing the effective normal stress and thus diminishing the restraining friction along the slip line. This is combined with increased soil weight due to the added groundwater. A 'shrinkage' crack (formed during prior dry weather) at the top of the slip may also fill with rain water, pushing the slip forward. At the other extreme, slab-shaped slips on hill sides can remove a layer of soil from the top of the underlying bedrock. Again, this is usually initiated by heavy rain, sometimes combined with increased loading from new buildings or removal of support at the toe (resulting from road widening or other construction work). Stability can thus be significantly improved by installing drainage paths to reduce the destabilizing forces. A weakness along the slip circle may remain at the reoccurrence of the next monsoon. If the forces available to resist movement are greater than the forces driving the movement, the slope is considered stable. Factor of safety is calculated by dividing the forces resisting movement by the forces driving movement. In earthquake-prone areas, the analysis is

typically run for static conditions and pseudo-static conditions, where seismic forces from an earthquake are assumed to add static loads to the analysis.

The slope stability analyses are performed to assess the safe and economic design of human-made or natural slopes (e.g. embankments, road cuts, open-pit mining, excavations, landfills etc.) and the equilibrium conditions [1]-[3]. The term slope stability may be defined as the resistance of inclined surface to failure by sliding or collapsing. The main objectives of slope stability analysis are finding endangered areas, investigation of potential failure mechanisms, determination of the slope sensitivity to different triggering mechanisms, designing of optimal slopes with regard to safety, reliability and economics, designing possible remedial measures, e.g. barriers and stabilization. Successful design of the slope requires information about site characteristics, e.g. properties of soil/rock mass, slope geometry, alteration of materials by faulting, joint or discontinuity systems, movements and tension in joints, earthquake activity, etc. Choice of correct analysis technique depends on both site conditions and the potential mode of failure, with consideration being given to the varying strengths, weaknesses and limitations inherent in each methodology. The hypothesis of this research is that analysis of slope stability can be more methodological using the information about probability distribution of the slope's characteristics to determine the slope stability from the output of the analysis. Knowledge of the probability distribution of the output allows the engineer to assess the probability of slope failure. Therefore, an allowable risk criterion can be used to establish a consistent target for the design process [4].

Scope and Objectives

Stability of slopes, natural or man-made, is particularly important for any hill road. Disturbance to slope can occur due to erosion by rainfall and run-off and consequent slides. During monsoons the hill roads experience slips, erosions and major and minor landslides at many places. Check for the stability of the slopes is very necessary in order to ensure the stability of the slope as it would affect the life of people directly as landslide causing life loss and indirectly as the hindrance to flow of the traffic. Since the profile is along the National Highway 21, so its failure can cause the closing of the highway and it has been observed many times that it has closed previously. Rainy season causes the maximum disturbance in its stability. Hence, slope stability is vital for prevention of landslides/slips [5]. If the cut slopes are not properly designed, it will fail and would causes huge loss to mankind in a direct or indirect way. Taking into consideration above factors and importance of the stability, essential remedial measures are required and should be properly designed. Moreover consideration of various uncertainties involved in the properties of the soil which ultimately determine the stability of slope should be taken into account. For that purpose, statistical analysis or reliability analysis of slope becomes necessary and should be performed for a particular slope to check the reliability index of that particular slope. Reliability analysis of slope stability has attracted considerable research attention in the past few decades [6]-[10]. Reliability of slope stability is frequently measured by ''reliability index,'' and slope failure probability, P_f, which is defined as the probability that the minimum factor of safety (FS) is less than unity (i.e., $P_f = P\ (FS < 1)$). Various solution methods have been proposed to estimate P_f and Reliability Index. Among the most widely used methods are the first order second moment (FOSM) method, first order reliability method (FORM, also referred to as the Hasofer-Lind method) [11] and direct Monte Carlo simulation [12]. The objective of this research is to develop a probabilistic model for slope analysis by (a) understanding the concept of reliability analysis and its application in slope stability analysis, (b) performing the reliability analysis of slope stability using Monte Carlo simulation (using

RiskAMP [13]; and Geo5 [14], (c) performing the slope stability analysis with the help of PLAXIS [15] and (d) comparing the results obtained from different methods.

Methodology

The methodology include the preparation of the contour map of the slope to determine the geometry and assessing the soil characteristics over the entire slope by collecting fairly representative sample and determining the input soil parameter in the laboratory.

Figure 1. Typical view of slope failure near Gambhar Bridge on NH 21

The slope stability was assessed using the deterministic analysis and commonly used methods of analysis along with the software SLOPEW and PLAXIS (including dynamic loading input). Finally, the results obtained from the two approaches are compared and their efficacy for slope stability is determined. The site selected for the study is located in district Bilaspur, Himachal Pradesh, India on NH-21 highway namely Gambhar bridge. The height of the site is 1230 meter above sea level respectively. The study area lies in earthquake zone IV at latitude 31° 20´ N and longitudes 76º 45’ E. Average annual rainfall of the area is around 135 cm. A typical view of the slope failure is shown in figure 1.

Determination of basic geometrical characteristics of the slope was done using total station survey. Total station surveying was done for both the sites in order to generate contour maps of the slopes. The reduced levels, horizontal distance, vertical and horizontal angle readings were recorded using total station. These are fed as input in the software LISCAD to generate contour map as shown in figure 2. Three predominant sections 1-1, 2-2 and 3-3 of slope failure have been identified on the basis of the field observations as indicated in figure 3.

Fairly large numbers of representative samples of soil were collected from soil slope considering the variability of soil strata throughout the extent of slope. The soil parameters for

the drained conditions were determined. The mean value of different properties was calculated. Typical results obtained for different properties are summarized in table 1.

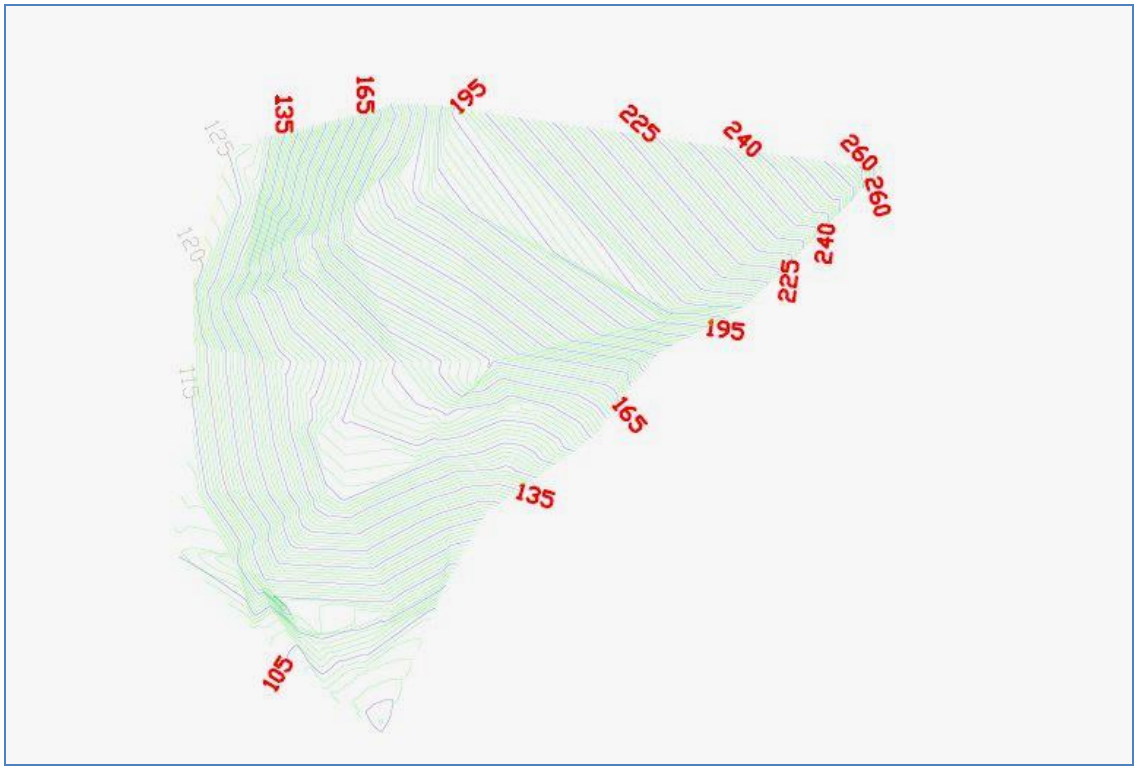


Figure 2. Contour map of site (near Gambhar bridge)

Table 1. Properties of different soil samples

S. No.	Water content (%)	Density (kN/m³)	Cohesion (kN/m²) (IS 2720 Part XIII, 1972)	Angle of internal friction (Φ) (IS 2720 Part XIII)[24]
1	6.80	17.63	27.16	9.85
2	3.69	18.97	19.03	24.96
3	14.79	17.39	7.57	21.4
4	14.85	19.79	18.24	15.97
5	13.95	22.40	5.13	21.03
6	12.56	20.06	17.72	21.76

Results and Analysis

Deterministic Approach

The traditional methods of slope stability normally use single valued parameters to analyze the characteristics of a slope. The output from traditional analysis methods yields single valued estimates of factor of safety of the stability of a slope. However, the parameters governing the stability of a slope vary considerably throughout the extent of the slope. Most commonly employed method of analysis of the stability a slope is Bishop's method [16] which yields the factor of safety as:

$$F= \frac{1}{\sum W \sin\alpha} \sum [cb + \tan\emptyset (W - Ub)] \frac{\sec\alpha}{1+\frac{\tan\emptyset \tan\alpha}{F}} \quad (1)$$

Where, F = Factor of safety, W = weight of slice, c = cohesion, b = width of slice, α = angle of inclination of slope, Ø = angle of internal friction and U = pore pressure at each slice.

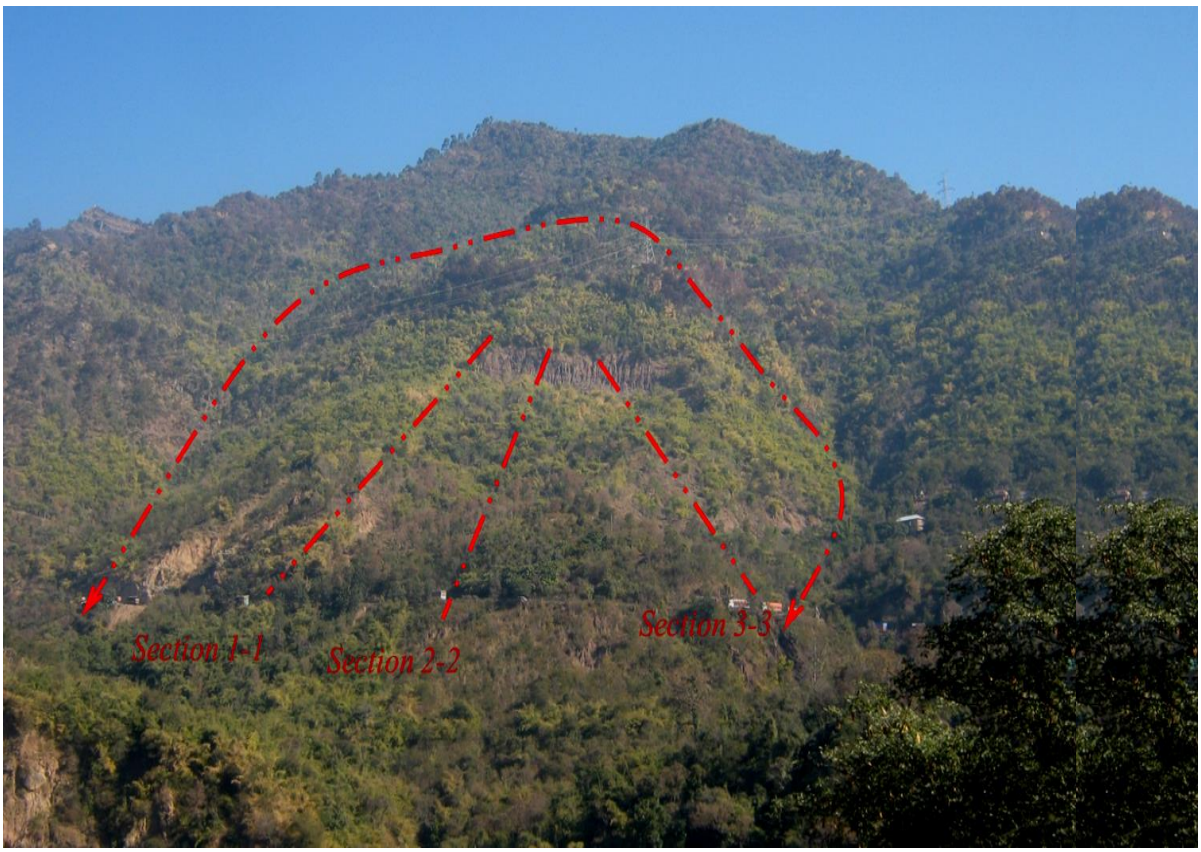


Figure 3. Three sections selected for slope stability analysis

An iterative analysis is necessary to obtain the factor of safety. Since this is a trial and error method, the assumed factor of safety F is entered with respect to which the new factor of safety is calculated and the iteration process is continued till the difference between the two values of factor of safety calculated is negligible. Three different sections namely 1-1, 2-2 and 3-3 were analyzed using SLOPE-W module of Geo Studio. The factor of safety for different sections was calculated with the help of different deterministic method namely ordinary method, Bishop's method [16], Janbu [17] method and Morgenstern Price Method [18][19]. Table 2 shows the values of factor of safety with the help of different methods. The results indicate that the slope is critically stable at sections 1-1 and 2-2 but the slope is unstable at section 3-3. The results show that the factor of safety values given by ordinary method of slices and Janbu method are in close proximity whereas the values indicated by Bishop's method and Morgenstern Price method are closer. However, the factor of safety determined using all methods for section 3-3 is nearly same which indicates that the factor of safety values is dependent upon slope geometry and characteristics.

Table 2. Factor of safety calculated for different sections using deterministic analysis

Sections	Ordinary Method	Bishop Method	Janbu Method	Morgenstern Price Method
1-1	1.041	1.071	1.039	1.069
2-2	1.091	1.245	1.086	1.128
3-3	0.839	0.840	0.818	0.839

Further, the slope sections have been analyzed as infinite slope using a MATLAB program. A MATLAB code was written for the slope stability considering the slope as infinite slope. The results obtained from code are represented through table 3. The results show that for an infinite slope the factor of safety values are very low even under dry condition and particularly very low under the condition when the tension crack is filled with water. The results, however, are not observed to be realistic as the slope is a finite one.

Table 3. Factor of safety for infinite slope

Section	Dry condition	Tension crack filled with water
1-1	0.90	0.44
2-2	0.88	0.43
3-3	0.78	0.38

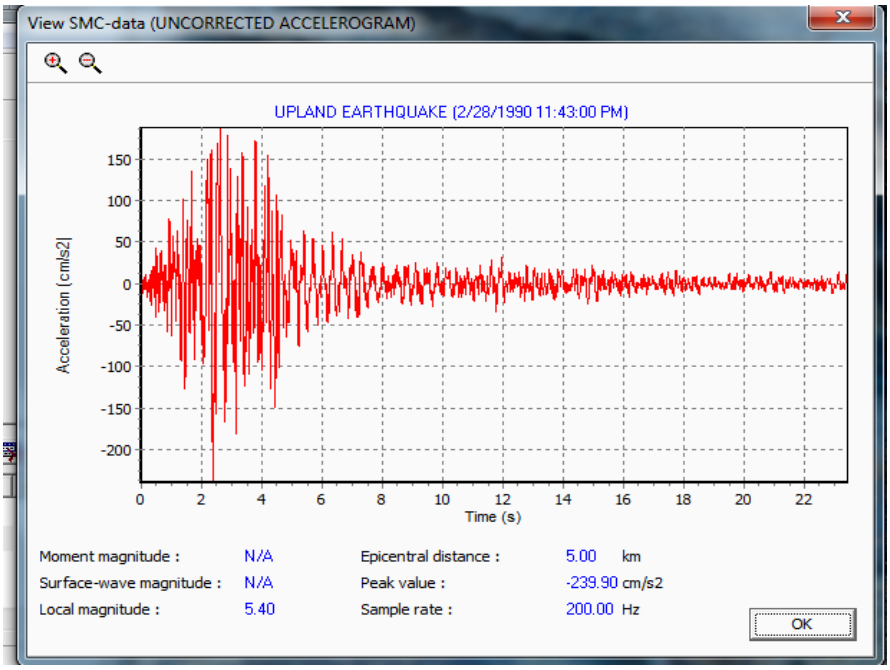


Figure 4. Accelerogram used to simulate dynamic loading input

PLAXIS version 8 has been used to carry out two-dimensional finite element analysis. A Plane strain model is used for geometries with a (more or less) uniform cross section and corresponding stress state and loading scheme over a certain length perpendicular to the cross section (z-direction). Displacements and strains in z-direction are assumed to be zero. However, normal stresses in z-direction are fully taken into account. In this software after defining geometry of the problem, assigning geotechnical specifications of soil layers, segment material and water table, settlement calculation and stress-strain analysis are done through two phases by stage construction capability of the software. The 15-node triangle is the default element which provides a fourth order interpolation for displacements and the numerical integration involving twelve Gauss points (stress points) has been used. The 15-node triangle is a very accurate element that has produced high quality stress results for difficult problems, as for example in collapse calculations for incompressible soils. Three different sections have been analyzed with the help of PLAXIS 8.2 for the following four different conditions: (i) slope is dry, (ii) tension crack filled with water, (iii) cohesion reduced to zero due to vibrations and (iv) Dynamic loading input. The accelerogram used to simulate the dynamic loading input used in the analysis is shown in figure 4.

The finite element modeling of the most critical failure plane at section 1-1 with simulation of dynamic loading is shown in figure 5. The deformed mesh at section 1-1 with simulation of dynamic loading at most critical plane is shown in figure 6.
The boundary elements, particularly at the sharp transitions are observed to incur appreciable displacements. The elements at the toe of the slope indicate large displacements and lead to stress concentrations as is observed from figure 7 showing the stress distribution across the cross-section 1-1. Similarly, the finite element modeling of the most critical failure plane along with the deformed mesh and the stress distribution at sections 2-2 and 3-3 for other conditions was performed to determine factor of safety. The factor of safety values computed using PLAXIS incorporate the consideration of all soil parameters and include the effect of tension crack filled with water, loss of soil cohesion due to vibrations as well as the effect of dynamic loading. The results obtained from PLAXIS for four different conditions are given in table 4.

Table 4. Factor of safety using PLAXIS

Section	1-1				2-2				3-3			
Case	I	II	III	IV	I	II	III	IV	I	II	III	IV
Factor of safety	1.117	1.061	0.983	0.890	1.072	0.886	0.743	0.652	0.934	0.927	0.549	0.456

The factor of safety values indicate that the slope is critically stable at section 1-1 under the two conditions for dry slope and when the tension crack filled with water; whereas at section 2-2 for dry condition of slope only. For the remaining conditions i.e. when cohesion is reduced to zero due to vibrations and under dynamic loading, the slope is unstable at section 1-1 and for the section 2-2 the slope is unstable for the remaining three conditions. At section 3-3, the slope is unstable for all the loading conditions which indicate that the slope stabilization measures have to be undertaken at this section.

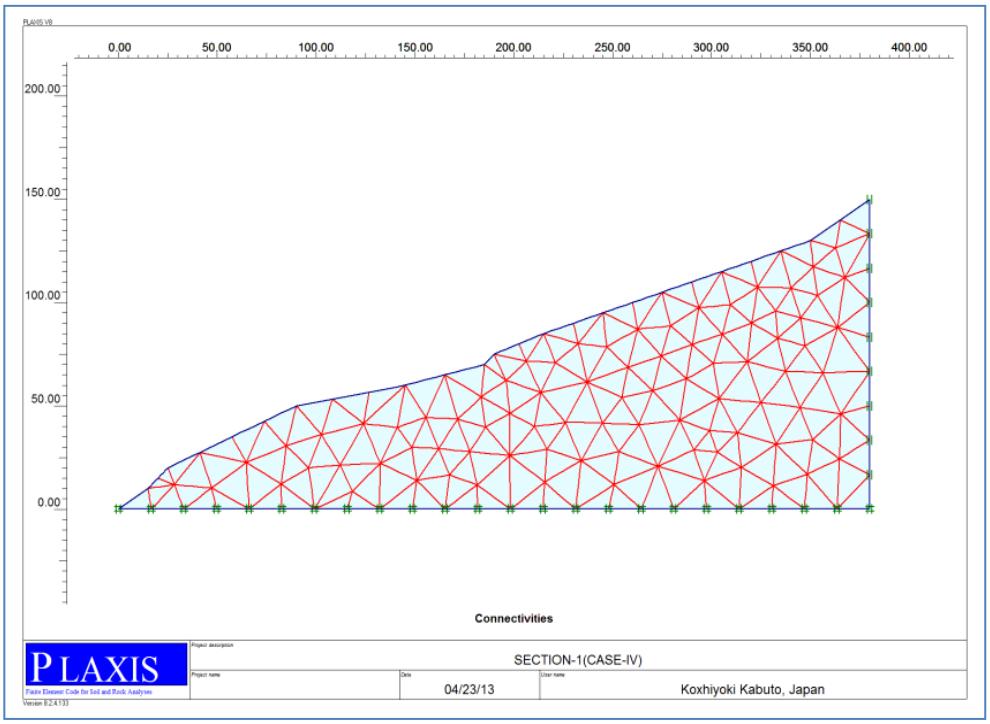


Figure 5. Finite element modeling at section 1-1 with dynamic loading at most critical plane

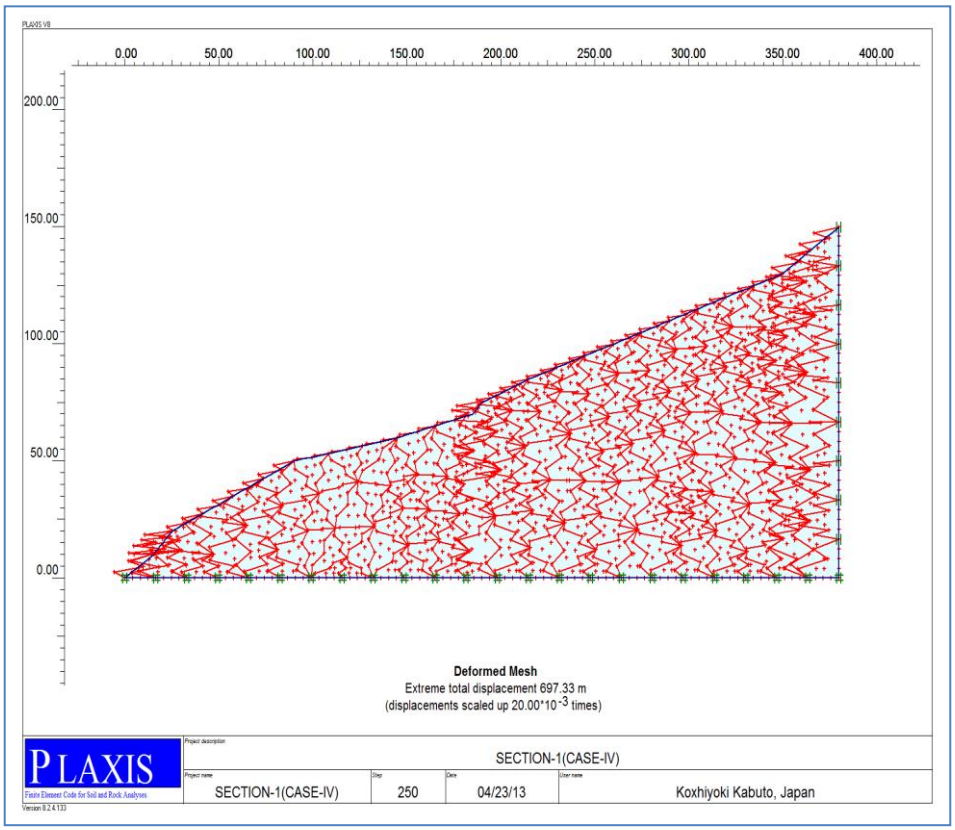


Figure 6. Deformed mesh at section 1-1 with dynamic loading at most critical plane

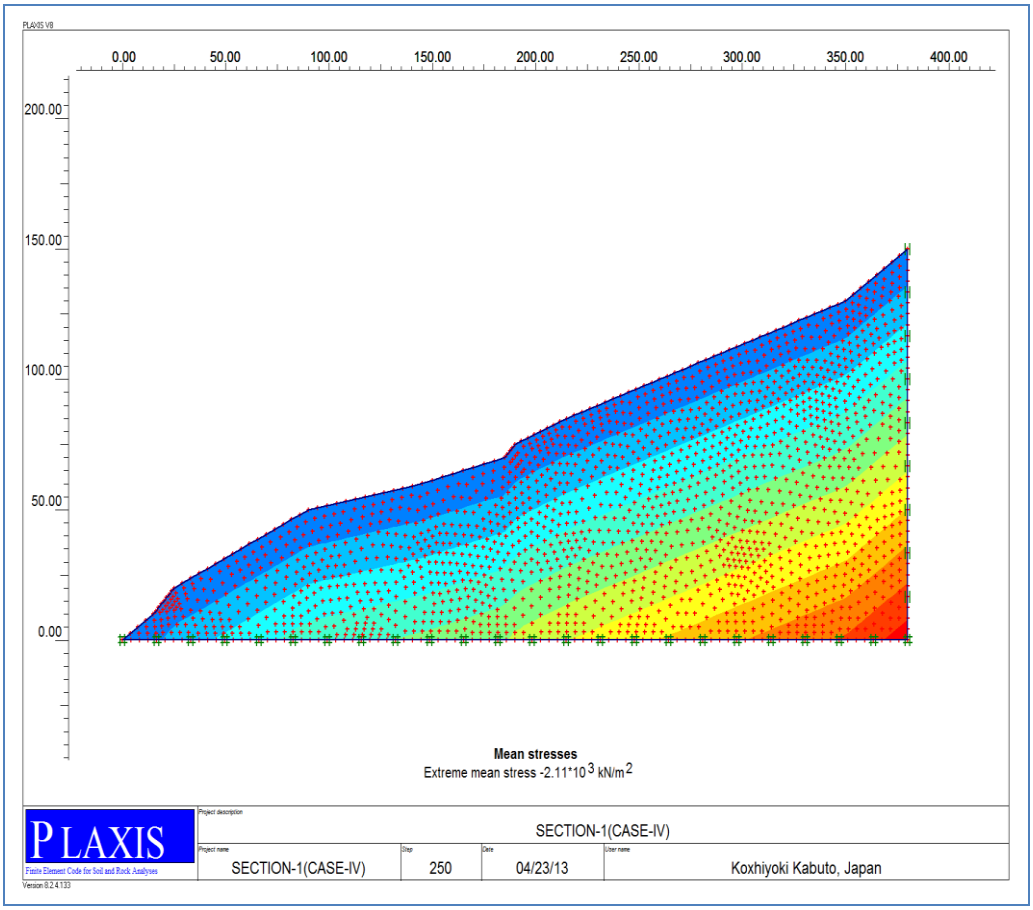


Figure 7. Stress distribution at section 1 with dynamic loading at most critical plane

Probabilistic Slope Stability Analysis Methods

Slope stability is one of the most important issues of concern to geotechnical engineers. Analysis of slope stability is composed of many uncertainties pertinent to lack of accurate geotechnical parameters, inherent spatial variability of geo-properties, change of environmental conditions, unpredictable mechanisms of failure, simplifications and approximations used in geotechnical models. Due to the importance of dam projects and its pertinent costs, determination of dam performance has a significant consequence to decision makers. With respect to the uncertainties of geotechnical parameters, utilizing risk analysis is inevitable in dam projects [20]. Conventional approaches do not take into account many uncertainties in their calculations quantitatively. Also, several conservative safety factors are using to cover some uncertainties which in most cases are more than required, and in some cases less than what is necessary. Actually, it is not possible to distinguish the accurate effect of these safety factors on safety level. By contrast, in probabilistic approaches the safety determination applies more accurately and clearly [21]. Uncertainties in soil properties, environmental conditions, and theoretical models are the reason for a lack of confidence in deterministic analyses [22]. Compared to a deterministic analysis, probabilistic analysis takes into consideration the inherent variability and uncertainties in the analysis parameters. Judgments are quantified within a probabilistic analysis by producing a distribution of outcomes rather than a single fixed value. Thus, a probabilistic analysis produces a direct estimate of the distribution of either the factor of safety or critical height associated with a design or analysis situation. There are several probabilistic techniques that can be used to evaluate geotechnical situations. Specifically, for geotechnical analysis, researchers have conducted probabilistic evaluations using Monte Carlo simulations, Point Estimate

method, and in conjunction with a probabilistic analysis a reliability assessment. Monte Carlo probabilistic analysis has been performed in this study.

Monte Carlo Simulation

The Monte Carlo method was developed in 1949 by John von Neumann and Stanislaw Ulam [23]-[25]. They designated the use of random sampling procedures for treating deterministic mathematical situations. The foundation of the Monte Carlo gained significance with the development of computers to automate the laborious calculation. The first step of a Monte Carlo simulation is to identify a deterministic model where multiple input variables are used to estimate a single value outcome. Step two requires that all variables or parameters be identified. Next, the probability distribution for each independent variable is established for the simulation model, (i.e., normal, beta, lognormal, etc.). Next, a random trial process is initiated to establish probability distribution function for the deterministic situation being modeled. During each pass, a random value from the distribution function for each parameter is selected and entered into the calculation.

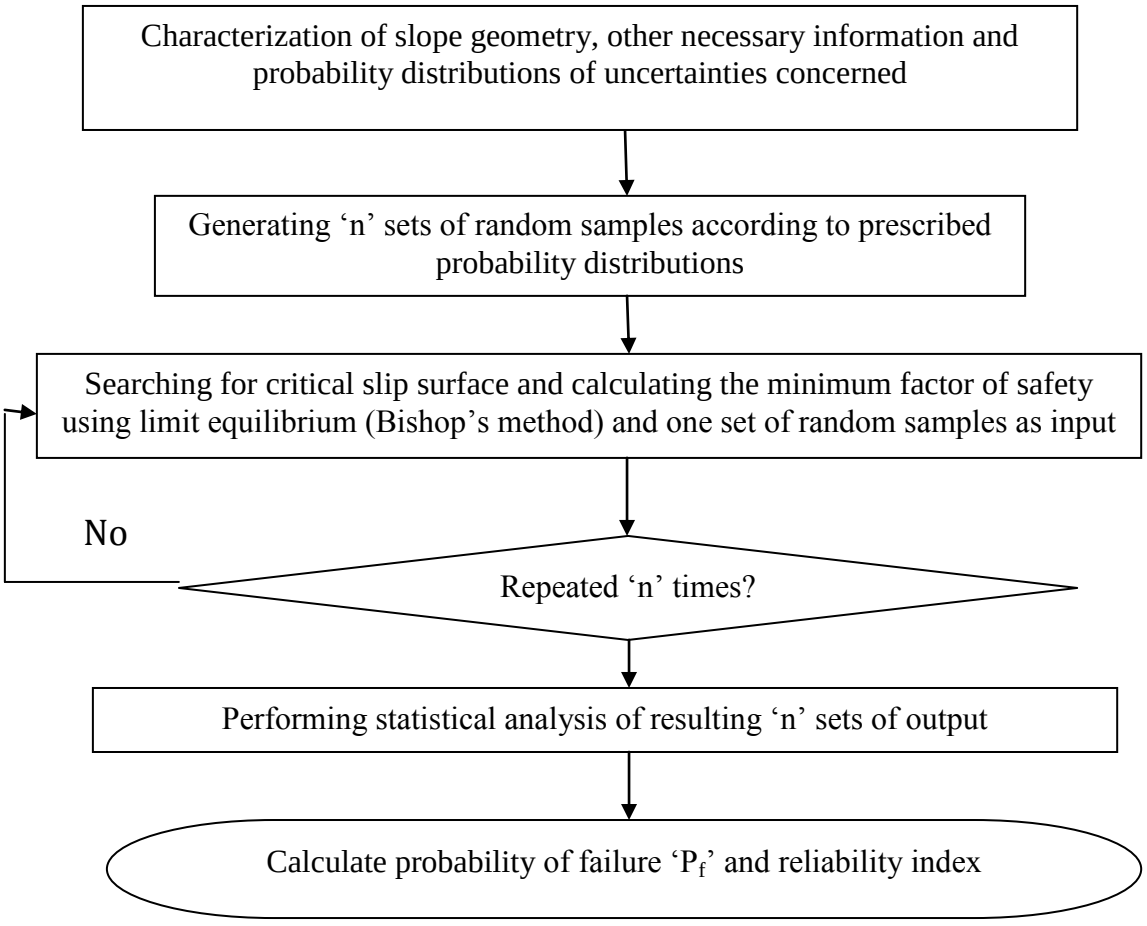


Figure 8. Steps involved in Monte Carlo simulation

Numerous solutions are obtained by making multiple passes through the program to obtain a solution for each pass. The appropriate number of passes for an analysis is a function of the number of input parameters, the complexity of the modeled situation, and the desired precision of the output. The final result of a Monte Carlo simulation is a probability distribution of the

output parameter. Monte Carlo simulation is a powerful tool for slope stability risk analysis. An iterative process using deterministic methods of slope stability analysis is applied in this technique. Monte Carlo simulation is a popular method of slope stability risk analysis among engineers because of its simplicity and no need of comprehensive mathematical and statistical knowledge. This method consists of four steps (figure 8) as below [26][27]: (a) choosing a random value for each input variable according to assigned probability density function, (b) calculating factor of safety by using a proper deterministic slope stability analysis method (such as Janbu, Bishop, Spencer, etc.)[16][17][28] based on selected values in step 1, (c) repeating steps 1 and 2 for many times as necessary and (d) determining distribution function of factors of safety and probability of failure. For the above mentioned sections, probabilistic analysis was performed using Monte Carlo simulations. According to Monte Carlo simulation method, a random value has been selected for each input parameter based on the assigned probability density function and its amplitude. Theoretically, more are Monte Carlo trials the more accurate the solution will be, but the number of required Monte Carlo trials is dependent on the level of confidence in the solution and the amount of variables being considered. Statistically, the following equation has been recommended [29]:

$$N = (\frac{d^2}{4(1-e)^2})\text{^}m \quad (1)$$

Where: N = number of Monte Carlo trials, d = the normal standard deviation corresponding to the level of confidence, e = desired level of confidence, and m = number of variables. The probability density functions of unit weight, cohesion and angle of internal friction, φ adopted in the analysis are shown in figures 9, 10 and 11 respectively. Based on equation (1) for three variables (unit weight, cohesion and phi) and for 90% confidence level 309610 trials have been done with respect to standard deviation of 1.645. The various variables involved in the study, their mean values and type of distribution adopted is summarized in table 5. Reliability index is a rational probabilistic criterion for safety level which can be calculated by the following equation:

$$\beta = (E(FS) - 1)/\sigma(FS) \quad (2)$$

Table 5. Variables involved in Monte Carlo simulations in this study

Variable	Mean value	Standard deviation	Distribution adopted
Unit weight	19.37	1.84	Normal
Cohesion	15.81	8.13	Normal
Phi	19.16º	5.40	Normal

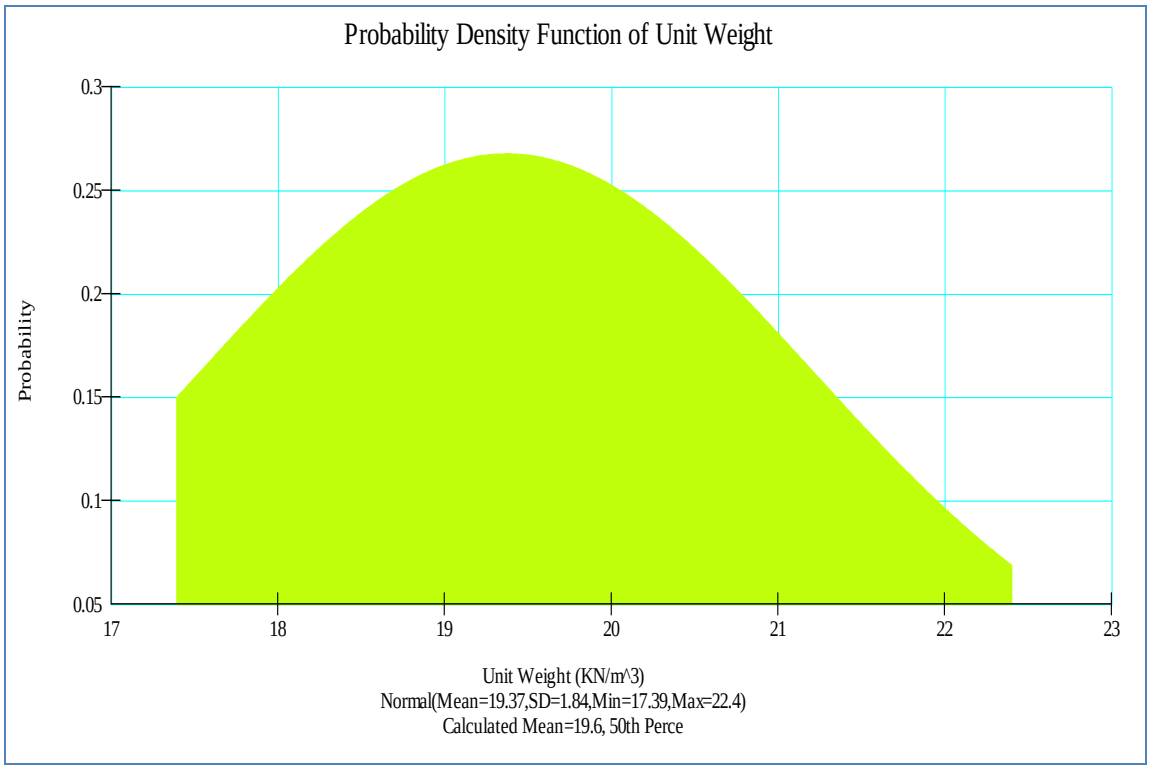


Figure 9. Probability density function of unit weight

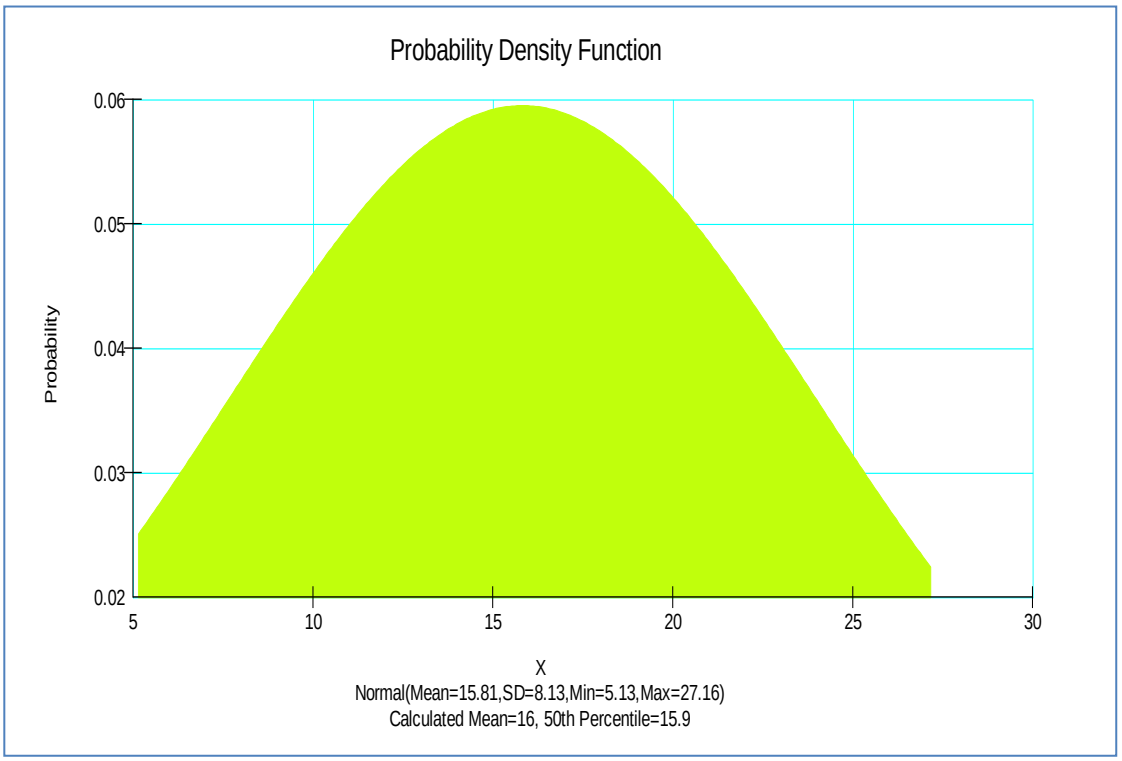


Figure 10. Probability density function of cohesion

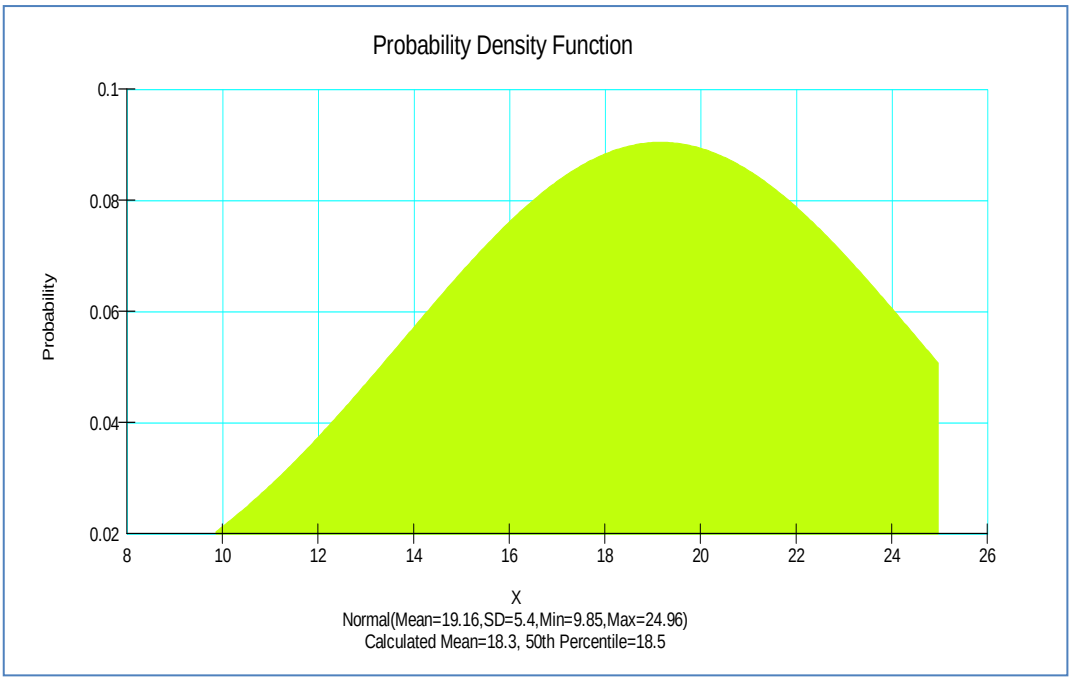


Figure 11. Probability density function for Phi, φ (angle of internal friction)

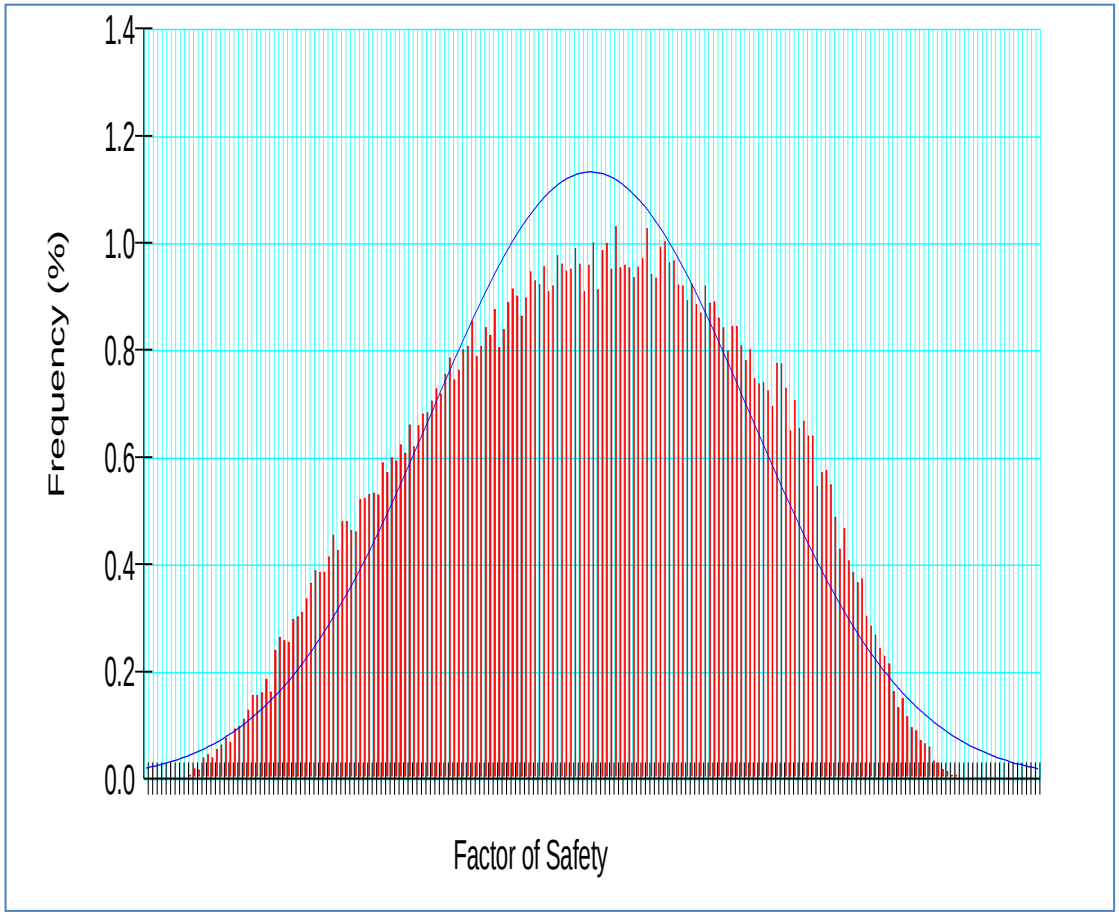


Figure 12. Probability distribution for factor of safety at section 1-1

Where E(FS) and $\sigma(FS)$ are average and standard deviation of safety factors respectively. Reliability index represents the level of reliability of an engineering system and reflects the effects of uncertain parameters on probabilistic analysis. The probability distribution for factor of safety at section 1 - 1, section 2 -2 and section 3 - 3 are shown in figures 12, 13 and 14. The results of probabilistic analysis are represented in Table. 6. As it appears from the table 6 that section 3-3 is most vulnerable towards failure. According to U.S. Army Corps of Engineers [20], for embankment dams, slopes with reliability index of more than 3 are stable. But from table 6, it can be observed that all three sections are having reliability index less than 3 so this slope is not reliable and requires slope stabilization techniques to stabilize it.

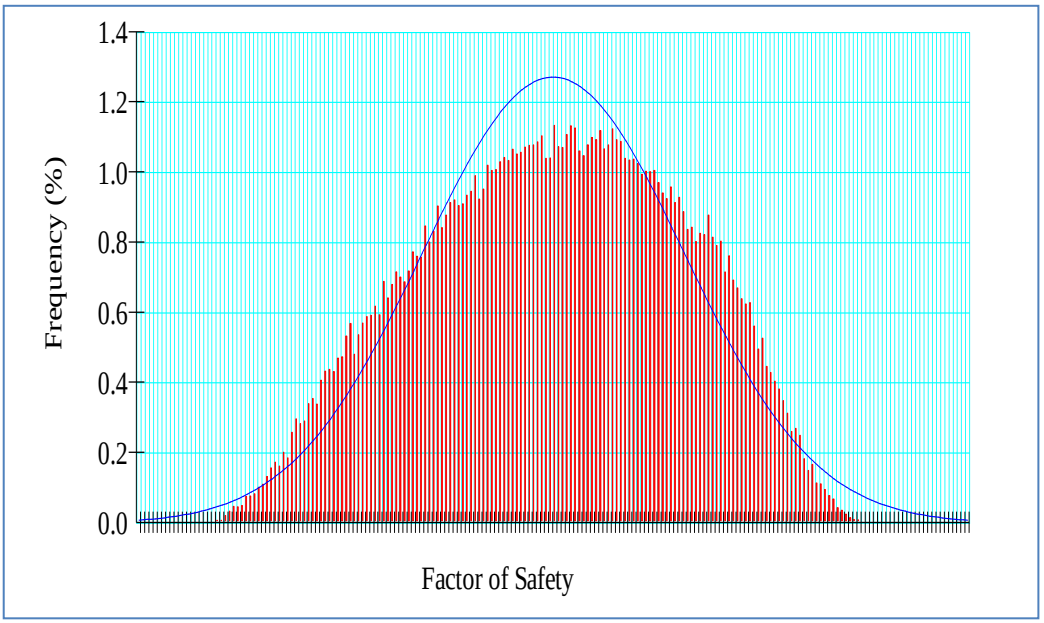


Figure 13. Probability distribution for factor of safety at section 2-2

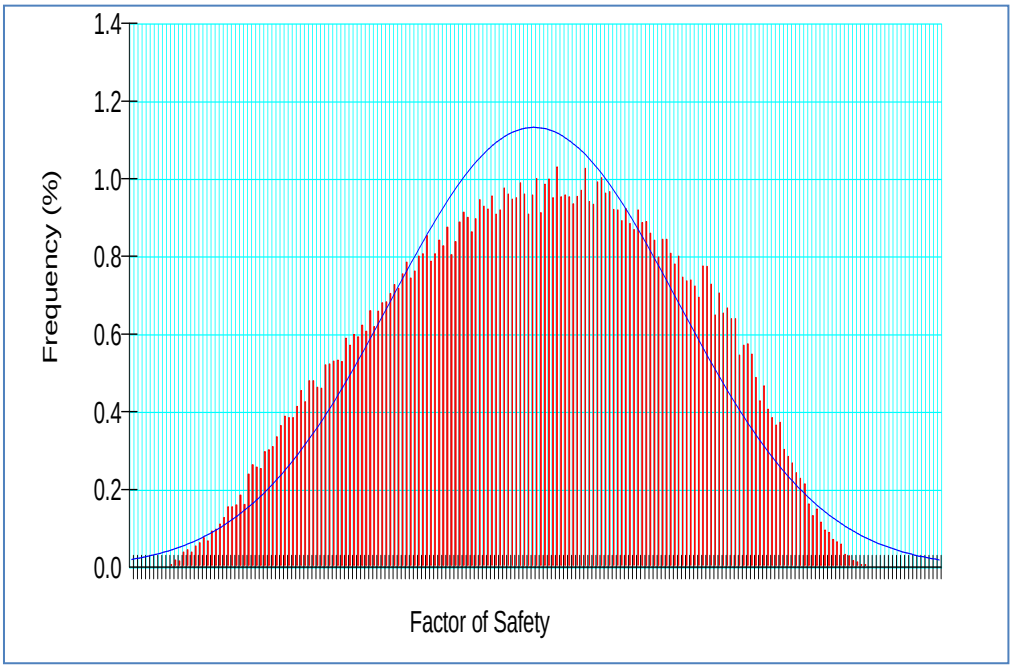


Figure 14. Probability distribution for factor of safety at section 3-3

The results of probabilistic analysis infer that, corresponding to the maximum factor of safety values, the slope at section 2-2 is stable but critically stable at sections 1-1 and 3-3. However, the minimum values of factor of safety indicate that the slope is highly unstable at all the three sections. Corresponding to mean value of factor of safety, slope is critically stable at sections 1-1 and 2-2 but unstable at section 3-3 (higher probability of failure). The results of probabilistic analysis are observed to be more realistic as compared to the results obtained from other methods. Further, the results obtained from probabilistic analysis can be used to determine the probability of failure corresponding to a particular of factor of safety. Therefore an allowable risk criterion can be used to establish a consistent target for the design process. The reliability of the proposed factor of safety can be assessed and the design of the cut slope can be decided accordingly.

Table 6. Results of probabilistic analysis

Section	Mean factor of safety	Min. factor of safety	Max. factor of safety	Reliability index	Probability of failure	Standard deviation
1-1	1.025	0.52	1.48	0.125	44.77	0.205
2-2	1.083	0.54	1.58	0.394	35.74	0.212
3-3	0.8059	0.40	1.17	-1.234	87.78	0.157

Conclusions

The deterministic approach considering different methods of stability analysis namely ordinary method, Bishop's method, Janbu's method and Morgenstern Price method using the iterative capabilities of software SLOPEW and PLAXIS (using dynamic loading input) have been used to assess the stability of a large natural slope. Deterministic approach generally yields conservative values of factor of safety since the input parameters assigned are single valued and the spatial variation of the input parameters is not accounted for. The results obtained from probabilistic approach can be used to determine the probability of failure corresponding to a particular of factor of safety and an allowable risk criterion can be used to establish a consistent target for the design process. The factor of safety obtained from the deterministic analysis indicates that Janbu's method gives the least factor of safety and Bishop's method giving the highest one with Morgenstern Price method yielding the values closer to Bishop's method. While considering the slope as an infinite slope, a smaller factor of safety was obtained which appears to be unrealistic. From probabilistic analysis, it is observed that section 3-3 is most vulnerable towards failure with reliability index of -1.234. Section 1-1 and section 2-2 too have reliability index less than 3 (recommended one for a slope for its stability). Thus, whole slope is vulnerable towards failure and that can be seen during rainy season when the slope faces failures and leads to disruption of traffic on the national highway. Further, the slope is vulnerable towards the dynamic loading with factor of safety reduced to nearly 0.5 under the dynamic loading input.

References

[1] Wright, S.G., Kulhawy, F.H. and Duncan, J.M. (1973). Accuracy of equilibrium slope stability analysis. *Soil Mechanics and Foundations Division*, American Society of Civil Engineers. **99**(10), 783-79.

[2] Anderson, M.G. and Richards, K.S. (1987). *Slope Stability: Geotechnical Engineering and Geomorphology*, New York, USA: John Wiley and Sons.

[3] Albataineh, N. (2006). *Slope Stability Analysis using 2D and 3D methods.* M.Sc. Thesis, University of Akron, Ohio, USA.

[4] Grocott, G., Horrey, P. and Riddolls, B. (1999). *Quantitative Assessment Methods for Determining Slope Stability Risk in the Building Industry.* Building Research Association of New Zealand (BRANZ), Report No. 83.
[5] Fell, R. and Hartford, D. (1997). Landslide risk management. *Proceedings of the International Workshop on Landslide Risk Assessment*, Honolulu, Hawaii, USA, 51-108.
[6] Li, K.S. and Lumb, P. (1987). Probabilistic design of slopes. *Canadian Geotechnical Journal*, **24**(4), 520-535.
[7] Oka, Y. and Wu, T. H. (1990): System reliability of slope stability. *Journal of Geotechnical Engineering*, ASCE, **116**, 1185–1189.
[8] Duncan, J. M. (2000). Factors of safety and reliability in geotechnical engineering. *Journal of Geotechnical and Geo-environmental Engineering*, ASCE, **126**(4):307–316.
[9] Whitman, R.V., (2000). Organizing and evaluating in geotechnical engineering. *Journal of Geotechnical and Geo-environmental Engineering*, ASCE, **126**(7):583–593.
[10] Griffiths, D. V., Huang, J. and Fenton G. A. (2009): Influence of spatial variability on slope reliability using 2-d random fields, *Journal of Geotechnical and Geo-environmental Engineering*, ASCE, **135**(10), doi:10.1061/(ASCE) GT.1943-5606.0000099.
[11] Hasofer, A., and Lind, N. (1974). An exact and invariant first-order reliability format. *Journal of the Engineering Mechanics Division*, ASCE, Vol. **100**, No. EM1, pp. 111-121.
[12] Arsham, H. (1998), "Techniques for Monte Carlo Optimizing," *Monte Carlo Methods and Applications*, vol. 4, pp. 181-229.
[13] Structured Data LLC, 2007, Risk AMP® User Guide and Reference Manual, RiskAMP® Version 2.5, Structured Data LLC, February 2007, http//www.riskamp.com.
[14] Geotechnical Software Suite and Fine Ltd. (2010), *Geo5 User's Guide*.
[15] PLAXIS B.V. (2006), *Reference Manual for PLAXIS 2D version 8.0*, Delft, Netherlands.
[16] Bishop, A.W. (1955). The use of slip circle in the stability analysis of slopes. *Geotechnique*, **5**(1), 7-17.
[17] Janbu, N. (1968). *Slope Stability Computations.* Soil Mechanics and Foundation Engineering Report, The Technical University of Norway, Trondheim, Norway.
[18] Morgenstern, N.R., and Price, V.E. (1965). The analysis of the stability of general slip surfaces. *Géotechnique*, **15**(1): 79–93.
[19] Morgenstern, N.R., and Price, V.E. (1967). A numerical method for solving the equations of stability of general slip surfaces. *Computer Journal*, **9**: 388–393.
[20] U. S. Army Corps of Engineers, (2006), Reliability analysis and risk assessment for seepage and slope stability failure modes for embankment dams, Engineering Technical Letter 1110-2-561, U.S. Army Corps of Engineers, Washington, D.C.
[21] Manafi Ghorabaei, S.M. and Noorzad, A. (2011). Role of risk assessment in repair and rehabilitation of earth dams. *Proceedings of the National Workshop on Operation, Maintenance and Rehabilitation of Dams and Hydropower Plants*, Tehran, Iran: Iranian National Committee on Large Dams (IRCOLD
[22] Alonso, E.E. (1976). Risk analysis of slope and its application to slopes in Canadian sensitive clays, *Geotechnique*, **26**(3):453-472.
[23] Eckhardt, Roger (1987). "Stan Ulam, John von Neumann, and the Monte Carlo method," *Los Alamos Science, Special Issue* (15): 131–137.
[24] Bureau of Indian Standards (1972), Methods of test for soils, Part XIII, Direct shear test. *B.I.S, New Delhi*, IS 2720.
[25] Fishman, G. S. (1995). *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer. ISBN 0-387-94527-X.
[26] Hammond, C.J., Prellwitz, R.W. and Miller, S.M. (1991). Landslide hazard assessment using Monte Carlo simulation. *Proceedings of the 6th International Symposium on Landslide*, Rotterdam, Netherlands, Vol. **2**, 959-964.
[27] Chandler, D.S. (1996). Monte Carlo simulation to evaluate slope stability. *Conference Proceedings on Uncertainty in the Geologic Environment*, Wisconsin, USA, Vol. **1**, 474-493.
[28] Spencer, E. (1967). A method of analysis of the stability of embankments assuming parallel inter-slice forces. *Geotechnique*. **17**(1), 11-26.
[29] Krahn, J. (2004). *Stability Modeling with SLOPE/W*. First Ed., Calgary, Alberta, Canada: GEO-SLOPE/W International Ltd.

Stiffness Based Assessment of Masonry Arch Bridges

Pardeep Kumar
Associate Professor, Department of Civil Engineering, National Institute of Technology Hamirpur (HP) India, 177005

Presenting author & Corresponding author: pardeepkumar.nit@gmail.com

Abstract

There are numerous methods available on date for the structural assessment of masonry arch bridges. Each of these methods has been developed, at different times and places, having its own limitations of use and none of these is a commonly putative method. Among the problems of development of such an approach, the problem of selection of a suitable failure criterion for the prediction of the collapse load is critical. Particularly for arch bridges, involving moments, normal thrust and tangential thrust, the interaction of the axial force and the moments play a vital role in the choosing failure criteria. In view of this, different axial force and moment interactions are reviewed, along with the implementation of the same through a developed stiffness approach based on mechanism method for the prediction of load carrying capacity of the masonry arch bridges. The application of the method has been demonstrated on the bridges tested in field, and the load carrying capacity has been compared.

Keywords: Masonry, Stiffness, Arch Bridges, Mechanism, MEXE

Notations

$[S_c]$	Complete structure stiffness matrix
$[\Delta_c]$	Complete joint displacement matrix
$[JL_c]$	Joint load matrix for complete structure
$[R_c]$	Complete support reaction matrix
$[K_i]$	Member stiffness matrix
P_o	Maximum concentric axial force
M_o	Maximum moment at an eccentricity of d/4
σ_t	Tensile strength of masonry

Introduction

Masonry arch bridges have been a legacy of past, but are built hardly now-a-days. The newer materials with better structural properties have overshadowed the use of masonry and the art of masonry arches has been kerbed to the papers. In most of the countries where masonry arch bridges exist on railway and road network, the first choice of the bridge owners is to use MEXE (Military Engineering Experimental Establishment) [1] method for assessment of such bridges. This mechanism approach to arch collapse, originated from the first work of Pippard and Ashby [2] and Pippard [3]. The identification of location of a number of hinges at the arch intrados and extrados to transform it into a mechanism yielded the minimum load. The limit load is obtained through the application of the kinematic theorem [4] that takes the position of the hinges as the unknowns of the problem. This approach finds its latest results in the work by different authors [5]-[11]. The method was originally developed, based on the minimum strain energy principle and later used it during Second World War to develop tables of allowable weights for wheeled and tracked vehicles for military use [12]. The original MEXE method was then developed from these basic tables in form of readily usable

nomogram. The referred MEXE method is empirical and is a working stress method based on the elastic analysis but provides little information in regard to the considerations of the serviceability. All the subsequent methods are marginal improvement over the original one and have tried to overcome the shortcomings in the earlier methods. More so, with the increasing advent of computers, many computer based assessment methods have developed in recent years, which are in use in different parts of the world. To list, such methods include, CTAP developed by Bridle and Hughes, which is based on Castigliano's elastic strain energy method, MINIPONT developed by Department of Transport is computerized version of MEXE method, program ARCHIE developed by Harvey and Smith and program ARCH developed by Cascade Software Ltd [13]. The program ARCHIE and ARCH are based on the mechanism method of assessment. Heyman has described in detail the development and use of mechanism method of assessment [4].

Assessment of existing structures is always considered more tedious than the design of new structure. The confidence in new design can be well achieved through properly designing well understood part and relying on unquantified additional safety for the remainder part. Existing structures often rely on behaviours that the engineer prefers to keep as safety factor. How those actions are used in assessment is a matter for individual judgement and any guidance that obscures the reliance on alternative load paths is inherently dangerous because it reduces the scope of the engineer's judgement [14].

The assessed load carrying capacities of bridges using different methods also vary widely, due to variety of assumptions underlying the idealisation, load application, material properties, hinge formation criteria and mechanism etc. In the proposed formulation, the four obvious hinge positions are not selected, but, instead based upon the interaction of bending moment and axial force present at the section at different instants of loading, the successive formation of the hinges takes place until a mechanism is formed. The approach has been fully computerized through a program written in Fortran [15]. From the assessed capacity of a bridge, the procedure to determine the load rating is also laid down.

Experimental Investigation of Moment –Axial Force Interaction

The control specimens were constructed in 1:4 cement sand mortar, having average cross-section 105 mm x 223 mm. Hand moulded class-A bricks of conventional size were used for the construction of prisms. The average height of the prisms was 681 mm, which was greater than the span of the test specimens. Three specimens were tested each at six different precompression levels. The arrangement was simply supported over the knife-edge supports to avoid any fixity. These were tested after curing of 28 days, under monotonically increasing two point load system at middle third points as shown in the Figure 1a. The test arrangement is shown in Figure 1b. The weight of the assembly was added to the load value. The experimental failure loads in masonry at different levels of precompression is given in Table 1.

The horizontal compressive force was applied along the centerline of the prism specimens. Three specimens each have been tested at different levels of the precompression corresponding to axial stress of 0%, 10%, 20%, 40%, 50%, and 60% of the crushing strength of the masonry from the uniaxial compression test on similar type of prism.

The bending tensile strength σ_t without any precompression can be found on equilibrating the internal and external moments, as given under.

$$\sigma_t = \frac{WL}{bd^2} \tag{1}$$

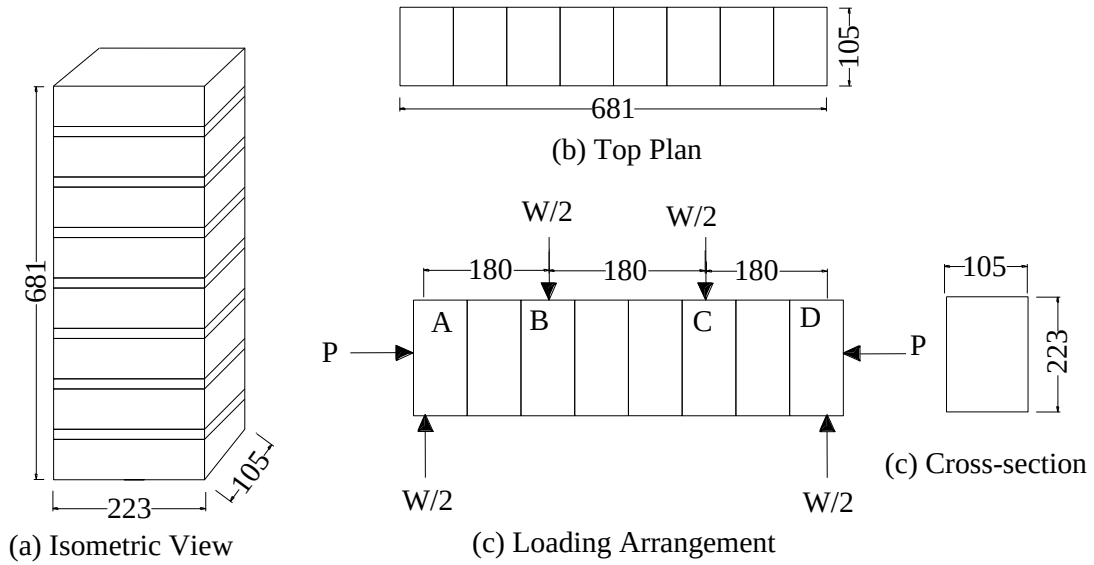


Figure 1a Test specimen and loading details (Dimensions in mm)

At 0% precompression, the bending tensile strength of the specimens tested has been determined as 0.29 N/mm2. The plot of non-dimensional parameters P/Po versus M/Mo is drawn (Figure 2) for the experimental values. The axial loads are normalized with respect to $P_o = \sigma_c\, b\, d$ and moments are normalized with respect to $M_o = 0.125\sigma_c\, b\, d^2$.

Table 1. Experimental failure loads in masonry at different levels of precompression

No. of specimens	Average size			Precompression		Exp. Failure Load, W (kN)
	Width, b (mm)	Depth, d (mm)	L (mm)	% of crushing strength	Load, P (kN)	
3	109.67	229.17	681	0	0.00	2.803
3	110.17	228.73	683	10	13.42	17.756
3	108.96	229.33	680	20	26.84	29.194
3	109.83	229.35	677	40	53.68	41.496
3	110.50	228.56	682	50	67.10	51.709
3	108.50	227.89	684	60	80.52	62.490

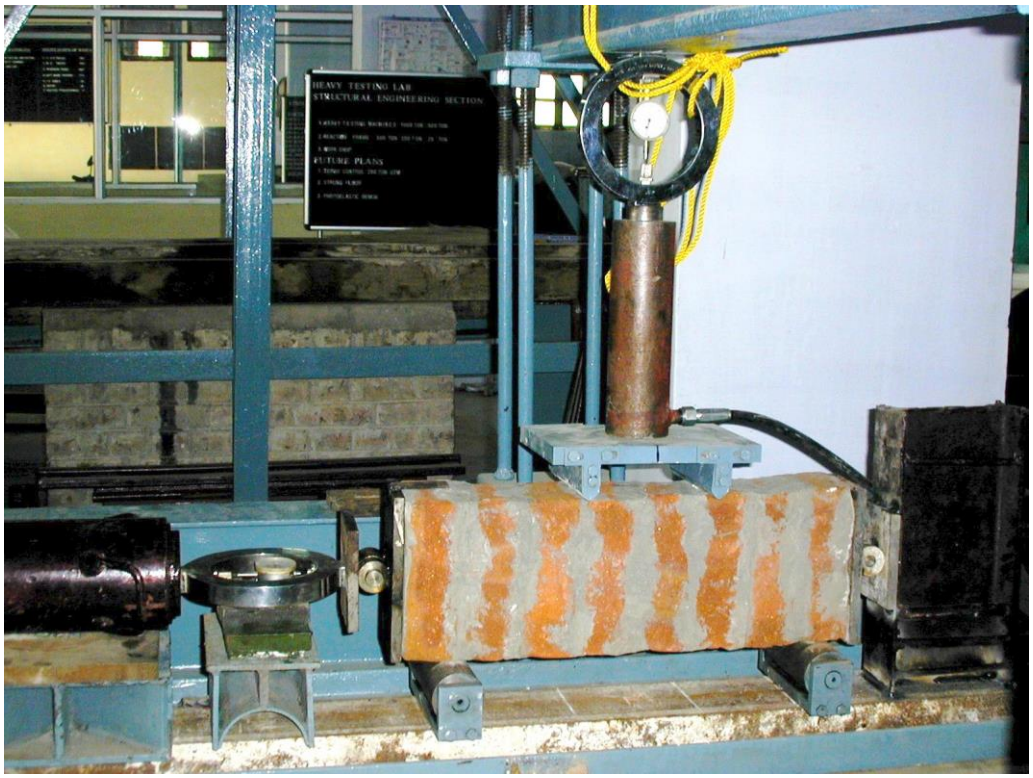


Figure 1b Test arrangements for determination of flexural bond strength of masonry

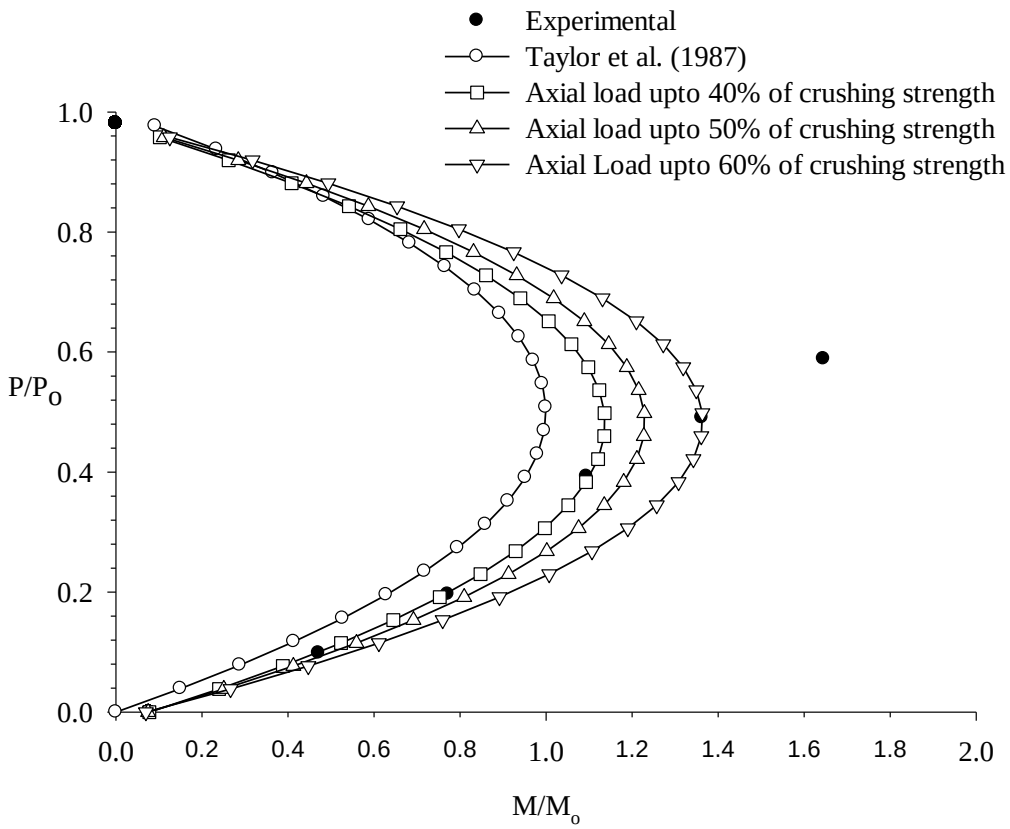


Figure 2 Comparison of experimental axial force-moment interaction and limit state interaction developed by Taylor and Malinder [16]

It has been observed during the testing that at higher precompression levels, with the increase in the transverse loads, the precompression automatically increased. Although, this has been taken care of by releasing the pressure in the load cell to maintain constant precompression. This may be one of the reasons that at precompression levels of 50% and 60% of crushing strength the transverse failure loads recorded are on the higher side, leading to M/Mo ratio greater than 1 as seen in Figure 2. The moment-axial force relationship has been extrapolated corresponding to all range of precompression levels and a parabolic equation is fitted to the normalized data as given under:

$$\frac{M}{M_o} = 0.0699 + 5.3476\left(\frac{P}{P_o}\right) - 5.5224\left(\frac{P}{P_o}\right)^2 \tag{2}$$

Keeping in view the problem encountered during the test program, the reliability of the point corresponding to precompression levels of 50% and 60% of crushing strength of masonry are low. Hence, discarding the points corresponding to these precompression levels modifies the best-fit equation and bring to the close proximity of that derived by Taylor and Mallinder [16]. Discarding the one point corresponding to precompression level of 60% of crushing strength would modify the equation as under:

$$\frac{M}{M_o} = 0.0751 + 4.7840\left(\frac{P}{P_o}\right) - 4.9550\left(\frac{P}{P_o}\right)^2 \tag{3}$$

Discarding two points corresponding to precompression levels of 50% and 60% of crushing strength modifies the equation as:

$$\frac{M}{M_o} = 0.0779 + 4.3994\left(\frac{P}{P_o}\right) - 4.5662\left(\frac{P}{P_o}\right)^2 \tag{4}$$

All these equations are plotted in Figure 2. Neglecting last two points provides an equation closely matching with the one available in the literature.

Taylor and Mallinder have reported the axial force/bending moment interaction for the limit state of rectangular masonry section. The strain distribution was assumed linear whereas a non-linear parabolic relation was assumed for the variation of stresses with strains. The moment-axial force interaction diagram represented by Eqns. 2, 3, and 4 has been compared in Figure 2 with that of analytically developed interaction (Eqn. 5) by Taylor and Mallinder [16].

$$\frac{M}{M_o} = 4\left(\frac{P}{P_o}\right) - 4\left(\frac{P}{P_o}\right)^2 \tag{5}$$

The proposed interaction equations take into account the masonry tensile strength, indicated by the presence of constant term in the equations. Despite the lack of the sophisticated equipment used in the present investigation, a reasonable correlation has been obtained.

The Basis of Proposed Method

Although the behaviour of the arches is fundamentally non-linear due to the axial force-moment interaction, the proposed method utilizes linear elastic theory. The linear elastic analysis under the action of unit live load is carried out and the load factors are computed by steering the analysis moments and axial forces to satisfy the axial force-moment interaction to incorporate plastic hinge at appropriate locations, until the formation of a collapse mechanism.

The moment-axial force interaction is the most important parameter to determine the load carrying capacity of the masonry arch bridges. Wherever the combination of moment and axial force developed in the section lies on this surface, a hinge shall be assumed to form at that section and the hinge will continue to rotate when further load is applied till the arch is converted to a mechanism.

Considering the unit width of the arch ring, it can be divided into a sufficient number of segments along the barrel centerline. Each segment can be assumed to be a straight line joining the two nodes. These segments can be represented by a beam element having appropriate material and sectional properties. The end nodes are fixed at the springing line to provide restraint against any horizontal, vertical, or rotational movement. The arch is analysed first under the dead loads imposed due to self-weight of the arch ring and the load of the overlaying fill. The weight of the fill is calculated over each segment and is applied as equivalent nodal loads at its two nodes. The arch is then analysed under a unit live load applied at quarter point. The obtained values of bending moment and axial force due to dead and live load so obtained are modified to satisfy the limit state envelope at every node. A step-by-step linear analysis is performed to locate the four hinge locations and the corresponding total load on the bridge is the failure load. The details of the method are reported elsewhere [15].

The Stiffness Method

The proposed formulation is based on stiffness approach, where a set of simultaneous equations in form of matrices are developed and solved. The representative set of equation can be expressed as

$$[S_c][\Delta_c] = [JL_c] + [R_c] \tag{6}$$

Defining $[S_c]$as the complete structure stiffness matrix, $[\Delta_c]$ as the complete joint displacement matrix, $[JL_c]$as the complete joint load matrix, and $[R_c]$ as complete support reaction matrix. In the development of the several matrices of Eqn. 6 all components of joint displacement, joint load and support reaction, which form the elements of respective matrices, must be described with respect to a same system of axes, i.e. the reference axes for the entire structure. The formulation of this method is given in many standard texts [17][18].

Each segment is modelled as a beam element that has either constant or variable moment of inertia over its length positioned in the local axis Xm-Ym, with origin at j-end of the member and Xm axis directed towards k-end of the member. If the beam element is subjected to general displacements θ_p, θ_q, δ_r, δ_s, δ_t and δ_u of its ends, the resulting end actions can be determined as shown in Figure 3. Hence, the force - displacement relationship in local system, for a prismatic member can be expressed as given by Eqn. 7.

The Beam Element

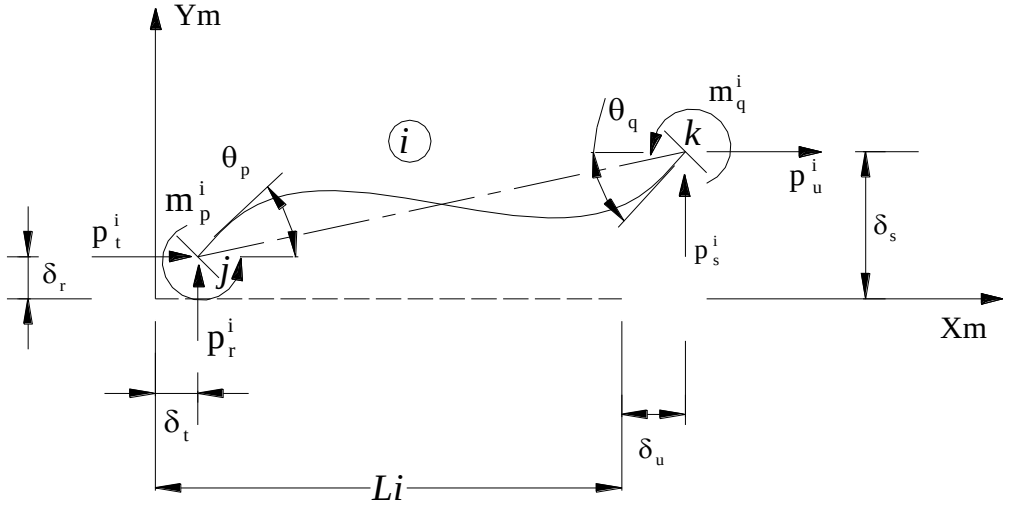


Figure 3 General displacement of a typical beam element with restrained ends.

$$\begin{Bmatrix} m_p^i \\ m_q^i \\ p_r^i \\ p_s^i \\ p_t^i \\ p_u^i \end{Bmatrix} = \begin{bmatrix} \frac{4EI}{L} & \frac{2EI}{L} & \frac{6EI}{L^2} & \frac{-6EI}{L^2} & 0 & 0 \\ \frac{2EI}{L} & \frac{4EI}{L} & \frac{6EI}{L^2} & \frac{-6EI}{L^2} & 0 & 0 \\ \frac{6EI}{L^2} & \frac{6EI}{L^2} & \frac{12EI}{L^3} & \frac{-12EI}{L^3} & 0 & 0 \\ \frac{-6EI}{L^2} & \frac{-6EI}{L^2} & \frac{-12EI}{L^3} & \frac{12EI}{L^3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{AE}{L} & \frac{-AE}{L} \\ 0 & 0 & 0 & 0 & \frac{-AE}{L} & \frac{AE}{L} \end{bmatrix} \begin{Bmatrix} \theta_p \\ \theta_q \\ \delta_r \\ \delta_s \\ \delta_t \\ \delta_u \end{Bmatrix} \tag{7}$$

$$\{p_i\} = [K_i]\{\delta_i\} \tag{8}$$

where the matrix $\{\delta_i\}$ represent the components of end displacements of member i, the matrix $\{p_i\}$ represents the components of end actions required to maintain equilibrium of member i when subjected to general end displacements and the matrix $[K_i]$ represent the components of member end actions resultant from independent application of unit values of the possible end displacements. This is also referred to as member stiffness matrix.

Transformation Matrix: Beam Element

For general end displacements of the restrained member, the components of end actions have been defined with respect to the local axis. The components of end actions in local axes can be transformed in terms of the components of end actions with in the frame of the global axes as

$$\{p_i\} = [T_i]\{\overline{p_i}\} \tag{9}$$

where the matrix $\{p_i\}$ represents the components of end actions for member i in local system; the matrix $\{\overline{p_i}\}$ represents the components of end action for member i in the frame of global axes and $[T_i]$ is the transformation matrix as given below.

$$[T_i]=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & C_x & 0 & -C_y & 0 \\ 0 & 0 & 0 & C_x & 0 & -C_y \\ 0 & 0 & C_y & 0 & C_x & 0 \\ 0 & 0 & 0 & C_y & 0 & C_x \end{bmatrix} \tag{10}$$

where C_x and C_y are direction cosines.

On the similar basis the relation between the components of end displacements of member i with respect to local axis Xm-Ym and the reference axes X-Y can be established as

$$\{\delta_i\}=[T_i]\{\overline{\delta_i}\} \tag{11}$$

where matrix $\{\delta_i\}$ represents the components of member end displacements in the system of local axes and the matrix $\{\overline{\delta_i}\}$ represents the components of member end displacements in a system of global axes.

Thus Eqn. 13 gives member stiffness matrix with respect to global axes.

$$\{\overline{p_i}\}=[T_i]^T[K_i][T_i]\{\overline{\delta_i}\} \tag{12}$$

$$[\overline{K_i}]=[T_i]^T[K_i][T_i] \tag{13}$$

is defined as the transformed member stiffness matrix expressed with respect to the arbitrary system of global axes X-Y. The Eqn. 11 describes the relationship between the components of the end displacements and the end actions of member i in the frame of global axes. Once the system of equilibrium equations are generated in the global axes, the independent components of the unrestrained joint displacements are evaluated by substituting the boundary conditions and solving the set of residual equations expressed as

$$[S_{uu}][\Delta_u]=[JL_u] \tag{14}$$

and the components of the support reactions are determined by solving that set of equations expressed as

$$[S_{ru}][\Delta_u]-[JL_r]=[R_r] \tag{15}$$

Assumptions of Method

The method is based on the following assumptions:

- At the point of hinge the axial force and shear force resisting capacity is not impaired and it continues to resist the axial force and shear force.
- The effect of the shear forces on moment-axial force interaction envelope has been ignored.
- The point where a hinge is formed will continue to rotate.

Comparison of Proposed Analysis to Field Results

In order to validate the proposed model for indigenous constructions, the two sets of arches were constructed and tested in the laboratory. The description of these arches can be found elsewhere [19]. The material properties and other input data used in the analysis are tabulated in Table 2. The obtained results with different axial force moment interaction have been compared with the experimental results in Table 3.

From the comparison of the results, it can be observed that inclusion of tensile strength of the masonry considerably improves the results. The loads predicted by using Eqn. 2 are in excess of test maximum loads. The estimated load carrying capacity is 40.94% in excess for first set of arches and 38.41% in excess for second set of arches in comparison to the experimental loads. The use of Eqn. 4 can reliably simulate the test results using the material properties given in Table 2. These properties have achieved through the experimental investigations on the indigenous masonry in the laboratory. The estimated load carrying capacity is only 11.96 % in excess for first set of arches and 5.52 % in excess for second set of arches in comparison to the experimental loads. On the other hand, using Eqn. 5 derived by Taylor and Mallinder [16], the estimated load carrying capacity is too low in comparison to the test results for both the sets of test arches. In view of this the equation derived from the experimental investigations after neglecting the points with unreliable data is proposed to be used for correctly predicting the load carrying capacity of a masonry arch bridge in fairly good condition with indigenous constructions.

Table 2. Material properties used in the analysis

Material	Property	Value	Units
Brick Masonry	Modulus of Elasticity	3723	N/mm^2
	Compressive Strength	5.84	N/mm^2
	Tensile Strength	0.29	N/mm^2

Table 3. Comparisons of failure loads (kN)

Bridge	Test Maximum Load (kN)	Load Predicted from Proposed Method (kN)		
		Using Eqn. 2	Using Eqn. 4	Using Eqn. 5
Arches AV1 & AV2	55.10 *	77.66	61.69	31.36
Arches AF1 & AF2	74.25 *	103.07	78.35	55.55

* The values are average for the two similar models.

Because of the difference in the method of analysis (Step-by-step linear) and the actual behaviour (non-linear) of the structures it is difficult to get the actual response of the structures from the proposed analysis. The load-deflection response achieved from the structures has been compared with the experimental behaviour of both the set of the arches. The comparison of the load deflection under the load point for test arches is shown in Figures 4 and 5 respectively. The predicted average deflections for test arches AV1 and AV2 are only 8.7 % of the experimentally observed value and 31.2 % for arches AF1 and AF2. It can be inferred from the comparison that the method can predict the load carrying capacity but not the deflections.

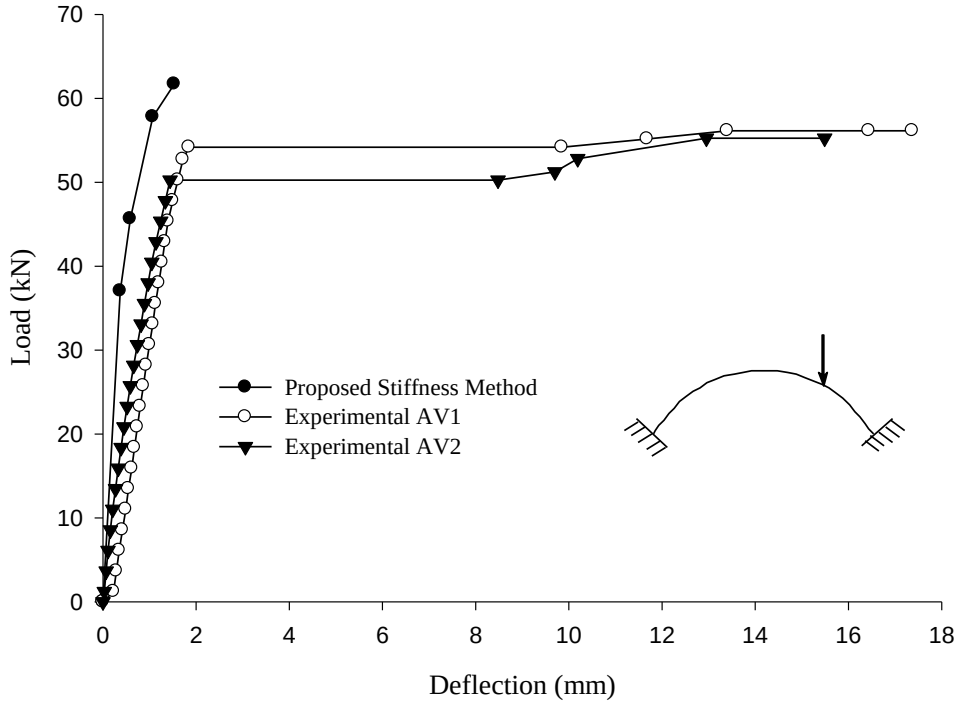


Figure 5 Comparison of predicted and experimental load-deflection behaviour of the arches AV1 and AV2

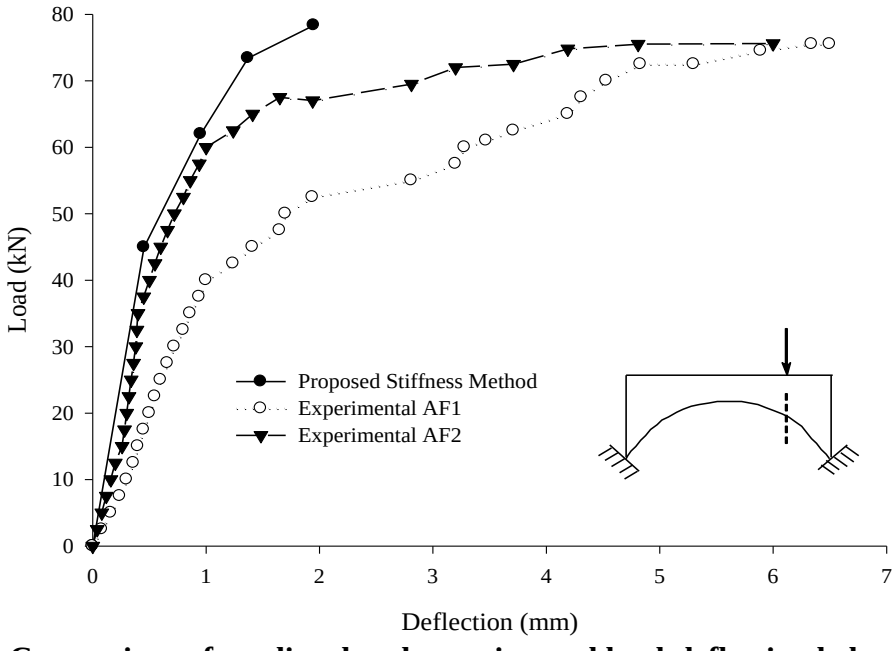


Figure 6 Comparison of predicted and experimental load-deflection behaviour of the arches AF1 and AF2

Conclusions

The axial force-moment interaction can be effectively used for the prediction of load carrying capacity of the masonry arch bridges. In the proposed method the experimentally determined axial force-moment interaction has been verified and implemented successfully to predict the

collapse load. The method can predict the load carrying capacity within an acceptable range of variation. For first set of test arches the predicted values differ by 11.56 % and by 5.52 % for second set of test arches. The predicted values are on higher side, which may be attributed to the use of material properties determined from the control specimens.

The proposed interaction, accounts for some minimum tensile strength of the masonry. The proposed method can predict the collapse load on the basis of formation of adequate number of hinges leading to conversion to a mechanism.

The frame analysis program automated for the formation of the hinges and further leading to failure on formation of the mechanism provides a sufficiently quick and simple method for determination of the load carrying capacity of the masonry arches assuming a unit width of the arch ring.

Acknowledgement

The help rendered by Professor N. M. Bhandari, Professor at IIT Jodhpur, is duly acknowledged for the development of the frame analysis program automated for the formation of the hinges and further leading to failure on formation of the mechanism.

References

[1] Department of Transport, (2001) Assessment of Masonry Arch Bridges by Modified MEXE Method", Vol. 3, Section 4, Part 4 BA 16/97, Amendment No. 2.
[2] A J S Pippard and R J Ashby, An Experimental Study of The Voussoir Arch. (Includes Photographs, Plates And Appendix), Volume 10 Issue 3, January 1939, pp. 383-404.
[3] Pippard, A. J. S., "The Approximate Estimation of Safe Loads on Masonry Arch Bridges", Civil Engineer in war, Vol. 1, 365, Institution of Civil Engineers, London, 1948.
[4] Heyman, J., (1982) The Masonry Arch, Ellis Horwood, Chichester.
[5] Crisfield, M.A., Packham, A.J., (1988). A mechanism program for computing the strength of masonry arch bridges, Transport and Road Research Laboratory, Dept. of Transport, Research Report 124.
[6] Harvey, W.E.J., (1988). Application of the mechanism analysis to masonry arches, The Structural Engineer, 66, 77-84.
[7] Blasi, C., Foraboschi, P., (1994). Analytical approach to collapse mechanisms of circular masonry arches, J. Struct. Engrg., ASCE, 120, 2288-2309.
[8] Falconer, R.E., (1994). Assessment of multi-span arch bridges, Proc. 3rd Int. Conf. on Inspection, Appraisal, Repair and Mainteinance of Buildings and Structures, Bangkok, 79-88.
[9] Gilbert, M. and Melbourne, C., (1994). Rigid-block analysis of masonry structures, The Str. Eng., 72, 356- 361.
[10] Hughes, T.G., (1995). Analysis and assessment of twin-span masonry arch bridges, Proc. Instn. Civ. Engrs., 110, 373-382.
[11] Como, M., (1998). Minimum and maximum thrusts states in Statics of ancient masonry bridges, Proc. II Int. Arch Bridge Conf., A. Sinopoli ed., Balkema, Rotterdam, 321-330.
[12] Pippard A. J. S. and Chitty L. A., (1951) Study of the voussior arch, National Building Studies Research Paper No. 11, HMSO London.
[13] Ng, K. H., Fairfield, C. A., (1999) Finite Element Analysis if Masonry Arch Bridges, Proc. Instn. Civ. Engrs. Structs. & Bldgs, Vol. 134, pp. 119-127.
[14] Harvey, Bill, Kumar, Pardeep and Bhandari, N. M., (2007) Mechanism Based Assessment of Masonry Arch Bridges SEI Discussion, Structural Engineering International, Vol. 17, Number 1, pp. 97-99.
[15] Kumar, Pardeep and Bhandari, N. M., (2006) Mechanism Based Assessment of Masonry Arch Bridges IABSE quarterly publication Structural Engineering International, Vol. 16, Number 3, pp. 226-234.
[16] Taylor, N. and Mallinder, P., (1987) On Limit State Properties of Masonry, Proceedings Institution of Civil Engineers, Part 2,Vol. 83, pp. 33-41.
[17] Weaver W. and Gere, J. M., (1986) Matrix Analysis of Framed structures, Second Edition, CBS Publishers and Distributors, Delhi, India.
[18] Beaufait, F. W. et al., (1970) Computer Methods of Structural Analysis, Prentice Hall.
[19] Kumar, Pardeep, (2005) Load Rating and Assessment of Masonry Arch Bridges, Ph.D. Thesis, Indian Institute of Technology, Roorkee.

Identification and Computation of Space Conflicts Using Geographic Information Systems

V.K. Bansal

Associate Professor, Department of Civil Engineering, National Institute of Technology (NIT), Hamirpur, Himachal Pradesh. India.

Presenting author & corresponding author: vijaybansal18@yahoo.com,vkb@nith.ac.in

Abstract

Various types of spaces for different purposes on various positions at various times are required to execute various construction activities on a construction site. Labors, equipment, materials, temporary facilities, and structure to be developed share the limited space available on a construction site. Planners use four-dimensional (4D) CAD modeling of the execution sequence to understand and generate the space requirements. The 4D CAD modeling simulates the construction process by linking execution schedule with a 3D model to visualize the construction sequence. 4D modeling is found helpful in the construction space planning. However, 4D CAD modeling lacks in considering the topography and surroundings when construction is in the hilly regions. In the present study, a geographic information system (GIS) has been utilized for the space planning. GIS facilitates the modeling of topography and existing surroundings. The components corresponding to different activities in the schedule and multiple types of spaces corresponding to various activities defined in the execution schedule have been generated in the *SketchUP*. A GIS-based procedure has been developed in *ArcGIS 10*, a GIS software, that enables identification and computation of the construction space conflicts before actual implementation of the schedule.

Keywords: Geographic Information System, Project Management, Workspace

Introduction

Deficiencies in the space planning results congested jobsite, loss of productivity, space conflicts, and schedule interference or delay [Guo (2002)]. Construction site engineers usually arrange daily activities on the jobsite according to the planned execution sequence. Existing literature suggests that like any other resource, construction activities also need execution space as a resource that need to be planned before the finalization of a schedule [Akinci et al. (2002a)]. It is impractical for a planner to visualize the dynamic multiple types of space requirements mentally because it changes with time/schedule like any other resource requirement in the construction industry. 4D CAD-based production models were used for the automated generation of spaces required by the construction activities to reduce time-space conflicts [Akinci et al. (2002b)].

Despite of many researches and applications of the 4D CAD technologies their use is not very common in the construction industry. After 4D CAD, there has been a major revolution of building information modeling (BIM) that also provides a mechanism to develop a conflict free construction schedule [Choi et al. (2014)]. BIM facilitates 3D modeling, scheduling, and linking them together to visualize the execution sequence that helps in the identification of space conflicts. However, construction space planning is not only related to the construction sequence visualization developed in CAD or BIM. For example, space planning for gravity dam construction where topography plays a major role cannot be done without geospatial

capabilities (available in GIS) which are missing in both, BIM and 4D CAD-based systems [Zhong et al. (2004)].

Keeping the importance of geospatial capabilities in view, contractors or organizations create, store, and share 3D modeling along with its surroundings [Bansal and Pal (2008)]. 3D model, topography, surroundings, 4D scheduling, and geospatial analyses capabilities together in a single platform help in the space planning much better way [Bansal (2011)]. In addition, modeling of the spatial relationships through GIS-topology is of great use in the spatial computing perspective because GIS-based topology has been matured in the last decade. However, recent efforts to represent topology in BIM still need further investigation [Borrmann et al. (2009)].

The use of 4D models in the GIS is found helpful in the space planning. The visualization of execution sequence in 4D along with its neighborhood supports space planning of a construction planning in hilly regions. A 3D model acts as an input in the development of a 4D model. However, the 3D modeling capabilities available in the GIS have not been developed like BIM or CAD-based systems [Bansal and Pal (2008)]. A few commercially available GIS tools offer 3D formats. In this respect, researchers have the challenging role to mature 3D GIS. The researchers have to show the GIS users the possibilities and constraints of 3D GIS in order to obtain a serious breakthrough of the 3D GIS. Therefore, at present, an alternative to the 3D modeling has been explored. The present study discusses how space-planning procedure in the 4D GIS has been designed for conflicts identification and computation.

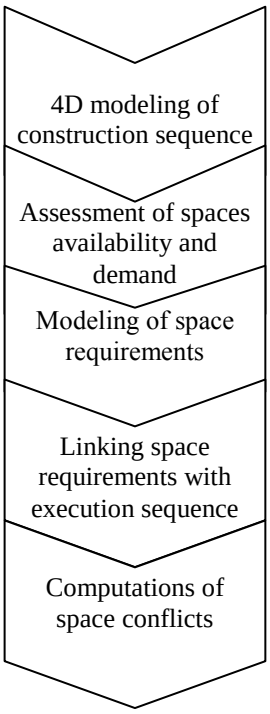


Figure 1: Process of space planning.

Purpose of GIS in Construction Planning

A construction either big or small cannot remain in isolation but is closely related to all other facilities in its surroundings. Even as a single entity, it creates a vast amount of information by its existence in in its surroundings. A construction cannot be planned as a single entity; careful consideration has to be given to the immediate neighborhood. Usually, this is done manually based upon previous experience. Software tools like building information modeling (BIM) and CAD mainly consider the inside geometry of a construction project, while, GIS is more concerned with the space outside a building. Therefore, any new construction using BIM and CAD systems can be planned in isolation only. GIS helps in efficient decision making with its capability to handle both spatial and attributes information which is queried, analyzed, and displayed together in various graphical and non-graphical forms. Spatial data describe features' geometry whereas; attributes stored in tabular form describe characteristics of different features. The 3D models of a construction project should be prepared along with topography to consider the surroundings. Layouts of existing utility services like: electric lines, gas supply, water distribution systems, sewerage network, etc. which play a major role in locating new facilities,

can easily be stored in GIS environment. Construction planning especially in hilly regions where topography plays a major role cannot be simulated without geospatial modeling and analysis capabilities which are available in GIS. The CAD and VR-based systems lack geospatial analytical capabilities such as evaluation of a location for flooding, drainage pattern, and route planning for vehicles carrying consignments from different access routes to construction site. Further, a planner also needs spatial information about the neighborhood of a facility to be developed to determine its dependence on project under consideration. Such dependence is not easily modeled in CAD and VR-based systems. The use of GIS allows a planner to view and analyze the effects of a new construction on existing facilities. GIS-based approach also helps in incorporating environmental aspects in the early phases of construction planning.

Process of Space Planning

In the space planning, to finalize a construction plan in terms of when, where, and how long a space is required on the jobsite, a link between workspace requirements and the execution schedule is found significant. 3D model along with its surroundings, a 4D sequence, and geospatial analysis capabilities into a single GIS platform helps in the space planning. The modeling of an area with spatial constraints using GIS-based topology contributes in the identification of space conflicts. Therefore, the main objective behind the present study was to explore the use of GIS in the space planning to identify spatial conflicts. The procedure for the identification of spatial conflicts was designed in which workspaces corresponding to various activities in the schedule were generated in the SketchUP. A link between workspaces and the 4D model of construction sequence was established in the ArcGIS 10. After the identification of space conflicts, their computation was done in ArcGIS 10 (Fig.1).

Identification of Space Conflicts

4D Modeling of Construction Sequence

Initially, the execution schedule of the project under construction was finalized. The modeling of building interior in 3D was done in the SketchUP. The terrain modeling around the building was done in ArcGIS 10 [Bansal (2014)]. The modeling of building interior depicts floor level detail whereas digital terrain model represents topographical condition of the jobsite. Linking of the project execution schedule with 3D components developed in the SketchUP [SketchUp (2010)] to make 4D construction sequence was done in ArcGIS 10 [Bansal and Pal (2008)]. The degree of detail in a 4D model depends upon the detail in the execution schedule. Hence, it is better to use full work breakdown structure. Detail in a schedule and division of a 3D model into small components have serious implication on the time required in the 4D modeling.

Assessment of Spaces Availability and Demand

Three categories of the available spaces were considered in the present study. These spaces includes: space provided by the jobsite on the ground, space provided by the temporary structures such as scaffolding or working platforms, and space provided by the structure to be constructed with time. The categories of the available spaces were characterized in terms of their sizes, locations, and time of availability. An activity requires working and path spaces for labors, equipment, and materials storage. Hence, various categories of space requirements for each activity were calculated. The spaces were positioned outward, inward, above, below, or around the reference component to be constructed. Site engineers describe each space requirement with respect to component to be constructed, component's location, size, and shape. The present study does not focus on the volumes and types of different spaces required,

for more details about this, readers are directed to the earlier studies [Akinci et al. (2002a; 2002b)].

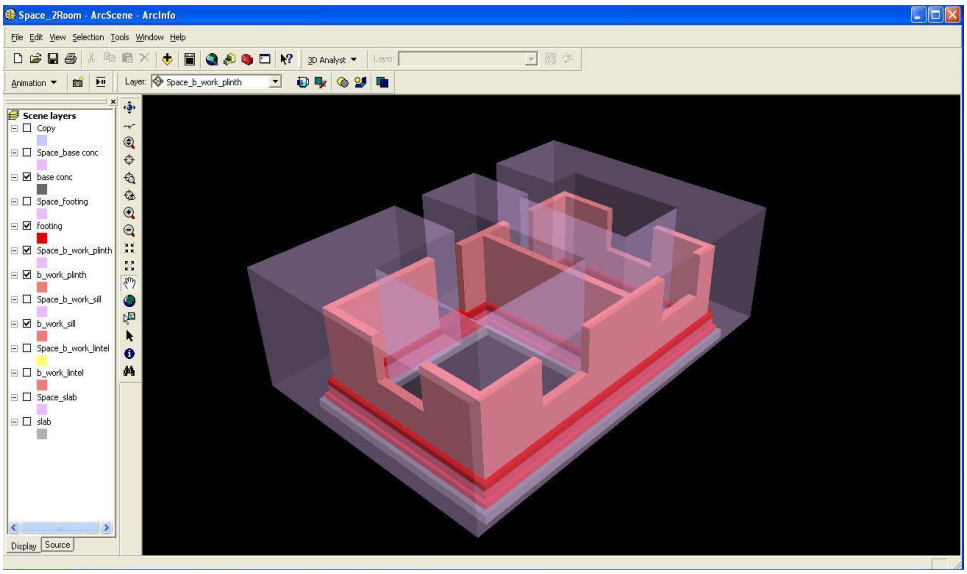


Figure 2: Work space requirement of the brick wall in the construction of a small

Modeling of Space Requirements

The position, size, shape, schedule, and reference component corresponding to each space decide its characteristics on the jobsite. SketchUP was used to model the spaces corresponding to each activities' space requirement estimated in the earlier step. Any shape of a space can be modeled in the SketchUP. The modeled spaces from the SketchUP were exported to ArcGIS 10 in the Multipatch format [ArcGIS (2014)], The Multipatch format supported in the ArcGIS 10 is used to represent spaces or components in 3D.

Linking Space Requirements with Execution Sequence

Project specific space requirements on a jobsite changes with time, therefore, the developed space requirements were linked with the execution sequence to generate dynamic space requirements. This link finds the start and finish times of each space corresponding to an activity defined in the execution schedule. 4D model of the execution sequence integrated with space requirements shows work space demand of various activities along with 3D components to be constructed along with its surrounding (Fig. 2). To finalize a plan, 4D model of the execution sequence integrated with space requirements was found helpful because the overlaps among various spaces were identified visually.

Computations of the space conflicts

The overlaps/conflicts between two spaces were identified visually with the help of integrated 4D model of the space requirements and execution sequence. The volume of an overlap/conflict between two spaces was computed in the ArcGIS 10 (Fig. 3). A closed space in the multipatch format is required for the analysis in ArcGIS 10 for finding an overlap; this is checked with *Is Closed 3D tool*. The *Enclose Multipatch tool* is used to eliminate gaps in multipatch features used to represent space requirements [ArcGIS (2014)]. Spaces in the SketchUP may be produce in extremely complex geometries. If one input is given, the *Intersection of features* in that multipatch dataset are computed, whereas if two were given,

the *Intersection of features* from both datasets are determined and intersections found in only one input get ignored.

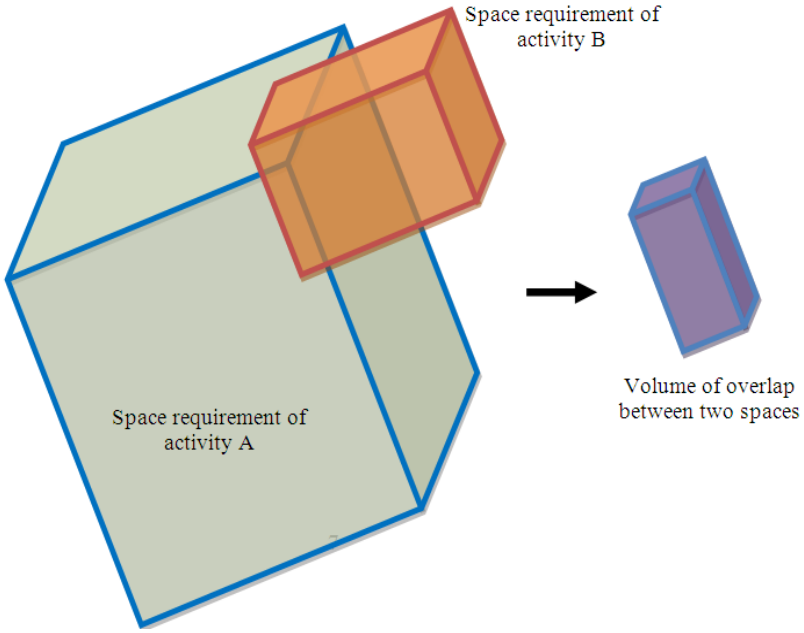


Figure 3: The volume of an overlap/conflict between two input spaces computed in the ArcGIS 10

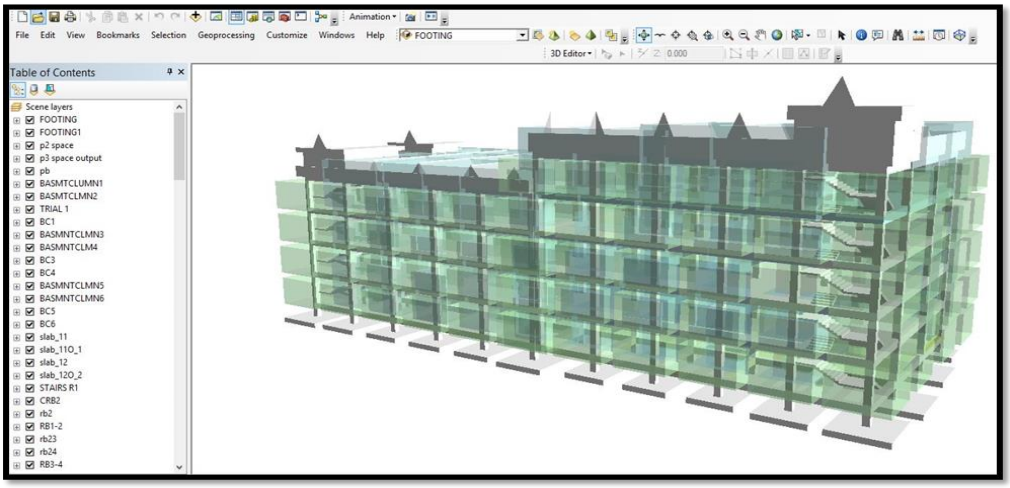


Figure 4: The case study building consists of two portions, left and right, left portion is of four floors and right is of five floors.

Case Study

The National Institute of Technology (NIT) Hamirpur, India is the premier technical institute of the region located in hilly terrain, covering an area of about 200 acres. The demand of buildings in the institute campus has been growing due to the increased academic and non-academic activities. The construction planning of a building is hilly region is highly influenced by site topography. Hence, construction of the building located in the hilly region of NIT campus was taken as the case study. The modeling of facilities/utilities around the case study building included institutional buildings, administrative block, health center, library, auditorium, and lecture hall complex. Other existing buildings modeled were food courts, water tanks, and stores. The existing public utility networks included were layouts of water distribution system consisting of main supply line and sub-mains, sewer network, road network, and overhead electric lines. Electric poles, telephone poles, and lamp posts were also modeled on their respective locations. The attributes corresponding to all existing facilities/utilities were kept in the relational database. For more detail about the modeling of surrounding readers are directed to the earlier study by [Bansal (2014)].

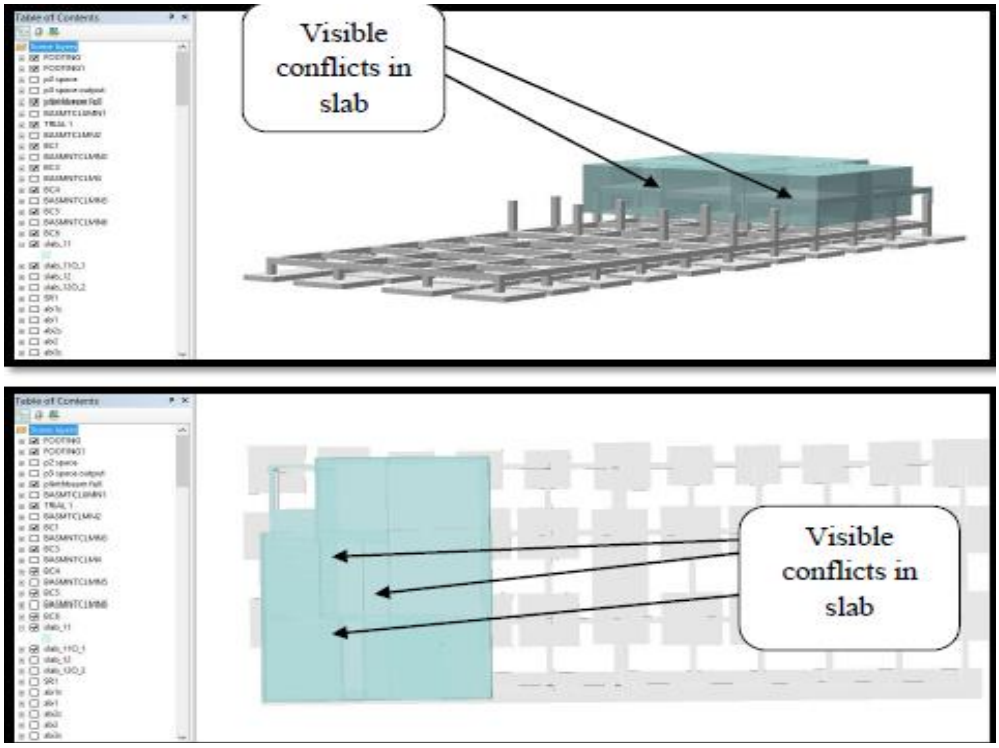


Figure 5: Identified space conflicts in the case study building through modeling in *ArcGIS.*

The case study building consists of two portions, left and right. The left portion is of four floors and right is of five floors as shown in figure 4. The construction plan of each portion was broadly divided into five parts. The construction of sub-structure was included in the first part of plan. It involved activities like: preparation of land, excavation, construction of foundation, and backfilling of foundation trenches. The second part of plan involved construction of reinforced cement concrete frame. The exterior walls, interior partition walls, and flooring were included in the third part of plan. The plastering, fixing of door and window

frames, and fixing of panels were included in the fourth part of plan. The electrical fitting, plumbing, and finishing works were included in the fifth part of plan. The identified space conflicts in the case study building through modeling in ArcGIS have been shown in figure 5.

Conclusions

Without considering the space requirements, execution schedule cannot be finalized. Displaying spaces required along with the corresponding components in the 4D helps in the detection of time-space conflicts and accordingly modification of the execution schedule to resolve conflicts before construction. In the space planning, if the execution schedule leads to space conflicts, it is changed until it becomes conflict free. This facilitates in the rapid generation of a conflict free schedule. Various graphical operations on spatial and non-graphical operations in GIS improve and speed up construction planning and space planning.

References

[1] Akinci, B., Fischer, M., Kunz, J. and Levitt, R. (2002a) Representing work spaces generically in construction method models, *Journal of Construction Engineering and Management* **128**(4), 296-305.

[2] Akinci, B., Fischer, M. and Kunz, J. (2002b) Automated generation of work spaces required by construction activities, *Journal of Construction Engineering and Management* **128**(4), 306-315.

[3] ArcGIS version 10 (2014) Environmental Systems and Research Institute. N.Y. Street, Redlands, California.

[4] Bansal, V. K. and Pal, M. (2008) Generating, evaluating, and visualizing construction schedule with Geographic Information Systems, *Journal of Computing in Civil. Engineering* **22**(4), 233-242.

[5] Bansal, V.K. (2011) Use of GIS and topology in the identification and resolution of space conflicts, *Journal of Computing in Civil. Engineering* **25**(2), 159-171.

[6] Bansal, V.K. (2014) Use of Geographic Information Systems in spatial planning: a case study of an institute campus, *Journal of Computing in Civil Engineering* **28**(4), 05014002-1-12.

[7] Borrmann A., Schraufstetter, S. and Rank, E. (2009) Implementing metric operators of a spatial query language for 3D building models: Octree and B-rep approaches, *Journal of Computing in Civil. Engineering* **23**(1), 34-46.

[8] Choi, B., Lee, H., Park, M., Cho, Y. and Kim, H. (2014), Framework for work-space planning using four-dimensional BIM in construction projects, *Journal of Construction Engineering and Management* **140**(9), 04014041.

[9] Guo, S. Y. (2002), Identification and resolution of work space conflicts in building construction, *Journal of Construction Engineering and Management* **128**(4), 287-295.

[10] SketchUp (2010) Google SketchUp Plugins, available at http://sketchup.google.com/intl/en/download/plugins.html, (accessed on May 15, 2010).

[11] Zhong, D., Li, J., Zhu, H. and Song, L. (2004) Geographic information system based visual simulation methodology and its application in concrete dam construction processes, *Journal of Construction Engineering and Management* **130**(5), 742-750.

Computation of vadoze zone moisture profiles for successive irrigation scheduling

Vijay Shankar

Department of Civil Engineering
National Institute of Technology, Hamirpur – 177005, Himachal Pradesh, India
Presenting author & corresponding author: vsdogra12@gmail.com

Abstract

A numerical simulation model has been developed, to compute vadoze zone soil moisture content profiles under transient field conditions by coupling soil moisture flow equation with a non linear root water uptake model. The model has been tested for the sensitivity of its non linear uptake parameter, for obtaining its optimal value. Computation takes into account a variable transpiration rate and a field measured initial moisture content. Rainfall, irrigation and evaporation have been treated as sources of non-uniform potential surface flux. Solutions to the computation have been obtained numerically by a fully implicit finite difference scheme, involving a non linear system of equations, which has been linearized using Picard's iterations. Field crop data of maize (Zea mays), which is among the most important crops in India and several other countries in the world, has been used to evaluate the results of the simulation. Determining the water requirements of crops is important for improved scheduling of irrigation, which in turn requires accurate measurement of crop evapotranspiration (ET_c). As the first objective, daily and seasonal ETc of maize are computed using Lysimeter set up in an experimental field from May 2006 to September 2006 at Roorkee, India. The average daily ETc of maize varied from a range of 1.4 to 3.4 mm day^{-1} in the early growing period to 8.3 mm day^{-1} at peak that occurred 9 weeks after sowing (WAS) at the silking stage of maize, when leaf area index (LAI) was 4.54. Average daily ETc declined sharply to 2.57 mm day^{-1} during late season stage of crop. The measured seasonal ET_c of maize was 495 mm. Development of computation based schedules of irrigation is the second objective of the study. Plant parameters like root depth and crop height have been continuously observed throughout the crop period. Top 0.3 m depth of root zone is considered to represent the soil moisture status governing the schedules of irrigation. Application of the computation technique to field conditions and comparison of the results with filed measured data shows very good agreement.

Introduction

The availability of water for plant roots is an important topic, which has been explored by a number of investigators (Feddes et al., 1978; Molz, 1981; Kang et al., 2001). Recently the attention is being given to irrigation management, by optimizing the frequency of irrigation, particularly in arid and semi-arid regions. Such management strongly depends upon knowledge of soil moisture movement through the root zone of the crops. Prediction of available moisture for plant roots also has significant effect on irrigation scheduling. The studies in this direction followed basically two approaches; microscopic, where a single root is assumed to be represented by a narrow infinitely long cylinder of constant radius which absorbs water (Afshar and Marino, 1978) and macroscopic, which focus on the removal of moisture from the differential volume of soil as a whole, without considering the effect of individual roots (Feddes et al., 1978). However, the basic assumptions along with the drawbacks and the difficulties involved in microscopic scale models under natural field conditions have restricted their applicability for field situations.

Soil moisture dynamics under cropped conditions are affected by soil, plant and climatic factors. The boundary between soil and the root system of plants is a major hydrologic interface across which well over 50% of evapotranspiration moves. Mathematical models of soil moisture dynamics on a macroscopic scale are mostly employed for predicting soil moisture distribution in the crop root zone on a day-to-day basis. Root water uptake in the crop root zone is represented as a sink term in the soil moisture flow equation. There are many different forms of sink term functions developed till date, of which, hypothetical linear distribution pattern of 40, 30, 20, 10 % moisture uptake in each quarter of root zone by Molz and Remson (1970), Feddes et al. (1978)'s constant rate model, Prasad (1988)'s linear rate model and Ojha and Rai (1996)'s non linear root water uptake model are the prominent ones. Precise estimation of soil moisture depletion in the crop root zone, accurately determines the soil moisture availability for the plant use. It has been established by many recent studies that plant moisture uptake involves considerable non-linearity owing to the non-linear root density distribution in the root zone (Ojha and Rai, 1996; Kang et al., 2001).

Present work couples Ojha and Rai (1996) non-linear root water uptake model, with Richards (1931) equation. A numerical simulation model is developed to compute the soil moisture dynamics in the crop root zone. Requisite soil and crop data is obtained by conducting the field crop experiments. Maize, which is a major crop in this region, has been grown during relevant crop season. Variation of crop evapotranspiration during the crop season has been determined. The first objective of the work is to accurately predict the soil moisture profiles in crop root zone. Based on the simulated soil moisture depletion in root zone, study also aims to compute optimal irrigation schedules for the crop grown in the field at different allowable moisture depletion levels.

Materials and Methods

Water Movement in Soil

The mixed form of Richards's equation governing water flow in the unsaturated zone, considering root water uptake can be written as

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\psi)\left(\frac{\partial \psi}{\partial z} + 1 \right) \right] - S(z,t) \qquad (1)$$

Where, θ is the volumetric moisture content of soil, ψ is the pressure head, t is the time, z is the vertical coordinate taken positive upwards, K is hydraulic conductivity, and S(z, t) is the water uptake by roots expressed as volume of water per unit volume of soil per unit time. Richards's equation is highly non linear due to changes in pressure head and hydraulic conductivity in unsaturated soils. In order to solve Richards's equation, it is required to specify constitutive relationships between the dependent variable (moisture content in this case) and the non linear terms (pressure head and hydraulic conductivity). Present study uses K-θ-ψ relationships proposed by Van Genuchten's (1980), given as:

$$\Theta = \left[\frac{1}{1 + \left\| \alpha \ \psi \right\|^{n}} \right]^{m} \quad \text{For } \psi \leq 0 \qquad (2)$$

$$= 1 \qquad \text{for } \psi > 0$$

In equation (2), α and n are unsaturated soil parameters with m = 1-(1/n) and Θ is effective saturation defined as

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{3}$$

Where, θ_s is saturated moisture content and θ_r is residual moisture content.

Based on Mualem's (1976) model the relation between moisture content and hydraulic conductivity is given by (Van Genuchten, 1980)

$$K = K_{sat}\Theta^{1/2}\left[1-\left(1-\Theta^{1/m}\right)^{m}\right]^{2} \tag{4}$$

Where K_{sat} = saturated hydraulic conductivity of soil

Root Water Uptake

Ojha and Rai (1996), non-linear root water uptake model, referred as O-R model hereafter, has been used to represent the sink term in Eqn (1). According to O-R model, for potential transpiration conditions, the potential rate of soil moisture extraction S_{max} is given by the relation

$$S_{max} = \left[\frac{T_j}{z_{rj}}(\beta+1)\left(1-\frac{z}{z_{rj}}\right)^{\beta}\right]; \qquad 0 \le z \le z_{rj} \tag{5}$$

Where, β is model parameter, z is depth below soil surface, and z_{rj} is root depth on the j^{th} day. For $z = z_{rj}$, S_{max} is zero as per (5) and at $z = 0$, S_{max} attains a maximum value. Thus (5) satisfies the desired extraction conditions, that extraction is maximum at the top and zero at the bottom of the root. It is to be noted that for $\beta = 0$, (9) reduces to a constant rate extraction model of Feddes et al. (1978) with $S_{max} = T_j/z_{rj}$ while for $\beta = 1$, (9) reduces to linear extraction model of Prasad (1988) with $S_{max} = 2T_j/z_{rj} - 2T_j\,(z/z_{rj}^2)$. Present work considers the moisture uptake under potential moisture condition.

Initial and Boundary Conditions

Measured pressure head values in the soil profile at the start of crop season have been used as the initial condition, i.e.

$$\psi = \psi_0(z, 0) \qquad 0 \le z \le L \tag{6}$$

Where ψ_0 is the measure pressure head value at corresponding soil depth. For intermediate depths values are linearly interpolated.

The upper boundary condition is a prescribed flux boundary condition accounting for the evaporation taking place from the top soil and a Drichlet boundary condition, during irrigation or rainfall. Thus

$$\psi(L, t) = \psi_s \qquad \text{during irrigation/rainfall} \tag{7a}$$

$$-K(\psi)\left(\frac{\partial\psi}{\partial z}+1\right)=E \qquad z = L, \text{ in absence of irrigation/rainfall} \qquad (7b)$$

Where ψ_s is the pressure head corresponding to the saturated soil moisture condition. E is the evaporation from the top soil.

At lower boundary gravity drainage type condition has been assumed, where a unit hydraulic gradient is considered.

$$-K(\psi)\left(\frac{\partial\psi}{\partial z}+1\right)=-K(\psi) \qquad \text{for } t \geq 0,\ z = 0 \qquad (8)$$

Numerical Model

A numerical model has been developed to solve equation (1) along with the sink term subjected to initial and boundary conditions (6) to (8), and employing the constitutive relationships (2) to (4). The numerical model is based on a mass conservative, fully implicit finite difference scheme proposed by Celia et al. (1990). The non linear system of equations is linearized using Picard's methods (Paniconi et al., 1991) and resulting system of equations are solved using Thomas algorithm. The model yields spatial distribution of pressure head and moisture content at successive advancing times in the soil. From the model computed moisture contents, the moisture depletion values at different zones of crop root at different times are computed by numerical integration.

Field Crop Experiments

Field crop experiments have been conducted at the field experimental station of Civil Engineering Department, Indian Institute of Technology, Roorkee, India, from April to September, 2006. The average annual rainfall at Roorkee is 1032 mm, of which about 75 % is usually received between July and September. The required meteorological data for the computation of corresponding crop evapotranspiration using crop coefficient approach is obtained from the Department of Hydrology, Indian Institute of Technology Roorkee. For measuring the soil moisture profile throughout the crop season soil moisture measurement sensors have been embedded at 0.15, 0.30, 0.45, 0.60, 0.75, 0.90, 1.05 and 1.20 m, however at the ground surface the moisture content is measured using TDR soil moisture meter.

Crop details

Maize (Variety K-99 HYBRID) was sown uniformly in Lysimeters and the surrounding field so that the field conditions could be simulated in and around the Lysimeters. Crop period of Maize lasted from May 20th to September 1st, 2006 (105 days). The sampling site for different plant parameters such as leaf area index (LAI) and root length is about 4 to 5 m away from the Lysimeter. The entire crop growth period for the crops is divided into four stages; I-Initial, II-Crop Development, III-Mid Season and IV-Late Season. Growth stages have been considered on the basis of study by Doorenbos and Pruit (1977). Initial stage corresponds to the germination and early growth when the soil surface is not or is hardly covered by the crop (ground cover < 10 %). Crop development stage starts from the end of initial stage to attainment of effective full ground cover (ground cover: 70-80 %). Mid season commences from the attainment of effective full ground cover to time of start of maturing as indicated by

discoloring of leaves or leaves falling off and late season stage begins from end of mid-season until full maturity or harvest. Duration of stage I, II, III and IV accordingly has been found to be 17, 30, 34 and 24 days respectively. Irrigations have been provided on 24th, 33rd and 42nd day of the crop period.

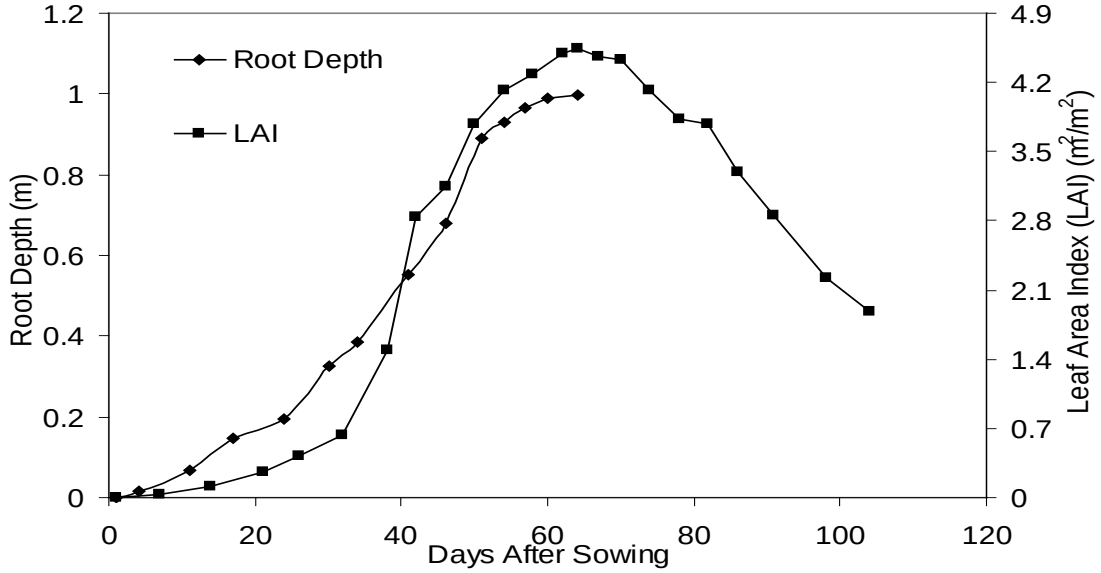


Figure 1. Field observed plant parameters for the Maize

Two major parameters; LAI, and root depth have been recorded at discrete time intervals throughout the growth period. Leaf area index (LAI) required for the partitioning of the crop evapotranspiration into plant transpiration and soil evaporation, was measured by direct method suggested by Jesus et al., (2001). Leaf area measurements are made once in a week during the initial stage, once in five days during development stage, twice a week during middle stage and once a week during last stage. Root depth has been measured by trench profile method described by Wolfgang (1979). At initial stages of crop growth root depth has been measured at 7-10 days interval, where as in later stages this interval has been reduced to 5 day interval. Figure 1 show the variation of root depth and LAI measurements with crop growth period for maize.

Soil parameters

Representative soil samples were obtained from the 0-0.3 m, 0.3-0.6 m, 0.6-0.8 m, 0.8-1.0 m and 1.0-1.2 m depths, in the experimental site for testing the soil properties. The cumulative particle size curves obtained through grain size and hydrometer analysis reveal that the soil profile up to 1.2 m is fairly uniform in texture. The upper 0-0.3 m depth however, shows a slight deviation from the general trend with higher silt and lower clay fractions being indicated, but it is within limits and hence a uniform soil textural classification is considered for 0-1.2 m depth. USDA soil textural class for the experimental field soil is sandy loam. The bulk density, particle density and porosity for the field soil are 1.62 g/cm^3, 2.61 g/cm^3 and 0.38 respectively.

Soil-moisture characteristic curve provides a convenient method for describing the moisture retention properties of different soils (Winter 1974). In-situ determination of SMC has been performed, which involves simultaneous measurement of soil matric potential (ψ) and

moisture content (θ) at 0.3, 0.6, 0.9 and 1.2 m depths below the ground level. No clear depth-wise relationship is discernible, indicating the similarity of the retention characteristics of the soil profile and as such a single SMC has been used for the entire zone. Van Genuchten Relationship (1980) described by Eqns (2)-(4) has been used to determine the soil hydraulic characteristics.

The saturated moisture content θ_s in eqn. (3) is assumed to be equal to the measured soil porosity (0.38 cm^3 cm^{-3}). A standard residual moisture content value equal to 0.065 cm^3 cm^{-3} (Carsel and Parrish, 1988) for sandy loam soil (soil type for experimental plot) has been considered. A non linear optimization algorithm E04FDF (N.A.G., 1990) has been used to estimate the Van Genuchten parameters α and n, which are 6.2 m^{-1} and 1.68 respectively. The value of average field saturated hydraulic conductivity (K_{sat}) determined at different depths using Guelph type Permeameter is 3.9 cm/hour. Experimentally obtained value of field capacity (θ_{fc} = 0.208) and SMC deduced value of wilting point (θ_{pwp} = 0.068) has been used in the present study. The available moisture which is the difference of θ_{fc} and θ_{pwp} is 0.14. The irrigation has been provided at 50% depletion of the available moisture in the effective root zone.

Computation of Crop Evapotranspiration (ET_c)

Crop evapotranspiration has been determined as the product of daily crop coefficient and reference evapotranspiration. Reference evapotranspiration (ET_0) is a complex phenomenon and depends on several climatological factors, such as temperature, humidity, wind speed, radiation, and, type and growth stage of crop. During the study period ET_0 (mm/day), has been computed by Penman Monteith method. The Penman-Monteith equation for the ET_0 is given as (Allen et al., 1998)

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (9)$$

Where, R_n = net radiation at the crop surface [MJ m^{-2} day^{-1}], G = soil heat flux density [MJ m^{-2} day^{-1}], T = mean daily air temperature at 2 m height [°C], u_2 = wind speed at 2 m height [m s^{-1}], e_s = saturation vapour pressure [kPa], e_a = actual vapour pressure [kPa], (e_s - e_a) = saturation vapour pressure deficit [kPa], Δ = slope vapour pressure curve [kPa $°C^{-1}$], γ = psychrometric constant [kPa $°C^{-1}$].

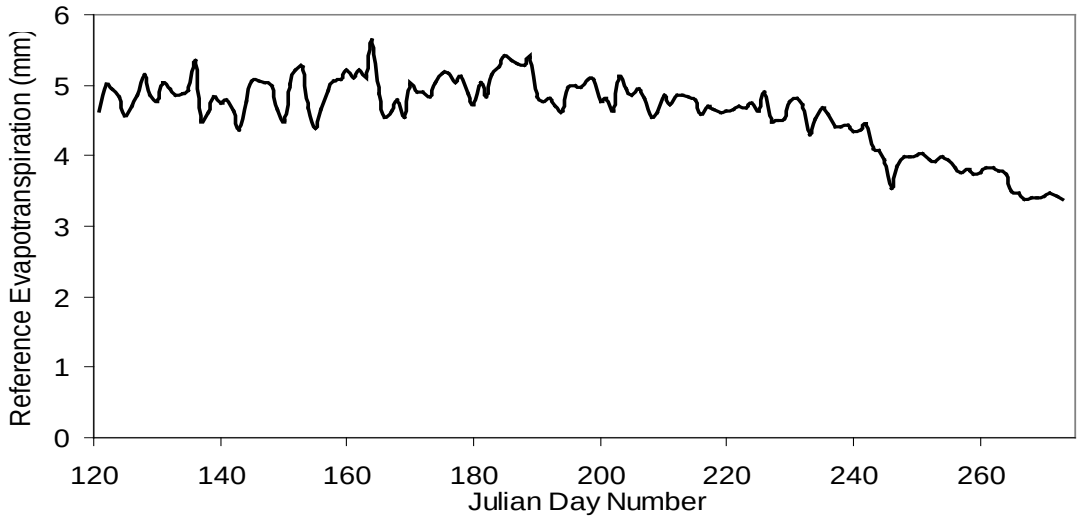


Figure 2. Daily reference evapotranspiration during study period

Different parameters involved have been computed using the mathematical formulations provided by Allen et al. (1998). Fig. 2 shows the daily ET_0 (mm/day) computed using Penman-Monteith method for the study period.

The crop coefficient (K_c) value represents crop-specific water use and is needed for accurate estimation of irrigation requirements of different crops. Comprehensive list of stage-wise crop coefficients is available in literature (Allen et al, 1998). The crop coefficients for initial, development, mid-season and end-season stages are denoted as $K_{c\ ini}$, $K_{c\ dev}$, $K_{c\ mid}$ and $K_{c\ end}$ respectively. In case the local calibration of the crop coefficients is not possible then a procedure has been outlined by Allen et al. (1998), to modify the reported crop coefficients for the local climatic conditions, and crop and irrigation practices. FAO proposed $K_{c\ ini}$, $K_{c\ mid}$ and $K_{c\ end}$ values are 0.3, 1.2 and 0.6 for Maize. These values have been modified for the local climatic, crop and soil characteristics according to the procedure outlined in FAO guidelines. The modified values of $K_{c\ ini}$, $K_{c\ mid}$ and $K_{c\ end}$ are 0.33, 1.126 and 0.55 respectively.

From the stage wise crop coefficients, daily K_c values during the growing period are determined either graphically or numerically (Allen et al., 1998). The daily crop coefficient depends on the plant characteristics as well as the meteorological factors, which are represented in the stage specific crop coefficients. Allen et al. (1998) had observed that K_c values remain constant for early and mid season stages. However, during the crop development and late season stage, K_c varies linearly between the K_c at the end of the previous stage ($K_{c\ prev}$) and the K_c at the beginning of the next stage ($K_{c\ next}$), which is $K_{c\ end}$ in the case of the late season stage. Following Allen et al. (1998), the crop coefficient for an i^{th} day in a particular stage is computed as:

$$K_{ci} = K_{c,prev} + \left[\frac{i - \sum\left(L_{prev}\right)}{L_{stage}}\right]\left(K_{c,next} - K_{c,prev}\right) \qquad (10)$$

Where, i is the day number within the growing season, $K_{c\ i}$ crop coefficient on day i, L_{stage} is length of the stage under consideration [days], and L_{prev} is the sum of the lengths of all previous stages [days]. Using equation (10) daily crop coefficients for Maize are determined.

Daily crop evapotranspiration is determined as the product of daily K_c value and reference evapotranspiration. Further, the daily crop evapotranspiration is partitioned into plant transpiration and soil evaporation using eqn. (11) method proposed by Belmans et al. (1983), where soil evaporation (E_s) is calculated as a fraction of the ET_c using the LAI of the soil surface.

$$E_s = f * EXP(- c * LAI)\, ET_c \qquad (11)$$

Where, f and c are regression coefficients, with f = 1.0, and c = 0.6. This relation gives an acceptable estimation of soil evaporation (Belmans et al., 1983). Plant transpiration is part of the ET_c, and it can be calculated after E_s is determined from Eqn. (12). Since $ET_c = E_s + T_p$, plant transpiration (T_p) is

$$T_p = ET_c - E_s \qquad (12)$$

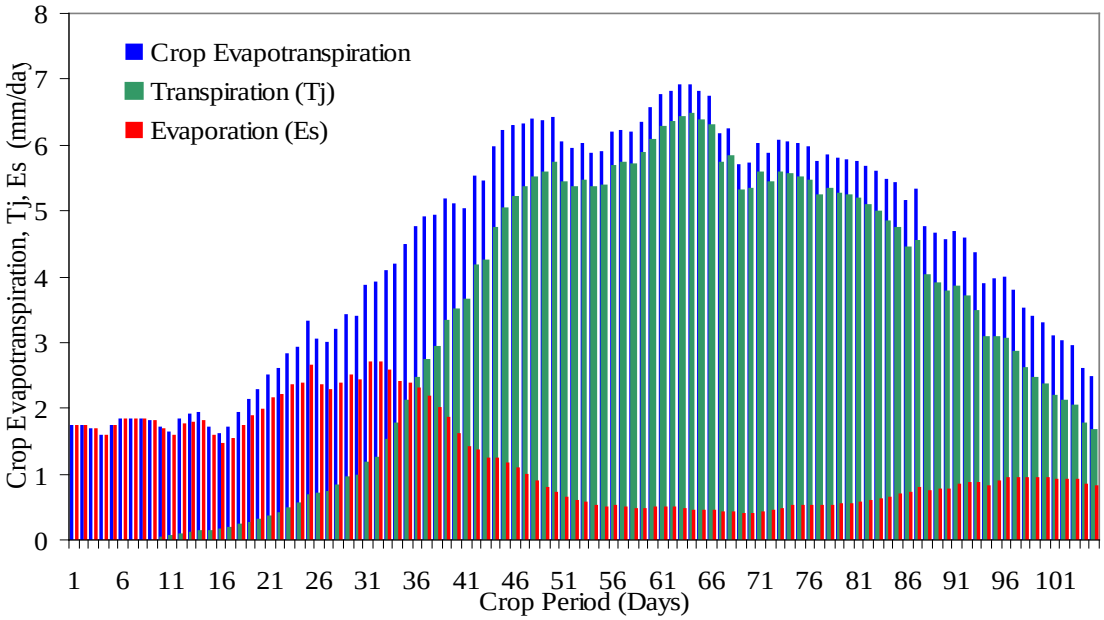


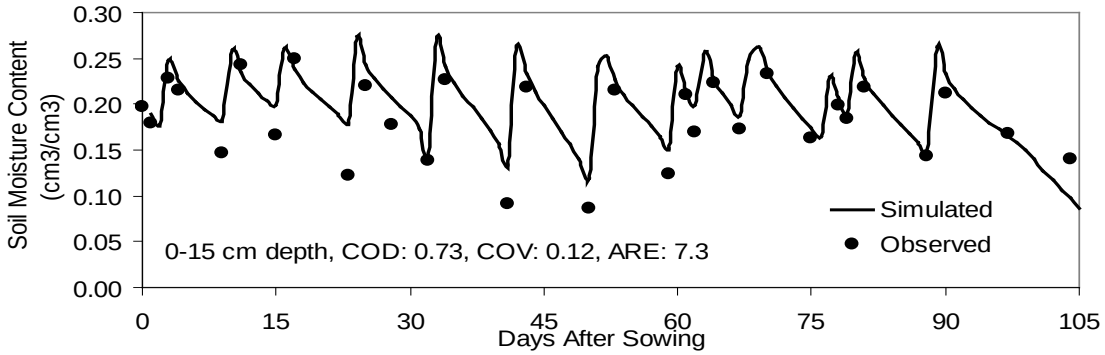
Figure 3. Daily Crop Evapotranspiration, Evaporation and Transpiration for Maize.

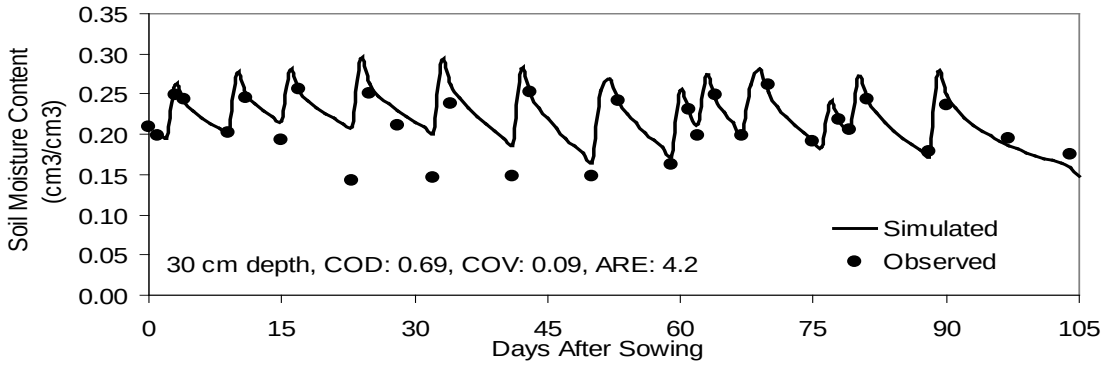
The plant transpiration is used as the sink term in the Richards equation and the soil evaporation is used as the boundary condition at the ground surface. Fig. 3 shows the variation of crop evapotranspiration and its components, evaporation and transpiration for Maize throughout the crop period. The average daily crop evapotranspiration of Maize varied from a range of 1.4 to 3.4 mm day^{-1} in the early growing period to 7.2 mm day^{-1} at peak that occurred 9 weeks after sowing (WAS) at the silking stage of maize, when leaf area index (LAI) was 4.54. Average daily ETc declined sharply to 2.57 mm day^{-1} during late season stage of crop.

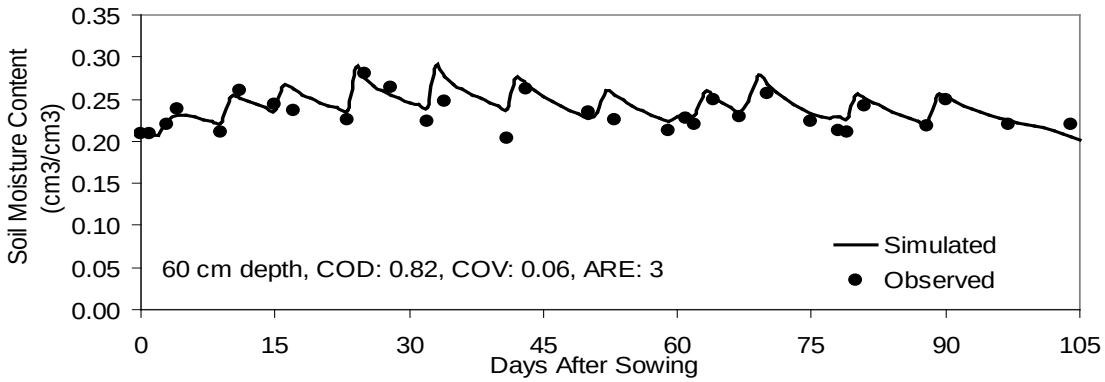
Results and Discussion

The obtained soil moisture characteristics, crop evapotranspiration and root depth variation over the crop period applied to the numerical model formulated by coupling Richards equation with O-R model to simulate plant moisture uptake. Initially the optimal value of the

non-linearity parameter β of O-R model is determined using observed and simulated soil moisture depletion pattern. The optimal value of β for Maize has been found to be 1.5. Observed and simulated soil moisture profiles in the vadoze zone on discrete days and soil moisture status during the crop period of Maize has been compared.







Figures 4, 5 and 6. Moisture status during crop period at 0-15, 30 and 60 cm depths in root zone

Figs 4, 5 and 6, show the observed and simulated soil moisture status during crop period, and Figs 7 and 8, show the observed and simulated soil moisture profiles on discrete days in crop period of Maize.

It can be observed from the Figs 4-8, that there exists a reliable agreement between simulated and observed values. However, for quantitative evaluation, error statistics e.g. coefficient of determination (COD), coefficient of variation (COV) and average relative error (ARE) (Ambrose and Roesch, 1982) are used for each set of values.

$$COD = 1 - \frac{\sum_{i=1}^{n}(\theta_{mi} - \theta_{si})^2}{\sum_{i=1}^{n}(\theta_{mi} - \theta_{avg})^2} \quad (13)$$

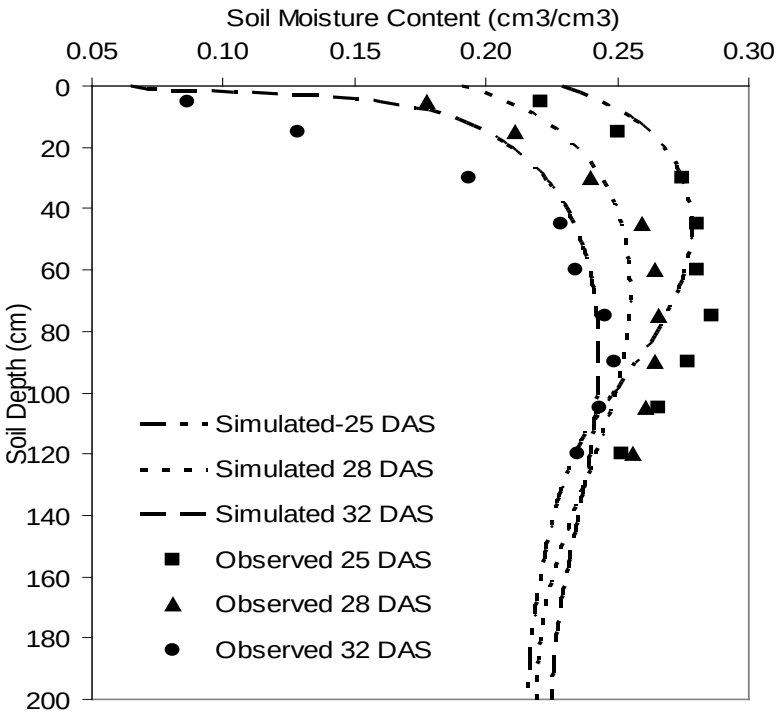
$$COV = \frac{\left[\sum_{i=1}^{n}\frac{(\theta_{si} - \theta_{mi})^2}{n}\right]^{0.5}}{|\theta_m|} \quad (14)$$

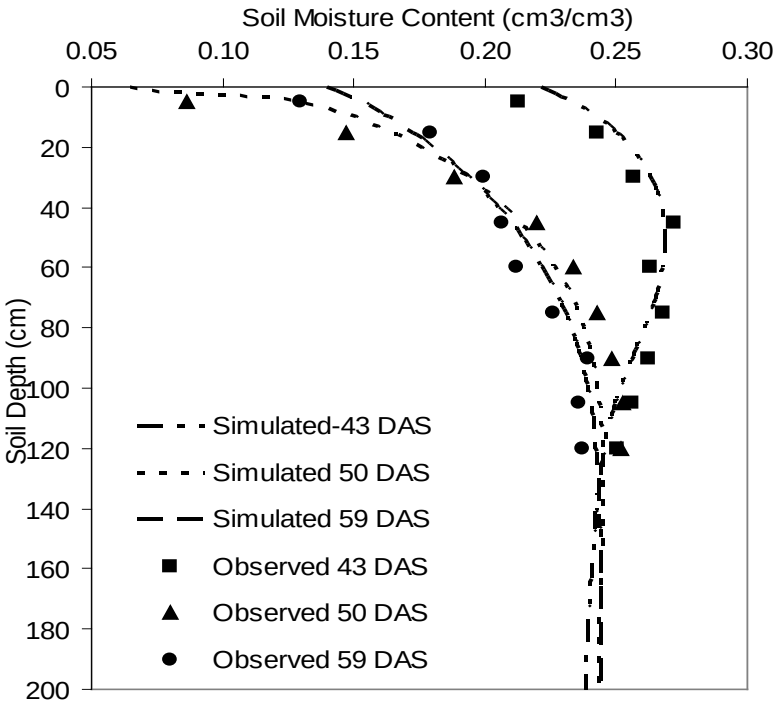
$$ARE\ (\%) = \frac{\sum_{i=1}^{n}(\theta_{si} - \theta_{mi})}{n|\theta_m|} \times 100 \quad (15)$$

Where, θ_{si} is the simulated sil moisture content at i^{th} point, θ_{mi} is the corresponding field observed value, θ_m is the average of the field measured values, and n is the number of observations. A value of COD close to the unity indicates a high degree of association between

The observed and simulated values, The COV quantifies the amount of "random scatter of the simulated and measured values about 1:1 line and ARE quantify the extent to which model simulations overestimate (positive ARE) or underestimate (negative ARE) the measured values. Corresponding values of error statistics for observed and simulated soil moisture at different depths are shown in the Figs 5-7. In case of observed and simulated soil moisture profiles the COD, COV and ARE values range between 0.74-0.92, 0.08-0.32 and -5.4-9.6 respectively. The values of error statistics fall in satisfactory-high agreement range.

It can be postulated from the above discussion that numerical model involving O-R model coupled with soil moisture flow equation, when applied to precisely determined soil parameters, crop data and crop evapotranspiration accurately simulates the soil moisture dynamics in the crop root zone. This provides the exact soil moisture availability for the plant moisture uptake in the crop root zone. Generally the irrigation is practiced when the average moisture content with in the root zone depth attains certain value between the field capacity and permanent wilting point (Prasad, 1988). This value of moisture content is called the allowable depletion level.

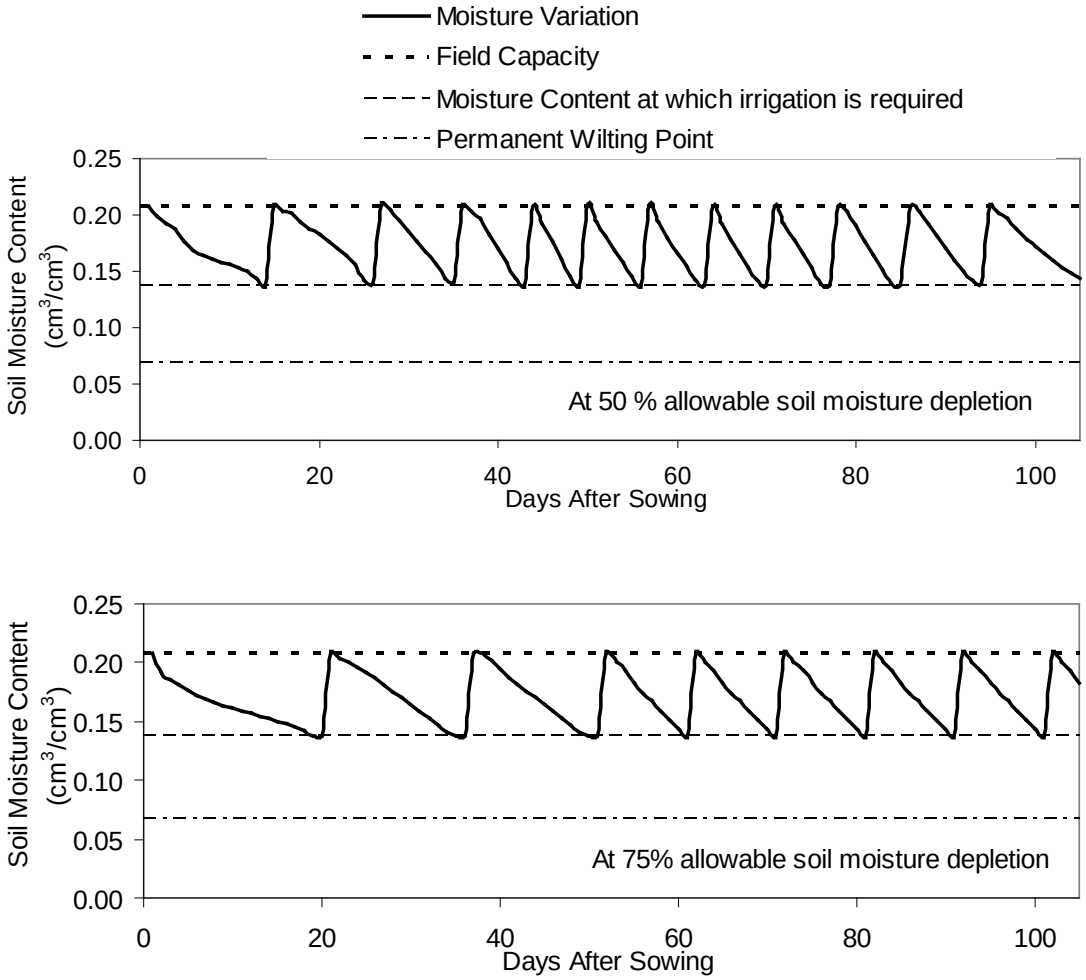




Figures 7 and 8. Vadoze zone soil moisture profiles on discrete days in the crop period

For different depletion levels required scheduling of irrigation is carried out. For optimal scheduling, adequate scheduling criterion is an important parameter in determining the frequency of irrigation events. The two parameters which contribute to assigning an adequate scheduling criterion are; allowable moisture depletion level and root depth considered for accounting the average soil moisture level. The hypothetical condition of no-rainfall is considered during the crop period of Maize. Though, allowable moisture depletion level is

dependent on the type of crop and the moisture retention capacity of the soil, 50 % and 75 % moisture depletion levels are considered in the present study. The effective root depth considered for accounting the average soil moisture status is 0.3 m. The optimal irrigation schedule at 50 and 75 % allowable moisture depletion level are given in Fig. 9 and 10.



Figures 9 and 10. Irrigation schedule for Maize at different allowable moisture depletion levels.

Summary and Conclusions

A numerical model has been formulated to compute the soil moisture content profiles under transient field conditions. A non-linear root water uptake model has been used as sink term to represent plant moisture uptake. Numerical model takes into account a variable transpiration rate and non-uniform initial soil moisture content. Rainfall, irrigation and evaporation are treated as sources of non-uniform potential surface flux. Plant control on water uptake when soil moisture is a limiting factor is not considered. The input parameters have been precisely determined using the field crop experiments.

Non-linear root water uptake model involving the optimal non-linearity coefficient has been found to represent the actual plant moisture uptake dependably. Application of the numerical model to field conditions and comparison of the results with field measured data showed good agreement. Precisely determined crop evapotranspiration is the dominant factor in predicting soil moisture dynamics. The practical significance of the study lies in the computation of

optimal irrigation schedules for field condition using the numerical model coupled with adequate scheduling criterion. Accurately computed soil moisture profiles result in generating optimal frequency of the irrigation and hence, results in irrigation water saving.

References

[1] Afshar, A., and Marino, M.A., (1978). Model for simulating soil water content considering evaporation. J. Hydrology, 37:309-322.

[2] Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., (1998). Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. FAO-Rome.

[3] Ambrose, Jr., R.B. and Roesch, S.E. 1982. “Dynamic estuary model performance”. J. Environ. Eng. Div. ASCE, 108, 51-57.

[4] Belmans, C., Wesseling, J.G. and Feddes, R.A. (1983). Simulation model of the water balance of a cropped soil: SWATRE. J. Hydrology, 63, 271-286.

[5] Carsel, R.F., and Parrish, R.S., (1988). Developing joint probability distributions of soil water retention characteristics. Water Resources Research, 24, 755-759.

[6] Celia, M.A., Bouloutas, E.T., and Zarba, R.L., (1990). A general mass conservative numerical solution for the unsaturated flow equation. Water Resources Research, 26:1483-1496.

[7] Doorenbos, J. and Pruitt, W.O., (1977). Guidelines for predicting crop water requirements. Irrigation and Drain. Div, FAO-Rome, Paper No. 24.

[8] Feddes, R.A., Kotwalik, P.J., and Zaradny, H., (1978). Simulation of field water use and crop yield. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

[9] Jesus, Waldir Cintra de Jr., Francisco Xavier Ribeiro do Vale, Reginaldo Resende Coelho, and Luiz Clau´ dio Costa, (2001). Comparison of two methods for estimating leaf area index on common bean. Agronomy J. 93:989–991.

[10] Kang, S., Zhang, F., and Zhang, J., (2001). A simulation model of water dynamics in winter wheat field and its application in a semiarid region. Agricultural Water Management, 49,115-129.

[11] Molz, F. J., (1981). Models of water transport in soil plant system: A review. Water Resources Research, 17, 1245-1260.

[12] Molz, F.J., and Remson, I., (1970). Extraction term models of soil moisture use by transpiring plants. Water Resources Research, 6(5), 1346-1356.

[13] Mualem, Y., (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research, 12 (3), 513-522.

[14] N.A.G., (1990). Routine name; E04FDF. The Numerical Algorithms Group FORTRAN Library Manual, Mark 14.

[15] Ojha, C.S.P., Rai, A.K., (1996). Non linear root water uptake model. Journal of Irrigation and Drainage Engineering, 122, 198-202.

[16] Paniconi, C., Aldama, A.A., and Wood, E.F., (1991). Numerical evaluation of iterative and numerical methods for the solution of the non-linear Richards equation. Water Resources Research, 27, 1147-1163.

[17] Prasad, R., (1988). A linear root water uptake model. J. of Hydrology, 99, 297-306.

[18] Richards, L.A., (1931). Capillary conduction of liquids through porous medium. Physics, 1:318-333.

[19] Van Genuchten, M.T., (1980). A closed form equation for predicting the hydraulic conductivity of unsaturated soil. Soil Science Society of America Journal, 44:892-898.

[20] Winter, E.J., (1974). Water, Soil and the Plant. The Macmillan Press Ltd. London, pp. 141.

[21] Wolfgang, B., (1979). Ecological studies series, 33, Methods of studying root systems. Eds: Billings, W.D., Goiley, F., Lange, G.L., Olson, J.S. and Ridge, O., pp188.

Non-intrusive POD-based Simulation for Heat Diffusion Systems

G.H. Zhang, *M.Y. Xiao, Y.F. Nie

Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an Shaanxi, P.R. China

*Presenting/Corresponding author: manyuxiao@nwpu.edu.cn

Abstract

Reduced order model constitutes an efficient option to decrease the high computational cost of dynamical systems governed by partial differential equations (PDE). The technique based on proper orthogonal decomposition (POD) was first presented in the article [1] to generate a reduced set of basis functions for Galerkin representation of PDEs which results in approximate the simulation at any time point by solving an ODEs of time dependent coefficients. Our approach in this article targets the development of a non-intrusive reduction technique. We keep the same manner of obtaining basis functions, while approximating the time dependent coefficients using Kriging based surrogate model. The proposed method is then illustrated with an application to the simulations of heat diffusion systems on a thin rod and on a square plate. The numerical results illustrate the simulation using the proposed idea.

Keywords: Proper Orthogonal Decomposition, Kriging surrogate model, heat diffusion system.

1 Introduction

Most of engineering problems may be presented as systems governed by partial differential equations. With the development of science, more rigorous device requirements arise to capture the characteristics of more complex systems, which are common for example in semiconductor manufacturing. The purpose, however, is not to provide an introduction to the complexity of such systems, Instead, we wish to propose a general methodology for implementation of one or two techniques based on surrogate models and apply them to a linear system of heat diffuse equation.

A widely used approach is performing a set of computer experiments 'a priori'. The data sampling is then used for construction of meta-models linking design variables with responses. The literature shows that a wide range of approximation methods that has been used for this purpose, such as polynomial response surfaces [3], least squares approximation [4], Kriging [5], radial basis functions [6] etc. In particular, surrogate model, developed by Krige [7] and then improved by Matheron [8], is emphasized here, as it is an exact interpolation method and a form of generalized linear regression for the formulation of an optimal estimator in a minimum mean square error sense. Due to the superiority of Kriging, it is widely used in structural reliability [9] and in optimization analysis [10].

Another class of among so-called physical based models, the popular one is Proper Orthogonal Decomposition (POD) also known as Karhunen-Loeve expansions in signal analysis and pattern recognition [11], or the Principal Component Analysis in statistics [12], or the method of empirical orthogonal functions in geophysical fluid dynamics [13,14]. Detailed description of the POD can be found in [15]. POD provides a useful tool for efficiently approximating a large amount of data. Lumley [16] first used POD to study turbulent flows. In 1987, Sirovich [17] incorporated the method of 'snapshots' into the POD framework and made important progress in this field. Other applications of POD are given in [18-20].

In this paper, a technique combining the advantages of Kriging surrogate model and POD model is proposed to represent heat diffusion on a one- or two- dimensional spatial domain. Suppose a given set of data sampling, discretization of PDE is approximately executed with the Galerkin method.

Then, we construct a basis of the finite dimensional function space of interest. In [2], the time dependent coefficients are obtained by solving an ODE. Here we propose a "non-intrusive" technique. Based on the original discrete data information, the approximated representation is built with Kriging surrogate model for the POD coefficients. It is finally applied to obtain the temperature field for any untried time point.

The paper is organized as follows: In section 2 we present the simulation of heat diffusion on a thin rod (one-dimensional spatial domain) and on a square plate (two-dimensional spatial domain) using infinite series expansion and finite difference scheme. Then we review the Galerkin projection and the POD model in the section 3. In section 4, a new method combining POD and Kriging surrogate model is described, and illustrates feasibility and efficiency of the proposed technique, followed by the numerical results in Section 5. The paper ends with conclusions and prospects.

2 Description of Heat Diffusion Equation

We consider an initial boundary value problem (IBVP) of heat equation. The methodology of "high fidelity" simulation is then explained to get the sampling data.
Case1: the one-dimension (1D) simulation of heat diffusion equation:

$$\begin{array}{lll} \text{PDE} & \boldsymbol{u}_t = \boldsymbol{u}_{xx} & x \in (0,1); \quad t > 0 \\ \text{BCs} & \boldsymbol{u}(0,t) = 0 = \boldsymbol{u}(1,t) & t > 0 \\ \text{IC} & \boldsymbol{u}(x,0) = 1 & x \in (0,1) \end{array} \tag{1}$$

where $\boldsymbol{u}(x,t)$ represents the temperature field on a thin rod.
Similarly, the case 2 is given by the following IBVP with two-dimension (2D) simulations:

$$\begin{array}{lll} \text{PDE} & \boldsymbol{u}_t = \boldsymbol{u}_{xx} + \boldsymbol{u}_{yy} & x \in (0,1); \quad y \in (0,1); \quad t > 0 \\ \text{BCs} & \boldsymbol{u}(0,y,t) = 0 = \boldsymbol{u}(1,y,t) & t > 0 \\ & \boldsymbol{u}(x,0,t) = 0 = \boldsymbol{u}(x,1,t) & t > 0 \\ \text{IC} & \boldsymbol{u}(x,y,0) = 1 & x \in (0,1); \quad y \in (0,1) \end{array} \tag{2}$$

where $\boldsymbol{u}(x,y,t)$ represents the temperature field on a flat plate.

2.1 Methodology

In order to obtain a set of "high fidelity" simulation data. A convenient method is to evaluate the infinite series solutions to the respective IBVPs at a set of spatial points and temporal values.
The infinite series solution to IBVP (1) is given by

$$\boldsymbol{u}(\boldsymbol{x},t) = \sum_{n=1}^{\infty} \boldsymbol{A}_n e^{-n^2\pi^2 t} sin(n\pi \boldsymbol{x}) \tag{3}$$

where $\boldsymbol{A}_n = \frac{2}{n\pi}(1-\cos(n\pi))$. And the infinite solution to IBVP (2) is given by

$$\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y},t) = \sum_{m,n=1}^{\infty} \boldsymbol{A}_{mn} e^{-(m^2+n^2)\pi^2 t} sin(m\pi \boldsymbol{x}) sin(n\pi \boldsymbol{y}) \tag{4}$$

where $A_{mn} = \frac{4}{mn\pi^2}(1-\cos(m\pi))(1-\cos(n\pi))$.

An alternative method is to solve this equation numerically. We approximate all the derivatives by finite differences with a second-order central difference scheme for the spatial derivative at position and the forward difference in time. The discrete form is then written as:

1D:

$$\frac{u_i^{n+1}-u_i^n}{\Delta t} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} \tag{5}$$

2D:

$$\frac{u_{i,j}^{n+1}-u_{i,j}^n}{\Delta t} = \frac{u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n}{\Delta x^2} + \frac{u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n}{\Delta y^2} \tag{6}$$

where Δt is time step, Δx and Δy are space steps in direction x and y respectively, $n = 0,1,2,\cdots$; $i = 0,1,\cdots I$; $j = 0,1,\cdots,J$.

So, with these recurrence relations, and knowing the values at time n, one can obtain the corresponding values at time $n+1$. $u_0^n, u_{0,0}^n$ and $u_I^n, u_{I,J}^n$ must be replaced by the boundary conditions. Furthermore, based on the initial conditions, $u_i^0, u_{i,j}^0$ are all given.

2.2 "High fidelity" simulation analysis

The aim in this section is to compare the difference of two "high fidelity" simulations. More precisely, above two methodologies are used to simulate the temperature field at each value of time in the set for IBVP (1) {0.00, 0.001, 0.002, …, 0.200} and in IBVP (2) in the set {0.00, 0.05, 0.10, …, 0.45, 0.50}. The space step is 0.01 in 1D and 0.01×0.01 in 2D. Several temperature distributions are shown in Figure 1. The data was stored for use as empirical data in the POD.

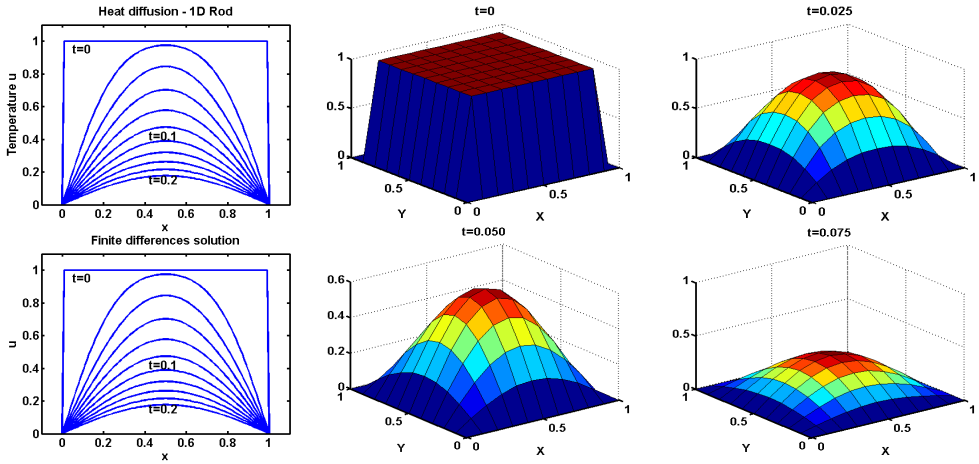


Figure 1. Top Left: Simulation of IBVP (1): time dependent heat diffusion on 1D rod with constant initial condition and zero boundary conditions. Bottom Left: The finite differences simulation of IBVP (1). Right: Simulation of IBVP (2): time dependent heat diffusion on 2D plate with constant initial condition and zero boundary conditions.

From Figure 1, we can observe that the simulations using analytical functions coincide with those obtained by the numerical method.

3 Approximate simulation based on the Galerkin Method and the POD

This section includes a brief overview of the Galerkin method for PDE discretization and implementation of POD to get an orthogonal basis of space domain.

3.1 Discretization with the Galerkin method

The Galerkin method is a discretization scheme for PDEs which is one type of spectral methods or methods of weighted residuals. The main idea is to separate variables and to represent a field with a truncated series expansion given by

$$\boldsymbol{u}(\boldsymbol{x},t)=\sum_{i=1}^{N}\alpha_i(t)\boldsymbol{\varphi}_i(\boldsymbol{x}) \tag{7}$$

where $\boldsymbol{\varphi}_i(\boldsymbol{x})$ are trial functions which can form an orthonormal basis for the approximate function space. $\alpha_i(t)$ are time dependent coefficients obtained by minimizing the residuals or errors between approximate and exact values. Equivalently, the residuals must be orthogonal to each one of the given trial functions. Thus, the original infinite dimensional system can be approximated by an *N*-dimensional one.

3.2 Construction of reduced basis function via the POD

As stated earlier, a set of "high fidelity" simulations is recorded yielding the snapshots of the heat equation solution for IBVP (1) at *M*=200 equally spaced sample times between $t=0$ and $t=0.200$, and at *M*=20 equally spaced sample times between $t=0$ and $t=0.5$ for IBVP (2) (the IC was also used as the first snapshot). These snapshots are used as the empirical data for computing a set of basis functions via the POD.

If we denote the set of original snapshots as $\{\boldsymbol{u}(\boldsymbol{x},t_k):k=1,2,...,M\}$ then the average snapshot is computed as

$$\bar{\boldsymbol{u}}(\boldsymbol{x})=\frac{1}{M}\sum_{k=1}^{M}\boldsymbol{u}(\boldsymbol{x},t_k) \tag{8}$$

and the centered snapshots are given by

$$\boldsymbol{v}^{(k)}=\boldsymbol{v}(\boldsymbol{x},\boldsymbol{t}_k)=\boldsymbol{u}(\boldsymbol{x},\boldsymbol{t}_k)-\bar{\boldsymbol{u}}(\boldsymbol{x}) \tag{9}$$

This adjustment leaves us with a new ensemble of data samples $\{\boldsymbol{v}(\boldsymbol{x},t_k):k=1,....M\}$. These snapshots are then used to compute the $M\times M$ empirical correlation matrix $\boldsymbol{C}$ whose entries are given by

$$(\boldsymbol{C})_{ij}=\frac{1}{M}\int_{\Omega}v^{(i)}(x)v^{(j)}(x)dx \quad i,j=1,\cdots,M \tag{10}$$

where Ω is the spatial domain ([0,1]). The problem is reduced to finding the eigenvectors and eigenvalues of $\boldsymbol{C}$, and the eigenvectors $\boldsymbol{A}^{(n)}$ of $\boldsymbol{C}$ and the corresponding eigenvalues λ_n satisfy

$$\boldsymbol{C}\boldsymbol{A}^{(n)}=\lambda_n\boldsymbol{A}^{(n)} \quad n=1,...,M \tag{11}$$

which can be solved for corresponding system of M eigenvalues and M eigenvectors. The numerical integration (10) is hard-coded using a simple approximation technique. The eigenvalues and eigenvectors of $\boldsymbol{C}$ are then used to compute the empirically determined eigenfunctions, and the basis functions are then computed as linear combinations of data samples using

$$\boldsymbol{\varphi}_n(\boldsymbol{x}) = \sum_{k=1}^{M} \boldsymbol{A}_k^{(n)} \boldsymbol{v}^{(k)}(\boldsymbol{x}) \quad n = 1,\ldots,M \tag{12}$$

It is easy to check

$$(\boldsymbol{\varphi}_l, \boldsymbol{\varphi}_m) = \begin{cases} 1 & l = m \\ 0 & l \neq m \end{cases}. \tag{13}$$

This completes the construction of the orthonormal set $\{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_M\}$.
By utilizing the properties of the POD one can specify an energy level e to be captured and then seek $N<<M$ such that

$$\frac{\sum_{i=1}^{N} \lambda_i}{\sum_{i=1}^{M} \lambda_i} > e .$$

Then, based on the Galerkin method, the approximation $\hat{\boldsymbol{v}}$ to the $\boldsymbol{v}(\boldsymbol{x},t)$ is given by the truncated series expansion

$$\hat{\boldsymbol{v}}(\boldsymbol{x},t) = \sum_{n=1}^{N} \alpha_n(t) \boldsymbol{\varphi}_n(\boldsymbol{x}). \tag{14}$$

The average snapshot $\bar{\boldsymbol{u}}$ is then added

$$\hat{\boldsymbol{u}}(\boldsymbol{x},t) = \bar{\boldsymbol{u}}(\boldsymbol{x}) + \hat{\boldsymbol{v}}(\boldsymbol{x},t) \tag{15}$$

to reconstruct the original data samples. The approximation order N can be varied to achieve the desired degree of accuracy.
The $\alpha_n(t)$ are time-dependent coefficients chosen to ensure the original PDE satisfied as closely as possible by (14). This is achieved by minimizing the residual. More details are discussed in the following section.

3.3 Calculation of the coefficients by solving an ODE

We suppose we have a system governed by the PDEs (in symbolic form)

$$\frac{\partial \boldsymbol{v}}{\partial t} = D(\boldsymbol{v}); \quad \boldsymbol{v} : D \times (0,\infty) \to \Re \tag{16}$$

with appropriate boundary conditions and initial conditions, where $D(\square)$ is a spatial operator , e.g. the Laplacian in the case of heat diffusion. Define the residual as

$$\boldsymbol{r}(\boldsymbol{x},t)=\frac{\partial\hat{\boldsymbol{v}}}{\partial t}-D(\hat{\boldsymbol{v}})\,. \tag{17}$$

We force the residual to be orthogonal to a suitable number of eigenfunctions, i.e.

$$\left\langle \boldsymbol{r}(\boldsymbol{x},t),\boldsymbol{\varphi}_n(\boldsymbol{x})\right\rangle=0\quad n=1,\ldots,N\,. \tag{18}$$

Substituting (14) into (17) yields,

$$\boldsymbol{r}(\boldsymbol{x},t)=\sum_{n=1}^{N}\dot{\alpha}_n(t)\boldsymbol{\varphi}_n(\boldsymbol{x})-D(\sum_{m=1}^{N}\alpha_m(t)\boldsymbol{\varphi}_m(\boldsymbol{x}))\,. \tag{19}$$

Applying the orthogonality condition (18) and using the orthonormality property of the set of eigenfunctions results in

$$\dot{\alpha}_i(t)=\int_D D(\sum_{m=1}^{N}\alpha_m(t)\boldsymbol{\varphi}_m(\boldsymbol{x}))\boldsymbol{\varphi}_i(\boldsymbol{x})d\boldsymbol{x}\quad i=1,\ldots,N \tag{20}$$

Thus, requiring the residual be orthogonal to the first N eigenfunctions yields a system of N ordinary differential equations in t (an N^{th} -order system)

$$\dot{\boldsymbol{\alpha}}=F(\boldsymbol{\alpha}) \tag{21}$$

where $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_N)$ and $F:\Re^N\to\Re^N$.
The initial conditions for the resulting system of ODEs are determined by a second application of the Galerkin approach. We force the residual $\boldsymbol{I}(\boldsymbol{x})=\boldsymbol{v}(\boldsymbol{x},0)-\hat{\boldsymbol{v}}(\boldsymbol{x},0)$ of the initial conditions to also be orthogonal to the first N eigenfunctions. We obtain a system of N linear equations

$$\alpha_i(0)=\int_D \boldsymbol{v}(\boldsymbol{x},0)\boldsymbol{\varphi}_i(\boldsymbol{x})d\boldsymbol{x}\quad i=1,\ldots,N\,. \tag{22}$$

The heat diffusion system dynamics are described by

$$\frac{\partial(\boldsymbol{v}+\overline{\boldsymbol{u}})}{\partial t}=D(\boldsymbol{v})=\nabla^2(\boldsymbol{v}+\overline{\boldsymbol{u}}) \tag{23}$$

Applying (20) yields the system of linear ODEs

$$\dot{\alpha}_i(t)=\sum_{j=1}^{N}\alpha_j(t)\int_D\boldsymbol{\varphi}_i(\boldsymbol{x})\nabla^2\boldsymbol{\varphi}_j(\boldsymbol{x})d\boldsymbol{x}+\int_D\nabla^2\overline{\boldsymbol{u}}(\boldsymbol{x})\boldsymbol{\varphi}_i(\boldsymbol{x})d\boldsymbol{x}\quad i=1,\ldots,N \tag{24}$$

with initial conditions

$$\alpha_i(0)=\int_D\boldsymbol{\varphi}_i(\boldsymbol{x})\boldsymbol{v}(0,\boldsymbol{x})d\boldsymbol{x}\quad i=1,\ldots,N \tag{25}$$

where $D=[0,1]$ for the rod. This results in linear system of ODEs

$$\dot{\boldsymbol{\alpha}}(t) = \boldsymbol{\Gamma}\boldsymbol{\alpha}(t) + \boldsymbol{b} \tag{26}$$

where $\boldsymbol{\alpha}(t)$ is an N-vector, $\boldsymbol{\Gamma}$ is the $N \times N$ matrix with entries

$$(\boldsymbol{\Gamma})_{ij} = \int_D \boldsymbol{\varphi}_i(\boldsymbol{x}) \nabla^2 \boldsymbol{\varphi}_j(\boldsymbol{x}) d\boldsymbol{x} \tag{27}$$

and $\boldsymbol{b}$ is an N-vector with elements

$$b_i = \int_D \nabla^2 \overline{u}(\boldsymbol{x}) \boldsymbol{\varphi}_i(\boldsymbol{x}) d\boldsymbol{x} \,. \tag{28}$$

The solution to (26) is given by the variation of constants formula

$$\boldsymbol{\alpha}(t) = e^{\boldsymbol{\Gamma} t}\boldsymbol{\alpha}(0) + \int_0^t e^{\boldsymbol{\Gamma}(t-\tau)} \boldsymbol{b} d\tau \tag{29}$$

where the IC $\boldsymbol{\alpha}(0)$ is an N-vector with entries given by (22). However, rather than hard-code the solution (29) we can numerically integrate (26) using Runge-Kutta method.
Once the ODE (26) is solved and evaluated at the desired values of t, the $\hat{\boldsymbol{u}}(\boldsymbol{x}, t)$ is known.

4 Calculation of the coefficients by Kriging interpolation

Once the set of snapshots $\{\boldsymbol{v}(\boldsymbol{x}, t_k) : k = 1, \ldots, M\}$ and reduced basis functions $\boldsymbol{\varphi}_n$ are obtained, the set of coefficients can be calculated by the projection of those snapshots on the basis fuctions:

$$\alpha_k^{(i)} = < \boldsymbol{v}(\boldsymbol{x}, t_k), \boldsymbol{\varphi}_i >, i = 1, \cdots, M; k = 1, \cdots, M \tag{30}$$

or

$$\alpha_k^{(i)} = < \boldsymbol{u}(\boldsymbol{x}, t_k) - \overline{\boldsymbol{u}}, \boldsymbol{\varphi}_i >, i = 1, \cdots, M; k = 1, \cdots, M \tag{31}$$

where $<\square, \square>$ denotes the inner product.
Then, any general approximation technique may be used to build surrogate response surfaces of each coefficient $\alpha_i(t), i = 1, \cdots, M$. Here, Kriging interpolation is used as it can capture the local phenomena. The simulation of heat diffusion is finally assembled at any time point:

$$\boldsymbol{u\_approx}(\boldsymbol{x}, t) = \overline{\boldsymbol{u}}(\boldsymbol{x}) + \sum_{n=1}^{M} \tilde{\alpha}_n(t) \boldsymbol{\varphi}_n(\boldsymbol{x}) \tag{32}$$

Same as the before, we can using the truncated expansion to evaluate $\boldsymbol{u}(\boldsymbol{x}, t)$ as

$$\boldsymbol{u\_tru}(x, t) = \overline{\boldsymbol{u}}(x) + \sum_{n=1}^{N} \tilde{\alpha}_n(t) \boldsymbol{\varphi}_n(x), \quad N << M \tag{33}$$

4.1 Kriging surrogate model

Kriging meta-model is an interpolation technique based on statistical theory, which consists of a parametric linear regression model and a non-parametric stochastic process. It needs a design of experiments to define its stochastic parameters and then predictions of the response can be completed at any unknown point. Given an initial design of experiments (initial DoE): $\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)}\}^T$, with observed responses, $\boldsymbol{Y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(n)}\}^T$. $\boldsymbol{Y}$ could be generated by high fidelity simulations or experiments.

Kriging surrogate model presumes the real function relationship between the DoE and the response as

$$y(\boldsymbol{X}) = \mu + Z(\boldsymbol{X}) \tag{34}$$

where μ is a hyperparameter which is determined part and $Z(\boldsymbol{X})$ is a Gaussian stochastic process with zero mean and covariance in the form of

$$\mathrm{Cov}(Z(\boldsymbol{X}^{(i)}), Z(\boldsymbol{X}^{(j)})) = \sigma_z^2 \boldsymbol{R}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}) \tag{35}$$

where $\boldsymbol{R}$ is the correlation function between two sample points and σ_z^2 the Gaussian process variance. For $\boldsymbol{R}$, most applications use Gaussian function

$$\boldsymbol{R}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}) = \exp(-d(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)})) \tag{36}$$

where $d(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)})$ is the distance function between $\boldsymbol{X}^{(i)}$ and $\boldsymbol{X}^{(j)}$. Usually it is a weighted distance function

$$d(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}) = \sum_{k=1}^{z} \theta_k \left| x_k^{(i)} - x_k^{(j)} \right|^2 \tag{37}$$

Hyperparameters θ_k control the degree of nonlinearity in kriging surrogate model. Sometimes we choose θ_k equal to 2. Through maximum likelihood prediction, the estimates for μ and σ_z^2 is given

$$\begin{cases} \hat{\mu} = \dfrac{\boldsymbol{I}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{Y}}{\boldsymbol{I}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{I}} \\ \sigma_z^2 = \dfrac{1}{n} (\boldsymbol{Y} - \boldsymbol{I}\hat{\mu})^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\boldsymbol{Y} - \boldsymbol{I}\hat{\mu}) \end{cases} \tag{38}$$

where $\boldsymbol{\Psi}$ is a $n \times n$ matrix $(\boldsymbol{\Psi})_{ij} = \boldsymbol{R}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)})$, $\boldsymbol{I}$ is the unit matrix. Thus the prediction model could be built as

$$y(\boldsymbol{X}) = \hat{\mu} + \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{X}) \boldsymbol{\Psi}^{-1} (\boldsymbol{Y} - \boldsymbol{I}\hat{\mu}) \tag{39}$$

Here $\boldsymbol{r}(\boldsymbol{X}) = [\boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{(1)}), \boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{(2)}), \ldots, \boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{(n)})]^{\mathrm{T}}$.

5 Numerical Analysis

Now, we present some results of the above computations and simulations. The whole process has been performed in four steps:
- A set of basis functions was determined using the POD (according to from Eq.(8) to Eq.(11)) for the 1D heat diffusion system with $M = 201$ snapshots and $M = 21$ in 2D
- Calculation of the coefficients α by projection of snapshots on the basis
- Based on the data obtained in step 1 and 2, the Kriging surrogate model can determine an approximation of coefficients α(t) for any time point
- The simulation is then approximated by Eq. (32) or Eq.(33)

5.1 Eigenvalues and corresponding eigenfunctions

As stated earlier, the eigenvalues measure the relative energy of the system dynamics. Figure 2 shows the resulting empirically determined eigenfunctions for the 1D and 2D heat diffusion systems corresponding to first four eigenvalues in decreasing order.

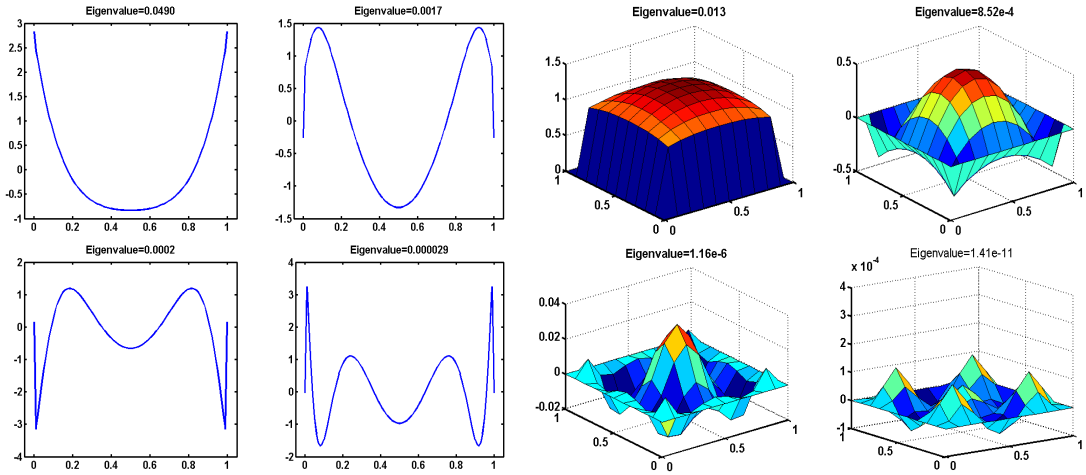


Figure 2. The first four basis functions of the system with corresponding eigenvalues for IBVP (1) (Left) and for IBVP (2)(Right).

From Fig.2, it is readily observed that the four modes contain virtually all of the energy.

5.2 Reconstruction error analysis

The reconstruction errors are calculated for the original snapshots. Figure 3 shows that the relative errors $\|\mathbf{u}-\tilde{\mathbf{u}}\|/\|\mathbf{u}\|$ on the temperature field (Figure 3,left for 1D, right is about 2D) decrease quickly with increasing the number of modes. Furthermore, we observe that the reconstruction errors at the initial time point are slightly smaller than those at other time points. So that'a why in the following Garlerkin approximations, only the first three modes are used (N=3).

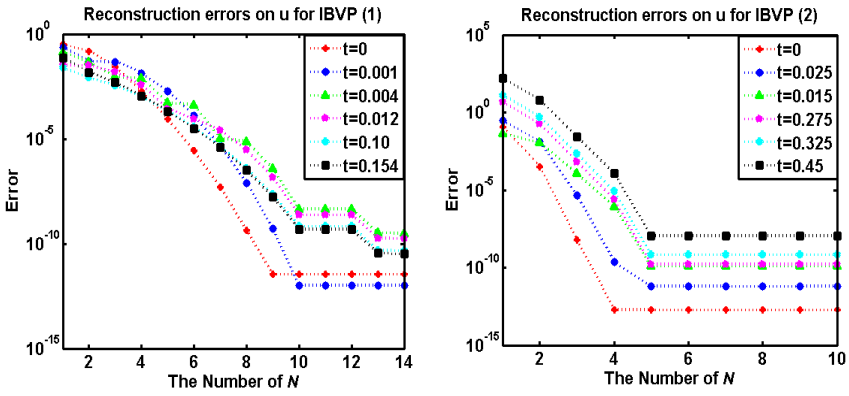


Figure 3. The construction errors on the temperature field u for both two IBVPs.

5.3 Comparison of exact temperature field and its reconstruction

From Figure 4, middle, we observe that the reconstruction temperature field is similar to the exact one. As expected, solution approximated with coefficients based on Kriging interpolation reproduces the original data when the number of POD modes N is chosen to equal the number of snapshots M. While a slight error with the truncation of POD modes, N =3. This can be seen more clearly in Figure 5.

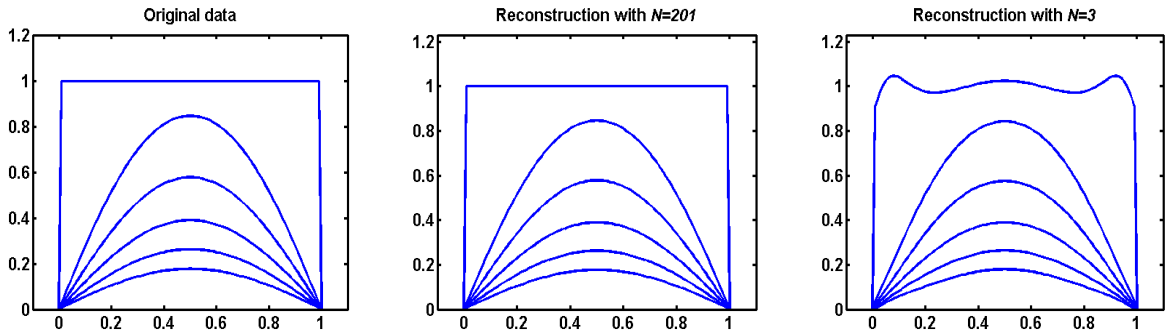


Figure 4. Original heat diffusion data u (top, left) from infinite series solution, and reconstruction temperature field of IBVP(1) using modes N=M=201 or empirical data determined eigenfunctions N(top, right) and ones with N=3 (bottom, left) for $t = 0$, 0.04, 0.08, 0.12, 0.16, 0.20 respectively.

Table 1. The reconstruction error with N=201 and N=3 respectively

Error	t=0.0	t=0.001	t=0.004	t=0.009
$\|u - u\_appox\|/\|u\|$	3.3829e-12	1.0317e-12	1.8553e-12	8.1031e-13
$\|u - u\_tru\|/\|u\|$	0.0270	0.0481	0.0091	0.0156

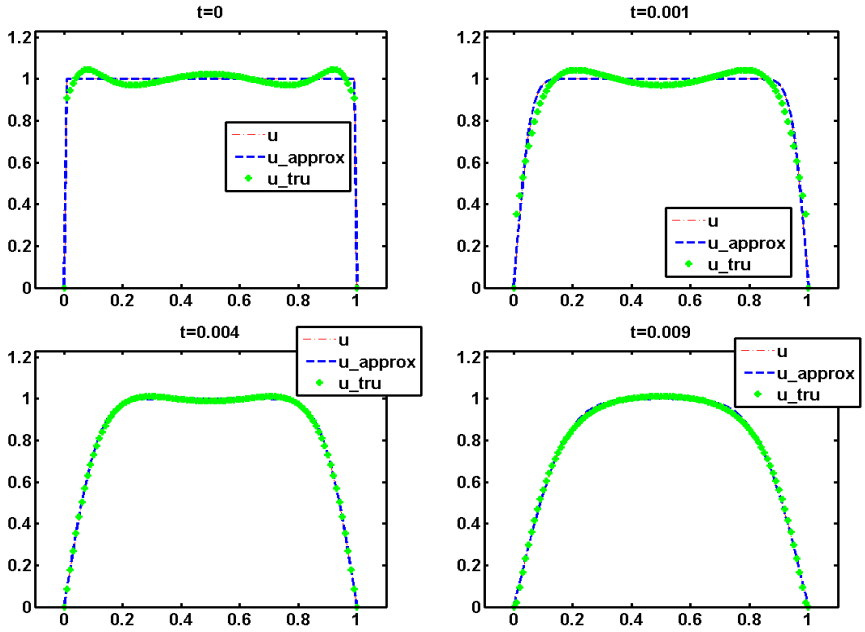


Figure 5. Heat diffusion profiles for $t = 0, 0.001, 0.004, 0.009$**.**

Figure 5 gives to the exact temperature field and reconstruction one at different time points. The errors are given in table 1. We observe that the approximations are accurate. That is to say, the approximation accuracy increases rapidly with time, although there is difficulty in reproducing the initial condition. This phenomenon is due to the fact that the solution progress from a discontinuous initial condition to smooth profiles requires fewer terms to get equivalent accuracy. Similar conclusion is observed for the 2D domain, Figure 6.

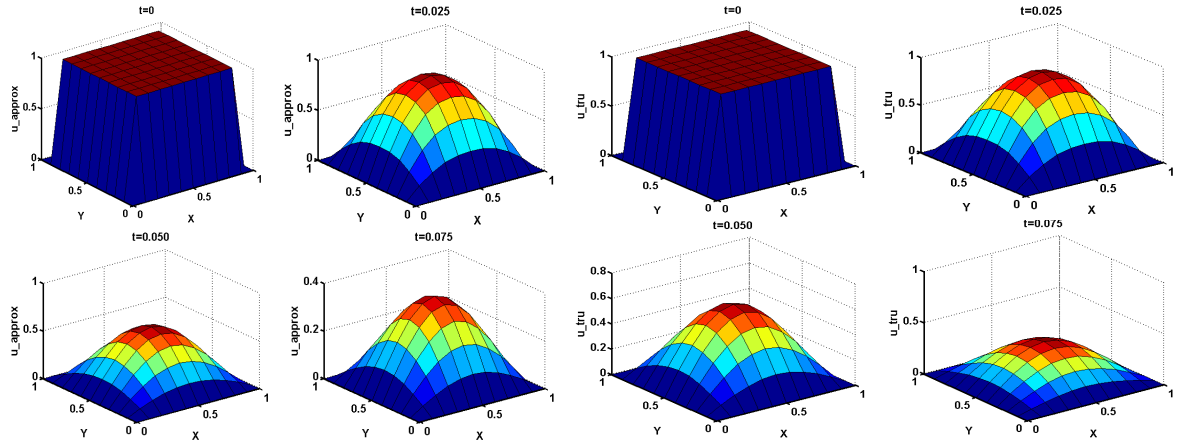


Figure 6. Approximate solutions of IBVP (2) using empirically determined eigenfunctions for *t*=0, 0.025, 0.05, 0.075. Left: 201 eigenfunctions are used. Right: 3 eigenfunctions are used.

As stated earlier, Kriging meta-model is a technique that can provide the predictions of the response at arbitrary point. Therefore, the advantage we used the Kriging to interpolation the coefficient $\boldsymbol{a}(t)$ of POD is that we can calculate the value of $\boldsymbol{u}$ at any time point different from the sampling points. Figure 7 shows the comparison of original data $\boldsymbol{u}$ and others two approximate values with N=201 denoting ***u_approx*** and N=3 for ***u_tru*** at $t = 0.0045, 0.1255, 0.201, 0.210$.

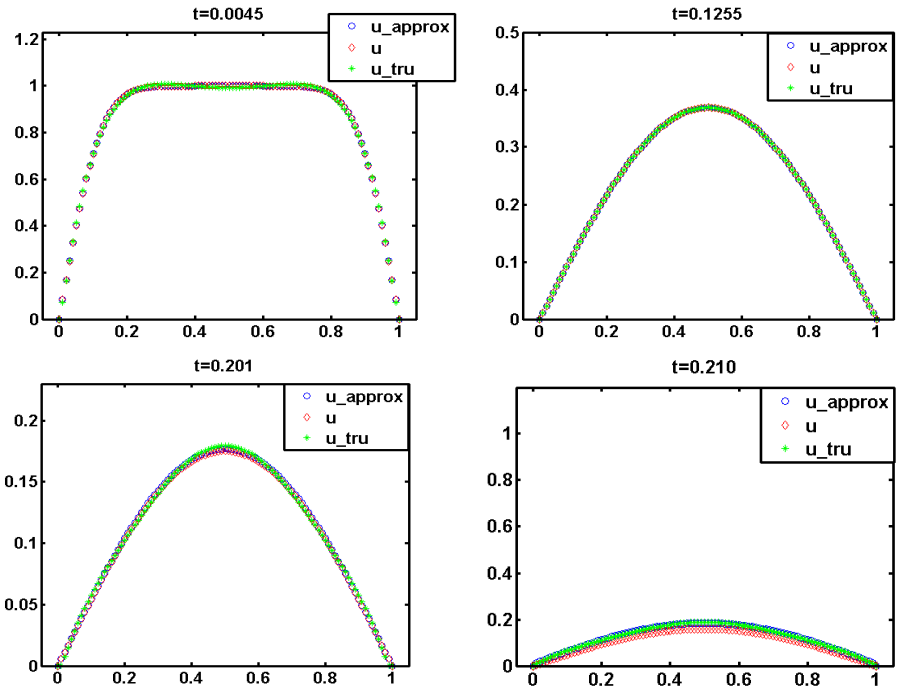


Figure 7. The values of u, *u_approx* ,and *u_tru* at $t = 0.0045, 0.1255, 0.201$ and $t = 0.210$.

From Figure 7, we can conclude that the prediction of u can get good accuracy when the time point t is in the region [0,0.2]. However, when t becomes larger than 0.2, the errors between exact field u and the approximations *u_approx* become much bigger with the time increasing. That is obviously due to the average field computed from the snapshots at the time points belonging in the region [0,0.2].

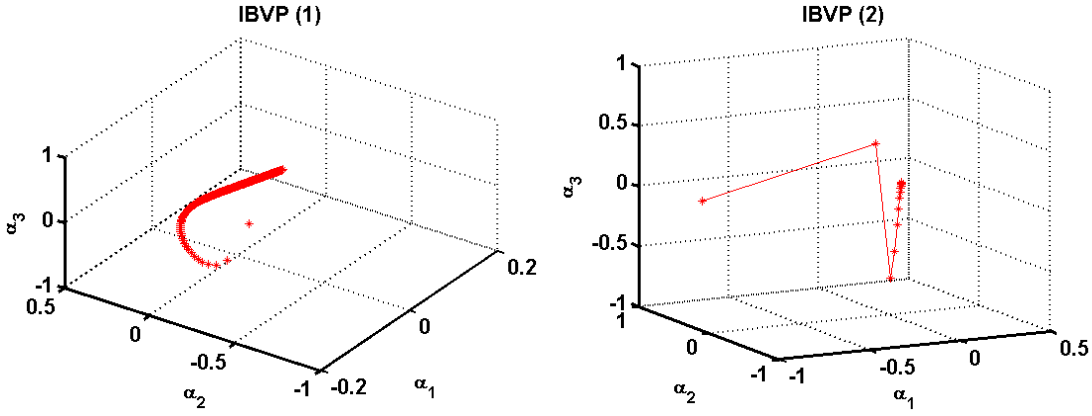


Figure 8. The first three $\alpha_i(t), i = 1,2,3$ for both two heat diffusion equation systems.

In Figure 8 we show the first three $\alpha_i(t)$,i=1,2,3 computed using inner product. It appears better to choose the second order polynomial function for regression in Kriging. While, the linear function for regression in IBVP(2) seeing from the right of Fig.8.

6 Conclusions

In this article, a technique combining the advantages of two types of surrogate models has been proposed to approximate the simulation of PDEs. After descritization of PDEs with the Galerkin method, the basis functions of space are first obtained by the standard POD. The second part consists in approximating the coefficients of Garlerkin discretization form of PDEs using Kriging surrogate model. The resulting reduced order model is then applied to simulate the heat diffusion in one-dimension rod and two-dimension plate. The numerical results show that reconstructed temperature field is efficiently approximated with the non-intrusive POD approach. The reconstruction errors are only controlled by the number of POD basis functions, as the Kriging interpolation of coefficients does not influence the precision of Garlerkin approximation.

In terms of future prospects, we will be interested in using this method to reconstruct the reduced order model for more complex systems and consider the multi-fidelity data at the same time.

7 Acknowledgment

This work has been supported by the National Natural Science Foundation of China (Grants No. 11302173).

References

[1] Andrew J. Newman. (1996) Model reduction via the Karhunen-Loeve expansion part I: An exposition. *Technical report, Institute for Systems Research.*

[2] Andrew J. Newman. (1996) Model reduction via the Karhunen-Loeve expansion part II: Some elementary examples. *Technical report, Institute for Systems Research.*

[3] A. Giunta, L. Watson, J. Koehler. (1998) A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models. *7th AIAA/USAF/NASA/IS SMO Symposium on Multidisciplinary Analysis & Optimization, Sept 2–4, St. Louis, MI, USA, AIAA*, 1998-4758.

[4] P. Breitkopf, H. Naceur, Rassineux. (2005) A moving least squares response surface approximation: formulation and metal forming applications. *Computers & Structures,* **83**:1411-1428.

[5] S.N. Lophaven, H.B. Nielsen, J. Sondergaard. (2002) DACE-A Matlab Kriging Toolbox. *Informatics and mathematical modeling, Denmark*.

[6] M. Samuelides. (2009) Surfaces de réponse et réduction de modèles, Optimisation multidisciplinaire en mécanique 2, Hermes Science Publications.

[7] M. Guarascio, M. David, C. Hyijbregts. (19750 Advanced geostatistics in mining industry, in: Proceedings of the NATO Advanced Study Institute, *Istituto di Geologia Applicata of the University, Rome, Italy*.

[8] G.Matheron. (1973) The intrinsic random functions and their applications. *Adv. Appl. Probab.* **5**(3): 439-468.

[9] I. Kaymaz. (2005) Application of Kriging method to structural reliability problems. *Struct. Saf.* **27**:133-151.

[10] M. Xiao, P. Breitkopf, R. Filomeno Coelho, C. Knopf-Lenoir, M. Sidorkiewicz, P. Villon. (2010) Model Reduction by CPOD and Kriging - application to the Shape Optimization of an intake port. *Structural and Multidisciplinary Optimization*, **41**(4):555-574.

[11] Holmes P, Lumley J L, Berkooz G. (1996) Turbulence, Coherent Structures, Dynamical Systems and Symmetry. *Cambridge: Cambridge University Press*.

[12] Fukunaga K. (1990) Introduction to Statistical Recognition. *New York: Academic Press*.

[13] Jolliffe I T. (2002) Principal Component Analysis, *Springer-Verlag*.

[14] Crommelin D T, Majda A J. (2004) Strategies for model reduction: comparing different optimal bases. *J Atmospheric Sci,* **61**: 2306-2317.

[15] S.S.Ravindran. (2000) A reduced-order approach for optimal control of fluids using proper orthogonal decomposition, *International Journal For Numerical Methods In Fluids,* **34**: 425-448.

[16] Lumley JL. (1967) The structure of inhomogeneous turbulence. In Atmospheric Turbulence and Radio Wave Propagation, Yaglom AM, Tatarski VA (eds). Nauka: Moscow, 166-178.

[17] Sirovich L. (1987) Tubulence and the dynamics of coherent structures: part I-III. *Quarterly of Applied Mathematics,* **45**(3): 561-590.

[18] Hung V. Ly, Hien T. Tran. (1999) Modeling and control of physical process using proper orthogonal decomposition. *Journal of Mathematical and Computer Modeling*.

[19] M. Xiao, P. Breitkopf, R. Filomeno Coelho, P. Villon, W. Zhang. (2014) Proper orthogonal decomposition with high number of linear constraints for aerodynamical shape optimization, *Applied Mathematics and Computation,* Volume **247**:1096-1112.

[20] M. Xiao, P. Breitkopf, R. Filomeno Coelho, C. Knopf-Lenoir, P. Villon, W. Zhang. (2013) Constrained Proper Orthogonal Decomposition based on QR-factorization for aerodynamical shape optimization. *Applied Mathematics and Computation* **22** (3):254–263.

An innovative approach to computational simulation of the functional characteristics of poroelastic materials illustrated with diffusion into articular cartilage

***†Jamal Kashani[1], Lihai Zhang[2], Yuantong Gu[1], and Adekunle Oloyede[1,3]**

[1]School of Chemistry, Physics and Mechanical Engineering, Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD, Australia.
[2]Department of Infrastructure Engineering, The University of Melbourne, Melbourne, VIC, Australia.
[3]Department of Mechanical Engineering, Elizade University, Ondo State, Nigeria.

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

Abstract

Collecting functional quantitative intra matrix data in experimental samples of articular cartilage is still challenging due to its delicate complex heterogeneous structure in which constituents are intermingled right up to the ultramicroscopic level. Any attempt to insert a transducer inside this material via piercing would damage the structure leading to unrepresentative data. Traditional non-invasive methods are technically difficult for obtaining precise functional data. This paper presents a novel computational approach, using the agent-based concept, to create a 'virtual microscope' that can be used to provide functional information throughout a heterogeneous complex medium, such as articular cartilage, in silico. The method involves two-dimensional cellular automata, a hybrid agent, new local agent rule and a traditional neighbourhood rule. The hybrid agent combines constituents of the system (solid and fluid) where the local rule determines intra-agent evolution. The proposed approach was validated by simulating diffusion into a model of cartilage matrix that was characterized with anisotropic permeability. The simulated results were then compared to magnetic resonance imaging (MRI) data. Spatial map of diffusion at different times and depth-dependent diffusion profiles were provided in colour-coded pictures. Qualitative and quantitative comparison of results with experimental data shows that this novel approach can accurately and efficiently represent diffusion of fluid into the cartilage matrix. It demonstrates the potential of hybrid agent and local rule to enhance agent-based techniques for porous materials and other areas of research. We conclude that the ability to establish a "virtual microscope" offers a viable opportunity for in-silico experiments that can extend our knowledge beyond the capability of traditional laboratory experiments, while also facilitating information for creating models for numerical methods such as finite element analysis, meshless and smoothed particle hydrodynamics. The combination of the approach presented here with conventional simulation methods can provide a framework for modelling and analysis of complex porous materials. We concluded that the hybrid agent and local rule concept introduced in this paper can also be potentially exploited to enhance many of the existing agent-based techniques.

Keywords: Articular cartilage, hybrid agent, local and global rules, porous materials, agent-based method

Introduction

Articular cartilage is a semipermeable porous biomechanically functional material that is saturated with an osmotically active fluid which occupies between 65 and 80 %, proteoglycans and collagen components that constitute its solid skeleton occupying 5-10% and 15-22% respectively of its matrix [2]. These components are intermixed right up to the molecular level [3] such that the tissue is highly heterogeneous and anisotropic in nature [4]. Quantitative observation and understanding of the underlying mechanisms of articular cartilage's functional characteristics at the microscopic and submicroscopic scales is still a major challenge due to the non-phasic nature of the tissue and the complex interactions between its components. Any physical interference, such as probing the matrix with a transducer via piercing can destroy the articular cartilage structure and lead to an unrepresentative tissue in experimental analysis. As a result, classical laboratory experiments are arguably deficient in their ability to provide functional information such as fluid dynamics with simultaneous osmotic activities which plays a significant role in the mechanical function of the tissue [5, 6].

The ability to probe the real time response of articular cartilage during function can provide a view beyond experimental curve-fitting that can only provide an estimated range of physical properties of the tissue [7]. Non-invasive methods, i.e. magnetic resonance imaging (MRI) and computed tomography (CT) scan, have been successfully used to obtain intra-matrix data from the tissue without disturbing its structure, where different components are distinguished based on their radio-densities or radio frequency signals contrast [8-11]. External contrast agents have been applied with MRI techniques to observe function-related properties of articular cartilage such as diffusion [1, 12, 13], however, it is still difficult and technically challenging to obtain accurate data such as time-varying diffusion and fluid percolation characteristics during deformation [14].

Methods based on continuum mechanics and physical laws have been developed to describe the behaviour of porous materials with respect to their phenomenological characteristics under known imposed external conditions [15, 16], while mechanical theories have also been employed to establish governing equations for cartilage behaviour where the tissue was described as a porous media or mixture [17-20]. These are usually represented as differential equations that determine characteristics of the medium as a function of parameters [21] and physical laws, e.g. Darcy's law. However, the solid skeleton and fluid components intermingle right up to the ultramicroscopic molecular level [3], leading to extremely complex responses that require a different approach beyond those available with current mathematical models and traditional experimental techniques. A close scrutiny of the results of such theoretical models demonstrates that they are inadequate for explaining the mechanisms behind observed material or system responses of this important tissue [22-24].

Agent-based methods (ABM) have recently improved capacity to simulate complex systems' behaviours [25, 26]. ABM is suitable for capturing complex emergent phenomena in which the "whole" seems to be more than the sum of its components because of the intricate interactions between the components [27, 28]. In our opinion further elucidation of the behavior of articular cartilage requires agent-based computational simulation, especially if we were to obtain critical insight into the micro-mechanisms underlying its complex responses under external stimuli. In this paper, we present a novel agent-based approach using an enhanced agent (hybrid agent) with local and global rules [29] that can be used to develop representative structural model of this tissue where the interactivities of the hybrid agent and

the neighbourhood rules provide a "virtual microscope" into the internal working of the system to provide critical knowledge in the area of cartilage biomechanics. This methodology would provide spatial and temporal functional data that could then facilitate other models such as finite element, mesh free, course-grained particle and smooth particle hydrodynamics. The method described below is a preliminary examination of the concept of the hybrid agent and use of a combination of local and global neighbourhood rules.

Material and methods

Adaptation of the hybrid agent for the articular cartilage

Hybrid agent contains within it the system's elements. It was adopted for articular cartilage in this study where fluid and solid skeleton are considered to be two major constituent components of the tissue. Hybrid agent (cell) consists of both solid and fluid within it, such that it is neither fully solid nor fluid while it can simultaneously exhibit the characteristics of both solid and fluid in time. Evolution of the hybrid agent occurs by changing and updating its solid and fluid proportion. Hybrid agent is also characterised by poroelastic material properties such as porosity and semi-permeability.

The matrix model

A two dimensional (2D) cellular automata (CA) lattice of hybrid cells, consisting of 29 x 46 cells, was employed to represent the extracellular matrix of the cartilage where all the hybrid cells in the lattice are equal and constant size since diffusion does not cause tissue deformation. Therefore a cell can be identified and characterized by the relevant fluid to solid ratio it contains (fs).

The distribution of fs in the lattice was determined based on known layered weight distribution of fluid and solid [30]. In this simulation diffusion was allowed from every direction except at the bottom of the lattice because of the assumed effect of the subchondral bone that results this region impervious. One layer of pseudo cells filled by marked fluid was added to the lattice at the left, right and top sides (figure. 1). This marked fluid penetrates into the lattice via fluid exchange between the pseudo and hybrid cells that represent the boundary of the cartilage matrix. The progression of the time-dependent flow (diffusion) within the matrix was followed by tracking marked fluid. The simulation ends when all initial fluid (unmarked) in hybrid cells has been replaced by marked fluid. A program in Matlab (Mathworks Inc, MA, USA) was developed to simulate the diffusion process over the time steps.

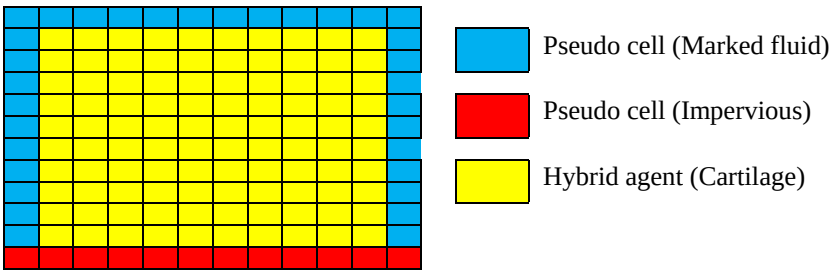


Figure 1: Schematic illustration of the lattice.

Rules

This simulation also incorporates a novel concept of simultaneous combination of local (intra-element) and global (inter-element) responses where intra-agent (local) and neighbourhood (global) rules apply. The local rule determines change within the hybrid cell (intra-agent change) in which the fluid-solid ratio (fs) of the agent changes and the global rule determines inter-agent interactions, e.g. interaction of a cell with its neighbours in the lattice.

Global rule: 2D van Neumann neighbourhood was implemented for interaction between neighbours in which each cell interacts with its orthogonally-adjacent neighbours as demonstrated in Figure 2. Van Neumann neighborhood defines a regular lattice that enables very efficient visualizations of diffusion processes [31].

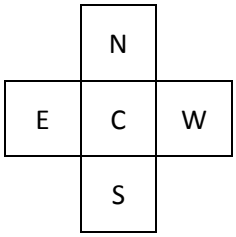


Figure 2. 2D van Neumann neighbourhood. Central cell (cell c) interacts with cells East (E), West (W), North (N) and South (S) at each time step.

Local rule: The following local rules were developed and used in this study:

- Cells are permeable and only fluid, including marked and unmarked, can move in and out of the cell. Since there is no deformation in the matrix, the amounts of fluid types that move in and out are equal.
- The amount of total fluid, marked and unmarked combined, in a cell is constant and does not change over time. Therefore, the ratio of total fluid to contained solid (fs) does not change in a cell. However, the proportion of marked and unmarked fluid may change as a result of fluid exchange.
- Only certain proportion of contained fluid in a cell can move out as a consequence of fluid exchange with neighbours at each time step. This proportion of exchangeable fluid depends on fs of the cell and location of the neighbours. The exchangeable fluid of a cell when interacts with another cell (a neighbour) is estimated as:
  Proportion of exchangeable fluid = k * fs , where k is constant.
  The parameter k depends on cell's neighbour location and indicates the direction of fluid movement. It is assumed that k in the horizontal direction is two times greater than in the vertical direction since hydraulic permeability of cartilage in the axial is half of that in radial direction when the tissue is unloaded [32]. In this simulation, k was set to 0.1 in the horizontal direction and 0.05 in the vertical direction. If a cell exchanges with more than one neighbour, the total exchangeable fluid proportion would be equal to the sum of the individual proportions. For example, when cell C in figure 1 interacts with all of its neighbours (W,E,N and S), the total fluid exchanged equals the sum of fluid exchanged with each neighbour:

$$FEP_C = k_N * fs_C + k_S * fs_C + k_W * fs_C + k_E * fs_C$$

Where, FEP_C is fluid exchange proportion of cell C, fs_C is ratio of fluid to solid content in cell C, and k_N, k_S, k_W and k_E are constant values of k in the N, S, W and E directions which are equal 0.05, 0.05, 0.1 and 0.1 respectively in this simulation.

The results obtained from cellular automata (CA) simulation were compared and validated with experimental data using contrast enhanced cartilage tomography (CECT) and peripheral quantitative computed tomography (pQCT) technique, taken from the literature [1] while 1800 time steps of the CA simulation is equivalent to 12 hours of diffusion. Therefore, each time step corresponds to 2.5 seconds. Width of the experimental samples is 2.5mm while the thickness is 4mm, corresponding to a width to thickness ratio of 0.625 and a simulation lattice dimensional ratio of 29 / 46 (approximately 0.63).

Results

The diffusion patterns of marked agents into the lattice at T=300, 600, 900 and 1800 are presented in figure 3A. The colour-coded map shows the spatial distribution of the ratio of marked fluid to total fluid content within the matrix based on percentage at a given time step. Each colour represents a certain percentage of concentration according to the legend attached to the pictures. Red colour illustrates regions where the initial fluid has almost been replaced by the marked fluid while blue represents areas with very little proportion of marked fluid. Initially (at T=0) concentration of the marked fluid in the lattice was zero (not shown in the figure). Then the marked fluid percolated into the lattice resulting in increased proportion of marked fluid over time (T=300, 600 and 900). The process of diffusion reaches equilibrium after about 1800 time steps, when all initial fluid was replaced by marked fluid.

Figure 3B presents experimental results [1] at time points 2, 4, 6 and 12 hours (left to right), corresponding to time steps in figure 3A. The legend on the right shows contrast agent concentration based on mM in which red illustrate maximum concentration (15 mM) that can be reached at equilibrium state (after 12 hours) and light blue demonstrate zero concentration. In order to compare the experimental with the simulated data, percentage of contrast agent concentration (left legend) was calculated based on ratio of contrast agent concentration to maximum concentration.

Comparison between CA and experimental data (micrographs) demonstrate similar patterns of diffusion into the cartilage. At T=300 and its experimental corresponding time (2 hours), the concentration at area near surface is high and fluid could not penetrate deep during this time. At T=600 (4 hours) concentration of marked fluid (or contrast agent) has been increased significantly up to the centre of the tissue along its thickness while at T=900 which is equal to 6 hours, only the region close to the bone did not undergo a significant concentration change. Both CA simulation and experimental test reached steady state condition at the same time (T=1800, T=12 hours).

The depth-dependent bulk concentration of marked fluid after 600 time steps and corresponding four-hour diffusion of contrast agents [1] (in percentage) are plotted in figure

4. The concentration is maximum at the surface and then drops gradually to about 40% near the bone with almost the same trend for both experimental and CA simulation results. The discrepancy between results in the middle region can be attributed by biological variation of tissue samples. The CA results compare resonably well with experimental data which substantiates the validity of the results of our CA simulation.

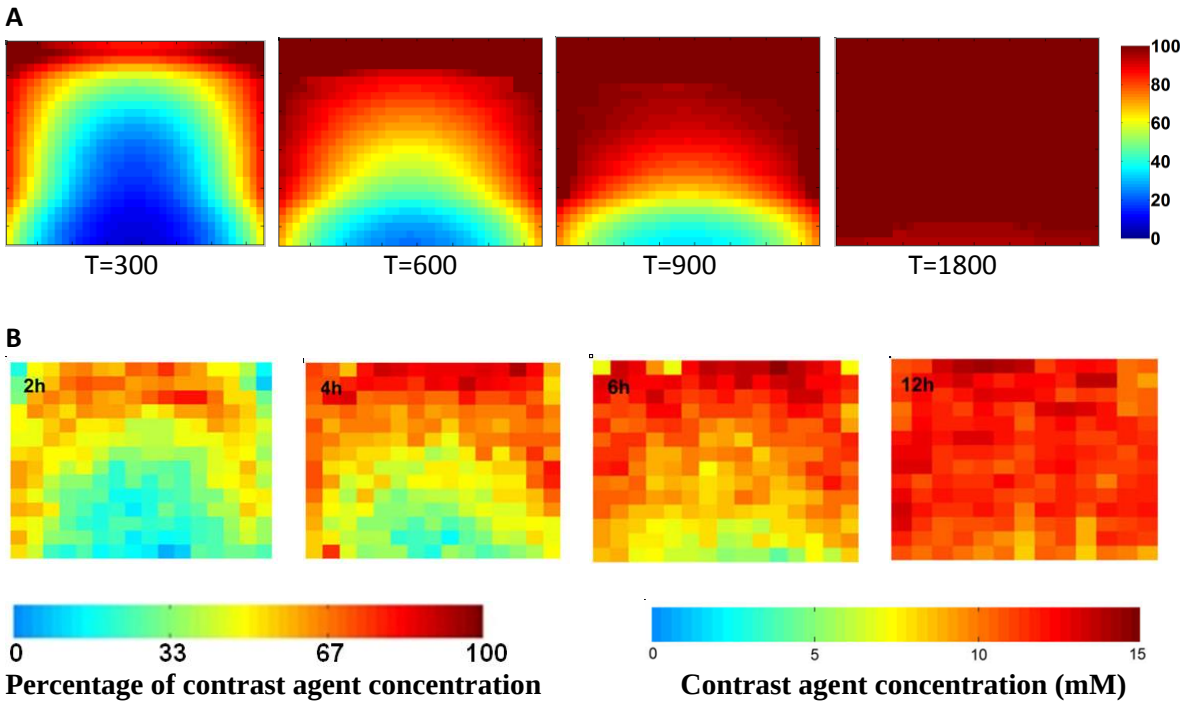


Figure 3. Diffusion into human articular cartilage at different times. A: Percentage of marked fluid in the lattice at time steps 300, 600, 900 and 1800. B: Contrast agent diffusion after 2, 4, 6 and 12 hours immersion [1].

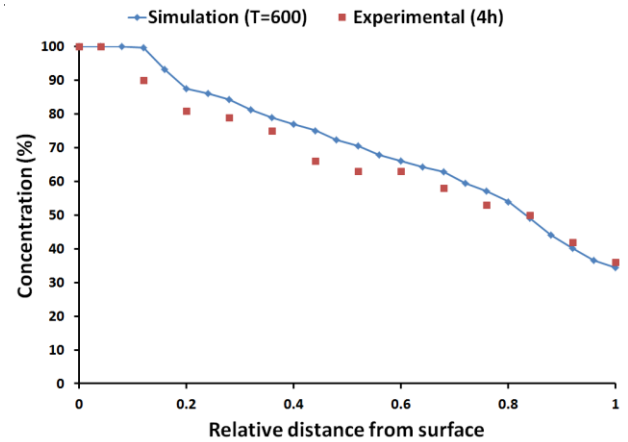


Figure 4. Percentage of depth concentration of marked agent (at T=600) and contrast agent in the human cartilage after 4 hours of immersion [1].

Figure 5 shows depth-dependent bulk concentration percentage of marked fluid collected at various areas in depth including surface, middle (½ thickness depth), ¾ thickness depth and bottom. Overall, the concentrations are lower towards the bottom regions (close to the bone) in time. The curve representing concentration at the surface illustrates that unmarked fluid is replaced by marked fluid rapidly and after about 400 time steps all of unmarked fluid move out of this region. The profile of concentration at the bottom layer follows different trends over time and takes significantly longer time to replace all initial fluid with marked fluid. All curves demonstrate growth of marked fluid over time while the rate of increase over time drops with depth.

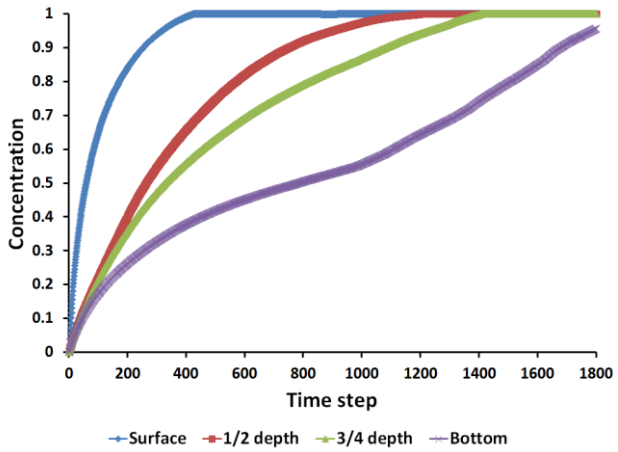


Figure 5. Depth- and time-dependent profiles of marked fluid concentration for surface, bottom, ½ and ¾ thickness depths.

Discussion

In the present paper, an enhanced agent-based approach involving the combination of a novel hybrid element, local intra agent and neighbourhood inter agent rules was applied to articular cartilage. This methodology provided a unique opportunity for investigating the transient intramatrix diffusion of the cartilage. For the first time, diffusion and percolation of fluid into cartilage as a non-phasic material was successfully investigated quantitatively (fig. 3) and qualitatively (figs 4 and 5) using an agent-based method. The comparison of results of this novel approach with corresponding experimental data shows a reasonably close agreement. The success of this approach suggests that it can be used for further investigation of the functional characteristics of loaded and deforming articular cartilage, and also tissues that are affected by degeneration and disease where current methods are technically or ethically inadequate.

The hybrid agent provides us an opportunity to create multi-component structures without any obligation to distinguish constituent components. This capability makes hybrid agent suitable for non-phasic porous materials such as biological tissues and articular cartilage in particular. In addition, as cells (agents) are micro-scale elements of an agent-based structure, hybrid agent, by means of local rule, is capable of intra-agent evolution that provides the feasibility

of studying system change in time at micro-scale level. Therefore, micro-scale spatial and temporal data can be obtained in a manner describable as using a "virtual microscope" in which tissue can be probed unlimitedly. It proposes a viable opportunity for in-silico experiments that can facilitate provision of input data for numerical methods such as finite element analysis, meshless and smoothed particle hydrodynamics. The combination of the approach presented here with numerical methods can prepare a framework for modelling and analysis of complex porous materials where the constituents of the system may be indistinguishable in the manner of known mixtures.

Acknowledgment
Authors would like to gratefully acknowledge the financial support provided by the Queensland University of Technology Postgraduate Research scholarship.

References

1. Silvast, T.S., et al., *pQCT study on diffusion and equilibrium distribution of iodinated anionic contrast agent in human articular cartilage – associations to matrix composition and integrity.* Osteoarthritis and Cartilage, 2009. **17**(1): p. 26-32.
2. Mow, V.C. and X.E. Guo, *Mechano-electrochemical properties of articular cartilage: their inhomogeneities and anisotropies.* Annual Review of Biomedical Engineering, 2002. **4**(1): p. 175-209.
3. Harrigan, T.P. and R.W. Mann, *State variables for modelling physical aspects of articular cartilage.* International Journal of Solids and Structures, 1987. **23**(9): p. 1205-1218.
4. Fox, A.J.S., A. Bedi, and S.A. Rodeo, *The basic science of articular cartilage: structure, composition, and function.* Sports Health: A Multidisciplinary Approach, 2009. **1**(6): p. 461-468.
5. Torzilli, P.A. and V.C. Mow, *On the fundamental fluid transport mechanisms through normal and pathological articular cartilage during function—I the formulation.* Journal of Biomechanics, 1976. **9**(8): p. 541-552.
6. Freeman, M.A.R., *Adult articular cartilage*. 1973, London: Pitman.
7. Grodzinsky, A.J., et al., *The significance of electromechanical and osmotic forces in the nonequilibrium swelling behavior of articular cartilage in tension.* Journal of biomechanical engineering, 1981. **103**(4): p. 221.
8. Hsieh, J. *Computed tomography: principles, design, artifacts, and recent advances*. 2009. SPIE Bellingham, WA.
9. Brown, M.A. and R.C. Semelka, *MRI: basic principles and applications*. 2011: John Wiley & Sons.
10. Nissi, M.J., et al., *T2 relaxation time mapping reveals age- and species-related diversity of collagen network architecture in articular cartilage.* Osteoarthritis and Cartilage, 2006. **14**(12): p. 1265-1271.
11. Kulmala, K.A.M., et al., *Diffusion coefficients of articular cartilage for different CT and MRI contrast agents.* Medical engineering & physics, 2010. **32**(8): p. 878-882.
12. Winalski, C.S. and P. Rajiah, *The evolution of articular cartilage imaging and its impact on clinical practice.* Skeletal Radiology, 2011. **40**(9): p. 1197-222.
13. Stewart, R.C., et al., *Contrast-enhanced CT with a high-affinity cationic contrast agent for imaging ex vivo bovine, intact ex vivo rabbit, and in vivo rabbit cartilage.* Radiology, 2013. **266**(1): p. 141-150.
14. Binks, D., et al., *Quantitative parametric MRI of articular cartilage: a review of progress and open challenges.* The British journal of radiology, 2013. **86**(1023): p. 20120163.
15. Bowen, R.M., *Incompressible porous media models by use of the theory of mixtures.* International Journal of Engineering Science, 1980. **18**(9): p. 1129-1148.
16. Whitaker, S., *Advances in theory of fluid motion in porous media.* Industrial & engineering chemistry, 1969. **61**(12): p. 14-28.
17. Wang, H., *Theory of linear poroelasticity with applications to geomechanics and hydrogeology*. 2000: Princeton University Press.
18. Biot, M.A., *Mechanics of deformation and acoustic propagation in porous media.* Journal of applied physics, 1962. **33**(4): p. 1482-1498.
19. Mow, V.C., et al., *Biphasic creep and stress relaxation of articular cartilage in compression? Theory and experiments.* Journal of biomechanical engineering, 1980. **102**(1): p. 73.

20. Lai, W.M., J.S. Hou, and V.C. Mow, *A triphasic theory for the swelling and deformation behaviors of articular cartilage.* Journal of biomechanical engineering, 1991. **113**(3): p. 245-258.
21. Wolfram, S., *Cellular automata as models of complexity.* Nature, 1984. **311**(5985): p. 419-424.
22. Brown, T.D. and R.J. Singerman, *Experimental determination of the linear biphasic constitutive coefficients of human fetal proximal femoral chondroepiphysis.* Journal of Biomechanics, 1986. **19**(6): p. 474-474.
23. Chen, A.C., et al., *Depth- and strain-dependent mechanical and electromechanical properties of full-thickness bovine articular cartilage in confined compression.* Journal of Biomechanics, 2001. **34**(1): p. 1-12.
24. Oloyede, A. and N. Broom, *The biomechanics of cartilage load-carriage.* Connective tissue research, 1996. **34**(2): p. 119-143.
25. Darley, V., *Emergent phenomena and complexity.* Artificial Life, 1994. **4**: p. 411-416.
26. Helbing, D., *Agent-based modeling*, in *Social self-organization*. 2012, Springer. p. 25-70.
27. Bonabeau, E., *Agent-Based Modeling: Methods and Techniques for Simulating Human Systems.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(10): p. 7280-7287.
28. Odell, J., *Agents and emergence.* Journal of Object Oriented Programming, 2000. **12**(9): p. 34-36.
29. Kashani, J., et al., *A hybrid agent for simulating porous fluid-saturated structures with complex microscale properties.* Submitted to journal of computational material science, 2016.
30. Rieppo, J., et al., *Spatial determination of water, collagen and proteoglycan contents by Fourier transform infrared imaging and digital densitometry.* Trans Orthop Res Soc, 2004. **29**: p. 1021.
31. Edmonds, B. and R. Meyer, *Simulating Social Complexity*. 2015: Springer.
32. Jurvelin, J.S., M.D. Buschmann, and E.B. Hunziker, *Mechanical anisotropy of the human knee articular cartilage in compression.* Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2003. **217**(3): p. 215-219.

Development and application of the 3D-SPH surface erosion model to simulate multiple and overlapping impacts by angular particles

* X.W. Dong[1], Zengliang Li[1]

[1]College of mechanical and electronic engineering, China University of Petroleum,
66 Changjiang Rd, Huangdao District, Qingdao, China

*Presenting author: dongxw139@163.com
†Corresponding author: dongxw139@163.com

Abstract

This paper presents a 3D smoothed particle hydrodynamics (SPH) modeling procedure to simulate surface erosion of ductile materials subjected to impacts of angular particles. Our SPH model is a meshfree, Lagrangian particle method, based on the standard SPH formulation, and the materials are discretized with a set of particles, in which the targeted ductile material is modeled as an elastic-plastic material, and the angular particles are modeled as a rigid bodys. The present SPH has been improved developing SPH formulations for the Johnson-Cook's plasticity model and failure model to describe plastic behavior and ductile fracture process. The particle interactions between the angular particles and targeted material are taken into account by employing a contact algorithm. Our SPH erosion model is applied to simulate multiple and overlapping impacts of particles on ductile targets. Two modified schemes in terms of density correction and kernel gradient correction are adopted to improve the accuracy of the SPH approximation. Besides, stabilities are ensured using artificial viscosity and density correction, and the numerical oscillations in conventional SPH method are effectively suppressed. The present SPH method and algorithm are then further performed to model solid particle erosion process. The results are compared with available experimental data, and good agreement has been achieved. It is demonstrated that the present SPH procedure is superior to the conventional numerical methods in treating problems of extremely large deformations and with breakages, which usually occurs in the surface erosion process by angular particles.

Keywords: 3D smoothed particle hydrodynamics (SPH); surface erosion; angular particles; multiple and overlapping impacts; kernel gradient correction

1.Introduction

The material removal caused by impacts of particles is generally described as surface erosion by impacts. Impact onductile materials using foreign particlesmay be viewed as either constructive useful engineering technique (e.g. shot blasting[1], abrasive jet[2]) or destructive harmful processes (e.g. impeller erosion[3], pipe erosion[4, 5]. Study of the mechanisms of surface erosion by impacts is helpful in promoting this engineering technique effectively or reducing possible erosive wear.

Material deformation and removal are two main material behaviors involved in surface erosion by impacts. For ductile material, the impacts of the foreign particles cause localized plastic strain[6, **15**] at the contact site on the surface and material is removed when the strain exceed a threshold value[7]. It has been known that material removal does not necessarily occur during the process of foreign particles impacting on ductile targets. It depends on many factors[9**,**5**,19,22**], some of which may individually or synthetically determine the erosion mechanisms, such as particle velocity, angle of attack, particle shape and size of particle, etc. Besides, these erosive factors also affect removal rate of targeted material, i.e., erosion rate. Usually, correlations between erosion rate and erosive factors are obtained through experiments by measuring mass loss or analyzing eroded surface. However, the interaction of these factors makes it difficult to take a close look at the mechanisms experimentally. For example, it is hard to observe the dynamic process of material removal (also called material spallation) or analyze the dependency on some single erosive factor through experiment due to the process is too fast and complex. Computer modeling allows studying the effect of factors separately. And, as a complement to experiment, it can obtain detail informations by controlling the simulation procedure, which can help to reveal the fundamental behaviors involved in the erosion process and predict the erosion performace with respect to different erosive factors.

Early computation models tried to build the correlations between erosion rate and concerned erosive variables[8−**14**,**19**,**21**]. These models simplified the eroded ductile targets as elastic–perfectly plastic materials, of which the yield stress is assumed to be constant. However, the targeted materials would endure high–strain–rate deformation during the short time of real process of surface erosion, especially by hard and angular particle[**15**−**17**]. The yield stress is rate–dependent rather than a constant[**18**, **20**]. Therefore, these models can only obtain correct results after tuning parameters by experiments, which then limited their developments and applications.

Finite element method (FEM) is an effective numerical method in solving completed problems in solid mechanics and has been applied widely to model the surface erosion impacted by spherical particles[7, **23**−**28**]. With appropriate constitutive material models, FEM is capable to simulate the relevant damage phenomena in surface erosion process. These models can be validated by experimental observations or analytical solutions. However, these FEM models mainly focused on predictions of erosion rate quantitatively or analysis of erosion mechanisms qualitatively. It is difficult to observe and reveal the erosion mechanisms for these FE models due to the poorly simulating of dynamic process of material removal. Moreover, actual foreign particles usually have complex geometry shape with angularity. Impacts of angular particle can cause large plastic deformation and rapid material removal, which may result in the heavily distorted elements with poor quality. Thereofore, standard FEM may be not suitable for modelling surface erosion by impacts involving large plastic deformation and material removal. Takaffoli[**12**] proposed a new model for modeling impact of angular particle on OFHC Copper. The model is able to handle these damage behaviors using techniques of adaptive re–meshing and element erosion. Although these techniques overcome the element distortion problems in FE model, they are computationally expensive and may lead to numerical instabilities, especially for multiple overlapping impacts. It can be concluded that these difficulties originate from grid limitation. Almost all the grid–based numerical methods have the difficulties to handle large deformation and material removal.

Smoothed particle hydrodynamics (SPH) is a Lagrangian meshfree particle method. It was initially developed for astrophysical problems[**29**−**31**]. Since its invention, SPH has been extensively applied in the many fields of science and engineering including fluid mechanics and solid mechanics, such as free surface flows [**32**,**33**], viscous flow[**34**,**35**], high velocity impacts[**36**−**38**], geophysical flows[**39**,**40**], etc. As a meshfree method, SPH does not need a mesh or elements to discretize computation domain. Instead of nodes, particles are adopted to carry the field variables such as mass, density, stress, and to approximate the governing equations. These particles have a spatial distance (named as the “smoothing length”), over which their properties are “smoothed” by a kernel function. SPH has great advantages over the grid–based numerical methods to deal with large deformation and material removal due to its adaptive nature[**40**]. Then, SPH method may be a better option to simulation of surface erosion by impacts.

In the past few years, several preliminary applications of SPH method to surface erosion by impacts have been performed andsome encouraging results have been obtained. For

example, Wang and Yang [**41**] investigated multiple impacts spherical particles on Ti–6AL–4V under the scheme of SPH method. The predicted erosion dependency on impact factors agrees well with the analytical and experimental results. However, this study focused on predictions of erosion ratewithoutdemonstrating the advantageous of SPH over conventional numerical method. Takaffoli[**42**] proposed a SPH model to simulate the impact of single angular particles on AL6061–T6 targets. This model implemented Johnson–Cook flow stress and failure model. The dynamic process of material removal caused by impacts was first revealed and the resluts showed that SPH method can account for both material deformation and chip separation. It demonstrated that the SPH method is able to capture the major fundamental dynamic behavior of surface erosion by impacts. However, the traditional SPH method encounters the problem of low accuracy as the accuracy is closely related to the distribution of particles[**43**, **44**]. Also, another crucial aspect is the phenomena of numerical oscillations, which highly affect the numerical stability of the SPH calculation[**38**, **45**].

This paper is to establish a general SPH framework for modeling surface erosion by impacts which comprises reproduction of material behavior in terms of both plastic deformation and material removal and improvement of numerical stability/accuracy. It is then necessary to extend of the SPH method to handle general material constitutive models with plastic flow rules and material failure. In Section 2, the general concepts of the SPH modelling for continuum material are given, and the SPH formulations are presented. Two modified schemes for density correction and kernel gradient correction are then implemented. This paper provides a general approach to resolve the material constitutive relations in SPH, in which small time step ensures the constitutive relations be computed correctly. In Section 3, the model is applied to simulate impacts of diamond particles on *OFHC Copper* and AL6061-T6 surface. Firstly, the SPH model is validated by reproducing the experimental data from published literature. Secondly, the validated model is used to simulate the multiple and overlapping impacts. The impact behaviors related to overlapping impacts are investigated by particularly selecting the impact points of the particles. Thirdly, the multiple and overlapping impacts are simulated by using randomly distributed impact points.

2. SPH surface erosion model

2.1 Model description

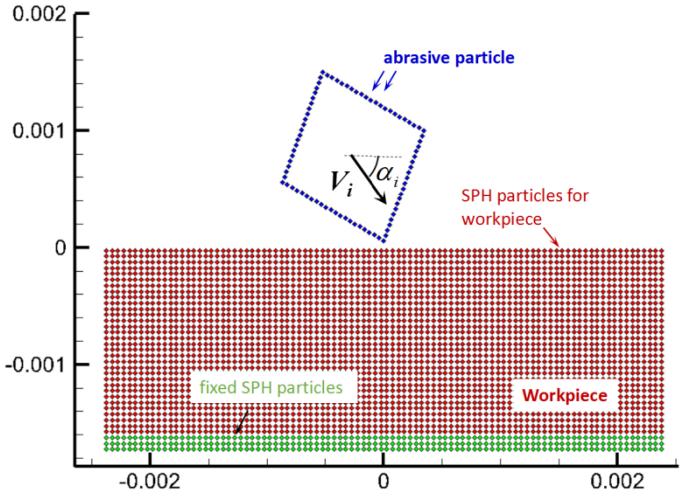


Fig1.Single angular particle impact on targeted material

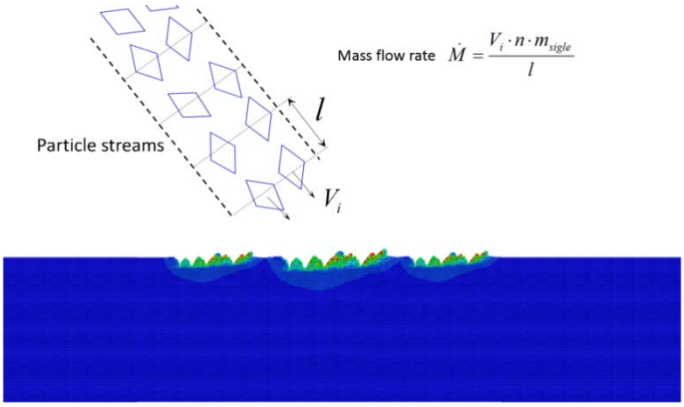


Fig2.Many angular particle impact on targeted material resulting in multiple and overlapping impacts

In this paper, surface erosion by impacts is modeled based on the rigid–plastic theory[57, 59]. Targeted materials which may have large deformation and chip separation are represented and discretized by SPH particles (not the ‘angular particle’), and the angular particle is treated as rigid body assuming it is hard enough to keep non–deformable during erosion process.

Fig1 shows the initial geometry of the two dimensional model of surface erosion by impact, in which angular particle is given a velocity and the targeted material is in steady state with the velocity and stress being zero at the initial time. The bottom particles are held fixed during the simulation to realize displacement boundary condition. Besides, in order to eliminate the effect of model width (L), periodic boundary conditions were prescribed on the side faces of the target block. As shown in Fig.1, the use of periodic boundary conditions assume an infinite plate in width direction. Moreover, the dimensions of the targeted block

(L,W) should also be determined so that the impact simulations would be not affected by edge effects.

The rigid foreign particle, as shown in Fig1, is discretized by one layer of 'surface particles'. The interaction between foreign particle and targeted material is considered by applying a particle contact algorithm developing for meshfree method. The proposed rigid–plastic SPH model allows the simulation of the entire event of particle impact with respect to different erosive factors (eg. impact velocity, angle of attack, particle shape etc.), including dynamic process of interaction between angular particle and targeted material, the particle kinematics in terms of rebound behavior and particle trajectory, the erosion performance.

2.2 Governing equations and SPH formulations

The governing equations of ductile targeted material which consist of mass and momentum conservation equations can be expressed following

$$\frac{D\rho}{Dt} = -\rho \frac{\partial v^{\alpha}}{\partial x^{\alpha}} \tag{1}$$

$$\frac{Dv^{\alpha}}{Dt} = \frac{1}{\rho}\frac{\partial \sigma^{\alpha\beta}}{\partial x^{\beta}} + f^{\alpha} \tag{2}$$

where α and β denote the Cartesian components *x*, *y* with the Einstein convention applied to repeated indices; ρ is the material density; *t* is the time; *v* is the velocity; $\sigma^{\alpha\beta}$ stands for the total stress tensor; the total stress tensor $\sigma^{\alpha\beta}$ has two parts, one is isotropic pressure *p* and the other one is deviatoric shear stress $\tau^{\alpha\beta}$; f^{α} is the component of acceleration caused by external force.

To solve the above governing equations in the SPH framework, one has to approximate these equations using SPH interpolation functions. Since the computation domain has been discretized by particles, the field function at a particle can be obtained simply through summations over all particles within the support domain of the particle using a kernel weighting function, of which the process is so–called particle approximation. The particle approximation for a function and its spatial derivatives at a particle *i*can be expressed in the form as

$$\langle f(\boldsymbol{x}_i)\rangle = \sum_{j=1}^{N} \frac{m_j}{\rho_j} f(\boldsymbol{x}_j) \cdot W(x_i - x_j, h) \tag{3}$$

$$\langle \nabla \cdot f(\boldsymbol{x}_i) \rangle = -\sum_{j=1}^{N} \frac{m_j}{\rho_j} f(\boldsymbol{x}_j) \cdot \nabla_i W_{ij} \tag{4}$$

where $W(x_i - x_j, h)$ the smoothing function or kernel function, and $\nabla_i W_{ij}$is gradient of kernel, $\nabla_i W_{ij} = \frac{\mathrm{x_i - x_j}}{\mathrm{r_{ij}}} \frac{\partial \mathrm{W_{ij}}}{\partial \mathrm{r_{ij}}} = \frac{\mathrm{x_{ij}}}{\mathrm{r_{ij}}} \frac{\partial \mathrm{W_{ij}}}{\partial \mathrm{r_{ij}}}$

According to the continuity equation (Eq.(1) and momentum equation (Eq.2), the governing equations can be expressed as[**46**]

$$\begin{cases} \dfrac{d\rho_i}{dt} = \rho_i \sum_{j=1}^{n} \dfrac{m_j}{\rho_j} v_{ji}^{\beta} \cdot \dfrac{\partial W_{ij}}{\partial x_i^{\beta}} \\ \dfrac{dv_i^{\alpha}}{dt} = \sum_{j=1}^{n} m_j \dfrac{\sigma_i^{\alpha\beta} + \sigma_j^{\alpha\beta}}{\rho_i \rho_j} \cdot \dfrac{\partial W_{ij}}{\partial x_i^{\beta}} + f_i^{\alpha} \\ \dfrac{de}{dt} = \dfrac{1}{2} \sum_{j=1}^{n} m_j \left(\dfrac{P_i + P_j}{\rho_i \rho_j} \right) v_{ij}^{\beta} \cdot \dfrac{\partial W_{ij}}{\partial x_i^{\beta}} + \dfrac{\tau_i^{\alpha\beta} \varepsilon_i^{\alpha\beta}}{\rho_i} \\ \dfrac{dx_i^{\alpha}}{dt} = v_i^{\alpha} \end{cases} \tag{5}$$

where *e* is internal energy, *p* is isotropic pressure, $\tau^{\alpha\beta}$ is deviatoric shear stress, $\varepsilon^{\alpha\beta}$is strain rate tensor.

In SPH, there are many possible choices of the smoothing function *W* in Eq(**3**)–(**5**). The cubic spline function, which was originally proposed by Monaghan and Lattanzio[**47**], has been the most widely used smoothing function in the published SPH literatures since it closely resembles a Gaussian function while having a narrower compact support[**37**]. The cubic spline function is used in this study

$$W_{ij} = \alpha_d \times \begin{cases} \dfrac{2}{3} - q^2 + \dfrac{1}{2} q^3, & 0 \le q < 1 \\ \dfrac{1}{6} (2 - q)^3, & 1 \le q < 2 \end{cases} \tag{6}$$

where α_d is the normalization factor, which is $15/(7\pi \mathrm{h}^2)$ for 2D problem and q is the normalized distance between particle *i* and *j* defined as $q = r_{ij}/h$. r_{ij} is the distance between particle *i* and *j*.

As discussed above, the total stress tensor $\sigma^{\alpha\beta}$ was decomposed into two parts: an volumetric part *p* (named 'pressure' in this paper) and a deviatoric shear stress $\tau^{\alpha\beta}$

$$\sigma^{\alpha\beta} = -p\delta^{\alpha\beta} + \tau^{\alpha\beta} \tag{7}$$

In this paper, the pressure (P) is computed by means of an equation of state (EOS). The Mie–Gruneisen equation, which has been shown to be suitable for solid materials under compressive shock loading[**38**], is employed to describe pressure–volume–energy behavior of ductile materials under particle impact. The pressure is related to density and internal energy in the form of $P = P(\rho, e)$.

For elastic solid of dynamics, the shear stress (τ) can be integrated by time following the incremental formulation of Hooke's law, in which the linear elastic relation between stress and deformation tensors has been derived in time. In order to guarantee the independence from rigid rotations, the Jaumann rate is adopted here with the following elastic constitutive equation as[**46**]

$$\frac{d\tau^{\alpha\beta}}{dt} = 2G\left(\dot{\varepsilon}^{\alpha\beta} - \frac{1}{3}\delta^{\alpha\beta}\dot{\varepsilon}^{\gamma\gamma}\right) + \tau^{\alpha\gamma}\cdot\dot{r}^{\beta\gamma} + \tau^{\gamma\beta}\cdot\dot{r}^{\alpha\gamma} \tag{8}$$

where G is the shear modulus of the concerned material, $\dot{\varepsilon}^{\alpha\beta}$ is the strain rate tensor given by

$$\dot{\varepsilon}^{\alpha\beta} = \frac{1}{2}\left(\frac{\partial v^{\alpha}}{\partial x^{\beta}} + \frac{\partial v^{\beta}}{\partial x^{\alpha}}\right) \tag{9}$$

$\dot{r}^{\alpha\beta}$ is the rotation rate tensor defined through

$$\dot{r}^{\alpha\beta} = \frac{1}{2}\left(\frac{\partial v^{\alpha}}{\partial x^{\beta}} - \frac{\partial v^{\beta}}{\partial x^{\alpha}}\right) \tag{10}$$

The above elastic constitutive relations can be extended to plastic behavior based on the von Mises yield criterion

$$f_y = \frac{\sigma_y}{\sigma_{vM}} < 1 \tag{11}$$

where σ_{vM} is von Mises equivalent stress, σ_y is yield stress. When the von Mises yield criterion is met ($f_y < 1$) the material is considered to be yielded and a plastic behavior is identified. Then the stress tensor is scaled back to the yield surface. For the elastic–perfectly plastic material, the yield stress is considered to be constant. However, the eroded targets can not be treated as elastic–perfectly plastic material due to the yield stress is rate–dependent. In this paper, the Johnson–Cook flow stress model[**55**], which is one of the most popular consititutive models for numerical simulations of impact, is adopted to account for rate–dependent plastic behavior of eroded ductile targets. Johnson–Cook flow stress model is a

purely empirical model and can accout for strain rate hardening and thermal softening. The yield stress in Johnson–Cook model can be written as

$$\sigma_y = \left[A + B\left(\varepsilon_{eff}^p\right)^N\right]\left[1 + Cln\left(\frac{\dot{\varepsilon}_{eff}^p}{\dot{\varepsilon}_0}\right)\right][1 - (T^*)^M] \tag{12}$$

where ε_{eff}^p is the equivalent plastic strain, $\dot{\varepsilon}_{eff}^p$ is the equivalent plastic strain rate, $\dot{\varepsilon}_0$ is reference equivalent plastic strain rate, and *A, B, C, N,* and M are material dependent constants. The normalized temperature (T^*) is given by

$$T^* = \frac{T - T_{ref}}{T_{melt} - T_{ref}} \tag{13}$$

where T_{ref} is reference temperature, T_{melt} is melting temperature of concerned material, and real temperature *T*is calculated by a simplified thermal mechanical coupling equation

$$T = \frac{\varphi W_p}{\rho C_p} + T_{ref} \tag{14}$$

where W_p is the plastic work, φ is the coefficient represents the fraction of the plastic work changing to heat, C_p is the specific heat of concerned material.

In order to model the material removal due to impact of angular particles, it is necessary to employ a failure model. Here, a cumulative–damage failure model, which was also proposed by Johnson and Cook [**56**], is adopted to simulate material removal during the impact process. In the failure model, a parameter *D* is introduced to measure the local damage state and given by

$$D = \sum \frac{\Delta\varepsilon_{eff}^P}{\varepsilon_{failure}} \tag{15}$$

where $\Delta\varepsilon_{eff}^p$ is the increment of equivalent plastic strain occurring during an integration cycle and $\varepsilon_{\text{failure}}$ is the equivalent strain to failure given by

$$\varepsilon_{failure} = [D_1 + D_2 exp(D_3\sigma^*)]\left[1 + D_4 ln\left(\frac{\dot{\varepsilon}_{eff}^p}{\dot{\varepsilon}_0}\right)\right][1 + D_5 T^*] \tag{16}$$

where $D_1 - D_5$ are material constants, $\sigma^*$is defined as the ratio of the mean stress σ_m to the von Mises equivalent stress σ_{vM}.

When parameter *D* is greater than 1, the material failure is considered to occur and the corresponding stress is reduced to zero, which considers the reduction of stress level due to the material failure.

To solve above constitutive relations, i.e. Eq. (**7)**~(**16)**, two steps are proposed. Firstly, the equations should be discretized into the SPH framework for every particle. For example, the strain and rotation rate tensors (Eq.(**9)**, Eq.(**10)**) of a particle are discretized into the SPH formulations given by

$$\varepsilon_i^{\alpha\beta} = \frac{1}{2}\sum_{j=1}^{N}\left(\frac{m_j}{\rho_j}v_{ji}^{\alpha}\frac{\partial W_{ij}}{\partial x_i^{\beta}} + \frac{m_j}{\rho_j}v_{ji}^{\beta}\frac{\partial W_{ij}}{\partial x_i^{\alpha}}\right) \tag{17}$$

$$r_i^{\alpha\beta} = \frac{1}{2}\sum_{j=1}^{N}\left(\frac{m_j}{\rho_j}v_{ji}^{\alpha}\frac{\partial W_{ij}}{\partial x_i^{\beta}} - \frac{m_j}{\rho_j}v_{ji}^{\beta}\frac{\partial W_{ij}}{\partial x_i^{\alpha}}\right) \tag{18}$$

Then, the discretized equations and corresponding variables are interpolated and updated following the updated Lagrangian formulations. Besides, the procedure of stress–rescaling and judgment of failure are performed during every integration cycle following the corresponding criterion we presented above. This paper adopt a very small timestep in the explicitly updated Largrangian procedure, which can reduce the inaccuracy of incrementally updating the stress state following the constitutive relations.

2.3 Corrective terms

In this paper, two modified schemes in terms of density correction and kernel gradient correction are adopted, which have been proved effectively to improve computational accuracy[**33**, **53**, **54**]. For the density correction, we adopt a so–called Moving Least Squares(MLS)[**49**] approach, which is a interpolation scheme on irregularly scattered points. This scheme has been applied successfully by Colagrossi and Landrini[**53**] in SPH dam break simulation. And the linear variation of the density field can be exactly reproduced by using this first order correction scheme to correct the density. Besides, it is found that for the cases with irregular particle distribution a smoother pressure field can be obtained through MLS density correction, which may be helpful in improving the stability in this simulation. Herein, we use MLS approach to correct the density field as

$$\langle\rho_i\rangle = \sum_j \rho_j W_{ij}^{MLS} V_j = \sum_j m_j W_{ij}^{MLS} \tag{19}$$

where the moving–least–square kernel W_j^{MLS} is computed through (for 3D problem)

$$\begin{cases} W_{ij}^{MLS} = \left[\beta_0(\boldsymbol{x}_i) + \beta_1(\boldsymbol{x}_i)x_{ij} + \beta_2(\boldsymbol{x}_i)y_{ij} + \beta_3(\boldsymbol{x}_i)z_{ij}\right]W_{ij} \\ \beta(\boldsymbol{x}_i) = [\beta_0\ \beta_1\ \beta_2\beta_3]^T = A(\boldsymbol{x}_i)[1\ 0\ 0\ 0]^T \\ A(\boldsymbol{x}_i) = \left[\sum_j W_{ij}\begin{bmatrix} 1 & x_{ij} & y_{ij} & z_{ij} \\ x_{ij} & (x_{ij})^2 & x_{ij}.y_{ij} & x_{ij}.z_{ij} \\ y_{ij} & x_{ij}.y_{ij} & (y_{ij})^2 & y_{ij}z_{ij} \\ z_{ij} & x_{ij}.z_{ij} & y_{ij}z_{ij} & (z_{ij})^2 \end{bmatrix}V_j\right]^{-1} \end{cases} \tag{20}$$

where $V_j(=\mathrm{m_j}/\rho_\mathrm{j})$ is the volume of particle j. It should be noted that the density is still integrated by time using continuity equation(Eq. **(1)**) and density correction is applied periodically.

As to kernel gradient correction, the accuracy is restored with the following correction on the kernel gradient by multiplying the original kernel gradient with a matrix $L(r_i)$, which is obtained from Taylor series expansion method [**33**]. In two dimensional spaces, the new kernel gradient can be obtained as follows

$$\nabla_i^{new}W_{ij} = L(r_i)\nabla_i W_{ij} \tag{21}$$

where $x_{ji} = x_j - x_i$, $y_{ji} = y_j - y_i$. It has been proved that the SPH particle approximation scheme with kernel gradient correction is of second order accuracyfor general cases with irregular particle distribution[**33**, **54**].

Then, the standard SPH formulation of momentum equation is rewritten based on our improved method in the following way

$$\frac{dv_i^\alpha}{dt} = \sum_{j=1}^{N} m_j\left[-\left(\frac{P_i + P_j}{\rho_i\rho_j}\right)\delta^{\alpha\beta} + \frac{\sigma_i^{\alpha\beta} + \sigma_j^{\alpha\beta}}{\rho_i\rho_j} + \Pi_{ij}\delta^{\alpha\beta}\right]\frac{\partial W_{ij}^{new}}{\partial x_i^\beta} + f_i^\alpha \tag{22}$$

where the last term(Π_{ij}) between brackets is called artificial viscosity and is used to reduce the unphysical oscillations in the numerical results around the shocked region[**46**]. Of several proposals for artificial viscosity developed so far, the most widely applied is derived by Monaghan[**31**]

$$\Pi_{ij} = \begin{cases} \dfrac{-\alpha\bar{c}_{ij}\mu_{ij} + \beta\mu_{ij}^2}{\bar{\rho}_{ij}} & \vec{V}_{ij}\cdot\vec{x}_{ij} < 0 \\ 0 & \vec{V}_{ij}\cdot\vec{x}_{ij} \geq 0 \end{cases} \tag{23}$$

where $\mu_{ij} = \frac{h_{ij}(\vec{V}_{ij}\cdot\vec{x}_{ij})}{|\vec{x}_{ij}|^2+0.01h_{ij}^2}$, $\bar{c}_{ij} = (c_i + c_j)/2$, $\bar{\rho}_{ij} = (\rho_i + \rho_j)/2$, $h_{ij} = (h_i + h_j)/2$, c is the speed of sound, h is the smoothing length; α, β are constants and should be chosen according to particular applications.

It should be note that for our improved SPH formulations only kernel and kernel gradient are modified. And a field function and its derivatives are approximated separately as the standard SPH method does, which means that there is no need to change the procedure of computation of previous standard SPH. The main structure of SPH code remains unchanged. Therefore, it is relatively convenient to implement above improved SPH formulations.

2.4 Time integration scheme

The discrete SPH formulations are generated for every particle in the form of ordinary differential equations as described above. In order to solve these ordinary differential equations, time integration scheme is used to integrate the field variables. In this work, the Leap Frog (LF) algorithm is adopted due to its low memory requirement and high efficiency. In LF algorithm, the field variables are updated by using the following equations:

$$\rho_{n+1/2} = \rho_{n-1/2} + \left(\frac{d\rho}{dt}\right)_n \cdot \Delta t \tag{24}$$

$$v^{\alpha}_{n+1/2} = v^{\alpha}_{n-1/2} + \left(\frac{dv^{\alpha}}{dt}\right)_n \cdot \Delta t \tag{25}$$

$$\tau^{\alpha\beta}_{n+1/2} = \tau^{\alpha\beta}_{n-1/2} + \left(\frac{d\tau^{\alpha\beta}}{dt}\right)_n \cdot \Delta t \tag{26}$$

$$x^{\alpha}_{n+1} = x^{\alpha}_{n} + v^{\alpha}_{n+1/2} \cdot \Delta t \tag{27}$$

where Δt is time step length.

The stability of the above LF integration scheme is governed by the CFL(Courant–Friedrichs–Levy) contidition

$$\Delta t \leq 0.2\frac{h}{c} \tag{28}$$

where c is sound speed of the concerned material.

According to basic principles presented above, a SPH procedure and code are established based on the SPH code written in Fortran[**46**].

3.Simulation of multiple and overlapping impacts using well-defined particles

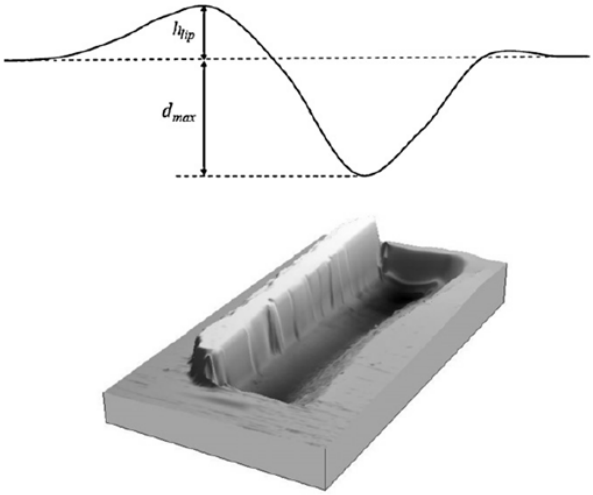


Fig. 1. Typical crater profile resulted by a well-defined angular particle[42]

3.1 Single impact and multiple impact

In this section, we simulate the impact of single angular particle on ductile surface (OFHC Copper and Al6061-T6). The Johnson-Cook parameters of two ductile materials are listed in Table.1. Smulation of single particle helps to validate the numerical model using available experimental results of single impact test. For example, M.Takaffoli and M.Papini[12] studied the single diamond particle impact on OFHC Copper. In their experiment, the launching device was specially designed to realize the adjustment of incident conditions of single particle such as initial orientation (θ_i), impact angle (α_i) and impact velocity (v_i). Figure 1 shows the definitions of incident parameters, geometry parameters and rebound parameters. In this section, we use the same test configuration as the experiment and the predicted results are compared with experimental data, then model validation could be performed.

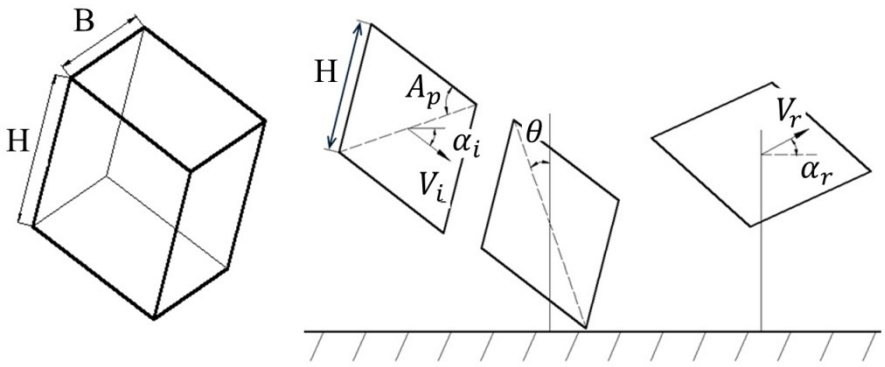


Fig. 2. Geometry, incident, rebound parameters of foreign particle

Table. 1 **Material parameters for Johnson-Cook model**

Material type	A (MPa)	B (MPa)	n	C	m
AL6061-T6	324	114	0.42	0.002	1.34
OFHC Copper	90	292	0.31	0.025	1.09

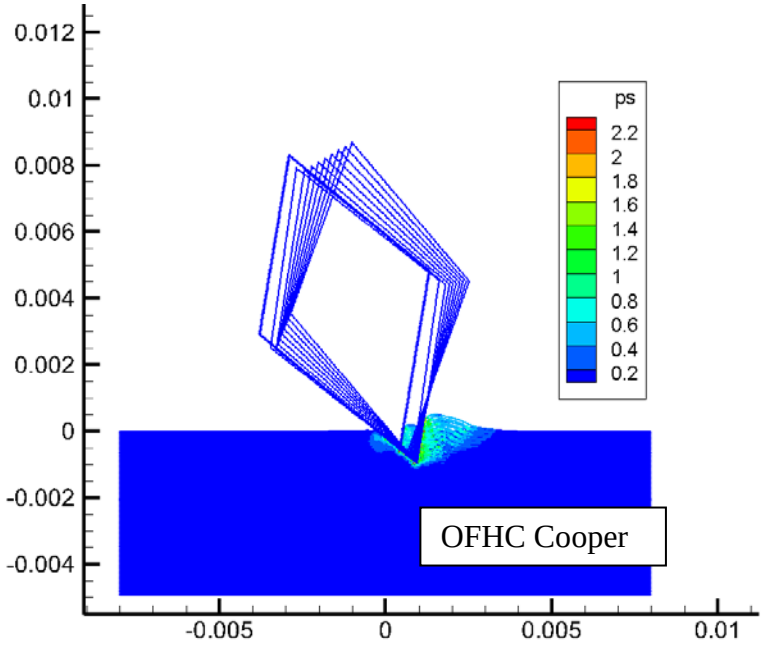


Fig. 3. Dynamic impact process of single angular particle (time interval 10μs)

$v_i = 81m/s\ \alpha_i = 60^\circ, \theta_i = 20^\circ$

Figure 2 shows the simulated impact process of diamond shaped particle on OFHC Copper. The length of the particle size is 5.46mm, the impact velocity is 81m/s. As shown in the figure, the particle impacts on the surface at an oblique impact angle (60°) resulting in an asymmetric erosive crater. In Fig. 3, the predicted crater is compared to measured crater profile, which shows that the predicted crater profile matches well with measured data. It illustrates the model could effectively and accurately obtained reliable results, which ensures further application on multiple and overlapping impact simulation.

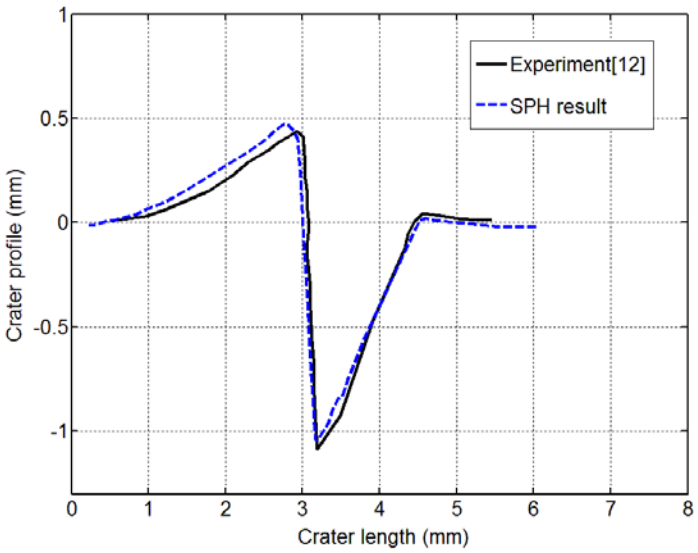


Fig. 4.Single particle impact on OFHC Copper surface at $v_i = 81\text{m/s}\ \alpha_i = 60^\circ, \theta_i = 20^\circ$

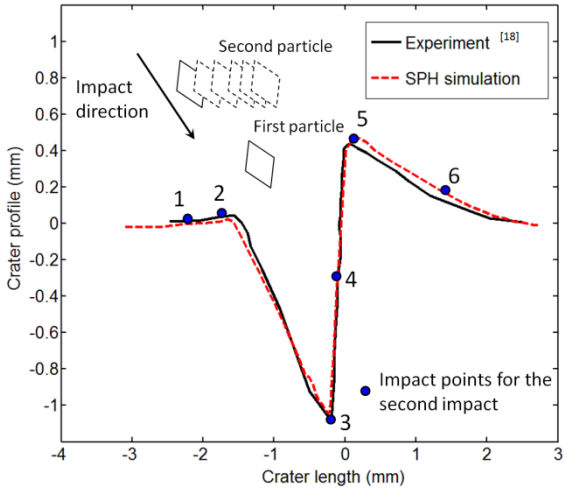


Fig. 5. Second particle impact on previous crater: illustration of different impact points for the second particles $(\theta_i = 20°, \alpha_i = 60°, V_i = 80\text{m/s})$

In surface erosion process, impact on piled-up material is usually considered as the main mechanism of material removal when particles repeatedly impact on the surface. In order to simplify the problem and reveal the fundamental process, two impacts are considered in one simulation. In other words, two angular particles given same incident conditions impact on the surface successively to make sure overlapping impact occur. Then, we investigate the effect of previously resulted crater on the impact behavior and erosion mechanism of subsequent impact. Figure 4 presents the predicted crater profile caused by the first impact and the

corresponding measured profile[12]. It shows good agreement both in crater shape and dimensions.

As shown in Fig.4, six impact points are particularly selected for the second particle along the crater surface resulted from the first impact. Accordingly, six predicted craters of overlapping impacts are obtained and shown in Fig.5. The crater profile of the first impact (black line) is also plotted in the figure for comparison purposes.

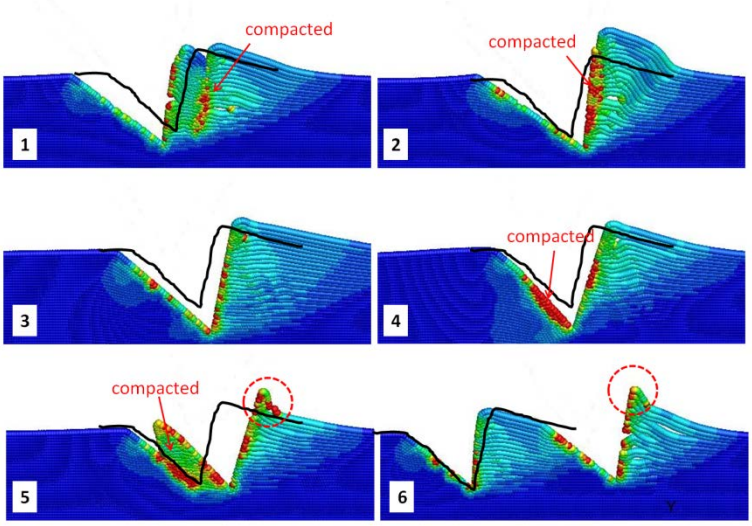


Fig. 6. Erosive craters by overlapping impacts of two particles (black line represents crater profile by the first impact)

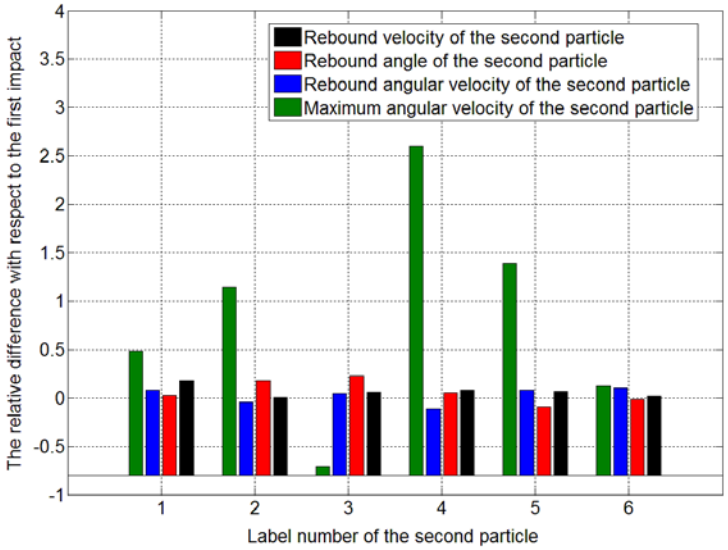


Fig. 7 Illustration of effect of location of impact point on the parameters related to particle motion

Figure 6 illustrates the effect of impact point on the predicted parameters of particle motion including v_r, α_r, ω_r and ω_{max}. It can be clearly seen that the influence of impact

point on the maximum angular velocity (ω_{max}) is bigger than that on any other predicted variables. It means that the existing crater (the first crater) highly influences the initially generated particle rotation, including not only the magnitude but also the rotation direction. For example, for the impact of number 4, the second particle impacts on the inner side of the crater, as shown in Fig.5, the actual θ_i relative to the contact surface is a negative value which results in particle tumbling forward with a far higher ω_{max}(up to 250% higher) than the first impact. Compared with ω_{max}, other variables (v_r, α_r, ω_r) have smaller change when changing the impact point. It should be noted that these three variables are all rebound parameters, of which α_r is mostly heavily affected with the maximum difference up to 25% (Number 3).

3.2 Multiple and overlapping impacts using random impact points

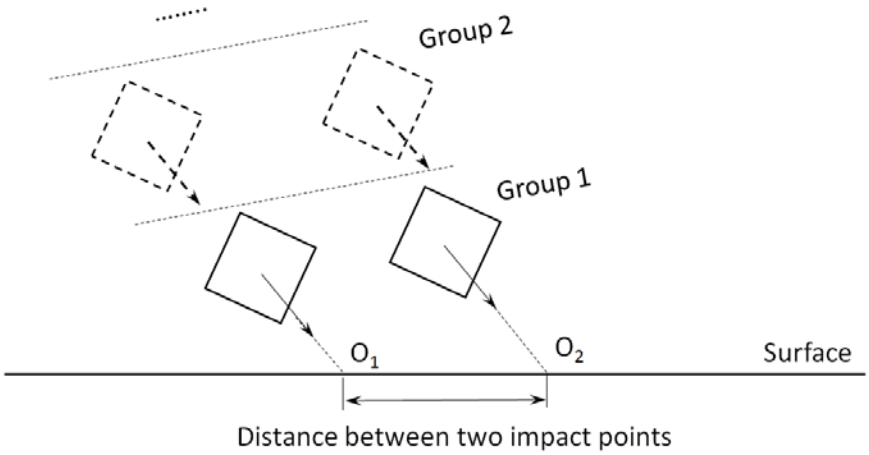


Fig. 8 Group of particles impact on the surface

Real particle erosion system usually involves many particles impact on component surface randomly. In order to reproduce the erosion process as realistic as possible, a random multiple impact model is proposed in this section. As shown in Fig.7, particles are launched to impact on surface group by group, each group contains several particles (two particle in this study). Total particles number is calculated by multiplying group number with particle number in one group. The random characteristic is realized through assigning a random impact point for each particle.

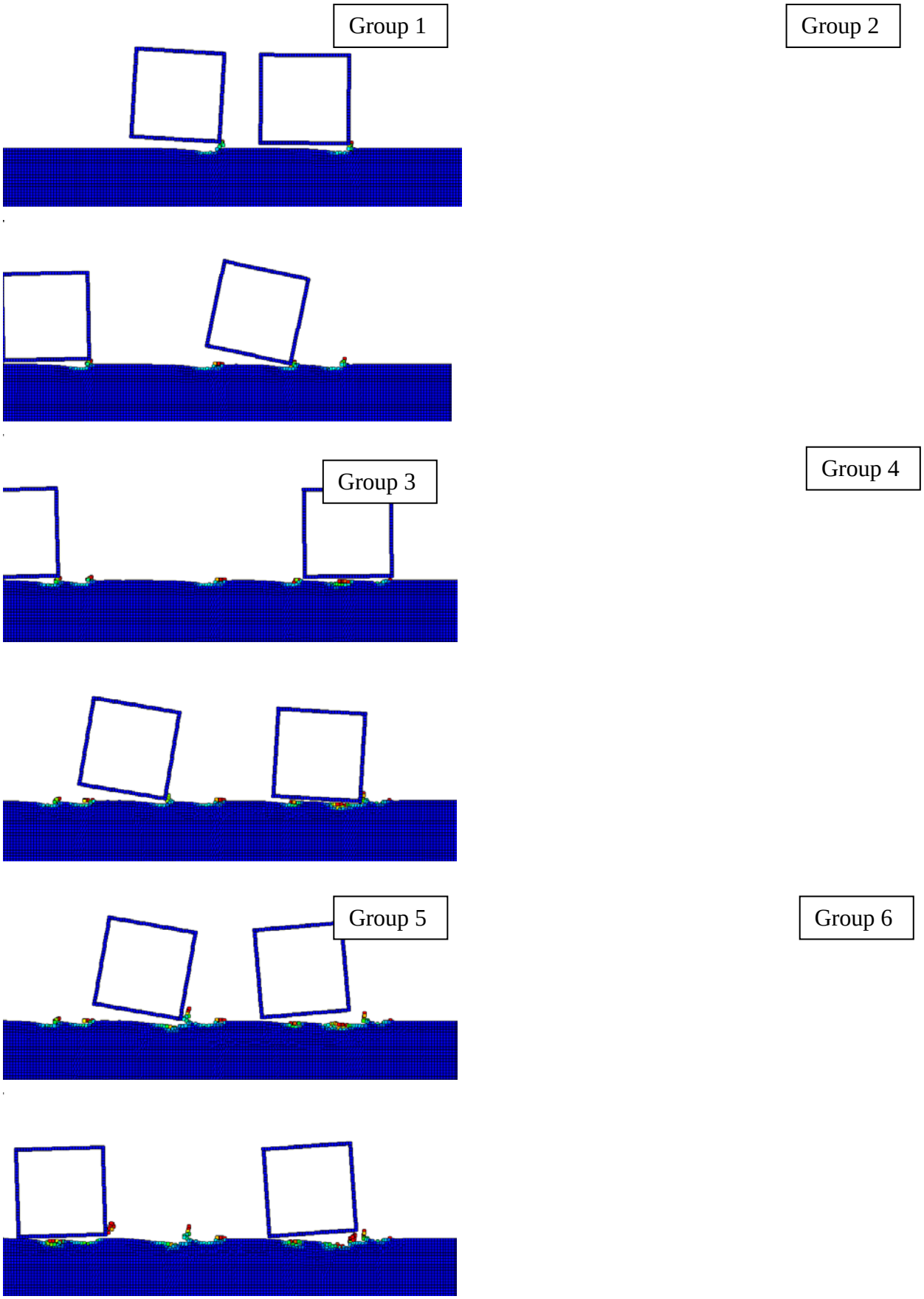


Fig. 9 Six group of particles impact on surface successively

In Fig.8, six group of particles impact on the surface successively. As discussed above, the impact point for each particle in one group is randomly selected. Therefore, overlapping impact may occur when successive particle just impacts on the craters caused by previous particles. Overlapping impacts make the surface materials continuously deform and damage is cumulated until failure occurs, which result in severe deformation on the surface. As shown in Fig.9, overlapping impacts increase the surface roughness. Besides, in the overlapping impact process, chip separation is likely to occur due to the gross failure of the chip materials.

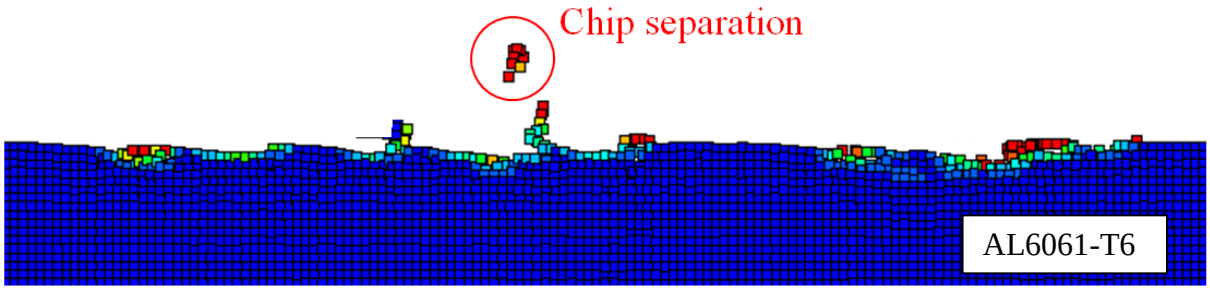


Fig. 10 Surface morphology resulted by 15 particles impact ($\theta_i = 39°$, $\alpha_i = 51°$, $V_i = 60\mathrm{m/s}$)

In Fig.9, same incident conditions ($\theta_i = 39°$, $\alpha_i = 51°$, $V_i = 60\mathrm{m/s}$) are assigned for all 15 particles. Even though incident conditions do not keep constant in real erosion process (such as θ_i), it is reasonable to assume the particles have same incident conditions (especially for impact angle and impact velocity) in order for comparative study.

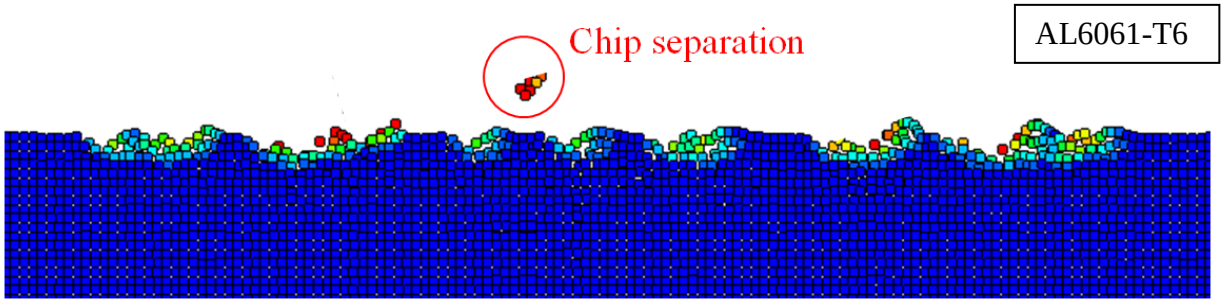


Fig. 11 Surface morphology resulted by 20 particles impact ($\theta_i = 0°$, $\alpha_i = 30°$, $V_i = 60\mathrm{m/s}$)

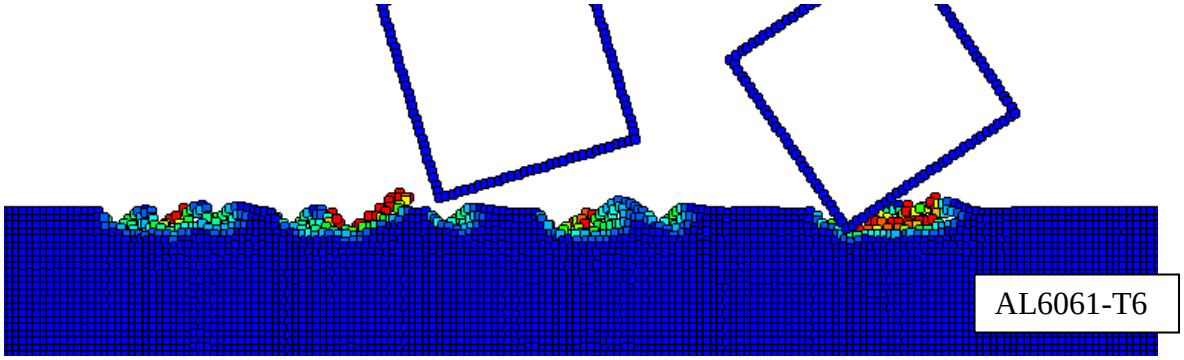
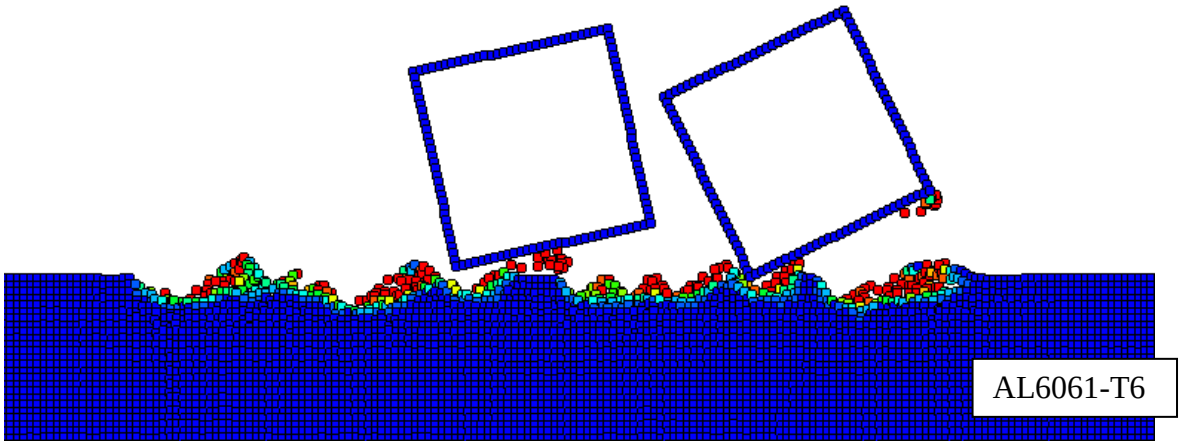
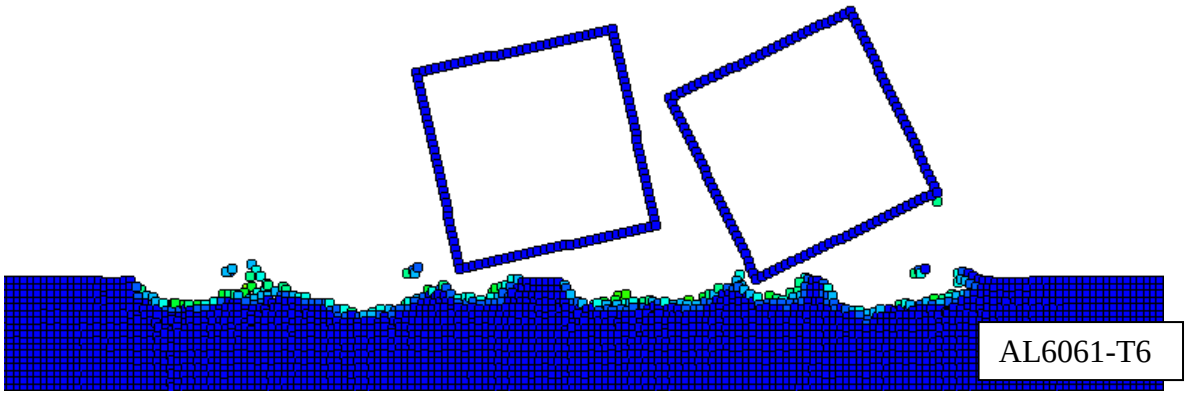


Fig. 12 Surface morphology resulted by 20 particles impact ($\theta_i = 0°, \alpha_i = 40°, V_i = 60\mathrm{m/s}$)

In Fig.10 and Fig.11, 20 particles impact on the surface at $\theta_i = 0°$, at $\alpha_i = 30°$ or $\alpha_i = 40°$ and at $V_i = 60\mathrm{m/s}$. Figure 12 (a) and (b) show 40 particles impact on the surface using the same incident conditions in Fig.11. Overlapping impacts make surface materials fail and the failed materials (SPH particles) are still maintained on the surface due to this study assume hydrostatic pressure could have negative value. The failed materials could be removed for better observation of the broken surface, as shown in Fig.12(b).



(a)



(b)

Fig. 13 Surface morphology resulted by 40 particles impact ($\theta_i = 0°, \alpha_i = 40°, V_i = 60\mathrm{m/s}$)

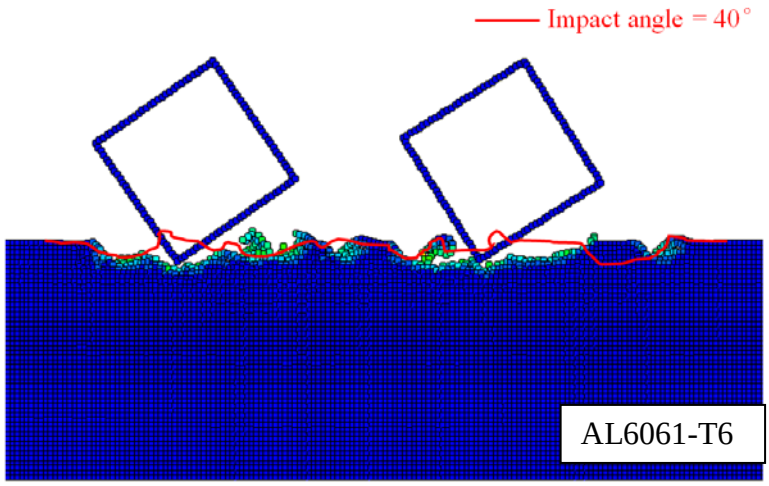


Fig. 14 Comparison of surface morphology between different impact angle

4.Discussion

The SPH has several advantages over element-based numerical methods, such as it can handle large deformation and material removal due to its Lagrangian and adaptive nature; it is relatively easy to incorporate complicated physics. For the present simulation, particle impact on ductile targets usually involves rapid deformation and quick damage, which may result in disordered particle distribution. As described in above sections, the SPH discretization

procedure based on the improved algorithm is employ. Two modified algorithms may help to improve the computational accuracy. Besides, there are many other aspects affecting the accuracy, efficiency and stability of the numerical solutions, such as the choice of the smoothing function, the artificial viscosity, and the neighbouring searching strategy, etc. These aspects degrade the repeatability of numerical test to some extent and make SPH not attractive as some element-based methods. Therefore, it is essential to properly address these issues before applying the method to particular applications.

In this study, the artificial viscosity is introduced into the momentum equation to damp out the undesirable oscillations. For the value of α, Monaghan[32] selected $\alpha = 0.01$ for the free surface flow; Libersky et al.[38] selected $\alpha = 2.5$ for solid mechanics problem. Monaghan also recommended that α close to 1 may be the best choice for most cases. The other term associated with parameter β is devoted to suppress particle interpenetration at high Mach number[40], which dose not have much effect in the present simulation since the velocity (<100m/s) is small compared with the speed of sound ($\sim 10^3$). Our tests give similar results for the value of β between 0 and 2.5, which is the commonly used range recommended by researchers[38, 46]. It has been found that $\alpha = 1.0$ and $\beta = 1.0$ are proper for the present simulation in terms of suppressing numerical oscillations on one hand and leading to less unphysical energy dissipations on the other hand.

Another important aspect affecting the efficiency of the computation is the neighbouring searching procedure. Generally, the easiest way to do this job is to calculate the distance between every two possible neighbouring particles in the computation domain. However, this direct way has low efficiency because it involves a number of interactions on the order of $N \times N$. In the present work, an efficient strategy named linked-list method is adopted. It is suitable for uniformly distributed particles, which is the case for this simulation. For more details on implementing this strategy one can refer to Ref. [49].

5.Conclusion

This paper developed a 3D–SPH model to simulate surface erosion of ductile materials subjected to impacts of angular particles travelling a given velocity. In the model, both the targeted material and the rigid angular particle are discretized bymeshfree particles. Once the rigid–target interaction has been detected, contact forces are imposed to particles close to the interface. In particular, the action of the rigid particle on the target is computed through particles contact algorithm based on penalty force approach. On the contrary, the action on the

rigid particle is computed by summing up all reaction forces from targeted particles which satisfy the action–reaction principle.

The SPH model, thanks to its Lagrangian and adaptive nature, has the great advantage of modeling large deformation and material removal, and does not need any specific treatment for the distorted computational domain. By incorporating the Johnson–Cook plasticity and failure model, the developed SPH model can capture the rate–dependent plastic behavior and damage behavior, which are the key components in erosion mechanisms of ductile material. Further on, chip separation caused by particle impacts is revealed and presented as a dynamic process, which is helpful in taking a close look at the fundamental mechanisms.

To solve the problem of low accuracy in standard SPH method, MLS density correction and kernel gradient correction are implemented into our SPH code. By using the density correction and artificial viscosity together, the stress oscillations in standard SPH model are effectively alleviated. And the unphysical energy dissipation of artificial viscosity is also significantly reduced by appropriately applying the MLS density correction.

The numerical analyses of angular particle impact on AL6026–T6 and OFHC Copper are applied to validate the capability and accuracy of the model.The obtained numerical results clearly demonstrate that the presented SPH model can effectively simulate particle erosion problems. The present work thus forms the basis from which the more realistic multiple particle impact erosion mechanisms can be simulated. However, the present work only simulates solid particle erosion on ductile materials. Future work will be applications in brittle materials using presented method.

References

1 P.Tangestanian, M.Papini, J.K. Spelt,Starch media blast cleaning of artificially aged paint films, Wear, 248 (1–2) (2001), 128–139.

2 Naresh Kumar, Mukul Shukla, Finite element analysis of multi–particle impact on erosion in abrasive water jet machining of titanium alloy. Journal of Computational and Applied Mathematics, 236 (18) (2012), 4600–4610.

3 S.Ariely, A.Khentov, Erosion corrosion of pump impeller of cyclic cooling water system. Engineering Failure Analysis, 13(6) (2006), 925–932.

4 Chong Y.Wong, Christopher Solnordal, Anthony Swallow, Experimental and computational modeling of solid particle erosion in a pipe annular cavity, Wear. 303 (1–2) 2013, 109–129.

5 SubhashN.Shah, Samyak Jain, Coiled tubing erosion during hydraulic fracturing slurry flow, Wear. 264 (3–4) (2008),279–290.

6 W. Zeng, J.M. Larsen, G.R. Liu,Smoothing technique based crystal plasticity finite element modeling of crystalline materials, International Journal of Plasticity. 65 (2015), 250–268.

7 M.S.EITobgy, E.Ng, M.A.Elbestawi,Finite element modeling of erosive wear. International Journal of Machine Tools and Manufacture. 45 (11) (2005), 1337–1346.

8 I.Finnie, The mechanism of erosion of ductile metals. Proceedings of the Third U.S.National Congress of Applied Mechanics. 1958, 527–532.

9 I.Finnie. Erosion of surface by solid particles. Wear.3 (2) 1960, 87–103.

10 J. Bitter, A study of erosion phenomena: part 1, Wear. 6 (1) (1963), 5–21.

11 J. Bitter, A study of erosion phenomena: part 2, Wear. 6(3) (1963), 161–190.

12 M.Takaffoli, M.Papini, Finite element analysis of single impacts of angular particles on ductile targets, Wear. 267 (1–4) (2009), 144–151.

13 M. Hashish, Modified model for erosion, Seventh International Conference on Erosion by Liquid and Solid Impact, Cambridge, England, 1987, 461–480.

14 J. Neilson, A. Gilchrist, Erosion by a stream of solid particles. Wear, 11 (2) (1968), 111–122.

15 Y.I. Oka, K.Nagahashi, Measurements of plastic strain around indentations caused by the impact of round and angular particles, and the origin of erosion, Wear.254 (12) (2003) ,1267–1275.

16 I.Hutchings, R. Winter, Particle erosion of ductile metals: a mechanism of material removal, Wear.27 (1) (1974) ,121–128.

17 H.M. Hawthorne, Y.Xie, S.K. Yick, A study of single particle–target surface interactions along a specimen in the Coriolis slurry erosion tester, Wear.253 (3–4) (2002) ,403–410.

18 G.Sundarajan, A comprehensive model for the solid particle erosion of ductile materials, Wear.149 (1–2) (1991) ,111–127

19 G.Sundararajan, P.G.Shewmon,The oblique impact of a hard ball against ductile, semi–infinite, target materials–experiment and analysis,International Journal of Impact Engineering. 6 (1) 1987, 3–22

20 Y.Tirupataiah,B.Venkataraman, G.Sundararajan,The nature of the elastic rebound of a hard ball impacting on ductile, metallic target materials, Material Science and Engineering: A. 124 (2) 1990, 133–140.

21 Rickerby DG, MacMillan NH,On the oblique impact of a rigid sphere against a rigid–plastic solid, Int. J.Mech.Sci..22 (8) (1980), 491–494.

22 G.Sundararajan, The depth of the plastic deformation beneath eroded surfaces: the influence of impact angle and velocity, particle shape and material properties. Wear. 149 (1–2) (1991), 129–153.

23 K.Shimizu, T.Noguchi, H.Seitoh, M.Okada, Y.Matsubara, FEM analysis of erosive wear, Wear. 250 (1–12) (2001), 779–784.

24 M.Junkar,B.Jurisevic, M.Fajdiga, M.Grah, Finite element analysis of single particle impact in abrasive water jet machining, International Journal of Impact Engineering. 32(7)(2006),1095–1112.

25 Griffin, A. Daadbin, S. Datta, The development of a three–dimensional finite element model for solid particle erosion on an alumina scale/MA956 substrate, Wear. 256 (9–10) (2004), 900–906.

26 P.J. Woytowitz, R.H. Richman, Modeling of damage from multiple impacts by spherical particles, Wear. 233–235 (1999), 120–133.

27 J.F. Molinari, M. Ortiz, A study of solid–particle erosion of metallic targets, International Journal of Impact Engineering. 27 (4) (2002), 347–358.

28 B. Zouari, M. Touratier, Simulation of organic coating removal by particle impact, Wear. 253 (3–4) (2002), 488–497.

29 Lucy. L. B. , A numerical approach to testing the fission hypothesis, Astronomical Journal. 82 (1977), 1013–1024.

30 Gingold. R. A, Monaghan. J. J ,Smoothed particle hydrodynamics– theory and application to non–spherical stars, Monthly Notices of the Royal Astronomical Society. 181 (1977), 375–389.

31 J. J. Monaghan, Smoothed particle hydrodynamics, Annual Review of astronomy and astrophysics. Astronomy and Astrophysics. 30 (1992), 543–574.

32 J. J. Monaghan, Simulating free surface flows with SPH, Journal of Computational Physics. 110 (2) (1994), 399–406.

33 J.R. Shao, H.Q. Li, G.R. Liu, M.B. Liu, An improved SPH method for modeling liquid sloshing dynamics,Computers & Structures, 100–101 (2012), 18–26

34 H. Takeda, S.M. Miyama, M. Sekiya, Numerical simulation of viscous flow by smoothed particle hydrodynamics, Progress of Theoretical Physics, 92 (5)(1994), 939–960.

35 J.P. Morris, P.J. Fox, Yi Zhu, Modeling low Reynolds number flows using SPH, Journal of Computational Physics.136 (1) (1997), 214–226.

36 G.R. Johnson, Robert A. Stryk, Stephen R. Beissel, SPH for high velocity impact computations, Computer Methods in Applied Mechanics and Engineering, 139 (1-4) (1996), 347–373.

37 M.B.Liu, G.R. Liu, Smoothed particle hydrodynamics(SPH): an overview and recent development, Arch Comput Methods Eng (2010)17: 25–76.

38 P.W.Randles, L.D. Libersky, Smoothed particle hydrodynamics: some recent improvement and applications,Computer Methods in Applied Mechanics and Engineering. 139 (1-4) (1996), 375–408.

39 Sonddong Shao, Edmond Y.M. Lo, Incompressible SPH method for simulating Newtonian and non–Newtonian flows with a free surface, Advances in Water Resources. 26 (7)(2003), 787–800.

40 Ha H.Bui, Ryoichi Fukagawa, KazunariSako, ShintaroOhno. Lagrangianmeshfree particles method (SPH) for large deformation and failure flows of geomaterial using elastic–plastic soil constitutive model, International Journal for Numerical and Analytical Methods in Geomechanics.32 (12) (2008), 1537–1570.

41 Yu–Fei Wang, Zhen–Guo Yang,A coupled finite element and meshfree analysis of erosive wear, Tribology International. 42 (2) (2009), 373–377.

42 M.Takaffoli, M.Papini. Material deformation and removal due to single particle impacts on ductile materials using smoothed particle hydrodynamics, Wear. (274–275) (2012), 50–59.

43 T. Belytschko, Y. Krongauz, J. Dolbow, C. Gerlach,On the completenesss of meshfree particle methods. International Journal for Numerical Methods in Engineering.43 (5) (1998), 785–819.

44 M.B. Liu, G.R. Liu, Restoring particle consistency in smoothed particle hydrodynamics, Applied Numerical Mathematics. 56 (1) (2006), 19–36.

45 R. Fatehi, M. T. Manzari, A remedy for numerical oscillations in weakly compressible smoothed particle hydrodynamics, International Journal for Numerical Methods in Fluids. 67 (9) (2011), 1100–1114.

46 G.R. Liu, M.B. Liu, Smoothed particle hydrodynamics: A meshfree particle method. World Scicentific: Singapore, 2004

47 Monaghan, J. J. ,Lattanzio. J. C., A refined particle method for astrophysical problems, Astronomy and Astrophysics, 149 (1) (1985),135–143.

48 M.B. Liu, W. P. Xie, G.R. Liu, Modeling incompressible flows using a finite particle method,Applied Mathematical Modelling . 29 (12) (2005), 1252–1270

49 G.A.Dilts, Moving–least–square–particles hydrodynamics –I. Consistency and stability, International Journal for Numerical Methods in Engineering. 44 (8) (1999), 1115–1155.

50 J. Bonet, S. Kulasegaram, Correction and stabilization of smooth particle hydrodynamics methods with applications in metal forming simulations, International Journal for Numerical Methods in Engineering. 47 (6) (2000), 1189–1214.

51 J.K. Chen, J.E. Beraun, T.C. Carney, A corrective smoothed particle method for boundary value problems in heat conduction, International Journal for Numerical Methods in Engineering. 46 (2) (1999), 231–252

52 J.K. Chen, J.E. Beraun,A generalized smoothed particle hydrodynamics method for nonlinear dynamic problems,Computer Methods in Applied Mechanics and Engineering. 190 (1–2) (2000), 225–239

53 Andrea Colagrossi, Maurizio Landrini,Numerical simulation of interfacial flows by smoothed particle hydrodynamics, Journal of Computational Physics. 191 (2) (2003), 448–475.

54 Man Hu, M.B Liu, M.W. Xie, G.R. Liu, Three–dimensional Run–out Analysis and Prediction of Flow–like Landslides using Smoothed Particle Hydrodynamics, Environmental Earth Science. August (2014).

55 Gordon. R. Johnson and William H. Cook, A constitutive model and data for metals subjected to large strains, high strain rates, and high temperatures. Proc. 7th Int. Symp.On Ballistics, Hague, Netherlands, April (1983).

56 Gordon. R. Johnson and William H. Cook,Fracture characteristics of three metals subjects to various strain, strain rate, temperatures and pressures, Engineering Fracture Mechanics, 21 (1)1985, 31–48.

57 M.Papini, J.K. Spelt,Impact of rigid angular particles with fully–plastic targets. Part I. Analysis,International Journal of Mechanical Sciences. 42 (5) (2000), 991–1006.

58 M.Papini, J.K.Spelt. Impact of rigid angular particles with fully–plastic targets, part II: Parametric study of erosion phenomena, International Journal of Mechanical Sciences. 42 (5)(2000), 1007–1025.

59 M.Papini, S.Dhar,Experimental verification of a model of erosion due to the impact of rigid single angular particles on fully plastic targets,International Journal of Mechanical Sciences. 48 (5) (2006), 469–482.

60 D.R.Lesuer, G. J. Kay, M. M. LeBlanc, Modeling large strain, high rate deformation in metals, in: Third Biennial Tri–Laboratory Engineering Conference Modeling and Simulation, Pleasanton, CA, November 3-5, 1999.

61 M. Meo, R. Vignevic, Finite element analysis of residual stress induced by shot peening process, Advances in Engineering Software, 34 (9) (2003), 569–575.

62 M.A.S Torres, H.J.CVoorwald, An evaluation of shot peening, residual stress and stress relaxation on the fatigue life of AISI 4340 steel, International Journal of Fatigue, 24 (8) (2002), 877–886.

Runge-Kutta discontinuous Galerkin method in solving compressible two-medium flow

***H. T. Lu[1], †N. Zhao[1]**

[1]College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, P.R. China.

*Presenting author: lhtgkzy@126.com
†Corresponding author: nzhao2000@hotmail.com

Abstract

In this paper, the Runge-Kutta discontinuous Galerkin method is used in solving compressible two-medium flow. The material interface is explicitly tracked by the front tracking method and the interface boundary conditions are defined via the real ghost fluid method. Several numerical examples are presented to show the accuracy and capacity of this method. It is found that the mass errors are smaller compared to the results obtained by the same order accurate finite difference method.

Keywords: Runge-Kutta discontinuous Galerkin method, front tracking method, real ghost fluid method, mass errors.

Introduction

One major difficulty in solving compressible two-medium flow is how to treat the material interface accurately. The front tracking method [3] provides an explicit way to track the moving interface and a sharp interface boundary is maintained during the computation. The ghost fluid method (GFM) [2] introduced by Fedkiw et al. presents a simple and flexible way of treating the material interface. However, when the pressure or the velocity experiences a large jump across the interface, the GFM can lead to inaccuracy or even incorrect solution. To better consider the effect of wave interaction and material property, the real ghost fluid method (RGFM) [9] is proposed to update the real fluid states and obtain the ghost fluid states by defining a Riemann problem at the interface. With these ghost fluid states, the mediums can be solved separately as if it is in a single medium.

In recent years, the Runge-Kutta discontinuous Galerkin (RKDG) method [1] performed very well and has been broadly applied to the simulation of single medium flow. For the RKDG method, the higher accuracy is easily obtained in smooth region and we can get the numerical solution everywhere from the solution polynomials. In many earlier works, the basic scheme used to solve the compressible multimedium flow is usually finite difference method [4]. For higher order accurate finite difference method, more ghost fluid states across the interface are solved. Since the geometrical information far from the interface is not solved precisely by the front tracking method, the corresponding ghost fluid states are less accurate especially for the complex interface in the later stage evolution [5][6]. However, due to the good compactness of the RKDG method, we only need to define the ghost fluid states in the ghost fluid cells which have the common edges with the real fluid cells. This is very simple but also favorable. The intention of this work is to apply the RKDG method in the simulation of compressible two-medium flow and compare the mass errors obtained by the same order accurate finite difference method. The material interface is explicitly tracked by several connected marker points and the RGFM is used to define the interface boundary conditions. A Riemann problem is constructed in the normal direction of each marker point, and the Riemann solutions are used to advance the interface and obtain the ghost fluid states directly.

Equations and interface treatment

Governing equations

The two-dimensional hyperbolic conservation laws can be written as follows:

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{U}) = 0 \tag{1}$$

where $\mathbf{U} = [\rho, \rho u, \rho v, E]^T$, $\mathbf{F}(\mathbf{U}) = [\mathbf{F}_1(\mathbf{U}), \mathbf{F}_2(\mathbf{U})]$, $\mathbf{F}_1(\mathbf{U}) = [\rho u, \rho u^2 + p, \rho uv, (E+p)u]^T$, $\mathbf{F}_2(\mathbf{U}) = [\rho v, \rho uv, \rho v^2 + p, (E+p)v]^T$. Here ρ is the density, u and v are the velocities, p is the pressure, E is the total energy per unit volume. The total energy is given as:

$$E = \rho e + \rho(u^2 + v^2)/2 \tag{2}$$

where e is the internal energy per unit mass. The stiffened gas equation of state is used:

$$p = (\gamma - 1)\rho e - \gamma B \tag{3}$$

here γ and B are characteristic parameters of material and can be treated as constants. For the ideal gas γ represents the ratio of the specific heats and B is zero.

Interface tracking

As indicated in Fig. 1, medium 1 and medium 2 are separated by the material interface. The marker points are represented by the intersections of the interface and the grid lines. $\vec{N}$ is the normal vector and $\vec{T}$ is the tangential vector of each marker point. Point $A(x_A, y_A)$ and point $B(x_B, y_B)$ are obtained by the same distance Δn [3] from the marker point $P(x_P, y_P)$:

$$\Delta n = \left[\left(\frac{N_{Px}}{\Delta x} \right)^2 + \left(\frac{N_{Py}}{\Delta y} \right)^2 \right]^{-\frac{1}{2}} \tag{4}$$

$$\begin{aligned} x_A &= x_P - \Delta n \cdot N_{Px},\ y_A = y_P - \Delta n \cdot N_{Py} \\ x_B &= x_P + \Delta n \cdot N_{Px},\ y_B = y_P + \Delta n \cdot N_{Py} \end{aligned} \tag{5}$$

where $\vec{N}_P = (N_{Px}, N_{Py})$ is the unit normal vector of the marker point P and Δx and Δy are the cell sizes. A Riemann problem is constructed at the marker point P with the initial conditions:

$$\mathbf{U}_0 = \begin{cases} \mathbf{U}_A \\ \mathbf{U}_B \end{cases} \tag{6}$$

where $\mathbf{U}_A$ and $\mathbf{U}_B$ are the fluid states at point A and point B and can be solved from the solution polynomials directly in the RKDG method [1]. An approximate Riemann problem solver (ARPS) based on a two shock structure can be employed to obtain the Riemann solutions. We denote the Riemann solutions by $\mathbf{R}_P = [\rho_I^L, \rho_I^R, u_I^N, p_I]^T$, where the subscript "$I$" refers to the interface, and the superscript "L" and "R" denote the left and right side of the interface, respectively. The tangential velocity of the marker point P depends on the sign of the normal velocity and is defined as:

$$v_I^T = \begin{cases} v_A^T, & \text{if } u_I^N \geq 0 \\ v_B^T, & \text{otherwise} \end{cases} \tag{7}$$

where v_A^T and v_B^T are the tangential velocity of point A and point B, respectively. After the velocity of each marker point has been solved, its new position is updated simultaneously:

$$\begin{aligned}\vec{x}_f^{(1)} &= \vec{x}_f^n + \Delta t \cdot \vec{v}_f(\vec{x}_f^n) \\ \vec{x}_f^{(2)} &= \frac{3}{4}\vec{x}_f^n + \frac{1}{4}\vec{x}_f^{(1)} + \frac{1}{4}\Delta t \cdot \vec{v}_f(\vec{x}_f^{(1)}) \\ \vec{x}_f^{n+1} &= \frac{1}{3}\vec{x}_f^n + \frac{2}{3}\vec{x}_f^{(2)} + \frac{2}{3}\Delta t \cdot \vec{v}_f(\vec{x}_f^{(2)})\end{aligned} \tag{8}$$

where $\vec{x}_f^n$ and $\vec{x}_f^{n+1}$ are the positions of the interface at time t^n and t^{n+1}, respectively. $\vec{v}_f$ is the interface velocity, and Δt is the time step.

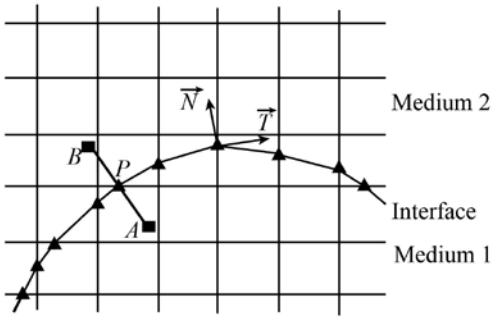


Figure 1. Construct the Riemann problem

RGFM

Since the Riemann problem has been solved at the marker point in the interface tracking, the Riemann solutions can be used directly to update the real fluid states and obtain the ghost fluid states. As shown in Fig. 2, points *R*, *S*, *P* and *Q* are the marker points near the grid cell *A*, $\vec{N}_P$ is the normal vector of the marker point *P* and $\vec{N}_A$ is the normal vector of the grid cell *A*. The flow states at the cell *A* can be updated by the marker point nearby. The marker point *P* is selected if the angle between $\vec{N}_P$ and $\vec{N}_A$ is the minimum compared with other marker points. We project the Riemann solutions at the marker point *P* to the base function space to obtain the average values in cell *A* while the tangential velocity in cell *A* remains unchanged. It is similar for other real fluid cells adjacent to the interface. The ghost fluid states are obtained by solving the advection equation:

$$\frac{\partial \phi}{\partial t} \pm \vec{N} \cdot \nabla \phi = 0 \tag{9}$$

where ϕ is the density, the normal velocity, the tangential velocity and the pressure, $\vec{N}$ is the unit normal vector of the ghost cells.

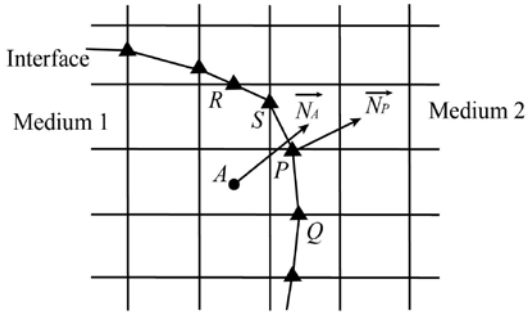


Figure 2. Update the fluid states adjacent to the interface

Numerical examples

In this section, several two dimensional compressible two-medium flow problems are simulated on uniform Cartesian meshes. The governing equations for each medium are solved by the P^2 (third-order accurate) RKDG method and the TVB limiter constant [1] is taken as 0.1. The time integration is solved by a third-order TVD Runge-Kutta scheme. The RKDG method combined with the front tracking method is named as RKDG-FT method for convenience.

Shock bubble interaction

The computational domain is shown in Fig. 3 and the geometrical parameters are: a=50 mm, b=25 mm, c=100 mm, d=325 mm, e=44.5 mm. A shock wave propagates to the left and hits a helium bubble with a Mach number of 1.22. Only the upper half domain is computed since the flow field is symmetric about the center axis. On the left and right boundaries, nonreflecting boundary condition is used and the upper boundary is treated as slip-wall. The speed of sound and the diameter of bubble are used for nondimensionalization. The computational domain is divided into 650×89 mesh cells. The initial conditions are: $\rho = 1, u = 0, v = 0, p = 1/1.4, \gamma = 1.4,$ for pre-shocked air, $\rho = 1.3764, u = -0.3336, v = 0, p = 1.5698/1.4,$ for post-shocked air, $\rho = 0.1819, u = 0, v = 0,$ $p = 1/1.4, \gamma = 1.648,$ for helium. The time histories of density field are shown in Fig. 4. The evolution of the bubble shape and the refracted shock wave can be seen clearly. In Fig. 5, it shows the space-time diagram for three characteristic points (Jet, Downstream, Upstream shown in the figure) with earlier results from [8]. In general, these results are in a relatively good agreement. To make quantitative comparisons with the finite difference method, here we replace the RKDG method by the third order accurate weighted essentially non-oscillatory (WENO) method and keep everything else unchanged in the code [4]. The WENO method combined with the front tracking method is named as WENO-FT method for convenience. The relative mass error of helium bubble is computed and shown in Fig. 6. It is found that the relative mass errors are limited within 7% before the helium bubble collapses for both methods. The general trends of the relative mass errors with time are similar but the error caused by the RKDG-FT method is much smaller.

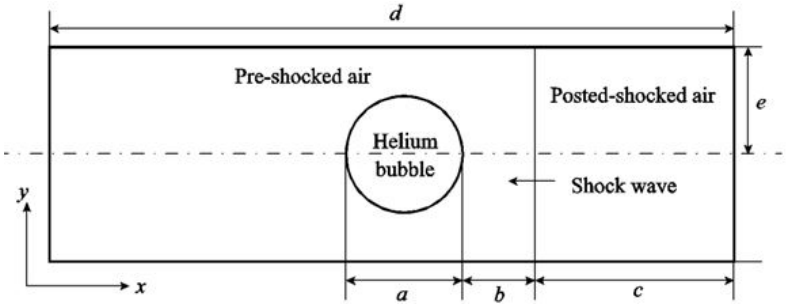


Figure 3. A schematic of computational domain (not to scale)

(a) t=102μsec

(b) t=427μsec

(c) t=674μsec

Figure 4. The evolution of density field (60 equally spaced density contours from 0.1 to 1.6)

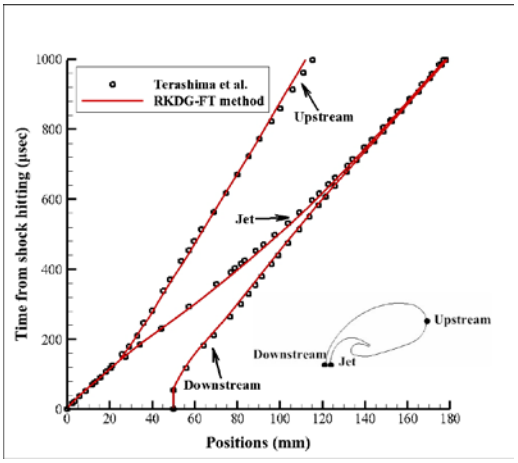


Figure 5. Space-time diagrams for three characteristic interface points

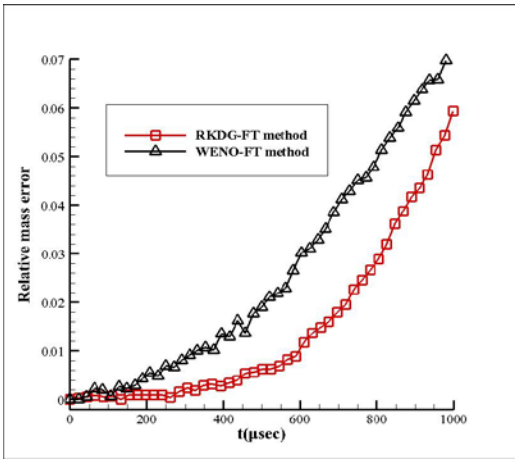


Figure 6. Comparison of relative mass error of helium bubble

Richtmyer-Meshkov instability

This example consists of two simulations of problems with gas-gas and gas-liquid interfaces. As indicated in Fig. 7, a computational domain of [0,4]×[0,0.5] is used and the initial location of the interface is represented by: $x = 2.9 - 0.1\sin(2\pi(y + 0.25)),\quad 0 < y < 0.5$. The upper and lower boundaries are periodic and the nonreflecting boundary condition is applied at the left and right boundaries. The computational domain is divided into 1000×125 mesh cells. The first one is a gas-

gas interface. At x=3.2 there is a planar shock wave with Mach number 1.24 in air propagating from the right to the left of the SF_6-air interface. The initial conditions are: $\rho = 5.04, u = 0, v = 0, p = 1,$ $\gamma = 1.093,$ for SF_6, $\rho = 1, u = 0, v = 0, p = 1, \gamma = 1.4,$ for pre-shocked air, $\rho = 1.411, u = -0.39,$ $v = 0, p = 1.628, \gamma = 1.4,$ for post-shocked air. The flow evolution in the density field is presented in Fig. 8. The interface is accelerated by a shock wave coming from the light-fluid to the heavy-fluid region. Fig. 9 presents the time evolution of the location of the spike and the leading edge of the bubble along with the results in [8]. It shows that these results are almost identical. The relative mass error of the SF_6 medium is shown in Fig. 10 before the shock wave transmits to the left boundary in order to make comparisons between the RKDG-FT method and the WENO-FT method. It is found that these errors are similar at the initial stage. Later, the error by the WENO-FT method increases quickly while the error curve by the RKDG-FT method is much smoother.

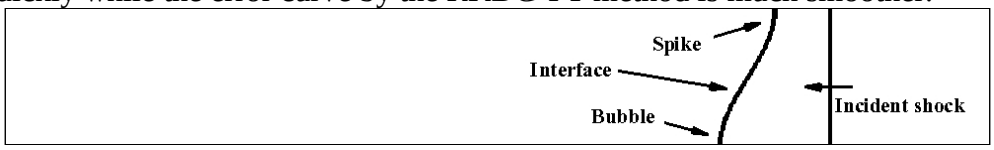
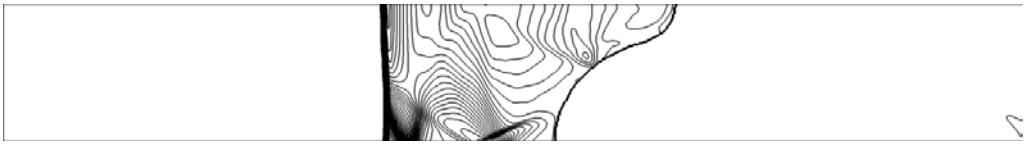
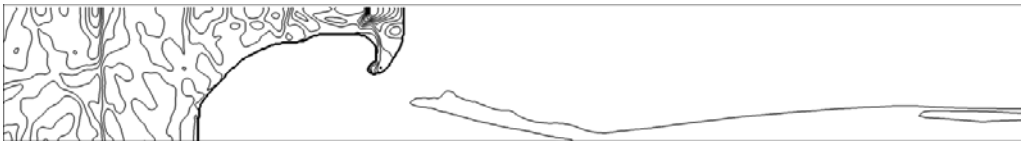


Figure 7. A schematic of flow field at t=0



(a) t=2.3099

(b) t=4.6062



(c) t=6.9045

Figure 8. Density field (230 equally spaced density contours from 0.5 to 9.5)

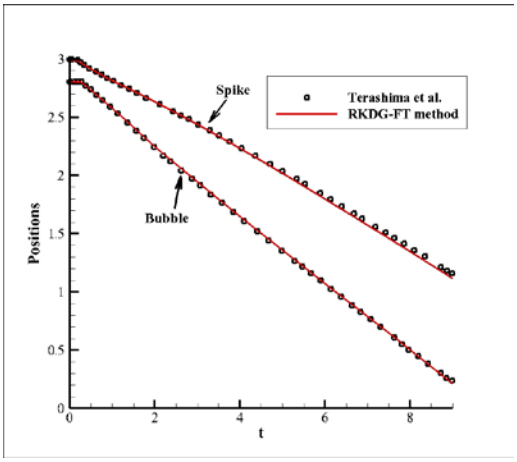


Figure 9. Comparison on time histories of characteristic positions

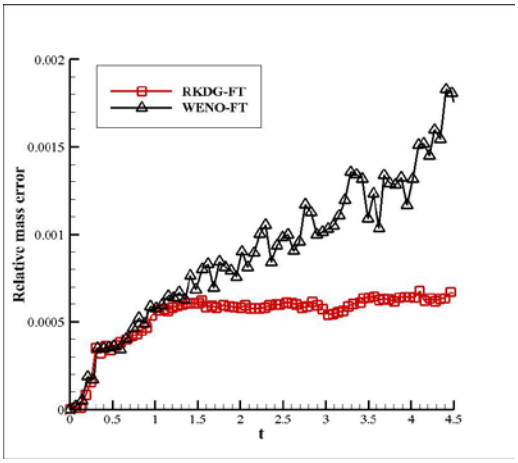
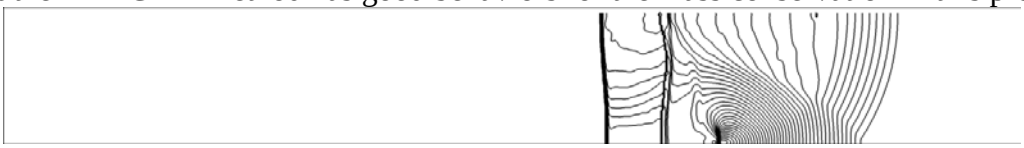


Figure 10. Comparison of relative mass error of SF_6

The second one is a gas-liquid interface that is interacting with a Mach number 1.95 shock wave at x=3.025 initially in liquid. The initial conditions are: $\rho = 1, u = 0, v = 0, p = 1, \gamma = 1.4,$ for air, $\rho = 5, u = 0, v = 0, p = 1, \gamma = 4, B = 1,$ for pre-shocked liquid, $\rho = 7.093, u = -0.7288, v = 0, p = 10,$ $\gamma = 4, B = 1,$ for post-shocked liquid. The density field is shown in Fig. 11 where the complex wave structure is once again presented and is relatively well captured. To check the correctness of the results, in Fig. 12 we compare the distributions of density and pressure along y=0.5 at t=0.5 with the results ('$\circ$') in [7]. Good agreement of the solutions is clearly observed. Similar to the gas-gas interface, the relative mass error of the air medium is measured and shown in Fig. 13. The error by the WENO-FT method increases quickly after the shock wave transmits into the air medium and it shows that the RKDG-FT method has good behaviors for the mass conservation in this problem.



(a) t =0.3

(b) t =0.5

(c) t =1.0

Figure 11. Density field (100 equally spaced density contours from 0.5 to 7.5)

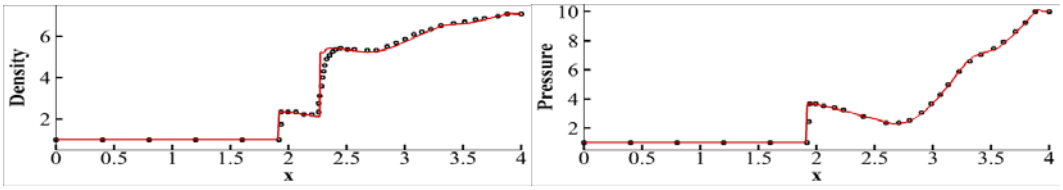


Figure 12. Comparison of density and pressure along y=0.5

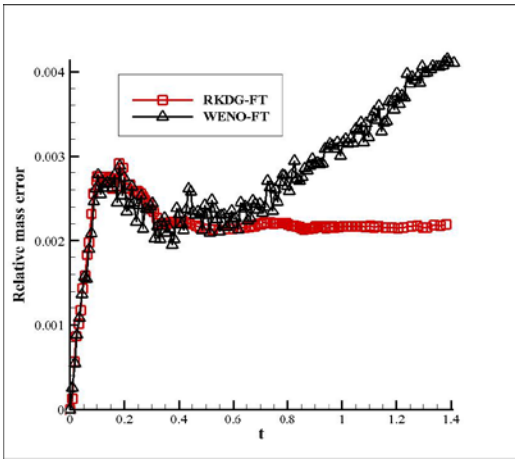


Figure 13. Comparison of relative mass error of air

Conclusions

In this paper, the RKDG method is applied to solve compressible two-medium flow. The interface is advanced by the front tracking method and the RGFM is used to define the interface boundary conditions. Due to the good compactness of the RKDG method, the ghost fluid states far from the interface which are less accuracy need not to be solved and used in the computation. Numerical results show that these procedures can work efficiently under different initial conditions. It also demonstrates that the RKDG-FT method has better mass conservation property compared to the WENO-FT method in general.

Acknowledgement

The research was supported by the National Basic Research Program of China ("973" Program) under grant No. 2014CB046200, NSFC grant 11432007.

References

[1] Cockburn, B. and Shu, C. -W. (1998) The Runge-Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems, *Journal of Computational Physics* **141**, 199–224.

[2] Fedkiw, R. P., Aslam, T., Merriman, B., Osher, S. (1999) A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method), *Journal of Computational Physics* **152**, 457–492.

[3] Glimm, J., Grove, J. W., Zhang, Y. (2002) Interface tracking for axisymmetric flows, SIAM Journal on Scientific Computing **24**, 208-236.

[4] Lu, H. T., Zhao, N., Wang, D. H. (2016) A front tracking method for the simulation of compressible multimedium flows, Communications in Computational Physics **19**, 124-142.

[5] Lu, H. T., Zhu, J., Wang, D. H., Zhao, N. (2016) Runge-Kutta discontinuous Galerkin method with front tracking method for solving the compressible two-medium flow, Computers and Fluids **126**, 1-11.

[6] Lu, H. T., Zhu, J., Wang, C. W., Zhao, N. (2016) Runge-Kutta discontinuous Galerkin method with front tracking method for solving the compressible two-medium flow on unstructured meshes, Advances in Applied Mathematics and Mechanics (accepted).

[7] Shyue, K. -M. (1998) An efficient shock-capturing algorithm for compressible multicomponent problems, *Journal of Computational Physics* **142**, 208–242.

[8] Terashima, H., Tryggvason, G. (2009) A front tracking/ghost fluid method for fluid interfaces in compressible flows, *Journal of Computational Physics* **228**, 4012–4037.

[9] Wang, C. W., Liu, T. G., Khoo, B. C. (2006) A real ghost fluid method for the simulation of multimedium compressible flow, SIAM Journal on Scientific Computing **28**, 278–302.

Multi-patches based B-Spline method for Solid and Structure

Yanan Liu* Bin Hu

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, the solution domain is divided into multi-patches on which B-Spline basis functions are used for approximation. The different B-Spline patches are connected by a transition region which is described by several elements. The basis functions in different B-Spline patches are modified in the transition region to ensure the basic polynomial reconstruction condition and the compatibility of displacements and their gradients. This new method is applied to the stress analysis of 2D elasticity problems in order to investigate its performance. Numerical results show that the present method is accurate and stable.

Keywords: B-Spline patches, Transition region, B-Spline basis functions.

Introduction

B-spline functions have been widely used in numerical analysis and simulation for decades. 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 [1]-[7]. The B-spline basis functions have compact support and lead to banded stiffness matrices. They can be used to construct piecewise approximations that provide higher order of continuity depending on the order of the polynomial basis. The B-spline basis functions form a partition of unity, which is an important property for convergence of the approximate solutions. As they are polynomials, accurate integration can be performed by using the Gauss quadrature. The B-spline approximation has good reproducing properties; thus, it is able to represent constant strains exactly. Compared to orthogonal or biorthogonal wavelets scaling functions and the shape functions constructed by meshless method, B spline functions are more simple and easy to work with for numerical analysis.

The main disadvantage of the general B-spline-based methods is that the scale used in approximation is usually uniform. In order to effectively simulate the local complicated deformation, the scale used in approximation should be very small. In this case, the computational efficiency will be very low. So it is desirable that the scales used for function approximation in solution domain are different. A more general approach that uses non-uniform rational B-splines (NURBS) [8]-[11] for the analysis has been developed. The method is referred to as the isogeometric analysis method because the geometry is also represented using NURBS basis functions to get an exact geometric representation. This method can achieve the traditional *h*- and *p*-adaptive refinement as well as *k*-refinement and get better solutions due to the superior basis functions. However, it is necessary to generate meshes that conform to the geometry of the analysis in this method.

In this paper, the solution domain is divided into multi-patches. The B-spline basis functions are directly used to approximate the unknown field functions in each patch. Thus, generation of conforming mesh is avoided in this approach. Different scales can be used in approximation for corresponding patch. The fine scale is used for function approximation in the patches where the deformation is complicated. The coarse scales are used for approximation in other patches where the deformation is relatively simple. A transition domain is used for combination of different B-spline patches. An algorithm is developed to modify the B-spline basis functions in the transition domain. The compatibility conditions on the interface between the different patches can be satisfied. The numerical examples of 2-D elasticity analysis are given to illustrate the stability and the effectiveness of the present method.

2. Approximation of 2-D functions by B-spline with single scale

The m degree B-spline is defined as

$$N_m(x) = N_{m-1} * N_1 = \int_0^1 N_{m-1}(x-t)dt, \ \ m \geq 2 \tag{1}$$

where

$$N_1(x) = \begin{cases} 1, x \in [0,1) \\ 0, \quad \text{else} \end{cases} \tag{2}$$

The major properties of B-spline are

$$\begin{gathered} \text{Supp} N_m = [0, m] \\ N_m'(x) = N_{m-1}(x) - N_{m-1}(x-1) \\ N_m(x) = \frac{x}{m-1} N_{m-1}(x) + \frac{m-x}{m-1} N_{m-1}(x-1) \end{gathered} \tag{3}$$

B-spline functions can be used as basis functions to approximate the function u defined on interval $[a,b]$.

$$u(x) = \sum_k c_k N_m^{i,k}(x) \tag{4}$$

where, $N_m^{i,k}(x) = N_m(1/i \cdot x - k)$ and i denotes the scale in approximation. According to properties of B-spline, the support of $N_m^{i,k}(x)$ is

$$\text{Supp} N_m^{i,k} = [ik, i(m+k)] \tag{5}$$

In approximation Eq.(4), the B-spline functions $N_m^{i,k}(x)$ should satisfy the following condition

$$\text{Supp} N_m^{i,k} \cap [a,b] \neq \varnothing \tag{6}$$

The basis functions for the higher-dimensional problems are constructed by taking the product of the basis functions for 1-D B-spline. In this case, the approximation of 2-D function $u(x,y)$ by B-spline function can be expressed as

$$u(x,y) = \sum_{k,l} c_{k,l} N_m^{i,k}(x) N_m^{j,l}(y) \tag{7}$$

where, $N_m^{i,k}(x)N_m^{j,l}(y)$ are 2-D B-spline basis functions, i and j are respectively the scales of x direction and y direction in approximation. For 2-D problems in general domains Ω, the 2-D B-spline basis functions which meet the following condition should be used in function approximation.

$$\text{Supp} N_m^{i,k}(x) N_m^{j,l}(y) \cap \Omega \neq 0 \tag{8}$$

Similar to finite element method and meshless methods, the approximation equation can be written as

$$u(x,y) = \sum_{h=1}^{N} \phi_h(x,y) c_h \tag{9}$$

where, $\phi_h(x,y) = N_m^{i,k}(x)N_m^{j,l}(y)$ is similar to shape functions in finite element method and meshless methods, c_h are the generalized displacement related to $\phi_h(x,y)$ and N is the number of 2-D B-spline functions used in approximation.

3. Coupling of different B-Spline patches

3.1 Basic formulations

The equations for the elastic problem are expressed as follows

$$\begin{aligned} &\sigma_{ij,j} + b_i = 0 \quad \text{in} \quad \Omega \\ &\sigma_{ij} n_j = \bar{t}_i \quad \text{on} \quad \Gamma_t \\ &u_i = \bar{u}_i \quad \text{on} \quad \Gamma_u \end{aligned} \tag{10}$$

where σ_{ij} is the stress tensor, b_i is the body force, $\bar{t}_i$ and $\bar{u}_i$ are respectively the prescribed boundary tension and displacement, and n_j is the unit outward normal to domain Ω. Consider the virtual displacement principle

$$\int_\Omega (\sigma_{ij,j} + b_i)\delta u_i d\Omega + \int_{\Gamma_t} (\sigma_{ij} n_j - \bar{t}_i)\delta u_i d\Gamma = 0 \tag{11}$$

where δu_i is the variation of real displacement. From Eq. (11), the weak form is

$$\int_\Omega \delta e_{ij} \sigma_{ij} d\Omega = -\int_{\Gamma_t} \delta u_i \bar{t}_i d\Gamma + \int_\Omega \delta u_i b_i d\Omega \tag{12}$$

where δu_i vanishes and $u_i = \bar{u}_i$ on Γ_u.

Consider a division with two patches and a transition region in a given region Ω as shown in Figure 1. In 2-D problem, the approximation for the displacement field u and v can be respectively written as

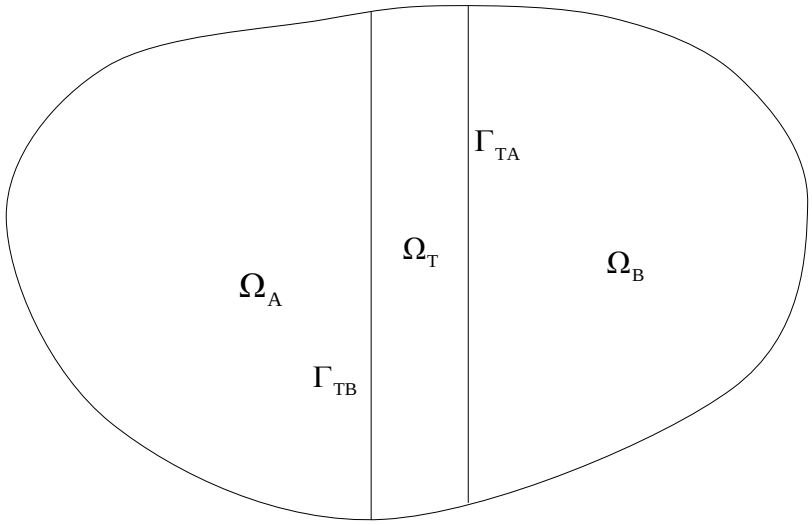


Figure 1. The problem domain is divided into two patches and a transition region

$$\begin{cases} u_A(x,y) = \sum_{i=1}^{n_W} a_i^{Au} \phi_i^A(x,y) \\ v_A(x,y) = \sum_{i=1}^{n_A} a_i^{Av} \phi_i^A(x,y) \end{cases} \quad (x,y) \in \Omega_A + \Omega_T \tag{13}$$

$$\begin{cases} u_B(x,y) = \sum_{i=1}^{n_F} a_i^{Bu} \phi_i^B(x,y) \\ v_B(x,y) = \sum_{i=1}^{n_B} a_i^{Bv} \phi_i^B(x,y) \end{cases} \quad (x,y) \in \Omega_B + \Omega_T \tag{14}$$

where, n_{A} and n_{B} is respectively the number of basis functions used in approximation. It is obvious that the two kinds of approximation functions are not compatible and should be modified in the transition region.

3.2 Transition region and modified basis functions

In 2D problems, the transition region can be described by several elements.

$$\begin{cases} x_k(\xi,\eta)=\sum_{i=1}^{n} N_i^{\mathrm{F}}(\eta,\xi)x_i^k \\ y_k(\xi,\eta)=\sum_{i=1}^{n} N_i^{\mathrm{F}}(\eta,\xi)y_i^k \end{cases} \quad -1\le\eta\le 1,-1\le\xi\le 1,1\le k\le n_{\mathrm{F}} \tag{15}$$

where, N_i^{F} is the shape function of four nodes plain element and n_{F} is the number of element. The basis functions should be modified in transition region. A weight function based on the transition region should be introduced into modification. The modified basis functions in the transition region can be expressed as

$$\begin{cases} \phi_{k,n}^{\mathrm{Am}}(\xi,\eta)=\phi_n^{\mathrm{A}}(x(\xi,\eta),y(\xi,\eta))*w(\eta) \\ \phi_{k,n}^{\mathrm{Bm}}(\xi,\eta)=\phi_n^{\mathrm{B}}(x(\xi,\eta),y(\xi,\eta))*(1-w(\eta)) \end{cases} \quad (x,y)\in\Omega_{\mathrm{T}} \tag{16}$$

The following functions can be chosen as weight function

$$w(\eta)=1-6*(\frac{\eta+1}{2})^2+8*(\frac{\eta+1}{2})^3-3*(\frac{\eta+1}{2})^4 \qquad -1\le\eta\le 1 \tag{17}$$

Then, the approximation in transition region can be expressed as

$$\begin{cases} u_k(\xi,\eta)=\sum_i a_{\mathrm{m},i}^{\mathrm{A}u}\phi_{k,i}^{\mathrm{Am}}(\xi,\eta)+\sum_j a_{\mathrm{m},j}^{\mathrm{B}u}\phi_{k,j}^{\mathrm{Bm}}(\xi,\eta) \\ v_k(\xi,\eta)=\sum_i a_{\mathrm{m},i}^{\mathrm{A}v}\phi_{k,i}^{\mathrm{Am}}(\xi,\eta)+\sum_j a_{\mathrm{m},j}^{\mathrm{B}v}\phi_{k,j}^{\mathrm{Bm}}(\xi,\eta) \end{cases} \quad -1\le\eta\le 1,-1\le\xi\le 1,1\le k\le n_{\mathrm{F}} \tag{18}$$

Then, the approximation formula (13) and (14) should be rewritten as

$$\begin{cases} u_{\mathrm{A}}(x,y)=\sum_{i=1}^{n_{\mathrm{A}}} a_i^{\mathrm{A}u}\phi_i^{\mathrm{A}}(x,y) \\ v_{\mathrm{A}}(x,y)=\sum_{i=1}^{n_{\mathrm{A}}} a_i^{\mathrm{A}v}\phi_i^{\mathrm{A}}(x,y) \end{cases} \quad (x,y)\in\Omega_{\mathrm{A}} \tag{19}$$

$$\begin{cases} u_{\mathrm{B}}(x,y)=\sum_{i=1}^{n_{\mathrm{B}}} a_i^{\mathrm{B}u}\phi_i^{\mathrm{B}}(x,y) \\ v_{\mathrm{B}}(x,y)=\sum_{i=1}^{n_{\mathrm{B}}} a_i^{\mathrm{B}v}N_i^{\mathrm{B}}(x,y) \end{cases} \quad (x,y)\in\Omega_{\mathrm{B}} \tag{20}$$

Eventually, a group of linear algebraic equations can be obtained by introducing the approximations formula (18), (19) and (20) into weak form (12).

$$\mathbf{Ka}=\mathbf{f} \tag{21}$$

where,

$$\mathbf{K}=\begin{bmatrix} \mathbf{K}_{\mathrm{A}} & \mathbf{K}_{\mathrm{AB}} \\ \mathbf{K}_{\mathrm{AB}}^{\mathrm{T}} & \mathbf{K}_{\mathrm{B}} \end{bmatrix} \tag{22}$$

$$\mathbf{a}=\begin{bmatrix} \mathbf{a}_{\mathrm{A}} \\ \mathbf{a}_{\mathrm{B}} \end{bmatrix} \tag{23}$$

$$\mathbf{f}=\begin{bmatrix} f_{\mathrm{A}} \\ f_{\mathrm{B}} \end{bmatrix} \tag{24}$$

and,

$$\mathbf{a}_{\mathrm{A}}=[a_1^{\mathrm{A}u},a_1^{\mathrm{A}v},\cdots,a_{n_{\mathrm{A}}}^{\mathrm{A}u},a_{n_{\mathrm{A}}}^{\mathrm{A}v}]^{\mathrm{T}} \tag{25}$$

$$\mathbf{a}_{\mathrm{B}}=[a_1^{\mathrm{B}u},a_1^{\mathrm{B}v},\cdots,a_{n_{\mathrm{B}}}^{\mathrm{B}u},a_{n_{\mathrm{B}}}^{\mathrm{B}v}]^{\mathrm{T}} \tag{26}$$

$$\mathbf{K}_{\mathrm{A}}=\int_{\Omega_{\mathrm{A}}}\mathbf{B}_{\mathrm{A}}{}^{\mathrm{T}}\mathrm{D}\mathbf{B}_{\mathrm{A}}d\Omega+\int_{\Omega_{\mathrm{T}}}\mathbf{B}_{\mathrm{A}}{}^{\mathrm{T}}\mathrm{D}\mathbf{B}_{\mathrm{A}}d\Omega \tag{27}$$

$$\mathbf{K}_{\mathrm{B}}=\int_{\Omega_{\mathrm{B}}}\mathbf{B}_{\mathrm{B}}{}^{\mathrm{T}}\mathbf{D}\mathbf{B}_{\mathrm{B}}d\Omega+\int_{\Omega_{\mathrm{T}}}\mathbf{B}_{\mathrm{B}}{}^{\mathrm{T}}\mathbf{D}\mathbf{B}_{\mathrm{B}}d\Omega \tag{28}$$

$$\mathbf{K}_{\mathrm{AB}}=\int_{\Omega_{\mathrm{T}}}\mathbf{B}_{\mathrm{A}}{}^{\mathrm{T}}\mathrm{D}\mathbf{B}_{\mathrm{B}}d\Omega \tag{29}$$

$$\mathbf{f}_{\mathrm{A}}{}^{\mathrm{T}}=\int_{\Omega_{\mathrm{A}}}\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{b}d\Omega+\int_{\Omega_{\mathrm{T}}}\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{b}d\Omega+\int_{\Gamma_{\mathrm{A}}}\boldsymbol{\varphi}^{\mathrm{T}}\bar{\mathbf{t}}d\Gamma \tag{30}$$

$$\mathbf{f}_{\mathrm{B}}{}^{\mathrm{T}}=\int_{\Omega_{\mathrm{B}}}\mathbf{N}^{\mathrm{T}}\mathbf{b}d\Omega+\int_{\Omega_{\mathrm{T}}}\mathbf{N}^{\mathrm{T}}\mathbf{b}d\Omega+\int_{\Gamma_{\mathrm{B}}}\mathbf{N}^{\mathrm{T}}\bar{\mathbf{t}}d\Gamma \tag{31}$$

$\mathbf{D}$ is the 2-D elasticity matrix.

$$\mathbf{D}=\frac{E_0}{(1-\nu_0^2)}\begin{bmatrix} 1 & \nu_0 & 0 \\ \nu_0 & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu_0}{2} \end{bmatrix}$$

Plain stress

$$E_0=E,\ \nu_0=\nu$$

Plain strain

$$E_0=\frac{E}{1-\nu^2},\ \nu_0=\frac{\nu}{1-\nu}$$

$$\mathbf{B}_{\mathrm{A}}=\mathbf{L}\boldsymbol{\phi}^{\mathrm{A}}$$

$$\mathbf{B}_{\mathrm{B}}=\mathbf{L}\boldsymbol{\phi}^{\mathrm{B}}$$

$$\mathbf{L}=\begin{bmatrix} \dfrac{\partial}{\partial x} & 0 \\ 0 & \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \end{bmatrix}$$

$$\boldsymbol{\phi}^{\mathrm{A}}=\begin{bmatrix} \phi_1 & 0 & \cdots & \phi_{n_{\mathrm{A}}} & 0 \\ 0 & \phi_1 & \cdots & 0 & \phi_{n_{\mathrm{A}}} \end{bmatrix}$$

$$\boldsymbol{\phi}^{\mathrm{B}}=\begin{bmatrix} \phi_1 & 0 & \cdots & \phi_{n_{\mathrm{B}}} & 0 \\ 0 & \phi_1 & \cdots & 0 & \phi_{n_{\mathrm{B}}} \end{bmatrix}$$

4 Numerical examples

In this part, numerical simulation of some 2-D plain elasticity problems is presented using the present 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.

Cantilever beam

A cantilever beam is analyzed by the presented method. As shown in Figure 2, the beam has a dimension of length $L=10$ and height $h=2$ and is subject to a parabolic traction with $P=-300$ and $p_y=-0.75P(1-(y-1)^2)$. The beam has a unit thickness and a plane strain problem is considered. The Young's modulus is set to $E=2.1\times10^4$ and Poisson's ration is set to $\nu=0.49$. In this problem, the analytical results of stress are expressed as follows

$$\sigma_x=\frac{P}{I}(L-x)(y-1),\quad \sigma_y=0,\quad \sigma_{xy}=-\frac{P}{2I}\left[(\frac{h}{2})^2-(y-1)^2\right]$$

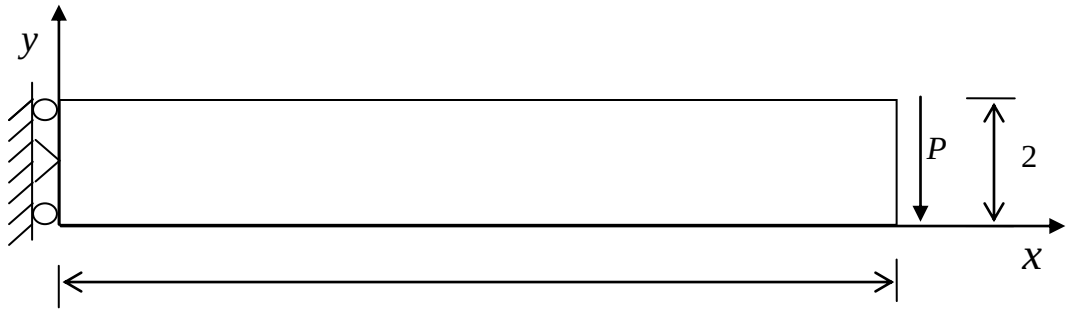


Figure 2. Cantilever beam under a parabolic traction at the free end

The problem domain is divided into two patches and a transition region as shown in Figure 3. Cubic B-Spline is used in this simulation. The scales used for approximation in two patches are denoted by $\mathrm{A}_x,\mathrm{A}_y$ and $\mathrm{B}_x,\mathrm{B}_y$, respectively. The width of transition region is expressed by t and the Γ_{TA} is fixed at $x=4$. The different parameters related to

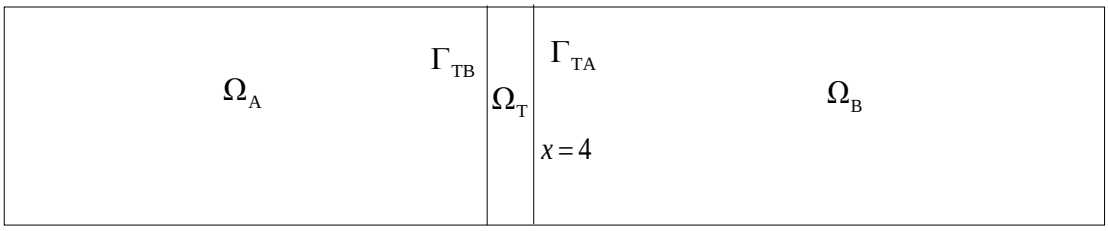


Figure 3. The patches in 2D beam problem

transition region are studied in this simulation. Figure 4 shows the results of σ_{xy} along $x=3$ and $x=5$ with $t=0.2$. It can be found that the results computed by the present method agree well with the analytical results.

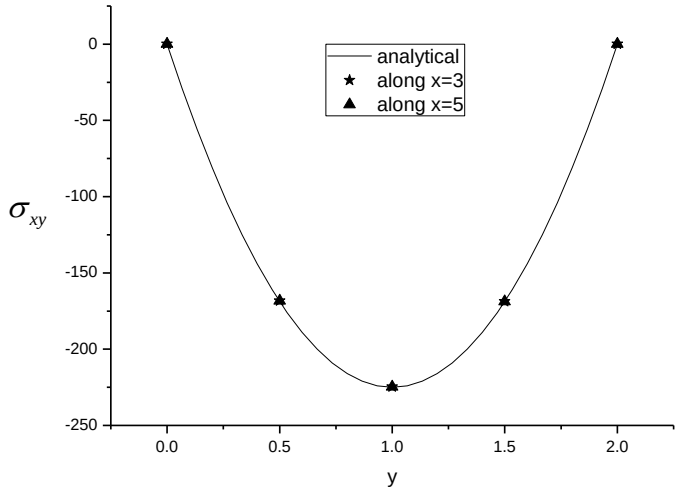


Figure 4. The comparison of shear stress

Conclusions

In this paper, the B-spline basis functions are directly used to approximate the unknown field functions in multi-patches. The generation of conforming mesh is avoided in this approach. Different scales are used in approximation for corresponding patch. A transition domain is used for combination of different B-spline patches. The B-spline basis functions are modified to satisfy the high-order compatibility conditions on the interface between the different patches. The computational efficiency of this method is much higher than single patch based single scale approach. Numerical examples for 2-D elasticity problems illustrate that this B-spline method is effective and stable for solving elasticity problems.

References

[1] Naginoa, H., Mikamia, T. and Mizusawa, T. (2008) Three-dimensional free vibration analysis of isotropic rectangular plates using the B-spline Ritz method, *Journal of Sound and Vibration* 317, 329–353.
[2] Caglar, N. and Caglar, H. (2009) B-spline method for solving linear system of second-order boundary value problems, *Computers Mathematics with Applications* 57, 757-762.
[3] Kagan, P., Fischer, A. and Bar-Yoseph, P.Z. (2003) Mechanically based models: Adaptive refinement for B-spline finite element, *International Journal for Numerical Methods in Engineering* 57, 1145–1175.
[4] Burla, R.K. and Kumar, A.V. (2008) Implicit boundary method for analysis using uniform B-spline basis and structured grid, *International Journal for Numerical Methods in Engineering* 76, 1993–2028.
[5] Lakestani, M. and Dehghan, M. (2009) Numerical Solution of Fokker—Planck Equation Using the Cubic B-Spline Scaling Functions, *Numerical Methods for Partial Differential Equations* 25, 418–429.
[6] Liu, Y., Sun, L., Xu, F., Liu, Y. and Cen, Z. (2011) B spline-based method for 2-D large deformation analysis, *Engineering Analysis with Boundary Elements* 35, 761-767.
[7] Liu, Y., Sun, L., Liu, Y. and Cen, Z. (2011) Multi-scale B-spline method for 2-D elastic problems, *Applied Mathematical Modelling* 35, 3685-3697.
[8] Hughes, T.J.R., Cottrell, J.A. and Bazilevs, Y. (2005) Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer Methods in Applied Mechanics and Engineering* 194, 4135–4195.
[9] Cottrell, J.A., Reali, A., Bazilevs, Y. and Hughes, T.J.R. (2006) Isogeometric analysis of structural vibrations, *Computer Methods in Applied Mechanics and Engineering* 195, 5257–5296.
[10] Cottrell, J.A., Hughes, T.J.R. and Reali, A. (2007) Studies of refinement and continuity in isogeometric analysis, *Computer Methods in Applied Mechanics and Engineering* 196, 4160–4183.
[11] Bazilevs, Y., Calo, V.M., Hughes, T.J.R. and Zhang, Y. (2008) Isogeometric fluid–structure interaction: theory, algorithms, and computations, *Computational Mechanics* 43, 3–37.

A spectral element analysis of sound transmission through metallic sandwich plates with adhesively-bonded corrugated cores

Hao Sen Yang[1], Heow Pueh Lee[2,*], Hui Zheng[1,]**

[1] Institute of Vibration, Shock & Noise, School of Mechanical Engineering
Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.
[2]Department of Mechanical Engineering, The National University of Singapore
9 Engineering Drive 1, Singapore 117575

*Presenting author: mpeleehp@nus.edu.sg
**Corresponding author: huizheng@sjtu.edu.cn

Abstract

This paper presents a spectral element based numerical method for calculating the vibro-acoustic response of sandwich plates with adhesively-bonded corrugated cores. The study is motivated by the need of optimal designs for improving the structural-acoustic performance of the considered structures. A two-dimensional plate model is firstly developed based on the spectral element method (SEM) for obtaining the frequency-domain vibration response of the whole structure subject to incident harmonic acoustic wave. Thereafter the Rayleigh integral formula is used to calculate the transmitted sound power via its structure-borne path. Comparing with the conventional finite element method, the SEM, since it is formulated in the frequency domain by using the exact wave solutions for the governing differential equation, provides exact frequency-domain solutions meanwhile using much fewer number of degrees-of-freedom. This is proven by the numerical results of structural vibration response. Furthermore, parametric studies are performed to investigate the influence of the inclined angle of bonded corrugated core and the thickness of face plates on the transmitted sound power of sandwich plates. Although these design parameters have different effects on the sound transmission loss in different frequency-bands of interest, the impacts of both of them become more evident with the increase of targeted sound-insulating frequency.

Keywords: Sandwich plate; corrugated core; spectral element; Sound transmission.

Introduction

Metallic sandwich plates with corrugated cores are used extensively in the high speed transportation engineering field for their lower area density, higher specific strength and stiffness than those of a homogeneous type. Vibro-acoustic response of this kind of structures subject to airborne excitation have been a concern in acoustic comfort design of high speed transportation systems such as airplanes and express trains.

Considering the wide usage of the sandwich structure, various theoretical studies have been performed aiming at understanding the mechanism of sound transmission through such kind of structure. The early studies of acoustic radiation problem for periodic stiffened structure are limited to single beam or plate[1–3], and the main concern of these works is the vib-acoustic response of periodic stiffened structure under mechanical forces. For the double layer structure, starting from the double-leaf partitions made up of homogeneous panels with no structural stiffener in the core, Pellicier and Trompette[4] reviewed various wave approach based methods for calculating the partitions transmission loss, and proposed a simple mechanism on the theory of sound transmission through such kind of structure. Considering the stiffened double layer structure, Wang et al.[5] studied the double-leaf lightweight partitions stiffened with periodically placed studs, and presented a theoretical model to predict the sound transmission loss of the structure. Legault et al.[6] studied the sound transmission

through an aircraft sidewall representative double panel structure theoretically by using space harmonic analysis, which was also used by Xin and Lu[7] to investigate the transmission loss of sound through infinite orthogonally rib-stiffened double-panel structures with cavity absorption.

According to the results of these theoretical analyses, both the structural topology and material properties have a great impact on the sound insulation capability. To balance the mechanical and acoustical properties of the sandwich structures, many researchers turn their attention focus on the structural-acoustic optimization problem[8–10], and in most of these work, the conventional finite element method is used to calculate the objective function. However, since most optimization and parametrical study requires tremendous computing workload, the computational time could be a bottleneck when a complex structural model is involved. Considering the drawback of the conventional finite element method[11], a more efficient alternative method is needed.

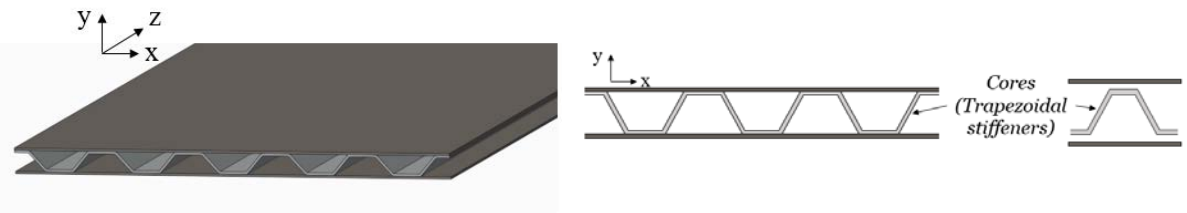
The spectral element method (SEM) is firstly proposed in the 80's[12]. Birgersson et al.[13] proposed a general theory for the analysis of structural vibration of an uniform plate under high frequency random excitation, Żak[14] presented a novel formulation of a spectral plate finite element for analysis of propagation of elastic waves in isotropic plate structures, and Wu et al.[15] studied the dynamic behavior of periodic plate structures by using SEM. All the results showed that the spectral element method appears to be an effective tool for modeling structural dynamic equations.

In this paper, a vibro-acoustic model of metallic sandwich plates with adhesively-bonded corrugated cores is presented in the first place. The governing dynamic equations are derived based on spectral element method and the structural vibration response of the sandwich plates subject to air borne sound excitation is calculated. By using the Rayleigh integral, the sound power radiated from the structure is obtained. Furthermore, parametric studies are performed to determine the influence of the inclined angle of the stiffener and the thickness of the face plate on the averaged radiated sound power.

Theoretical formulation

Structural configuration

As illustrated in Figure. 1(a), the sandwich plates considered here consists of two metallic face plates and a trapezoidal corrugated core while the core is press-formed and glued on both of the two face plates. Because of the simple manufacturing process, this kind of sandwich structure is favored by the transportation industry, like being used in the carriages of air planes and high speed trains.



(a) 3-demensional model (b) 2-dimensional cross section

Figure. 1 Geometric schematic of the sandwich plates

As shown in Figure 1(a), the sandwich structure is periodically stiffened by the adhesively-bonded corrugated cores along the z-direction, the sandwich plate can be considered as a one-way stiffened structure. Following the traditional method of modeling the vibration and sound

transmission of sandwich structures with cellular cores or truss-like periodic panels[10,16,17], the sandwich plate with corrugated core is assumed infinite along the z-axis. Thus the three-dimensional sandwich plates can be simplified a sandwich beam structure represented by the cross section as shown in Figure 1(b).

Spectral element method modeling

For spectral element method, the governing equation of motion of the global system is assembled by all the spectral elements, and it is given in the frequency-domain as

$$\boldsymbol{S}_g(\omega)\boldsymbol{D}_g(x,\omega) = \boldsymbol{F}_g \tag{1}$$

where $\boldsymbol{S}_g$, $\boldsymbol{D}_g$ and $\boldsymbol{F}_g$ represent the global dynamic stiffness matrix, the global spectral nodal DOFs vector, and the global nodal force vector, respectively. For beam structure, the exact dynamic stiffness matrix is formulated based on the exact wave solutions to the governing differential equations[18]. Thus, theoretically, SEM can provide accurate solution to the dynamic response of the beam structure, but, comparing to the conventional finite element method, SEM only uses a minimum number of DOFs, which makes SEM much more computationally efficient[19].

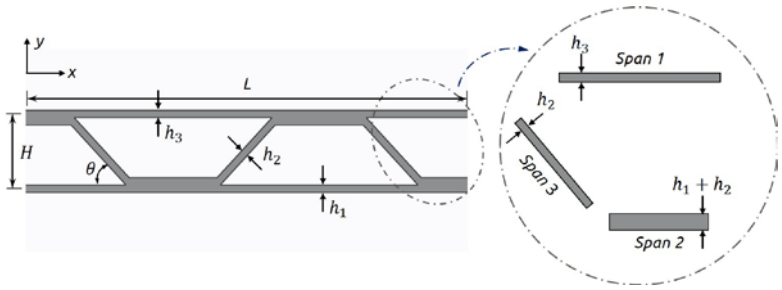


Figure. 2 Geometric configuration of sandwich beam

Figure. 2 gives the details of the sandwich beam with a finite total length of L and total height of H. h_1, h_2 and h_3 represent the thickness of the lower face plate, core plate and the upper face plate. The inclined angle of the stiffened core is defined as θ. Both the face plates and core are made of the same isotropic, homogeneous material with the elasticity modulus E and density ρ. In order to apply the spectral element method, the sandwich beam structure is firstly divided into a number of spans, and each span can be treated as a single spectral element.

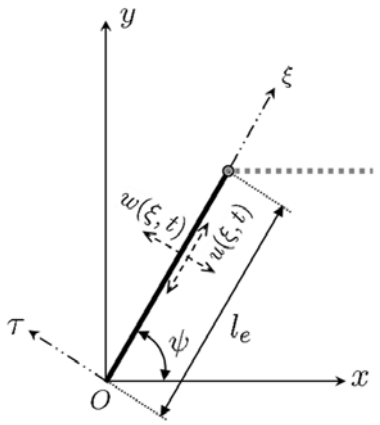


Figure. 3 Local & global coordinate system of a beam element

An illustration of a single spectral beam element in the local coordinates system is shown in Figure 3, the classical governing equations of free longitudinal and flexural vibration are：

$$\frac{\partial^2 u(\xi,t)}{\partial \xi^2} - \frac{\rho}{E}\cdot\frac{\partial^2 u(\xi,t)}{\partial t^2} = 0, \quad \frac{\partial^4 w(\xi,t)}{\partial \xi^4} + \frac{\rho A}{EI}\cdot\frac{\partial^2 w(\xi,t)}{\partial t^2} = 0 \tag{2}$$

where $u(\xi,t)$, $w(\xi,t)$ are respectively, the longitudinal and flexural displacement in the local coordinate system. *A* is the cross-sectional area, *I* is the cross sectional moment of inertia. The solution to Eq. (2) is assumed in the spectral form as:

$$u(\xi,t) = \frac{1}{N}\sum_{n=1}^{N} U_n(\xi;\omega_n)\cdot e^{i\omega_n t}, \quad w(\xi,t) = \frac{1}{N}\sum_{n=1}^{N} W_n(\xi;\omega_n)\cdot e^{i\omega_n t} \tag{3}$$

at a specific discretization frequency $\omega = \omega_n$, the general solution to Eq. (2) can be written as:

$$\begin{cases} U_n(\xi;\omega_n) = \{e^{-ik_l\xi} \quad e^{+ik_l\xi}\}\cdot\{\boldsymbol{C}_l\}, \quad k_l = \omega_n\sqrt{\rho/E} \\ W_n(\xi;\omega_n) = \{e^{ik_f\xi} \quad e^{-ik_f\xi} \quad e^{-k_f\xi} \quad e^{k_f\xi}\}\cdot\{\boldsymbol{C}_f\}, \quad k_f = \sqrt[4]{\rho A\omega_n^2/EI} \end{cases} \tag{4}$$

where $\{\boldsymbol{C}_l\}$ and $\{\boldsymbol{C}_f\}$ are both constant column vectors. The longitudinal and transverse nodal displacement at both ends of the beam element can be written as:

$$\begin{cases} \boldsymbol{d}_l = \{u_1|_{\xi=0} \quad u_2|_{\xi=l_e}\} = \boldsymbol{B}_l\cdot\{\boldsymbol{C}_l\} \\ \boldsymbol{d}_f = \{w_1|_{\xi=0} \quad \theta_1|_{\xi=0} \quad w_2|_{\xi=l_e} \quad \theta_2|_{\xi=l_e}\} = \boldsymbol{B}_f\cdot\{\boldsymbol{C}_f\} \end{cases} \tag{5}$$

u, *w* and θ are, respectively, longitudinal displacement, deflection and slope. Considering the force-displacement relation, the internal axial force $F(\xi)$, shear force $T(\xi)$ and moment $M(\xi)$ are given by:

$$F(\xi) = EA\frac{dU(\xi)}{d\xi}, \quad T(\xi) = -EI\frac{d^3W(\xi)}{d\xi^3}, \quad M(\xi) = EI\frac{d^2W(\xi)}{d\xi^2} \tag{6}$$

According to the compatibility condition, using Eq. (4) and Eq. (6), the external spectral nodal forces and moments acting on the two nodes of the beam element can be given as the form of:

$$\boldsymbol{f}_l = \{\widetilde{F}_1 \quad \widetilde{F}_2\}^T = \boldsymbol{H}_l\cdot\{\boldsymbol{C}_l\}, \quad \boldsymbol{f}_f = \{\widetilde{T}_1 \quad \widetilde{M}_1 \quad \widetilde{T}_2 \quad \widetilde{M}_2\}^T = \boldsymbol{H}_f\cdot\{\boldsymbol{C}_f\} \tag{7}$$

Eliminate the constant vectors using Eq. (5) and Eq. (7), it gives:

$$\underbrace{\boldsymbol{H}_l\cdot\boldsymbol{B}_l^{-1}}_{\boldsymbol{S}_l(\omega_n)}\cdot\boldsymbol{d}_l = \begin{bmatrix}\kappa_{11}^L & \kappa_{12}^L \\ \kappa_{21}^L & \kappa_{22}^L\end{bmatrix}\cdot\boldsymbol{d}_l = \boldsymbol{f}_l, \qquad \underbrace{\boldsymbol{H}_f\cdot\boldsymbol{B}_f^{-1}}_{\boldsymbol{S}_f(\omega_n)}\cdot\boldsymbol{d}_f = \begin{bmatrix}\kappa_{11}^F & \kappa_{12}^F & \kappa_{13}^F & \kappa_{12}^F \\ \kappa_{21}^F & \kappa_{22}^F & \kappa_{23}^F & \kappa_{22}^F \\ \kappa_{31}^F & \kappa_{32}^F & \kappa_{33}^F & \kappa_{32}^F \\ \kappa_{41}^F & \kappa_{42}^F & \kappa_{43}^F & \kappa_{42}^F\end{bmatrix}\cdot\boldsymbol{d}_f = \boldsymbol{f}_f \tag{8}$$

where $\boldsymbol{S}_l(\omega_n)$ and $\boldsymbol{S}_f(\omega_n)$ are known as spectral element matrix for longitudinal and transvers vibration of a single beam element. Combining the longitudinal and transvers equation into one single spectral equation, it gives:

$$\underbrace{\begin{bmatrix} \kappa_{11}^{L} & 0 & 0 & \kappa_{12}^{L} & 0 & 0 \\ 0 & \kappa_{11}^{F} & \kappa_{12}^{F} & 0 & \kappa_{13}^{F} & \kappa_{14}^{F} \\ 0 & \kappa_{21}^{F} & \kappa_{22}^{F} & 0 & \kappa_{23}^{F} & \kappa_{24}^{F} \\ \kappa_{21}^{L} & 0 & 0 & \kappa_{22}^{L} & 0 & 0 \\ 0 & \kappa_{31}^{F} & \kappa_{32}^{F} & 0 & \kappa_{33}^{F} & \kappa_{34}^{F} \\ 0 & \kappa_{41}^{F} & \kappa_{42}^{F} & 0 & \kappa_{43}^{F} & \kappa_{44}^{F} \end{bmatrix}}_{\boldsymbol{S}(\omega_n)} \underbrace{\begin{bmatrix} u_1 \\ w_1 \\ \theta_1 \\ u_2 \\ w_2 \\ \theta_2 \end{bmatrix}}_{\boldsymbol{d}} = \boldsymbol{f} = \begin{bmatrix} \widetilde{F}_1 \\ \widetilde{T}_1 \\ \widetilde{M}_1 \\ \widetilde{F}_2 \\ \widetilde{T}_2 \\ \widetilde{M}_2 \end{bmatrix} \tag{9}$$

where $\boldsymbol{S}(\omega_n)$ is the general spectral element matrix of a single beam. The continuous displacement field can be represented by the spectral nodal displacements $\boldsymbol{d}$ as

$$\boldsymbol{G}(\xi) = \widetilde{\boldsymbol{N}} \cdot \boldsymbol{d} \tag{10}$$

$\boldsymbol{G}(\xi)$ represents both the longitudinal and flexural displacements, and $\widetilde{\boldsymbol{N}}$ is the dynamic shape function which can be obtained from Eq. (4) and Eq. (5). When the structure is exposed to a distributed force loading, the distributed force $\boldsymbol{P}(\xi, \omega_n)$ must be transferred to each node of the spectral elements by using the virtual work principle in the frequency-domain.

$$\boldsymbol{f} = \int_0^{l_e} \widetilde{\boldsymbol{N}}^T(\xi, \omega_n) \boldsymbol{P}(\xi, \omega_n) \mathrm{d}\xi \tag{11}$$

Now the governing equation of motion of a single spectral element can be symbolically represented by

$$\boldsymbol{s}(\omega_n)_{local} \cdot \boldsymbol{d}_{local} = \boldsymbol{f}_{local} \tag{12}$$

With the coordinate transformation matrix T:

$$\boldsymbol{d}_{global} = \boldsymbol{T} \cdot \boldsymbol{d}_{local}, \quad \boldsymbol{f}^d_{global} = \boldsymbol{T} \cdot \boldsymbol{f}^d_{local}, \quad \boldsymbol{s}(\omega)_{global} = \boldsymbol{T}^{-1} \cdot \boldsymbol{s}(\omega)_{local} \cdot \boldsymbol{T} \tag{13}$$

The spectral equations of the whole sandwich structure can be written as the assembly of the coordinate transformed local equations:

$$\mathbb{S}(\omega_n) \cdot \mathbb{D} = \mathbb{F} \tag{14}$$

Acoustic radiation

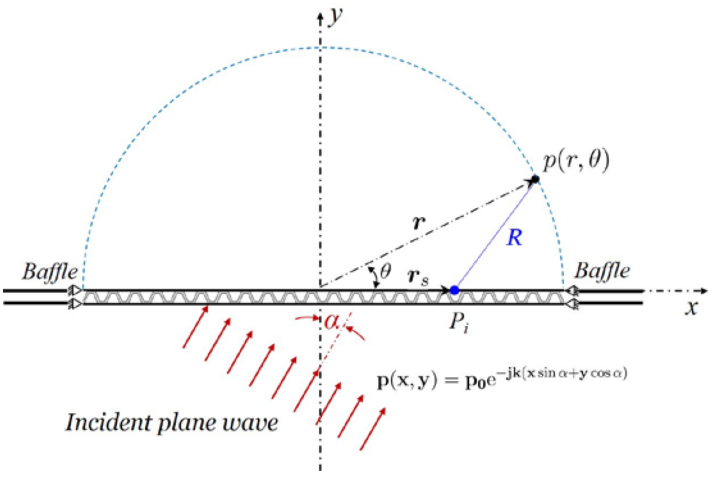


Figure 4. The acoustic transmission model

It is assumed that the sandwich beam is baffled at both the top and bottom surface, as shown in Figure 4. Considering a unit magnitude acoustic plane wave impinged on the bottom beam surface with an incident angle of α. The acoustic pressure is transmitted to the top beam via the structure-borne path and radiates sound to the semi-infinite space. The half-circle illustrated in Figure 4 is the observation surface where the acoustic power radiated from the top beam surface is calculated.

The transmitted acoustic pressure $p(r,\theta,\omega)$ at a specific observation point $\boldsymbol{r}$ due to the surface normal velocity v_i on the top beam can be calculated using the Rayleigh's integral[20]:

$$p(\boldsymbol{r},\theta,\omega)=\int_L \frac{\rho_0\omega}{2}H_0^{(2)}\left[kR(x)\right]\cdot v(x)\mathrm{d}x \tag{15}$$

where ρ_0 is the air density, $R=|\boldsymbol{r}-\boldsymbol{r}_s|$ and $H_0^{(2)}$ is the Hankel function of the second kind. The acoustic power radiated from the baffled beam at frequency ω can be obtained by an integration over the receiver surface (the half round surface *S'* shown in Figure 4):

$$W(\omega)=\frac{1}{2\rho_0 c_0}\cdot\int_{S'}p^2(r,\theta,\omega)\cdot r\ \mathrm{d}\theta \tag{16}$$

Numerical validation

To verify the accuracy of present spectral element method, both the conventional FEM and SEM are used to calculate the vibration response of the sandwich beam structure. The structural drawing of the validation model has been given in Figure 3, in order to be more rigorous, two sets of design parameters are chosen to test the present method. The details of the two models are listed in Table 1. Another is worth mentioning, the whole structure is made of aluminum, with the modulus of elasticity is 7.1×10^{10} Pa, structural damping factor is 0.01, and the mass density is 2700 kg/m3.

Table 1. Validation model parameters

Parameter	Model 1	Model 2
Total length of the sandwich structure (L)	1200.0 mm	
Total height of the sandwich beam (H)	50.0 mm	
Number of inclined stiffener (N)	11	
Inclined angle of the stiffener (θ)	40°	60°
Thickness of the top face beam (h_3)	2.0 mm	
Thickness of the bottom face beam (h_1)	2.0 mm	
Thickness of the core beam (h_2)	2.0 mm	

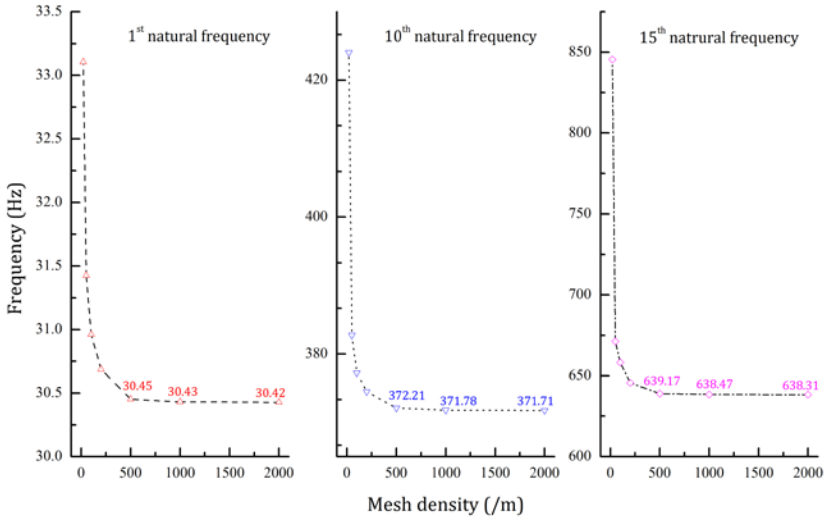


Figure 5. FEM mesh convergency

To ensure the reliablity of the FEM results, a mesh convergency study is performed based on model 2 given in Table 1. The mesh convergency diagram is shown in Figure 6, the 1st , 10th and 15th natural frequency of the sandwich structure are chosen to be criteria of the convergency study. As shown in Figure 5, high frequency analysis requires high mesh density, in consideration of the computational efficiency, under 1000Hz, the mesh density of 1000/m is chosen to perform the FEM harmonic analysis.

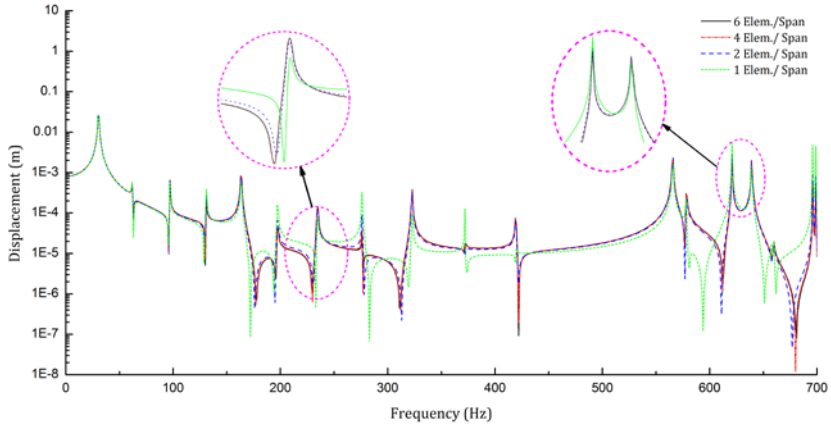
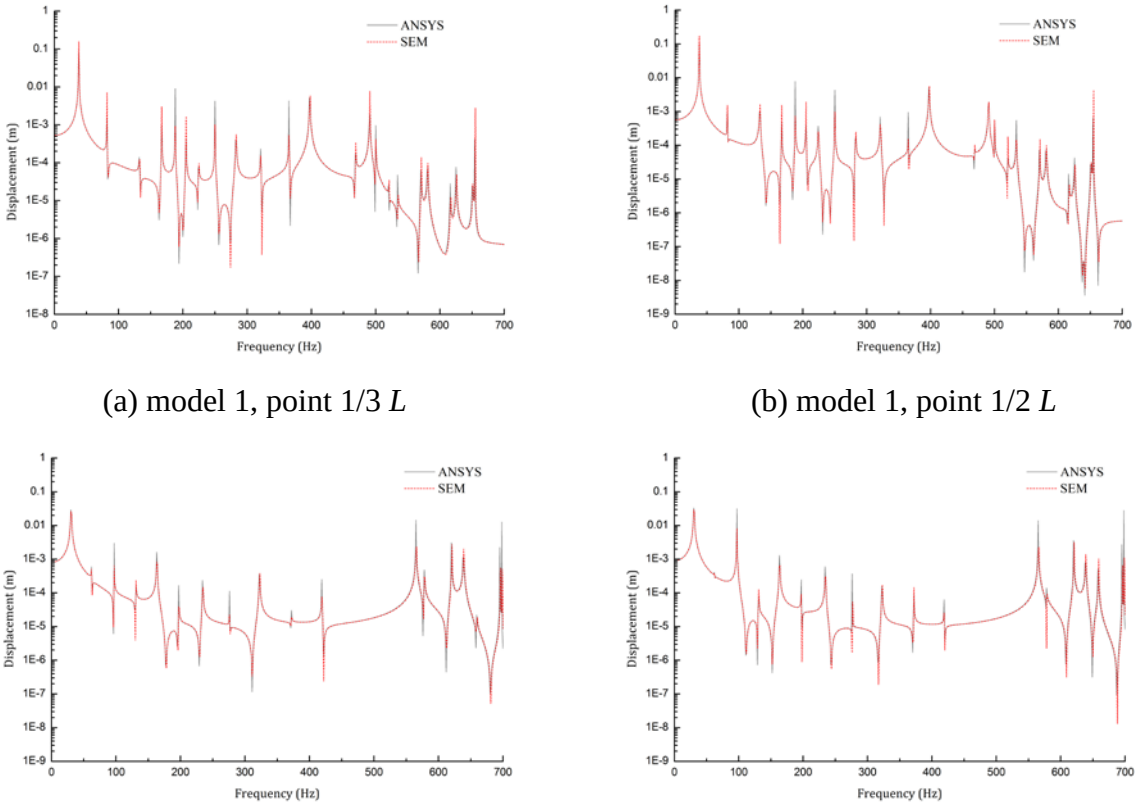


Figure 6. SEM element convergency

As for the present spectral element method, the convergency study is also performed as shown in Figure 6, the test model is also model 2 given in Table 1 and the observation point is located at 1/3 *L* of the top beam. The results indicate that when the element density is bigger than 2/span, the difference between the curves of the dynamic response is quite small, thus, the element density of 4/span is used in this paper which is much smaller than the FEM.

Considering a uniform distributed acoustic pressure with a unit magnitude is acting on the whole bottom beam surface, and the simply supported boundary conditions are applied at both ends of the top and bottom beams. The displacement response of two observation points, respectively, located at 1/3*L* and 1/2*L* of the top beam are given in Figure 7.



(a) model 1, point 1/3 *L*

(b) model 1, point 1/2 *L*

(c) model 2, point 1/3 *L*

(d) model 2, point 1/2 *L*

Figure 7. Point displacement of the validation model

According to Fingure 7, the numerical results of present SEM agree very well with the results provided by the conventional FEM. It should be emphasized that when SEM is used to compute the vibration response of the validation model, a single computational run takes 21.7 seconds only. comparing with 401.3 seconds of FEM, SEM uses only 5.4% computaional time of the conversional FEM to obtain the same accurate results.

Parametric study

Benefit from the computational time and accuracy of present spectral element method, the vibration response of the sandwich structure subject to external sound wave excitation with wide frequency range can be calculated more effectively. Furthermore, by using the Rayleigh's integral, as given by Eq. (18), the transmitted sound power can be easily obtained. To provide a reference for the structural-acoustical design of the sandwich structure with adhesively-bonded corrugated cores given in this paper, parametric studies are implemented to reveal the effect of the thickness of the face plates and the inclined angle of stiffener on the transmitted sound power. The details of the reference model for the parametric study is tabulated in Table 2.

Table 2. Reference model for parametric study

Parameter	value
Total length of the sandwich structure (*L*)	1200.0 mm
Total height of the sandwich beam (*H*)	50.0 mm

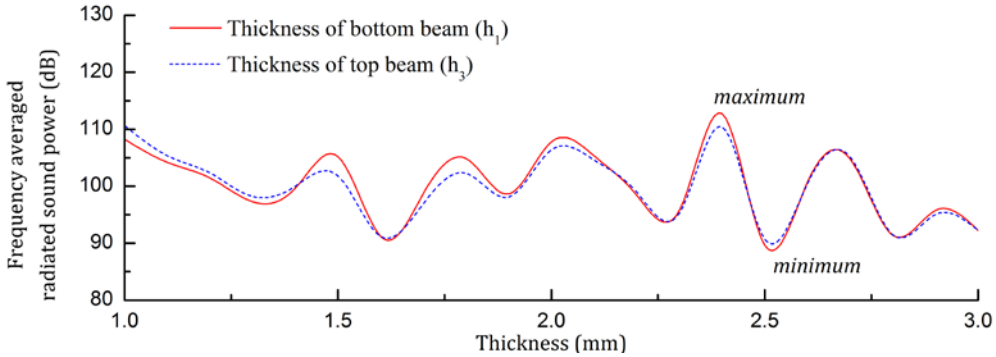
Number of inclined stiffener (N)	11
Inclined angle of the stiffener (θ)	**Variate, 30° ~ 90°**
Thickness of the top face beam (h_3)	**Variate, 1.0 ~ 3.0 mm**
Thickness of the bottom face beam (h_1)	**Variate, 1.0 ~ 3.0 mm**
Thickness of the core beam (h_2)	3.0 mm

As illustrated in Figure 4, considering a unit magnitude plane wave of acoustic pressure is impinged on the bottom beam surface with an incident angle of α=30°. The frequency range of the excitation is 1~800Hz, and the frequency averaged sound power is introduced here as an evaluation index for the acoustic performance of the sandwich structure, which is:

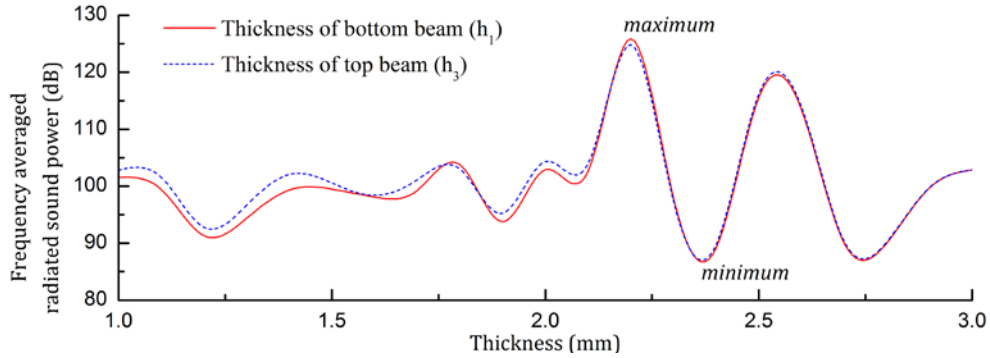
$$W_a = \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} W(\omega) \mathrm{d}\omega \tag{18}$$

Effect of the thickness of the face plate

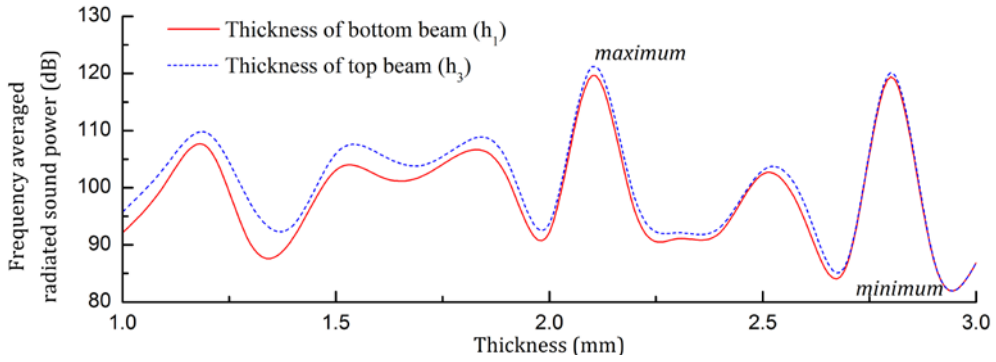
Figure 8 gives the effect of the thickness of the face plates on the radiated sound power with the target frequency range from 0 to 800Hz. As is shown in the graph, for the univariate study, because of the periodicity of the sandwich structure, h_1 and h_3 affect the frequency averaged sound power in much the same way.



(a) Inclined angle of stiffener θ=30°



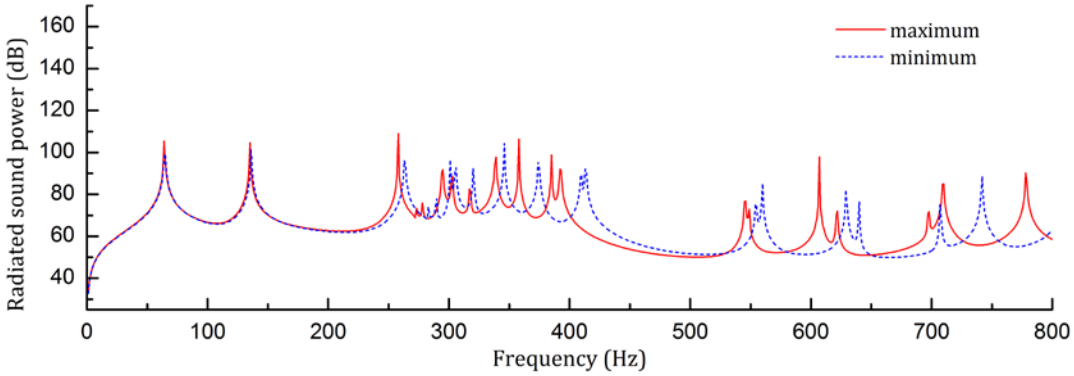
(b) Inclined angle of stiffener θ=40°



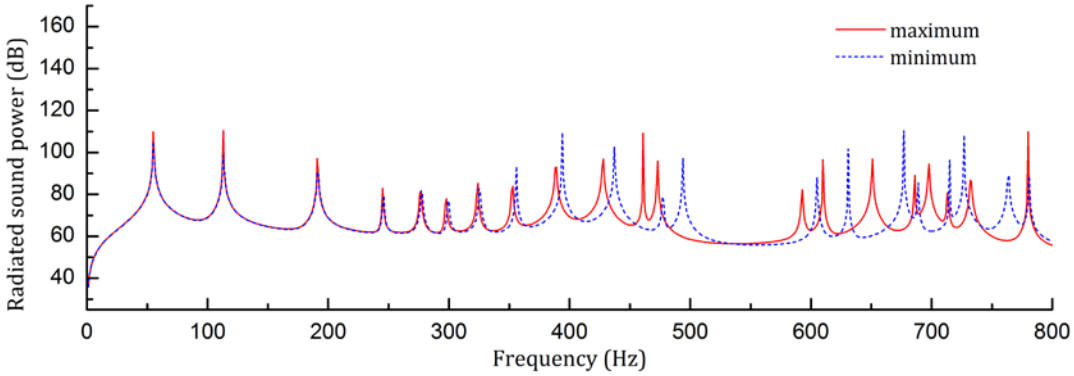
(c) Inclined angle of stiffener θ=50°

Figure 8. Influence of the thickness of the face plates

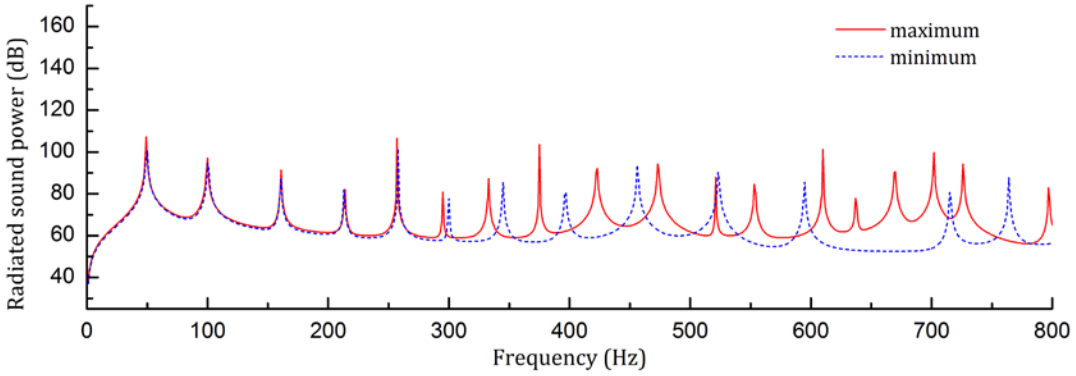
Just for one specific inclined angle, as the increase of the thickness, there are less differences between the two curves in each sub-figure, and also, the influence become more obvious and the variation tendency become more violent. The spectral distribution of structural radiated sound power of the maximum and minimum points in Figure 8 are plotted in Figure 9. It can be seen that the main difference between the maximum curve and minimum curve is in the high frequency range, the changing of the thickness of the face plates has a limited impact on the radiated sound power in the low frequency range.



(a) Inclined angle of stiffener θ=30°



(b) Inclined angle of stiffener θ=40°



(c) Inclined angle of stiffener θ=50°

Figure 9. Spectral distribution of the radiated sound power of the maximum and minimum point in Figure 8

Effect of inclined angle θ

By fixing both the thickness of the top and bottom beam at 3mm, the effect of the inclined angle of the stiffener is illustrated in Figure 10. In order to avoid the interference between two adjacent stiffeners, the variation range of the inclined angle is limited from 30 to 90 degrees.

Due to the structural inhomogeneity of the adhesively-bonded corrugated core layer, as there is any change of the inclined angle, not only the structure layout, but also the structural mass and stiffness are changed simultaneously. These complex relationships eventually lead to an erratic curve as shown in Figure 10 with the target frequency range from 0 to 500Hz.

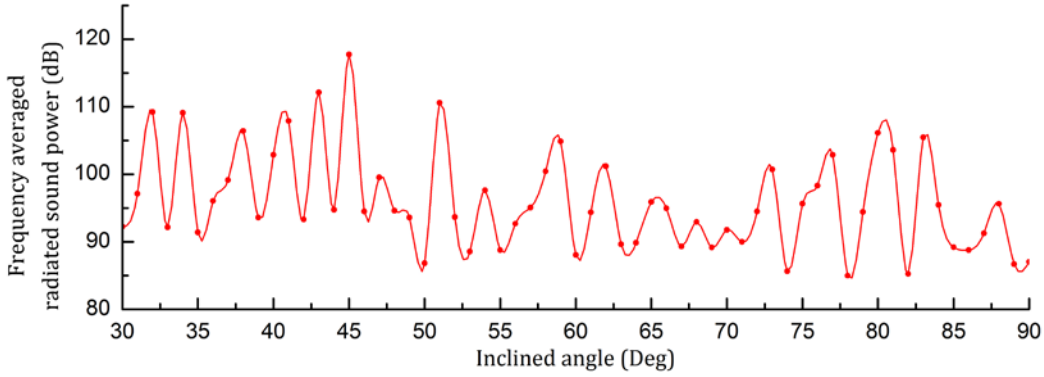


Figure 10. Influence of the inclinde angle of the stiffener

Since h_1 and h_3 have the same effect on the radiated sound power, either one could be used to perform the thickness and angle conbined influence study on the radiated sound power. As shown in Figure 11, in zone 1, with big thickness and less inclination of the stiffeners, the change tendency of the sound power is relatively gently. To the contrary, in zone 2, small thickness and small inclined angle lead to less structural stiffness and dramatic flactuation of sound power in this zone.

According to the data of Figure 11, the maximum and minimum value are 128.8dB and 89.1dB, which appear when h_1=2.0mm, θ=87° and h_1=1.5mm, θ=84°, respectively. The spectral distribution of radiated sound power is given in Figure 12, obviously, the main difference between the two curves is high frequency range. In fact, for the sandwich structure, the sound radiation in low frequency range gives the greatest contribution to the averaged radiated sound power and it is hard to control.

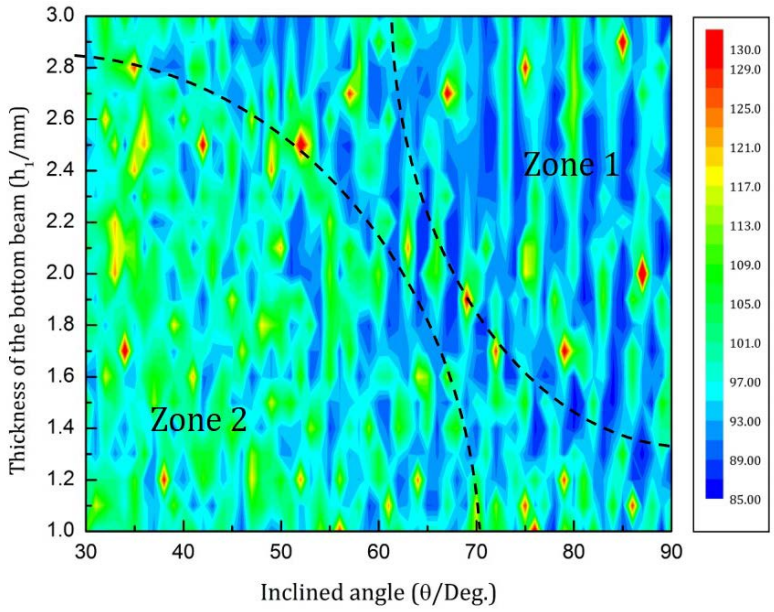


Figure 11. Frequency averaged sound power related to both the thickness of the face plate and the inclined angle of the stiffener

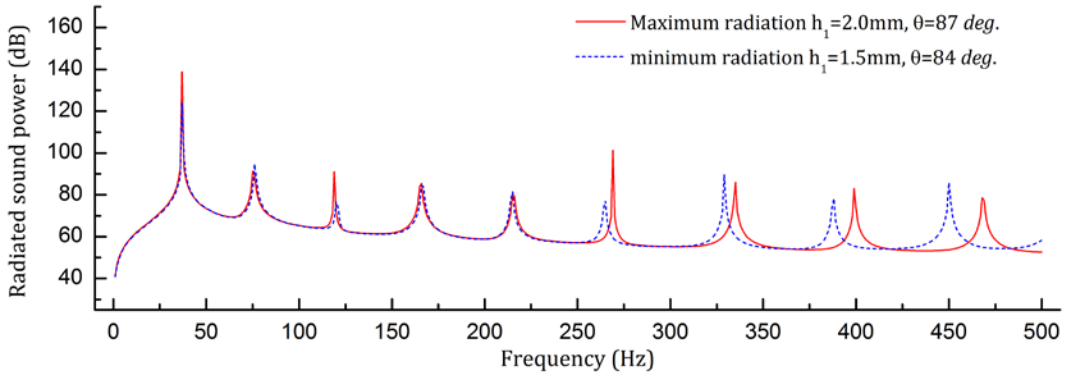


Figure 12. Spectral distribution of the radiated sound power

Conclusions

In this paper, a finite-length numerical 2D model of sandwich plates with adhesively-bonded corrugated cores is developed using spectral element method. The numerical result shows that SEM has a much higher computational efficiency than the traditional FEM without losing any accuracy, which makes SEM an efficient method for the vibro-acoustic analysis of such periodic structure. By using the Rayleigh's integral, parametric studies are performed to test the influence of two main design parameters, the thickness of the face plates and the inclined angle of the stiffener, on the radiated sound power when the structure is subjected to external sound wave excitation. The result shows that for the periodic structure, both the thickness of the top and bottom beam have almost the same effect on the frequency averaged radiation sound power. Comparing with the thickness of the face plates, the averaged sound power is more sensitive to the inclined angle of the stiffener. From the perspective of reducing the radiated sound power, the sound radiation in high frequency range can be affected more easily by changing these two parameters. Since the sound radiation in low frequency range gives the greatest contribution to the averaged radiated sound power, suppressing the sound radiation in low frequency range would be more important to reduce sound radiation of the whole structure.

Acknowledgments

This research is supported by the National Science Foundation of China (Grant No. 51275289).

References

[1] Rumerman, M. L., 1975, "Vibration and wave propagation in ribbed plates," J. Acoust. Soc. Am., **57**(2), pp. 370–373.

[2] Mace, B. R., 1980, "Sound radiation from a plate reinforced by two sets of parallel stiffeners," J. Sound Vib., **71**(3), pp. 435–441.

[3] Mead, D. M., 1996, "Wave Propagation in Continuous Periodic Structures: Research Contributions From Southampton, 1964-1995," J. Sound Vib., **190**(3), pp. 495–524.

[4] Pellicier, A., and Trompette, N., 2007, "A review of analytical methods, based on the wave approach, to compute partitions transmission loss," Appl. Acoust., **68**(10), pp. 1192–1212.

[5] Wang, J., Lu, T. J., Woodhouse, J., Langley, R. S., and Evans, J., 2005, "Sound transmission through lightweight double-leaf partitions: Theoretical modelling," J. Sound Vib., **286**(4-5), pp. 817–847.

[6] Legault, J., and Atalla, N., 2010, "Sound transmission through a double panel structure periodically coupled with vibration insulators," J. Sound Vib., **329**(15), pp. 3082–3100.

[7] Xin, F. X., and Lu, T. J., 2011, "Transmission loss of orthogonally rib-stiffened double-panel structures with cavity absorption.," J. Acoust. Soc. Am., **129**(4), pp. 1919–34.

[8] Thamburaj, P., and Sun, J. Q., 2002, "Optimization of Anisotropic Sandwich Beams for Higher Sound Transmission Loss," J. Sound Vib., **254**(1), pp. 23–36.

[9] Franco, F., Cunefare, K. a., and Ruzzene, M., 2007, "Structural-Acoustic Optimization of Sandwich Panels," J. Vib. Acoust., **129**(3), p. 330.

[10] Denli, H., and Sun, J. Q., 2007, "Structural-acoustic optimization of sandwich structures with cellular cores for minimum sound radiation," J. Sound Vib., **301**(1-2), pp. 93–105.

[11] Fahy, F., and Walker, J., 2004, Advanced Applications in Acoustics, Noise and Vibration, CRC Press, New York.

[12] Patera, A. T., 1984, "A spectral element method for fluid dynamics: laminar flow in a channel expansion," J. Comput. Phys., **54**, pp. 468–488.

[13] Birgersson, F., Finnveden, S., and Nilsson, C. M., 2005, "A spectral super element for modelling of plate vibration. Part 1: General theory," J. Sound Vib., **287**, pp. 297–314.

[14] Zak, A., 2009, "A novel formulation of a spectral plate element for wave propagation in isotropic structures," Finite Elem. Anal. Des., **45**(10), pp. 650–658.

[15] Wu, Z., Li, F., and Wang, Y., 2013, "Study on vibration characteristics in periodic plate structures using the spectral element method," Acta Mech., **1101**, pp. 1089–1101.

[16] El-Raheb, M., and Wagner, P., 1997, "Transmission of sound across a trusslike periodic panel; 2-D analysis," J. Acoust. Soc. Am., **102**(4), pp. 2176–2183.

[17] Ruzzene, M., 2004, "Vibration and sound radiation of sandwich beams with

honeycomb truss core," J. Sound Vib., **277**(4-5), pp. 741–763.

[18] Lee, U., 2009, Spectral Element Method in Structural Dynamics, J. Wiley & Sons Asia, Hoboken, N.J.

[19] Doyle, J. F., 1997, Wave Propagation in Structures: Spectral Analysis Using Fast Discrete Fourier Transforms, Springer, New York.

[20] Fahy, F., and Gardonio, P., 2007, Sound and Structural Vibration:Radiation, Transmission and Response, Academic Press, London.

Stochastic homogenization in the framework of domain decomposition to evaluate effective elastic properties of random composite materials : application to a 2D case of fiber composites

P. Karamian-Surville[1,2,3,a)] and W. Leclerc[1,2,3]

[1]Normandie Univ, France

[2]UNICAEN, LMNO, F-14032 Caen, France

[3]CNRS, UMR 6139, F-14032, Caen France

[a)]Corresponding author: philippe.karamian@unicaen.fr. Presenting author: philippe.karamian@unicaen.fr

ABSTRACT

The paper deals with the setting up of a stochastic homogenization method in the framework of domain decomposition. We focus our investigation on the random fibre composites in the elasticity field. We generate a random representative volume elements (RVE) of the composite and evaluate its elastic properties by the double-scale homogenization. We propose an adaptation of this latter in the domain decomposition framework in order to drastically reduce the calculation cost which is important in this context.

Keywords: Domain decomposition, RVE, Random composites materials,Stochastic homogeneization.

Introduction

Random fibre composites are difficult to model and study. The complexity of their strongly entangled network of fibres leads to technical drawbacks related to the mesh generation. In addition, their study requires the generation of large and numerous RVE during the numerical evaluation of the effective properties. Domain decomposition methods are efficient tools to decrease the calculation time which is important (give a value for example gain of 50% or 30%) in this context. Two adaptations of the homogenization method are proposed in this paper: a modified Schur complement method, and a combination of the FETI-1 method, and the method of Schur complement. In this article, we present both concepts and provide some relevant results demonstrating their ability in the context of random fiber composites. First, a 2D square RVEs with the help of random parameters describing the morphology of the network of fibres is generated. A meshing process, according to voxelisation approach of RVEs is made: the model with an n-order approximate geometry [2, 4, 6]. Then a finite element study is realized in order to estimate elastic properties with the help of the double-scale homogenization [1, 7]. In order to use the double-scale homogenization method we had to make two main adaptations. First, when generating the RVEs we take care of the continuity of fibers between each sub-domains, second we have to eliminate redundant information over the edges. The calculation is performed according to one of two proposed domain decomposition methods. The paper is outlined as follows. First, we present the minimization problem associated to the double-scale homogenization and describe both modified domain decomposition methods. Second, we provide some numerical results in effective properties.

Domain decomposition methods

This section is devoted to the implementation of the homogenization method in order to adapt it to the domain decomposition method. A brief recall of the equations governing a linear elastic boundary value problem is done. Two approaches of domain decomposition are proposed to solve it. The first method is the modified Schur complement method, whereas the second one, is a mixture of FETI-1 method and the Schur complement.

Setting up of the problem

Let Ω be a open bounded set of R^2 or R^3 . Let $\partial\Omega = \partial_1\Omega \cup \partial_2\Omega$ designates the border of Ω. The periodic multi-scale homogenization is a powerful tool to evaluate effective properties [1, 7]. The method consists in expanding some constitutive

equations according to several scales of the medium. In the present contribution, we consider the framework of the linear elasticity. Thus,

$$\begin{cases} -div\sigma(\mathbf{u}^\epsilon) = \mathbf{f} \quad \text{a.e in} \quad \Omega \\ \sigma_{ij}(\mathbf{u}^\epsilon) = C^\epsilon_{ijkh} e_{khx}(\mathbf{u}^\epsilon) \\ e_{khx}(\mathbf{u}^\epsilon) = \frac{1}{2}(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) \\ \sigma_{ij}(\mathbf{u}^\epsilon) n_j|_{\partial_1\Omega} = F_j \\ \mathbf{u}^\epsilon|_{\partial_2\Omega} = \mathbf{O} \end{cases} \tag{1}$$

where σ is the stress tensor, e is the strain-displacement, C^ϵ_{ijkh} is the local stiffness tensor, $\mathbf{f}$ is the loading and $\mathbf{u}^\epsilon$ is the displacement which is expanded according to the ϵ parameter. We suppose a pseudo-periodic medium and consider a two-scale expansion of Equation 1. The first scale called macroscopic is denoted as x, and the second one called microscopic is denoted as y. Variational considerations lead to a new formulation of the equations at the macro-scale which take into account the local disruptions related to the heterogeneities. Hence, we can extract the following formulation of the effective stiffness tensor,

$$\tilde{C}_{ijkl} = \frac{1}{|Y|} \int_Y C_{ijmn} \left[\delta_{mk}\delta_{nl} + e_{mny}\left(\omega^{kl}(y)\right) \right] dy \tag{2}$$

Y denotes the periodic cell and $|Y|$ its volume. C_{ijmn} is the local stiffness tensor which depends on both the medium (heterogeneity or matrix) and the corresponding behaviour law. ω^{kl} is a local solution in the cell with periodic boundary conditions. Thus, the effective tensor turns out to be the sum of the mean of properties and a corrective term related to the local disruption at the microscopic scale.

Partitioning of the RVEs

We generate 2D square RVEs according to a set of random parameters describing the complex microstructure of a random fibre composite. The RVEs are conceived and meshed according to the technique outlined in [4] and with the help of CASTEM we generate properly the RVEs. The basic idea consists in approximating the real geometry according to a grid of quadrangular elements. Such a concept turns out to be suitable in the framework of domain decomposition due to the uniformity of the elements. Thus, we evenly subdivide the RVEs into several square subdomains without remeshing. The similarity between the RVE and each subdomain enables us to denote them as sub-RVEs. Figure 1 illustrates an example of partitioning of a RVE into four sub-RVEs. One can notice that we consider non-overlapping domains, both the periodicity and the continuity of fiber at the interfaces are ensured by taking a special care to maintain the geometrical continuity so that they match together once the sub-domains are together. Γ^i designates the set of inner boundaries, and Γ^e the set of outer boundaries. $\Gamma = \Gamma^i \cap \Gamma^e$ represents the gathering of the both previous sets. Ω_n represents the area of the nth subdomain.

Modified Schur complement method

In a first approach, we adapt the Schur complement method to the calculation of effective properties by the double-scale homogenization. Once Equation 1 is descretized using finite elements, for the considered example (see Figure 1) in which we consider four subdomains, this one leads to a discrete system which reads as follows:

$$\underbrace{\begin{bmatrix} K_{11} & 0 & 0 & 0 & \tilde{K}'_{\Gamma 1} \\ 0 & K_{22} & 0 & 0 & \tilde{K}'_{\Gamma 2} \\ 0 & 0 & K_{33} & 0 & \tilde{K}'_{\Gamma 3} \\ 0 & 0 & 0 & K_{44} & \tilde{K}'_{\Gamma 4} \\ \tilde{K}_{\Gamma 1} & \tilde{K}_{\Gamma 2} & \tilde{K}_{\Gamma 3} & \tilde{K}_{\Gamma 4} & \tilde{K}_{\Gamma\Gamma} \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} \omega_1^{kl} \\ \omega_2^{kl} \\ \omega_3^{kl} \\ \omega_4^{kl} \\ \tilde{\omega}_\Gamma^{kl} \end{bmatrix}}_{\mathbf{u}} = \underbrace{\begin{bmatrix} \mathbf{f}_1^{kl} \\ \mathbf{f}_2^{kl} \\ \mathbf{f}_3^{kl} \\ \mathbf{f}_4^{kl} \\ \tilde{\mathbf{f}}_\Gamma^{kl} \end{bmatrix}}_{\mathbf{f}} \tag{3}$$

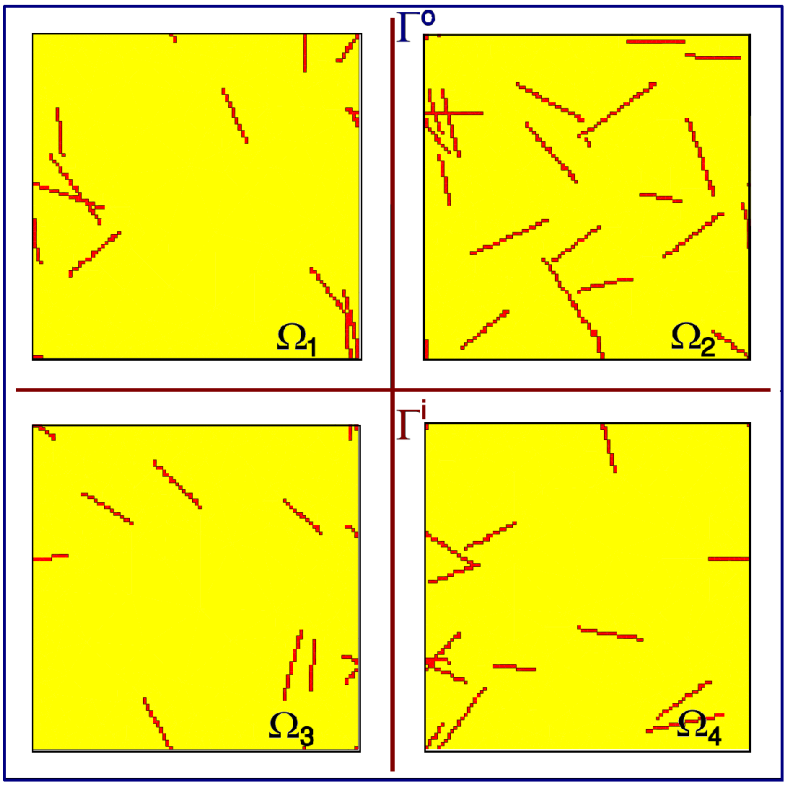


Figure 1: Partitioning of a RVE in 4 subdomains

K_{ii} designates the stiffness tensor of the ith subdomain, ω_i^{kl} is the microscopic displacement and $\mathbf{f}_i^{kl}$ the applied strength which is zero in the present context. $\tilde{K}_{\Gamma i}$ is typically the tensor of the nodes located on the boundary Γ for each subdomain i. $\tilde{\omega}_\Gamma^{kl}$ is the vector of solutions in both displacement on the boundary and homogenized coefficients. Generally speaking, the $\tilde{}$ denotes 3 additional terms in 2D (6 in 3D) relative to the assessment of elastic coefficients. Practically, the solving is realized with the help of a new system $S\,\mathbf{u}_\Gamma = \mathbf{b}$ where S is the Schur complement and $\mathbf{b}$ its corresponding second member which is equal to $\tilde{\mathbf{f}}_\Gamma^{kl}$. We have,

$$S = \tilde{K}_{\Gamma\Gamma} - \sum_i \tilde{K}_{\Gamma i} K_{ii}^{-1} \tilde{K}'_{\Gamma i} \tag{4}$$

Mixed FETI-1 and Schur complement method

We propose an adaptation of the FETI method in the framework of the double-scale homogenization. The method is the dual of the Schur complement one in the sense that the interface problem is formulated in Lagrange multipliers and not in displacements. We consider the basic form of the process called FETI-1 [2, 3]. Different modifications have to be performed to adapt the approach to the double-scale homogenization. First, the hypothesis of periodicity leads to practical difficulties related to an excessive number of rigid body modes when taking into account by Lagrange multipliers. A possible way to get round the drawback is to rewrite the problem in another base which leads to the appearance of unsuitable coupling terms between subdomains. Our choice is to consider the periodicity on the outer boundaries Γ^e by the primal Schur complement. Hence, we talk about a mixed FETI-1 and Schur complement method. Second, we must consider additional terms related to the homogenized coefficients. The terms are added to the tensor describing the connections on the outer boundaries and consequently taken into account by the Schur complement as well. Thus, the only connections on the inner boundaries are described by Lagrange multipliers. Under these hypotheses, the matrix-vector system to solve is similar to the previous one,

$$\begin{bmatrix} K_{11} & 0 & 0 & 0 & R'_1 \\ 0 & K_{22} & 0 & 0 & R'_2 \\ 0 & 0 & K_{33} & 0 & R'_3 \\ 0 & 0 & 0 & K_{44} & R'_4 \\ R_1 & R_2 & R_3 & R_4 & K_R \end{bmatrix} \begin{bmatrix} \omega_1^{kl} \\ \omega_2^{kl} \\ \omega_3^{kl} \\ \omega_4^{kl} \\ \Lambda^{kl} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1^{kl} \\ \mathbf{f}_2^{kl} \\ \mathbf{f}_3^{kl} \\ \mathbf{f}_4^{kl} \\ \mathbf{f}_R^{kl} \end{bmatrix} \tag{5}$$

where R_i and K_R are two tensors describing the connections at the interfaces and, $\mathbf{f}_i^{kl}$ and $\mathbf{f}_R^{kl}$ the second members corresponding to the subdomains and the interfaces respectively. Such a system can not be directly solved by the conjugate gradient method because of the floating subdomains. A classical FETI interface problem has to be performed on the Lagrange multipliers, for which a second level of multipliers are provided by the rigid body modes. The new system is then solved by a preconditioned conjugate gradient and leads us to a direct assessment of the homogenized coefficients.

Numerical results

Algorithms of the two previous methods have been implemented in C++ language. The present section provides some numerical results obtained from the modified Schur complement method. Effective elastic properties are assessed and compared with a direct calculation.

Hypotheses

We consider a set of 2D unit RVEs for which the fibres are randomly oriented and distributed. A special care is carried out to guarantee the periodicity, this treatment is ensured during the generation of the RVE with the help of the n-order approximate geometry to build meshes. The length and the width of each heterogeneity is fixed at 0.2 and 0.01 respectively, and we suppose no curvature. Each RVE is subdivided into 4, 9, 16 and 36 non-overlapping subdomains. We suppose the continuity of the medium as well as the connection of the meshes on the boundaries of each part. Thus, one heterogeneity can be located on several domains and crosses several inner boundaries. The density of fibres is randomly distributed between 0 and 30 fibres per unit cell. The meshes are generated according to the concept of 0-order approximate geometry. In other words, each inclusion is approximated by a grid of quadrangular elements the size of which is equal to the diameter of the heterogeneities. The elastic properties are assessed in two steps. First, we evaluate some coefficients for each RVE with the help of the modified Schur complement method. Second, we take the average of the results obtained from a complete set of representative patterns. The suitable number of realizations is estimated according to statistical considerations. Each fibre is supposed to follow a transverse isotropic behaviour law. The longitudinal and transverse Young's modulus are set at 1050 and 600 GPa respectively. The shear modulus is set at 450 GPa. The matrix is an isotropic polymer resin with Young's and shear moduli set at 4.2 and 1.55 GPa respectively. We deliberately choose a high-contrast of properties to maximise the conditioning of the matrix in the solving which one is realized by a preconditioned conjugate gradient.

Elastic moduli

A sample of 86 RVEs is built according to the previous hypotheses. Figure 2 exhibits the evolution of the effective Young's modulus depending on the density of fibres for different levels of partitioning. A comparison is realized with a usual direct calculation performed on the same sample of representative patterns without partitioning. One must keep in mind that we consider the same grid of quadrangular elements whatever the level of subdivision is such that the degrees of freedom number is constant. Globally the differents curves fit together which highlights the consistency of the method. However, the greater both the number of fibres and subdomains are the more some small discrepancies are observable between the different curves. Thus, the relative error is 3.48% between the calculation realized with 36 subdomains and the direct calculation for a density of fibres set at 30.

"Once the RVEs are split into several subdomains in order to guarantee the continuity of fibres through the interfaces, we have to replace some elements labelled as matrix in fibre. This process ensures the continuity of cross fibres, but modifies the rate of matrix for the global RVES, especially for a high contrast of properties of fibre and a large number of subdomains, what leads an effect on the calculation of the effective mechanical properties and explains why we observe a discrepancies on numerical results between the global RVEs, and RVEs, themselves divided into several subdomains."Figure 2 illustrates

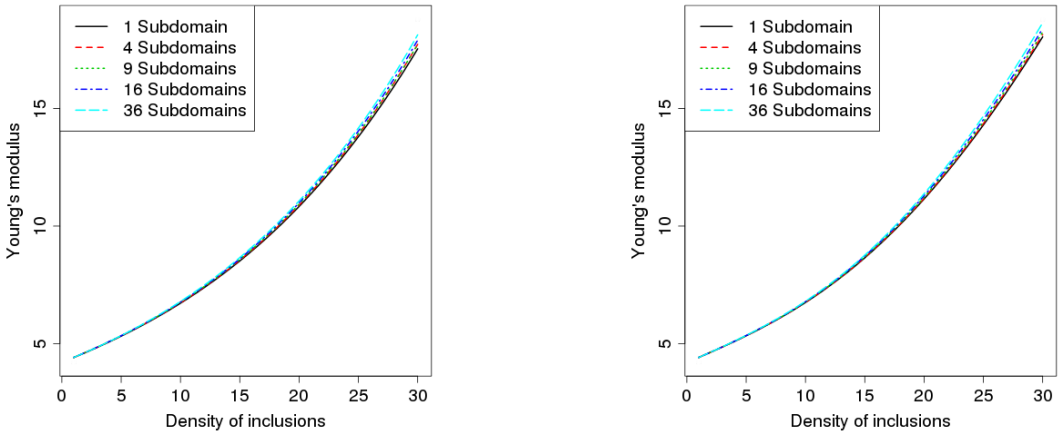


Figure 2: Influence of the density of fibres on the effective Young's modulus in the case of a direct calculation

the same results in the case of a direct calculation of the matrix-vector system of Equation 3 with partitioning. One can observe the same discrepancies as previously seen in the case of a domain decomposition calculation.

Conclusion

Two domain decomposition methods have been adapted and set up to evaluate elastic properties of a random fibre composite with the help of the double-scale homogenization. Both modified Schur complement method and mixed FETI-1 Schur complement method take into account some additional tensors related to both the homogenized coefficients and the hypothesis of periodicity, but are solved as classical ones. Numerical results highlight the consistency of the modified Schur complement method in the framework of a high entanglement of fibres and a high contrast of properties between the matrix and the heterogeneities.

Acknowledgement

The authors wish to thank the High-Performance Computing Centre of Normandie (CRIAAN) for the computing means put at their disposal.

References

[1] A. Bensoussan, J.L. Lions and G.C. Papanicolaou. Asymptotic analysis for periodic structures. Springer, North Holland, Amsterdam edition, 1978.

[2] C. Fahrat and F.X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. International Journal of Numerical Methods in Engineering, 32:1205 – 1227, 1991.

[3] C. Fahrat and F.X. Roux. Implicit parallel processing in structural mechanics. Computational Mechanics Advances, North Holland, 2(1):1 – 124, 1994.

[4] W. Leclerc, P. Karamian and A. Vivet. An efficient stochastic and double-scale model to evaluate the effective elastic properties of 2D overlapping random fibre composites. Computational Science of Materials, 481 – 493, 2013.

[5] W. Leclerc, P. Karamian, A. Vivet and A. Campbell. Numerical evaluation of the effective elastic properties of 2D overlapping random fibre composites. Technische Mechanik, 32:358 – 368, 2012.

[6] L. Mishnaevsky Jr. Automatic voxel-based generation of 3D microstructural FE models and its application to the damage analysis of composites. Material Science and Engineering A, 407(1-2):11 – 23, 2005.

[7] E. Sanchez-Palencia. Non-homogeneous media and vibration theory. Springer, Berlin, 127, 1980.

Study on post-failure evolution of underwater landslide with SPH method

***Y. An[1], C.Q. Shi[1], †Q.Q. Liu[1,2], and S.H. Yang[1]**

[1] Key Laboratory for Mechanics in Fluid Solid Coupling Systems, Institute of Mechanics, Chinese Academy of Sciences, China.
[2] School of Aerospace Engineering, Beijing Institute of Technology, China.

*Presenting author: anyi@imech.ac.cn
†Corresponding author: qqliu@imech.ac.cn

Abstract

Underwater landslide is a serious nature hazard which could occur at both sea floor and reservoir banks and results in massive destruction. It generally involves large deformation of landslide and water body, especially the interface between them. A numerical model for describing the soil-water interface and its large deformation in the framework of smoothed particle hydrodynamics (SPH) method is employed to simulate the evolution of underwater landslides. The elasto-plastic-viscous model with Dracker-Prager plastic yield rule is used for soil deformation simulation. And the direct forces exchange between interfacial soil and water particles is implemented to characterize the interface deformation. Both quasi-steady and dynamic behaviors of soil and soil-water interface during underwater landslide post-failure stage are revealed. Simulated results shows that the landslide body experiences strong deformation during the impact process.

Keywords: underwater landslide, soil-water coupling, interface, smoothed particle hydrodynamics.

Introduction

Underwater landslide could occur at both sea floor and reservoir banks, causing massive destruction. In this hazard, the underwater landslide failures during earthquake or underwater excavation or all kinds of porous pressure accumulation. The plastic bands will form in the landslide body and cause fast movement of the landslide like it in subaerial landslides. However fast opposite movement of landslide body and water surrounding it is a unique feature comparing with subaerial landslides. As the density of landslide body and the water is basically in the same magnitude order comparing with the subaerial landslide, the water resistance effect is much stronger than it of subaerial landslides. Thus the interface between slide and water must be considered in the simulation. What is more, this process also involves very large deformation of the landslide and water body, especially the interface between them. If we ignore the seepage force in the landslide body which might always be true for fine grain soil, the problem is basically dealing with a gravity controlled deformable interface between soil and water. So, two crucial characteristics, i.e. the soil-water interfacial coupling and large deformation of interface, should be well addressed in the simulation of the underwater landslides evolution process.

However these two features raise great challenge to classical mesh based simulation methods such as FEM due to the large deformation nature. Many studies simplify this problem into the interaction between two fluids, i.e. Newtonian fluid and non-Newtonian fluid, and thus both mesh based FVM with VOF interfacial model and mesh-free methods such as SPH could be employed. Rzadkiewicz et al. [1] have used such an approach to simulate the landslide generated wave problem. However, the non-Newtonian model for granular flow is not designed for quasi-steady problems as the static stress state could not be represent truly due to its fluid nature, thus could not been used to predict not only the failure form of the slope but also localized particle-solid state during the slide evolution. As geo-engineers often prefer

elasto-plastic models for landslide simulation which could describe the quasi-steady state of soil very well, a coupled model including elasto-plastic-viscous soil model and Naiver-Stokes equation based fluid model with an interaction model between soil and water is the best choose. Thus a recently developed mesh free soil-water coupling model which could deal with both quasi-steady and dynamic behaviors of soil, water and the interface is employed in the framework of smoothed particle hydrodynamics method.

Numerical Model for Soil

The model is constituted of three parts: model for soil, model for water flow, and model for the interface between. Different with non-Newtonian models which are commonly used for granular flow simulation, this study employs the elasto-plastic-viscous model for soil deformation simulation as the latter could reproduce more phenomena in landslide evolution, i.e. from stable state to granular flow. The steady state is ruled by elastic model, while the quasi-steady state is ruled by plastic model with Dracker-Prager plastic yield criterion. The post-plastic behavior of soil (particle-fluid state) is modelled with the plastic-viscous model which is similar with non-Newtonian models. This model could also profit from extensive existing constitutive laws and plastic yield rules which all have large amount of experimental data to support. Pioneering work of Bui et al. [2] has introduced SPH method into the elasto-plastic simulation of soil slopes. Following Bui's work, this study uses similar approach for landslide simulation, the detailed equation is list below:

$$\frac{D\rho_i}{Dt} = \sum_{j=1}^{N} m_j (v_i^{\alpha} - v_j^{\alpha}) \frac{\partial W_{ij}}{\partial x_i^{\alpha}} \tag{1}$$

$$\frac{Dv_i^{\alpha}}{Dt} = \sum_{j=1}^{N} m_j \left(\frac{\sigma_i^{\alpha\beta} + \sigma_j^{\alpha\beta}}{\rho_i \rho_j} - \Pi_{ij}\delta^{\alpha\beta} + F_{ij}^{n} R_{ij}^{\alpha\beta}\right) \frac{\partial W_{ij}}{\partial x_i^{\beta}} + g^{\alpha} \tag{2}$$

where $F_{ij}^{n} R_{ij}^{\alpha\beta}$ is the artificial stress term, helping to remove the tensile instability when soil is stretched; $F_{ij} = W_{ij} / W(\Delta x, h)$, and the exponent n is set as 2.55 in this paper. $R_{ij}^{\alpha\beta} = R_i^{\alpha\beta} + R_j^{\alpha\beta}$ where $R_i^{\alpha\beta}$ and $R_j^{\alpha\beta}$ are the components of the artificial stress tensor for particles i and j, respectively. $\sigma^{\alpha\beta}$ is the total stress tensor, while the elastic–plastic soil constitutive model with the Drucker–Prager criterion can be expressed as:

$$\frac{D\sigma_i^{\alpha\beta}}{Dt} = \sigma_i^{\alpha\gamma}\dot{\omega}^{\beta\gamma} + \sigma_i^{\gamma\beta}\dot{\omega}_i^{\alpha\gamma} + 2G\dot{e}_i^{\alpha\beta} + K\varepsilon_i^{\gamma\gamma}\delta_i^{\alpha\beta} - \dot{\lambda}_i \left[3\alpha_{\psi} K\delta^{\alpha\beta} + \frac{G}{\sqrt{J_2}} s_i^{\alpha\beta}\right] \tag{3}$$

where $\dot{e}^{\alpha\beta}$ is the deviatoric shear strain rate tensor, $\dot{s}^{\alpha\beta}$ is the deviatoric shear stress rate tensor, $\delta^{\alpha\beta}$ is Kronecker's delta. $\dot{\lambda}$ is the rate of the plastic multiplier λ dependent on the state of stress and load history:

$$\dot{\lambda}_i = \begin{cases} \dfrac{3\alpha_{\phi} K\dot{\varepsilon}_i^{\gamma\gamma} + (G/\sqrt{J_2}) s_i^{\alpha\beta}\dot{\varepsilon}_i^{\alpha\beta}}{9\alpha_{\phi}\alpha_{\psi} K + G} & f(I_1, J_2) = 0 \\ 0 & f(I_1, J_2) < 0 \end{cases} \tag{4}$$

where the $\dot{\varepsilon}^{\alpha\beta}$ and $\dot{\omega}^{\alpha\beta}$ are the elastic strain rate tensor and the spin rate tensor, respectively. $f(I_1, J_2)$ is the yield function, I_1 and J_2 are the first and second invariants of the stress tensor, respectively; α_{φ} and k_c are Drucker–Prager's constants, which are related to the Mohr–Coulomb material constants c (cohesion) and φ (internal friction), and α_{ψ} is a dilatancy factor related with the dilatancy angle.

Numerical Model for Water and Soil-Water Coupling

The traditional weak compressible SPH model is used for modeling water flow. The artificial viscosity model calibrated with Viroulet et al.'s [3] experiment is employed to describe the viscous effect as we found that laminar+SPS model could not give better results for this complex problem with limited particle sizes. For example, if we choose the sub-particle scale viscosity model, we need at least 0.04mm spatial resolution in the first case to make the first grid space from boundaries located in logarithmic zone (y+~[10-100]). This resolution will make the calculation cost unbearable even if we can handle the numerical viscosity properly so as to not overestimate the viscosity in other zones. Secondly, although the viscosity of water is important in the underwater landslide evolvement problems, its influences concentrate in the shear stress between water and soil. While the normal stress might play a more important role in describing the soil deformation, especially when the soil is modeled as an elasto-plastic-viscous material which is "stiffer" than Bingham fluid. Thus, although introducing the artificial viscosity may not be elegant enough, but neither it is notably worse than other choices nor it alters the significance of interactions between soil and water, especially in the situation of this study.

The interfacial coupling method is crucial for this problem. We use explicit time evolution scheme and the consistency of both the displacement of the interface and the pressure on the interface to setup the coupling model (Fig. 1). The displacement of interfacial particles is determined by the soil phase calculation while the stress on the interface is corrected to represent the effect of water pressure. Then the obtained displacement is used as the displacement of the interfacial moving wall for water phase calculation. As dynamic boundary condition is employed for wall boundaries, we simply use the interfacial soil particles to act as boundary particles for water phase to support the water calculation. Thus, we can directly obtained force pairs between water and boundary particles. These force pairs is exactly the same as them between water particles and the interfacial soil particles. So we can use them to correct the interfacial stress for the soil phase calculation. In this way, a direct coupling model is implemented in the framework of SPH method. Although this method is very simple, but due to its external interfacial coupling nature, it is robust and easy to extend to more complex situation such as considering three phases or rigid stones.

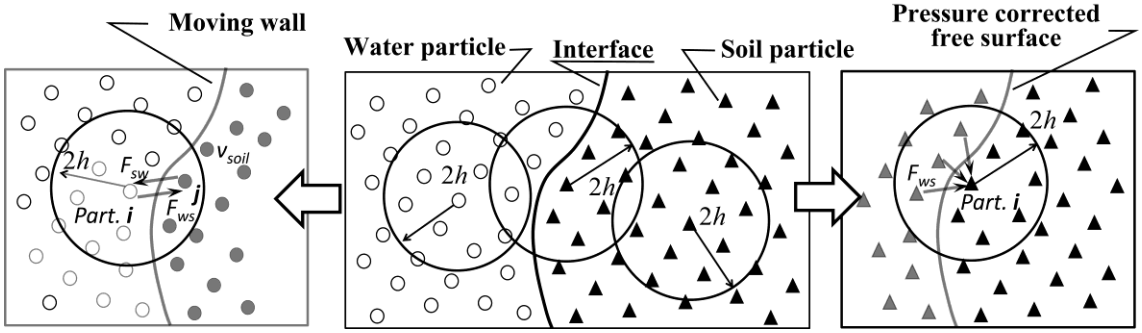


Figure 1. Illustration of the coupling model

Model Validation and Application on Underwater Landslide

Fritz et al.'s (2009) [4] laboratory experiment is used to validate this model and good agreements between simulated results and experimental data are obtained on slide shape evolutions, as shown in Fig. 2. We can see that although the simulated result is slightly different from experiment at the bottom of the slide: the simulated result has water cushion at the head of slide while it is not observed in the experiments, the simulated slide head position

and the thickness is similar to the experimental data which proves the validity of this model applying on the soil-water coupling simulation.

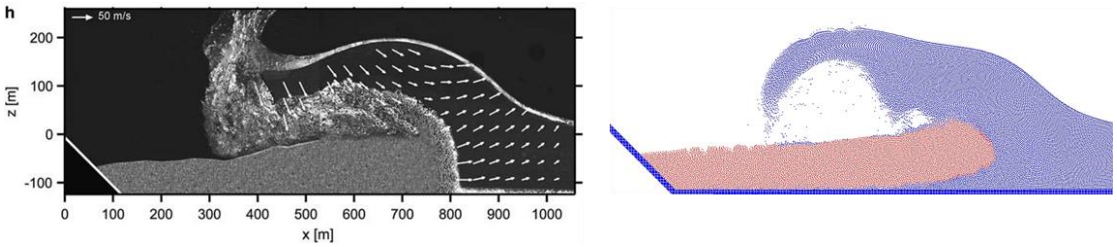


Figure 2. Comparison of simulated result and Fritz's experiment

The proposed model is use to study the subaqueous landslide evolution. We choose a typical experiment work on underwater landslide evolution by Rzadkiewicz et al. (1997) [1]. A submerged triangle slide made of fine-grain sands is placed on a 45° slope and a vertical board is placed at one end of the slide to keep it steady. When the board is suddenly removed, the landslide body collapses. The lower part of the landslide body deforms firstly, and the left part of the landslide body moves afterwards. In this process, the interaction force between grain landslide and water phase results large deformation of the head of the landslide body.

The comparison of experimental snapshots and results from the proposed model is shown in Fig. 3a,b. The accumulated plastics strain (ADPS) which could be considered as indication of the shear induced plastics band are shown in simulated results. It is clear that the plastics zones are located at two interfaces: (1) interface between the slide and the bottom due to the shear of the wall, which also leads many inner plastics zones, (2) interface between the slide surface and the water. As the time goes, the inner plastics zone gets larger which is a conjunct result of the slide slowdown and the slide getting thinner.

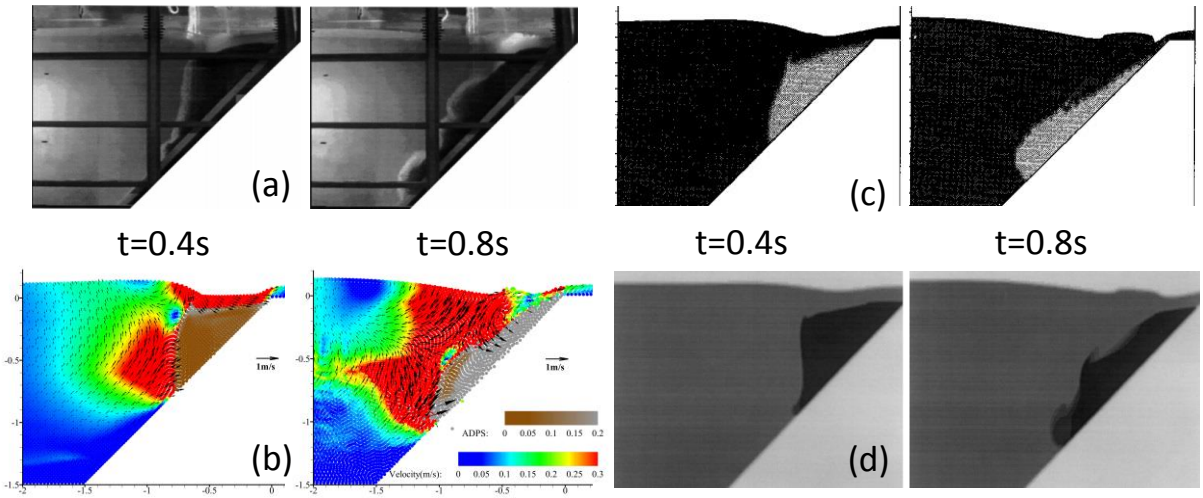


Figure 3. Simulated underwater landslide evolution against experimental data and previous numerical studies

Previous studies on the same problem, which are carried out by Rzadkiewicz et al. (1997) [1] and Mariotti et al. (1999) [6] respectively, are also presented (Figure 3c,d). It should be note that although these models could all reproduce the main features in the experiments, the parameters used in their studies are very different. How to choose suitable values of parameters is one of the difficulties when using non-Newtonian fluid models. Besides, most of non-Newtonian fluid based model could not represent the shear band in the landslide body, not even the initial failure prediction.

Discussion

Due to the mesh free characteristics, this model do not need mesh and remesh, and is robust enough in very large deformation situation which is not easy to achieve in traditional FEM methods. And because the SPH method is a mesoscale model, parameters in this method is easy to obtained. Comparing with previous studies using non-Newtonian fluid models to describe soil deformation, all parameters in our model have their physical meaning in soil mechanics and can be obtained from conventional soil mechanics experiments. Besides, as we use the elasto-plastic-viscous constitutive law and Drucker-Prager yield model of soil, the deformation of the soil is better represented than it of non-Newton models.

In Fig. 3b, a different shape of the landslide leading edge can also been seen between the experiment and simulation. That is because the velocity of water phase is large and the confining pressure of soil grain is weak at the leading edge of the landslide, which leads to rolling up of the grains in the experiments, while the proposed model cannot reproduce this mechanism properly, which results a smaller thickness. However, dense fluid as these rolling up grains are, they will have little influence on neither internal stress of soil nor the leading wave.

Conclusions

The post-failure evolution of underwater landslides is numerically studied based on a soil-water interfacial coupled smoothed particle hydrodynamics method. The elasto-plastic-viscous model with Dracker-Prager plastic yield rule instead of traditional non-Newtonian model is used for soil deformation simulation and good agreement between simulated results and experiment are obtained. Simulation results show that the landslide body experiences strong deformation during the impact process.

Acknowledgement

This study was financially supported by the National Basic Research Program of China (No.2014 CB04680202) and the National Natural Science Foundation of China (No.11372326, No.11432015).

References

[1] Rzadkiewicz SA, Mariotti C, Heinrich P. Numerical simulation of submarine landslides and their hydraulic effects. J Waterw Port Coast Ocean Eng 1997;123(4):149-57.
[2] Bui HH, Fukagawa R, Sako K, Wells JC. Slope stability analysis and discontinuous slope failure simulation by elasto-plastic smoothed particle hydrodynamics (SPH). Géotechnique, 2011;61(7): 565-74.
[3] Viroulet S, Cébron D, Kimmoun O, Kharif C. Shallow water waves generated by subaerial solid landslides. Geophys J Int 2013;193(2):747-62.
[4] Fritz HM, Mohammed F, Yoo J. Lituya Bay landslide impact generated mega-tsunami 50th anniversary. Pure Appl Geophys 2009;66(1-2):153-75.
[5] Rzadkiewicz SA, Mariotti C, Heinrich P. Numerical simulation of submarine landslides and their hydraulic effects. J Waterw Port Coast Ocean Eng 1997;123(4):149-57.
[6] Mariotti C, Heinrich P. Modelling of submarine landslides of rock and soil. Int J Numer Anal Meth Geomech 1999;23:335-54.

Under the strong earthquake conditions, this paper use the GPU acceleration technology and discrete element method of continuous media mechanics to study the soil along the metro line and it's shock absorption stability, in order to play a guiding role in metro long-term safety operation.

Discontinuous deformation theory

The block in the calculation of discontinuous deformation is formed by one or more finite element units, continuous structure is used in the block, and discontinuous structure is used on the block boundary.

Governing equation

The governing equation of the discontinuous deformation calculation theory is the motion equation, the block body is subjected to internal force and external force. Internal force include the force which is caused by the deformation of the block and the damping force, external force include the out boundary force and the force between springs. In mechanics, because of the block body is regarded as a continuous, isotropic linear elastic body, so its mechanical properties are described by the basic differential equations of three-dimensional elastodynamics theory, That is:

Equilibrium equation: $\sigma_{ij,j} + f_i - \rho u_{i,n} - \mu u_i = 0$

Geometric equation: $\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$

Physical equation: $\sigma_{ij} = \lambda \delta_{ij} \varepsilon_{kk} + 2G\varepsilon_{ij}$

Boundary condition: $u_i = \overline{u}_i$ (on the displacement boundary of Γ_u), $\sigma_{ij} n_j = T_i$ (On the force boundary of Γ_σ)

In the formula, σ_{ij}, u_i, f_i and T_i respectively represent stress, displacement, volume force and area force; Ω and Γ respectively represent the rock block region and its boundary, $\Gamma=\Gamma_u \cup \Gamma_\sigma$, λ and G are lame constant, ρ and μ respectively represent mass density and damping coefficient, δ_{ij} is Kronecker delta symbol. Based on the elastic variational principle, the governing equation of the calculation is the motion equation of block body:

$$[\boldsymbol{M}]\{\ddot{\boldsymbol{u}}(t)\} + [\boldsymbol{C}]\{\dot{\boldsymbol{u}}(t)\} + [\boldsymbol{K}]\{\boldsymbol{u}(t)\} = \{\boldsymbol{Q}(t)\} \tag{1}$$

In the formula, $\{\dot{u}(t)\}$, $\{u(t)\}$, $\{\ddot{u}(t)\}$ respectively represent acceleration array, speed array, displacement array of all the nodes of block body. $[M]$, $[C]$, $[K]$, $[Q]$ respectively represent mass matrix, damping matrix, stiffness matrix and nodal load array.

The calculation of each time step for solving the governing equations is divided into two parts. The first step is to loop each deformable block body, and complete the corresponding continuous deformation calculation. The Second step is to calculate the force of contact surface. Firstly, from the stiffness matrix and the nodal displacement obtain the elastic force, then, from damping matrix and nodal velocity obtain damping force, finally, combining the direct integral method and external force to solve motion equation. Specific equations are:

Elastic force:

$$\begin{bmatrix} K_{1,1} & K_{1,2} & \cdots & K_{1,n} \\ K_{2,1} & K_{2,2} & \cdots & K_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ K_{n,1} & K_{n,2} & \cdots & K_{n,n} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{bmatrix} \tag{2}$$

Damping force:

$$\begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ C_{n,1} & C_{n,2} & \cdots & C_{n,n} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix} = \begin{bmatrix} f_1' \\ f_2' \\ \cdots \\ f_n' \end{bmatrix} \tag{3}$$

Combining the direct integral method and external force to solve motion equation:

$$\begin{Bmatrix} a_i = (f_i + f_i' + f_i^{out}) / m_i \\ v_i = v_i^{t-1} + a_i t \\ u_i = u_i^{t-1} + v_i t \end{Bmatrix} \tag{4}$$

As shown in the formula (4), through the resultant force to obtain the acceleration, velocity and displacement of block body nodes. f_i^{out} include the forces of boundary surface and the forces of contact surface, boundary conditions provide boundary force.

Model boundary

Figure 1 show the normal and tangential spring of the interface. F_n^j and F_s^j are normal and tangential forces of springs, K_n^j and K_s^j are normal and tangential stiffness of springs, Δd_n^j and Δd_s^j are normal and tangential displacements of springs.

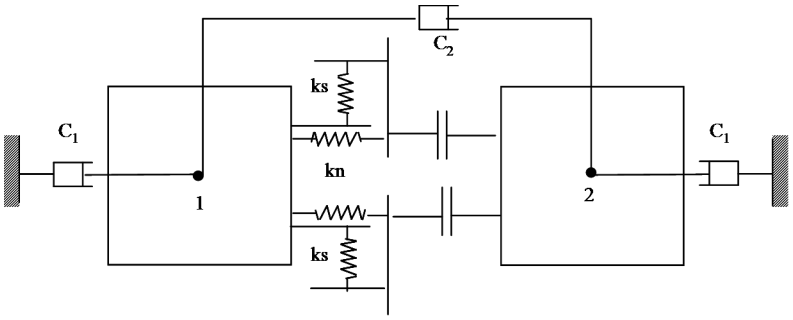


Figure. 1 The normal and tangential spring of the interface

Three dimensional calculation model and parameter selection

The study object is an excavation section of metro engineering, its numerical calculation model size is 24m × 17m × 17m. Circular cross section is adopted to calculate, and its size is ∅ 3m×15m. Elastic plastic model as the calculation model, and the calculation model is divided into four layers，from top to bottom: gravel-boulder bed (5m), roof layer (3m), excavation layer (6m), bottom layer (3m) [4,5]. A total of nine measuring points set on the top plate, the bottom plate and the two sides of model, (The distance between the measuring points is 0.5m. From left to right, the number of the measurement points on both sides of the model are respectively No.1 to No.6.,the bottom plate measuring point is from top to bottom for 7 to 12), row spacing of U-shaped Steel is 2m. The three-dimensional numerical computation model is shown in figure 2. The local geological data is

the reference of parameters of the calculation model, and its values are shown in Table 1.

The boundary conditions of model are respectively: bottom surface is full constraint, flank is horizontal constraint, and the top surface is free. Considering the surrounding building load (200-meter- high building, overhead bridges and traffic load), the initial stress of the model boundary as follow: the maximum horizontal stress is 20 MPa, the minimum horizontal stress is 18 MPa, the vertical stress is 17 MPa.

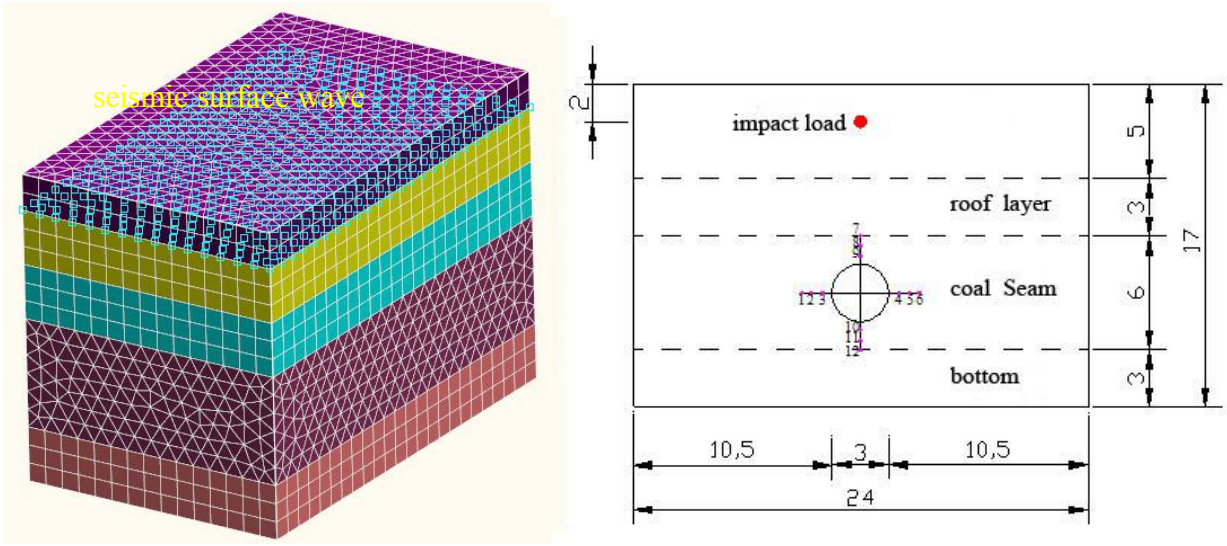


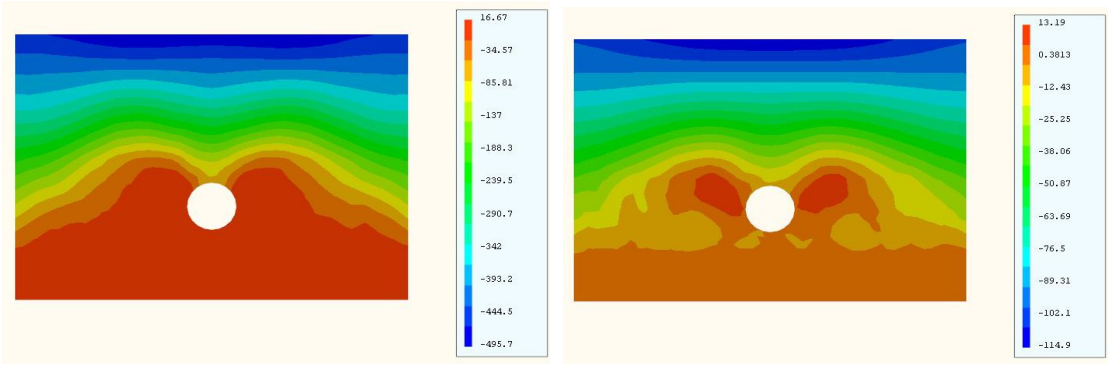
Fig.2 3-D calculation model

Table1 Computing model parameters

Material name	Elastic modulus /E(GPa)	Poisson ratio / μ	Density / ρ (kg/m3)	cohesive strength /C(kPa)	Internal friction angle / φ (°)	Yield strength /(MPa)
Gravel-boulder bed	80	0.25	2300	50	30	42
Roof layer	100	0.2	2440	55	40	60
Excavation layer	3.5	0.28	1700	29	25	20
Bottom layer	90	0.22	2200	52	35	57
Concrete energy absorbing layer	68	0.35	400	—	—	25
Duct piece	210	0.31	7850	—	—	350

Study on the propagation law of the vibration stress wave in soil

Figure 3 shows the calculation results of vertical velocity at different time. From the results we can know that the metro is strong affected by the vibration load. When t=0.5s, the influence of vibration load on metro has enhanced. The vibration load has an upward pushing influence to the floor and both sides of the metro, it also has an downward influence to the metro roof. When t=1s, the vibration influence continue to increase, the influence of vibration load on the metro roof is approximated to the shape of sheep horns, the whole metro have an upward tendency.



(a) 0.5s (b) 1.0s

Fig. 3 The results of vertical velocity at different time

Analysis of the concrete segment support action under the vibration load

Figure 4 shows the relationship between vertical stress and time, as well as the vertical stress curves of the monitoring points. From the data analysis we can know that the vertical stress of monitoring points 7, 8 and 9 are basically negative. The vertical stress of monitoring point 8 fluctuates between positive and negative, and the positive value is about 100MPa. The curve of monitoring point 7 has the largest fluctuation, it vertical stress is negative which the average value is about 200MPa. The vertical stress of monitoring point 10, 11 and 12 also greatly fluctuate between positive and negative. The vertical stress of measuring points indicated that the U steel protection has improved the passive support strength, but under the condition of vibration load, it is easy to produce stress concentration [6-8]. Under the effect of vibration load, the rigid support as an energy storage body will produce serious stress concentration. Once the damage, it will have a great influence on the deformation of the metro.

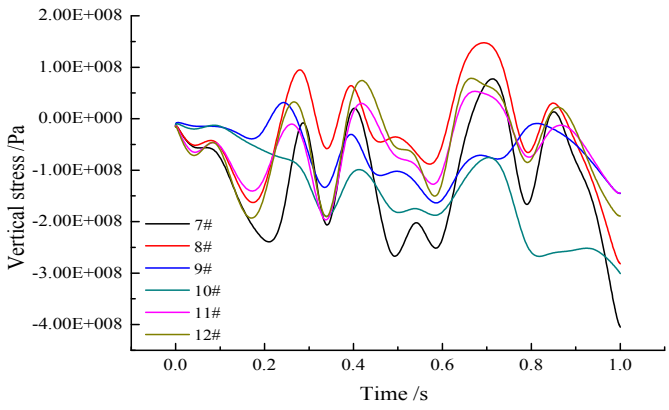
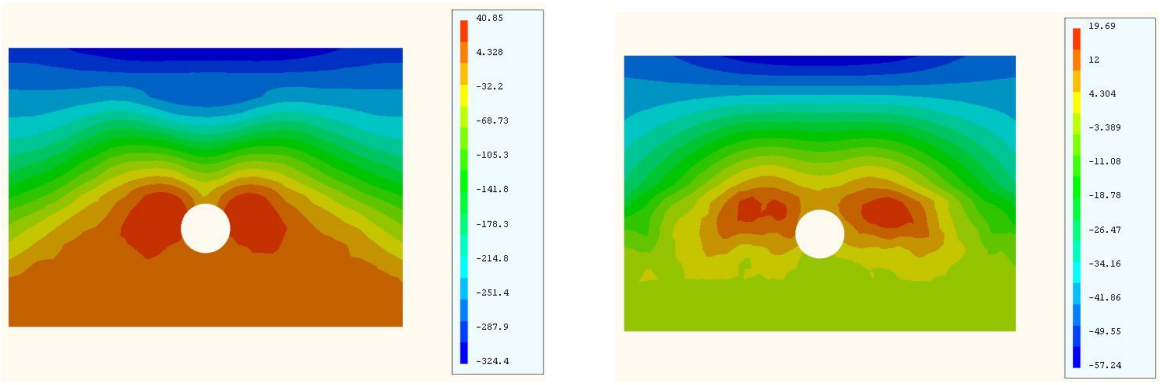


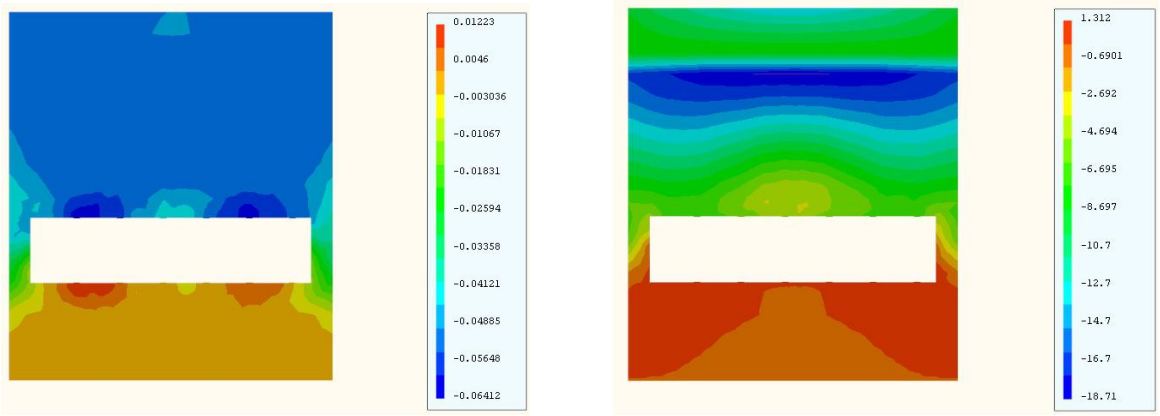
Fig. 4 Vertical stress curves with time

Figure 5 and Figure 6 show the calculation results of vertical velocity and vertical displacement at different times. The results of vertical speed show that the rigid support metro is obviously influenced by vibration load. when t=0.5s, the action of vibration load on metro has enhanced. It has downward action to the metro roof plate, and upward action to the metro bottom plate. Vibration load on the metro both sides has a local concentrate phenomenon, and its distribution is similar to the "bat wing". When t=1s, the effect of vibration load on the metro top roof continue to increase, the distribution of vibration load is similar to the "helmet". Vibration load on both sides of the metro has enhanced, the "bat wing" area is obviously increased. The result of vertical displacement shows that the rock and soil around the metro obviously affected by vibration load, the metro top plate has downward trend, and the metro two sides are squeezed toward inside. The result of metro level profile shows that there are lots of severe displacement deformation area on the metro top and bottom plate, which have a significance influence to the metro deformation failure.



(a) 0.5s (b) 1.0s

Fig.5 vertical velocity results at different times (perpendicular to the metro profile)



(a) 0.1s (b) 0.5s

Fig. 6 vertical displacement results at different time (parallel to the metro profile)

Stability analysis of underground concrete absorption energy layer

Figure 7 shows the relationship between vertical stress and time, as well as the vertical stress curves of the monitoring points. From the data analysis we can know that the vertical stress of monitoring points 7, 8 and 9 are basically negative. The vertical stress of monitoring point 8,9 fluctuates

between positive and negative, and the positive value is about 100MPa. The monitoring point 7 data is negative, its curve fluctuation is the largest and the average value is more than 200MPa. Compared with the monitoring point 7, the value of the monitoring points 8 and 9 greatly reduced, which shows that the concrete energy absorbing layer can effectively reduce the strength of the seismic source wave. The monitoring point 10 data fluctuation is small, monitoring points 11 and 12 data fluctuation is greater, which shows that the seismic wave near the metro bottom plate has weakened [9-11]. Compared with the u-steel support metro, the concrete energy absorbing layer has a larger deformation space, which shows that the deformation of the R-F-R protection metro has obviously reduced, and the metro stability also enhanced[12,13].

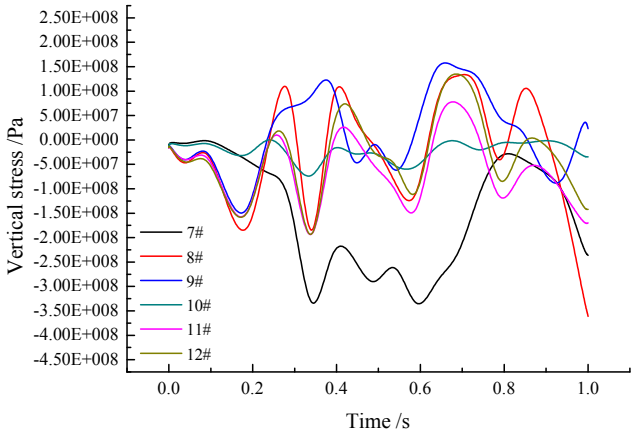
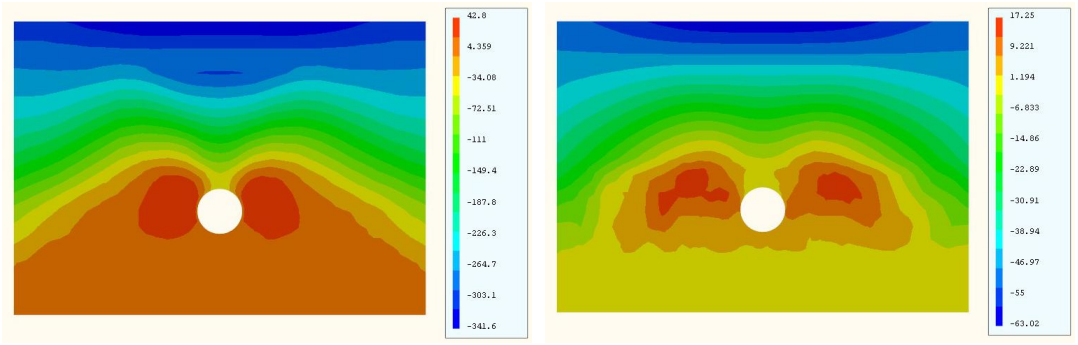


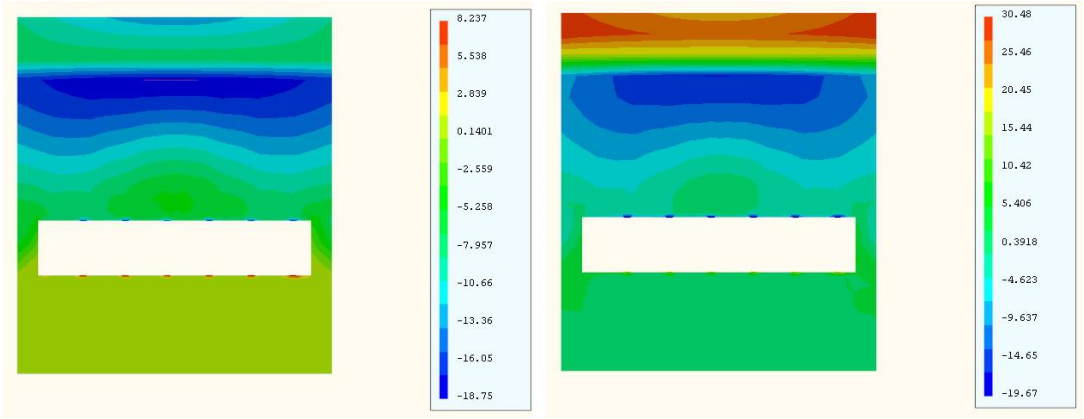
Fig.7 The relationship between vertical stress and time

Figure 8 and figure 9 show the calculation results of vertical speed and vertical displacement at different times. The result of vertical speed shows that the effect of vibration load on the metro has weakened after the concrete energy absorbing layer set up. when t=0.5s, the effect of vibration load on metro has enhanced, the effect of vibration load on the roof is downward, on the bottom plate is upward. When t=0.5s, the vibration load on both sides of the subway is similar to the "bat wing". The distribution area is larger, but the concentrate phenomenon is not obvious. when t=1s, the effect of vibration load continues to increase. The vibration load has wave action to the metro, but the concentrate phenomenon is not obvious, the distribution of vibration load on the metro roof plate is similar to the "helmet", on the metro both sides is similar to the "bat wing". The vertical displacement results shows that the rock and soil around the metro obviously affected by vibration load, the metro top plate has downward trend, and the metro two sides are squeezed toward inside. However, this change has little influence on the metro deformation, this is due to the coordinated deformation of concrete absorbing layer can reduce the surrounding rock deformation. The metro vertical stress and horizontal displacement have obviously reduced, this is due to the concrete energy absorbing layer good coordination deformation performance enable metro can reduce the vibration and vibration intensity, and maintain itself stability [14,15]. Compared with the data of rigid support metro, the vertical force curve volatility decreases and the vertical force of measuring point 8 and 9 also reduced, but the horizontal displacement almost the same. In addition, because the concrete energy absorbing layer has large deformation, so the overall deformation of metro have significantly reduced and there is no obvious deformation concentrated area in the surrounding rock vicinity. This is indicated that the rigid flexible coupling support can fully coordinate deformation, which can reduce the vibration load, improve the impact resistance of deep underground projects, but also conducive to maintaining the stability of the metro.



(a) 0.5s (b) 1.0s

Fig.8 Vertical speed results of vertical roadway section



(a) 0.1s (b) 0.5s

Fig.9 Vertical displacement results of parallel roadway section

Conclusions

Based on the continuum mechanics for discrete element calculation method, the shock resistance stability under the action of seismic wave of rigid support and rigid-flexible coupling support deep buried chamber was analyzed. The strengthening and damping action of different protective bodies on the chamber are studied. And the main conclusions are obtained.

Under the condition of vibration load, rigid support as a strong energy storage body can improve the strength of the passive support, but the damage of shock instability becomes easier, and these will lead to the metro extrusion deformation even the overall closure failure;

Concrete energy absorbing layer can effectively the attenuated seismic wave, in the propagation process or near the metro bottom. Under the action of coordinated deformation, the surrounding rock soil deformation tends to be mild and reduced, the stability of the metro enhanced.

Under the action of strong vibration load, the safety of deep buried chamber can be greatly improved by increasing the strength of rigid support and setting up the flexible deformation buffer layer.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China, project number: 51504029. and Beijing Nova program, project number: xx2016041. It is also supported by the sustainable development of the project in Xicheng District City, Beijing science and Technology Commission, project number: SD2014-36. The comments and suggestions from editors and reviewers helping the improvement of the quality of this paper are appreciated very much by the author.

References:

[1] Wang Sijing, Yang Zhifa, Liu Zhuhua. Rock mass stability analysis for underground engineering[M]. Beijing: Science Press, 1984.

[2] Zheng Yingren, Liu Xinghua. Modern nonlinear science and rock mechanic problems[J]. Chinese Journal of Geotechnical Engineering, 1996(1):98-100.

[3] Qian Qihu. Challenges Faced by Underground Projects Construction Safety and Counter Measures[J]. Chinese Journal of Rock Mechanics and Engineering,2012,31(10):1945-1956.

[4] Ren GingWen. Rigid body element method and its application in stability analysis of rock mass[J].Journal of Hohai University, 1995(1):1-7.

[5] Zhiyin Wang, Yunpeng Ling. Sijing Wang. Numerical simulation of the geomechanical processes in rock engineering[J]. International Journal of Rock Mechanics and Mining Sciences,2000,37:439-507.

[6] MS Diederichs, P K Kaiser. Tensile strength and abutment relax-ation as failure control mechanisms in underground excavations[J]. International Journal of Rock Mechanics and Mining sciences,1999,36:69-96.

[7] Lv Qin, Zhang Dingli, Huang Jun. Mechanism of Stratum Deformation and Its Control Practice in Tunneling Urban Subway at Shallow Depth[J]. China Safety Science Journal,2003,13(7):29-34.

[8] Wang Mengshu, Luo Qiong. Construction of shallow buried excavation method in Beijing Metro[J]. Journal of Railway Engineering Society, 1988,(4):107-117.

[9] Sun Jun. Design theory and practice of underground engineering[M]. Shanghai: Shanghai science and Technology Press, 1996:36-42.

[10] Kong Heng, Wang Mengshu, Yao Haibo, etc. Distribution Laws of Stratum Stress With Working Face Excavation in Subway Tunnels[J]. Chinese Journal of Rock Mechanics and Engineering, 2005,24(3):485-489.

[11] P.A. Fotiu, H. Irschik, F. Ziegle. Modal analysis of elastic-plastic plate vibrations by integral equations. Engineering Analysis with Boundary Elements[J].1994, Volume 14, Issue 1:81-97.

[12] Y. Yang, K.P. Kou, C.C. Lam, V.P. Iu. Free vibration analysis of two-dimensional functionally graded coated and undercoated substrate structures. Engineering Analysis with Boundary Elements[J].2015,Volume 60:10-17.

[13] J. R. Shao, S. M. Li, M. B. Liu. Numerical Simulation of Violent Impinging Jet Flows with Improved SPH Method[J]. International Journal of Computational Methods[J].2016, DOI: 10.1142/S0219876216410012.

[14] Liu Daohua, Niu Shixin, Peng Guangsheng, et al. Construction techniques in tackling the collapse in excavating a highway tunnel[J]. Journal of Geological Hazards and Environment Preservation, 2013,24(1):93-98.

[15] Lv Xiangfeng. Excavation Instability Study of Deep Buried Tunnel Consider the Disturbance Action of Earthquake[J]. Tunnel Construction, 2014,34(2):129-133.

Suspension stability analysis of soil along the metro lines impact by strong vibrations traffic load

LV Xiangfeng[1,2], YANG Dongbo[1], ZHOU Hongyuan[1,2]

[1]Beijing Municipal Engineering Research Institute, Geotechnical Engineering Technology Research Center, Beijing, China.
[2]Beijing Municipal Construction Engineering Quality Third Test Institute, Beijing, China.

*Presenting author: szgcyjylvxiangfeng@163.com

†Corresponding author: 973087860@qq.com

Abstract

Under the action of strong vibration load, the safety threat of deep buried chamber greatly increased, this bring serious challenges to the excavation of deep underground engineering. Therefore, it is urgent to carry out the research on the reinforcement and vibration reduction of deep buried chamber. Based on the continuum mechanics of discrete element method, the vibration reduction and reinforcement of the rigid support and rigid flexible coupling support of deep buried chamber were studied. The calculation results show that the traffic vibration stress wave have a wave effect on the whole metro area, when it act on the metro top, it's distribution will approximate to "horns"; Under the vibration load conditions, the concrete segment as a strong energy storage body can improve the passive support strength, but can also lead to the cracks which is caused by the metro extrusion deformation. After set the concrete energy absorbing layer, the seismic wave which is in the vicinity of the metro bottom and the propagation process have been effectively attenuation, the deformation of the soil around metro also reduced. In addition, increasing the strength of concrete segment can greatly improve metro operation safety.

Key words: Strong vibration transportation load, Soil along the metro lines, Concrete energy-absorbing layer, Shock absorption, Numerical studies

Introduction

Along with the increasing of ground traffic shock load, the influence of the strong vibration traffic load on the soil along the metro line became a technical problem to be solved badly [1-3]. A lot of practice proves that under the complicated geological conditions, the large surrounding rock deformation combined with the influence of ground traffic load let the metro control become more difficulty and even let the metro surrounding rock deformation or broken. Therefore, the further analysis of the influence of strong vibration traffic load on the soil along the metro lines, and the shock absorption stability of metro will provide a new method for the metro operation safety, also have important practical significance.

Domestic and foreign researchers have carried out a lot of research on the ground deformation, which is caused by Underground engineering excavation and traffic load. Qian Qihu have studied the challenges faced by underground projects construction safety and it's corresponding measures. Chen li have investigated the mechanism of deformation body of fill subgrade and the treatment engineering measures. Dahl F et al. studied the classifications of properties influencing the drill ability of rocks based on the test method. However, previous studies generally consider the structural stability of layer based on single factor. In fact, the deformation and damaging process is closely related to the coupling effects of high building, overhead bridges and traffic load, which are still in the infant stage.

Under the strong earthquake conditions, this paper use the GPU acceleration technology and discrete element method of continuous media mechanics to study the soil along the metro line and it's shock absorption stability, in order to play a guiding role in metro long-term safety operation.

Discontinuous deformation theory

The block in the calculation of discontinuous deformation is formed by one or more finite element units, continuous structure is used in the block, and discontinuous structure is used on the block boundary.

Governing equation

The governing equation of the discontinuous deformation calculation theory is the motion equation, the block body is subjected to internal force and external force. Internal force include the force which is caused by the deformation of the block and the damping force, external force include the out boundary force and the force between springs. In mechanics, because of the block body is regarded as a continuous, isotropic linear elastic body, so its mechanical properties are described by the basic differential equations of three-dimensional elastodynamics theory, That is:

Equilibrium equation: $\sigma_{ij,j} + f_i - \rho u_{i,n} - \mu u_i = 0$

Geometric equation: $\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$

Physical equation: $\sigma_{ij} = \lambda \delta_{ij} \varepsilon_{kk} + 2G\varepsilon_{ij}$

Boundary condition: $u_i = \overline{u}_i$ (on the displacement boundary of Γ_u), $\sigma_{ij} n_j = T_i$ (On the force boundary of Γ_σ)

In the formula, σ_{ij}, u_i, f_i and T_i respectively represent stress, displacement, volume force and area force; Ω and Γ respectively represent the rock block region and its boundary, $\Gamma=\Gamma_u \cup \Gamma_\sigma$, λ and G are lame constant, ρ and μ respectively represent mass density and damping coefficient, δ_{ij} is Kronecker delta symbol. Based on the elastic variational principle, the governing equation of the calculation is the motion equation of block body:

$$[\boldsymbol{M}]\{\ddot{\boldsymbol{u}}(t)\} + [\boldsymbol{C}]\{\dot{\boldsymbol{u}}(t)\} + [\boldsymbol{K}]\{\boldsymbol{u}(t)\} = \{\boldsymbol{Q}(t)\} \tag{1}$$

In the formula, $\{\dot{u}(t)\}$, $\{u(t)\}$, $\{\ddot{u}(t)\}$ respectively represent acceleration array, speed array, displacement array of all the nodes of block body. $[M]$, $[C]$, $[K]$, $[Q]$ respectively represent mass matrix, damping matrix, stiffness matrix and nodal load array.

The calculation of each time step for solving the governing equations is divided into two parts. The first step is to loop each deformable block body, and complete the corresponding continuous deformation calculation. The Second step is to calculate the force of contact surface. Firstly, from the stiffness matrix and the nodal displacement obtain the elastic force, then, from damping matrix and nodal velocity obtain damping force, finally, combining the direct integral method and external force to solve motion equation. Specific equations are:

Elastic force:

$$\begin{bmatrix} K_{1,1} & K_{1,2} & \cdots & K_{1,n} \\ K_{2,1} & K_{2,2} & \cdots & K_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ K_{n,1} & K_{n,2} & \cdots & K_{n,n} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{bmatrix} \tag{2}$$

Damping force:

$$\begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ C_{n,1} & C_{n,2} & \cdots & C_{n,n} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix} = \begin{bmatrix} f_1' \\ f_2' \\ \cdots \\ f_n' \end{bmatrix} \tag{3}$$

Combining the direct integral method and external force to solve motion equation:

$$\begin{Bmatrix} a_i = (f_i + f_i' + f_i^{out}) / m_i \\ v_i = v_i^{t-1} + a_i t \\ u_i = u_i^{t-1} + v_i t \end{Bmatrix} \tag{4}$$

As shown in the formula (4), through the resultant force to obtain the acceleration, velocity and displacement of block body nodes. f_i^{out} include the forces of boundary surface and the forces of contact surface, boundary conditions provide boundary force.

Model boundary

Figure 1 show the normal and tangential spring of the interface. F_n^j and F_s^j are normal and tangential forces of springs, K_n^j and K_s^j are normal and tangential stiffness of springs, Δd_n^j and Δd_s^j are normal and tangential displacements of springs.

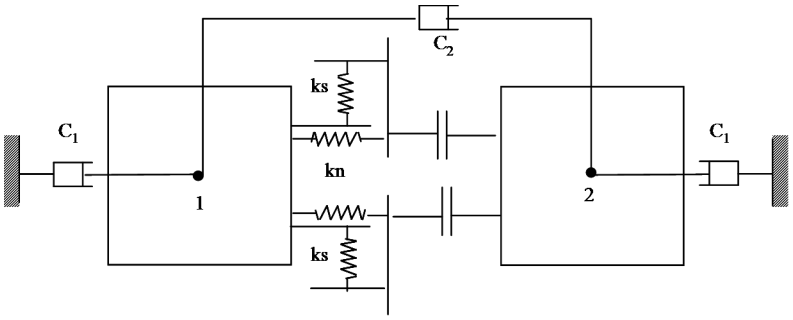


Figure. 1 The normal and tangential spring of the interface

Three dimensional calculation model and parameter selection

The study object is an excavation section of metro engineering, its numerical calculation model size is 24m × 17m × 17m. Circular cross section is adopted to calculate, and its size is ⌀ 3m×15m. Elastic plastic model as the calculation model, and the calculation model is divided into four layers，from top to bottom: gravel-boulder bed (5m), roof layer (3m), excavation layer (6m), bottom layer (3m) [4,5]. A total of nine measuring points set on the top plate, the bottom plate and the two sides of model, (The distance between the measuring points is 0.5m. From left to right, the number of the measurement points on both sides of the model are respectively No.1 to No.6.,the bottom plate measuring point is from top to bottom for 7 to 12), row spacing of U-shaped Steel is 2m. The three-dimensional numerical computation model is shown in figure 2. The local geological data is

the reference of parameters of the calculation model, and its values are shown in Table 1.

The boundary conditions of model are respectively: bottom surface is full constraint, flank is horizontal constraint, and the top surface is free. Considering the surrounding building load (200-meter- high building, overhead bridges and traffic load), the initial stress of the model boundary as follow: the maximum horizontal stress is 20 MPa, the minimum horizontal stress is 18 MPa, the vertical stress is 17 MPa.

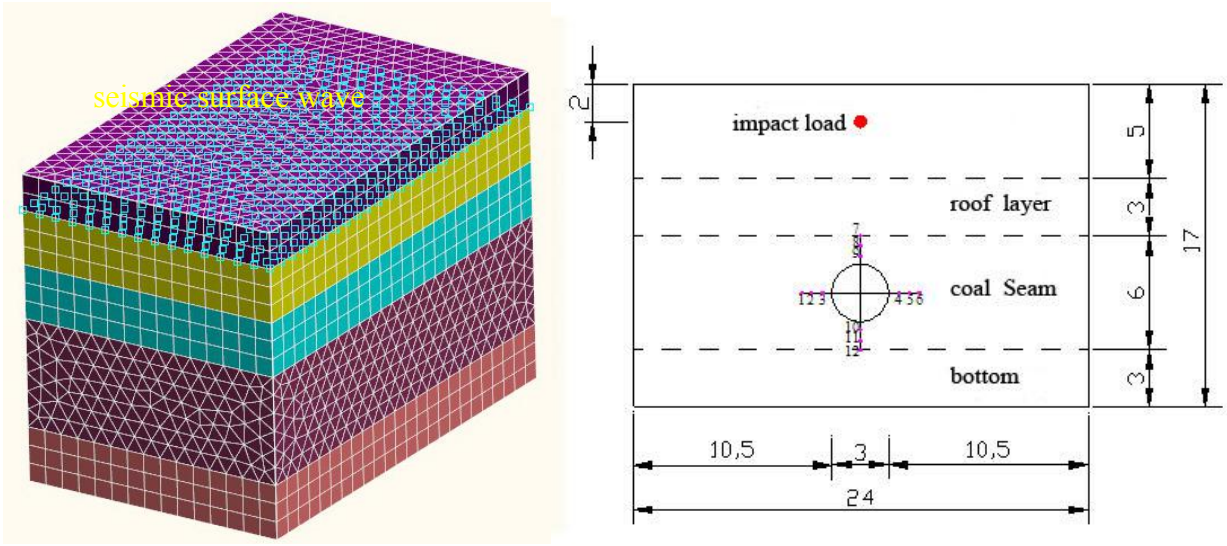


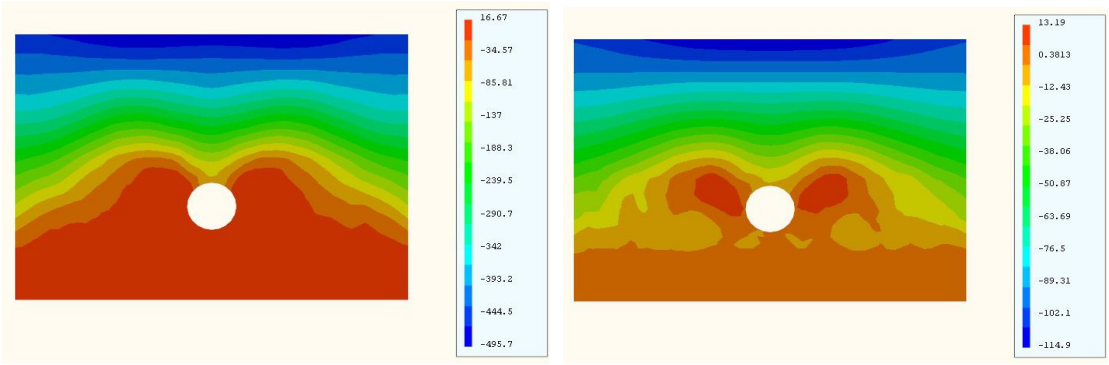
Fig.2 3-D calculation model

Table1 Computing model parameters

Material name	Elastic modulus /E(GPa)	Poisson ratio / μ	Density / ρ (kg/m3)	cohesive strength /C(kPa)	Internal friction angle / φ (°)	Yield strength /(MPa)
Gravel-boulder bed	80	0.25	2300	50	30	42
Roof layer	100	0.2	2440	55	40	60
Excavation layer	3.5	0.28	1700	29	25	20
Bottom layer	90	0.22	2200	52	35	57
Concrete energy absorbing layer	68	0.35	400	—	—	25
Duct piece	210	0.31	7850	—	—	350

Study on the propagation law of the vibration stress wave in soil

Figure 3 shows the calculation results of vertical velocity at different time. From the results we can know that the metro is strong affected by the vibration load. When t=0.5s, the influence of vibration load on metro has enhanced. The vibration load has an upward pushing influence to the floor and both sides of the metro, it also has an downward influence to the metro roof. When t=1s, the vibration influence continue to increase, the influence of vibration load on the metro roof is approximated to the shape of sheep horns, the whole metro have an upward tendency.



(a) 0.5s (b) 1.0s

Fig. 3 The results of vertical velocity at different time

Analysis of the concrete segment support action under the vibration load

Figure 4 shows the relationship between vertical stress and time, as well as the vertical stress curves of the monitoring points. From the data analysis we can know that the vertical stress of monitoring points 7, 8 and 9 are basically negative. The vertical stress of monitoring point 8 fluctuates between positive and negative, and the positive value is about 100MPa. The curve of monitoring point 7 has the largest fluctuation, it vertical stress is negative which the average value is about 200MPa. The vertical stress of monitoring point 10, 11 and 12 also greatly fluctuate between positive and negative. The vertical stress of measuring points indicated that the U steel protection has improved the passive support strength, but under the condition of vibration load, it is easy to produce stress concentration [6-8]. Under the effect of vibration load, the rigid support as an energy storage body will produce serious stress concentration. Once the damage, it will have a great influence on the deformation of the metro.

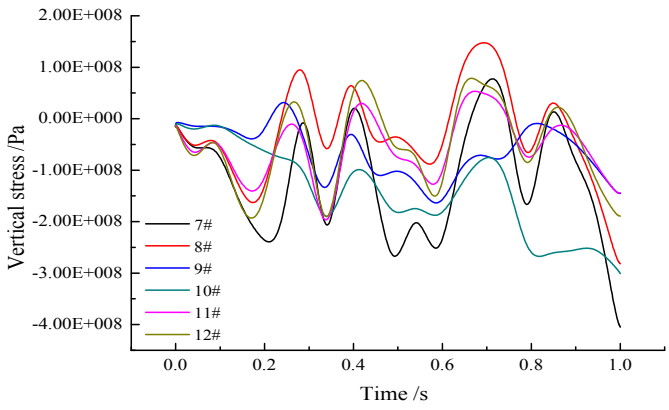
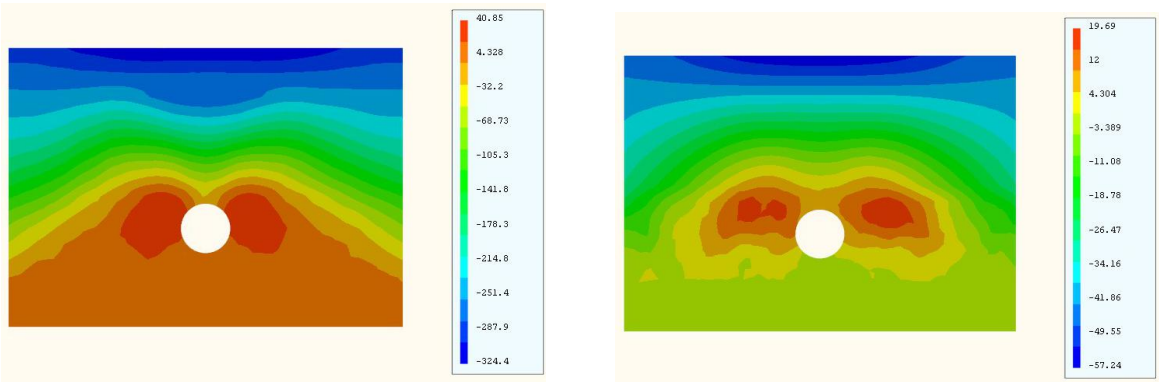


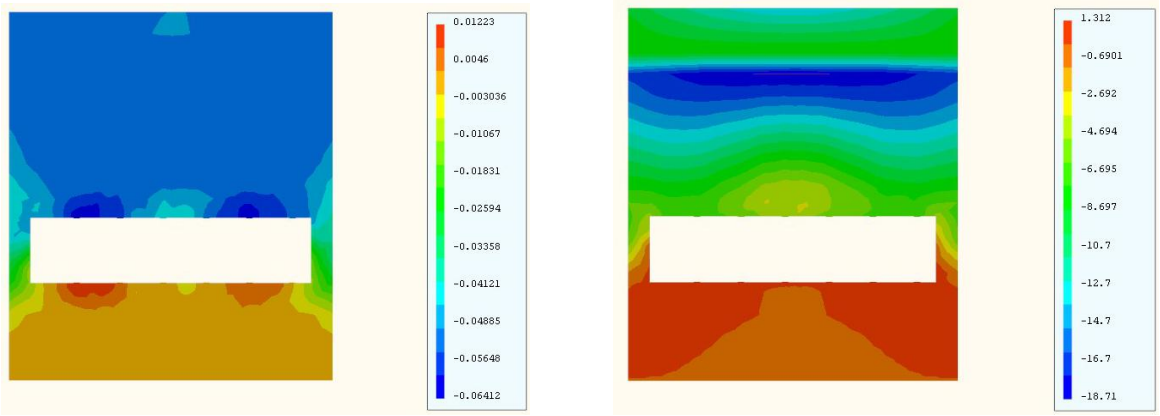
Fig. 4 Vertical stress curves with time

Figure 5 and Figure 6 show the calculation results of vertical velocity and vertical displacement at different times. The results of vertical speed show that the rigid support metro is obviously influenced by vibration load. when t=0.5s, the action of vibration load on metro has enhanced. It has downward action to the metro roof plate, and upward action to the metro bottom plate. Vibration load on the metro both sides has a local concentrate phenomenon, and its distribution is similar to the "bat wing". When t=1s, the effect of vibration load on the metro top roof continue to increase, the distribution of vibration load is similar to the "helmet". Vibration load on both sides of the metro has enhanced, the "bat wing" area is obviously increased. The result of vertical displacement shows that the rock and soil around the metro obviously affected by vibration load, the metro top plate has downward trend, and the metro two sides are squeezed toward inside. The result of metro level profile shows that there are lots of severe displacement deformation area on the metro top and bottom plate, which have a significance influence to the metro deformation failure.



(a) 0.5s (b) 1.0s

Fig.5 vertical velocity results at different times (perpendicular to the metro profile)



(a) 0.1s (b) 0.5s

Fig. 6 vertical displacement results at different time (parallel to the metro profile)

Stability analysis of underground concrete absorption energy layer

Figure 7 shows the relationship between vertical stress and time, as well as the vertical stress curves of the monitoring points. From the data analysis we can know that the vertical stress of monitoring points 7, 8 and 9 are basically negative. The vertical stress of monitoring point 8,9 fluctuates

between positive and negative, and the positive value is about 100MPa. The monitoring point 7 data is negative, its curve fluctuation is the largest and the average value is more than 200MPa. Compared with the monitoring point 7, the value of the monitoring points 8 and 9 greatly reduced, which shows that the concrete energy absorbing layer can effectively reduce the strength of the seismic source wave. The monitoring point 10 data fluctuation is small, monitoring points 11 and 12 data fluctuation is greater, which shows that the seismic wave near the metro bottom plate has weakened [9-11]. Compared with the u-steel support metro, the concrete energy absorbing layer has a larger deformation space, which shows that the deformation of the R-F-R protection metro has obviously reduced, and the metro stability also enhanced[12,13].

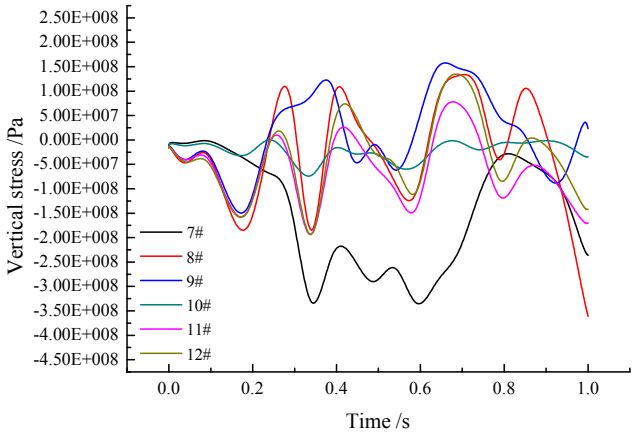
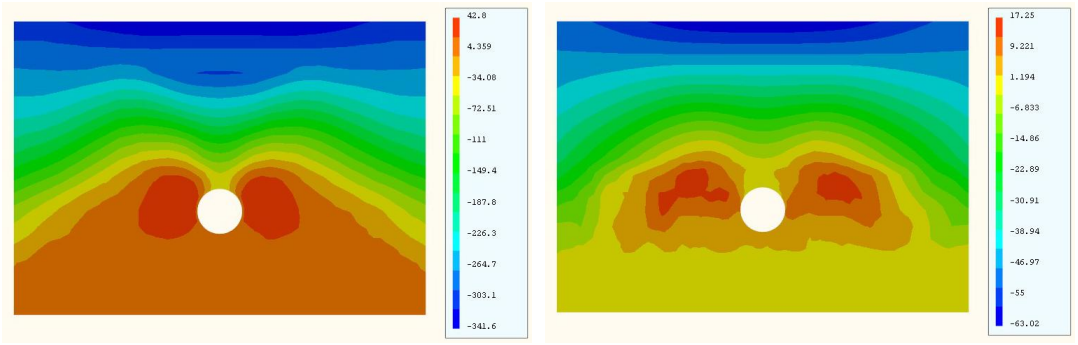


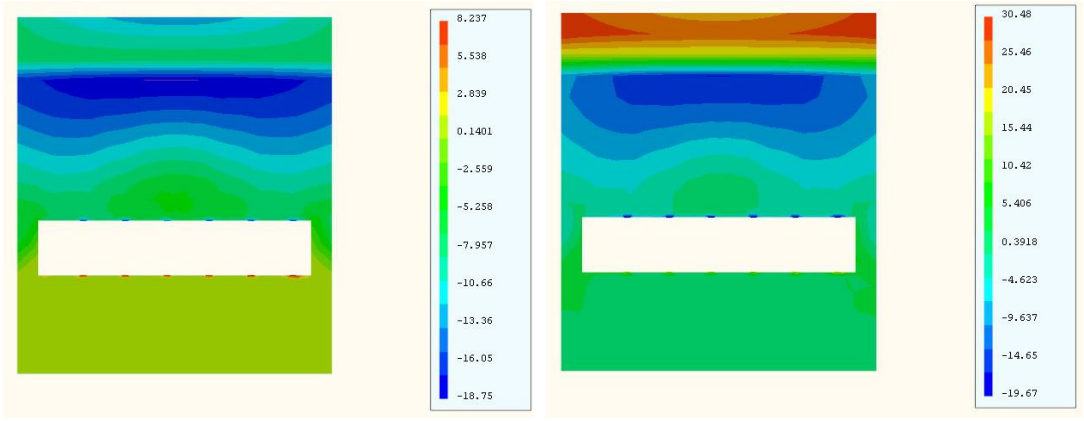
Fig.7 The relationship between vertical stress and time

Figure 8 and figure 9 show the calculation results of vertical speed and vertical displacement at different times. The result of vertical speed shows that the effect of vibration load on the metro has weakened after the concrete energy absorbing layer set up. when t=0.5s, the effect of vibration load on metro has enhanced, the effect of vibration load on the roof is downward, on the bottom plate is upward. When t=0.5s, the vibration load on both sides of the subway is similar to the "bat wing". The distribution area is larger, but the concentrate phenomenon is not obvious. when t=1s, the effect of vibration load continues to increase. The vibration load has wave action to the metro, but the concentrate phenomenon is not obvious, the distribution of vibration load on the metro roof plate is similar to the "helmet", on the metro both sides is similar to the "bat wing". The vertical displacement results shows that the rock and soil around the metro obviously affected by vibration load, the metro top plate has downward trend, and the metro two sides are squeezed toward inside. However, this change has little influence on the metro deformation, this is due to the coordinated deformation of concrete absorbing layer can reduce the surrounding rock deformation. The metro vertical stress and horizontal displacement have obviously reduced, this is due to the concrete energy absorbing layer good coordination deformation performance enable metro can reduce the vibration and vibration intensity, and maintain itself stability [14,15]. Compared with the data of rigid support metro, the vertical force curve volatility decreases and the vertical force of measuring point 8 and 9 also reduced, but the horizontal displacement almost the same. In addition, because the concrete energy absorbing layer has large deformation, so the overall deformation of metro have significantly reduced and there is no obvious deformation concentrated area in the surrounding rock vicinity. This is indicated that the rigid flexible coupling support can fully coordinate deformation, which can reduce the vibration load, improve the impact resistance of deep underground projects, but also conducive to maintaining the stability of the metro.



(a) 0.5s (b) 1.0s

Fig.8 Vertical speed results of vertical roadway section



(a) 0.1s (b) 0.5s

Fig.9 Vertical displacement results of parallel roadway section

Conclusions

Based on the continuum mechanics for discrete element calculation method, the shock resistance stability under the action of seismic wave of rigid support and rigid-flexible coupling support deep buried chamber was analyzed. The strengthening and damping action of different protective bodies on the chamber are studied. And the main conclusions are obtained.

Under the condition of vibration load, rigid support as a strong energy storage body can improve the strength of the passive support, but the damage of shock instability becomes easier, and these will lead to the metro extrusion deformation even the overall closure failure;

Concrete energy absorbing layer can effectively the attenuated seismic wave, in the propagation process or near the metro bottom. Under the action of coordinated deformation, the surrounding rock soil deformation tends to be mild and reduced, the stability of the metro enhanced.

Under the action of strong vibration load, the safety of deep buried chamber can be greatly improved by increasing the strength of rigid support and setting up the flexible deformation buffer layer.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China, project number: 51504029. and Beijing Nova program, project number: xx2016041. It is also supported by the sustainable development of the project in Xicheng District City, Beijing science and Technology Commission, project number: SD2014-36. The comments and suggestions from editors and reviewers helping the improvement of the quality of this paper are appreciated very much by the author.

References:

[1] Wang Sijing, Yang Zhifa, Liu Zhuhua. Rock mass stability analysis for underground engineering[M]. Beijing: Science Press, 1984.

[2] Zheng Yingren, Liu Xinghua. Modern nonlinear science and rock mechanic problems[J]. Chinese Journal of Geotechnical Engineering, 1996(1):98-100.

[3] Qian Qihu. Challenges Faced by Underground Projects Construction Safety and Counter Measures[J]. Chinese Journal of Rock Mechanics and Engineering,2012,31(10):1945-1956.

[4] Ren GingWen. Rigid body element method and its application in stability analysis of rock mass[J].Journal of Hohai University, 1995(1):1-7.

[5] Zhiyin Wang, Yunpeng Ling. Sijing Wang. Numerical simulation of the geomechanical processes in rock engineering[J]. International Journal of Rock Mechanics and Mining Sciences,2000,37:439-507.

[6] MS Diederichs, P K Kaiser. Tensile strength and abutment relax-ation as failure control mechanisms in underground excavations[J]. International Journal of Rock Mechanics and Mining sciences,1999,36:69-96.

[7] Lv Qin, Zhang Dingli, Huang Jun. Mechanism of Stratum Deformation and Its Control Practice in Tunneling Urban Subway at Shallow Depth[J]. China Safety Science Journal,2003,13(7):29-34.

[8] Wang Mengshu, Luo Qiong. Construction of shallow buried excavation method in Beijing Metro[J]. Journal of Railway Engineering Society, 1988,(4):107-117.

[9] Sun Jun. Design theory and practice of underground engineering[M]. Shanghai: Shanghai science and Technology Press, 1996:36-42.

[10] Kong Heng, Wang Mengshu, Yao Haibo, etc. Distribution Laws of Stratum Stress With Working Face Excavation in Subway Tunnels[J]. Chinese Journal of Rock Mechanics and Engineering, 2005,24(3):485-489.

[11] P.A. Fotiu, H. Irschik, F. Ziegle. Modal analysis of elastic-plastic plate vibrations by integral equations. Engineering Analysis with Boundary Elements[J].1994, Volume 14, Issue 1:81-97.

[12] Y. Yang, K.P. Kou, C.C. Lam, V.P. Iu. Free vibration analysis of two-dimensional functionally graded coated and undercoated substrate structures. Engineering Analysis with Boundary Elements[J].2015,Volume 60:10-17.

[13] J. R. Shao, S. M. Li, M. B. Liu. Numerical Simulation of Violent Impinging Jet Flows with Improved SPH Method[J]. International Journal of Computational Methods[J].2016, DOI: 10.1142/S0219876216410012.

[14] Liu Daohua, Niu Shixin, Peng Guangsheng, et al. Construction techniques in tackling the collapse in excavating a highway tunnel[J]. Journal of Geological Hazards and Environment Preservation, 2013,24(1):93-98.

[15] Lv Xiangfeng. Excavation Instability Study of Deep Buried Tunnel Consider the Disturbance Action of Earthquake[J]. Tunnel Construction, 2014,34(2):129-133.

Damage Location Identification of Simply Supported Steel Truss Bridge Based on Displacement

Shaopu Yang, Jianying Renand Shaohua Li

Shijiazhuang Tiedao University, 050043, Shijiazhuang, China

Abstract

Bridge structure damage identification is an important step in bridge structure health monitoring system, but all kinds of damage identification method at present are all complicated and have poor applicability. Therefore, this paper will propose a simple and applicable method of damage identification based on displacement. This has important significance to realize the real-time and exact warning and forecasting the bridge structural health situation. The damage identification indexes are the change percentages of the lower chord panel points maximum deflections and the beam end maximum displacement. The identification model are established respectively using C-Support Vector Classification (C-SVC) and Probabilistic Neural Network (PNN) to identify the damage location, and the two models results are analyzed. The numerical example results show that: (1) The damage identification method based on the bridge deflection is feasible. (2) PNN model and SVC model all have good anti-noise capacity and generalization(3) SVC model is more suitable to be used in site.

Index Terms: displacement, damage location identification, SVM, PNN, railway double-track simply supported steel truss bridge

Introduction

Large-scale civil engineering structures (such as: long-span bridge, high-rise buildings, ocean platform, large span space structure and dam) are very important to the social economy development. But in their working life, because of the environment factors, human factors and natural hazard, successive damages accumulate in the large-scale civil engineering structures, these damages can cause potential safety hazard, and then impact the structure normal use. For real-time mastering the structures health condition, there are many large-scale structures established health monitoring system, such as: Tsing Ma Bridge, Sutong Bridge, Wuhu Yangtze River Bridge, etc.. Structural damage identification is the critical step in the structure health condition assessment, and is one of the research hotspot in academic world and engineering world. The bridge structure damage identification methods include two mainly methods: model-based damage identification method and no-model-based damage identification method[1]. Model-based damage identification method include: pattern matching method[2], damage index method[3], adjustment model method[4]. No-model-based damage identification method include: frequency domain identification method[5], time domain identification method[6], time-frequency analysis method[7]. These methods are initially successfully used in the damage identification of mechanical, and also have a large number of applications in the field of civil engineering in recent ten years[1]. This paper proposes a novel damage location identification method[8-9], which is combined the model-based damage identification method and no-model-based damage identification method. A number example for a 64 m railway double-track simply supported steel truss bridge is provided to verify the feasibility of the method. And the intelligence algorithms

are respectively using C-Support Vector Classification (C-SVC) and Probabilistic Neural Network (PNN) to establish the damage location identification model.

Displacement-based Damage Identification Method

A. Damage identification index

There are two possibilities during the structure damaged. One is the structure mass changed, the other is the structure stiffness decreased[3]. In view of Mechanics of Materials[10] and General code for design on railway bridges and culverts [11], the structure displacement can reflect the structure stiffness. And, in Finite Element Method, the structure node displacements are calculated by equation (1).

$$\{\Delta\} = [K]^{-1}\{P\} \tag{1}$$

Where,

$\{\Delta\}$—structure node displacement vector,

$[K]$—structure stiffness coefficient matrix,

$\{P\}$—node load vector.

In equation (1), if the node load vector$\{P\}$ is constant, the structure node displacement vector$\{\Delta\}$ will be as the change of the structure stiffness coefficient matrix $[K]$. That is to say, the nodes displacement can reflect the structure stiffness.

When a train is travelling on a railway bridge, the train and the bridge compose a complicated train-bridge time-varying system. The bridge structure nodes displacement will change along with the change of the trains location. In view of the bridge structure nodes being very many, this paper constructs the damage identification index based on the bridge certain nodes maximum displacement, that is, the damage identification index is the change percentages of the bridge certain nodes maximum displacement,

$$\Delta x_i = \frac{x_{imax} - x_i}{x_i} \times 100\% \tag{2}$$

Where,

x_i—in certain load case,the maximum displacement of node i, when the structure stiffness isnt damaged,

x_{imax}—in the same load case, the maximum displacement of node i, when the structure stiffness is damaged,

Δx_i—in the same load case, the change percentages of the maximum displacement of node i.

B. Intelligent Algorithm

(A) Artificial Neural Networks

Artificial Neural Networks (ANNs) is a kind of mathematic model by simulating biology neural networks to process information. Artificial neuron is the ANNs information processing unit and the ANNs design fundamental. A large number of artificial neurons are organized by a certain topological structure to constitute a colony parallel mode processing computation structure, which is called ANNs. According to the topological structure, ANNs is divided into the forward neural network and

feedback neural network. Probabilistic Neural Networks (PNN), which is used in this paper, is a forward neural network. PNN is usually applied to research pattern classification problems.

(B) Support Vector Machine

Support Vector Machine (SVM) is a powerful method to solve the tradition problems, such as Curse of dimensionality and Over learning etc. This paper use Matlab and LIBSVM, which is developed by Taiwan University PhD Lin Chih-Jen and his team members, to train the damage location identification model. The C- Support Vector Classification Machine (C-SVC)[12] algorithm flow chart, which is used in this paper, is shown in Fig.1.

In this paper, the kernel function is Gauss radial direction kernel function,

$$K(x, x') = \exp(-\|x - x'\|^2/\sigma^2) \tag{3}$$

C. Damage Location Identification

Damage identification includes 2 steps: damage location identification and damage degree identification. For the paper length limited, this paper only studies the damage location identification.

Fig. 2 shows the damage location identification flow chart.

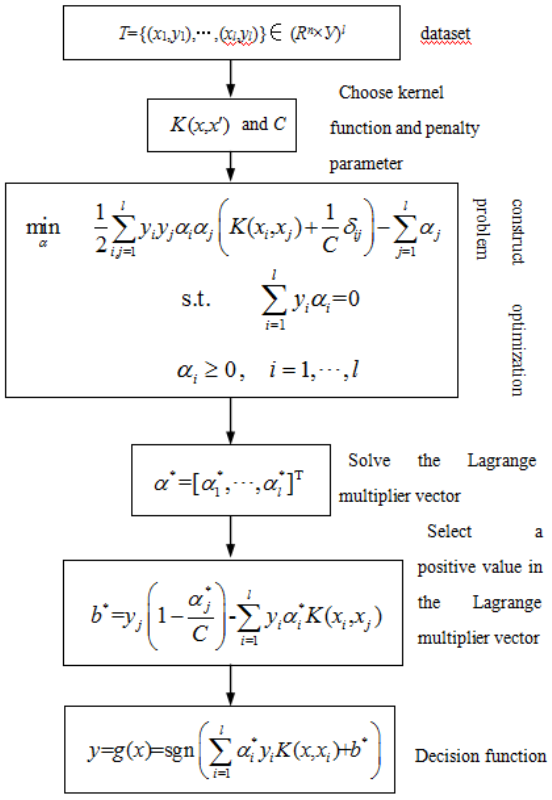


Figure 1. C-SVC flow chart

In this flow chart, there are two ways to add noise.

One way is that a certain data vector added noise according to Equation (4). This way can expand the data set. If the original data have n sets data, and $j \in [1, m]$ in Equation (4), the expanded data will have $n \times m$ data sets. The purpose is to increase the damage identification accuracy, anti-noise ability

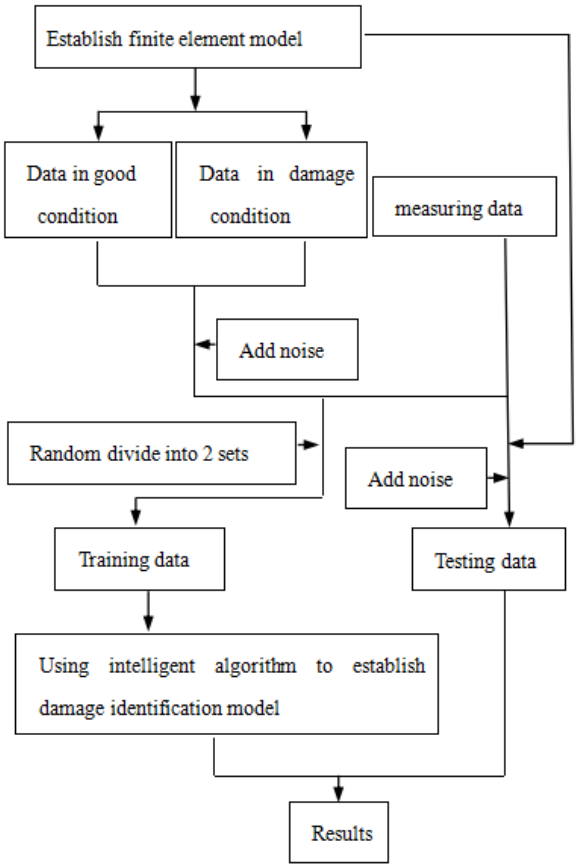


Figure 2. damage location identification flow chart

and generalization ability.

$$\{x\}_{jtest} = \{x\}_{calculate} \times (1 + \varepsilon R_j) \tag{4}$$

Where,

$\{x\}_{jtest}$—the j th simulate test data vector after a certain calculation data vector is expanded,

$\{x\}_{calculate}$—a certain calculation data vector,

R_j—the j th datum of the normal distribution random data, which the mean value is 0 and the mean square deviation is 1,

ε—noise level.

The other way is that a certain element in a certain data vector is added noise according to Equation (5) [13].

$$\{x\}_{ktest} = \{x\}_{kcalculate} \times (1 + \varepsilon R_k) \tag{5}$$

Where,

x_{ktest}—the k th independent variables simulate test data,

$x_{kcalculate}$—the k th independent variables calculation data,

R_k— the k th datum of the normal distribution random data, which the mean value is 0 and the mean square deviation is 1,

ε—the noise level.

64 m Simply Supported Steel Truss Bridge Numerical Example

A. Finite element model

This bridge is a 64 m simply supported steel truss bridge. The finite element model is established using space bar element, there are 32 nodes and 116 bar elements (Fig. 3). The x direction, y direction and z direction linear displacement are restrained on the node 1 and the node 10 to simulate fixed hinged support, and the y direction and z direction linear displacement are restrained on the node 9 and the node 18 to simulate activity hinged support. The coordinate system is shown in Fig.3.

The main truss node numbers, the upper chord unit number and the lower chord unit number are shown in Fig.4.

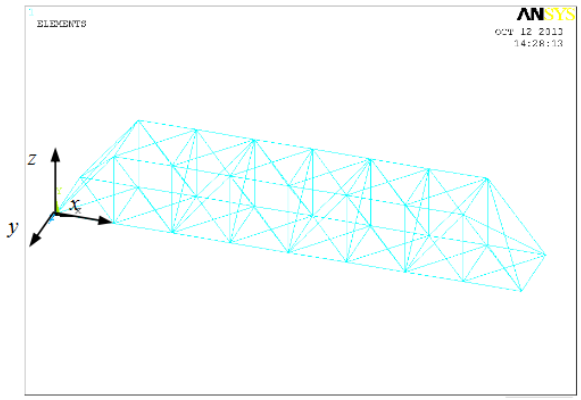
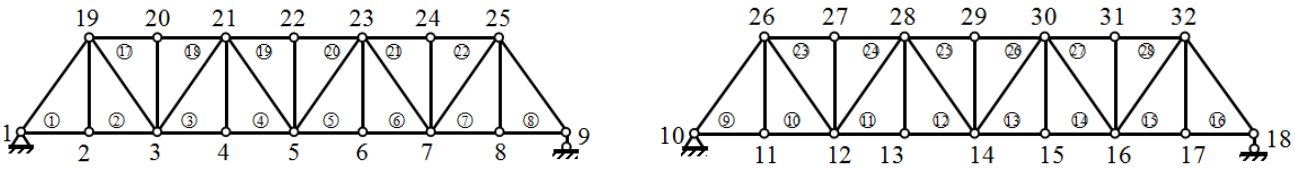


Figure 3. 64 m simply supported steeltruss bridge finite element model



（a） the main truss where y=0

（b） the main truss where y= -10

Figure 4. The main truss node numbers, the upper chord unit number and the lower chord unit numbert

B. Data preparation

The train load is considered as moving dead load. The load cases include one locomotive up-run on the bridge, one locomotive down-run on the bridge, one locomotive simultaneously from the bridge two ends run on the bridge, a train with one locomotive up-run on the bridge, a train with one locomotive down-run on the bridge and two trains with one locomotive simultaneously from the bridge two ends run on the bridge. Where, the locomotive is Dongfeng 4 locomotive, the axle load is 23 t, the vehicle is C62the axle load is 20.15 tthe wheel bases are respectively shown in Fig.5 and Fig.6[14].

Under the train loads, the upper chord and the lower chord internal forces are larger, easily damaged. Consequently, in this paper, The extensional rigidity EA of the element ①, ③,⑤ ,,⑦ ,⑨ ,⑪ ,⑬ ,⑮ ,⑰ ,⑲ ,㉑ ,㉓,㉕ and ㉗respectively discount 5%, 10%, 15%, 20%, 30%, 50% to simulate damage. When

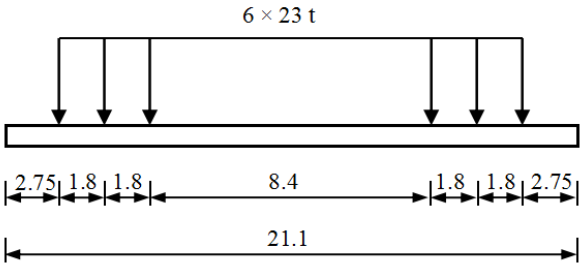


Figure 5. Dongfeng 4 locomotive axle load and wheel base (unit: m)

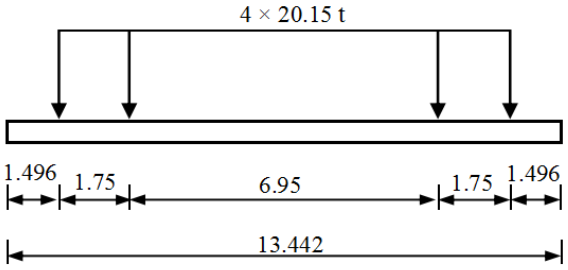


Figure 6. C62 vehicle axle load and wheel base (unit: m)

the 6 load cases are respectively on the bridge, the lower chord panel points maximum deflections and the beam end maximum displacement are calculated using the finite element model, and 504 sets data are obtained. Then according to Equation (2), the damage location identification indexes are obtained.

C. Data expand

Firstly, the 504 sets data are added noise according to Equation (4), where $\varepsilon = 1\%$, j =1,2,...,5. Then, 2520 sets data are obtained.

Secondly, the 2520 sets data are added noise according to Equation (5), where $\varepsilon = 1\%$, k =1,2,...,16.

D. Normalization processing

For increasing the classification and regression accuracy rate, and reducing the error, the indexes and the damage degrees are normalization processed. The normalization algorithm is

$$f : x_l \to y_l = \frac{x_l - x_{min}}{x_{max} - x_{min}} \tag{6}$$

Where,

x and $y \in R^n$, $x_{min} = min(x)$, $x_{max} = max(x)$.

The normalization results is that the original data are normalized in $[0, 1]$, that is $y_l \in [0, 1], l = 1, 2, ..., n$[15].

The 2520 sets data are normalized according to Equation (6). Then the training data are obtained.

E. Testing data

In order to test the model generalization and anti-noise capacity, the testing data is obtained by the following method. Firstly, the finite element model calculation data is selected, when the element① ,③ ,⑤ ,⑦ , ⑨,⑪ ,⑬ ,⑮ , ⑰,⑲ , ㉑,㉓ ,㉕ ,㉗ are respectively damage 25% and 40% (these damage degree arent included in the training set). The load cases are a train with one locomotive up-run on

the bridge, a train with one locomotive down-run on the bridge and two trains with one locomotive simultaneously from the bridge two ends run on the bridge. Then 84 sets data are obtained. Secondly, the damage identification indexes are obtained according to Equation (2). Thirdly, the 84 sets data are added noise according to Equation (5), where =0.1%, 0.5%, 1%, 5%, 10%, 20%, 30%, 50%, 80%, and k =1,2,..., 16. Last, the 84 sets data are normalized according to Equation (6). Then the testing data are obtained.

F. Damage Location Identification

(A) PNN identification model

Damage location identification model is established by MATLAB neural network toolbox function, which is newpnn(P,T,SPREAD). Where P is the input vector, T is the goal vector, SPREAD is expansion rate of the radial basis function, in this paper SPREAD=0.2.

Firstly, 2520 sets training data are used to establish and train the PNN model. Secondly, 1500 sets data are randomly selected to check the PNN model result. Lastly, the 84 sets testing data are inputted in the PNN model to check the models anti-noise ability and the generalization ability.

(B) SVC identification model

The 2520 sets training data are randomly divided into two groups, one is to train the SVC identification model, the other is to test the model. Using k–fold cross-validation method (Deng N.Y. et al., 2009), the penalty parameter C and the kernel function parameter σ are selected,$C = 32$ and $\sigma = 2$. Then, the 570 sets original data is considered as the training data to establish the damage location identification model. Then, based the 2520 sets training data, the damage location identification model is established using LIBSVM software package. Lastly, the 84 sets testing data are inputted in the SVC model to check the models anti-noise ability and the generalization ability.

(C) Damage location identification result

Table 1 shows the results of the PNN model and the SVC model, when input the testing data added various noise levels.

Fig.7 and Fig.8 respectively show the damage location identification result of the PNN model and the SVC model, when the noise level is 30%.

Table 1. The comparison results of the PNN model and the SVC model

Noise level	The number of mis- identification		Accuracy rate (%)		Elapsed time (s)	
	P NN	SVC	PNN	SVC	PNN	SVC
1%	0	0	100	100	3.85	0.35
5%	0	0	100	100	3.835	0.37
10%	0	0	100	100	3.85	0.37
15%	0	0	100	100	3.90	0.37
20%	0	0	100	100	3.91	0.36
30%	7	0	91.6667	97.619	3.72	0.36
50%	9	0	89.2857	89.2857	3.76	0.37
80%	26	0	69.0476	70.2381	3.95	0.37

From Table 1, when the noise level is less than 20%, the PNN model and the SVC model identification

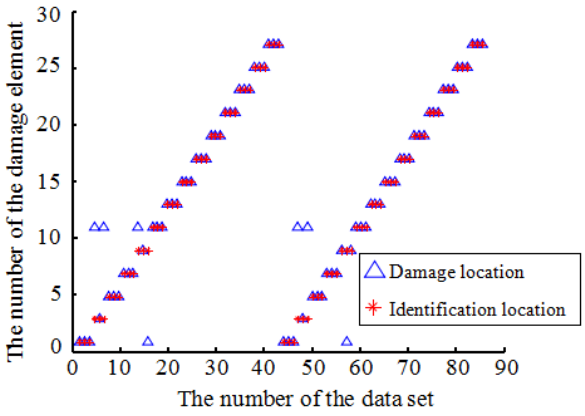


Figure 7. The damage location identification result of the PNN model, when the noise level is 30%

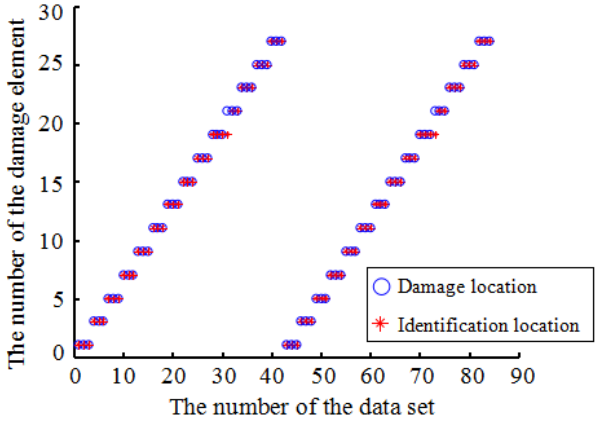


Figure 8. The damage location identification result of the SVC model, when the noise level is 30%

accuracy rates are all 100%. When the noise level is 30%, the PNN model has 7 mis-locations(Fig.7), the identification accuracy rate is 91.6667%. And, When the noise level is 30%, the SVC model has 2 mis-locations (Fig.8), the identification accuracy rate is 97.619%. And when the noise level is 50%, the identification accuracy rates decrease to 71.4%. When the noise level is 50% and 80%, these two model identification accuracy are almost equal. Meanwhile, for all noise level, the PNN model identification elapsed times are all in $3.7s \sim 4.0s$, and the SVC model identification elapsed times are all in$0.35s \sim 0.38s$, only is 10% of the PNN model. This indicates that the SVC method more can satisfy the requirement of real time, fast and accurate identification damage location, and has strong anti-noise ability and good generalization ability.

Conclusions

(1) It is feasible that the 64 m steel truss bridge lower chord panel nodes maximum deflections and the beam end maximum horizontal displacement act as the damage identification indexes.

(2) The PNN method and the SVC method all have strong anti-noise ability and good generalization ability.

(3) In the training and identification process, C-SVC algorithm is faster than PNN algorithm, more

suitable applied in damage location identification, and more can satisfy the job site requirements which require it can real-time fast and accurately identify the damage.

Acknowlegement

This study is supported by National Natural Science Foundation of China (NSFC)(11472180), the New Century Talent Foundation of Ministry of Education under Grant (NCET-13-0913).

References

[1] Zhu Hongping, Yu Jing, Zhang Junbing. A summary review and advantages of vibration-based damage identification methods in structural health monitoring [J]. *Engineering Mechanics* **,** vol.28, No. 2, 2011, PP. 1-11,17.

[2] Zhao Qilin, Zhai Kewei, Zhang Zhi, Hu Yeping. Structure damage location model matching method based on static information[J]. Chinese Journal of Computational Mechanics , **v**ol. 23, No. 6, 2006, PP. 789-793.

[3] Chang Jun, Ren Yonghui, Chen Zhonghan. Experimental investigation of structural damage identification by combination index method under ambient exitation[J]. *Engineering Mechanics* **v**ol. 28, No. 7, 2011, PP. 130-135.

[4] Zong Zhouhong, Chu Fupeng, Niu Jie. Damage identification methods of bridge structures using response surface based on finite element model updating[J]. *China Civil Engineering Journal,*, **v**ol. 46, No. 2, 2013, PP. 115-122.

[5] Gu Peiying, Deng Chang. Modal parameters identification of strain modes under ambient excitation with frequency domain method[J]. *Journal of Vibration and Shock,* **v**ol. 27, No. 8, 2008, PP. 68-70,178.

[6] Liu Tao, Li Aiqun, Ding Youliang. A larming method for cable damage of long-span cable-stayed bridges based on wavelet packet energy spectrum[J]. Journal of Southeast University (Natural Science Edition), **v**ol. 37, No. 2, 2007, PP. 270-274.

[7] Yao Jingchuan, Yang Yiqian, Wang Lan. The damage alarming method for bridge based on Hilbert-Huang transform[J]. China Railway Science, **v**ol. 31, No. 4, 2010, PP. 46-52.

[8] Ren Jianying, Su Mubiao, Zeng Qingyuan. Damage identification of railway simply supported steel truss bridge based on support vector machine[J]. Journal of Applied Sciences, **v**ol. 13, No. 17, 2013, PP. 3589-3593.

[9] Ren Jianying, Su Mubiao, Zeng Qingyuan. Railway Simply Supported Steel Truss Bridge Damage Identification Based on Deflection[J]. Information Technology Journal, **v**ol. 12, No. 7, 2013, PP. 3946-3955.

[10] Sun Xunfang, Fang Xiaoshu, Guan Laitai. Mechanics of materials[M]. Beijing: Higher Education Press, 2009.

[11] The Ministry of Railways of the People's Republic of China. General code for design on railway bridges and culverts[S]. Beijing: China Railway Press, 2005.

[12] Deng Naiyang, Tian Yingjie. Support vector machinetheory, algorithm and expanding[M]. Beijing: Science Press2009.

[13] Jiang Shaofei, Wu Zhaoqi. Structural health nonitoring and intelligent information processing technology and application[M]. Beijing: China Building Industry Press, 2011.

[14] Ge Limei. Rail wagon made in China[S]. Beijing: China Railway Press, 1996.

[15] Shi Feng, Wang Xiaochuan, etc.. Analysis of 30 Matlab neural network cases[M]. Beijing: Beihang University Press, 2010.

Simulation and Experimental Validation of Mining Induced Bed Separation of Overlying Strata with Realistic Failure Process Analysis （RFPA）*

† G.M. Yu [1], S.B. Lu [1], G.Y. Wang [1], X.Y. Hu [1], W.R. Mi [1]

[1]Civil Engineering School, Qingdao University of Technology, China

†Corresponding author& Presenting author: yu-guangming@263.net

Abstract

In the mining process of underground coal, the bed separation of overlying strata is inevitable. The developing process of the bed separation has an important influence on mining subsidence. It is very significant to study the developing regular patterns of the bed separation for understanding and perfecting the mining subsidence theory further. In this paper, Realistic Failure Process Analysis (RFPA) is used to research the distributing patterns of mining induced bed separation of overlying strata. The strata are sedimentary coal strata. And the similar material simulation experiments are used to test the results. The study shows that the growing height of bed separation is increasing as the advance of working face. At the beginning of coal mining, the height of bed separation increases slowly. As the distance of advance increases, the growing rate of separation height becomes faster gradually. After the working face advances a certain length, the growing rate of separation height decreases and closes to zero. After arrive a limit height, the growing height of bed separation will not increase after arriving a limit height. At last, the bed separation will distribute in a range of trapezoid with a 60 degree bottom angle above the goaf.

Keywords: Rock Fracture, Computation Method,Coal Mining,Bed Separation,Developing Regular Pattern.

Introduction

As deeply researched the theory of mining subsidence, people have realized that there is a bed separation phenomenon in the mining induced damage process of overlying strata. Many scholars have discussed the existence, forming cause, growing process, distributing patterns of the mining induced bed separation in the overlying strata with different method. They have also researched all kinds of factors influencing the development of bed separation. Germany scholar, H.Kratzsch, introduced the bed separation phenomenon in his book of "Mining Damage and Protection",[1]. In 1984, based on the pressure arch theory, American scholar S.S.Peng explained the unload state of direct roof of coal seam and the bed separation phenomenon[2]. In 1986, based on many similar material simulation experiments, Chinese scholar, Zhao Deshen, researched the distribution regular patterns of bed separation and put forward the 'Arch Beam Balance Theory' of mining induced bed separation in the overlying strata. In this theory, the mechanics structure of bed separation in the overlying strata is defined in macroscopic view[3]. In 1990, Russian scholar, В.Л.Самарин [4], invested the cause of bed separating, place of bed separation and related factors of influencing bed separating in the fracture zone, etc. In 2011, Chinese scholar, Dai Huayang, researched the distribution discipline of rock fractures after coal mining with numerical simulation and probability integral method. And the distribution of bed separation in the overlying strata in the process of mining is determined [5].

In summary, for the question of mining induced bed separation, scholars have got a lot of research results. But, as a nonlinear damage phenomenon, the growing of bed separation in the overlying strata has a certain space complexity [6]. From begin to finish, the bed separation is changing with time. Because the rock layer properties and places are different, the change rates are different in the whole growing process of bed separating. To the questions of different growing rate and spatial variations in the mining process, more work will be done in the future.

So, in this paper, Realistic Failure Process Analysis (RFPA) is used to discover the mechanics mechanism of mining induced bed separating and different growing rate in the overlying strata which are sedimentary coal strata and to research the time and spatial distribution patterns of bed separations. Furthermore, the relationship between the distribution patterns and mining subsidence is studied. The research results about mining induced bed separation will be applied perfectly to control the mining subsidence.

Introduction of Realistic Failure Process Analysis (RFPA)

In 1995, based on the algorithm idea of basic theory of finite element and new material damage process, professor Tang Chun'an put forward the new numerical simulation method 'RFPA' (Realistic Failure Process Analysis). He fully considers the nonlinear, non-uniformity and anisotropy characters in the rupture process of rock or concrete. This theory which is based on the finite element and statistics damage theories is used to analyze the rock rupture process with elastic damage theory and amended Coulomb failure criterion.

The basic principle of RFPA is to discrete the material into a large number primitives their mechanics properties are supposed to obey some statistical distribution. Then the stress and strain state of the material can be got with responding solve methods. Through analyzing the phrases of these primitives and with related failure criterions and damage principles, the material rupture process can be clear [7].

The functions of RFPA are following: (1) To simulate the rock rupture process. Especially to study the influence of the local damage induced stress re-distribution to further deformation and damage process. (2) To simulate the acoustic emission in the process of rock rupture in order to invest the omens of the rock failure and the relationship between the frequency of acoustic emission and the magnitude. (3)Consider the non-uniformity distribution of material mechanics parameters (strengths, elastic models), the nonlinear deformation of rock can be tested basically through all kinds of statistical functions in the software, such as Weibull distribution and normal distribution. (4)Micro faults and macro fault such as joints and fractures can be simulated. (5)Damage process induced by loading and failure process induced by weight can be simulated. (6)To simulate the tunnel digging process, mining subsidence and coal seam roof falling, etc. [8]

Simulation Design of mining induced bed separation with RFPA

The simulation background is the sedimentary coal strata. The time-space distribution laws of bed separation in the mining process are tested in this part.

Mining Geological Conditions

The mechanics parameters of rock in the strata and the geological conditions used in the numerical simulation are in the Table 1. And considering the difference of mechanics parameters of different layers and the change of mechanics parameters of rock in the falling zone, weak planes are set between two different planes. The parameters of these planes are in the Table 2.

Table 1. Rock Mechanics Parameters of the Model

Layer	Elastic Modulus */MPa*	Compressive strength */MPa*	Bulk Density */(KN/m^3)*	Angle of internal friction */(°)*	Poisson's ratio	Thickness */m*
Sandstone	6000	60	26.5	30	0.25	30
Fine Sandstone	4000	55	25.5	35	0.30	42
Sand-shale	1500	30	25.0	37	0.30	8
coal	1000	20	14.0	38	0.35	4
Floor sandstone	10000	100	28.0	30	0.25	16

Table 2. Mechanics Parameters of Weak Planes

Elastic Modulus */MPa*	Compressive Strength */MPa*	Bulk Density */(KN/m^3)*	Poisson's ratio
500	10	20.0	0.25

Numerical Mode

The mechanics parameters of rock in the strata and the geological conditions used in the numerical simulation are in the Table 1. And considering the difference of mechanics parameters of different layers and the change of mechanics parameters of rock in the falling zone, weak planes are set between two different planes. The parameters of these planes are in the Table 2.
A two-dimensional model of RFPA is made. Its length is 190*m*, height 100*m*. There are 200 horizontal split lines and 100 vertical split lines to make 20000 units. According to the geological conditions of prototype strata, the depth of coal seam is 80*m* and the thickness of coal seam is 4*m*. The load is strata weight. Five steps are set to simulate the mining process. Every step of digging is 10m. The mining width is 50*m*. It is 70*m* from the cup open to the left side of model. Figure 1 shows the numerical model.

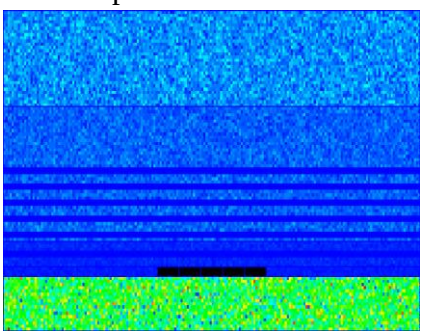


Fig. 1 Numerical model

Results analysis of RFPA

Five steps are simulated in the model, every step is 10m. From figure 2 to figure 6 are the simulating results of bed separating process in the overlying strata.

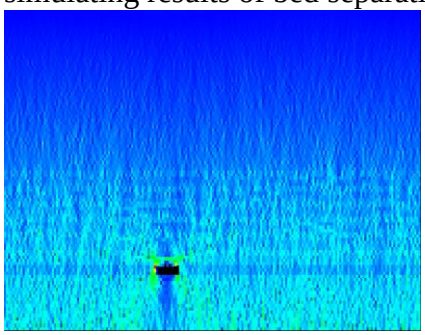


Fig.2 Develop state (10*m*)

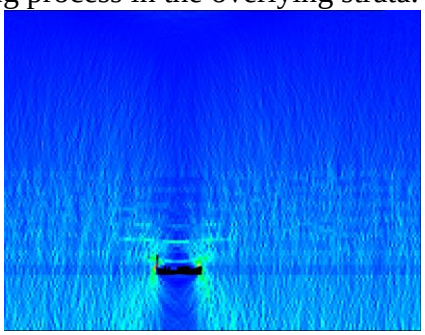


Fig.3 Bed Separation State (20*m*)

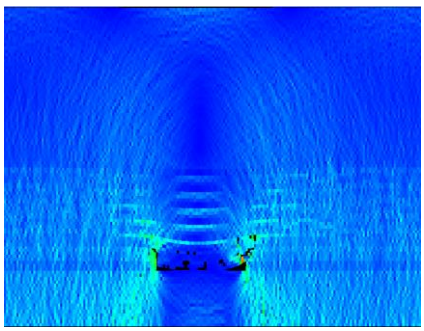


Figure 4. Bed Separation State(30*m*)

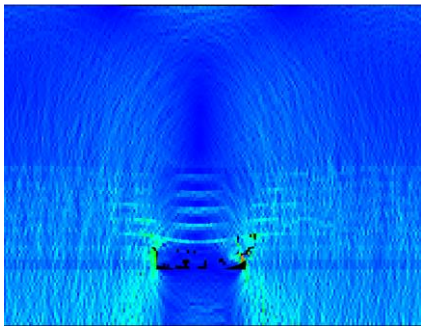


Fig.5 Bed Separation State (40*m*)

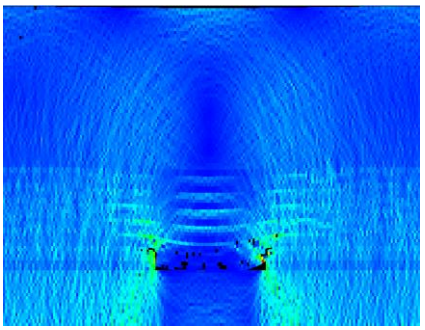


Fig.6 Bed Separation State (50*m*)

Analyzing figures from 2 to 6 shows that the plate girder structure of seam roof bends under the vertical load applied by overlying strata and weight when work face advanced 10*m*. When the face moves 20*m*, bed separation will appear in overlying strata as the range increase of naked roof and increasing bend of layers. The most growing height is 8.03*m* in this digging step. When the work face moves 30m, the most growing height of bed separation is 12.7*m*. When moving 40*m*, the height is 21.9*m*. When moving 50*m*, the height is 26.6*m*. The distribution zone of bed separation is a trapezoid with the bottom angles are 63 degrees (left) and 61 degrees (right) after mining stop, as figure 7 shows. The corresponding relationships

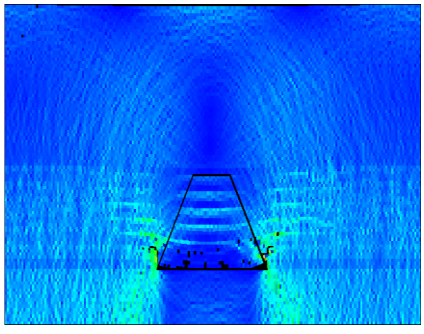


Fig.7 Distribution of bed separation numerical simulation

between work face movement and the most growing height of bed separation are in table 3.

Table 3. The corresponding relationships

Work face movement /(m)	20	30	40	50
The most growing height of bed separation /(m)	8.03	12.7	21.9	26.6

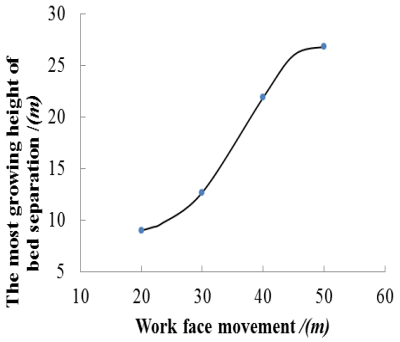


Fig.8 Relationship curve between work face movements and bed separation heights

According to the table 3, the relationship curve between work face movements and the most growing height of bed separation is shown in figure 8.
The equation of relationship curve between work face movements and bed separation heights is:

$$H=-0.0015D^3+0.1581D^2-4.5785D+48.4 \quad (1)$$

In the formula, H represents the bed separation height, D represents the distance of the work face movement.
According to the analysis, the growing height of bed separation increases as the work face moves. At the beginning of coal mining, the height of bed separation increases slowly. As the distance of moving increases, the growing rate of separation height becomes faster gradually. After the working face advances a certain length, the growing rate of separation height decreases and closes to zero. After arrive a limit height, the growing height of bed separation will not increase after arriving a limit height. In the coal mining process, the growing rates of bed separation height are different in different periods.
In summary, the main conclusions obtained by numerical simulation of bed separation in overlying strata with RFPA are following. The bed separation constantly grows forward and upward as the work face moves. At last, the bed separation distributes in a trapezoid zone with the bottom angles about 60 degrees above the goaf. The height of bed separation grows slowly at early stage and grows faster at later stage until the limit height. At the end, the height doesn't grow. At different stages, the increase rates are different. The limit height of be separation is sixty percent of the work face moving distance.

Verification on similar material experiment based on the results of numerical calculation for bed separation in mining overburden

Simulation conditions of the model and experimental purposes

In order to verify the numerical simulation results of bed separation in mining overburden based on the calculation method of rock failure process (RFPA for short), the similar material simulation experiment is adopted. Parameters of the model material are same between numerical simulation experiment and similar material simulation experiment. Coal seam dip angle is 0 °. The depth of coal seam is 80m under the ground. Mining width is 50*m*. Mining thickness is 4*m*. The coal bulk density is $1.4\times10^{-3}kg/cm^3$ and the uniaxial compressive strength of coal is 20*Mpa*. The average bulk density of overlying strata is 2.5×10-3kg/cm^3. The uniaxial compressive strength of overlying strata is 40*Mpa*.

Selection of similar constants of simulation experiment

The ratio between the physical quantities corresponding to the experimental model (m for short) and the prototype (P for short) is called the similarity constant (c for short). The similarity constants in the simulation experiment must be determined reasonably. So the deformation and failure of the whole simulation mining process is more close to the actual situation. Similar constants in this experiment can be shown as followed.

Geometric similarity constant can be calculated by Eq. (1).

$$a_l = l_m / l_p = 1:100 \tag{1}$$

Time similar constant can be calculated by Eq. (2).

$$a_t = t_m / t_p = \sqrt{a_l} = 1:10 \tag{2}$$

Speed similar constant can be calculated by Eq. (3).

$$a_u = u_m / u_p = \sqrt{a_l} = 1:10 \tag{3}$$

Acceleration of gravity similar constant can be calculated by Eq. (4).

$$a_g = g_m / g_p = 1:1 \tag{4}$$

Displacement similar constant can be calculated by Eq. (5).

$$a_s = a_l = 1:100 \tag{5}$$

Bulk density similar constant can be calculated by Eq. (6).

$$a_r = r_m / r_p = 3:5 \tag{6}$$

Strength elastic modulus bond force similar constant can be calculated by Eq. (7).

$$a_R = a_E = a_C = a_l \cdot a_r = 3:500 \tag{7}$$

Internal friction angle similar constant can be calculated by Eq. (8).

$$a_f = f_m / f_p = a_g \cdot a_r \cdot a_l^3 = 0.6 \times 10^{-6} \tag{8}$$

Selection and mix ratio of similar materials

Model is mixed with different types and properties of materials in order to meet the mechanical properties of coal seam overburden. Similar materials usually consist of cementing material and filler. Different mechanical properties of overburden strata will be obtained by adjusting the ratio of the different materials

In this experiment, similar materials with quartz sand, barite and mica are used as filler. Lime and gypsum are used as cementing materials. Borax is used as retarder. Bulk density of similar materials is shown as Table 4.

Table4. Bulk density of similar materials

Material name	quartz sand	barite	Mica	Gypsum	Lime
Bulk density（g/cm^3）	1.4	4.0	0.5	0.8	0.8

The bond force of the model can be controlled by adjusting the ratio of lime and gypsum. Internal friction angle of the model can be controlled by changing the structure of quartz sand. The material mix ratio is shown as Tab.5 according to similar constants and overburden property.

Table5. The material mix ratio

Material name	binder-aggregate ratio	aggregate ratio	Compounds ratio
		quartz sand: mica: barite	Gypsum- lime ratio
Sandstone	1:4	2:1:1	1:1
coal	1:6	6:1:1	3:7

Model dimension and model making

Coal is mined with the roof all Collapsed. The width of mining is 0.5*m*. The Thickness of mining is 0.04*m*. The dimension of the model is shown as Figure 9.

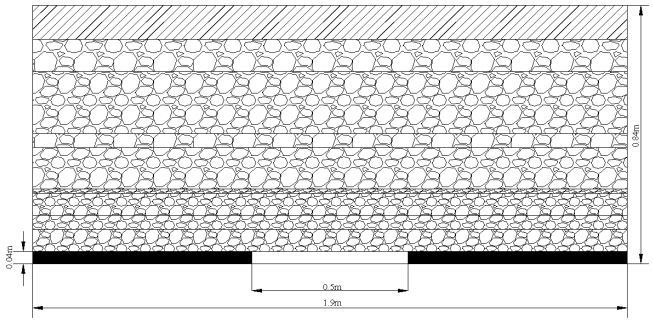


Fig.9 The dimension of similar material model **Fig.10 Model for similar simulation experiment**

The amount of materials of every layer is calculated according to Table 5 and then the model is made stratified. Mica as joint plane is drop between the two layers. So the model made by this method has a good integrity and the strength of the material is easy to be maintained. The model is made on the model desk. First channel Steel is placed in both sides of model. Then the materials are weighed according to material mix ratio and are stirred evenly by water. The materials are layered hierarchical. The thickness of strata is 0.02m and the thickness of coal layer is 0.01m. It need for 3 up to 5 days to remove template and start test after the materials are compacted uniform. The model is shown as Figure 10.

Analysis of similar material experiment results

Coal is mined with the roof all collapsed in this experiment. The mining process is divided into five steps according to numerical similar experiment. The length of every excavation is 10m. The width of mining is 50m. The morphology of bending, fracture, overburden caving and bed separation in overburden strata are shown from Figure 11 to Figure 15 corresponding to every excavation step. The following figures respectively show the development status of bed separation when the mining face advanced 10m,20m,30m,40m,50m.

Fig.11 the development status(10*m*) **Fig.12 the development status(20*m*)** **Fig.13 the development status(30*m*)**

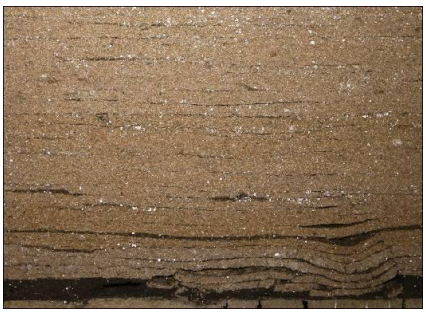


Fig.14 the development status(40*m*)

Fig.15 the development status(50*m*)

The movement and fracture of overburden strata are not occurred because of stress balance according to from Figure11 to Figure15. Goaf is formed after coal is mined in layers before coal excavation. The roof above goaf forms beam structure because it loses the support when the mining face advanced 10m. The small bed separated fissures are produced because bending is not synchronized between adjacent strata. Then the roof is collapsed rapidly. At this moment the first excavation step is completed. The strata under the initial bed separation are further collapsed and caving zone is developed upward when the mining face advanced 20m. The bed separation space near the coal seam roof experienced a process from generation to quickly disappear. The existence time is very short for the bottom bed separation.

Mining area increased gradually with mining face advanced. The scope of overlying strata bending is also enlarged. Strata are broken after bed separation formation. Caving zone and fractured zone are formed in the lower strata. A large number of bed separations are developed in fractured zone. Bend zone is formed from fractured zone up to the ground. Strata of bending zone are continuous and stable. Mining is over when work face advanced 50m. Most bed separations undergo the process of crack initiation, development and closure and the height of bed separation reaches the maximum. Bed separation are distributed in the area of "eight" shape which is roughly trapezoidal symmetry. The fracture angle of strata at open-off cut of coal is 63 degrees. The fracture angle of strata at stopping line is 62 degrees. The height of bed separation is nonlinear growth with the increase of mining working face advance distance according to the experimental data. The maximum height of bed separation is 0.6 times of the advancing distance of mining face. The above experimental results are same to the numerical simulation results by RFPA for mining overburden separated strata.

Conclusions

In this paper, the bed separation in mining overburden is simulated based on RFPA. The development of bed separation in mining overburden during mining is studied. The conclusions are showed as followed.

Firstly, tensile failure and shear failure occur in weak formation under the joint action of transversal shear and gravity stress during the process of mining. Interlayer dislocation and vertical separation are generated in strata. The separation space is formed.

Secondly, the bed separation is developed forward with the mining face advanced. The growth rate of the bed separation height during different stages is different. Fracture initiation time of bed separation is different. So the growth state of different bed separation is different at the same time.

Thirdly, the distribution laws of bed separation in mining overburden are from below and from back to front with the continuous advance of the mining face. It has a certain timelines for the growth process of bed separation.

Finally, the bed separations are distributed within the scope of trapezoid under the specific conditions. The angle of trapezoid base is 62 °. The bed separations are located just above the goaf. The growth rate of the bed separation is from fast to slow and gradually approach to zero.

Characteristics of different speed for different bed separation are shown in different stages. The maximum height of the bed separation is 0.6 times of the distance of mining face advance.

* The project is supported by the Cooperative Innovation Center of Engineering Construction and Safety in Shandong Blue Economic Zone and Nation Natural Science Foundation of China (No. 51374135, 51179080), Qingdao science and technology plan projects (SDSITC-0108310), and China Scholarship Council.

References

[1] Kratcsch, H, *Mining Subsidence Engineering Springer Verlag*. Berlin, 1983.
[2] S.S Peng. *Strata control in mines*. China Coal Industry Publishing House, 1984.
[3] Zhao Deshen. *The propagation laws of mining space in overlying strata*. Fuxin Mining Institute, 1986.
[4] В.Л.Самарин, образвание полости расслоения вподра батьываемом массиве горыx пород изв. горныйжурнал, 1990.
[5] Dai H.Y., Deng Z.Y., Yan Y.G., etc. Study on distribution laws of the normal fractures and overburden separation in deep mining of Tangshan mine. *Coal Mining*, **2011, 16(2)**:8-11.
[6] Tang C., Wang S.H.,Fu Y.F. *Numerical experiments of rock fracture process.* Science Press, 2003.
[7] Tang C., Yu G.M., Liu H.Y., Rui Y.Q. *Fracture of rock mass induced by mining and strata movement numerical experiments.* Jilin University Press, 2003.

An Euler-Lagrange Approach to Model the Dynamics of Particulate Phase Exposed to Hot Gas Injection into Packed Bed Reactors

E. Rabadan Santana[1,a] and B. Peters[1]

[1]Research Unit of Engineering Science RUES, University of Luxembourg, Luxembourg

[a]Corresponding author: edder.rabadan@uni.lu

ABSTRACT

In the present work a coupled Euler-Lagrange approach is used to model the dynamics of a particulate phase and its interaction with hot gas injection into particle bed reactors. The proposed numerical approach is based on the Discrete Element Method (DEM) to model the granular phase. The in-house DEM solver has been extended to account for heat and mass transfer within the gas phase by coupling it with the governing Navier-Stokes equations in the Eulerian Computational Fluid Dynamics (CFD) gas model. This coupling has been done by using the CFD OpenFoam library. As a result the numerical simulation framework called the Extended Discrete Element Method (XDEM) has being developed. The present case uses the XDEM as a numerical tool to study a generic small scale packed bed reactor where hot gas is injected laterally into a packed bed of coke particles. The interaction between solids and different fluid phases in packed bed reactors represents a challenging phenomenon for numerical simulation. In order to represent more accurately such processes the XDEM code has being adapted and several features like particle gasification, chemical reaction and diverse particle shapes have been implemented. The XDEM Euler-Lagrange approach showed the ability to track the positions of the coke particles in the simulation domain allowing an in-depth study of the particle-gas interaction. Since hot air at 1200 K was injected, the effects of gasification, reactions inside the particles, and shrinking were considered. Comparison between measured and predicted data was made for char coal particles.

Keywords: Extended Discrete Element Method, Euler-Lagrange, Hot Blast, Packed Bed, Gas-Particle interaction, Gasification.

Introduction

Injection of preheated air at high speed or blast injection is being used extensively in many industrial applications such as packed bed reactors. When air is injected laterally into a particle bed it causes the formation of granular circulation region within the bed. This process increases the interaction between the solid and gas phases resulting in a more efficient heat and mass transfer within the reactor. Blast injection is widely use in different petrochemical and metallurgical processes such as catalyst, gasification, and combustion. One of the main applications of blast injection is found in the Blast Furnace (BF) reactors. Blast furnace reactors are widely used in the ironmaking industry and are one of the largest operational reactors. Typical dimensions of BF area about 40 m high and 15 m wide for a production over 10 000 t/d of pig iron [1]. The nature of the blast furnace operation includes several types of flow, a packed bed of solids descending, liquid dripping and gas with powder ascending through the packed bed [2]. In a BF liquid iron is produce from ferrous oxides and carbon reductants. Ore is normally used as a ferrous oxide and coke as a carbon reductant. Ore and coke are charged in layers from the top of the furnace. At the bottom part of the furnace, hot air is injected at high velocity through a tuyere. The fast stream of blast gas entering into the packed bed forces the coke particles to displace back and upwards forming a circulation region around the injected gas. This circulation zone forms a cavity called raceway. A schematic drawing of the blast furnace and the raceway are shown in Fig. 1. In the raceway the carbon from the coke or other auxiliary fuels for instance the pulverized coal (PC) reacts with the oxygen to provide heat and to form the main reductant gas, CO. The reductant gas rises through the void space towards the top of the furnace. Chemical reactions take place as the solid material moves downward and interacts with the reductant gas producing liquid iron and carbon dioxide. The final products in form of melted iron and slag are tapped from the bottom of the blast furnace and the flue gas is removed from the top.

Since the stability of the BF operation is related to the motion of coke particles in and around the raceway, many studies have been conducted focusing on the gas flow and the formation of the cavity [4, 5]. However, modeling the raceway

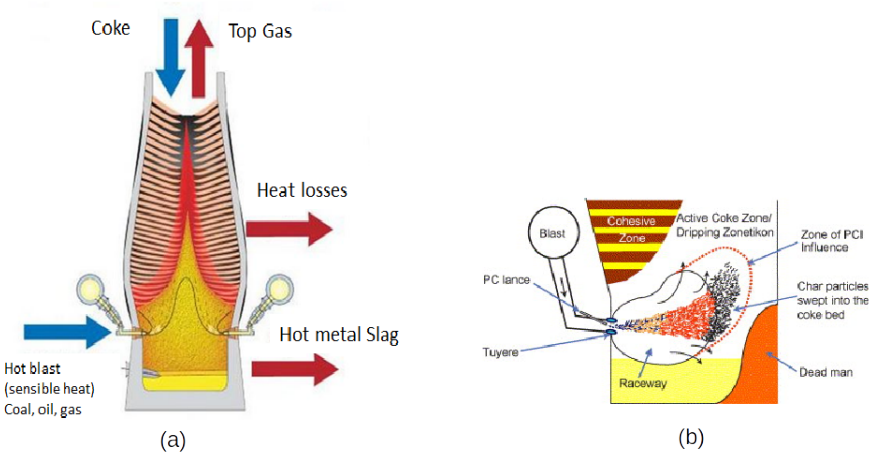


Figure 1: (a) General scheme of a blast furnace [2]; (b) the raceway [3].

is a challenging task due to the high velocities of the blast gas and its interaction with the particles, existence of high temperatures, chemical reactions, and shrinkage of particles [1, 6]. Currently, two models are used to investigate the in-furnace phenomena: continuum model and Discrete Element Method (DEM). In the first approach the geometry/shape of the raceway is fixed and the bed of coke is consider to be a porous media. In such approach the Navier-Stokes equations accounting for porosity are solved over the entire domain [7, 8]. This method is restricted to the dynamics of the gas flow therefore, the formation of the raceway cannot be investigated. Consequently, is not possible to determine the boundaries of the raceway. In the second approach, the Discrete Element Method (DEM), the solid phase is considered as a discrete part while the flow (liquid and/or gas) in the void space between the particles is treated as a continuum phase. This approach also referred to as the Combined Continuum and Discrete Model (CCDM) [9] can be able to predict the distribution of the particulate phase under the influence of lateral gas injection.

In the present work the developed simulation framework *Extended Discrete Element Method (XDEM)* is used to model the motion of the particulate phase and its interaction within the gas. For that purpose, an over-simplified geometry of a generic packed bed reactor is used to evaluate the XDEM. The geometry includes lateral gas injection, as in the case of BF reactors. However, at this stage, the intention of this work is to study the Euler-Lagrange approach of the XDEM to model such type of solid-flow configurations and not the in-furnace phenomena. In a next step, simulations with more realistic geometries and operational conditions have to be conducted to address the in-furnace phenomena.

Numerical Framework: the eXtended Discrete Element Method (XDEM)

The proposed numerical approach is based on the Discrete Element Method (DEM) to model the dynamics of granular matter and the Eulerian Computational Fluid Dynamics (CFD) model for the fluid phase. A coupling between both modeling approaches allows to track the individual motion of the particles and the dynamics of the fluid phase. For that purposes, the in-house DEM solver has been coupled with the open source library OpenFoam. The coupling between both solvers resulted in the development of the numerical simulation framework called the eXtended Discrete Element Method (XDEM) [10]. As a result the XDEM solver accounts for heat and mass transfer within the solid and fluid phases. Within the XDEM solver, the CFD gas phase is solved using OpenFoam. The coupling algorithm between the in-house DEM solver and OpenFoam allows to exchange information between the discrete and gas phases at each time step. In this way, the current position of individual particles can be tracked. In addition, particles are allowed to exchange heat and mass transfer with its environment. This allows to determine, for each particle, its temperature, porosity, reaction degree, shrinking, and species distribution in conjunction with the surrounding gas phase.

A schematic representation of the XDEM modular structure is showed in Fig. 2. The XDEM is composed by two modules: dynamics and conversion modules. The Lagrangian concept of the XDEM-Dynamics module is used to predict the motion of solid particles. The movement of particles is characterized by the motion of a rigid body through six degrees of freedom for translation along the three directions in space and rotation about the centre-of-mass. Thus, the entire motion of each particle is describe by these degrees of freedom. This method is widely accepted and effective to address engineering problems in granular and discontinuous materials, especially in granular flows, rock mechanics, and powder mechanics

[11, 12, 13, 14]. Chemical conversion at each discrete particle is computed by the XDEM-Conversion module. A discrete particle may consist of different phases like solid, liquid, gas or inert material. Since particles can be porous, gas diffusion within the pore volume is accounted for. A particle is allowed to exchange heat with its environment depending on the specified boundary conditions for its surrounding gas. The distribution of temperature is accounted for by system of one dimensional and transient conservation equations for energy [15, 16]. For the particle energy balance, local thermal equilibrium between gas phase and the porous solid is assumed. Thermal energy is transferred from the fluid to the particles and/or from particles to fluid as a heat source. The XDEM-Conversion calculates for each CFD cell the corresponding heat source value depending on the particles properties within the specific cell. The modular structure of the XDEM allows to use the dynamics and conversion modules in a de-coupled mode for better adaptability to the modeling requirements [17]. Figure 2 shows the modular structure of the XDEM solver.

For the present case-study all modules of the XDEM including the coupling with the CFD tool are used. The gas phase is modeled in an Eulerian approach solving the Navier-Stokes equations for compressible fluid in porous media implemented in OpenFoam. The position, orientation, and heat interaction between particles (conduction, radiation) as well as between particles and their environment (conduction, convection) is resolved with the above mentioned XDEM modules.

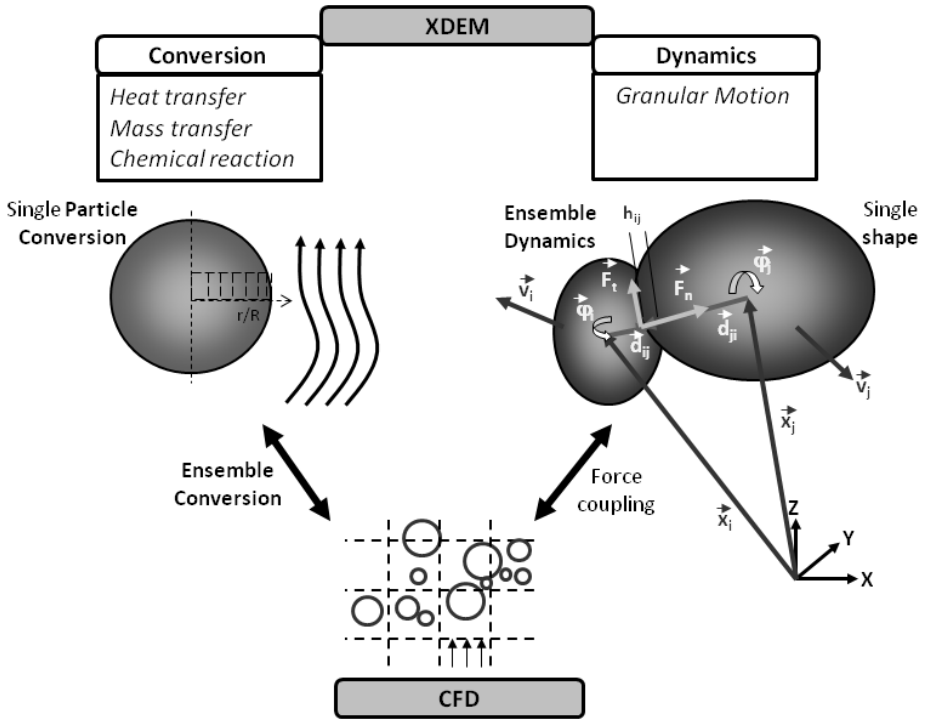


Figure 2: Interaction modules in the XDEM.

The complete set of equations and detailed description of the XDEM numerical framework can be found in [18, 10, 16]. For purpose of clarity a brief description of the main equations is given next. Heat interaction between particles as well as heat and mass transfer between particles and their environment is solved with the XDEM-Conversion module. Thus, the conservation equations of mass, momentum, energy, and species are solved for a porous particle:

$$\frac{\partial(\epsilon_p \rho_f)}{\partial t} + \nabla \cdot \left(\epsilon_p \rho_f v_f\right) = \dot{m}_{s,f} \tag{1}$$

$$-\frac{\partial(\epsilon_p p)}{\partial x} = \frac{\mu_f \epsilon_p}{K} v_f \tag{2}$$

$$\frac{\partial(\rho c_p T)}{\partial t} = \frac{1}{r^n}\frac{\partial}{\partial r}\left(r^n \lambda_{\text{eff}} \frac{\partial T}{\partial r}\right) + \sum_{k=1}^{l} \dot{\omega}_k H_k \tag{3}$$

$$\frac{\partial(\epsilon_p\,\rho_{f,i})}{\partial t} + \nabla\cdot\left(\epsilon_p\,\rho_{f,i}\,v_f\right) = \frac{1}{r^n}\frac{\partial}{\partial r}\left(r^n\epsilon_p D_i\frac{\partial\,\rho_{f,i}}{\partial r}\right) + \dot{m}_{s,f,i} \tag{4}$$

The term on the right hand side in the mass conservation equation, Eq. (1), accounts for the mass transfer between the fluid within the pore of the particle or the solid phase with gas as a result of the chemical reactions, ϵ_p denotes the particle porosity, v_f the advective velocity, and ρ_f the density of the gas phase. Equation (2) is the momentum equation based on Darcy's law for the transport of gaseous species within the pore space of the particle; here K represent the permeability, p the pressure, and μ the dynamic viscosity. Since the thermal mass in the solid and fluid phase are significantly greater than the thermal mass in the gas phase (ρc_p), the heat transported through the bulk motion or diffusion of the gaseous species within the pore space can be neglected. Thus, the energy balance equation, Eq.(3), is based on the homogeneous model for a porous medium as described by Faghri [19] where λ_{eff} is the effective thermal conductivity evaluated as [20]

$$\lambda_{\text{eff}} = \epsilon_p\lambda_f + \sum_{i=1}^{k}\eta_i\,\lambda_{i,solid} + \lambda_{rad} \tag{5}$$

which takes into account heat transfer by conduction in the gas, solid, and radiation in the pore. The later is evaluated as

$$\lambda_{rad} = \frac{\epsilon}{1-\epsilon}\,\sigma\,4.0T^3 \tag{6}$$

where T and σ stand for the temperature and the Boltzmann constant, respectively. The source term $\dot{\omega}$ represents the production or consumption of heat due to chemical reactions where H_k is the enthalpy of reaction k. The formulation of Eq. (3) allows to represent different geometries based on a radial coordinate r: infinite plate ($n = 0$), infinite cylinder ($n = 1$), and sphere ($n = 2$). Equation (4) is the conservation equation of species which accounts for convection in conjunction with diffusive transport to describe the distribution of gaseous species i in the porous particle. The effective diffusion coefficient $D_{i,\text{eff}}$ of species i is derive from the influence of tortuosity τ and the molecular diffusion coefficient D_i [21, 22]:

$$D_{i,\text{eff}} = D_i\frac{\epsilon_p}{\tau} \tag{7}$$

Depending on the rate-limiting process, the depletion of the solid material results in either a decreasing particle density or a reduction of the particle size [23, 24]. The distribution of the porosity and the specific inner surface S are determined by the following equations:

$$\frac{\partial\epsilon}{\partial t} = \frac{M}{\rho\delta}\dot{\omega} \tag{8}$$

$$\frac{\partial S}{\partial t} = \frac{1-\epsilon_0}{C_0}\dot{\omega} \tag{9}$$

where M is the molecular weight of the particle, δ is the characteristic pore length and C_0 and $\dot{\omega}$ are the concentration and reaction of the solid material, respectively. The subindex 0 indicates the initial values of the appropriate variable.

Since geometries are consider to be either infinite plate, infinite cylinder or sphere, a symmetric boundary condition is applied at the center of the particle for the effective thermal conductivity

$$-\lambda_{\text{eff}}\frac{\partial T}{\partial r}\bigg|_{r=0} = 0 \tag{10}$$

and for the heat and mass transfer at the surface of the particle

$$-\lambda_{\text{eff}}\frac{\partial T}{\partial r}\bigg|_{r=R} = \alpha(T_R - T_\infty) + \dot{q}_{\text{rad}} + \dot{q}_{\text{cond}} \tag{11}$$

$$-D_{\text{i,eff}}\frac{\partial\rho_i}{\partial r}\bigg|_{r=R} = \beta_i(\rho_{i,R} - \rho_{i,\infty}) \tag{12}$$

where T_∞ is the gas temperature, $\rho_{i,\infty}$ the ambient density, D_i, α_i and β_i are the diffusion, heat, and mass transfer coefficients of species i, respectively. The heat fluxes $\dot{q}''$ in Eq. (12) account for potential radiative heat exchange with the surrounding and/or conductive heat transport through physical contact with other bodies. Thermodynamic equilibrium within the intra-particle fluid is assumed and the thermal equation of state is used to close the above set of equations $p = \rho RT$ and $h = c_pT$, both used in their formulation for multi-species flow.

The XDEM-Dynamics module is used to predict the motion of solid particles based on Newton's Second Law for conservation of linear and angular momentum

$$m_i \frac{d^2\,\vec{r}_i}{dt^2} = \sum_{i=1}^{N} \vec{F}_{ij}\left(\vec{r}_j, \vec{v}_j, \vec{\phi}_j, \vec{\omega}_j\right) + \vec{F}_{\text{extern}} \tag{13}$$

$$\bar{I}_i \frac{d^2\,\vec{\phi}_i}{dt^2} = \sum_{i=1}^{N} \vec{M}_{ij}\left(\vec{r}_j, \vec{v}_j, \vec{\phi}_j, \vec{\omega}_j\right) + \vec{M}_{\text{extern}} \tag{14}$$

where $\vec{F}_{ij}\left(\vec{r}_j, \vec{v}_j, \vec{\phi}_j, \vec{\omega}_j\right)$ and $\vec{M}_{ij}\left(\vec{r}_j, \vec{v}_j, \vec{\phi}_j, \vec{\omega}_j\right)$ are the forces and torques acting on a particle i of mass m and $\bar{I}_i$ is the tensor moment of inertia. Equations (13) and (14) show that forces and torques exerted on particle i depend on the position $\vec{r}_j$, velocity $\vec{v}_j$, orientation $\vec{\phi}_j$, and angular velocity $\vec{\omega}_j$ of its neighbor particles j. External forces may be include by moving grate bars, fluid forces and contact forces between particles in contact with a bounding wall. Within the XDEM-Dynamics module, forces within particles are present only during mechanical contact. The repulsive force between particles in contact are calculated based on the rigidity of the particles. The interaction between particle-particle and particle-wall is calculated by the contact model linear spring-dashpot and the fluid drag force by the Di Felice's correlation [25].

In the present formulation the deformation of two particles in contact is approximated by its overlapping [11]. The resulting force $\vec{F}_{ij}$ due to contact is calculated by its normal and tangential components

$$\vec{F}_{ij} = \vec{F}_{n,ij} + \vec{F}_{t,ij} \tag{15}$$

where the normal n and tangential t components additionally depend on displacements and velocities normal and tangential to the point of impact between the particles.

The XDEM conversion and dynamics modules are coupled to an implemented CFD solver in OpenFoam for compressible porous media. The last is based on the PIMPLE (PISO-SIMPLE) solution for time-resolved and pseudo-transient simulations allowing equation under-relaxation for better convergence of the equations at each time-step [26]. The CFD equations are the set of the Navier-Stokes equations comprising the mass, momentum, and energy equations for multi-species flow adapted for a porous media [19, 27]

$$\frac{\partial(\epsilon_f\,\rho_f)}{\partial t} + \nabla\cdot(\epsilon_f\,\rho_f\,v_f) = \dot{m}_{s,f} \tag{16}$$

$$\frac{\partial(\epsilon_f\,\rho_f^K\,v_f)}{\partial t} + \nabla\cdot(\epsilon_f\,\rho_f\,v_f\,v_f) = \nabla\cdot(\epsilon_f\,\tau_f) - \frac{\mu_f}{K}\epsilon_f^2\,v_f - C\,\rho_f\,\epsilon_f^3\,|v_f|\,v_f \tag{17}$$

$$\frac{\partial(\epsilon_f\,\rho_f\,h_f)}{\partial t} + (\epsilon_f\,\rho_f\,v_f\,h_f) = \frac{\partial p_f}{\partial t} + \epsilon_f\cdot v_f\cdot\nabla p_f + \sum_{i=1}^{M}\frac{S_p}{V_{REV}}\alpha\Delta T_i \tag{18}$$

$$\frac{\partial(\epsilon_f\,\rho_{f,i})}{\partial t} + \nabla\cdot(\epsilon_f\,\rho_{f,i}\cdot v_f) = \sum_{i=1}^{M} m'''_{s,f,i} \tag{19}$$

The porous media formulation is based on an averaging process over a Representative Elementary Volume (REV) approach [28, 29]. The momentum equation, Eq. (17), is expressed in the formulation for a gas flow within a porous media [19, 27] where K represents the permeability of the packed bed and C the dimensionless drag coefficient. For spherical particles of diameter D_p, K and C can be obtained from [19, 17]

$$K = \frac{D_P^2\epsilon_f^3}{150(1-\epsilon_f)^2} \tag{20}$$

$$C = \frac{1.75(1-\epsilon_f)}{D_P\epsilon_f^3} \tag{21}$$

The intensity of heat exchange between the solid and fluid phases in the energy equation (Eq. (18)) is subjected to the thermal boundary conditions at the interface where S_p is the heat source term responsible for transferring the thermal energy from the fluid to the particles and/or from the particles to the flow

$$S_p\, h_{p,f}(\Delta T_p) = q'''_{sf}\, V_{REV} \tag{22}$$

The last term of the right hand side in Eq. (18) represent the coupling between DPM and CFD for heat transfer simulations.

Simulation Domain and Boundary Conditions

A cylindrical shape can be used to represent the geometry of a generic reactor. In the present case, the generic geometry is further simplified by considering one quarter of it filled with particles. The two investigated geometries are shown in Figs. 3 and 4. The computational domain 1 shown in Fig. 3 is used to study chemical conversion by employing the XDEM-Conversion module. The dimensions of this domain are 300 mm in height with a radius of 225 mm. The domain is discretized with a structured grid which contains 3176 cells with an average size of 20 mm × 22 mm × 22 mm. The second geometry is presented in Fig.4. This geometry is used to study the motion of the particles around the gas injection as well as the heat-up process due to the hot gas. The dimensions of this domains are 800 mm in height with a radius of 400 mm. The domain contains 9144 cells with an average size of 25 mm × 25 mm × 23 mm. Both computational domains are bounded by two side-walls, an outer wall which shapes the cylindrical form of the reactor, an inlet, an outlet (top wall) and a bottom wall. The inlet (showed in solid color) is located at center of the outer wall. In geometry 1 (Fig. 3) is located one cell above the bottom wall whereas in geometry 2 (Fig. 4) is placed at the third cell above the bottom wall. The packed bed is represented by an ensemble of particles with an inhomogeneous void space between them due to their packing. Each sphere has a diameter of 20 mm and is considered to be a coke particle. The coke particles are at rest and randomly settled at the bottom of the cylinder. This was done in a preliminary simulation by placing the particles at the center of the domain and let them fall until they reached their random and steady position. The computational domain in Fig. 3 contains 300 coke particles where as the domain in Fig. 4 contains 3000 particles. Coke properties are shown in Table 1. Hot air at 1200 K in a standard mass fraction composition, $Y_{O_2} = 0.21$ and $Y_{N_2} = 0.79$, is injected through the inlet at a velocity of $u = 20\,\mathrm{m\,s^{-1}}$. For simplicity the inflow profile is considered to be uniform. For the side-walls a cyclic boundary condition is used. At the outlet (top wall), the conservative variables are simply extrapolated from the inside domain. An initial temperature of 600 K is set for the gas and the particles. Temperature is extrapolated from the inside domain to the walls. The standard $k - \epsilon$ model is used with an initial turbulence values set to 3 %.

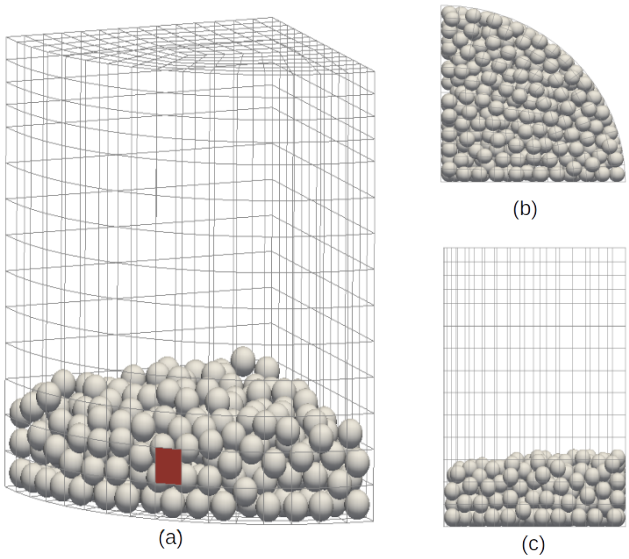


Figure 3: Geometry 1 with 300 particles, different views of the computational domain: (a) isometric, (b) top, and (c) lateral.

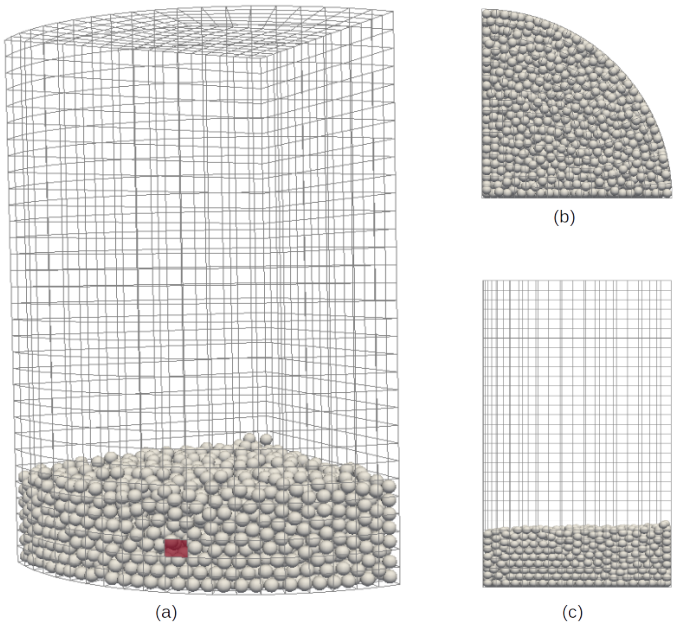


Figure 4: Geometry 2 with 3000 particles, different views of the computational domain: (a) isometric, (b) top, and (c) lateral.

Table 1: Thermodynamic and mechanical properties of coke.

Density [kg/m^3]	1050	Young modulus [Pa]	22×10^9
Porosity [-]	0.2	Poisson ratio [Nm^{-1}]	0.3
Tortuosity [-]	1.0	Friction coefficient [-]	1

Results and Discussion

The XDEM has been validated for spherical particles of diverse materials by comparing predicted and experimental results [16]. For instance, Fig. 5 shows the experimental and predicted data for gasification of spherical char particles of 10 mm and 15 mm diameter exposed to a heating temperature of 773 K. The measurements were obtained from experiments conducted by Schäffer and Wyrsch [30]. The comparison between the measured and predicted reduction in the particle mass fraction shows a good agreement. The particle mass fraction and radius decreases linearly indicating a shrinking behavior. The high char reactivity limits the reaction regime and the transfer of oxygen through the boundary layer represents the rate-limiting step [16]. Therefore, the obtained agreement shows that heat and mass transfer are evaluated with sufficient accuracy.

For the present test-case, the gasification of coke particles inside the reactor is approximated by the following reaction

$$C + \frac{1}{2}O_2 \leftrightarrow CO \tag{23}$$

As observed from Fig. 5 gasification occurs if the particle is exposed to a heating temperature for a considerable amount of time. In terms of simulation this represents a large computational time. In order to account for gasification effects and to test the XDEM capability while simulating multiple particles, the simulation domain shown in Fig. 3 is used. Since large exposure of time is required to observed the heating up and gasification processes, an static simulation using only the XDEM-Conversion module and the CFD coupling is computed first. Accordingly, the governing equations for the flow and particles are solved taking into account conduction between particles and the surrounding flow field. Deactivating the XDEM-Dynamics module allows to increase the time step and achieve a faster solution accounting for heat and mass transfer within the particles and gas, while keeping the particles static. Figure 6 shows the predicted heat-up of the packed bed via the temperature distribution at surface of the particles and in the flow. The increase in temperature can

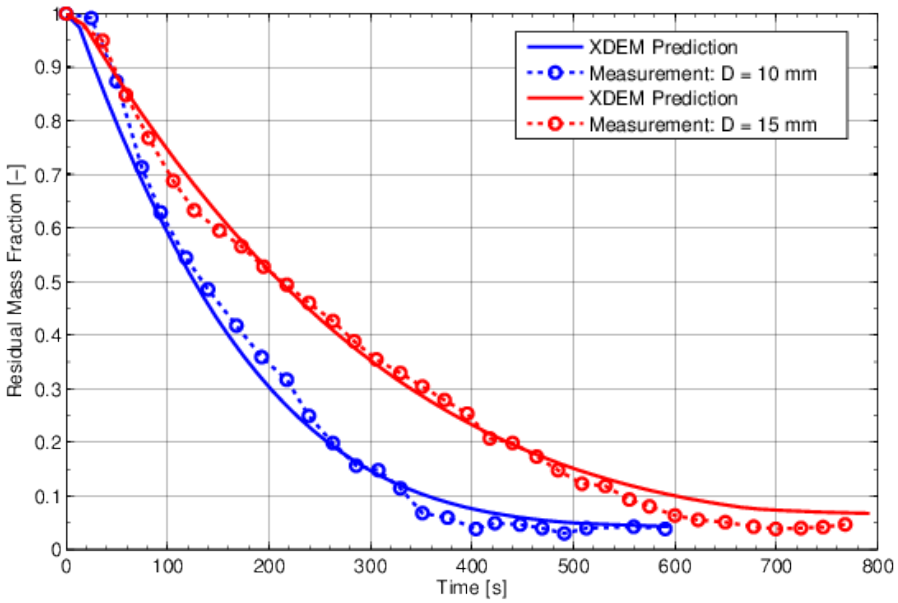


Figure 5: Comparison between measurements and predictions for gasification of char particles of 10 mm and 15 mm diameter exposed to a heating temperature of 773 K [16].

be observed through different instants in time. At the initial time, $t = 0$ s, the temperature of the domain is given by the ambient conditions $T = 600$ K. As air is injected through the inlet, the hot stream at $T = 1200$ K heats up the particles and the ambient air. The heat propagation can be appreciated through the different instants in time $t = 120$ s, $t = 400$ s, and $t = 800$ s. The particles located in front of the inlet are the first to receive the blast of hot air and to rise its temperature. The hot air moving through the void space of the packed bed continue heating up the particles. As observed from Fig. 6, the heat propagates from the inlet towards the side-walls and to the center of the reactor. At the last time, $t = 800$ s, temperatures over 1000 K are reached, principally at the particles located around the inlet. The CO concentration in the flow and the shrinking of the particles are the result of gasification, heat and mass transfer between the solid and gas phases. The carbon and oxygen react to form CO according to Eq. (23) just after the necessary temperature to activate the reaction has been reached. CO is then transported through the void space of the packed bed and to the flow above the particles. The effects of gasification are visible by the decrease in the diameter of the particles and the production of CO at $t = 800$ s, Figs. 7 and 8. A mass fraction reduction of approximately 20 % is observed at the particles that are directly located at the inlet and have been exposed to a longer heating period, $t = 800$ s.

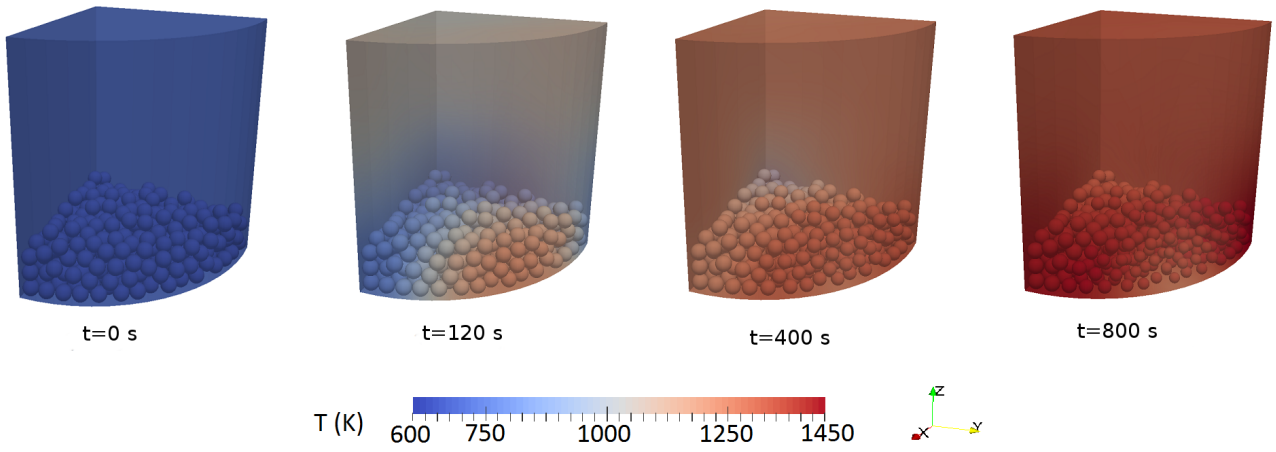


Figure 6: Temperature distribution at the particles surface and the flow field at different times.

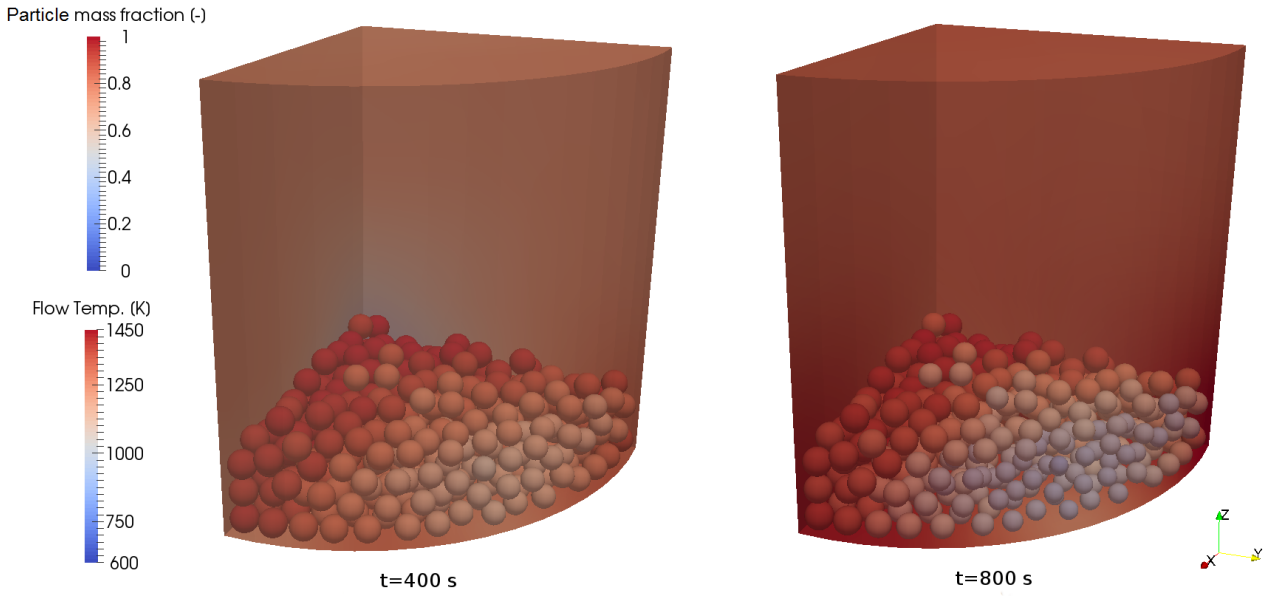


Figure 7: Particle mass fraction and flow temperature distribution in the reactor.

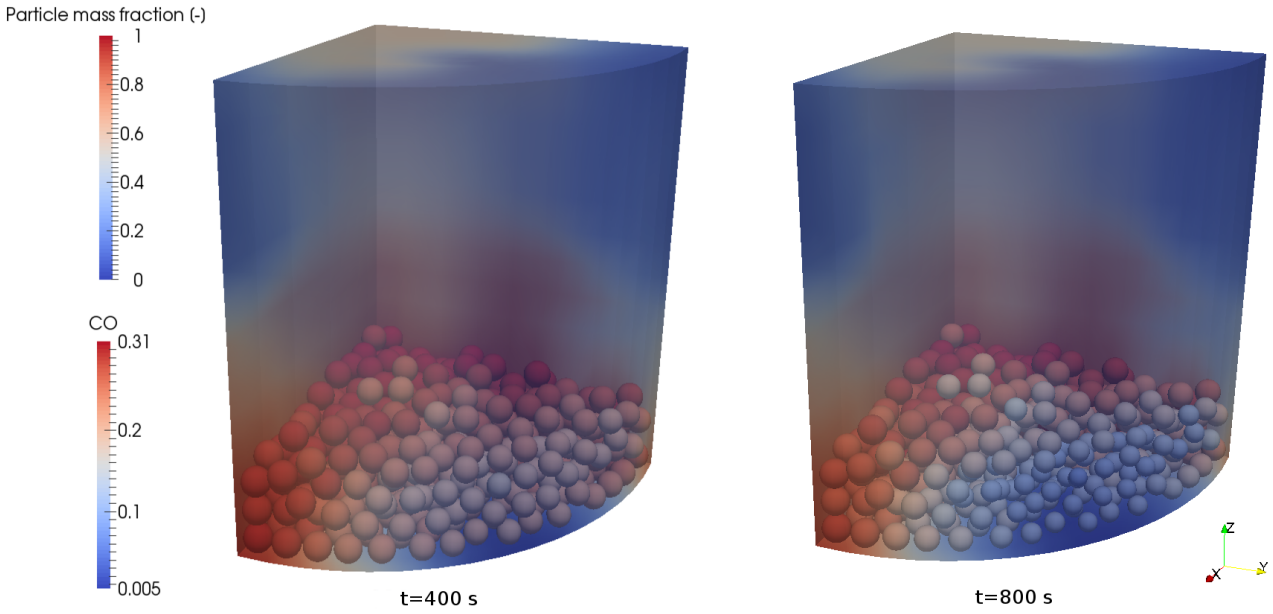


Figure 8: Particle mass fraction and CO concentration.

The XDEM-Dynamics module is used to account for the motion of the coke particles. A combination of the conversion and dynamics modules coupled with the CFD gas phase allows the XDEM to predict the motion and chemical conversion of each particle. The computational domain shown in Fig. 4 is used to simulate the motion of the coke particles inside the reactor by using the XDEM-Dynamics. Since the particles are also exposed to hot air, chemical conversion is solved with the conversion module. As observed from the previous case, gasification of coke particles require large exposure time to a heating source. This results in highly computational costs. In order to avoid such large computing time, the following test-case runs for a simulation time of 10 s. This time is enough to analyze the motion of the particles and the the heat-up process. Based on the previous case it is assumed that if the run time were increased, gasification process would take place. Figure 9 shows the temperature distribution in the flow and at the surface of the particles. Streamlines colored by the flow temperature are displayed to better visualize the flow going through the particles and the reactor. From Fig. 9 it

can observed how the injected flow at $T = 1200\,\mathrm{K}$ progressively heats-up the particles. Right after the injection, $t = 0.2\,\mathrm{s}$, the hot jet penetrates through the packed bed pushing forward the particles. Due to the packing of the particles and the location of the inlet, the jet immediately impacts on the solid bodies and it divides in two main streams. One stream continues in direction to the center of the reactor and the other continues in radial direction around the outer-walls. The particles located in front and around the inlet are pushed back and upwards by the stream traveling to the center.

The inject hot gas moves through the void space of the packed bed transferring its thermal and kinetic energy to particles. As it can be observed from Fig. 9 the streamlines represent the increase of the temperature in the flow field from the initial time $t = 0\,\mathrm{s}$ up to $t = 10\,\mathrm{s}$. Most of the thermal energy contained in the injected gas is absorbed by the particles. As the time increases the heat-up in the particles is more visible. As expected, the particles that are exposed directly to the hot gas are the first to heat-up and reach the highest temperatures. Figure 9 also shows the temperature at the surface of the particles. From the ambient condition, $T = 600\,\mathrm{K}$, it takes around 10 seconds to heat-up few particles to a more than 1000 K. Since the particles are moving and changing location not all of them close to the inlet show the same the temperature. However, the particles with larger temperatures are located around the inlet.

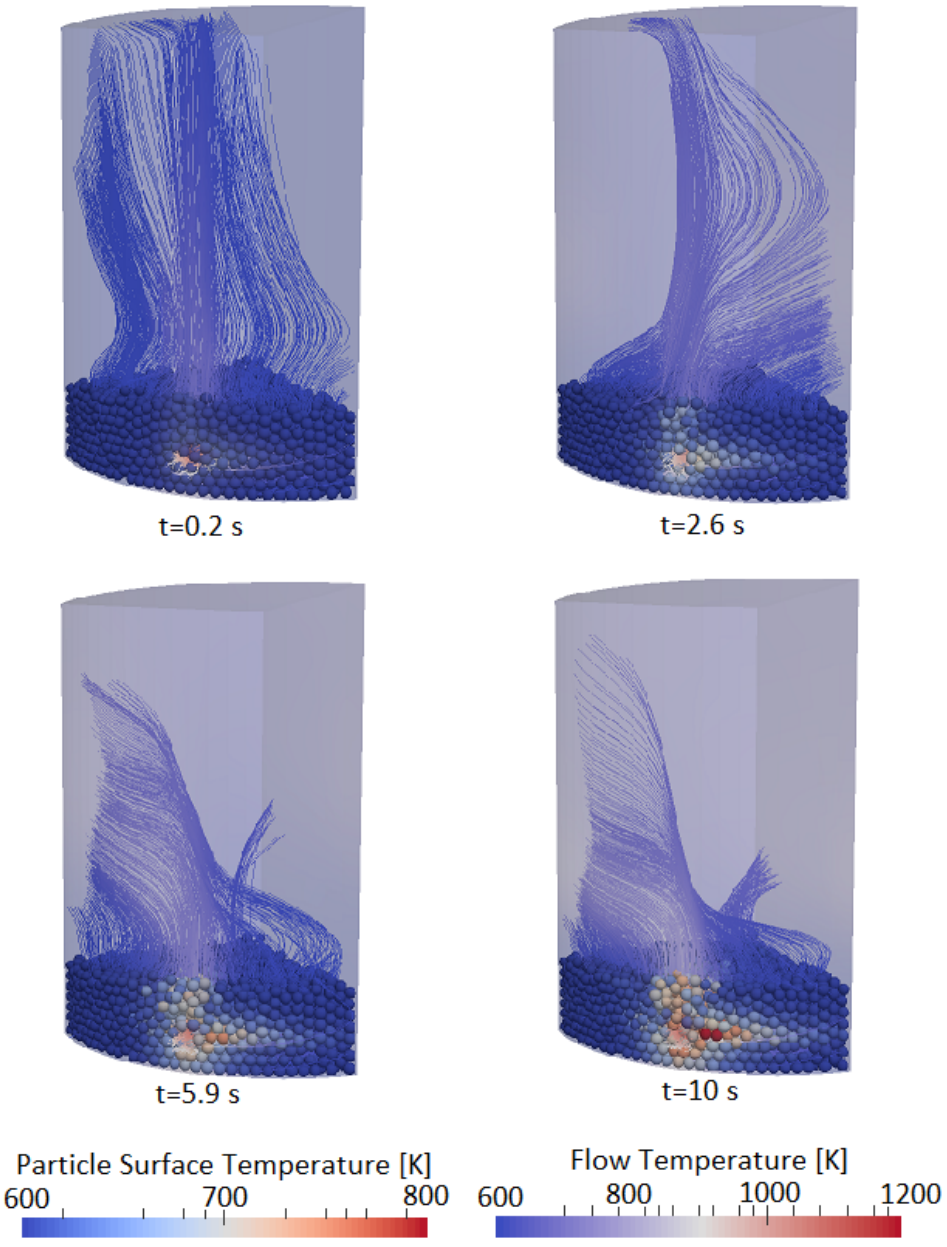


Figure 9: Temperature distribution in the flow field and at the particles surface at different times.

The air injected into the packed bed generates a circulation region with the particles moving around of it. This can be appreciated in Fig. 10 where a close-up around the inlet is shown. At the initial time, $t = 0\,\mathrm{s}$, the particles are static inside the reactor. When the air is injected at $u = 20\,\mathrm{m\,s}$ the particles located right at the inlet are pushed inwards. As the air continue penetrating into the packed bed, the particles are further displaced towards the center leaving a small cavity occupied only by the incoming air. Due to the displacement of these particles and the jet penetrating deeper into the packed bed and displacing more entities. The particles above this layer fall into the cavity to fulfill the free space. Since air is supplied continuously, the particles falling into the cavity and then been pushed inwards is repeated. Since the air speed is not large enough to penetrate deeper and push the particles further towards the center, these particles are rotating around the incoming air as shown in Fig. 11. The effect of the injected air into the particles extends up to approximately one third of the radius of the cylindrical reactor, 130 mm. It seems that the momentum of the jet is not enough to significantly displace the particles located beyond this distance. However, the thermal energy of the jet is enough to heat-up those particles. The rotational movement of the particles located near the inlet is responsible for the non uniform temperature distribution observed in Fig. 9. In particular, the particles with higher temperatures are located above and below the inlet due to this vertical rotational movement. The formation of the cavity and the recirculation region around the jet is a common pattern to be expected in such typed of gas-particle beds [31, 32].

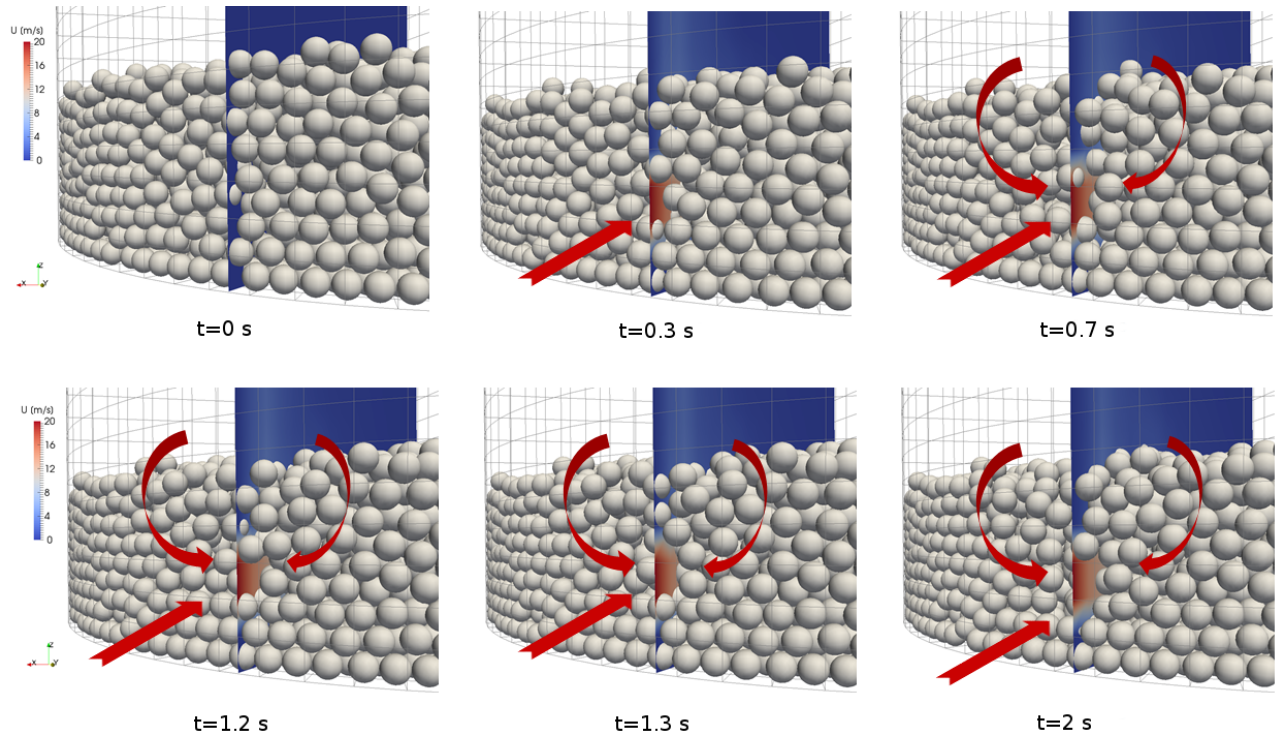


Figure 10: Velocity distribution and displacement of the particles around the inlet.

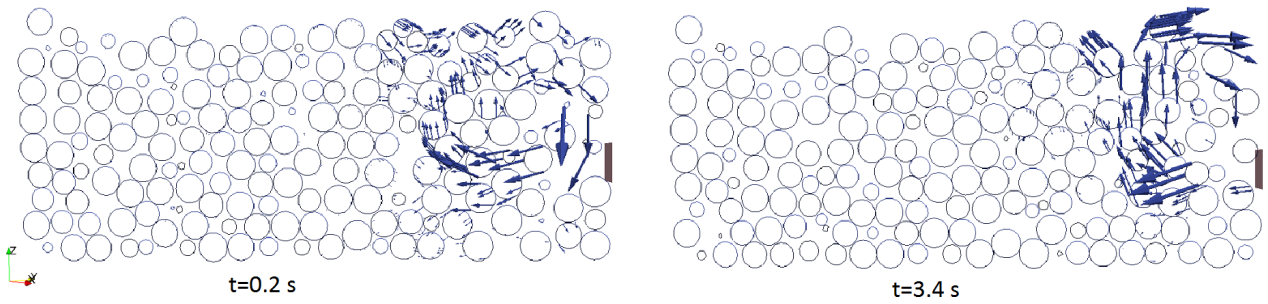


Figure 11: Velocity vector indicating the trajectory of the particles. 2D-cut at the cross-section in front of the inlet.

Conclusions

In the present case-study the developed numerical framework eXtended Discrete Element Method (XDEM) has been used to simulate the injection of gas into a packed bed of particles. The XDEM conversion and dynamics modules coupled to an implemented CFD solver in OpenFoam have been used to account for the motion and chemical conversion of particles subject to lateral gas injection. In a first step, the XDEM-Conversion module was validated for a single coke particle exposed to a heating source. The available experimental data was compared with the predicted values for the reduction of mass fraction and the particle diameter. Since good agreement between the measured and predicted values was obtained, the XDEM-Conversion was applied to a multiple coke particles to predict the heat-up and gasification process. The results shown how the particles shrink and loss their mass as the heating time increases. Due to the internal reaction part of the mass of the coke particles was transformed into CO and transported into the flow. The XDEM conversion and dynamics modules together with the CFD solver were used to predict the motion of the particles exposed to the lateral gas injection. In this case, the expected formation of a cavity and circulating region around the gas was observed. The transfer of thermal energy from the hot gas to the particles was observed by the increase in the temperature of the particles. The circulating region was also confirmed by the temperature distribution within the packed bed. The XDEM showed the ability to predict the motion and chemical conversion of particles in a packed bed subject to lateral gas injection. However, for better representation of the case in a real industrial application, such as a blast furnace, it is necessary to account for more realistic geometries and boundary conditions.

Acknowledgments

The authors would like to thank the Luxembourg National Research Foundation (FNR) for the support of this project.

References

[1] Adema, A. (2014) DEM-CFD Modelling of the ironmaking blast furnace. *PhD Thesis* TU Delft, The Netherlands.

[2] Geerdes, M., Toxopeus, H., van der Vliet, C., Chaigneau, R., and Vander, T. (2009) Modern Blast Furnace Ironmaking: An Introduction. *IOS Press*.

[3] Mathieson, J. G., Truelove, J. S., and Rogers, H. (2005) Toward an understanding of coal combustion in blast furnace tuyere injection. *Fuel* **84(10)**, 1229-1237.

[4] Burgess, J. M. (1985) Fuel combustion in the blast furnace raceway zone. *Progress in Energy and Combustion Science* **11(1)**, 61-82.

[5] Khairil, K., Kamihashira, D., and Naruse, I. (2002) Interaction between molten coal ash and coke in raceway of blast furnace. *Proceedings of the Combustion Institute* **29(1)**, 805-810.

[6] Peters, B., Dziugys, A., and Navakas, R. (2012) A shrinking model for combustion/gasification of char based on transport and reaction time scales. *Mechanika* **18(2)**, 177-185.

[7] Gidaspow, D. (1994) Multiphase Flow and Fluidization: Continuum and Kinetic Theory Descriptions. *Academic Press*

[8] Kuipers, J. A. M., and van Swaaij W. P. M. (1997) Application of Computational Fluid Dynamics to Chemical Reaction Engineering. *Reviews in Chemical Engineering*.

[9] Xu, B. H., and Yu, A. B. (1998) Comments on the paper numerical simulation of the gas-solid flow in a fluidized bed by combining discrete particle method with computational fluid dynamics - reply. *Chemical Engineering Science*. **53(2)**, 2646-2647.

[10] Peters, B. (2013) The extended discrete element method XDEM for multi-physics applications. *Scholarly Journal of Engineering Research* **2(1)**, 1-20.

[11] Cundall, P. A. and Strack, O. D. L. (1979) A discrete numerical model for granular assemblies. *Geotechnique* **29**, 47-65.

[12] Gallas, J. A. C., Herrmann, H. J., and Sokolowski, S. (1992) Convection cells in vibrating granular media. *Physical Review Letters* **69(9)**, 1371-1374.

[13] Walton, O. R. and Braun, R. L. (1986) Viscosity, granulartemperature, and stress calculations for shearing assemblies of inelastic, frictional disks. *Journal of Rheology* **30(5)**, 949-980.

[14] Džiugys, A. and Peters, B. (2001) An approach to simulate the motion of spherical and non-spherical fuel particles in combustion chambers. *Granular Matter* **3(4)**, 231-266.

[15] Mahmoudi, A. H., Hoffmann, F., and Peters, B. (2014) Application of XDEM as a novel approach to predict drying of a packed bed. *International Journal of Thermal Sciences* **75**, 65-75.

[16] Peters, B. (2003) Thermal Conversion of Solid Fuels. *Internal Series on Developments in Heat Transfer, WIT press*

[17] Peters, B., Dziugys, A., and Navakas, R. (2010) A discrete approach to thermal conversion of solid fuel by the discrete element method (dpm). *Modern Building Materials, Structures and Techniques*

[18] Mahmoudi, A. (2015) Prediction of Heat-up, Drying and Gasification of Fixed and Moving Bed by the Discrete Particle Method (DPM). *PhD Thesis, University of Luxembourg*.

[19] Faghri, A. and Zhang, Y. (2006) Transport Phenomena in Multiphase Systems. *Elsevier Academic Press*

[20] Grønli, M. (1996) A theoretical and experimental study of the thermal degradation of biomass. *PhD Thesis, NTNU Trondheim*.

[21] Shih-I, P. (1977) Two-Phase Flow. *Vieweg Tracts in Pure and Applied Physics*

[22] Dullien, F. A. L. (1979) Porous Media Fluid Transport and Pore Structure. *Academic Press, San Diego*

[23] Peters. B. (1997) Numerical simulation of heterogeneous particle combustion accounting for morphological changes. *27th International Conference on Environmental Systems* **SAE paper 972562**, USA.

[24] Peters, B. (1999) Classification of combustion regimes in a packed bed based on the relevant time and length scales. *Combustion and Flame* **116**, 297-301.

[25] Di Felice, R. (1994) The voidage function for fluid particle interaction system. *International Journal of Multiphase Flow* **20**, 153-159.

[26] OpenFoam User Guide, V.2.0.0. (2011)

[27] Bird, R. B., Stewart, W. E., and Lightfoot E. N. (1960) Transport Phenomena. *John Wiley & Sons.*

[28] Teng, H. and Zhao, T. S. (2000) An extension of Darcy's law to non-stokes flow in porous media. *Chemical Engineering Science* **55**, 2727-2735.

[29] Lage, J. L., Lemos, M. D., and Nield, D. (2002) Transport Phenomena in Porous Media II. *Modeling Turbulence in Porous Media, Pergamon*, **Chap. 8**, 198-230.

[30] Schäffer, B. and Wyrsch, F. (2000) Untersuchungen der Produktzusammensetzung bei thermochemischen Konversionsprozessen von Biomasse. *Diplomarbeit, ETH Zürich, Institut für Energietechnik, LTNT.*

[31] Hilton, J. E. and Cleary, P. W. (2012) Raceway formation in laterally gas-driven particle beds. *Chemical Engineering Science* **80**, 306-316.

[32] Nogami, H., Yamaoka, H., and Takatani, K. (2004) Raceway Design for the Innovative Blast Furnace. *ISIJ International* **44**, 2150-2158.

An original DEM bearing model with electromechanical coupling

C. Machado[1,a)], S. Baudon[1,4], M. Guessasma[1], V. Bourny[1,2], J. Fortin[1,2], R. Bouzerar[3]and P. Maier[4]

[1]Laboratoire des Technonologies Innovantes (LTI EA3899), Université de Picardie Jules Verne, France

[2]ESIEE Amiens, 4 quai de la Somme, 80082 Amiens Cedex 2, France

[3]Laboratoire de Physique de la Matiére Condensée (LPMC EA2081), Université de Picardie Jules Verne, France

[4]société EREM, ZA Sud, Rue de la sucrerie, 60130 Wavignies, France

[a)]Corresponding and presenting author: charles.machado@u-picardie.fr

ABSTRACT

Rolling bearings are one of the most important and frequently components encountered in domestic and industrial rotating machines. Statistical studies show that these bearings are considered as critical mechanical parts which represent between 40% and 50% of malfunction in rotating machineries. We performed an electrical monitoring of a bearing and numerical aspects using smooth contact dynamic are studied. An original elastic 2D modelling by discrete elements (DEM) reproduces the dynamic and the mechanical behavior of a bearing [1]. An electromechanical coupling is introduced to provide monitoring solutions [2]. This study proposes an original method of simulating the bearings to analyze dynamic stress in rings and to detect malfunctions (defects or unusual load) in the impedance of a ball bearing over time. The bearing is seen as a polydisperse granular chain where rolling elements and cage components interact with a Hertzian contact model. Moreover, rings (and housing) are also taken into account using a cohesive model [3]. Indeed, while many studies have been conducted on bearing simulation using FEM and multibody approaches, this new discrete model gives relevant information on physical phenomena in the contact interface. Roller-race contacts are analyzed in detail with an electromechanical coupling. One of our objectives is to investigate the sensitivity of the electrical measurement due to the variation of mechanical loading.

Keywords: bearing, electromechanical coupling, DEM, electrical transfer, contact model, roughness

Introduction

Rolling elements bearings are among the most important components in rotating machinery. In order to ensure the industrial systems availability and the safety of goods and persons, the monitoring and diagnosis of bearing defects have to be considered with prime importance and the challenges in terms of productivity are non-negligible. Recently, maintenance has led to extensive research with the development of new methods. Usually, vibrations and sounds of the machine are followed over time with a sensor and coupled analysis of time domain and frequency domain give information about the bearing state [4, 5]. Thermal and current motor analysis may be implemented to confirm the presence of abnormalities. A detected defect means that the damage is already sufficiently pronounced to be corrected and the bearing has to be changed. Finally, a defect is often due to mounting problem which implies unbalanced load, excessive load, misalignment... In this paper, a mechanical comparison between rigid housing and elastic housing is discussed. The rigid assumption becomes unrealistic when the machine design is optimized to minimize congestion and to reduce costs in raw material. Knowledge of the state of load bearing is particularly important but in practice, it is difficult to determine the loading bearing accurately. A new tool based on electrical measurements is presented in order to monitor the loads before problems occurs. Moreover, this electrical measurement has proven itself for a low speed application where other measurements are difficult to implement. A test bench has allowed to find some sensitivities relative to a charge status in the electrical measurement[6]. Although, the presence of an electric current through the ball bearing is ordinary harmful but the current densities required to perform a relevant experimental measurement are sufficiently low to cause damage [7, 8]. While most numerical studies on the bearing are carried out with finite

elements (FEM) [9, 10] and multibody approaches [11, 12], an original numerical approach using discrete element method is described in this paper and dynamic electromechanical simulations are studied. These developments are a straight continuation of previous work initiated in [13, 14] with notable advances in rings modelling.

DEM Mechanical modelling

A ball bearing is made of rolling elements constrained by two rings. A cage ensures a constant space between each rolling element. This component is seen as a granular chain. To distinguish themselves from multibody and FEM approaches, some important mechanical considerations are modelled in this paper and a dynamical resolution is proposed. The cage component is made of discrete elements moving freely along the pitch radius ($R_{pich} = \frac{R_{race}^{inner} + R_{race}^{outer}}{2}$) and the rings and the housing are elastic. The rings can be deformed under mechanical load if the housing is sufficiently flexible. In this original description, a bearing is represented by a collection of polydispersed (cylindrical or spherical) rigid particles. Contact interactions with particles are given with a contact model for the interface description and with a cohesive model for the continuum media (figure 1). A

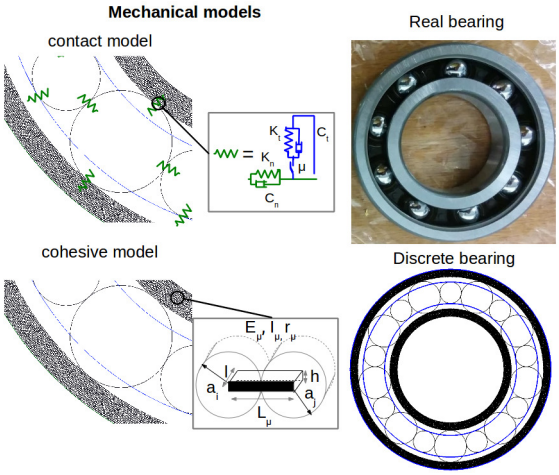


Figure 1. General mechanical modelling

commonly used radial ball bearing type 6208 is selected for modelling and its dimensions are given in table 2 :

$R_{rolling}$	R_{cage}	R_{race} inner	R_{race} outer	inner ring thickness	outer ring thickness
0.0063 m	0.0042 m	0.024 m	0.0366 m	0.003 m	0.003 m

Figure 2. Dimensions of a radial ball bearing 6208

Contact description

By using the smooth contact DEM, developed by Cundall and Strack [15, 16], the contact forces in a bearing are described with a contact model depending on elastic force displacement law, Coulomb's friction and viscous damping. The principle of the calculation is based on dynamic considerations and the contact occurs only when particles penetrate which means that a contact between a rolling element and a ring or a contact between a rolling element and a cage component is proved. The equivalent model of the contact is given in figure 1 using analogies with damped springs mass systems (K_n, K_t, C_n and C_t) and the dry friction coefficient μ, set to $\mu = 0.1$ are introduced. The lubricant effect is not taken into account and a rough interface is modelled [17]. The force $\vec{F_i}$ between particles at the interface includes the inter-particle interaction forces and the external forces.

$$\vec{F_i} = \sum_{j \neq i} \vec{F_{ij}} + \vec{F_{ext,i}} \tag{1}$$

Where $\vec{F_{ij}}$ is the force exerted by particle j to particle i. $\vec{F_{ext,i}}$ are the external forces on particle i (gravity, loading, ...). The contact force $\vec{F_{ij}}$ is deduced from analogies with damp-spring. From figure (1), this model includes a normal component and a tangential component. $\vec{F_{ij}}$ is then decomposed as follow :

$$\vec{F_{ij}} = F_n\vec{n} + F_t\vec{t} \tag{2}$$

F_n is the contact force in the normal direction and F_t is the contact force in the tangential direction. By considering the analogies with a damped spring mass system, where K_n,C_n and K_t, C_t represent the stiffness and the viscous damping coefficient, in the normal direction $\vec{n}$ and in the tangential direction $\vec{t}$. The overlap between particles $\vec{u} = u_n\vec{n} + u_t\vec{t}$ gives the contact force :

$$\begin{cases} F_n = K_n \times u_n + C_n \times \vec{u}.\vec{n} \\ F_t = K_t \times u_t + C_t \times \vec{u}.\vec{t} \end{cases} \tag{3}$$

where $\vec{u}$ is the relative velocity of the contact point between particles. The tangential overlap u_t can be approximated by the expression : $u_t = \vec{u}.\vec{t}\Delta_t$, where Δ_t is the time step. F_t is a candidate force because the slider μ, due to dry friction is considered. Coulomb's friction law is written in equation 4 and determines whether the contact is slipping or sliding :

$$F_t = -min(F_t, \mu F_n) \times sgn(\vec{u}.\vec{t}) \tag{4}$$

The expressions of normal and tangential stiffness are given from the elastic solid mechanics analysis of Hertz-Mindlin theory [18, 19] :

$$\begin{cases} K_n = 4E\dfrac{a_i a_j}{a_i + a_j} \\ K_t = K_n \dfrac{1-\nu}{1-\frac{1}{2}\nu} \end{cases} . \tag{5}$$

K_n and K_t are related to mechanical properties the Young's modulus E, the Poisson's ratio ν and the dimensions of particles in contact (a_i, a_j). The harmonic behaviour of linear model with constant parameters is well known and adapted in a first approximation for a description of a roller bearing. A general load-deflection relationship without damping is written as $F_n = K_n U_n^N$, where N and K_n depend on the bearing type ($N = 10/9 \sim 1$ for a roller-raceway contact, $N = 3/2$ for a ball-raceway contact, ...). The role of interactions at the contact plays an important role in the distribution of efforts. Harris [20] offers similar stiffness models derived from Hertz's theory. A critical viscous damping ratio $C_{n,t}$ is introduced by equation 6, where m^* is the reduced mass :

$$C_{n,t} = 2\sqrt{K_{n,t}m^*} \tag{6}$$

Other viscous damping coefficients can be introduced if lubricant effects are considered [21, 22]. A simple bearing is made of $2Z + 1$ discrete elements where Z are dedicated to rolling elements, Z others are dedicated to cage components and the last one represents the inner race/ring or shaft.

Cohesive description

Rings and the housing may be deformed under mechanical loadings. In order to simulate a 2D continuous material with DEM, the rings are discretized by a dense polydisperse granular assembly. The generation is controlled with Lubachevsky-Stillinger's algorithm [23] so as to satisfy the following properties :

- Isotropic contact orientation
- Local homogeneous properties (coordination number, local porosity, ...)
- Compacity close to 86-87 % (Random close packing [24])

In order to reflect the mechanical behaviour of continuous medium, contacts must be persistent and a cohesive contact law is considered (figure 1). In the proposed DEM formulation, the interaction between two particles in contact is modelled with a beam of length L_μ, Young's modulus E_μ, cross-section A_μ and quadratic moment $I\mu$ (figure 1) [25]. Therefore, the cohesive contacts are maintained by a vector of three-component generalized forces acting as internal forces. The normal component acts as an attractive or repulsive force, the tangential component allows to resist to the tangential relative displacement and the moment component counteracts the

bending motion.
From figure 1, A_μ is rectangular with depth $l = 1cm$ and h, the height of the cross section defined by :

$$h = r_\mu \frac{a_i + a_j}{2} \tag{7}$$

where $r_\mu \in]0, 1]$ is a dimensionless radius, a_i and a_j are respectively the radius of particles i and j. The cohesive forces and moments between two particles i and j are given as follow:

$$\begin{cases} m_i \ddot{u}_i = F_i^{ext} + \sum_j F^{i \to j} \\ I_i \ddot{\theta}_i = M_i^{ext} + \sum_j M^{i \to j} \end{cases} \tag{8}$$

where m_i is the elementary matrix and I_i is the quadratic moment of intertia of the particle i. $F^{i \to j}$ and $M^{i \to j}$ are respectively the force and the moment of interaction of particle j on i. F^{ext} et M^{ext} are respectively the external force and moment of acting on particle i.
The local cohesion forces between particles i and j are deduced from the following linear system :

$$\begin{bmatrix} F_n^{i \to j} \\ F_t^{i \to j} \\ M_{i \to j}^{int} \end{bmatrix} = \begin{bmatrix} \frac{E_\mu A_\mu}{L_\mu} & 0 & 0 & 0 \\ 0 & \frac{12 E_\mu I_\mu}{L_\mu^3} & \frac{6 E_\mu I_\mu}{L_\mu^2} & \frac{6 E_\mu I_\mu}{L_\mu^2} \\ 0 & \frac{6 E_\mu I_\mu}{L_\mu^2} & \frac{4 E_\mu I_\mu}{L_\mu} & \frac{2 E_\mu I_\mu}{L_\mu} \end{bmatrix} \begin{bmatrix} u_n^i - u_n^j \\ u_t^i - u_t^j \\ \theta_i \\ \theta_j \end{bmatrix}$$

where θ_i and θ_j are respectively the rotations of particles i and j. $u_n^{i,j}$ and $u_t^{i,j}$ are respectively the normal and tangential displacements. The numerical resolution is based on an explicit time integration with a formulation based on a Verlet scheme. In the ball bearing context, the rings are made of steel and the identification of model parameters E_μ and r_μ is correlated with macroscopic Young's modulus $E_M = 210GPa$ and Poisson's ratio $\nu_M = 0.3$. A procedure based on a uniaxial quasi-static tensile test [26] suggests to choose $E_\mu = 505GPa$ and $r_\mu = 0.5$. The rings and the housing are composed of millimeter polydisperse particles.

Electromechanical modelling

The electrical transfer in a bearing in operation is a complex mechanism depending on intrinsic mechanical, electrical properties of materials in contact and on properties of the interface (roughness, lubricant film, oxydation, temperature, ...). Electrical response depends also on mechanical load and on rotation speed. For moderate rotation speeds ω and heavy loads F_r, the lubricant thickness in the interface may be neglected [6] and a simple electrical model, based on analogies with resistors is considered [2, 13, 14]. We assumed that the temperature is constant and the oxyde layer on the surface of particles is neglected but the effect of roughness is considered. An electrical macroscopic resistance is associated to each rolling element in contact with both rings using expression 9 :

$$\frac{1}{R_{ij}^k} = \frac{\gamma S_i S_j}{2V_b}(1 - cos\theta) \tag{9}$$

where γ is the electrical conductivity of steel ($\gamma = 5.8 \times 10^7 S.m^{-1}$), V_b is the volume of the rolling element, θ is the angle formed by the points i and j ($\theta = \pi$ for radial bearings). The coupling between the mechanical and electrical computation is carried out by Hertz's theory where S_i and S_j denote contact areas. The elements of the cage are insulating (made of polyamide) therefore only the rolling elements are involved in current transfer. Considering a cohesive model implies rough races depending on the discretization. The contact between a rolling element and a race is supported by small cohesive particles. This description of rough contact using spherical caps can be found in the Greenwood's work [27, 28]. Unlike previous works where a perfect contact was considered leading to overly conductive simulations [14], the contact area responsible of the electrical transfer is the sum of spots using Hertz's theory, as suggested by figure 3(b). The surface of rolling elements is supposed perfect but in practice the arithmetical rugosity of a rolling element ($Ra_b = 10^{-8}$ m) is about ten time smaller than the arithmetical rugosity of races ($Ra_{race} = 10^{-7}$ m). The roughness is numerically overestimated for reasons of time computing.

For a rolling element k, at the angular position ψ^k, the radial local load Q_ψ^k is distributed over several "micro-contacts" or spots on the inner race (or outer race), as shown in figure 3(b). The contact area responsible of

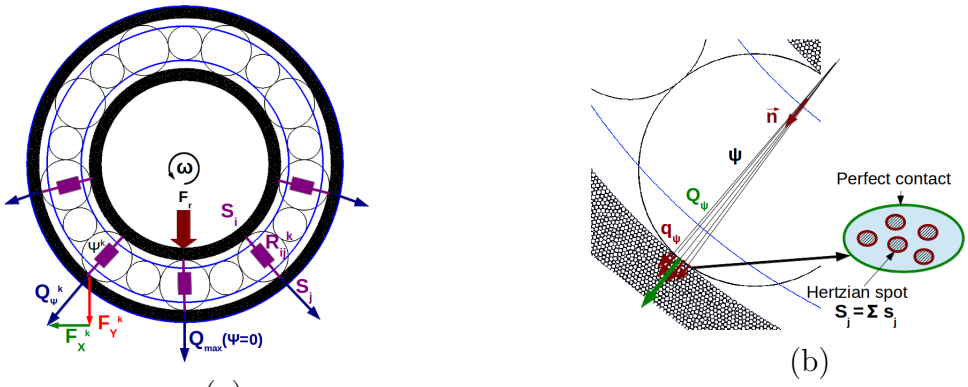


Figure 3. (a) Load projection and electrical circuit (b) interface description

the electrical transfer is written as follow :

$$S = \sum_{i=1}^{m_c} s_i = \sum_{i=1}^{m_c} \pi \left(\frac{3 \times q_\psi^i R^*}{4E^*} \right)^{2/3} \tag{10}$$

where m_c is a number of "micro-contact" within a contact between a rolling element and a ring, depending on the discretization. q_ψ^i denotes the radial load transmitted by the "micro-contact" i. R^* and E^* respectively characterize the relative radius of curvature and the reduce modulus.

Simulation results

The electromechanical results are obtained for a fixed rotation speed $\omega = 500$ rad/s and the time step is $\Delta_t = 10^{-8}$ s. The considered bearing has no clearance which means that only 50 % of rolling elements are implicated in the electrical determination. The rolling components in the load zone form a parallel electrical resistor circuit, as shown in figure 3(a).

Mechanical analysis

Consider a bearing with rigid rings or deformable rings involves particular mechanical behaviours that affect the bearing fatigue lifetime and the electrical determination. The rigid description is made of $26100 + 19$ discrete elements and the elastic description is made of $36100 + 19$ discrete elements (figure 4(a)).

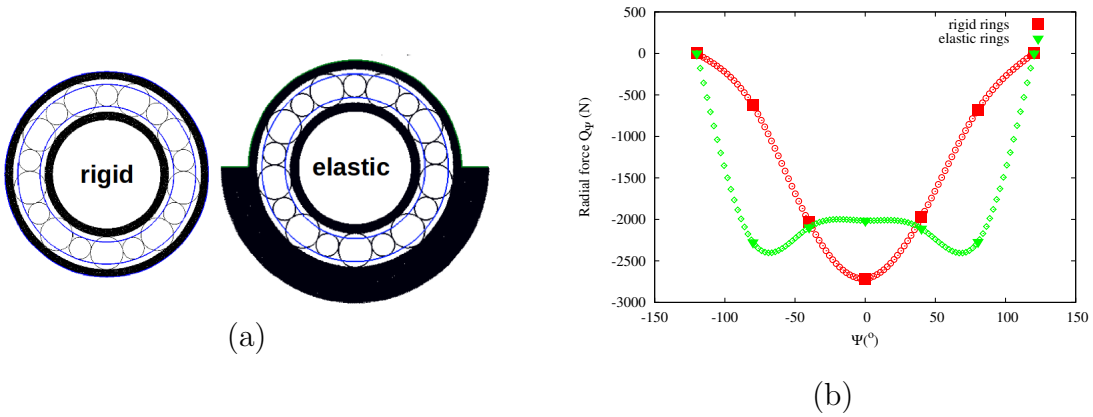


Figure 4. (a) DEM models (b) Static radial load distribution for $F_r = 6$ kN

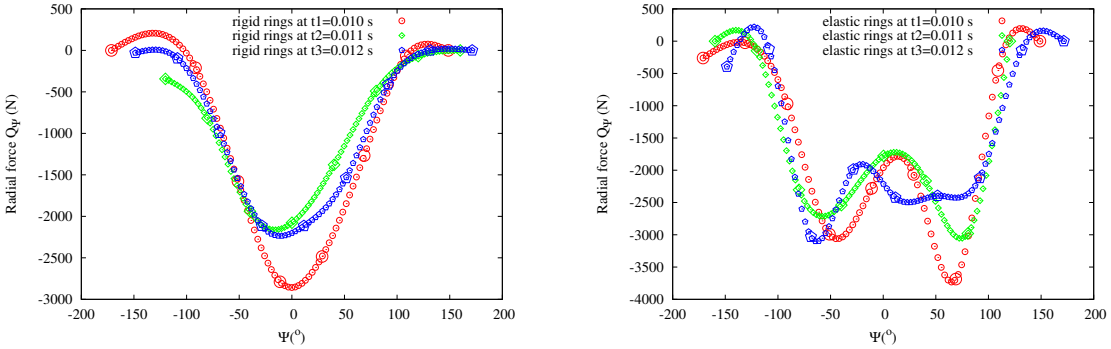
A radial load $\vec{F_r}$ = -6 $\vec{j}$ kN is applied on the inner ring according to the vertical direction. From figure 4(b), at the static equilibrium ($\omega = 0$), the local load distribution Q_ψ is represented according to angular position ψ, in rigid and elastic cases. Each point (■ or ▼) gives the position and the radial local load supported by a

rolling element and points ($\diamondsuit$ or $\odot$) give typical trends. The radial load distribution with rigid rings describes a sinusoid function which matches with the classical rigid theory [20] :

$$Q_\psi = Q_{max}\left(1 - \frac{1}{2\epsilon}(1 - cos\psi)\right)^N \tag{11}$$

where $Q_{max} \sim -2728$ N denotes the maximum radial local load at $\psi = 0°$ and for a roller bearing Q_{max} may also be determined using radial integral $J_r(\epsilon)$ with expression $Q_{max} = \dfrac{J_r(\epsilon = 0.5) \times F_r}{Z} = \dfrac{4.08 \times -6000}{9} = -2720$ N. The dimensionless load parameter ϵ describes the state of load. When no clearance or preload is considered, $\epsilon = 0.5$ means over half of rolling elements is involved in the radial distribution. N is relative to the stiffness model ($N = 3/2$ for ball bearing and $N \sim 1$ for roller bearing. The radial load distribution with elastic rings shows a symmetrical function about the vertical axis where $Q_{max} \sim -2250$N is found at angle $\psi \pm 80°$, caused by the roundness of the rings. There is no theoretical expression associated with this elastic distribution.

The following simulations are obtained by considering a non-zero rotational speed $\omega = 500 rad/s$. The static equilibrium described in figure 4(b) is replaced by complex dynamic regime where deformation modes of the rings and rough interfaces disturb the load distribution over time as suggested by figures 5.



(a) (b)

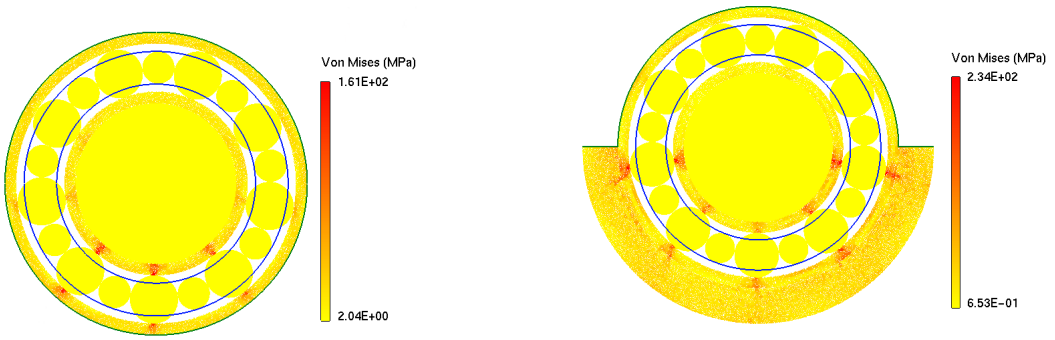
Figure 5. Radial load distribution for several instants at $F_r = 6$ kN and $\omega = 500$ rad/s (a) for rigid rings (b) for elastic rings

The rigid results over time show that the sinusoidal profile conserves the same shape and Q_{max} is time varying due to micro-contact variations. A similar analysis could be done for elastic results. These remarks demonstrate that even if a constant load F_r and a constant rotation speed ω are applied to the system, the mechanical response in the bearing is time varying. The main difference between rigid and elastic analysis, in dynamic or static attempts to show that areas prone to damage are dependent on the rigidity of the montage. As proof, the dynamical study of mechanical stress fields in the rings at same time (figure 6) shows that in the rigid case, the area or contact interface near the south pole ($\psi = 0$) is more prone to damages. In the elastic case, this area is pushed towards the embedding conditions close to $\psi \pm 90°$.

Electrical analysis

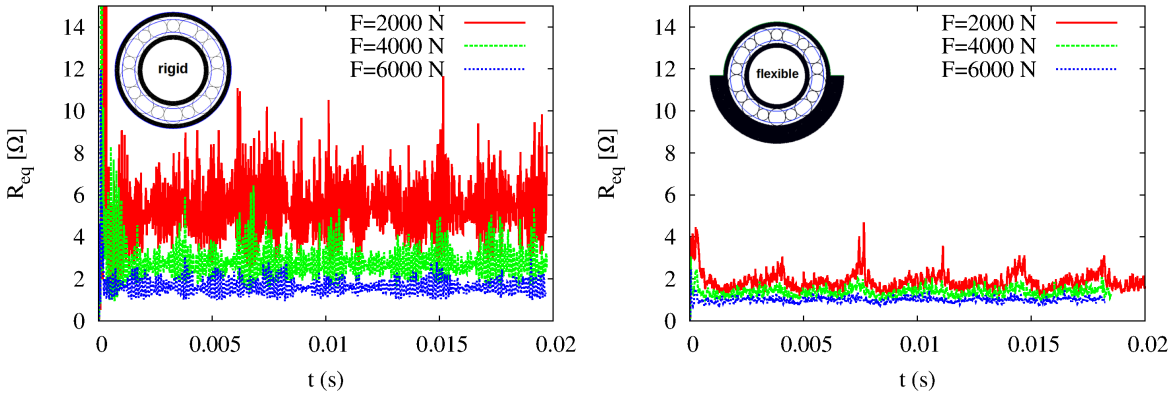
The electrical sensitivity over time is simulated for several radial load F_r at 50 kHz. The overall electrical resistance is given in figure 7(a) for rigid rings and in figure 7(b) for elastic rings.

In both cases, when the radial load F_r increases, the electrical resistance decreases with a non linear dependency according to Hertz's theory [14]. Typical values of the electrical resistance computed which may be assimilated to the mean resistances give the order of Ω. In rigid considerations, the resistance shows substantial variations in amplitude depending on load at high frequencies due to micro-contact variation (figure 7(a)). In flexible considerations, the micro-contact variation still exists but due to the elasticity of the system, the resistance is less noisy and a typical low frequency appears close to 300 Hz (figure 7(b)). This low frequency is assumed to



(a) (b)

Figure 6. Von Mises stress in a bearing at $F_r = -6$ kN and $\omega = 500$ rad/s (a)with rigid rings (b) with elastic rings



(a) (b)

Figure 7. Electrical resistance versus time with different radial loads F_r at $\omega = 500$ rad/s (a) for rigid rings (b) for elastic rings

be related to the system deformation modes. For an identical radial load, the flexible mounting systematically gives a lower resistance than the rigid case.

Conclusion

An original description of the dynamic behaviour of bearings with DEM is described and interesting electromechanical results are discussed. This type of modelling provides access to new quantities for understanding the mechanisms of damage (figure 6). Load distribution of the bearing is determined with a contact law based on analogies with damped springs and deformations of rings using a cohesive model are considered. In a static case with rigid rings, the contact model K_n verifies Harris's theory [20] and taking account of the rigidity of the rolling bearing implies significant effects. An electrical measurement is proposed to diagnose the state of load in operation. The electrical sensitivity of this measurement allows us to distinguish several radial loads. Subsequently, abnormal loads, misalignments and defects generated with decohesion will be imposed on ball bearings and their electrical signatures will be analysed. For now, the electrical model considers a rough contact but we could improve this model by taking into account the effect of lubricant with the theory of elastohydrodynamic lubrication [29]. In this case, the lubricant acts as a capacitor and an electrical model based on impedance spectroscopy has to be developed. In a future work, simulation results will be compared with experimental measurements for moderate speeds. Other simulations on a elementary rolling contact will introduce realistic roughness.

Acknowledgements : *This study has been carried out under project EROLLING2 (2015-2018) using the univer-*

sity Chair program on electrical transfer. Thanks to the "Région Nord-Pas de Calais-Picardie" for its financial support.

References

[1] K. Bourbatache, M. Guessasma, E. Bellenger, V. Bourny, and A. Tekaya. Discrete modelling of electrical transfer in multi-contact systems. *Granular Matter*, 14 (1):1–10, 2012.

[2] C. Machado, M. Guessasma, E. Bellenger, K. Bourbatache, V. Bourny, and J. Fortin. Diagnosis of faults in the bearing by electrical measures and numerical simulations. *Mechanics and Industry*, 15(5):383–391, 2014.

[3] H. Haddad. *Couplage MED-MEF : modélisation numérique du transfert thermique dans les interfaces de contact.* PhD thesis, Université de Picardie Jules Verne, 2013.

[4] R. B. Randall and J. Antoni. Rolling element bearing diagnostics - a tutorial. *Mechanical Systems and Signal Processing*, 25:485–520, 2011.

[5] N. Tandon and A. Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32(8):469–480, 1999.

[6] C. Machado. *Modélisation et simulation éllectromélcaniques par la MED des systèmes multi-contacts : application à la surveillance des roulements par une mesure électrique.* PhD thesis, Université de Picardie Jules Verne, 2015.

[7] J.R. Stack, T.G. Habetler, and R.G. Harley. Experimentally generating faults in rolling element bearings via shaft current. *IEEE transactions on industry applications*, 41 (1):25–29, 2005.

[8] D.F. Busse, J.M. Erdman, R.J. Kerkman, D.W. Schlegel, and G.L. Skibinski. The effects of pwm voltage source inverters on the mechanical performance of rolling bearings. *IEEE transactions on industry applications*, 33 (2):567–576, 1997.

[9] N. Demirhan and B. Kanber. Stress and displacement distributions on cylindrical roller bearing rings using fem. *Mechanics Based Design of Structures and Machines*, 36:86–102, 2008.

[10] A. E. Azianou, K. Debray, F. Bolaers, P. Chiozzi, and F. Palleschi. Modeling of the behavior of a deep groove ball bearing in its housing. *Journal of Applied Mathematics and Physics*, 1:45–50, 2013.

[11] M. Tiwari and K. Gupta. Dynamic response of an unbalanced rotor supported on ball bearings. *Journal of Sound and Vibration*, 238 (5):757–779, 2000.

[12] L. Xu, Y. Yang, Y. Li, C. Li, and S. Wang. Modeling and analysis of planar multibody systems containing deep groove ball bearing with clearance. *mechanism and machine theory*, 56:69–88, 2012.

[13] K. Bourbatache, M. Guessasma, E. Bellenger, V. Bourny, and J. Fortin. Dem ball bearing model and defect diagnosis by electrical measurement. *mechanical systems and signal processing*, 41:98–112, 2013.

[14] C. Machado, M. Guessasma, and E. Bellenger. Electromechanical modelling by dem for assessing internal ball bearing loading. *Mechanism and Machine Theory*, 92:338–355, 2015.

[15] O. D. L. Cundall and P. A. Strack. A discrete numerical model for granular assemblies. *Géotechnique*, 29:235–257, 1979.

[16] P. A. Cundall. Formulation of three-dimensional distinct element mode part 1. a scheme to detect and represent contacts in a system composed of many polyhedral blocks. *J. Rock Mech., Min. Sci. and Geomech*, 25:107–116, 1988.

[17] R. Stribeck. Ball bearings for various loads. *Trans. ASME*, 29:420–463, 1907.

[18] H. Hertz. "uber die beruhrung fester elastischer korper" on the contact of elastic solids. *reprinted in Miscellaneous Papers, Macmillan*, pages 146–162, 1896.

[19] R. D. Mindlin and H. Deresiewicz. Elastic spheres in contact under varying oblique force. *ASME journal of applied mechanics*, 20:327–344, 1953.

[20] T. A. Harris and M. N. Kotzalas. *Rolling Bearing Analysis : Essential concepts of Bearing Technology.* 2006.

[21] D. Downson and G. R. Higginson. *Elasto-hydrodynamic lubrication.* Pergamon Press, 2nd ed, 1977.

[22] B. J. Hamrock and W. J. Anderson. Analysis of an arched outer race ball bearing considering centrifugal forces. *ASME Journal of Tribology*, 95 (3):265–276, 1973.

[23] B. D. Lubachevsky and F. H. Stillinger. Geometric properties of random disk packings. *Journal of Statistical Physics*, 60 (5):561–583, 1990.

[24] G. D. Scott and D. M. Kilgour. The density of random packing of spheres. *Appl. Phys.*, 2:863–866, 1969.

[25] D. André, I. Iordanoff, J.-L. Charles, and J. Néauport. Discrete element method to simulate continuous material by using the cohesive beam model. *Comput. Methods Appl. Mech. Engrg*, 213-216:113–125, 2012.

[26] H. Haddad, W. Leclerc, M. Guessasma, C. Pélegris, E. Bellenger E., and N. Ferguen. Application of dem to predict the elastic behavior of particulate composite materials. *Granular Matter*, 17 (2):459473, 2015.

[27] J. A. Greenwood and J. B. P. Williamson. Contact of nominally flat surfaces. *Proc. of the Royal A.*, 295:300–319, 1966.

[28] J. A. Greenwood. Constriction resistance of the real area of contact. *Brit. J. Appl. Phys.*, 17, 1966.

[29] B. J. Hamrock and W. J. Anderson. *Rolling-element bearings.* National Technical Information Service, 1983.

High-order algorithms for nonlinear problems and numerical instability

***José Elias Laier[1]**

[1]Department of Structural Engineering, Engineering School of São Carlos, University of São Paulo, Brazil.

*Presenting author: jelaier@sc.usp.br

Abstract

The objective of this paper is to study the numerical behavior (accuracy and numerical instability) of two high-order order single step direct integration algorithm for nonlinear dynamic. These algorithms are formulated in terms of two Hermitian finite difference operators of fifth-order local truncation error. In addition, these algorithms are unconditionally stable with no numerical damping for linear dynamic problems. The attention is devoted to the classical second-order Duffing and Van der Pol equations, as well the non-linear elastic pendulum, including the first-order Lorenz and Lotka-Volterra equations. Numerical applications compare the results including with those obtained by the second-order Newmark method

Keywords: Numerical instability, nonlinear dynamic, Hermitian finite difference algorithms

Introduction

The objective of this paper is to study the numerical behavior of two high-order order single step direct integration algorithm for nonlinear dynamic. The first one has been developed by the author [1] and the second is the classical cubic Hermitian Algorithm developed by Argyris and Mlejek [2]. These algorithms are formulated in terms of two Hermitian finite difference operators [3] of fifth-order local truncation error. In addition, these algorithms are unconditionally stable with no numerical damping for linear dynamic problems. As the analytical treatment of the numerical instability of the resultant nonlinear difference equation (i.e. the numerical version of the differential equation) is quite complex, just numerical investigation is performed.

As the high-order algorithms takes into account the repeated differentiation of the governing equation, additional nonlinear terms are required to solve nonlinear structural dynamic problems. Thus, it is interesting to consider, for example, the classic iterative procedures presented by Argyris and Mlejek [2]. Although the presence of these additional nonlinear terms increases the number of operations in the iterative operations and introduces some numerical noise in comparison to the Padè-P_{22} algorithm family [4], the reduction obtained in the matrix factorization and higher orders of the relative radii errors are interesting attributes of the proposed algorithm. Numerical applications compare the results including with those obtained by Newmark method. The results show that the accuracy of both third-order algorithms is quite similar for refined mesh, but the numerical instability (that occurs for coarse mesh) is not similar.

Hermitian Operators

The step-by-step integration algorithm to be considered in this paper takes into account the following Hermitian operators [1] [3]:

$$\begin{aligned} &Ay_i + By_{i+1} + C\Delta t\dot{y}_i + D\Delta t\dot{y}_{i+1} + E\Delta t^2\ddot{y}_i + F\Delta t^2\ddot{y}_{i+1} + G\Delta t^3\dddot{y}_i + H\Delta t^3\dddot{y}_{i+1} = 0 \\ &A_1y_i + B_1y_{i+1} + C_1\dot{y}_i + D_1\dot{y}_{i+1} + E_1\Delta t^2\ddot{y}_i + F_1\Delta t^2\ddot{y}_{i+1} + G_1\Delta t^3\dddot{y}_i + H_1\Delta t^3\dddot{y}_{i+1} = 0 \end{aligned} \quad (1)$$

where Δt is the time step, i and i+1 indicate the step, y is the function to be integrated, $\dot{y}$, $\ddot{y}$ and $\dddot{y}$ are derivatives of the function with respect to time; A, B … G_1, H_1 are combination non-dimension parameters that define the order of accuracy (local truncation error) [3]. Table 1 presents the combination parameters for the algorithms herein considered.

Duffing equation

The Duffing equation and its first time derivative can be expressed as

Table 1. Combination Parameters

	A	B	C	D	E	F	G	H
Laier [1]	12	-12	6	6	1	-1	0	0
Argyris [3]	1	-1	1	0	21/60	9/60	3/60	-2/60
Newmark	0	0	1	-1	1/2	1/2	0	0
	A_1	B_1	C_1	D_1	E_1	F_1	G_1	H_1
Laier [1]	0	0	12	-12	6	6	1	-1
Argyris [3]	0	0	1	-1	6/12	6/12	1/12	-1/12
Newmark	1	-1	1	0	1/4	1/4	0	0

$$\begin{aligned} &\ddot{y}+\delta\dot{y}+\alpha y+\beta y^3=p\cos(\omega t)\\ &\dddot{y}+\delta\ddot{y}+\alpha\dot{y}+3\beta y^2\dot{y}=-p\omega\sin(\omega t)\end{aligned} \tag{2}$$

where α, β, δ, p and ω are parameters of the equation. The second and third derivatives present in equation (2) can be explicitly written by

$$\begin{aligned} &\ddot{y}=-\delta\dot{y}-\alpha y-\beta y^3+p\cos(\omega t)\\ &\dddot{y}=-\delta\left(-d\dot{y}-\alpha y-\beta y^3+p\cos(\omega t)\right)-\alpha\dot{y}-3\beta y^2\dot{y}-p\omega\sin(\omega t)\end{aligned} \tag{3}$$

Now, taking into account equation (3) and Hermitian operators (1) the following nonlinear recurrence first-order difference equation can be written:

$$\begin{aligned} &F(y_{i+1},\dot{y}_{i+1})=Ay_i+By_{i+1}+C\Delta t\dot{y}_i+D\Delta t\dot{y}_{i+1}+E\Delta t^2\ddot{y}_i+F\Delta t^2\ddot{y}_{i+1}+G\Delta t^3\dddot{y}_i+H\Delta t^3\dddot{y}_{i+1}=0\\ &G(y_{i+1},\dot{y}_{i+1})=A_1y_i+B_1y_{i+1}+C_1\dot{y}_i+D_1\dot{y}_{i+1}+E_1\Delta t^2\ddot{y}_i+F_1\Delta t^2\ddot{y}_{i+1}+G_1\Delta t^3\dddot{y}_i+H_1\Delta t^3\dddot{y}_{i+1}=0\end{aligned} \tag{4}$$

And so, the corresponding Newton iterative formula can be expressed as:

$$\begin{Bmatrix} y_{i+1}\\ \dot{y}_{i+1}\end{Bmatrix}_{j+1}=\begin{Bmatrix} y_{i+1}\\ \dot{y}_{i+1}\end{Bmatrix}_{j}-\begin{vmatrix} F_{y_{i+1}}(y_{i+1},\dot{y}_{i+1}) & F_{\dot{y}_{i+1}}(y_{i+1},\dot{y}_{i+1})\\ G_{y_{i+1}}(y_{i+1},\dot{y}_{i+1}) & G_{y_{i+1}}(y_{i+1},\dot{y}_{i+1})\end{vmatrix}_{j}^{-1}\begin{Bmatrix} F(y_{i+1},\dot{y}_{i+1})\\ G(y_{i+1},\dot{y}_{i+1})\end{Bmatrix}_{j} \tag{5}$$

were subscript y_{i+1} and $\dot{y}_{i+1}$ indicate the partial derivative with respect to these discrete variables and the subscript j and j+1 indicate the iteration step . Table 1 compares the first displacement peak results for three practical time-steps and the instable time step Δt limit for $\delta=0.4$, $\alpha=1.0$, $\beta=0.5$, p=0.5 and $\omega=0.5$.

Table 1. First peak displacement and instability limit

Δt	LAIER[1]	ARGYRES[3]	NEWMARK
0.2s	0.5050	0.5050	0.5041
0.1s	0.5220	0.5220	0.5217
0.05s	0.5303	0.5303	0.5302
Instability	stable	0.6622s	stable

The results show that these two third-order algorithms present the same accuracy, but the cubic algorithm presents conditional numerical stability.

Van der Pol equation

The Van der Pol equation and its first time derivative are given by

$$\begin{aligned} &\ddot{y}-\mu\left(y_0^2-y^2\right)\dot{y}+\omega_0^2 y=0 \\ &\dddot{y}-\mu\ddot{y}\left(y_0^2-y^2\right)+2\mu y\dot{y}^2+\omega_0^2\dot{y}=0 \end{aligned} \tag{6}$$

where μ, y_0 and ω_0 are parameters of the equation. Table 2 compares the first displacement peak results for three practical time-steps and the instable time step Δt limit for $\mu=1.5$, $y_0=1$ and $\omega_0=1$.

Table 2. First peak displacement and instability limit

Δt	LAIER[1]	ARGYRES[3]	NEWMARK
0.2s	-0.3193	-0.3193	-0.3127
0.1s	-0.3193	-0.3193	-0.3177
0.05s	-0.3199	-0.3193	-0.3189
Instability	3.773s	11.83s	stable

The results show that also these two third-order algorithms present the same accuracy, but just Newmark method presents unconditional numerical stability.

Nonlinear pendulum

Figure 1 depicts the nonlinear pendulum that has been extensively analyzed by Argyris and Mlejek [2]. The equation of motion and its first derivative are written as:

$$\begin{aligned} &m\ddot{y}+c\dot{y}+2ky+\left(2N_0-2ak\right)\left(a^2+y^2\right)^{-0.5}y=f(t) \\ &m\dddot{y}+c\ddot{y}+2k\dot{y}+\left(2N_0-2ak\right)\left(a^2+y^2\right)^{-0.5}\dot{y}-2\left(2N_0-2ak\right)\left(a^2+y^2\right)^{-1.5}y^2\dot{y}=\dot{f}(t) \end{aligned} \tag{7}$$

where m is the mass of the pendulum, c is the damping, k is the stiffness, N_0 is the pre-tension force of the string and $f(t)$ is the excitation force. Table 3 compares the first displacement peak results for three practical time-steps and the instable time step Δt limit for $m=500\text{Kg}$, $N_0=500\text{N}$, $a=1\text{m}$, $k=10^7\,\text{N/m}$ and $f(t)=50\left(1-\cos(23.73t)\right)$.

Table 3. First peak displacement and instability limit

Δt	LAIER[1]	ARGYRES[3]	NEWMARK
0.022648s	0.024645	0.024645	0.024624
0.0052958s	0.024656	0.024656	0.024656
0.0022648s	0.24656	0.024656	0.024656
Instability	stable	2.2648s	stable

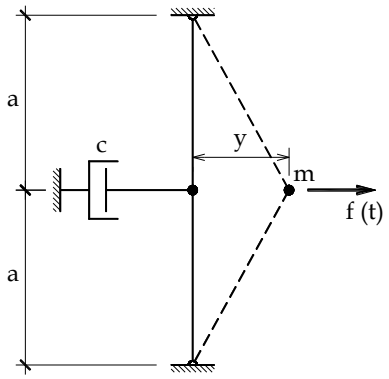


Figure 1. Nonlinear pendulum

The results shown in Table 3 indicate that the considered two third-order algorithms present the same accuracy, but the Newmark method and the algorithm developed by the author [1] present unconditional numerical stability.

Lotka-Volterra equation

The predator-prey Lotka-Volterra equation and its second and third time derivatives are given by

$$\begin{aligned}
\dot{x} &= kx - axy \\
\dot{y} &= -ly + bxy \\
\ddot{x} &= k\dot{x} - a\dot{x}y - ax\dot{y} \\
\ddot{y} &= -l\dot{y} + b\dot{x}y + bx\dot{y} \\
\dddot{x} &= k\ddot{x} - a\ddot{x}y - 2a\dot{x}\dot{y} - ax\ddot{y} \\
\dddot{y} &= -l\ddot{y} + b\ddot{x}y + 2b\dot{x}\dot{y} + bx\ddot{y}
\end{aligned} \tag{8}$$

where k, a, l and b are positive constant. As the Lotka-Volterra is of first-order just the first Hermitian operator given by equation (1) is involved. Table 4 compares the first displacement peak results for three practical time-steps and the instable time step Δt limit for k=a=l=b=1.

Table 4. First minimum peak for x function and instability limit

Δt	LAIER[1]	ARGYRES[3]	NEWMARK
0.01s	$0.560288\ 10^{-6}$	$0.560287\ 10^{-6}$	$0.549775\ 10^{-6}$
0.001s	$0.560280\ 10^{-6}$	$0.560280\ 10^{-6}$	$0.560174\ 10^{-6}$
0.0001s	$0.560280\ 10^{-6}$	$0.560280\ 10^{-6}$	$0.560279\ 10^{-6}$
Instability	0.157s	0.0952s	0.119s

The results shown in Table 4 indicate that the two third-order algorithms present again the same accuracy, but these three algorithms are not unconditional stable.

Lorenz equation

The atmospheric convection Lorenz model is governed by the equation

$$\begin{aligned} \dot{x} &= -\sigma(x - y) \\ \dot{y} &= rx - y - xz \\ \dot{z} &= xy - bz \end{aligned} \tag{9}$$

where σ, r and b are constant. The second and third derivatives of equation (9) are given by

$$\begin{aligned} \ddot{x} &= -\sigma(\dot{x} - \dot{y}) \\ \ddot{y} &= r\dot{x} - \dot{y} - \dot{x}z - x\dot{z} \\ \ddot{z} &= \dot{x}y + x\dot{y} - b\dot{z} \\ \dddot{x} &= -\sigma(\ddot{x} - \ddot{y}) \\ \dddot{y} &= r\ddot{x} - \ddot{y} - \ddot{x}z - 2\dot{x}\dot{z} - x\ddot{z} \\ \dddot{z} &= \ddot{x}y + 2\dot{x}\dot{y} + x\ddot{y} - b\ddot{z} \end{aligned} \tag{10}$$

As the Lorenz is of first-order just the first Hermitian operator given by equation (1) is involved Table 5 compares the first displacement peak results for three practical time-steps and the instable time step Δt limit for $\sigma = 10.0$, $r = 28.0$ and $b = 8/3$.

Table 5. First minimum peak for x function and instability limit

Δt	LAIER[1]	ARGYRES[3]	NEWMARK
0.01s	$0.203652\ 10^2$	$0.198015\ 10^2$	$0.135838\ 10^2$
0.001s	$0.200112\ 10^2$	$0.198099\ 10^2$	$0.200108\ 10^2$
0.0001s	$0.199781.10^2$	$0.198100\ 10^2$	$0.199781\ 10^2$
Instability	stable	0.00510s	0.0976s

The results presented in Table 5 show that the two third-order algorithms present quite similar accuracy, but in this case just the algorithm developed by the author is unconditionally stable.

Conclusions

The numerical applications show that the third-order algorithm developed by the author [1] and the cubic Hermitian developed by Argyris and Mlejek present as expected quite similar accuracy for refined mesh and little discrepancy for coarse mesh. The Newmark method also presents similar accuracy for refined mesh, but the discrepancy of the accuracy increase for coarse mesh. The time integration algorithm developed by the author is conditionally stable for Van der Pol and Lotka-Volterra equations. On the other hand, the Newmark method is conditionally stable for Lotka-Volterra and Lorenz equation. Finally, one has to note that the cubic Hermitian is conditionally stable for these five equations.

Acknowledgement

The author acknowledges the support of this work by the São Paulo Research Foundation (FAPESP), under grant#2011/15731-5.

References

[1] Laier, J. E. (2011) Spectral analysis of high-order Hermitian algorithm for structural dynamics , *Applied Mathematical Modelling* **35**, 965–971.
[2] Argyris, J. and Mlejek, H. P. (1991) *Dynamics of Structures in Text on Computational Mechanics* North-Holland, Amsterdam.
[3] Collatz, L. (1966) *The Numerical Treatment of Differential Equations*, 2nd edn, Springer´Verlag.Ruge, P. (2001) Restrict Padè scheme in computational structural dynamics, *Computer & Structures* **79**, 1913-1921.
[4] Ruge, P., (2001) Restrict Padè scheme in computational structural dynamic, *Computer & Structures* **79**, 1913-1921

The implementation of multi-block lattice Boltzmann method on GPU

***Ya Zhang[1], †Guang Pan[1], and Qiaogao Huang[1]**

[1] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China.

*Presenting author: zhangya9741@163.com

†Corresponding author: panguang601@163.com

Abstract

A straightforward implementation of multi-block lattice Boltzmann method (MB-LBM) on a graphical processing unit (GPU) is presented to accelerate simulations of complex fluid flows. The characteristics of MB-LBM algorithm are analyzed in detail. The algorithm is tested in terms of accuracy and computational time with the benchmark cases of lid driven cavity flow and the flow past a circular cylinder, and satisfactory results are obtained. The results show the performance on GPU is consistently better than that on CPU, and the greater the amount of data, the larger the acceleration ratio. Moreover, the arrangement of computational domain has significant effects on the performance of GPU. These results demonstrate the great potential of GPU on MB-LBM, especially for the calculation with large amounts of data.

Keywords: Multi-block, Lattice Boltzmann method, Graphical processing unit, Ratio of acceleration.

Introduction

During recent decades, the lattice Boltzmann method (LBM) has developed into an alternative method for simulating complex fluid flow [1]. LBM is based on the statistical physics and originally came from the Boltzmann equation. A direct connection between the lattice Boltzmann equation and Navier-Stokes equations has been established under the nearly incompressible condition [2]. The fact that LBM evolves rather locally makes it more suitable for parallel computing compared to the conventional computation method.

Graphical processing unit (GPU) is designed to process large graphics data sets for rendering tasks, so it has exceeded the computation speed of PC-based central processing unit (CPU) by more than one order of magnitude while being available for a comparable price. Another advantage for GPU application is that Compute Unified Device Architecture (CUDA) provided by NVIDIA, a standard C language extension for parallel application development on a GPU, reduces the development threshold of GPU programming greatly. Due to the inherent parallelism of LBM, a significant speedup of GPU-based computation on LBM has been reported in different areas. Fan et al. [3] implemented the LBM simulations on a cluster of GPUs with message passing interface (MPI). Tolke and Krafczyk [4] implemented a three-dimensional LBM and achieved near teraflop computing on a single workstation. Zhou et al. [5] provided an efficient GPU implementation of flows with curved boundaries, leading to nearly an 18-fold speed increase. Tubbs et al. [6] implemented LBM for solving the shallow water equations and the advection dispersion equation on GPU-based architectures, and the results indicate the promise of the GPU-accelerated LBM for modeling mass transport phenomena in shallow water flows. GPU has tremendous potential to accelerate LBM computation owing to the parallel nature of LBM.

The traditional LBM is often employed on uniform grids, which makes the evolution explicit and the algorithm simple, but at the same time could increase the computational effort dramatically on the road to high resolution. To solve this problem, a multi-block lattice Boltzmann method (MB-LBM) is designed and applied over the flow area where relatively high resolution is needed. As a useful tool of grid refinements in LBM, the multi-block technique has been investigated in recent years. In 1998, Filippova et al. [7] introduced a local second order refinement scheme and provided the theoretical foundation for multi-block techniques. In 2000, Lin and Lai [8] designed a composite block-structured scheme by placing the fine grid blocks on needed area for the mesh refinement. In 2002, Yu et al. [9] proposed a multi-block scheme, where the fine block is partially overlapped at the interfacial lattices, increasing the model efficiency greatly. The model has been successfully applied to various areas. Yu and Girimaji [10] extended this model to 3D turbulence simulations. Y. Peng et al. [11] applied it in the immersed boundary lattice Boltzmann method with multi-relaxation-time collision scheme. Liu et al. [12] validated the multi-block lattice Boltzmann model coupled with the large eddy simulation model in transient shallow water flows simulation. Farhat et al. [13][14] extended the single phase MB-LBM to the multiphase Gunstensen model, in which the grid was free to migrate with the suspended phase, and validated a 3D migrating multi-block model. Following from this, the present study aims to develop an efficient and straightforward algorithm for the GPU implementation of MB-LBM, and test it in terms of accuracy and computational time.

Multi-block lattice Boltzmann method

In the present study, the BGK lattice Boltzmann method is used with a two-dimensional nine-velocity (D2Q9) discrete velocity model [2], as shown in Fig. 1. The lattice Boltzmann method formulates as the following evolution equation:

$$f_\alpha(\boldsymbol{x}+\boldsymbol{e}_\alpha\delta t, t+\delta t) = f_\alpha(\boldsymbol{x},t) - \frac{1}{\tau}\left[f_\alpha(\boldsymbol{x},t) - f_\alpha^{eq}(\boldsymbol{x},t)\right] \tag{1}$$

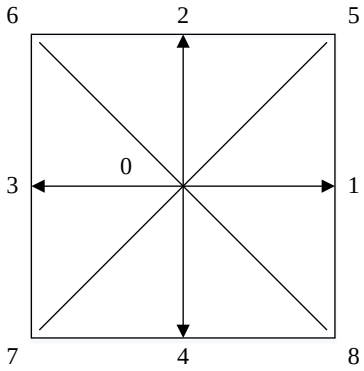


Figure 1. Lattice pattern: D2Q9

where f_α is the particle distribution functions representing the probability of particles at position $\boldsymbol{x}$ and discrete velocity $\boldsymbol{e}_\alpha$ at time t; δt is the time step; τ is the single-relaxation-time,

depending on the kinematic viscosity ν, $\tau = 3\nu + 0.5$; e_α is the α th discrete velocity, the discrete velocity model is

$$\boldsymbol{e} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 & 1 & -1 & -1 \end{bmatrix} \tag{2}$$

f_α^{eq}, the approximate of the Maxwell-Boltzmann equilibrium distribution function at low numbers, is expressed as follow:

$$f_\alpha^{eq} = \rho w_i \left[1 + \frac{\boldsymbol{e}_\alpha \cdot \boldsymbol{u}}{c_s^{\,2}} + \frac{(\boldsymbol{e}_\alpha \cdot \boldsymbol{u})^2}{2c_s^{\,4}} - \frac{u^2}{2c_s^{\,2}} \right] \tag{3}$$

where w_α is the weighting coefficient, valued by $w_0 = 4/9$, $w_1 = w_2 = w_3 = w_4 = 1/9$ and $w_5 = w_6 = w_7 = w_8 = 1/36$; the sound speed is $c_s = 1/\sqrt{3}$; ρ and $\boldsymbol{u}$ are the macroscopic density and velocity, which can be calculated from the distribution function respectively by:

$$\rho = \sum_{\alpha=0}^{8} f_\alpha = \sum_{\alpha=0}^{8} f_\alpha^{eq} \tag{4-1}$$

$$\rho \boldsymbol{u} = \sum_{\alpha=0}^{8} \boldsymbol{e}_\alpha f_\alpha = \sum_{\alpha=0}^{8} \boldsymbol{e}_\alpha f_\alpha^{eq} \tag{4-2}$$

This paper uses the multi-block method proposed by Yu et al. [9], which satisfies the continuity of mass, momentum and stresses across the interface. To illustrate the basic idea, a two-block system consisting of a coarse block and a fine block is shown in Fig. 2.

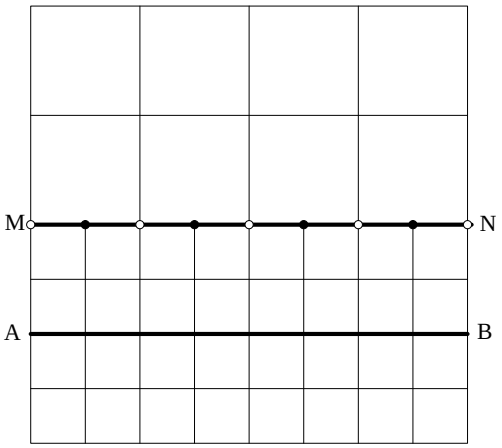


Figure 2. Interfaces structure between two blocks

The ratio of the lattice space between coarse blocks and fine blocks *m* is defined as:

$$m = \delta x_c / \delta x_f = m_c m_f \tag{5}$$

where the subscript *c* refers to the coarse block while *f* refers to the fine block, δx_c and δx_f are the lattice space, $m_c = 1$ and $m_f = \delta x_c / \delta x_f$ are the lattice space parameters. To maintain a consistent viscosity across blocks, the relaxation time τ_f on the fine block and τ_c on the coarse block have to satisfy the following equation:

$$\tau_f = 0.5 + m_f(\tau_c - 0.5) \tag{6}$$

The transfer of the post-collision distribution functions between different blocks happens after the collision step. Since each interface grid consists of overlapping two sets of coarse and fine nodes, the information of coarse boundary nodes can be obtained after m_f steps of evolution on the fine grid, where the post-collision distribution $\tilde{f}_\alpha^c$ for the coarse block is written as:

$$\tilde{f}_\alpha^c = f_\alpha^{eq,f} + m_f \frac{\tau_c - 1}{\tau_f - 1}(\tilde{f}_\alpha^f - f_\alpha^{eq,f}) \tag{7}$$

Similarly, when transferring the data from the coarse block to the fine block, one follows:

$$\tilde{f}_\alpha^f = f_\alpha^{eq,c} + \frac{\tau_f - 1}{m_f(\tau_c - 1)}(\tilde{f}_\alpha^c - f_\alpha^{eq,c}) \tag{8}$$

As shown in Fig. 2, the line MN is the fine block boundary, while the line AB is the coarse block boundary. The information on the nodes noted by solid symbol can be obtained through spatial interpolation based on the information at the open nodes on the line MN.

To eliminate the possibility of spatial asymmetry caused by interpolations, a symmetric cubic spline fitting is used to calculate the unknown nodes on the fine blocks [9], which is done by

$$\begin{aligned} &\tilde{f}(x) = a_i(x_i - x)^3 + b_i(x - x_{i-1})^3 + c_i(x_i - x) + d_i(x - x_{i-1}) \\ &x_{i-1} \le x \le x_{i+1} \end{aligned} \tag{9}$$

where according to the continuity of the nodal condition of $\tilde{f}$ and $\tilde{f}'$ (the first order derivation of $\tilde{f}$), and suitable end condition, the coefficients (a_i, b_i, c_i, d_i) in Eq. (9) are computed as follows:

$$
\begin{aligned}
a_i &= \frac{M_{i-1}}{6h_i} \\
b_i &= \frac{M_i}{6h_i} \\
c_i &= \frac{\tilde{f}_{i-1}}{h_i} - \frac{M_{i-1}h_i}{6} \\
d_i &= \frac{\tilde{f}_i}{h_i} - \frac{M_i h_i}{6}
\end{aligned}
\tag{10}
$$

where M_i is the second order derivatives of $\tilde{f}_i$, following the equation

$$0.5M_{i-1} + 2M_i + 0.5M_{i+1} = 3(2f_i - f_{i-1} - f_{i+1}) \tag{11}$$

The natural spline end condition is stipulated with $M_0 = M_n = 0$.

A three-point Lagrangian scheme is used in the temporal interpolation of the post-collision distribution function on the interface grid at the specific time:

$$\tilde{f}_i^{\,f}(t) = \sum_{k=-1}^{1} \tilde{f}_i^{\,f}(t_k) \left(\prod_{\substack{k'=-1 \\ k \neq k'}}^{1} \frac{t - t_{k'}}{t_k - t_{k'}} \right) \tag{12}$$

So the function for the *n*th evolution of the fine block is expressed as

$$\tilde{f}_i^{\,f}(t) = 0.5\frac{n}{m_f}\left(\frac{n}{m_f} - 1\right)\tilde{f}_i^{\,f}(t_{-1}) - \left(\frac{n}{m_f} - 1\right)\left(\frac{n}{m_f} + 1\right)\tilde{f}_i^{\,f}(t_0) + 0.5\frac{n}{m_f}\left(\frac{n}{m_f} + 1\right)\tilde{f}_i^{\,f}(t_1) \tag{13}$$

where the present time is $t = t_0 + \frac{n}{m_f}$.

The flow chart of the computational sequence for the MB-LBM in the two-block system is shown in Fig. 3.

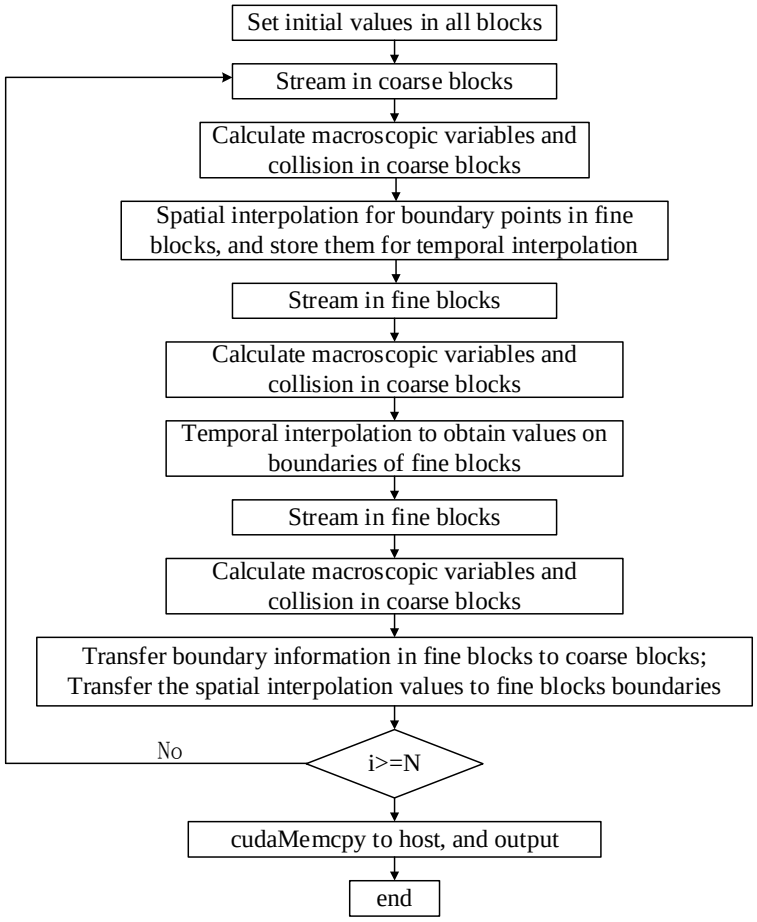


Figure 3. The flow chart of the computational sequence for MB-LBM

In this paper the momentum-exchange method [15] is used to calculate the force exerted onto the obstacle considering its simplicity. In order to differentiate the nodes in the computational domain, node type is employed to donate the fluid node, boundary node of computational domain, boundary node of blocks and solid node. If particles in the solid node $\boldsymbol{x}_b(i, j)$ of the fine block, will move to a fluid node along the direction $\boldsymbol{e}_\alpha$ in the next step, values (i, j, α) should be stored in an array. The force can be calculated by

$$\boldsymbol{F} = \frac{1}{m} \sum_{All\ (i,j,\alpha)} \left[\boldsymbol{e}_\alpha f_\alpha(\boldsymbol{x}_b) - \boldsymbol{e}_{\bar{\alpha}} f_{\bar{\alpha}}(\boldsymbol{x}_b + \boldsymbol{e}_\alpha \delta t) \right] \tag{14}$$

GPU implementation

A graphical processing unit (GPU) is specifically designed to process large graphics data sets for rendering tasks. As GPU has a number of processing cores, so besides graphic rendering tasks, it also is used to implement other parallel computing tasks. In this work, the simulation is carried out on a CPU platform of Intel Xeon(R) W3550, 3.07GHz) with RAM of 24.0 GB and a NVIDIA GPUs device (Geforce 980ti), programming using CUDA (Compute Unified Device Architecture).

In the CUDA programing architecture, CPU is considered to be the host, while GPU is considered

to be the device. The code is split up into a CPU and GPU part, the latter is called kernel, compiled by NVIDIA C-Compiler (NVCC). When a kernel function is launched with required parameters, the number of blocks and the number of threads in each block (256 in this paper), it is executed by these threads on a device. In one block, each thread is indexed by a thread identification. Threads from different blocks cannot communicate, while threads from the same block are independent, but can communicate via shared memory and have synchronize execution. A kernel is executed in a grid of thread blocks indexed by a block identification. The grid terminates when all threads of a kernel complete their execution, and the execution continues on the host until another kernel is launched.

The memory access of the kernel has a great influence on the implementation performance. The registers are trace buffer on GPU, and can be accessed with nearly no time delay, but is rather small, so excessive local variables used in kernel should be avoided. The global memory is a device memory and is the largest memory device in GPU, but not as fast as the registers. In this work, each node requires nearly 200 bytes of memory for double precision computation, so most of the data will be stored in the global memory. Besides, there is a share memory for each multiprocessor, allowing communication between threads, and can be accessed as fast as the registers. The constant memory, which can also be fast accessed, is used to store the constants that are read only and are accessed frequently.

The LBM code is highly parallelizable since it can be separated into two main steps, streaming and collision [2]. In the collision step, the distribution functions of a certain node will not exchange with its neighbor, and the post-collision function is given by

$$\tilde{f}_\alpha(\boldsymbol{x},t) = f_\alpha(\boldsymbol{x},t) - \frac{1}{\tau}\left[f_\alpha(\boldsymbol{x},t) - f_\alpha^{eq}(\boldsymbol{x},t) \right] \tag{15}$$

The streaming step is related to the distribution functions of the surrounding nodes according to Eq. (1) and Eq. (15). Considering the fact that misaligned read is faster than misaligned write[16], the streaming is carried out with the following equation

$$f_\alpha(\boldsymbol{x},t+\delta t) = \tilde{f}_\alpha(\boldsymbol{x}-\boldsymbol{e}_\alpha\delta t,t) \tag{16}$$

To increases the efficiency of data communication, the collision and the streaming step are combined into one kernel to avoid repeated access of global memory for distribution functions.

For systems containing multi-level blocks, according to the flow chart in Fig. 3, the computation can be expressed with a recursive function shown in Fig. 4.

```
void evolution(int level)
{
	for (int i=0; i<m[level]; i++)
	{
		if (level == LEVEL)
			return;
		if (!(level == 0 || i == 0))
		{
			//temporal interpolation
		}
		//stream, calculate macroscopic variables and collision
		//information exchange between the present level blocks
		if (level != LEVEL-1)
		{
			//spatial interpolation to prepare for blocks, level+1
		}
		evolution(level+1);
		if (level != 0 && i == m[level]-1)
		{
			//Transfer boundary information in blocks level+1 to level;
			//Transfer the spatial interpolation values to level+1
		}
	}
}
```

Figure 4. Program of the recursive function for MB-LBM

Since there are always the same data types of variables needed to be record in each node, a struct body, including pointers to node type, position, density, velocity, distribution functions and post-collision distribution functions, is created to store variable information. With these pointers, memory in host and device is allocated dynamics for the variables.

In the stage of the spatial interpolation, it is needed to obtain M_i in Eq. (10) and Eq. (11). In serial processing, the tridiagonal matrix in Eq. (11) is solved with the Thomas algorithm, which is almost unfeasible in parallel algorithm. The cuSPARSE library presented by NVIDIA contains a set of basic linear algebra subroutines used for handling sparse matrices in parallel mode. The function cusparseDgtsv() is employed in this paper. It can be used by cusparseDgtsv(cusparseHandle_t *handle*, int *m*, int *n*, const double **dl*, double **d*, double **du*, double **B*, int *ldb*), where *handle* is the handle to the cuSPARSE library context; *m* is the size of the linear system (must be larger than or equal to three); *n* is the columns of matrix *B*, which means M_i for different variables can be solve in a single call; array *dl*, *d*, *du* contain the lower, the main, the upper diagonal of the tridiagonal linear system, respectively; *B* is the right-hand-side array, *ldb* is the leading dimension of *B*. The solution will be written in array *B* before the function completes.

It is obvious that the spatial interpolation in parallel is much more complex than the temporal interpolation, so it is suggested that the largest ratio of the lattice space between adjacent levels should be placed on the finest level. And in this work, the arrangement of levels is expressed in form of m_1-…-m_i-…-m_n in coarse-fine order, where m_1 is always 1, m_i is the ratio of the lattice space of level *i* to that of level *i*-1. So as mentioned, the arrangement of levels 1-2-3 is better than 1-3-2.

In this work, all the procedures but output are completed on the GPU directly to eliminate the unnecessary copy between host and device. At the same time due to the fact that the atom function atomicAdd() in the CUDA toolkit provided by NVIDIA can only be used for Integer and Long, the parallel reduction is used to calculate the force in Eq. (14) after loading the position and direction.

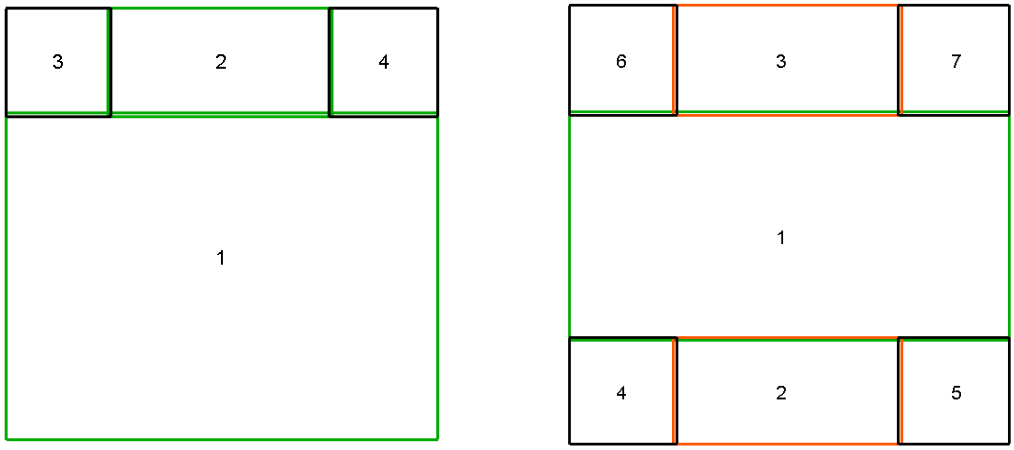
Presentation of test cases and discussion

Lid driven cavity flow

The lid driven cavity flow has been extensively used as a benchmark problem to test the accuracy of a numerical method. The computations are carried out using the multi-block computational domains, whose schematic diagrams are shown in Fig. 5.

In all the arrangements, the finest blocks are placed on the areas of singularity points or changing sharply. As shown in Fig. 5, the finest blocks is located in the two upper corner regions. In Fig. 5(a), there are two levels of blocks and four separate blocks in the calculation. Block 1 and block 2 belong to the first level; block 3 and block 4 belong to the second level; the diagram in Fig. 5(b) contains three levels and seven blocks, while block 1 belong to the first level, block 2 and block 3 belong to the second level, and block 4 to block 7 belong to the third level.

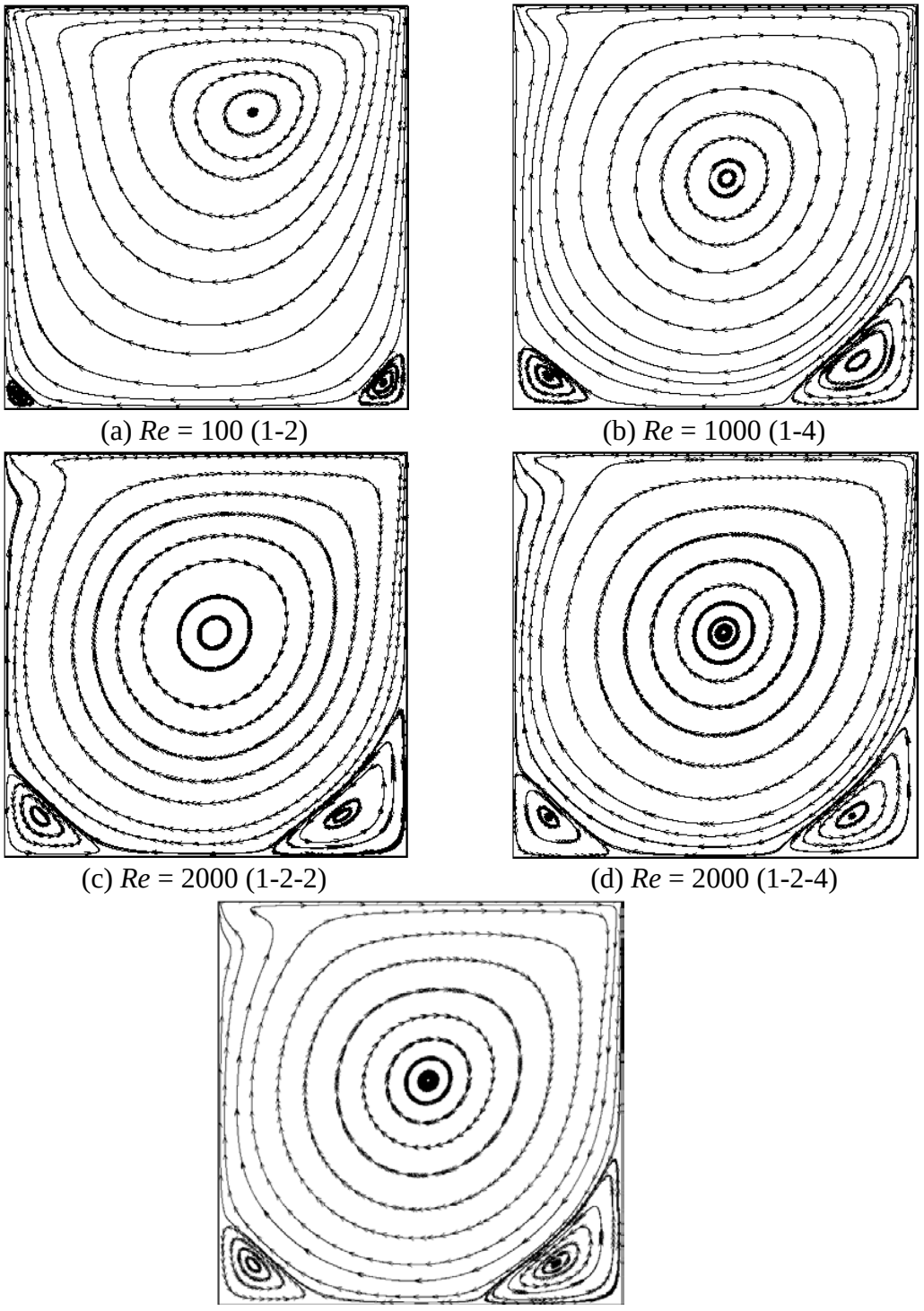
The simulation region is 128-128. The initial condition for density is unity and that for velocity is zero. The upper wall velocity is U = 0.1. All the boundaries uses the moving boundary half-way bounce-back scheme.



(a) (b)

Figure 5. Arrangements of blocks for the lid driven cavity flow

To assess the results, the solutions of Ref. [17] and Ref. [18] are used for comparison. The dimensionless locations of the centers of the primary vortex, the lower left vortex and the lower right vortex of present work and of previous literatures are listed in Table 1. As shown in Table 1, all the results show a good agreement with previous researches. And for Re = 2000, different arrangements of blocks appear identical results.



(a) *Re* = 100 (1-2) (b) *Re* = 1000 (1-4)

(c) *Re* = 2000 (1-2-2) (d) *Re* = 2000 (1-2-4)

(e) *Re* = 2000 (1-4-2)

Figure 6. Streamlines for the lid driven flow

Table 1 Comparison of the vortex centers with previous literatures [17][18]

Re	Arrangement	Primary vortex	Lower left vortex	Lower right vortex
100				
Present	1-2 (Fig. 5(a))	(0.6142, 0.7402)	(0.0354, 0.0394)	(0.9370, 0.0669)
Ref. [17]		(0.6172, 0.7344)	(0.0313, 0.0391)	(0.9453, 0.0625)

1000				
Present	1-4 (Fig. 5(a))	(0.5276, 0.5669)	(0.0866, 0.0787)	(0.8504, 0.1181)
Ref. [17]		(0.5313, 0.5625)	(0.0859, 0.0781)	(0.8594, 0.1094)
2000				
Present	1-2-2 (Fig. 5(b))	(0.5238, 0.5555)	(0.0873, 0.1032)	(0.8413, 0.0992)
	1-2-4 (Fig. 5(b))	(0.5238, 0.5555)	(0.0873, 0.1032)	(0.8413, 0.0992)
	1-4-2 (Fig. 5(b))	(0.5238, 0.5555)	(0.0873, 0.1032)	(0.8413, 0.0992)
Ref. [18]		(0.5250, 0.5500)	(0.0875, 0.1063)	(0.8375, 0.0938)

Flow past a circular cylinder

A flow past a circular cylinder is simulated to implement the parallel algorithm in simulation domain that has more levels and blocks.

The arrangement of the computational domain is shown in Fig. 7. There are four levels of blocks in the simulation. Block 1 to block 4 belong to the first level; block 5 and block 6 belong to level two; block 7 to block 9 belong to level 3; block 10 belong to level 4, the finest level. The ratio of the lattice space between adjacent levels is 1-2-2-2.

In this calculation, the cylinder diameter D is set to 6. The length of the simulation region is 320, and the width is 128. The center of the cylinder is at (64, 64), which makes it located in the finest block, as shown in Fig. 7. The slip boundary scheme is implemented on the top and bottom boundaries. The standard bounce back scheme is used on the cylinder surface. The velocity and the pressure scheme of Zou and He are applied on the inlet and the outlet boundaries, respectively, where the far field velocity is U_0=0.1 and the initial density is unity. The relaxation time for the first level grid is computed by Re=100, based on the far field velocity and the diameter of the cylinder.

Drag coefficient, lift coefficient and Strouhal number are the benchmark dimensionless numbers for the flow past a circular cylinder. The drag and the lift coefficients are calculated using the following formulae, $C_D = \frac{2F_D}{\rho U^2 D}$ and $C_L = \frac{2F_L}{\rho U^2 D}$, and the Strouhal number is defined as $St = \frac{aD}{U}$, where F_L the lift force, F_D the drag force, D the cylinder diameter, a the frequency of vortex-shedding, obtained by processing F_L with Fast Fourier Transform.

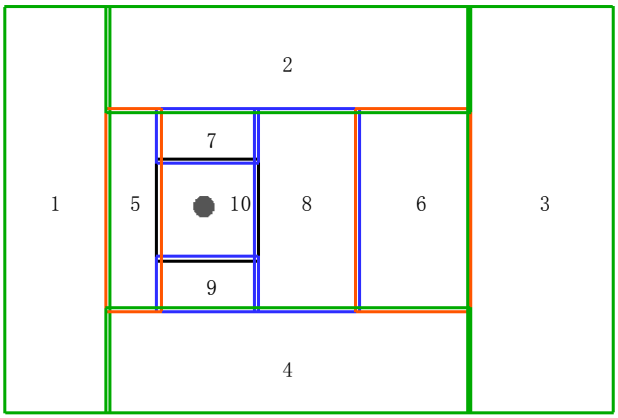


Figure 7. Arrangement of blocks for the flow past a circular cylinder

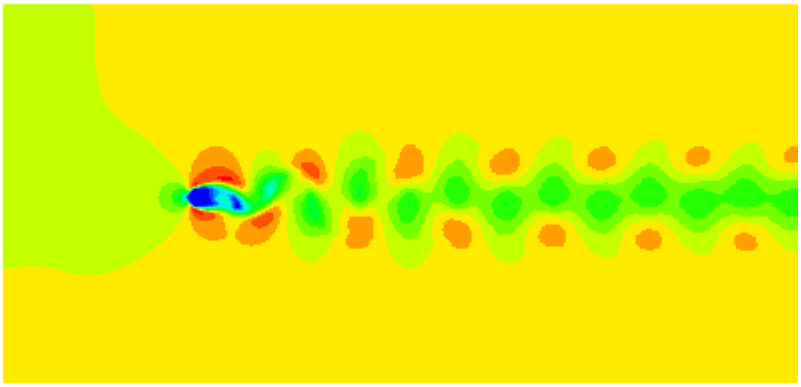


Figure 8. Velocity contour for the flow past a circular cylinder

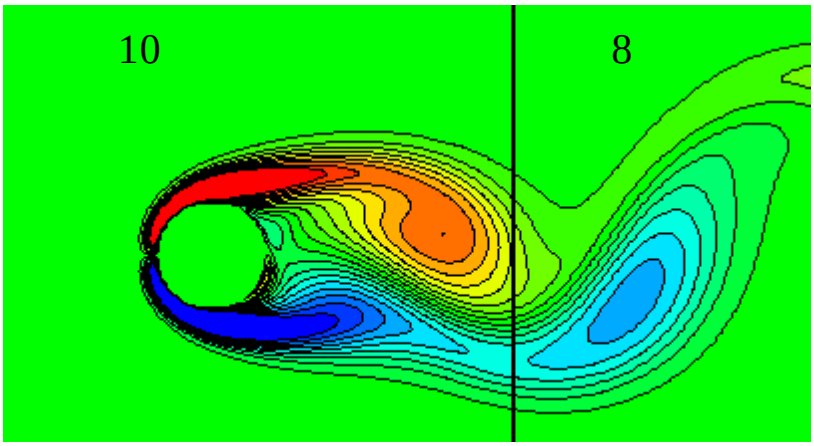


Figure 9. Vorticity contour for the flow past a circular cylinder

Table 3 Comparison of results at Re = 100 with previous literatures [19][20][21]

Author	C_D	C_L	S_t
Silva [19]	1.39	-	0.16
Zhou [20]	1.428	0.315	0.172
Xu [21]	1.423	0.34	0.171
This work	1.381	0.304	0.168

The velocity contour for the flow past a circular cylinder is shown in Fig. 8. The instantaneous vorticity contours of vortex shedding are plotted in Fig. 9. It can be seen clearly that the vorticity is rather smooth across the block interface. This shows that the implementation of multi-block scheme functions well for unsteady flow. Table 3 shows our numerical results compare well with the previous results, despite little differences.

Assess the performance of MB-LBM code on GPU

The parameters of performance of MB-LBM on CPU and on GPU is shown in Table 2, including the time spending for evolution of 10^4 steps (in second), the number of lattice updates per step in an arrangement (LUPS), million lattice updates per second (MLUPS), and the acceleration ratio of GPU to CPU. In general, LUPS represents the amount of data, and a large MLUPS means a high data processing speed.

Table 2 Performance of CPU and GPU for 104 steps

Case	Arrangement	LUPS	CPU		GPU		Acceleration
			Time	MLUPS	Time	MLUPS	ratio
1	1-2 (Fig. 5(a))	31267	205.66	1.52	64.52	4.85	3.19
2	1-4 (Fig. 5(a))	145691	1083.28	1.34	86.63	16.82	12.50
3	1-2-2 (Fig. 5(b))	300688	1894.23	1.59	245.59	12.24	7.71
4	1-2-4 (Fig. 5(b))	2090912	19543.09	1.07	498.00	42.00	39.24
5	1-4-2 (Fig. 5(b))	2326104	21736.04	1.07	659.12	35.29	33.00
6	1-2-2-2 (Fig. 7)	790392	5835.10	1.35	578.71	13.66	10.08

It can be seen from Table 2 that the ratio of acceleration is not a constant, and performance on GPU is always better than that of CPU. To be specifically, as the amount of data increases, roughly the speedup is more obvious. Besides, the arrangement of computational domain has great impact on the performance of GPU. In case 2 and case 3, the resolution of upper corners is the same, but on GPU the performance of case 2 is much better while with a smaller LUPS, so it is not recommended to employ more levels for the same resolution. In addition, according to the performance of case 4 and case 5, considering the time consumed by spatial interpolation in MB-LBM, it is verified that the largest ratio of the lattice space between adjacent levels should be placed on the finest level.

Conclusion

In this paper, a straightforward multi-block LBM parallel algorithm based on a single GPU has been presented. The characteristics of MB-LBM algorithm are analyzed in detail. The benchmark cases of the lid driven cavity flow and the flow past a circular cylinder are investigated as the test cases for the GPU-based implementation, and satisfactory results are obtained. Performance on GPU is always better than that of CPU, and the greater the amount of data, the larger the acceleration ratio. And arrangement of computational domain has significant effects on the performance. The largest acceleration ratio 39.24 are achieved by now, however that still leaves room for a large rise in computation with large amounts of data.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (11502210, 51279165).

References

[1] Aidun, C. K. and Clausen, J. R. (2010) Lattice-Boltzmann method for complex flows, *Annual review of fluid mechanics* **42**, 439-472.
[2] Mohamad, A. A. (2011) Lattice Boltzmann method: fundamentals and engineering applications with computer codes, Springer-Verlag, London, UK.
[3] Fan, Z., Qiu, F., Kaufman, A. and Yoakum-Stover, S. (2004) GPU cluster for high performance computing, *Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, **47**.
[4] Tölke, J. and Krafczyk, M. (2008) TeraFLOP computing on a desktop PC with GPUs for 3D CFD, *International Journal of Computational Fluid Dynamics* **22**, 443-456.
[5] Zhou, H., Mo, G., Wu, F., Zhao, J., Rui, M. and Cen, K. (2012) GPU implementation of lattice Boltzmann method for flows with curved boundaries, *Computer Methods in Applied Mechanics and Engineering* **225**, 65-73.
[6] Tubbs, K. R. and Tsai, F. T. C. (2011) GPU accelerated lattice Boltzmann model for shallow water flow and mass transport, *International Journal for Numerical Methods in Engineering* **86**, 316-334.
[7] Filippova, O. and Hänel, D. (1998) Grid refinement for lattice-BGK models, *Journal of Computational Physics* **147**, 219-228.
[8] Lin, C. L. and Lai, Y. G. (2000) Lattice Boltzmann method on composite grids, *Physical Review E* **62**, 2219-2225.
[9] Yu, D., Mei, R. and Shyy, W. (2002) A multi‐block lattice Boltzmann method for viscous fluid flows, *International journal for numerical methods in fluids* **39**, 99-120.
[10] Yu, D. and Girimaji, S. S. (2006) Multi-block lattice Boltzmann method: extension to 3D and validation in turbulence, *Physica A: Statistical Mechanics and its Applications* **362**, 118-124.
[11] Peng, Y., Shu, C., Chew, Y. T. Niu, X. D. and Lu, X. Y. (2006) Application of multi-block approach in the immersed boundary–lattice Boltzmann method for viscous fluid flows, *Journal of Computational Physics* **218**, 460-478.
[12] Liu, H., Zhou, J. G. and Burrows, R. (2010) Lattice Boltzmann simulations of the transient shallow water flows, *Advances in Water Resources* **33**, 387-396.
[13] Farhat, H. and Lee, J. S. Fundamentals of migrating multi-block lattice Boltzmann model for immiscible mixtures in 2D geometries, *International Journal of Multiphase Flow* **36**, 769-779.
[14] Farhat, H., Choi, W. and Lee, J. S. (2010) Migrating multi-block lattice Boltzmann model for immiscible mixtures: 3D algorithm development and validation, *Computers & Fluids* **39**, 1284-1295.
[15] Mei, R., Shyy, W. and Yu, D. and Luo, S. L. (1999) Force Evaluation in the Lattice Boltzmann Method, *APS Division of Fluid Dynamics Meeting Abstracts* **1**.
[16] Obrecht, C., Kuznik, F., Tourancheau, B. and Roux, J-J. (2011) A new approach to the lattice Boltzmann method for graphics processing units, *Computers and Mathematics with Applications* **61**, 3628-3638.
[17] Ghia, U., Ghia, K. N. and Shin, C. T. (1982) High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method, *Journal of computational physics* **48**, 387-411.
[18] Vanka, S. P. (1986) Block-implicit multigrid solution of Navier-Stokes equations in primitive variables, *Journal of Computational Physics* **65**, 138-158.
[19] Silva, A. L. E., Silveira-Neto, A. and Damasceno, J. J. R. (2003) Numerical simulation of two-dimensional flows over a circular cylinder using the immersed boundary method, *Journal of Computational Physics* **189**, 351-370.
[20] Zhou, H., Mo, G., Wu, F., Zhao, J., Rui, M. and Cen, K. GPU implementation of lattice Boltzmann method for flows with curved boundaries, *Computer Methods in Applied Mechanics and Engineering* **225**, 65-73.
[21] Xu, S. and Wang, Z. J. (2006) An immersed interface method for simulating the interaction of a fluid with moving boundaries, *Journal of Computational Physics* **216**, 454-493.

Analyzing and predicting the criteria pollutants over a tropical urban area by using statistical models

***S. Dey[1], P. Sibanda[1], S. Gupta[2] and A. Chakraborty[3]**

[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01 Scottsville 3209, Pietermaritzburg, South Africa

[2]Department of Environmental Science, The University of Burdwan, Golapbag, Burdwan 713104, West Bengal, India

[3]Center for Rivers, Oceans, Atmosphere and Land Sciences (CORAL), Indian Institute of Technology, Kharagpur,
Kharagpur -721302, West Bengal, India

*Corresponding author: sharadiadey1985@gmail.com

Abstract

Modeling and prediction of criteria pollutants over the urban areas is essential for the formulation and improvisation of urban air quality management strategies. Various statistical techniques have been employed worldwide for accurate prediction of the air pollutants. This study focuses on the analysis and prediction of the criteria pollutants over a tropical urban area (Durgapur, 23° 30′ 34.58″ N and 87° 21′ 03.42″ E) performed by using statistical models *viz.* multiple linear regression (MLR) and principal component regression (PCR). Multiple linear regression analyses have been performed using the original variables and principal components (PCs) as the inputs. On the basis of the performance indicators, MLR model is found to perform better than the PCR in most cases. The R^2 values obtained by MLR are 0.962, 0.945, 0.898, 0.937, 0.603, 0.874, 0.871, 0.837, 0.858, 0.868, 0.842 and 0.825 for PM_{10}, $PM_{2.5}$, sulphur dioxide, nitrogen dioxide, carbon monoxide, ammonia, ozone, benzene, benz(a)pyrene, arsenic, lead and nickel respectively which are greater than the respective R^2 values obtained by PCR model. Results of the two models reveal that use of PCA could not enhance the MLR performance. The predictive equations proposed by the statistical models suggest that the meteorological parameters (temperature, relative humidity, wind speed and cloud cover) have significant influence on the concentration of the criteria pollutants.

Key words: PCR, MLR, Performance Indicators, criteria pollutants

1. Introduction

Escalating air pollution and deteriorating air quality status of urban areas is a matter of concern worldwide. In this era of rapid urbanization and industrialization, air pollutants containing toxic substances like particulate matters, heavy metals, polycyclic hydrocarbons (PAH), volatile organic compound (VOC) and other gaseous substances (like SO_2, NO_2, CO, NH_3, tropospheric O_3 etc.) have an increasing impact on urban air quality. Actually, air pollution risk is a function of the hazard of the pollutant and exposure to the pollutant. Carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter, and sulfur dioxide have identified as criteria pollutants by Clean Air Act (CAA) of 1970. Central Pollution Control Board (CPCB) has identified 12 health based parameters [namely particulate matters (PM_{10} & $PM_{2.5}$), benzene, benzo(a)pyrene, nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon monoxide (CO), ammonia (NH_3),ozone (O_3), lead (Pb), nickel (Ni) and arsenic (As)] for assessing the air quality status across the country in 2009 under the provision of Air (Prevention & Control of Pollution) Act, 1981.

The complexities and difficulties in continuous measurement of air pollutant concentrations have led to the development of modeling techniques which enable the researchers to predict the pollutant concentration with acceptable accuracy [1]. Accurate knowledge of pollutant sources, emission inventories and proper description of the physico - chemical processes are essential for minimizing biasness and errors of the outputs of the deterministic models. These are quick and easy empirical techniques for predicting the ambient air pollutant concentration as a function of several input parameters. In air quality modeling, one of the most common models available for predicting outdoor and indoor air pollutant concentrations are statistical regression methods [2]. Statistical models are suitable for the description of the complex site-specific relationship between air pollutants and explanatory variables, and they often make predictions with a higher accuracy than mechanistic models [3]. Multiple linear regression (MLR) is a widely used multivariate statistical technique for expressing the dependence of a response variable on several independent (predictor) variables. Awang *et al.* [4] compared the multivariate methods (MLR and PCR) for predicting the surface O_3 concentration during daytime, nighttime and critical conversion time in Shah Alam, Malaysia. The concentration PM_{10}, $PM_{2.5}$, CO and CO_2 concentrations and meteorological variables (wind speed, air temperature, and relative humidity) were employed by Elbayoumi *et al*. [5] for predicting the annual and seasonal indoor concentration of PM_{10} and $PM_{2.5}$ at Gaza Strip (Palestine) using multivariate statistical methods. Luvsan *et al.* [6] used multiple linear regression models for exploring the association of concentration of SO2 with temperature, relative humidity and wind speed in Mongolia. Sayegh et al.[7] employed several approaches including linear, non-linear, and machine learning methods are evaluated for the prediction of urban PM_{10} concentrations in the City of Makkah, Saudi Arabia.

In the present work, we predict the concentration of various criteria pollutants by using multiple linear regression (MLR) and principal component regression (PCR) models, the performance of both the statistical models is evaluated in terms of the performance indicators. Deterministic models require a large number of input data which are difficult to provide whereas statistical models are relatively simple and sufficiently reliable tools for predicting the concentration of different air pollutants. Moreover, application of multivariate statistical methods for the prediction of the air pollutants is a new piece of work over this eastern part of India.

2. Method

2.1 Description of the study area

Durgapur (chosen urban area) is situated in the Burdwan district of West Bengal, India. It is located on the bank of River Damodar. This area is covered with Red and Yellow Ultisols soil and the topography of this area is undulating, with an average elevation of 65 m MSL. This area experiences a transitional climate between the tropical wet and dry climate and the more humid subtropical climate.

2.2 Data used

The data of concentration of all the criteria pollutants such as ammonia, arsenic, benzene, benzo(α)pyrene, carbon monoxide, lead, nickel, nitrogen dioxide, ozone, sulphur dioxide, PM_{10} and $PM_{2.5}$ at Bidhannagar, India (23° 30′ 34.58″ N and 87° 21′ 03.42″ E) were collected for the duration of June, 2013 to May, 2015 from the archived data set of WBPCB (Bidhannagar unit of Durgapur). These parameters are monitored twice a week at this location by WBPCB [www.wbpcb.gov.in]. The data of meteorological parameters [Temperature (T), relative humidity (RH), wind speed (WS) and cloud cover (CC) are collected from the NOAA Air Resources Laboratory (ARL) website. (http://ready.arl.noaa.gov/READYamet.php).

The air pollutants and the meteorological parameters data were divided into two sets: model development set and the model validation set. The model development set comprises of the 24 average values of criteria air pollutants and meteorological parameters recorded from June, 2013 to December, 2014 while the data set of January, 2015 to May, 2015 is used for data validation. The accuracy and errors in the MLR and PCR models were evaluated in terms of performance indicators (PIs)

2.3 Statistical analysis

Data analysis was carried using the statistical software XLSTAT 2015. Step wise multiple regression (MLR) and Principle Component Regression (PCR) analyses have been used for finding the predictive equations of the criteria pollutants.

2.3.1 Principal component analysis (PCA)

Among multivariate techniques, Principal Components Analysis (PCA) is designed to classify variables based on their correlations with each other. The goal of PCA and other factor analysis procedures, is to consolidate a large number of observed variables into a smaller number of factors (components) that can be more readily interpreted as these underlying processes. It is often used as an exploratory tool to identify the major sources of air pollutant emissions [8] [9]. For physical interpretation of the components, loadings of variables on the component are estimated. Loading represents the degree and direction of relationship of the variables with a factor. An analysis of the PC loadings on the chosen variables allows the identification of the PCs as pollution sources affecting the data. The number of factors (PCs) is selected such that the cumulative percentage variance explained by all the chosen factors is more than 70%. As the normalized variables each carry one unit of variance, so the factors with eigen value more than 1 are chosen for the study. The factors with eigen values less than one are discarded as they are assumed to contain less information [10]. To undertake PCA, the XLSTAT 2015 statistical software was used, specifying the principal components method with varimax rotation [11]. The rotation of the component axis is performed so that components are clearly defined by high loadings for some variables and low loadings for others, facilitating the interpretation in terms of original variables.

The principal components of the predictor variables are obtained using a data reduction technique by means of finding linear combinations of the original variables. In general, PCs are expressed by the following equation

$$PC_i = A_{1i} V_i + A_{2i} V_2 + \ldots\ldots + A_{ni} V_n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where,

PC_i is principal component i and

A_{ni} is the loading (correlation coefficient) of the original variable V_n.

As the scores of high loading components with an eigen value greater than or equal to 1 account for most of the variations in the data set, it is ideal to use them as independent or predictor variables in regression analysis. Thus, principal component regression (PCR) establishes relationship between dependent variables and the selected PCs of the independent variables [12].

2.3.2 Multiple Linear Regression (MLR)

Multiple linear regression attempts to model the relationship between two or more explanatory variables (independent variables) and a response variable (dependent variable) by fitting a linear equation to observed data. This multivariate statistical technique finds wide application in the field of atmospheric science, especially air pollution studies. The MLR

technique has the capability of exploring the contribution of selected variables to chosen air pollutant concentration. The general equation of MLR is expressed as [12]

$$y = b_0 + \sum_{k=1}^{p} \mathbf{b_i}\, \boldsymbol{x_i} + \xi \quad \text{.....(2)}$$

Where,
b_i is the regression coefficient,
x_i is the independent variable, and
ξ is the stochastic error associated with the regressions.

2.3.3 Principal Component Regression (PCR)

Principal Component Regression (PCR) is a combination of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR). The PCs obtained in PCA are used as the inputs in MLR. The selected variables with high loadings from PCA ensure the inclusion of the majority of the original variances in the statistical model and they are ideal for use as independent variables in MLR **[12]**. The use of PCs as the independent variables of MLR reduces the problem of multicolinearity.

2.3.4 Performance Indicators (PIs)

The performance of MLR and PCR models are assessed on the basis of the performance indicators (PIs). Good prediction models should have minimal errors (closer to 0 for NAE and RMSE) and high accuracy (closer to 1 for IA, PA, and R^2). The following PIs are used in this study -

- Normalized Absolute error (NAE) - It measures the average difference between predicted and observed values in all cases divided by observed values **[5]** and is expressed as:

$$\text{NAE} = \frac{\sum_{i=1}^{n} |P_i - O_i|}{\sum_{i=1}^{n} O_i} \quad \text{.....(3)}$$

where *n* is the sample size, P_i is the predicted concentration of the criteria pollutant and O_i is the observed value of the pollutant concentration.

- Root Mean Square Error (RMSE) - It measures the success of numerical prediction.

RMSE is calculated by the equation [13] [14]

$$\text{RMSE} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (P_i - O_i)^2} \quad \text{.....(4)}$$

where n is the number of sample, O_i is the observed concentrations of the pollutants and P_i is the predicted concentration of the pollutants.

- Prediction accuracy (PA) - The prediction accuracy is computed using by the following equation [15]:

$$\text{PA} = \frac{\sum_{i=1}^{n} (P_i - \bar{P})^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2} \quad \text{.....(5)}$$

where n is the number of sample, O_i is the observed concentrations of the pollutants and P_i is the predicted concentration of the pollutants.

- Index of agreement (IA) - a measure of accuracy, was calculated using Equation (6) [16].

$$\text{IA} = 1 - \left[\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right] \quad \text{.....(6)}$$

where n is the number of sample, O_i is the observed concentrations of the pollutants and P_i is the predicted concentration of the pollutants

- Coefficient of determination (R^2) - The coefficient of determination explains how much the variability in the predicted data can explain by the fact that they are related to the observed values. R^2 is expressed by the following equation [15] :

$$R^2 = \left(\frac{\sum_{i=1}^{n} (P_i - \overline{P})(O_i - \overline{O})}{n\, S_{pred}\, S_{obs}} \right)^2 \quad \text{..... (7)}$$

where n is the number of sample, O_i is the observed concentrations of the pollutants and P_i is the predicted concentration of the pollutants, $\overline{P}$ is the average of predicted value , $\overline{O}$ is the average of observed values, S_{pred} is a standard deviation of the predicted pollutant concentration, S_{obs} is a standard deviation of the observed pollutant concentration.

3. Result and discussion

3.1 MLR model development

MLR modeling (stepwise method) has been performed for finding the predictive equations of the criteria pollutants with the regression assumptions approximately satisfied. During this statistical analysis, the distribution of residuals was approximately with zero mean and constant variance. Variance Inflation Factor (VIF) was mostly below 10 except on very few occasions when the VIF value exceeded 10. Therefore, the MLR predictor variables have negligible collinearity problem.

3.2 PCR model development

PCA was applied for variable reduction and for providing most relevant variable for understanding the pollutant variation. Varimax rotation was applied in PCA for maximizing the loading of a predictor variable on one component. The adequacy of input data for the PCA was assessed using the Kaiser–Meyer–Olkin (KMO) test. The results obtained from application of KMO test on the input data set were more than 0.5 which indicated that the input data set were sufficient for PCA.

Before extraction using PCA, 16 linear components (twelve criteria pollutants, temperature, humidity, cloud cover and wind speed) were used. After performing PCA, three linear factors were considered as principal components (PCs) on the basis of their eigen values. In PCA, the eigen value provides the amount of variation explained by each PC. As the normalized variables each carry one unit of variance, the factors with eigen value more than 1 were chosen for the study. The factors with eigen values less than one are discarded as they provide less information [10]. Occasionally, eigen values smaller than unity are considered as they are very close to one **[17].**The variability of PCs obtained after varimax rotation are summarized in Table 1.The obtained PCs are used as the independent variables (explanatory variables) and the original criteria pollutant as the dependent variables in stepwise multiple linear regression analysis in PCR model. The use of PCs as input in MLR is intended to reduce the complexity and multicollinearity problems of the models.

Table 1. Total variance for different criteria pollutants after varimax rotation

Sl.No.	Parameter	Components	Eigen value	Variability (%)	Cumulative %
1	PM_{10}	PC1	8.880	35.416	35.416
		PC2	1.938	15.687	51.103
		PC3	1.064	23.157	74.260
2	$PM_{2.5}$	PC1	8.962	34.053	34.053
		PC2	1.933	15.327	49.379
		PC3	1.094	25.556	74.935
3	Sulphur dioxide	PC1	8.960	29.705	29.705
		PC2	1.959	14.692	44.397
		PC3	1.091	30.666	75.063
4	Nitrogen dioxide	PC1	8.922	32.114	32.114
		PC2	1.903	15.527	47.641
		PC3	1.087	26.808	74.449
5	Carbon monoxide	PC1	9.596	33.732	33.732
		PC2	1.550	31.157	64.889
		PC3	1.031	11.214	76.104
6	Ammonia	PC1	9.054	32.690	32.690
		PC2	1.897	15.845	48.535
		PC3	1.067	26.579	75.114
7	Ozone	PC1	9.115	30.682	30.682
		PC2	1.848	15.140	45.822
		PC3	1.110	29.638	75.460
8	Benzene	PC1	9.128	40.932	40.932
		PC2	1.960	22.655	63.588
		PC3	0.941	11.591	75.178
9	Benz(a)Pyrene	PC1	9.162	39.499	39.499
		PC2	1.951	23.602	63.101
		PC3	0.999	12.597	75.699
10	Arsenic	PC1	8.962	33.076	33.076
		PC2	1.952	15.137	48.213
		PC3	1.090	26.814	75.027
11	Lead	PC1	9.434	31.759	31.759
		PC2	1.620	12.760	44.520
		PC3	1.103	31.461	75.980
12	Nickel	PC1	9.018	33.232	33.232
		PC2	1.959	15.636	48.868
		PC3	1.097	26.593	75.461

3.3 Comparison of MLR and PCR models

MLR and PCR models provide an estimate of 24 hour average concentration of all the criteria pollutants (Table 2).

Table 2. Summary of models of all the criteria pollutants using Multiple Linear Regression (MLR) and Principal Component Regression (PCR)

Sl.No.	Parameter	Method	R^2	Model
1	PM_{10}	MLR	0.962	PM_{10} = 0.135+4.138*As + 17.633*BAP + 1.788*Ni + 1.156*$PM_{2.5}$
		PCR	0.918	PM_{10} = 102.688 + 23.684*PC1 + 17.671PC2 + 35.397*PC3
2	$PM_{2.5}$	MLR	0.945	$PM_{2.5}$ = 6.22 - 4.01*BAP -0.13*O_3 + 0.529*PM_{10} + 1.745*SO2 - 2.455*WS
		PCR	0.852	$PM_{2.5}$ = 59.280 + 11.535*PC1 + 12.138*PC2 + 18.645*PC3
3	Sulphur dioxide (SO_2)	MLR	0.898	SO_2 = 1.184 + 0.093*NH_3 - 0.283*As +3.553*Pb + 0.108*NO_2 - 0.042*O_3+ 0.018*$PM_{2.5}$
		PCR	0.779	SO_2 = 8.22 + 1.01*PC1 + 0.746*PC2 + 1.231*PC3
4	Nitrogen dioxide (NO_2)	MLR	0.937	NO_2 = 28.993 - 20.994*CO + 0.298*O_3 + 3.837*SO_2 - 0.723*RH
		PCR	0.888	NO_2 = 53.871 + 9.939*PC1 +1.281*PC2 + 11.993*PC3
5	Carbon monoxide (CO)	MLR	0.603	CO = 0.743 + 0.498*Pb - 0.005*Ni - 0.004*T
		PCR	0.439	CO = 0.665 + 0.017*PC2 + 0.046*PC3
6	Ammonia (NH_3)	MLR	0.874	NH_3 = 3.957 + 1.039*As - 2.153*C_6H_6 + 7.971*CO - 17.578*Pb + 0.410*Ni +0.142*O3 + 1.085*SO_2
		PCR	0.799	NH_3 = 25.773 + 2.498*PC1 + 4.244*PC3
7	Ozone (O_3)	MLR	0.871	O_3 = 14.02 + 1.255*NH_3 + 4.316*As + 0.659*NO_2 -4.083*SO_2 - 0.066*CC
		PCR	0.77	O_3 = 53.433 + 7.695*PC1 - 1.579*PC2 + 12.798*PC3
8	Benzene (C_6H_6)	MLR	0.837	C_6H_6 = 1.093 + 0.536*BAP - 0.447*CO + 0.002*PM_{10}
		PCR	0.698	C_6H_6 = 1.352 + 0.229*PC1 + 0.21*PC2
9	Benzo(a)pyrene (BAP)	MLR	0.858	BAP = -0.643 + 0.067*As + 0.755*C_6H_6 + 0.003*PM10 - 0.038*SO_2
		PCR	0.71	BAP = 0.579 + 0.265*PC1 +0.318*PC2
10	Arsenic (As)	MLR	0.868	As = -0.306 + 0.05*NH_3 + 0.538*BAP - 2.130*CO + 3.853*Pb - 0.086*Ni + 0.026*O_3 + 0.006*PM_{10} + 0.106*WS
		PCR	0.796	As = 2.000 + 0.674*PC1 +0.114*PC2 + 0.695*PC3
11	Lead (Pb)	MLR	0.842	Pb = -0.457 - 0.009*NH_3 + 0.05*As - 0.054*BAP + 0.646*CO +0.015*Ni + 0.001*PM_{10} + 0.01*SO_2 +0.007*RH - 0.01*WS
		PCR	0.59	Pb = 0.163 + 0.076*PC2 + 0.056*PC3
12	Nickel (Ni)	MLR	0.825	Ni = 2.925 + 0.205*NH_3 - 0.949*As - 9.310*CO + 14.681*Pb + 0.049*O_3 + 0.037*PM_{10}
		PCR	0.749	Ni = 8.814 + 1.769*PC1 + 1.158*PC2 + 2.197*PC3

* Temperature (T), relative humidity (RH), wind speed (WS) and cloud cover (CC)

The MLR models were found to perform better than the corresponding PCR models as the R^2 values of the MLR models are higher than those of PCR models (Table 2). The predictive equations suggested by the statistical models suggest that meteorological factors (temperature, relative humidity, cloud cover and wind speed) play an important role in the prediction of concentration of the criteria pollutants. For example, cloud cover is negatively associated with ozone concentration in the predictive equation proposed by the MLR model which is in agreement with the mechanism of photochemical formation of tropospheric ozone. In general, high wind speed flushes out the air pollutants near the earth's surface thereby leading to lower concentration of ground level air pollutants. Such a result is reflected in the predictive equations of the MLR model. The PCR has more degrees of freedom and offers variable combinations for the principal components in choosing multiple components but the use of PCs as the inputs in the MLR could not improve the performance of the model. Actually, the PCA is an unsupervised dimension reduction methodology which does not consider the correlation among the dependent and independent variables. This might be a reason for the

failure of the PCR model. Elbayoumi *et al*. [5] also concluded that the use of PCR could not improve the accuracy in predicting indoor PM_{10} and $PM_{2.5}$ in the Gaza Strip (Palestine) over MLR. Awang *et al*. [4] also reported the optimal performance of MLR model for daytime ground level ozone in terms of normalized absolute error, index of agreement, prediction accuracy, and coefficient of determination (R^2). The R^2 for the correlation between the observed and the predicted concentration of the criteria pollutants for MLR and PCR models are shown in Figures 1 to 4. The performances of the two models are further compared on the basis of the performance indicators namely normalized absolute error (NAE), root mean square error (RMSE), prediction accuracy (PA), index of agreement (IA) and coefficient of determination (R^2) (Table 3). Good prediction models should have minimal errors (closer to 0 for NAE and RMSE) and high accuracy (closer to 1 for IA, PA, and R2). On the basis of this principle, MLR models for prediction of air pollutants are found to give better performance than the corresponding PCR model.

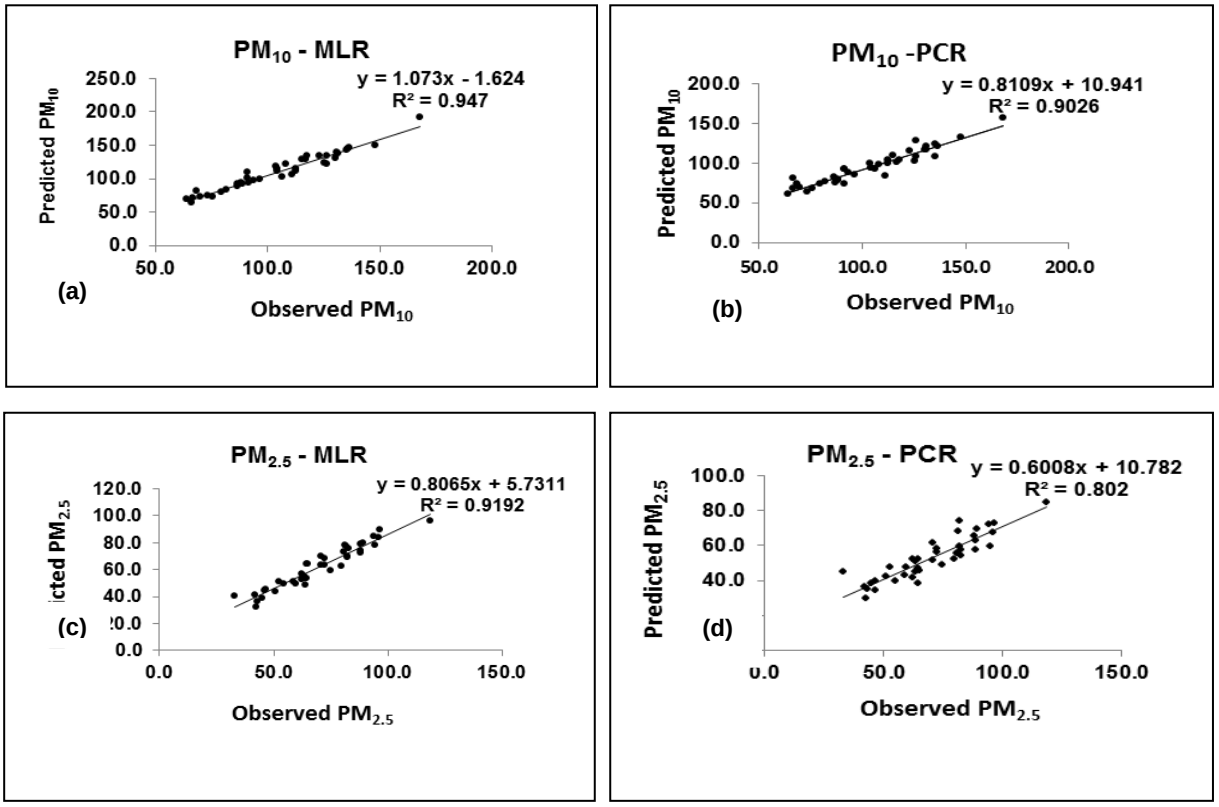
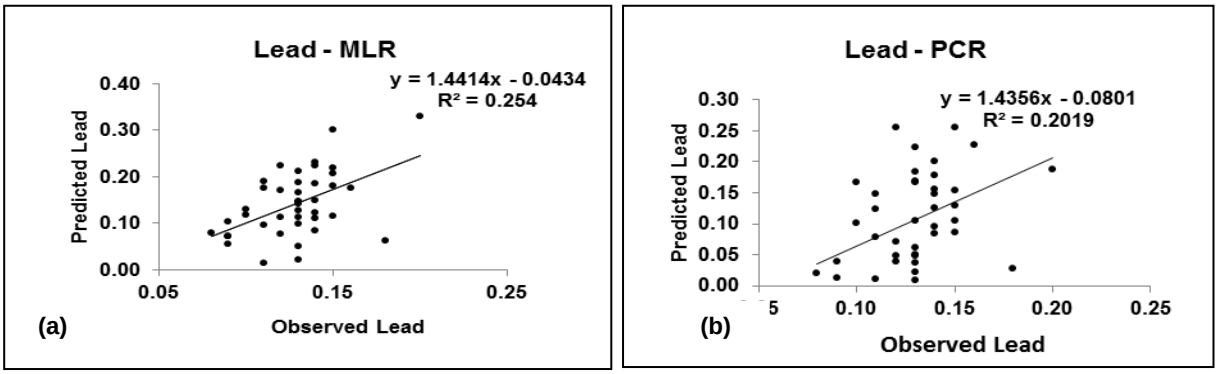


Figure 1. Scatter plots of observed and predicted values of (a) PM_{10} by MLR method, (b) PM_{10} by PCR method, (c) $PM_{2.5}$ by MLR method, (d) $PM_{2.5}$ by PCR method



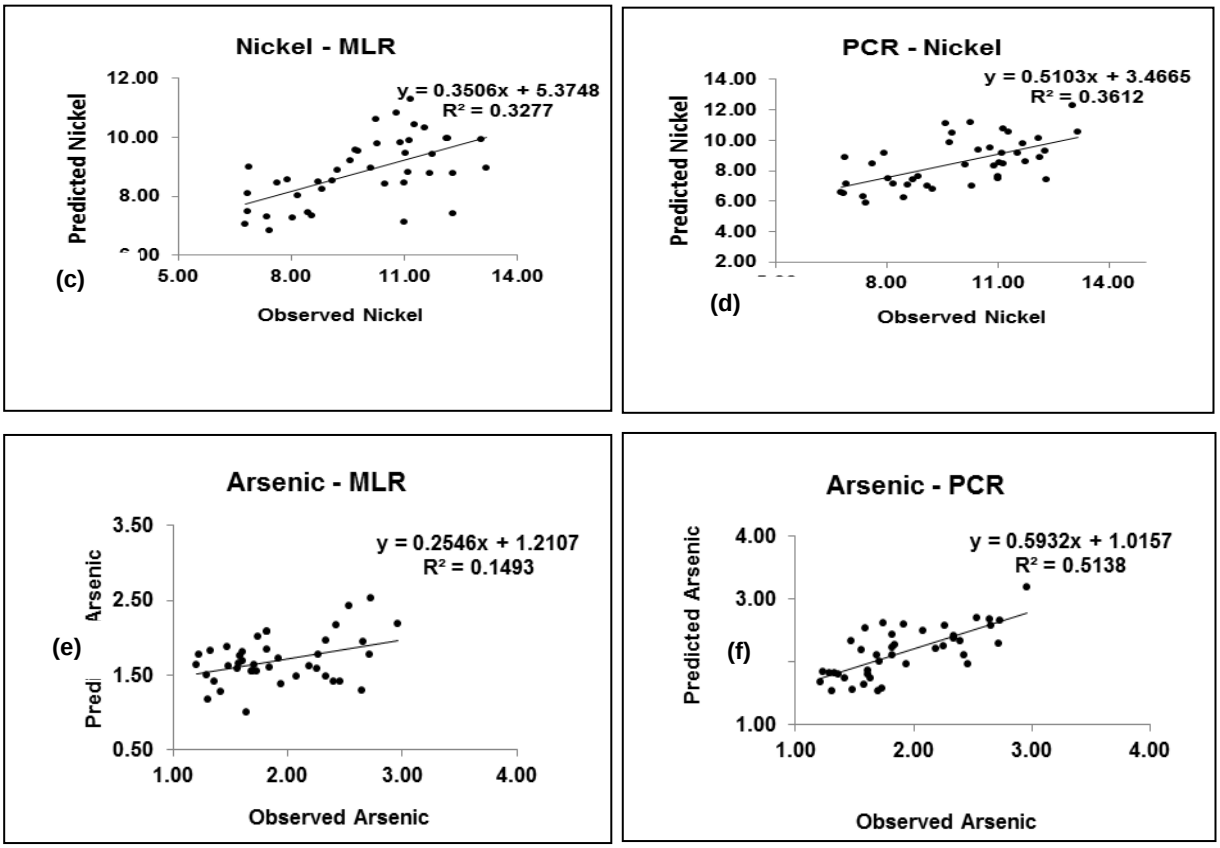
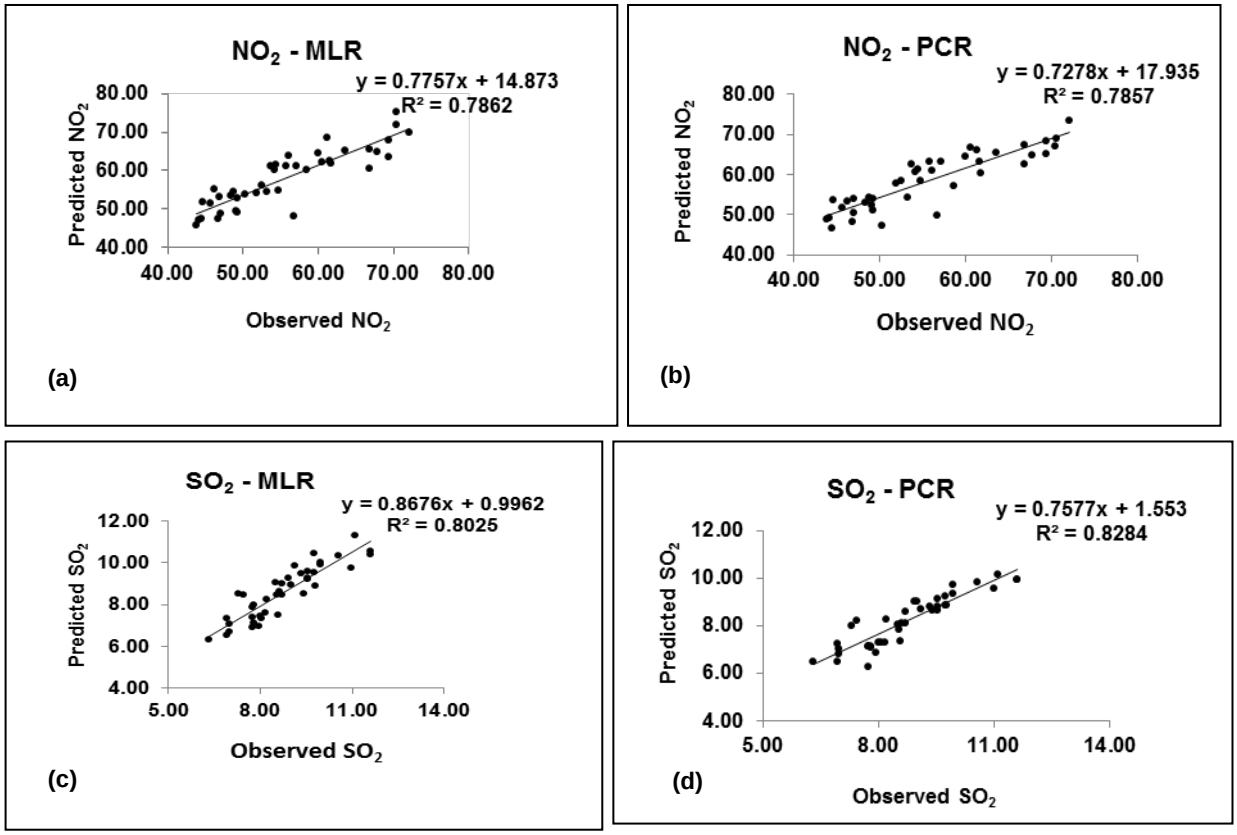


Figure 2. Scatter plots of observed and predicted values of (a) Lead (Pb) by MLR method, (b) Lead (Pb) by PCR method, (c) Nickel (Ni) by MLR method, (d) Nickel (Ni) by PCR method, (e) Arsenic (As) by MLR method and (f) Arsenic (As) by PCR method



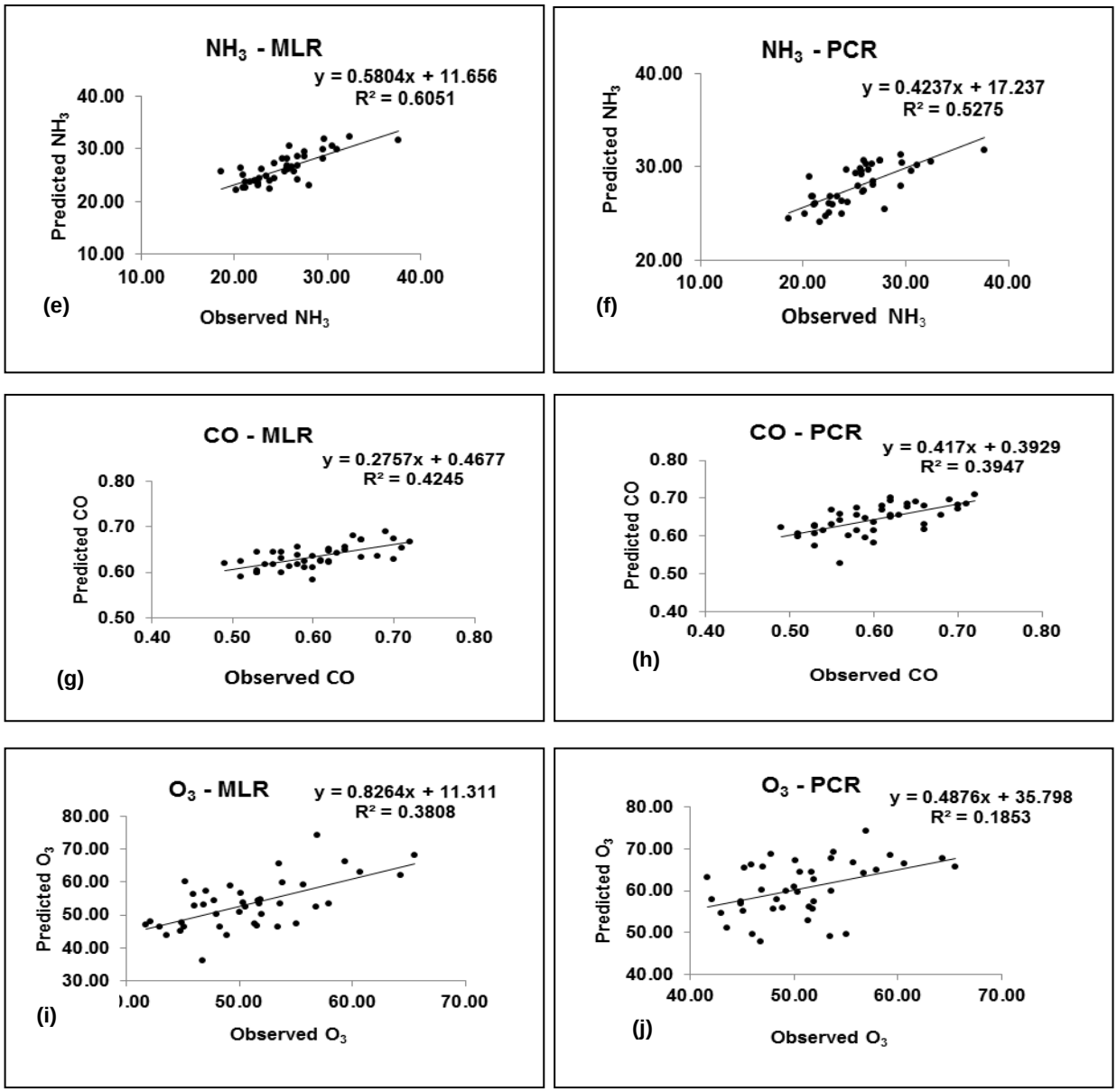


Figure 3. Scatter plots of observed and predicted values of (a) NO_2 by MLR method, (b)NO_2 by PCR method, (c) SO_2 by MLR method, (d) SO_2 by PCR method, (e) NH_3 by MLR method , (f) NH_3 by PCR method, (g) CO by MLR method, (h) CO by PCR method, (i) O_3 by MLR method and (j) O_3 by PCR method

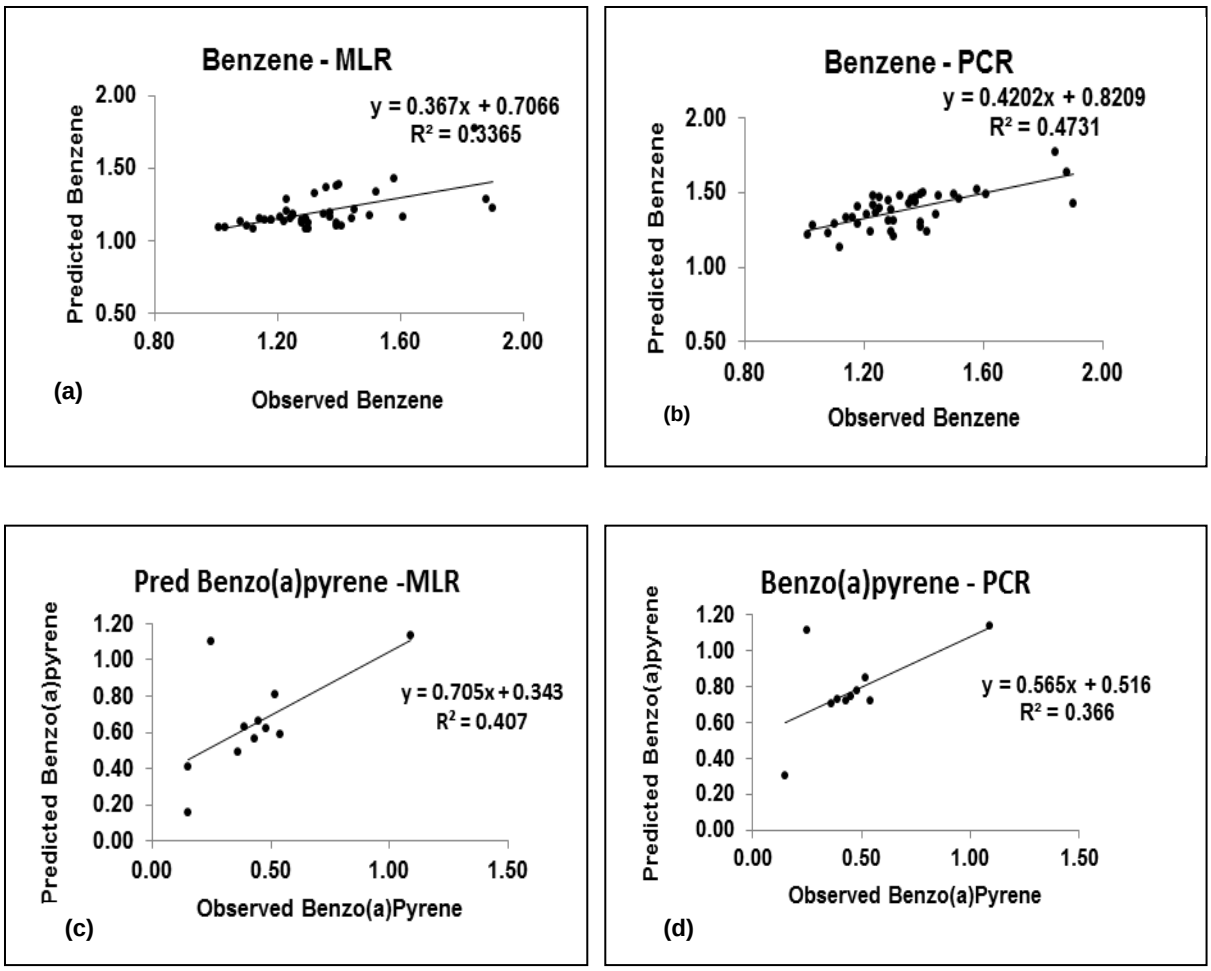


Figure 4. Scatter plots of observed and predicted values of (a) Benzene by MLR method, (b)Benzene by PCR method, (c) Benzo(a)pyrene by MLR method and (d) Benzo(a)pyrene by PCR method

It appears from Table 3 that the error indicators (NAE and RMSE) are minimum and accuracy indicators (IA, PA and R^2) are maximum in case of each criteria pollutant by using MLR model (except Benzene and Arsenic). This observation suggests that the physico-chemical characteristics and the interaction of Benzene and Arsenic with other substances in the atmosphere should be explored for understanding these outcomes of these statistical models.

Table 3. Summary of performance indicators (PIs) of the models

Sl.No.	Parameters	Method	NAE	RMSE	IA	PA	R^2
1	PM_{10}	MLR	0.068	8.981	0.971	0.823	0.902
		PCR	0.097	12.126	0.934	0.728	0.859
2	$PM_{2.5}$	MLR	0.117	9.8	0.924	0.708	0.875
		PCR	0.254	19.718	0.725	0.451	0.763
3	Sulphur dioxide	MLR	0.054	0.613	0.941	0.938	0.764
		PCR	0.076	0.799	0.891	0.693	0.789
4	Nitrogen dioxide	MLR	0.068	4.634	0.915	0.765	0.748
		PCR	0.076	4.889	0.902	0.674	0.748
5	Carbon monoxide	MLR	0.076	0.057	0.635	0.181	0.423
		PCR	0.088	0.063	0.643	0.438	0.732
6	Ammonia	MLR	0.075	2.59	0.839	0.557	0.576
		PCR	0.130	3.752	0.678	0.34	0.504
7	Ozone	MLR	0.102	6.593	0.744	0.558	0.362
		PCR	0.202	11.842	0.488	0.779	0.176
8	Benzene	MLR	0.115	0.219	0.645	0.400	0.32
		PCR	0.034	0.157	0.762	0.373	0.453
9	Benz(a)Pyrene	MLR	1.326	0.361	0.529	0.702	0.278
		PCR	1.899	0.469	0.428	0.955	0.283
10	Arsenic	MLR	0.205	0.511	0.605	0.433	0.141
		PCR	0.170	0.415	0.784	0.685	0.488
11	Lead	MLR	0.379	0.062	0.52	0.124	0.251
		PCR	0.484	0.074	0.448	0.100	0.196
12	Nickel	MLR	0.138	1.863	0.670	0.375	0.312
		PCR	0.178	2.092	0.690	0.721	0.344

4. Conclusion

In this study, multiple linear regression analyses have been performed using the original variables and principal components (PCs) as the inputs. MLR can encounter the complexity of multicollinearity as the environmental variables are correlated to each other. MLR using the PCs as the inputs is known as principal component regression (PCR) and the use of this technique is expected to reduce the problem of multicollinearity. Both models provide an estimate of 24 hour average concentration of all the criteria pollutants. On the basis of the performance indicators, the MLR model was found to perform better than the PCR in most cases (except Benzene and Arsenic). Analysis of the physico - chemical properties and mode of interaction of Benzene and Arsenic with other substances present in the ambient

environment may further clarify the characteristics of these two criteria pollutants. Meteorological parameters, particularly temperature, relative humidity and cloud cover are found to influence the concentration of the air pollutants over that region. The use of characteristics of boundary layer processes and traffic may further improve the accuracy of prediction of the criteria pollutants over urban areas.

References

[1] Chaloulakou, A. and Mavroidis, I. (2002) Comparison of indoor and outdoor concentrations of CO at a public school. Evaluation of an indoor air quality model. *Atmospheric Environment* **36**, 1769 – 1781.

[2] Özbay, B. (2012) Modeling the effects of meteorological factors on SO_2 and PM_{10} concentrations with statistical approaches, *CLEAN- Soil, Air, Water* **40**, 571 – 577.

[3] Hrust, L., Klai´c, Z.B., Križan, J., Antoni´c, O. and Hercog, P. (2009) Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations, *Atmospheric Environment* **43**, 5588–5596.

[4] Awang, N. R., Ramli, N. A., Yahaya, A. S. and Elbayoumi, M. (2015) Multivariate methods to predict ground level ozone during daytime, nighttime, and critical conversion time in urban areas, *Atmospheric Pollution Research* **6**, 726 - 734.

[5] Elbayoumi, M., Ramli, N.A., Yusof, N.F.F.M., Yahaya, A.S.B., Madhoun, W.A. and Ul-Saufie, A.Z. (2014). Multivariate methods for indoor PM_{10} and $PM_{2.5}$ modelling in naturally ventilated schools buildings, *Atmospheric Environment* **94**, 11 – 21.

[6] Luvsan, M.E., Shie, R.H., Purevdori, T., Badarch, L., Baldorj, B. and Chan, C.C. (2012) The influence of emission sources and meteorological conditions on SO_2 pollution in Mongolia. *Atmospheric Environment* **61**, 542 – 549.

[7] Sayegh, A. S., Munir, S. and Habeebullah, T. M (2014) Comparing the Performance of Statistical Models for Predicting PM10 Concentrations, *Aerosol and Air Quality Research* **14**, 653–665.

[8] Bruno, P., Caselli, M., Gennaro, G. and Traini, A. (2001) Source apportionment of gaseous atmospheric pollutants by means of an absolute principal component scores (APCS) receptor model, *Fresenius Journal* of *Analytical Chemistry* **371**, 1119 – 1123.

[9] Guo, H., Wang, T. and Louie P.K.K. (2004) Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model, *Environmental Pollution* **129**, 489 – 498.

[10] Maenhaut, W. and Cafmeyer, J. (1987) Particle induced X-ray emission analysis and multivariate techniques: An application to the study of the sources of respirable atmospheric particles in Gent, Belgium, *Journal of* Trace *and* Microprobe *Techniques* **5**, 135 – 158.

[11] Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis, *Psychometrika* **23**, 187 – 200.

[12] Gvozdic, V., Kovac – Andric, E. and Brana, J.(2011) Influence of meteorological factors NO_2, SO_2, CO and PM_{10} on the concentration of O_3 in the urban atmosphere of Eastern Croatia, *Environmental Modeling & Assessment* **16**, 491 – 501.

[13] Alshitawi, M., Awbi, H. and Mahyuddin, N. (2009) Particulate matter mass concentration (PM_{10}) under different ventilation methods in classrooms. *International Journal of Ventilation* **8**, 93 – 108.

[14] Karppinen, A., Kukkonen, J., Elolähde, T., Konttinen, M., Koskentalo, T. and Rantakrans, E., (2000) A modelling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area, *Atmospheric Environment* **34**, 3723 – 3733.

[15] Gervasi, O. (2008) Computational Science and its Applications - ICCSA 2008, Springer, Italy.

[16] Yusof, N.F.F.M., Ramli, N.A., Yahaya, A.S., Sansuddin, N., Ghazali, N.A. and al Madhoun, W. (2010) Monsoonal differences and probability distribution of PM_{10} concentration, *Environmental Monitoring and Assessment* **163**, 655 – 667.

[17] Ul–Saufie, A.Z., Yahaya, A.S., Ramli, N.A., Rosaida, N. and Hamid, H.A. (2013) Future daily PM_{10} concentrations prediction by combining regression models and feed forward back propagation models with principle component analysis (PCA), *Atmospheric Environment* **77**, 621 – 630.

The extended Timoshenko beam element in finite element analysis for the investigation of size effects

D. Lu[1], Y.M. Xie[1], Q. Li2, X. Huang[1], Y.F. Li[1] and † S.W. Zhou[1]

1Centre for Innovative Structures and Materials, School of Engineering, RMIT University, GPO Box 2476, Melbourne 3001, Australia
2 School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, NSW 2006, Australia
*Presenting author: dingjie.lu@rmit.edu.au
†Corresponding author: shiwei.zhou@rmit.edu.au

Abstract

In this paper an extended Timoshenko beam element is developed for the investigation of size effect via finite element analysis. The surface effect derived from initial surface stress and surface elasticity is considered as external pressure in terms of the generalized Young-Laplace equation and the virtual displacement principle. We find the size effect highly relies on the geometrical model considered in numerical simulation. For a cantilever nanowire the stocky beam suddenly becomes strengthened provided the diameter is less than a critical size, while it is weakened for slender case. These abnormal changes of stiffness can be supported by static bending tests. This method will bring useful insight into the size effect and is of importance to some engineering applications like nanofabrication and nano sensors.

Keywords: Size effects, Surface effects, Timoshenko beam, FEM

Introduction

Size effects broadly refer to the abnormal changes of mechanical properties as the structure size approaches to tens of nanometer.[1,2] Over the last couple of decades, increasing attentions have been drawn to these behaviors because nanostructures have emerged as one of the most attractive topics and size effect at nanoscale has great potential to design lightweight material and sensors.[3] Previous studies have indicated size effects are attributed to the large ratio of surface area to the material volume, in which case the interactions of superficial atoms become extremely active. However the inherent mechanism is still a challenging problem.

In general, investigations of size effects on mechanical properties can be divided into two major groups, namely experimental validation and theoretical analysis based on simple beam theories. There have been many literatures report that surge of effecitveYoung's moduli is observed experimentally as the characteristic size approaching to nanometers.[4,5] Classical continuum theory cannot formulate this size dependent characteristic since it lacks of mechanism to account for the size effects on the mechanical properties of material.[6] Many efforts have been dedicated to build analytical framework to including size effect due to the difficult in manufacturing and controlling of materials at a length scale of several tens or hundreds nanometers.[7-10] Method like classical molecular dynamics simulation[11], nonlocal theory of elasticity[12] are effective in predicting size-dependence of mechanical properties at nano scale. However, the computational cost is intensive and paradoxes arise.[13-16]

Recently by incorporating surface elasticity[17] and generalized Young-Laplace equation,[18] the analytical solutions that predict the size-dependence of effective Young's modulus of nanomaterials show a good agreement with experimental data.[8,9] A recursive algorithm that breaks the constraint on model complexity for analytical solution successfully captured size effects of continuum nanoscale solid with complex 3D topology and obtained result that matches with experimental data.[7] But computation cost and convergence issue still remain. Given the fact that a great portion of nanomaterials that ubiquitously existing in nature is consist of beam like

ligaments,[14] it is very attractive to establish an extended beam element which can formulate the size dependence of the mechanical properties and overcomes those aforementioned drawbacks.
To incarnate the size effects in the simulation model, an extended Timoshenko beam element is developed so that surface elasticity theory and generalized Young-Laplace equation are well integrated into finite element analysis. Shape-dependent pressure is introduced in the model to serve as the external loading under which the nanostructure is deformed. Theoretical prediction is verified by two numerical tests which show the softened and strengthened beams below critical dimensional size.

Surface elasticity theory and generalized Young-Laplace equation

Surface elasticity theory [19] states that the surface stress $\sigma^s_{\alpha\beta}$, a symmetric 2×2 tensor in the tangent plane, is:

$$\sigma^s_{\alpha\beta} = \frac{\partial G\left(\varepsilon^s_{\alpha\beta}\right)}{\partial \varepsilon^s_{\alpha\beta}} + \tau_0 \delta_{\alpha\beta} \quad (\alpha, \beta = 1,2,3) \tag{1}$$

here $\varepsilon^s_{\alpha\beta}$ denotes the surface strain tensor, $G(\varepsilon^s_{\alpha\beta})$ is the surface energy and $\delta_{\alpha\beta}$ is the Kronecker delta. The initial surface stress is represented by τ_0. With assumption that the surface is homogeneous, isotropic, and linearly elastic, the overall surface stress tensor can be further simplified to:

$$\sigma^s_{\alpha\beta} = \tau_0 + E_s \varepsilon^s_{\alpha\beta} \tag{2}$$

where E_s is the surface stiffness.
From generalized Young-Laplace equation [18], a stress jump normal to the interface which depends on the curvature $\kappa_{\alpha\beta}$ and surface tension $\tau_{\alpha\beta}$ occurs on the curved material surface as:

$$\Delta\sigma_{ij} n_i n_j = \tau_{\alpha\beta} \kappa_{\alpha\beta} \tag{3}$$

here n_i and n_j is the unit normal vectors of the material surface.
Equations (1), (2) and (3) formulated the surface effects as a curvature-dependent distributed load along the normal direction of beam surface, as show in Fig. 1.

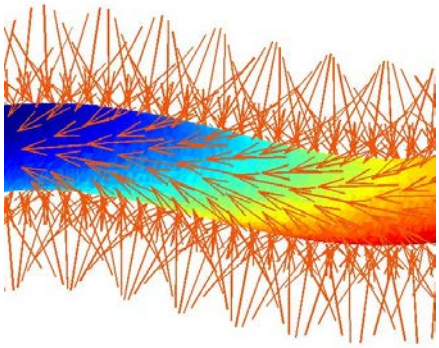


FIG. 1. A schematic of size-effect-induced pressure (red arrows) on the beam surface.

Timoshenko beam with surface effect

The formulation of the element stiffness matrix for extended Timoshenko beam element comprises contributions from axial compression, torsional and bending. Axial and torsional effects are considered in the conventional manner.
The bending contribution is formulated under Timoshenko beam theory. Element stiffness is derived from a 2D circular cross-section beam model with only bending considered for simplicity, extending to 3D is straightforward.

the axial $u(x, y)$ and transverse $v(x, y)$ displacements in the x-y plane is used to describe the motion of an arbitrary material point on the beam. Here motion in z direction is not considered. The assumption of Timoshenko beam theory can be represented as:

$$\begin{aligned} u(x,\mathrm{y}) &= -y(\frac{\partial v(x)}{\partial x}+\gamma) \\ v(x,y) &= v(x) \\ \theta_z &= \frac{\partial v}{\partial x}+\gamma = v'+\gamma \\ \gamma &= \frac{V}{GA_s} \end{aligned} \tag{4}$$

here θ_z is the rotation angle and γ is angle of shearing.
The strain can be determined by differentiating the displacement of (5) as:

$$\begin{aligned} \varepsilon_{11} &= -y\frac{\partial \theta_z}{\partial x} \\ \varepsilon_{13} &= \frac{1}{2}(\theta_z - \frac{\partial v}{\partial x}) = \frac{1}{2}\gamma \end{aligned} \tag{5}$$

The stress component given by Hooke's law is,

$$\begin{aligned} \sigma_{11} &= E\varepsilon_{11} \\ \sigma_{13} &= G\gamma \end{aligned} \tag{6}$$

here E is the elastic modulus, G is the shear modulus. The bending moment M over the cross section is the integral,

$$M = \int_A -y\sigma_{11}dA \tag{7}$$

According to the principle of virtual displacements, the virtual external work of real external forces moving through collocated virtual displacements equals the internal virtual work of real stresses in equilibrium with real forces with the virtual strains compatible with the virtual displacements integrated over the volume of the solid[20] and can be mathematically expressed as:

$$\begin{aligned} &\delta W_I = \delta W_E \\ &\int_v \sigma\cdot\delta\varepsilon dV = \int_v \delta U^T f^B dV + \int_S \delta U^{ST} f^s dV + \sum_i \delta U^{iT} F^i \end{aligned} \tag{8}$$

where δW_I is the total internal virtual work and δW_E is the total external virtual work. σ is the actual stress, $\delta\varepsilon$ is the virtual strains. f^b, f^s and F^i are the actual external body force, surface traction and concentrated force and δU^T, δU^{ST} and δU^{iF} are the corresponding virtual displacement.
The overall internal virtual work of Timoshenko beam including surface effect can be express as:

$$\delta W_I = \delta W_{IC} + \delta W_{IS} \tag{9}$$

where δW_I denote the overall internal virtual work and it is consist of the contribution from the conventional bending and shearing effects and the contribution of surface effects from initial surface tension and surface stiffness, denoted as δW_{IC} and δW_{IS} correspondingly. The Timoshenko beam theory assumes that the internal energy of beam member is due to bending and shearing which can be expressed as:

$$\delta W_{IC} = \int_0^L EI(\frac{\partial \theta_z}{\partial x})^2 dx + \int_0^L \kappa AG(\theta_z + \frac{\partial v}{\partial x})^2 dx \tag{10}$$

here I denotes the moment of inertia of the cross section. EI is the flexure rigidity, κ denotes the shear area coefficient and κ = 10/9 for solid circular sections, A is the cross section area.

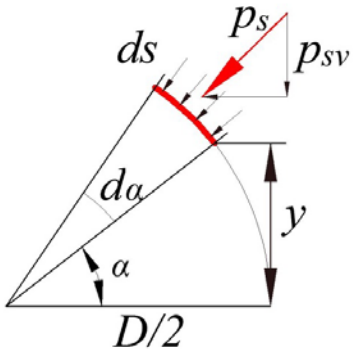


FIG. 2. Circular cross section of beam with surface effects considered.

For a representative infinitesimal surface element on the surface of the cross section display as red arc in Fig. 2, according to Eq. (6) the longitudinal strain, which is perpendicular to the cross sectional plane, is,

$$\varepsilon_s = -\frac{D}{2}\sin\alpha\frac{\partial\theta_z}{\partial x} \tag{11}$$

Base on Eq. (2), the surface stress along beam axis can be expressed as:

$$\begin{aligned}\tau_{axial} &= \tau_0 + E_s\varepsilon_s \\ &= \tau_0 + E_s(-\frac{D}{2}\sin\alpha\frac{\partial\theta_z}{\partial x})\end{aligned} \tag{12}$$

This surface stress along beam axis introduces an extra moment on the infinitesimal surface element

$$\begin{aligned}dM_s &= -\tau_{axial}yds \\ &= -(\tau_0 + E_s(-\frac{D}{2}\sin\alpha\frac{\partial\theta_z}{\partial x}))\frac{D}{2}\sin\alpha\frac{D}{2}d\alpha\end{aligned} \tag{13}$$

By integration along the edge of the cross section the overall moment of the surface effect at this cross section is,

$$\begin{aligned}M_s &= \int dM_s \\ &= \int_0^{2\pi} -(\tau_0 + E_s(-\frac{D}{2}\sin\alpha\frac{\partial\theta_z}{\partial x}))\frac{D}{2}\sin\alpha\frac{D}{2}d\alpha \\ &= \frac{\pi D^3}{8}E_s\frac{\partial\theta_z}{\partial x}\end{aligned} \tag{14}$$

The contribution of the surface effects to the internal work is then determined as:

$$\delta W_{IS} = \int_0^L M_s\frac{\partial\theta_z}{\partial x}dx = \int_0^L \frac{\pi E_s D^3}{8}(\frac{\partial\theta_z}{\partial x})^2 dx \tag{15}$$

The overall internal virtual work is obtained as:

$$\delta W_I = \delta W_{IC} + \delta W_{IS} = \int_0^L EI(\frac{\partial\theta_z}{\partial x})^2 dx + \int_0^L \kappa AG(\theta_z + \frac{\partial v}{\partial x})^2 dx + \int_0^L \frac{\pi E_s D^3}{8}(\frac{\partial\theta_z}{\partial x})^2 dx \tag{16}$$

The external virtual work δW_E is also composed of the conventional and the surface effect part. The conventional part is the work done by the external load as:

$$\delta W_{EC} = \int_0^L q_c v dx \tag{17}$$

where q_c is the transverse force per unit length that acts on the beam.

From generalized Young-Laplace equation the surface tension alone beam longitudinal direction causes a force normal to the surface, as shown in Fig. 2 as red arrow, which can be expressed as:

$$p_s = \tau_{axial}\frac{\partial\theta_z}{\partial x}\frac{D}{2}\sin\alpha d\alpha \tag{18}$$

Only the force component acting in the flexure plane contributes to the external virtual work and can be obtained by decomposition as,

$$p_{s\_flexure} = p_s\sin\alpha = \tau_{axial}\frac{\partial\theta_z}{\partial x}\frac{D}{2}\sin\alpha d\alpha\sin\alpha \tag{19}$$

By integrating $p_{s\_flexure}$ around the edge of the cross section total surface effects induced transverse load at this cross section can be obtained as:

$$\begin{aligned} q_s &= \int \tau_{axial} \frac{\partial \theta_z}{\partial x} \sin\alpha ds \\ &= \int_0^{\pi} \frac{D}{2} \frac{\partial \theta_z}{\partial x} (\tau_0 - E_s \sin\alpha \frac{D}{2} \frac{\partial \theta_z}{\partial x}) \sin\alpha d\alpha - \int_{\pi}^{2\pi} \frac{D}{2} \frac{\partial \theta_z}{\partial x} (\tau_0 - E_s \sin\alpha \frac{D}{2} \frac{\partial \theta_z}{\partial x}) \sin\alpha d\alpha \\ &= 2\tau_0 D \frac{\partial \theta_z}{\partial x} \end{aligned} \tag{20}$$

the surface effects part for the external work is,

$$\begin{aligned} \delta W_{ES} &= \int_0^L q_s v dx \\ &= \int_0^L 2\tau_0 D \frac{\partial \theta_z}{\partial x} v dx \end{aligned} \tag{21}$$

The total external virtual work is then determined as:

$$\delta W_E = \delta W_{EC} + \delta W_{ES} = \int_0^L q_c v dx + \int_0^L 2\tau_0 D \frac{\partial \theta_z}{\partial x} v dx \tag{22}$$

The virtual displacement principle of Timoshenko beam with surface effect including is then obtained as:

$$\int_0^L EI(\frac{\partial \theta_z}{\partial x})^2 dx + \int_0^L \kappa AG(\theta_z + \frac{\partial v}{\partial x})^2 dx + \int_0^L \frac{\pi E_s D^3}{8} (\frac{\partial \theta_z}{\partial x})^2 dx = \int_0^L q_c v dx + \int_0^L 2\tau_0 D \frac{\partial \theta_z}{\partial x} v dx \tag{23}$$

Compared with that of ordinary beam element,[21] after some rearrangement of Eq. (24),the controlling equation that correspond to the beam element with surface effect considered becomes,

$$\int_0^L EI(\frac{\partial \theta_z}{\partial x})^2 dx + \int_0^L \kappa AG(\theta_z + \frac{\partial v}{\partial x})^2 dx + \int_0^L \frac{\pi E_s D^3}{8} (\frac{\partial \theta_z}{\partial x})^2 dx - \int_0^L 2\tau_0 D \frac{\partial \theta_z}{\partial x} v dx = \int_0^L q_c v dx \tag{24}$$

By introducing the displacement interpolation matrix and strain displacement matrix, stiffness matrix of the extended Timoshenko beam element with surface effects can be obtained. The detailed finite element implementation is out of the scope of this paper.

Case study

Here we compare the deflection of cantilever obtain from proposed extended Timoshenko element with the analytical solution with and without surface effect. For cantilever beam with unit diameter, E_s=3.63 N/m and τ_0=1.22 N/m, the deflections predicted by analytical solution which do not have surface effect and the proposed new beam element for slenderness ratio equals 5 and 16 are plotted in Fig. 3.

It can be seen from Fig. 3(a) that the deflection obtained by proposed element, as shown with red solid line, have size effect since with the size approaching to nm scale the deflection decreases which indicate a strengthen effect. As the size extending to macro scale the deflection converges to that of conventional result. The analytical solutions obtained by both Timoshenko and Euler-Bernoulli beam, on the other hand, cannot capture this size effect as the deflection is constant with the change of scale. The difference between the green and blue line here is due to the difference between beam theories for stocky beam for which the Timoshenko beam theory is more physically realistic. Fig. 3(b) is the same simulation for slender beam with L/D = 16, it can been seen that the deflection at macro scale converges to the same value which is consistent with the theory that for beam with L/D > 16 the shear effect is negligible and both Timoshenko and Euler-Bernoulli beam theory obtain the same result. Meanwhile the deflection prediction by proposed element increases with decreasing of scale which indicates a softening effect occurs.

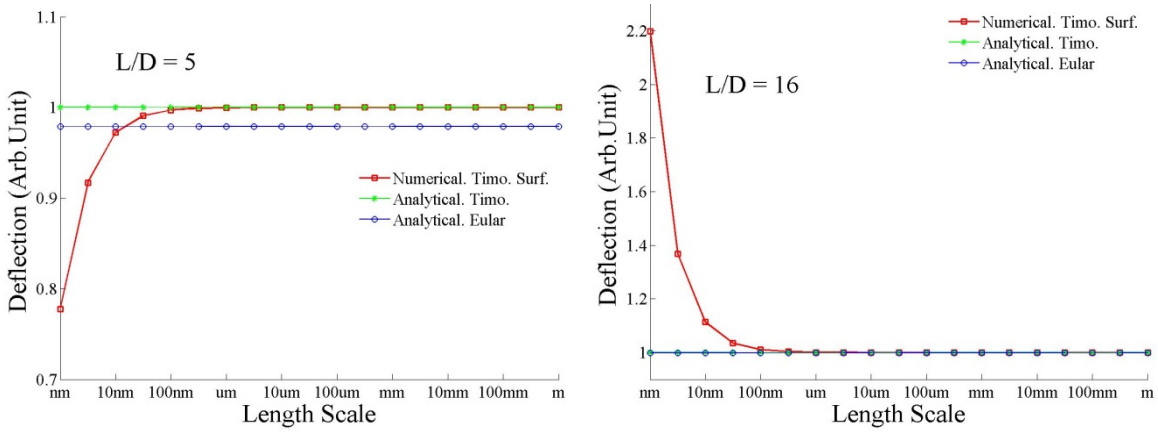


FIG. 3. Deflection prediction of proposed element and the analytical solution obtain using Timoshenko and Euler-Bernoulli beam for L/D = 5 and 16.

Conclusions

Based on the principle of virtual displacements, we derive the weak form for Timoshenko beam element with surface effects considered. Two characteristic parameters, the surface stiffness and initial surface tension, are introduced to be responsible for the size effect. Numerical simulation results successfully captured the size effect, strengthening and softening effects as the size decrease to nano meter is also observed which is consistent with theoretical prediction and experimental observation.

References

[1] A. M. Hodge, J. Biener, J. R. Hayes, P. M. Bythrow, C. A. Volkert, and A. V. Hamza, Acta Mater. **55**, 1343 (2007).
[2] R. Xia, X. Q. Feng, and G. F. Wang, Acta Mater. **59**, 6801 (2011).
[3] L. R. Meza, S. Das, and J. R. Greer, Science **345**, 1322 (2014).
[4] J. P. Salvetat, G. A. D. Briggs, J. M. Bonard, R. R. Bacsa, A. J. Kulik, T. Stöckli, N. A. Burnham, and L. Forró, Phys. Rev. Lett. **82**, 944 (1999).
[5] T. W. Tombler, C. Zhou, L. Alexseyev, J. Kong, H. Dai, L. Liu, C. S. Jayanthi, M. Tang, and S. Y. Wu, Nature **405**, 769 (2000).
[6] J. M. Gere and S. Timoshenko, *Mechanics of materials* (PWS-KENT Pub. Co., 1990).
[7] D. Lu, Y. M. Xie, Q. Li, X. Huang, and S. Zhou, J. Appl. Phys. **118**, 204301 (2015).
[8] D. Lu, Y. M. Xie, Q. Li, X. Huang, and S. Zhou, Appl. Phys. Lett. **105**, 101903 (2014).
[9] J. He and C. M. Lilley, Nano Lett. **8**, 1798 (2008).
[10] X. Q. Feng, R. Xia, X. Li, and B. Li, Appl. Phys. Lett. **94**, 011916 (2009).
[11] C. Mi, S. Jun, D. A. Kouris, and S. Y. Kim, Phys. Rev. B **77**, 075425 (2008).
[12] A. C. Eringen, J. Appl. Phys. **54**, 4703 (1983).
[13] C. Li and T. W. Chou, Appl. Phys. Lett. **84**, 121 (2004).
[14] A. Sears and R. C. Batra, Phys. Rev. B **69**, 235406 (2004).
[15] Q. Wang and K. M. Liew, Phys. Lett. A **363**, 236 (2007).
[16] N. Challamel and C. M. Wang, Nanotechnology **19**, 345703 (2008).
[17] M. E. Gurtin, J. Weissmüller, and F. Larché, Philos. Mag. A **78**, 1093 (1998).
[18] T. Chen, M. S. Chiu, and C. N. Weng, J. Appl. Phys. **100**, 074308 (2006).
[19] M. Gurtin and A. Ian Murdoch, Arch. Ration. Mech. An. **57**, 291 (1975).
[20] C. Lanczos, *The variational principles of mechanics* (Courier Corporation, 1970), Vol. 4.
[21] O. C. Zienkiewicz, R. L. Taylor, O. C. Zienkiewicz, and R. L. Taylor, *The finite element method* (McGraw-hill London, 1977), Vol. 3.

MPS-FEM Coupled Method for Interaction between Sloshing Flow and Elastic Structure in Rolling Tanks

Youlin Zhang, Zhenyuan Tang, Decheng Wan*

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author: dcwan@sjtu.edu.cn

Abstract

A coupling improved Moving Particle Semi-Implicit (MPS) method and the finite element method (FEM) is developed and applied to the problem of interaction between elastic structures and the violent sloshing flow in rolling tanks. The MPS method and the FEM, used to calculate the fluid field and structural deformation respectively, are introduced firstly. Then, the coupling strategy is also presented. To validate accuracy of the proposed algorithm for deformation of an elastic structure, two benchmarks are investigated and present results show good agreement with published data. Finally, cases about the sloshing with thin elastic baffles mounted in the partially filled rolling tanks are numerically studied. Both profiles of free surface and deflections of the baffles are in good agreement with experimental data.

Keywords: Particle method; Moving Particle Semi-Implicit (MPS); finite element method (FEM); Fluid structure interaction (FSI); Sloshing; Roll motion

Introduction

Fluid structure interaction (FSI) problems are commonly existent in ship and ocean engineering, such as sloshing in liquid containers while vessels sailing on very rough sea. Due to the impact loads induced by the periodic motion of inner liquid, the bulkheads or baffles mounted inside the tank may be deformed or even damaged. Hence, the investigation about interaction between the violent sloshing flow and the structures is useful for the assessment of safety of liquid containers.

For a typical FSI problem, the whole computational domain contains the fluid domain and the structural domain. Accurate prediction of the fluid computational domain is one of the key aspects for FSI problems. Generally speaking, numerical algorithms for the fluid domain simulation can be divided into two categories, the grid based methods and the meshless methods [1]. The grid based methods, such as the finite difference method (FDM), finite volume method (FVM), and finite element method (FEM), are much popular in the simulation of fluid domain. However, the main challenges of these approaches include inefficient process of grids generation for complex shape of structure, complex technology of dynamic mesh for moving boundary or structural deformation, simulation of free surface with large deformation or breaking, etc [2]. On the contrary, the meshless methods are in good performance to settle these challenges. One representative Lagrangian particle method for free surface flows is the MPS method which is originally proposed by Koshizuka and Oka [3] for incompressible flow. Since lots of improvements were proposed to suppress the numerical unphysical pressure oscillation [4]-[10], the MPS method can be employed to deal with kinds of hydrodynamic problems. Such as dam-breaking flow [11], water-entry flow [12]-[14], wave-float interaction

problem [2][15][16], sloshing in liquid tank [1][17], impinging jet flow [18], etc. In this paper, the MPS method is employed for the computation of fluid domain in FSI problem.

For the calculation of structural domain, deformation of structure is commonly computed based on the modal superposition analysis or the FEM method. Though the modal superposition analysis is easy to formulate and programming [19], it's incapable of solving large and nonlinear deformation of structure. Relatively, FEM method is widely employed to deal with structural deformation [20]-[25] and adopted in many commercial software, such as ABAQUES, ANSYS, MSC.NASTRAN, etc. In present research, both linear and nonlinear deformations of baffles inside in the rolling tank will be investigated based on the FEM method.

In the FSI simulations, the coupling strategies between fluid solver and structural solver can be classified into two groups: the strong coupling approach and the weak coupling approach. In the strong coupling approach, a single system equation involving all variables related to both the fluid and structure dynamics is solved simultaneously [26]. However, the equation is much difficulty to form without any modification for complex engineering problems [27] and much expensive to be solved [28]. On the contrary, the fluid and structure fields are self-governed by different equations and solved separately in the weak coupling approach. Interfacial information communicates explicitly between the fluid and structure solution. This approach allows the use of separated fluid and structure codes or established software for each computational domain [23], and it is suitable to deal with engineering problems with large deformation. Hence, the weak coupling approach is utilized in the present paper.

The main object of this study is to develop a MPS-FEM coupled method which can be applied in nonlinear FSI problems, such as the interaction between sloshing flow in a rolling tank and elastic structure. The paper is organized as follows. Firstly, the MPS method is briefly reviewed. Next, the FEM method and the coupling strategy are described. Accuracy of the structure solver is validated by two benchmarks of dynamic oscillating beams. Then, the MPS-FEM coupled solver is applied to the problem of liquid sloshing in a tank interacting with baffles which will deform nonlinearly. Accuracy of the proposed method are verified by comparison against experimental data and simulation data from Idelsohn et al [21].

Numerical methods

In present study, the fluid domain is calculated by our in-house particle solver MLParticle-SJTU based on improved MPS method. Details about the improvements and validation of the solver can be find in the published literatures [1][11][17][18]. In this section, a brief review about the structure solver and the MPS-FEM coupling strategy is described as fellow.

Structure solver based on FEM

Based on Hamilton's principle, deformation of structure should satisfy

$$\delta H = 0, \quad H = \int_{t_1}^{t_2} [T - \Pi_s + \Pi_p] \mathrm{d}t \tag{1}$$

where T is the kinetic energy, Π_s is the strain energy, Π_p is the potential energy of external force and damping force.

According to previous literatures [29], the structural dynamic equations, which governing the motion of structural elements, can be derived from Eq. (1) and expressed as

$$\mathbf{M}\,\ddot{\boldsymbol{y}} + \mathbf{C}\,\dot{\boldsymbol{y}} + \mathbf{K}\,\boldsymbol{y} = \boldsymbol{F}(t) \tag{2}$$

$$\mathbf{C} = \alpha_1 \mathbf{M} + \alpha_2 \mathbf{K} \tag{3}$$

where **M**, **C**, **K** are the mass matrix, the Rayleigh damping matrix, the stiffness matrix of the structure, respectively. ***F*** is the external force vector acting on structure, and varies with computational time. ***y*** is the displacement vector of structure. α_1 and α_2 are coefficients which are related with natural frequencies and damping ratios of structure.

To solve the structural dynamic equation, another two group functions should be supplemented to set up a closed-form equation system. Here, Taylor's expansions of velocity and displacement developed by Newmark [30] are employed:

$$\dot{\boldsymbol{y}}_{t+\Delta t} = \dot{\boldsymbol{y}}_t + (1-\gamma)\ddot{\boldsymbol{y}}_t \Delta t + \gamma \ddot{\boldsymbol{y}}_{t+\Delta t}\Delta t \quad , \quad 0 < \gamma < 1 \tag{4}$$

$$\boldsymbol{y}_{t+\Delta t} = \boldsymbol{y}_t + \dot{\boldsymbol{y}}_t \Delta t + \frac{1-2\beta}{2}\ddot{\boldsymbol{y}}_t \Delta t^2 + \beta \ddot{\boldsymbol{y}}_{t+\Delta t}\Delta t^2 \quad , \quad 0 < \beta < 1 \tag{5}$$

where β and γ are important parameters of the Newmark method, and selected as β=0.25, γ=0.5 for all simulations in present paper. From Eq. (2-5), the displacement at t=t+Δt can be solved by the following formula [31]:

$$\bar{\mathbf{K}}\,\boldsymbol{y}_{t+\Delta t} = \bar{\boldsymbol{F}}_{t+\Delta t} \tag{6-a}$$

$$\bar{\mathbf{K}} = \mathbf{K} + a_0\mathbf{M} + a_1\mathbf{C} \tag{6-b}$$

$$\bar{\boldsymbol{F}}_{t+\Delta t} = \boldsymbol{F}_t + \mathbf{M}(a_0\boldsymbol{y}_t + a_2\dot{\boldsymbol{y}}_t + a_3\ddot{\boldsymbol{y}}_t) + \mathbf{C}(a_1\boldsymbol{y}_t + a_4\dot{\boldsymbol{y}}_t + a_5\ddot{\boldsymbol{y}}_t) \tag{6-c}$$

$$\begin{aligned} &a_0 = \frac{1}{\beta\Delta t^2},\ a_1 = \frac{\gamma}{\beta\Delta t},\ a_2 = \frac{1}{\beta\Delta t},\ a_3 = \frac{1}{2\beta} - 1, a_4 = \frac{\gamma}{\beta} - 1, \\ &a_5 = \frac{\Delta t}{2}\left(\frac{\gamma}{\beta} - 2\right),\ a_6 = \Delta t(1-\gamma),\ a_7 = \gamma\Delta t \end{aligned} \tag{6-d}$$

where $\bar{\mathbf{K}}$ and $\bar{\boldsymbol{F}}$ are so-called effective stiffness matrix and effective force vector, respectively. Finally, the accelerations and velocities corresponding to the next time step are updated as follows.

$$\ddot{\boldsymbol{y}}_{t+\Delta t} = a_0(\boldsymbol{y}_{t+\Delta t} - \boldsymbol{y}_t) - a_2\dot{\boldsymbol{y}}_t - a_3\ddot{\boldsymbol{y}}_t \tag{7}$$

$$\dot{\boldsymbol{y}}_{t+\Delta t} = \dot{\boldsymbol{y}}_t + a_6\ddot{\boldsymbol{y}}_t + a_7\ddot{\boldsymbol{y}}_{t+\Delta t} \tag{8}$$

To validate the accuracy of present structural solver, two test cases are carried out. In the first case, response of the undamped cantilever beam under a ramp-infinite duration load is studied. The sketches of beam geometry and load history are shown as Fig. 1 and Fig. 2. The Young's modulus, density, moment of inertia, and cross area of the structure are 30×10^6 psi, 4.567×10^{-3} lb s^2/in^4, 100 in^4 and 21.9 in^2, respectively. Time history about the displacement at the tip of the undamped cantilever is shown as Fig. 3. According to the comparison between present result and Behdinan's data [34], good agreement can be achieved.

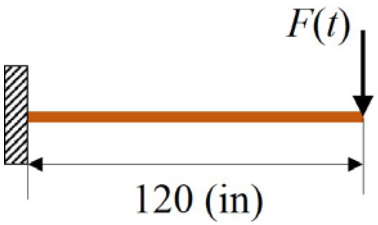


Figure 1. Beam geometry

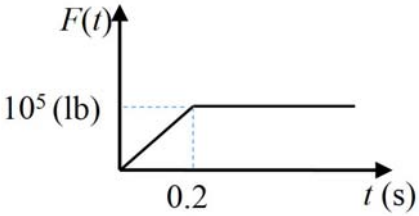


Figure 2.Load history of test 1

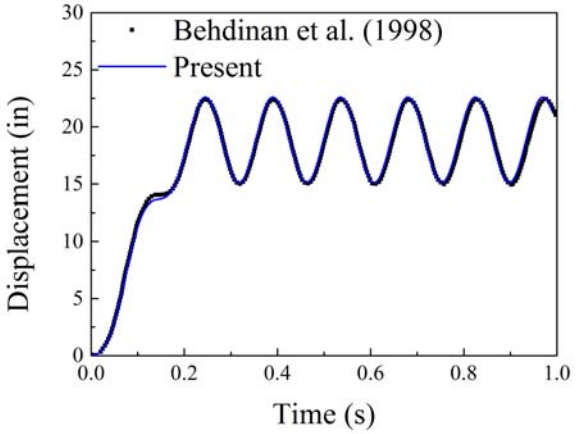


Figure 3. Time response of the tip (test 1)

In the FEM method, response of structure is obviously related to viscous damp coefficients. Hence, the second test case about the damped cantilever beam under a ramp-ramp duration load is studied. The sketches of beam geometry is same as that in test case 1 and shown as Fig. 1. The load history is shown as Fig. 4. The Young's modulus, density, moment of inertia, and cross area of the structure are all same as the first test case. However, the effect of damp is considered and the Rayleigh's coefficients are set $\alpha_1 = 0.0$, $\alpha_2 = 0.003$. Time history about the displacement at the tip of the damped cantilever is shown as Fig. 5. Present result and Behdinan's data are in good agreement. So, present structural solver is suitable to solve deformation of structure.

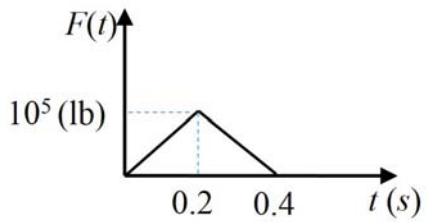


Figure 4. Load history of test 2

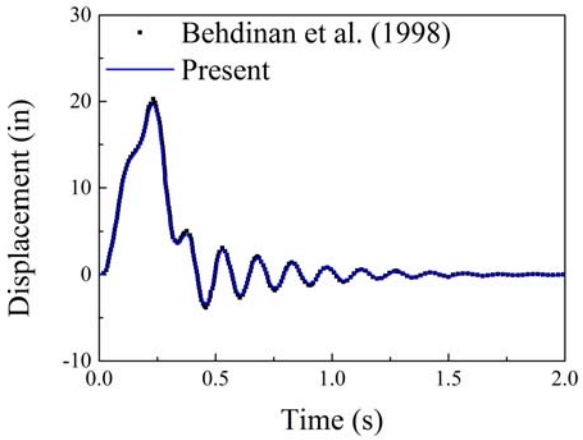


Figure 5. Time response of the tip (test 2)

MPS-FEM coupling strategy

In present study, the weak coupling between MPS and the FEM method is implemented. Flowchart of solution procedure is shown as Fig. 6. Sizes of time step for structure analysis and fluid analysis are Δt_s and Δt_f, respectively. Here, Δt_s is k multiples of Δt_f, where k is an integer. The procedure of interaction can be summarized as below.

(1) The fluid field would be calculate k times based on MPS method. Pressure of fluid wall boundary particle is calculated as follows:

$$\overline{p}_{n+1} = \frac{1}{k}\sum_{i=1}^{k} p_{n+i} \tag{9}$$

where p_{n+i} is pressure of the fluid particle on wall boundary at the instant $t + i\Delta t_f$, $\overline{p}_{n+1}$ is averaged pressures of fluid particle within Δt_s.

(2) Determine the values of structural nodal position y_t, velocity $\dot{y}_t$ and acceleration $\ddot{y}_t$ based on the results of previous time step.

(3) Calculate external force vector $\boldsymbol{F}_{t+\Delta t_s}$ of structural boundary particles based on pressure of fluid wall boundary particles $\overline{p}_{n+1}$.

(4) Calculate the new values of structural nodal displacements and velocities based on the Newmark method described in the previous section.

(5) Update velocity and position of both structural boundary particles and fluid particles.

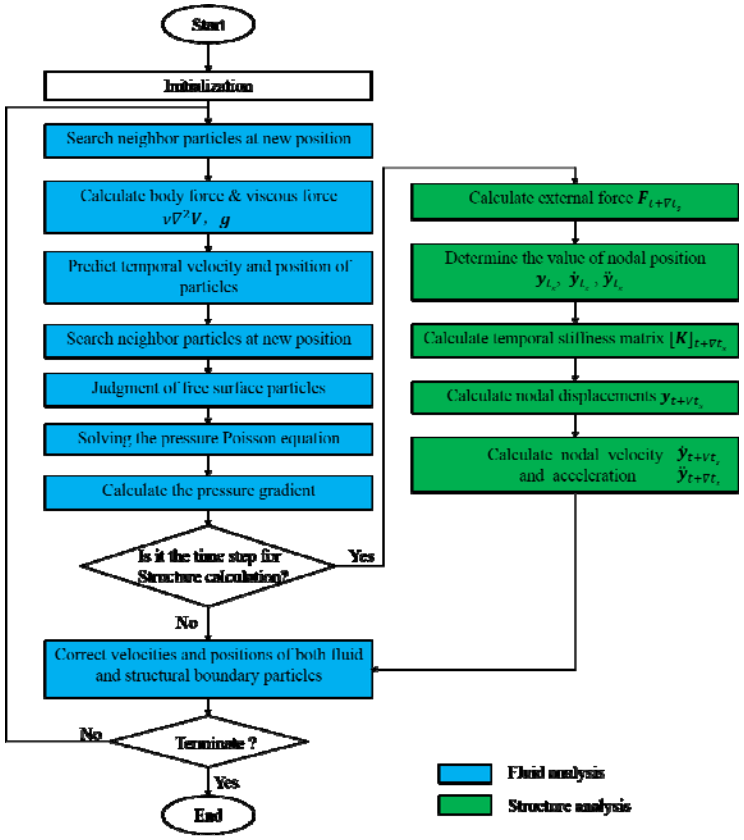


Figure 6. Flowchart of MPS-FEM coupling procedure

Numerical Simulations

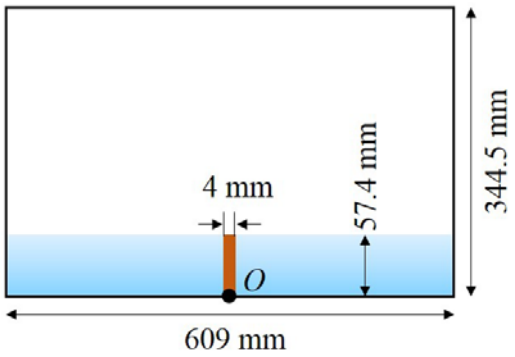
In present study, the MPS-FEM coupled method is used to simulate the interaction between sloshing flow and elastic structure in a 2D rolling tank. The experimental data published by Idelsohn et al. [21] and the numerical result published by Paik [32][33] are used for comparison study and validation of the capability of present numerical method.

Numerical setup

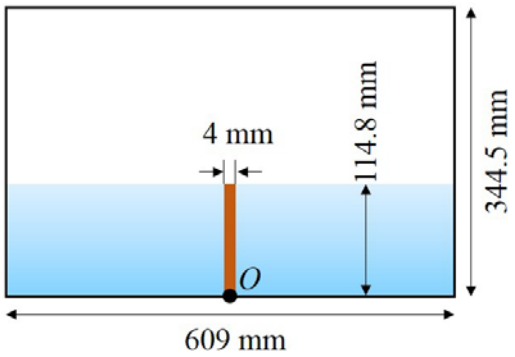
According to the experiments carried out by Idelsohn [21], three cases are numerically investigated in this paper. Elastic baffles are mounted at the bottom or top of the two-dimensional tank and related sketches about the geometry setup are shown as Fig. 7. The tank, with a length of 609 mm and a height of 344.5mm, is free to roll around the point O which is the center of bottom of the container. The tank is forced to roll harmoniously with the governing equation of motion defined as

$$\theta(t) = \theta_0 \sin(\omega t) \tag{10}$$

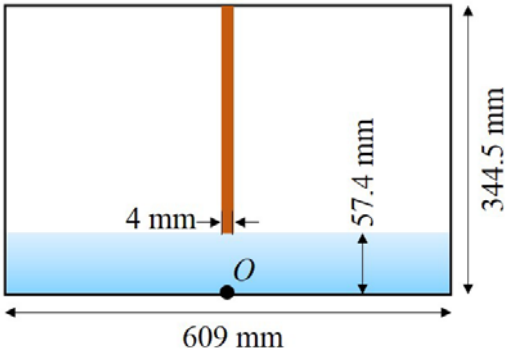
where $\theta(t)$ is the rotation angle of the tank, θ_0 is the excitation amplitude, ω is the angular frequency.



(a) Case1: Elastic beam interacting with shallow water flow



(b) Case2: Elastic beam interacting with deep water flow



(c) Case3: Hanging elastic beam interacting with shallow water flow

Figure 7. Sketches of the rolling tank with elastic beams

Table 1. Fluid parameters of numerical cases

Parameters	Case 1	Case 2	Case 3
Fluid density (kg/m^3)	917	917	998
Kinematic viscosity (m^2/s)	5×10^{-5}	5×10^{-5}	1×10^{-6}
Gravitational acceleration (s/m^2)	9.81	9.81	9.81
Fluid depth (mm)	57.4	114.8	57.4
Rolling frequency (Hz)	0.61	0.83	0.61
Rolling amplitude (degree)	4	4	2
Particle spacing (mm)	2	2	2
Time step size (s)	2×10^{-4}	2×10^{-4}	2×10^{-4}

Table 2. Structure parameters of numerical cases

Parameters	Case 1	Case 2	Case 3
Structure density (kg/m^3)	1100	1100	1900
Young's modulus (Pa)	6×10^6	6×10^6	4×10^6
Length (mm)	57.4	114.8	287.1
Clamped position	Bottom	Bottom	Top
Number of elements	29	58	145
Damping coefficients α1	0	0	0
Damping coefficients α2	0.05	0.025	0.025
Time step size (s)	2×10^{-3}	2×10^{-3}	2×10^{-3}

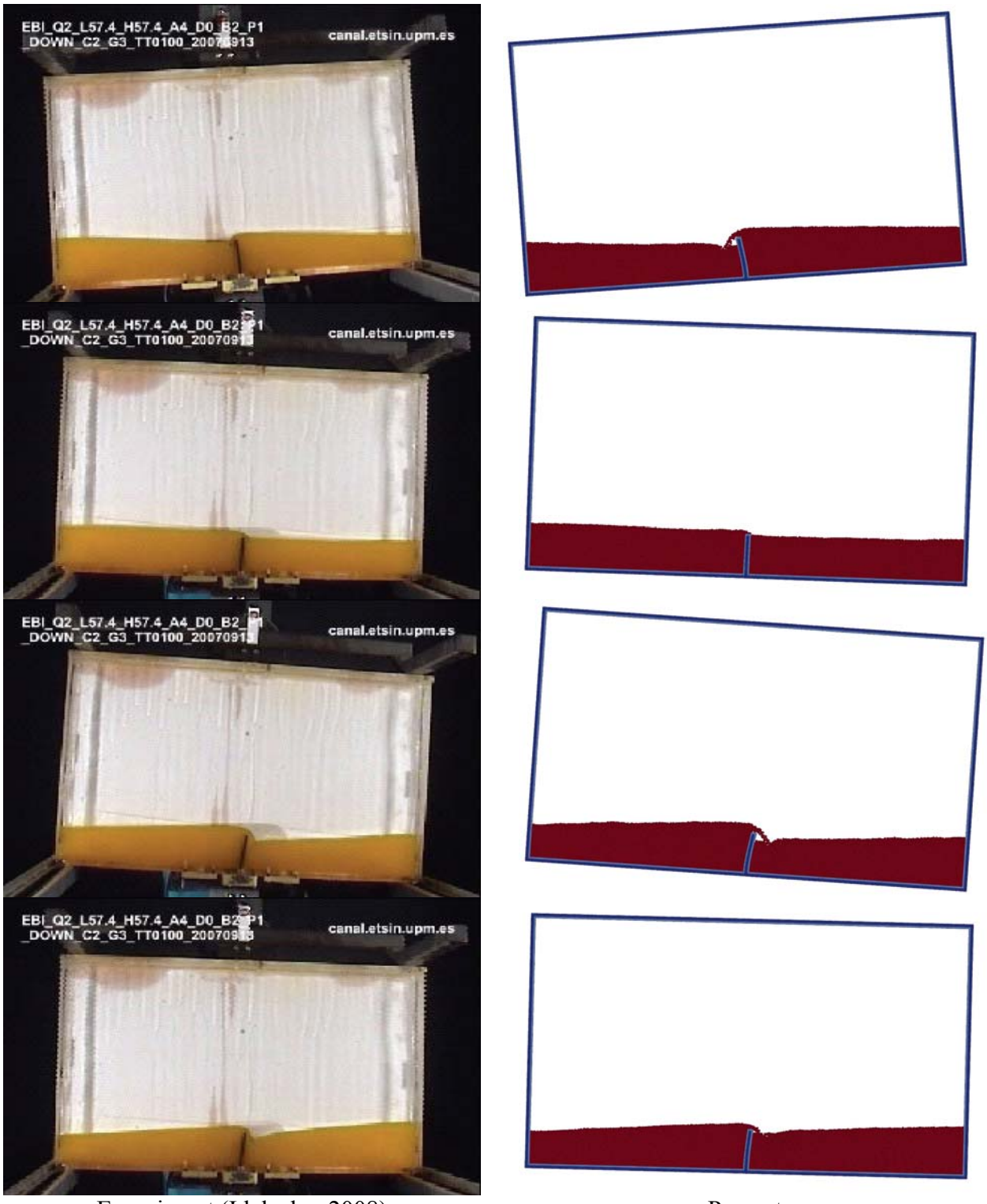
Elastic beam interacting with shallow water flow

In present case, the tank, rolling with the amplitude of 4 degrees and frequency of 0.61 Hz, is partially filled with fluid of 57.4 mm depth. Density and Kinematic viscosity are 917 kg/m^3 and 5×10^{-5} m^2/s, respectively. A short baffle is mounted at the rolling center point *O*. Length and width of the baffle are 57.4 mm and 4 mm. Density and the Young's modulus of the baffle are 1100 kg/m^3 and 6×10^6 Pa, respectively. The models of both fluid and structure are dispersed by particles with spacing of 2 mm. The baffle is simplified as a beam and dispersed by 29 elements. The coefficients of $\alpha_1 = 0.0$ and $\alpha_2 = 0.05$ are used to compose the structural Rayleigh damping matrix $\boldsymbol{C}$ which is an important part of the dynamic equations. The size of time steps is 0.0002 s for the calculation of fluid domain while that is 0.002 s for the structural domain.

Snapshots about deformation of baffle and elevation of free surface are shown in Fig. 8. Numerical data is compared with experiment at four instants, t=0.95, 1.35, 1.62, and 1.88 s. Profiles of the deformed baffle and free surface are coincident with that of experiment. However, a bubble cavity, which doesn't exist in the experiment, forms near the top of baffle while the fluid flows over the structure in present simulation. As mentioned in Paik et al. [33], the possible reason about the babble cavity is the three dimensional nature that the channel is open and air is able to escape for the real flow. Generally, the agreement between the numerical results and the experimental ones are acceptable.

Time histories of the horizontal displacement at the top tip of baffle are shown as Fig. 9. Present numerical result based on MPS-FEM method is compared with experimental data of Idelsohn [21] and simulation results from both Idelsohn and Paik [33]. The trend of numerical curve evolves harmonically and with a period similar to experiment. Though the amplitude of

present numerical curve is larger than experiment, it's similar to the simulation results published by Paik et al. [33].



Experiment (Idelsohn, 2008) Present

Figure 8. Deformation of baffle and elevation of free surface for Case 1: t=0.95, 1.35, 1.62, and 1.88 s.

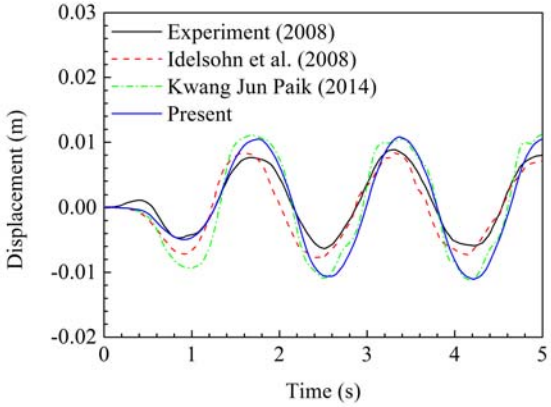


Figure 9. Comparison of the horizontal displacement at the tip of baffle (Case 1)

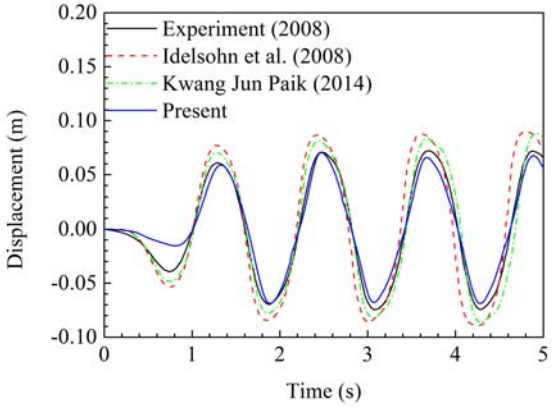


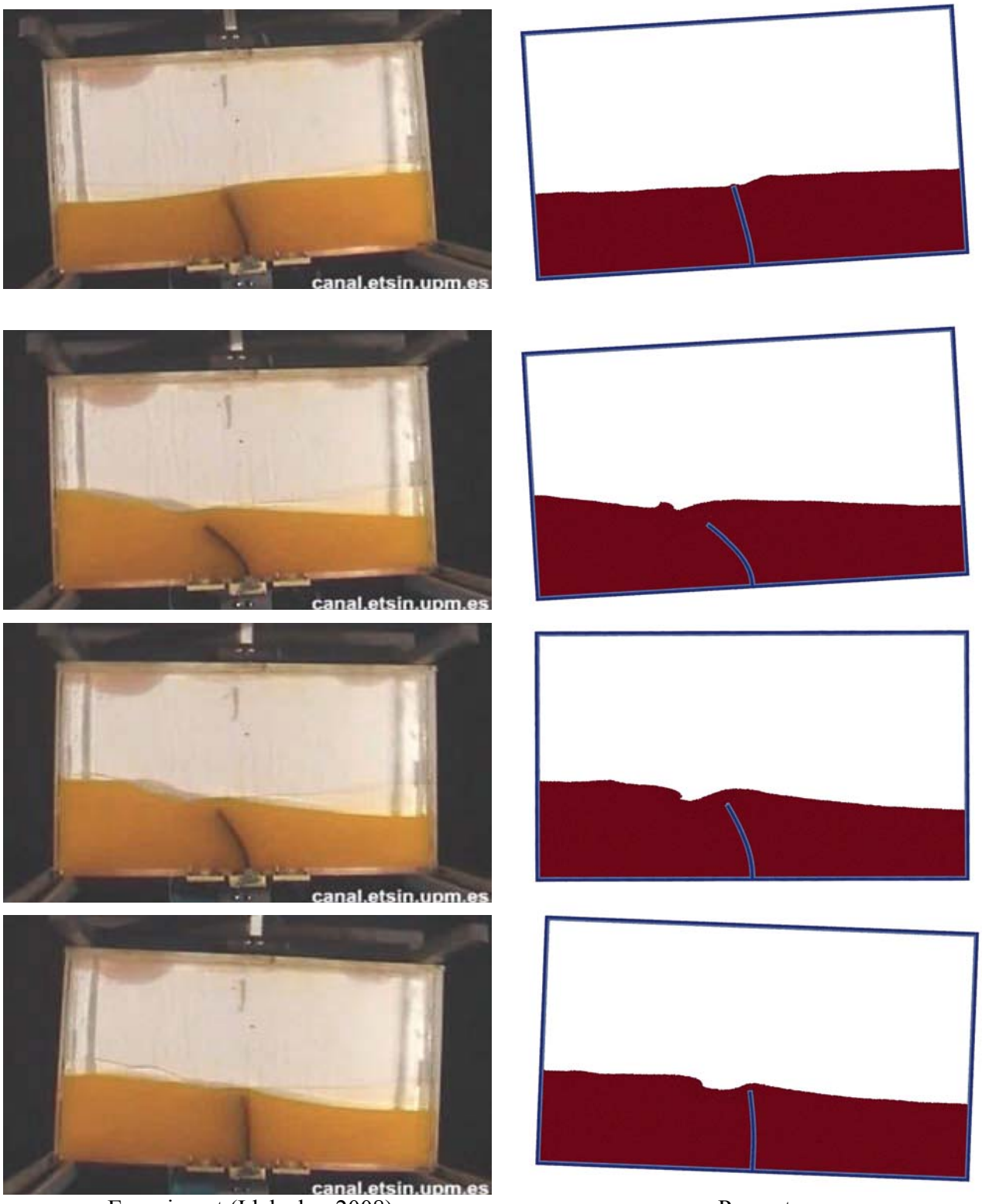
Figure 10. Comparison of the horizontal displacement at the tip of baffle (Case 2)

Elastic beam interacting with deep water flow

In present case, most parameters of simulation are same as that of Case 1. The tank rotates with an amplitude same as that of case 1 but a higher frequency of 0.83 Hz. Level of fluid filled in the tank is twice the depth of case 1. A longer baffle with the length of 114.8 mm is also mounted at the rolling center. The baffle is dispersed by 58 beam elements. The coefficients of $\alpha_1 = 0.0$ and $\alpha_2 = 0.025$ are used in this case. Detailed parameters of the simulation are shown in table 1 and table 2.

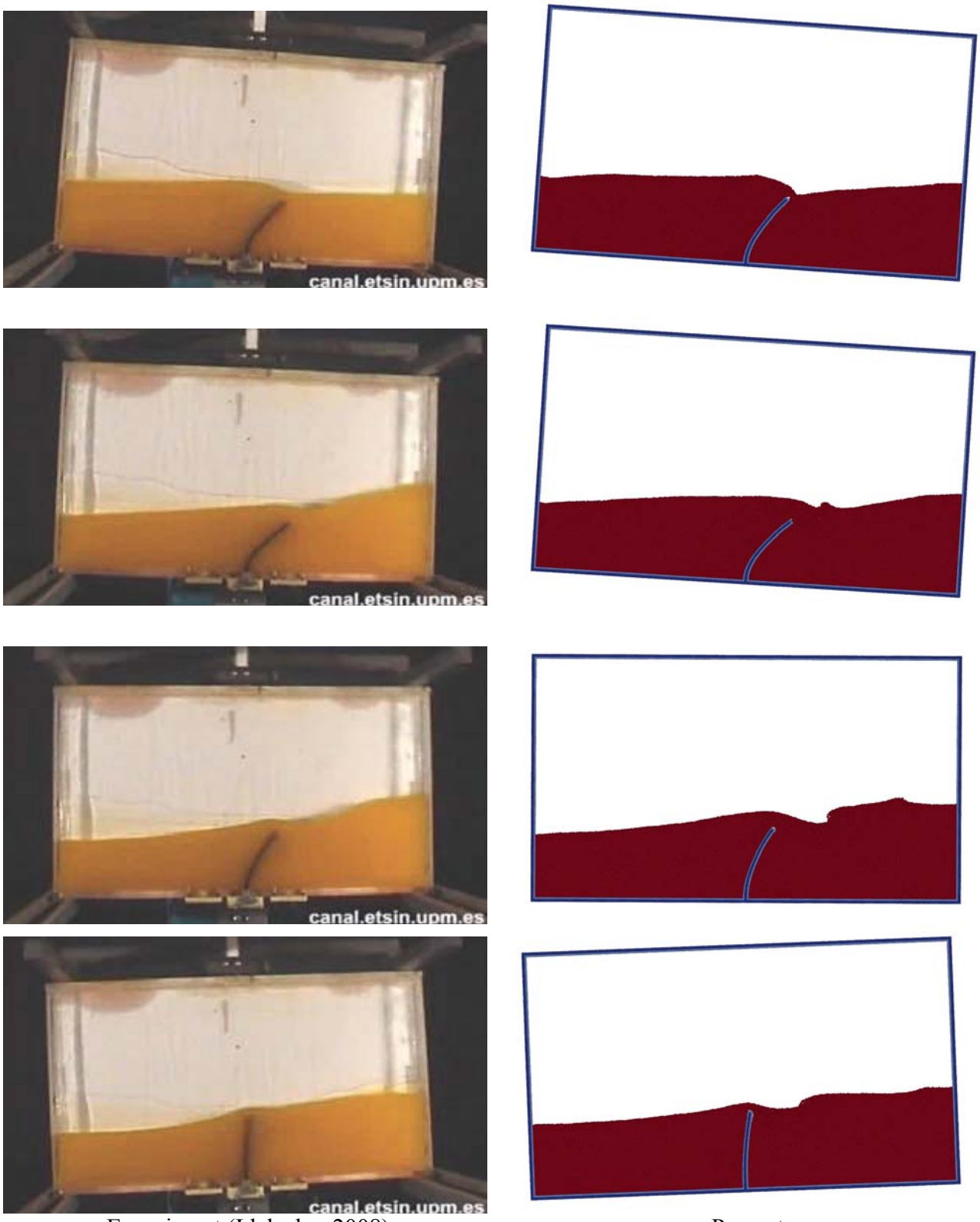
Fig. 10 shows the comparison of time histories of the horizontal displacement at the top tip of baffle. Forms of the curves are similar to those in previous case but with larger amplitudes due to a much deeper fluid filled in the tank. According to the figure, both amplitude and period are in good agreement with experimental data.

Snapshots about deformation of the baffle and elevation of free surface are shown in Fig. 11. Numerical data is compared with experiment at eight instants, t=1.69, 1.96, 2.09, 2.23, 2.36, 2.56, 2.69, 2.83 s. The baffle deforms obviously and keep submerged after the instant t=1.69 s. Though the interaction between fluid and the elastic baffle is very strong, both numerical shapes of baffle and free surface are in good agreement with experiment.



Experiment (Idelsohn, 2008) Present

Figure 11. Deformation of baffle and elevation of free surface for Case 2: t=1.69, 1.96, 2.09, 2.23, 2.36, 2.56, 2.69, 2.83 seconds



Experiment (Idelsohn, 2008) Present

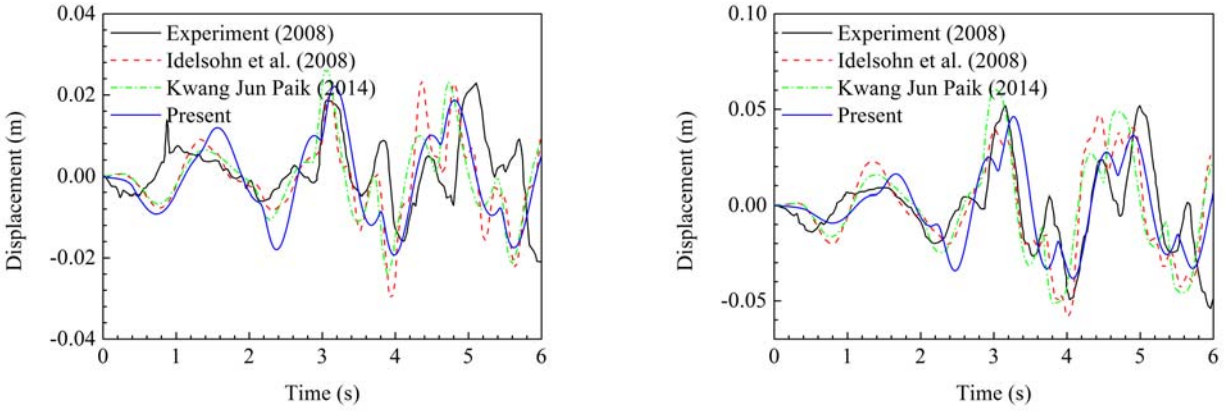
Figure 11. Continued

Hanging elastic beam interacting with shallow water flow

This case is much different from Cases 1 and 2. Unlike the arrangements of baffles in previous two cases, the longest baffle is hanging at the top of tank and the end tip reaches to the surface of fluid. So, the deformation of baffle is only caused by the impact force of free surface waves. In this case, the tank is forced to roll with the amplitude of 2 degrees and the frequency of 0.61 Hz. Level of fluid is same as that in case 1. Density and Kinematic viscosity are 998 kg/m^3 and 1×10^{-6} m^2/s, respectively. The baffle is dispersed by 145 beam elements. Density and the Young's modulus of the baffle are 1900 kg/m^3 and 4×10^6 Pa, respectively. The coefficients of $\alpha_1 = 0.0$ and $\alpha_2 = 0.025$ are used in this case. Detailed parameters of the simulation are shown in table 1 and table 2.

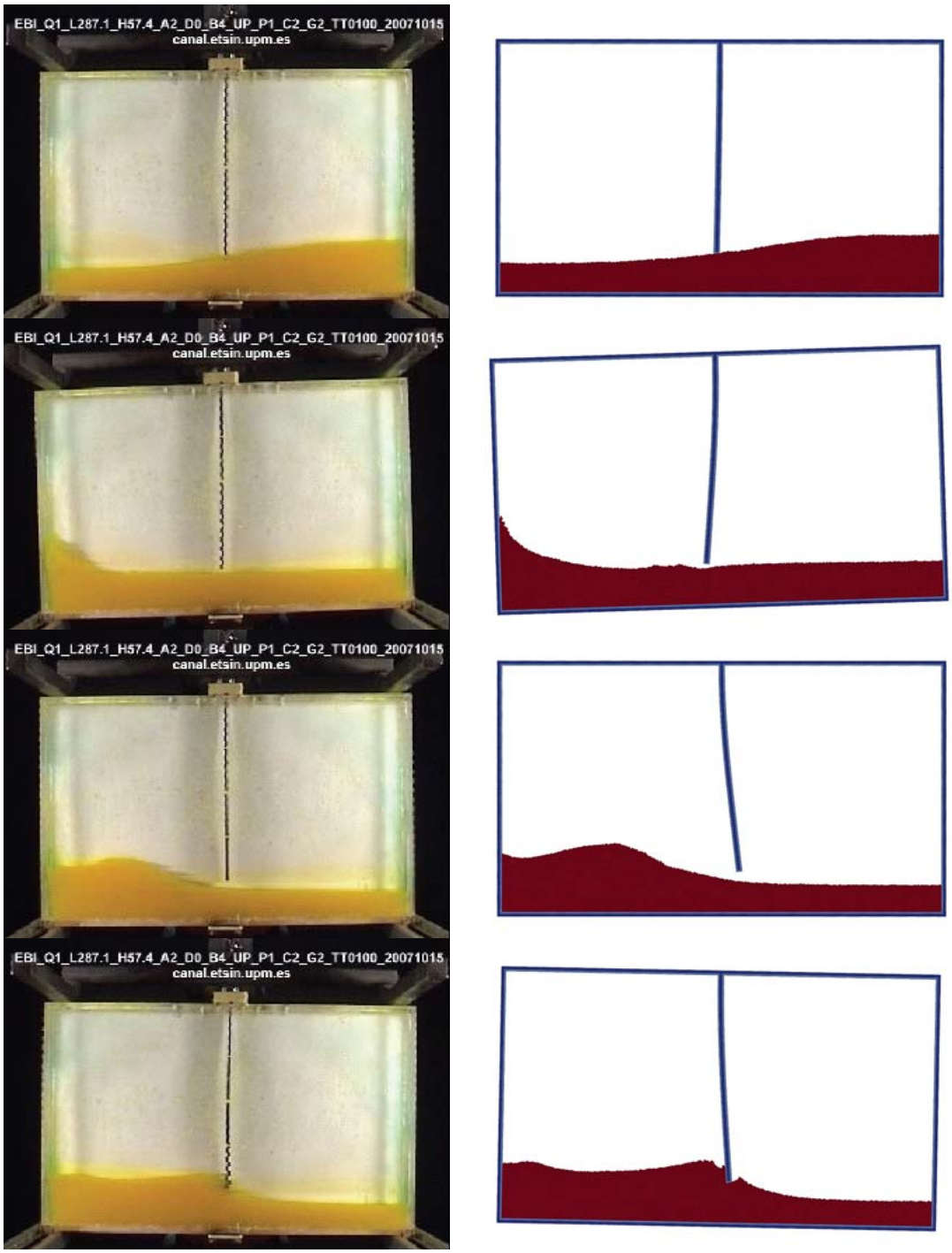
Fig. 12 shows the comparison of time histories of the horizontal displacement at the middle and end tip of baffle. According to both experimental and numerical data, deformation of the baffle is highly nonlinear. High frequency oscillation is observed after t=2 s for both middle and tip of the baffle. Though it's much more challenging to obtain the accurate solution, the agreement between present result and experiment is acceptable.

Snapshots about deformation of baffle and elevation of free surface are shown in Fig. 13. Numerical data is compared with experiment at nine instants, t=1.95, 2.42, 2.69, 2.82, 3.02, 3.29, 3.69, 3.89, 4.09 s. Both numerical shapes of baffle and free surface are quite similar to experiment results during the whole process of wave propagation. However, spray around the tip of baffle, caused by the impact between baffle and wave crest, exists at the instances 3.02 and 3.89 s. This phenomenon is not obviously observed from the experimental figures. Possible reasons for the discrepancy between present results and the experiment could be the three dimensional characters. Besides, the effect of rough boundary of the elastic baffle in experiment shouldn't be neglected.



(a) Displacement at the middle of baffle (b) Displacement at the tip of baffle

Figure 12. Comparison of the horizontal displacement (Case 3)



Experiment (Idelsohn, 2008) Present

Figure 13. pressure contours (middle) and velocity vectors (right) for Case 3: t=1.95, 2.42, 2.69, 2.82, 3.02, 3.29, 3.69, 3.89, 4.09 seconds (Continued)

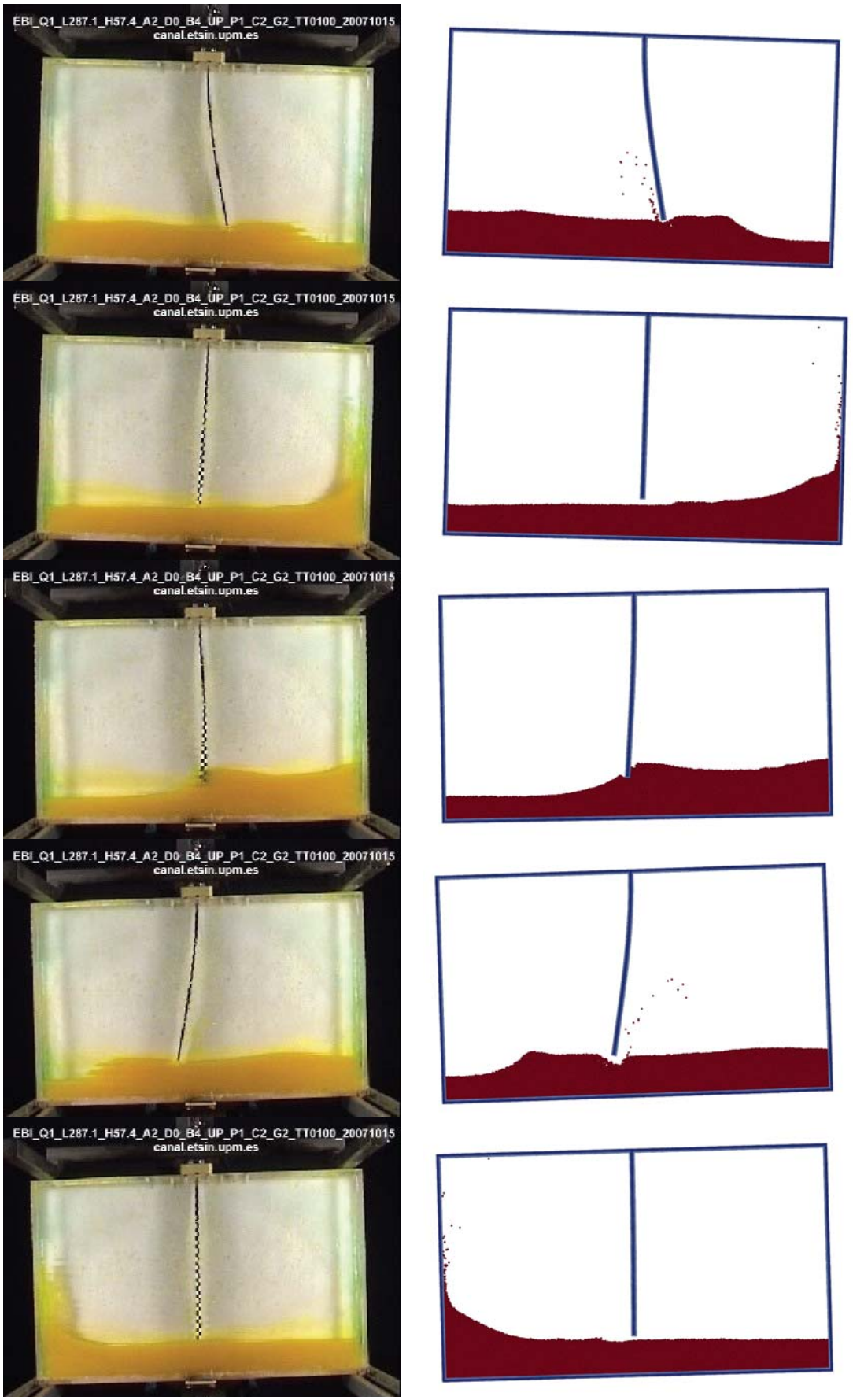


Figure 13. Continued

Conclusions

The aim of this paper is to develop a MPS-FEM coupled method for fluid structure interaction problems and validate the capability of this method. Mathematical equations for the MPS and FEM method, together with the coupling strategy, are described firstly. According to two dynamic tests, the proposed structural solver is accurate enough for structural deformation problems. Then, the FSI problems of sloshing with elastic baffles are numerically studied by the MPS-FEM coupled method. Deformations of the baffles, include the linear and nonlinear responses, are quite coincident between present numerical results and experiment. Present numerical results show that the proposed MPS-FEM coupled method is capable of simulating problems about structural deformation interaction with violent free surface flow.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (51379125, 51490675, 11432009, 51579145, 11272120), Chang Jiang Scholars Program (T2014099), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

References

[1] Zhang, Y. X., Wan, D. C., Hino, T. (2014) Comparative study of MPS method and level-set method for sloshing flows, *Journal of hydrodynamics* **26**(4), 577-585.

[2] Zhang, Y. L., Tang, Z. Y., and Wan, D. C. (2016) Numerical Investigations of Waves Interacting with Free Rolling Body by Modified MPS Method, *International Journal of Computational Methods* **13**(4), 1641013.

[3] Koshizuka, S., and Oka, Y. (1996) Moving particle Semi-implicit Method for Fragmentation of Incompressible Fluid, *Nuclear Science and Engineering* **123**, 421-434.

[4] Khayyer, A., and Gotoh, H. (2009) Modified Moving Particle Semi-implicit methods for the prediction of 2D wave impact pressure, *Coastal Engineering* **56**, 419-440.

[5] Khayyer, A., and Gotoh, H. (2010) A higher order Laplacian model for enhancement and stabilization of pressure calculation by the MPS method, *Applied Ocean Research* **32**, 124-131.

[6] Khayyer, A., and Gotoh, H. (2011) Enhancement of stability and accuracy of the moving particle semi-implicit method, *Journal of Computational Physics* **230**, 3093-3118.

[7] Khayyer, A., and Gotoh, H. (2012) A 3D higher order Laplacian model for enhancement and stabilization of pressure calculation in 3D MPS-based simulations, *Applied Ocean Research* **37**, 120-126.

[8] Kondo, M., and Koshizuka, S. (2011) Improvement of stability in moving particle semi-implicit method, *Int. J. Numer. Meth. Fluids* **65**, 638-654.

[9] Tanaka, M., and Masunaga, T. (2010) Stabilization and smoothing of pressure in MPS method by Quasi-Compressibility, *Journal of Computational Physics* **229**, 4279-4290.

[10] Ikari, H., Khayyer, A., and Gotoh. H. (2015) Corrected higher order Laplacian for enhancement of pressure calculation by projection-based particle methods with applications in ocean engineering, *J. Ocean Eng. Mar. Energy* **1**(4), 361-376.

[11] Zhang, Y. X., and Wan, D. C. (2011) Application of MPS in 3D Dam Breaking Flows, *Sci. Sin. Phys. Mech. Astron.* **41**, 140-154.

[12] Lee, B. H., Park, J. C., Kim, M. H., Jung, S. J., Ryu, M. C., and Kim, Y. S. (2010). Numerical simulation of impact loads using a particle method, *Ocean Engineering* **37**, 164-173.

[13] Yokoyama, M., Kubota, Y., Kikuchi, K., Yagawa, G., and Mochizuki, O. (2014). Some remarks on surface conditions of solid body plunging into water with particle method, *Advanced Modeling and Simulation in Engineering Sciences* **1**(1), 1-14.

[14] Hwang, S. H., Khayyer, A., Gotoh, H., and Park, J. C. (2015). *Simulations of Incompressible Fluid Flow-Elastic Structure Interactions by a Coupled Fully Lagrangian Solver*, Proc 25th Int Offshore and Polar Eng Conf, Hawaii, ISOPE, 1247-1250.

[15] Koshizuka, S., Shibata, K., Tanaka, M. and Suzuki, Y. (2007) *Numerical analysis of fluid-structure and fluid-rigid body interactions using a particle method*, Proceedings of FEDSM2007, San Diego, California USA.
[16] Sueyoshi, M., Kashiwagi, M., and Naito, S. (2008) Numerical simulation of wave-induced nonlinear motions of a two-dimensional floating body by the moving particle semi-implicit method, *J. Mar. Sci. Technol* **13**, 85-94.
[17] Zhang, Y. X., and Wan, D. C. (2012) *Apply MPS Method to Simulate Liquid Sloshing in LNG Tank*, Proc. 22nd Int. Offshore and Polar Eng. Conf., Rhodes, Greece, 381-391.
[18] Tang, Z. Y., and Wan, D.C. (2015) Numerical simulation of impinging jet flows by modified MPS method, *Engineering Computations* **32**(4), 1153-1171.
[19] Sun, Z., Xing, J. T., Djidjeli, K. and Cheng, F. (2015) *Coupling MPS and Modal Superposition Method for Flexible Wedge Dropping Simulation*, Proceedings of the Twenty-fifth International Ocean and Polar Engineering Conference, Kona, Big Island, Hawaii, USA, 144-151.
[20] Onate, E., Idelsohn, S. R., Celigueta, M. A., Rossi, R. (2006) *Advances in the particle finite element method for fluid-structure interaction problems*, In: Proceedings of 1st South-East European Conference on Computational Mechanics, SEECCM-06, Kragujevac, Serbia and Montenegro.
[21] Idelsohn, S. R., Marti, J., Limache, A. and Onate, E. (2008) Unified Lagrangian formulation for elastic solids and incompressible fluids: Application to fluid-structure interaction problems via the PFEM, *Comput. Methods Appl. Mech. Eng* **197**, 1762-1776.
[22] Ryzhakov, P. B., Rossi, R., Idelsohn, S. R. and Onate, E. (2010) A monolithic Lagrangian approach for fluid-structure interaction problems, *Comput. Mech* **46**, 883-899.
[23] Liao, K. P., Hu, C. H. (2013) A coupled FDM-FEM method for free surface flow interaction with thin elastic plate, *J. Mar. Sci. Technol* **18**, 1-11.
[24] Mitsume, N., Yoshimura, S., Murotani, K., Yamada, T. (2014a) MPS-FEM partitioned coupling approach for fluid-structure interaction with free surface flow, *International Journal of Computational Methods* **11**(4), 4157-4160.
[25] Mitsume, N., Yoshimura, S., Murotani, K., Yamada, T. (2014b) Improved MPS-FE Fluid-Structure Interaction Coupled Method with MPS Polygon Wall Boundary Model, *Comput. Model. Eng. Sci* **101**(4), 229-247.
[26] Hou, G., Wang, J., Layton, A. (2012) Numerical Methods for Fluid-Structure Interaction - A Review, *Commun. Comput. Phys* **12**(2), 337-377.
[27] Longatte, E., Verremana, V., Souli, M. (2009) Time marching for simulation of fluid-structure interaction problems, *Journal of Fluids and Structures* **25**, 95-111.
[28] Heil, M., Hazel, A. L., Boyle, J. (2008) Solvers for large-displacement fluid-structure interaction problems: segregated versus monolithic approaches, *Computational Mechanics* **43**(1), 91-101.
[29] Iura, M., Atluri, S. N. (1995) Dynamic analysis of planar flexible beams with finite rotations by using inertial and rotating frames, *Computers and Structures* **55**(3), 453-462.
[30] Newmark, N. M. (1959) A method of computation for structural dynamics, *Journal of the engineering mechanics division* **85**(3), 67-94.
[31] Hsiao, K. M., Lin, J. Y., Lin, W. Y. (1999) A consistent co-rotational finite element formulation for geometrically nonlinear dynamic analysis of 3-D beams, *Comput. Methods Appl. Mech. Engrg* **169**, 1-18.
[32] Paik, K. J. (2010) Simulation of fluid-structure interaction for surface ships with linear/nonlinear deformations, University of Iowa, thesis for PhD degree.
[33] Paik, K. J., and Carrica, P. M. (2014) Fluid-structure interaction for an elastic structure interacting with free surface in a rolling tank, *Ocean Engineering* **84**, 201-212.
[34] Behdinan, K., Stylianou, M. C., Tabarrok, B. (1998) Co-rotational dynamic analysis of flexible beams, *Comput. Methods Appl. Mech. Eng* **154**, 151-161.

A novel immersed boundary method for the strongly coupled fluid-structure interaction

*,†**Shang-Gui Cai**[1] **and Abdellatif Ouahsine**[1]

[1]Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7337 Roberval, Centre de Recherche Royallieu, CS 60319, 60203 Compiègne Cedex, France

*Presenting author: shanggui.cai@utc.fr

†Corresponding author: shanggui.cai@utc.fr

ABSTRACT

In the present work a novel non-body conforming mesh method, termed as the moving immersed boundary method, is proposed for the strongly coupled fluid-structure interaction. The immersed boundary method enables solids of complex shape to move arbitrarily in an incompressible viscous fluid, without fitting the solid boundary motion with dynamic meshes. A boundary force is usually employed to impose the no-slip boundary condition at the solid surface. In the novel method, an additional equation is derived to compute the boundary force implicitly. The coefficient matrix is formulated to be symmetric and positive-definite, so that the conjugate gradient method can solve the resulting system very efficiently. The current immersed boundary solver is integrated into the fluid projection method as another operator splitting. Finally an efficient fixed point iteration scheme is constructed for the strongly coupled fluid-structure interaction.

Keywords: Immersed boundary method, Fluid-structure interaction, Strongly coupled algorithm, Projection method, Fractional step method.

Introduction

The fluid-structure interaction (FSI) is of great importance in many scientific and engineering fields. The difficulties of its numerical simulation lie in the facts that the interaction interface is often complicated, time-dependent and the two physical domains are strongly coupled. The FSI problem has been extensively studied in the past with body-conforming mesh methods, such as the arbitrary Lagrangian-Eulerian (ALE) method, where the mesh is deformed or renewed in order to fit the novel interface (e.g. [1]). This procedure however is usually time-consuming and it is very difficult to maintain the mesh quality when solids undergo large displacements.

The immersed boundary method (IBM) emerged in 1970s by the work of Peskin [8] as an effective tool to circumvent the dynamic mesh issues. A boundary force is introduced to the fluid momentum equation to account for the solid effects, hence the fluid equations are solved on a fixed Eulerian grid. The original method is developed for the simulation of blood flow over an elastic beating heart. Its direct extension to rigid boundary poses a lot of difficulties, since the stiffness value approaches infinity. The time step is also kept very small in order to maintain the stability. This method has been successfully extended to moving rigid bodies by the work of Uhlmann [9] by using the direct forcing concept of [3]. No artificial constants and additional time constraint are introduced for the rigid body formulation. However, fully explicit schemes were adopted for the force evaluation and the interface coupling in [9]. Consequently, the no-slip boundary condition is never satisfied and the calculation will not be stable when the solid density is smaller or even close to the fluid density ($\rho_s/\rho_f \lesssim 1.05$ for

circular disks as reported in [9]).

Therefore, implicit schemes should be considered for obtaining accurate and stable results. In this work we extend the implicit immersed boundary method of [2] to two-way fluid-structure interactions in the next section. We will demonstrate the stability and the accuracy of present scheme in the numerical examples.

Numerical method

Governing equations

In the present study, we consider the rigid body motion in an incompressible fluid. The fluid-structure interaction problem is illustrated in Figure 1, where the fluid and the rigid body occupy the domain Ω_f and Ω_s respectively. The interaction takes place at the their common boundary $\partial\Omega_i = \Omega_f \cap \Omega_s$. The whole system is subjected to the gravitational acceleration $\mathbf{g}$.

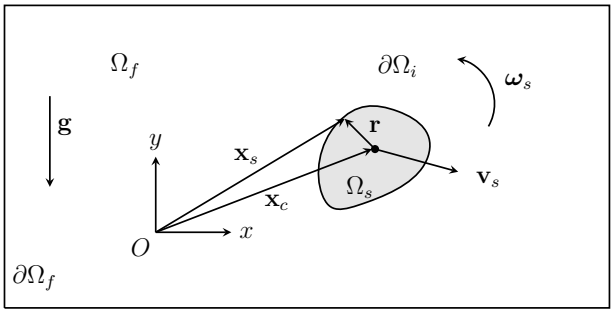


Figure 1: Sketch of the fluid-structure interaction problem.

The fluid motion is governed by the Navier-Stokes equations

$$\frac{\partial \mathbf{v}_f}{\partial t} + \nabla \cdot (\mathbf{v}_f \otimes \mathbf{v}_f) = \nabla \cdot \boldsymbol{\sigma}_f + \mathbf{g} \tag{1a}$$

$$\nabla \cdot \mathbf{v}_f = 0 \tag{1b}$$

where $\mathbf{v}_f$ is the fluid velocity vector and the fluid stress tensor $\boldsymbol{\sigma}_f$ is given by

$$\boldsymbol{\sigma}_f = -\frac{p}{\rho_f}\mathbf{I} + \nu(\nabla \mathbf{v}_f + (\nabla \mathbf{v}_f)^{\mathrm{T}}) \tag{1c}$$

where p is the fluid pressure, ρ_f the fluid density, ν the fluid kinematic viscosity. Appropriate initial and boundary conditions are assumed to the fluid Navier-Stokes equations to ensure that the problem is well posed.

The rigid body motion is governed by the Newton-Euler equations

$$m_s \frac{d\mathbf{v}_s}{dt} = \rho_f \int_{\partial\Omega_i} \boldsymbol{\sigma}_f \cdot \mathbf{n} ds + m_s(1 - \frac{\rho_f}{\rho_s})\mathbf{g} \tag{2a}$$

$$I_s \frac{d\boldsymbol{\omega}_s}{dt} = \rho_f \int_{\partial\Omega_i} \mathbf{r} \times \left(\boldsymbol{\sigma}_f \cdot \mathbf{n}\right) ds \tag{2b}$$

where m_s, ρ_s, I_s represent the solid mass, the solid density and the moment of inertia respectively. $\mathbf{v}_s$, $\boldsymbol{\omega}_s$ designate the translational velocity and the angular velocity of the solid. $\mathbf{r} = \mathbf{x}_s - \mathbf{x}_c$ is the position

vector of the surface point with respect to the solid mass center, where $\mathbf{x}_s$ is the solid position vector at the surface and $\mathbf{x}_c$ is the solid gravity center vector (see Figure 1). $\mathbf{n}$ represents the outward-pointing normal vector to the surface $\partial\Omega_i$. The position of the rigid body can be obtained by the integration of the following kinematic equations

$$\frac{d\mathbf{x}_c}{dt} = \mathbf{v}_s \tag{3a}$$

$$\frac{d\boldsymbol{\theta}_c}{dt} = \boldsymbol{\omega}_s \tag{3b}$$

where $\boldsymbol{\theta}_c$ designates the rotation angle around the solid mass center.

On the fluid-structure interface $\partial\Omega_i$ the following no-slip boundary condition

$$\mathbf{v}_f = \mathbf{v}_s + \boldsymbol{\omega}_s \times \mathbf{r} \tag{4}$$

needs to be satisfied in order to take the fluid-structure interaction into account.

The immersed boundary method approximates the above fluid-structure interaction problem by replacing the solid domain with the surrounding fluid. To account for the presence of the immersed solid, a boundary force $\mathbf{f}$ is introduced and added into the fluid momentum equation. Therefore the fluid is simply simulated in a fixed domain $\overline{\Omega} = \Omega_f(t) \cup \Omega_s(t)$ irrespective to the movement of the immersed solid. Following Glowinski *et al.* [4], we write the entire fluid-structure interaction problem in the immersed boundary formulation as

$$\frac{\partial \mathbf{v}_f}{\partial t} + \nabla \cdot (\mathbf{v}_f \otimes \mathbf{v}_f) = -\frac{1}{\rho_f}\nabla p + \nu \nabla^2 \mathbf{v}_f + \mathbf{f} \quad \text{in } \overline{\Omega} \tag{5a}$$

$$\nabla \cdot \mathbf{v}_f = 0 \quad \text{in } \overline{\Omega} \tag{5b}$$

$$\mathbf{v}_f = \mathbf{v}_s + \boldsymbol{\omega}_s \times \mathbf{r} \quad \text{on } \partial\Omega_i \tag{5c}$$

$$m_s \frac{d\mathbf{v}_s}{dt} = -\rho_f \int_{\Omega_s} \mathbf{f} dV + m_s(1 - \frac{\rho_f}{\rho_s})\mathbf{g} \tag{5d}$$

$$I_s \frac{d\boldsymbol{\omega}_s}{dt} = -\rho_f \int_{\Omega_s} \mathbf{r} \times \mathbf{f} dV \tag{5e}$$

$$\frac{d\mathbf{x}_c}{dt} = \mathbf{v}_s \tag{5f}$$

$$\frac{d\boldsymbol{\theta}_c}{dt} = \boldsymbol{\omega}_s \tag{5g}$$

where the effect of gravity in the fluid momentum equation is from now on incorporated into the pressure.

Moving immersed boundary method for strongly coupled FSI

We first discretize the governing equations as

$$\frac{{\mathbf{v}_f}^{n+1} - {\mathbf{v}_f}^{n}}{\Delta t} + \frac{3}{2}\mathcal{N}({\mathbf{v}_f}^{n}) - \frac{1}{2}\mathcal{N}({\mathbf{v}_f}^{n-1}) = -\frac{1}{\rho_f}\mathcal{G}p^{n+1} + \frac{\nu}{2}\mathcal{L}({\mathbf{v}_f}^{n+1} + {\mathbf{v}_f}^{n}) + \mathcal{S}\mathbf{F}^{n+1} \tag{6a}$$

$$\mathcal{D}{\mathbf{v}_f}^{n+1} = 0 \tag{6b}$$

$$\mathcal{T}{\mathbf{v}_f}^{n+1} = {\mathbf{v}_s}^{n+1} + {\boldsymbol{\omega}_s}^{n+1} \times \mathbf{r}^{n+1} \tag{6c}$$

$$m_s \frac{{\mathbf{v}_s}^{n+1} - {\mathbf{v}_s}^{n}}{\Delta t} = -\rho_f \mathbf{F}^{n+1} + m_s(1 - \frac{\rho_f}{\rho_s})\mathbf{g} \tag{6d}$$

$$I_s \frac{{\boldsymbol{\omega}_s}^{n+1} - {\boldsymbol{\omega}_s}^{n}}{\Delta t} = -\rho_f \mathbf{r} \times \mathbf{F}^{n+1} \tag{6e}$$

$$\frac{{\mathbf{x}_c}^{n+1} - {\mathbf{x}_c}^{n}}{\Delta t} = {\mathbf{v}_s}^{n+1} \tag{6f}$$

$$\frac{{\boldsymbol{\theta}_c}^{n+1} - {\boldsymbol{\theta}_c}^{n}}{\Delta t} = {\boldsymbol{\omega}_s}^{n+1} \tag{6g}$$

where $\mathcal{L}$, $\mathcal{N}$, $\mathcal{D}$, $\mathcal{G}$ are the discrete Laplacian, convective, divergence, gradient operators respectively. Since the fluid mesh in general does not coincident with the solid mesh, $\mathcal{T}$ and $\mathcal{S}$ are the interpolation and spreading operators to exchange the flow quantities on both meshes, which can be constructed from the discrete delta functions as in [8]. $\mathbf{F}$ designates the boundary force defined on the solid surface and thus we have $\mathbf{f} = \mathcal{S}\mathbf{F}$. $n+1$ represents the time level to be solved. Here the convection is treated explicitly with a second order Adams-Bashforth scheme but the diffusion is handled implicitly with a second order Crank-Nicolson scheme. Hence the overall scheme is stable under the standard CFL condition.

To solve above coupled fluid-structure system, we perform the following fractional step scheme:

(1) Prediction step for ${\hat{\mathbf{v}}_f}^{n+1}$

$$\frac{{\hat{\mathbf{v}}_f}^{n+1} - {\mathbf{v}_f}^{n}}{\Delta t} + \frac{3}{2}\mathcal{N}({\mathbf{v}_f}^{n}) - \frac{1}{2}\mathcal{N}({\mathbf{v}_f}^{n-1}) = -\frac{1}{\rho_f}\mathcal{G}p^n + \frac{\nu}{2}\mathcal{L}({\hat{\mathbf{v}}_f}^{n+1} + {\mathbf{v}_f}^{n}) \tag{7}$$

(2) Immersed boundary forcing step for the interface coupling

$$\frac{{\tilde{\mathbf{v}}_f}^{n+1} - {\hat{\mathbf{v}}_f}^{n+1}}{\Delta t} = \mathcal{S}\mathbf{F}^{n+1} \tag{8a}$$

$$\mathcal{T}{\tilde{\mathbf{v}}_f}^{n+1} = {\mathbf{v}_s}^{n+1} + {\boldsymbol{\omega}_s}^{n+1} \times \mathbf{r}^{n+1} \tag{8b}$$

Applying (8b) to (8a), we obtain

$$\mathcal{M}\mathbf{F}^{n+1} = \frac{{\mathbf{v}_s}^{n+1} + {\boldsymbol{\omega}_s}^{n+1} \times \mathbf{r}^{n+1} - \mathcal{T}{\hat{\mathbf{v}}_f}^{n+1}}{\Delta t} \tag{9a}$$

$${\tilde{\mathbf{v}}_f}^{n+1} = {\hat{\mathbf{v}}_f}^{n+1} + \Delta t \mathcal{S}\mathbf{F}^{n+1} \tag{9b}$$

where $\mathcal{M}$ is termed as the moving force matrix ($\mathcal{M} = \mathcal{T}\mathcal{S}$) in [2], which is found to be symmetric and positive-definite.

For the interface coupling, the solid velocity and position are solved with this moving force equation through a fixed point iteration, namely iterating (6d)-(6e)-(6f)-(6g)-(9a) until convergence. At each subiteration, the moving force equation is solved with the conjugate gradient method.

(3) Projection step for obtaining a divergence free velocity ${\mathbf{v}_f}^{n+1}$

$$\frac{{\mathbf{v}_f}^{n+1} - {\tilde{\mathbf{v}}_f}^{n+1}}{\Delta t} = -\mathcal{G}\phi^{n+1} \tag{10a}$$

$$\mathcal{D}{\mathbf{v}_f}^{n+1} = 0 \tag{10b}$$

where ϕ is the pseudo pressure. Applying the divergence operator to (10a) along with the divergence free condition (10a) gives

$$\mathcal{L}\phi^{n+1} = \frac{1}{\Delta t}\mathcal{D}{\tilde{\mathbf{v}}_f}^{n+1} \tag{11a}$$

$$\mathbf{v}_f{}^{n+1} = \tilde{\mathbf{v}}_f{}^{n+1} - \Delta t \mathcal{G}\phi^{n+1} \tag{11b}$$

The final pressure is advanced by

$$p^{n+1} = p^n + \phi^{n+1} - \frac{\nu}{2}\mathcal{D}\hat{\mathbf{v}}_f{}^{n+1} \tag{12}$$

where the last term is the splitting error resulted from velocity prediction and now is absorbed into the pressure. This type of projection method yields a consistent pressure boundary condition and thus free of numerical boundary layer, termed as the rotational incremental pressure correction projection method in [5].

The novel strongly coupled scheme is computational inexpensive, since the time-consuming pressure Poisson equation is not evolved in the interface coupling and the moving force equation is very easy to solve. We will demonstrate the novel scheme in the following numerical examples.

Results

Freely falling and rising cylinder in an infinite quiescent fluid

We first consider a circular cylinder freely falling and rising in an infinite quiescent fluid. This phenomenon happens frequently in nature and a large amount of work can be found in the literature. Here we compare our numerical results with the data of [6][7]. Namkoong *et al.* [7] performed the simulation using a body-fitted ALE formulation while Lacis *et al.* [6] employed the immersed boundary projection method.

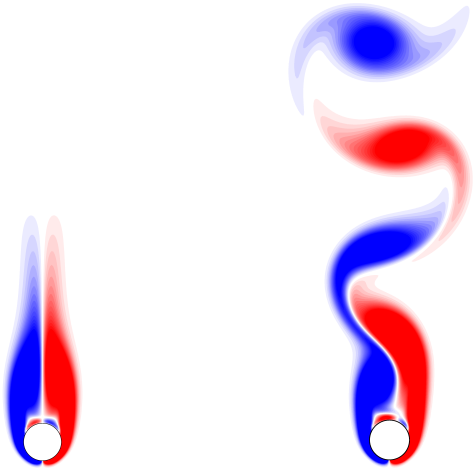


Figure 2: Vorticity fields for a freely falling cylinder in an open domain: (Left) $tV_t/D = 10$ and (right) $tV_t/D = 90$. The contour level is set from -6 (blue) to 6 (red) with an increment of 0.4.

Two density ratios are considered in this study, i.e. $\rho_s/\rho_f = 1.01$ for the falling case and $\rho_s/\rho_f = 0.99$ for the rising simulation. A large computational domain is taken as $[-5D, 5D] \times [-70D, 70D]$ with free-slip boundary conditions applied at all exterior boundaries, where $D = 0.5$ cm is the cylinder diameter. A uniform mesh is employed to cover the computational domain, and the mesh resolution is kept to $0.04D$ in order to compare with Lacis *et al.* [6]. Initially the cylinder is located at $\pm 65D$, depending on the situation ($65D$ for the falling case, $-65D$ for the rising case). The Reynolds number

$Re = V_t D/\nu_f$ is 156, where V_t is the terminal velocity. Note that the Reynolds number depends on the Galileo number $G = (|\rho_s/\rho_f - 1|gD^3)^{1/2}/\nu_f$ (here $G = 138$) and the density ratio ρ_s/ρ_f.

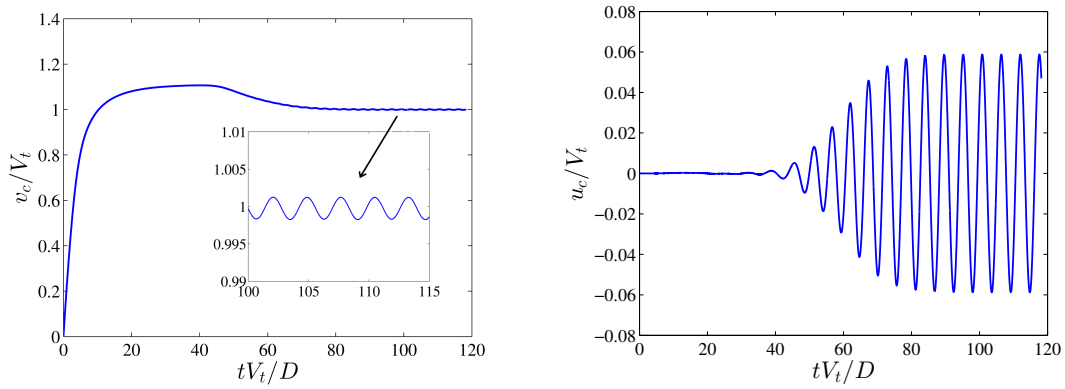


Figure 3: Time histories of the vertical and horizontal velocity for the freely rising cylinder $\rho_s/\rho_f = 0.99$.

Table 1: The drag, lift coefficients and the Strouhal number for the freely falling and rising circular cylinder in an open domain.

		C_D	max$\|C_L\|$	St
$\rho_s/\rho_f = 1.01$	Present	1.35	0.10	0.189
	Lacis *et al.* [6]	1.29	0.14	0.17185
	Namkoong *et al.* [7]	1.23	0.15	0.1684
$\rho_s/\rho_f = 0.99$	Present	1.35	0.10	0.189
	Lacis *et al.* [6]	1.29	0.14	0.17188
	Namkoong *et al.* [7]	-	-	0.1687

The vorticity fields are presented in Figure 2 for the falling cylinder case. Initially symmetric vortex pair forms behind the cylinder in the beginning of falling. After that the numerical error accumulates and breaks the symmetry. At around $tV_t/D = 40$, the flow becomes unsteady and periodic vortex shedding occurs. The time histories of the velocity components of the cylinder are plotted in Figure 3. Table 1 shows the Strouhal number $St = fD/V_t$ (f is the shedding frequency) and the coefficients of drag and lift. Present results are compared to those of [6][7]. Good agreements have been obtained.

Elliptical particle sedimentation in a confined channel

Next we consider the sedimentation of an elliptical particle in a narrow channel, to demonstrate the ability of current FSI algorithm for handling non-circular object. This example was studied previously by Xia *et al.* [10] for the boundary effects on the sedimentation mode. In their work, a multi-block lattice Boltzmann method is used and compared to the traditional ALE formulation.

To compare with Xia *et al.* [10], the computational domain is selected to be $[0, L] \times [0, 7L]$ with $L = 0.4$ cm. The aspect ratio of the ellipse is $\alpha = a/b = 2$, where a and b are the major and minor

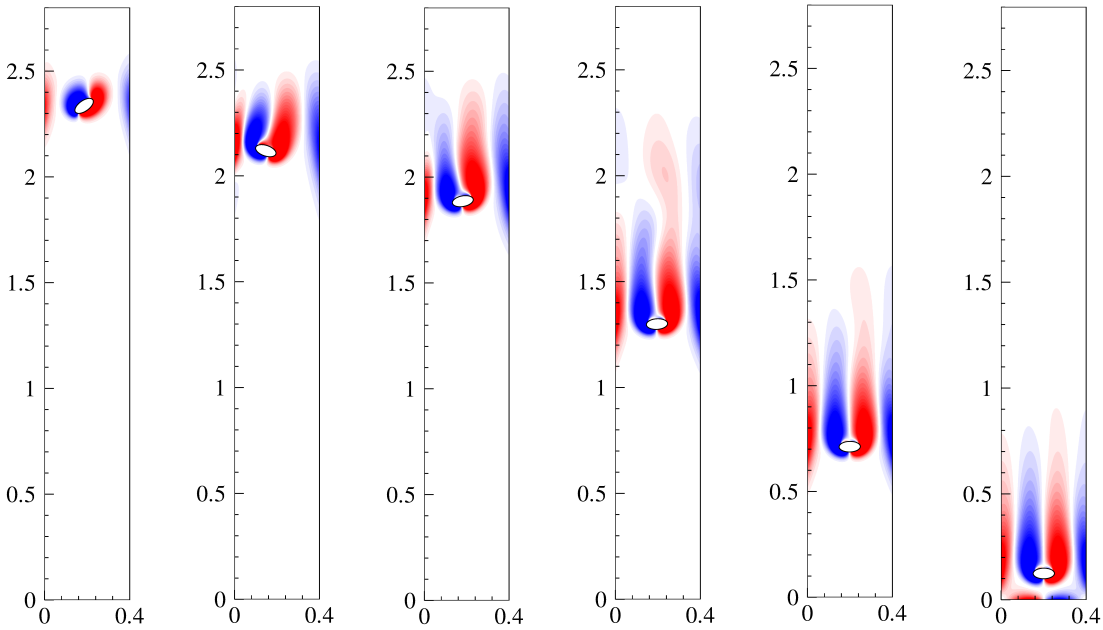


Figure 4: Vorticity fields at different times: (from left to right) t = 0.1 s, 0.3 s, 0.5 s, 1.0 s, 1.5 s, 2.0 s. The contour levels are set from -15 (blue) to 15 (red).

axes respectively. The blockage ratio is defined as $\beta = L/a = 4$. The density ratio is $\rho_s/\rho_f = 1.1$. The kinematic viscosity of fluid is set to $\nu = 0.01$ cm^2/s. The particle starts falling in a quiescent fluid from the centroid at $(0.5L, 6L)$ with an initial angle of $\pi/4$ to break the symmetry.

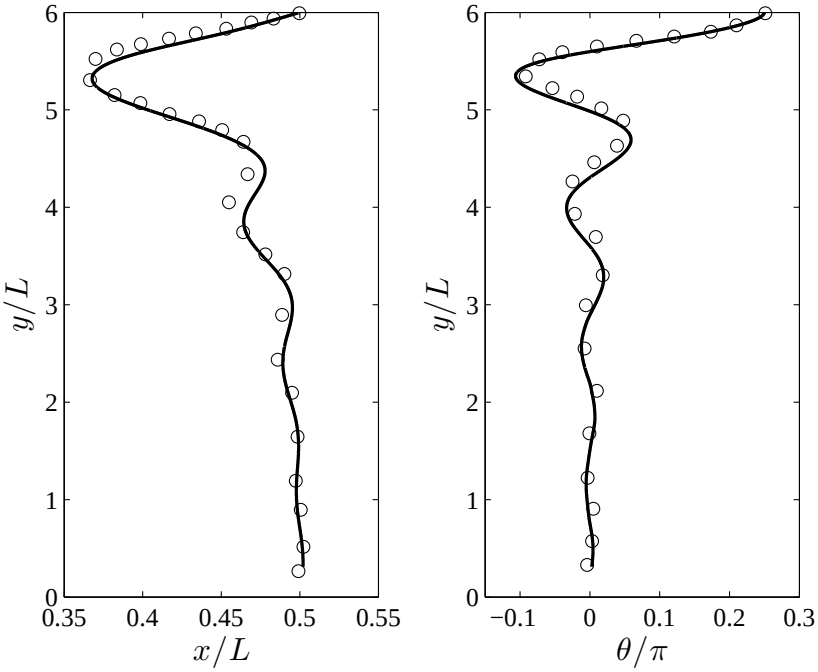


Figure 5: Particle trajectory and orientation of the elliptical particle. ”—”, present results; ”∘”, results of [10].

No-slip boundary conditions are applied at four boundaries. A uniform mesh is employed with a gird resolution of 0.0027 cm. The time step is chosen such that the CFL condition is satisfied. Figure 4 shows the vorticity fields at different times at $t = 0.1$ s, 0.3 s, 0.5 s, 1.0 s, 1.5 s, 2.0 s. The trajectory and orientations are compared to the results of [10] in Figure 5. Good agreements have been obtained.

Conclusions

In this work an efficient strongly coupled fluid-structure interaction scheme was proposed in the context of the moving immersed boundary method. To accurately impose the no-slip boundary condition at the immersed interface, a moving force equation was derived and solved with the conjugate gradient method. The global scheme follows a fractional step manner while the interface coupling was accomplished between the solid motion equations with the moving force equation in the immersed boundary forcing step. Stable results were obtained even when the solid density is smaller than the fluid density. Numerical results have demonstrated the accuracy of the proposed method.

Acknowledgements

The support of the China Scholarship Council is greatly acknowledged.

References

[1] Cai, S.-G., Ouahsine, A. and Sergent, P. (2016) Modelling wave energy conversion of a semi-submerged heaving cylinder. *In Ibrahimbegovic, A. editor, Computational Methods for Solids and Fluids: Multiscale Analysis, Probability Aspects, Model Reduction and Software Coupling*, 67–79.

[2] Cai, S.-G., Ouahsine, A., Favier, J. and Hoarau, Y. (2016) Improved implicit immersed boundary method via operator splitting. *In Ibrahimbegovic, A. editor, Computational Methods for Solids and Fluids: Multiscale Analysis, Probability Aspects, Model Reduction and Software Coupling*, 49–66.

[3] Fadlun, E.A., Verzicco, R., Orlandi, P., and Mohd-Yusof, J. (2000) Combined immersed boundary finite-difference methods for three-dimensional complex flow simulations. *Journal of Computational Physics* **161**, 35–60.

[4] Glowinski, R., Pan, T.W., Hesla, T.I., Joseph, D.D., and Périaux, J. (2001) A fictitious domain approach to the direct numerical simulation of incompressible viscous flow past moving rigid bodies: application to particulate flow. *Journal of Computational Physics* **169**, 363–426.

[5] Guermond, J.L., Minev, P., and Shen, J. (2006) An overview of projection methods for incompressible flows. *Computer Methods in Applied Mechanics and Engineering* **195**, 6011–6045.

[6] Lacis, U., Taira K. and Bagheri, S. (2016): A stable fluid-structure-interaction solver for low-density rigid bodies using the immersed boundary projection method. *Journal of Computational Physics* **305**, 300–318.

[7] Namkoong, K., Yoo, J.Y., and Choi, H.G. (2008) Numerical analysis of two-dimensional motion of a freely falling circular cylinder in an infinite fluid. *Journal of Fluid Mechanics* **604**, 33–53.

[8] Peskin, C.S. (1972) Flow patterns around heart valves: A numerical method. *Journal of Computational Physics* **10**, 252–271.

[9] Uhlmann, M. (2005) An immersed boundary method with direct forcing for the simulation of particulate flows. *Journal of Computational Physics* **209**, 448–476.

[10] Xia, Z., Connington, K.W., Rapaka, S., Yue, P., Feng, J.J., and Chen, S. (2009) Flow patterns in the sedimentation of an elliptical particle. *Journal of Fluid Mechanics* **625**, 249–272.

3D Point Cloud Data and Triangle Face Compression by a Novel Geometry Minimization Algorithm and Comparison with other 3D Formats

***M. M. Siddeq[1], †M. A. Rodrigues[2]**

[1,2]GMPR-Geometric Modeling and Pattern Recognition Research Group,
Sheffield Hallam University, Sheffield, UK

*Presenting author:mamadmmx76@gmail.com
†Corresponding author: M.Rodrigues@shu.ac.uk

Abstract

Polygonal meshes remain the primary representation for visualization of 3D data in a wide range of industries including manufacturing, architecture, geographic information systems, medical imaging, robotics, entertainment, and military applications. Because of its widespread use, it is desirable to compress polygonal meshes stored in file servers and exchanged over computer networks to reduce storage and transmission time requirements. 3D files encoded by OBJ format are commonly used to share models due to its clear simple design. Normally each OBJ file contains a large amount of data (e.g. vertices and triangulated faces) describing the mesh surface. In this research we introduce a novelalgorithm to compress vertices and triangle faces called Geometry Minimization Algorithm (GM-Algorithm). First, each vertex consists of (*x, y, z*) coordinates that are encoded into a single value by the GM-Algorithm. Second, triangle faces are encoded by computing the differences between two adjacent vertex locations, and then coded by theGM-Algorithm followed byarithmetic coding. We tested the method on large data sets achieving highcompression ratios over90% while keeping the same number of vertices and triangle faces as the original mesh. The decompression step is based on a Parallel Fast Matching Search Algorithm (Parallel-FMS) to recover the structure of the 3D mesh. A comparative analysis of compression ratios is provided with a number of commonly used 3D file formats such as MATLAB, VRML, OpenCTM and STL showing the advantages and effectiveness of our approach.

Keywords: 3D Object Compression and Reconstruction, Data Compression, GM-Algorithm, Parallel-FMS Algorithm

1. Introduction

Polygonal meshes are the primary representation used in the manufacturing, architectural, and entertainment industries for the visualization of 3D data, and they are central to Internet and broadcast multimedia standards such as MPEG-4 [1,2,4] and VRML [3]. In these standards, a polygonal mesh is defined by the position of its vertices (geometry); by the association between each face and its sustaining vertices (connectivity); and optional colour, normal and texture coordinates (properties). Deering [5] introduced the first geometry compression scheme to compress the bit stream sent by a CPU to a graphics adapter, generalizing the popular triangle strips and fans. Motivated by Deering's work, but optimized for transmission over the internet instead, Taubin and Rossignac introduced the Topological Surgery (TS) method[6], the first connectivity preserving single-resolution manifold triangular mesh compression scheme. TS was later extended to handle arbitrary manifold polygonal meshes with attached properties, and proposed as a compressed file format to encode VRML files [9]. With a more efficient encoding, Topological Surgery is now part of the MPEG-4 standard.

Several closely related methods were subsequently developed by Touma and Gotsman[12], Gumhold and Strasser [7], Li and Kuo[8] and Rossignac[10]. The methods proposed by

Gumholdand Strasser, and by Rossignac only capable of encoding connectivity. The method proposed by Touma and Gotsman, predicts geometry and properties better, and the method proposed by Li and Kuo improves on the entropy encoding of prediction errors. More recently, Bajaj *et al*. [11] proposed yet another method to encode single-resolution triangular meshes. It is based on a decomposition of the mesh into rings of triangles originally used by Taubin and Rossignac in their compression algorithm, but with a different and more complex encoding. All of these schemes require *O(n)* total bits of data to represent a single-resolution mesh in compressed form.

While single resolution schemes can be used to reduce transmission bandwidth, it is frequently desirable to send the mesh in progressive fashion. A progressive scheme sends a compressed version of the lowest resolution level of a level-of-detail (LOD) hierarchy, followed by a sequence of additional refinement operations. In this manner, successively finer levels of detail may be displayed while even more detailed levels are still arriving. To prevent visual artefacts, sometimes referred to as popping, it is also desirable to be able to transition smoothly from one level of the LOD hierarchy to the next by interpolating the positions of corresponding vertices in consecutive levels of detail as a function of time [11].

The Progressive Mesh (PM) scheme introduced by Hoppe [13] was the first method to address the progressive transmission of multi-resolution manifold triangular mesh data. PM is an *adaptive refinement* scheme where new faces are inserted in between existing faces. Every triangular mesh can be represented as a base mesh followed by a sequence of *vertex split* refinements. Each vertex split is specified for the current level of detail by identifying two edges and a shared vertex. The mesh is refined by cutting it through the pair of edges, splitting the common vertex into two vertices and creating a quadrilateral hole, which is filled with two triangles sharing the edge connecting the two new vertices. The PM scheme is not an efficient compression scheme. Since the refinement operations perform very small and localized changes, the scheme requires $O(V \log2(V))$ bits to double the size of a mesh with V vertices. Later on Hoppe proposed a more efficient implementation based on changing the order of transmission of the edge split operations [14].

In progressive representations discussed above,multi-resolution polygonal models are represented in compressed form. However, as compression schemes, these are not as efficient as the single-resolution schemes described earlier. Taubin*et al*.[15] recently introduced a method to compress any multi-resolution mesh produced by a vertex clustering algorithm with compression ratios comparable to the best single resolution schemes. In this scheme, the connectivity of the LOD hierarchy is transmitted from high resolution to low resolution, followed by the geometry and properties from low resolution to high resolution. The main contribution of this scheme is a method to compress the *clustering* mappings which relate consecutive levels of detail, from high to low resolution. The method achieves high compression ratios but is not progressive.

The MPEG-4 3D Mesh Coding scheme is based on the Topological Surgery and Progressive Forest Split schemes. But it incorporates improvements to connectivity encoding for progressive transmission proposed by Bossen[16], non-manifold encoding proposed byGuéziec*et al*. [17], error resiliency proposed by Jang *et al*. [18], parallelogram prediction proposed by Touma and Gotsman[15], and error encoding proposed by Li and Kuo[8]. It allows the encoding of any polygonal mesh (including non-manifolds) with no loss of connectivity information and no repetition of geometry and property data associated to singular vertices as a progressive single-resolution bit stream, and any manifold polygonal mesh in hierarchical multi-resolution mode. Extensive experimentation performed during the course of the MPEG-4 process has shown that the resulting methods are state-of the-art.

Siddeq and Rodrigues proposed a new way to compress vertices by using a Geometry Minimization Algorithm (paper submitted to a journal and under review – for more information please contact the authors). In this paper we introducea new concept for geometry and mesh connectivity compression.The proposed method encodes both the point cloud data representing the integer vertices (geometry) and the triangulated faces (connectivity). Thereafter, the encoded output is subjected to arithmetic coding. We demonstrate the approach by performing a comparative analysis with a number of 3D data file formats focusing on compression ratios.

This remainder of this paper is organized as follows: Section 2 introduces geometry coding and describes the proposedGeometry Minimization (GM-Algorithm) applied to the vertices. Section 3 describes mesh connectivity lossless coding bythe GM-Algorithm, while section 4 describes theParallel Fast Matching Search algorithm (PFMS), used to reconstruct vertices and triangulated faces. Section 5 describes experimental results with a comparative analysis followed by conclusions in Section 6.

2. Geometry Compression

Geometry compression combines quantization and statistical coding. Quantization truncates the vertex coordinates to a desired accuracy and maps them into integers that can be represented with a limited number of bits. The quantization parameter, α, is a scale parameter that normally moves the decimal place of each vertex to the right. A tight (min-max) axis aligned bounding box around each object is computed. The minima and maxima of the (x, y, z) coordinates, which define the box, together with the parameter α are encoded and transmitted with the compressed representation of each object. In this way, the 3D structure can be reconstructed in the same units and scale as the original.

The quantization by α transforms each (x, y, z) coordinates into integers ranging from 0 to $2B$–1, where B is the maximum number of bits needed to represent the quantized coordinates. Normally, 12bit integers are sufficient to ensure geometric fidelity for most applications and most models. Thus, suchlossy quantization step reduces the storage cost of geometry from 96-bits to less than 36-bits. The quantization of vertices (x, y, z) is defined as:

$$V_{x,y,z} = floor(V_{x,y,z}\ \alpha) \quad (1)$$

Where $2 \leq \alpha \leq 10{,}000$ .In addition to reducing the storage cost of geometry, we reduced the number of bits for each vertex to less than 16-bit by calculating the differences between two adjacent coordinates for increased redundancy data and thus, more susceptible to compression. The differential process defined in Eq. (2) belowis applied toaxes X, Y and Z independently [19].

$$D(i) = D(i) - D(i+1) \quad (2)$$

Where $i=1, 2, 3... m-1$ and m is the size of the list of vertices.

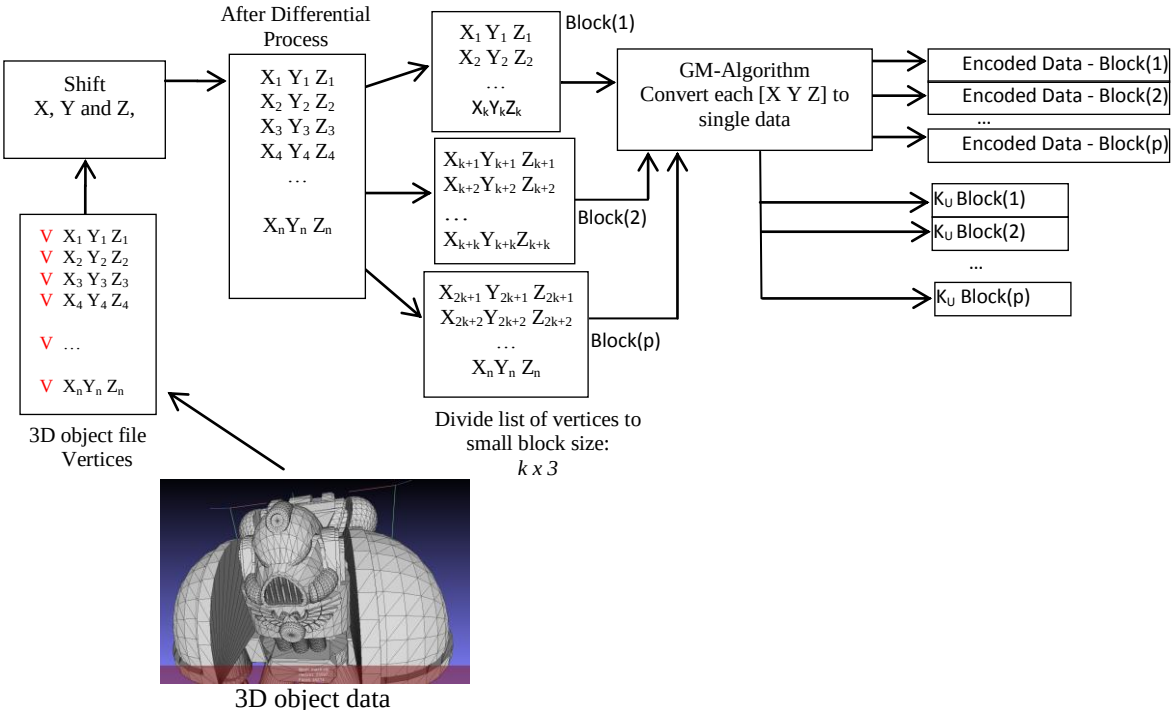


Figure 1. TheGM-Algorithm applied to each block of vertices

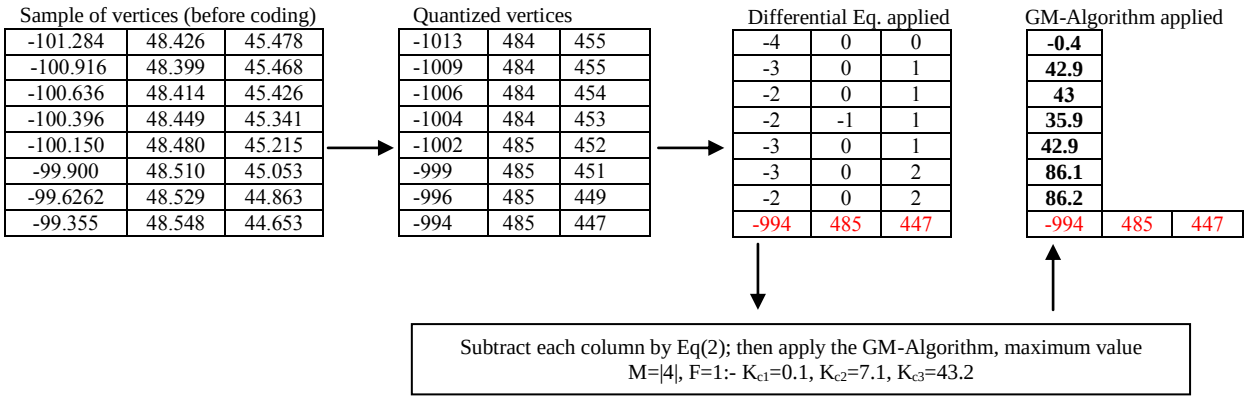
Once the differential process isapplied to the vertices, thelist of vertices is divided into blocks, and the GM-Algorithm is applied to each block of vertices (i.e. the vertex matrix from 3D object file is divided into *k* non-overlapping blocks) as illustrated in Figure 1. The main reason for placing vertices into separate blocks is to speed up the compression and decompression steps. Each *k*block isreduced to an encoded data array. The GM-Algorithm is defined as taking three key values and multiplying theseby three geometry coordinates (*x*, *y*, *z*) from a block of vertices which are then summed overto asingle integer value. A 3-value compression key K_C is generated from vertex data as follows:

$$M = \max(V_X, V_Y, V_Z) + \frac{\max(V_X, V_Y, V_Z)}{2} \quad \% \textit{ Define M as a function of maximum}$$
$$K_{C1} = random(0,1) \quad \% \textit{ First weight}\leq 1 \textit{ defined by random between 0 and 1}$$
$$K_{C2} = (K_{C1} + M) + F \quad \% \textit{ F is an integer factor F=1,2,3,...}$$
$$K_{C3} = (M * K_{C1} + M * K_{C2}) * F$$

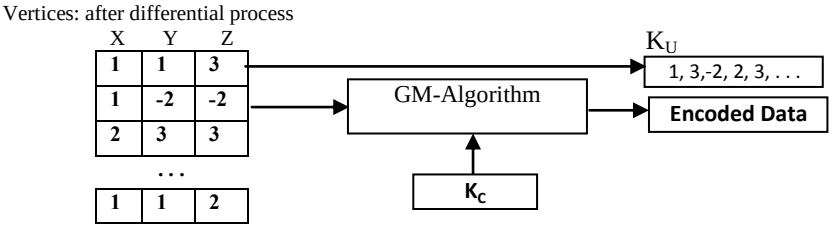
Where$F$ is a positive factor multiplier, each vertex is then encoded as:

$$V(i) = V_x(i)K_{C1} + V_y(i)K_{C2} + V_z(i)K_{C3} \qquad (3)$$

Figure 2(a) illustrates the GM-Algorithm by applying Equation (3) to a sample of vertices. After this operation, the likelihood for each block of vertices is selected from which a *Ku (unique Key)* is generated to be used in the decompression stage as illustrated in Figure 2(b) with a numerical example.



(a) Floating point vertices



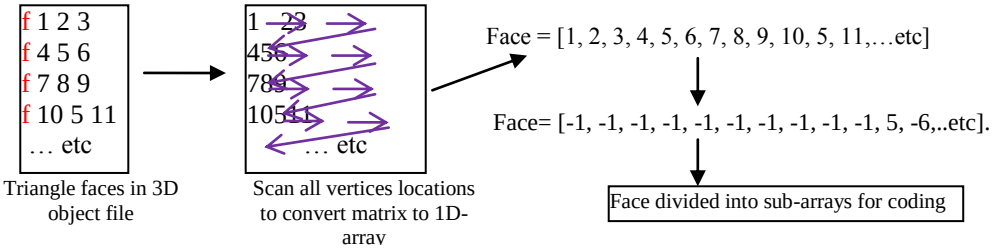
(b) Unique Key

Figure2: (a): Sample of vertices compressed by GM-Algorithm, (b) The set of K_U values generated from a block of vertices

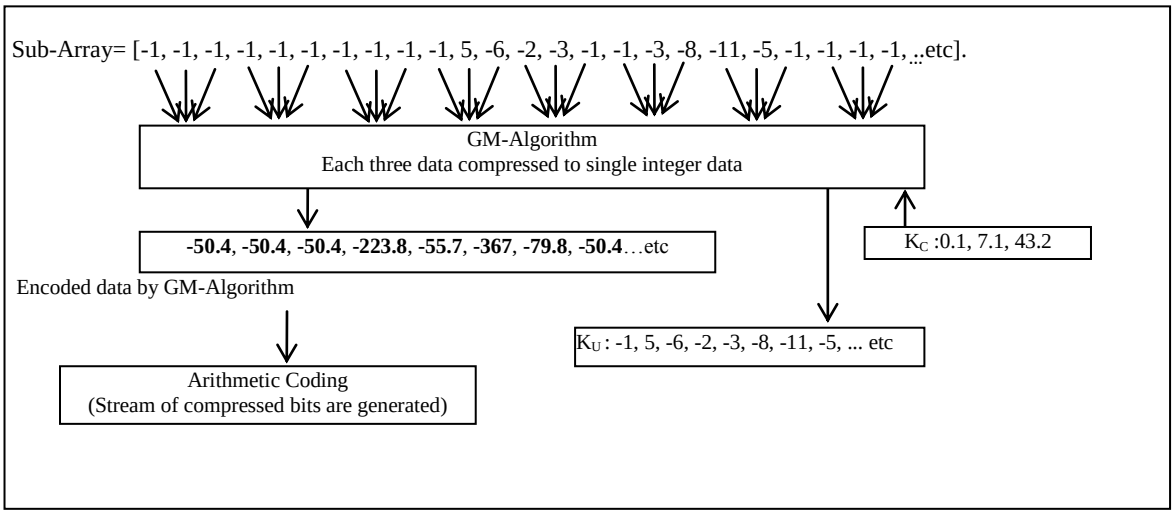
3. Connectivity Compression

Several algorithms have been developed to address the problem of compactly encoding the connectivity of polygonal meshes, both as the theoretical problem of short encodings of embedded graphs and as a practical problem of compressing the incidence table of the triangle mesh in a 3D model.

Triangulated meshes represent geometric connectivity. In a 3D OBJ file, each triangle is followed by reference numbers representing the index of the vertices in the 3D file. These reference numbers arearranged in ascending order in most 3DOBJ files. We refer to these as *regular triangles*. One of regular triangles' advantages is that they can be losslesscompressed ina few bits by applying a differential process (e.g. the differential processed finedby Equation(2) applied to all reference numbers). The resulting 1D-array is divided into sub-arrays, and each sub-array encoded independently by the GM-Algorithm followed by arithmetic coding as illustrated in Figure 3.TheGM-Algorithm works in thesame way as applied to the vertices: three key values are generated and multiplied by three adjacent values which are then summed to a singlevalue by Equation (3).



(a) Triangle Face are scannedrow-by-row



(b), 1D-array divided into sub-arrays, each sub-array encoded independently

Figure 3. (a) and (b): Lossless Triangle Mesh Compression by GM-Algorithm and Arithmetic Coding

4. Data Decompression: Parallel Fast-Matching-Search Algorithm (Parallel-FMS)

The decompression algorithm represents theinverseofcompression usingtheParallel-Fast-Matching-Search Algorithm (Parallel-FMS) to reconstruct vertices and meshconnectivity. First, theParallel-FMS isapplied to encoded block of vertices to reconstruct the originalvertices as a point cloud. Second, the Parallel-FMS isapplied to each encoded sub-array resulting in the reconstructed triangle mesh sub-array.Thereafter, all the sub-arrays arecombined together to recover theincidence table of triangulated facesof the 3D model.Figure 4 shows the layout of the decompression algorithm.

The Parallel-FMS provides the means for fast recovery of both vertices and triangulated meshes, which has been compressed by three different keys (K_C) for each three entries. The header of the compressed file contains information about the compressed data namely K_C and K_U followed by streams of compressed encoded data. The Parallel-FMS algorithm picks up in turn each block of encoded data to reconstruct the vertices and the triangle sub-array. The Parallel-FMS uses a binary search algorithm and is illustrated through the following steps A and B:

A) Initially, K_U is copied three times to sepatared arrays to estimates coordinates (X,Y,Z), that is X1=Y1=Z1, X2=Y2=Z2, X3=Y3=Z3 the searching algorithm computes all possible combinations of X with $K_U(1)$, Y with $K_U(2)$ and Z with $K_U(3)$ that yield a result R-Array illustrated in Figure 5(a).As a means of an example consider that $K_U(1)$=[X1 X2 X3] ,

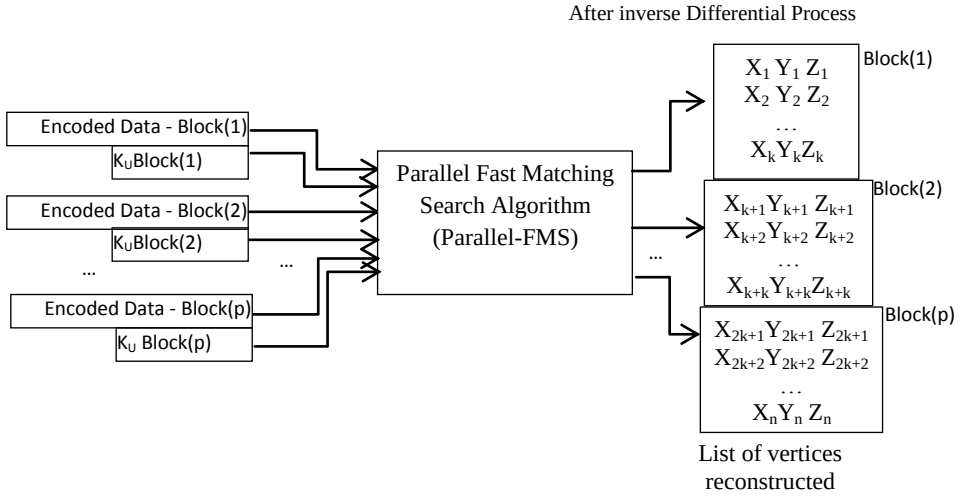
$K_U(2)$=[Y1 Y2 Y3] and $K_U(3)$=[Z1 Z2 Z3]. Then, Equation (3) is executed 27 times to build the R-Array, as described in Figure 5(a). The match indicates that the unique combination of X, Y and Z are represented in theoriginal vertex block.

B) A *Binary Search algorithm* **[**21] is used to recover an item in an array**.** In this research we designed a parallel binary search algorithm consisting of *k*-Binray Search algorithms working in parallel to reconstruct *k* block of vertices in the list of vertices, as shown in Figure 5(b). In each step *k*-Binary Search Algorithms compare*k*-Encoded Data (i.e. each binary search algorithm takes a single compressed data item) with the middle of the element of the R-Array, If the values match, then a matching element has been found and its R-Array's relevant (X,Y,Z) returned. Otherwise, if the search is less than the middle element of the R-Array, then the algorithms repeats its action on the sub-array to the left of the middle element or, if the value is greater, on the sub-array to the right. All *k*-Binary Search algorithms are synchronised such that the correct R-Array is returned. To illustrate our decompression algorithm, the compressed samples in Figure 2(a) (by our GM-Algorithm) can be used by our decompression algorithm to reconstruct X, Y and Z values as shown in Figure 5(c).

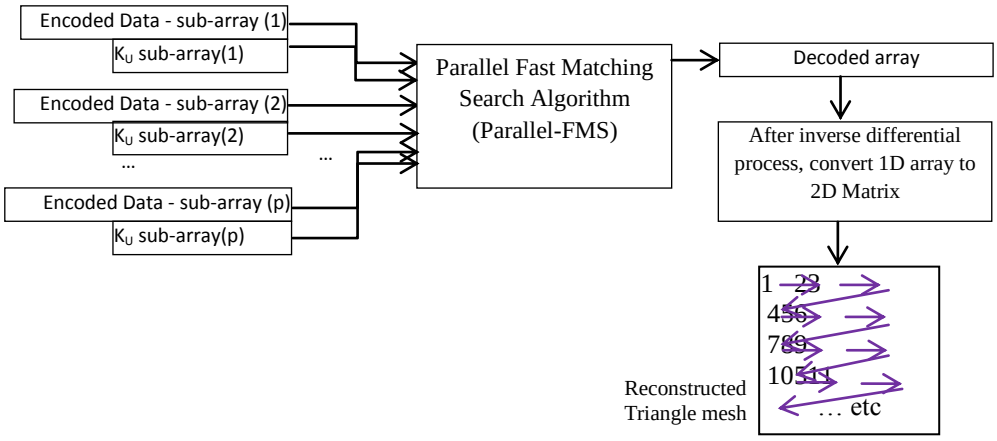
In order to Decode Triangle Faces and Vertices, reverse the differential process of Equation (2) by addition such that the encoded values in the triangle faces and vertices return to their original values. This process takes the last value at position *m*, and adds it to the previous value, and then the total adds to the next previous value and so on. The following equation defines the addition decoder [20].

$$A(i-1) = A(i-1) + A(i) \qquad (4)$$

where $i = m, (m-1), (m-2), (m-3), \ldots, 2$

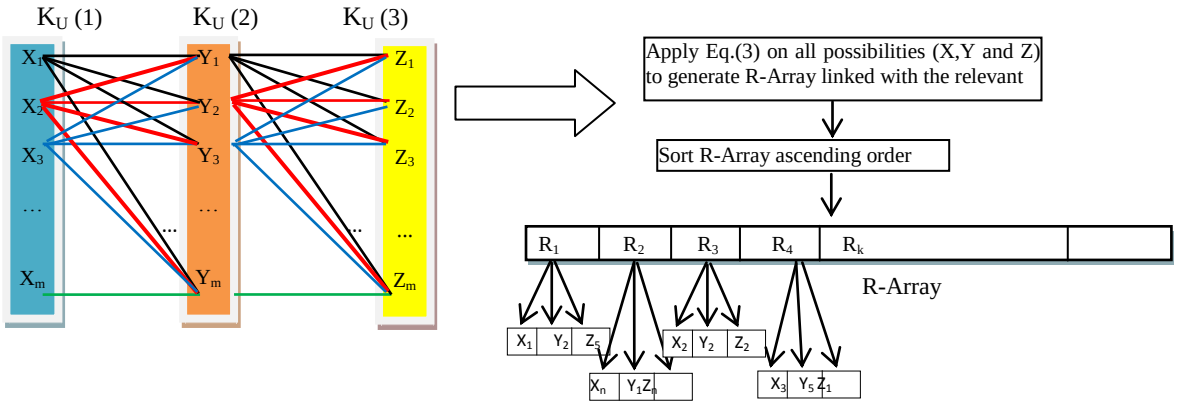


(a) Vertices (X,Y and Z) reconstructed

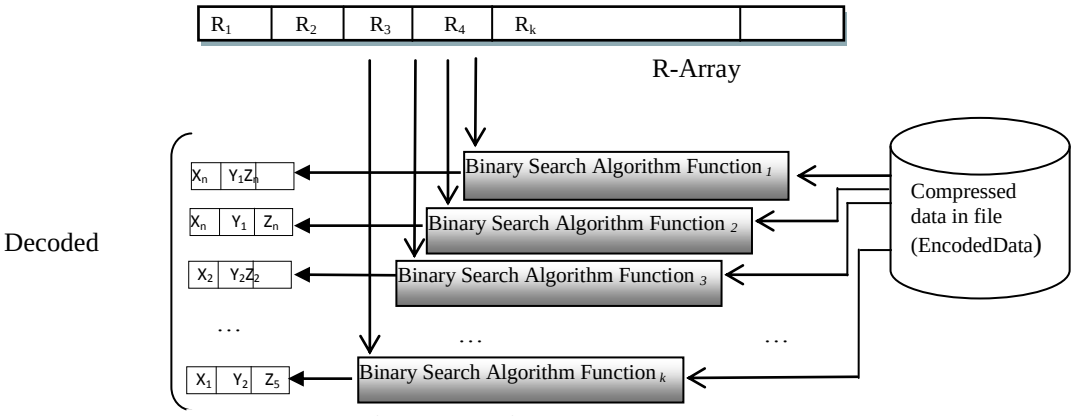


(b) Triangle mesh reconstructed

Figure 4. (a) and (b): Parallel-FMS Algorithm applied on encoded vertices and encoded triangle mesh

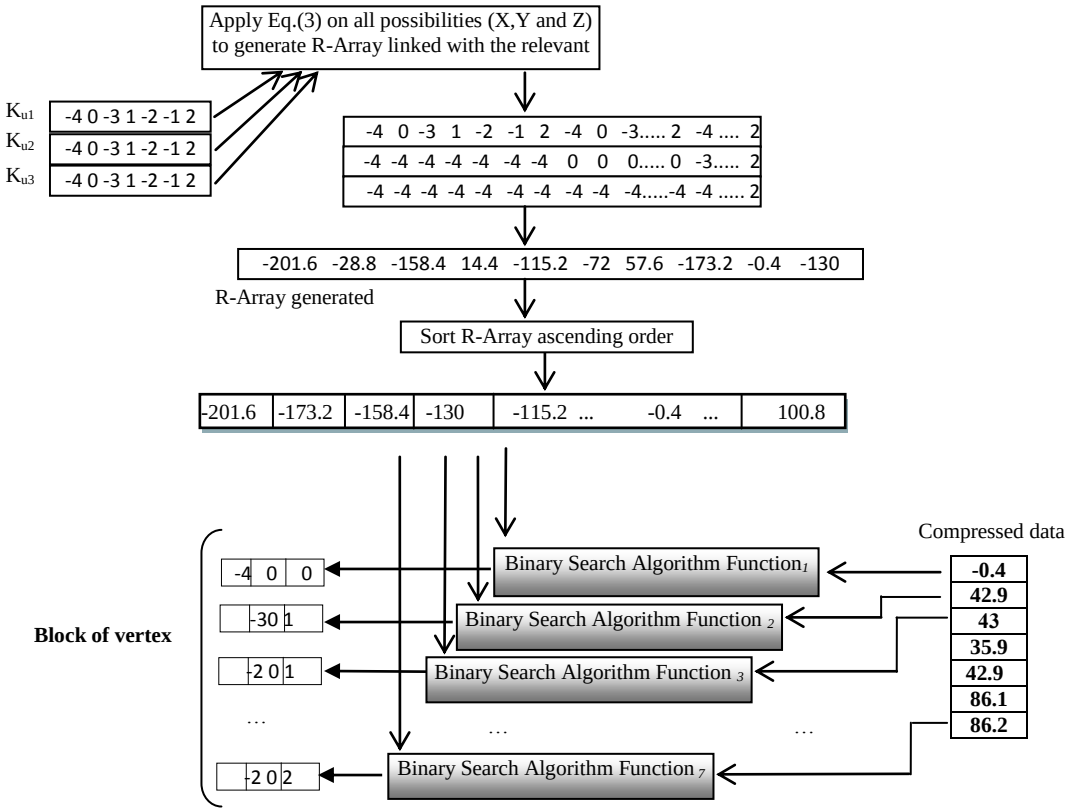


(a) Compute all the probabilities for compute all possible *k*-Encoded Data for reconstruct*k*-block of data



Each Binary Search find Location of the "R-Array" corresponding to the compressed data, output is relevant [X,Y,Z], which represents a original data

(b) All Binary Search algorithms work in Parallel to find group of decompressed data approximately at thesame time.



(c) All Binary Search Algorithm run in Parallel to recover the sampleof vertices, approximately at same time.

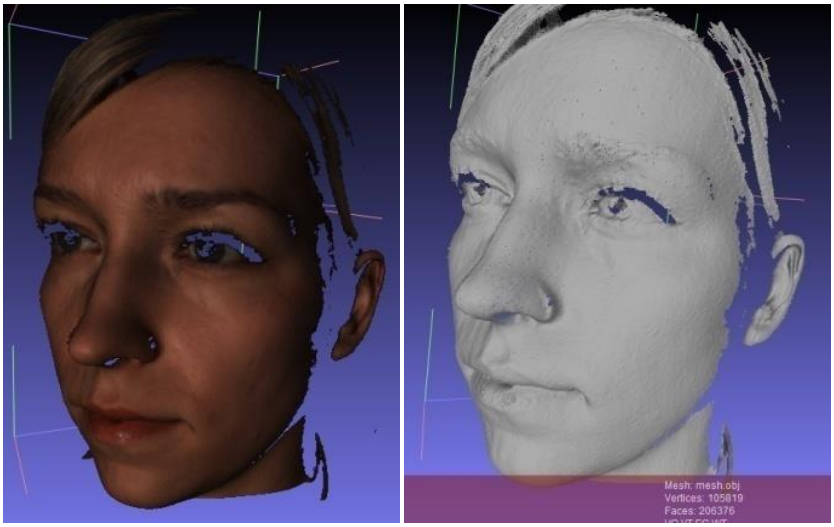
Figure 5. Parallel-FMS algorithm to reconstruct the reduced array (a) Compute all the probabilities for all possible k-Encoded Data (R-Array) by using KC combinations with KU. (b) All Binary Search Algorithm run in Parallel to recover the decompressed 3D data approximately at the same time. (c) Sample of data recovered.

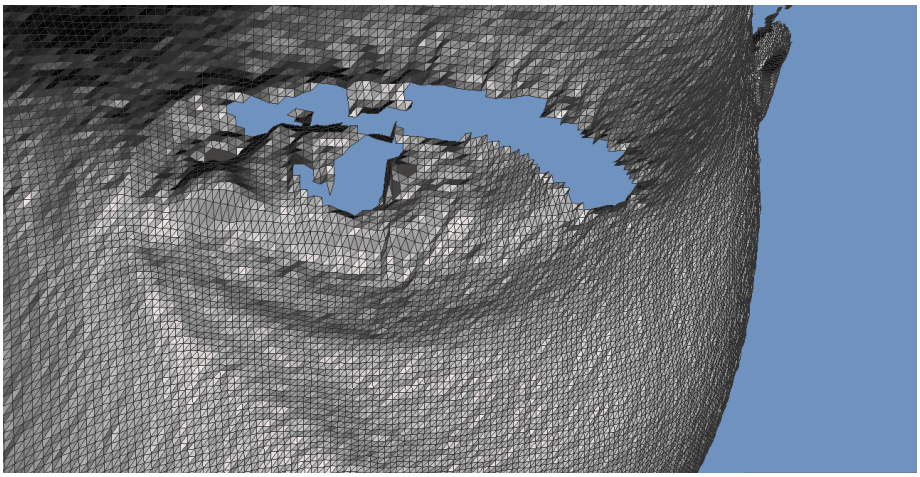
5. Experimental Results

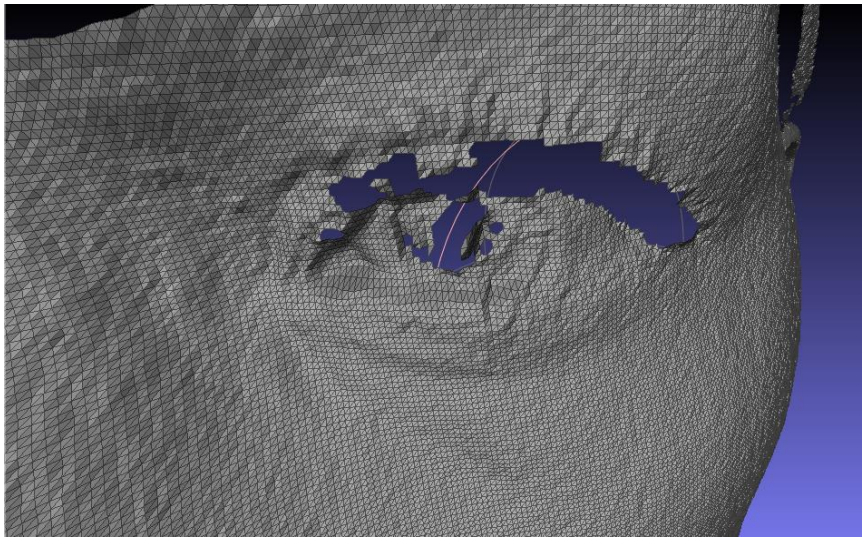
The algorithms were implemented in MATLAB R2013a and Visual C++ 2008 running on an AMD Quad-Core microprocessor. We applied the compression and decompression algorithms to 3D data object generatedby 3dsmax, CAD/CAM, 3D camera or other devices/software. Table 1 shows our compression algorithm applied to each 3D OBJ file, and Figure 6 shows the visual properties of the decompressed 3D object data for 3D images respectively. Additionally, 3D RMSE are used to compare 3D original file sizeswith the recovered files.TheRoot Mean Square Error (RMSE) is used to refer to 3D mesh quality mathematically [22, 23] and can be calculated very easily by computing the differences between thegeometry of the decompressed and the original 3D OBJ files.

Table 1. Our compression approach results

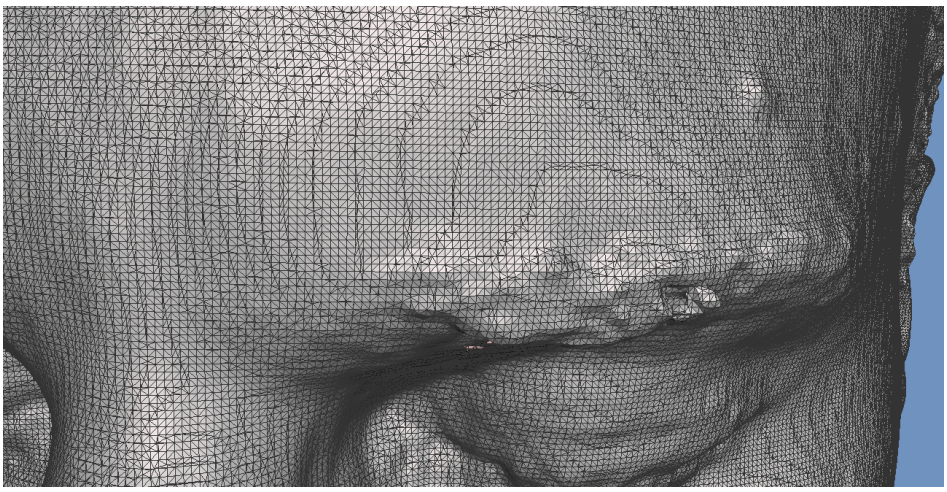
3D object Name	Original file size	Quantization value	Compressed file size	No. of Vertices (Compressed Size)	No. of Triangle faces (Compressed size)	3D RMSE (X Y Z)	Compression ratio
Face1	13.3MB	10	213 KB	105819 (187 KB)	206376 (26 KB)	0.288	98%
Face2	96MB	10	3.7 MB	621693 (1.8MB)	1216249 (1.9MB)	0.289	96%
Angel	23.5 MB	20	1.75MB	307144 (1.055MB)	614288 (715 KB)	0.288	93%
Robot	1.5 MB	400	88.9KB	23597 (56.3KB)	45814 (32.6KB)	0.289	94%
Cup	57KB	2	3.5 KB	594 (2.13 KB)	572 (1.36KB)	0.263	91%
Knot	178 KB	2	7.94KB	1440 (7.4 KB)	2880 (553 Bytes)	0.027	96%

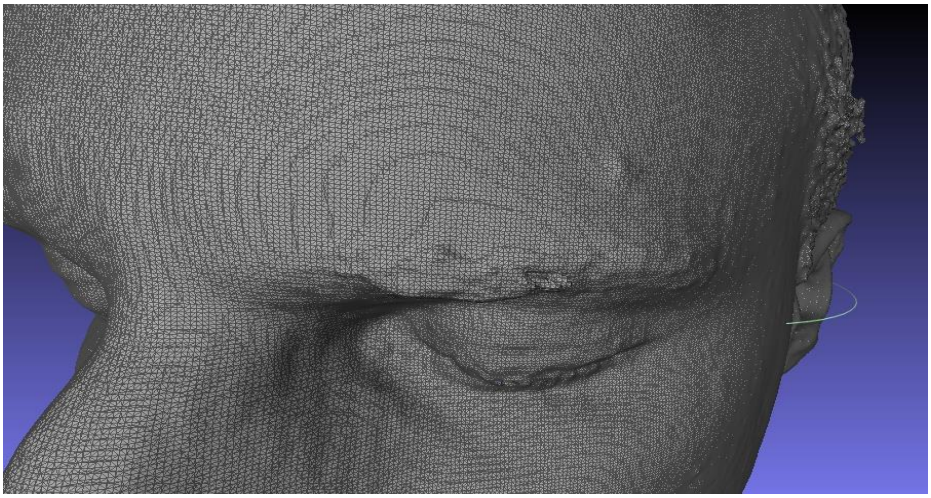




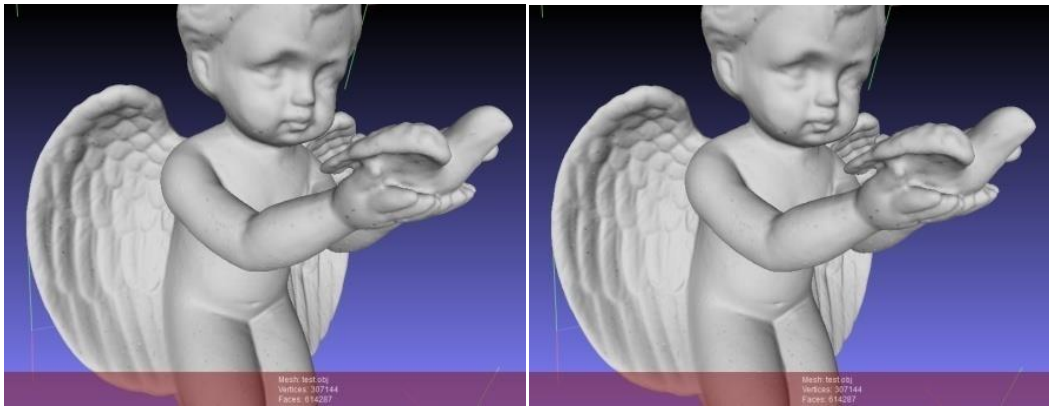


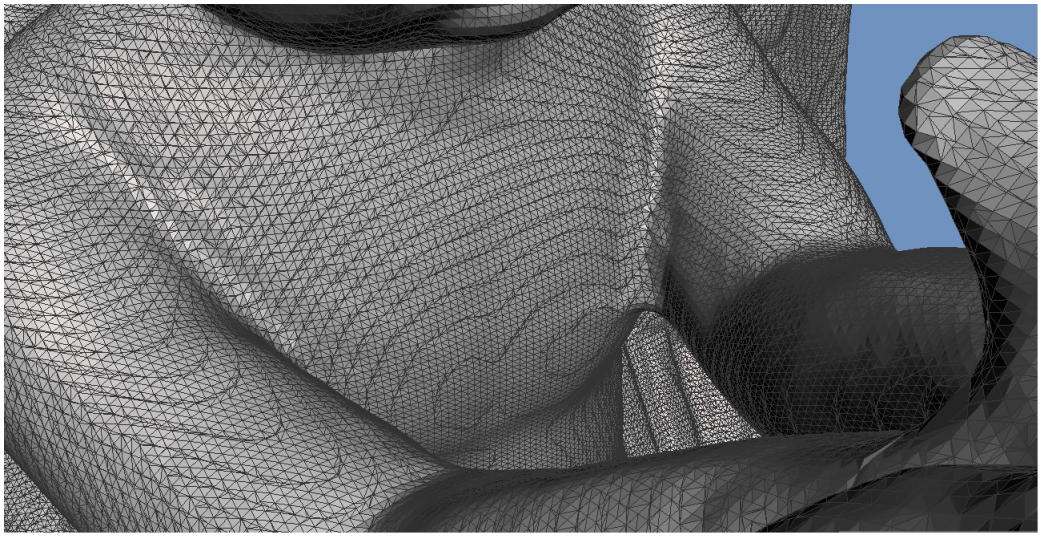
(a) (Top left) original 3D FACE1 object, (Top Right) reconstructed 3D mesh FACE1 without texture, compressed size: 213 KB, (middle) original 3D mesh zoomed by Autodesk application,(bottom) reconstructed 3D mesh zoomed by Meshlab application.

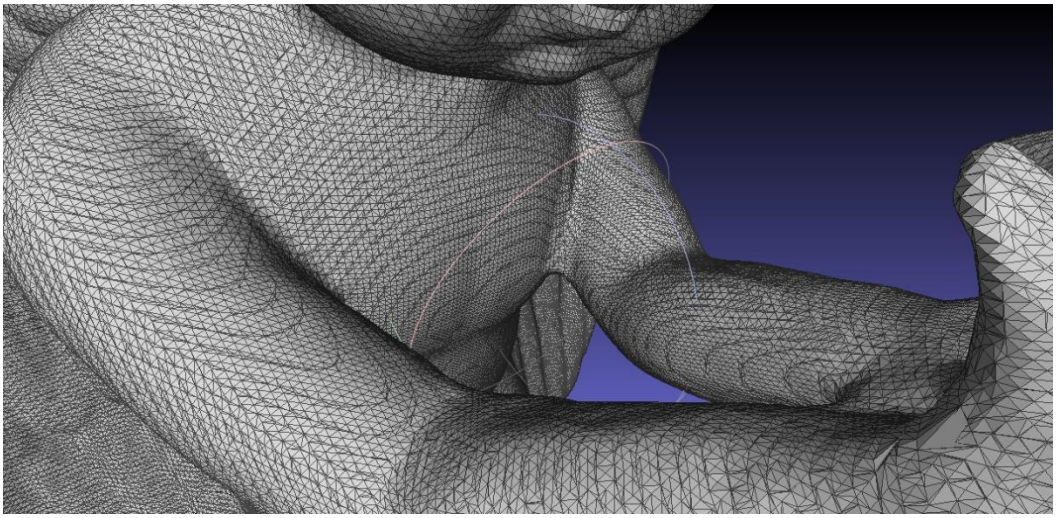




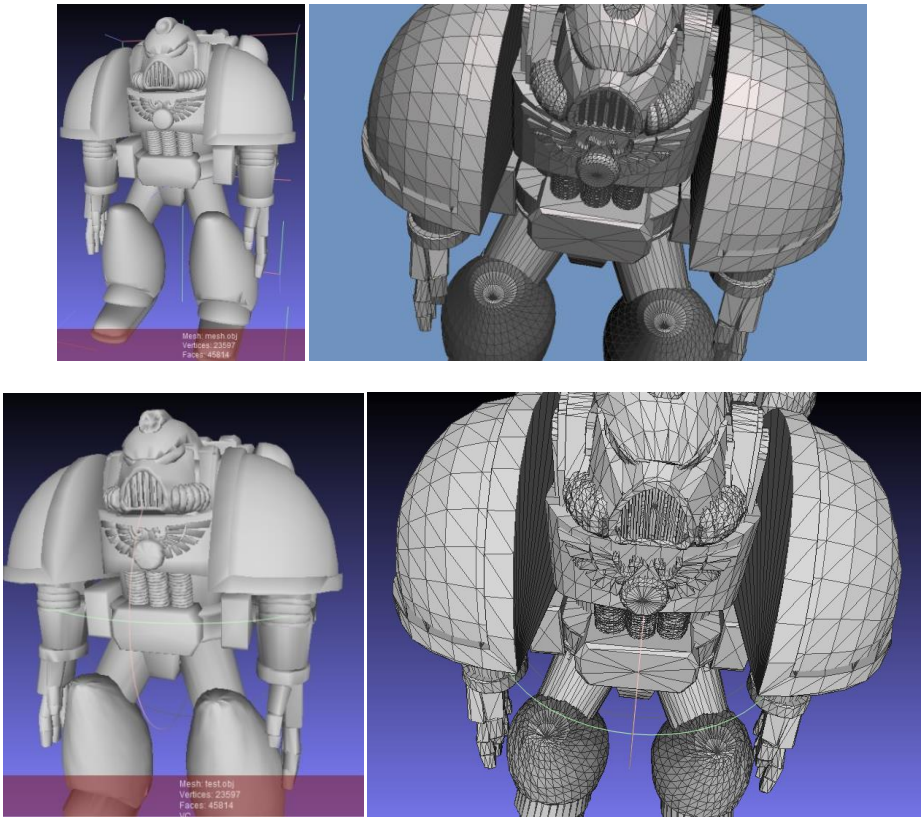
(b) (Top left) original 3D FACE2 object, (Top Right) reconstructed 3D mesh FACE2 without texture, compressed size: 3.7 MB, (middle), original 3D mesh zoomed by Autodesk application, (bottom) reconstructed 3D mesh zoomed by Meshlab application.



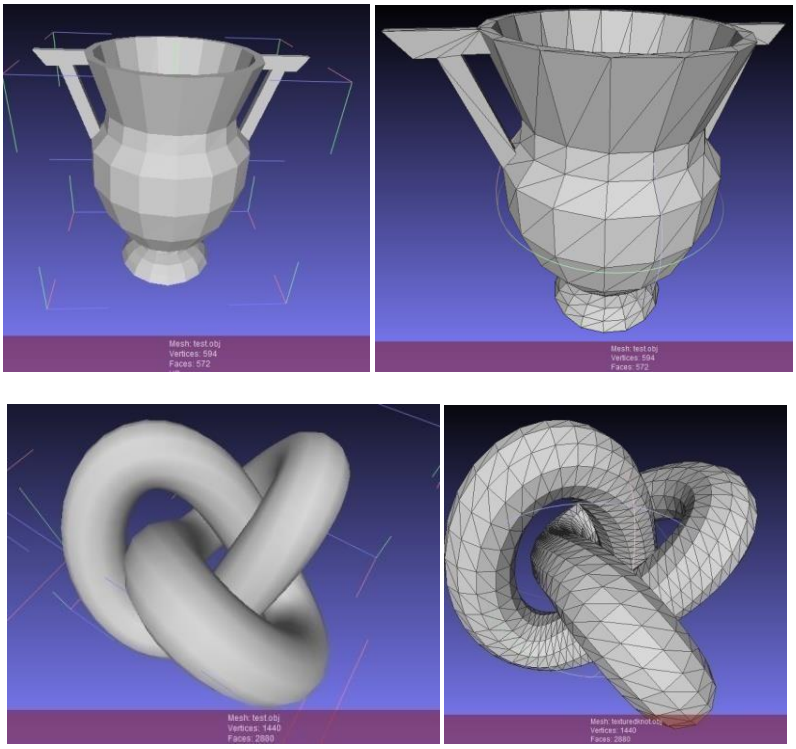




(c) (Top left) original 3D Angel object, (Right left) reconstructed 3D mesh Angel at compressed size: 1.75 MB, (middle) original 3D zoomed by Autodesk application, (bottom) reconstructed 3D mesh zoomed by Meshlab application.



(d) (Top) original 3D Robot object, (bottom) reconstructed 3D mesh Robot, at compressed size: 88.9KB



(e) (Top) original and reconstructed 3D mesh cup, at compressed size: 3.5 KB, (bottom)original and reconstructed 3D mesh Knot, at compressed size: 7.94 KB

Figure 6: (a – e) shows decompressed 3D objects by the proposed algorithms

Tables 2 and 3 showa comparisonofthe proposed method with the 3D file formats: VRML, OpenCTM and STL. In this research we also used a new simple fileformat referred here as MATLAB format. This format savesthegeometry, texture and triangle faces as lossless data, in separated matrices and all the matrices are collected into a single file. We investigate this format obtaining compression ratios over 50% for most of 3D OBJ files. In comparison, our approach uses aunique format to compress 3D files over 98% in the best case; this is mostly dependent on the triangle face details.

Table 2. Our approach compared with other encoding 3D data format according to compressed size

3D object Name	Original file size	**Proposed Algorithm**	MATLAB format	VRML format	OpenCTM	STL
Angel	23.5 MB	**1.75MB**	5.31 MB	23.2 MB	1.92 MB	29.2 MB
Face1	13.3 MB	**213 KB**	4.04 MB	9.19 MB	808 KB	9.84 MB
Face2	96MB	**3.7MB**	23.3 MB	47.7MB	3.7MB	57.9MB
Robot	1.5 MB	**88.9 KB**	449 KB	1.7 MB	151 KB	2.18 MB
Cup	57 KB	**3.5 KB**	12 KB	25.2 KB	3.24 KB	28 KB
Knot	178 KB	**7.94 KB**	23.6KB	95.4KB	14.2KB	140 KB

Total Compressed Size	**5.75 MB**	33.12 MB	81.9 MB	6.57 MB	99.28 MB
Mean Compression Ratio	**95.7 %**	75.3 %	39.4%	95.1 %	26.2 %

Table 3. Our approach compared with other encoding 3D data format according to 3D RMSE

3D object	**Proposed Method**	MATLAB	VRML	OpenCTM	STL
Angel	**0.288**	0	0.0002	44.86	46.32
Face1	**0.289**	0	0.00021	64.79	42.05
Face2	**0.288**	0	0.000109	82.23	43.44
Robot	**0.289**	0	0	0.0587	0.137
Cup	**0.263**	0	0.00000075	37.7	39.2
Knot	**0.027**	0	0.000105	47.65	12.62

6. Conclusion

This research has presented and demonstrated a new method for 3D data compression and compared the quality of compression through 3D reconstruction, 3D RMSE and the perceived quality of the 3D visualisation. The method is based on minimization of geometric values to a stream of new integer data by theGM-Algorithm.Meshconnectivity is partitioned into groups of data,whereeach group iscompressed by theGM-Algorithm followed byarithmetic coding. We note that some of the existing 3D file formats do not efficiently encode geometry and connectivity,as a simple format developed in MATLAB showedhighercompression ratios than STL and VRML. The results show that our approach yields high quality encoding of 3D geometryand connectivity with high compression ratios compared toa number of standard 3D data formats. The slight disadvantage is a larger number of steps for decompression, leading to increased execution time at decoding stage,making the method slower than 3D standard compression methods. Further research includes investigation of methods to speed up decoding, possibly by sorting the*R-Array*entries by frequency.Also, a comparative analysis with a larger number of 3D file formats and compression technique is forthcoming.

References

[1] R. Koenen. (1999) Mpeg-4: Multimedia for our time. *IEEE Spectrum*, 36(2):26–33.

[2] Mpeg-4 overview Seoul revision, (1999). ISO/IEC JTC1/SC29/WG11 Document No. W2725

[3] The Virtual Reality Modeling Language (1999) . http://www.web3d.org, September 1997. ISO/IEC 14772-1.

[4] M.M. Chow. Optimized geometry compression for real-time rendering. In *IEEE Visualization'97 Conference Proceedings*, pages 347–354, 1997.

[5] M. Deering. (1995) *Geometric Compression*. *In Siggraph'95 Conference Proceedings*, pages 13–20,

[6] G. Taubin and J. Rossignac.(1998) Geometry Compression through Topological Surgery.*ACM Transactions on Graphics*, 17(2):84–115.

[7] S. Gumhold and W. Strasser (1998). *Real time compressions of triangle mesh connectivity - In Siggraph'98 Conference Proceedings*, pages 133–140, July 1998.

[8] J. Li and C.C. Kuo (1998). *Progressive Coding of 3D Graphics Models - Proceedings of the IEEE*, 86(6):1052–1063.

[9] G. Taubin, W.P. Horn, and F. Lazarus (1997) . The VRML Compressed Binary Format, June 1997http://www.research.ibm.com/vrml/binary.

[10] J. Rossignac. Edgebreaker (1999) : Connectivity compression for triangular meshes. *IEEE Transactions on Visualization and Computer Graphics*, 5(1):47–61.
[11] C. Bajaj, V. Pascucci, and G. Zhuang.Single(1999) *Resolution compression of arbitrary triangular meshes with properties -InIEEE Data Compression Conference Proceedings*.
[12] C. Touma and C. Gotsman (1998).*Triangle mesh compression* - In *Graphics Interface Conference Proceedings*, Vancouver.
[13] H. Hoppe (1996)*Progressive meshes - In Siggraph'96 Conference Proceedings*, pages 99–108, August 1996.
[14] H. Hoppe (1998) Efficient implementation of progressive meshes. *Computers& Graphics, 1998*.
[15] G. Taubin,W. Horn, and P. Borrel (1999). Compression and transmission of multi-resolution clustered meshes. *Technical Report RC-21398, IBM Research*, February 1999.
[16] F. Bossen (1999) On The Art Of Compressing Three-Dimensional Polygonal Meshes And Their Associated Properties.PhD thesis, ÉcolePolytechniqueFédérale de Lausanne (EPFL), June 1999.
[17] A. Guéziec, G. Taubin, F. Lazarus, and W.P. Horn (1998).*Converting sets of polygons to manifold surfaces by cutting and stitching. In IEEE Visualization'98 Conference Proceedings*, pages 383–390.
[18] E.S. Jang, S.J. Kim, M. Song, M. Han, S.Y. Jung, and Y.S. Seo (1998). Results of ce m5 error resilient 3D mesh coding. *ISO/IEC JTC 1/SC 29/WG 11 Input Document No. M4251*.
[19] M. M. Siddeq, M. A. Rodrigues (2014) A Novel Image Compression Algorithm for high resolution 3D Reconstruction, *3D Research. Springer Vol. 5 No.2*.DOI 10.1007/s13319-014-0007-6
[20] M. M. Siddeq, M. A. Rodrigues (2015) A Novel 2D Image Compression Algorithm Based on Two Levels DWT and DCT Transforms with Enhanced Minimize-Matrix-Size Algorithm for High Resolution Structured Light 3D Surface Reconstruction, *3D Research. Springer Vol. 6 No.3*.DOI 10.1007/s13319-015-0055-6
[21] Knuth, Donald (1997). Sorting and Searching: Section 6.2.1: Searching an Ordered Table, *The Art of Computer Programming (3rd Ed.), Addison-Wesley*. pp. 409–426. ISBN 0-201-89685-0
[22] I.E. G.Richardson (2002)Video Codec Design, *JohnWiley&Sons*.
[23] K. Sayood, (2000) Introduction to Data Compression, *2nd edition, Academic Press, Morgan Kaufman Publishers*.

Self-propulsion Simulation of ONR Tumblehome Using Dynamic Overset Grid Method

J.H. Wang, W.W. Zhao, and †D.C. Wan

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

†Corresponding author: dcwan@sjtu.edu.cn

Abstract

With the great progress in supercomputers and the numerical methods, the application of computational fluid dynamics are advancing rapidly in the field of ship hydrodynamics. And the dynamic overset grid method makes it possible for computing complex ship motions, e.g. ship self-propulsion with moving propellers and rudders. In the present work, CFD-based method coupling with dynamic overset grid technique is applied to investigate the hydrodynamic performance of the fully appended ONR Tumblehome ship model during self-propulsion condition. Open water performance of propeller and towing condition of bare hull are computed before the self-propulsion simulation. The ship model is fitted with twin rotating propellers and twin static rudders, achieving self-propulsion model point at Fr=0.2 and Fr=0.3, respectively. All the computations are carried out by our in-house CFD solver naoe-FOAM-SJTU, which is developed on the open source platform OpenFOAM and mainly composed of a dynamic overset grid module and a full 6DOF motion module with a hierarchy of bodies. The CFD code naoe-FOAM-SJTU solves the Navier-Stokes equations for unsteady turbulent flows with VOF method capturing free surface around the complex geometry models. During the self-propulsion simulation, a feedback controller is used to update the rate of revolutions of the propeller to achieve the target speed. Detailed information of the flow field during the self-propulsion condition is presented and analyzed. In addition, predicted results, i.e. ship motions and force coefficients, are also presented and compared with the available experimental data. Good agreements are achieved which indicates that the present approach is applicable for the self-propulsion simulation.

Keywords: Overset grid, naoe-FOAM-SJTU solver, self-propulsion, ONR Tumblehome ship

Introduction

Self-propulsion is a key standard to examine a ship's powering performance and is closely bound up with energy consumption. With the coming out of energy efficiency design index (EEDI) proposed by IMO, more attention is devoted to the research of ship self-propulsion character. Thus how to evaluate the self-propulsion characteristics at the design stage is of great importance and the studies in this area have been extensively progressed. However, great challenges show up with the complexity of the flow field and interaction between hull, moving rudders and rotating propellers. When dealing with the fully appended ship, the vortical structures separated from the hull and appendages can be even more complicated. Among the available approaches to perform CFD simulation of self-propulsion, direct self-propulsion simulation with discretized module of fully appended hull, rotating propeller and

moving rudder is the one least reliant on geometries. Furthermore, self-propulsion requires capabilities of 6DOF module of a hierarchy of bodies in a free surface environment. All the above aspects increase the difficulties in direct simulating the self-propulsion problems.

Up to now, the main approach for predicting self-propulsion still strongly relies on the experimental results, in which model scale experiments in a conventional towing tank account for the main part. It can give high accurate results for the experiments but conversely at high cost. Nowadays, the use of CFD based method for self-propulsion prediction is becoming more and more popular as numerical algorithms improve and computers gain in power. Increasing demand of high accuracy for ship self-propulsion prediction has made it essential to develop full numerical simulation model for ship hull, propeller and rudder. In addition, the dynamic overset grid method, including a hierarchy of bodies that enable computation of ship motions with moving components, makes it possible to directly compute self-propulsion with rotating propellers and moving rudders. So far, overset grid method has been applied to the computations of ship hydrodynamics, especially for the direct simulation of hull-propeller-rudder interaction. Carrica et al. (2010)[1] use a speed controller and a discretized propeller with dynamic overset grids to directly perform the self-propulsion computations. Three ship hulls are evaluated, i.e. the single-propeller KVLCC1 tanker appended with a rudder, the twin propeller fully appended surface combatant model DTMB 5613, and the KCS container ship without a rudder, and good agreements with experimental data show that direct computation of self-propelled ships is feasible. Castro et al. (2011)[2] investigate the full-scale computations for self-propelled KRISO container ship KCS using discretized propeller model, and give the conclusion that the propeller operates more efficiently in full scale and is subject to smaller load fluctuations. Shen et al. (2015)[3] implement dynamic overset grid module to OpenFOAM and apply to the KCS self-propulsion and zigzag maneuvering simulation. Direct simulated results show good agreements with the experimental data, which show that the fully discretized model with overset grid method is applicable for the computations of ship hull, propeller and rudder interaction.

The present paper shows our recent progress in the numerical prediction of self-propulsion for fully appended ONR Tumblehome using overset grid method. Discretized model for rotating propellers and moving rudders are used in the simulation. Emphasis is put on the hydrodynamic performance for self-propulsion in different Froude numbers, i.e. $F_r = 0.20, F_r = 0.30$. The main framework of this paper goes as following. The first part is the numerical algorithm and solver, where naoe-FOAM-SJTU solver and overset grid method are presented. The second part is the geometry model and grid distribution. Then comes the simulation part, where towing condition, open water calculation and self-propulsion will be presented systematically. In this part, extensively comparisons are performed against the experimental data including ship motions and hydrodynamic coefficients. Following this part is the grid uncertainty study for towing condition at $F_r = 0.30$. Finally, a conclusion of this paper is drawn.

Numerical algorithm and solver

naoe-FOAM-SJTU solver

The in-house CFD code naoe-FOAM-SJTU applied in this study solves the Navier-Stokes equations for unsteady turbulent flows and using VOF method to capture free surface around the complex geometry models. The main framework and features of naoe-FOAM-SJTU solver are only briefly described here; detailed information can be referred to Shen et al.

(2014, 2015)[3,4], Cao et al. (2014)[5], and Wang et al. (2015a, 2015b)[6,7]. The solver is based on the open source platform OpenFOAM and consists of self-developed modules, i.e. a velocity inlet wave-making module, a full 6DOF module with a hierarchy of bodies and a mooring system module. The solver has the capability of handling varies problems in naval architecture and ocean engineering, i.e. large motion response prediction for ship and platforms in ocean waves; ship resistance, seakeeping prediction; direct simulations of self-propulsion and free maneuvering with moving rudders and rotating propellers.

The unsteady RANS equations and VOF transport equation are discretized by the finite volume method (FVM). The merged PISO-SIMPLE (PIMPLE) algorithm is applied to solve the coupled equations for velocity and pressure field. The Semi-Implicit Method for Pressure-Linked Equations (SIMPLE) algorithm allows to couple the Navier-Stokes equations with an iterative procedure. And the Pressure Implicit Splitting Operator (PISO) algorithm enables the PIMPLE algorithm to do the pressure-velocity correction. Detailed description for the SIMPLE and PISO algorithm can be found in Ferziger and Peric (1999)[8] and Issa (1986)[9]. Near wall treatment wall functions are applied to the moving wall boundary, which can reduce computational grid with coarse layer near the ship (y^+ can be more than 30). In addition, several built-in numerical schemes in OpenFOAM are used in solving the partial differential equations (PDE). The convection terms are discretized by a second-order TVD limited linear scheme, and the diffusion terms are approximated by a second-order central difference scheme. Van Leer scheme (Van Leer, 1979)[10] is applied for VOF equation discretization and Euler scheme is used for temporal discretization.

Overset Grid Method

The overset grid method is of great importance for direct simulating the full coupled hull, propeller and rudder system. Here a brief introduction for the utilization of overset grid module in naoe-FOAM-SJTU solver is presented. Overset grid is a grid system that made up of blocks of overlapping structured or unstructured grids. By using dynamic overset grid technique, the overlapping grids can move independently without any constraints. To this aim, the cells in the computational domain are classified into several types, i.e. fringe, hole, donor etc. The information of cell types is stored in the domain connectivity information (DCI) file. In our present solver, Suggar++ (Noack et al., 2009)[11] is utilized to generate the domain connectivity information (DCI) for the overset grid interpolation. To combine OpenFOAM with Suggar++, a communication, which is responsible for DCI exchange between OpenFOAM and Suggar++, has been implemented using the Message passing interface (MPI) library (Shen et al., 2015)[3]. Other features consist of a full 6DOF motion module with a hierarchy moving components and several modifications for sparse matrix solvers and MULES solver to excluded non-active cells. The flowchart of the parallel calculation between OpenFOAM processor and Suggar++ processor is shown in Figure 1.

By using overset grid method, the full 6DOF motion solver allows the ship and its appendages as well as the moving components to move simultaneously. Two coordinate systems are used to solve the 6DOF equations. One is the inertial system (earth-fixed system) and the other is non-inertial system (ship-fixed system). The inertial system can be fixed to earth or move at a constant speed with respect to the ship (here we only apply the horizontal motion for the moving of inertial system). The non-inertial system is fixed to the ship and can translate or rotate according to the ship motions. More information of the 6DOF motion solver with overset grid module implementation can be followed Shen et al. (2015)[3]. In our present study, the computational domain is decomposed into several overlapping grids, where each moving component has its own grid to deal with complex motion problems.

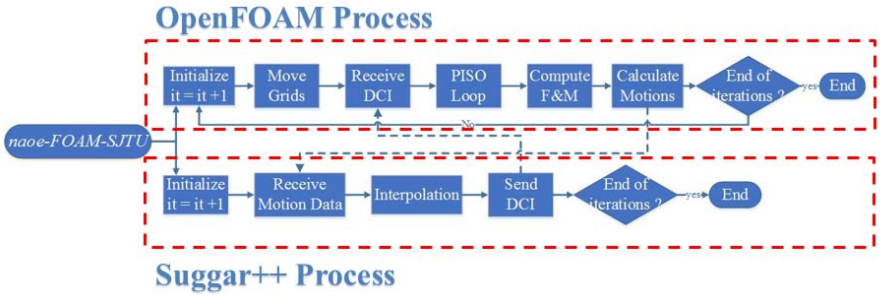


Figure 1 Flowchart of the calculation procedure

Geometry, grid and test conditions

Geometry model and computational domain

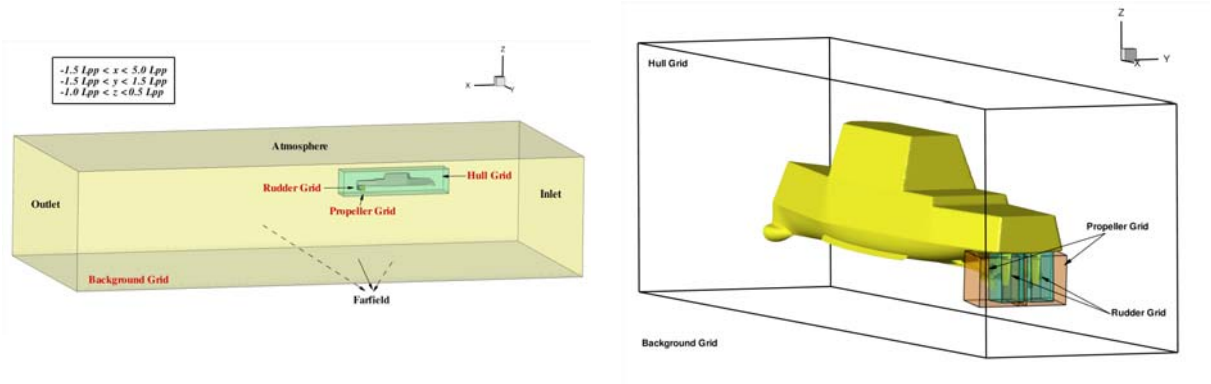
The present numerical simulations are carried out for the ONR Tumblehome model 5613, which is a preliminary design of a modern surface combatant fully appended with skeg and bilge keels. The ship model also involves rudders, shafts and propellers with propeller shaft brackets. The geometry model of ONR Tumblehome without propellers and shaft brackets is shown in Figure 2, and its principle parameters are listed in Table 1. The ship model is used as one of the benchmark cases in Tokyo 2015 CFD workshop in ship hydrodynamics. Experiments were widely performed in IIHR wave basin for this ship model and the available experimental data can be used to validate our present computational results.

Figure 2 Geometry model of ONR Tumblehome (from Tokyo 2015 CFD Workshop)

Table 1 Principle dimensions of fully appended ship

Main particulars		Model scale	Full scale
Length of waterline	$L_{WL}(m)$	3.147	154.0
Maximum beam of waterline	$B_{WL}(m)$	0.384	18.78
Depth	$D(m)$	0.266	14.50
Draft	$T(m)$	0.112	5.494
Displacement	$\Delta(kg)$	72.6	8.507e6
Wetted surface area (fully appended)	$S_0(m^2)$	1.5	NA
Block coefficient (CB)	$\nabla/(L_{WL}B_{WL}T)$	0.535	0.535
LCB	$aft.of\ FP(m)$	1.625	NA
Vertical center of gravity (from keel)	$KG(m)$	0.156	NA
Metacentric height	$GM(m)$	0.0422	NA
Moment of inertia	K_{xx}/B_{WL}	0.444	0.444
	K_{yy}/L_{WL}, K_{zz}/L_{WL}	0.246	0.25
Propeller diameter	$D_P(m)$	0.1066	NA
Propeller shaft angle (downward pos.)	$\varepsilon(^\circ)$	5	NA
Propeller rotation direction (from stern)		inward	inward

Using dynamic overset grid technique, here we have four parts of the computational grids, i.e. grid around ship hull, propeller grid, rudder grid and background grid. Background grid is the root element during the hole-cutting procedure, and the hull grid is at parent motion level with children grid of twin propellers and rudders. The four grid blocks have overlapping areas, which can move independently without restrictions and build connections among them by interpolation at appropriate cells or points. The computational domain arrangement in global and local view is shown in Figure 3. For the self-propulsion computation, the background domain extends to $-1.5L_{pp} < x < 5.0L_{pp}$, $-1.5L_{pp} < y < 1.5L_{pp}$, $-1.0L_{pp} < z < 0.5L_{pp}$, and the hull domain has a much smaller region with a range of $-0.15L_{pp} < x < 1.2L_{pp}$, $-0.13L_{pp} < y < 0.13L_{pp}$, $-0.2L_{pp} < z < 0.2L_{pp}$.



a) Global view b) Local view

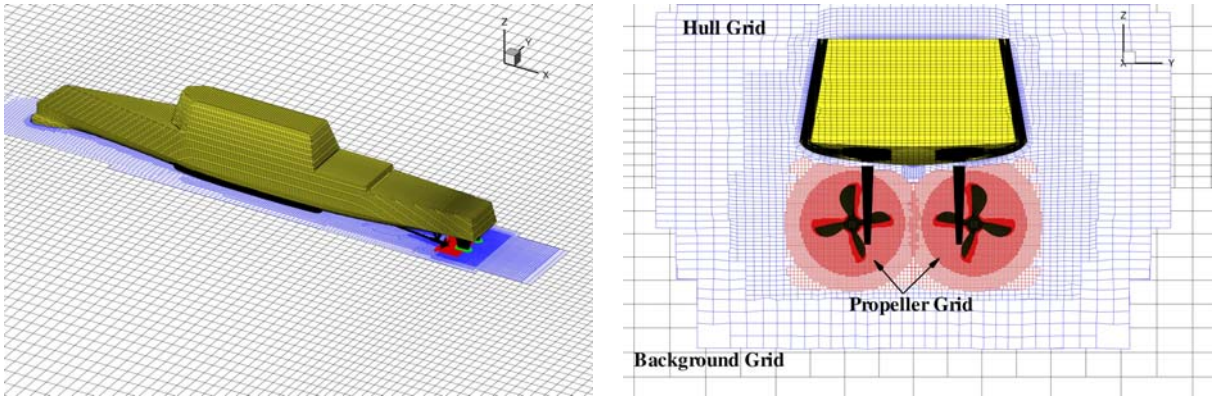
Figure 3 Computational domain for self-propulsion computation

Grid distribution

Fully unstructured grids used in this paper are generated by *snappyHexMesh* with the background grid generated by *blockMesh*, both are pre-processing utility provided by OpenFOAM. The total grid number for the self-propulsion simulation is 6.81M and the detailed grid information in each part is shown in Table 2. Considering the grid quality in overlapping areas, several refinement regions are applied to offer enough donor cells for interpolation. Grids in gaps should be handled specifically, i.e., the grid dimensions of different grid blocks in overlapping areas should be similar. Good grid quality at overlapping areas can resolve better flow information and reduce the computational cost. The global and local grid distribution around ship hull is shown in Figure 4.

Table 2 Grid distribution in each part

Grid	Total	Port	Starboard	Level
Background	1.34M	NA	NA	Highest
Hull	2.61M	NA	NA	Parent
Propeller	2.28M	1.14M	1.14M	Children
Rudder	0.58M	0.29M	0.29M	Children
Total	6.81M	NA	NA	NA



a) Global view b) Local view

Figure 4 Grid distribution around ship hull

Test conditions

The present work is for self-propulsion computation of ONR Tumblehome model. According to the experimental setup, the fully appended ship is set to advance at model point in calm water with rotating propellers and rudders. In the present simulation, two approaching speeds, i.e. U=1.110m/s and $U=1.667m/s$, corresponding to Froude number of $F_r=0.20$ and $F_r=0.30$, are taken into account to further investigate the self-propulsion performance of the fully appended ONR Tumblehome model. Note that both simulations are performed with the same overlapping grids, since wall functions can allow the y^+ in the range of 30-200.

Simulation results and analysis

When dealing with self-propulsion problems, the initial condition for the computation is interpolated from the final solution of the towing condition with the utility *mapFields* supported by OpenFOAM. This pre-processing step can save large amount of computational time by starting with a developed flow field and boundary layer. The initial ship speed was set to the target cruise speed and the rate of resolutions of propeller is static at the beginning. A feedback proportional-integral (PI) controller is applied to adjust the rotational rate of the propeller to achieve the target ship speed. Detailed information for the PI controller can be referred to Shen et al. (2015)[3]. The proportional and integral constants were set to 800 with the consideration of large PI constants can accelerate the convergence of the propeller revolution rate and reduce the total computation time.

Towing condition

The simulation of towing condition is followed by the experimental setup, and the advancing speeds are U=1.110m/s and $U=1.667m/s$, corresponding to $F_r=0.20$ and $F_r=0.30$. The computations are carried out without appendages and moving components. Overset grid approach is also applied in this simulation, and the computational domain is separated into the hull grid and background grid. The total grid number is 1.87M, with 0.82M for hull grid and 1.05M for background grid. Boundary conditions are identical with zero velocity and zero gradient of pressure imposed on inlet and far-field boundaries, while the boundary condition of interface between two grids is set to overlap for flow information interpolation.

During the computation, the ship model is advancing at the desired speed while the remaining 5 freedoms of degree are constrained. Through this way, the calculated flow field can be used as an initial state for the self-propulsion simulation. As a consequence, this step can save large amount of computational time by starting the calculation with a developed flow field

and boundary layer. Cook (2011)[12] investigate the appendages effect on the total resistance for ONR Tumblehome model at different Froude numbers, and the comparison between the present numerical results and the experimental data as well as CFD results by IIHR are listed in Table 3. An obvious phenomenon can be observed from the table that the total resistance of bare hull without bilge keels is much smaller than the fully appended model. The present results for the bare hull resistance show good agreement with the EFD data performed at INSEAN and the CFD results from IIHR. Figure 5 shows the convergence curves of the three components of ship resistance, i.e. $F_t, F_v, and\ F_p$ at $F_r = 0.30$ in 50s. Satisfactory agreements for the towing condition are achieved and high accuracy result can give a better initial state of the self-propulsion simulation. In addition, to further validate our numerical results, grid uncertainty analysis is performed for the towing condition, which will be described in the grid uncertainty analysis part.

Table 3 Total resistance comparison with bare hull simulation

F_r	IIHR EFD fully appended	INSEAN EFD bare hull w/o BK	IIHR CFD bare hull w/o BK	naoe-FOAM-SJTU bare hull w/o BK
0.20	4.54 N	-18.6%	-15.7%	-17.9%
0.30	11.30 N	-19.0%	-20.8%	-18.3%

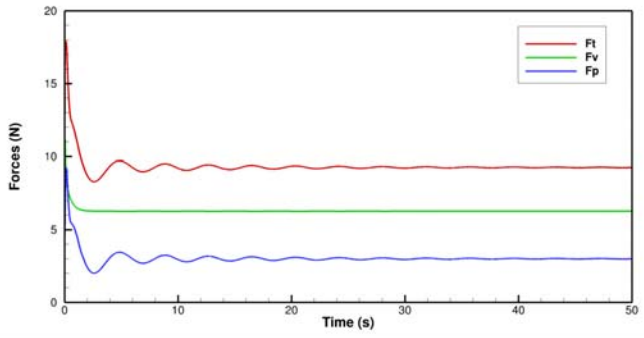
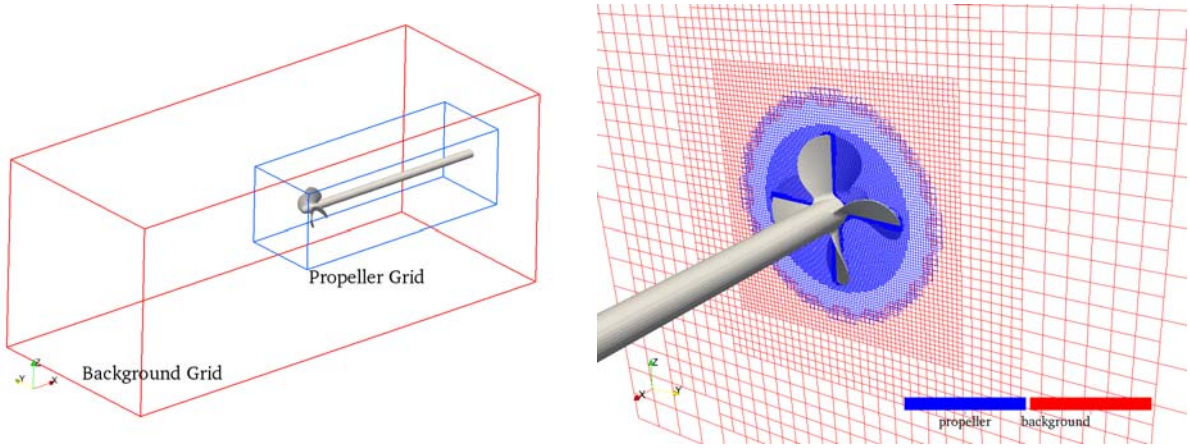


Figure 5 Time histories of the ship resistance at Fr=0.3

Open water calculations

Open water calculations for the propeller is carried out before the self-propulsion simulation. In the present study, open water curves are obtained by the single-run procedure described in Xing et al. (2008)[13]. As for the single-run procedure, the propeller is towing at a small acceleration to fulfil a wide range of advancing velocities in one turn. Based on overset grid, the computational domain is separated into two parts, i.e. background grid and propeller grid (Figure 6a). When doing the calculation, the propeller grid rotates with the rotating propeller while the background grid moves forward with the propeller advancing velocity. The total number of the computational grids is 1.13M with 0.51M for propeller grid and 0.62M for background grid. The grid distribution around propeller disk is shown in Figure 6b. Calculated open water curves are compared to the experimental results performed by IIHR (available at Tokyo 2015 CFD Workshop). The comparison between the numerical results and experimental data can be used to validate the current dynamic overset grid method coupled with single-run approach in simulating rotating propellers.



a) Computational domain b) Grid distribution around propeller

Figure 6 Computational domain and grid distribution for open water calculation

During the procedure, the rate of resolutions of propeller is set to fixed value *n*=8.97 r/s according to the test model point for self-propulsion at $F_r = 0.20$. Note that open water curves can be obtained by different rotating speed of propeller using single-run approach, and here we use the model point at $F_r = 0.20$ with the consideration of larger time step can be applied at low speed of propeller. Large range of advancing speed is performed to achieve the desired advance coefficient J. Thrust coefficients K_T, torque coefficient K_Q and efficiency η_0 for each advance coefficient are obtained from the calculated thrust and torque. The propulsive coefficients mentioned above are defined as:

$$J = \frac{V_A}{nD_P} \tag{1}$$

$$K_T = \frac{T}{\rho n^2 D_P{}^4} \tag{2}$$

$$K_Q = \frac{Q}{\rho n^2 D_P{}^5} \tag{3}$$

$$\eta_0 = \frac{JK_T}{2\pi K_Q} \tag{4}$$

where T and Q are the propeller thrust and torque, D_P is the diameter of propeller, n is the RPS and V_A is the advancing speed. The propeller accelerates from $V_A = 0\,m/s$ to $V_A = 1.721 m/s$ in 10 seconds with advance coefficient various from J=0 to J=1.8. The predicted results of the open water curves are shown in Figure 7 and overall agreement is achieved according to the comparison with the experiment. However, the numerical results for torque coefficient K_Q and thrust coefficient K_T are not so good at both the beginning and the end. Figure 8 shows the vortical structures using isosurface of Q=200 and colored by the axial velocity at three advancing coefficients, i.e. $J = 0.9, J = 1.0$ *and* $J = 1.1$. Tip vortices of the propeller are resolved clearly, and a decreasing with the strength of vortices is experienced with the increasing of advance coefficient. This phenomenon is mainly due to the angle of attack decreases when the advance coefficient becomes larger.

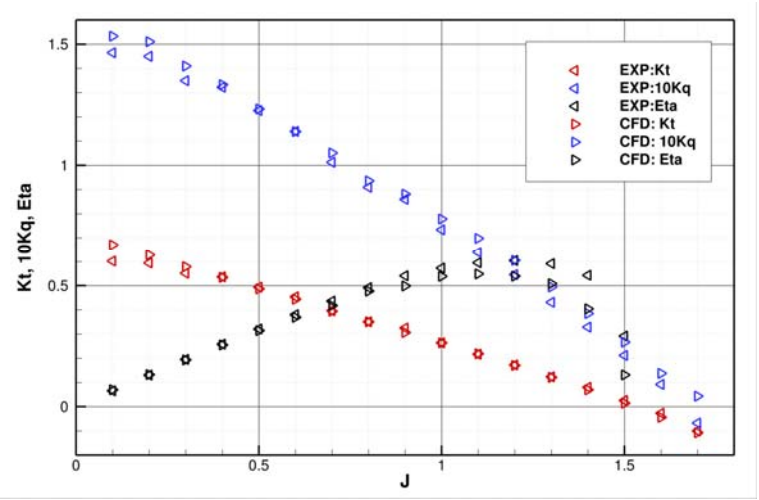
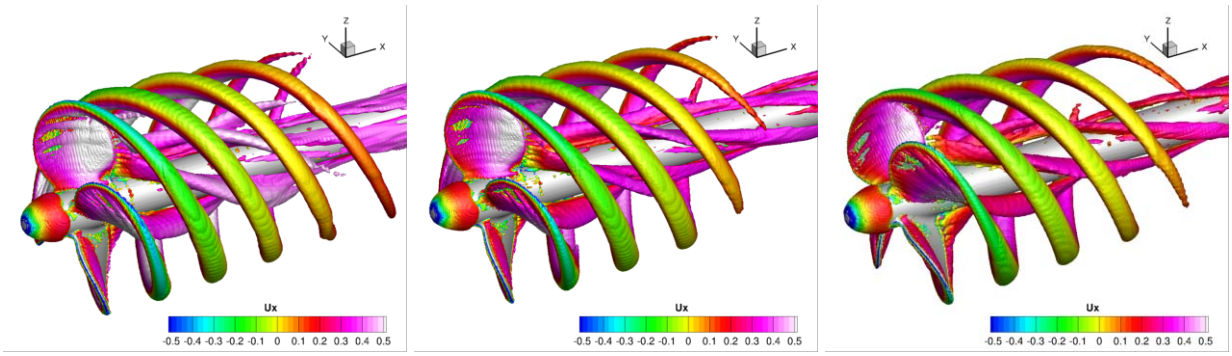


Figure 7 Open water results by experiment (left triangle) and CFD (right triangle)

The hub vortices of the propeller experienced the same trend with the strength decreasing. With rather coarse mesh and the RANS turbulence model, the evolution of vortical structures is relatively stable. In spite of the limitation of RANS model, the calculated coefficients K_T, K_Q and η_0 show overall agreement with the experimental data by the present dynamic overset grid method coupled with single-run approach.



a) J=0.9 b) J=1.0 c) J=1.1

Figure 8 Isosurfaces of Q=200 at different advance coefficients colored by axial velocity

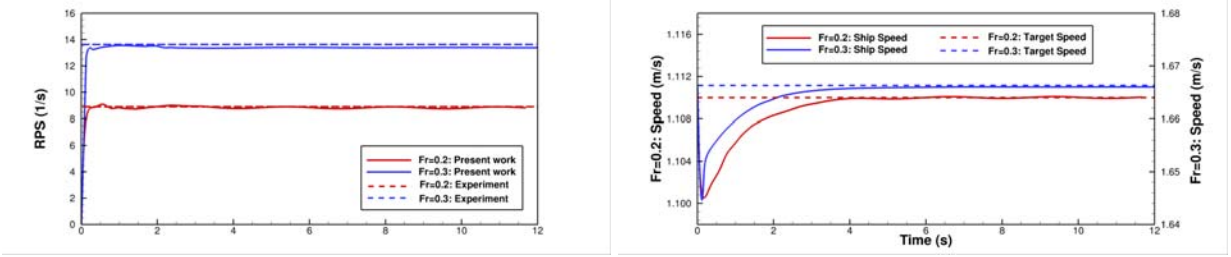
Self-propulsion simulation

As mentioned in test conditions, two approaching speeds, i.e. U=1.110m/s, U =1.667m/s, are performed for the self-propulsion simulation. The former situation is one of the benchmark cases (case 3.9) in Tokyo 2015 Workshop on CFD in ship hydrodynamics. And the experiment data for the latter one is also available in Elshiekh (2014)[14]. According to the experimental setup, the fully appended ship is set to approaching at model point in calm water. The twin rotating propellers, updating RPS by a feedback PI controller, provide the thrust for the ship to move forward. Overset grid arrangement and mesh distribution is described in Figure 3-4, and the size of each part grid is shown in Table 2.

The initial state of the simulation is obtained by interpolating data from the final flow field of towing condition to accelerate the convergence of the calculation. The interpolation is

conducted by the *mapFields* utility, which is a pre-processing tool supported by OpenFOAM. During the self-propulsion simulation, the twin propellers start from static state and speed up the rotational velocity to provide enough thrust. The proportional and integral coefficients P and I are set to 800 and the detailed process of the PI controller can be referred to Shen et al. (2015)[3].

The time histories of the rate of resolutions (RPS) of propellers and ship model advancing speed for both conditions are shown in Figure 9.



a) RPS b) Ship speed

Figure 9 Time histories of RPS and ship speed

Both time histories of the RPS start from zero and increase quickly and the curves of the RPS converge to the desired the value in about 5 seconds at model scale. According to Figure 9b, the ship speed first decreases due to less thrust provided by the rotational propellers and with the increasing RPS of propellers, the available thrust can prompt the ship speed comes back to the target value. In addition, the time histories of ship speed describe the characters at the beginning of different conditions, where the increasing rate of speed as well as the speed loss for $F_r = 0.30$ are larger than that of $F_r = 0.20$. This is mainly due to the fact that larger target speed requires larger thrust, thus more speed loss at beginning with static propeller and larger increasing rate with higher RPS of propeller. Figure 9 also presents the test results for the rate of resolutions of propeller (RPS) and target ship speed. Numerical results of both RPS and ship speed can finally achieve a stable desiring state.

Table 4 lists the numerical results of ship motions and self-propulsion coefficients. All the predicted force coefficients are in non-dimensional format using the provided wetted surface area at rest S_0, fluid density ρ and ship advancing speed U. The force coefficients are defined as follows:

$$C_T = \frac{R_T}{\frac{1}{2}\rho U^2 S_0} \tag{5}$$

$$C_V = \frac{R_V}{\frac{1}{2}\rho U^2 S_0} \tag{6}$$

$$C_P = \frac{R_P}{\frac{1}{2}\rho U^2 S_0} \tag{7}$$

Table 4 Numerical results for ship motions and self-propulsion coefficients

Parameters	$F_r = 0.20$			$F_r = 0.30$		
	CFD	EFD*	Error	CFD	EFD*	Error
u (m/s)	1.109	1.125	-1.4%	1.664	1.667	-0.2%
sinkage $\sigma \times 10^2$ (m)	2.41E-1	2.26E-1	6.5%	5.78E-1		
trim τ(deg)	4.64E-2	3.86E-2	20.3%	7.81E-2		
$C_T \times 10^3$	5.291			5.465		
$C_V \times 10^3$	1.539			3.310		
$C_P \times 10^3$	3.752			2.155		
$n(RPS)$	8.819	8.97	-1.7%	13.389	13.684	-2.16%
K_T	0.242			0.246		
K_Q	0.616			0.673		

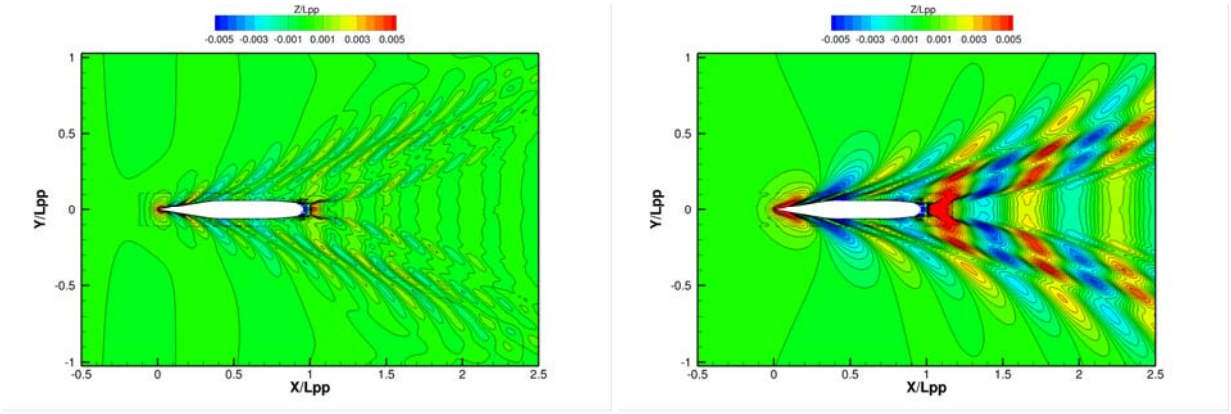
*the sinkage and trim of experimental data at $F_r = 0.20$ is available at Tokyo 2015 CFD Workshop and is not available for $F_r = 0.30$， so only numerical results are presented.

Note that the computation is carried out to predict the self-propulsion model point and the propulsion coefficients are obtained by the predicted results, none of the coefficients except *n* can be compared with the measured data. So only parts of the results are compared with the experiment. Table 4 gives a general comparison for ship motions and force coefficients. It shows that the present CFD approach can precisely achieve the desired ship speed and the computational results of ship motions can also give a general performance compared with the experiments. The sinkage and trim are overpredicted in high Froude number, while the thrust coefficient K_T and torque coefficient K_Q are at the same level. In addition, according to the simulated force coefficients, the viscous coefficient C_V accounts for the main part of the total resistance at $F_r = 0.30$, while the pressure coefficient C_P occupies a dominant place at $F_r = 0.20$. This further confirms that viscous effect plays an important role at high Froude numbers, especially when $F_r > 0.30$.

The rate of revolutions of the propeller *n* computed by our own solver naoe-FOAM-SJTU is 8.819 and 13.389 for $F_r = 0.20$ and $F_r = 0.30$, respectively. Both results are underestimated within 3% compared with the experimental data. The high accuracy of the predicted rate of resolutions of propeller confirms that the present dynamic overset grid approach is applicable to predict the model point for free running ship model.

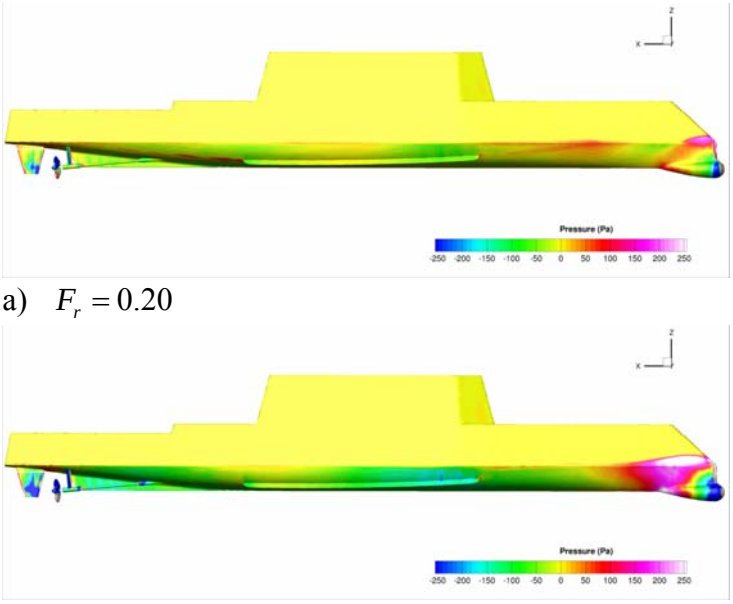
Figure 10 shows the wave patterns for self-propulsion at different conditions. The flow region and velocities are non-dimensioned by the ship model length L_{WL} and magnitude velocity U. Both the wave height and wave length at F_r=0.30 is significantly larger. Pressure distribution around ship hull, twin propellers and rudders is shown in Figure 11. The distribution at different Froude number has a consistent relationship to the wave patterns. Larger bow wave results in larger pressure in the forehead of the ship hull. As for the pressure distribution around the twin propellers and rudders, pressure distribution experience

the same trend with the bow pressure. This is mainly due to the higher rotating speed at $F_r = 0.30$.



a) $F_r = 0.20$ b) $F_r = 0.30$

Figure 10 Wave patterns at different Froude number colored by nondimensional wave height Z / L_{pp}



a) $F_r = 0.20$

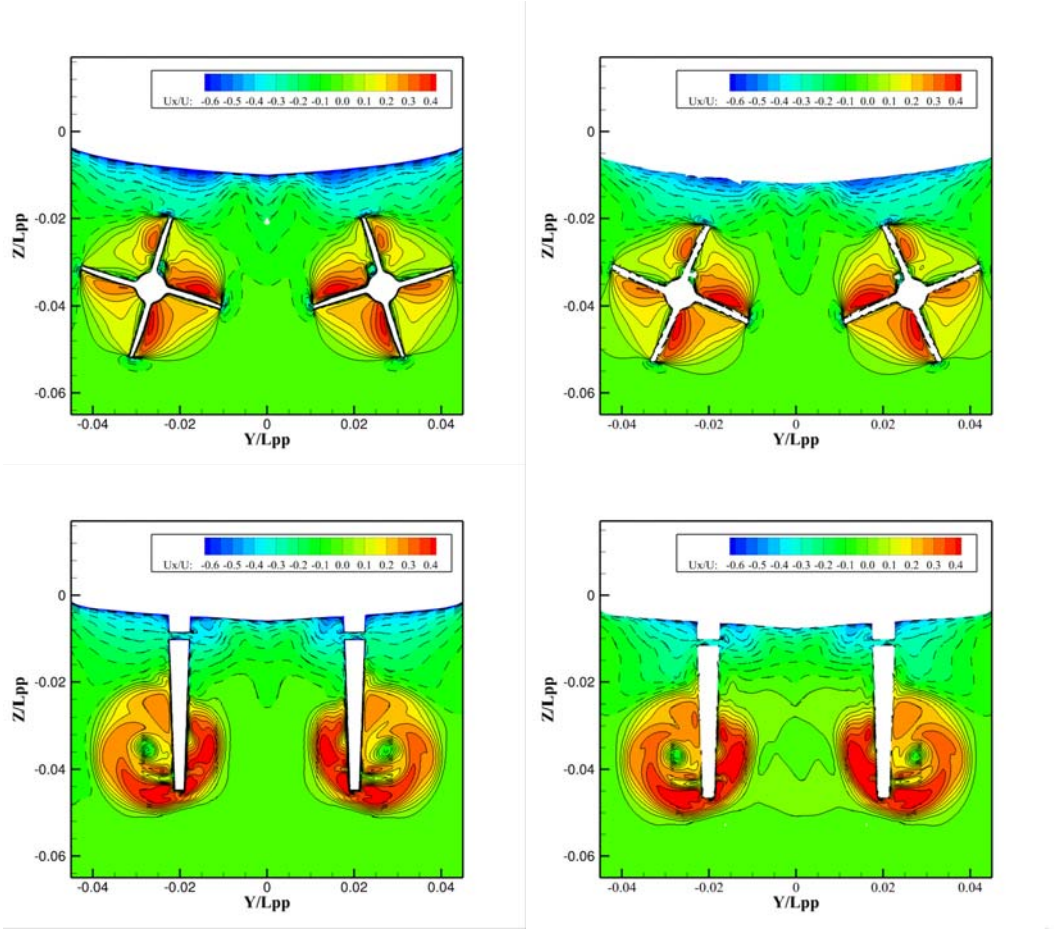
b) $F_r = 0.30$

Figure 11 Pressure distribution around ship hull, propellers and rudders

Figure 12 presents the detailed flow information at wake region, i.e. propeller disk ($X / L_{pp} = 0.909$) and the rudder section ($X / L_{pp} = 0.965$). From the figure we can see that the boundary layer around ship hull at high Froude number is thinner and the non-dimensional axial velocity is approximately the same, which can further explain the thrust coefficients are at the same level in different conditions. Little discrepancy is found for the wake distribution at the rudder section due to different vortex strength, which will be described later.

Figure 13 shows a profile view of vortical structures displayed as isosurface of Q=200 colored by axial velocity. According to the stern view of the vortical structure, tip vortices of the propellers are clearly resolved even when passing through the rudders, but dissipate quickly within the coarser mesh downstream. In addition, the strength of tip vortices is

stronger in higher Froude number, which can be clearly seen in the figure. The hub vortex observed is stronger and has a much larger size so that it is still somewhat resolved by the coarser grid downstream of the refinement. Another obvious phenomenon can be seen from the figure is that the vortices after the rudder root, which is caused by the artificial gap between the rudder and rudder root, and it will not appear in the real test.



a) $F_r = 0.20$ b) $F_r = 0.30$

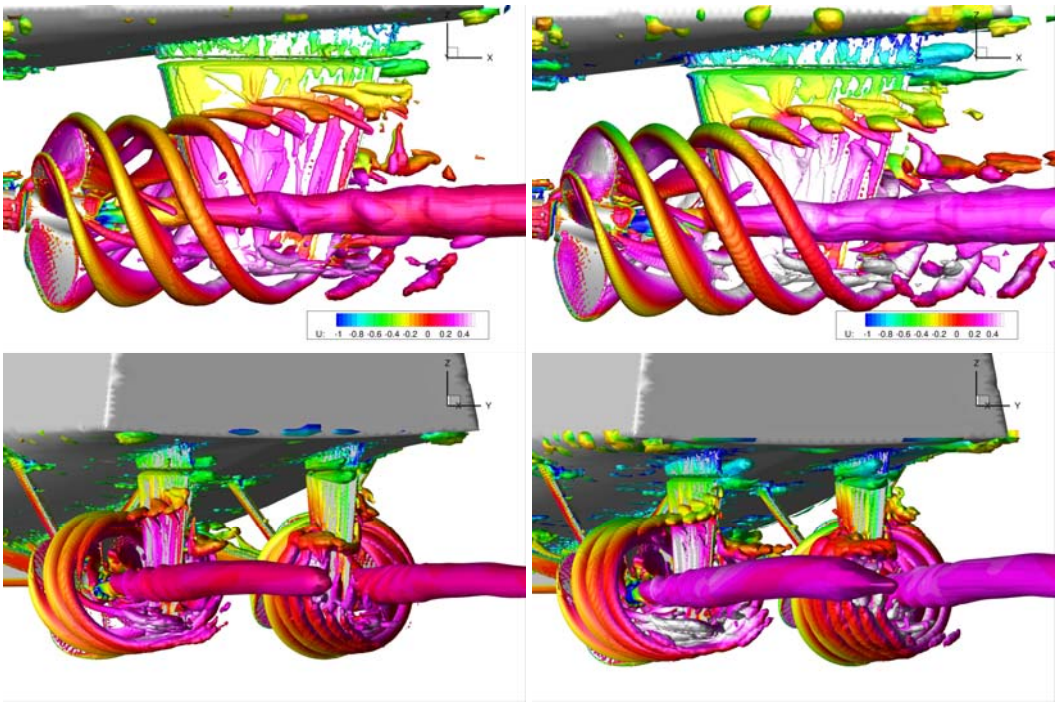
Figure 12 Wake distribution at different Froude number (upper for slice $X / L_{pp} = 0.909$ /propeller disk; lower for slice $X / L_{pp} = 0.965$ /rudder)

Figure 13 also shows the 3D view of vortical structure, where strong interaction between the propeller vortex and the rudder geometry is occurred. The strong hub vortex of the propeller is rarely affected by the following rudder, which is due to the fact that the axis of rudder has a distance away from the axis of propeller. An interesting effect occurs when the tip vortices of blades pass through the rudders, where the vortices are strongly affected by the rudder geometry both at the inward and outward side. In addition, little flow interaction is observed between the port side propeller and starboard side propeller (Figure 12, Figure 13). Furthermore, the strength of the hub vortex in higher Froude number is also stronger than the lower one at the same grid size. The strong flow interaction between the propellers and rudders can result in complex hydrodynamic performance of ship hull.

Grid uncertainty analysis

With the fact of the large amount of computing time required by the self-propulsion, grid uncertainty analysis is only conducted on the towing condition in the present work with

consideration of the simplicity of the overset grid arrangement in towing condition for bare hull calculation (only two part of grid is applied).



a) $F_r = 0.20$ b) $F_r = 0.30$

Figure 13 Profile and 3D view of vortical structures around twin propellers and rudders

Grid convergence study in the present work follows the verification methodology described in Stern et al. (2006)[15]. The convergence solution (R_G) of the different solutions (S_i, at least three) is defined as:

$$R_G = \frac{S_2 - S_1}{S_3 - S_2} \tag{8}$$

where $S_i, i = 1,2,3$, correspond to solutions with fine, medium, and coarse grid, respectively. Three convergence conditions are possible:

$$\begin{aligned} &(i) \quad 0 < R_G < 1 \quad \textit{Monotonic convergence} \\ &(ii) \quad R_G < 0 \quad \textit{Oscillatory convergence} \\ &(iii) \quad R_G > 1 \quad \textit{Divergence} \end{aligned} \tag{9}$$

For condition (*i*), generalized Richardson extrapolation (RE) is used to estimate grid uncertainty U_G. For condition (*ii*), uncertainties are estimated simply by attempting to bound the error based on oscillation maximums S_U and minimums S_L, i.e. $U_G = 1/2(S_U - S_L)$. While for condition (*iii*), errors and uncertainties cannot be estimated.

The grid convergence study is carried out for towing condition with bare hull at $F_r = 0.30$. Three grids with a refinement ratio of $\sqrt{2}$ in each direction are carried out for the grid convergence study. Considering the grids used in the present calculation is fully unstructured, the systematic refinement in three directions is very difficult. In order to do the grid

convergence study, an alternative approach is applied as follows. The background grid required by the snappyHexMesh is refined by splitting cells. Three systematic background grids with specified refinement ratio are taken into account. The final generated grids are approximately refined (not exactly the same) according to the grid convergence study. The results of the grid uncertainty is listed in Table 5.

Table 5 Grid uncertainty results for towing condition at $F_r = 0.30$

Grid	ID	Grid Size (M)	$C_P\,(10^{-3})$	$C_V\,(10^{-3})$	$C_T\,(10^{-3})$	Error
EFD					4.639	
Fine	S_1	3.65	1.549	3.098	4.647	0.17%
Medium	S_2	1.87	1.503	3.076	4.579	-1.29%
Coarse	S_3	0.68	1.690	3.124	4.814	3.77%
R_G			-0.246	-0.458	-0.289	
$U_G(\%S_2)$			4.691	4.226	1.824	
Convergence type			Oscillatory	Oscillatory	Oscillatory	

The force coefficients, i.e. $C_P, C_V, and\ C_T$, are used to estimate the grid uncertainty of the towing condition. The results have good convergence as shown in Table 5. All coefficients show oscillatory convergence with R_G of -0.246, -0.458, and -0.289, respectively. The C_P meets the maximum grid uncertainty with $U_G = 4.691\%$ and the grid uncertainty of total resistance coefficient C_T is only 1.824%, which confirms that the grid density has limited effect on the resistance in the current range of grid size.

Conclusions

This paper presents the self-propulsion simulations of fully appended ONR Tumblehome. Numerical simulations at two different speeds, i.e. $F_r = 0.20$, $F_r = 0.30$, are performed using in-house CFD solver naoe-FOAM-SJTU. During the simulation, the moving objects are handled by the dynamic overset grid method, and a feedback proportional-integral (PI) controller is employed to adjust the rotational rate of the propeller to achieve the desired ship speed.

Towing condition for bare hull model at different Froude numbers are carried out to give an approximate initial state of the self-propulsion computation. Predicted total resistance are compared with the experimental results and satisfactory agreement for bare hull is achieved. Furthermore, grid uncertainty analysis is performed with the bare hull towing condition at $F_r = 0.30$. All the predicted force coefficients show oscillatory convergence and the grid uncertainty of C_T is 1.824%, indicating that the grid density has limited effect on the resistance in the current range of grid size. Open water calculations are also carried out beforehand using the single-run method and the numerical results show an overall agreement with the experiment performed at IIHR.

The time histories of RPS and ship speed are converged to the target value in about 5s, and the increasing rate of speed as well as the speed loss for $F_r = 0.30$ is larger than that of

$F_r = 0.20$. In addition, according to the simulated force coefficients, the viscous coefficient C_V accounts for the main part of the total resistance at $F_r = 0.30$, while the pressure coefficient C_P occupies a dominant place at $F_r = 0.20$, which further confirms that viscous effect plays an important role with high Froude number, especially when $F_r > 0.30$. Predicted model point at different Froude number of self-propulsion simulation are underestimated by 1.7% and 2.16%, respectively. Detailed information of the flow field around twin propellers and rudders, i.e. wave patterns, wake distribution, pressure distribution, and vortical structures, at different Froude number are depicted and analyzed to explain the strong interaction among the ship hull, propellers and rudders.

Future work will focus on self-propulsion simulation in waves. Difficulties will be the direct simulating of moving propellers with large ship motions. More work will be done to do the free maneuvering simulation depending on the computed self-propulsion results.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (51379125, 51490675, 11432009, 51579145, 11272120), Chang Jiang Scholars Program (T2014099), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23) and Lloyd's Register Foundation for doctoral students, to which the authors are most grateful.

References

[1] Carrica, P.M., Castro, A.M., and Stern, F. (2010) Self-propulsion computations using a speed controller and a discretized propeller with dynamic overset grids, *Journal of Marine Science and Technology* **15**(4), 316–330.

[2] Castro, A.M., Carrica, P.M., and Stern, F. (2011) Full scale self-propulsion computations using discretized propeller for the KRISO container ship KCS, *Computers & Fluids* **51**(1), 35–47.

[3] Shen, Z., Wan, D., and Carrica, P.M. (2015) Dynamic overset grids in OpenFOAM with application to KCS self-propulsion and maneuvering, *Ocean Engineering* **108**, 287–306.

[4] Shen, Z., Zhao, W., Wang, J., and Wan, D. (2014) Manual of CFD solver for ship and ocean engineering flows: naoe-FOAM-SJTU, Shanghai Jiao Tong University.

[5] Cao, H., and Wan, D. (2014) Development of Multidirectional Nonlinear Numerical Wave Tank by naoe-FOAM-SJTU Solver, *International Journal of Ocean System Engineering* **4**(1), 52–59.

[6] Wang, J., Liu, X., and Wan, D. (2015) Numerical Simulation of an Oblique Towing Ship by naoe-FOAM-SJTU Solver, *Proceedings of 25th International Offshore and Polar Engineering Conference*, Big Island, Hawaii, USA.

[7] Wang, J., Liu, X., and Wan, D. (2015) Numerical prediction of free runing at model point for ONR Tumblehome using overset grid method, *Proceedings of CFD Workshop 2015*, Tokyo, Japan, 3, 383–388.

[8] Ferziger, J.H., and Peric, M. (2012) *Computational methods for fluid dynamics* Springer Science & Business Media.

[9] Issa, R.I. (1986) Solution of the implicitly discretised fluid flow equations by operator-splitting, *Journal of Computational Physics* **62**(1), 40–65.

[10] van Leer, B. (1979) Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method, *Journal of Computational Physics* **32**(1), 101–136.

[11] Noack, R.W., Boger, D.A., Kunz, R.F., and Carrica, P.M. (2009) Suggar++: An improved general overset grid assembly capability, *Proceedings of the 47th AIAA Aerospace Science and Exhibit* 22–25.

[12] Cook, S.S. (2011) Effects of headwinds on towing tank resistance and PMM tests for ONR Tumblehome, The University of IOWA.

[13] Xing, T., Carrica, P., and Stern, F. (2008) Computational towing tank procedures for single run curves of resistance and propulsion, *Journal of Fluids Engineering* **130**(10), 101102.

[14] Elshiekh, H. (2014) Maneuvering characteristics in calm water and regular waves for ONR Tumblehome, The University of IOWA.

[15] Stern, F., Wilson, R., and Shao, J. (2006) Quantitative V&V of CFD simulations and certification of CFD codes, *International Journal for Numerical Methods in Fluids* **50**(11), 1335–1355.

Hull form optimization based on a NM+CFD integrated method for KCS

***Aiqin Miao, Jianwei Wu and †Decheng Wan**

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Presenting author: maq046@163.com
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

It is a definite trend and hot topics of hull form optimal design based on computational fluid dynamics(CFD). Hull form optimization is carried out in this paper which combines the Neumann-Michell (NM) theory with CFD technology (NM+CFD integrated method) to OPTShip-SJTU, an optimization tool. The Free Form Deformation (FFD) method adopted for automatically modifying the hull form are illustrated. In order to reduce the overall highly computational effort, not only the surrogate model is established based on the samples produced by OLHS method and is used to directly predict the total resistance in optimization process, but also a NM+CFD integrated method, the NM theory for evaluating wave resistance and CFD technology based on RANS for evaluating viscous resistance of double body, are discussed to evaluate the total resistance of ships. In addition, NSGA-II, a multi-objective genetic algorithm, is implemented to produce pareto-optimal front. In the present paper the KRISO 3600TEU container ship model (KCS) is chosen as initial ship and optimal solutions with obvious total resistance coefficient reductions at specific speeds(at Fr=0.2, 0.26) are obtained. Eventually, one typical optimal hull is analyzed by a RANS-based CFD solver naoe-FOAM-SJTU. Numerical results confirm the availability and reliability of this multi-objective optimization tool.

Keywords: Hull form optimization, total resistance coefficient, FFD, OPTShip-SJTU, naoe-FOAM-SJTU solver.

Introduction

Ship designers often design a new ship mostly by their own experience in accordance with the requirements proposed by shipping companies[1]. Generally, designers can attempt to transform several initial ships with similar usage, similar shapes and as well as with satisfaction of ship owners during the operation and then predict and check performances of the new ship over and over. The above design process is a single-threaded circle, which mainly depends on designers' experience and intuition.

With the development of computer technologies and computational fluid dynamics(CFD), ship optimization design has raised the interest of researchers and designers, which is a converse process absolutely different from the traditional ship design process mentioned above. It is a process where to achieve the best performances of a new ship directly drives ship design. During the last several decades, a rapidly increasing number of papers devoted to ship optimization design based on hydrodynamic performance have been yielded with the advantage of optimization techniques and high-performance computer(HPC), resulting in the huge development of ship design[1]-[8]. Among these papers, the initial ships often adapted are Wigley[9][10], an internationally common and mathematical ship type, and S60, both with simple hull forms and a great quantity of experimental data. However, the increasing complexity of a real-life optimization problem in ship industry has raised the challenges for designers[], because hull form optimization of a complex geometry typically involve a large number of variables, different disciplines and conflicting objectives, requiring hundreds or thousands of function evaluations to converge to an optimal design .

Thanks to the development of some excellent modern optimization algorithms[11]-[14] such as the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) and particle swarm optimization (PSO), multi-objective optimization of ship hulls makes a significant breakthrough[17], and ongoing research is still much concerned about this topic[18][30].
Actually, how to quickly and accurately evaluate objective functions or the hydrodynamic performance during the optimization process is an important segment. Both of the potential-flow theory and the advanced RANS-based CFD method had been employed to predict the hydrodynamic performance during the hull optimization. If high-fidelity solvers based on CFD are used as analysis tools (e.g., RANS solvers), many conditional optimization methods become more and more expensive. However, the potential-flow theory can be used in evaluating the wave-making resistance in calm water because of the efficiency[31][32], and a RANS-based CFD method can be just used as predicting the viscous resistance with a double-model. Furthermore, the total resistance can be expressed as the sum of the wave-making resistance and viscous resistance[19].
In the present paper, KCS is chosen as the initial hull form to locally optimize its bow and its stern, respectively, based on the minimum total resistance coefficients at two specific speeds. First, the Design of Experiment is used to select a reasonable optimal design space. Specifically, optimized Latin hypercube sampling (OLHS) method is applied here which satisfies the requirements of orthogonality and uniformity to obtain different design variables, that is to say, different ship samples. Then, these ship samples are deformed by free form deformation(FFD). Next step is to evaluate their total resistances at two specific speeds, so called objective functions, where wave-making resistances are evaluated by NM theory and viscous resistances are evaluated by CFD-based naoe-FOAM-SJTU solver. So far, the surrogate model [23] is adopted to describe the complex relationship between the design variables and multi-objective funcitons, which largely decreases the optimization difficulty and computational cost. Last but not least, a vital multi-objective optimization process is completed by NSGA-II, a series of optimal ship hull obtained. The whole optimization frame can be seen as follows (Fig. 1) .

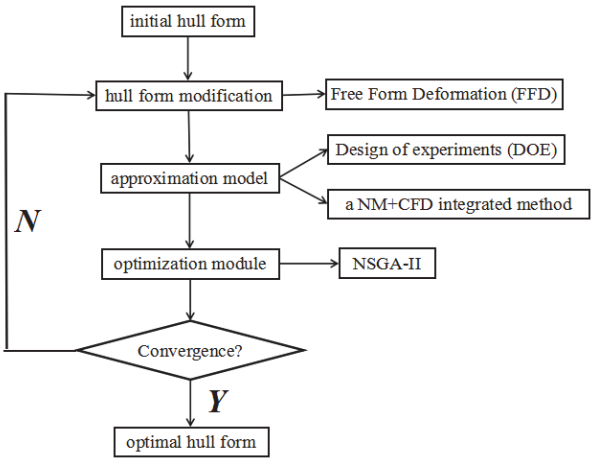


Figure 1. The flow chart of the iterative optimization process

Hull form deformation

An effective and rational method for hull form deformation is indispensable and crucial in the optimization process of ship design. One Hull form should be quickly and reasonably transformed to another new one. And there should be as less as possible deformation parameters involved in the optimization design, otherwise it will increase the complexity of the problem and lead to vast computational cost in multiples. Here FFD method is chosen to modify hull form locally, based on the idea of enclosing the ship within a cube, and transforming the hull form within the cube as the cube is deformed. FFD method was first

described by Thomas W. Sederberg and Scott R. Parry in 1986[21][22],and was based on an earlier technique by Alan Barr[20]. It is widely used in the optimization of ships, because less design variables related are involved in the optimization, surfaces are flexibly transformed, it is easily realized by making a program and so on. It is strongly dominant among many deformation methods that the main dimension of the initial ship can be limited and any new shape obtained by FFD method can be more reasonable and practical.
In this paper, FFD method is applied to locally modify the bulb bow and the stern of KCS, respectively.

Total resistance evaluation

The total resistance of ships can be solved according to two methods of division. One is according to the assumption of Froude, through his experiments, Froude realized that the ship resistance problem had to be broken into two different parts: residuary resistance (mainly wave making resistance) only related to Froude number (*Fr*)and frictional resistance only related to Reynolds number (*Re*). However, the influence of the two parts is ignored by using this method. Actually, especially for the fat full ship type, ΔC_f will be negative.
So in 1950s, Hughes proposed another method—three dimensional conversion, which was recommended as the standard conversion at ITTC in 1978. Through this conversion, total resistance(R_t) is broken into two new parts: wave-making resistance (R_w) related to Froude number and the viscous resistance (R_v) (the sum of the viscous pressure resistance and friction resistance) related to Reynolds number.

$$R_t = R_w + R_v \tag{1}$$

In this paper, the above standard conversion is chosen to predict the total resistance, wave-making resistance calculated by NM theory and viscous resistance calculated by simulating the flow field around the double ship model based on RANS, which is abbreviated as the NM+CFD integrated evaluation. Nobless et al. [23] present an efficient potential theory, Neumann-Michell (NM) theory, which provides more accurate prediction of wave-making resistance and wave profiles than the Hogner slender-ship approximation, with no appreciable increase in computational cost (seconds on a PC) for the classical Wigley parabolic hull. Besides, there are lots of research about comparison of experimental measurements of wave-making resistance with numerical predictions obtained using a preliminary version of the NM theory for the Wigley hull, the Series 60 and DTMB 5415 model[24]-[26]. A RANS-based CFD solver naoe-FOAM-SJTU, which is developed under the framework of the open source code, OpenFOAM, and has been validated in computation of a ship with heave and pitch motion in head waves[18].
The validation study for the NM+CFD integrated method is carried out before the optimization. For KCS, the results calculated by naoe-FOAM-SJTU and the NM+CFD integrated method and experimental data are respectively shown in Table 1.

Table 1. Total resistance coefficients predicted by the NM+CFD integrated method, CFD and experimental data.

Comparison	Speed	Fr=0.2	Fr=0.26
Ct	NM+CFD	3.72E-03	3.82E-03
	CFD	3.58E-03	3.84E-03
	EXP	3.46E-03	3.75E-03
Deviation	NM+CFD-EXP	-7.09%	-1.74%
	CFD-EXP	3.47%	2.40%

As shown in Tab. 1, the results based on the NM+CFD integrated method are within the error allowed, (-7.09% at Fr =0.2 and -1.74% at Fr=0.26), which is a little bit worse than the results totally based on CFD. Even so, this integrated method is still worth to be adopted because of its lower computational time cost. It's a huge advantage for hull form optimization design.

The definition of multi-objective optimization

Multi-objective optimization problem is a problem of multiple criteria decision making, that is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously. Multi-objective optimization problem has been applied in many fields of science, including engineering, economics and logistics where optimal decisions need to be taken in the presence of trade-offs between two or more conflicting objectives.

In mathematical terms, a multi-objective optimization problem can be formulated as

$$\min(f_1(x), f_2(x), ..., f_k(x))$$
$$s.t. x \in X$$

Where the integer $k \geq 2$ is the number of objectives and the set X is the feasible set of decision vectors.

In the ship industry, there is still a problem about the trade-offs between each performance of a new ship during the ship design process. The following content will clearly describe a complete multi-objective optimization of ship design.

The establishment of the optimization problem

For an entire optimization problem to be solved, the following basic items must be specified in detail: (1)an initial hull form to be optimized and the region(s) to be modified;(2)the objective function to be minimized and the design variables to be used;(3)the constraints to be defined. All of these items will be described in terms of the ship optimization presented by this paper.

Initial hull form

The initial hull form is the KRISO 3600TEU container ship model (KCS), which was conceived to provide data for both explication of flow physics and CFD validation for a modern container ship with bulb bow and stern. There is a large experimental database for KCS due to an international collaborative study on experimental/numerical uncertainty assessment between NMRI, MOERI and SVA[29].
The geometry of the initial model is presented in Fig.8 and the principal dimensions of KCS in table 2.

Figure 2. The geometry of KCS

Table 2. The principal dimensions of KCS

Principal Dimensions	full-scale ship	ship model
Length between perpendiculars L_{pp}/m	230	7.28

Length of waterlines L_{wl}/m	232.5	7.36
Breadth moulded B/m	32.2	1.019
Depth moulded D/m	19	0.6013
Draught T/m	10.8	0.3418
Block coefficient C_b	0.651	0.651

Multi-objective function and design variables

The multi-objective functions to be minimized is the total resistance coefficient of KCS sailing in calm water at two speeds of Fr=0.2, Fr=0.26. This condition corresponds to using a reference length of 7.36m, that is the length of the ship's model used in the experimental validation.

$$C_t = C_w + C_v \tag{2}$$

$$C_w = \frac{R_w}{0.5\rho U^2 S} \tag{3}$$

$$C_v = \frac{R_v}{0.5\rho U^2 S} \tag{4}$$

The deformation region is only the foremost part of the ship (x=3.45~3.99m)and the stern of the ship (x=-3.44~-0.44), with the origin of coordinates at the midship in Fig. 4. As explained in the introduction, this is the typical redesign problem of some part of an existing complex system, a necessity which often arises in real industrial applications. At the stern of the ship, two control boxes are used in order to modify the origin shape of the stern to any practical new one. In Fig. 5, some certain movable control points and the other fixed control points are clearly grouped into two kinds of colors, red and green.

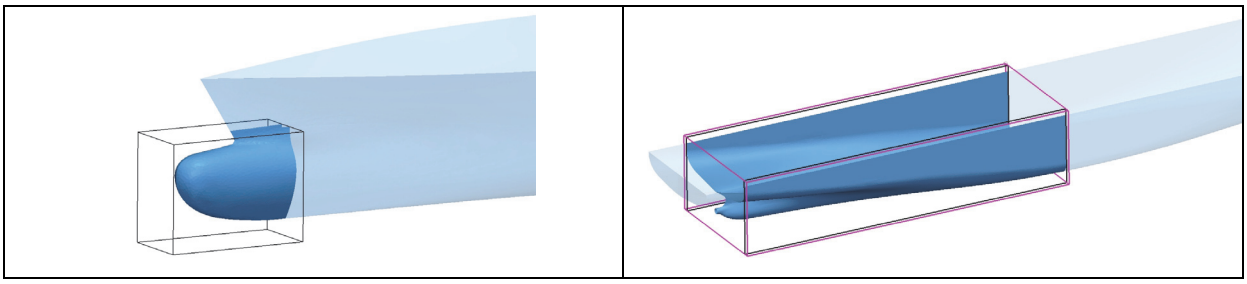
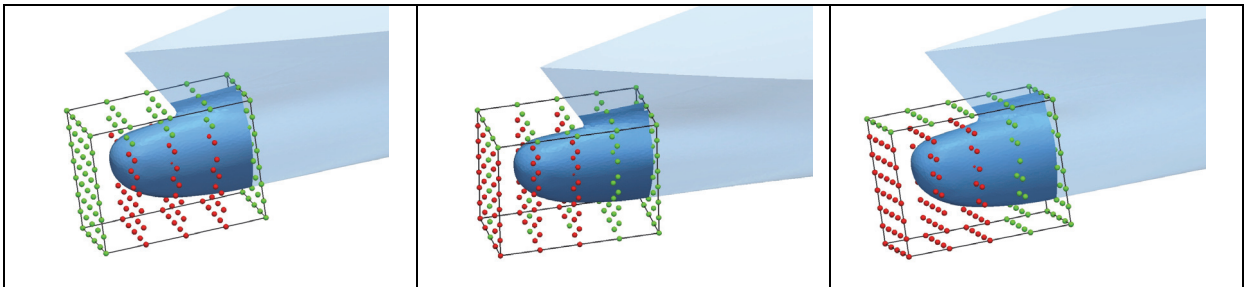
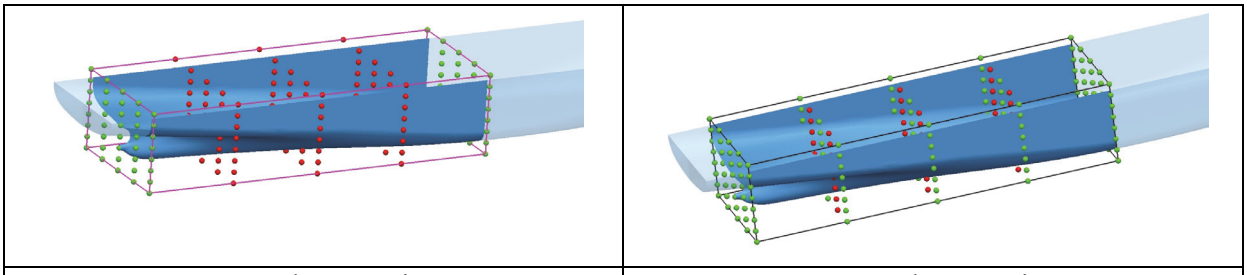


Figure 3. The modification regions by FFD method



Move along x axis	Move along y axis	Move along z axis



Move along x axis | Move along y axis

Figure 4. The modification regions by FFD method

Additionally, some geometric constraints are imposed on the design variables, the displacement (∇), the wetted surface area (S_{wet}) and the principal dimensions of the ship. Detail information regarding these constraints is reported in Tab. 4.

Table 3. Definition of the optimization problem

Type	Definition	Note
Initial hull	the KRISO 3600TEU container ship model (KCS)	
Objective functions	$f_{obj}^1 = C_t = C_w + C_v$, $at\ \ Fr = 0.2$ $f_{obj}^2 = C_t = C_w + C_v$, $at\ \ Fr = 0.26$	Bare hull Aim is to search for hull forms with potential drag reduction at given speeds
Design variables		
Δx_1 (Variable1)	[-0.0736, 0.0736]	Displacement in x direction in the fore-part region
Δy_1 (Variable2)	[-0.0368, 0.0368]	Displacement in y direction in the fore-part region
Δz_1 (Variable3)	[-0.04784, 0.04784]	Displacement in z direction in the fore-part region
Δx_2 (Variable4)	[-0.05152, 0.05152]	Displacement in x direction in the aft-part region
Δy_2 (Variable5)	[-0.0736, 0.08832]	Displacement in y direction in the aft-part region
Geometric constraints		
Main dimensions	L_{pp}, D, B are fixed	
Displacement (∇)	Maximum variation ±1%	
Wetted surface area (S_{wet})	Maximum variation ±1%	
Experimental design	OLHS method	Generate 100 sample points
Approximation model	Kriging model	
Optimizer	NSGA-II	
Size of population	200	
Number of generations	300	

Numerical results: the optimal design

Based on the optimal Latin hypercube design method (OLHS), 100 sample points about five design variables Δx_1 (displacement of control points in x direction in the fore part), Δy_1 (displacement of control points in y direction in the fore part), Δz_1 (displacement of control points in z direction in the fore part), Δx_2 (displacement of control points in x direction in the aft part), Δy_2 (displacement of control points in y direction in the aft part) are generated, then the corresponding values of multi-objective function, total resistance coefficients, are obtained using the NM+CFD integrated method.

Additionally, the ANOVA test is used to reflect the effects of each design variable on the objective functions. Denote by V1~V5 the five design variables (Δx_1, Δy_1, Δz_1, Δx_2, Δy_2) (see Fig. 6). The effects of design variables on different objective functions vary widely. V2(Δy_1) and V1(Δx_1) have main effects on f_{obj}^1, while V2(Δy_1), V5(Δy_2) and V1(Δx_1) on f_{obj}^2. But the total effect of the others is not neglected, and the computational cost considering all of five design variables is adequately affordable, thus, all of which are adopted in optimization.

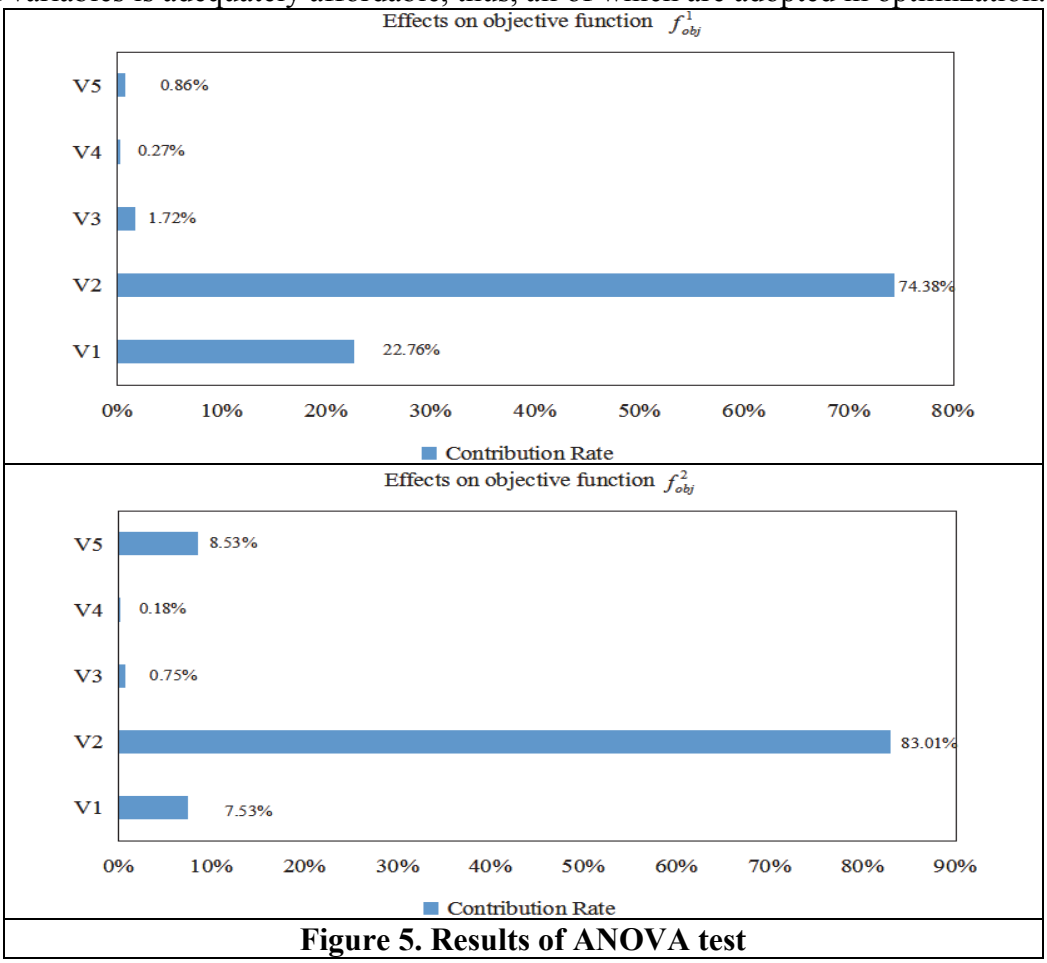


Figure 5. Results of ANOVA test

The Pareto front is reported in the function space in Fig. 7, where each red point represents an optimal solution while one typical example is marked in red to be analyzed further. Fig. 6 shows the Pareto optimal set from multi-objective optimization with NSGA-II algorithm. A reduction in resistance coefficients can be seen from Fig. 7. As an example of the Pareto optimal ships, one optimal configuration detected by the procedure is reported as the mark of the green point in Fig. 7.

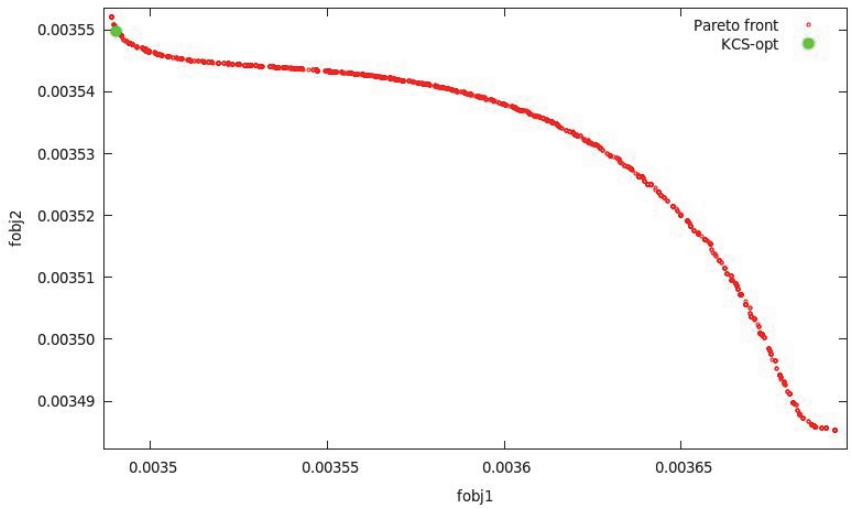


Figure 6. Pareto optimal points and optimal cases in objective functions space

Although the control modification regions are small, quite different configurations are readily yielded, all the Pareto optimal solutions, and different alternatives may be considered at this stage. KCS-opt represents the optimal hull form selected in this paper. As shown in Fig. 8 and Fig. 9, the bulb bow of the optimal hull form is evidently upturned than the initial one, and the stern lines are slightly changed, which corresponds directly with the ANOVA results presented before.

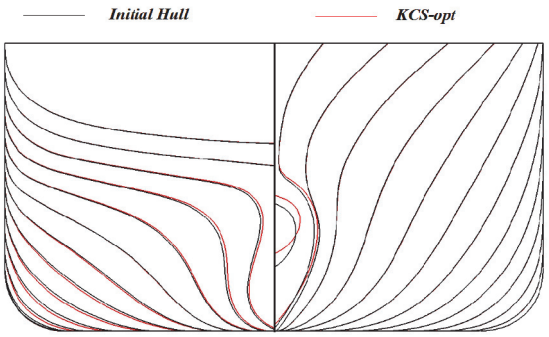


Figure 7. Body plans between the initial hull form and the optimal hull form

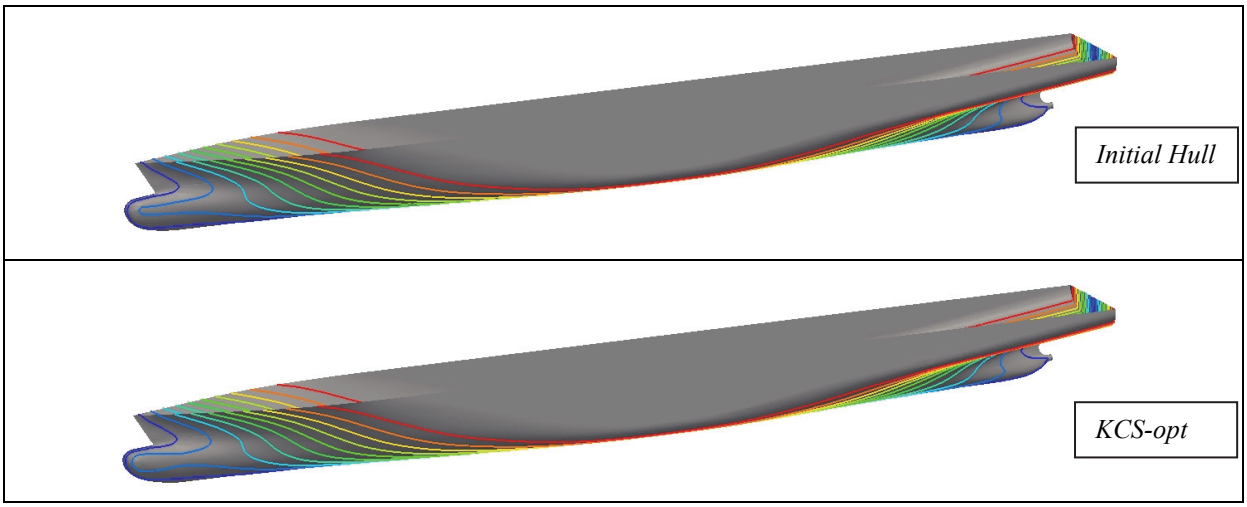


Figure 8. Buttock lines and 3D models between the initial hull form and the optimal hull form

The following table 4 shows the comparison of the results between the optimal hull form and the initial one. The respective reduction of the two total resistances is clearly seen, 4.73% reducing at Fr=0.2 and 8.32% reducing at Fr=0.26. However, there is a strange phenomenon that the total resistance of the KCS-opt at Fr=0.26 is even lower than at Fr=0.2. Further to understand, bulbous bow is first designed only to produce the positive effects on the resistance performance at the design speed. If so, it appears that the resistance is higher at other speeds.

Table 4. The prediction results for the initial and optimal hull forms based on NM+CFD integrated method

Comparison / Speed		Fr=0.2	Fr=0.26
C_t	Initial Hull	3.72E-03	3.82E-03
	KCS-opt	3.55E-03	3.50E-03
	Reduction	4.73%	8.32%

Validation with naoe-FOAM-SJTU solver

A high-fidelity numerical computation tool, naoe-FOAM-SJTU solver, is used to provide more accurate validation of the optimal hull form considering viscous effect, based on RANS method. Here, the total resistances between the initial hull form and KCS-opt mentioned above are only predicted at the design speed Fr=0.26. And the numerical results are presented in Tab. 5. KCS-opt displays a decrease of the total resistance coefficients of 3.39% at Fr=0.26.

Table 5. Numerical results for the initial and optimal hull forms by naoe-FOAM-SJTU solver

	Design Speed	Fr=0.26
C_t	Initial Hull	3.84E-03
	KCS-opt	3.71E-03
	Reduction	3.39%

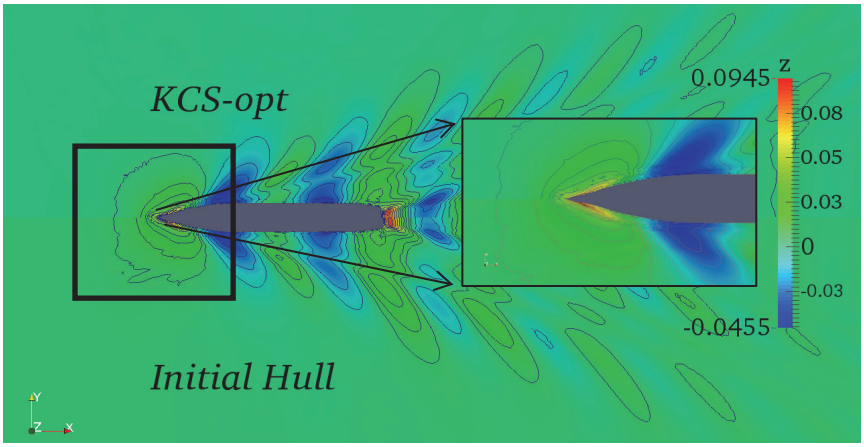


Figure 9. Wave patterns of free surfaces between the initial hull form and the optimal hull form

The computed wave patterns are reported in Fig. 8 The wave field caused by KCS-opt is with a smaller bow wave, a clear sign that the wave component of the ship's resistance has been reduced. A typical changes in the foremost wave pattern is enlarged, KCS-opt slightly reduces the amplitude of the bow wave. However, the wave pattern along the aft part of KCS-opt is a little bit changed, which is corresponding to the ANOVA test. It is partly illustrated that the multi-objective optimization, using OPTShip-SJTU solver, is reliable.

Conclusions

1. A numerical multi-objective optimization tool, OPTShip-SJTU, has been developed and tested in present work. the KRISO 3600TEU container ship model (KCS) is adopted as initial hull form, and the aim is to search for optimal hull forms with improved resistance performances at two given speeds (Fr = 0.20, 0.26).

2. During the procedure of optimization, the regions of bulb bow and stern are deformed with free-form deformation (FFD) method. FFD method is sufficiently flexible to generate a series of realistic alternative hull forms with a few number of design variables involved.

3. OPTShip-SJTU solver based on the integrated method of Neumann-Michell (NM) theory and Reynolds Average Navier Stokes (RANS) as the hydrodynamic performance evaluation module to predict the total resistance, turns out to be applicable for a real optimization problem.

4. The optimizer based on a multi-objective genetic algorithm, NSGA-II, and pareto-optimal front is obtained eventually.

5. The validation of the optimization problem is also carried out by naoe-FOAM-SJTU, a solver based on OpenFOAM source code. It shows the multi-objective optimization is acceptable and useful. and the results of OPTShip-SJTU solver should be further validated and verified by experimental data.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (51379125, 51490675, 11432009, 51579145, 11272120), Chang Jiang Scholars Program (T2014099), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

References

[1] Lee, S.-S., et al. (2014) A STUDY ON OPTIMIZATION OF SHIP HULL FORM BASED ON NEURO-RESPONSE SURFACE METHOD (NRSM). *Journal of Marine Science and Technology-Taiwan* **22**(6), 746-753.
[2] Kim, H. (2009) Multi-objective optimization for ship hull form design(Doctoral dissertation, George Mason University).
[3] Peri, D., & Campana, E. F. (2003)Multidisciplinary design optimization of a naval surface combatant. *Journal of Ship Research***47**(1), 1-12.
[4] Yang, C., et al. (2014) Hydrodynamic optimization of a triswach. *Journal of Hydrodynamics* **26**(6): 856-864.
[5] Tahara, Y., et al. (2011) Single- and multi-objective design optimization of a fast multihull ship: numerical and experimental results. *Journal of Marine Science and Technology* **16**(4): 412-433.
[6] Zhang, B.-J. and Z.-X. Zhang (2015) Research on theoretical optimization and experimental verification of minimum resistance hull form based on Rankine source method. *International Journal of Naval Architecture and Ocean Engineering* **7**(5): 785-794.
[7] Zhang, B.-j., et al. (2015) Research on design method of the full form ship with minimum thrust deduction factor. *China Ocean Engineering* **29**(2): 301-310.

[8] Zhang, B.-J. and A.-q. Miao (2015) THE DESIGN OF A HULL FORM WITH THE MINIMUM TOTAL RESISTANCE. *Journal of Marine Science and Technology-Taiwan* **23**(5): 591-597.
[9] Choi, H.-J., et al. (2015) STUDY ON OPTIMIZED HULL FORM OF BASIC SHIPS USING OPTIMIZATION ALGORITHM. *Journal of Marine Science and Technology-Taiwan* **23**(1): 60-68.
[10] Bagheri, H. and H. Ghassemi (2014) Optimization of Wigley hull form in order to ensure the objective functions of the seakeeping performance. *Journal of Marine Science and Application* **13**(4): 422-429.
[11] Deb, K., et al. (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6**(2): 182-197.
[12] Deb, K., et al. (2000) A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. Parallel Problem Solving from Nature PPSN VI: 6th International Conference Paris, France, September 18–20, 2000 Proceedings. Springer Berlin Heidelberg: 849-858.
[13] Eberhart, R. and J. Kennedy (1995) A new optimizer using particle swarm theory. *Micro Machine and Human Science* MHS '95.
[14] Shi, Y. and R. Eberhart (1998) A modified particle swarm optimizer. Evolutionary Computation Proceedings, *IEEE World Congress on Computational Intelligence*.
[15] Srinivasan, N., & Deb, K. (1994) Multi-objective function optimization using non-dominated sorting genetic algorithm. *Evolutionary Comp* **2**(3), 221-248.
[16] Campana, E. F., et al. (2006) Particle Swarm Optimization: efficient globally convergent modifications. *III European Conference on Computational Mechanics: Solids, Structures and Coupled Problems in Engineering*.
[17] Huang, F., et al. (2016) A new improved artificial bee colony algorithm for ship hull form optimization. *Engineering Optimization* **48**(4): 672-686.
[18] Kim, H.-J., et al. (2016) Hull-form optimization using parametric modification functions and particle swarm optimization. *Journal of Marine Science and Technology* **21**(1): 129-144.
[19] Barr, A. H. (1984) Global and local deformations of solid primitives. *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, ACM: 21-30.
[20] Sederberg, T. W. and S. R. Parry (1986) Free-form deformation of solid geometric models. *SIGGRAPH Comput. Graph.* **20**(4): 151-160.
[21] Coquillart, S. (1990) Extended free-form deformation: a sculpturing tool for 3D geometric modeling." *SIGGRAPH Comput. Graph.* **24**(4): 187-196.
[22] Noblesse, F., et al. (2012) The Neumann–Michell theory of ship waves. *Journal of Engineering Mathematics* **79**(1): 51-71.
[23] Noblesse F, Huang FX, Yang C. (2013) The Neumann-Michell Theory of Ship Waves. *Journal of Engineering Mathematics***79**(1): 51-71.
[24] Huang, F., Yang, C., & Noblesse, F. (2013) Numerical implementation and validation of the neumann–michell theory of ship waves. *European Journal of Mechanics - B/Fluids,* **42**(6), 47-68.
[25] Huang, Fuxin. (2013) A practical computational method for steady flow about a ship. *Dissertations & Theses - Gradworks*.
[26] Yang, C., Delhommeau, G., & Noblesse, F. (2007). The neumann-kelvin and neumann-michell linear models of steady flow about a ship.*International Congress of the International Maritime Association of the Mediterranean Imam*.
[27] Shen ZR, Jiang L, Miao S, Wan DC, Yang C. RANS simulations of benchmark ships based on open source code. In: *7th International Workshop on Ship Hydrodynamics (IWSH'2011)*, Shanghai, China, 2011.
[28] Lee, S.-J., et al. (2003) PIV velocity field measurements of flow around a KRISO 3600TEU container ship model. *Journal of Marine Science and Technology* **8**(2): 76-87.
[29] Choi, H.-J., et al. (2015) STUDY ON OPTIMIZED HULL FORM OF BASIC SHIPS USING OPTIMIZATION ALGORITHM. *Journal of Marine Science and Technology-Taiwan* **23**(1): 60-68.
[30] Bagheri, H. and H. Ghassemi. (2014) GENETIC ALGORITHM APPLIED TO OPTIMIZATION OF THE SHIP HULL FORM WITH RESPECT TO SEAKEEPING PERFORMANCE. *Transactions of Famena* **38**(3): 45-58.
[31] Dawson CW. (1997) A practical computer method for solving ship-wave problems. *2nd International Conference on Numerical Ship Hydrodynamics*, Berkeley, USA.
[32] Suzuki K, Kai H, Kashiwabara S. (2005) Studies on the optimization of stern hull form based on a potential flow solver. *Journal of Marine Science and Technology* **10**(2): 61-69.

Numerical Validation and Analysis of the Semi-submersible Platform of the DeepCwind Floating Wind Turbine based on CFD

Ke Xia, Decheng Wan*

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

*Corresponding author, E-mail: dcwan@sjtu.edu.cn.

Abstract

With the rapid development of the ocean engineering and the renewable energy, more and more attention are paid to the floating wind turbine. In recent years, researchers do much work on the floating wind turbine while most of the researchers investigate the problem by experiment or 3D potential theory but not the CFD, which has its own advantages in some aspects. A numerical simulation of motion performance of the DeepCwind floating wind turbine is investigated in the present study. In this paper, motion responses of the platform with mooring system under regular wave conditions are investigated numerically by a viscous flow solver naoe-FOAM-SJTU based on the open source toolbox OpenFOAM.
The motion performance of the platform under five different regular wave conditions are presented and compared with the data of the model test to validate the accuracy of the solver. The motion curves are presented in both the time domain and the frequency domain to research the response characteristics of the platform. Subsequently, the investigation about the parameter sensitive is conducted, and the results indicate that the motion performance would be better with the decrease of the height of COG and the increase of the draft within a reasonable range. And the broken mooring line has a huge impact on the platform to which should be pay more attention for the safety of the platform.

Keywords: semi-submersible platform, motion performance, parameter sensitive, naoe-FOAM-SJTU solver.

Introduction

As to the energy crisis and the environmental issues like pollution and global warming, the exploration for renewable and clean energies becomes crucial. Some potential resources become more and more significant, just like the wind, wave, solar and tidal, that numerous researchers are devoted to them (Ma and Hu, 2014) [1]. The wind energy is the fastest growing renewable energy resource which can never be exhausted, so it's attracting more and more attention worldwide (Tang and Song, 2015) [2].

As the main part of the floating wind turbine, the motion performance of the floating platform is significant for the wind turbine, and the motion performance of the platform has obvious effects on the aerodynamic performance of the wind turbine as well as the electricity generating capacity (Zhao and Yang, 2016) [3]. One challenge of the floating wind turbines is the wave induced platform tilt motion, which will heavily increase the displacements and load on turbine structure due to high inertial and gravitational forces and will bring severe fatigue and ultimate loads at tower bottom and blades roots (Hu and He, 2015) [4].

To research the motion performance and the wave loads of the floating platform, predecessors have done much work. A reasonable assumption is put forward by Hooft (2002) [5] that

Morison equation can be used to calculate the wave force of the platform, which is a semi empirical formula and wave force around the platform can be is divided into two parts, one is inertia force and the other one is drag force. This formula is widely used in the calculation of small scale component of the platform whose cross section is relatively simple (Lee and Incecik, 2005) [6]. Frank (2005) [7] find that the pulse source can be discretely distributed on the surface of the floating structure, so that people can calculate the wave force of the floating structure with arbitrary cross section shape, which is superior to the Morison equation, and this method is suitable for the compiling of the program. In the recent research about the hydrodynamic performance of the floating platform. Nowadays, most of the researchers investigate the motion response of the platform in the wave environment by the 3d potential theory. In this theory, platform is solved as a whole part not several sections, and the surface of the physical model of the platform below the waterline will be replaced by the mesh model so that the Green function can be got to calculate the velocity potential, and the distribution of the wave pressure can be calculated, as well as the motion response (Shi and Yang, 2010) [8]. At present, most popular hydrodynamic software such as AQWA, Seasam, Hydrostar and FAST are all based on the 3d potential theory to do the hydrodynamic calculation of the platform (Shi and Yang, 2011) [9]. 3D potential theory has several advantages that the calculated results are relatively accurate when compared with the results of the experiment, and it is very convenient which can get a satisfactory statistical results in a short period of time. However, the disadvantages of this method is extremely obvious. 3d potential theory is based on an assumption that the water is potential flow which is non-viscous, irrotational and incompressible. This is a simplification of the actual phenomenon which leads to obvious error from the results of the experiment. Actually, the viscidity of the water shouldn't be ignored in the motion of the platform, for it has significant effect on the motion of response, especially when the period of the coming wave is close to the natural period of the platform. Also the potential theory can't deal with a strongly nonlinear problem (Wang and Cao, 2015) [10]. On the contrary, Computational Fluid Dynamics (CFD) methods might be employed to obtain a better result via employing a more realistic model. The most prominent advantage of the CFD is that result of the simulation is more authentic and with the consideration of viscous flow, some more complex problems can be simulated such as green water, slamming, wave run up and other strongly nonlinear issues, which can't be solved by the potential flow method (Liu and Wan, 2015) [11]. In this paper, a viscous flow solver (naoe-FOAM-SJTU) (Shen and Wan, 2013; Zhou and Wan, 2013; Cao and Wan, 2014; Zha and Wan, 2014; Zhao and Wan, 2015) [12] which is developed and based on the popular open source toolbox OpenFOAM for predicting dynamics of floating structures with mooring systems is presented. The solver is adopted to study motion responses of a floating semi-submersible platform with mooring system under regular wave conditions.

The outline of this paper is as follows. Mathematical equations and numerical methods are first described concerning fluid flow, floating platform and mooring systems. Parameters of the platform and mooring system studied here together with computational domain are then presented. Then the validation work is done to compare the calculated results with the data of the model test which is conducted by the University of Maine DeepCwind program at Maritime Research Institute Netherlands' offshore wind/wave basin, located in the Netherlands (Robertson, 2012) [13]. Also, the calculation result of the same issue simulated by the FAST which is a software who is based on the 3d potential flow theory is added to the comparison (Alexander, 2013) [14]. The floating wind turbine used in the tests was a 1/50th-scale model of the NREL 5-MW horizontal-axis reference wind turbine with a 126 m rotor diameter. Subsequently, the research of parameter sensitive is done to investigate the effect of different height of the gravity center and draft on the motion performance of the platform. In

addition, the mooring line may be broken when the wave or wind is too large, so one of the mooring line is removed in this paper to study the motion response of the platform in a dangerous condition. Results and conclusions are made at the end.

Methods

The present solver naoe-FOAM-SJTU adopted for numerical simulation is based on a built-in solver in OpenFOAM named interDyFoam, which can be used to solve two-phase flow which is incompressible, isothermal and immiscible. To deal with common fluid-structure interaction problems in ship hydrodynamics and offshore engineering, several modules are further developed and integrated into the solver, such as a wave generation/damping module, a six-degrees-of-freedom (6 DOF) module and a mooring system module. Laminar Reynolds model are carried out in all the calculations. Mathematical formulae related to the solver are described as follows in detail.

1. Governing equations

For transient, incompressible and viscous fluid, flow problems are governed by Navier-Stokes equations:

$$\nabla \cdot U = 0 \tag{1}$$

$$\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla(\rho(\mathbf{U} - \mathbf{U}_g)\mathbf{U}) = -\nabla p_d - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \nabla(\mu \nabla \mathbf{U}) + \mathbf{f}_\sigma \tag{2}$$

Where $\mathbf{U}$ and $\mathbf{U}_g$ represent velocity of flow field and grid nodes separately; $p_d = p - \rho \mathbf{g} \cdot \mathbf{x}$ is dynamic pressure of flow field by subtracting the hydrostatic part from total pressure p; $\mathbf{g}, \rho$ and μ denote the gravity acceleration vector, density and dynamic viscosity of fluid respectively; $\mathbf{f}_\sigma$ is surface tension which only takes effect at the free surface and equals zero elsewhere. The Laminar model means that the Navier-Stokes equation will be solved directly and the turbulence model is not been considered in the calculation.

2. Wave generation/damping

For a floating platform, wave loading is a most important environment loads. So that, wave generation must be implemented numerically. The wave generation module of the naoe-FOAM-SJTU can make various types of waves such as linear wave and Stokes 2nd order waves which will be adopted in the following paper. The linear wave (3) and Stokes 2nd order wave theory (4) are adopted in this paper and the equation used to describe free surface is:

$$\eta = A\cos\theta \tag{3}$$

$$\eta = a_1 \cos(kx - \omega t) + a_2 \cos 2(kx - \omega t) \tag{4}$$

Where A and $H=2A$ denote wave amplitude and wave height; a_1 is the amplitude of the first order item and the a_2 is the amplitude of the 2nd order item.

Once the wave is generated, reflection has to be considered when wave propagates towards outlet boundary which will travels in an opposite direction that will interfere the incident wave. So that, the wave damping module is developed in this solver. Sponge layer takes effect by adding an additional artificial viscous term to the source term of the momentum equation. The new term is expressed as:

$$\mathbf{f}_s = -\rho \mu_s \mathbf{U} \tag{5}$$

Where μ_s is the artificial viscosity calculated by the following equation:

$$\mu_s(x) = \begin{cases} \alpha_s \left(\frac{x - x_0}{L_s}\right)^2, & x > x_0 \\ 0, & x \le x_0 \end{cases} \tag{6}$$

Where α_s is a dimensionless quantity defining damping strength for the sponge layer. Other parameter can be easily understood by reading the following figure.

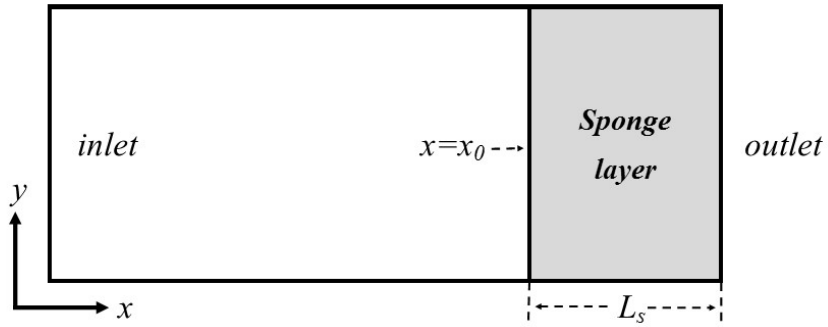


Figure 1. Overlooking of the calculation domain and sponge layer

3. Mooring system

To simulate the actual condition and the interaction problem of the mooring line and floating platform, the code of mooring line module is developed and added to the existing solver. The mooring line used in this paper is based on the PEM (piecewise extrapolating method) which is implemented to calculating the statics of mooring lines and it could take into account line elongation as well as the drag force induced by the fluid. With this method, mooring lines are divided into a number of segments, and a typical example of these is shown in Figure 2. Equations of static equilibrium are established in both horizontal and vertical directions:

$$\begin{cases} T_{xi+1} = T_{xi} + F_i ds \cos\varphi_{i+1} + D_i ds \sin\varphi_{i+1} \\ T_{zi+1} + D_i ds \cos\varphi_{i+1} = T_{zi} + F_i ds \sin\varphi_{i+1} + w_i dl \end{cases} \tag{7}$$

Where T_x, T_z *and* φ represent horizontal and vertical components of tension at a cross section of one segment and the angle between tension and T_x; *dl* and *ds* are length of the segment before and after elongation respectively; *w* is net submerged weight of lines per unit length; *D* and *F* denote normal and tangential components of drag force acting on the segment which are calculated by Morison's equation.

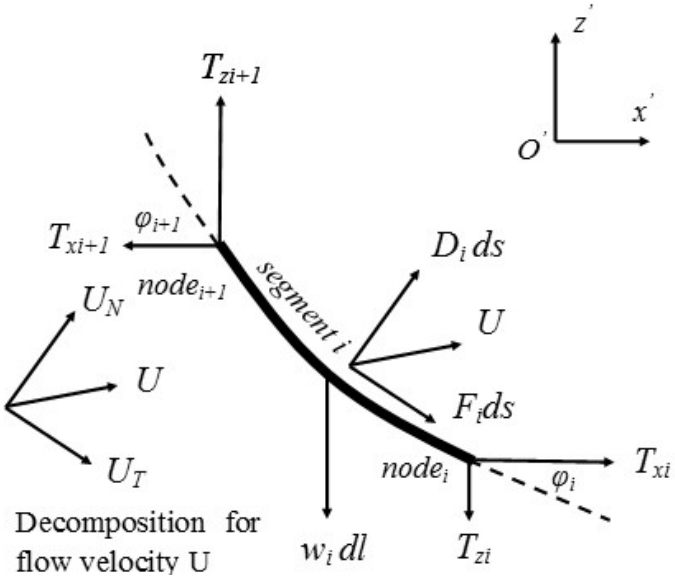


Figure 2. Force analysis of a mooring line segment for PEM

Computational model

A deep water semi-submersible platform of the DeepCwind with a catenary mooring system is presented in this paper, which is investigated both experimentally and numerically by Alexandwer (2012). Parameter of this platform and mooring system are respectively given in the section 1 and 2, as well as the computational domain.

1. Platform parameter

This floating platform for this model is semi-submersible which is downloaded from the website of National Renewable Energy Laboratory (NREL). It's standard mode named OC4 that researchers all over the world are investigating it. The platform is made up of three offset columns with larger diameter lower bases, one center support column for the turbine and a series of horizontal and diagonal cross bracing, and for the purpose of simplify the calculation, the diagonal cross bracing which is not very important in the seakeeping calculation is removed. The drawing of the DeepCwind semi-submersible platform are given in Figure 3 and the gross properties are presented in the Table 1. Furthermore, the Figure 4 give out the coordinate system of the platform in this study.

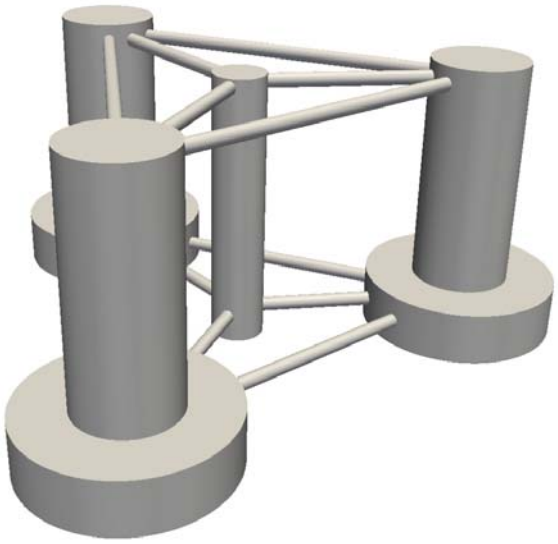


Figure 3. Overlooking of the 3d model platform

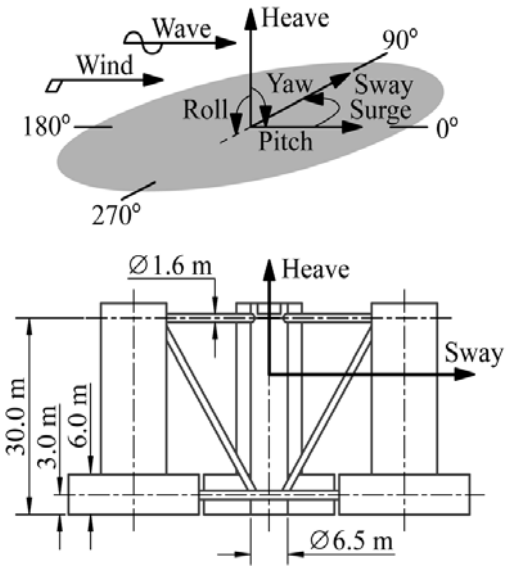


Figure 4. Coordinate system of the

2. Mooring system configuration

The mooring system of this semi-submersible platform is composed of 3 lines which interval between adjacent mooring lines is 120 degrees. And the fairleads of all lines are positioned at the surface of the base column. And the arrangement of the mooring system is shown in the Figure 5. And the parameter of the mooring system is presented in the Table 2.

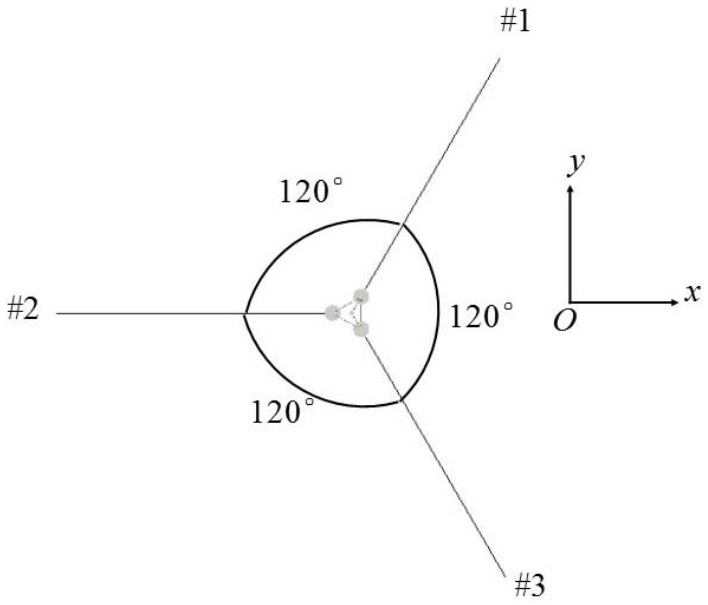


Figure 5. Configuration of the mooring system and platform

Table 1. Gross parameters of the semi-submersible platform of DeepCwind

Primary parameter	Unit	Value
Depth of platform base below SWL (total draft)	m	20
Elevation of main column (tower base) above SWL	m	10
Elevation of offset columns above SWL	m	12
Spacing between offset columns	m	50
Length of upper columns	m	16
Length of case columns	m	6
Depth to top of base columns below SWL	m	14
Diameter of main column	m	6.5
Diameter of offset (upper) columns	m	12
Diameter of base columns	m	24
Diameter of pontoons and cross braces	m	1.6
Displacement	m^3	13986.8
Center of mass location below SWL along platform center line	m	9.936

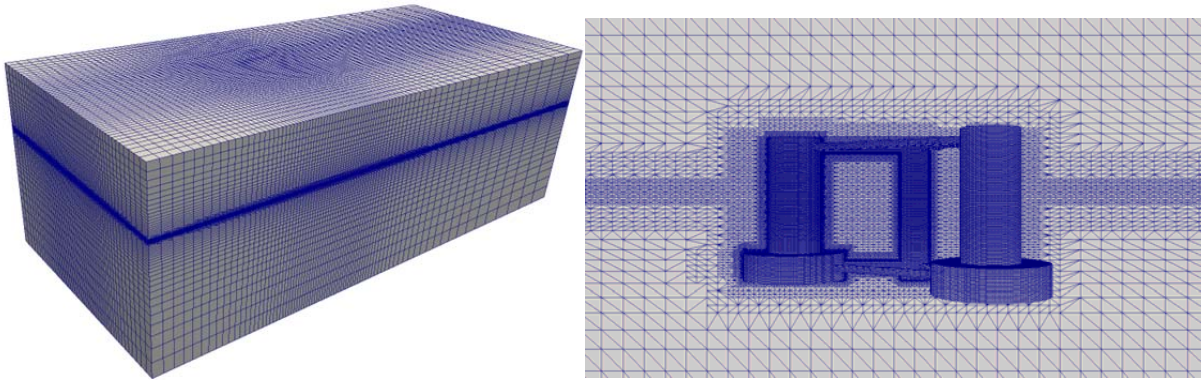
Table 2. Primary parameters of the mooring system

Primary parameter	Unit	Value
Number of mooring lines		3
Angle between adjacent lines	°	120
Depth to anchors below SWL (water depth)	m	200

Depth to fairleads below SWL	m	14
Radius to fairleads from platform centerline	m	4.0868
Radius to anchors from platform centerline	m	837.6
Equivalent mooring line mass in water	kg/m	108.63
Equivalent mooring line extensional stiffness	N	7.536E+8

3. Calculation domain

The solver used in this paper is based on the OpenFOAM who provides users a very powerful and convenient utility named snappyHexMesh (OpenFOAM, 2013) [15] to create the computational mesh with high quality in relatively short time. The overview of the computational mesh is shown in the Figure 6 (a), and the local refinement of the mesh near the platform is given in the Figure 6 (b). The model is located in the center of the computational domain. The totally cell number is about 1.3 million. And the principal dimension of the calculation domain is shown in the Figure 7.



(a) Overview of all the mesh (b) Local view near the platform

Figure 6. Computational mesh of the platform

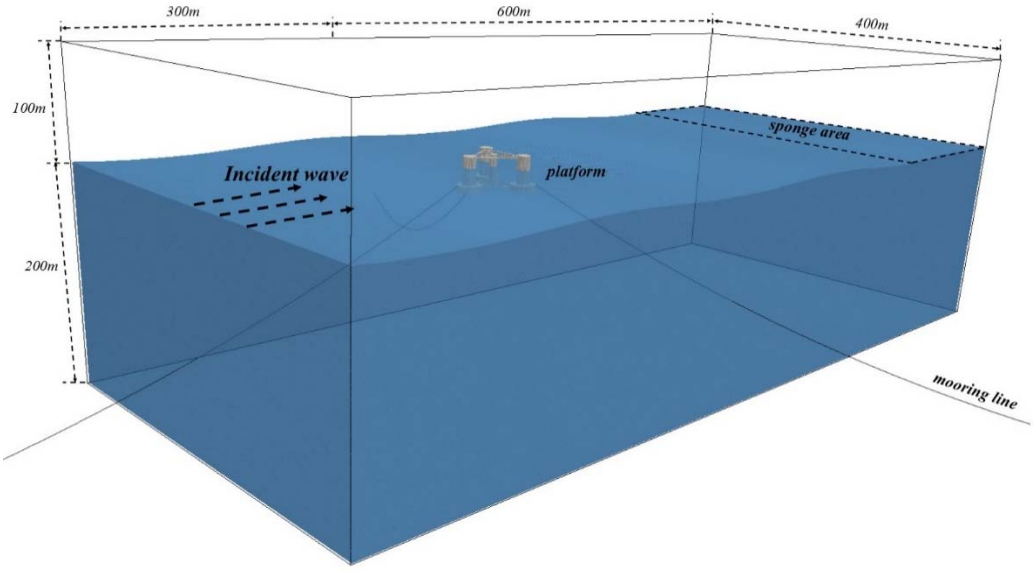


Figure 7. Overview of the calculation domain with principle dimension

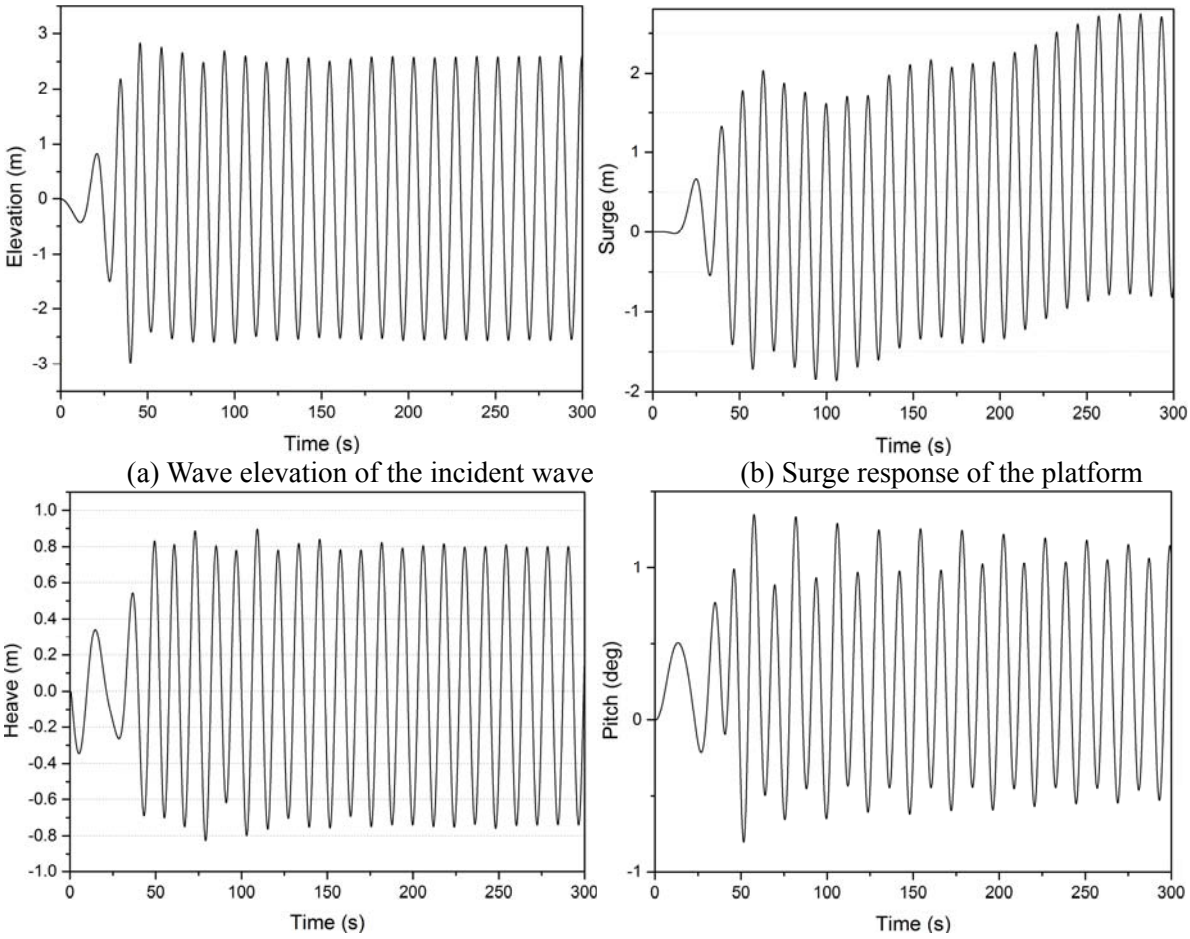
Validation

The follow research are all based on the solver naoe-FOAM-SJTU which is composed of a wave generation module, 6dof motion module, mooring system module and wave damping module. With these powerful module, several research can be done such as the problem of ship hydrodynamics and offshore engineering in various condition. But before doing the research about the platform of the DeepCwind, the validation work should be done to verify the correctness of the solver. So that, the validation work is done to compare the calculated results with the experimental data which is conducted by the Maritime Research Institute Netherlands' offshore wind/wave basin. The response of the DeepCwind semi-submersible platform to regular waves in the absence of wind is investigated in the validation of this paper. Different regular waves are considered, the amplitudes and periods of which are given in the Table 3. All waves propagated in the positive surge direction. It should be noted that two distinct amplitudes were investigated for periods of 14.3 and 20.0 s for the purpose of assessing any nonlinearity in system response. The motion performance of the DeepCwind platform is characterized by response amplitude operators (RAOs) magnitudes, which normalize the amplitude of a periodic response of a field variable by the amplitude of the regular waves.

Since the wind is not considered in this study, the weight, height of gravity and the moment of inertia of the whole wind turbine are converted and added to the parameters of the platform. Before the calculation of the motion, the work of wave generation should be done in the empty computational domain without the platform. The wave probe is set at the longitudinal min-section of the domain near the inlet. And the elevation of the wave whose wave height is 5.15 m and the period is 12.1 s is shown in the Figure 8 (a). The time step used is fixed at 0.05s and the overall time simulated is set as 300s. Figure 8 (b), (c) and (d) shows the surge, heave and pitch response of the platform within the wave whose height is 5.15 m and the period is 12.1 s. For the limitation of the length of the paper, the waveform figure of other waves and the motion response of the platform under other wave conditions are not given in the paper.

Table 3. Calculated regular wave amplitudes and natural periods

Amplitude (m)	Period (s)
3.57	14.3
3.79	20
5.15	12.1
5.37	14.3
5.56	20



(a) Wave elevation of the incident wave (b) Surge response of the platform

(c) Heave response of the platform (d) Pitch response of the platform

Figure 8. Elevation of the incident wave and motion response of the platform

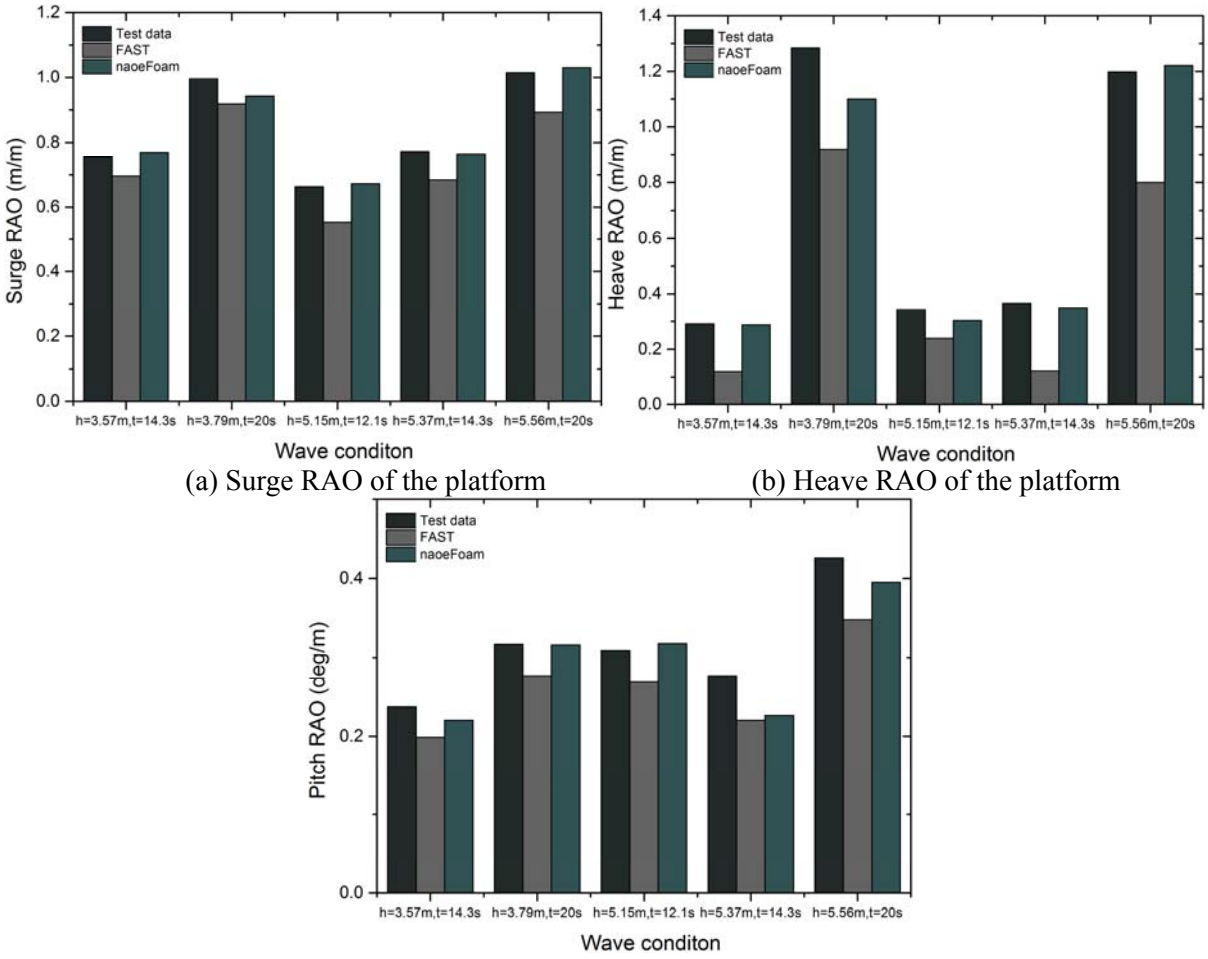
Since the platform is calculated in the regular waves and according to the International Towing Tank Conference (ITTC) that motion data should be collected at least for 10 quasi-steady cycles under regular wave conditions to ensure accuracy of results (ITTC, 2002) [16]. So that in this paper, the last ten period of the motion response are considered and calculated the average value. And then the RAO, the average amplitude of the motion response divided by the wave amplitude, can be counted.

Before showing the validation result, it is a wonderful job to analyze the motion response results given in the Figure 8, whose incident wave height is 5.15 m. It's evident that the height of the made wave highly agree with the requested wave and it's very steady after 100s. Therefor the following analysis can be based on the 100-300 s of the calculation. The average surge amplitude is 1.73 m, and the average heave amplitude is 0.78 m. In addition, the average pitch amplitude is 0.83°. From the figure above, several conclusion can be drawn.

Firstly, it's evident that the surge motion of the platform is nonlinear, and under the action of the wave force, the platform drifts about 1 m along the direction of the wave during 300 s. Subsequently, the heave response is relatively steady that it's almost a linear motion. Finally, a conclusion can be drawn that the pitch motion natural period of the platform is definitely larger than the period of incident wave which is 12.1 s, because it's obvious that the amplitude of the pitch motion of the platform presents a periodic variation, one large and one small, which means that in the motion of the platform, the second wave acts on the platform before the first natural period of the pitch motion over, and then causes the phenomenon of nonlinear pitch motion, as well as the increased pitch motion center. After the following study, it is

found that the nonlinear phenomenon is gradually stabilized after 350 s, and the average pitch displacement is about 0.25° instead of 0°.

The RAO magnitude for surge, heave and pitch are given in the Figure 9 for the five regular waves investigated, and the comparison is conducted between the experiment, naoe-FOAM-SJTU and FAST. FAST is a professional software for calculation of performance of wind turbine, who is based on the 3d potential theory, and the calculation with FAST is conducted by Alexander (2013).



(a) Surge RAO of the platform (b) Heave RAO of the platform

(c) Pitch RAO of the platform

Figure 9. Comparisons of RAOs from experimental test, FAST and naoe-FOAM-SJTU

Almost all the comparisons in the Figure 9 between the test data and the solver that used in this paper are extraordinarily good, which is much better than the FAST in these comparisons. The large discrepancy is likely a result of the damping system that the quadratic damping model employed in the FAST, which over-predicts the damping in large amplitude heave scenarios at the expense of properly modeling the damping for small to moderate motions. The first conclusion can be drawn that the response of the platform to the low frequency wave is more intense that the RAOs of the wave condition whose period is 20 s are obviously larger than that of 12.1s and 14.3 s. At the same time, another conclusion can be concluded from the Figure 9 that there exists nonlinear phenomenon in the test and calculation that the RAOs of the platform are exactly different from each other with the dame wave period but distinct wave height which should be the same regardless of the viscosity.

As everybody knows, there is no viscosity and vortex in the 3d potential theory, so in the calculation with FAST, it is difficult to confirm a right damping coefficient. However, it is precisely the advantage of CFD, which is based on the Navier-Stokes equations so that the viscosity and vortex are take into consideration. As shown in the Figure 10, it present the vortex and the radiation or reflected wave generated in the procedure of motion of the platform under the 5.37 m wave height condition. Moreover, CFD can do some nonlinear problems, such as the wave run up and fracture in the process of calculation, which is more close to the actual situation. Therefore, the solver used in this paper is validated.

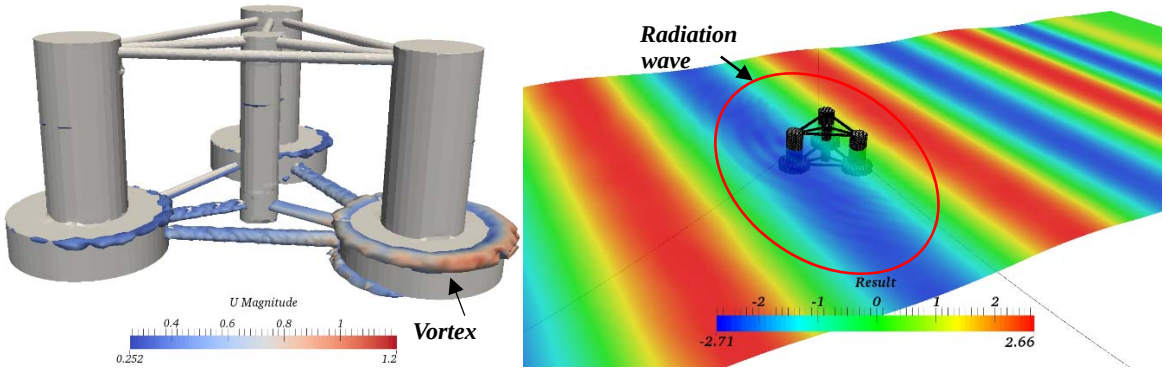


Figure 10. Vortex and radiation wave generated by the motion response of the platform

Results

In order to evaluate the influence of the parameter sensitive on the motion response and other hydrodynamic parameters. And through the investigation, several laws can be drawn to get the best motion performance of the platform in the design procedure to ensure the stability of electricity generating of the wind turbine. Following three aspects are investigated: 1. The research about the effect of different height of gravity center of the platform on the motion performance of the platform. 2. The influence of different draft of the platform on the performance of the platform. 3. A dangerous condition of the platform is considered that one mooring line is removed to investigate the influence on the motion performance of the platform.

1. Effect of height of gravity center

The height of gravity center is a vital parameter for an offshore platform, and it directly affects the stability of the platform. At the same time, the gravity height will exactly affect the sea-keeping performance of the platform by affecting both the motion period and the motion amplitude. Therefore, in this section, three distinct height of COG (center of gravity) is considered, which includes the original height -9.9 m, and the others. The sketch of the distribution of the different COG calculated in this study is shown in the Figure 11.

The 5.15 m wave height and 12.1s wave period wave condition is selected in the investigation of effect of the COG. And in this section, the overtime of calculated condition are set as 200s, which is due to the large amount of computation and aimed to save time. And the comparison of -7m, -13m and the original height -9.9m are given in the Figure 12.

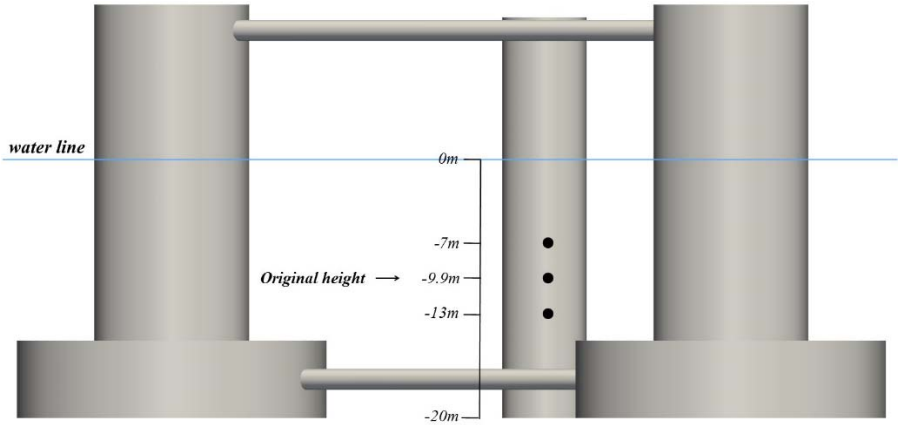
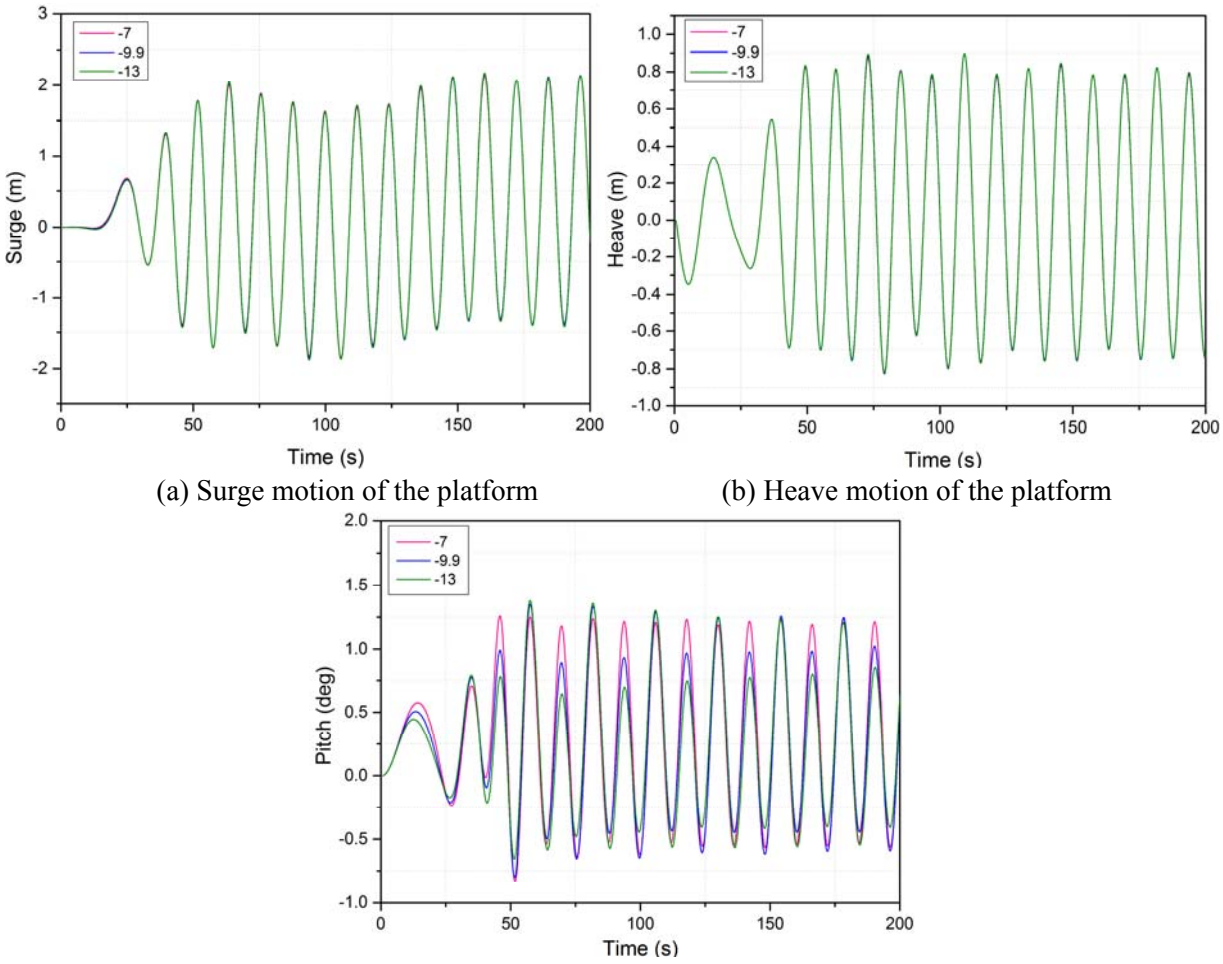


Figure 11. Distribution of the calculated height of gravity center



(a) Surge motion of the platform (b) Heave motion of the platform

(c) Pitch motion of the platform

Figure 12. Comparison of the effect of different height of COG on the platform

A conclusion can be drawn that the motion performance of the platform is better with the decrease of the height of the COG within a reasonable range. The surge and pitch motions of different height of COG didn't show a significant difference that the surge and heave amplitude of these three height of COG is almost the same. The difference of the pitch motion between the calculated conditions is quite obvious. The first conclusion can be gotten that the pitch performance becomes better with the decrease of the height of the COG. Subsequently, the pitch motion trends are consistent, which follows a similar increase or decrease law and

also the strong nonlinearity is very evident in this process. Moreover, the pitch motion performance is relatively steady than other conditions when the height is -7 m which is almost a linear motion.

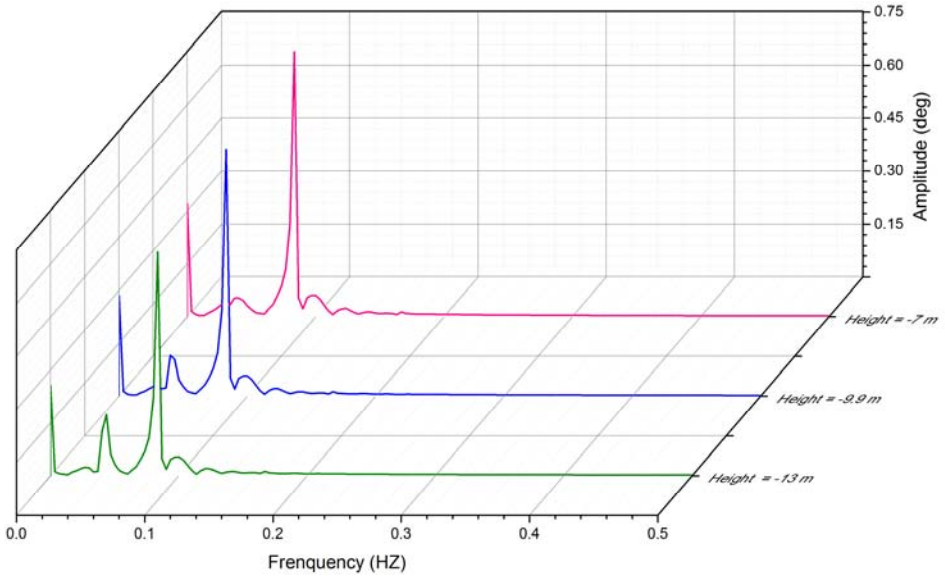


Figure 13. FFT of the pitch motion with different height of COG

For the strong nonlinear pitch motion, the FFT (Fast Fourier Transform) about the motion amplitude is conducted, for the strong nonlinear phenomenon, to investigate the characteristics of the pitch motion performance of the platform on the frequency domain. The results are given in the Figure 13, and to do a further analysis, the specific value about the first order term and the second order term of the amplitude of the pitch motion is given in the Table 4.

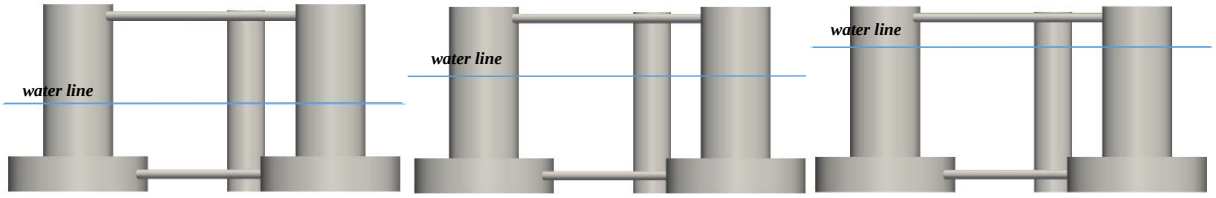
Table 4. Specific numerical analysis in frequency domain

Height of COG (m)	Orders	Frequency (HZ)	Amplitude value (deg)
-7	First order	0.083	0.782
	Second order	0.039	0.053
-9.9	First order	0.083	0.710
	Second order	0.040	0.116
-13	First order	0.083	0.654
	Second order	0.043	0.170

By the Table 4, some obvious rules can be summed up that besides the first order motion, there also exists second order pitch motion which can be easily found in the Figure 13. The frequency of the first order motion is about 0.083 HZ which is the frequency of the incident wave. And the frequency of second order of different COG are distinct which is related to the natural period of the platform and the mooring system, for the period of the platform come up from theory say that it is affected by the height of the COG. The frequency of the second order

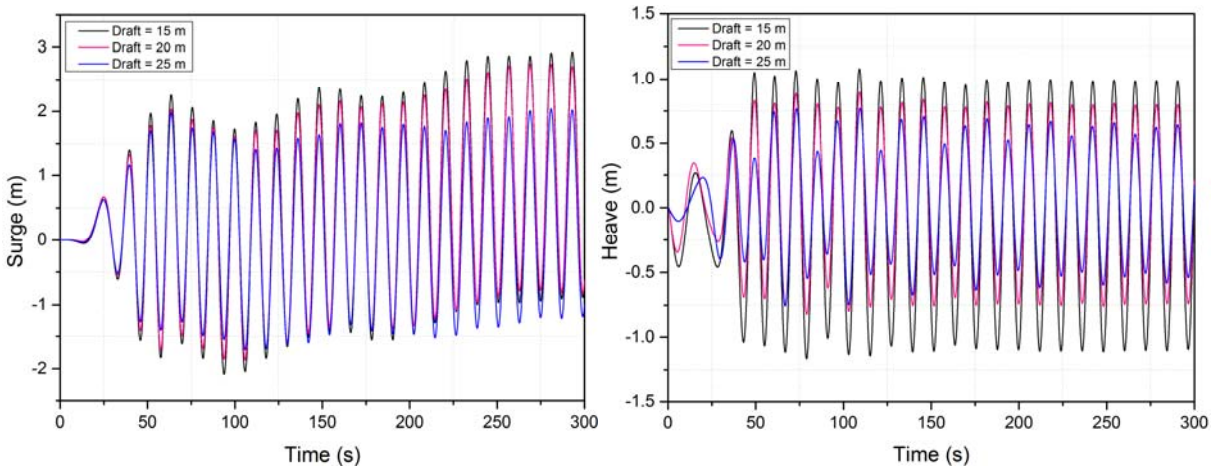
pitch motion decrease with the COG rise up. And it is also evident that the second order of the pitch motion of the height -7 m is quite small which means that it is nearly a linear motion in this condition. The amplitude of the first order plus that of the second order is almost the amplitude of the pitch motion amplitude of the platform, so that it is easy to find the law of the effect of the COG on the pitch motion of the platform by add the first order's amplitude and the second order one. Then the law mentioned above can be confirmed that the motion performance will be better with the decrease of the height of the COG within a certain range.

2. Effect of draft

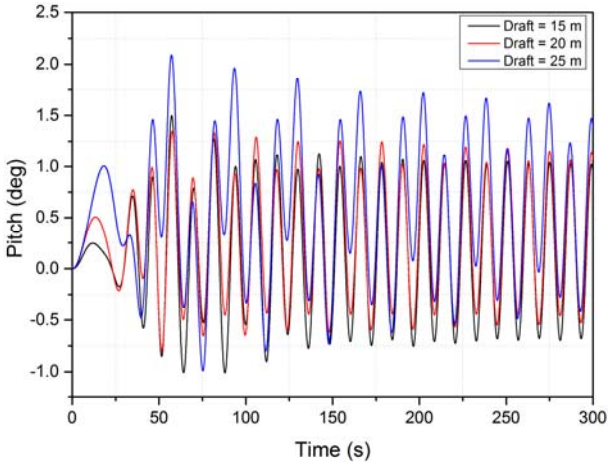


(a) Draft = 15 m (b) Draft = 20 m (original draft) (c) Draft = 25 m

Figure 14. Different draft condition of the platform



(a) Surge motion of the platform (b) Heave motion of the platform



(c) Pitch motion of the platform

Figure 15. Effect of draft on the motion response of the platform

Draft is also a significant parameter for a floating platform who not only affects the displacement of the hull but also affect the motion performance. And for a platform for the floating wind turbine, the draft would change if the mass or the force that act on the blade

changed, and the draft would be different in different condition. So it is meaningful to investigate the effect of draft on the motion performance of platform. In this section, three different drafts are conducted including 15 m, 25 m and the original one 20 m. The Sketch is given in the Figure 14 and the curves of response results are shown in the Figure 15.

The average motion amplitude of each condition can be calculated easily when the motion are quasi-steady. The three kinds of average motion amplitude of the platform with 25 m draft are that surge amplitude 1.629 m, heave amplitude 0.558 m and the pitch amplitude 0.85°. The average motions amplitude of the platform with 20 m draft are given in the validation that surge 1.73 m, heave 0.78 m and pitch 0.83°. At last, the average motions amplitude of the platform with 15 m draft are that surge 1.925 m, heave 1.043 m and pitch 0.862°. So an evident law can be found from the comparison of the value that the motion performance of the platform is better with the increase of the draft within a reasonable range, which can be easily observed from the Figure 15. Also a strange point can be found after analysis that the pitch motion of the 25 m draft is larger than that of 20 m, which is inconsistent with the rules summed up before and a likely reason for this problem is that it is obvious that the pitch motion of the platform with 25 m draft has obvious oscillation which is not steady for the short calculation time, thus the average value is probably larger than the pitch motion amplitude in the steady condition. The Figure 16 show the motion of the platform with different draft in the same time 295s that platforms reach the motion amplitude. It is easily to find that the motion degree is larger with the draft decrease and it can be also found that the free surface has an obviously nonlinear up and down near the platform.

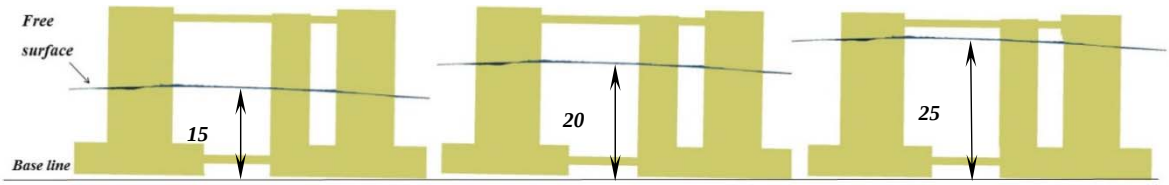


Figure 16. Posture of the platforms with different draft in the 295 s

3. One mooring line broken condition

The mooring system play a vital role in both working and survive condition of the platform, which not only maintains the position of the platform in the horizontal direction but also provides restoring force and conductive to the performance of the platform. In this section, a dangerous condition is investigated that one mooring line, the #3 line, of the mooring system is removed to represent the broken one. The model diagram of the configuration of the mooring system is shown in the Figure 17, and the comparison between this condition and the normal condition is shown in the Figure 18.

As can be seen in the Figure 18, the x-axis direction motion appeared a huge mutation when one mooring line is broken down and the platform moves in negative direction of the x-axis in which the most distance is about 15 m at 50 s and pulling back by the rest lines after that and finally steady at about 6 m against the x-axis. As shown in the Figure 19, which is comparison between the position of the platform at 50 s and the original position of the platform. Subsequently, in the z-axis direction, the platform float up 0.21 m after one mooring line is broken for the lack of pretension force. Finally, in the rotation direction, the platform also has an obvious change on the floating condition which skews about 0.39° overall.

Besides the movement mentioned above, it is also necessary to analyze the surge, heave and pitch motion response of the platform. As shown in the Figure, the surge amplitude of the normal condition is 1.73 m and 1.78 m at the one line broken condition. And the platform in

these two condition presents the same heave and pitch amplitude that heave 0.78 m and pitch 0.83° although the large difference between the displacements of these two kinds of situation.

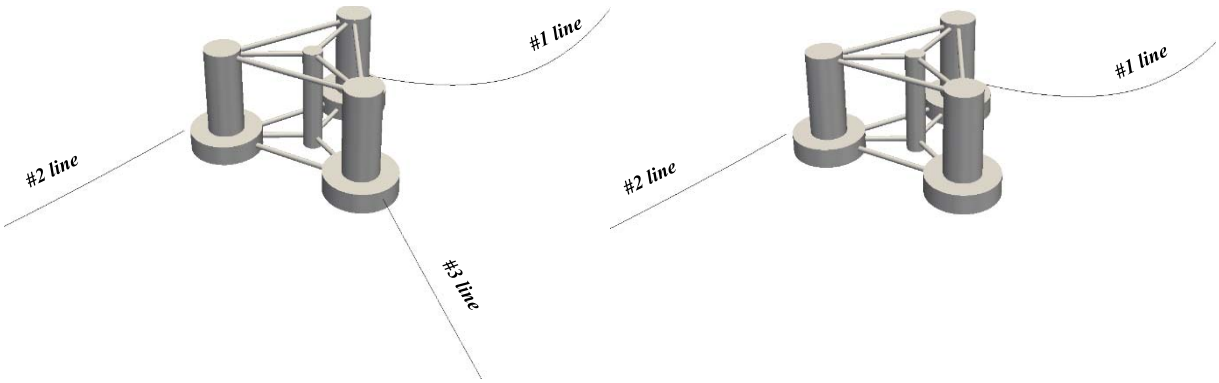
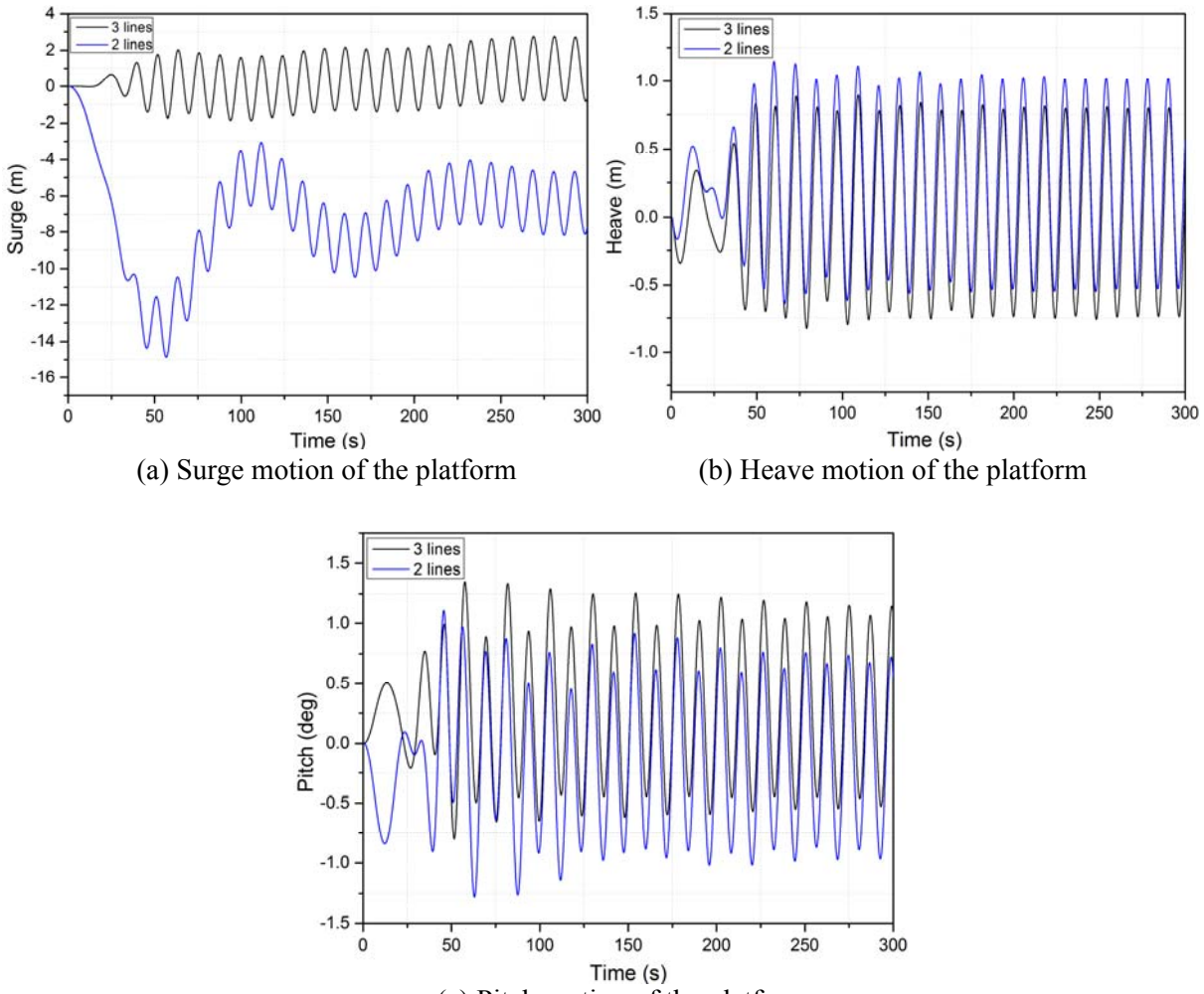


Figure 17. The diagram of complete mooring system and broken mooring system



(a) Surge motion of the platform

(b) Heave motion of the platform

(c) Pitch motion of the platform

Figure 18. Response of working condition and one mooring line broken condition

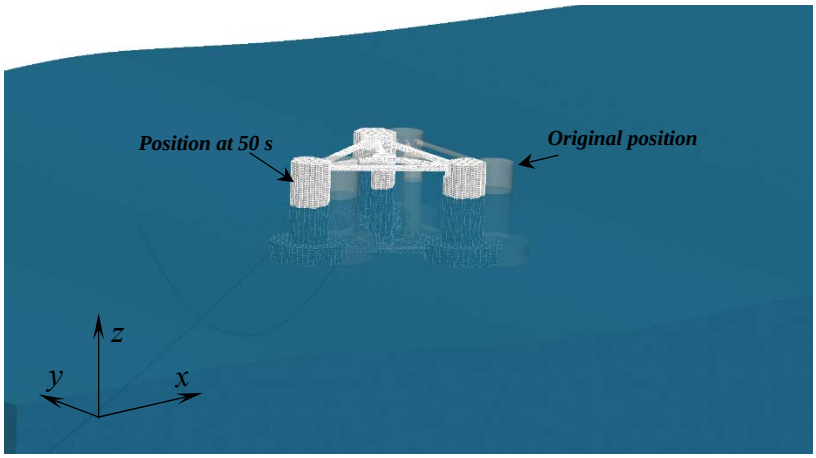
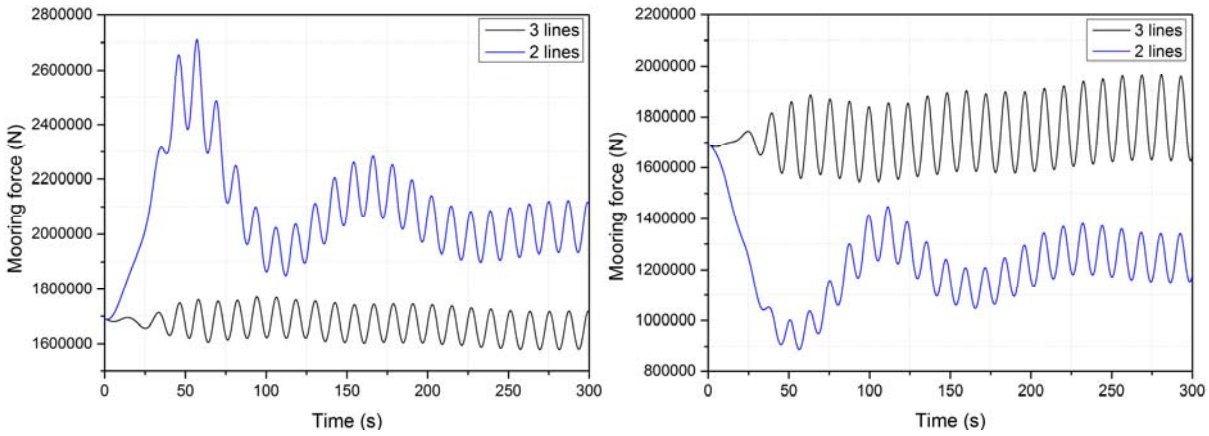


Figure 19. Position of the platform at 50 s and 0 s

In addition to the content above, the mooring forces of #1 and #2 are also be investigated. As shown in the Figure 20, the blue line represents the one line broken condition and the other one represent the normal condition. It is easy to find that the mooring force of the #1 becomes larger after the #3 broken and the mooring force of #2 smaller. At the same time, the amplitude of the oscillation of the mooring force #1 becomes larger and the amplitude of #2 smaller in this procedure. All above analyses indicate that the load acting on mooring line #1 become larger and it plays a more significant role in the motion of the platform and the mooring system, and the #2 quite the contrary. So that it is the more dangerous one and more attention should be paid to the #1 when the #3 is broken.



(a) Mooring force of #1 line in two conditions (b) Mooring force of #2 line in two conditions

Figure 20. Mooring force of #1 and #2 mooring line in these two conditions

Conclusion

In this paper, a viscous flow solver naoe-FOAM-SJTU based on the open source toolbox OpenFOAM is developed and presented. By comparing numerically calculated results with the experimental test data and the results of the FAST, the ability of present solver to handle hydrodynamic problems of floating structures with mooring system with various wave condition is validated. The solver is then adopted to investigate the parameter sensitive of the platform including the height of gravity center and the draft. In this section, several conclusion are drawn that the motion performance would be better with the decrease of the height of COG within a suitable range. Subsequently, it is found that the performance is better with the increase of the draft within a reasonable range. Moreover, to investigate the motion performance of the platform in a dangerous condition, one of the mooring line is removed to

simulate the condition that on mooring line is broken by the wave or flow force. Results indicates that the platform would move a certain distance along the direction against the x-axis and float up because of the lack of pretention force of the mooring line and cause the unbalanced force, as well the rotational movement. The mooring force of the rest two lines are investigated as well, which indicate that one of the two lines would be very generous whose mooring force increase immediately after the third line is broken. So that, people should pay more attention to this dangerous one in this condition. Although the present work are all based on the regular wave, the regular one can analyze the characteristic of the platform better, and the irregular wave would be carried out in the future work. The work done in this paper can serve as foundation for the design and working of the DeepCwind wind turbine platform which can ensure a more steady motion performance as well as the stability of the electricity generation. And the solver used in this paper can do more complex issues like VIV and wind-wave-current coupling issues in the future.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (51379125, 51490675, 11432009, 51579145, 11272120), Chang Jiang Scholars Program (T2014099), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23) and Lloyd's Register Foundation for doctoral student, to which the authors are most grateful.

References

[1] MA, Y., Z. HU and L. XIAO, Wind-wave induced dynamic response analysis for motions and mooring loads of a spar-type offshore floating wind turbine. Journal of Hydrodynamics, Ser. B, 2015. 26(6): 865-874.
[2] Tang, Y., K. Song and B. Wang, Experiment study of dynamics response for wind turbine system of floating foundation. China Ocean Engineering, 2015. 29(6): 835-846.
[3] Zhao, Y., et al., Dynamic response analysis of a multi-column tension-leg-type floating wind turbine under combined wind and wave loading. Journal of Shanghai Jiaotong University (Science), 2016. 21(1): 103-111.
[4] Yang, H.Y.H.E., Optimization Design of TMD for Vibration Suppression of Offshore Floating Wind Turbine. International Journal of Plant Engineering and Management, 2015. 1(20): 13-27.
[5] Hooft. Coupled Effects of Risers/Supporting Guide Frames on Spar Responses [A]. Proc. 12th International Offshore and Polar Engineering Conf. [C], Kitakyushu, Japan, 2002: 231-236.
[6] Lee Y W, Incecik A, Chan H S and Kim Z k. Design Evaluation in the Aspects of Hydrodynamics on a Prototype Semi-Submersible with Rectangular Cross-Section Members[A].Proceedings of the Fifteenth (2005) International Offshore and Polar Engineering Conference[C]. Seoul, Korea, ISOPE2005: 320-327.
[7] Frank, Lee D.Y., Choi Y.H., etc. An Experimental Study on the Extreme Motion Responses of a SPAR Platform in the Heave Resonant Waves [A]. Proc. International Off-shore and Polar Engineering Conf. [C], Seoul, Korea, 2005: 225-232.
[8] SHI Qi-qi, YANG Jian-min. Research on hydrodynamic characteristics of a semi-submersible platform and its mooring system. The Ocean Engineering, 2010. 28(4):1-8.
[9] SHI Qi-qi, YANG Jian-min , XIAO Long-fei. Research on motion and hydrodynamic characteristics of a deepwater semi-submersible by numerical simulation and model test. The Ocean Engineering, 2011. 29(4):29-42.
[10] Wang, S., et al., Hydrodynamic performance of a novel semi-submersible platform with nonsymmetrical pontoons. Ocean Engineering, 2015. 110: 106-115.
[11] Yuanchuan Liu, Y.P.D.W., Numerical Investigation on Interaction between a Semi-submersible Platform and Its Mooring System. Proceedings of the ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering OMAE2015 May 31-June 5, 2015, St. John's, Newfoundland, Canada.
[12] Shen, Z. R., Zhao, W. W., Wang, J. H. and Wan, D. C. 2014. "Manual of CFD solver for ship and ocean engineering flows: naoe-FOAM-SJTU." Technical Report for Solver Manual, Shanghai Jiao Tong University.
[13] A. Robertson, J. Jonkman, M. Masciola, H. Song, A. Goupee, A. Coulling, and C. Luan. Definition of the Semisubmersible Floating System for Phase II of OC4. 2012. Available from: http://www.nrel.gov/
[14] Coulling, A.J., et al., Validation of a FAST semi-submersible floating wind turbine numerical model with DeepCwind test data. Journal of Renewable and Sustainable Energy, 2013. 5(2): 023116.
[15] OpenFOAM. Mesh generation with the snappyHexMesh utility. 2013. Available from: http://www.openfoam.org/ docs/user/snappyHexMesh.php#x26-1510005.4.

Numerical Study on Ship Motion Coupled with LNG tank Sloshing Using Dynamic Overset Grid Approach

***Y. Zhuang, C.H. Yin and †D.C. Wan**

State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China

***** Presenting author: nana2_0@sjtu.edu.cn.

†Corresponding author: dcwan@sjtu.edu.cn.

Abstract

In this paper, numerical simulations of ship motion coupled with LNG tank sloshing in waves are considered. The fully coupled problems are performed by our in-house RANS/DES solver, naoe-FOAM-SJTU, which is developed based on the open source tool libraries of OpenFOAM. The internal tank sloshing and external wave flow are solved simultaneously. The considered models are LNG FPSO and a modified KVLCC2 coupled with two LNG tanks respectively. Three degrees of freedom is released in the regular waves. The ship motion responses of LNG FPSO are carried out both in head and beam waves to compare with existing experimental data to validate this solver. Next, the modified KVLCC2 coupled with two LNG tanks and a propeller is simulated with a forward-speed in the head wave using dynamic overset method. Two filling ratios of tanks: 30% and 60% are considered, and results are compared with that without sloshing.

Keywords: LNG sloshing, OpenFOAM, nonlinear coupled motion, dynamic overset grids, naoe-FOAM-SJTU solver

Introduction

The sailing performance of the ship equipped with liquid tanks is different from that without tanks. The sloshing flow in tanks which is excited by ship motion would affect ship performance in return. This coupling effect not only causes impact pressure which may damage the cargo, but also changes ship motion in waves. It is especially essential for FLNG or FPSO, for these kinds of vessels suffer from both external wave force and internal force when they transport liquid cargoes on the sea. Therefore, the maneuvering of ship equipped with liquid cargoes in waves is still a researchable issue. Since the coupling effect is nonlinear and viscosity in sloshing flow is ignorable, the numerical simulation has its advantages to treat this problem. Computational Fluid Dynamics (CFD) is an effective method to simulate the ship motion in waves coupled with LNG tank sloshing, and with the assistant of overset grid technology, the coupling effect of large-amplitude motion such as self-propulsion in waves with partially filled tanks can be solved effectively.

Several researches about ship motion coupled with tank sloshing have been done. Nam, B.W. et al[1] carried out both numerical and experimental studies of LNG FPSO model. The

impulse-response-function (IRF) was used to simulate ship motion and finite-difference method was adopted to solve nonlinear tank sloshing. In recent decades, many studies used viscous flow theory in order to solve the nonlinearity of the tank sloshing. Li, Y. L. et al[2] applied both potential flow theory and viscous flow theory under OpenFOAM. Jiang, S. C. et al[3] also used OpenFOAM to simulate the coupling effect, and applied VOF to capture the interface, and the paper still considered ship response in IRF method. Shen, Z. R. et al[4] achieved fully coupled of ship motion and tank sloshing by the unsteady RANS solver, naoe-FOAM-SJTU. Considering the ship performance with sloshing tanks at forward speed in the sea, Kim, B. et al[5] studied the coupled seakeeping and sloshing tanks in frequency domain. A forward-speed seakeeping theory was implemented to investigate the coupling effects. Mitra, S. et al[6] investigated the coupling effect in six degrees of freedom, solving the sloshing tank in potential flow equation and finite element method. The hybrid marine control system was applied to simulate the maneuvering of the ship.

In this paper, ship motion coupled with LNG tanks is simulated by CFD method. The internal sloshing tank and external sea waves are treated as an entire computational region, and both solved by RANS solver simultaneously. The Volume of fluid (VOF) method is applied to capture both outside wave surface and sloshing liquid. The computations are solved by our in house solver naoe-FOAM-SJTU with dynamic overset grid capability[7]. SUGGAR++ is used to obtain DCI[8], which connects the information of overset component grids. The solver contains 6DOF module, wave generation and damping module for various wave types.

To validate the current CFD method, five different filling ratios of LNG FPSO with two tanks in waves are selected. The simulation is compared with existing experimental results to prove the ability of our solver. To observe the coupling effect on large-amplitude motion, a benchmark ship KVLCC2 equipped with two LNG tanks and propeller is also considered. The simulation conditions include three different filling ratios (0%, 30% and 60%) in head waves with forward-speed.

Numerical Methods

The incompressible Reynolds-Averaged Navier-Stocks equations are adopted in this paper to investigate the viscous flow. The governing equations are:

$$\nabla \bullet \mathbf{U} = 0 \tag{1}$$

$$\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla \bullet (\rho(\mathbf{U} - \mathbf{U}_g)\mathbf{U}) = -\nabla p_d - \mathbf{g} \bullet \mathbf{x} \nabla \rho + \nabla \bullet (\mu_{eff} \nabla \mathbf{U}) + (\nabla \mathbf{U}) \bullet \nabla \mu_{eff} + f_\sigma + f_s \tag{2}$$

Where $\mathbf{U}$ is velocity field, $\mathbf{U}_g$ is velocity of grid nodes; $p_d = p - \rho \mathbf{g} \bullet \mathbf{x}$ is dynamic pressure; $\mu_{eff} = \rho(\nu + \nu_t)$ is effective dynamic viscosity, in which ν and ν_t are kinematic viscosity and eddy viscosity respectively. f_σ is the surface tension term in two phases model. The

solution of momentum and continuity equations is implemented by using the pressure-implicit spit operator (PISO) algorithm. A *k-ω* SST model is selected for turbulence closure[9].

The Volume of fluid (VOF) method with bounded compression techniques is applied to control numerical diffusion and capture the two-phase interface efficiently. The VOF transport equation is described below:

$$\frac{\partial \alpha}{\partial t} + \nabla \bullet [(\mathbf{U} - \mathbf{U}_g)\alpha] + \nabla \bullet [\mathbf{U}_r(1-\alpha)\alpha] = 0 \quad \phi_I = \sum_{i=1}^{n} \omega_i \bullet \phi_i \tag{3}$$

Where α is volume of fraction, indicating the relative proportion of fluid in each cell and its value is always between zero and one:

$$\begin{cases} \alpha = 0 & \text{air} \\ \alpha = 1 & \text{water} \\ 0 < \alpha < 1 & \text{interface} \end{cases} \tag{4}$$

The overset grid technique is implemented into OpenFOAM to handle the large-amplitude motion of ship and complex hierarchical motion of appendages such as rotating propeller and moving rudder[10]. The overset grid method allows separate overlapping grids to move independently without restrictions. The hole cells located outside the domain or of no interest, such as inside a body, are excluded from computation. The cells around hole cells are marked as fringe or receptor cells, they receive the information from other component grids by interpolation. Donor cells provide information to the fringe cell from donor grid. The value for fringe cell is obtained by the summation of weight coefficients and the values of donor cells

$$\phi_I = \sum_{i=1}^{n} \omega_i \bullet \phi_i \tag{5}$$

Where ω_i are the weight coefficients and ϕ_i is the value of donor cell; ϕ_I is the resulting value in the interpolated fringe cell; *n* is the number of donor cells and it is equal to eight if the structured grid is employed. And then the values of fringe cells need update with interpolated ones. The suitable approach for implicit scheme is to modify the matrix in the linear algebraic system after discretizing the equations

A fully 6DOF module with hierarchy of bodies are implemented. This module allows ship to move independently in the computational domain and in the meanwhile, the propeller is rotating around the propeller axis. Two coordinate systems, earth-fixed and ship-fixed systems are adopted in this 6DOF module. The forces and moments on ship hull and propeller are computed in earth-fixed system and then they are projected to ship-fixed system. The ship motions for the next time step are predicted by the projected forces and moments in ship fixed

system. For the movements of hierarchal objects, the propeller grid rotates first about a fixed axis in the ship coordinate system, and then both ship and propeller grids translate and rotate in the earth-fixed system according to the predicted motions, as shown in Fig.1. In the meanwhile, SUGGAR++ library is called to compute the DCI based on the new grid positions. OpenFOAM processors receive the new data right after the movements of the overset grids and start the computation for the next time step. For the details of the implementations of overset capability and 6DOF module can be referred to [8].

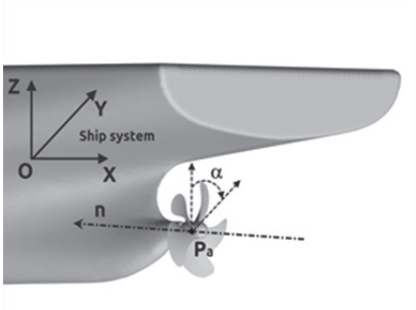


Fig.1 Demonstration of propeller rotating in the ship system

The incoming regular wave is generated by imposing the boundary conditions of α and **U** at the inlet. The linear Stokes wave in deep water is applied for the wave generation.

$$\xi(x,t) = a\cos[k(x-x_{cg})-\omega_e t] \tag{6}$$

$$u(x,y,z,t) = U_0 + a\omega e^{kz}\cos[k(x-x_{cg})-\omega_e t] \tag{7}$$

$$w(x,y,z,t) = a\omega e^{kz}\sin[k(x-x_{cg})-\omega_e t] \tag{8}$$

Where ξ is the wave elevation; a is the wave amplitude; k is the wave number; U_0 is the ship velocity; ω is the natural frequency of wave; ω_e is the encounter frequency, given by $\omega_e=\omega_e+kU_0$ in head waves; x_{cg} is the longitudinal gravity center of the ship model, it is used to adjust the phase of the incident wave to make the wave crest reach the gravity center of ship at $t = 0$.

Validation

Geometry and Condition

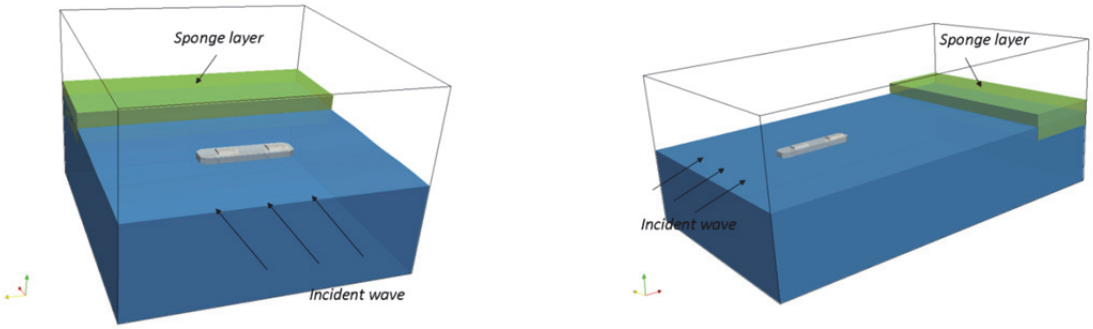
To validate the current method, a LNG FPSO model with two prismatic tanks is selected. The main particulars of LNG FPSO are shown in Table 1. To compare with experiments which have been done by Nam, B. W. et al[1], the LNG FPSO model is 1/100 scale of the full scale ship. The length, breadth and height of the fore tank and the aft tank are 49.68m, 46.92m, 32.23m and 56.62m, 46.92m, 32.23m respectively. The distance from the bottom of tank to

the keel line is 3.3m. The geometry of experimental ship model and numerical ship model are illustrated in Fig 3.

Five different filling conditions and two wave directions are included to verify the computations. The settings of numerical computation for head wave and beam wave are illustrated in Fig.2. To compare with experiment data, the filling ratios carried out as the same with those in experiments: 0%-0% (fore tank-aft tank), 20%-20%, 30%-30%, 57.5%-43.3% and 82.6%-23.5%. Those filling conditions are shown in Fig 4. The daft at each condition were kept the same, as well as longitudinal moment inertia.

Table 1 Main particular of LNG FPSO

Main particulars		Full Scale	Model
Scale factor	—	1	1/100
Length between perpendiculars	L_{PP} (m)	285	2.85
Maximum beam of waterline	B_{WL} (m)	63	0.63
Draft	T (m)	13	0.13
Displacement	Δ (m^3)	220017.6	220.0176
Natural period of roll	T_{ϕ} (s)	13	1.3
Vertical Center of Gravity (from keel)	KG (m)	16.5	0.165
Radius of gyration	K_{xx}	19.45	0.1945
	K_{yy}	71.25	0.7125

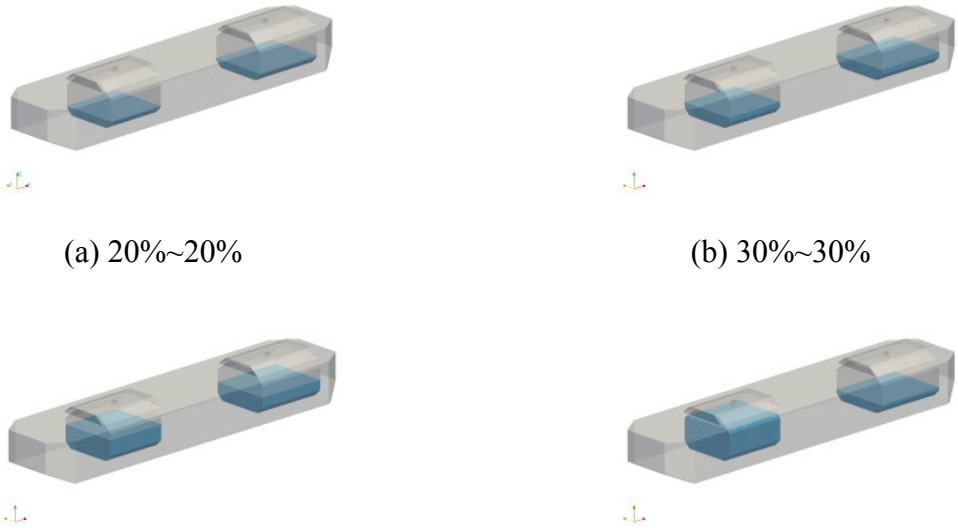


(a) Beam wave (b) Head wave

Fig.2 Setup of numerical computation

(a) Experimental ship model (b) Numerical simulation model

Fig. 3 geometry of LNG FPSO



(a) 20%~20% (b) 30%~30%

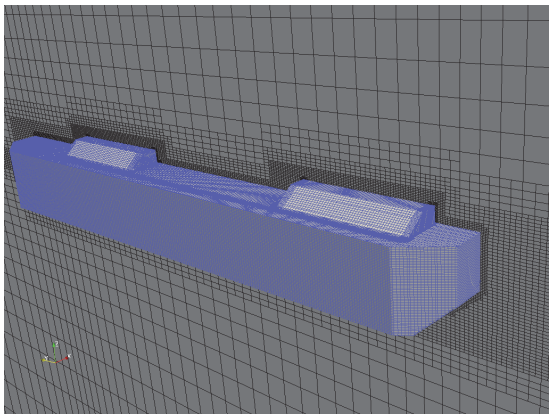
(c) 57.5%~43.3% (d) 82.6%~23.5%

Fig. 4 filling ratios of LNG FPSO equipped with LNG tanks

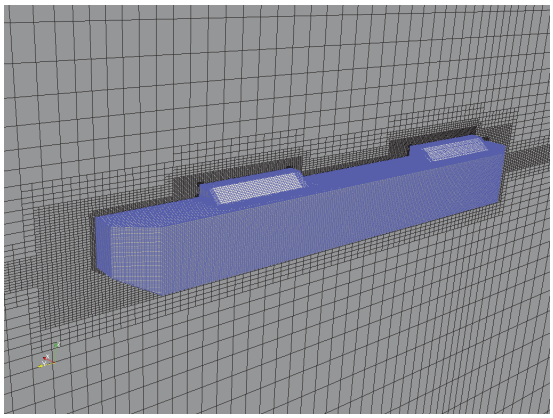
Considering the large-amplitude motion of LNG FPSO, the length of regular wave is chosen as 2.865m, 1.005 times length of the ship. Same to the experiment, the wave height is fixed to 0.025m, and encounter frequency is 4.6382.

Mesh

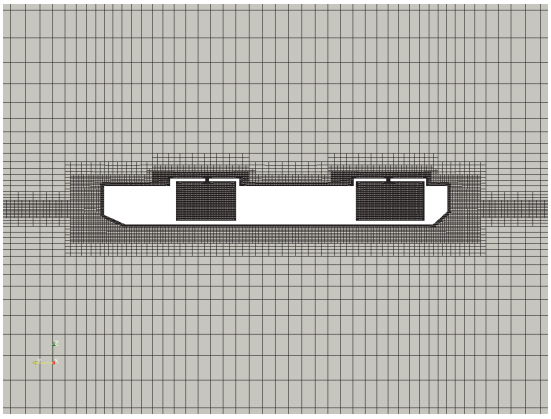
There are two computational domain in beam wave condition and head wave condition. The selected domain is described as $-1.0L_{pp}<x<2.0L_{pp}$, $-1.5L_{pp}<y<1.5L_{pp}$, $-1.0L_{pp}<z<1.0L_{pp}$ in beam wave condition; and $-2.0L_{pp}<x<4.0L_{pp}$, $-1.5L_{pp}<y<1.5L_{pp}$, $-1.0L_{pp}<z<1.0L_{pp}$ in head wave condition. The meshes are generated by *snappyHexMesh*, an auto mesh generation utility provided by OpenFOAM. The total cell numbers are around 2.1M, and the LNG tanks require additional 0.5M cells. The mesh details are shown in Fig. 5. Two small tunnels are used to connect the LNG tanks to the external region, which can keep the pressure inside the tanks same to the external region, and simplify the computations.



(a) bow



(b) stern

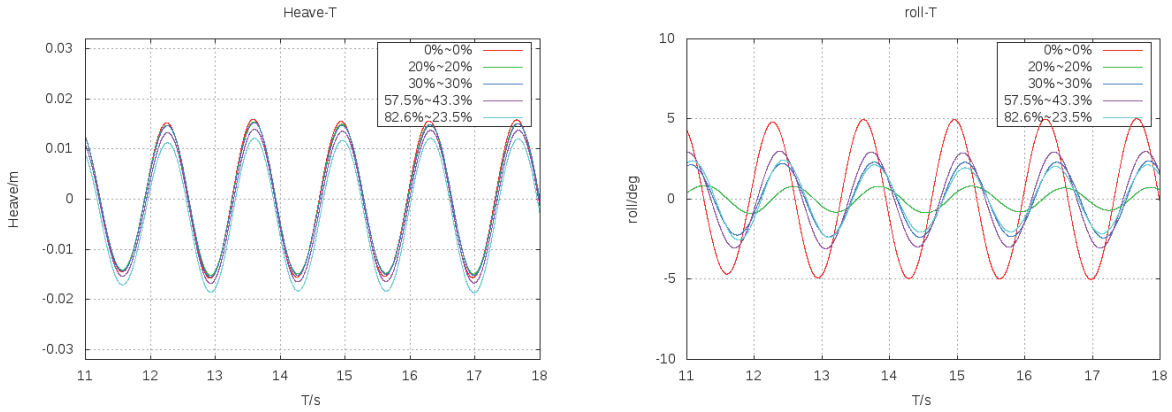


(c) ship hull with tanks

Fig.5 Demonstrations of meshes

Results

The ship motion was restricted to three degree-of-freedom, heave, pitch and roll. Beam wave conditions are analyzed first. Time histories of heave and roll motion are shown in Fig.6. The normalized motion amplitude and natural frequency were considered to compare with experimental data. The normalized roll motion is given as: $R_1 = \theta B/2A$, which θ is maximum degree of roll motion, B is beam of ship and A is wave amplitude; The normalized heave motion is given as: $H_1 = \xi/A$, which ξ is the maximum value of heave motion; and normalized natural frequency is given as: $T = \omega(L/g)^{(1/2)}$, which ω is natural frequency of water, L represents length of ship. Computations in this paper uses T=2.5 when the wave length is close to ship length.



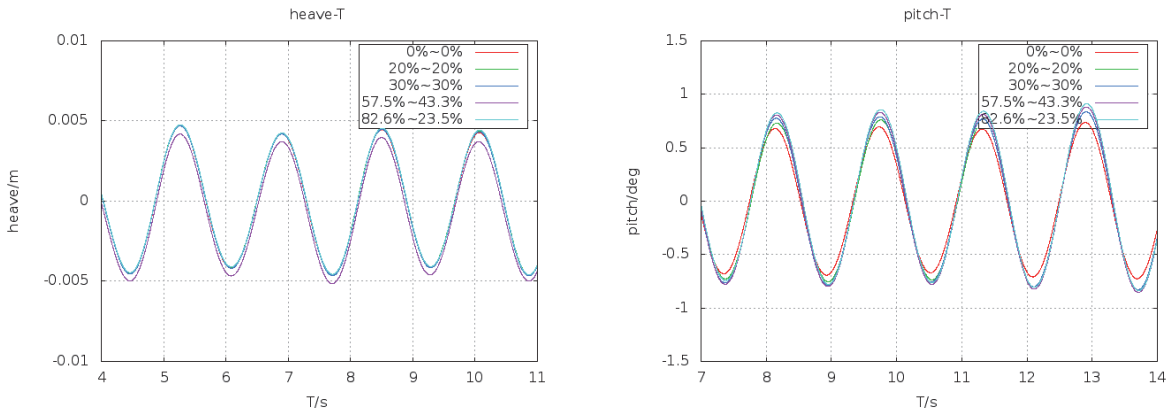
(a) Heave motion (b) Roll motion

Fig6. Time history of heave and roll motion with different filling ratios in beam wave

Table 2 Comparison of ship motion between CFD and experiments in beam wave

No	Filling ratio	EFD($\boldsymbol{R_1}$)	CFD($\boldsymbol{R_1}$)	EFD($\boldsymbol{H_1}$)	CFD($\boldsymbol{H_1}$)
1	0%~0%	1.85	2.00(8%)	1.25	1.28(2.4%)
2	20%~20%	0.60	0.53(-12%)	-	1.22
3	30%~30%	1.25	1.12(-10%)	-	1.21
4	57.5%~43.3%	1.30	1.28(-1.5%)	-	1.10
5	82.6%~23.5%	1.20	1.10(-8.3%)	-	1.04

Table 2 shows the comparison of roll motion and heave motion between current computation and experiments in beam waves. Five different filling conditions were considered, and the results fairly agree with those in experiments. The head wave conditions were also selected to validate the method in head waves. Fig.7 illustrates the time history of heave and pitch motion.



(a) Heave motion (b) Pitch motion

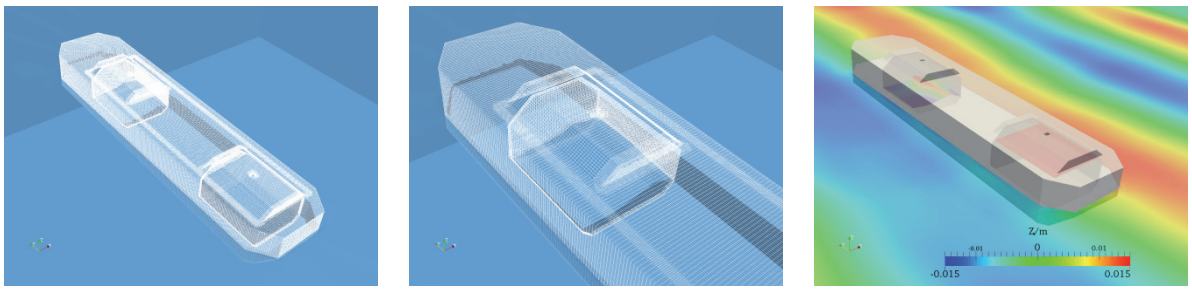
Fig7. Time history of heave and pitch motion with different filling ratios in head wave

Table 3 shows the results comparison between current simulation and experimental results. The dimensionless parameters of ship motion are considered. The normalized pitch motion is given as:$P_1 = \theta L/2A$, which L is ship length. Five filling conditions were considered and compared to the existing experimental data, the simulation results fairly agree with the experimental results.

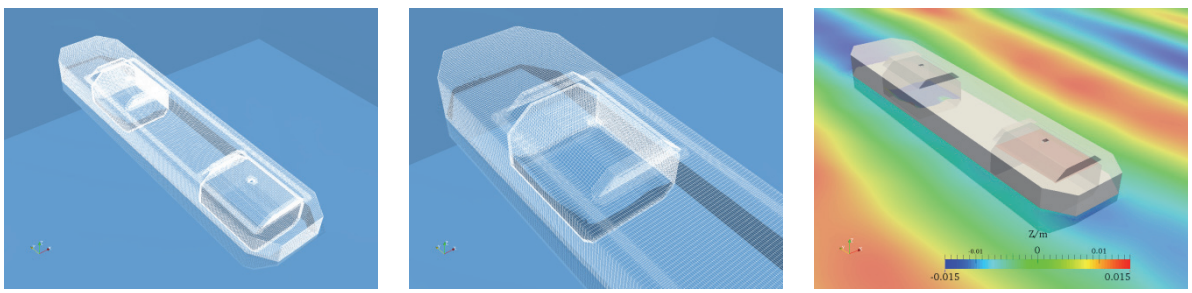
Table 3 Comparison of ship motion between CFD and experiments in head wave

No	Filling ratio	EFD($\mathbf{P_1}$)	CFD($\mathbf{P_1}$)	EFD($\boldsymbol{H_1}$)	CFD($\boldsymbol{H_1}$)
1	0%~0%	1.20	1.31(9.3%)	0.12	0.14(16%)
2	20%~20%	1.13	1.30(15%)	-	0.14
3	30%~30%	1.30	1.45(11%)	0.12	0.14(16%)
4	57.5%~43.3%	-	1.60	-	0.138
5	82.6%~23.5%	-	1.63	-	0.14

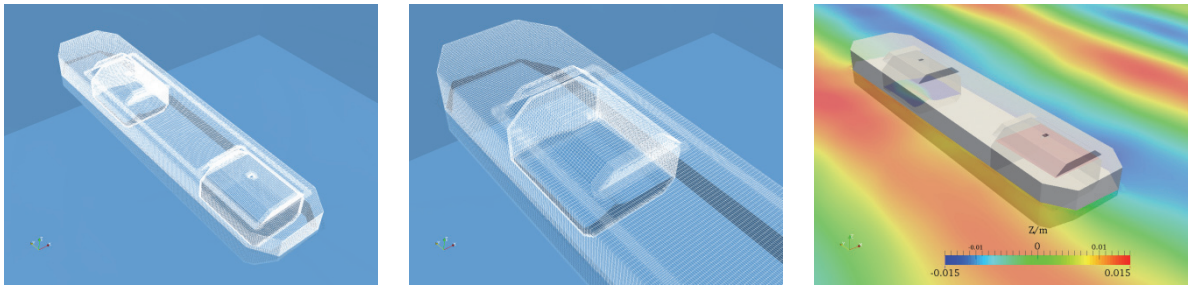
Fig.6 and Fig.7 indicates that the ship exhibits sinusoidal motion both in head and beam waves. The coupling effects are limited in head wave. In beam wave condition, the coupling effects of ship motion and tank sloshing are not obvious in heave motion, shown in Fig.6(a), but quite significant in roll motion, shown in Fig.6(b). The four partially filling conditions of sloshing tanks all reduce the roll amplitude of ship motion; on the contrary, although not obvious in head waves, the filling conditions increase the amplitude of ship motion. In beam wave, for low-filling condition, like 20%~20%, the decrease in amplitude of roll motion is evident and thus shows great coupling effect. For the water in tanks is shallow, the sloshing in tanks is more violent and influence ship motion more.



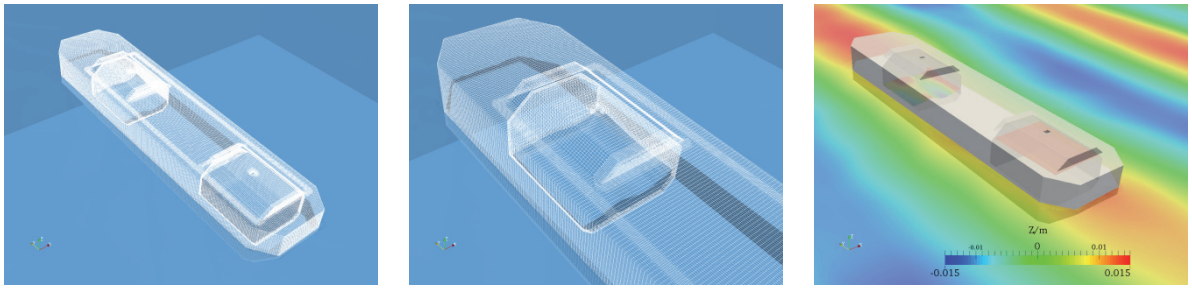
t/T=0



t/T=1/4



t/T=1/2



t/T=3/4

Fig.8 Four snapshots of LNG FPSO motion in beam regular waves. From left to right: global view, detail view of the aft tank sloshing and view of free surface in turn. 82.6%~23.5% filled, wave propagates from left to right.

Fig.8 illustrates four snapshots of ship motion coupled with 82.6%-23.5% filling ratio in one period (1.35s), wave propagates from left to right. The fore tank has insignificant coupling effects, so the aft tank is studied in details. The flow in aft tank shows different phase to that of ship motion. In the time t/T=0, ship stays in balance position, the sloshing liquid in aft tank starts to move from right to left. At t/T=1/4, the ship is in the region of wave trough, and begins to roll to the left (towards the wave direction); the peak of the tank liquid reaches the left bulkhead. At t/T=1/2, the ship returns to the balance position, the peak of the sloshing liquid moves to the right. At t/T=3/4, the wave crest reaches the ship and ship begins roll to the right, the peak of the in-tank flow arrives at the right bulkhead.

Coupling Effects on KVLCC2 with a Propeller

Geometry and Conditions

To figure out the coupling effects on large-amplitude motion, KVLCC2, a benchmark ship in Gothenburg Workshop 2010 (G2010)[11] equipped with two LNG tanks and a propeller is considered. Fig.9 shows geometry of the ship and its LNG tanks and Table 4 illustrates the principle dimensions of KVLCC2. The modified KVLCC2 equipped with two identical LNG tanks, and the main particulars and settings of those tanks are shown in Fig. 10. The tanks are in model scale, and all the numbers are in mm.

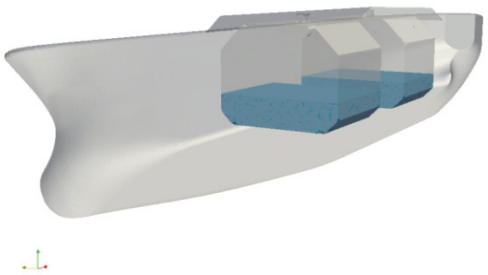
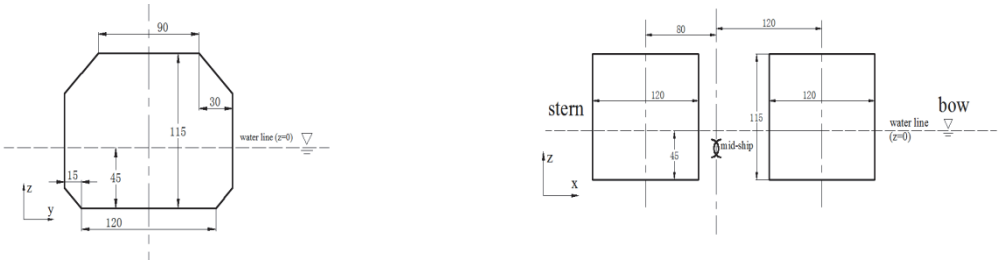


Fig.9 Geometry of KVLCC2 with two LNG tanks



(a) Transverse section of a LNG tank (b) Longitudinal section of two LNG tanks

Fig. 10 Geometry of LNG tanks

Table 4 Main dimensions of KVLCC2

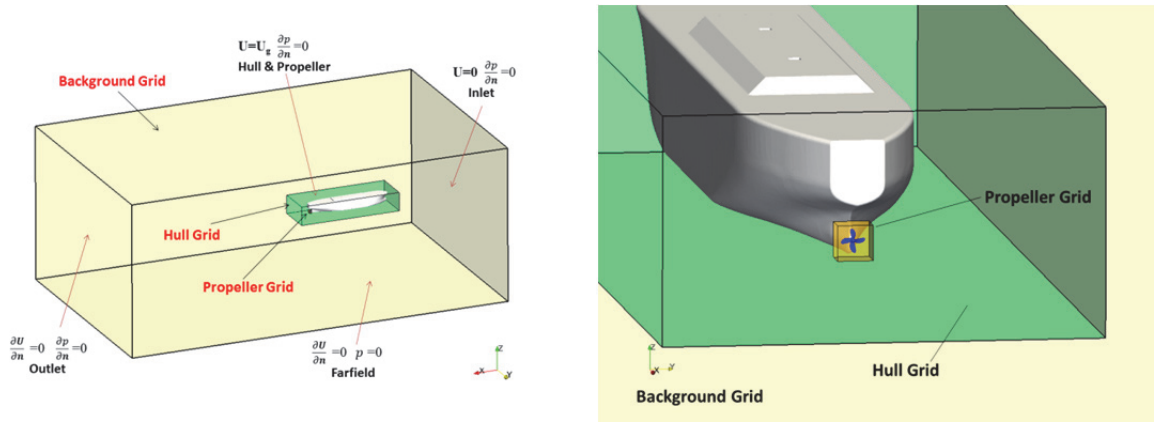
Main particulars		Full Scale	Model
Length between perpendiculars	L_{pp} (m)	320	3.200
Maximum beam of waterline	B_{WL} (m)	58	0.580
Depth	D (m)	30	0.300
Draft	T (m)	20.8	0.208
Displacement	Δ (m^3)	312622	0.313
Wetted area	SW (m^2)	27194	2.719
Vertical Center of Gravity (from keel)	KG (m)	18.6	0.186
Moment of Inertia	K_{xx}/B	0.4	0.400
	K_{yy}/L_{pp}, K_{zz}/L_{pp}	0.25	0.250

To evaluate the coupling effects on self-propulsion, KVLCC2 with a certain velocity in head wave is calculated. The motions are allowed for roll, heave and pitch. Three different filling ratios are considered, 30%, 60% and no sloshing, respectively. The wave length is equal to 3.2m, and wave height is 0.12m. The ship has a forward-speed of Fr=0.179.

Mesh and Computational Domain

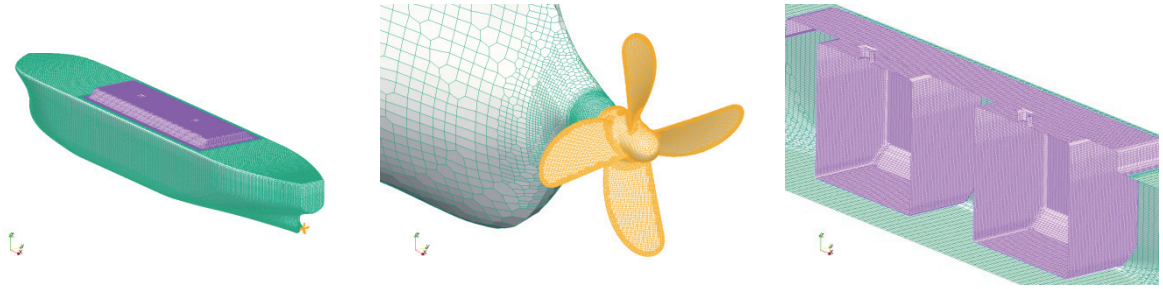
The space coordinate range of computational domain is $-1.0L_{pp}<x<4.0L_{pp}$, $-1.5L_{pp}<y<1.5L_{pp}$, $-1.0L_{pp}<z<1.0L_{pp}$. The mesh is generated by automatic mesh generation tool *snappyHexMesh*. The overset grids consist of hull, background and propeller grids. The computational domain

contains around 3.9M grid cells, in which hull uses 2.65M grid cells and propeller possesses 0.68M grid cells. Some regions have been refined to capture free surface, violent flow and vortex structure. Boundary conditions and layout of overset grid systems are displayed in Fig. 11, and the surface mesh of hull, tanks and propeller are shown in Fig. 12.



(a) Overall view (b) Stern view

Fig. 11 Layout of overset grid system



(a) Hull mesh (b) Propeller mesh (c) Tank mesh

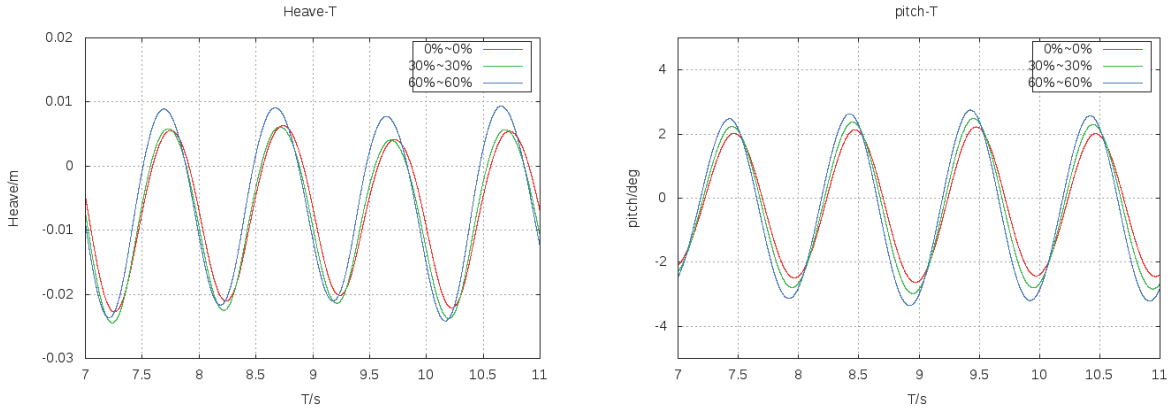
Fig. 12 Mesh generation

Results

The pitch and heave motion of ship with a forward-speed in head wave are shown in Fig.13. Two different filling conditions are considered and compared to that without sloshing. The coupling effects are observed in pitch and heave motion but they are not prominent. The 60% filled ship has more violent motion than ship with 30% filling ratio and without sloshing.

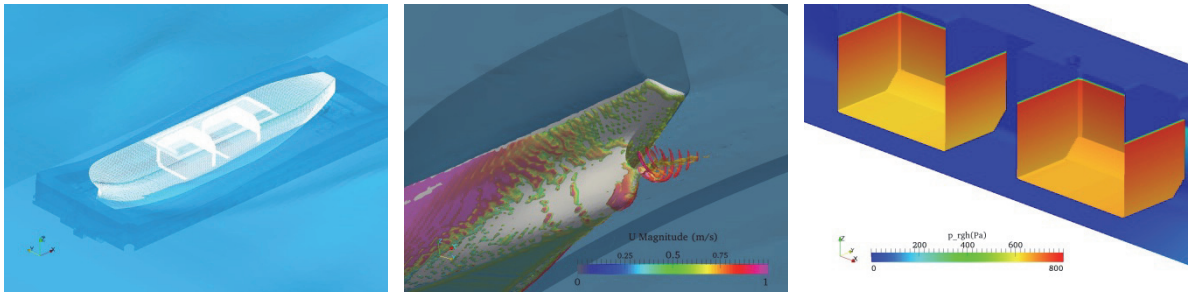
Fig.14 illustrates four snapshots of ship motion and the dynamic pressure on bulkhead with 60% filling ratio in an encounter period. Propeller vortices behind ship stern are illustrated by iso-surfaces of Q=100. The iso-surfaces are colored by velocity magnitude. At t/T=0, ship stays at balanced position, the dynamic pressure on bulkhead stays the same in fore and aft tank. The value of dynamic pressure near in-tank liquid surface is larger than that near bottom bulkhead, which means liquid slosh more violent near surface. At t/T=1/4, the wave crest reaches and ship bow nearly buries into wave. The dynamic pressure on bulkhead decreases in fore tank and it increases in aft tank. At t/T=0 and 1/4, the velocity of propeller vortices is high, for the ship stern is near the surface, the load on the propeller blades is small

correspondingly. At t/T=1/2, ship returns back to balanced position, and the dynamic pressure on bulkhead stays the same in fore and aft tank. However, the dynamic pressure at this time is smaller than that at t/T=0 and tank liquid slosh more violent near bottom than that near in-tank water surface. At t/T=3/4, the trough of wave reaches and ship bow nearly comes out of the surface. The dynamic pressure on bulkhead decreases in aft tank while it increases in fore tank. The tank liquid is not affected by ship motion intensively, for the surface in tanks slosh slight. Moreover, the dynamic pressure on bulkhead in fore and aft tank shows phase difference in an encounter period. At t/T=1/2 and 3/4, the ship stern buries into water, thus the loads on the propeller blades increase, and the velocity of propeller vortices is lower than that at t/T=0 and 1/4.

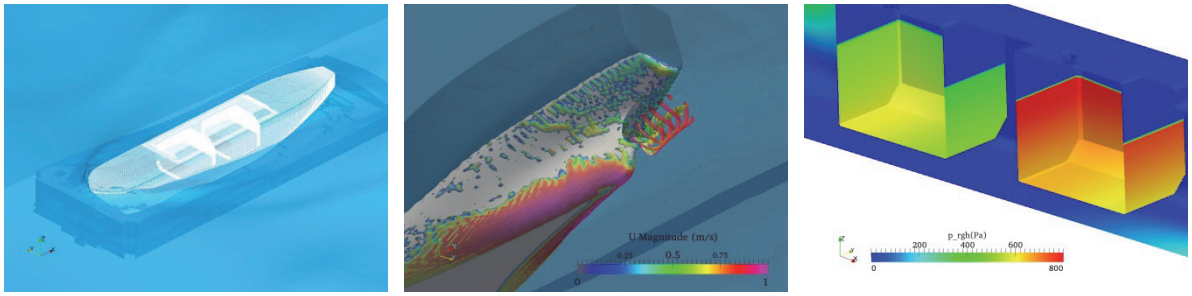


(a) Heave motion (b) Pitch motion

Fig.13 Time history of heave and pitch motion at forward speed with different filling ratios in head wave.



t/T=0



t/T=1/4

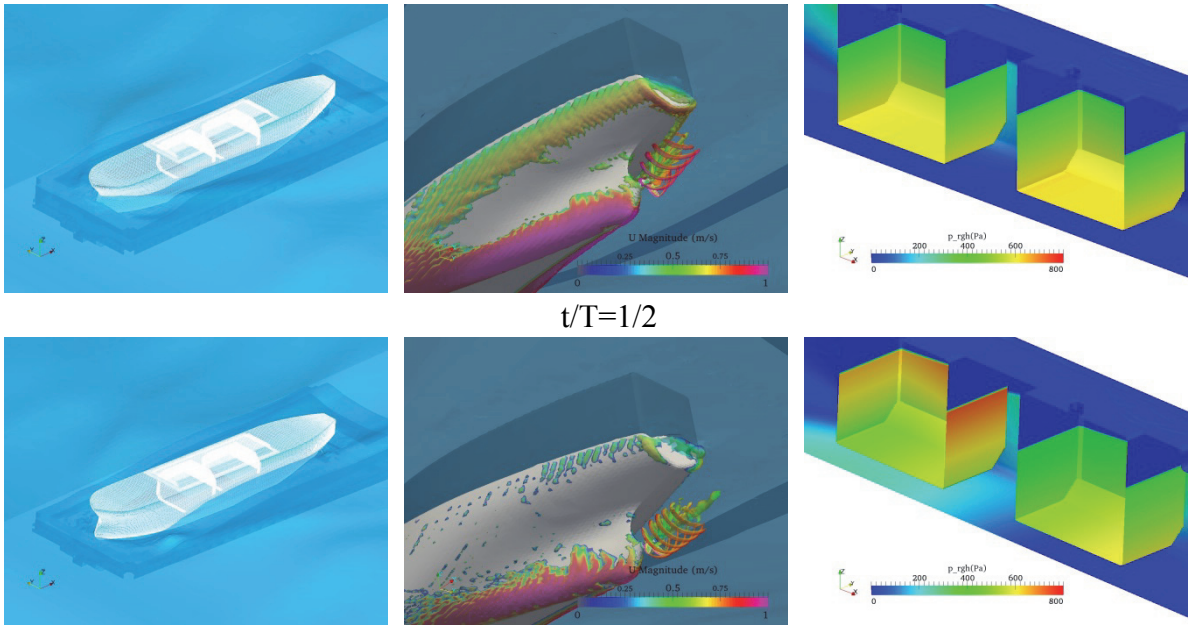


Fig. 14 Snapshots of ship motion, Q iso-surfaces and dynamic pressure on bulkhead in one period

Conclusion

In this paper, the large-amplitude of ship motion fully coupled with internal sloshing tanks is studied. The numerical simulations are performed by the solver naoe-FOAM-SJTU, which is developed based on open source CFD package OpenFOAM and implemented with dynamic overset grid technique. The internal tank sloshing and external wave excitation are computed simultaneously by solving RANS equations. Two phase interface is captured by VOF method. To validate the current method, LNG FPSO is chosen to compare with existing measurements data. Five different filling conditions are considered both in the head and beam wave. The results show fairly agreement with those in experiments. At the meantime, the coupling effects are investigated. With the wave length equal to 1.005 times ship length, the sloshing has little effect on the heave and pitch motion both in the head and beam wave. However, the sloshing has remarkable effect on roll motion in the beam wave condition. The comparison between four different filling ratios with non-filling ratio indicates that all these four kinds of sloshing reduce the roll amplitude of ship motions, especially the low filling ratios, like 20% filled tanks.

To make a further study of large-amplitude ship motion coupled with sloshing tanks with a forward-speed in head waves, a KVLCC2 model equipped with two LNG tanks and a propeller is chosen. In the condition of wave length equal to ship length, two filling ratios are considered to compare with non-filling ratio condition in head wave. With the forward speed, the coupling effect can be observed but it is not obvious. Unlike coupling effect on roll motion in the beam wave, the tank sloshing increase the amplitude motion both in heave and pitch motion, especially for the 60% filled tanks.

However, in this stage, only one wave condition is considered in the simulation, thus more wave conditions need to be computed in the future work to fully investigate the ship motion coupled with LNG tank sloshing.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (51379125, 51490675, 11432009, 51579145, 11272120), Chang Jiang Scholars Program (T2014099), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (2013022), Innovative Special Project of Numerical Tank of Ministry of Industry and Information Technology of China (2016-23) and Lloyd's Register Foundation for doctoral students, to which the authors are most grateful.

Reference

[1] Nam, B. W., Kim, Y., Kim, D. W., and Kim, Y. S. (2009). Experimental and numerical studies on ship motion responses coupled with sloshing in waves, *Journal of Ship Research* **53**(2), 68-82.

[2] LI, Y. L., ZHU, R. C., MIAO, G. P., and Ju, F. A. N. (2012). Simulation of tank sloshing based on OpenFOAM and coupling with ship motions in time domain, *Journal of Hydrodynamics* **24**(3), 450-457.

[3] Jiang, S. C., Teng, B., Bai, W., and Gou, Y. (2015). Numerical Simulation of Coupling Effect between Ship Motion and Liquid Sloshing under wave action, *Ocean Engineering* **108**, 140-154.

[4] Shen, Z., and Wan, D. C. (2012). Numerical Simulations of Large-Amplitude Motions of KVLCC2 With Tank Liquid Sloshing in Waves, In *Proc 2nd Int Conf Violent Flows, Nantes, France, Ecole Centrale Nantes* (pp. 149-156).

[5] Kim, B., and Shin, Y. S. (2008, January). Coupled seakeeping with liquid sloshing in ship tanks, In *ASME 2008 27th International Conference on Offshore Mechanics and Arctic Engineering* (pp. 247-257). American Society of Mechanical Engineers.

[6] Mitra, S., Wang, C. Z., Reddy, J. N., and Khoo, B. C. (2012). A 3D fully coupled analysis of nonlinear sloshing and ship motion, *Ocean Engineering* **39**, 1-13.

[7] Shen, Z., Carrica, P. M., and Wan, D. (2014, June). Ship Motions of KCS in Head Waves With Rotating Propeller Using Overset Grid Method, In *ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering* (pp. V002T08A043-V002T08A043). American Society of Mechanical Engineers.

[8] Shen, Z., Wan, D., and Carrica, P. M. (2015). Dynamic overset grids in OpenFOAM with application to KCS self-propulsion and maneuvering, *Ocean Engineering* **108**, 287-306.

[9] Dhakal, T. P., and Walters, D. K. (2009, January). Curvature and rotation sensitive variants of the K-Omega SST turbulence model, In *ASME 2009 Fluids Engineering Division Summer Meeting* (pp. 2221-2229). American Society of Mechanical Engineers.

[10] Shen, Z., and Wan, D. (2014, August). Computation of Steady Viscous Flows around Ship with Free Surface by Overset Grids Techniques in OpenFOAM, In *The Twenty-fourth International Ocean and Polar Engineering Conference*. International Society of Offshore and Polar Engineers.

[11] *CFD Workshop in Gothenburg 2010*. : http://www.gothenburg2010.org

COMPRESSIBLE MULTIMATERIAL FLOWS

F. Bernard[1], A. de Brauer[2], A. Iollo[1,a)], T. Milcent[3]and H. Telib[4]

[1]IMB, University of Bordeaux, UMR CNRS 5251; INRIA Memphis Team, F-33400 Talence, France

[2]IIHR, University of Iowa, Iowa City, Iowa 52242-1585, USA

[3]I2M, University of Bordeaux, UMR CNRS 5295; Arts et Métiers Paristech, F-33600 Pessac, France

[4]Optimad Engineering, I-10143, Turin, Italy.

[a)]Corresponding and presenting author: angelo.iollo@inria.fr

1 ABSTRACT

We consider hyperbolic models of gas flows past unsteady or elastic-plastic solids. The numerical framework is based on hierarchical cartesian grids, implicit representation of fluid-solid interfaces, stable and accurate discretization schemes. We present examples relative to compressible flows in unsteady aerodynamics, high-speed elastoplastic impacts and rarefied re-entry flows.

2 MODELS

2.1 Elasto-plastic materials

We consider two models. The first is relative to a compressible elastic-plastic continuum medium. This model was introduced in the literature thanks to several authors [6, 10, 9, 3, 5]. We follow here the formulation presented in [7, 4] and extend it to plasticity modelling. The equations of mass, momentum, deformation and energy conservation are given by

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho u) = 0 \\ \partial_t(\rho u) + \mathrm{div}_x(\rho u \otimes u - \sigma) = 0 \\ \partial_t(\nabla_x Y) + \nabla_x(u \cdot \nabla_x Y) = 0 \\ \partial_t(\rho e) + \mathrm{div}_x(\rho e u - \sigma^T u) = 0 \end{cases} \tag{1}$$

Here $Y(x,t)$ are the backward characteristics that for a time t and a point x in the deformed configuration, give the corresponding initial point.

We assume that the internal energy per unit mass $\varepsilon = e - \frac{1}{2}|u|^2$ is the sum of a term accounting for volume deformation that depends on ρ and entropy s, and a term accounting for isochoric deformation depending on the modified left Cauchy-Green tensor $\overline{B}$ given by $\overline{B}(x,t) = [\nabla_x Y]^{-1}[\nabla_x Y]^{-T}/J^{\frac{2}{3}}(x,t)$, $J(x,t) = \det([\nabla_x Y])^{-1}$. A general constitutive law that models gas, fluids and elastic solids is then given by

$$\varepsilon(\rho, s, \nabla_x Y) = \frac{\kappa(s)\rho^{\gamma-1}}{\gamma - 1} + \frac{p_\infty}{\rho} + \frac{\chi}{\rho_0}(\mathrm{Tr}(\overline{B}) - 3) \tag{2}$$

where the first term accounts for a perfect gas, the second for a stiffened gas (e.g. water) and the third for a neohookean elastic solid. The Cauchy stress tensor is obtained from the above constitutive law. Here $\kappa(s) = \exp(s/c_v)$ and c_v, γ, p_∞, χ are positive constants that characterize a given material. Compressible Euler equations are included in this model.

Plasticity describes the deformation of a material undergoing non-reversible changes of shape in response to applied forces. The deformation can be modeled by the composition of a plastic and an elastic deformation [8]. We introduce the backward characteristics for elastic and plastic deformations denoted by Y^e and Y^p, respectively. Let us define the deviatoric part of the stress tensor $\mathrm{dev}(\sigma) = \sigma - \frac{\mathrm{Tr}(\sigma)}{3} I$. Experimentally plasticity occurs when the stress exceeds a critical value. The yield function of von Misses $f_{VM}(\sigma) = |\mathrm{dev}(\sigma)|^2 - \frac{2}{3}(\sigma_y)^2$ defines a yield surface $f_{VM}(\sigma) = 0$ where σ_y is the plastic yield limit. We restrict ourselves to the case of perfect plasticity where σ_y is a constant.

A constitutive law for plasticity [9, 1] is defined by

$$\partial_t(\nabla_x Y^e) + \nabla_x(u \cdot \nabla_x Y^e) = \frac{1}{\chi\tau}[\nabla_x Y^e]\mathrm{dev}(\sigma) \tag{3}$$

where χ is the shear modulus and τ is the relaxation time of the plastic process. Beyond yield, plasticity appears as a source term in the equation of deformations and can be seen as a penalization of the deviatoric part of σ.

2.2 *Rarefied polyatomic flows*

We consider a BGK model for polyatomic gases. Going from monoatomic to polyatomic gases implies that additional energy degrees of freedom are considered. In the classical BGK model, only translational energy degrees of freedom are taken into account. We now consider a more general case with d energy degrees of freedom including rotational and vibrational energy degrees of freedom. The idea is to consider these additional energy degrees of freedom in the expression of the maxwellian distribution function. Moreover, we consider a general case where the energy is not equally distributed between the energy degrees of freedom. Let $\boldsymbol{\eta} \in \mathbb{R}^d$ the vector of the energy degrees of freedom ($\boldsymbol{\eta} = \boldsymbol{\xi}$ for the BGK model), $\overline{\boldsymbol{\eta}} \in \mathbb{R}^d$ the macroscopic value on which the equilibrium function is centered ($\mathbf{U}$ in the case of the BGK model), $\boldsymbol{\lambda} \in \mathbb{R}^d$ the vector of the coefficient giving the distribution of the energy between the degrees of freedom ($1/2T$ in the case of the BGK model for the three translational energy degrees of freedom). The model reads:

$$\frac{\partial f}{\partial t} + \boldsymbol{\xi} \cdot \nabla_{\mathbf{x}} f = \frac{1}{\tau}\Big(M_f - f\Big) \tag{4}$$

$$M_f(\mathbf{x}, \boldsymbol{\eta}, t) = \rho(\mathbf{x}, t) \prod_{k=1,d} \Big(\frac{\lambda_k}{\pi}\Big)^{1/2} \exp\Big(- (\lambda_k(\eta_k - \overline{\eta}_k)^2\Big) \tag{5}$$

The evolution of $\boldsymbol{\lambda}$ is governed by the equation of energy conservation and by a relaxation of the equilibrium temperature of the rotational degrees of freedom, Θ, towards the equilibrium temperature of the translational degrees of freedom denoted Λ:

$$\partial_t\Theta + \mathbf{U} \cdot \nabla\Theta = \frac{1}{Z_r\tau}(\Lambda - \Theta) \tag{6}$$

where Z_r is a given parameter corresponding the rotation frequency of the gas molecules.

3 NUMERICAL ILLUSTRATIONS

3.1 *Plastic impact*

We have extended the scheme described in [7, 4] to model elasto-plastic flows. The scheme is based on a sharp-interface locally non-conservative approximate Riemann solver that has been validated in 2D and 3D.

Here we show a 2D test case where an iron circular projectile is impacting onto an aluminium flat plate fixed to the upper and lower boundaries of the computational domain. The initial horizontal velocity of the iron projectile is $1000m.s^{-1}$. The physical parameters for the different materials are found in the literature and the computational domain is $[-0.3, 0.7]m \times [-0.4, 0.4]$m. The computation is performed on a 2000×1600 mesh with 144 processors. Homogeneous Neumann conditions are imposed on the left and right borders and embedded on the top and bottom.

The results are presented in Fig. 2 depicting a Schlieren image and the value of the von Mises criteria $|\mathrm{dev}(\sigma)|^2 - \frac{2}{3}(\sigma_y)^2$ at a time steps corresponding to an early impact stage and to an highly deformed plastic state. A longitudinal wave propagating in the plate is followed by a shear wave that causes the plasticity of the material. We can observe that the plate, initially straight, is strongly deformed and forms a long filament; the projectile, initially round, is considerably flattened. Shock waves and contact discontinuities characterise the air flow.

3.2 *Capsule re-entry*

A capsule based on Apollo design is immersed in a rarefied gas flow at Mach 5. Free flow conditions are imposed on the boundaries of the domain except at the inlet where the state is imposed. On the capsule we enforce a zero velocity (with respect to the capsule) and a temperature equal to 1 in dimensionless variables. The capsule will then move according to the torque due to the fluid force on the body until an equilibrium depending on the position of the center of mass is

attained. The numerical scheme for a monoatomic gas is described in [2]. We extend that scheme to polyatomic gases. The computation is performed on a 80x80x80 grid in space and a 21x21x21 grid in velocity with a Knudsen number of 10^{-2} with 128 processors. The simulation took about 2 days.

4 CONCLUSIONS

In the proposed presentation we will describe the hierarchical schemes used for these simulations. In particular, we will detail how to recover consistency and accuracy at the unsteady interfaces that arbitrarily cross the grid. Additional results in 3D aerodynamics will be presented.

References

[1] P.T. Barton, R. Deiterding, D. Meiron, and D. Pullin. Eulerian adaptive finite-difference method for high-velocity impact and penetration problems. Journal of Computational Physics, 240(C):76–99, 2013. (Cited on page 2.)

[2] F. Bernard, A. Iollo, and G. Puppo. Accurate Asymptotic Preserving Boundary Conditions for Kinetic Equations on Cartesian Grids. Journal of Scientific Computing, January 2015. (Cited on page 3.)

[3] G.H. Cottet, E. Maitre, and T. Milcent. Eulerian formulation and level set models for incompressible fluid-structure interaction. ESAIM: Mathematical Modelling and Numerical Analysis, 42:471–492, 2008. (Cited on page 1.)

[4] A. de Brauer, A. Iollo, and T. Milcent. A cartesian scheme for compressible multimaterial models in 3d. Journal of Computational Physics, 313(C):121–143, May 2016. (Cited on pages 1 and 2.)

[5] N. Favrie, S.L. Gavrilyuk, and R. Saurel. Solid-fluid diffuse interface model in cases of extreme deformations. Journal of Computational Physics, 228(16):6037–6077, 2009. (Cited on page 1.)

[6] S.K. Godunov. Elements of continuum mechanics. Nauka Moscow, 1978. (Cited on page 1.)

[7] Y. Gorsse, A. Iollo, T. Milcent, and H. Telib. A simple cartesian scheme for compressible multimaterials. Journal of Computational Physics, 272:772–798, 2014. (Cited on pages 1 and 2.)

[8] E.H. Lee and D.T. Liu. Finite-strain elastic-plastic theory with application to plane-wave analysis. Journal of Applied Physics, 38(1):19–27, 1967. (Cited on page 1.)

[9] G.H. Miller and P. Colella. A high-order eulerian godunov method for elastic-plastic flow in solids. Journal of Computational Physics, 167(1):131–176, 2001. (Cited on pages 1 and 2.)

[10] B.J. Plohr and D.H. Sharp. A conservative eulerian formulation of the equations for elastic flow. Advances in Applied Mathematics, 9:481–499, 1988. (Cited on page 1.)

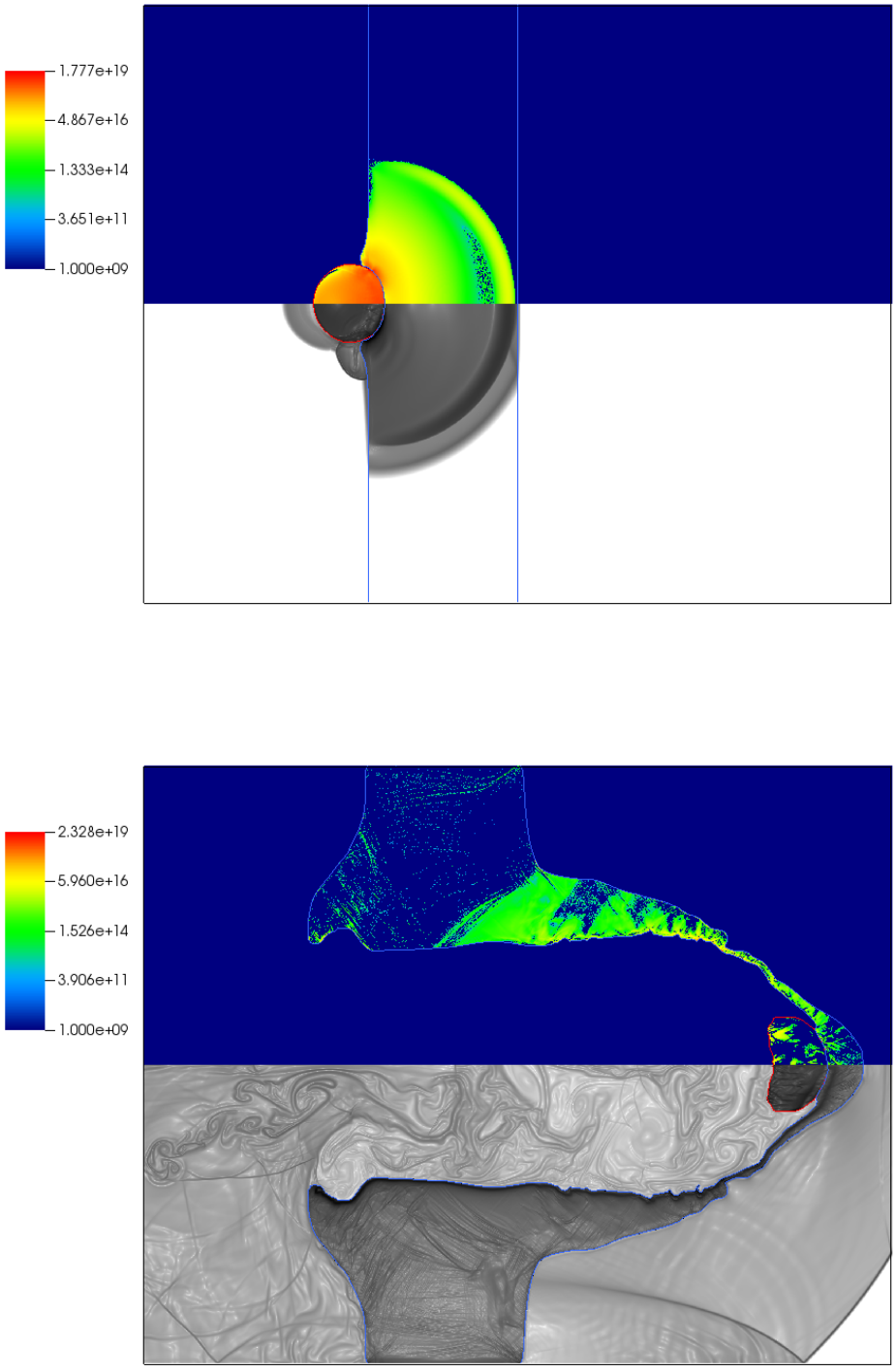


Figure 1. Iron round projectile on an aluminium shield in air. Schlieren image and von Mises criterium at $t = .03$ms and $t = 1.04$ms

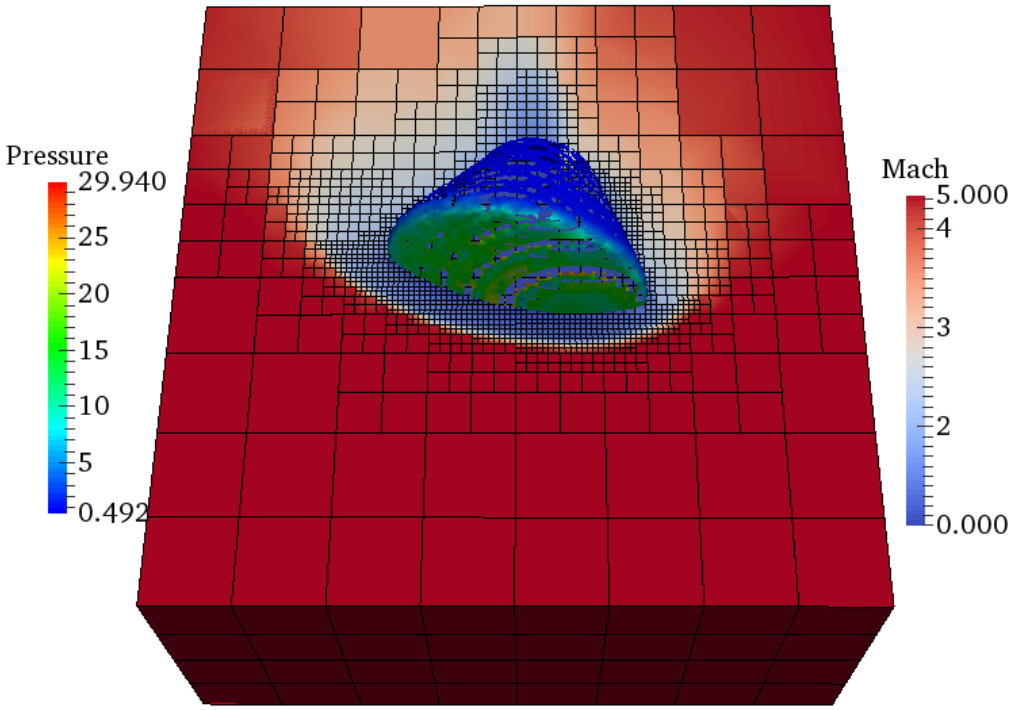


Figure 2. Octree simulation of a re-entry capsule in a rarefied polyatomic gas.

The traffic jerk for the full velocity different car-following model

Y Liu[1,2,3], H. X. Ge[1,2,3], KL Tsui[4], [*,†]KK Yuen[4], S. M. Lo[4]

[1] Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China

[2] Jiangsu Province Collaborative Innovation Center for Modern Urban Traffic Technologies, Nanjing 210096, China

[3] National Traffic Management Engineering and Technology Research Centre Ningbo University Sub-centre, Ningbo 315211, China

[4]Departement of Civil and Architectural Engineering, City University of Hong Kong, Kowloon, Hong Kong 999077, China

[*]Presenting author: bckkyuen@cityu.edu.hk

[†]Corresponding author: bckkyuen@cityu.edu.hk

Abstract

Considering sudden change in vehicle's acceleration, an improved car-following model with a feedback control signal jerk was studied and presented in this paper. Stability analysis of the modified model was achieved according to the control theory method. Through theoretical analysis, the modified model may provide insights for developing management strategy to improve traffic jams.

Keyword: car-following models, feedback control, traffic jerk

1. Introduction

Traffic jams have been studied by many traffic simulation models namely, the car following models, the hydrodynamic models, the cellular automation models and the gas kinetic models [1-15]. In 1999, Konishi et al. [16] put forward a chaotic car-following model by setting the time delay feedback control, and researched single-lane traffic operation without reverse phenomenon under an open boundary condition. In 2006, Zhao et al. [17] put forward a control method for the suppression of the traffic jam. They gave a control signal which included the effect of velocity difference between the preceding and the considered vehicle. In 2007, Han and Ge [18] presented a coupled map car-following model for traffic flows with the consideration of the application of intelligent transportation systems. The control signal uniform to the velocity difference between the i-th vehicle in front and the (i+1)-th vehicle, and the developed model can improve the stability of traffic flow. Other research was connected with the control signal has been carried out recently [19, 20, 21, 22].

In 1961, Newell [2] put forward a car-following model with a differential equation and give graphic description of the optimal velocity (OV) function. In 1995, Bando et al. [3] proposed optimal velocity model (OVM) for car-following model. In the OVM, the acceleration of the n-th vehicle at time is identified by the difference between the actual velocity and an optimal velocity, which depends on the headway to the car in the front. In 2001 Jiang et al. [23] presented full velocity different model for car-following theory (FVDM) by considering both negative and positive velocity difference, which can give a better description of starting process than OVM. In 2012, Yu et al. [24] proposed a full velocity difference and acceleration model (FVDAM).The following cars in FVDAM react more quickly than those in FVDM and the stability of FVDAM is more stable than that of FVDM. However, no studies have ever tried to consider control method with jerk for the FVDM. Based on

previous work, this paper investigates a new control scheme considering jerk. As we known, the vehicle's velocity changes are its acceleration, which means how quickly the vehicle increases and loses speed. Furthermore, abrupt change in vehicle's acceleration is called 'jerk', and it will affect the stability of traffic flow, so FVDM with the traffic jerk is studied in this paper.

In section 2, the FVDM is recovered and stability analysis is carried out. In section 3, the car-following model including a feedback control signal is put forward and the feedback control method is used to analyze the stability conditions. Conclusions are given in section 4.

2. Car-following model and its stability analysis

2.1. Full velocity different model

The dynamic equations of FVDM [23] are given by:

$$\begin{cases} \dfrac{d^2 x_n(t)}{dt^2} = a\left[V^{OP}(y_n(t)) - v_n(t)\right] + \lambda \Delta v_n(t), \\ \dfrac{dy_n(t)}{dt} = v_{n+1}(t) - v_n(t), \end{cases} \tag{1}$$

where $a = 1/\tau$ is the sensitivity of a driver, $y_n(t) = x_{n+1}(t) - x_n(t)$ and $\Delta v_n(t) = v_{n+1}(t) - v_n(t)$ are the headway and the velocity difference between the n-th considering vehicle and the preceding one, and $V^{OP}(y_n(t))$ is the optimal velocity function, which is written as follows:

$$V^{OP}(y_n(t)) = \frac{v_{\max}}{2}[\tanh(\Delta x_n(t) - h_c) + \tanh(h_c)], \tag{2}$$

where h_c is the safety headway distance.

2.2. Stability analysis

We assume that the leading vehicle runs constantly at speed v_0, so the steady state of the following vehicles are

$$(v, y) = (v^*, y^*). \tag{3}$$

Then, consider an error system around steady state (1), that is,

$$\begin{cases} \dfrac{dv_n^o(t)}{dt} = a\left[\Lambda y_n^o(t) - v_n^o(t)\right] + \lambda \Delta v_n^o(t), \\ \dfrac{\mathrm{d}y_n^o(t)}{dt} = v_{n+1}^o(t) - v_n^o(t), \end{cases} \tag{4}$$

where $\Lambda = \left.\dfrac{\partial V^{OP}(y_n^o(t))}{\partial y_n(t)}\right|_{y_n(t)=y_n^o(t)}$, $\Delta v_n^o(t) = v_{n+1}^o(t) - v_n^o(t)$, $v_n^o(t) = v_n(t) - v^*$,

$y_n^o(t) = y_n(t) - y^*$.

The Laplace transformation for Eq. (4) leads to

$$\begin{cases} \mathrm{sV_n}(s) - V_n(0) = a\left[\Lambda Y_n(s) - V_n(s)\right] + \lambda \Delta V_n(s), \\ sY_n(s) - Y_n(0) = V_{n+1}(s) - V_n(s), \end{cases} \tag{5}$$

where $\mathrm{V_n}(s) = L(\delta v_n(t))$, $\mathrm{Y_n}(s) = L(\delta y_n(t))$, $\mathrm{L}(\cdot)$ denotes the Laplace transformation, s is a complex variable. Form Eq. (5), we have

$$\mathrm{V_n}(s) = \frac{a\Lambda + \lambda s}{s^2 + (a+\lambda)s + a\Lambda} V_{n+1}(s) + \frac{a\Lambda}{s^2 + (a+\lambda)s + a\Lambda} Y_n(0) + \frac{s}{s^2 + (a+\lambda)s + a\Lambda} V_n(0) \tag{6}$$

Let $p(s) = s^2 + (a+\lambda)s + a\Lambda$ and the transfer function can be obtained as

$$G(s) = \frac{a\Lambda + \lambda s}{s^2 + (a+\lambda)s + a\Lambda} \tag{7}$$

Based on stability theory, the traffic jam will never occur in the traffic flow system as long as the characteristic function $p(s) = s^2 + (a+\lambda)s + a\Lambda$ is stable and $\|G(s)\| \le 1$.

In order to make $p(s)$ stable, that $a > 0$ and $a\Lambda > 0$ should be confirm. According to the Hurwitz stability criterion, the OV function is monotonic increase (i.e. $\Lambda > 0$) and $a > 0$, we obtain that $p(s)$ is stable.

Then, we consider $\|G(s)\| \le 1$ which can be expressed as

$$\left|G(j\omega)\right|^2 = \left|G(j\omega)G(-j\omega)\right| = \frac{(a\Lambda)^2 + (\lambda\omega)^2}{(a\Lambda - \omega^2)^2 + (a+\lambda)^2\omega^2} \le 1. \tag{8}$$

The sufficient condition can be obtained as

$$\omega^4 + a^2\omega^2 - 2a\Lambda\omega^2 + 2a\lambda\omega^2 \ge 0, \omega \in [0, +\infty), \tag{9}$$

which can be rewritten as

$$\lambda \ge \Lambda - \frac{a}{2}. \tag{10}$$

If the condition $\lambda \ge \Lambda - \frac{a}{2}$ is satisfied, the traffic system will be stable.

3. Control scheme

The aim of this paper is to purpose a control scheme for suppression of congested traffic in the car-following model. A feedback control signal $u_n(t)$ is designated as follows:

$$u_n(t) = k\left(\frac{dv_n(t)}{dt} - \frac{dv_n(t-1)}{dt}\right) \tag{11}$$

where k is the feedback gain, which can be adjusted. The control signal term is added to Eq. (1) as

$$\begin{cases} \dfrac{d^2x_n(t)}{dt^2} = a\left[V^{OP}(y_n(t)) - v_n(t)\right] + \lambda\Delta v_n(t) + u_n(t), \\ \dfrac{dy_n(t)}{dt} = v_{n+1}(t) - v_n(t), \end{cases} \tag{12}$$

The control signal $u_n(t)$ is traffic jerk.

Similarly, we assumed that the leading vehicle runs with constant speed v_0, the steady state of the following vehicles are the same of Eq.(3). Then, consider an error system around steady state (12), that is

$$\begin{cases} \dfrac{dv_n^0(t)}{dt} = a\left[\Lambda y_n^0(t) - v_n^0(t)\right] + \lambda\Delta v_n^0 + u_n^0(t), \\ \dfrac{dy_n^0(t)}{dt} = v_{n+1}^0(t) - v_n^0(t), \end{cases} \tag{13}$$

where $\Lambda = \left.\dfrac{\partial V^{OP}(y_n^0(t))}{\partial y_n(t)}\right|_{y_n(t)=y_n^0(t)}$, $v_n^0(t) = v_n(t) - v^*$, $y_n^0(t) = y_n(t) - y^*$,

$$u_n^0(t) = k\left[\frac{dv_n^0(t)}{dt} - \frac{dv_n^0(t-1)}{dt}\right].$$

The Laplace transformation for Eq. (13) leads to

$$\begin{cases} sV_n(s) - V_n(0) = a\left[\Lambda Y_n(s) - V_n(s)\right] + \lambda\Delta V_n(s) + U_n(s) \\ sY_n(s) - Y_n(0) = V_{n+1}(s) - V_n(s) \end{cases} \tag{14}$$

where $U_n(s) = k\left[(sV_n(s) - V_n(0)) - e^{-s}(sV_n(s) - V_n(0))\right]$, $V_n(s) = L(\delta v_n(t))$,

$Y_n(s) = L(\delta y_n(t))$, $L(.)$ denotes the Laplace transformation, s is a complex variable. Form Eq. (14), we have

$$\begin{aligned} V_n(s) &= \frac{a\Lambda + \lambda s}{a\Lambda + (a+\lambda)s + s^2 - k(1-e^{-s})s^2}V_{n+1}(s) + \frac{a\Lambda}{a\Lambda + (a+\lambda)s + s^2 - k(1-e^{-s})s^2}Y_n(0) \\ &= \frac{\left[1-k(1-e^{-s})\right]s}{a\Lambda + (a+\lambda)s + s^2 - k(1-e^{-s})s^2}V_n(0) \end{aligned} \tag{15}$$

Let $1-e^{-s} = s$, substituting it into Eq. (15), which leads to

$$V_n(s) = \frac{a\Lambda + \lambda s}{a\Lambda + (a+\lambda)s + s^2 - ks^3}V_{n+1}(s) + \frac{a\Lambda}{a\Lambda + (a+\lambda)s + s^2 - ks^3}Y_n(0) + \frac{(1-ks)s}{a\Lambda + (a+\lambda)s + s^2 - ks^3}V_n(0) \tag{16}$$

Let $p^*(s) = a\Lambda + (a+\lambda)s + s^2 - ks^3$ and the transfer function can be obtained as

$$G^*(s) = \frac{a\Lambda + \lambda s}{a\Lambda + (a+\lambda)s + s^2 - ks^3} \tag{17}$$

Thus, traffic jams will never occur in the traffic flow system if $p(s)$ is stable and $\left\|G^*(S)\right\|_\infty \le 1$. Similarly to the second part of the analysis, the sufficient condition is given as

$$\left|G^*(j\omega)\right|^2 = \left|G^*(j\omega)G^*(-j\omega)\right| = \frac{(a\Lambda)^2 + (\lambda\omega)^2}{(a\Lambda - \omega^2)^2 + \left[(a+\lambda) + k\omega^2\right]^2 \omega^2} \le 1 \qquad (18)$$

Then, we can obtain the sufficient condition through the above analysis, that is

$$k^2\omega^4 + \left[2(a+\lambda)k + 1\right]\omega^2 + a(a + 2\lambda - 2\Lambda) \ge 0\,,\ \omega \in [0,+\infty) \qquad (19)$$

The sufficient condition for Eq. (19) is

$$\begin{cases} (2\lambda k + 1)^2 + 4ak(1 + 2k\Lambda) \le 0, if\ \dfrac{2(a+\lambda)k + 1}{2k^2} < 0 \\ a(a + 2\lambda - 2k) \ge 0, if\ \dfrac{2(a+\lambda)k + 1}{2k^2} > 0 \end{cases} \qquad (20)$$

4. Summary

In this paper, a new feedback control signal 'jerk' is added to FVDM. The stability condition of developed model is analyzed by using feedback control theory. Through theoretical analysis, the range of reaction parameter λ for the model with and without feedback control signal obtained.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grant No. 11372166]; the Scientific Research Fund of Zhejiang Provincial, China [Grant Nos. LY15A020007, LY15E080013]; the Natural Science Foundation of Ningbo [Grant Nos. 2014A610028, 2014A610022]; the project Hong Kong RGC TBRS Project: T32-101/15-R; and the K.C. Wong Magna Fund in Ningbo University, China.

References

[1] Pipes, L. A. (1953) an operational analysis of traffic dynamics [J], J. *APPl. Phy* **24**, 274-281.
[2] Newell G.F. (1961) Nonlinear Effects In The Dynamics Of Car Following [J], *Oper. Res* **9**, 209-229.
[3] Bando M. et al. (1995) Dynamical model of traffic congestion and numerical simulation [J], *Phys. Rev. E* **51**, 1035.
[4] Ge, H.X., Cui, Y., Zhu, K.Q., Cheng, R.J. (2015) The control method for the lattice hydrodynamic model. *Commun Nonlinear Sci Numer Simulat* **22**, 903–908.
[5] Li, Z.P., Liu, F.Q. (2006) An Improved Car-Following Model for Multiphase Vehicular Traffic Flow and Numerical Tests. *Communications in Theoretical Physics A* **46**, 367-373.
[6] Tang, T.Q., Huang, H.J., Gao, Z.Y. (2005) Stability of the car-following model on two lanes. *Phys. Rev. E* **72**, 066124.
[7] Peng, G.H., Cai, X.H., Liu, C.Q., Cao, M.X., Tuo, M.X. (2011) Optimal velocity difference model for a car-following theory. *Physics Letters A* **375**, 3973-3977.
[8] Peng, G.H. (2013) A new lattice model of two-lane traffic flow with the consideration of optimal current difference. *Communication Nonlinear Science Numerical Simulation* **18**, 559-566.

[9] Li, Z.P., Li, X.L., Liu, F.Q. (2008) Stabilization analysis and modified KdV equation of lattice models with consideration of relative current. *International Journal of Modern Physics C* **19**, 1163-1173.
[10] Tang, T.Q., Huang, H.J., Wong, S.C., Xu, X.Y. (2007) A new overtaking model and numerical tests. *Physical A* **376**, 649-657.
[11] Peng GH, Cai XH, Liu CQ, Cao MX, Tuo MX. (2011) Optimal velocity difference model for a car-following theory. *Phys Lett A* **375**, 3973–7.
[12] E De Angelis, (1999) Nonlinear hydrodynamic models of traffic flow modelling and mathematical problems. *Mathematical and Computer Modelling* **7**, 83-95.
[13] Sapna Sharma, (2015) Lattice hydrodynamic modeling of two-lane traffic flow with timid and aggressive driving behavior. *Physica A* **421**, 401-411.
[14] Biham O., A. Middleton & D.Levine (1995) Self-organization & a dynamic transition in traffic flow models [J], *Phys. Rev. E* **51(1)**, 772-774.
[15] Gu G.Q., Zhong, J.X., Xu, B.M.(1995) Two-dimensional traffic flow problems in inhomogeneous lattice [J], *Physca A* **217(3-4)**, 339-347.
[16] Konishi K, Kokame H and Hirata K. (1999) Coupled map car-following model and its delayed-feedback control. *Physical Review E* **60**, 4000-4007.
[17] Zhao, X.M., Gao, Z.Y. (2006) A control method for congested traffic induced by bottlenecks in the coupled map car-following model. *Physics A* 366, 513-522.
[18] Han, X.L., Jiang, C.Y., Ge, H.X., et.al. (2007) A modified coupled map car-following model based on application of intelligent transportation system and control of traffic congestion. *Acta Physica Sinica* **56**, 4383-4392.
[19] Ge, H.X., Yu, J., Lo, S.M. (2012) A control method for congested traffic in the car-following model. *Chin.Phys. lett* **29**, 050502.
[20] Ge, H.X., Meng, X.P., Zhu, H.B., Li, Z.P. (2014) Feedback control for car following model based on two-lane traffic flow. *Physica A* **408**, 28-39.
[21] Yu, X.W., Shi, Z.K.: An improved car-following model considering headway changes with memory. *Physica A*, 2015, 421: 1-14.
[22] Zheng, Y.Z., Zheng, P.J., Ge, H.X. (2014) An improved car-following model with consideration of the lateral effect and its feedback control research. *Chin.Phys.B* **23**, 020503.
[23] Jiang, R, Wu, Q.S., Zhu, Z.J. (2001) Full Velocity different model for car-following theory. *Phys. Rev. E* **64**, 017101-017105 .
[24] Yu, S.W., Liu, Q.L., Li, X.H. (2013) Full velocity difference and acceleration model for a car-following theory. *Commun Nonlinear Sci Numer Simulat* **18**, 1229–1234.

The effect of stray grains on the mechanical behavior of nickel-based single crystal superalloy

H.B. Tang, †*H.D. Guo, X.G. Liu, S.H. Yang, L. Huang

Jiangsu Province Key Laboratory of Aerospace Power System, Collaborative Innovation Center for Advanced Aero-Engine, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

*Presenting author: ghd@nuaa.edu.cn
†Corresponding author: ghd@nuaa.edu.cn

Abstract

In this paper，a new bicrystal model, consists of primary and stray grains, is proposed to simulate the weakening effect of stray strains generated at geometric discontinuities of single crystal (SC) superalloy. A constitutive model considered crystallographic orientations is introduced, and then the bicrystal model under uniaxial loading is built and analyzed in commercial finite element software ABAQUS. The numerical simulation results indicate that yield strength and elastic modulus of stray grains, which can be determined by the crystallographic orientation, have a significant effect on the deformation of the bicrystal model. To evaluate the local stress rise at the sub-boundary of primary and stray grains, a critical stress based on the yield criterion of SC material is proposed. In the elastic stage, as the elastic modulus difference between primary and stray grains increases, the local stress rise would be more severe. In the elastic-plastic stage II, while the yield strength of primary grains is greater than that of stray grains, the lower the yield strength of stray grains is, the smaller load the bicrystal structure can sustain. Finally an evolution equation of critical stress is constructed with consideration of stray grains under uniaxial loading conditions.

Keywords: Nickel-based single crystal (SC) superalloy, Stray grains, Local stress rise, Critical stress.

1. Introduction

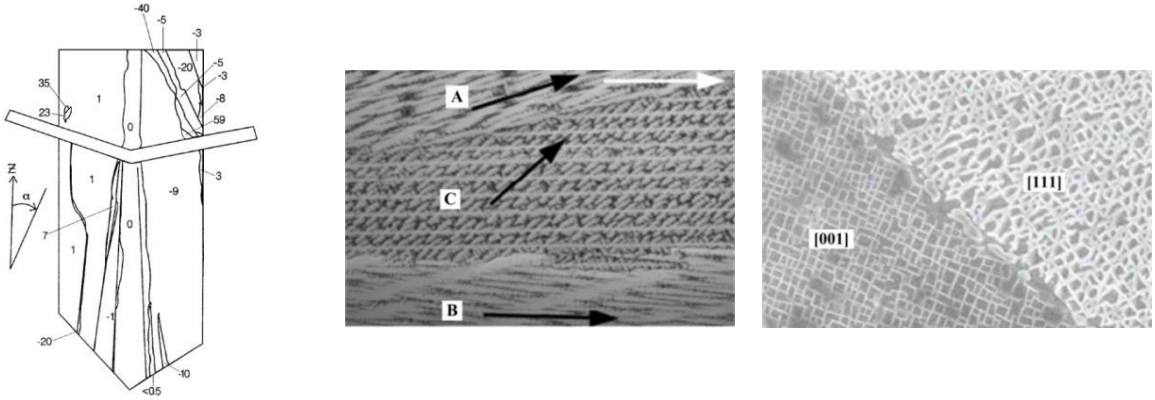
Compared with polycrystalline materials, nickel-based single crystal (SC) superalloy has better mechanical properties at elevated temperature in the absence of weak traditional grain boundaries. During the manufacture of complex structures such as turbine blades, stray grains can be generated in the SC superalloy casting by directional solidification [1-5]. It has been found that thermal condition and mold geometry have a significant impact on the formation of stray grains [5-6]. The disordered temperature distribution at geometric discontinuities, e.g. blade shrouds, turbine blade platforms and turbine blade rabbets, can lead to distortions in the crystal lattice [7]. Usually, stray grains are observed in critical areas with complex stress state, and the mechanical and fatigue characteristics of SC materials can be greatly influenced by stray grains, so the effect of stray grains on SC complex structures should be considered. The basic material properties of SC containing stray grains have been experimentally studied [7, 10-13]. However, there is still a lack of numerical modeling and theoretical analysis of the stray grains effect on SC Materials. In this paper, a SC partition model is built, and the effect of stray grains on the mechanical behaviors of SC materials is further investigated by the bicrystal model through finite element analysis.

2. The SC model considered stray grains

2.1. *The SC partition model*

X-ray topography of SC material Rene N5 containing stray grains were performed by Napolitano et al. [6], and major grain defects were categorized as low-angle boundaries, high-angle boundaries, and spurious grains, shown in Fig. 1(a). The crystal morphology of SC material AM3 containing stray grains was investigated by Zhao et al. [10]. The results showed

that stray grains can be divided into several groups by crystallographic orientation, and stray grains in each group were approximately along the same direction. Besides, high-angle boundaries were observed, shown in Fig. 1(b). The casting microstructure of SC material DD6 containing stray grains was studied by Shi et al. [14], and the stray grains along [111] were observed, shown in Fig. 1(c). On the basis of X-ray topography and predictions, the crystallographic orientation and locations of stray grains can be obtained, and then the stray grains can be divided into several partitions. Given a single partition of stray grains, for the case with grains orientated in almost the same direction, it can be still modeled as SC material, and for the case with spurious grains, isotropic model can be used instead.



(a) Rene N5 **(b) AM3** **(c) DD6**

Figure 1. SC materials containing stray grains

2.2. *The bicrystal model*

The SC structures containing stray grains will be simplified in the following discussion. Given the crystallographic orientations of primary and stray grains, a bicrystal model containing one group of stray grains is proposed, as detailed in Fig. 2. In Fig.2, the left and right part of the bicrystal model represents the primary and stray grains, respectively. The bicrystal model is fixed at one end and applied with a concentrated load at other end horizontally. β is the angle between the interface of sub-boundary and the orientation of primary grains, and a wide range of β can be expected in the real SC structures.

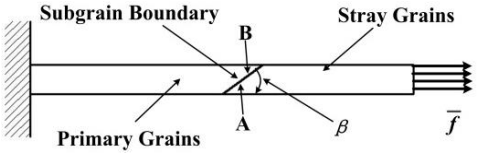


Figure 2. A bicrystal model

3. Material Property Prediction for the bicrystal model

The tensile tests of SC material DD3 along [001], [011] and [111] at 680°C were conducted by Ding [15], and basic material properties are summarized in Table 1. The elastic-plastic parameters of DD3 (ID is QX#.) along each direction, as shown in Table 2, is calculated from the constitutive model given in Appendix A. Since the angle α between crystallographic orientations of primary and stray grains, which is different from β, cannot clearly distinguish the stray grains in different crystallographic orientations, the actual orientation [hkl] is introduced.

Table 1. The tensile test results of DD3 (680 ℃) along [001], [011] and [111]

Temperature (°C)	G (GPa)	μ	$E_{[001]}$ (GPa)
680	113	0.322	129.7
$\phi'_{[001]}$ (GPa)	$\sigma_{y[001]}$ (MPa)	$\sigma_{y[011]}$ (MPa)	$\sigma_{y[111]}$ (MPa)
1.328	943	896	1085

where $\sigma_{y[001]}$, $\sigma_{y[011]}$ and $\sigma_{y[111]}$ are the yield strengths of DD3 along [001], [011] and [111], respectively. $\phi'_{[001]}$ is the plastic modulus of DD3 along [001].

Table 2. Basic material properties of DD3 in selected directions (680 ℃)

ID	Orientation [*hkl*]			α (°)	σ_{ym} (MPa)	E (GPa)	ϕ' (GPa)
	h	k	l				
QX1	0	0.105	1	6	935.50	131.83	1.308
QX2	0	0.177	1	10	924.50	135.65	1.278
QX3	0	0.268	1	15	909.50	143.08	1.237
QX4	0.189	0.189	1	15	911.81	143.35	1.244
QX5	0	0.577	1	30	888.31	180.34	1.183
QX6	0	1	1	45	895.8	207.21	1.204
QX7	0.707	0.707	1	45	1031.18	243.6	1.594
QX8	1	1	1	54.74	1084.8	258.75	1.764

where σ_{ym}, E and ϕ' are the yield strength, elastic modulus and plastic modulus of DD3, respectively.

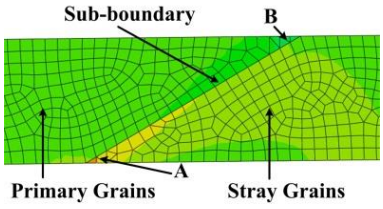
4. Influence analysis based on the bicrystal model

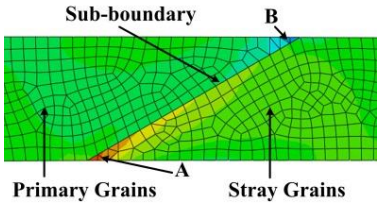
4.1. *The FE model*

Based on the DD3 bicrystal model as illustrated in Fig. 2, the stress distribution at critical region will be discussed in this section to analyze the effect of stray grains on the mechanical properties of SC material. An 3D model is built in ABAQUS 6.13, and the dimension of the model is $60mm \times 10mm \times 4mm$, and $\beta = 30°$. The primary grains are along [001], and the stray grains are labeled as from QX1 to QX8 for different orientations, as detailed in Table 2. Both primary and stray grains are under axial tension, thus the constitutive relationship of DD3 along corresponding directions should be used. The non-linear large deflection algorithm is enabled through the finite element analysis to accurately capture the local plastic deformation.

4.2. *The critical stress in the bicrystal model*

According to the finite element analysis results, non-uniform stress distribution and the critical region (A or B) can be always observed under axial tensile load. The stress contours obtained from two typical configurations under two different loads are illustrated from Fig. 3(a) to Fig. 3(d).





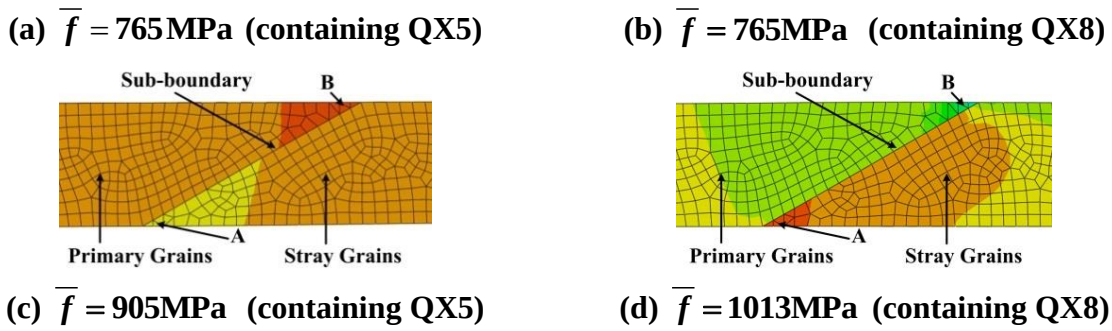
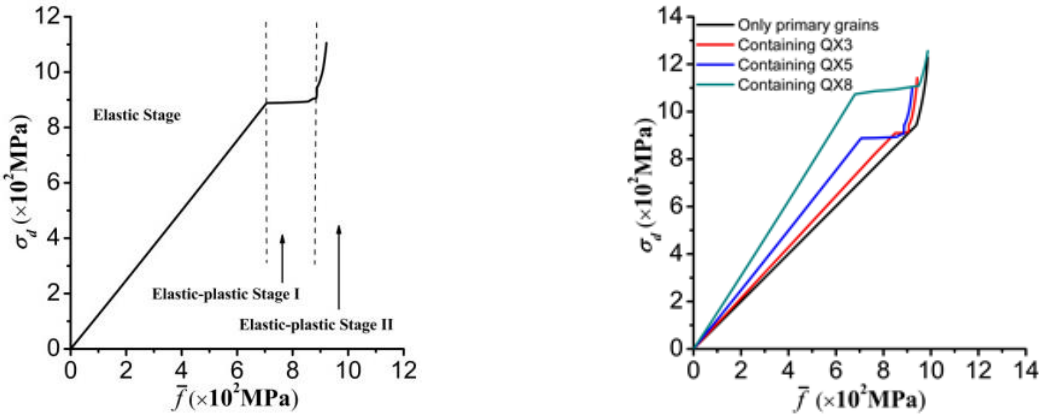


Figure 3. Stress contour of the DD3 bicrystal model containing stray grains

where $\overline{f}$ is the applied load.

To evaluate the local stress rise near the sub-boundary, a SC critical stress σ_d is proposed, which can be calculated by Eq. (B.1) from Appendix B. Fig. 4(a) presents the relationship between critical stress σ_d and applied load $\overline{f}$ (nominal surface force) of the DD3 bicrystal model containing stray grains QX5; Fig. 4(b) shows the relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model containing different stray grains (partly). According to Fig. 4(a)-(b), the whole loading process of the model can be divided into three stages:
(1) Elastic stage
(2) Elastic-plastic stage I
(3) Elastic-plastic stage II



(a) containing QX5 **(b) containing different stray grains**

Figure 4. Relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model

In the elastic stage, the variation of σ_d depends on the elastic modulus of primary grains E_0 and the elastic modulus of stray grains E_{m}. Local high stress is observed obviously near region A, as shown in Fig. 3(a)-(b). As the load increases, the material at region A will yield firstly. In the elastic-plastic stage I, the local stress rise location is still within region A. With the increase of load, the local stress rise tends to less distinct and more stray grains reach the initial yield stress. It could be found that, in the elastic-plastic stage II, the variation of σ_d is mainly determined by the yield strength of primary grains σ_{y0} and the yield strength of stray grains σ_{ym}. While σ_{ym} is smaller than σ_{y0} (e.g. QX5), the critical region will be transferred from A to B, as shown in Fig. 3(c), and with the increase of load, more primary grains reach

the initial yield stress; while σ_{ym} is larger than σ_{y0} (e.g. QX8), Region A will always be the most critical location, as shown in Fig. 3(d). By the end of Elastic-plastic stage II, the maximum equivalent stress of the grains with lower yield strength will increase rapidly due to necking, while high stress location around sub-boundary is still a critical region in consideration of the fragility of sub-boundary.

In the elastic stage, as the elastic modulus difference between primary and stray grains increases, the local stress concentration would be more severe. When the grain defect is either a low-angle boundary ($\alpha \leq 15°$) or a high-angle boundary ($\alpha > 15°$), the local stress rise will be unremarkable or significant, respectively. In the Elastic-plastic stage II, while σ_{ym} is smaller than σ_{y0}, the lower σ_{ym} is, the smaller load the bicrystal structure can sustain; while σ_{ym} is greater than σ_{y0}, the maximum load, which the bicrystal structure can sustain, is nearly the same.

4.3. *Evolution equation of the critical stress*

The critical stress observed near the sub-boundary of primary and stray strains has been discussed in previous section, and the evolution equation of the critical stress will be built with considerations of the effect of stray strains.

4.3.1. *Elastic stage*

There is a significant linear correlation between σ_d and $\overline{f}$, as shown in Fig. 4(b). As the slope of the linear relationship k_1 depends on E_0 and E_m, $k_1 = 4.57\times10^{-6}(E_m - E_0)+1$ can be calculated by regression analysis. Thus, the evolution equation of σ_d in the elastic stage can be given as,

$$\sigma_d = [4.57\times10^{-6}(E_m - E_0)+1]\overline{f} \tag{1}$$

Since $\sigma_d \leq \min\{\sigma_{y0}, \sigma_{ym}\}$, the range of load can be obtained in Eq. (2).

$$0 \leq \overline{f} \leq \frac{\min\{\sigma_{y0}, \sigma_{ym}\}}{4.57\times10^{-6}(E_m - E_0)+1} \tag{2}$$

4.3.2. *Elastic-plastic stage I*

The range of load can be expressed as,

$$\frac{\min\{\sigma_{y0}, \sigma_{ym}\}}{4.57\times10^{-6}(E_m - E_0)+1} < \overline{f} < \min\{\sigma_{y0}, \sigma_{ym}\} \tag{3}$$

With increase of applied load, σ_d is nearly the same, as detailed in Fig. 4(b). Thus, the evolution equation of σ_d in this stage has the following form,

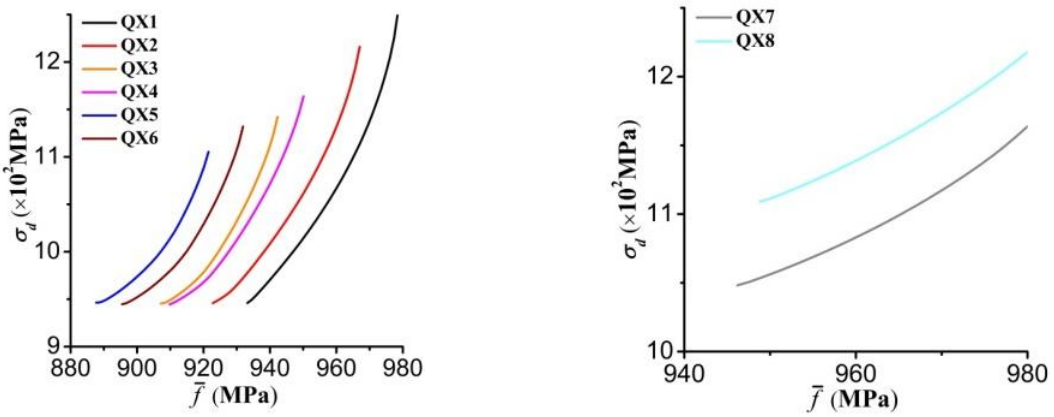
$$\sigma_d = \sigma_{ym} \tag{4}$$

4.3.3. *Elastic-plastic stage II*

The range of load can be written as,

$$\overline{f} > \min(\sigma_{y0}, \sigma_{ym}) \tag{5}$$

In the DD3 bicrystal model, the difference in the magnitudes of E_0 and E_m will result in a different critical region, and σ_d will change in a different way, too. Fig. 5(a)-(b) present the relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model containing different stray grains in the elastic-plastic stage II.



（a） $\sigma_{ym} < \sigma_{y0}$ （b） $\sigma_{ym} > \sigma_{y0}$

Figure 5. Relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model containing different stray grains in the elastic-plastic stage II

While $\sigma_{ym} < \sigma_{y0}$, the evolution equation of σ_d in this stage has the following form:

$$\sigma_d = 0.9(\overline{f} - \sigma_{ym})^{1.49} + \sigma_{y0} \tag{6}$$

While $\sigma_{ym} > \sigma_{y0}$, the evolution equation of σ_d in this stage has the following expression:

$$\sigma_d = 1.55(\overline{f} - \sigma_{y0})^{1.23} + \sigma_{ym} \tag{7}$$

Finally, the evolution equation of σ_d is summarized by Eq. (8).

$$\sigma_d = \begin{cases} [4.57\times10^{-6}(E_0 - E_m)+1]\overline{f}, & 0 \le \overline{f} \le \dfrac{\min\{\sigma_{y0}, \sigma_{ym}\}}{4.57\times10^{-6}(E_0 - E_m)+1} \\ \sigma_{ym}, & \dfrac{\min\{\sigma_{y0}, \sigma_{ym}\}}{4.57\times10^{-6}(E_m - E_0)+1} < \overline{f} < \min\{\sigma_{y0}, \sigma_{ym}\} \\ 0.9(\overline{f} - \sigma_{ym})^{1.49} + \sigma_{y0}, & \overline{f} > \min\{\sigma_{y0}, \sigma_{ym}\} \text{ and } \sigma_{ym} < \sigma_{y0} \\ 1.55(\overline{f} - \sigma_{y0})^{1.23} + \sigma_{ym}, & \overline{f} > \min\{\sigma_{y0}, \sigma_{ym}\} \text{ and } \sigma_{ym} > \sigma_{y0} \end{cases} \tag{8}$$

When $\sigma_{ym} < \sigma_{y0}$ and $\overline{f}$ is nearby σ_{ym}, the critical region will be transferred and σ_d will

increase dramatically, which is not included in Eq. (8).

4.4. *Influence analysis of the angle between sub-boundary and orientation of primary grains*

The effect of β on SC material containing stray grains is also analyzed by the model shown in Fig 2. The primary grains are along [001], and the stray grains are QX5. β equals to $20°$, $30°$ and $45°$, respectively. Fig. 6 presents the relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model with different β. The result shows that the local stress rise will be more distinct in elastic stage and the loading process will be longer, with decrease of β. However, β has little influence on mechanical behavior of SC material containing stray grains in the elastic-plastic stage II.

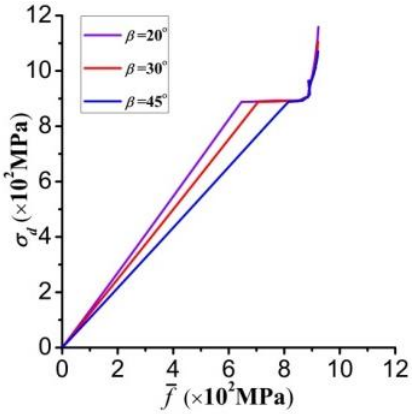


Figure 6. Relationship between σ_d and $\overline{f}$ of the DD3 bicrystal model with different β

5. Discussions

1. The proposed SC partition model can be used to simulate the SC materials containing several groups of stray grains. The local high stress can be found near the sub-boundary of primary and stray grains. The local stress distribution and critical stress will also be influenced by the geometry of SC structure.
2. As the applied load increases, the local high stress region is always observed near the sub-boundary. Given the fragility of sub-boundary, the effect of stray grains should be considered in the analysis of the mechanical behavior and fatigue characteristics of SC complex structures.

6. Conclusions

In this paper，a new bicrystal model, consists of primary and stray grains, is proposed to simulate the weakening effect of stray strains generated at geometric discontinuities of SC material. A constitutive model considered crystallographic orientations is introduced, and then the bicrystal model under uniaxial loading is built and analyzed. The numerical simulation results indicate that yield strength and elastic modulus of stray grains, which can be determined by the crystallographic orientation, have a significant effect on the deformation of the bicrystal model. To evaluate the local stress rise at the sub-boundary of primary and stray grains, a critical stress based on the yield criterion of single crystal material is proposed. In the elastic stage, as the elastic modulus difference between primary and stray grains increases, the local stress rise would be more severe. In the elastic-plastic stage II, while the yield strength of primary grains is greater than that of stray grains, the lower the yield strength of stray grains is, the smaller load the bicrystal structure can sustain. Hence, the effect of stray

grains on the mechanical and fatigue characteristics of SC complex structures should not be neglected. Finally an evolution equation of critical stress is constructed with consideration of stray grains under uniaxial loading conditions.

Appendix A. Constitutive model of SC superalloy

T-G criterion [16] can be used to describe the yield behavior of SC superalloy:

$$Y^2 = [(P_1 + P_2)^2 + P_6]^{\frac{1}{2}} \tag{A.1}$$

$$P_1 = \frac{A_1}{6}(S_{ii} - S_{jj})^2; P_2 = B_1 S_{ij}^2; P_6 = C_1[(S_{kk} - S_{jj})^2 + (S_{kk} - S_{ii})^2]S_{ij}^2$$

Based on Drucker postulation and associated flow rule [17], T-G criterion can be used as plastic potential. The constitutive model of SC superalloy can be constructed by isotropic hardening model. As $\bar{\sigma} = Y$, plastic potential has the following form:

$$U = \bar{\sigma}^2 \tag{A.2}$$

Since isotropic hardening model is adopted, hardening parameter is given as,

$$\mathrm{K} = \bar{\varepsilon}_p \tag{A.3}$$

Hence, the elasto-plastic matrix can be finally derived as,

$$[C]_{ep} = [C]_e - \frac{[C]_e \left\{\frac{\partial \bar{\sigma}}{\partial \sigma}\right\} \left\{\frac{\partial \bar{\sigma}}{\partial \sigma}\right\}^{\mathrm{T}} [C]_e}{\mathrm{H}' + \left\{\frac{\partial \bar{\sigma}}{\partial \sigma}\right\}^{\mathrm{T}} [C]_e \left\{\frac{\partial \bar{\sigma}}{\partial \sigma}\right\}} \tag{A.4}$$

Appendix B. Critical stress of SC superalloy

According to the yield criterion of SC superalloy presented in Appendix A, stress of the critical region can be constructed as,

$$\bar{\sigma} = [(P_1 + P_2)^2 + P_6]^{\frac{1}{4}} \tag{B.1}$$

Reference

[1] Zhang, X. L., Zhou, Y. Z., Jin, T. and Sun, X. F. (2012) Study on the tendency of stray grain formation of Ni-based single crystal superalloys, *Acta Metallurgica Sinica* **48**, 1229-1236.

[2] Li, B., Zhu, S. Z., Li, X. and Pan, G. Z. (2012) Research on gating system of a directional solidified blade, *Foundry* **61**, 81-83.

[3] Xuan, W. D., Ren, Z. M., Ren, W. L., Yu, J. B. and Chen, C. (2011) Effect of seed crystal orientation on the stray grain of directional solidified Ni-based superalloy, *Journal of Iron and Steel Research* **23**, 368-372.

[4] D'Souza, N., Jennings, P. A., Yang, X. L., Dong, H. B., Lee, P. D. and Mclean, M. (2005b) Seeding of single-crystal superalloys role of constitutional undercooling and primary dendrite orientation on stray-grain nucleation and growth, *Metallurgical and materials transactions* **36**, 657-666.

[5] Yang, X. L., Dong, H. B., Wang, W. and Lee, P. D. (2004a) Microscale simulation of stray grain formation in investment cast turbine blades, *Materials Science and Engineering* **396**,129-139.
[6] Napolitano, R.E., Schaefer, R. J. (2000) The convergence-fault mechanism for low-angle boundary formation in single-crystal castings, *Journal of Materials Science* **35**, 1641-1659.
[7] Yan, X. J., Deng, Y., Sun, R. J. and Xie, J. W. (2011) Study of fatigue property variation at different regions on a DS turbine blade, *Acta Aeronautica et Astronautica Sinica* **32**, 1930-1936.
[8] A. de Bussac, C. A. Gandin. (1997a) Prediction of a process window for the investment casting of dendritic single crystals, *Materials Science and Engineering* **237**, 35-42.
[9] Zhao, X. B., Liu, L., Zhang, W. G., Qu, M., Zhang, J. and Fu, H. Z. (2011) Analysis of competitive growth mechanism of stray grains of single crystal superalloys during directional solidification process, *Rare Metal Materials and Engineering* **40**, 9-13.
[10] Zhao, J. Q., Li, J. R., Liu, S. Z. and Yuan, H. L. (2008) Effects of low angle grain boundaries on tensile properties of single crystal superalloy DD6, *Journal of Materials Engineering* **8**, 73-76.
[11] Shi, Z. X., Li, J. R., Liu, S. Z. and Zhao, J. Q. (2009) Tensile properties of twist low angle boundary of DD 6 single crystal superalloy, *Journal of Aeronautical Materials* **6**, 88-92
[12] Shi, Z. X., Li, J. R., Liu, S. Z. and Zhao, J. Q. (2012) Effect of LAB on the stress rupture properties and fracture characteristic of DD6 single crystal superalloy, *Rare Metal Materials and Engineering* **41**, 962-966.
[13] Shi, Z. X., Liu, S. Z., Zhao, J. Q. and Li, J. R. (2015) Effect of low angle boundary on high cycle fatigue properties of single crystal superalloy, *Transactions of materials and heat treatment* **6**, 52-57.
[14] Shi, Z. X., Liu, S. Z., Li, J. R. (2015) Study on the stray grain microstructure of DD6 single crystal superalloy, *Foundry* **64**, 153-156.
[15] Ding, Z. P. (2005) *Study on multiaxial low cycle fatigue damage of single crystal nickel-based superalloy*, Central South University.
[16] Tang, H. B. (2014) *Effect of geometric discontinuities on the fatigue behavior of single crystal superalloy,* Nanjing University of Aeronautics and Astronautics.
[17] Shi, F. (2012) *Re-development of ANSYS and application examples*, China Water & Power Press, Beijing.

A Numerical Study of Compressible Two-Phase Flows Shock and Expansion Tube Problems

Dia Zeidan[1,a)] and Eric Goncalves[2]

[1]School of Basic Sciences and Humanities, German Jordanian University, Amman, Jordan

[2]Ensma - Pprime, Department of Aeronautical Engineering, Poitiers, France

[a)]Corresponding and presenting author: dia.zeidan@gju.edu.jo

ABSTRACT

A compressible and multiphase flows solver has been developed for the study of one-dimensional shock and expansion tube problems. This solver has a structure similar to those of the one-fluid Euler solvers, differing from them by the presence of a void ratio transport-equation. The model and the system of equations to be simulated are presented. Results are displayed for shock and expansion tube problems. Close agreement with reference solutions, obtained from explicit finite volume approaches, is demonstrated for all of the examples. Different numerical methods are additionally displayed to provide comparable and improved computational efficiency to the model and the system of equations. The overall procedure is therefore very well suited for use in general two-phase fluid flow simulations.

Keywords: Two-phase flows, shock and expansion tube problems, homogeneous model, Riemann problem, finite volume, inviscid simulation

Introduction

Theoretical and numerical modeling of two-phase fluid flow problems is of practical importance in many areas of industry such as thermal power generation plants and other interesting phenomena occurring in environmental applications. Despite their relevance in industrial and environmental applications, compressible two-phase flow investigations have remained complex and challenging areas of applied mathematics and computational methods. The most widely used modeling approach is based on averaged two-phase fluid flow model such as the one-fluid formulation. Within such averaged model, there are different approaches according to the physical assumptions of interest made on the local mechanical and thermodynamical equilibrium and to the slip condition between phases. This has resulted in the development of diverse models and system of equations ranging from seven to three equations only. There also have been a number of significant contributions in different areas and applications relevant to two-phase flows. These are very well acknowledged in the scientific literature for which we refer the reader to [2, 5, 9, 19] for further details.

A critical aspect for two-phase simulations concerns the numerical methods of interest and their accuracy problems. The hyperbolic nature of such flows and their characteristic analysis makes the simulation very stiff and challenging. In addition to that, the volume fraction variation across acoustic waves causes difficulties for the Riemann problem resolution particularly in the derivation of approximate Riemann solvers. This is due to the occurrence of the large discontinuities of thermodynamic variables and equations of state involved at material interfaces. As a result, numerical instabilities and spurious oscillations appear through the complete wave structure [1]. The reason for such unusual behavior lies in the numerical dissipation of the methods which reproduce a thermodynamic path that is not correct. This also implies computational failure for Godunov methods which is due to the large decrease of the pressure up to vacuum ghost.

In the present paper, modeling and computer simulations are performed on the basis of Navier-Stokes applications. A four-equation model of the two-fluid model type is considered for the current purpose. The set of equations includes three conservation laws for mixture quantities along with a void ratio transport-equation [6, 7]. This set of equations is solved by means of explicit finite volume techniques based on Jameson, Rusanov, AUSM-type, VF Roe and HLLC Riemann solvers methods. This is followed by computational simulations on one-dimensional inviscid problems to study the behavior of the performed numerical methods. Computational results are then displayed for shock tube and rarefaction problems, including problems of large depression. These test cases establish the ability, accuracy and efficiency of our computational treatment.

Models and Methods

The homogeneous mixture approach is used to model two-phase flows. In addition, the phases are assumed to be in thermal and mechanical equilibrium, that is, both phases share the same temperature T and the same pressure P. The evolution of the two-phase flow can be described by the conservation laws that employ the representative flow properties as unknowns just as in a single-phase problem.

A four-equation model

We consider a reduction form of the five-equation Kapila model [9] under thermal equilibrium between phases. We also assume that the liquid phase is in a saturation state. The model consists of three conservation laws for mixture quantities and an additional equation for the void ratio. The governing equations under consideration are then governed by the following set of partial differential equations:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0 \tag{1}$$

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial (\rho u^2 + P)}{\partial x} = 0 \tag{2}$$

$$\frac{\partial (\rho E)}{\partial t} + \frac{\partial (\rho u H)}{\partial x} = 0 \tag{3}$$

$$\frac{\partial \alpha}{\partial t} + u\frac{\partial \alpha}{\partial x} = \underbrace{\left(\frac{\rho_l c_l^2 - \rho_v c_v^2}{\frac{\rho_l c_l^2}{1-\alpha} + \frac{\rho_v c_v^2}{\alpha}} \right)}_{=K} \frac{\partial u}{\partial x} \tag{4}$$

The individual variables are ρ mixture density, u velocity, P pressure, α void fraction, E and H are total energy and enthalpy of the two-phase flow. The source term K involves the speed of sound, c_k, and densities, ρ_k, of pure phases, $k = l, v$. The subscripts v and l indicate the vapor and the liquid phase, respectively. The four equations model form a system of conservation laws having a hyperbolic nature. The eigenvalues of the system are found to be:

$$\lambda_1 = u - c_{wallis}, \quad \lambda_2 = u = \lambda_3, \quad \lambda_4 = u + c_{wallis} \tag{5}$$

where c_{wallis} is the the propagation of acoustic waves without mass and heat transfer [17]. This speed of sound is expressed as a weighted harmonic mean of speeds of sound of each phase:

$$\frac{1}{\rho c_{wallis}^2} = \frac{\alpha}{\rho_v c_v^2} + \frac{1-\alpha}{\rho_l c_l^2} \tag{6}$$

Equation of state

To close the system, an equation of state (EOS) is necessary to link the pressure and the temperature to the internal energy and density. For the pure phases, we have employed the convex stiffened gas EOS. An expression for the pressure and the temperature can be deduced from the thermal and mechanical equilibrium assumption (see [13], and references therein, for details).

Numerical methods

In this short paper, the finite volume techniques are performed on the basis of the Riemann problem. In one-dimensional space, the conservative part of the four-equation model can be represented in a matrix form as:

$$\frac{\partial W}{\partial t} + \frac{\partial F(W)}{\partial x} = 0 \tag{7}$$

$$\frac{\partial \alpha}{\partial t} + u\frac{\partial \alpha}{\partial x} = S(W) \tag{8}$$

Here W is the vector of conserved variables, F and S are the convective flux and the source term that includes the void ratio equation given in (4). These vectors are defined by

$$W = \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} \rho u \\ \rho u^2 + P \\ \rho u H \end{pmatrix}$$

Based on finite volume techniques, the computational cells involve the discretization of the spatial domain x into regular meshes of length Δx and the temporal domain t into intervals of duration Δt. A discrete form of equations (7) and (8) can be written as:

$$\Delta x \frac{W_i^{n+1} - W_i^n}{\Delta t} + F_{i+1/2}^n - F_{i-1/2}^n = S_i^n \Delta x \tag{9}$$

where the time step should fulfill the CFL condition in order to guarantee stability requirement and $F_{i+1/2}^n$ is the numerical flux through the cell interface. This numerical flux can be computed using the solution of the Riemann problem or any other numerical method of interest where the resolution of the Riemann problem is fully numerical.
Various formulations of numerical flux have been proposed to solve multiphase compressible flows. See for instance [18] or [14], and references therein, for such formulations and extensions. In the present study, we have tested and compared five documented formulations, namely, the Jameson-Schmidt-Turkel scheme [8], an AUSM-type scheme [4], the Rusanov scheme [12], the HLLC scheme [16] and a VF Roe non-conservative scheme [3].

Computational Results on One-Dimensional Two-Phase Flow Problems

In this section we exhibit the ability of the current four-equation model, convergence and computational performance of the proposed numerical methods on two groups of two-phase flow problems. In the first group, we considered two shock tube problems to validate the current numerical tool. A comparison with solutions provided with a seven-equation model using the Discrete Equations Method (DEM) is proposed [15]. In DEM approach, the pure fluids are first integrated at the microscopic level and then the discrete formulae are averaged. The obtained continuous model of multiphase flow is equivalent to the Baer-Nunziato model. The infinite rate relaxation procedures are used to correctly treat the full system. The second group tests the expansion tube, double rarefaction, problems which are very stiff cases for numerical methods. Results of the expansion tube problems are validated with other models as we shall see later.

Water-gas mixture shock tube

This test case is proposed in [10], computed with five- and seven-equation models. A one meter shock tube involves a discontinuity of the volume fraction. For $x < 0.7$ the gas volume fraction is 0.2, while it is 0.8 otherwise. The fluids are governed by the stiffened gas EOS and are initially at rest. The left chamber contains high pressure fluids (10^9 Pa) while the right one contains low pressure fluids (10^5 Pa). The parameters of EOS are:

$$\begin{pmatrix} \gamma \\ P_\infty \\ \rho \end{pmatrix}_{Liq} = \begin{pmatrix} 4.4 \\ 6.\,10^8 \\ 1000 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \gamma \\ P_\infty \\ \rho \end{pmatrix}_{Gas} = \begin{pmatrix} 1.4 \\ 0 \\ 1 \end{pmatrix}$$

Computations have been performed with a mesh of 1000 cells and with a time step of 10^{-7} s. Results are shown at time 0.2 μs in Fig. 1 for all numerical methods. Profiles of void ratio and pressure. Near discontinuities, the Jameson scheme produced small oscillations of the solution. For the void ratio profile, we observe a small discrepancy in the post-shock area around $x = 0.85$ m. The solution obtained with the Rusanov and AUSM methods present a small variation, not captured by other methods.

In comparison with the seven-equation model, the pressure curve is quite similar. Yet, we notice some differences between the solutions in the volume fraction profile. In particular, the post-shock values of the void ratio are not the same and the seven-equation model shows an oscillation near the contact discontinuity zone. This behaviour was also noted in simulations presented in [10].

Epoxy-spinel mixture shock tube

In [11] a one meter tube contains two chambers separated at $x = 0.6$ m. A mixture of epoxy and spinel fills both chambers. The initial volume fraction of epoxy is 0.5954 everywhere. The left chamber pressure is $2\,10^{11}$ Pa, while the right chamber is at atmospheric pressure. The fluids are initially at rest. The parameters of EOS are:

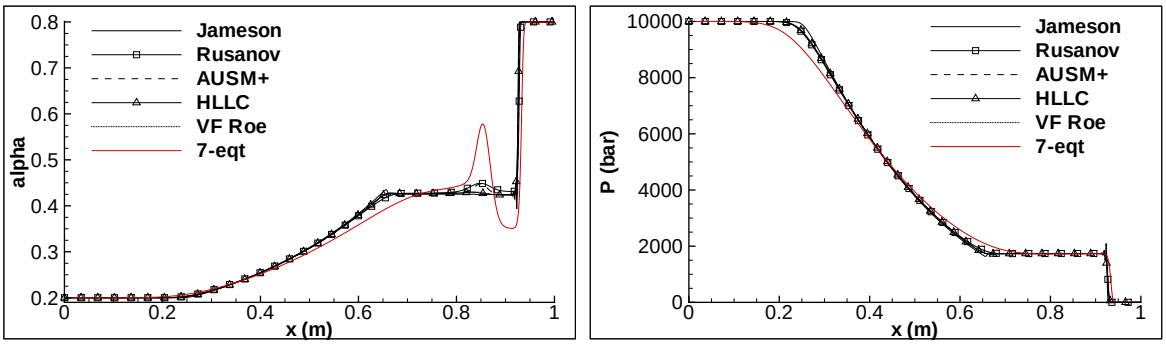


Figure 1. Water-gas shock tube problem. Comparison of different numerical methods comparison on a mesh of 1000 cells at a time of $t = 0.2$ ms. Void ratio and pressure profiles.

$$\begin{pmatrix} \gamma \\ P_\infty \\ \rho \end{pmatrix}_{Epoxy} = \begin{pmatrix} 2.43 \\ 5.3\,10^9 \\ 1185 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \gamma \\ P_\infty \\ \rho \end{pmatrix}_{S\,pinel} = \begin{pmatrix} 1.62 \\ 141\,10^9 \\ 3622 \end{pmatrix}$$

Computations have been performed with a mesh of 1000 cells and with a time step of 10^{-7} s. Numerical solutions computed with the 4-equation model at time $t = 29\,\mu s$ are shown in Figure 2. The analytical solution of the equilibrium model proposed in [11] is incorporated for the sake of comparison and validation. Differences between solutions are weak. For the void ratio profiles, the plateau after the shock is less intense with the Rusanov scheme. As previously indicated, the solution computed with the Jameson scheme presents small oscillations near discontinuities. For all methods, the pressure profiles are in close agreement with the analytical solution. Discrepancies appear on the void ratio jump at shock front, which is underestimated by all models, especially the seven-equation model.

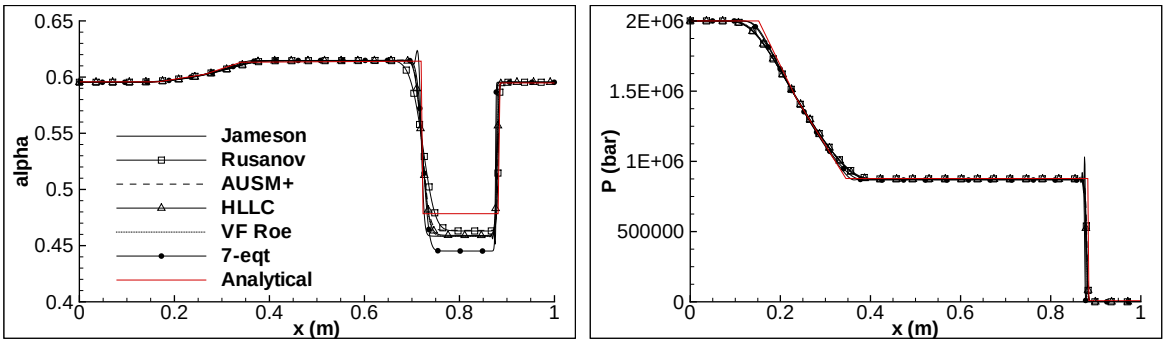


Figure 2. Epoxy-spinel shock tube problem. Different numerical methods comparison on a mesh 1000 cells at time $t = 0.29\,\mu s$. Void ratio and pressure profiles.

Water-gas mixture expansion tube, | u |= 2 m/s

An expansion tube problem is considered with an initial velocity discontinuity located at the middle of the tube. This test consists in a one meter long tube filled with liquid water at atmospheric pressure and with density ρ_l =1150 kg/m^3. A weak volume fraction of vapor α =0.01 is initially added to the liquid. The initial discontinuity is set at 0.5 m, the left velocity is -2 m/s and the right velocity is 2 m/s. The solution involves two expansion waves. As gas is present, the pressure cannot become negative. To maintain positive pressure, the gas volume fraction increases due to the gas mechanical expansion and creates a pocket [13].

In Figure 3, the solution obtained is presented at time $t = 3.2$ *ms*. The mesh contains 1000 cells. The time step is set to 10^{-7} s. The pressure evolution marks large discrepancies. Solutions provided by the Jameson, Rusanov and AUSM methods are in close agreement with the two-fluid solution computed in [20]. With the approximate Riemann solvers, the rarefaction waves are badly predicted. A CPU time of 14h was necessary for the two-fluid simulation. With our 4-equation model, using the Rusanov or Jameson scheme, the CPU time is less than five minutes.

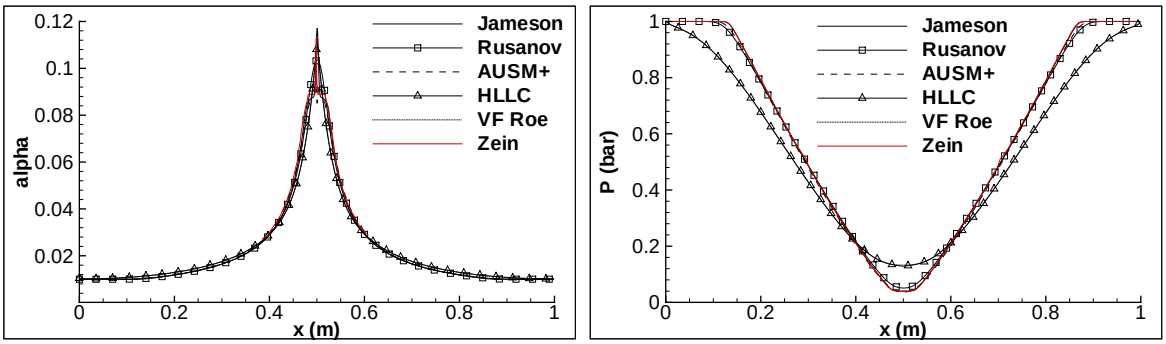


Figure 3. Water-gas expansion tube $|u| = 2$ m/s. Different numerical methods comparison on a mesh of 1000 cells at time $t = 3.2\ \mu s$. Void ratio and pressure profiles.

Water-gas mixture expansion tube, $|\ u\ | = 100$ *m/s*

In [13], an expansion tube, double rarefaction, test is considered. A one meter tube filled with pure water is at atmospheric pressure. The density for water is 1000 kg/m^3. An initial velocity discontinuity is located at $x = 0.5$ m. The velocity of the right part is set as 100 m/s, and the left part as -100 m/s. The EOS parameters are similar to those used for the previous test case. A small volume fraction of gas (1 kg/m^3) is initially present in the water. This case is stiffer than the previous one because of the high value of the initial velocity. Computations are performed on a 1000-cell mesh with a time step set to 10^{-7} s. The approximate Riemann solvers (HLLC and VF Roe) were not able to provide a solution. An anti-diffusive term can be added to the HLLC dissipation to improve the scheme. It has been not tested in the present study.
Figure 4 presents results obtained with the 4-equation model at time $t = 1.85$ ms. The pressure evolution given by the AUSM scheme is not correct. With a grid refinement, the solver leads to divergence. We observe also oscillations on the velocity profiles near the initial discontinuity position. On the contrary, the solutions provided by both the Jameson and Rusanov scheme are in very good agreement with solutions presented in [13].

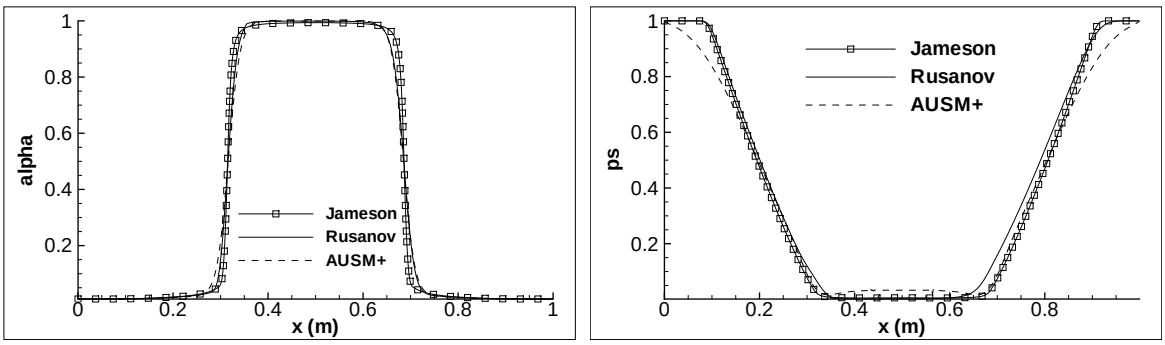


Figure 4. Water-gas expansion tube $|u| = 100$ m/s, numerical methods comparison, mesh 1000 cells, $t = 1.85\ \mu s$. Void ratio and pressure profiles.

Concluding Remarks

This paper provides a comparison of various numerical methods for compressible two-phase flow four-equation model. In its present form, these methods include the AUSM-family, approximate Riemann solvers (VF Roe, HLLC), a simple Godunov approach (Rusanov) and a space-centered scheme with artificial dissipation (Jameson). We then extensively investigated the proposed methods in the existing system of equations on the basis of shock and expansion tube problems. The simulation results suggest the rarefaction waves near the vacuum apparition is more than hard situation for both the approximate Riemann solvers and the AUSM scheme. More specifically, it is not possible to obtain a resolution using these methods. Only the Jameson and Rusanov methods facilitated the simulation of large rarefaction cases.
The presented computational results give considerable confidence in our four-equation model and methods for use as a robust and reliable approach in shock and expansion tube problems of two-phase flows. Room is still available for further work on such problems. For instance, investigation of anti-diffusive terms in needed towards homogeneous and

non-equilibrium two-phase flows.

Acknowledgments

The authors gratefully acknowledge the German Jordanian University and Ensma - Pprime for supporting the current work.

References

[1] Abgrall, R. (1996) How to prevent pressure oscillations in multicomponent flow calculations : a quasi conservative approach. *Journal of Computational Physics*, **125**, 150–160.

[2] Baer, MR. and Nunziato, JW. (1986) A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. *International Journal of Multiphase Flow*, **12**, 861–889.

[3] Buffard, T., Gallouet, T. and Herard, JM. (2000) A sequel to a rough Godunov scheme: application to real gases. *Computers & Fluids*, **29**, 813–847.

[4] Chang, CH. and Liou, MS. (2007) A robust and accurate approach to computing compressible multiphase flow: stratified flow model and AUSM$^+$-up scheme. *Journal of Computational Physics*, **225**, 840–873.

[5] Clerc, S. (2000) Numerical simulation of the homogeneous equilibrium model for two-phase flows. *Journal of Computational Physics*, **161**, 354–375.

[6] Goncalves, E. (2013) Numerical study of expansion tube problems: Toward the simulation of cavitation. *Computers & Fluids*, **72**, 1–19.

[7] Goncalves, E. and charriere, B. (2014) Modelling for isothermal cavitation with a four-equation model. *International Journal of Multiphase Flow*, **59**, 54–72.

[8] Jameson, A., Schmidt, W. and Turkel, E. (1981) Numerical solution of the Euler equations by finite volume methods using Runge-Kutta time stepping methods, AIAA Paper 81–125.

[9] Kapila, A.K., Menikoff, R., Bdzil, J.B., Son S.F. and Stewart, D.S. (2001) Two-phase modeling of deflagration-to-detonation transition in granular materials: reduced equations. *Physics of fluids*, **13**, 3002–3024.

[10] Murrone A. and Guillard, H. (2005) A five equation reduced model for compressible two phase flows problems. *Journal of Computational Physics*, **202**, 664–698.

[11] Petitpas, F., Franquet, E., Saurel, R. and Le Metayer, O. (2007) A relaxation-projection method for compressible flows. Part II: artificial heat exchanges for multiphase shocks. *Journal of Computational Physics*, **225**, 2214–2248.

[12] Rusanov, V.V. (1961) Calculation of interaction of non-steady shock waves with obstacles. *Journal of Computational Mathematics and Physics*, **1**, 267–279.

[13] Saurel, R., Petitpas, F. and Abgrall, R. (2008) Modelling phase transition in metastable liquids: application to cavitating and flashing flows. *Journal of Fluid Mechanics*, **607**, 313–350.

[14] Saurel, R., Boivin, P. and Le Métayer, O. (2016) A general formulation for cavitating, boiling and evaporating flows. *Computers & Fluids*, **128**, 53–64.

[15] Tang, K. (2012) Combining discrete equations method and upwind downwind-controlled splitting for non-reacting and reacting two-fluid computations. Ph.D thesis, University of Grenoble.

[16] Toro, E.F. (1999) *Riemann solvers and numerical methods for fluid dynamics*, 2nd edition, New York: Springer.

[17] Wallis, G. (1967) *One-dimensional two-phase flow*. New York, McGraw-Hill.

[18] Zeidan, D. (2009) The Riemann problem for a hyperbolic model of two-phase flow in conservative form. *International Journal of Computational Fluid Dynamics*, **25**, 299–318.

[19] Zeidan, D. (2016) Assessment of mixture two-phase flow equations for volcanic flows using Godunov-type methods. *Applied Mathematics and Computation*, **272**, 707–719.

[20] Zein, A., Hantke, M. and Warnecke, G. (2010) Modeling phase transition for compressible two-phase flows applied to metastable liquids. *Journal of Computational Physics*, **229**, 2964–2998.

Parametric Study on the Effects of Catenary Cables and Soil-Structure Interaction On Dynamic Behavior of Pole Structures Using the Finite Elements Method & Exprimental Validation

R. Khosravian[1,a)], M. Steiner[1] and C. Koenke[2]

[1]Research Training Group 1462, Bauhaus-University Weimar, Germany

[2]Institute of Structural Mechanics, Bauhaus-University Weimar, Germany

[a)]Corresponding & presenting author: Reza.Khosravian.Champiri@uni-weimar.de

ABSTRACT

Being a typical structural element in infrastructure of transportation systems, the poles are one of the key parts of almost any railway system, carrying the required electricity wires and further side-supplies. On the other hand, numerical simulations have become an inseparable part of any modern engineering task, such that they lead to a deeper insight into the problem and its various perspectives. Pole structures, not being an exception, have attracted significant attention in this regard, especially due to the increase in utilization of railway systems. Therefore, a deep study on diverse modeling aspects of such structures is a necessity to obtain trustable simulation results.

The current study is a survey aimed at investigating the effects of two factors, namely the catenary cables and soil-structure interaction (SSI), on the dynamic behavior of the pole structures which are used in a high-speed train line connecting the cities of Leipzig and Erfurt in the eastern region of Germany. The study is conducted using 3D Finite Element models (FEM). The final goal is to gain an understanding of how the two mentioned factors, from a modeling point of view, affect the eigenfrequencies of the structure.

Initially, the modeling aspects and assumptions used in the study are clarified, and the methods which were used to model the catenary cables and the SSI are briefly explained. Henceforth, the simulation results are presented and discussed. Finally, a parameter study is performed in order to identify the most decisive parameters of the model when calculating the eigenfrequencies, while simultaneously observing the behavior of the model when only one parameter changes. Last but not least, the eigenfrequencies calculated using the acceleration data which are extracted from the sensors installed on an in-service pole are presented, so that a comparison between the modeling results and those of the real-world model would further assist in making a judgment about the prognosis capability and accuracy of the simulations. Such a comparison especially proves to be useful in order to decide about the boundary conditions and the modeling assumptions concerning the SSI and the cables. It is also worth noting that in the course of the parameter studies, the so called *metamodeling* techniques are used after being shortly introduced, to accelerate the analyses.

Keywords: Soil-Structure Interaction, Pole, Catenary Cables, Dynamic Behavior, Sensitivity Analysis, Metamodeling.

Introduction

Poles, either the ones used for luminary posts or electricity cables along railway systems, are nearly identical structures that despite their relative simpleness in shape and dimensions, are subject to various experimental and numerical investigations. Among various reasons, one could name the possibility of consideration of stochastic properties since the experimental data could stem from multiple structures which are commonly considered to be identical when numerically modeled; however, the necessity of accurate simulation of such structures is undeniable due to their importance and vast utilization. Being diverse in dimensions, usage, structural characteristics and building material, the poles investigated in this study are a part of a high-speed railway system which connects the cities of Leipzig and Erfurt in the eastern region of Germany to each other. Made of reinforced concrete, the investigated structures are prestressed spun-cast poles with a length of 10 meters and outer diameter of 40 and 25 centimeters at the bottom and the top respectively. Unlike statically-cast concrete poles, the spun-cast concrete poles are cenrifugally spun with embedded high-strength, prestressed steel strands which are totally enclosed within the concrete. This technique allows the poles to be extremely strong, while improving the resistance against corrosion. Each pole has a floating pile underneath as the foundation, with a length of almost 6 meters.

Being a critical issue in order to ensure the continues supply of energy, the stability of these pole-pile systems is provided using the direct embedment method. This involves creating a cylindrical hole in the ground by a drill and inserting the concrete pole into the open hole, whose gap is then to be filled using grouting materials. This is a common technique especially in cases which are subject to high overturning moments but only moderate vertical loads. The pole studied in this survey, caries only the catenary cables, but not the full electricity system yet.

Figure 1. The pole with the catenary cable

The goal is to initially investigate the effects of the catenary cables and SSI on the eigenmodes of the structure when modeled using FEM. After a brief introduction to the modeling assumptions, techniques, and the method to model the SSI, the simulation results are presented and discussed. Henceforth, a parameter study is performed in order to identify the most decisive parameters when calculating the eigenfrequencies, while simultaneously observing the behavior of the model when only one parameter changes. Finally, the eigenfrequencies calculated using the acceleration data which are extracted from the sensors installed on an in-service pole are presented, so that a comparison between the modeling results and those of the real-world model would further assist in making a judgment about the prognosis capability and accuracy of the simulations. This will also lead to the conclusion, which boundary conditions in the model simulate the reality in a better manner. The survey comes to an end after making the final conclusions, and indicating the open areas related to the problem, which are to be further investigated.

Modeling Aspects

General

Despite the resemblance of the general behavior of the structure to that of a cantilevered beam, in the absence of closed-form solutions, FEM was used to calculate the natural frequencies of the structure. The structure is modeled using the FEM, having almost 55000 quadratic tetrahedral mesh elements. The prestressing effect was neglected due to the fact that the prestressed load is considerably lower than the buckling load of the pole, a fact that makes this effect negligible when

calculating the eigenfrequencies [1].

The soil behavior is assumed to stay in linear range since the main concern is the calculation of the eigenfrequencies, although the linear springs modeling the soil are calculated considering the soil characteristics. The methodology which also accounts for SSI is briefly explained. The concrete is also modeled using a linear elastic constitutive law, based on the properties attributed to C80/95 concrete in Eurocode 2.

In the course of the study, a naming convention was used to differentiate the seven different models (six numerical and one experimental) more clearly:

- Ref: The pole with clamped support, no catenary cables, no SSI
- Ca: The pole with clamped support, with catenary cables, no SSI
- S1: The pole with spring support counting for SSI (Constant soil profile), no catenary cables
- CaS1: The pole with spring support counting for SSI (Constant soil profile), with catenary cables
- S2: The pole with spring support counting for SSI (Parabolic soil profile), no catenary cables
- CaS2: The pole with spring support counting for SSI (Parabolic soil profile), with catenary cables
- Exp: The eigenmodes calculated using the acceleration data of the sensors installed on an in-service pole

Last but not least, throughout the entire paper, the X direction represents the direction in which the catenary cables are extended, while Y is its perpendicular direction and Z is aligned with the length of the pole.

Catenary Cables

The catenary cables at service during this phase of the project, were the ones used for electricity grounding only (Figure 1). Having a cross sectional area of $242.5mm^2$, the cables are made of aluminum type 243 AL1 (DIN EN 50182). The distance between each pair of poles is 65 meters which results in a sag of 1 meter in the middle based on field observations.

Modeled using linear elastic material, the cables were assumed to behave geometrically nonlinear, such that after the deformation due to their self weight, they resembled a hyperbolic cosine function ($f(x) = a.\cosh(\frac{x}{a})$); the function's shape is decided by a constant parameter, the so-called *catenary constant* (a), which is the ratio of the horizontal tension to the weight of the cable in the middle and is to be calculated in an iterative procedure since it is initially unknown [2]. A circular cross section with the mentioned cross sectional area was used to model the cables; however, its bending and torsion stiffness was supposed to be only 20% of that of a rigid cross section with the same material properties and cross sectional area.

For modeling simplifications, the cables were substituted by a spring-mass system; however, the stiffness and mass of the system were numerically calculated based on the entire cable system modeled separately. In order to calculate the stiffness ($K = \frac{\Delta F}{\Delta X}$) of the spring which substituted the cable system, only two cables were modeled as explained before, and the change of reaction force (ΔF) was calculated when a known displacement (ΔX) was applied at their intersection point. As illustrated in Figure 2, different values of applied displacement led to different values of stiffness.

Although the stiffness with regard to the displacement of 0.1 meter was adopted for the rest of the calculations, the effect of the chosen value will be discussed in the parameter studies. Finally, the entire cable system was modeled using the linear spring and the mass of the cables acting in X and Z directions respectively.

Soil-Structure Interaction (SSI)

As long as the model concerns the calculation of the eigenmodes, the assumption of not violating the linear range stays valid [3]. Hence, the substructure part in models which simulated the SSI was modeled using seven elastic springs, such that the springs would also count for the interaction between the soil and the floating pile underneath the pole. Proposed by Novak [4], the stiffness values of the springs (three translational, two rotational and two translational-rotational coupling springs) depend on the soil's shear modulus and density, as well as on the pile's modulus of elasticity and radius. Furthermore, the slenderness and bottom condition (floating or end-bearing) of the pile are decisive only in the vertical direction, which is not of great importance in this case due to the relatively low loading in the vertical direction, as was also proven in the parameter studies to be presented. The method eventually leads to four spring stiffness values for horizontal degrees of freedom (DOF), vertical and rotational DOFs as well as the coupling between the rotational and horizontal DOFs. In

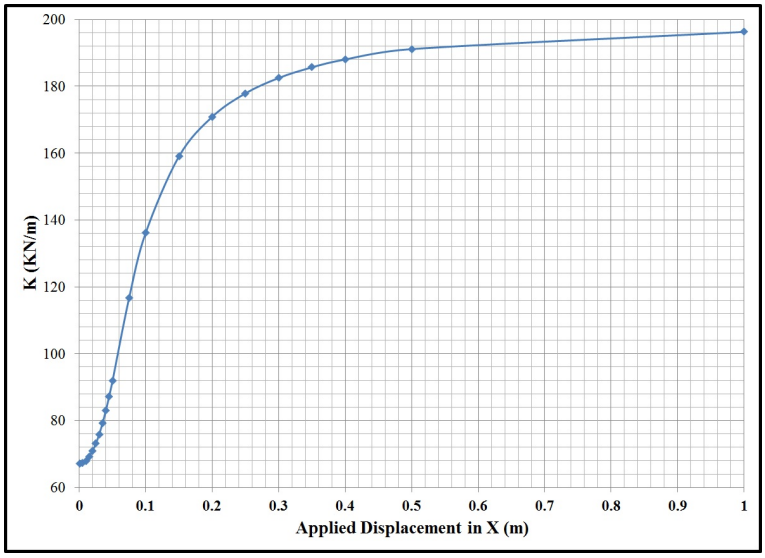


Figure 2. Stiffness of the spring substituting the cables

their work, Novak neglects the torsional behavior around the pile's axis since he states that this motion is not only strongly frequency dependent, but also consequential just for caisson foundations or groups of massive piles.

Moreover, one last critical assumption in the mentioned methodology concerns the soil profile. The soil's shear modulus is considered to be either constant or varying with depth according to a quadratic parabola. Parabolic variation of soil's shear modulus (models S2 and CaS2), versus a constant modulus in the entire soil profile (models S1 and CaS1), represents the physically homogeneous soil stratum with its shear modulus increasing by depth, as the confining pressure enlarges. Each assumption leads to a set of spring parameters which were studied in this work. Full details on this method and the exact formulations could be found in [4].

The spring stiffness values calculated for this problem are shown in Table 1.

Table 1. Stiffness Values of the SSI Springs (GN/m)

DOF	Constant Soil	Parabolic Soil
Vertical (V)	2.472	1.601
Horizontal (H)	1.049	0.382
Rotational (R)	1.548	1.241
H-R Coupling	-0.906	-0.543

It is no surprise that the springs representing the soil with parabolic stiffness profile are softer compared with their constant soil counterparts, due to the loss of stiffness in top layers of the soil.

Test Results: Model vs. Experiment

In order to separately understand the effects of the two factors, the cables and the SSI, on the dynamic behavior of the pole, solely the results of the simulations are initially presented and discussed. Eventually, the experimental results are presented as a measure to judge the precision of the models.

Figure 3 and Table 2 represent the mode shapes and their respective eigenfrequencies for the clamped model (Ref). Moreover, the even modes (2, 4, 6 and 8) represent the modes in the X direction (along the cables), while the odd modes represent those of the Y direction.

To identify the effects of the cables and the SSI, the five numerical models (Ca, S1, CaS1, S2, CaS2) are compared with

Table 2. Eigenfrequencies of the Clamped Pole (Ref)

Mode	f (Hz)
1^{st} bending (1 & 2)	3.4693
2^{nd} bending (3 & 4)	17.113
3^{rd} bending (5 & 6)	44.160
4^{th} bending (7 & 8)	83.423

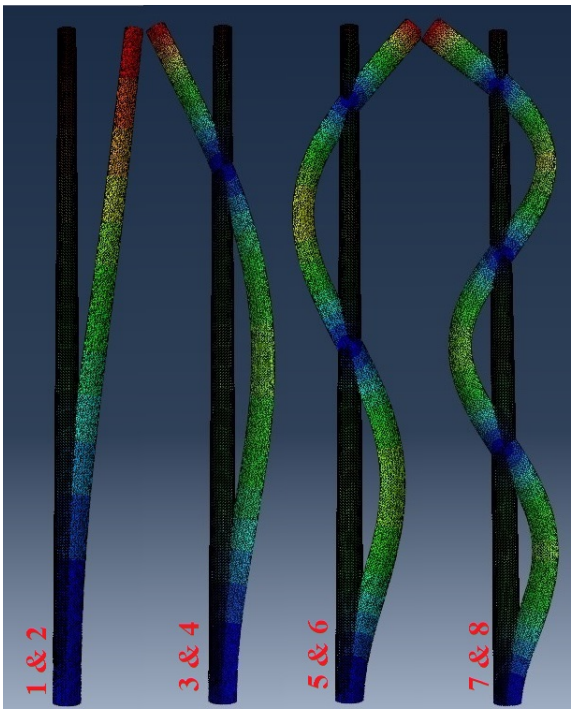


Figure 3. Mode Shapes of the Pole

the model "Ref". Figure 4 illustrates in percentage, how much modeling each phenomenon affects the values of the natural frequencies in the model. The diagram clarifies the effect coming from the inclusion of only SSI (S1 and S2) or the cables (Ca) in the model, while simultaneously showing the overall effect of modeling both phenomena (CaS1 and CaS2).

Based on the diagram in Figure 4, it is possible to state that adding the cable system to the clamped model of the pole results in the reduction of the eigenfrequencies due to the increase in the mass of the system; however, there is a significant increase (over 30%) in the natural frequency of mode 2 (1^{st} bending in X direction) due to the stiffness of the cable system, an effect which is not as influential in higher modes since the action point of the cable stiffness becomes close to the zero-displacement point of the modes (Figure 3). Furthermore, the parabolic soil profile assumption (S2 and CaS2) leads to a softer behavior than the constant soil profile assumption, and the intensity of this difference in behavior becomes more detectable in higher modes, as the structure responds with its stiffer manner.

Analogous to Figure 4, Figure 5 demonstrates the percentage of the difference in eigenfrequencies of the four models (Ref, Ca, CaS1 and CaS2) when compared to those of the experimental data. Having in mind that a positive value in this diagram indicates a stiffer behavior of the model compared to the experimental data, it could be concluded that except the mode 2 and the 4^{th} bending modes, the Ref model (clamped at the bottom) is too stiff to ideally represent the real behavior of the pole, necessitating the simulation of the substructure part and the cable system. Moreover, the cable system has a more dominant effect, similar to Figure 4, compared with the SSI; however, one should not forget that the stiffness of the spring which substituted the cable system plays a significant role here, a parameter which exhibits a large uncertainty due to its nonlinear nature. Furthermore, Figure 5 is a decent basis to judge that the methodology used in this work

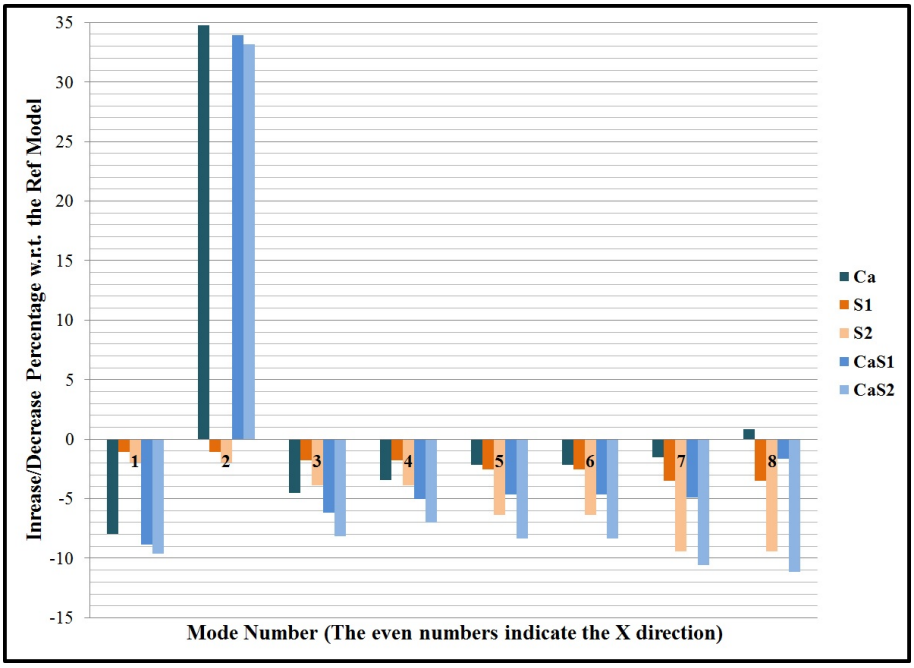


Figure 4. Effect of Catenary Cables & SSI on the Eigenfrequencies

to model the SSI leads to a relatively softer behavior compared to the reality (experimental data), especially with the parabolic assumption for the soil's stiffness profile. It is nevertheless reminded that this methodology has a large field of uncertainties too. Hence, these issues will also be addressed in the parameter studies in order to help to reach a balance between the parameters of the problem, such that a better compromise takes place.

In order to make more supported conclusions, further interpretation of the results is left to be done in the "Conclusions" section, after a more general viewpoint is obtained from the parameter studies.

Parameter Study

Among all the possible factors each of which could be considered as an uncertain parameter in this problem (e.g. the dimensions, density of aluminum etc.), the following 7 parameters were initially assumed to be the most influential and uncertain ones, with their possible ranges of variation shown in Table 3. The model subjected to the parameter studies accounts for both the cable system and the SSI.

Table 3. The Problem's Initial Parameters (SI Units)

Parameter	Abbreviation	Minimum	Maximum
Concrete's Density	Con_Dens (1)	2200.0	2600.0
Concrete's Young's Modulus	Con_E (2)	$3.36 * 10^{10}$	$5.04 * 10^{10}$
Cable Spring's stiffness	K_Spring (3)	$67.06 * 10^{3}$	$196.28 * 10^{3}$
SSI Spring, Horizontal DOF	SSI_H (4)	$4.71 * 10^{8}$	$1.15 * 10^{9}$
SSI Spring, Rotational DOF	SSI_R (5)	$1.19 * 10^{9}$	$1.67 * 10^{9}$
SSI Spring, Vertical DOF	SSI_V	$1.28 * 10^{9}$	$2.97 * 10^{9}$
SSI Spring, H-R Coupling DOF	SSI_HR (6)	$5.07 * 10^{8}$	$9.69 * 10^{8}$

In order to perceive the general influence of each individual parameter, each parameter was changed from its minimum to its maximum in 15 steps, while the rest were kept constant at their mean value. The foremost conclusion was that the change in SSI_V does not affect the eigenfrequencies at all, hence it was omitted from the list of variables for the upcoming sensitivity analyses.

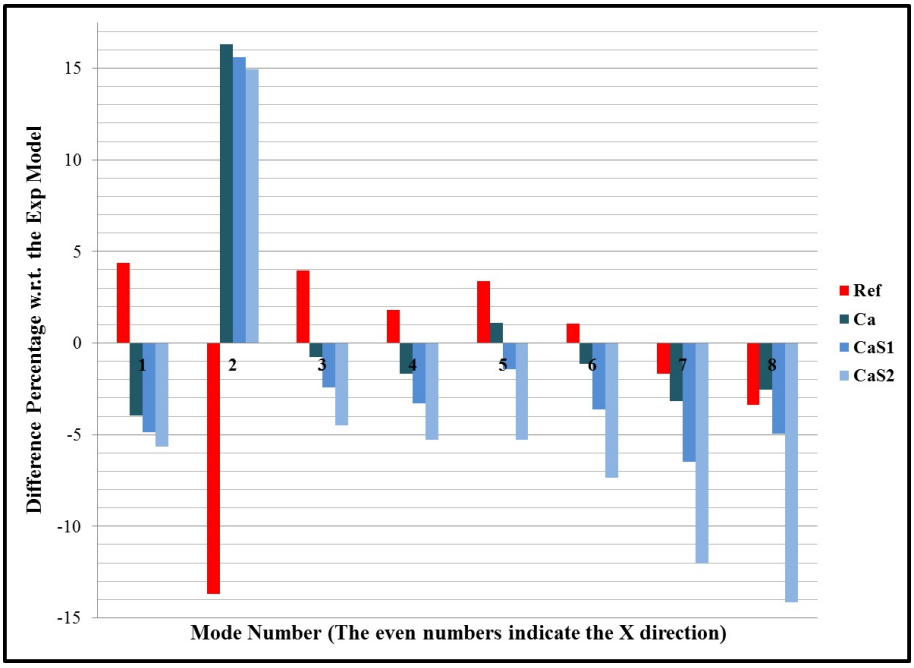


Figure 5. Comparison of the Models with the Experimental Data

Figure 6 shows the results of the parametric study, when the parameters change individually step-by-step in the mentioned range. It is possible to interpret the graph either mode-wise (i.e. judging which parameters affect a specific mode more intensely) or parameter-wise (i.e. concluding which modes a certain parameter affects).

It is understandable from the figure, that K_Spring affects only the mode 2 (1^{st} bending in X direction) majorly, and mode 4 slightly. Moreover, SSI_R's variation proves to have the least effect on the output, among the soil stiffness parameters, and SSI_H is the most decisive parameter in all modes except 2, in which the K_Spring plays a more significant role. Eventually, while the Con_E varies the output significantly more than Con_Dens when varying in their mentioned ranges, both parameters remain to be influential in all of the modes.

Despite the benefits of the conclusions made, the complex nature of this problem triggers the need to a sensitivity analysis, since the behavior of the model is highly nonlinear with respect to some parameters (mainly the SSI parameters) on one hand, and the response also depends on the interaction between the parameters (e.g. the relative stiffness of the SSI and the cable system springs etc.) on the other hand. This would allow the simultaneous, but yet random variation of all the parameters, in order to gain a deeper insight into the problem. Last but not least, this would lead to quantitative measures based on which one could judge on the importance of the parameters, rather than just making qualitative comparisons.

Brief Introduction to Sensitivity Analysis

The need to identify the most significant parameters in a multidimensional problem triggers the efforts to develop methods addressing this issue. Sensitivity analysis, as a popular methodology, is a commonly used approach to fulfill this aim. The final output of sensitivity analysis is a measure, based on which one can judge which parameters are more decisive in the final output of the problem, hence a more efficient orientation of time and cost investment could be done to accurately determine only the crucial parameters.

There are various approaches to perform this analysis. The methodology adopted in this work is a variance-based sensitivity analysis proposed in [5]. In this approach, the first-order effect (S_i) and total effect (S_{T_i}) of the parameter X_i on the output Y is calculated for each i in order to get a general impression about the parameter prioritization.

Being a value theoretically always between 0 and 1, S_i is in fact a measure indicating what would happen to the uncertainty of Y if the i'th parameter would be fixed. Hence, a high value represents an important parameter while a small value does not necessarily signal a low importance for the parameter. In fact, to achieve a thorough understanding of the sensitivity pattern for a model with n parameters, one needs the total set of first-order and total effect indices of the parameters.

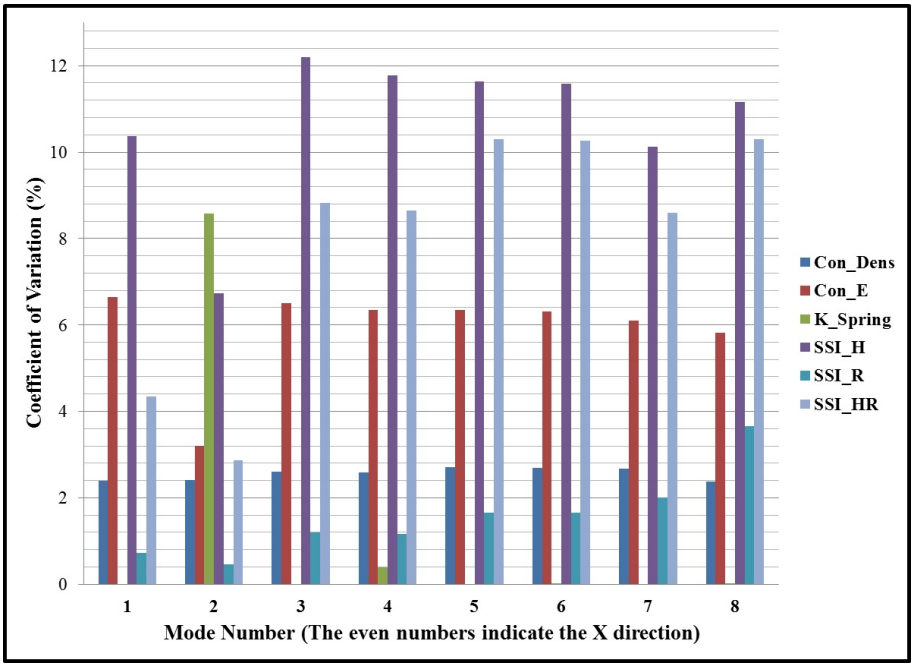


Figure 6. CoV of the Modes When Only One Parameter Changes

Accounting for the total contribution to the output variation due to parameter X_i (that is, its first order effect plus all higher order effects due to interactions among parameters), total effect of this parameter is calculated using a formula which depends on the variances of both the input and the output. The condition $S_i = 0$ is necessary but insufficient to identify parameter i as non-effective, while $S_{T_i} = 0$ is a necessary and sufficient condition for it being non-influential. Accordingly, if $S_{T_i} = 0$, X_i can be fixed at any value within its range of uncertainty without remarkably affecting the variance of the output [5].

Based on the definitions mentioned, S_{T_i} is larger than or equal to S_i, where the latter case happens only when X_i is not involved in any interaction with other parameters. Therefore, the difference, i.e. $S_{T_i} - S_i$ indicates how much parameter i is involved in interactions with other parameters. It is worth mentioning that $1 - \Sigma_i S_i$ is an indicator of presence of interactions among the model's parameters. Moreover, $\Sigma_i S_{T_i}$ is always greater than 1 or, in case that the model is perfectly additive w.r.t its variance, equal to 1 [5].

In spite of the efficiency of variance-based methods to perform sensitivity analysis, high computational costs due to the relatively large number of required samples remains a major drawback of such methodologies, a disadvantage which is to be addressed in this work using the metamodeling techniques.

Brief Introduction to Metamodeling

A common approach to reduce the computational cost of calculating the required outputs for a sensitivity analysis is the application of the so-called *Metamodels*. A metamodel is generally an approximation function which adapts the behavior of a set of input-output data (in this case, the parameters of the FEM model as the input, and the eigenfrequencies as the output). In order to build a metamodel, a support data set $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^k$ and the respective evaluations $\mathbf{y} = [y_1, ..., y_n]^T = [f(\mathbf{x}_1), ..., f(\mathbf{x}_n)]^T$ of the original function $f(\mathbf{x})$ are used. Polynomial Regression [6][7], Moving Least Squares [10] and the Kriging approximation [8][9] are examples of common metamodeling approaches mentioned in the literature. The various techniques differ significantly in their calculation time and approximation quality, which, however, depends on the nature of the observed problem. Accordingly, the optimal metdamodel choice is mainly a case-dependent issue to be addressed.

In this research, various metamodeling techniques were compared and the two optimal models were used as approximation functions. The first one, the Polynomial Regression, is a common and simple approach to adapt a function. There, the original function is approximated by a polynomial function that results in the best fit with respect to the sum of least

squares criterion [11]. This results in the approximation function

$$\hat{f}(\mathbf{x}) = p^{\mathrm{T}}(\mathbf{x})\hat{\mathbf{w}} = p^{\mathrm{T}}(\mathbf{x})(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{1}$$

with $p(\mathbf{x})$ denoting the g-dimensional polynomial basis of $\mathbf{x}$ and $\mathbf{X} = [p^{\mathrm{T}}(\mathbf{x}_1), ..., p^{\mathrm{T}}(\mathbf{x}_n)]^{\mathrm{T}}$ being the matrix containing the basis vectors of the support points. The approximation can be optimized by varying the degree g of the polynomial function. A higher polynomial degree often leads to a better result; nevertheless, there is a risk of over-fitting and extreme increase of the computation time.

The second approach, Kriging approximation, uses a completely different concept since it interprets the data as the output of a stochastic process $\hat{f}(\mathbf{x}_i) = \mu + \varepsilon(\mathbf{x}_i)$ with an unknown constant trend μ and correlated residuals $\varepsilon(\mathbf{x}_i)$. By application of the maximum Likelihood criterion [9][11] the approximation function

$$\hat{y} = \hat{\mu} + \boldsymbol{\psi}(\mathbf{x})\,\boldsymbol{\Psi}^{-1}\,(\mathbf{y} - \mathbf{1}\hat{\mu})\,, \tag{2}$$

can be reached, where $\boldsymbol{\Psi}$ describes the correlation matrix of the support points and $\boldsymbol{\psi}(\mathbf{x})$ is the correlation vector between the support points and the examination point $\mathbf{x}$. [8] and [9] contain full details on derivations and specific formulations of this method.

Compared with the Polynomial Regression, the Kriging method is much more flexible in the fitting procedure, so that usually a higher approximation quality could be expected; however, it is one of the most complex, and hence expensive metamodeling approaches.

In order to make a decent model selection between the possible metamodeling approaches, a meaningful error criterion should be chosen. For the observed data set related to the problem studied here, the coefficient of determination (R^2) [12] with a validation data set was taken as a reference for the model selection. This error measure could be determined with

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{m}\left(y_i^{val} - \hat{f}(\mathbf{x}_i^{val})\right)^2}{\sum_{i=1}^{m}\left(y_i^{val} - \bar{y}^{val}\right)^2}, \tag{3}$$

where $\bar{y}^{val}$ is the mean value of the functions' evaluations $y_1^{val}, ..., y_m^{val}$. To avoid an over rating of the model quality, a set of untrained data is used.

During the comparison process of various metamodels, the results of different methods with different number of support points (number of samples, $n = 200, 500$ and 1000) were tested and the coefficient of determination was calculated with a validation data set of $m = 4000$ points. Based on the calculated values, a separate decision for each of the first four eigenfrequencies was made. Eventually, for the frequencies $f1$, $f2$ and $f4$ the Polynomial Regression with $g = 2$ and $n = 200$, and for $f3$ the Kriging method based on $n = 500$ support points were used. In this work, these metamodels were used to calculate the sensitivity indices of the parameters in the CaS1 model, when the first four eigenfrequences are considered to be the output.

Results of the Sensitivity Analysis (Using the Metamodels)

Using the values mentioned in Table 3, different numbers of samples (n), each containing the mentioned six parameters, were produced using a random procedure, such that they obeyed a uniform distribution. The responses (namely the $f1$, $f2$, $f3$ and $f4$) were calculated using the mentioned metamodels. It is worth mentioning that the same calculation using the original FEM model with the available computation power takes around 10 days for a sample size of only $n = 500$. Increasing the number of samples in this method of conducting the sensitivity analysis, leads each sensitivity index (and hence the sum of the indexes) to converge to a certain value. Figure 7 illustrates this convergence trend for the first order sensitivity indexes, when the output is $f2$ (the first bending in X direction).

Furthermore, Table 4 contains the sensitivity values for $f1$ and $f2$ calculated using the metamodels.

Based on the values in Table 4, it is concluded that the most decisive parameter on the frequency of the first bending mode in X direction ($f2$) is the *K_S pring*, a conclusion which is also consistent with the results of Figure 6. Furthermore, the concrete's Young's modulus is a more important parameter compared to the density of the concrete in calculation of both frequencies; It is also observed that in the Y direction, *Con_E* has a relatively large first order effect on the response, a conclusion that physically makes sense due to the absence of the large influence from the cable system; however, the

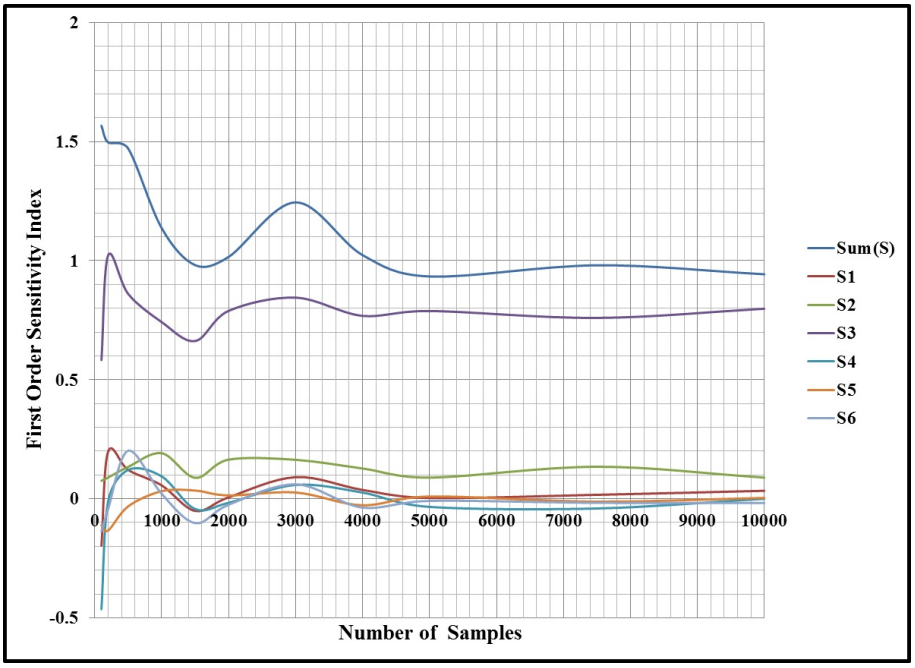


Figure 7. First Order Sensitivity Indexes vs. n for $f2$

Table 4. Sensitivity Indexes for $f1$ & $f2$

Parameter	1	2	3	4	5	6	Sum
$f1(S)$	0.03	0.70	0.00	0.01	0.03	0.00	0.77
$f1(S_T)$	0.07	1.00	0.06	0.09	0.06	0.10	1.38
$f2(S)$	0.03	0.09	0.80	0.00	0.00	0.00	0.92
$f2(S_T)$	0.05	0.21	0.80	0.03	0.02	0.04	1.15

conclusion that *Con_E* is the most influential parameter on $f1$ contradicts the results shown in Figure 6. Moreover, it was also observed during the studies that the converging trend for the sensitivity indexes would not happen for $f3$ and $f4$, besides the fact that such contradictions in the results of the sensitivity analyses with the metamodels continued to exist. This was nevertheless expected, due to the low quality of the created metamodels (R^2 value of around 0.4).

Conclusion & Outlook

The dynamic behavior of a pole structure used in a high-speed railway system in Germany was studied using the FEM, to identify the effects of the catenary cables and the soil-structure interaction (SSI), and to propose a suitable model for simulation of this structure. The numerical results were compared with results extracted from the data acquired from the sensors installed on an in-service pole, to provide a trustable measure against which the numerical results could be judged. The SSI was modeled using two assumptions for the soil profile, namely a constant and a parabolic profile for the stiffness of the soil, while the cable system was modeled using a spring-mass system. A comparison between Figure 4 and Figure 5 shows that while a clamped boundary condition for the pole (no SSI effect and no cables included) is not a suitable approach, the assumption of a parabolic stiffness profile for the soil also leads to a large modeling error in this case. Based on Figure 6 it is concluded that the spring stiffness values of the SSI for the horizontal direction (SSI_H) and the coupling DOF (SSI_HR) are the most decisive parameters of this problem in all modes, except mode 2 (1^{st} bending in X direction) in which the cable stiffness plays the most crucial role. The results of this parameter study were also supported by sensitivity analyses conducted using the metamodels, although the quality of the metamodels led to shortcomings in some areas. Taking these facts into consideration, one can conclude using Figure 5 that the CaS1 model (the model accounting for both the SSI and the cable system, with a constant soil profile assumption) is the best compromise among

all the six numerical approaches used in the study, to simulate the real response of the structure; however, the stiffness of the spring which substituted the cable system should be reduced, while at the same time that of the SSI springs should be increased in order to match the eigenfrequencies of the real pole in service.

However, the eigenfrequencies calculated from the data (the "Exp" model in this work) exhibit different uncertainties due to existing obstacles in conducting in-site measurements and also the quality of the acquired data. Moreover, despite the structural resemblance of the poles used in this railway system, nonidentical boundary conditions for various poles are practically expected to exist; hence, having a higher number of poles with their eigenfrequencies extracted from the acceleration data would significantly increase the trustability of the conclusions made here. Furthermore, this problem triggers the need to use a more advanced metamodeling strategy, e.g. a combination of metamodels for different ranges of the various parameters, in order to cover a wider range of conclusions. Therefore, overcoming these issues and calibrating the model using the results would remain an open problem to be addressed in an extensive work.

Acknowledgment

This research is supported by the German Research Foundation (DFG) via Research Training Group "Evaluation of Coupled Numerical Partial Models in Structural Engineering (GRK 1462)", which is gratefully acknowledged.

References

[1] Blevins, R.D. (1995) *Formulas for Natural Frequency and Mode Shape*, Krieger Publishing.

[2] Costello, E. (2011) Length of a Hanging Cable. *Undergraduate Journal of Mathematical Modeling: One + Two* **4**, Article 4.

[3] Clough, R. W., Penzien, J. (1993) *Dynamics of Structures*. McGraw-Hill Publishing Company, New York, USA.

[4] Novak, M., El Sharnouby, B. (1983) Stiffness Constants of Single Piles. *Journal of Geotechnical Engineering* **109**, 961-974.

[5] Saltelli, A., Ratto, M., Andres, T. (2008) *Global Sensitivity Analysis: The Primer*, John Wiley and Sons, England.

[6] Myers, R.H.(1971) *Response Surface Methodology*. Allyn and Bacon, Boston.

[7] Box, G.E.P., Draper, N.R.(1987) *Empirical Model-Building and Response Surfaces*. John Wiley and Sons, New York.

[8] Krige, D.G. (1951) A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**, 119–139.

[9] Forrester, A.I.J., Sobester, A., Keane, A.J. (2008) *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley and Sons, UK.

[10] Lancaster, P., Salkauskas, K. (1981) Surface Generated by Moving Least Squares Methods. *Mathematics of Computation* **37**, 141–158.

[11] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1986) *Numerical Recipes - The Art of Scientific Computing*. Cambridge University Press.

[12] Schoelkopf, B., Smola, A. (2002) *Learning with Kernels*. MIT Press.

Elevated temperature fatigue and failure mechanism of 2.5D T300/QY8911-IV woven composites

†J. Song[1,2,3], *H. T. Cui[1,2,3], and W. D. Wen[1,2,3,4]

[1] College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, China
[2] Jiangsu Province Key Laboratory of Aerospace Power System, China
[3]Collaborative innovation center of advanced aero-engine, China
[4] State Key Laboratory of Mechanics and Control of Mechanical Structures, China

*Presenting author: cuiht@nuaa.edu.cn
†Corresponding author: cuiht@nuaa.edu.cn

Abstract

Static tensile and tension-tension fatigue tests were conducted on 2.5D woven composites at room and elevated temperatures. Macro-Fracture morphology and SEM micrographs were examined to understand the corresponding failure mechanism. The results show that the stress-strain curves and the fractured morphology are significantly different in the room and elevated temperature environments. Furthermore, the static tensile properties decrease sharply with increasing the temperature due to the weakness of fiber/matrix interfacial adhesion. The fatigue life and damage progression at elevated temperature are also substantially different compared with those at room temperature. Meanwhile, a damage mechanism, called rotation deformation mechanism, was proposed to explain the elevated fatigue behavior.

Keywords: 2.5D woven composites, Elevated temperature, Stress-strain behavior, Fatigue behavior, Scanning electron microscopy, Damage progression

Introduction

Textile composite materials are widely used in advanced aerospace industry, owing to their good comprehensive mechanical performance. However, numbers of structural components exposed to long-term temperatures in 100-200℃, such as aero-engine casing, require that polymer matrix composite materials have an advantage of elevated temperature resistance performance. A new generation of high glass-transition temperature (T_g) polymers such as QY8911-IV[1] has enabled this progressive development, which can be easily used to manufacture the composites by resin transfer modeling (RTM). Additionally, compared to the relatively complex 3D braided or woven structure, a new class of 2.5D angle-interlock woven composites has been proposed. Therefore, it is of great importance to understand the mechanical behavior, especially the fatigue behavior of the materials at various temperatures.

Unfortunately, due to the high-cost and difficult-to-test at elevated temperatures, the related researches related to the static behavior and fatigue life at elevated temperature are relatively backward and most of investigations were focus on FRP[2-4] or the mechanical properties of 2.5D woven composites at room temperature (RT)[1]. Selezneva et al.[5] investigated the failure mechanism in off-axis 2D woven laminates at elevated temperature by experiment, and found that the woven yarns began to straighten out and rotated towards the loading direction just prior to failure. Vieille and Taleb[6] studied the influence of temperature and matrix ductility on the behavior of 2D woven composites with notch and unnotched, and the results revealed that the highly ductile behavior of thermoplastic laminates was quite effective to accommodate the overstresses near the hole at temperatures higher than their T_g. Several static and fatigue tests were conducted by Montesano et al.[7,8] to investigate the fatigue behavior

of a triaxially braided composites at elevated temperatures, and the corresponding stiffness degradation model was proposed based on the measurements of actual observed damage mechanisms.

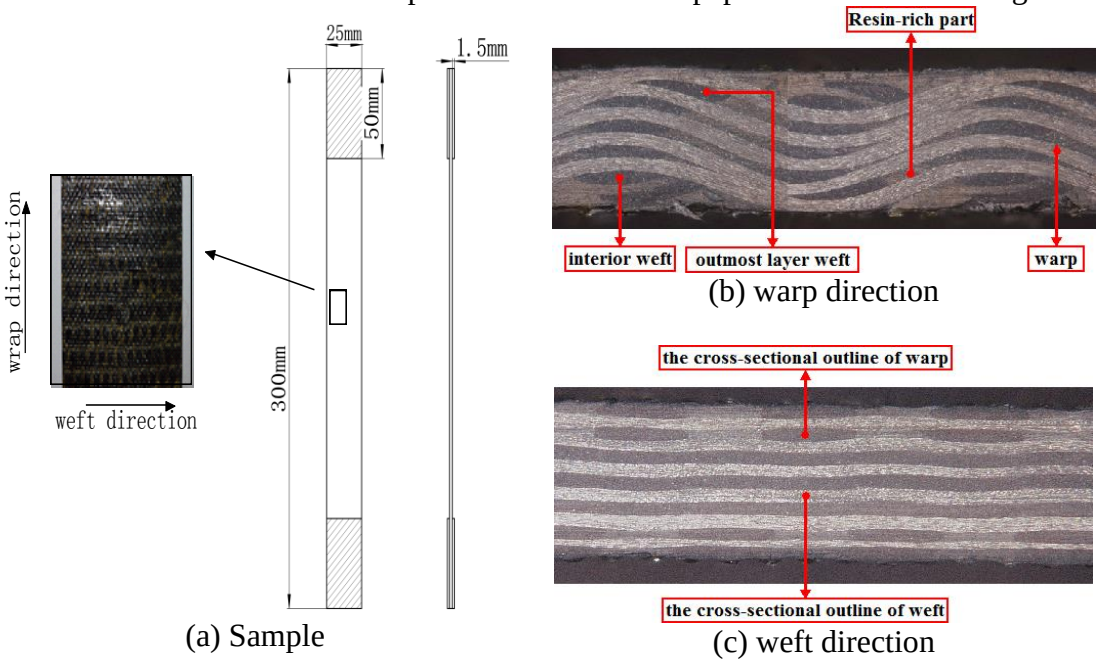
This study aims to investigate the static tension and fatigue behaviors of 2.5D angle-interlock woven carbon fiber/ QY8911-IV composites at room and elevated temperatures by experiments. In the first part, the corresponding elevated experiments are conducted. After that, scanning electron microscopy (SEM) is employed to study the failure mechanism subjected to the static or fatigue loading at different temperatures. Finally, some useful conclusions are presented.

According to this basic research, the database related to the elevated temperature performances and fatigue behavior of 2.5D woven composites can be established at room and elevated temperatures.

Materials and experimental procedure

2.5D woven fabric was prepared using T300 carbon fiber yarns that consist of 3K filaments per bundle, and the matrix is QY8911-IV with a glass transition temperature 256℃.

The flat composite panels with six plies of weft yarns were manufactured by the resin transfer molding (RTM) process. The static and fatigue test specimens with a fiber volume fraction of 42.94% were obtained (see Fig. 1). In addition, the microstructure is actually a spatial net-shape fabric, which is formed by interlacing binding threads in the thickness direction to join adjacent layers of warp and weft together, and cured with matrix under certain conditions. The architecture of 2.5D woven composites studied in this paper is also shown in Fig. 1.



(a) Sample (b) warp direction (c) weft direction

Figure 1. Static tensile/ fatigue samples and the corresponding internal microstructure

As there are no standards of static tensile and tension-tension fatigue tests for the 2.5D woven composites at elevated temperature, the corresponding test procedures were followed by ASTM D 3039[9] and ASTM D 3479[10], respectively. All of the tests were conducted by an MTS 810 hydraulic servo dynamic material test machine (see Fig.2a) with a 25.4mm MTS-634-25 extensometer (see Fig.2b) used to monitor the strain continuously during the static and fatigue tests. Moreover, an MTS809 furnace with an integrated temperature controller was

used, which can ensure the temperature in the chamber is consistent throughout the duration of all tests within 2℃.

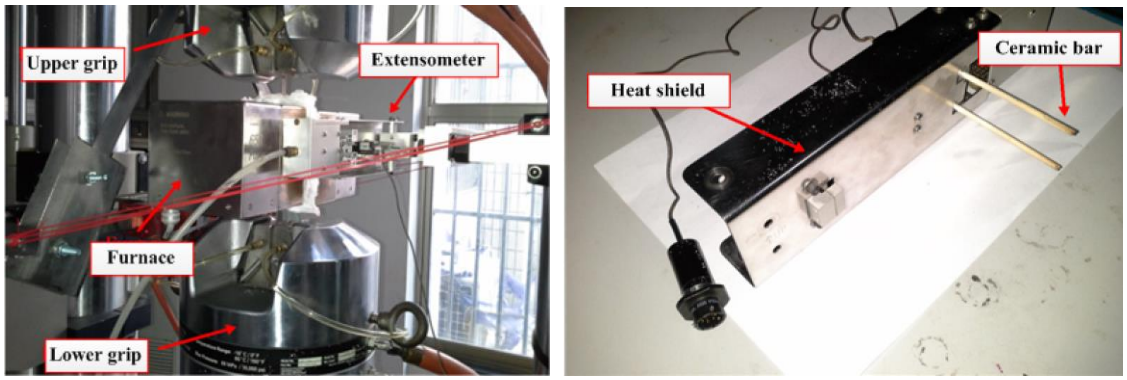


Figure 1. Photograph of MTS-810 test machine (a) and 634-25 extensometer (b)

Results

3.1 Typical stress-strain behavior at different temperatures

Fig. 3(a) shows the typical stress vs. strain curves of 2.5D woven composites tested at 20℃ and 180℃. At room temperature (20℃), the materials behave almost in a linear manner up to approximately 1%, after which an obvious nonlinear behavior can be observed up to the failure. At 180℃, the slope of the curves reduces significantly due to the resin matrix softening, interfacial debonding or sliding, resulting in a nonlinear response up to ultimate fracture.

Fig. 3(b) summarizes the modulus and UTS of the composites at 20℃ and 180℃. Comparing the properties at RT with that at 180℃, the average moduli are 48.39GPa and 40.78GPa, respectively, and the modulus at 180℃ decreases by 15.73%. Meanwhile, the average tensile strengths are 515.09MPa and 431.89MPa, respectively, and the property at 180℃ decreases by 16.15%. The results indicate that the mechanical properties are very sensitive to temperature (180℃).

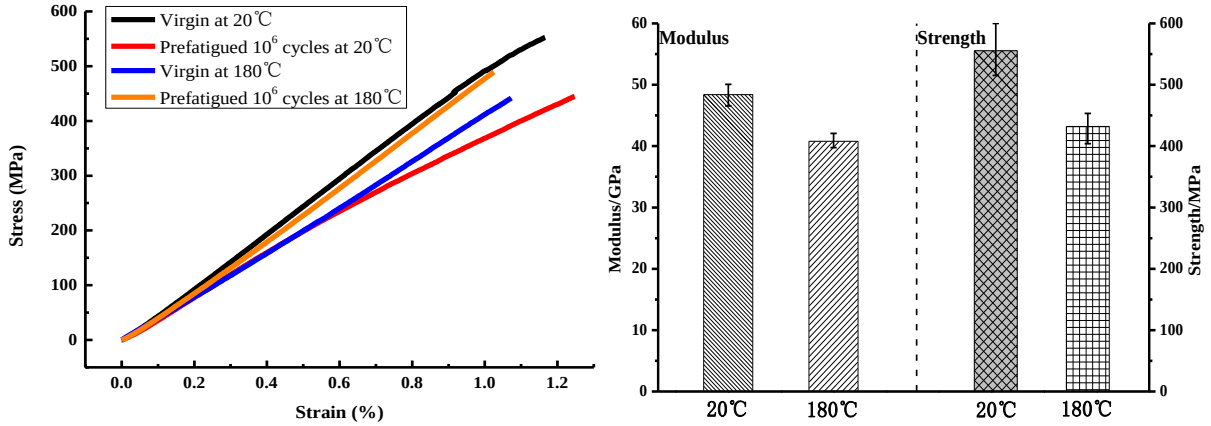


Figure 3.(a) Representative tensile stress-strain curves of 2.5D woven composites for the virgin and fatigued specimens at 20℃ and 180℃;(b) Tensile properties of 2.5D woven composites at RT and 180℃

3.2 S-N curves at various temperatures

Fig. 4(a) shows the normalized stress-fatigue life curves of 2.5D woven composites at RT and 180℃ and the corresponding values are listed in Table 1 and 2. There are significant differences in fatigue behavior between RT and 180℃. The elevated temperature causes a reduction in the fatigue life, and the fatigue strength for the specimens tested at RT is about 1.2 times of that at 180℃. Additionally, it seems that there is a threshold for the elevated S-N curves. The elevated specimens subjected to the maximum fatigue loading in the range of 73%-80% have a quite short fatigue life (less than 1×10^4 cycles). Nevertheless, when the stress levels are lower than 70%, the fatigue life reaches the predefined infinite life. This phenomenon was also observed by Zhu[11], who studies the fatigue behavior of 3D braided composites at RT.

Table 1. Fatigue life (cycles) test result of 2.5D woven composites at room temperature

Stress level	No.	Peak load/N	Valley load/N	Fatigue life	Average life
90%σ_u	1	23.13	2.32	9303	9303
87%σ_u	2	22.31	2.23	27658	18909
	3	21.83	2.18	10159	
83%σ_u	4	21.33	2.13	44149	44149
80%σ_u	5	21.21	2.12	73918	103545
	6	21.36	2.14	133171	
78%σ_u	7	20.25	2.033	10^{6*}	10^{6*}
75%σ_u	8	20.33	2.04	10^{6*}	10^{6*}

Table 2 Fatigue life (cycles) test result of 2.5D woven composites at 180℃

Stress level	No.	Peak load/N	Valley load/N	Fatigue life	Average life
80%σ_u	1	18.06	1.81	1511	1221
	2	18.60	1.86	931	
75%σ_u	3	16.81	1.68	2672	2411
	4	17.34	1.73	2150	
73%σ_u	5	16.13	1.61	4125	4221
	6	16.44	1.64	4317	
70%σ_u	7	15.65	1.57	10^{6*}	10^{6*}
	8	15.31	1.53	10^{6*}	10^{6*}
	9	15.70	1.58	10^{6*}	10^{6*}

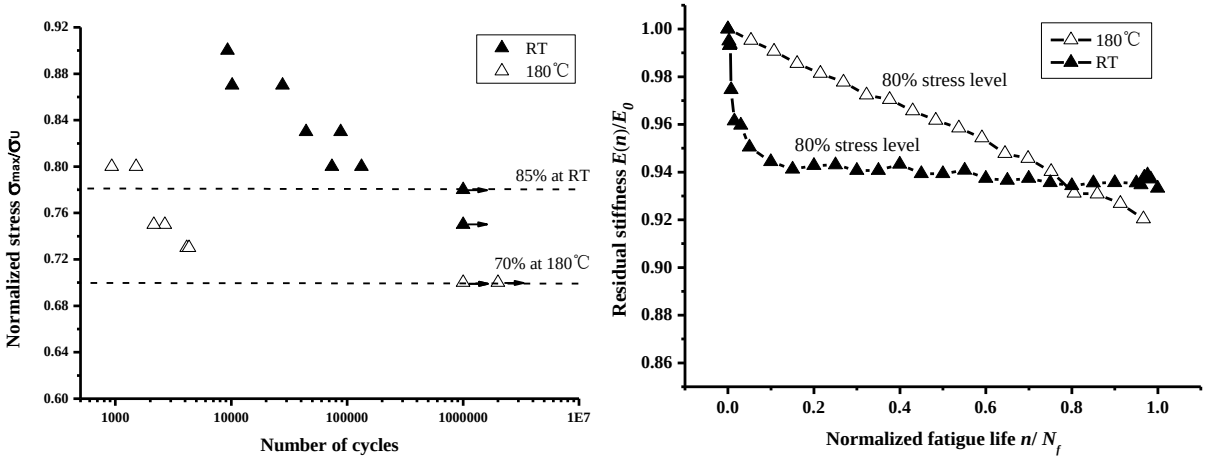


Figure 4(a). S-N curves of 2.5D woven composites at RT and 180℃;(b) Normalized stiffness for maximum applied stress of 80% UTS at RT and 180℃

Additionally, from the view of residual strength (Fig .3(b)), it can be found that the residual strength at elevated temperature is higher than the virgin strength at the corresponding temperature, which can result in an infinite life is reached.

3.3 Stiffness degradation behavior at various temperatures

Fig. 4(b) shows the normalized dynamic stiffness vs. cycle curves for test specimens cycled with maximum applied stress level of 80% at RT and 180℃. The stiffness degradation behavior for the specimens tested at RT can be characterized by a rapid stiffness degradation trend during the first stage of cycling, followed by a gradual stiffness degradation trend during the subsequent stage and a rapid stiffness drop occurs prior to final fracture. The stiffness degradation feature obtained at RT is similar with that for laminated composites tested at RT[12]. Whereas, the notably difference in stiffness degradation behavior is relative to the elevated temperature specimens compared to the room temperature specimens. Compared with the room temperature stiffness behavior, a more gradual stiffness degradation characteristic is observed at elevated temperature environment (Fig. 4(b)). This may result from the duration of matrix affected by elevated temperature.

3.4 Residual strength behavior

In order to investigate on the abnormal fatigue behavior tested at 180℃ mentioned above, residual strength tests were performed subjected to 80% stress level at RT and 180℃. After reaching a certain cycle number (1×10^6 cycles), fatigue tests were terminated, and then the as-fatigue strength (defined residual strength) was measured. The corresponding results have been plotted in Fig. 3(a). It can be seen that the tensile stress vs. strain behaviors at RT or 180℃ after the cyclic loading are clearly different from those for the virgin specimens at the corresponding temperatures. The room temperature strength and modulus are both lower than the corresponding fatigued composites, however, the elevated strength and modulus are higher than the fatigued composites conducted at the same temperature, which suggests that although the elevated temperature specimen has been experienced 1,000,000 cycles, the elevated mechanical properties can be strengthened instead. The increase in strength is in agreement with those observed on other composites[13, 14].

3.5 Fractured surface morphology

Fig. 5 and Fig. 6 display the morphology of fractures observed from the macroscopic and microscopic views for the static tensile samples. From Fig. 5, there is no obvious necking phenomena observed at the fractured surface at RT and 180℃, indicating a brittle-natured fracture. The fracture mainly occurs in the warp bundles at the crossover points of warp and weft bundles. Although larger damage regions and more delamination cracks are observed at RT (see Fig. 5a, b), there is relatively less fiber pull-out for the specimen tested at RT (see Fig. 6).

Fig. 7 and Fig. 8 show the magnified SEM photomicrographs of the fractured surface of the 2.5D woven composites tested at RT and 180℃. From Fig. 7, for the room temperature failure fractures, the failure behaves as interfacial debonding between fibers and matrix and the localized fibers bundles loosen within each other observed near the fractured surface. Whereas, for the elevated temperature failure fractures, the material maintains good integrity and less inter-

facial debonding damage. Similar fractured morphology with the static tests at 180℃, it is noticeable that the presence of fiber pull-out for the specimens conducted at 180℃ is revealed by the brushy appearance of the fracture surface (see Fig. 8).

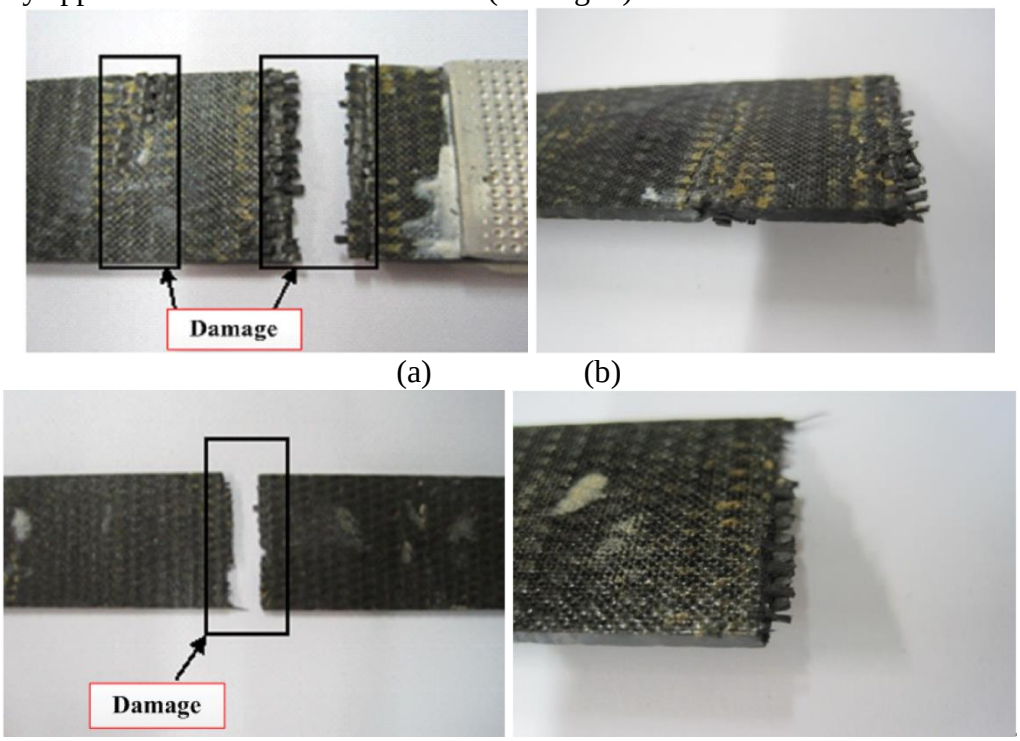


Figure 5. The fracture photographs of static tension samples at (a)-(b) room temperature and (c)-(d) 180℃

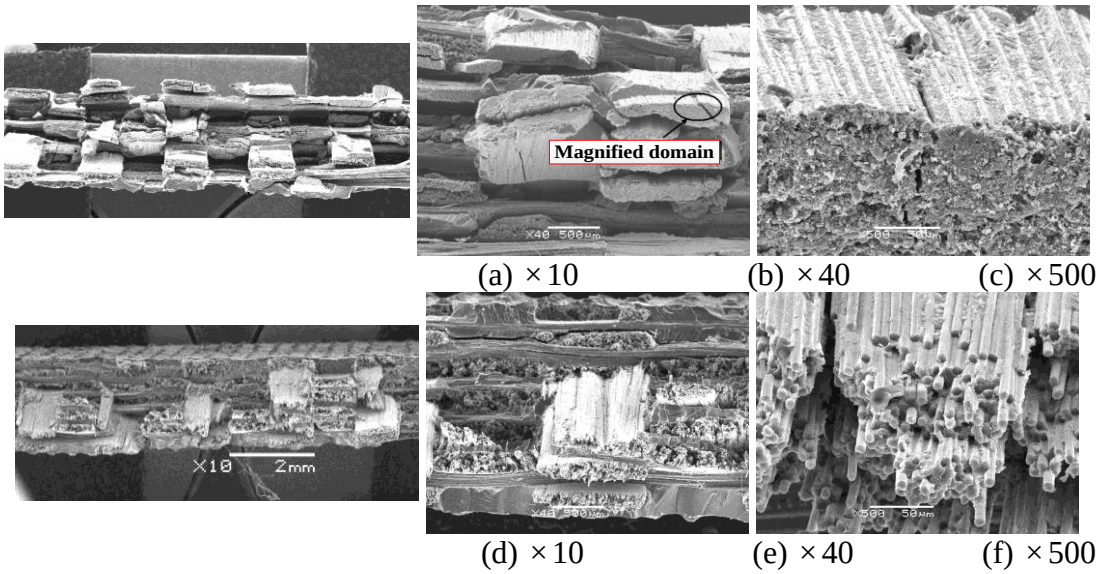
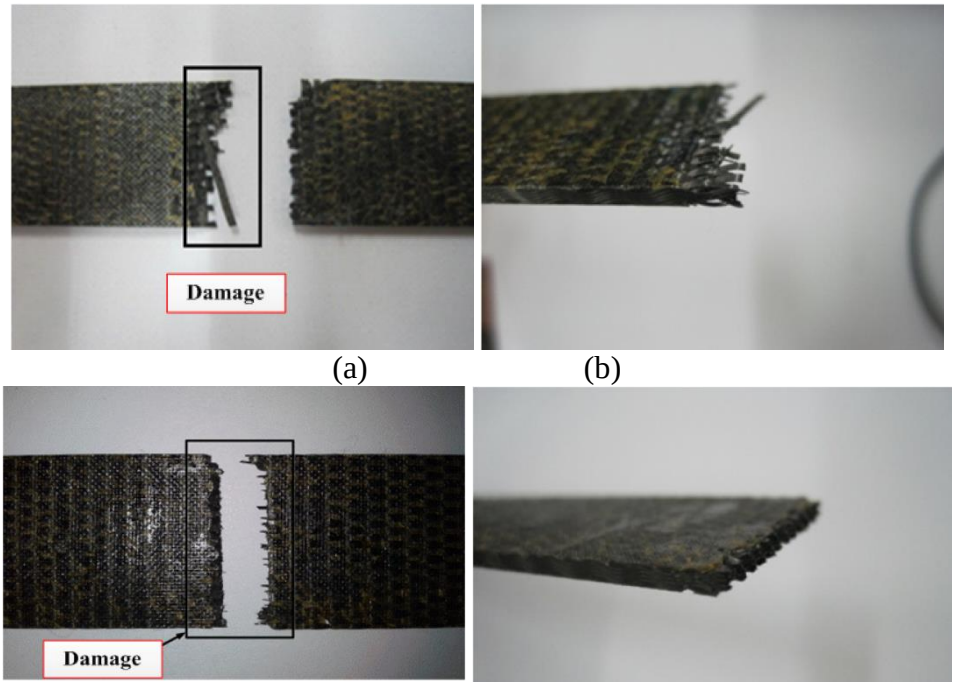


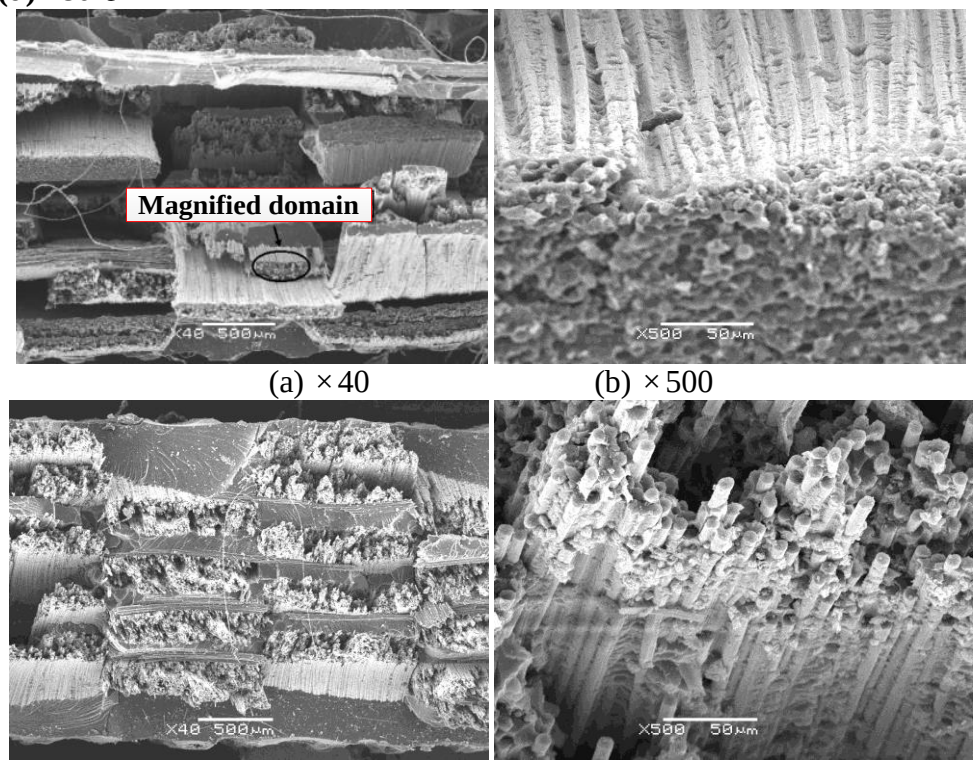
Figure 6. SEM photomicrographs of fracture surface taken from specimens subjected to static loadings, (a)-(c), RT, and (d)-(f), 180℃



(a) (b)

(c) (d)

Figure 7. Fracture surfaces of the fatigue composites subjected to 80% UTS. (a), (b) RT, and (c), (d) 180℃



(a) ×40 (b) ×500

(c) ×40 (d) ×500

Figure 8. SEM photomicrographs of the fatigue composites subjected to 80% UTS. (a), (b) RT, and (c), (d) 180℃

Conclusions

An investigation on the static and fatigue mechanical behavior of 2.5D woven composites at room and elevated temperatures was accomplished. The influence of temperature on the stress vs. strain curves, tensile modulus, strength, fatigue behaviors, stiffness degradation behaviors

and residual strength at RT and 180℃ were analyzed and discussed in detail. The damage mechanisms were revealed by observing the fractured morphology and measuring the residual tensile properties. Several useful conclusions were made as following:

(1) The results show the room temperature stress-strain curve has an initial linear behavior, followed by a non-linear feature, while the curves at 180℃ show an obvious non-linear feature. But both of the curves exhibit a brittle fracture feature.

(2) The fatigue life and fatigue strength at 180℃ decrease significantly compared with those at RT subjected to the same stress level. However, the residual strength at 180℃ can be strengthened by fatigue.

(3) The fracture morphology examinations indicate the damage and failure patterns of composites vary with the environmental temperatures. When the temperature is 180℃, there are little indication of large-scale debonding, but the presence of fiber pull-out is revealed by the brushy the bare fibers under the fatigue loading.

References

[1] Shimokawa, T., Kakuta, Y., Aiyama, T. (2008) Static and fatigue strengths of a G40-800/5260 carbon fiber/bismaleimide composite material at room temperature and 150,℃, *J Compos Mater* **6**, 55-79.

[2] Berthe, J., Brieu, M., Deletombe, E., Portemont (2014) Temperature effects on the time dependent viscoelastic behaviour of carbon/epoxy composite materials: Application to T700GC/M21, *Mater Design* **62**, 241-246.

[3] Ludovico, M. D., Piscitelli, F., Prota, A., Lavorgna, M., Mensitieri, G., Manfredi, G. (2012) Improved mechanical properties of CFRP laminates at elevated temperatures and freeze‐thaw cycling, *Constr Build Mater* **31**, 273-283.

[4] Bai, Y,, Keller, T., Vallée, T. (2008) Modeling of stiffness of FRP composites under elevated and high temperatures, *Compos Sci Technol* **68**, 3099-106.

[5] Selezneva, M., Montesano, J., Fawaz, Z. (2011) Behdinan K, Poon C. Microscale experimental investigation of failure mechanisms in off-axis woven laminates at elevated temperatures, *Composites Part A: Applied Science and Manufacturing* **42**, 1756-1763.

[6] Vieille, B., Taleb, L. (2011) About the influence of temperature and matrix ductility on the behavior of carbon woven-ply PPS or epoxy laminates: Notched and unnotched laminates, *Compos Sci Technol* **71**, 998-1007.

[7] Montesano, J., Fawaz, Z., Behdinan, K., Poon, C. (2013) Fatigue damage characterization and modeling of a triaxially braided polymer matrix composite at elevated temperatures, *Compos Struct* **101**, 129-37.

[8] Montesano, J., Fawaz, Z., Poon, C. (2014) Behdinan K. A microscopic investigation of failure mechanisms in a triaxially braided polyimide composite at room and elevated temperatures, *Mater Design* **53**,1026-36.

[9] ASTM D3039 (2008) Standard test method for tensile properties of polymer matrix composite materials, *ASTM International*

[10] ASTM D 3479 (2007) Standard test method for tension-tension fatigue of polymer matrix composite materials, *ASTM International*

[11] Zhu, Y, L. (2012) Research on prediction of damage failure and fatigue life for C/C composites. *Nanjing: Nanjing University of Aeronautics and Astronautics*.

[12] Cheng, Y. Z., Xuan, W. W., Yong, S. L., Bo, W., Dong, H. (2013) Tensile fatigue of 2.5D-C/SiC composites at room temperature and 900 C, *Materials and Design* **49**, 814-819.

[13] S.F. Shuler, J.W. Holmes, X. Wu, D. Roach (1993) Influence of Loading Frequency on the Room‐Temperature Fatigue of a Carbon‐Fiber/SiC‐Matrix Composite, *J Am Ceram Soc* **76,** 2327-2336.

[14] Liu, Z., Zhang, H., Lu, Z., Li, D. (2007) Investigation on the thermal conductivity of 3-dimensional and 4-directional braided composites, *Chinese Journal of Aeronautics* **20,** 327-331.

Predicting stability of a prototype un-bonded fibre-reinforced elastomeric isolator by finite element analysis

†Thuyet Van Ngo[1], *Anjan Dutta[2], and Sajal K. Deb[2]

[1]PhD student, Department of Civil Engineering, IIT Guwahati-781039, Assam, India.
[2]Professor of Civil Engineering Department, IIT Guwahati-781039, Assam, India.

*Presenting author: adutta@iitg.ernet.in
†Corresponding author: ngothuyetxd@gmail.com

Abstract

Fibre-reinforced elastomeric isolator (FREI) in an un-bonded application is an improved device for seismic mitigation of low-rise buildings. It is expected to reduce the cost, weight and provide easier installation in comparison to the conventional elastomeric isolator, which consists of elastomeric layers interleaved with steel plate as reinforcement. The horizontal response of un-bonded isolator is nonlinear due to rollover deformation and the horizontal stiffness is a function of both vertical load and horizontal displacement. Most previous studies have been focused to develop the model for predicting stability of the bonded conventional elastomeric isolators with low shape factors. In the present study, predicting stability of a prototype un-bonded FREI is presented based on the dynamic response utilizing finite element (FE) analysis. A prototype isolator is investigated under the variation of vertical loads and cyclic horizontal displacement to evaluate the performance and the effect of the vertical load on the behaviour of the isolator. FE analysis result shows that the critical load capacity of the isolator is significantly higher than the design vertical load, and the effective horizontal stiffness decreases with the increase in the vertical loads. Furthermore, the horizontal response of the isolator is also conducted under the design vertical load and increasing horizontal displacement up to $2.00t_r$ to observe the rollout instability.

Keywords: Fibre reinforced elastomeric isolator, un-bonded isolator, rollout instability, dynamic stability, buckling, critical load, analytical model.

Introduction

Seismic isolation is a well-known earthquake mitigation technique, where a layer of low horizontal stiffness is introduced between the foundation and superstructure. As a result, the natural period of vibration of the structure changes beyond the high-energy period range of earthquakes, and hence the seismic energy transferred to the structure is significantly reduced. Conventional steel reinforced elastomeric isolators (SREIs) have become a widely accepted technique in the structure over the past four decades for protecting the buildings from strong ground motion. They consist of alternating layers of rubber bonded to intermediate steel shims with two steel end plates at top and bottom. In general, SREIs are often applied for large, important buildings like hospitals and emergency centres, in countries such as Japan, New Zealand, United States, Mexico, Italy, etc. This limited use is largely due to the high material, manufacturing and installation costs. It is expected that the use of seismic isolators can be extended to ordinary low-rise housing if the weight and cost of the isolators are reduced. In view of this, fibre reinforced elastomeric isolators (FREIs) are proposed by replacing steel shims in conventional isolators by multi-layer of fibre fabric as reinforcement sheets to reduce their weight and cost. An un-bonded fibre reinforced elastomeric isolator (U-FREI) is a significant effort to improve FREI by removing two steel end plates and installing directly between the foundation and superstructure without any connection to these boundaries. Using U-FREI would reduce the weight and cost, easier installation, and can be made as a long strip and then easily cut to the required size. It means that the U-FREIs can be used for low-rise buildings subjected to earthquake loading in the developing countries.

The stability of elastomeric isolators is an important parameter for the design of seismic isolation systems. Elastomeric isolators are used in the structure to resist strong ground motion of earthquake with large displacement. Study on stability of elastomeric isolators refers to the determination of critical load carrying capacity while undergoing large horizontal

displacement. Generally, the critical load carrying capacity of isolator reduces with increasing horizontal displacement due to the reduction of the effective horizontal stiffness. The critical load in an elastomeric isolator is defined as the vertical load for which the horizontal stiffness is reduced to zero.

Procedures to evaluate critical loads of elastomeric isolators are based on an extension of Euler buckling load theory by Southwell [1932] to experimentally determine the buckling load in the flexible columns and a theoretical approach by Haringx [1948, 1949(a,b)] to predict the stability of rubber rods. Later, Buckle and Kelly [1986] carried out experimental studies to evaluate stability of SREIs under quasi-static loading using Southwell's procedure and under dynamic loading on a scaled model of bridge deck using shaking table test. Stable rollover of isolators could be observed in this study. These studies were however conducted with linear model and under small imposed displacement. In general, the behaviour of elastomeric isolators is nonlinear when subjected to large horizontal displacement under strong ground motion.

Some extensive analytical and numerical studies were performed to analyze the stability limit in elastomeric isolators and model their behaviour. Koh and Kelly [1989] proposed a two-spring mechanical model and visco-elastic stability model based on extension of Haringx's theory. The influence of vertical load on the horizontal stiffness of SREIs was evaluated. Stanton, et al. [1990] studied the stability of steel laminated elastomeric bearings using a modified linear model from Haringx's theory with configuration accounting for nonlinearity. When an elastomeric bearing was simultaneously subjected to vertical load and increasing lateral displacement, the shear force on bearing was observed to have passed through a maximum value. This point is the location of zero tangential stiffness, which is considered as the stability limit. Buckle and Liu [1993, 1994] experimentally determined the critical buckling behaviour of SREIs at high shear strains and proposed a simple reduced-area formula to estimate the critical load in bearings by overlapping area method. However, this method predicted a simple linear (for rectangular bearings) or nearly linear (for circular bearings) reduction in critical load with lateral displacement independent of material or geometric parameters of bearings. Actually, this reduction is not linear as observed in experimental tests. A nonlinear analytical model consisting of two-spring systems was proposed by Nagarajaiah and Ferrell [1999] in an effort to more accurately predict the critical load capacity of SREIs of different sizes and shape factors at a certain lateral displacement. The model was developed from two-spring model by Koh and Kelly with large displacement, large rotations and nonlinearities in shear and rotational stiffness of the bearing. The model was shown to predict a reduction in the critical load capacity with increasing lateral displacement, and the critical load capacity was not equal to zero at a lateral displacement equal to width of bearing. Buckle, et al. [2002] validated the nonlinear analytical solutions proposed by Nagarajaiah and Ferrell [1999] and determined the effect of lateral displacement on critical load by experimental tests with a series of low-shape-factor elastomeric bearings. Iizuka [2000] proposed a macroscopic model based on the two-spring model by Koh and Kelly, where the linear springs were replaced by nonlinear springs for predicting the stability of laminated rubber bearings at large deformations and under different vertical loads. The nonlinear parameters of the shear and rotational springs were determined from basic load test. Detailed nonlinear finite element analysis and an improved analytical formulation for predicting the reduced load-carrying capacity of bearings based on overlapping area method were also presented by Weisman and Warn [2012]. A recent study by Sanchez, et al. [2013] focused on experimental tests to examine the behaviour of steel reinforced elastomeric bearings at and beyond their stability limits. Three methods (two quasi-static tests and one dynamic loading test) were conducted to predict the stability limits of bearings and compared with the reduced-area formulation. Han, et al. [2013] proposed a modified analytical model based on the sensitivity analysis using Iizuka's model for the prediction of critical load capacity of bearings. Vemuru, et al. [2014] presented an enhanced analytical model based on a nonlinear analytical model by Nagarajaiah and Ferrell for application beyond stability limit. Thus, most previous studies were focused to improve the analytical model for predicting stability of elastomeric isolators and these models were developed for bonded conventional elastomeric isolators. Therefore, it is necessary to study on the stability of FREIs in un-bonded application.

As a result of un-bonded application, isolators undergo large deformation due to stable rollover under large horizontal displacements. Some regions of the top and bottom surfaces of isolator lose contact with the superstructure and substructure when the isolator is displaced horizontally. The reduction of the effective horizontal stiffness due to rollover deformation increases the seismic mitigation efficiency of isolator; but stability of isolator must be maintained. If an un-bonded FREI with a certain shape factor, S (defined as the ratio of the loaded area to load free area of a rubber layer) and aspect ratio, R (as the ratio of width to total height of the isolator) is able to achieve positive incremental load-resisting capacity during the course of cyclic loading, the isolator is assumed to be stable. On the other hand, the effective stiffness of an un-bonded isolator may also increase due to the initiation of contact between the vertical faces of the elastomer layers with the support surfaces, when they undergo very large displacement. Thus, a transition region between the decrease and increase in the effective stiffness is observed, and at certain value of displacement within this region, the increase in the effective stiffness of isolator due to contact exceeds the decrease in the stiffness due to rollover, and a hardening behaviour is occurred. This hardening behaviour observed in an un-bonded FREI is considered to be an advantageous characteristic since it can limit the maximum horizontal displacement of the isolation system in situations beyond the design basis seismic events. Studies related to the prediction of stability of un-bonded FREIs under cyclic loading were carried out experimentally by Raaf, et al. [2011]. In this study, authors proposed a method of fitting a polynomial to experimental force-displacement hysteresis data to predict the critical load capacity of isolator. This method was used to determine the fitted backbone curve and horizontal tangential stiffness. Additional studies for the buckling behaviour of un-bonded isolators were conducted using theoretical analysis by Kelly, et al. [2011, 2012].

From the above-mentioned literature review, it is observed that most of the models for predicting stability of elastomeric isolators are developed for conventional isolators in bonded application. There are very few studies for ascertaining the stability of FREIs in an un-bonded application. In addition, scaled sizes of elastomeric isolators were considered in these studies with low shape factors and aspect ratio, e.g. Nagarajaiah and Ferrell [1999], Buckle, et al. [2002] considered isolators with S = 1.67 to 10; Sanchez, et al. [2013] with S = 5.51 to 10.16; Han, et al. [2013] with S = 5 to 10.2; Vemuru, et al. [2014] with S = 10.64. Experimental studies were conducted for isolators with larger shape factors such as *Raaf, et al.* [2011] with S = 11 but for a scaled size of 70x70x24 *mm*; *Weisman and Warn* [2012] with S = 10 to 12. Therefore, it is necessary to carry out the studies for predicting the stability of a prototype U-FREI with high shape factor.

This paper presents studies related to predicting stability of prototype un-bonded FREI by FE analysis. Determination of the stability limit of an prototype isolator by experimental tests is relatively accurate, but it is difficult to investigate in laboratory due to constraints of experimental facility. In this study, predicting stability of a prototype un-bonded isolator is investigated by FE analysis and the accuracy of the response of the isolator under design vertical load and increasing horizontal displacement up to $0.89t_r$ (80 *mm*) is validated by comparing with the experimental results. A prototype FREI with size of 250x250x100 *mm*, shape factor of 12.5 and aspect ratio of 2.50 is investigated under the variation of vertical load and cyclic horizontal displacement to determine the critical load capacity and the effect of the vertical load on the behaviour of this isolator. Further, the horizontal response of the un-bonded isolator is also evaluated under the design vertical load and increasing horizontal displacement up to $2.00t_r$ (180 *mm*) to observe the rollout instability of the isolator.

Procedure for determination the critical load capacity of un-bonded FREI

As observed from literature survey, stability of an elastomeric isolator is evaluated based on the relation of shear force with horizontal displacement. The critical load capacity of the elastomeric isolator is defined as the vertical load for which the horizontal stiffness is reduced to zero (or zero tangential stiffness). When the elastomeric isolator is subjected to simultaneously the vertical load, P, and increasing horizontal displacement, u, shear force may pass through a maximum value, as illustrated in Fig. 1. The point of maximum shear force is considered the stability limit defined by the critical horizontal displacement, u_{cr}, and corresponding vertical load referred to herein as the critical load, P_{cr}.

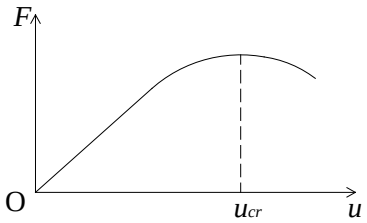


Fig. 1. Shear force versus horizontal displacement

From theoretical analysis, the critical load is defined as the point, where the shear force reaches a maximum value:

$$K_h = \frac{\partial F}{\partial u} = 0 \tag{1}$$

Using chain rule,

$$K_h = \frac{\partial F}{\partial u} = \frac{\partial F}{\partial P} \times \frac{\partial P}{\partial u} = 0 \tag{2}$$

where *F, u, P* are shear force, horizontal displacement and vertical load, respectively.
There is no requirement that $\partial F / \partial P$ must be equal to zero. Therefore,

$$\frac{\partial P}{\partial u} = 0 \tag{3}$$

where $\partial P / \partial u$ = derivative of the vertical load with respect to the horizontal displacement.

For a conventional elastomeric isolator in bonded application, the prediction of critical load capacity is often conducted by two quasi-static methods. In the first method, the isolator is subjected to a constant vertical load, *P*, and a monotonically increasing horizontal displacement, *u*, until the isolator reaches its stability limit ($K_h = 0$). The point of equilibrium is determined directly from shear force-horizontal displacement response as the point where the slope equals zero. The second method includes shearing the isolator to a constant horizontal displacement, *u* and applying monotonically increasing vertical load, *P*, while monitoring a reduction in shear force *F*. Repeating this procedure for different horizontal displacement levels provides unique equilibrium trajectories (*F vs P*) from which the point of neutral equilibrium, thus critical point (u_{cr}, P_{cr}) can be indirectly obtained.

However, for a FREI in un-bonded application subjected simultaneously to vertical load and horizontal dynamic displacements, the evaluation of critical load needs to be appropriately considered. Particularly for performance-based design, it is important to extend the theoretical understanding on the stability of isolators based on static/quasi-static methods to dynamic behaviour and enhance the ability to predict their response when subjected to extreme earthquake loading. Thus, it should use a dynamic method under cyclic loading to evaluate the critical load capacity of isolator in an un-bonded application.

In dynamic method, the un-bonded isolators undergo simultaneously a variation of the vertical load and cyclic horizontal displacement. Two important parameters such as the effective horizontal stiffness and damping factor are obtained from the hysteresis loops. The effective horizontal stiffness of isolator at a amplitude of horizontal displacement is defined as

$$K_{eff}^{h} = \frac{F_{\max} - F_{\min}}{u_{\max} - u_{\min}} \tag{4}$$

where, F_{max}, F_{min} are maximum and minimum value of the shear force,
u_{max}, u_{min} are maximum and minimum value of the horizontal displacement.

The equivalent viscous damping of isolator (damping factor, β) is computed by measuring the energy dissipated in each cycle (W_d), which is the area enclosed by the hysteresis loop. The formula to computed β is given by

$$\beta = \frac{W_d}{2\pi K_{eff}^h \Delta_{\max}^2} \tag{5}$$

where Δ_{max} is the average of the positive and negative maximum displacements.

Horizontal stiffness of an un-bonded FREI has two components, namely, horizontal secant stiffness and tangential stiffness. The present study is intended to determine the critical load at which the tangential stiffness becomes zero. In order to calculate the critical buckling load from the hysteresis loops obtained from the dynamic method, a curve is fitted to shear force-displacement hysteresis. According to the previous studies by Toopchi-Nezhad, et al. [2008] and Raaf, et al. [2011], a method of fitting a polynomial to shear force-displacement hysteresis data is developed. The fitted curve, denoted as backbone curve, represents an idealized evaluate of horizontal response of an un-bonded FREI with the damping forces removed (Fig. 2).

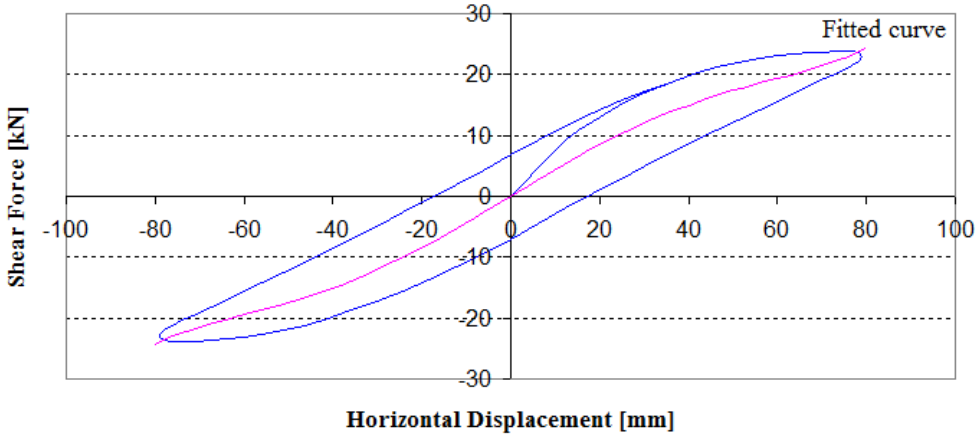


Fig. 2. Illustration of a fitted backbone curve in a hysteresis loop

The total horizontal load, $f_{b,i}$, experienced by the i^{th} isolator is described as:

$$f_{b,i}(t) = f_{sb,i}(t) + f_{db,i}(t) \tag{6}$$

where, $f_{sb,i}$ is stiffness force and $f_{db,i}$ is the corresponding force due to damping.

In a simple approach, the stiffness force can be modelled as a polynomial of order 5 given by:

$$\begin{aligned} f_{sb,i}(t) &= k_{b,i}(v_b(t)) \times v_b(t) = [b_0 + b_1 v_b(t) + b_2 v_b^2(t) + b_3 v_b^3(t) + b_4 v_b^4(t)] \times v_b(t) \\ f_{sb,i}(t) &= b_0 v_b(t) + b_1 v_b^2(t) + b_2 v_b^3(t) + b_3 v_b^4(t) + b_4 v_b^5(t) \end{aligned} \tag{7}$$

where, $v_b(t)$ is horizontal displacement and $k_{b,i}(v_b(t))$ is the horizontal secant stiffness as a function of horizontal displacement:

$$k_{b,i}(v_b(t)) = b_0 + b_1 v_b(t) + b_2 v_b^2(t) + b_3 v_b^3(t) + b_4 v_b^4(t) \tag{8}$$

The five parameters b_0 to b_4 are determined by applying a least squares fit to shear force-displacement hysteresis data.

The corresponding force due to damping, $f_{db,i}$ represents an idealized Rayleigh damping:

$$f_{db,i}(t) = c_{b,i}(t) \times \dot{v}_b(t) \tag{9}$$

where $c_{b,i}(t)$ is damping coefficient dependent on a equivalent viscous damping ratio ξ, tributary mass of structure on each isolator (m_i) and the horizontal secant stiffness $k_{b,i}(v_b(t))$:

$$c_{b,i}(t) = 2\xi\sqrt{k_{b,i}(v_b(t))m_i} \tag{10}$$

The tangential stiffness of the i^{th} isolator, $k_{tb,i}(v_b(t))$, as a function of horizontal displacement is

$$k_{tb,i}(v_b(t)) = \frac{df_{sb,i}(t)}{dv_b(t)} = b_0 + 2b_1 v_b(t) + 3b_2 v_b^2(t) + 4b_3 v_b^3(t) + 5b_4 v_b^4(t) \tag{11}$$

where the parameter b_0 is the tangential stiffness of the i^{th} isolator at $v_b(t)$=0.

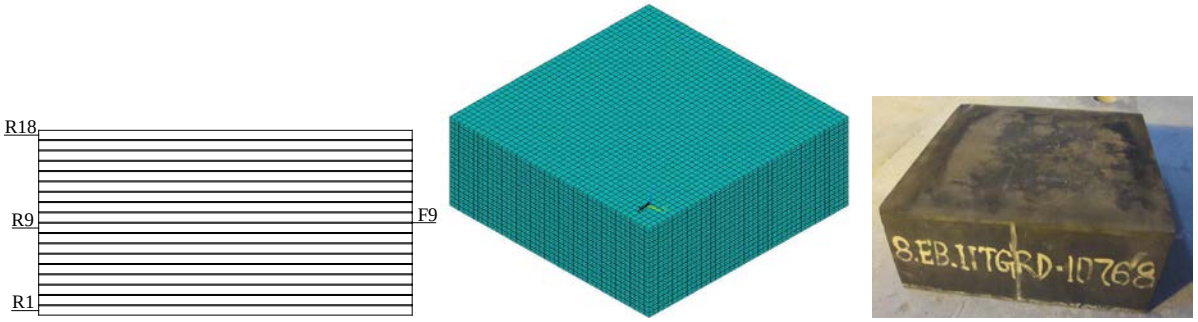
According to the remark of the previous study by Stanton, et al. [1990], the tangential stiffness at zero horizontal displacement in a shear force-displacement hysteresis is referred to as the transverse stiffness (K_t) of the isolator. The transverse stiffness is not necessarily the minimum tangential stiffness in every fully reserved hysteresis loop under constant vertical load. However, the transverse stiffness represents the tangential stiffness at which zero horizontal stiffness first occurs under increasing vertical load. The vertical load corresponding to a transverse stiffness of zero (K_t = 0) is defined as the critical buckling load under cyclic loading for an un-bonded FREI.

Prototype un-bonded fibre-reinforced elastomeric isolator

Prototype FREI considered in this study were manufactured by METCO Pvt. Ltd., Kolkata, India. These are already in use in an actual building in Tawang, India. Figure 3 shows the view of a typical prototype isolator with component layers and finite element model. The isolator comprises of 17 layers of fibre reinforcement sheets interleaved and bonded between 18 layers of rubber. Natural rubber and bi-directional ($0^0/90^0$) carbon fibre fabric are used in the isolator with the thickness of 5.0 and 0.55 *mm* for each layer of rubber and fibre, respectively. The physical dimensions and material properties of the isolator are shown in Table 1.

Finite element modelling

In this paper, fibre reinforced elastomeric isolator is numerically simulated using FE method in Ansys (v.14). The isolator is subjected to a variation of the vertical load and cyclic horizontal displacement to predict the stability of the isolator in an un-bonded application. FE analysis can address many issues which are rather difficult in closed-form solution. Analysis of isolator using FE method has some prominent advantages for the description of the detailed stress and strain of layers. Further, FE analysis can easily evaluate the response of the prototype isolator under high vertical load and large horizontal displacement, which is very difficult experimentally due to limitation of capacity in experimental facility.



a) Cross section of isolator b) FE model of isolator c) Prototype FREI

Fig. 3. The component of isolator

Table 1. Geometrical details and material properties of the square isolator

Description	Values
Size of specimen, *mm*	250x250x100
Number of rubber layer, (n_e)	18
Thickness of single rubber layer, (t_e), *mm*	5
Total height of rubber, (t_r), *mm*	90
Number of fibre layer, n_f	17
Thickness of single fibre layer, (t_f), *mm*	0.55
Shape factor, (S)	12.5
Aspect ratio, (R)	2.50
Initial shear modulus of elastomer, (G_o), *MPa*	0.90
Elastic modulus of carbon fibre reinforcement, (E_f), *GPa*	40
Poisson's ratio of carbon fibre reinforcement, (υ_f)	0.2

General description of the model

In this study, the isolator is modelled by elements having capabilities like large strain, incompressibility of material and nonlinear solution convergence. Incompressible material may lead to some difficulties in numerical simulation, such as volumetric locking, inaccuracy of solution, checkerboard pattern of stress distributions, or occasionally, divergence. Lagrange multiplier-based mixed *u-P* element is used to overcome incompressible material problems. These elements are designed to model material behaviour with high incompressibility such as fully or nearly incompressible hyper-elastic materials and nearly incompressible elasto-plastic materials (high Poisson's ratio or undergoing large plastic strain). Largange multipliers extend the internal virtual work so that the volume constraint is included explicitly. Further, an updated Lagrangian approach has been used in this study to update the local coordinate system on the deformed configuration of element when the isolator is subjected to very large horizontal displacement.

In the FE model of FREI, the elastomer is natural rubber which exhibits nonlinear behaviour. It is modelled using SOLID185 which is an eight-node structural solid element having three degrees of freedom at each node such as translations in the nodal x, y, and z directions. The fibre reinforcement is modelled using SOLID46 which is a 3-D eight-node layered structural solid designed to model layered thick shells or solid. Fibre-reinforcements are provided in the form of bi-directional ($0^0/90^0$) layers and bonded between rubber layers. Two rigid horizontal plates are considered at the top and bottom of the isolator to represent the superstructure and foundation. Vertical load and horizontal displacement are applied at the top plate which is allowed to move both in the vertical and horizontal directions, while all degrees of freedom of bottom plate are constrained. In order to study un-bonded FREI, surface-to-surface contact elements are used. Contact element CONTA173 is used to define the exterior rubber surfaces and target element TARGE170 is used to define the interior surface of top and bottom rigid plates. The contact element supports the Coulomb friction model to transfer the shear forces at the interface of contact and target surface. The model is meshed using hexagonal volume sweep.

Material models used for the rubber and fibre reinforcement

Material properties of isolator shown in Table 1 are used in FE model. Elastomer is modelled with hyper-elastic and visco-elastic parameters. Hyper-elasticity refers to materials which can experience large elastic strain that is recoverable. Rubber-like and many other polymer materials fall in this category. The constitutive behaviours of hyper-elastic materials are usually derived from the strain energy potentials. Further, hyper-elastic materials generally have very small compressibility. This is often referred to as incompressibility. Hyper-elastic materials have a stiffness that varies with the stress level.

In this study, Ogden 3-terms model has been adopted to model the hyper-elastic behaviour of the rubber which is characterized by shear (G_e) and bulk (k_e) modulus of the rubber and the vico-elastic behaviour is modelled by Prony Visco-elastic Shear Response parameter. The material parameters used are [Holzapfel , 1996].
Ogden (3-terms): $\mu_1 = 1.89\text{x}10^6$; $\mu_2 = 3600$; $\mu_1 = -30000$;
$\alpha_1 = 1.3$; $\alpha_2 = 5$; $\alpha_3 = -2$;

Details of input loading

The isolator is subjected to a variation of the vertical load to determine the effect of the vertical load on the dynamic properties and the predicting stability of un-bonded isolator under cyclic horizontal displacement. Elastomeric isolator is loaded simultaneously to the design vertical load of 350 *kN*, which is equal to the axial force in the column of the actual building and two fully reversed sinusoidal cycles of horizontal displacement of amplitude 80 *mm* ($0.89t_r$) (seen in Fig. 4) applied at the top steel plate. Amplitude of horizontal displacement is increased up to 135 *mm* ($1.50t_r$). The vertical load is subsequently increased and the process is repeated starting at the displacement amplitude of 80 *mm*. The complete simulation is considered for three displacement amplitudes of 80, 112.5 and 135 *mm* ($0.89t_r$, $1.25t_r$ and $1.50t_r$) and four vertical loads of 350, 550, 700 and 850 *kN*. In addition, the horizontal response of the un-bonded isolator is also conducted under the design vertical load of 350 *kN* and increasing horizontal displacement up to $2.00t_r$ (180 *mm*) to investigate the rollout instability.

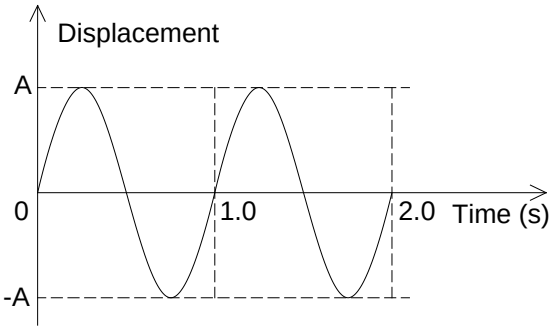
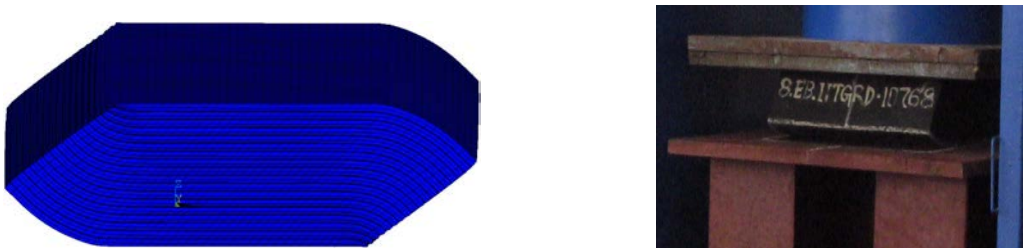


Fig. 4. Imposed horizontal displacement history versus time

Finite element model validation

For the finite element model validation, the numerical results are compared with experimental findings from test conducted at the structural laboratory in IIT Guwahati, India for a prototype un-bonded isolator. This specimen with the same size, component layers and material properties as given in Table 1 is checked here before using in an actual building in Tawang, India. In this test, the specimen is subjected simultaneously to a constant vertical load of 350 *kN* and three fully reversed sinusoidal cycles of horizontal displacement of amplitude 20 *mm*

($0.22t_r$), 40 *mm* ($0.44t_r$), 60 *mm* ($0.67t_r$) and 80 *mm* ($0.89t_r$). Comparisons of numerical and experimental results are conducted to evaluate the accuracy of FE model.



a) Deformed shape from numerical simulation b) Deformed shape from experiment

Fig. 5. Deformed shapes of an un-bonded isolator at displacement amplitude of 80 mm

Deformed shapes of isolator as obtained from both numerical and experimental result at the horizontal displacement amplitude of 80 *mm* are shown in Fig. 5. The top and bottom surfaces of un-bonded FREI exhibit stable roll off the contact surfaces without any damage and resulting in development of very low tensile stresses in that zone. This leads to reduction of the effective horizontal stiffness of the isolator. It can be seen from this Fig.5 that the deformed shapes of the isolator from FE analysis are observed to be in very good agreement with that from experimental test.

Fig. 6 shows the back bone curve for horizontal load-displacement relationships of the un-bonded isolator for displacement up to $0.89t_r$ (80 *mm*) as obtained from both experiment and FE analysis. Good agreement is observed between the experimental and FE analysis results. It can be seen from the figure, the horizontal load-displacement relation is nearly linear in the range of small displacement. Slope of this line is the effective horizontal stiffness of the isolator. When displacement increases, the response of un-bonded isolator becomes nonlinear due to the rollover. Consequently, the horizontal stiffness decreases with the increasing horizontal displacement. Fundamental period of un-bonded isolator thus increases with the decrease in stiffness, which result in increasing seismic mitigation capacity of isolator.

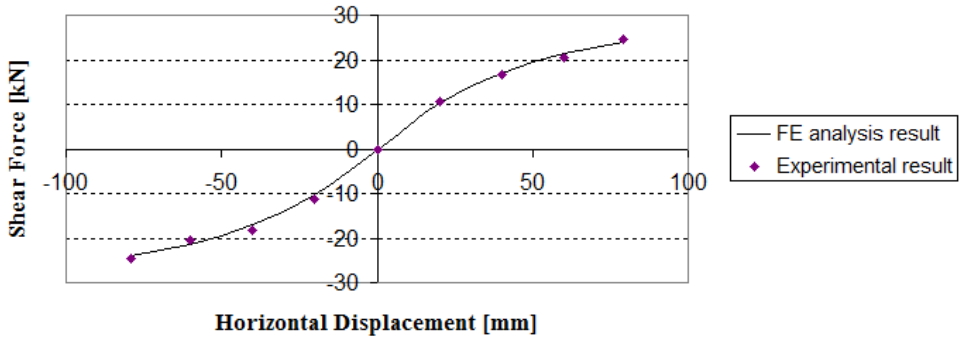


Fig. 6. Horizontal load versus displacement of the un-bonded FREI

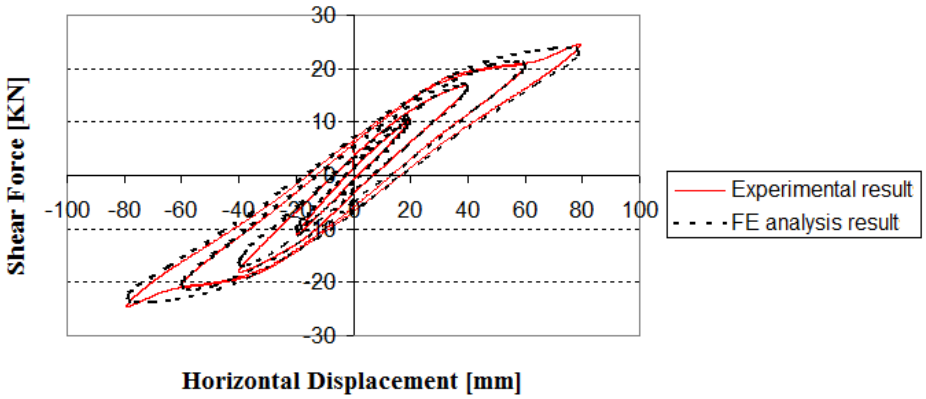


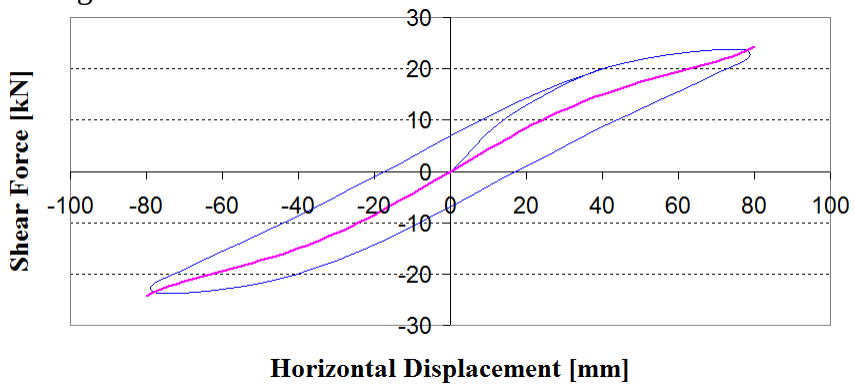
Fig. 7. Comparison of hysteresis loops for the un-bonded isolator by FEA and experimental results

Comparison of the hysteresis loops of the un-bonded isolator obtained as from experiment and FE analysis is presented in Fig. 7, which shows the discrepancy to be quite less. Thus, the adopted finite element analysis strategy is really effective in evaluating the dynamic response of un-bonded FREI under cyclic loading.

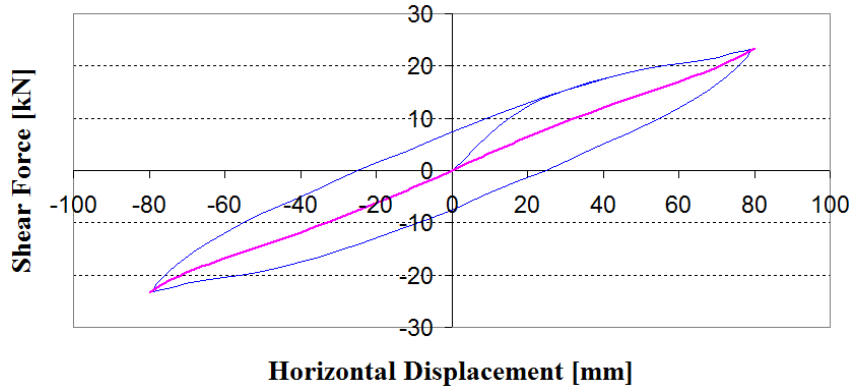
Finite element analysis and discussion

Critical buckling load capacity

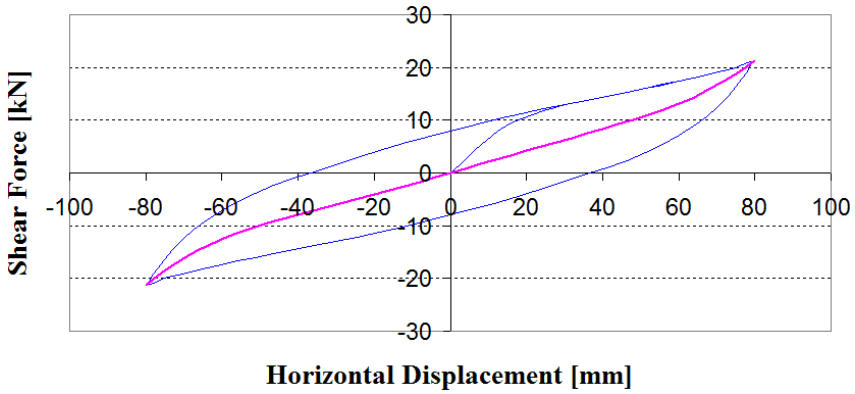
The objective of the dynamic stability analysis is to determine the critical buckling load at which the tangential stiffness becomes zero or the isolator would no longer be able to maintain positive incremental force resisting capacity. As noted above, the isolator is subjected to a variation of the vertical loads under cyclic horizontal displacement. According to the fitting method, the fitted backbone curves and corresponding hysteresis loops of the un-bonded isolator for each vertical load and displacement amplitude up to 80 *mm* as obtained from FE analysis are shown in Fig. 8. The fitted backbone curve is obtained from the average value of shear forces at any given horizontal displacements in the corresponding hysteresis loop and described by a polynomial. It can be seen from the figure, each cycle of FE analysis result maintains both symmetric and comparable hysteresis loops for all the vertical loads investigated. Similarly, considering other displacement amplitudes ($1.25t_r$ and $1.50t_r$), the fitted backbone curves of the isolator under different vertical loads (350, 550, 700 and 850 *kN*) are shown in Fig. 9.



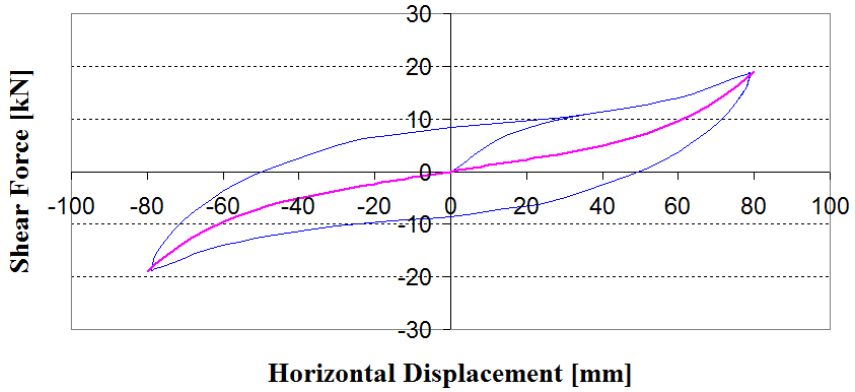
a) The vertical load of 350 kN



b) The vertical load of 550 kN

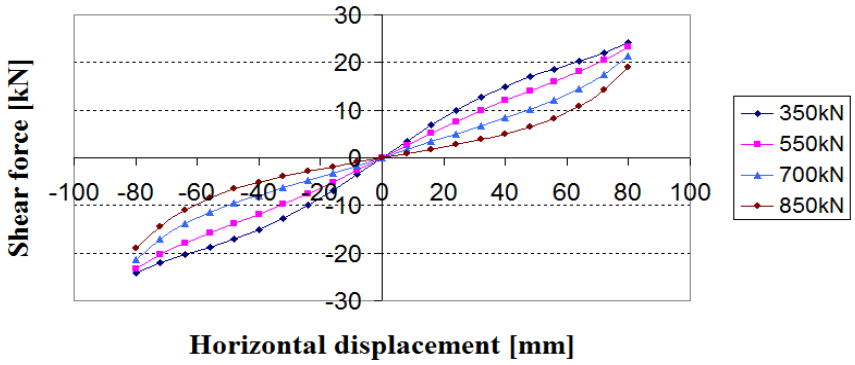


c) The vertical load of 700 kN

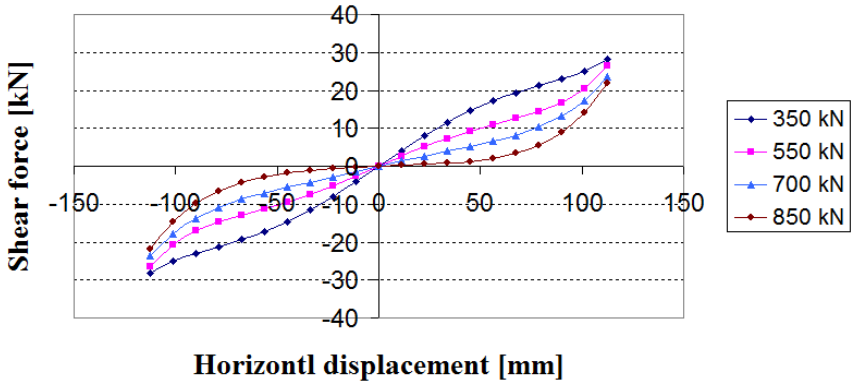


d) The vertical load of 850 kN

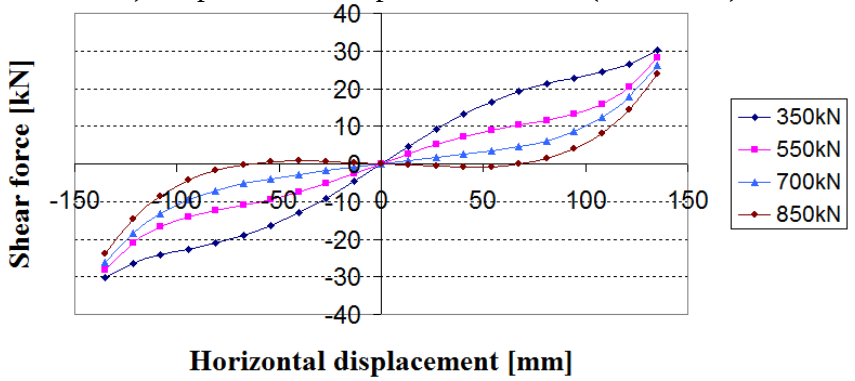
Fig. 8. Hysteresis loops with backbone curves of the isolator at displacement amplitude of 80 *mm*



a) Displacement amplitude of 0.89t_r (80 mm)

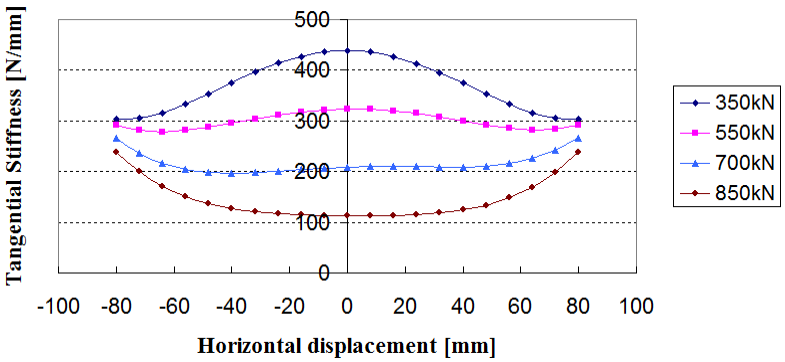


b) Displacement amplitude of 1.25t_r (112.5 mm)

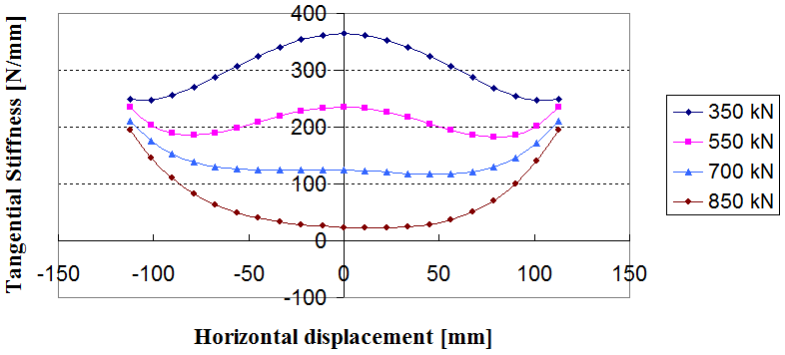


c) Displacement amplitude of 1.50t_r (135 mm)

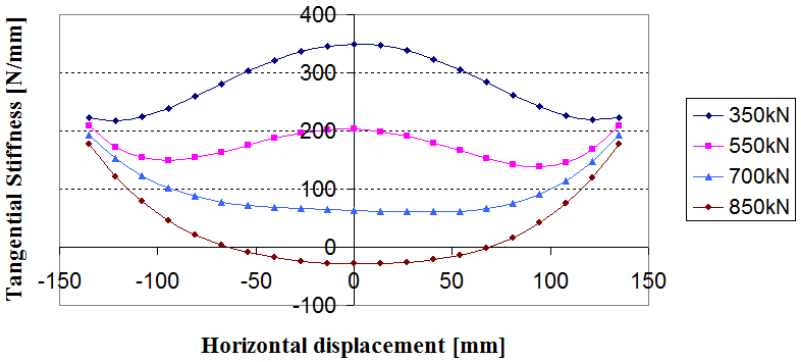
Fig. 9. Fitted backbone curves of the un-bonded isolator at the horizontal displacement amplitudes of 0.89t_r, 1.25t_r and 1.50t_r



a) Displacement amplitude of 0.89t_r (80 mm)



b) Displacement amplitude of 1.25t_r (112.5 mm)



c) Displacement amplitude of 1.50t_r (135 mm)

Fig. 10. Tangential stiffness obtained from the first derivative of the fitted backbone curve at the horizontal displacement amplitudes of 0.89t_r, 1.25t_r and 1.50t_r

The values of tangential stiffness results are evaluated from Eq. (11) and are presented in Fig. 10. The tangential stiffness at zero horizontal displacement, does not represent the minimum effective stiffness in a fully reversed sinusoidal cycle of horizontal displacement under low vertical loads of 350 and 550 kN. As the vertical load increases, the minimum slope of the backbone curve (tangential stiffness) occurs at zero horizontal displacement. At the large horizontal displacement and under large vertical loads, the transverse stiffness may acquire a negative value (Fig. 10c). Consequently, the vertical load corresponding to zero transverse stiffness is predicted by the approximation method.

As discussed above, the vertical load corresponding to zero transverse stiffness is defined as the critical buckling load for an un-bonded FREI. The points corresponding to zero transverse stiffness of the un-bonded isolator for different amplitudes of horizontal displacement as obtained by approximation method are shown in Fig. 11. As expected, the transverse stiffness decreases with the increase of the vertical load. The critical buckling loads are obtained from the points which have zero transverse stiffness. The relation of these critical buckling loads versus the horizontal displacement amplitude is shown in Fig. 12.

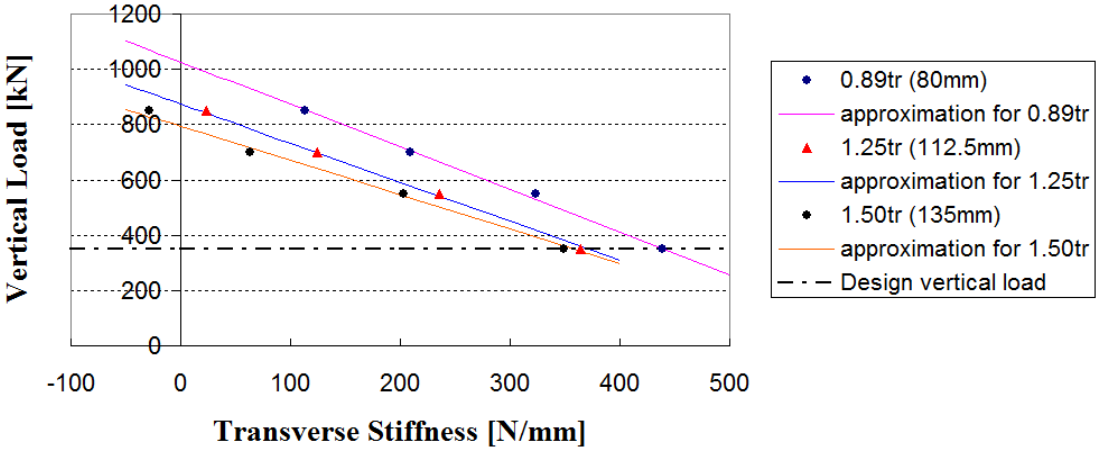


Fig. 11. Influence of the vertical load on transverse stiffness for the un-bonded isolator

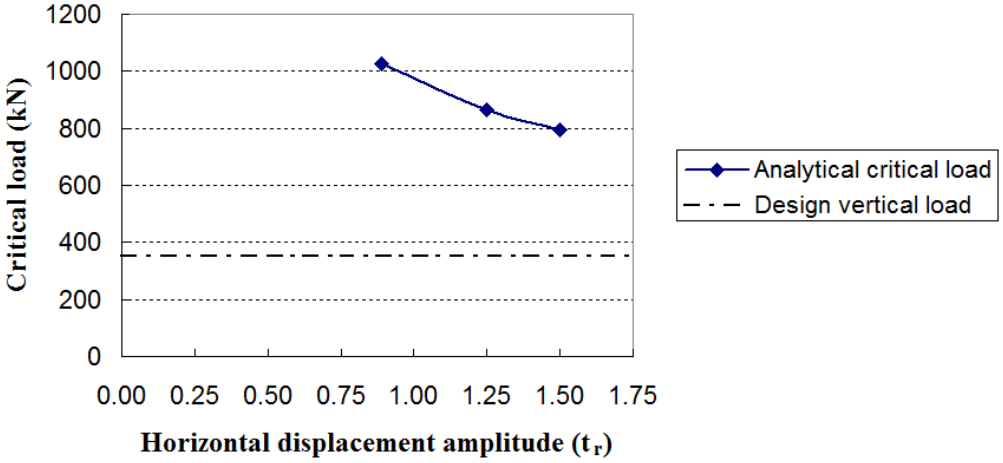


Fig. 12. The critical buckling load capacity of the un-bonded FREI.

It can be seen from the Fig. 12 that the critical buckling load decreases with the increase of the horizontal displacement amplitude, and it is relatively great at low displacement amplitude. The critical load capacity as obtained from FE analysis is significantly higher than the design vertical load, as example, the critical loads are found to be 2.9, 2.5 and 2.3 times higher than the design vertical load at displacement amplitude of $u = 0.89t_r$, $1.25t_r$ and $1.50t_r$ respectively. It is similar to the observation made by Raaf, et al. [2011] based on the experimental critical load carrying capacity of a scaled un-bonded isolator. From these results, it is thus realized that the prototype un-bonded specimen in the experimental tests didn't obviously show any sign of damage and susceptibility to buckling under the design vertical load.

The influence of vertical load on dynamic properties of the isolator

During the course of evaluation of critical load carrying capacity of the isolator, the effect of the vertical loads on the characteristic properties of the un-bonded isolator under cyclic horizontal displacement is also investigated. The effective horizontal stiffness and damping factor of the isolator under the variation of the vertical loads and amplitudes of displacement obtained from equation (4) and (5) are provided in Table 2 and plotted in Fig. 13.

Table. 2. Characteristic properties of un-bonded isolator

Vertical load (kN)	Amplitude of horizontal displacement					
	$0.89t_r$ (80mm)		$1.25t_r$ (112.5mm)		$1.50t_r$ (135mm)	
	K_{eff}^h (kN/m)	β (%)	K_{eff}^h (kN/m)	β (%)	K_{eff}^h (kN/m)	β (%)
350	301.67	13.46	247.09	14.58	222.03	15.42
550	288.09	15.61	233.67	16.92	209.04	17.87
700	267.53	18.19	218.00	19.90	189.40	21.04
850	238.72	21.69	194.13	24.45	165.19	27.00

It can be seen from Fig. 13 that the effective horizontal stiffness of the un-bonded isolator decreases, while the equivalent viscous damping increases with the increase in the vertical load at a given amplitude of horizontal displacement. The decreases of the effective stiffness are found to be 20.9%, 21.4% and 25.6% under the vertical load ranging from 350 kN to 850

kN at the displacement amplitudes of $0.89t_r$, $1.25t_r$ and $1.50t_r$, respectively. At a given vertical load, the effective horizontal stiffness decreases and the damping factor increases with the increasing horizontal displacement amplitudes. It is presented in more detail later. Despite the reduction in the effective horizontal stiffness at high vertical loads, the un-bonded isolator could maintain symmetric force-displacement hysteresis under cyclic loading.

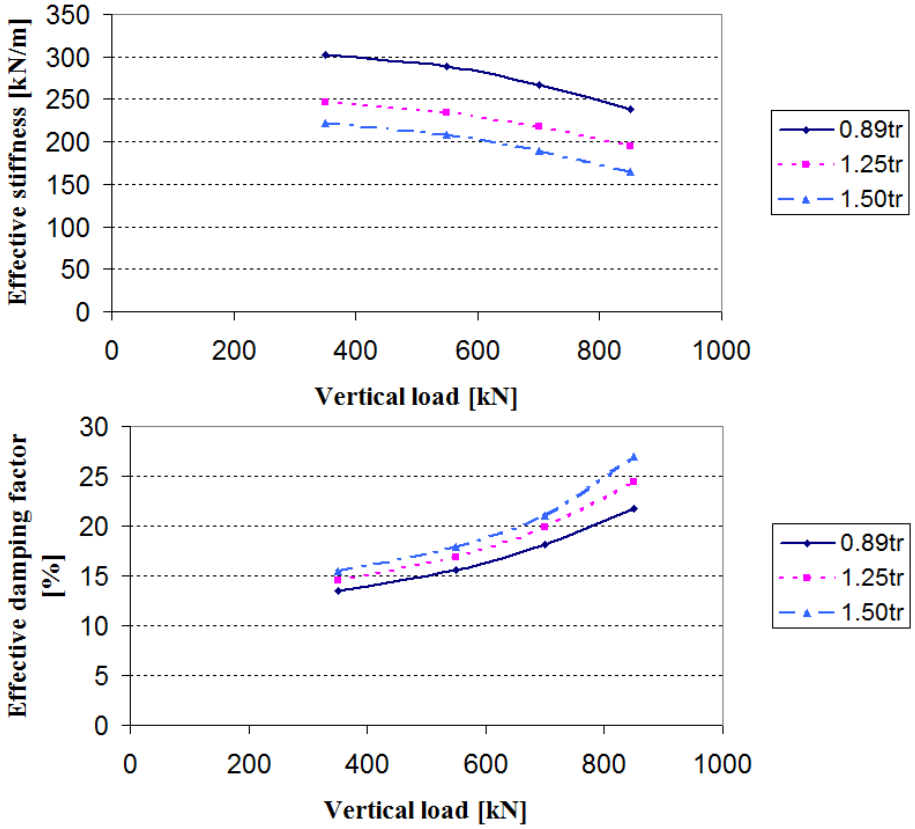


Fig. 13. The relation of the effective horizontal stiffness and damping factor versus vertical load

The rollout instability of the un-bonded FREI under design vertical load

As observed, the un-bonded isolator is not susceptible to buckling under the design vertical load at the amplitude of displacement less than $1.50t_r$. In this case, it is necessary to investigate the horizontal response of the un-bonded isolator under design vertical load of 350 kN and increasing horizontal displacement such that the original vertical faces of isolator establish full contact with the support surfaces, herein up to $2.00t_r$ (180 mm). At the large horizontal displacement, rollover deformation of the un-bonded isolator occurs and the rollout instability may be observed. Rollout is defined as the instability of a recessed isolator under shear displacement. The objective is to determine the horizontal displacement amplitude at which the tangential stiffness will be zero under design vertical load.

The shear force-displacement curve and horizontal secant stiffness-displacement relationship of the un-bonded isolator under the design vertical load and increasing horizontal displacement up to $2.00t_r$ are shown in Figs. 14 and 15. It can be seen from these figures that positive force resisting capacity is observed throughout the displacement range between zero to $2.00t_r$, and hence the isolator remains stable. Thus, the rollout instability of the un-bonded isolator is not observed here, although the results provide a shear profile having four stages of the horizontal response of the un-bonded isolator.

As observed in Fig.14, the horizontal stiffness of the un-bonded isolator is nearly linear under small horizontal displacement from zero to a displacement level at which the upper and lower contact surfaces of the isolator start to roll off the supports, denoted by u_r, is at 18 mm ($0.20t_r$). As the horizontal displacement is further increased, rollover deformation is observed in the isolator and the slope of force-displacement curve decrease to induce the reduction in the effective stiffness. At a certain displacement, portions of originally vertical faces of the isolator come in contact with the support surfaces. From these results from FE analysis, at $u = u_c = 1.40t_r$ (126 mm) the appearance of initial contact is observed. More numbers of originally vertical faces make contact with the support surfaces under the additional increase in horizontal displacement. At $u = u_f = 1.88t_r$ (169.2 mm), all the originally vertical faces of the un-bonded isolator are observed to be fully in contact with the supports. When displacement increases from u_r to u_c, the response of shear force-displacement is nonlinear, the effective horizontal stiffness of the isolator decreases due to rollover (seen in Fig. 15). Meanwhile, at the increasing displacement from u_f to $2.00t_r$, the effective stiffness of the isolator increases due to the contact between the originally vertical faces of isolator and the support surfaces. When the displacement changes in u_c to u_f range, the effective horizontal stiffness is affected by two things: a reduction due to rollover deformation and an increase due to the contact between the originally vertical faces of isolator and the support surfaces. Thus, there exists a transition point in the range of u_c and u_f in which the increase in the effective horizontal stiffness of the isolator due to contact exceeds the decrease in the stiffness due to rollover, and here a hardening behaviour is observed at displacement $u_h = 1.70t_r$ (153 mm). As seen from Fig. 15, the horizontal stiffness get the minimum value at the hardening point. At larger horizontal displacement $u > 2.00t_r$, the increase in horizontal stiffness is very less and the deformed shape of the isolator maintains full contact between the originally vertical faces of the isolator and the supports. The horizontal stiffness of the isolator is found to increase by approximately 32% as the horizontal displacement increases from u_h to $2.00t_r$. This hardening behaviour is advantageous as it can limit the horizontal displacement of the isolation system when subjected to extreme horizontal excitation events. The deformed shapes of the un-bonded isolator at different horizontal displacements as obtained from FE analysis results are shown in Fig. 16.

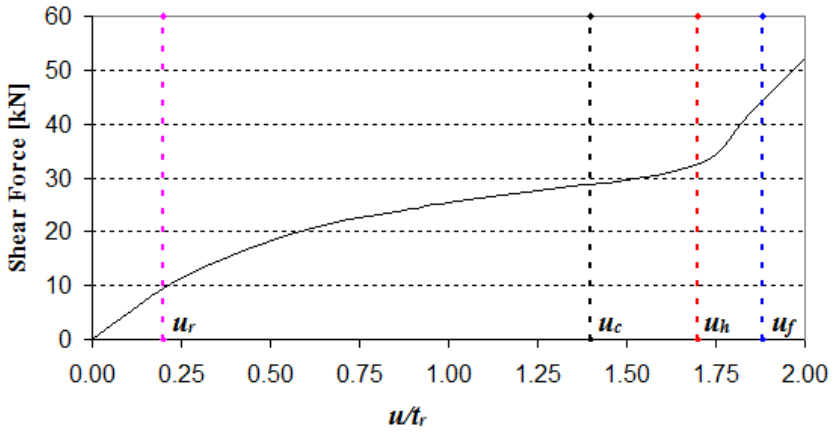


Fig. 14. Horizontal load–displacement curve of the un-bonded isolator.

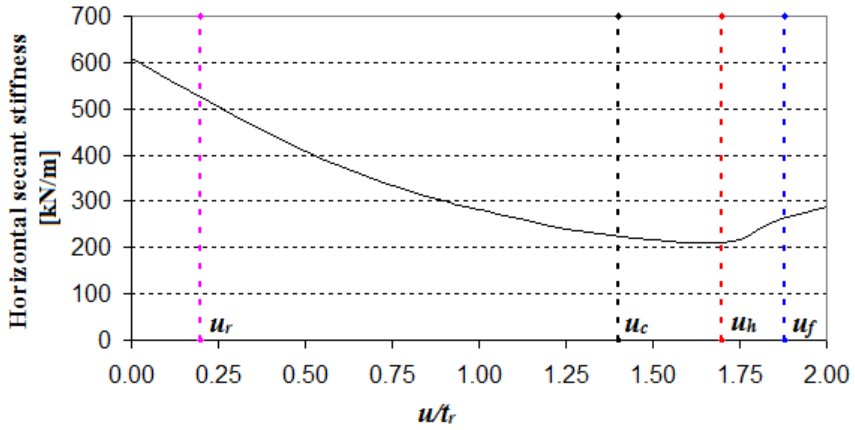
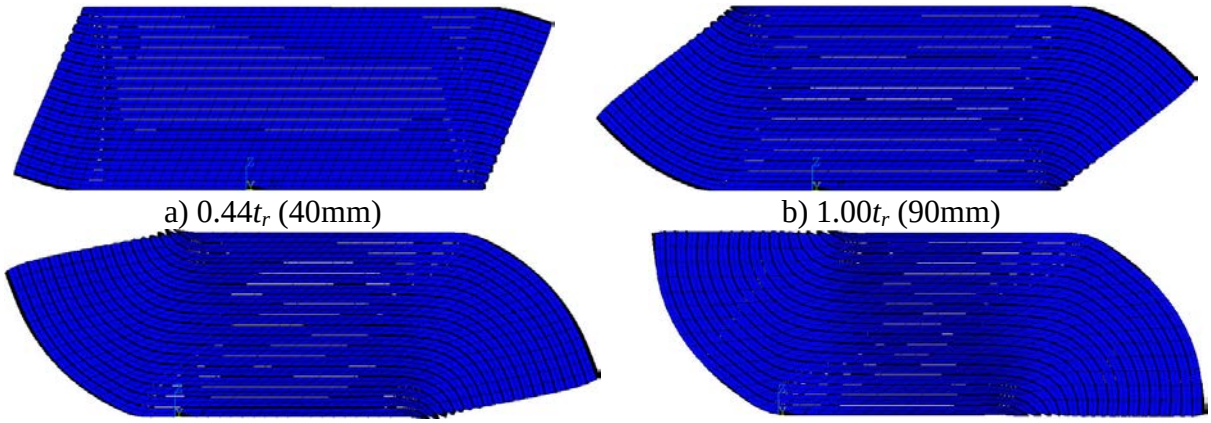


Fig. 15. Horizontal secant stiffness versus displacement



a) 0.44t_r (40mm) b) 1.00t_r (90mm)

c) 1.50t_r (135mm) d) 2.00t_r (180mm)

Fig. 16. Deformed shapes of un-bonded isolator obtained from FE analysis results

Conclusions

This paper presents the prediction of stability of a prototype un-bonded fibre-reinforced elastomeric isolator based on response from finite element analysis. The prototype isolators with the same dimensions, component layers and material properties are in use in an actual building in Tawang, India. Size of the isolator is 250 x 250 x 100 mm with the shape factor of 12.5 and aspect ratio of 2.50. In this study, the isolator is subjected to a variation of the vertical loads under cyclic horizontal displacement to determine the effect of the vertical load on the dynamic properties and the predicting stability of the isolator in an un-bonded application. In addition, the horizontal response of the isolator is also gradually increased to investigate the rollout instability under the design vertical load. The concluding remarks are as follows.

- The critical buckling load of the isolator as obtained by dynamic stability analysis corresponds to the point in which tangential stiffness is reduced to zero. The critical buckling load of the isolator decreases with the increase of the horizontal displacement amplitude.
- The critical load carrying capacity of the prototype isolator as obtained from FE analysis is significantly higher than the design vertical load. The critical loads are found to be 2.9, 2.5 and 2.3 times higher than the design vertical load at displacement amplitude of $u = 0.89t_r$, $1.25t_r$ and $1.50t_r$ respectively. It establishes the observation that the actual isolator in experimental testes didn't show any sign of damage and susceptibility to buckling under the design vertical load.

- The effective horizontal stiffness of the un-bonded isolator decreases, while the damping factor increases with the increase in the vertical load at a given amplitude of horizontal displacement.
- The effective horizontal stiffness of the un-bonded isolator decreases, while the damping factor increases with the increase in amplitude of horizontal displacement at a given value of applied vertical load.
- In the behaviour of the isolator under design vertical load, the effective horizontal stiffness decreases at the increasing horizontal displacement. However, under larger displacement up to $2.00t_r$ the horizontal stiffness starts to increase due to the contact between the vertical faces of the isolator with the support surfaces.

Acknowledgements

These authors would like to acknowledge the contribution of METCO Pvt. Ltd., Kolkata, India, for manufacturing FREI and Structural Engng. Laboratory, IIT Guwahati, India for extending facility for experimental investigation.

References

[1] Buckle I.G., Kelly J.M. [1986], "Properties of Slender Elastomeric Isolation Bearings During Shake Table Studies of a Large-Scale Model Bridge Deck", *Joint Sealing and bearing systems for concrete structures, ACI*, Detroit, Mich., Vol. 1, pp. 247–269.

[2] Buckle I.G., Liu H. [1993], "Stability of elastomeric seismic isolation systems", *Proc. Sem. Seismic Isolation, Passive Energy Dissipation and Active Control, Applied Technology Council*, Report ATC17-1, pp. 293-305.

[3] Buckle I.G., Liu H. [1994], "Experimental Determination of Critical Loads of Elastomeric Isolators at High Shear Strain", *NCEER Bulletin*, Vol. 8(3), pp. 1-5.

[4] Buckle I., Nagarajaiah S., Ferrell K. [2002], "Stability of Elastomeric Isolation Bearings: Experimental study", *Journal of Structural Engineering, ASCE*, Vol. 128(1), pp. 3-11.

[5] Han X., Kelleher C.A., Warn G.P., Wagener T. [2013], "Identification of the Controlling Mechanism for Predicting Critical Loads in Elastomeric Bearings", *Journal of Structural Engineering, ASCE*, Vol. 139(12), 04013016.

[6] Haringx J.A. [1948], "One highly compressible helical springs and rubber rods and their application for vibration-free mountings. I.", *Philips Res. Rep.*, Vol. 3, pp. 401-449.

[7] Haringx J.A. [1949a], "One highly compressible helical springs and rubber rods and their application for vibration-free mountings. II.", *Philips Res. Rep.*, Vol. 4, pp. 49-80.

[8] Haringx J.A. [1949b], "One highly compressible helical springs and rubber rods and their application for vibration-free mountings. III.", *Philips Res. Rep.*, Vol. 4, pp. 206-220.

[9] Holzapfel G.A. [1996], "On large strain viscoelasticity: Continuum formulation and finite element applications to elastomeric structures", *International Journal for Numerical Methods in Engineering*, Vol. 39, pp. 3903-3926.

[10] Iizuka M. [2000], "A macroscopic model for predicting large-deformation behaviours of laminated rubber bearings", *Engineering Structures, ELSEVIER*, Vol. 22(4), pp. 323-334.

[11] Kelly J.M. [1999], "Analysis of Fibre-Reinforced Elastomeric Isolators", *Earthquake Engineering Research Center, University of California, Berkeley, USA, JSEE*, Vol. 2(1), pp. 19-34.

[12] Kelly J.M., Konstantinidis D.A. [2011], "Mechanics of Rubber Bearings for Seismic and Vibration Isolation", *John Wiley & Sons, Ltd*, Publication.

[13] Kelly J.M., Calabrese A. [2012], "Mechanics of Fibre Reinforced Bearings", *PEER Report*, 2012/101, Pacific Earthquake Engineering Research Center, University of California, Berkeley, USA.

[14] Koh C.G., Kelly J.M. [1989], "Viscoelastic Stability Model for Elastomeric Isolation Bearings", *Journal of Structural Engineering, ASCE*, Vol. 115(2), pp. 285-302.

[15] Nagarajaiah S., Ferrell K. [1999], "Stability of Elastomeric Seismic Isolation Bearings", *Journal of Structural Engineering, ASCE*, Vol. 125(9), pp. 946-954.

[16] Osgooei P.M., Tait M.J., Konstantinidis D. [2014], "Finite element analysis of unbonded square fibre-reinforced elastomeric isolators (FREIs) under lateral loading in different directions", *Composite Structures, ELSEVIER*, Vol. 113, pp. 164-173.

[17] Raaf M.G.P.D, Tait M.J., Toopchi-Nezhad H. [2011], "Stability of Fibre-reinforced Bearings in an Un-bonded Application", *Journal of Composite Materials, SAGE*, Vol. 45(18), pp. 1873-1884.

[18] Sanchez J., Masroor A., Mosqueda G., Ryan K. [2013], "Static and Dynamic Stability of Elastomeric Bearings for Seismic Protection of Structures", *Journal of Structural Engineering, ASCE*, Vol. 139(7), pp. 1149-1159.
[19] Southwell, R.V. [1932], "On the analysis of experimental observations in problems of elastomer stability", *Proc. R. Soc. Lond. A*, Vol. 135(828), pp. 601-616.
[20] Stanton J.F., Scroggins G., Taylor A.W., Roeder C.W. [1990], "Stability of Laminated Elastomeric Bearings", *Journal of Engineering Mechanics, ASCE*, Vol. 116(6), pp. 1351-1371.
[21] Toopchi-Nezhad H., Tait M.J., Drysdale R.G. [2008a], "Testing and Modelling of Square Carbon Fibre-reinforced Elastomeric Seismic Isolators", *Structural Control and Health Monitoring*, Vol. 15, pp. 876-900.
[22] Toopchi-Nezhad H., Tait M.J., Drysdale R.G. [2008b], "A Noval Elastomeric Base Isolation System For Seismic Mitigation of Low-rise Buildings", *Proceedings of the 14th World Conference on Earthquake Engineering, October 12-17*, Beijing, China.
[23] Toopchi-Nezhad H., Drysdale R.G., Tait M.J. [2009a], "Parametric Study on the Response of Stable Unbonded-Fibre Reinforced Elastomeric Isolator (SU-FREIs)", *Journal of Composite Materials, SAGE*, Vol. 43(15), pp. 1569-1587.
[24] Toopchi-Nezhad H., Tait M.J., Drysdale R.G. [2009b], "Simplified Analysis of a Low-rise Building Seismically Isolated with Stable Un-bonded Fibre Reinforced Elastomeric Isolators", *Canadian Journal of Civil Engineering*, Vol. 36(7), pp. 1182-1194.
[25] Toopchi-Nezhad H., Tait M.J., Drysdale R.G. [2011], "Bonded versus Unbonded Strip Fibre Reinforced Elastomeric Isolators: Finite Element Analysis", *Composite structures, ELSEVIER*, Vol. 93, pp. 850-859.
[26] Vemuru V.S., Nagarajaiah S., Masroor A., Mosqueda G. [2014], "Dynamic Lateral Stability of Elastomeric Seismic Isolation Bearings", *Journal of Structural Engineering, ASCE*, Vol. 140(14), A4014014.
[27] Weisman J., Warn G.P. [2012], "Stability of Elastomeric and Lead-Rubber Seismic Isolation Bearings", *Journal of Structural Engineering, ASCE*, Vol. 138(2), pp. 215-223.

LARGE EDDY SIMULATION OF MIXED JET IN CROSSFLOW AT LOW REYNOLDS NUMBER

Jianlong Chang[1,2], †Guoqing Zhang[2] and Xudong Shao[3]

[1]College of Mechatronic Engineering, North University of China
[2]School of Aerospace Engineering, Beijing Institute of Technology
[3]Beijing Institute of Space Systems Engineering

*Presenting author:changjianlong1989@126.com
†Corresponding author: zhanggq@bit.edu.cn

Abstract: Large eddy simulation for a jet in crossflow at very low Reynolds number (Re=100) is performed for different jet-to-crossflow velocity ratios (r) ranging from 1 to 4.5, and the corresponding streamlines, vortex characters and interaction between the vortices have been analyzed. The results show that the streamlines for the jet in crossflow are closely related to the velocity ratios. The evolution of three-dimensional vorticity for displaying the formation of large-scale vortices has also been investigated. Near the nozzle of the jet, the stable mixed vortices including the counter-rotating vortex pair (CRVP), the horseshoe-vortex (HSV), the wake vortices (WV), the upright-vortices (UV) and the ring-like vortices come into being. The presence of the CRVP and RLV structures can maintain a quite long distance even to the flow exits at lower velocity ratio. However, the RLV are in destruction soon at larger velocity ratios under the interaction for the mixed jet in crossflow. Vortex evolution at velocity ratio r=1.5 have been displayed to explain the mechanism of the regular vortex under the interaction of UV and WV. The velocity streamlines have been obtained computationally and analyzed in details.

Keywords: jet in crossflow (JICF); large eddy simulation (LES); vortex interaction; vortex.

1. Introduction

A jet in crossflow (JICF) is an important flow phenomenon that is defined as the flow field where a jet of fluid enters and interacts with a crossflowing fluid [Muppidi (2007)]. There are various applications in the engineering problems such as the aerodynamic flow control, film cooling of turbines and combustors, control of separated flows over an airfoil, industrial mixing, and pollutant dispersion from effluent stacks [Lim (2006)].

In the past 70 years, numbers of experimental and computational research for JICF have been conducted. These researches mainly focus on the development and evolution of large-scale vortex structures, trajectory and other related flow phenomena. The interaction between the jet and the crossflow can generate the coherent vortex structures: the counter-rotating vortex pair (CRVP), the horseshoe-vortex (HSV), the wake vortices (WV), the upright-vortices (UV) and the ring-like vortices (RLV) [Cárdenas (2007)]. Fig. 1 has shown a side view of the corresponding vortex in the flow field. An experimental investigation on the

effects of jet velocity profiles on the flow field of a round jet in cross-flow (JICF) using laser-induced fluorescence and digital particle-image velocimetry techniques (DPIV) have been conducted [New (2006)] in 2006. The results had shown that the parabolic JICF not only can exhibit a faster velocity recovery, it can also register a higher magnitude of the peak average vorticity. The establishment [Camussi (2002)] of different behaviours at various velocity ratios is interpreted physically as an effect of the Reynolds number of the jet. This means that the Reynolds number has an essential effect on the destabilization mechanisms for the formation of the mixed vortex. The CRVP undulates in the turbulent flow and interacts with the intermittent wake vortices which in turn interact with the boundary layer and the vortices therein [Salewski (2008)].

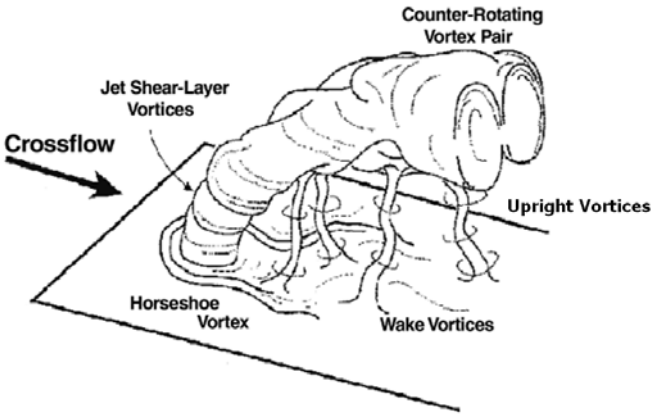


Fig. 1 Vortex Structures of a JICF [Jouhaud (2007)]

A new model [Mashayek (2011)] for atomization of a turbulent liquid jet in a subsonic crossflow had been developed. The corresponding results had shown that the droplets stripping apart from the jet body can make a great contribution to the formation of the vortical structures along the wake of the jet. It was also shown that the spreading of the jet into a sheetlike shape strengthened the extent of the vortical structures in the JICF, which will affect the droplet dynamics downstream of the jet. The problem of the proper choice of the turbulent Schmidt number in the Reynolds-averaged Navier-Stokes (RANS) jet in crossflow mixing simulations had been summarized [Ivanova (2013)]. The mainly conclusion was that the turbulent Schmidt numbers ranging from 0.2~0.3 used in JICF simulations for obtaining the optimal mixing predictions were not in agreement with the physical reality. More accurate prediction of mixing in JICF is significantly important to the development of combustion systems. The turbulent mixing of a jet in crossflow performed at the Reynolds number of 6930 by using the large eddy simulation method [Esmaeili (2015)]. The velocity profile for the jet pulsation substantially affected the JICF structures, and relatively low Strouhal numbers ranging from 0.0075 to 0.05 can develop an optimal condition for mixing, entrainment and penetration in the corresponding JICF. The effects of pulsing of high-speed subsonic jets (Ma=0.47~0.77) on mixing and jet trajectory in turbulent subsonic crossflows by using large-eddy simulation had been investigated [Srinivasan (2012)]. The regime of pulsed JICF had shown both the similarities and differences to the earlier experimental work. At the larger Strouhal number, the vortex interaction will still increase. However, the vortex ring will

be broken down in a relative short time, resulting in reduced penetration at a larger Strouhal number.

One leading parameter determining the development and evolution of large-scale vortex structures in JICF is the velocity ratio *r* [Yuan (1999)], if the densities in the jet and the crossflow are the same, *r* can be defined as:

$$r = \frac{V_{jet}}{V_{crossflow}} \tag{1}$$

otherwise, the effective velocity ratio *r* can be obtained by the square root of the momentum flux ratio as [Gutmark (1999)].

$$r = \sqrt{\frac{(\rho V^2)_{jet}}{(\rho V^2)_{crossflow}}} \tag{2}$$

The simulation and analysis for a jet in crossflow at very low Reynolds number (*Re*=100) will be performed for different jet-to-crossflow velocity ratios (*r*) ranging from 1 to 4.5.

2. Numerical methods and flow configuration

2.1.*Numerical methods*

At present research, the large eddy simulation (LES) has been adopted, because the LES turbulence model is different from Reynolds Averaged Navier-Stokes Equation (RANS) and Direct Numerical Simulation (DNS). The aim of the LES is to resolve the large scale of turbulence, and the smaller ones are modeled based on the universality. By filtering process in the large eddy simulation, the vortices less than a certain scale are filtered from the flow field, large eddy is calculated firstly. Then the solution of small eddy by solving additional equation will be obtained. Consequently, LES is more suitable for industrial configurations, in which large scales are known to be essential. In the case of the JICF, the unsteady behavior of the various flow structures is expected to be more important, the unsteady LES approach that provides spatiotemporal resolution should be used [Jouhaud (2007)].

In the LES method, the whole flow will be divided into large-scale eddy and small-scale eddy. Basic equations of LES are obtained after filtering Navier-Stokes equation and the continuity equation [Chen (2010)]:

$$\frac{\partial(\rho\overline{u_i})}{\partial t} + \frac{\partial(\rho\overline{u_i}\overline{u_j})}{\partial x_j} = -\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial}{\partial x_j}(\mu\frac{\partial \overline{u_i}}{\partial x_j}) - \frac{\partial \tau_{ij}}{\partial x_j} \tag{3}$$

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho\overline{u_i})}{\partial x_i} = 0 \tag{4}$$

where ρ is the density of fluid , u_i and u_j are the velocity components, p is the pressure , μ is the kinematic viscosity coefficient, the variables of formula with an overline are the field variables filtered.

Component of subgrid-stress tensor (SGS) is obtained as $\overline{\tau_{ij}} = -\rho(\overline{u_i u_j} - \overline{u_i}\,\overline{u_j})$. And u_i is defined as $u_i = \overline{u}_i + u_i'$, therefore the SGS can be decomposed into three parts:

$$\begin{aligned} \tau_{ij} &= \overline{u_i u_j} - \overline{u_i}\,\overline{u_j} = \overline{(\overline{u}_i + u_i')(\overline{u}_j + u_j')} - \overline{u_i}\,\overline{u_j} \\ &= \overline{\overline{u}_i \overline{u}_j} - \overline{u_i}\,\overline{u_j} + \overline{\overline{u}_i u_j'} + \overline{u_i' \overline{u}_j} + \overline{u_i' u_j'} = L_{ij} + C_{ij} + R_{ij} \end{aligned} \tag{5}$$

where, L_{ij} is Leonard stress of the SGS, which can be obtained by $L_{ij} = \overline{\overline{u}_i \overline{u}_j} - \overline{u_i}\,\overline{u_j}$. $C_{ij} = \overline{\overline{u}_i u_j'} + \overline{u_i' \overline{u}_j}$, and it is named cross stress of the SGS. R_{ij} captured by $R_{ij} = \overline{u_i' u_j'}$ represents the Reynold stress of the SGS. L_{ij} shows the motion effect among the solvable large eddy, C_{ij} stands for the motion effect between the solvable large eddy and the unsolvable small eddy, and R_{ij} is the interaction among the unsolvable small eddy, respectively.

Based on the assumption of Boussinesq, the relationship between $\overline{\tau_{ij}}$ and $\overline{S_{ij}}$ can be expressed as:

$$\overline{\tau_{ij}} - \frac{1}{3}\delta_{ij}\overline{\tau_{kk}} = -2\mu_T \overline{S_{ij}} \tag{6}$$

where μ_T is turbulent viscosity, δ_{ij} is Kroneker symbol, $\overline{S_{ij}}$ is strain rate tensor after filtering,

$$\overline{S_{ij}} = \frac{1}{2}\left(\frac{\partial \overline{u_i}}{\partial x_j} + \frac{\partial \overline{u_j}}{\partial x_i}\right) \tag{7}$$

Turbulent viscosity μ_T can be configured as product between length scale l and velocity scale q. By assuming that the magnitude of small-scale is in equilibrium, length scale and velocity scale can be defined as $l = C_s\overline{\Delta}$, $q = \overline{\Delta}\left|\overline{\mathbf{S}}\right|$, then the turbulent viscosity μ_T can be expressed as:

$$\mu_t = l^2 \left|\overline{\mathbf{S}}\right| \tag{8}$$

where C_s is constant of Smagorinsky, the approximation of the constant is

$$C_s \approx \frac{1}{\pi}\left(\frac{3C_k}{2}\right)^{-3/4} \tag{9}$$

The value measured in the atmosphere for Kolmogorov constant is 1.4, thereby $C_s \approx 0.18$. However, the value of C_s is usually taken as 0.1 in practical application. $\overline{\Delta}$ is the scale of grid filter, and it is obtained by $\overline{\Delta} = (\Delta x \Delta y \Delta z)^{1/3}$. For unstructured grids, $\overline{\Delta}$ could be acquired by extracting a cube root for the unit volume. $\overline{\mathbf{S}}$ can be captured by

$$\left|\overline{\mathbf{S}}\right| = \sqrt{2\overline{\mathbf{S}}_{ij}\overline{\mathbf{S}}_{ij}} \tag{10}$$

2.2.*Flow configuration and grid distribution*

Fig.2 shows the flow configuration and the size of the computational domain. The flow region is rectangular, the length, breadth and height are 65D×20D×16D. The jet channel is circular and Reynolds number of the fluid is 100, D presents the diameter of round jet. The structured grids will be adopted in the calculation, and the corresponding mesh have also been refined near the entrance of jet.

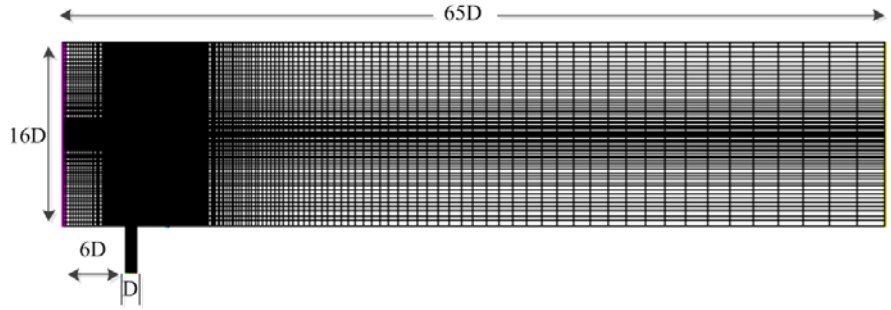


Fig. 2 View of the grid

3. Results and discussion

3.1.*Time-averaged particle trajectory*

The connection of the jet center has been defined as the time-averaged particle trajectory for the JICF. Fig.3 shows different trajectories obtained from the stream traces. These time-averaged particle trajectories show that by the action of the crossflow the jet is deflected downstream. In the proximity of the jet exit, it is noticed that the trajectories are almost vertical up to the main flow indicating that the ability of vertical penetration for the cases is about the dimension of the jet [Saha (2012)]. As shown in Fig.3, the jet has a larger kinetic energy and a stronger penetration compared with the crossflow near the nozzle, therefore the jet can quickly flow across the boundary layer. Once the jet reaches the crossflow in the flow channel, it will experience drastic exchange of energy and momentum, and penetration capability of the jet decreases leading to a deflected jet thereby. This phenomenon can produce more complex vortices, such as the counter-rotating vortex pair, upright vortex, vortex ring, horseshoe vortex and wake vortices, respectively.

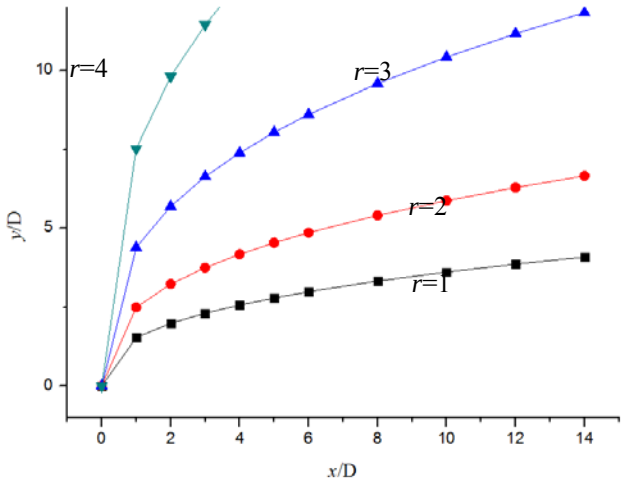


Fig. 3 Time-averaged particle trajectories for the JICF

Comparison of various cases in Fig.3, it has revealed that the jet penetration is highest for the highest r and it is almost followed by velocity ratio r. The upstream for the inflow with boundary layer near the nozzle can generate more kinetic energy loss. This will result in the higher pressure gradient in the vertical direction which rises fluid upwards to a higher extent. Thereby, the more momentum loss, the deeper penetration of jet in crossflow will reach for a given jet profile.

The time-averaged particle trajectories in the JICF can be approximately expressed as [Chassaing (1974) and Camussi (2002)]:

$$y = A(x)^n \tag{11}$$

where y is the height of the jet penetration, x is the streamwise position. In the presented cases, the corresponding A and n are listed in Tab. 1. The studied cases in this paper have shown a good agreement with the experiment results (Camussi 2002) within 5% error. This demonstrates that the large eddy simulation has higher accuracy.

Table 1 Coefficient A and n

veolicity ratio	A	A (Camussi 2002)	relative error of A (%)	n	n (Camussi 2002)	relative error of n (%)
1	1.5377	1.5962	3.66%	0.3708	0.39	4.92%
2	2.4926	2.5158	0.92%	0.3723	0.39	4.54%
3	4.3904	4.566	3.85%	0.3755	0.39	3.72%
4	7.5213	7.865	4.37%	0.3821	0.39	2.03%

3.2.*Evolution of three-dimensional vortex*

Fig.4 is an overall view of vortex structure in JICF at Reynolds number Re=100 with a velocity ratio r=1.5. It is apparent that the spatial evolution for large-scale such as CRVP and RLV can be clearly observed, the HSV, UV and WV have been also shown below the large scale vortices (CRVP and RLV). Near the nozzle, the CRVP and RLV have appeared due to

the action of shear layer. While once produced, the vortices will not immediately fall off, but stretch along the flow at a certain frequency. The CRVP and RLV generate a gradually rising in the interaction among the small scale vortices (HSV, UV and WV) and boundary layer. As shown in Fig.4(a), the vortices are generated near the nozzle, with the increasing distance from the entrance of the jet, scale and strength of vortex rings are enhancing (Fig.4(a)~(e)). With the further development of the JICF, the intensity of vortices will be evolved stronger, the CRVP and RLV start to fall at a certain time after deformation and distortion. The falling frequency of vortices is much faster than the formation. After completing the process of falling, the JICF will generate more stable CRVP far away from the nozzle (Fig.4(f)).

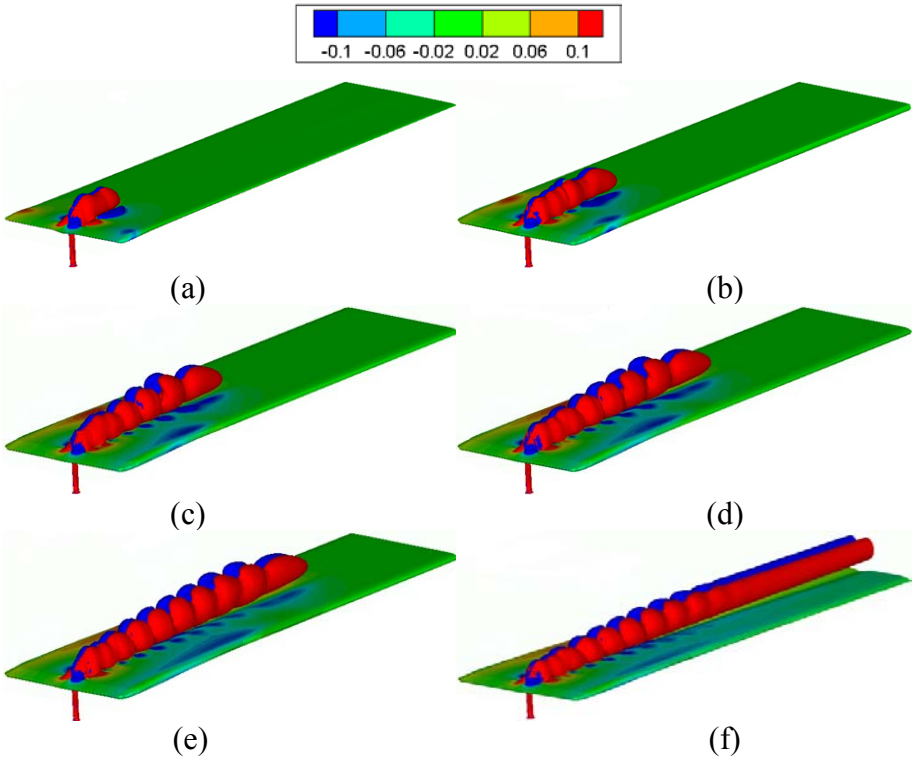


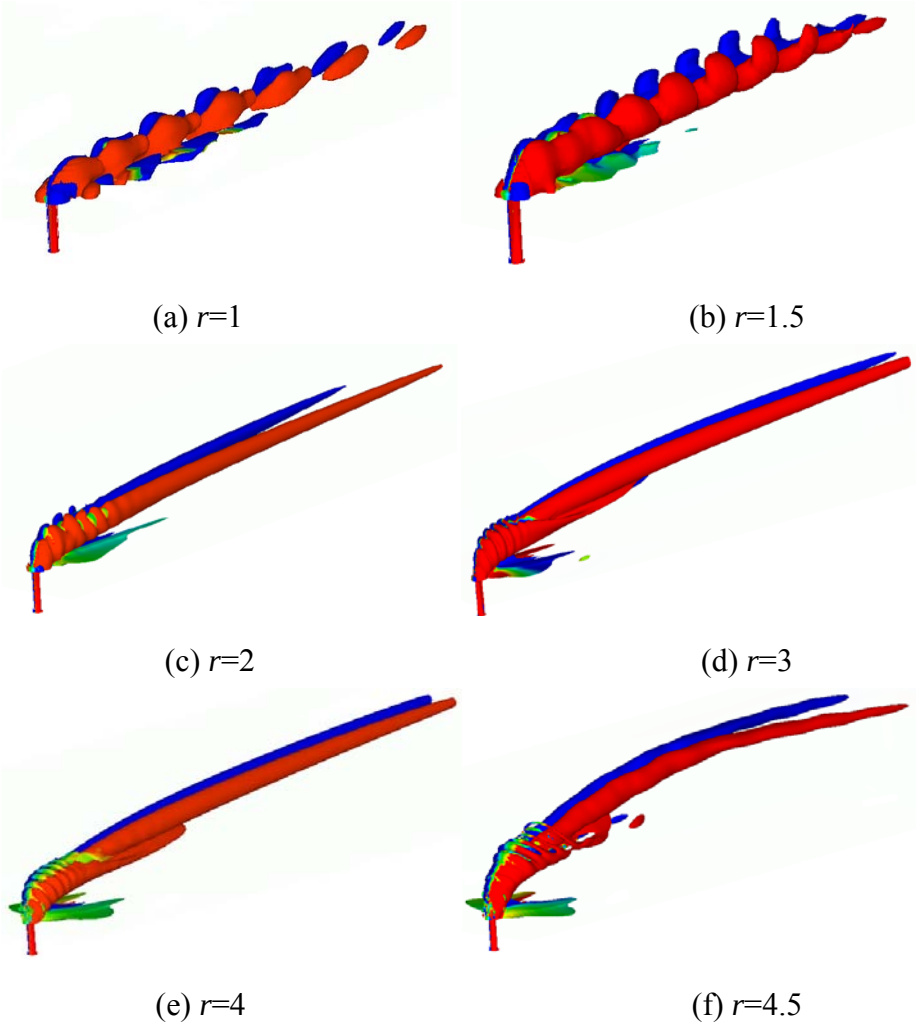
Fig. 4 Evolution of vorticity for JICF with a circle nozzle
(a) t=1.9s; (b) t =3.74s; (c) t =6.24s;(d) t =7.8s; (e) t =9.86s; (f) t =15.72s

3.3.*Analysis of three-dimensional vorticity*

After flowing into the crossflow, the jet will generate three processes under the interaction of the jet and crossflow: the initial phase, the curved phase and penetration phase. The JICF will make an access to the full development phase after shearing, wrapping and other effects. The small-scale structure in the jet core area surrounding is ongoing for stretch, rupture merged into the large scale vortices. The CRVP and RLV will be broken down into crossflow. Far away from the entrance, the previous vortices will gradually decline and evolve into CRVP at higher r.

The evolution and interactions for the vortices structures are significantly affected by the variations of velocity ratio. In particular, the vortex content of the CRVP and RLV are

analyzed, showing a vortex flow phenomena which strongly depends on *r*. The most important effect of *r* on the flow behavior is the changing of the CRVP and RLV structures.



(a) *r*=1 (b) *r*=1.5

(c) *r*=2 (d) *r*=3

(e) *r*=4 (f) *r*=4.5

Fig. 5 Three-dimensional vorticity

As shown in Fig.5, the presence of CRVP and RLV structures can last quite a long distance even to the flow exits at lower velocity ratio (Fig.5(a) and (b)). The interval between CRVP and RLV is relatively large. However, when the velocity ratio becomes larger (Figs.5(c)~(f)), the RLV could only maintain a short distance. Forming frequency of RLV increases with the increasing velocity ratio *r*. While with the augment of velocity ratio, the jet kinetic energy increases, gap of the RLV near the nozzle will be generated closer, and the diameter for the RLV will become smaller. However, due to the strong interaction of the WV, HSV, UV, RLV and shear layer, the RLV will be destroyed soon.

3.4.*Vortex Interaction*

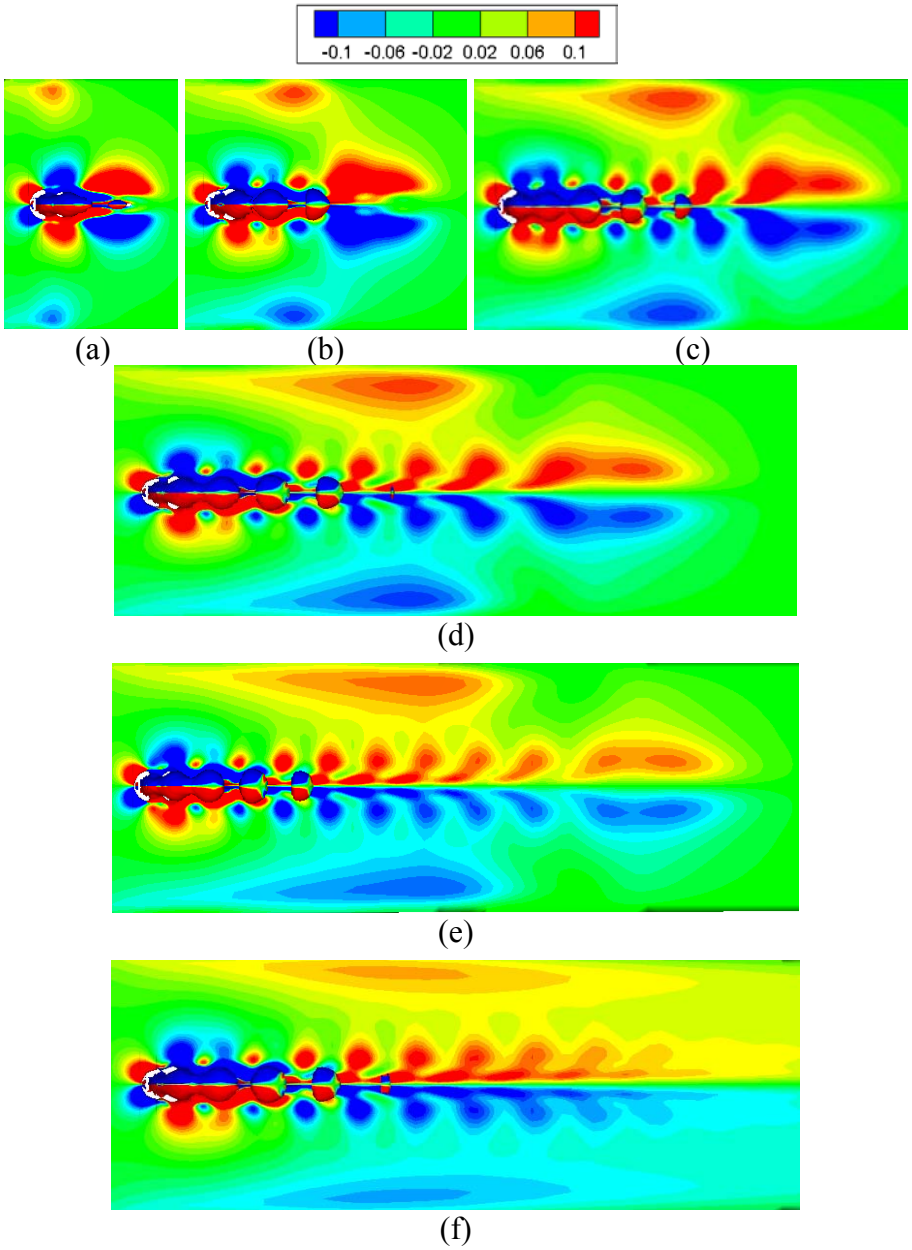


Fig. 6 Evolution for interaction of HSV, WV and UV at r =1.5
(a) t=1.9s; (b) t =3.74s; (c) t =6.24s;(d) t =7.8s; (e) t =9.86s; (f) t =15.72s

It is essential to investigate the interaction of the WV, HSV, UV, RLV and shear layer, which is the main reason for the disappearing of the RLV. A scheme for clarifying this feature has been displayed in Fig.6, which has shown the evolution of interaction between the UV, HSV and WV at r=1.5. The formation of the HSV is near the nozzle (Fig.6(a)), which is seemed to show a fixed shape during the evolution. Once affected by UV in the initial phase, the HSV can generate a stable status through a period of development (Fig.6(a)~(f)). Since the HSV is formed earlier than the CRVP, the CRVP can only affect WV.

As shown in Fig.6, the most significant phenomenon is the evolving regular vortex, which can be attributed to the interaction between the UV and WV. From the outset, the core area of the jet is strongly influenced by the interaction between the boundary layer and the UV system, therefore, the jet is restrained which can result in lifting up from the jet core area. There is a clear exhibition about the flow of the oscillation along the direction of crossflow. The regular vortex will be generated below the CRVP and RLV under the interaction of the UV and WV. Due to the combined effect of the jet, boundary layer, the WV, HSV, UV in the jet core area, vortices including the regular vortices and large scale eddy will be soon decomposed to a series of shocking eddy wrapped into the wake zone (shown in Fig.6(f)) [Guan (2007)].

3.5.*Spanwise Velocity Streamlines*

As shown in Fig.7, the CRVP have been clearly generated based on different velocity ratios (r=1~4) by adopting the spanwise velocity streamlines, which is formed due to the high velocity ratio. The annular area affected by the CRVP (marked with red circle) at lower velocity ratio has an approximately elliptical shape with a smaller acreage. The velocity stream is seemed to converge at one point above the CRVP. However, with the increasing the velocity ratio, the eccentricity for the annular area affected by CRVP is nearly generated at zero, which means that the shape of the annular area is almost in the circle shape. As a matter of course, the corresponding area is followed bigger by r.

When r=1 (Fig.7(a)), to be same to the other studied cases (Figs.7(b)~(d)), the relatively symmetric vortices pair have been formed at the upper boundary. However, the corresponding intensity of the crimping and winding for the formed vortices at low ratio (r=1) are much weaker than the other cases. In addition, the whole rotating region is also much more flat. As the velocity ratio further increases, the whole former formed vortex cores have begun to move up gradually (shown in Figs.7(b)~(d)). And the corresponding strength have also been boosted up. The whole vortex regions have become more circular and greater. When r=4 (shown in Fig.7(d)), the intensity and region of the formed vortices have reached the maximum value. The spanwise position for the vortex core has located nearly five times height compared with the lowest velocity ratio (shown in Fig.7(a)). Except this, as shown in Fig.7(d), the second vortices have also been induced by the formed CRVP at relatively higher velocity ratios (r=4). The higher velocity ratio got, the stronger second vortices can be obtained. And the corresponding position is also gradually moving up, which will be swallowed up by the CRVP ultimately at a certain downstream position.

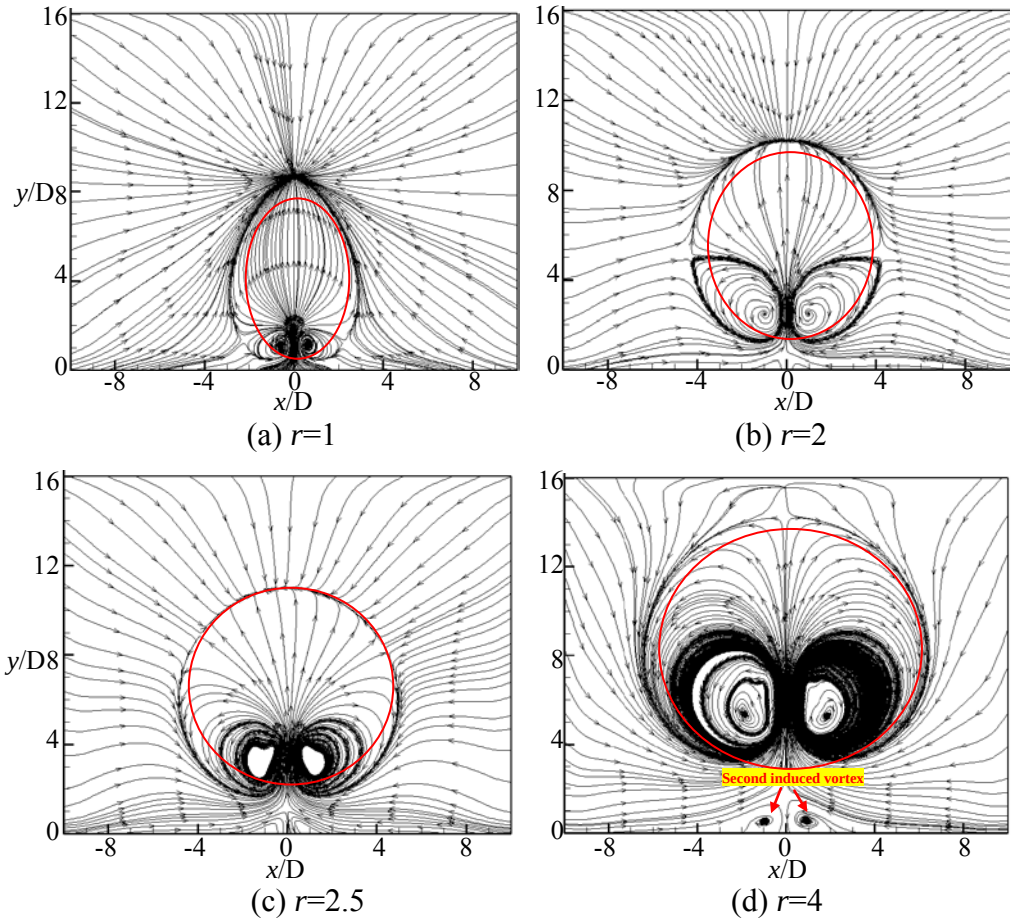


Fig. 7 The spanwise velocity streamlines for the JICF

4. Conclusions

The jet in crossflow at Reynolds number (*Re*=100) have been performed based on the LES method, the corresponding conclusions have been listed as follows:

(1) Three-dimensional streamlines are closely related to the velocity ratios, the higher velocity ratios become, the deeper the penetration can reach. The more stable mixed vortices including the CRVP, RLV, UV, HSV and WV can be generated. The RLV will move up at the beginning, and then start to fall after a certain period of time near the nozzle of the jet.

(2) The presence of the CRVP and RLV structures can maintain quite a long distance even the flow exits at lower velocity ratio, but the RLV will be in destruction soon at larger velocity ratios under the interaction for the mixed jet in crossflow. The relative regular vortices can be generated below the CRVP and RLV under the interaction of UV and WV, which will be decomposed to a series of shocking eddy wrapped into the crossflow.

(3) The relatively symmetric vortices pair can be generated at the upper boundary. The whole vortex regions can become more circular and greater with the increasing velocity ratio.

The second vortices are also induced by the previous formed CRVP at relatively higher velocity ratios

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The grant support from the National Science Foundation of China (No. U1430113) and Academic Start Program for Young Teachers of Beijing Institute of Technology (No. 3010012261502) is greatly acknowledged.

References

Camussi R, Guj G, Stella A. (2002). Experimental study of a jet in a crossflow at very low Reynolds number. *Journal of Fluid Mechanics*, **454**: 113-144.

Cárdenas C, Suntz R, Denev J A, et al. (2007). Two-dimensional estimation of Reynolds-fluxes and-stresses in a Jet-in-Crossflow arrangement by simultaneous 2D-LIF and PIV. *Applied physics B*, **88**: 581-591.

Chassaing P, George J, Claria A, et al. (1974). Physical characteristics of subsonic jets in a cross-stream. *Journal of Fluid Mechanics*, **62**: 41-64.

Chen A R, Ai H L. (2010). Computional Bridge Aerodynamics: Large Eddy Simulation. China Communications Press, Beijing. (in Chinese)

Esmaeili M, Afshari A, Jaberi F A. (2015). Large-eddy simulation of turbulent mixing of a jet in cross-flow. *Journal of Engineering for Gas Turbines and Power*, **137**: 091510.

Guan H, Wu C J. (2007). Large-eddy simulations and vortex structures of turbulent jets in crossflow. *Science in China Series G: Physics, Mechanics and Astronomy*, **50**: 118-132.

Gutmark E J, Grinstein F F. (1999). Flow control with noncircular jets. *Annual review of fluid mechanics*, **31**: 239-272.

Ivanova E M, Noll B E, Aigner M. (2013). A numerical study on the turbulent Schmidt numbers in a jet in crossflow. *Journal of Engineering for Gas Turbines and Power*, **135**: 011505.

Jouhaud J C, Gicquel L Y M, Enaux B, et al. (2007). Large-eddy-simulation modeling for aerothermal predictions behind a jet in crossflow. *AIAA journal*, **45**: 2438-2447.

Lacarelle A, Paschereit C O. (2012). Increasing the Passive Scalar Mixing Quality of Jets in Crossflow With Fluidics Actuators. *Journal of Engineering for Gas Turbines and Power*, **134**: 021503.

Lim T T, New T H, Luo S C. (2006). Scaling of trajectories of elliptic jets in crossflow. *AIAA journal*, **44**: 3157-3160.

Mashayek A, Behzad M, Ashgriz N. (2011). Multiple Injector Model for Primary Breakup of a Liquid Jet in Crossflow. *AIAA journal*, **49**: 2407-2420.

Muppidi S, Mahesh K. (2007). Direct numerical simulation of round turbulent jets in crossflow. *Journal of Fluid Mechanics*, **574**: 59-84.

New T H, Lim T T, Luo S C. (2006). Effects of jet velocity profiles on a round jet in cross-flow. *Experiments in Fluids*, **40**: 859-875..

Saha A K, Yaragani C B. (2012). Three-dimensional numerical study of jet-in-crossflow characteristics at low Reynolds number. *Heat and Mass Transfer*, **48**: 391-411.

Salewski M, Stankovic D, Fuchs L. (2008). Mixing in circular and non-circular jets in crossflow. *Flow, Turbulence and Combustion*, **80**: 255-283.

Srinivasan S, Pasumarti R, Menon S. (2012). Large-eddy simulation of pulsed high-speed subsonic jets in a turbulent crossflow. *Journal of Turbulence*, **13**: N1.

Yuan L L, Street R L, Ferziger J H. (1999). Large-eddy simulations of a round jet in crossflow. *Journal of Fluid Mechanics*, **379**: 71-104.

Car-following model with considering vehicle's backward looking effect and its stability analysis

Y. N. Wang[1,2,3], H. X. Ge[1,2,3], *,†S. M. Lo[4], K. L. Tsui[4], and K. K. Yuen[4]

[1] Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China
[2] Jiangsu Province Collaborative Innovation Center for Modern Urban Traffic Technologies, Nanjing 210096, China
[3] National Traffic Management Engineering and Technology Research Centre Ningbo University Sub-centre, Ningbo 315211, China
[4] Department of Civil and Architectural Engineering, City University of Hong Kong, Kowloon, Hong Kong 999077, China.

*Presenting author: bcsmli@cityu.edu.hk
†Corresponding author: bcsmli@cityu.edu.hk

Abstract

In this paper, an extended car-following model is derived by considering vehicle's backward looking effect which is based on the optimal velocity model and the optimal velocity (OV) function is extended by introducing variable safety distance. Also, a new control signal including more comprehensive information is introduced on the viewpoint of feedback control. Furthermore, the stability condition for the model is derived and the numerical simulation is carried out to investigate the advantage of the proposed model with control signal which can alleviate the traffic jams efficiently. The results are also consistent with the theoretical analysis correspondingly.

Keywords: Car-following model, Feedback control method, Stability condition, Variable safety distance.

Introduction

In recent decades, traffic flow theories have attracted much attention of scientists' and researchers' in the study of mathematical physics and control theory. Because the traffic congestion has closely influenced human's daily life up to present, such as traffic accident, fuel consumption and air pollution. As for traffic behavior, many approaches have been introduced to investigate the properties of traffic flow, and obtained some significant results [1-5].

Modern traffic is one of the most significant symbols of social modernization which provides much convenience for our daily life. However, traffic congestion problem is also being increasingly deteriorated because of the huge traffic flux. Back to 1953, Pipes [6] developed a car following model to restrain the traffic congestion and provided some relevant results through theoretical analysis, which assumed that the behind vehicle adjusted its behavior following the preceding vehicle's action in the same lane. After that, Newell [7] proposed a car-following model with a differential equation and gave some graphic description for the optimal velocity (OV) function in 1961. Then it's worth pointing out that an vital extended car-following model called optimal velocity model (OVM) was introduecd by Bando et al. [8]. In the OVM, the acceleration of the vehicle at the same time was determined by the difference between actual velocity and an optimal velocity. Based on this (OVM), a great deal of car-following models have been extended by adding more comprehensive information into the real traffic system [9-12].

In 1999, Konishi et al. [13] developed a chaotic car-following model by setting time delay feedback control signals, and studied single-lane traffic operation without reverse phenomenon under an open boundary condition. In 2007, Han et al. [14] put forward to a modified CM car-following model and found that their model could promote the stability of traffic flow. Recently, Zheng et al. [15] presented an improved car-following model with considering lateral effect and its feedback control research, and the obtained results were correspond to the theoretical analysis. Additionally, other researches related to the control scheme have been carried out in a piecemeal form gradually [16][17].

Even in the physical community, the car-following model is still a hot topic. But up to now, we can hardly see studies concerning car-following in a viewpoint of control methods. So in this paper, it's necessary to provide a modified car-following model considering vehicle's backward looking effect based on the control theory which means a new control scheme that takes more comprehensive information into account is proposed. Detail definitions are in the section 3.

The outline of this paper is organized as follows. In Sec. 2, the modified car-following model considering vehicle's backward looking effect is presented, and its stability condition is analyzed via control method. In Sec. 3, the model including control signal is established and feedback control theory is used to analyze the stability conditions. In Sec. 4, several numerical simulations are carried out to verify the theoretical results. Conclusions are given in Sec. 5.

Car-following model and its stability analysis

Modified model

This research is based on OVM [8] in 1995. The dynamic equation is described as

$$\frac{dv_n(t)}{dt} = \frac{1}{\tau}\left[pV_p\left(\Delta x_n(t), v_n(t)\right) + qV_b\left(\Delta x_{n-1}(t), v_{n-1}(t)\right) - v_n(t)\right] \tag{1}$$

Where $\Delta v_n(t) = v_{n+1}(t) - v_n(t)$ is the velocity difference between $\Delta x_n(t) = x_{n+1}(t) - x_n(t)$, the $n\text{-}th$ considering vehicle and the preceding vehicle; $x_n(t)$ is the real position of the $n\text{-}th$ considering car at time t; $a = \frac{1}{\tau}$ is the sensitivity of driver and is the inverse of delay time τ. $V_p\left(\Delta x_n(t), v_n(t)\right)$ is the improved optimal velocity (OV) function for forward looking and $V_b\left(\Delta x_{n-1}(t), v_{n-1}(t)\right)$ is the modified optimal velocity (OV) function for backward looking; $p, q(p \geq q)$ stands for the relative weights of two OV functions. The two OV functions are given as:

$$V_p(\Delta x_n(t), v_n(t)) = \frac{v_{max}}{2}[\tanh(\Delta x_n(t) - h_n^v) + \tanh(h_n^v)] \tag{2}$$

$$V_b(\Delta x_{n-1}(t), v_{n-1}(t)) = \frac{v_{max}}{2}[\tanh(\Delta x_{n-1}(t) - h_{n-1}^v) + \tanh(h_{n-1}^v)] \tag{3}$$

$$h_n^v = d_1 T_s v_n(t) + h_c ; h_{n-1}^v = d_2 T_s v_{n-1}(t) + h_c \tag{4}$$

where v_{max} is the maximum velocity and h_c is the traditional safety distance; T_s is

the time step unit, and d is the reaction coefficient for $v_n(t)$.

Stability analysis

The dynamical equation is rewritten as follows:

$$\begin{cases} \dfrac{dv_n(t)}{dt} = a\left[pV_p\left(y_n(t), v_n(t)\right) + qV_b\left(y_{n-1}(t), v_{n-1}(t)\right) - v_n(t)\right], \\ \dfrac{dy_n(t)}{dt} = v_{n+1}(t) - v_n(t), \end{cases} \tag{5}$$

where $y_n(t) = \Delta x_n(t)$.

We suppose the desired velocity of vehicles and comprehensive distance are v^* and y^*, so the steady state of the following vehicles is

$$[v_n(t), y_n(t)]^T = [v^*, y^*]^T. \tag{6}$$

Then, consider an error system around steady state (6), that is,

$$\begin{cases} \dfrac{d\delta v_n(t)}{dt} = a\left[p\delta y_n(t)\Lambda_1 + p\delta v_n(t)\Lambda_2 + q\delta y_{n-1}(t)\Lambda_3 + q\delta v_{n-1}(t)\Lambda_4 - \delta v_n(t)\right] \\ \dfrac{d\delta y_n(t)}{dt} = \delta v_{n+1}(t) - \delta v_n(t) \end{cases} \tag{7}$$

where $\delta v_n(t) = v_n(t) - v^*$, $\Lambda_1 = \dfrac{\partial V\left(y_n(t), v_n(t)\right)}{\partial y_n(t)}\Big|_{y_n(t)=V_p^{-1}(v_0)}$, $\Lambda_2 = \dfrac{\partial V\left(y_n(t), v_n(t)\right)}{\partial v_n(t)}\Big|_{v_n(t)=v_0}$, $\Lambda_3 = \dfrac{\partial V\left(y_{n-1}(t), v_{n-1}(t)\right)}{\partial y_{n-1}(t)}\Big|_{y_{n-1}(t)=V_b^{-1}(v_0)}$, $\Lambda_4 = \dfrac{\partial V\left(y_{n-1}(t), v_{n-1}(t)\right)}{\partial v_{n-1}(t)}\Big|_{v_{n-1}(t)=v_0}$, $\delta y_n(t) = y_n(t) - y^*$.

After Laplace transformation for traffic system (7), we can get

$$\begin{bmatrix} V_n(s) \\ Y_n(s) \end{bmatrix} = \frac{1}{p(s)} \begin{bmatrix} s & ap\Lambda_1 \\ -1 & s + a - ap\Lambda_2 \end{bmatrix} \begin{bmatrix} aq\Lambda_4 & 0 & aq\Lambda_3 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_{n-1}(s) \\ V_{n+1}(s) \\ Y_{n-1}(s) \end{bmatrix} \tag{8}$$

$$p(s) = s^2 + a(1 - p\Lambda_2)s + ap\Lambda_1 \tag{9}$$

where $V_n(s) = L(\delta v_n(t))$, $Y_n(s) = L(\delta y_n(t))$, $L(.)$ denotes the Laplace transform and s is a complex variable.

In reality, based on the control theory, we obtain the transfer function $G(s)$, that is

$$G(s) = \frac{(aq\Lambda_3 + aq\Lambda_4)s + ap\Lambda_1}{s^2 + a(1 - p\Lambda_2)s + ap\Lambda_1} \tag{10}$$

Thus, traffic jams will never occur in the traffic flow system if $p(s)$ is stable and $\|G(s)\|_\infty \le 1$. In fact, based on the Hurwitz stability criterion, we can get that $p(s)$ is stable. So, the stability condition is given by

$$a \ge \frac{2p\Lambda_1 + q(\Lambda_3 + \Lambda_4)}{(1 - p\Lambda_2)^2} \tag{11}$$

Feedback control scheme

In this part, an extended feedback control signal including more comprehensive information is added into system (1), so we have

$$\frac{dv_n(t)}{dt} = a\left[pV_p(\Delta x_n(t), v_n(t)) + qV_b(\Delta x_{n-1}(t), v_{n-1}(t)) - v_n(t)\right] + u_n(t) \tag{12}$$

$$u_n(t) = \lambda \Delta v_n(t) + \gamma^2 H(y_n(t) - h_n^v)(y_n(t) - h_n^v), \tag{13}$$

where λ is the reaction coefficient for the relative velocity $\Delta v_n(t)$ and γ is another reaction coefficient for the $H(y_n(t) - h_c)(h_c - y_n(t))$. Function $H(.)$ is described as

$$H(y_n(t) - h_n^v) = \begin{cases} 0, & y_n(t) - h_n^v > 0, \\ 1, & y_n(t) - h_n^v \le 0, \end{cases} \tag{14}$$

As $y_n(t) - h_n^v \le 0$, our feedback control signal $u_n(t)$ is

$$u_n(t) = \lambda \Delta v_n(t) + \lambda^2 (y_n(t) - h_n^v), \tag{15}$$

Under this condition, the dynamical Eq.(12) can be described as

$$\begin{cases} \dfrac{dv_n(t)}{dt} = a\left[pV_p(y_n(t), v_n(t)) + qV_b(y_{n-1}(t), v_{n-1}(t)) - v_n(t)\right] \\ +\lambda(v_{n+1}(t) - v_n(t)) + \gamma^2 (y_n(t) - h_n^v), \\ \dfrac{dy_n(t)}{dt} = v_{n+1}(t) - v_n(t), \end{cases} \tag{16}$$

Similar to the analysis of second part, the transfer function $\tilde{G}(s)$ can be obtained after Laplace transform.

$$\tilde{G}(s) = \frac{(aq\Lambda_3 + aq\Lambda_4 + \lambda)s + (ap\Lambda_1 + \gamma^2)}{s^2 + (a + dT_s\gamma^2 + \lambda - ap\Lambda_2)s + ap\Lambda_1 + \gamma^2} \tag{17}$$

$$\tilde{p}(s) = s^2 + (a + dT_s\gamma^2 + \lambda - ap\Lambda_2)s + ap\Lambda_1 + \gamma^2 \tag{18}$$

In fact, the traffic jams will be weaken if $\tilde{p}(s)$ is stable and $\|\tilde{G}(s)\|_\infty \le 1$. Furthermore, $\tilde{G}(j\omega)$ must be smaller than 1 for all positive ω^2 to ensure stability. Hence, the stability criterion of the extended mode is given by

$$Aa^2 + Ba + C \geq 0 \tag{19}$$

where $A=(1-p\Lambda_2), B=2(1-p\Lambda_2)(dT_s\gamma^2+\lambda^2)-(2p\Lambda_1+q\Lambda_3+q\Lambda_4), C=(dT_s\gamma^2+\lambda^2)^2-(2\gamma^2+\lambda).$

Numerical simulations

In this simulations, the parameters for the improved car-following model are set as $y^* = 5.0m$, $a = 2s^{-1}$, $v^* = 20m/s$, $p = 0.8$, $q = 0.2$, $d = 0.3$ and $T = 0.1s$. It is assumed that all vehicles have the same parameters. The initial condition is the steady state for the model, and the initial positions and speeds are set as $y_n(0) = y^*$, $v_n(0) = v^*$, and $N = 120$ is the total number of vehicles. We consider a case where the leading vehicle stops suddenly for $v_n(0) = 0, t = nT = 100 - 103$.

Figure. 1 shows the velocity-time patterns of the 1st, the 25th and the 50th vehicles with different parameter values of γ. It can be seen from Fig. 1 that with the control signal, as the reaction coefficient γ decreases from 0.85 to 0.35, the stability of the traffic system is strengthened. And we can find that vehicles can reach steady running state in relatively short time as the reaction coefficient γ decreases. The amplitude of the velocity for the 25th vehicle decreases and the 50th vehicle runs smoothly.

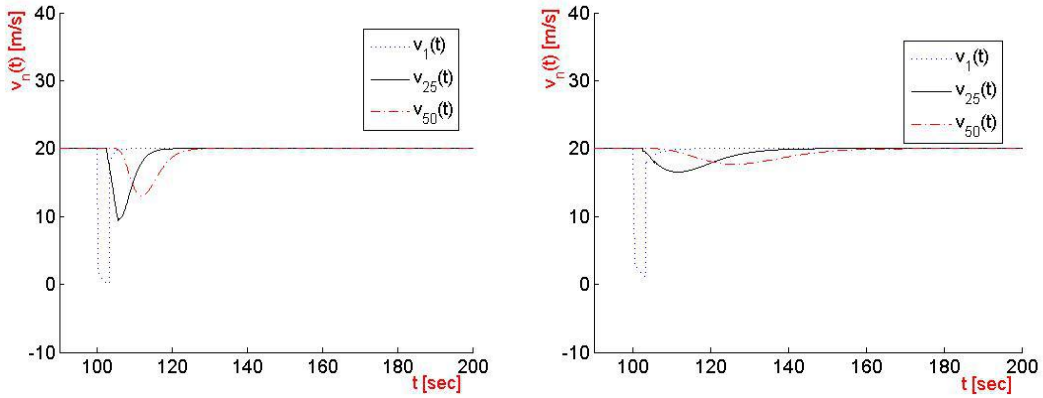


Figure 1. Numerical simulations for the modified car-following model with $\lambda = 0.65$, $v_{max} = 25m/s$, $\gamma = 0.85$ (left); $\lambda = 0.65$, $v_{max} = 25m/s$, $\gamma = 0.35$ (right)

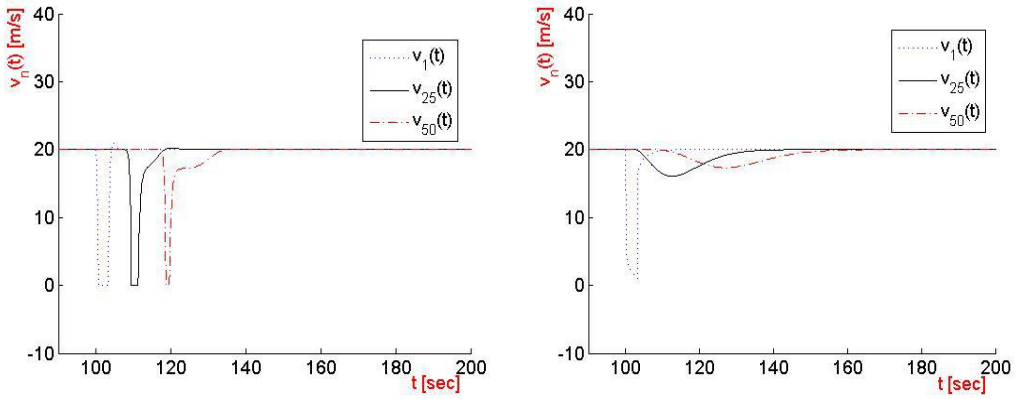


Figure 2. Numerical simulations for the modified car-following model with $\gamma = 0.35$, $v_{max} = 25m/s$, $\lambda = 0.15$ (left); $\gamma = 0.35$, $v_{max} = 25m/s$, $\lambda = 0.65$ (right)

Figure. 2 shows the velocity-time patterns of the 1st, the 25th and the 50th vehicles with different parameter. It can be seen from Fig.2 that with the control signal, as the reaction coefficient increases from 0.15 to 0.65, the stability of the traffic system is strengthened. And we can find that vehicles can reach steady running state in relatively short time as the reaction coefficient increases. The amplitude of the velocity for the 25th vehicle decreases and the 50th vehicle runs placidly. The simulation results of Fig. 1 and Fig. 2 illustrate that feedback control plays an vital role in vehicle dynamic driving behavior.

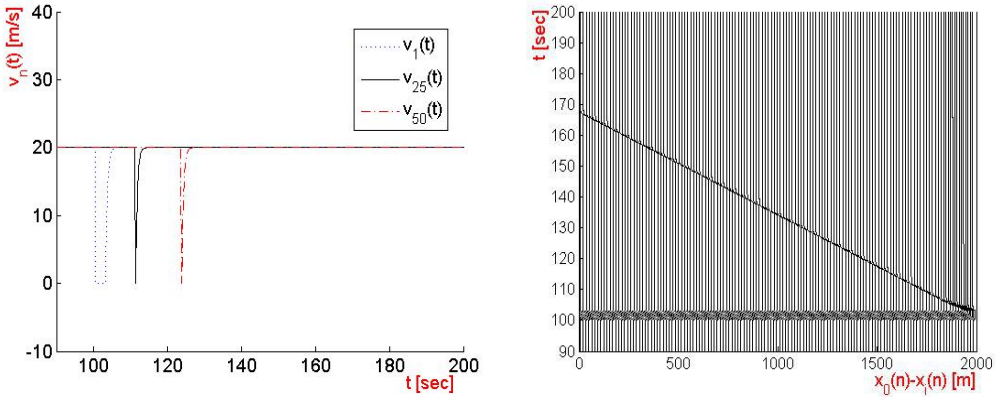


Figure 3. (a) Space-time plot of the traffic system (b) Temporal velocity behavior of the first,25th and 50th vehicles ($\gamma = 0, \lambda = 0, v_{max} = 20m/s$)

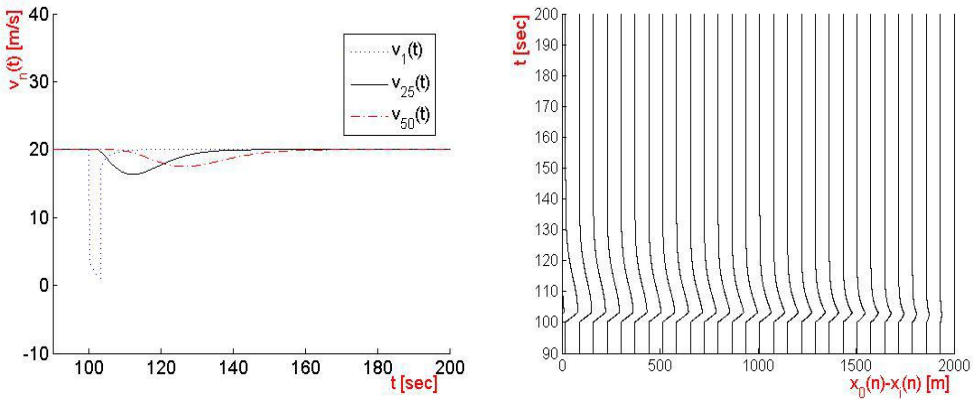


Figure 4. (a) Space-time plot of the traffic system (b) Temporal velocity behavior of the first, 25th and 50th vehicles ($\lambda = 0.75, \gamma = 0.35, v_{max} = 25m/s$)

Then, we simulate the system with the modified control scheme. As the stability condition in Eq. (11) and Eq. (18) is met, a comparison between the results in Figs. 3-4 illustrate that with control signal, although the maximum speed is larger compared with Fig. 3, as we choose the right parameters ($\lambda = 0.75, \gamma = 0.35, v_{max} = 25m/s$), it can be seen that vehicles can reach more steady running state in relatively short time. The amplitude of the velocity for the 25th vehicle decreases and the 50th vehicle runs more smoothly. Thus, it can be concluded that the proposed car-following model is useful for suppressing the increasingly serious traffic jams.

Conclusions

In this paper, an extended car-following model is established considering vehicle's backward looking effect. The optimal velocity (OV) function is extended by introducing variable safety distance. The effect of some important information (such as the relative velocity and the

difference between safety variable distance and headway) on the traffic current and the jamming transition has been investigated with the use of numerical and analytic methods. The stability condition is obtained for the new model via control method. The numerical simulation is used to show the advantage of the proposed model with control scheme. The results are consistent with the theoretical analysis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grant No. 11372166]; the Scientific Research Fund of Zhejiang Provincial, China [Grant Nos. LY15A020007, LY15E080013]; the Natural Science Foundation of Ningbo [Grant Nos. 2014A610028, 2014A610022]; the project T32-101/15-R; and the K.C. Wong Magna Fund in Ningbo University, China.

References

[1] Kerner, B. S. and Rehborn, H. (1996) Experimental properties of complexity in traffic flow, *Phys. Rev. E* **53**, 4275–4278.
[2] Li, Z. P. and Liu, Y. C. (2006) Analysis of stability and density waves of traffic flow model in an ITS environment, *Eur. Phys. J B* **53**, 367–374.
[3] Tang, T. Q., Li, J. G., Wang, Y. P. and Yu, G. Z. (2013) Vehicle's fuel consumption of car-following models, *Sci China Tech Sci*, **56**, 1307–1312.
[4] Li, Y. F., Sun, D. H., Liu, W. N., Zhang, M., Liao, X. Y. and Tang L. Modeling and simulation for microscopic traffic flow based on multiple headway, velocity and acceleration difference, *Nonlinear Dynamics* **66**, 15–28.
[5] Zhu, W. X. (2008) A backward looking optimal current lattice model. *Commun. Theor. Phys* **50**, 753–756.
[6] Pipes, L. A. (1953) An operational analysis of traffic dynamics *J. Appl.Phys* **24**, 274–281.
[7] Newell, G. F. (1961) Nonlinear effects in the dynamics of car-following, *Oper.Res* **9**, 209–229.
[8] Bando, M., Hasebe, K., Nakayama, A., Shibata, A. and Sugiyama, Y. (1998) Analysis of optimal velocity model with explicit delay, *Phys. Rev. E* **58**, 5429–5435.
[9] Ge, H. X., Cheng, R. J. and Dai, S. Q. (2005) KdV and kink-antikink solitons in car-following models, *Physica A* **357**, 466–476.
[10] Tang, T. Q., Wu, Y. H., Caccetta, L. and Huang, H. J. (2011) A new car-following model with consideration of roadside memorial, *Phys. Lett. A* **375**, 3845–3850.
[11] Ge, H. X., Dai, S. Q., Xue, Y. and Dong, L. Y. (2005) Stabilization analysis and modified Korteweg-de Vries equation in a cooperative driving system, *Phys. Rev. E* **71**, 066119.
[12] Peng, G. H. and Sun, D. H. (2010) A dynamical model of car-following with the consideration of the multiple information of preceding cars, *Phys. Lett. A* **374**, 1694–1698.
[13] Konishi, K., Kokame, H. and Hirata, K. (2000) Decentralized delayed-feedback control of an optimal velocity traffic model, *Eur. Phys. J B* **15**, 715–722.
[14] Han, X. L., Jiang, C. Y., Ge, H. X. and Dai, S. Q. (2007) A modified coupled map car-following model based on application of intelligent transportation system and control of traffic congestion, *Acta Phys. Sin* **56**, 4383–4392.
[15] Zheng, Y. Z., Zheng, P. J. and Ge, H. X. (2014) An improved car-following model with considering lateral effect and its feedback control research, *Chin. Phys. B* **23**, 020503.
[16] Sun, D. H., Zhou, T., Liu, W. N. and Zheng, L. J. (2013) A modified feedback controlled car-following model considering the comprehensive information of the nearest-neighbor leading car, *Acta Phys. Sin* **62**, 170503.
[17] Jiang, R., Hu, M. B., Zhang, H. M., Gao, Z. Y., Jia, B. and Wu, Q. S. (2015) On some experimental features of car-following behavior and how to model them, *Transportation Research Part B* **80**, 338–354.

Numerical study on effectiveness of continuum model box used in shaking table test under non-uniform excitation

† Zhiyi Chen[1], *Sunbin Liang[1]

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

*Presenting author: 550507206@qq.com

† Corresponding author: zhiyichen@tongji.edu.cn

Abstract

Unlike superstructure, it is necessary to bury underground structure model into model soil carried by a continuum model box, when conducting shaking table test under non-uniform excitation. The problem, how to transmit dynamic effectively between different shaking tables is need to be solved firstly. The present paper is devoted to study the effectiveness of continuum model box using when conducting non-uniform excitation shaking table test. A full-scale 3D entity finite element model of soil and model boxes is simulated. In order to avoid the randomness of calculation result, three conventional coherency models are adopted to synthetize non-uniform ground motions respectively. In order to evaluate the effectiveness of continuum model box, the calculation results, including time history and frequency spectrum of soil acceleration responses, are contrasted with those of 2D free field analysis. The calculation results show that the distribution of peak acceleration response of soil cased in the continuum model box is almost the same as that of 2D free field analysis. The Fourier Amplitudes of the surface acceleration responses of soil state that the frequency spectrum components of soil acceleration response have little difference between 3D dynamic analysis and 2D free field analysis. Thus, it is rational to adopt continuum model box with rigid connection to conduct shaking table test of underground structure under non-uniform excitation.

Keywords: Effectiveness; Continuum model box; Shaking table test; Non-uniform excitation.

1. Introduction

Observations from earthquake strong-motion arrays show notable differences among the records of ground motions at different locations within the dimensions of typical extended structures [1]. That is called spatially varying ground motions, which is caused by the wave passage effect, the incoherence effect and the site-response effect [2]. Unlike the small-scale structure, it is necessary to conduct non-uniform excitation analysis for extended structures, such as tunnels, bridges and pipelines, since spatially varying ground motions may have significant influence on seismic response.

In the last few decades, researches on seismic responses of tunnel induced by non-uniform excitations are mainly limited to numerical analysis. Hashash et al. [3] and Anastasopoulos et al. [4] performed 3-D dynamic analysis to study seismic responses of the San Francisco bay tunnel and Greece Rion-Antirrion strait tunnel under spatially varying ground motions, respectively. A consistent conclusion stated that spatially varying ground motions increased the seismic responses of the immersed tunnels significantly. Park et al. [5] conducted pseudo-static 3-D finite element analysis to investigate seismic responses of a tunnel under non-uniform excitations. Yu et al. [6] proposed a multi-scale method to simulate a water

delivery tunnel constructed by shield method and studied the influence of wave passage effect on seismic responses. Li and Song [7] developed a 3-D finite element model in time domain to provide feasible computational modeling technique for the tunnels under asynchronous excitations. However, few experimental investigations are conducted to study the seismic responses of tunnels under non-uniform excitations.

Experimental method plays an important role in geotechnical engineering researches. It provided a realistic way to test and verify the results derived from theoretical analyses, and potentially to identify novel phenomena that are inaccessible by theoretical analysis alone. In recent years, centrifuge and shaking table tests are conducted to study the seismic performance and reveal failure mechanism of underground structure [8]-[10]. Since centrifuge test can reproduce the in situ stress state of soil, it is commonly believed that the is an attractive way to study seismic performance of underground structure [11]. However, shaking table test is precise in seismic loading, control and observation [12]. Moreover, shaking table array provides a feasible way to study the dynamic response of the extended underground structure, like tunnel, under non-uniform excitations. Unlike superstructure, it is necessary to bury underground structure model into model soil carried by a continuum model box, when conducting shaking table test under non-uniform excitation. Extremely limited shaking table test of underground structure under non-uniform excitations has been conducted. Chen et al. [13] performed a shaking table test of utility tunnel to study the effect of non-uniform earthquake wave excitations. However, two separating model boxes were adopted, and it ignored the continuum of soil. It is believed that this ignorance affects the evaluations of seismic performance since the deformation of the surrounding soil rather than structural dynamic characteristic is the control factor of response of underground structure. Thus, in order to represent the reality of dynamic response of line-like underground structure as far as possible, some efforts should devoted to develop a continuum model box before conducting non-uniform shaking table tests. Therefore, the problem, verifying the effectiveness of continuum model box connecting different shaking tables, is need to be solved.

Aiming this goal, a full-scale 3D entity finite element model of soil and model box is simulated to verify the effectiveness of continuum model box in this paper. To avoid the randomness of calculation results, three conventional coherency models are adopted to synthetize non-uniform ground motions as input excitations, respectively. The conclusions of the presented paper could be valuable to the non-uniform excitation shaking table test of underground structure.

2. Numerical modeling of shaking table tests

The prototype shake table array is consisting of two Quanser Company shake tables, named Shake Table Π, at the Structural Engineering Laboratory in Tongji University. As shown in Fig.1, the dimension of each table stage is 46cm × 46cm in plane. The maximum acceleration is 2.5g with the maximum payload 7.5kg. The frequency of the input ground motion covers the range 0.1–20 Hz. Finite element model of the soil-continuum box system is established in

this section based on the prototype shake table array.

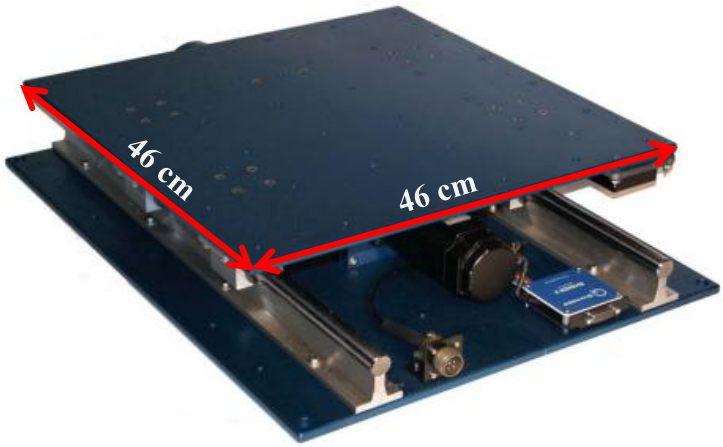


Fig. 1. Prototype of Shake Table Π

In the presented paper, dynamic time-history analyses are carried out using the general-purpose commercial ABAQUS software [14]. Element C3D8R is adopted to simulate model soil, and the soil density, elastic modulus and Possion's ratio are set as 700kg/m^3, 4.89MPa and 0.35, respectively. Mohr-Coulomb model and Rayleigh damping are used to take the plasticity and nonlinear dynamical characteristics into account. The detailed information of soil is listed in Table 1.

Table 1. Properties of the soil

Description	Parameter	Value
Density	ρ(kg/m^3)	700
Elastic modulus	E(MPa)	2000
Possion's ratio	υ	0.35
Friction angle	φ(°)	33
Cohesion	c(kPa)	10.6
Rayleigh damping	α	0.288043
	β	0.045054

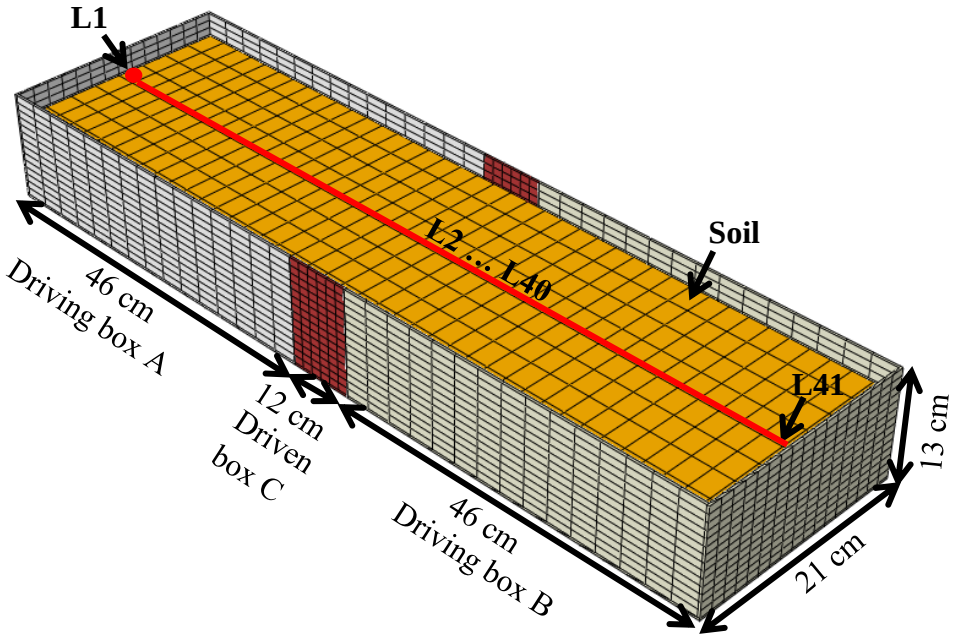


Fig. 2. 3D finite element model of the whole soil-continuum box system

Fig. 2 illustrates the finite element model of the whole soil-continuum box system. There are two driving model box, consisting of driving box A and box B that are fixed on two shaking tables and a driven model box. The model box will be fabricated by organic glass in the future physical shaking table test, which is a homogeneous material with a stable mechanical property. Element C3D8R is also employed to simulate model box. The density, elastic modulus and Possion's ratio of model box are set as 1120kg/m^3, 3150MPa and 0.3, respectively. As shown in Fig. 2, the whole continuum box is with the length of 104cm, consisting of two driving model box (box A and B) are both with the length of 46cm and a driven model box (box C) is with the length of 12cm. Since the materials of driving and driven boxes are the same, the model box with the length of 104cm is established as one whole. The transverse dimension of the model box is 21cm (width) × 13 cm (height). The thickness of model box is 3mm. Due to capability limitation of the prototype shaking table, the height of soil cased in the model box, which is denoted as *H*, is set as 9cm. The surface interaction of the soil and the sidewalls of the model box are all set as Finite Slip with the friction and the slip tolerance factors of 0.2 and 0.005, respectively. Tie Constraint is adopted to simulate the surface interaction of the soil and the bottom of the model box.

3 Analysis process and calculation cases

3.1 Analysis process

To verify the effectiveness of continuum model box used in shaking table test under non-uniform excitation, the following analysis process is used.

1. As stated above, a full-scale 3D entity finite element model of soil and model boxes is established. Three conventional coherency models are adopted to synthetize non-uniform ground motions as input excitations to avoid the randomness of calculation results.
2. In order to evaluate the validity of the above-mentioned 3D dynamic analysis, 2D free field analysis, as a reference standard, is performed under three different non-uniform excitations. The finite element model of 2D free field analysis is depicted in Fig. 3. There are three parts of the free field with the length of 46, 12 and 46cm, which are corresponding to the soil cased in boxes A, B and C in 3D dynamic analysis. The infinite element is adopted in two sides of the free field model to consider the boundary effect. Element CPE4R is used to simulate the soil with density, elastic modulus and Possion's ratio of 700kg/m^3, 4.89MPa and 0.35, respectively. Same as 3D dynamic analysis, Mohr-Coulomb model and Rayleigh damping are used to consider the plasticity and nonlinear dynamical characteristics. As shown in Table 1, the soil characteristics are the same as 3D dynamic analysis.
3. After the aforementioned two steps, the soil acceleration responses in longitudinal direction, which emphasize the peak values and the Fourier Spectrum, of 3D dynamic analysis and 2D free field analysis are compared to each other. There are some conclusions drawn from the calculation results.

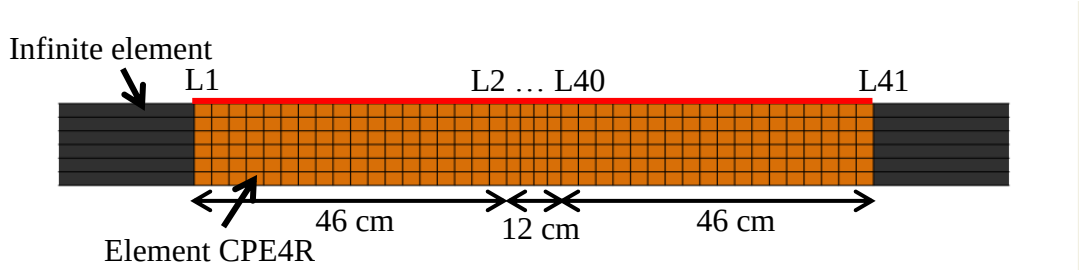


Fig. 3. 2D finite element model of free field analysis

3.2 Calculation cases

In the presented paper, the effectiveness of continuum model box is studied by full-scale 3D dynamic analysis. 2D free field analysis is conducted as a reference standard. In order to avoid the randomness of calculation results, three conventional coherency models are used to synthetize non-uniform ground motions as input excitations, respectively. The selected coherency models are described as following.

1) Hindy and Novak coherency model: When conducting a stochastic analysis of the pipeline, Hindy and Novak [15] firstly introduced the coherency model into earthquake engineering to describe the spatial variation of the ground motion. Based on wind engineering, the expression is relatively simple with only two parameters, that is:

$$\left|\gamma\left(\omega,d\right)\right|=exp\left(-\alpha\left(\omega d\right)^{\beta}\right) \tag{1}$$

Where, ω and d are the angular frequency and distance respectively; and the model parameters are α=3.007×10^{-4}, β=0.9. H-N model is depicted in Fig. 4(a).

2) Harichandran and Vanmarcke coherency model: Basing on the study of four events recorded by SMART-1 array in Taiwan, Harichandran and Vanmarcke [16] proposed an empirical coherency model, which has been widely applied. The expression of this coherency model is shown as follows:

$$|\gamma(\omega,d)| = A\exp\left[-\frac{2d}{\alpha\theta(\omega)}(1-A+\alpha A)\right] + (1-A)\exp\left[-\frac{2d}{\theta(\omega)}(1-A+\alpha A)\right] \quad (2)$$

Where, $\theta(\omega)=k[1+(\omega+\omega_o)^b]^{-0.5}$; basing on Event 20 recorded by the SMART-1 array, the model parameters are A=0.636, α=0.0186, k=31,200 m, ω_o=9.49 rad/s, b=2.95 [17]. H-V model is shown in Fig. 4(b).

3) Qu-Wang-Wang coherency model: From the standpoint of coherency model in engineering application, Qu et al. [18] referenced to the method of determining the design response spectrum in seismic code, averaged the collected coherence value of the empirical coherency model for several earthquakes, and proposed a coherency model. It is beneficial for practical application to put forward a mean coherency model referencing the determination of design response spectrum. The function is shown as:

$$|\gamma(\omega,d)| = \exp\left[-a(\omega)^{b(\omega)}d\right] \quad (3)$$

Where, $a(\omega)=a_1\omega^2+a_2$; $b(\omega)=b_1\omega+b_2$; the parameters are a_1=0.00001678, a_2=0.001219, b_1=-0.0055 and b_2=0.7674. Q-W-W model is depicted in Fig. 4(c).

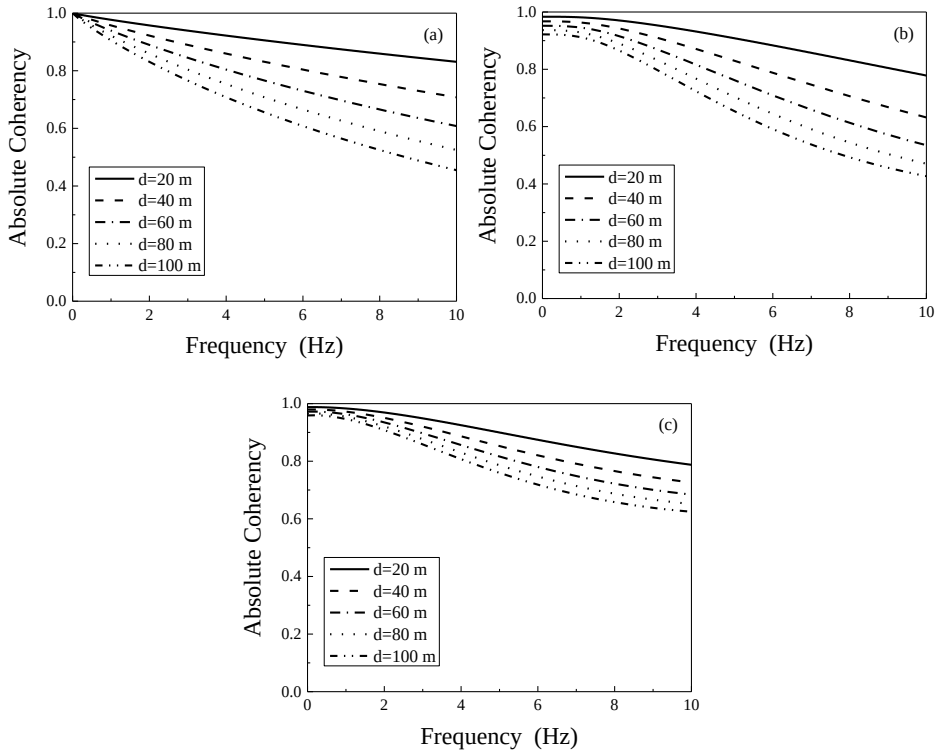


Fig. 4. Coherency Models: (a) H-N; (b) H-V; (c) Q-W-W

As shown in Table 2, there are four test cases for both 3D dynamic analysis and 2D free field analysis, which are consisted of uniform excitation (Case 1) and three cases for three models, including Case 2 is of H-N model, Case 3 is of H-V model and Case 4 is of Q-W-W model, respectively. Fig. 5 depicts the time histories of the synthetic ground motions. The peak ground motion is 0.1g. In this paper, trigonometric series simulation algorithm put forward by Hao [19] to simulate multi-support ground motion time histories are adopted. The power spectrum model $S(\omega)$ (Eq. (4)) proposed by Clough and Penzien [20] is adopted to simulate ground motions. The expression of this model is shown as:

$$S(\omega)=\frac{\omega_g^4+4\xi_g^2\omega_g^2\omega^2}{\left(\omega_g^2-\omega^2\right)^2+4\xi_g^2\omega_g^2\omega^2}\bullet\frac{\omega^4}{\left(\omega_f^2-\omega^2\right)^2+4\xi_f^2\omega_f^2\omega^2}S_0 \tag{4}$$

Where, S_0 is spectral intensity factor; ω is the angular frequency; ω_g and ξ_g are the resonant frequency and damping ratio of the first filter, which are relative to the site condition; ω_f and ξ_f are those of the second filter. The filter parameters corresponding to this soil type of Clough and Penzien power spectrum model are determined: S_0=0.0123347; ω_g= 9.67; ξ_g=0.9; ω_f=1.934; ξ_f=0.9. To consider the non-stationary of ground motion, the envelope function adopted in this paper was proposed by Amin and Ang [21], and its expression shown as following:

$$f(t)=\begin{cases}(t/t_1)^2, 0\le t\le t_1;\\ 1, t_1<t\le t_2;\\ exp\left[-c\left(t-t_2\right)\right], t>t_2\end{cases} \tag{5}$$

Where, c is the attenuation coefficient; t_1 and t_2 are the beginning and the ending moment of the stationary vibration stage, respectively. The parameters in Eq. (5) can be obtained as c=0.15, t_1=1.6s, t_2=12s.

Table 2. Detailed information of numerical analysis cases

Case name	Type of excitation	Coherency model
Case 1	Uniform	—
Case 2	Incoherent	Hindy and Novak coherency model
Case 3	Incoherent	Harichandran and Vanmarcke coherency model
Case 4	Incoherent	Qu-Wang-Wang coherency model

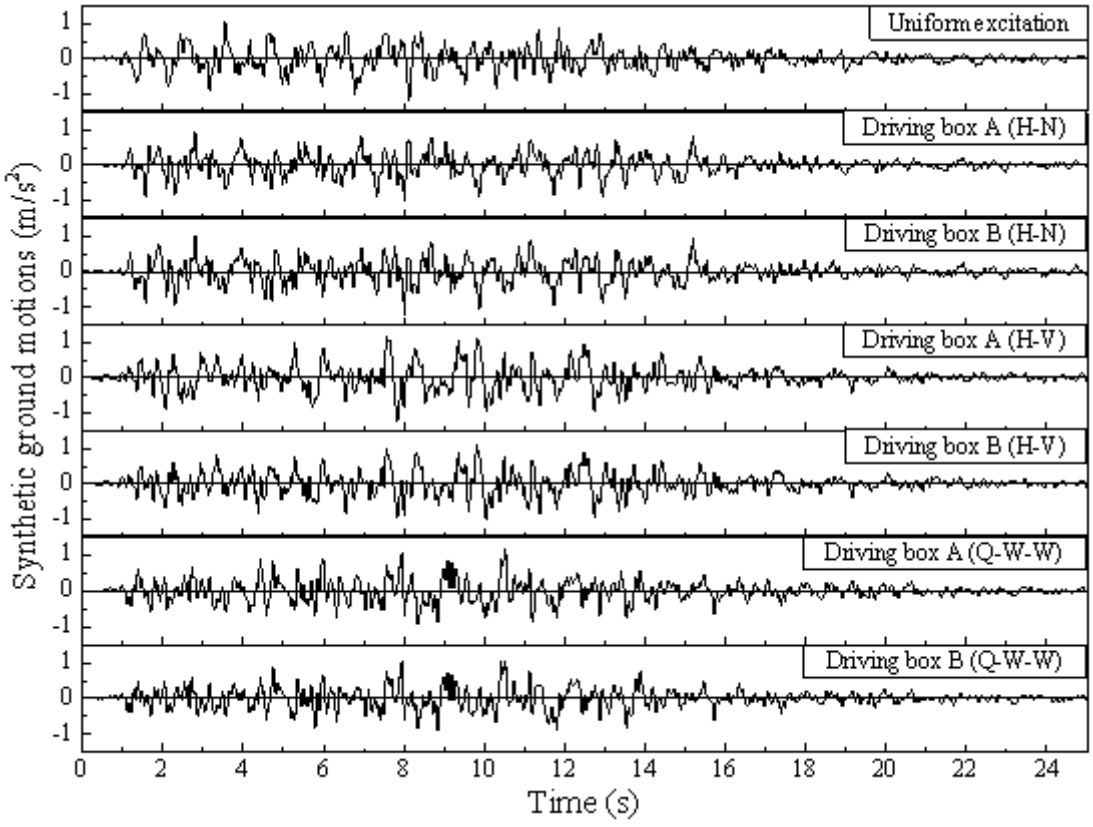


Fig. 5. Time histories of the synthetic ground motions

4 Numerical analysis results and discussions

Fig. 6 depicts the profile of longitudinal distribution of the peak acceleration response of soil

on ground surface. Peak acceleration responses of points L1-L41, whose locations are shown in Fig. 2, are selected to study. Non-uniform excitation causes differentia of acceleration response of soil among different locations in longitudinal direction. As shown in Fig. 6, the peak acceleration responses of soil are almost the same under uniform excitation (Case 1), while the profiles of distribution of the peak acceleration response of soil are asymmetric under non-uniform excitation (Case 2, Case 3 and Case 4).

No matter under uniform excitation or non-uniform excitation, the profile of distribution of the peak acceleration response of soil of 3D dynamic analysis basically overlap that of 2D free field analysis, which the soil is cased in the continuum model box. It illustrates that a continuum model box has almost no influence on the acceleration response of soil in longitudinal direction. The effectiveness of continuum model box used in shaking table test under non-uniform excitation is verified. More results and discussions are shown from different aspects to verify the effectiveness of continuum model box in the following.

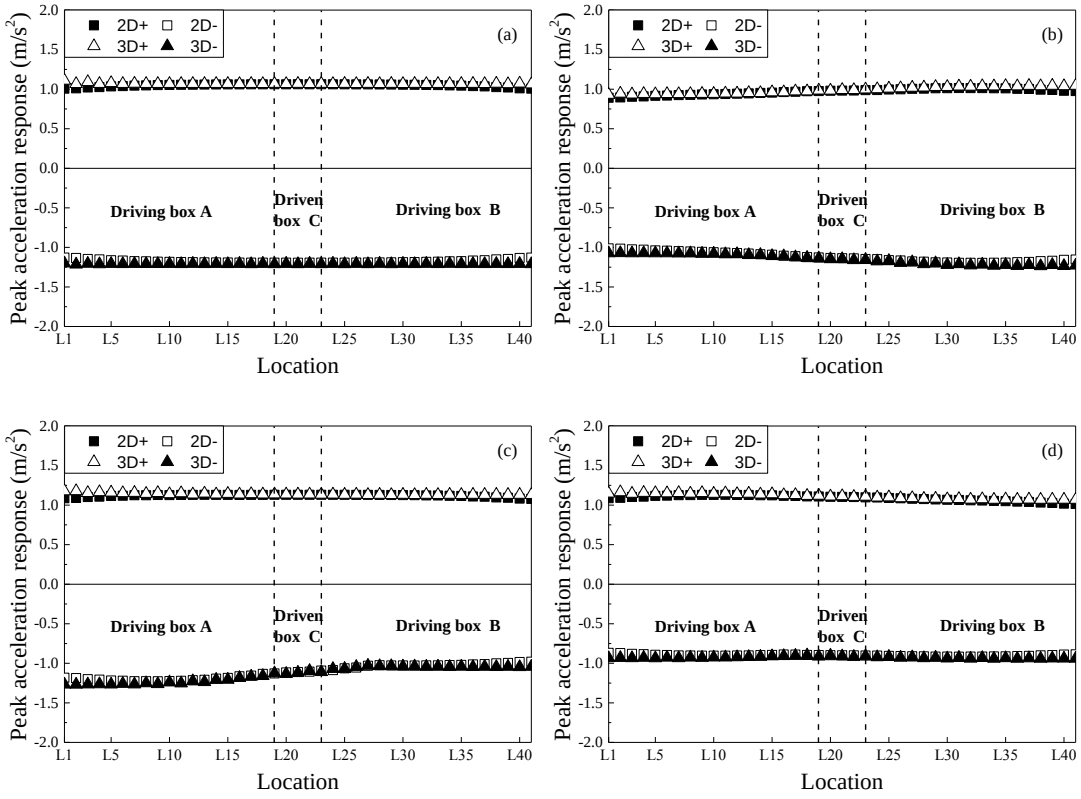


Fig. 6. Profile of longitudinal distribution of the peak acceleration response of soil: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4

Fig. 7 shows the profile of vertical distribution of the peak acceleration response of soil. It should be noted that the peak acceleration response is normalized to the peak value of the input ground motion. Totally seven equidistant locations in vertical above each the middle point of the driving box A, driven box C and driving box B are selected to studied. In Fig. 7, *H* represents the height of the soil cased in the continuum model box as stated before. There is an amplification effect of soil acceleration response. The maximum amplification factor is

1.03, which means there is 3% larger than the peak value of the input ground motion, since the height of the soil is too small of only 9cm.

Like in longitudinal direction, the profile of vertical distribution of the peak acceleration response of soil of 3D dynamic analysis is almost consistent with that of 2D free field analysis, especially for driven box C. Although it seems there exists great difference between the calculation results of 3D and 2D analysis in driving model box A (Fig. 7(c)), the greatest differential is less than 3% actually. Thus, it states that continuum model box has limited influence on the acceleration response of soil in vertical direction, and the effectiveness of continuum model box is also verified.

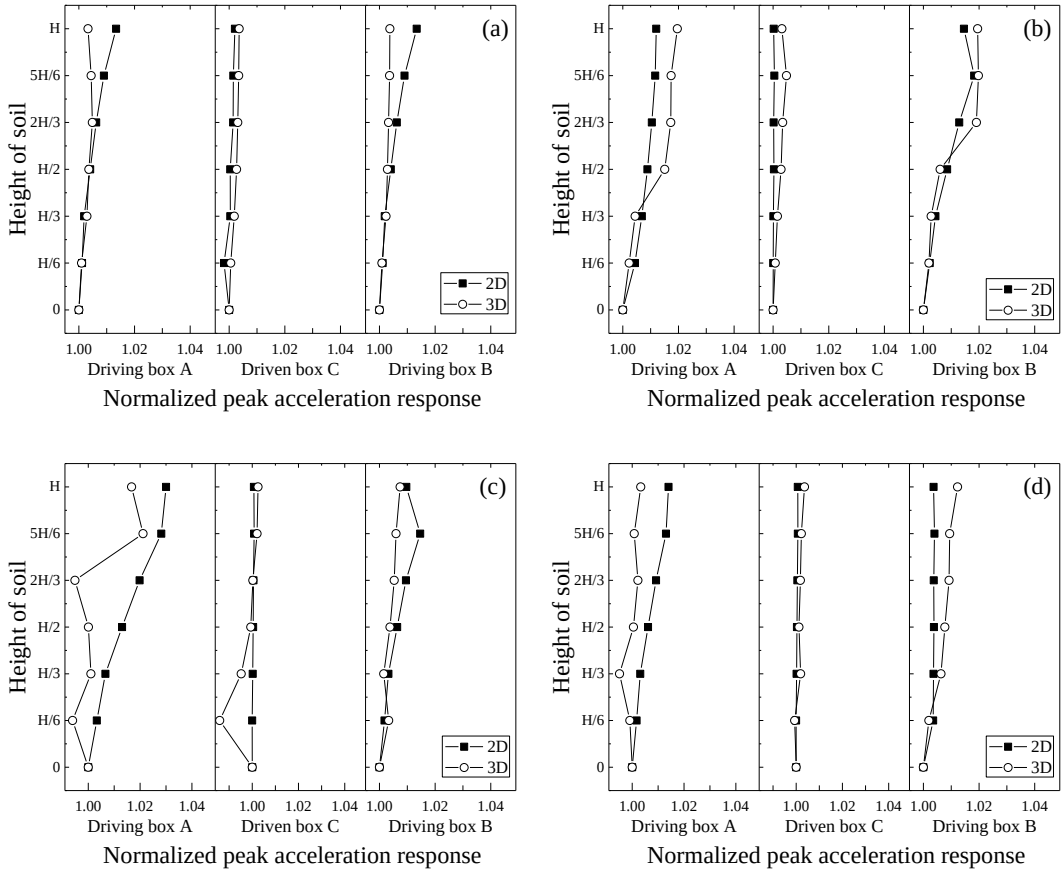


Fig. 7. Profile of vertical distribution of the peak acceleration response of soil: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4

Fourier Spectrum is used to study the differential of frequency contents of soil acceleration response between 3D dynamic analysis with continuum model box and 2D free field analysis. Due to space limitation, only the Fourier Spectrum of surface soil acceleration responses above the middle point of driving box A, driven box C and driving box B under Case 1 (uniform excitation) and Case 2 (non-uniform excitation) are depicted in this presented paper. Fig. 8 and Fig. 9 show the Fourier Spectrum of soil acceleration responses under Case 1 and Case 2, respectively. Under uniform excitation, the frequency contents of soil acceleration response in different locations are identical along the longitudinal direction (Fig. 8). There are

some differences among the frequency contents of soil acceleration response in different locations due to non-uniform excitation (Fig. 9). For example, the predominant frequencies of soil acceleration responses of driving box A and driving box B are 1.56 and 0.73Hz, respectively.

Under both uniform excitation and non-uniform excitation Cases, the frequency contents of soil acceleration response of 3D dynamic analysis with continuum model box are basically identical with that of 2D free field analysis. It means continuum model box has little influence on the frequency contents of soil acceleration response. The effectiveness of continuum model box is demonstrated.

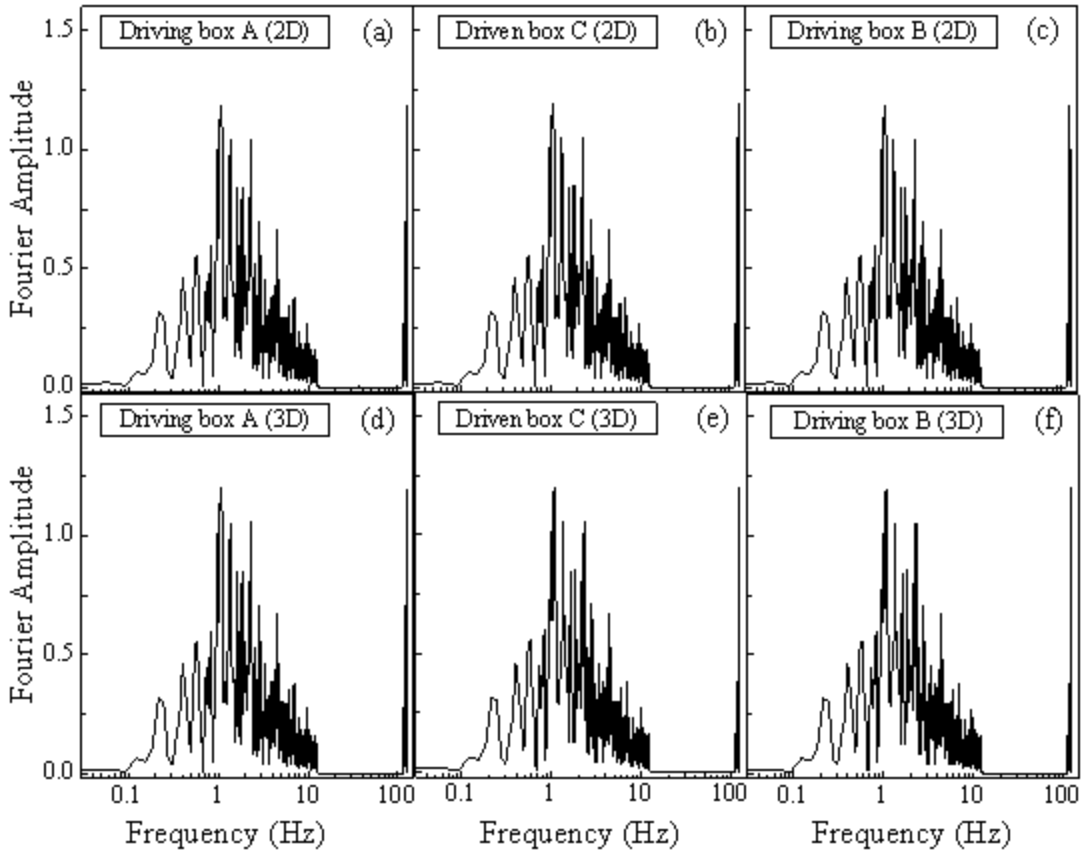


Fig. 8. Fourier Spectrum of soil acceleration response under uniform excitation (Case 1)

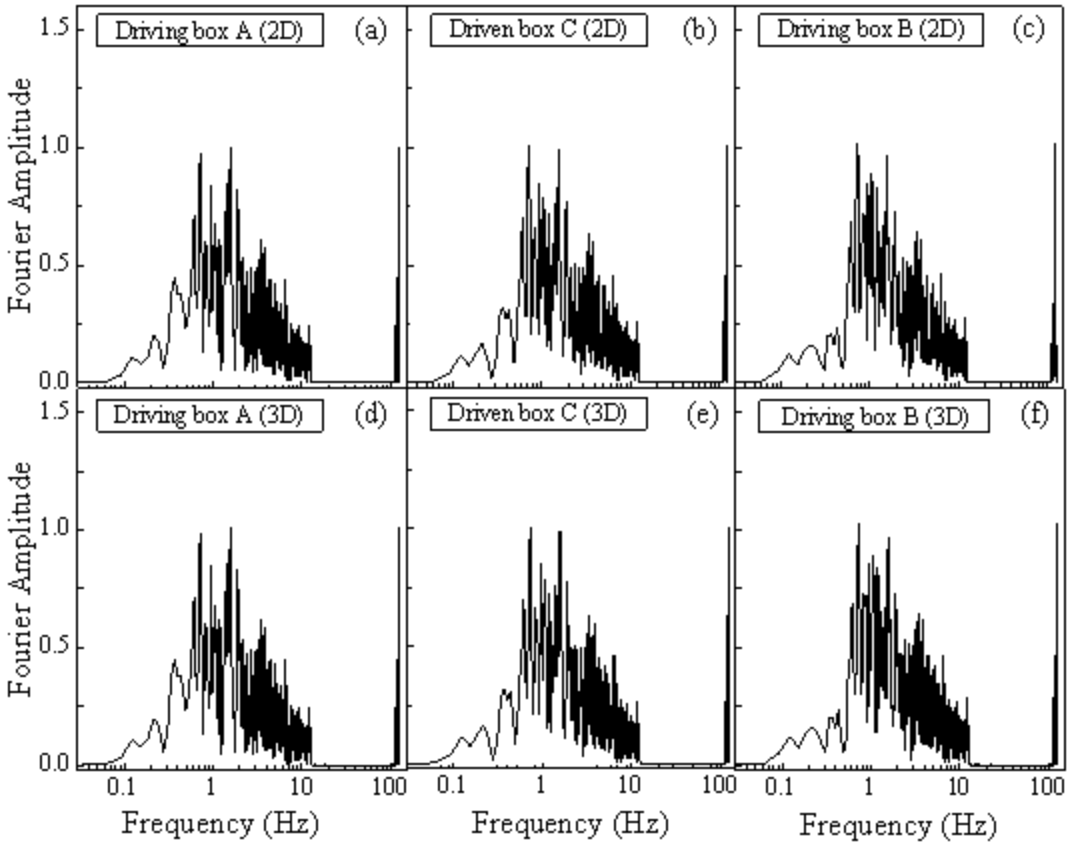


Fig. 9. Fourier Spectrum of soil acceleration response under non-uniform excitation (Case 2)

5 Conclusion

The goal of this presented paper is to verify the effectiveness of continuum model box connecting different shaking tables. A full-scale 3D entity finite element model of soil and model box is simulated to study, and 2D free field analysis is conducted as a reference standard. To avoid the randomness of calculation results, three conventional coherency models are adopted to synthetize non-uniform ground motions as input excitations, respectively. The calculation results, including the distributions of peak acceleration response in longitudinal and vertical direction and Fourier Spectrum of soil acceleration response, show that continuum model box has very limited influence on soil acceleration responses. The effectiveness of continuum model box connecting different shaking tables is verified. In the end, the conclusion of the presented paper could be valuable to the non-uniform excitation shaking table test of underground structure.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 41472246) and the Fundamental Research Funds for the Central Universities. All support is gratefully acknowledged.

References

[1] Santa-Cruz, S., Heredia-Zavoni, E. and Harichandran, R. S. (2000) Low-frequency behavior of coherency for strong ground motions in Mexico City and Japan, In: *Proceedings 12th World Conference on Earthquake Engineering*, Auckland, New Zealand, Paper No. 0076.

[2] Der, Kiureghian, A. (1996) A coherency model for spatially varying ground motions, *Earthquake Engineering and Structural Dynamic* **25**, 99-111.

[3] Hashash, Y., Tseng, W.S. and Krimotat, A. (1998) Seismic soil-structure interaction analysis for immersed tube tunnels retrofit, ASCE *Geotechnical Earthquake Engineering and Soil Dynamics* **III**, 1380-1391.

[4] Anastasopoulos, I., Gerolymos, N., Drosos, V., Kourkoulis, R., Georgarakos, T. and Gazetas, G. (2007) Nonlinear response of deep immersed tunnel to strong seismic shaking, *Journal of Geotechnical and Geoenvironmental Engineering* **133**, 1067-1090.

[5] Park, D., Sagong, M., Kwak, D. Y. and Jeong, C. G. (2009) Simulation of tunnel response under spatially varying ground motion, *Soil Dynamics and Earthquake Engineering* **29**, 1417-1424.

[6] Yu, H. T., Yuan, Y., Qiao, Z. Z., Yun, G., Yang, Z. H. and Li, X. D. (2013) Seismic analysis of a long tunnel based on multi-scale method, *Engineering Structures* **49**, 572-587.

[7] Li, P. and Song, E. X. (2015) Three-dimensional numerical analysis for the longitudinal seismic response of tunnels under an asynchronous wave input, *Computers and Geotechnics* **63**, 229-243.

[8] Tamari, Y. and Towhata, I. (2003) Seismic soil–structure interaction of cross sections of flexible underground structures subjected to soil liquefaction, *Soils and Foundations* **43**, 69-87.

[9] Chen, G. X., Wang, Z. H., Zuo, X., Du, X. L. and Gao, H. M. (2013) Shaking table test on the seismic failure characteristics of a subway station structure on liquefiable ground, *Earthquake Engineering and Structural Dynamics* **42**, 1489-1507.

[10] Moss, R. E. S. and Crosariol, V. A. (2013) Scale model shake table testing of an underground tunnel cross section in soft clay, *Earthquake Spectra* **29**, 1413-1440.

[11] Gopal, Madabhushi, S. P. (2004) Modelling of earthquake damage using geotechnical centrifuges, *Geotechnics and Earthquake* **87**, 10-25.

[12] Pitilakis, D., Dietz, M., Wood, D. M., Clouteau, D. and Modaressi, A. (2008) Numerical simulation of dynamic soil-structure interaction in shaking table testing, *Soil Dynamics and Earthquake Engineering* **28**, 453-467.

[13] Chen, J., Shi, X. J. and Li, J. (2010) Shaking table test of utility tunnel under non-uniform earthquake wave excitation, *Soil Dynamics and Earthquake Engineering* **30**, 1400-1416.

[14] ABAQUS, Inc. (2010) *ABAQUS/Analysis user's manual—version 6.9., Providence*, **RI 02909-2499**, USA.

[15] Hindy, A. and Novak, M. (1980) Pipeline response to random ground motion, ASCE, *Journal of Engineering Mechanics* **106**, 339-360.

[16] Harichandran, R. S. and Vanmarcke, E. H. (1986) Stochastic variation of earthquake ground motion in space and time. ASCE, *Journal of Engineering Mechanics* **112**, 154-174.

[17] Harichandran, R. S. (1991) Estimating the spatial variation of earthquake ground motion from dense array recordings, *Structural Safety* **10**, 219-233.

[18] Qu, T. J., Wang, J. J. and Wang, Q. X. (1996) A Practical Model for the Power Spectrum of SpatiallyVariant Ground Motion, *Acta Seismologica Sinica* **9**, 69-80.

[19] Hao, H. (1989) Effects of spatial variation of ground motions on large multiple supported structures, University of California, Berkeley, USA.

[20] Clough, R. W. and Penzien, J. (1993) *Dynamics of Structures*, **2nd edition**, McGraw-Hill, Inc., New York.

[21] Amin, M. and Ang, A. H. S. (1968) Non-stationary stochastic model of earthquake motions, ASCE, *Journal of Engineering Mechanics* **94**, 559-583.

A reliability optimization allocation method considering differentiation of functions Based on Goal Oriented method

X. J. Yi[1, 2], *†N. H. Mu[1, 3], P. Hou[1], and Y. H. Lai[1]

[1]School of Mechatronical Engineering, Beijing Institute of Technology, China
[2] Department of Mechanical Engineering, University of Ottawa, Ottawa, Canada
[3] Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA

*Presenting author: mhnzhy@126.com
†Corresponding author: mhnzhy@126.com

Abstract

A new reliability optimization allocation for multifunction systems considering differentiation of functions based on GO methodology is proposed in this paper. First, constraints considering differentiations of functions are proposed based on GO method, which are function importance factor constraint, and system reliability constraint, respectively. Then, the objective function of optimization allocation problem is built to minimize the system cost. Based on above, the mathematic model of reliability optimization allocation problem for multifunction systems considering differentiations of functions is established. In addition, an improved Ant Colony Optimization (ACO) is proposed to solve this mathematic model. Furthermore, the process of the new method is formulated. Finally, the new method is applied in reliability optimization allocation of Power-Shift Steering Transmission whose goal is to minimize the system cost. Compared with the results by using basic ACO, it is shown that the new method is reasonable, advantageous, and feasible for the reliability optimization allocation problem with differentiation of functions. Clearly, this study solves the disadvantages of the existing reliability optimization allocation methods efficiently so that it can quickly, efficiently, and directly allocate the system reliability index to design units for complex systems. All in all, this paper not only provides a new approach to conduct reliability optimization allocation for multifunction systems considering differentiation of functions, but also improves the theory and widens the application of GO methodology. In addition, this paper can also provide guidance for the similar reliability optimization problem

Keywords: reliability optimization, differentiation of functions, importance factor, multifunction systems, Ant Colony Optimization

Introduction

The aim of reliability optimization allocation is that the system reliability index is allocated to design units considering restrictions, which are cost, size, and weight etc., in order to provide guidance for reliability design of product. Nowadays, a large number of studies on reliability optimization allocation are mainly as follows: (i) fault-tolerance mechanism, (ii) active and cold-standby redundancy, (iii) optimization techniques, (iv) multi-objective optimization, (v) optimization techniques: [Kuo et al., (2007)]. With development of technology, the multifunction systems are often applied in Engineering, and have a key role. While, a large

number of research works of reliability allocation for multifunction systems are only considered one single main function of system, and ignored other functions. Clearly, it will lead to an unreasonable and a bias of reliability allocation result. Thus, some researchers are focus on the reliability optimization of multifunction systems. Lim et al. proposed the allocation of the equipment path in a multi-stage manufacturing process: [Lim et al. (2015)]. An improved AGREE method is proposed to solve reliability allocation of multi-mission networked avionic system without considering resource constraint and system structure: [Li et al. (2015)]. For reliability optimization allocation of multifunction systems, Yi et al. proposed the reliability optimization allocation method for units designed and units selected versions [Yi et al. (2015a-b); Yi et al. (2016a)]. While, above reliability optimization allocation methods have three disadvantages, as follows: (i) The product is finalized production through multiple design revisions, but the above method is difficult to conduct reliability re-allocation quickly and efficiently at the situation of design changes, (ii) The reliability models used in above methods are hard to reflect product structure, working principles, (iii) It is difficult to quickly, efficiently, and directly allocate the system reliability index to design units for complex systems containing series structure and redundant structure. In addition, the optimization technologies, such as genetic algorithm, ant colony algorithm, and neural network algorithm etc. are used to solve the problem of reliability optimization allocation effectively. And Kuo et al. overviewed the optimization techniques for reliability optimization allocation: [Kuo and Rin (2007)]. And there are three concerns of the optimization technologies for reliability optimization allocation problem, as follows: (i) To obtain satisfactory convergence effects and efficiency, (ii) To avoid local extremum problem, (iii) For specific problems, the basic optimization algorithm need to improve. It is meaningful to improve the basic algorithm so that it is applicable for specific problems and can obtain the optimal solution efficiently: [Wang and Lee (2015); Alavidoost et al. (2015); Zhao et al. (2015); Yi et al. (2016a)].

Above all, a reliability optimization allocation method for multifunction systems is described through develop reliability optimization allocation problem, and solve this optimization allocation problem by optimization technologies. Goal Oriented (GO) methodology is a success-oriented method for reliability analysis of complex systems [Yi et al. (2014a-b); Yi et al. (2015c-e); Yi et al. (2016b)]. Moreover, the reliability analysis results can be obtained through GO operation according to GO algorithm and GO model. The GO model is developed directly using product schematic diagrams, its structure, and its functional hierarchy. And GO algorithm has a high efficiency and easy to operate. Thus, GO method can be suitable for reliability optimization allocation to overcome above disadvantages of the existing reliability optimization allocation method. Furthermore, Ant Colony Optimization (ACO) has been used widely.

In view of advantages of GO method in aspects of establishing system model and reliability analysis, a reliability optimization allocation for multifunction systems considering differentiation of functions based on GO methodology is firstly proposed in this paper. First, constraints considering differentiations of functions are proposed, which are function importance factor constraint, and system reliability constraint, respectively. The function

importance factor constraint is consist of the predicted function importance factors by using allocated reliability of unit based on GO method, and the allocated function importance factors. And the system reliability constraint function is consist of the target reliability of system, and the predicted reliability of system by using allocated reliability of unit based on GO method. Then, the objective function of optimization allocation problem is built to minimize the system cost. Based on above, the mathematic model of reliability optimization allocation problem for multifunction systems considering differentiations of functions is established. In addition, an improved ACO is proposed to solve this mathematic model. Furthermore, the process of the new method is formulated. Finally, the new method is applied in reliability optimization allocation of Power-Shift Steering Transmission (PSST) whose goal is to minimize the system cost. To verify the advantages and engineering applicability of the new method, the results of the new method are compared with the results by using basic ACO.

Reliability optimization allocation method considering differentiation of functions based on GO method

A reliability optimization allocation for multifunction systems considering differentiation of functions based on GO method is proposed in aspects of description of reliability optimization allocation problem, and solving algorithm.

Description of reliability optimization allocation problem

The reliability optimization allocation problem is described through corresponding mathematic model, which contains constraints considering the differentiation of functions, objective function.

Constraints considering the differentiation of functions

(1) Reliability of function and system based on GO operation

For reliability optimization allocation of multifunction systems considering differentiation of functions, the reliabilities of function and system can be obtained by using the success probability of design unit and GO algorithm to conduct GO operation according to GO model. Thus, GO model and GO operation are key elements in GO method. GO model is directly using product schematic diagrams, its structure, and its functional hierarchy, and is consist of GO operator and signal flow. GO operator represents design unit and logical relation in system, and signal flow represents the specific fluid, logical process, and the direction of GO operation. GO algorithm is key element of GO operation, and there are two kinds of GO algorithm, which are suitable for GO model with shared signal flow [Shen et al. (2000)], and GO model without containing shared signal flow [Shen et al. (2001)].

Therefore, the reliability of signal flow based on GO method for reliability optimization allocation is defined, as follows:

$$R_x = F_x(R_{x1}, R_{x2}, \cdots, R_{xy}) \tag{1}$$

where, R_x is the predicted reliability of *xth* signal flow, R_{xy} is allocated reliability of *yth* unit for calculating *xth* signal flow, $F_x(\cdot)$ is the success probability of *xth* signal flow obtained by using allocated reliability of unit to conduct GO operation according to GO model and GO algorithm.

In GO model, the reliability of output signal flows for functions and system represents the reliability of system functions and system.

(2) Constraints considering the differentiation of functions

To deal with the differentiation of functions in the process of reliability allocation, the function importance factor constraint is proposed combined the predicted function importance factors by using allocated reliability of unit based on GO method with the allocated function importance factors, and the system reliability constraint are proposed combined the predicted reliability of system by using allocated reliability of unit based on GO method with target reliability of system. Assumed that a multifunction system is consists of *m* units, and can execute *n* functions.

The higher importance factor of function, the higher reliability should be allocated. The function importance factor constraint indicates predicted function importance factor by using allocated reliability of unit based on GO method should meet the allocated function importance factor, as shown in Eq. (2).

$$\begin{cases} R_{Gw} = F_w(R_{w1}, R_{w2}, \cdots, R_{wj}) \\ g(R_{Gw}) = e^{(R_{Gw}-1)} \geq g(R_{Gw}^*) = e^{(R_{Gw}^*-1)} \end{cases} \tag{2}$$

where, R_{Gw} is the predicted reliability of *wth* function, R_{wj} is allocated reliability of *jth* unit for *wth* function, $F_w(\cdot)$ is the reliability of output signal flow represented *wth* function by using allocated reliability of unit based on GO method, $g(R_{Gw})$ is the predicted function importance factor of *wth* function, $g(R_{Gw}^*)$ is the allocated function importance factor by the estimation of function importance factors [Yi et al. (2016a)], $w = 1, 2, \cdots, n$, $1 \leq j \leq m$.

The system reliability constraint indicates predicted reliability of system by using allocated reliability of unit based on GO method should meet the target reliability of system, as shown in Eq. (3).

$$\begin{cases} R_S = F_S(R_1, R_2, \cdots, R_m) \\ R_S \geq R_S^* \end{cases} \tag{3}$$

where, R_S is the predicted reliability of system, R_i is the allocated reliability of *ith* unit, $F_S(\cdot)$ is the reliability of output signal flow of system by using allocated reliability of unit based on GO method, R_S^* is the target reliability of system, $i = 1, 2, \cdots, m$.

Objective function of reliability optimization allocation

The cost of system is greatly concerned, so the objective function of reliability optimization allocation in this paper is to minimize the cost, as follows:

$$\min C_S(R) = \sum_{i=1}^{m} c_i\left(P_i, R_i, R_{i,\min}, R_{i,\max}\right) \tag{4}$$

where, $C_S(\cdot)$ is the cost function of system; $c_i\left(P_i, R_i, R_{i,\min}, R_{i,\max}\right)$ is the cost function of design unit, i.e. $c_i\left(P_i, R_i, R_{i,\min}\right) = P_i e^{\left(\frac{R_i - R_{i,\max}}{R_{i,\min} - R_{i,\max}}\right)}$, P_i is the basic cost of *ith* unit, R_i is allocated reliability of *ith* unit, $R_{i,\min}$ is the lower limit value of reliability of *ith* unit.

Mathematic model of reliability optimization allocation

Combining above the objective function and constraints, the reliability optimization allocation problem with differentiation of functions can be given by

$$\begin{cases} \min C_S(R) = \sum_{i=1}^{m} P_i e^{\left(\frac{R_i}{R_{i,\min}} - 1\right)} \\ s.t. \\ \qquad R_{i,\min} \le R_i \le R_{i,\max} \qquad i = 1, 2, \cdots, m \\ g(R_{Gw}) = g(F_w(R_{w1}, R_{w2}, \cdots, R_{wj})) \ge g(R_{Gw}^*) \quad w = 1, 2, \cdots, n \\ R_S = F_S(R_1, R_2, \cdots, R_m) \ge R_S^* \end{cases} \tag{5}$$

Improved AOC for reliability optimization allocation problem

In order to obtain the satisfactory results effectively, the basic AOC is improved for solving the reliability optimization allocation problem with differentiation of functions. Its steps are as follows:

(1) Establishing ant colony path diagram

All of the directed paths allowed the ant individual to walk constitute the ant colony path diagram. Each directed path corresponds to an optimization allocated results, and each node

value corresponds to an allocated reliability of design unit. The ant colony path diagram is shown in Figure 1.

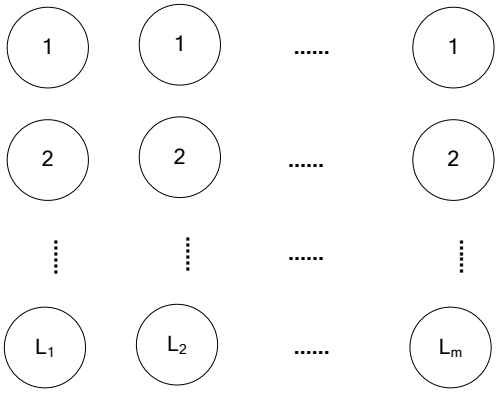


Figure 1. Ant colony path diagram

In Figure 1, the node values of each column represent the selectable allocated results of corresponding design unit. And the number of node can be obtained by Eq. (6).

$$N_i = \frac{R_{i,\max} - R_{i,\min}}{L} \tag{6}$$

where, N_i is the number of node of *i-th* column, L is the step length of node interval.

The ant colony path diagram can be represented by the cell array, as follows:

$$R = \begin{Bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,\mathrm{m}} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ R_{L1,1} & R_{L2,2} & \cdots & R_{Lm,m} \end{Bmatrix} \tag{7}$$

where, $R_{j,i}$ is the *j-th* selectable allocated result of *i-th* unit, $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, L_i$.

(2) Initializing pheromone path diagram

According to ant colony path diagram, the corresponding pheromone path diagram can be established. The carrier of pheromone is the moving path of ant individual. And the pheromone concentration of the moving path for ant individual correspond the quality of objective function value for such moving path. The pheromone diagram will update with the change of the number of iterations, and the pheromone path diagram can be represented by the cell array, as follows:

$$\tau(Loop) = \begin{Bmatrix} \tau_{1,1} & \tau_{1,2} & \cdots & \tau_{1,\mathrm{m}} \\ \tau_{2,1} & \tau_{2,2} & \cdots & \tau_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ \tau_{L1,1} & \tau_{L2,2} & \cdots & \tau_{Lm,m} \end{Bmatrix} \tag{8}$$

where, $\tau_{j,i}$ is the pheromone element of the *j-th* selectable allocated result of *i-th* unit; Loop is the iterations times, when $Loop = 1$, the pheromone path diagram is the initialization pheromone path diagram, and $\tau_{j,i} = 1$, $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, l_i$.

(3) Ant colony moving

The process of formation path for each ant is defined as ant colony moving. Each path correspond a reliability allocated result. And the reliability allocated result is determined by the pheromone path diagram and the cost of each node. The reliability allocated result is obtained as follows:

(i) To establish the node probability diagram in order to represent the selected probability for ant individual in the node. The node probability diagram can be represented by the cell array, as follows:

$$P = \begin{Bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,m} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ P_{L1,1} & P_{L2,2} & \cdots & P_{Lm,m} \end{Bmatrix} \tag{9}$$

where, $P_{i,j}$ is the selected probability of the *j-th* selectable allocated result of *i-th* unit, $P_{i,j} = \dfrac{\tau_{i,j} \cdot \dfrac{1}{C_{i,j}}}{\sum_{j=1}^{L_i} \tau_{i,j} \cdot \dfrac{1}{C_{i,j}}}$, $C_{i,j}$ is the cost of *j-th* selectable allocated result of *i-th* unit.

(ii) To obtain the reliability allocated result by using the roulette wheel method to select node of each column in ant colony path diagram.

(4) Constraint IF and solving the objective function

After the ant colony moving, the reliability allocated result obtained by each ant individual needs to judge if it meets the constraints based on GO method. If it meets the constraints, setting constrain value is 1, i.e., $constrain = 1$; otherwise, setting $constrain = 0$. Then, the ant individual corresponding to the minimum value of objective function is determined among the ant individuals Satisfied the constraint.

(5) Updating pheromone path diagram

When making the next iteration computation, the pheromone path diagram needs to update in order to improve the convergence efficiency and obtain the satisfactory results. The approach of updating pheromone path diagram is as follows:

(i) For the ant individual corresponding to the minimum value of objective function in the previous iteration, the formula of updating pheromone is given by

$$\tau_{i,j}(Loop+1) = constrain \cdot \tau_{i,j}(Loop) + constrain \cdot (\frac{X}{C}) \quad (10)$$

where, X is the convergence operator, C is the cost of such ant individual.

(ii) For other ant individuals, the formula of updating pheromone is given by

$$\tau_{i,j}(Loop+1) = constrain \cdot \tau_{i,j}(Loop) \quad (11)$$

(6) Judging the termination condition

The iteration times as the termination condition of the improve ACO. If it meets the termination condition, the optimal allocation results and system cost will be output. And if it does not meet the termination condition, it will operate the algorithm from Step (3): Ant colony moving.

Above all, the operation process of improved ACO is shown in Figure 2.

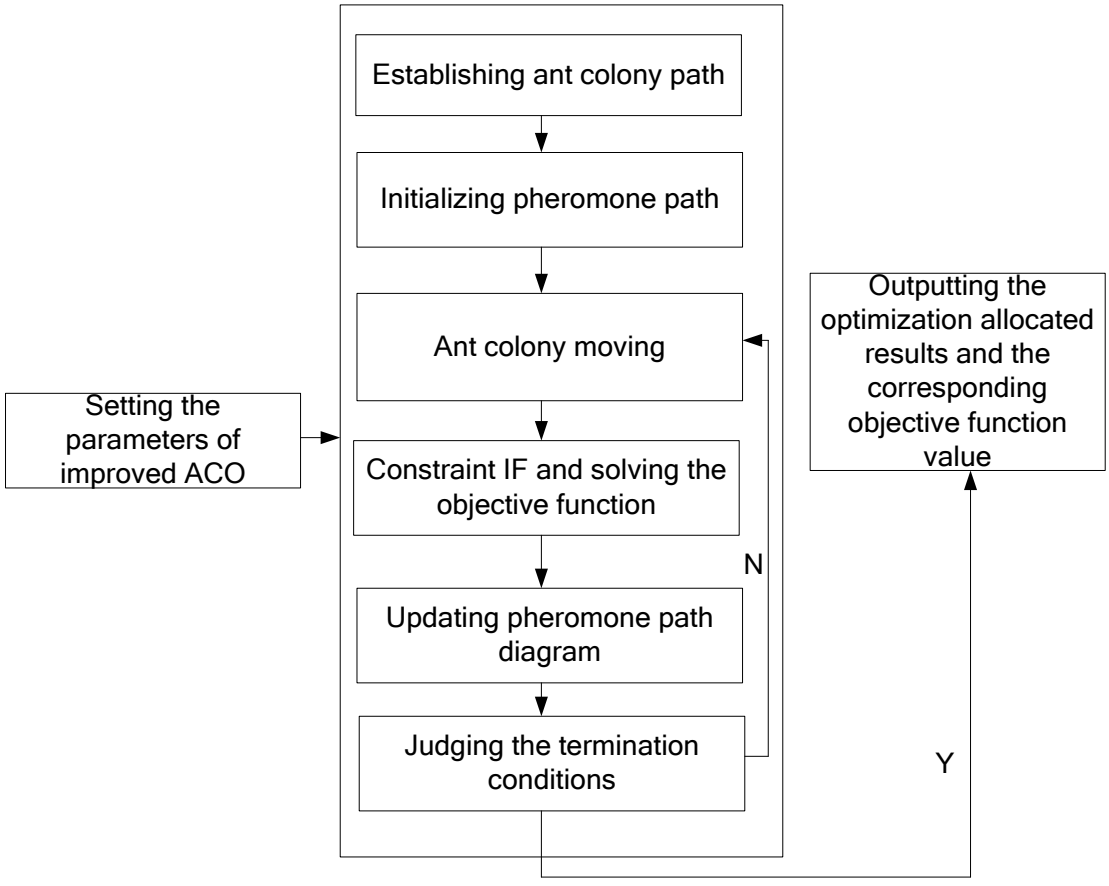


Figure 2. Operation process of improved ACO.

Reliability optimization allocation process considering differentiation of functions based on GO method

For systems with differentiation of functions, the process of reliability optimization allocation under the goal of minimizing the system cost, based on GO method, are formulated, as shown in Figure 3.

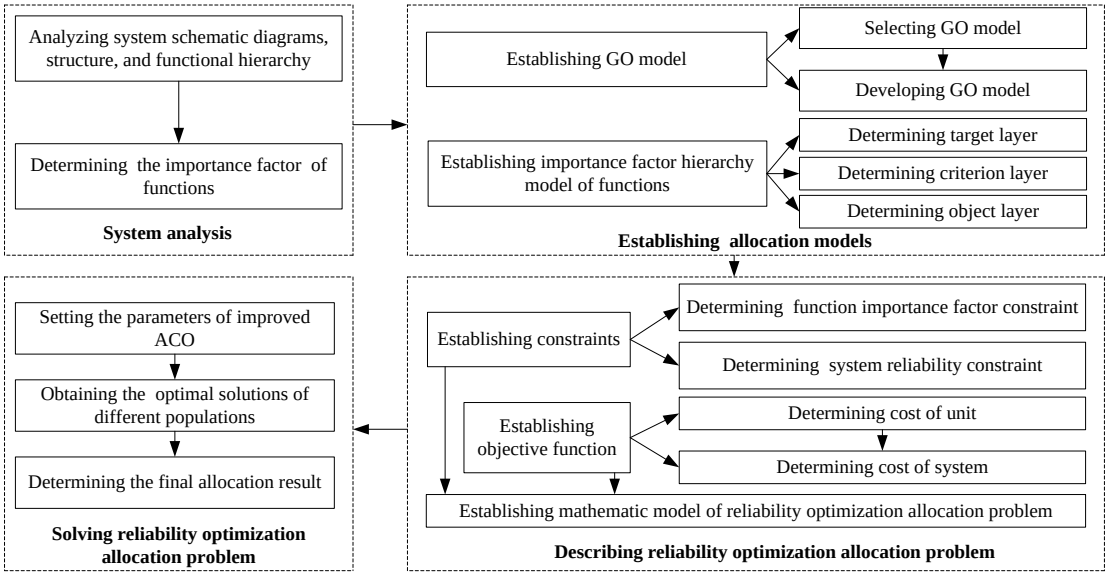


Figure 3. Reliability optimization allocation process for systems with differentiation of functions under the goal of minimizing the system cost based on GO method.

Example

The reliability optimization allocation of PSST considering differentiations of functions under the goal of minimize the system cost is conduct by the new method proposed in this paper in order to illustrate its feasibility and advantage. In order to show conveniently and compare with other results, we assume that:

(1) The PSST is very complex system, which is consist of hundreds of units in 16 subsystems, so the system reliability is allocated to units of the oil supple systems, and other subsystems in this paper. In addition, the tube and interface of system is not considered in the oil supple systems.

(2) The four main functions of PSST is considering in this paper. They are straight driving function, steering function, braking function, and fan cooling function, respectively.

(3) The basic costs of unit are set 15.

System analysis of PSST

(1) Analyzing system working principle and function

The PSST is consist of 16 subsystems, as shown in Figure 4.

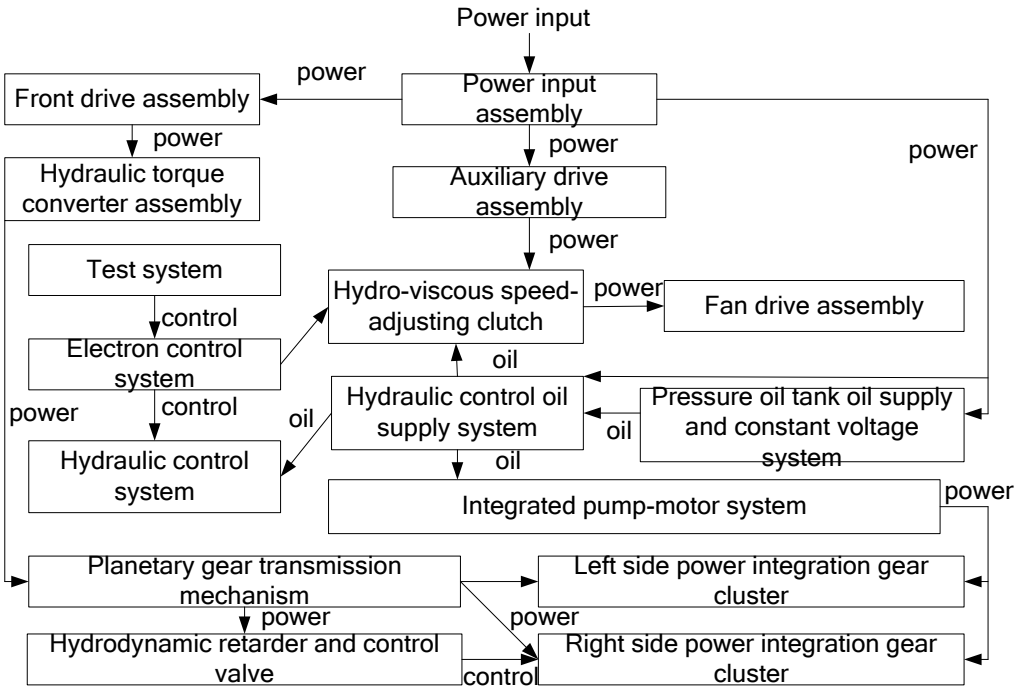


Figure 4. Working principle diagram of PSST.

The oil supply systems contain pressure oil tank oil supply and constant voltage system, and hydraulic control oil supply system. And the structure diagram of oil supply systems is shown in Figure 5.

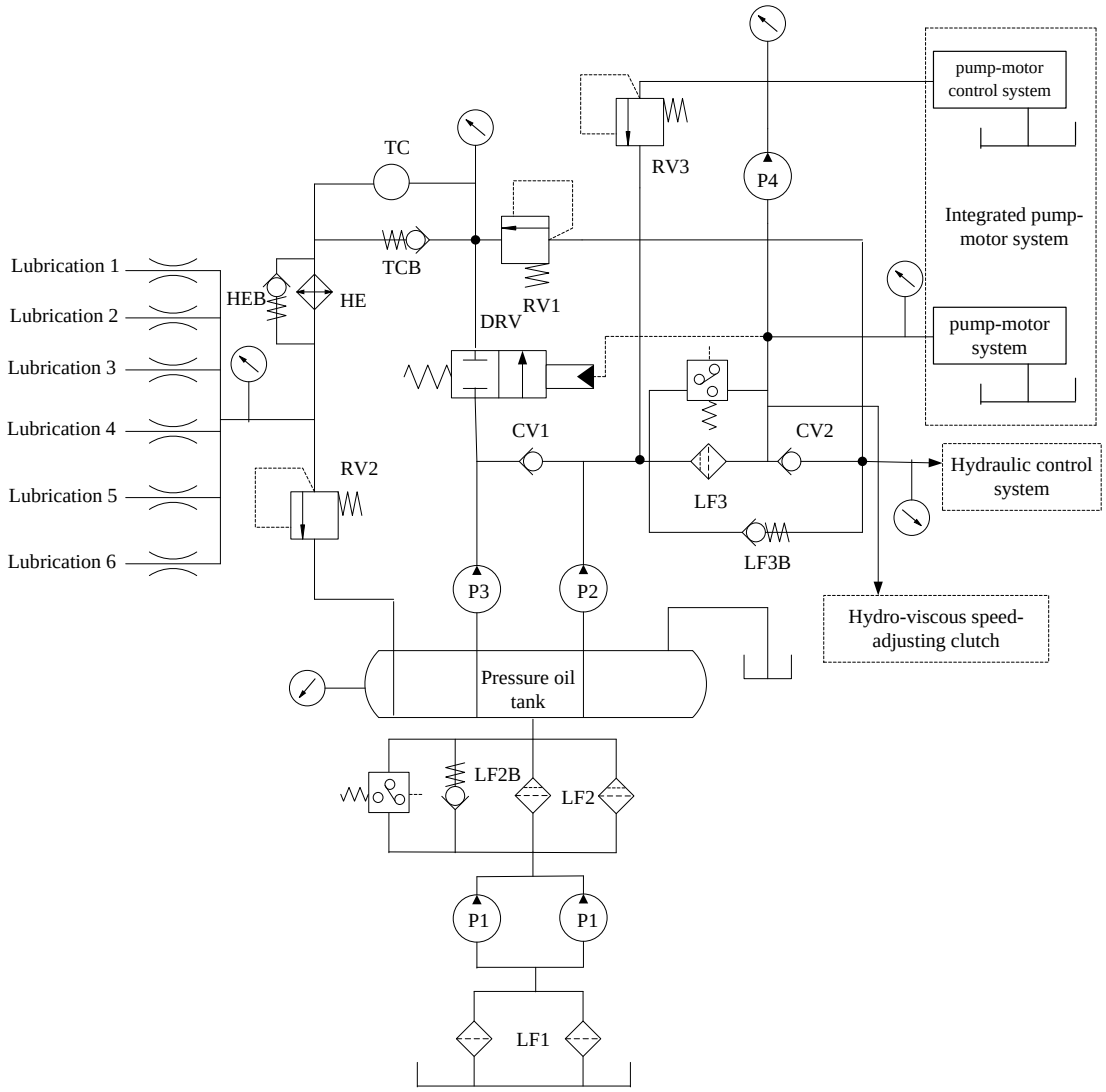


Figure 5. Structure diagram of oil supply systems.

(2) Determining the important factor of function

Only the four functions of PSST all meet the requirements in terms of importance factor, the system can be denoted as success. The reliability of function is affected by working time, functional property, and the design level. Thus, the target layer A corresponds to the system reliability index, the criterion layer C corresponds to the various factors, and the object layer P corresponds to the functions of system.

Establishing allocation models of PSST

(1) Establishing GO model of PSST

According to the analysis result of PSST, the GO operator types are corresponding function description are presented in Tab. 1. In Tab. 1, there are 6 basic GO operators, Type 5' represents virtual input signal, whose success probability is 1, Type 5 represents input unit, Type 1 represents two-state unit, Type 6 represents unit controlled by two signals, Type 2 represents logical relation of OR, Type 10 represents logical relation of AND: [Shen et al.

(2002)]. In Tab. 1, Type 22 represents multiple-Input and multi-function unit, Type 15B represents multi-conditions control signal of multiple-Input and multi-function unit: [Yi et al. (2015b)].

Table 1. GO operator type in GO model

NO. (operator)	NO. (unit)	Component	Type	NO. (operator)	NO. (unit)	Component	Type
1	1	Power input	5'	25	22	HE	1
2	2	Power input assembly	1	26	23	HEB	1
3	3	Front drive assembly	1	27	—	OR gate	2
4	4	Hydraulic torque converter assembly	6	28	24	RV2	1
5	5	Planetary gear transmission mechanism	6	29	—	AND gate	10
6	6	Auxiliary drive assembly	1	30	25	LF3	1
7	7	Hydro-viscous speed-adjusting clutch	22	31	26	LF3B	1
8	8	Fan drive assembly	1	32	27	CV2	1
9	9	Oil pan	5	33	—	OR gate	2
10	10	LF1	1	34	28	RV1	1
11	11	LF1	1	35	29	P4	6
12	—	OR gate	2	36	30	RV3	1
13	12	P1	6	37	31	Integrated pump-motor system	6
14	13	P1	6	38	32	Hydraulic control system	6
15	—	OR gate	2	39	33	Test system	5
16	14	LF2	1	40	34	Electron control system	1
17	15	LF2	1	41	—	Auxiliary operator	15B
18	16	LF2B	1	42	35	Hydrodynamic retarder and control valve	1
19	—	OR gate	2	43	—	Auxiliary operator	15B
20	17	Oil tank	1	44	36	Left side power integration gear	22

21	18	P3	6	45	37	cluster Right side power integration gear cluster	22
22	19	P2	6	46	—	AND gate	10
23	20	DRV	1	47	—	AND gate	10
24	21	TCB	1	48	—	AND gate	10

According to Figure 4, Figure 5, and Tab.1, the GO model of PSST is developed from system input to system output, as shown in Figure 6. In operators of the GO model, the former number is type of operator, and the latter number is a serial number. The number on a signal flow is serial number of signal flow. Signal flows 8, 40, 46, 47, and 48 represent output of fan cooling, output of breaking, output of steering function, output of straight driving function, and system output, respectively.

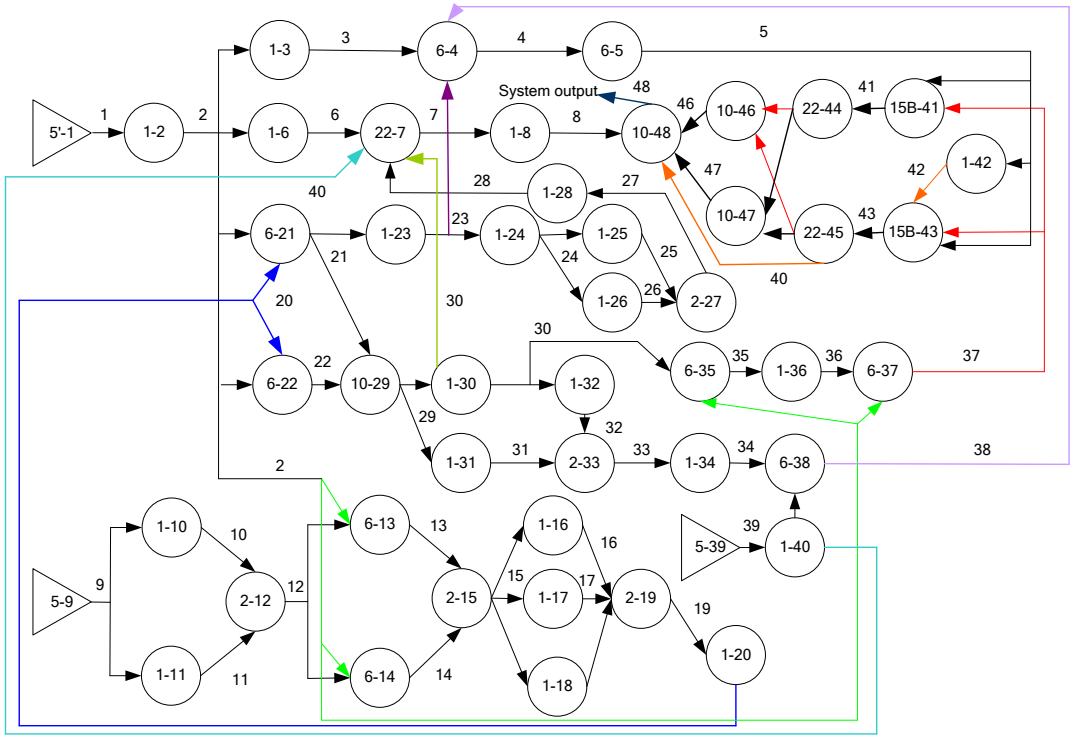


Figure 6. Structure diagram of oil supply systems.

(2) Establishing importance factor hierarchy model of functions

According to important factor analysis of function, the corresponding importance factor hierarchy model of functions is shown as Figure 7.

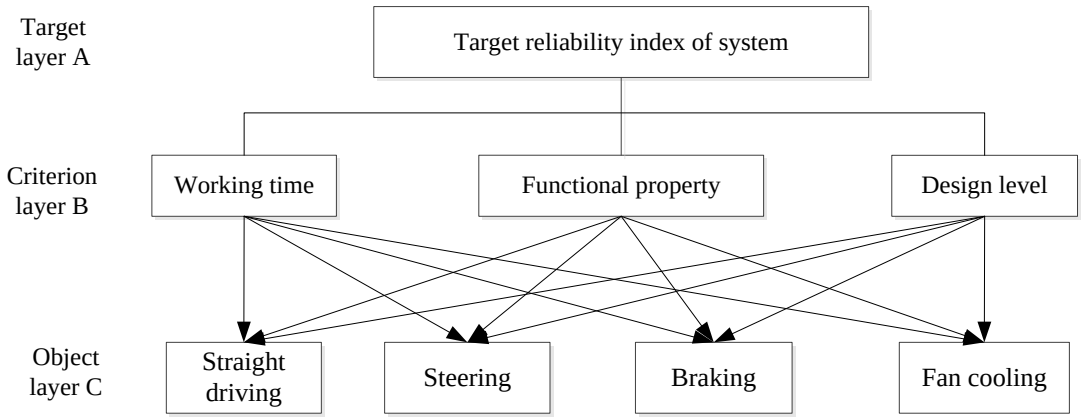


Figure 7. Importance factor hierarchy model of functions.

Describing reliability optimization allocation problem

(1) Establishing constraints

According to Eq. (2), Figure 6 and Figure 7, the function importance factor constraint based on GO method is obtained by Eq. (12).

$$\begin{cases} g(R_{G1}) = e^{(P_{S8}-1)} \geq g(R_{G1}^*) = 0.993 \\ g(R_{G2}) = e^{(P_{S40}-1)} \geq g(R_{G2}^*) = 0.9905 \\ g(R_{G3}) = e^{(P_{S46}-1)} \geq g(R_{G3}^*) = 0.9819 \\ g(R_{G4}) = e^{(P_{S47}-1)} \geq g(R_{G4}^*) = 0.9818 \end{cases} \tag{12}$$

where, $g(R_{G1})$, $g(R_{G2})$, $g(R_{G3})$, and $g(R_{G4})$ are the predicted function importance factors of fan cooling function, breaking function, steering function, and straight driving function, respectively; P_{S8}, P_{S40}, P_{S46}, and P_{S47} are reliability of signal flow 8, 40, 46, 47 in GO model, respectively; $g(R_{G1}^*)$, $g(R_{G2}^*)$, $g(R_{G3}^*)$, and $g(R_{G4}^*)$ are the allocated function importance factors by the estimation of function importance factors [Yi et al. (2016a)].

According to Eq. (3), the system reliability constraint based on GO method is obtained by Eq. (13).

$$R_S = P_{S48} \geq R_S^* = 0.951 \tag{13}$$

where, R_S is the predicted reliability of system based on GO method, i.e. P_{S48}, R_S^* is the target reliability of system.

(2) Establishing objective function

The objective function of reliability optimization allocation in this paper is to minimize the cost, as follows:

$$\min C_S = \sum_{i=1}^{37} c_i \tag{14}$$

where, C_S and C_i are the cost of system and unit, respectively.

(3) Establishing mathematic model of reliability optimization allocation problem

According to Eq. (12), (13), and (14), the reliability optimization allocation problem with differentiation of functions is described by Eq. (15).

$$\begin{cases} \min C_S = \sum_{i=1}^{37} c_i \\ s.t. \\ g(R_{G1}) = e^{(P_{S8}-1)} \geq g(R_{G1}^*) = 0.993 \\ g(R_{G2}) = e^{(P_{S40}-1)} \geq g(R_{G2}^*) = 0.9905 \\ g(R_{G3}) = e^{(P_{S46}-1)} \geq g(R_{G3}^*) = 0.9919 \\ g(R_{G4}) = e^{(P_{S47}-1)} \geq g(R_{G4}^*) = 0.9918 \\ R_S = P_{S48} \geq R_S^* = 0.957 \\ 0.999 \leq R_i \leq 0.99999, i = 1, 2, \cdots, 37 \end{cases} \tag{15}$$

Solving reliability optimization allocation problem

According to the improved AOC proposed in this paper, the parameters of improved AOC are presented in Tab. 2, and the system cost of different iterative times are shown in Figure 8.

Table 2. The parameters of improved AOC for solving Eq. (9)

Parameter	Node	Iterative times	Ant individuals	convergence operator
Value	20	500	150	500

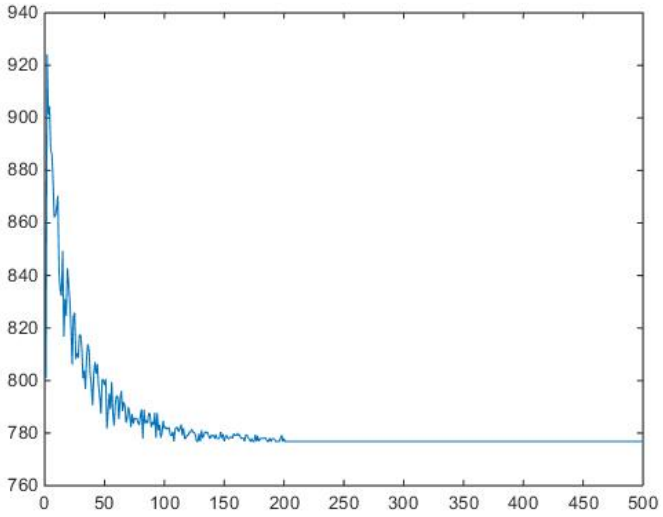


Figure 8. The system cost of different iterative times by improved ACO.

According to Figure 8, the solution convergence is at the 200th convergence time, and the system cost is 7.768215814845887e+02. The allocated reliabilities of corresponding design units are presented in Tab. 3.

Table 3. The optimization allocated reliabilities of design units by improved ACO

NO. (unit)	Reliability	NO. (unit)	Reliability	NO. (unit)	Reliability
1	0.999885789	14	0.999104211	27	0.999260526
2	0.999312632	15	0.999052105	28	0.999521053
3	0.999468947	16	0.999104211	29	0.999100000
4	0.999208421	17	0.999521053	30	0.999729474
5	0.999156316	18	0.999364737	31	0.999052105
6	0.999208421	19	0.999729000	32	0.999312632
7	0.999000000	20	0.999261000	33	0.999208421
8	0.999625263	21	0.999417000	34	0.999260526
9	0.999364737	22	0.999000000	35	0.999208421
10	0.999260526	23	0.999000000	36	0.999364737
11	0.999468947	24	0.999469000	37	0.999260526
12	0.999000000	25	0.999208000		
13	0.999521053	26	0.999469000		

4 Result Analysis

In order to illustrate feasibility and advantage of the new method, the result the results by the new method is compared with the results by using basic AOC [Nahas N. et al. (2005)]. First, setting the parameters of the node, iterative times, and ant individuals are 20, 500, and 150, respectively. Second, the reliability allocated results for each ant individual are obtained by operating algorithm according to the node transition rule. Furthermore, the feasible solution is improved according to the heuristic rule, and the non-feasible solution is improved by local

search to let it become the feasible solution as much as possible. The optimal allocated results are obtained in once iterative. Then, the pheromones are updated. If it meets the termination condition, the optimal allocation results and system cost will be output. And if it does not meet the termination condition, it will operate the algorithm again. The system cost of different iterative times is shown in Figure 9.

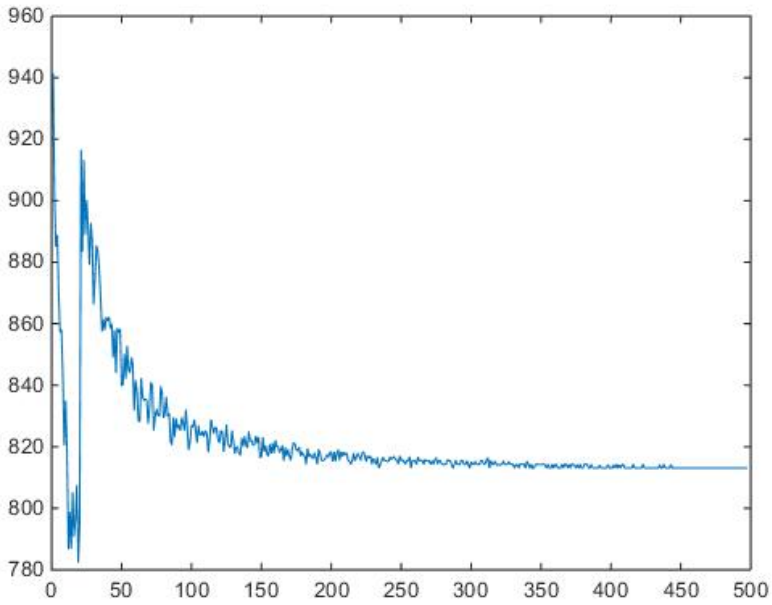


Figure 9. The system cost of different iterative times by basic ACO.

According to Figure 9, the solution convergence is at the 450th convergence time, the system cost is 8.130765482215386e+02. The allocated reliabilities of corresponding design units are presented in Tab. 4.

Table 4. The optimization allocated reliabilities of design units by improved ACO

NO. (unit)	Reliability	NO. (unit)	Reliability	NO. (unit)	Reliability
1	0.999468947	14	0.999104211	27	0.999104211
2	0.999208421	15	0.999910000	28	0.999521053
3	0.999416842	16	0.999364737	29	0.999156316
4	0.999416842	17	0.999312632	30	0.999312632
5	0.999260526	18	0.999521053	31	0.999625263
6	0.999500000	19	0.999416842	32	0.999915632
7	0.999952105	20	0.999625263	33	0.999468947
8	0.999416842	21	0.999677368	34	0.999941684
9	0.999573158	22	0.999416842	35	0.999208421
10	0.999312632	23	0.999416842	36	0.999156316
11	0.999625263	24	0.999416842	37	0.999573158
12	0.999915632	25	0.999364737		
13	0.999312632	26	0.999931263		

Tab. 3, Tab. 4, Figure 8, and Figure 9 show that:

(1) The system cost by using the basic ACO is larger than the system cost by using the improved ACO proposed in this paper. In addition, the solution convergence of improved ACO is faster than that of basic ACO. Thus, it is illustrated that the improved ACO is more effective, and more reasonable for solving the reliability optimization allocation problem with differentiation of functions.

(2) The reliability allocated results of 7^{th} unit, 15^{th} unit, 26^{th} unit, 32^{th} unit, 12^{th} unit, and 34^{th} unit by using basic ACO exceed 0.9999, which is hard to design in engineering. While, the reliability allocated results of all units by using improved ACO are less than 0.9999. Thus, it is illustrated that the improve ACO can obtain more satisfactory results, which meet the engineering practice.

(3) The analysis process of this new method shows that the new reliability optimization allocated method proposed in this paper can overcome the aforementioned disadvantages of the existing reliability optimization allocation methods efficiently so that it can quickly, efficiently, and directly allocate the system reliability index to design units for complex systems.

Conclusion

This study proposes a new reliability optimization allocation for multifunction systems considering differentiation of functions based on GO method. First, the description of reliability optimization allocation problem is proposed in aspects of constraints considering differentiations of functions based on GO method, the objective function of optimization allocation problem whose goal is minimize the system cost, and the mathematic model of reliability optimization allocation problem. Then, an improved ACO is proposed to solve above mathematic model. Furthermore, the process of the new method is formulated. Finally, the new method is applied in reliability optimization allocation of PSST whose goal is to minimize the system is cost. In order to verify the advantages and engineering applicability of the new method, the results by using improved ACO are compared with the results by using basic ACO. And the comparison results show that the new method is reasonable, advantageous, and feasible for the reliability optimization allocation problem with differentiation of functions. Clearly, this study solves the aforementioned disadvantages of the existing reliability optimization allocation methods efficiently so that it can quickly, efficiently, and directly allocate the system reliability index to design units for complex systems.

All in all, this paper not only provides a new approach to conduct reliability optimization allocation for multifunction systems considering differentiation of functions, but also improves the theory and widens the application of GO methodology. In addition, this paper can also provide guidance for the similar reliability optimization problem.

Acknowledgement

This project is supported by National Natural Science Foundation of China (NSAF Joint Funds, U1530135) in the years 2016-2018, the Ministry of Industry and Information Technology (China) in the years 2011-2014 (ZQ092012B003), and Chinese Scholarship Council in the years 2015-2016 ([2015]3012 & [2015]3022). We are grateful to the chief editor, editor and reviewers for the suggestions which improve the draft of this paper.

Reference

[1] Kuo, W. and Rin, W. (2007) Recent advances in optimal reliability allocation, *Ieee Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* **37**, 143–156.
[2] Lim, Y. B., Chung, J. and Park, C. (2015) Allocation of the equipment path in a multi-stage manufacturing process, *Journal of the Korean Statistical Society* **44**, 366–375.
[3] Li, R. Y., Wang, J. F. Liao, H. T. and Huang, N. (2015) A new method for reliability allocation of avionics connected via an airborne network. *Journal of Network and Computer Applications* **48**, 14–21.
[4] Yi X. J., Dong H. P. and Jiang J. P. (2014a) Reliability Analysis of Hydraulic Transmission Oil Supply System of Power-Shift Steering Transmission Based on GO Methodology, *Journal of Donghua University (Eng. Ed.)* **31**, 785-788.
[5] Yi, X. J., Dong, H. P., Shi, J., Bao, K. and Jiang, J. P. (2014b) Reliability analysis of hydraulic transmission oil supply system considering maintenance correlation of parallel structure based on GO methodology. *In: Reliability, Maintainability and Safety (ICRMS), 2014 International Conference on IEEE*, 508-513.
[6] Yi X. J., Lai Y. H., Dong H. P. and Hou P. (2015a) A reliability optimization allocation method considering differentiation of functions, *Proceedings of the International Conference on Computational Methods*, **2**, 515-525.
[7] Yi X. J., Hou P., Dong H. P., Lai Y. H. and Mu H. N. (2015b) A reliability optimization allocation method for systems with differentiation of functions, *Proceedings of ASME 2015 International Mechanical Engineering Congress & Exposition*, IMECE2015-52928.
[8] Yi X. J., Dhillon B. S., J. Shi, Mu H. N. and Dong H. P. (2015c) Reliability Analysis Method on Repairable System with Standby Structure Based on Goal Oriented Methodology, *Quality and Reliability Engineering International*, (in press).
[9] Yi X. J., J. Shi, H. N. Mu, Dong H. P. and Zhang Z. (2015d) Reliability Analysis of Repairable System with Multiple-Input And Multi-Function Component Based on GO Methodology, *Proceedings of ASME 2015 International Mechanical Engineering Congress & Exposition*, IMECE2015-51289.
[10] Yi X. J., Shi J., Mu H. N., Dong H. P. and Guo S. W. (2015e) Reliability Analysis of Hydraulic Steering System with DICLFL Considering Shutdown Correlation Based on GO Methodology, *In: Reliability Systems Engineering (ICRSE), 2015 First International Conference on, IEEE*.
[11] Yi X. J., Lai Y. H., Dong H. P. and Hou P. (2016a) A reliability optimization allocation method considering differentiation of functions, *International Journal of Computational Methods*, (Accepted).
[12] Yi, X. J., Shi, J., Dong, H. P. and Lai, Y. H. (2016b) Reliability Analysis of Repairable System With Multiple Fault Modes Based on Goal-Oriented Methodology, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* **2**, 011003.
[13] Wang, K. J. and Lee. C. H. (2015) A revised ant algorithm for solving location–allocation problem withrisky demand in a multi-echelon supply chain network, *Applied Soft Computing* **32**, 311–321.
[14] Alavidoost, M. H., Tarimoradi, M. and Zarandi, M. H. (2015) Fuzzy adaptive genetic algorithm for multi-objective assembly line balancing problems, *Applied Soft Computing* **34**, 655-677.
[15] Zhao, Y., Zheng, S. J., Wang, H. T. and Yang, L. Q. (2015) Simultaneous optimization of photostrictive actuator locations, numbers and light intensities for structural shape control using hierarchical genetic algorithm, *Advances in Engineering Software* **88**, 21-29.
[16] Shen Z. P., Gao J. and Huang X. R. (2000) An new quantification algorithm for the GO Methodology, *Reliability Engineering and System Safety*, **67**, 241-247.
[17] Shen Z. P., Gao J. and Huang X. R. (2001) An Exact Algorithm Dealing with Shared Signals in the GO Methodology, *Reliability Engineering and System Safety*, **73**, 177-181.
[18] Shen Z. P., and Huang X. R. (2004) Principle and Application of GO Methodology, Tsinghua University press, Beijing.
[19] Nahas N., and Nourelfath M. (2005). Ant system for reliability optimization of a series system with multiple-choice and budget constraints, *Reliability Engineering & System Safety*, **87**, 1-12.

Stress/Displacement Field Calculation for Bolted Joint Based on State Space Theory

Q.C. Sun*, Y.J. Jiang, X. Huang, W.Q. Huang, Z.Y. Sun, X.K. Mu
School of Mechanical Engineering, Dalian University of Technology, China
*Corresponding author: sqc_dut@163.com

Abstract.
Stress/displacement field analyzing of mechanical assembly is important for predicting mechanical property, and optimizing structural parameters and assembly technology parameters of mechanical assembly. However the structural discontinuity and material difference of mechanical assembly determines the complexity of stress function, it is difficult for analytically computing stress/displacement field of mechanical assembly. In this paper, taking bolted joint under the action of normal load as the research object, a stress/displacement field layered mapping and calculating model of mechanical assembly is proposed, with considering the stress/displacement transmission characteristics of mechanical assembly, combining state space method and elastic mechanics theory. The model divides mechanical assembly as the layered structures, and determines layered constraint conditions according to structural discontinuity or continuity in different positions, such as the structure at the junction surfaces is discontinuous. Considering the difference between bolted joint and the common axsymmetric structure, taking the stresses σ_z, τ_{zr} and the displacements u, w as the state variables, the state equations based on Fourier-Bessel series was built to express the stress/displacement transmission relationships. Linearizing the stress/displacement transmission rules, the relationships between state variables at arbitrary and external load were determined by accumulated calculating, and stress/displacement characteristics at arbitrary positions of bolted joint structure were obtained. Finally, the pressure distribution of the bolted joint interface, and stress/ displacement distribution of the whole bolted joint structure was calculated, the comparison among the analytically calculation, FEA and the test data proves the effectiveness of the model.

Keywords: Structural discontinuity, Bolted joint, Stress/displacement field, State space method, Elastic mechanics

1. Introduction

Mechanical systems are usually composed of multiple parts, which were assembled according to the specific requirements. The contact surfaces among these parts are known as the "joint surfaces" or "interfaces", such as bolted joint surfaces, guide contact surfaces and the mating surface between hole and shaft. The joint surfaces, together with the influence area of stress/deformation in the connected mechanical components, are known as "joint" [1].

Joint surfaces / joints have remarkable influence on the statics, dynamics and thermodynamic characteristics of mechanical systems, and obtaining the stress / displacement distribution in joints is the basis for accurately analyzing the characteristics of mechanical systems. Joint surfaces / joints stiffness, which is closely related to the stress / displacement distribution in the joints, is a key factor affecting the accuracy of the mechanical systems [2, 3]. Bolted joints and hole-shaft mating surfaces often occur fretting fatigue under the external alternating loads, which leads to premature parts failure, and the stress / displacement distribution in joints is the main factor affecting fretting fatigue [4]. The dynamics characteristics are the key

features of mechanical assemblies, the stiffness distributions in joints are the important factors influencing dynamics characteristics, and obtaining the stress / displacement fields is the premise of calculating stiffness of joints and revealing the dynamics of assemblies [5-7]. Moreover, determining the contact area distribution and elastic-plastic contact state in joints have also great significance for analyzing heat transfer mechanism of assemblies [8, 9]. However, it is difficult to measure directly the stress/displacement distribution in joints, theoretical analysis and calculation are the primary means of obtaining stress / displacement fields information.

Compared to a single part, the structural discontinuity and material difference of mechanical assemblies makes it difficult to calculate the stress / displacement field based on the traditional elasticity theory. The traditional elasticity theory based on the continuity, uniformity and other basic assumptions, and one component is composed of the same material, the stress, deformation and displacement characteristics in one component are completely continuous. The mechanical assemblies have the discontinuous structure characteristics, the stress /displacement distribution of joint surfaces is unknown. Because lacking mature stress/ displacement distribution function under the unknown boundary conditions[10-12], it is difficult to accurately calculate the stress/displacement field of the mechanical assembly in the traditional elastic mechanics system.

Finite element method (FEM) is the current main method of calculating the stress / displacement field in mechanical assembly [4, 6, 13-16]. The stress / displacement field calculation of joint surfaces / joints belongs to contact nonlinear problem, which requires large computation memory but embodies a low computational efficiency. Additionally, the FEM computation results depend on the high quality of grids, especially need dense grids in the contact area, which also limits the efficiency of solving such problems.

Combining with elastic mechanics, the state space methods have been used to calculate exactly the stress / displacement fields of laminated plates, functionally graded plates, and multi-layer civil structures, etc., these studies provide references for the stress/displacement field calculation of bolt joint. Such as, Xiang and Wang[17] obtained the exact buckling and vibration solutions of unidirectional ladder rectangular plates by combining the Levy method and the state space theory. Chen and Ding[18] derived two independent state equation with variable coefficients, and analyzed the freedom vibration of transversely isotropic piezoelectric material rectangular plate on the basis of three-dimensional theory equations of transversely isotropic piezoelasticity. Ying et al. [19] put forward the exact solutions of bending and free vibration for functionally graded beams placing on a Winkler–Pasternak elastic foundation, based on the two-dimensional elasticity theory and state space method. Adopting state space method, Hongyu and Jiarang [20, 21] obtained the analytical solutions of bending problem for clamped or simply supported thick laminated circular plate, as well as thick laminated circular plate on elastic foundation with free edges.

Taking the external load of joint surfaces/joints as the input information and stress/displacement distribution as the output information, and expressing the transmission characteristics of the stress/displacement as state transition matrix, the stress/displacement field can be calculated based on state space theory. The

stress/displacement field calculation of mechanical assemblies have the similar theory basis to the previous research objects[17-21], with discarding any assumptions about displacement pattern and stress distribution, and constructing the stress/displacement transfer matrix of mechanical assembly by adopting the state space differential equation.

The bolted joint under the action of normal load was selected as research object in this paper, the structure, material and loading mode of bolted joint are different from the laminated plates, etc., a new calculation model for bolted joint was built. And the traditional axial symmetry stress/displacement state equations do not completely match with the structural characteristics of bolted joint, the state equations for bolted joint was built.

2. State Space Method in Elastic Mechanics Problem

State space method is a method to analyze and synthesize control systems based on the state variable description in modern control theory. State space method describes the state of the system with the state variables, and establishes the relationship between the state variables within the system and the external input/output variable. State equation is the mathematical description which reflects the causal relationship between state variables and input variables in state space method. Because state space method uses matrix representation, the increase in the number of state variables, input variables or output variables, does not increase the complexity of the system description, which makes it especially suitable for dealing with multiple input, multiple output and multivariable system problems.

Using the state space method to solve the elastic mechanics problem, first of all, should select the key unknown variables as state variables, and then set up the mechanics model according to the actual problem. The number and the type of the state variables depend on the specific problem. For example, the stresses $\sigma_z, \tau_{xz}, \tau_{yz}$ and the displacements u, v, w can be selected as state variables，and constitute the state vector $S=\left[\sigma_z \quad \tau_{xz} \quad \tau_{yz} \quad u \quad v \quad w\right]^T$. The ordinary differential state equation of elastic mechanics problem can be obtained by physical equation, equilibrium equation, and a series of changes, such as series expansion, Laplace transform, Hankel transform, etc. Generally the form of ordinary differential state equation as follows:

$$\frac{d}{dz}S(z)=DS(z)+\tilde{\Phi}(z) \tag{1}$$

where $S(z)$ is the state vector, D is a square matrix related to material parameters, $\tilde{\Phi}(z)$ is an array related to boundary conditions. The state equation is obtained by solving **Eq. (1)** as follows:

$$S(z)=T(z)S(0)+\Phi(z) \tag{2}$$

where square matrix $T(z)$ is the state transition matrix. Thus, the state $S(z)$ that

transfers any distance along the z direction is obtained from **Eq. (2)**, with the known initial state $S(0)$.

The state space method is an effective way to deal with the discontinuous structure problem of mechanical joints, which divides the matching components into different "chains" and sets the boundary conditions so as to adapt to the material discontinuous characteristic in structure and material. And it calculate the stress/ displacement exactly by dividing a single component (corresponds to a "chain") into different virtual "layers". Moreover, the assembly that is divided into virtual layers, is an end-to-end chain system, whose state variables can be obtained from the simple accumulation of the state transition matrix. Since the number of the state variables don't vary with the number of "chains" or "layers", to a great extent, complex problems can be simplified.

3. Calculation Model for Bolted Joint

3.1 Model Assumption

To research the stress/displacement field in bolted joints, the analytical model is assumed to be the axisymmetric mechanics problem as shown in **Fig. 1**. The two bolted plates are expressed as hollow cylinder Ⅰ and Ⅱ with inner diameter $2a$ and outer diameter $2b$, whose thicknesses are h_1 and h_2 respectively. Preload is an axisymmetric distributed pressure $p(r)$ on the upper surface of hollow cylinder Ⅰ over an annular region $a \le r \le c$. Hollow cylinder Ⅱ corresponds to the bolted member with a fixed lower surface. All of the cylindrical surfaces of hollow cylinder Ⅰ and Ⅱ are free boundaries. Take the center of upper surface of hollow cylinder Ⅰ as the coordinate origin O. Take the central axis of the hollow cylinders as the symmetry axis z, whose direction is vertical downward, and axis r is in the horizontal direction, the global cylindrical coordinate system (r,θ,z) is established as shown in **Fig. 1(a)**. in the same way, the local cylindrical coordinate systems (r,θ,z_1) and (r,θ,z_2) are established on hollow cylinder Ⅰ and Ⅱ, respectively.

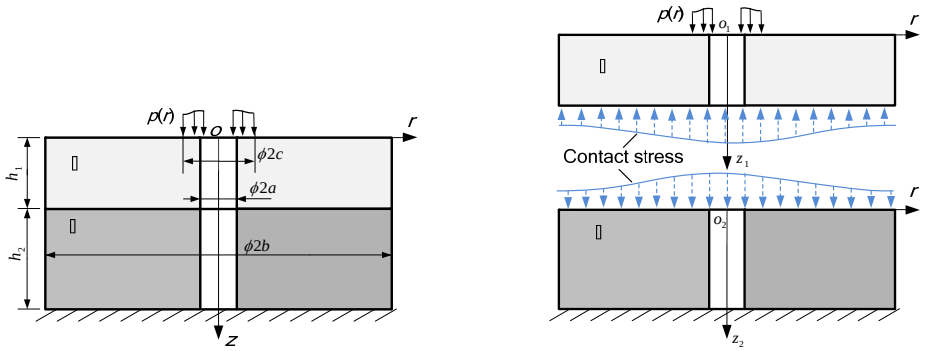
To make the model as simple as possible, following basic assumptions are used:

(1) The material of each hollow cylinder is assumed to be ideal elastic, continuous, homogeneous, and isotropic.

(2) Body forces are ignored.

(3) The roughness, flatness and other practical machining errors of the contact surface are ignored, the joint surface is absolutely smooth and flat.

(4) The points in the contact surface of the two pieces of hollow cylinders always maintain contact during the loading process.



(a) (b)

Fig. 1 Analytical model of bolted joint

3.2 Boundary Conditions

To simplify the calculation, the distributed pressure $p(r)$ in **Fig. 1** is assumed to be a uniformly distributed load whose resultant force is F:

$$p(r)=\begin{cases} p=\dfrac{F}{\pi(c^2-a^2)} & (a\le r\le c) \\ 0 & (c<r\le b) \end{cases} \tag{3}$$

In this paper, the superscripts (1) and (2) denote the mechanical characteristics of hollow cylinder Ⅰ and Ⅱ respectively, and u and w denote horizontal and vertical displacement respectively. The boundary conditions are given by **Eq. (4) ~ (7).**

(1) The upper surface of hollow cylinder Ⅰ:

$$z_1=0:\quad \sigma_z^{(1)}=-p(r),\ \tau_{zr}^{(1)}=0\ , \tag{4}$$

(2) The lower surface of hollow cylinder Ⅱ:

$$z_2=h_2:\quad u^{(2)}=w^{(2)}=0 \tag{5}$$

(3) The contact surface of hollow cylinder Ⅰ and Ⅱ:

$$z_1=h_1\ ,\ z_2=0:\quad \sigma_z^{(1)}=\sigma_z^{(2)},w^{(1)}=w^{(2)},\tau_{zr}^{(1)}=\tau_{zr}^{(2)}=0 \tag{6}$$

(4) All of the cylindrical surfaces:

$$r=a,b:\quad \tau_{zr}^{(1)}=\tau_{zr}^{(2)}=0,\ \sigma_r^{(1)}=\sigma_r^{(2)}=0 \tag{7}$$

3.3 State Equations

The stress/displacement field calculation in bolted joint under the normal load belongs to axisymmetric problem, whose physical equations are expressed as follows:

$$\begin{cases} \sigma_r = (\lambda + 2G)\dfrac{\partial u}{\partial r} + \lambda \dfrac{u}{r} + \lambda \dfrac{\partial w}{\partial z} \\ \sigma_\theta = \lambda \dfrac{\partial u}{\partial r} + (\lambda + 2G)\dfrac{u}{r} + \lambda \dfrac{\partial w}{\partial z} \\ \sigma_z = \lambda \dfrac{\partial u}{\partial r} + \lambda \dfrac{u}{r} + (\lambda + 2G)\dfrac{\partial w}{\partial z} \\ \tau_{zr} = G(\dfrac{\partial w}{\partial r} + \dfrac{\partial u}{\partial z}) \end{cases} \tag{8}$$

where λ is Lame Constant, G is the shear modulus.

The equilibrium equations of axisymmetric problem are expressed as follows:

$$\begin{cases} \dfrac{\partial \sigma_r}{\partial r} + \dfrac{\partial \tau_{zr}}{\partial z} + \dfrac{\sigma_r - \sigma_\theta}{r} = 0 \\ \dfrac{\partial \tau_{zr}}{\partial r} + \dfrac{\partial \sigma_z}{\partial z} + \dfrac{\tau_{zr}}{r} = 0 \end{cases} \tag{9}$$

Let

$$C_1 = -\frac{\lambda}{\lambda + 2G},\ C_2 = \lambda + 2G - \frac{\lambda^2}{\lambda + 2G}$$
$$C_3 = \lambda - \frac{\lambda^2}{\lambda + 2G},\ C_4 = \frac{1}{\lambda + 2G},\ C_5 = \frac{1}{G}$$

Eliminating σ_r and σ_θ from **Eq. (8)**, we obtain

$$\sigma_r = C_2 \frac{\partial u}{\partial r} + C_3 \frac{u}{r} - C_1 \sigma_z \tag{10}$$

$$\sigma_\theta = C_3 \frac{\partial u}{\partial r} + C_2 \frac{u}{r} - C_1 \sigma_z \tag{11}$$

Selecting u, w, τ_{zr} and σ_z as the state variables, the following is obtained from **Eq. (8)** and **(9).**

$$\frac{\partial}{\partial z}\begin{Bmatrix} u \\ w \\ \tau_{zr} \\ \sigma_z \end{Bmatrix} = \begin{bmatrix} 0 & -\dfrac{\partial}{\partial r} & C_5 & 0 \\ C_1\left(\dfrac{\partial}{\partial r}+\dfrac{1}{r}\right) & 0 & 0 & C_4 \\ C_2\left(\dfrac{\partial^2}{\partial r^2}+\dfrac{1}{r}\dfrac{\partial}{\partial r}-\dfrac{1}{r^2}\right) & 0 & 0 & C_1\dfrac{\partial}{\partial r} \\ 0 & 0 & -\left(\dfrac{\partial}{\partial r}+\dfrac{1}{r}\right) & 0 \end{bmatrix} \begin{Bmatrix} u \\ w \\ \tau_{zr} \\ \sigma_z \end{Bmatrix} \tag{12}$$

This paper expands the solution of **Eq. (9)** into following Fourier-Bessel series

$$\begin{cases} u(r,z) = \sum_{m=1}^{\infty} U_m(z) V_1(\alpha_m r) + r\tilde{U}(z) \\ w(r,z) = W_0(z) + \sum_{m=1}^{\infty} W_m(z) V_0(\alpha_m r) \\ \tau_{zr}(r,z) = \sum_{m=1}^{\infty} R_m(z) V_1(\alpha_m r) \\ \sigma_z(r,z) = Z_0(z) + \sum_{m=1}^{\infty} Z_m(z) V_0(\alpha_m r) \end{cases} \tag{13}$$

The form of Fourier - Bessel series Hongyu and Jiarang [20,21] proposed, can meet the boundary conditions of circular plate, but cannot meet the boundary conditions of bolted joint structure (hollow cylinder). To solve this problem, the form of the function $V_\mu(\alpha_m r)$ is structured as follows:

$$V_\mu(\alpha_m r) = J_\mu(\alpha_m r) - \frac{J_\mu(\alpha_m b)}{Y_\mu(\alpha_m b)} Y_\mu(\alpha_m r)$$

where $J_\mu(\alpha_m r)$ and $Y_\mu(\alpha_m r)$ are the first type and the second type μ-order Bessel functions separately. $\tilde{U}(z)$ is an unknown function for z. U_m, W_m, R_m and Z_m $(m = 0,1,2,3,\cdots)$ are respectively the coefficients of Fourier-Bessel series of u, w, τ_{zr} and σ_z. $\alpha_m = \beta_m / a$, $\beta_m (m = 1,2,3,\cdots)$ is the m-th positive root satisfying the following equation

$$J_1(\beta_m) Y_1\left(\frac{b}{a}\beta_m\right) - J_1\left(\frac{b}{a}\beta_m\right) Y_1(\beta_m) = 0, \quad (0 < \beta_1 < \beta_2 < \beta_3 \cdots) \tag{14}$$

$V_1(\alpha_m r)$ satisfies $V_1(\alpha_m a) = V_1(\alpha_m b) = 0$, therefore **Eq. (13)** satisfies the boundary conditions τ_{zr}=0 in **Eq. (7)**. In addition, to satisfy boundary conditions of cylindrical surfaces, there should be $\sigma_r = 0$ at $r = a$ or b. Substituting the first and the forth equation of **Eq. (13)** into **Eq. (10)** and setting $\sigma_r = 0$, the following two equations can be obtained, and the unknown function $\tilde{U}(z)$ can be determined from **Eq. (15)** and **(16)**.

$$\left(C_3+C_2\right)\tilde{U}\left(z\right)+\sum_{m=1}^{\infty}\left[C_2\alpha_m U_m\left(z\right)-C_1 Z_m\left(z\right)\right]V_0\left(\alpha_m a\right)-C_1 Z_0\left(z\right)=0\text{ ,at }\ r=a \tag{15}$$

$$\left(C_3+C_2\right)\tilde{U}\left(z\right)+\sum_{m=1}^{\infty}\left[C_2\alpha_m U_m\left(z\right)-C_1 Z_m\left(z\right)\right]V_0\left(\alpha_m b\right)-C_1 Z_0\left(z\right)=0\text{ ,at }\ r=b \tag{16}$$

Performing the following series expansion

$$\begin{cases} r=\sum_{m=1}^{\infty}\tilde{A}_m V_1\left(\alpha_m r\right) \\ \left(\frac{\partial}{\partial r}+\frac{1}{r}\right)r=\tilde{B}_0+\sum_{m=1}^{\infty}\tilde{B}_m V_0\left(\alpha_m r\right) \\ \left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}\right)r=\sum_{m=1}^{\infty}\tilde{C}_m V_1\left(\alpha_m r\right) \end{cases} \tag{17}$$

The coefficients are obtained according to related knowledge of Fourier - Bessel series as follows

$$\begin{aligned} &\tilde{A}_m=\frac{c^2V_2\left(\alpha_m c\right)-a^2V_2\left(\alpha_m a\right)}{\alpha_m M_m} \\ &\tilde{B}_0=2 \\ &\tilde{B}_m=\tilde{C}_m=0,\left(m=1,2,3,\cdots\right) \end{aligned}$$

Where

$$M_m=\frac{b^2V_0^{\ 2}\left(\alpha_m b\right)-a^2V_0^{\ 2}\left(\alpha_m a\right)}{2}$$

Substituting **Eq. (13)** and **(17)** into **Eq. (12)**, the following equation can be obtained

$$\frac{d}{dz}S\left(z\right)=DS\left(z\right)+\tilde{\Phi}\left(z\right) \tag{18}$$

Where

$$S\left(z\right)=\left[U_m\left(z\right)\quad W_m\left(z\right)\quad R_m\left(z\right)\quad Z_m\left(z\right)\right]^T \tag{19}$$

$$D=\begin{bmatrix} 0 & \alpha_m & C_5 & 0 \\ C_1\alpha_m & 0 & 0 & C_4 \\ C_2\alpha_m^2 & 0 & 0 & -C_1\alpha_m \\ 0 & 0 & -\alpha_m & 0 \end{bmatrix} \tag{20}$$

$$\tilde{\Phi}(z)=\begin{bmatrix} -\tilde{A}_m \dfrac{d\tilde{U}(z)}{dz} & C_1\tilde{B}_m\tilde{U}(z) & -C_2\tilde{C}_m\tilde{U}(z) & 0 \end{bmatrix}^T \tag{21}$$

Eq. (17) is a nonhomogeneous ordinary differential equation, and solving it yields the state equation as follows

$$S(z)=T(z)S(0)+\Phi(z) \tag{22}$$

Where

$$T(z)=e^{Dz} \tag{23}$$

$$\Phi(z)=\int_0^z e^{D(z-t)}\tilde{\Phi}(t)\,dt \tag{24}$$

In **Eq. (22)**, $S(z)$ is the state vector at z, namely the coefficient terms of Fourier-Bessel series, $S(0)$ is the initial state vector on the upper surface. For a certain m, the matrix D is a constant matrix, so $T(z)$ can be obtained. The parameters of $\Phi(z)$ are known except $\tilde{U}(z)$. Therefore, For a certain m, the state vector $S(z)$, namely the coefficients of Fourier-Bessel series U_m, W_m, R_m and Z_m at z, can be obtained with the initial state vector $S(0)$ in **Eq. (22)**, only if the function $\tilde{U}(z)$ is determined.

Particularly, there are the following relation for $m=0$.

$$\begin{cases} \dfrac{d}{dz}Z_0(z)=0 \\ \dfrac{d}{dz}W_0(z)=C_4Z_0(z)+C_1\tilde{B}_0\tilde{U}(z) \end{cases} \tag{25}$$

3.4 Coefficients of Fourier-Bessel Series for $m\geq 1$

Determining the function $\tilde{U}(z)$ is the key to Stress/displacement field calculation. As shown in **Fig. 2**, the j-th hollow cylinder is divided into N_j virtual sublayers averagely, and the thickness of each sublayer is $d_j=h_j/N_j$. Let $x_{j,i}$ and $x_{j,i+1}$ be the end-values of the upper surface and the lower surface of the i-th sublayer in the j-th hollow cylinder, respectively. As shown in **Fig. 3**, provided that the sublayer is

thin enough, it is reasonable to consider that the unknown function $\tilde{U}(z)$ in the sublayer is linear distributed along z direction [20]. So in the i-th sublayer of the j-th hollow cylinder, function $\tilde{U}(z)$ can be denoted by $\tilde{U}_{j,i}(z_{j,i})$ as follows, in the local coordinate system whose origin of axis $z_{j,i}$ is on the upper surface of the sublayer.

$$\tilde{U}_{j,i}(z_{j,i}) = \frac{x_{j,i+1} - x_{j,i}}{d_j} \cdot z_{j,i} + x_{j,i},\ (0 \le z_{j,i} \le d_j, i = 1,2,\ldots,N_j, j = 1,2) \tag{26}$$

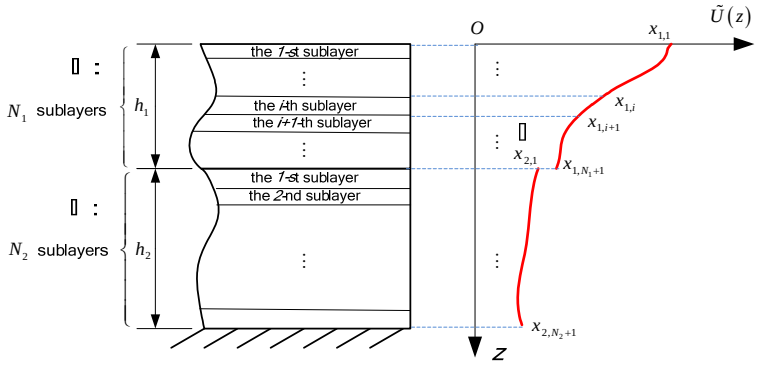


Fig. 2 Sublayers and corresponding $\tilde{U}(z)$

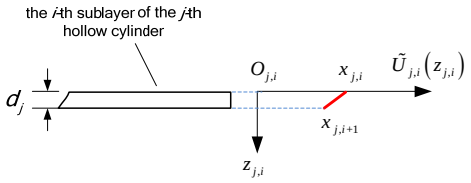


Fig. 3 Linear distribution assumption

Linear distribution assumption (26) causes calculation error, but if the number of the sublayers N_j increases gradually, the error will decrease. So the error is controllable and N_j can be determined based on the accuracy requirement. For any sublayer, the ordinary differential state equation is obtained according to **Eq. (18)**, **(21)** and **(26)**

$$\frac{d}{dz} S_{j,i}(z_{j,i}) = D_j S_{j,i}(z_{j,i}) + \tilde{\Phi}_{j,i}(z_{j,i}) \tag{27}$$

Where

$$\tilde{\Phi}_{j,i}(z_{j,i}) = \left[\tilde{A}_m \frac{x_{j,i} - x_{j,i+1}}{d_j} \quad 0 \quad 0 \quad 0 \right]^T \tag{28}$$

According to **Eq. (22)** ~ **(24)**, the solution of **Eq. (27)** is obtained

$$S_{j,i}\left(z_{j,i}\right)=T_j\left(z_{j,i}\right)S_{j,i}\left(0\right)+\Phi_{j,i}\left(z_{j,i}\right) \tag{29}$$

Setting $z_{j,i}=d_j$ in **Eq. (29)**, the solutions of adjacent sublayers within the same part as follows.

$$\begin{cases} S_{j,i}\left(d_j\right)=T_j\left(d_j\right)S_{j,i}\left(0\right)+\Phi_{j,i}\left(d_j\right) \\ S_{j,i+1}\left(d_j\right)=T_j\left(d_j\right)S_{j,i+1}\left(0\right)+\Phi_{j,i+1}\left(d_j\right) \end{cases} \tag{30}$$

The continuity condition between the sublayers is

$$S_{j,i}\left(d_j\right)=S_{j,i+1}\left(0\right) \tag{31}$$

Perform **Eq. (30)** and **(31)** successively for all the sublayers, and finally the relationship between the state vectors of the lower surface of the j-th hollow cylinder $S_{j,N_j}\left(d_j\right)$ and the upper surface $S_{j,1}\left(0\right)$ can be expressed as the following formula:

$$S_{j,N_j}\left(d_j\right)=\left[T_j\left(d_j\right)\right]^{N_j}S_{j,1}\left(0\right)+\pi_{j,N_j}\left(d_j\right),\quad\left(j=1,2\right) \tag{32}$$

Where

$$\begin{aligned} \pi_{j,N_j}\left(d_j\right)=&\left[T_j\left(d_j\right)\right]^{N_j-1}\Phi_{j,1}\left(d_j\right)+\cdots+\left[T_j\left(d_j\right)\right]^{2}\Phi_{j,N_j-2}\left(d_j\right)\\ &+T_j\left(d_j\right)\Phi_{j,N_j-1}\left(d_j\right)+\Phi_{j,N_j}\left(d_j\right) \end{aligned} \tag{33}$$

In the local coordinate system of each sublayer, the state vector is

$$S_{j,i}\left(z_{j,i}\right)=\left[U_m^{(j,i)}\left(z_{j,i}\right)\quad W_m^{(j,i)}\left(z_{j,i}\right)\quad R_m^{(j,i)}\left(z_{j,i}\right)\quad Z_m^{(j,i)}\left(z_{j,i}\right)\right]^T \tag{34}$$

where $z_{j,i}=0$ denots sublayer's upper surface, $z_{j,i}=d_j$ denots sublayer's lower surface.

According to the boundary condition **(4)**, the following is given:

$$R_m^{(1,1)}\left(0\right)=0 \tag{35}$$

In addition, the distributed pressure $p(r)$ is known, so $-p(r)$ can be expressed as the form of Fourier-Bessel series according to **Eq. (30)**

$$-p(r)=Z_0^{(1,1)}(0)+\sum_{m=1}^{\infty}Z_m^{(1,1)}(0)V_0(a_m r) \tag{36}$$

According to the boundary condition **(5)**, the following is given:

$$U_m^{(2,N_2)}(d_2)=W_m^{(2,N_2)}(d_2)=0 \tag{37}$$

According to the boundary condition **(6)**, the following is given:

$$W_m^{(1,N_1)}(d_1)=W_m^{(2,1)}(0),\quad Z_m^{(1,N_1)}(d_1)=Z_m^{(2,1)}(0),\quad R_m^{(1,N_1)}(d_1)=R_m^{(2,1)}(0)=0 \tag{38}$$

Regarding the variables of $S_{j,N_j}(d_j)$ and $S_{j,1}(0)$ $(j=1,2)$ in **Eq. (33)** as unknown, there are 16 unknown variables in total, eight of which can be eliminated by **Eq. (35) ~ (38)**. Therefore, the expression of other unknown variables can be solved from **Eq. (32)**. Obviously, the expression of $S_{j,1}(0)$ is also obtained. It is important to note that the expression of $S_{j,1}(0)$ also contains the undetermined constants $x_{j,i}$ $(i=1,2,\cdots,N_j+1;\ \ j=1,2)$. After determining the expression of $S_{j,1}(0)$, by repeating the derivation process of **Eq. (32)** , the stress/displacement of the i-th sublayer in the j-th hollow cylinder can be calculated, matrix $\Pi_{j,i}(z_{j,i})$ and vector $\pi_{j,i}(z_{j,i})$ are not difficultly to obtain.

$$S_{j,i}(z_{j,i})=\Pi_{j,i}(z_{j,i})S_{j,1}(0)+\pi_{j,i}(z_{j,i}) \tag{39}$$

If global coordinate z locates in the i-th sublayer of the j-th hollow cylinder, $z_{j,i}$ is given by

$$z_{j,i}=\begin{cases}z-(i-1)d_1, & (j=1)\\ z-h_1-(i-1)d_2, & (j=2)\end{cases} \tag{40}$$

3.5 Coefficients of Fourier-Bessel Series for $m=0$

The following formulas is obtained from **Eq. (25)** and **(26)**:

$$\begin{cases}Z_0^{(j,i)}(z_{j,i})=Z_0^{(j,i)}(0)=Z_0^{(1,1)}(0)\\ W_0^{(j,i)}(z_{j,i})=W_0^{(j,i)}(0)+C_4Z_0^{(j,i)}(0)z_{j,i}+C_1(x_{j,i}+x_{j,i+1})z_{j,i}\end{cases} \tag{41}$$

Particularly, setting $z_{j,i} = d_j$, the following formulas can be obtained:

$$W_0^{(j,i)}(0) = W_0^{(j,i)}(d_j) - C_4 Z_0^{(1,1)}(0) d_j - C_1 (x_{j,i} + x_{j,i+1}) d_j \tag{42}$$

According to the boundary conditions, the following is given:

$$\begin{aligned} W_0^{(2,N_2)}(d_2) &= 0 \\ W_0^{(2,1)}(d_2) &= W_0^{(1,N_1)}(d_1) \\ W_0^{(j,i+1)}(0) &= W_0^{(j,i)}(d_j) \end{aligned}$$

Therefore, all of the $W_0^{(j,i)}(0)$ can be solved from **Eq. (42)**, and then the expression of $W_0^{(j,i)}(z_{j,i})$ at any position can be obtained from **Eq. (41)**.

3.6 Solving the Undetermined Constants

There are $N_j + 1$ undetermined constants $x_{j,i}$ in the j-th hollow cylinder, add up to $N_1 + N_2 + 2$ undetermined constants in hollow cylinder Ⅰ and Ⅱ, to solve the undetermined constants, $N_1 + N_2 + 2$ equations were needed. Set $N_1 = 2n_1$, $N_2 = 2n_2$,where n_1, n_2 are positive integer. By solving **Eq. (26)**, **(39)**, **(40)** and **(41)**, $\tilde{U}_{j,i}(z)$, $U_m^{(j,i)}(z)$ and $Z_m^{(j,i)}(z)$ at the position of **Eq. (43)** can be determined, and substituting them into **Eq. (15)** and **(16)**, $N_1 + N_2 + 2$ linear equations are obtained, and all the undetermined constants can be solved. The state variables at any position in the two hollow cylinders can be determined according to **Eq. (39)** and **(41)**, and the corresponding stress/displacement can be obtained by substituting the coefficients into **Eq. (13).**

$$z = \begin{cases} 2k_1 d_1, & (k_1 = 0,1,2,3,\cdots,n_1) \\ 2k_2 d_2, & (k_2 = 0,1,2,3,\cdots,n_2, but\ \ k_2 \neq n_2 - 1) \end{cases} \tag{43}$$

4. Calculation Example

4.1 Comparison of Three Methods

To verify the effectiveness of the above method, a specific example is designed, as shown in **Fig. 4**, the contact stress of the joint surface is extracted, and comparing with the experimental measurement and the finite element analysis result is carried out.

Both the material of the two hollow cylinders are Q235, Young's modulus $E_1 = E_2 = 2 \times 10^5 \text{MPa}$, Poisson's ratio $\nu_1 = \nu_2 = 0.3$. the parameters in **Fig. 1** are $a = 6.3mm$, $b = 45mm$, $c = 12mm$, $h_1 = 10mm$, $h_2 = 20mm$. The normal load on the surface

is $F=4500N$, which can be expanded into the form of Fourier-Bessel series according to **Eq. (36)**. The coefficients is given by

$$Z_m^{(1,1)}(0)=\begin{cases}-\dfrac{c^2-a^2}{b^2-a^2}p, & (m=0)\\ -\dfrac{pcV_1(\alpha_m c)}{\alpha_m M_m}, & (m=1,2,3,\cdots)\end{cases}$$

A satisfactory results was obtained by setting N_1=N_2=80 and selecting first 10 items of the Fourier-Bessel series.

In the experimental, two hollow cylinders of Q235 with ϕ12.6 through-hole were placed on worktable and connected by M12 bolt and ϕ24 gasket. The normal load reached $4500N$, which was measured by a pressure sensor. The contact stress was measured by means of the pressure-sensitive film, as shown in **Fig. 4**.

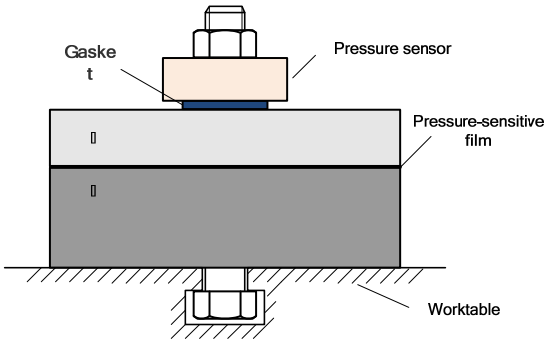
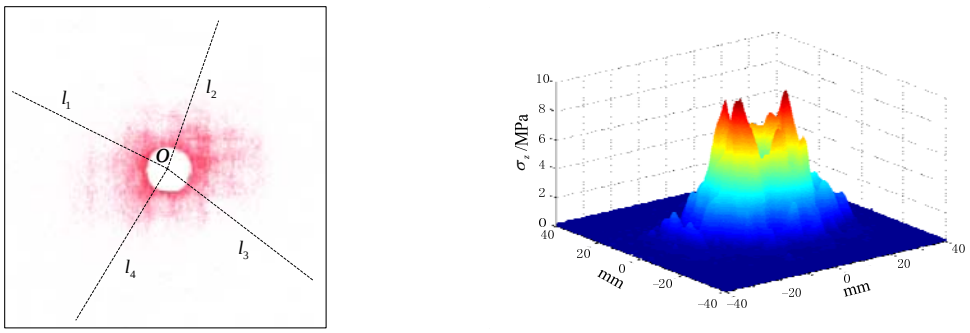


Fig. 4 Experimental setup for joint contact stress test

The white pressure-sensitive film turns red under pressure, and the red concentration increases with the increase of intensity of pressure. So the contact stress can be measured by evaluating the color concentration of the film. **Fig. 5(a)** shows the scanning image of the pressure-sensitive film after experiment. **Fig. 5(b)** shows the contact stress distribution with a three-dimensional figure. The figure clearly shows that the contact stress presents "steep peak" shape distribution, the maximum contact stress appears near the center of the load, and the stress decreases rapidly from the center to the edge until reduces to zero.



(a) Scanning image of the film **(b)** 3-D distribution figure

Fig. 5 Contact stress distribution in bolted joint

As shown in **Fig. 5**, because of the machining error of specimens, the position deviation of bolt installation and the measurement error of pressure-sensitive film, measurement result is not absolutely axisymmetric. In order to eliminate the impact of these factors on the measurement result, four straight paths along radial direction, $l_1 \sim l_4$, are set up on the film, as shown in **Fig. 5(a)**. The pressure value of several points of the paths are extracted, and the average value of the same radial position are obtained. Thus, the contact stress distribution along radial direction are obtained. It should be noted that because the measurement error is large near the edge of the hole, the experimental data at the position isn't extracted.

Fig. 6 shows the contact stress distribution curves of state space method (SSM), experimental measurement and finite element method (FEM). The negative value denotes compressive stress. It can be seen that three distribution curves have a good consistency, so the state space method of this paper is accurate and reliable.

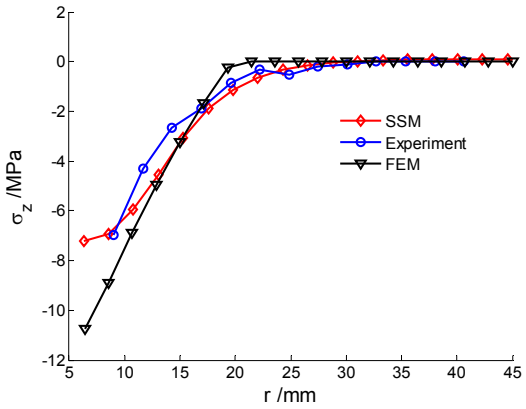
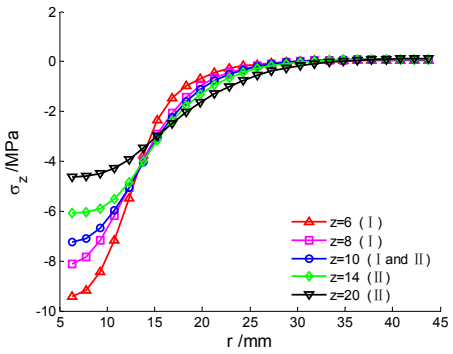


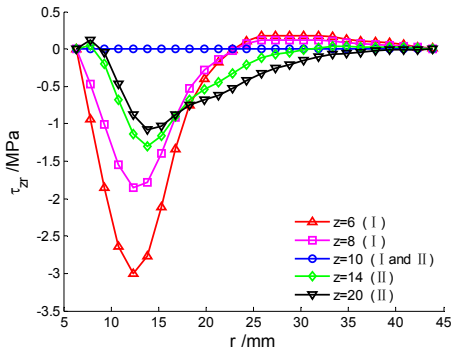
Fig. 6 Data comparison of contact stress

4.2 Stress / Displacement Field in Bolted Joint

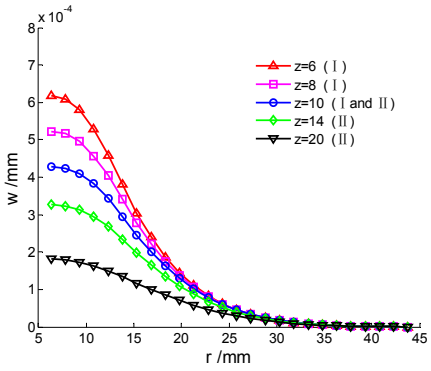
The stress and displacement information of bolted joint are extracted on the basis of the state space method calculate model proposed in this paper. Some stress and displacement distribution curves along radial direction are shown in **Fig. 7**. It can be seen that normal stress σ_z and vertical displacement w both own considerable variation gradient at $r = c$, but tangential stress τ_{zr} and horizontal displacement u appear to be bigger values at $r = c$ than other positions. Moreover, with the increase of coordinate r, all the stresses and displacements tend to zero as shown in **Fig. 7**.



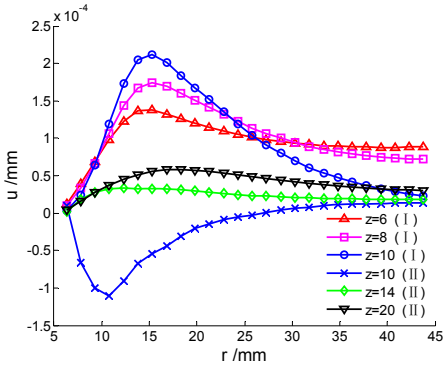
(a) Normal stress σ_z



(b) Tangential stress τ_{zr}



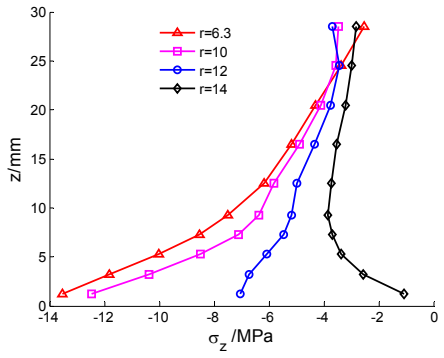
(c) Vertical displacement w



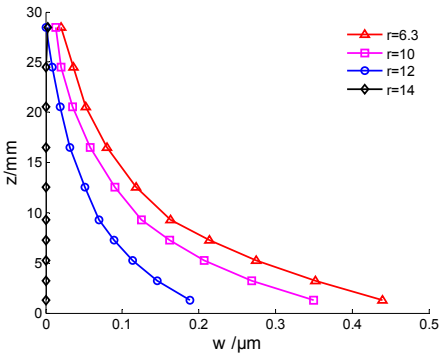
(d) Horizontal displacement u

Fig. 7 Stress and displacement distribution curves along radial direction

For mechanical discontinuous structure problem, normal stress σ_z and vertical displacement w are likely to be the mechanics characteristics people pay more attention to. Some stress and displacement distribution curves along z direction are shown in Fig. 8. It can be seen that σ_z and w decrease nonlinearly with the increase of z. In order to display the distribution of bolted joint under the normal load more visually, the contour map of σ_z and w are drawn, as shown in Fig. 9. The σ_z and w in the bolted joint subjected to a normal load, can transmit swimmingly from the upper plate to the lower plate, and shows a good continuity. The influence region of the external load is mainly on the surface region $a \le r \le c$, as well as its lower region. With the increase of z, the influence region spreads gradually, nevertheless, the stress and displacement decrease rapidly.

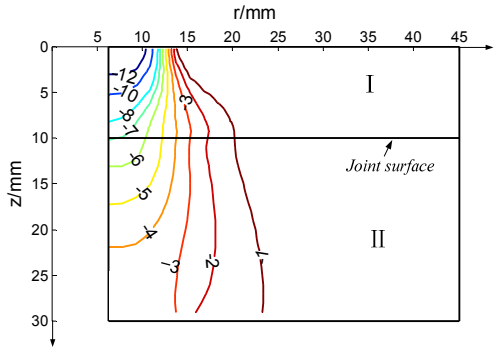


(a) Normal stress σ_z

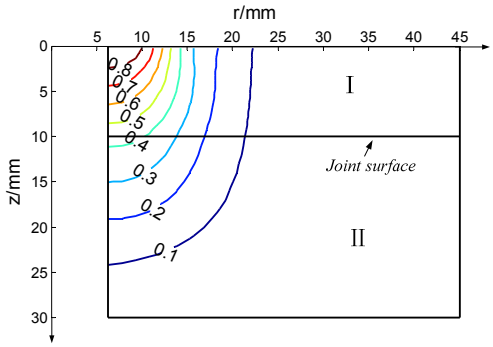


(b) Vertical displacement w

Fig. 8 σ_z and w distribution curves along z direction



(a) Normal stress σ_z (MPa)



(b) Vertical displacement w (μm)

Fig. 9 Contour map of σ_z and w

5. Conclusions

A stress/displacement field calculation model combining elastic mechanics with state space theory is established to solve the mechanical calculation problem associated with the discontinuity of structure and material in bolted joints. The stress / displacement distribution regularities of the joint surface and the components are obtained accurately, and the transfer characteristic of mechanics characteristics in bolted joint structure is analyzed.

The calculation model based on state space theory is a new way to calculate the stress / displacement field in bolted joints rapidly. It can rapidly and accurately obtain the relationship between the mechanics characteristics distributions in bolted joint and the factors such as structure, material, load, and so on, and has a wide application prospect in the design and optimization process of bolted joints.

This model still has some shortcomings. For example, because of ignoring the flatness, waviness and roughness of the contact surfaces, there will be a deviation between the calculation results and the actual situation to some extent. The object of the model is only limited to simple geometric shapes and force conditions. The analytical model of the mechanics characteristics of the complex geometry parts subjected to non-axisymmetric loads or horizontal load (unidirectional load, rotational load) needs further study.

Acknowledgement

This work supported by National Natural Science Foundation of China(No.51475077,51005038) and Science and Technology Foundation of Liaoning, China (No.201301002, 2014028012).

References

[1] Honglin Z, Qingxin D, Ming Z, et al. Theoretic analysis on and application of behaviors of machine tool joints[J]. Chinese Journal of Mechanical Engineering, 2008, 44(12): 208-214.

[2] Nassar S A, Abboud A. An improved stiffness model for bolted joints[J]. Journal of Mechanical Design, 2009, 131(12): 121001.

[3] Mi L, Yin G, Sun M, et al. Effects of preloads on joints on dynamic stiffness of a whole machine tool structure[J]. Journal of Mechanical Science and Technology, 2012, 26(2): 495-508.

[4] Benhamena A, Talha A, Benseddiq N, et al. Effect of clamping force on fretting fatigue behaviour of bolted assemblies: Case of couple steel–aluminium[J]. Materials Science and Engineering: A, 2010, 527(23): 6413-6421.

[5] Tian H, Li B, Liu H, et al. A new method of virtual material hypothesis-based dynamic modeling on fixed joint interface in machine tools[J]. International Journal of Machine Tools and Manufacture, 2011, 51(3): 239-249.

[6] Wang L, Du R, Jin T, et al. Matching Design for Bolted Joints Based by Effective Contact Radius Maximization[J]. Journal of Xi'an Jiaotong University, 2013, 7: 013.

[7] Namazi M, Altintas Y, Abe T, et al. Modeling and identification of tool holder–spindle interface dynamics[J]. International Journal of Machine Tools and Manufacture, 2007, 47(9): 1333-1341.

[8] Singhal V, Litke P J, Black A F, et al. An experimentally validated thermo-mechanical model for the prediction of thermal contact conductance[J]. International Journal of Heat and Mass Transfer, 2005, 48(25): 5446-5459.

[9] Voller G P, Tirovic M. Conductive heat transfer across a bolted automotive joint and the influence of interface conditioning[J]. International Journal of Heat and Mass Transfer, 2007, 50(23): 4833-4844.

[10] Chandrashekhara K, Muthanna S K. Analysis of a thick plate with a circular hole resting on a smooth rigid bed and subjected to axisymmetric normal load[J]. Acta Mechanica, 1979, 33(1-2): 33-44.

[11] Sawa T, Kumano H, Kobayashi F, et al. On the characteristics of bolted joints with gaskets: stress analysis of a metal flat gasket interposed between two hollow cylinders[J]. Bulletin of JSME, 1984, 27(227): 900-908.

[12] Sawa T, Kumano H, Morohoshi T. The contact stress in a bolted joint with a threaded bolt[J]. Experimental Mechanics, 1996, 36(1): 17-23.

[13] Gray P J, McCarthy C T. A global bolted joint model for finite element analysis of load distributions in multi-bolt composite joints[J]. Composites Part B: Engineering, 2010, 41(4): 317-325.

[14] Guo Y, Parker R G. Stiffness matrix calculation of rolling element bearings using a finite element/contact mechanics model[J]. Mechanism and Machine Theory, 2012, 51: 32-45.

[15] Sethuraman R, Kumar T S. Finite element based member stiffness evaluation of axisymmetric bolted joints[J]. Journal of Mechanical Design, 2009, 131(1): 011012.

[16] Li S. Finite element analyses for contact strength and bending strength of a pair of spur gears with machining errors, assembly errors and tooth modifications[J]. Mechanism and Machine Theory, 2007, 42(1): 88-114.

[17]Xiang Y, Wang C M. Exact buckling and vibration solutions for stepped rectangular plates[J]. Journal of sound and Vibration, 2002, 250(3): 503-517.

[18] Chen W Q, Ding H J. On free vibration of a functionally graded piezoelectric rectangular plate[J]. Acta Mechanica, 2002, 153(3-4): 207-216.

[19] Ying J, Lü C F, Chen W Q. Two-dimensional elasticity solutions for functionally graded beams resting on elastic foundations[J]. Composite Structures, 2008, 84(3): 209-219.

[20] Hongyu S, Jiarang F. Axisymmetric bending for thick laminated circular plate under a concentrated load[J]. Applied Mathematics and Mechanics, 2000, 21(1): 95-102.

[21] Hongyu S. Analytical solution of bending problem for thick laminated circular plate on elastic foundation with free edges[J]. Chinese Journal of Geotechnical Engineering, 2000, 3: 011.

Seismic Response of Structure under Nonlinear Soil-Structure Interaction Effect

***Narith Prok[1], †Yoshiro Kai[2]**

[1]Department of Infrastructure System Engineering, Kochi University of Technology, Japan.
[2] Department of Infrastructure System Engineering, Kochi University of Technology, Japan.

*Presenting author: 178010z@gs.kochi-tech.ac.jp
†Corresponding author: kai.yoshiro@kochi-tech.ac.jp

Abstract

Seismic response of structure under soil-structure interaction effect (SSI) is an impressive subject in earthquake engineering domain. Many analytical models and methods have been proposed and utilized. These methods can be categorized as direct and substructure (indirect) approach. Due to the simplicity requirement, substructure approach is frequently utilized in practical work and research field. In this approach, the analysis procedure is distinguished into three steps: foundation input motion (FIM), dynamic impedance (Spring-Dashpot), and seismic response of structure. However, the state of problem in this approach was found and needed to improve. In the existing analytical model under substructure approach, SSI problem is performed with equivalent-linear of soil material and motion in frequency domain (FD). This restriction can lead to mismatched response results between SSI analysis and actual response of structure during earthquake disaster.

Therefore, the objective of this paper is to propose an analytical model considering nonlinear response of soil material and motion in time domain (TD), which leads to perform the seismic response of structure under nonlinear SSI effect using substructure approach.

In this paper, the proposed analytical model procedure considering nonlinear response of soil material and motion were presented. An example was provided to validate this proposed analytical model. Moreover, the seismic response of structure under existing analytical model and proposed analytical model considering nonlinear response of soil material and motion were conducted. The seismic response of structure was performed under linear response of base-shear, overturning-moment, acceleration, and relative-displacement. Furthermore, the foundation stiffness-damping and hysteretic curve were also provided.

According to the nonlinear response motion in TD from the proposed analytical model, this motion showed a good agreement compared to the linear and equivalent-linear response of ground motion in FD. This agreement confirmed about the validation of this proposed analytical model. Furthermore, the seismic response of structure, it was showed that the response results under existing analytical model were larger than the responses under the proposed analytical model. These discrepancies showed about the overestimated results of using existing model compared to the actual response of structure under earthquake disaster.

Keywords: Soil-structure interaction, substructure approach, nonlinear response of soil material, nonlinear SSI effect.

Introduction

Soil-Structure Interaction (SSI) problem is regarded as a crucial major in earthquake engineering domain. SSI analysis permits evaluating the seismic response of structure and foundation system including the interaction effect of soil medium. This analysis leads to an understanding the actual response of structure under earthquake disaster and controlling the damage response of structural elements.

In order to perform SSI analysis, there are three significant interaction effects that have to consider: kinematic interaction effect, inertial interaction effect, and soil-foundation flexibility effect [1]. To evaluate these interaction effects, various methods have been proposed and

utilized such as Finite Element Method (FEM), Boundary Element Method (BEM), the coupling of FEM-BEM, Discrete Element Method (DEM), etc. However, these methods can be categorized as direct and substructure (indirect) approach [2]. In direct approach, the structure and soil are simulated within the same model and analyzed as a complete solution. This approach can deal with complicated structural geometry and soil condition. Many studies have been conducted base on this approach. However, this approach is rarely used in practical work, especially for complex geometrical structure and nonlinearity behavior of soil medium as a result of large computer-storage, running time, and cost consumption [3]. In substructure approach, SSI problem is commonly distinguished into three evaluation steps, which are combined to a complete solution of the seismic response of the whole structure base on the law of superposition [1]. These evaluation steps include foundation input motion (FIM), dynamic impedance (Spring-Dashpot), and seismic response of structure. This approach is widely used in research and practical works due to the simplicity, time, and cost consumption. However, this approach is commonly performed with equivalent-linear response of soil material and motion in FD. This restriction can cause mismatched structural responses between analysis and actual response of structure under earthquake disaster.

In order to deal with this restriction, the objective of this paper is to propose an analytical model considering nonlinear response of soil material and motion, which facilitates performing seismic response of structure under nonlinear SSI effect using substructure approach. In order to obtain this objective, the 3D RC frame structural model was used in this study. This structure was assumed as a rigid surface foundation and supported by uniform soil medium. The Kobe earthquake record data was used as input motion in this study. Other relevant parameters were presented in the following sections.

Existing Analytical Model of SSI Effect under Substructure Approach

Foundation Input Motion (FIM)

FIM can be derived from the relationship of free field ground motion (FFGM) and transfer function. FIM component is composed by translation and rotation motion that can be expressed in Eq. (1) and (2) [4] [5], respectively.

$$u_{FIM} = f\left(H_u, u_g\right) \tag{1}$$

$$\phi_{FIM} = f\left(u_g, I_\phi, B\right) \tag{2}$$

Where

u_{FIM}, ϕ_{FIM} : translation and rotation of FIM

u_g : free field ground motion

H_u, I_ϕ : translation and rotation of transfer function

B : foundation dimension

In this step, FFGM is performed in FD using equivalent-linear method, which is extensively described in the geotechnical earthquake engineering [6]. This method has been used and described in many programs such as SHAKE [7], EERA [8], etc. According to this method, the equivalent-linear response of soil material (G_{EL}, ξ_{EL}) and the corresponding FFGM at the ground surface (u_g) were achieved.

Regarding for the transfer function (H_u, I_ϕ), various expressions under relationship of foundation and wave motion types were described in NIST guideline for SSI problem [5], Mylonakis et al [4], Nikolaou et al [9], etc.

Based on the description above, the FIM can be achieved corresponding to soil conditions, wave motions, and foundation types.

Dynamic Impedance (Spring-dashpot)

Dynamic impedance function is an interaction function between foundation and soil medium. This function is represented by spring and dashpot of soil-foundation interaction system as

shown in Fig. 1. The equation of this function is composed by the stiffness (real part) and damping (imaginary part) as expressed in Eq. (3) and (4) [3]-[5].

$$\bar{K}_i = k_i + i\omega c_i \tag{3}$$

$$\bar{K}_i = k_i\left(1 + i2\beta_i\right) \tag{4}$$

$$\beta_i = \frac{\omega c_i}{2k_i} \tag{5}$$

Where

β_i: radiation damping ratio of foundation

$\bar{K}_i$: complex-valued impedance function

k_i, c_i: frequency-dependent foundation stiffness and dashpot

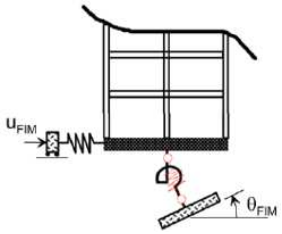


Figure 1. Soil-foundation system [5]

In the Eq. (4), the foundation stiffness k_i can be expressed in function of constant equivalent-linear soil material values (G, υ) from the FFGM analysis in FD and foundation dimensions while the foundation damping can be expressed in function foundation stiffness and radiation damping ratio as shown in Eq. (5).

Various expressions have been proposed for both functions (k_i, c_i) related to different types of foundation and soil conditions. For instance, Mylonakis et al. [4], Gazatas [11] [12], Pais et al. [13] have proposed the solutions for surface and embedded foundation with different types of soil condition while the solution for single pile and group of pile have been discussed in NIST [5].

Seismic Response of Structure

For the seismic response of structure, the structure was assumed to support by spring and dashpot that computed in the second step and subjected to FIM in the first step, as shown in Fig. 2. The seismic response of the structure under SSI effect can be solved under both FD and TD [25] as expressed in Eq. (6) and (7), respectively.

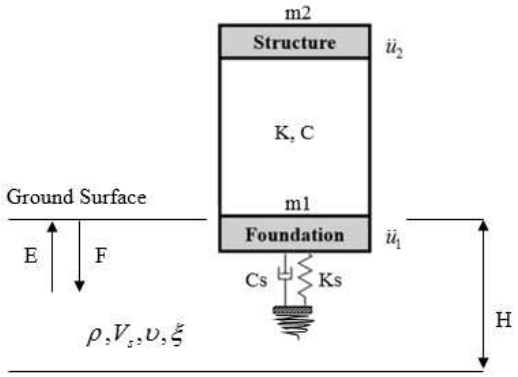


Figure 2. Structure model under SSI effect

-FD Equation:

$$\left(-\omega^2[M]+i\omega[C]+[K]\right)\{U\}=\omega^2[M]\{1\}U_0 \tag{6}$$

-TD Equation:

$$[M]\{\ddot{u}\}+[C]\{\dot{u}\}+[K]\{u\}=-[M]\{1\}\ddot{u}_0 \tag{7}$$

Where

$[M],[C],[K]$: mass, damping, stiffness of the whole structure

U,U_0: structure displacement and foundation input motion

$\ddot{u},\dot{u},u$: acceleration, velocity and displacement of structure

$\ddot{u}_0$: foundation input motion

According to the description in the existing analytical model above, the seismic response of structure considering SSI effect is solved under equivalent-linear of soil material and FFGM in FD. However, due to this condition, this analytical model might not represent the actual response of structure under earthquake disaster. Therefore, an analytical model of SSI effect considering the nonlinear response of soil material and FFGM in TD was proposed as in the following sections.

Proposed Analytical Model of Nonlinear SSI Effect using Substructure Approach

Free Field Ground Motion Analysis in TD

As described above, the existing analytical model of SSI problem can be performed only with equivalent-linear response of soil material, which can lead to mismatched response of analysis results compared to the actual response of structure under earthquake disaster. Thus, the objective of this paper is to propose an analytical model considering the nonlinear response of soil material and motion that facilitated performing the seismic response of structure under nonlinear SSI effect.

In order to perform FFGM analysis in TD, the cooperation of FFGM analysis in FD was necessary. FFGM analysis in TD was performed by using Newmark's equation [14]-[18], as expressed in Eq. (8). Furthermore, the modified Ramberg-Osgood model [19] was used for hysteretic rule of nonlinear response of soil material. In each layer, soil model can be represented by consistent mass, dashpot, and nonlinear spring, as shown in Fig. 3.

$$[M]\{\ddot{u}\}+[C]\{\dot{u}\}+[K]\{u\}=-[M]\{I\}\ddot{u}_g \tag{8}$$

Where

$[M],[C],[K]$: mass, damping, stiffness matrix

$\{\ddot{u}\},\{\dot{u}\},\{u\}$: acceleration, velocity, displacement vector

$\{\ddot{u}_g\}$: acceleration at the base of column

$\{I\}$: unit vector

The soil element matrix in each layer can be expressed from the Eq. (9) to (11):

$$[M]=\frac{\rho h}{6}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \tag{9}$$

$$[K]=\frac{G}{h}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{10}$$

$$[C] = \alpha_R [M] + \beta_R [K] \tag{11}$$

Where

ρ : unit weight of soil, h : thickness in each layer

G :shear stiffness of soil, α_R, β_R : coefficient of Rayleigh damping

According to Rayleigh [20], α_R and β_R coefficient can be computed using two significant modes m and n:

$$\frac{1}{2}\begin{bmatrix} 1/\omega_n & \omega_m \\ 1/\omega_m & \omega_n \end{bmatrix}\begin{Bmatrix} \alpha_R \\ \beta_R \end{Bmatrix} = \begin{bmatrix} \xi_m \\ \xi_n \end{bmatrix} \tag{12}$$

This matrix can be solved as the following expressions:

$$\alpha_R = 2\omega_m\omega_n\left(\frac{\omega_m\xi_n - \omega_n\xi_m}{\omega_m^2 - \omega_n^2}\right) \qquad \beta_R = 2\left(\frac{\omega_m\xi_m - \omega_n\xi_n}{\omega_m^2 - \omega_n^2}\right)$$

If the damping ratio is frequency independent, α_R and β_R coefficient becomes:

$$\alpha_R = 2\xi\left(\frac{\omega_m\omega_n}{\omega_m + \omega_n}\right) \qquad \beta_R = 2\xi\left(\frac{1}{\omega_m + \omega_n}\right)$$

Where

ξ : damping ratio ω_m, ω_n : two significant frequency modes

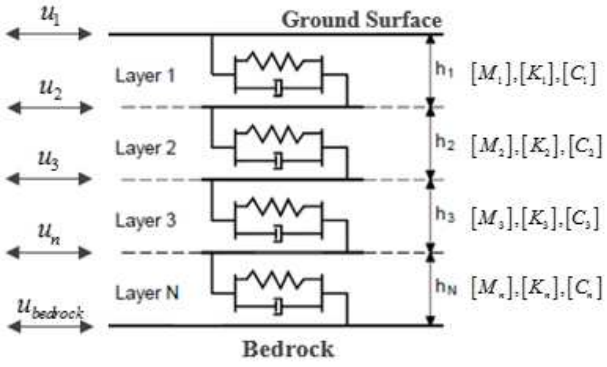


Figure 3. Soil deposit model for TD analysis

As mentioned above, the nonlinear response of soil material was conducted using the modified Ramberg-Osgood model, as shown in Fig. 4, which was proposed by Tatsuoka et al. [19]. The skeleton and hysteretic curve equations were expressed in Eq. (13) and (14), respectively.

$$\gamma = \frac{\tau}{G_0}\left(1 + \alpha|\tau|^{\beta}\right) \tag{13}$$

$$\frac{\gamma \pm \gamma_a}{2} = \frac{\tau \pm \tau_a}{2G_0}\left(1 + \alpha\left|\frac{\tau \pm \tau_a}{2}\right|^{\beta}\right) \tag{14}$$

Where

$$\beta = \frac{2\pi h_{\max}}{2 - h_{\max}}, \qquad \alpha = \left(\frac{2}{\gamma_{0.5} G_0} \right)^{\beta}$$

$\gamma_{0.5}$: corresponds to $G/G_0 = 0.5$ α, β: parameter of modified R-O

γ_a, τ_a: reversal shear strain and stress G_0: initial shear soil stiffness

$h_{\max}$: maximum soil damping

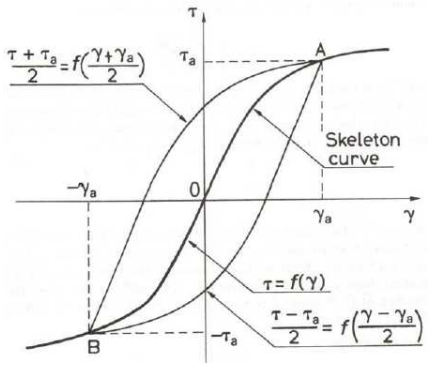


Figure 4. Stress-strain relationship of Ramberg-Osgood model [21]

According to hysteretic rule, the nonlinear response of shear stiffness $G_i(t)$ can be derived from Eq. (15) and shown in Fig. 5.

$$G_i(t) = \frac{\tau_i - \tau_{i-1}}{\gamma_i - \gamma_{i-1}} \tag{15}$$

Where

τ_i, τ_{i-1}: reversal shear stress of point i and i-1

γ_i, γ_{i-1}: reversal shear strain of point i and i-1

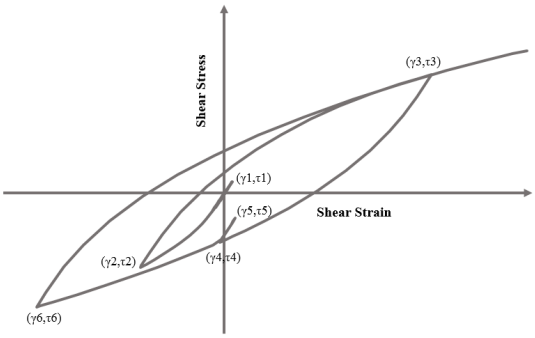


Figure 5. Reversal points of shear stress-strain

Besides this, in order to perform nonlinear response analysis of FFGM in TD, the analytical procedure of input motion at the base of soil column was very important in order to obtain a properly output motion at the ground surface.

Analytical Procedure for Input Motion in TD

In this step, the FFGM in FD was needed to obtain the input motion for FFGM analysis in TD.

In linear analysis (LN), the target earthquake motion was input at the base of soil column (or surface layer) as an outcrop motion (*2E*). Then, the within output motion (*E+F*) was extracted at the base of soil column and applied this motion at the same layer of soil column (as input motion) for TD analysis, as shown in Fig. 6. The within motion (*E+F*) of any location is an actual motion of that location.

In nonlinear analysis (NL), the procedure was the same as linear analysis but it was required to perform in both linear and equivalent-linear (EL) analysis in FD and the within output motion (*E+F*) of both analyses were significant to be the same or almost the same. Due to this requirement, some extra layers might be needed. Then, this within output motion (*E+F*) was applied at the same layer of soil column for TD analysis, as shown in Fig. 6.

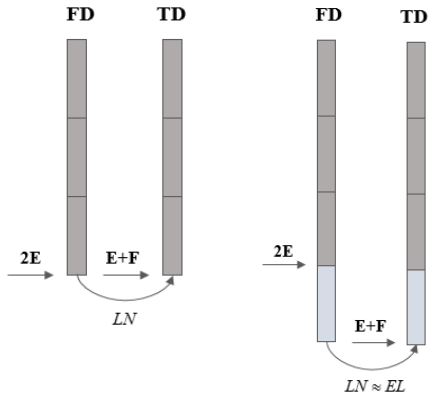


Figure 6. Linear and nonlinear input motion procedure for TD analysis

Furthermore, in order to validate the nonlinear response output motion at the ground surface in TD, a comparison of this motion with linear and equivalent-linear analysis at the ground surface were necessary. This comparison can lead to an understanding how correctly of this nonlinear response motion.

Example of FFGM Analysis in TD

In this study, the uniform soil column in depth 20m was assumed rested on the bedrock. This uniform soil column consisted the same properties as in class E of IBC [22], as shown in table 1. Kobe earthquake record data were assumed as input motion at the base of soil column. The motion in X and Y direction were assumed as the same as the motion in EW and NS of record data as shown in Fig. 7 while the motion in UD direction was ignored in this study.

Table 1. Uniform soil column properties

H (m)	Vs (m/s)	γ (kN/m^3)	ξ (%)
0.00-20.0	180	18.0	5
Bedrock	900	21.0	1

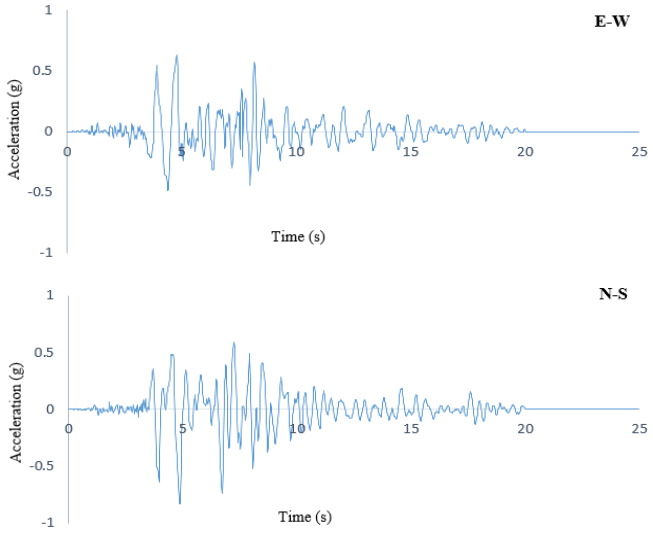


Figure 7. Kobe earthquake record motion data

For linear analysis, according to the procedure described above, the output result at the ground surface for both analyses was shown in Fig. 8.

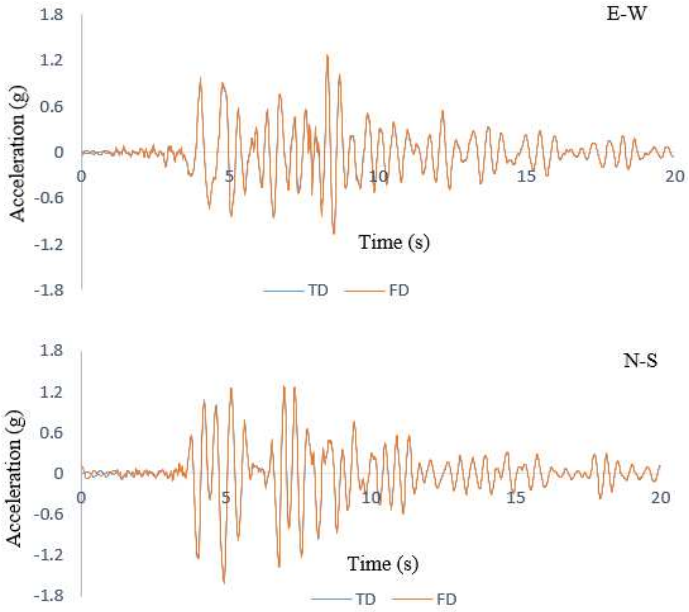


Figure 8. Linear analysis of FFGM in FD and TD

For nonlinear analysis, according to the procedure described above, two extra layers were needed for this study case, as shown in Fig. 9. The first layer consisted 5m in depth and 900 m/s for shear velocity while the second layer consisted 800m in depth and 8km/s for shear velocity above the bedrock, which consisted 8 km/s for shear velocity. The within output motion in FD was shown in Fig. 10 and the output motion at the ground surface in TD was shown in Fig. 11.

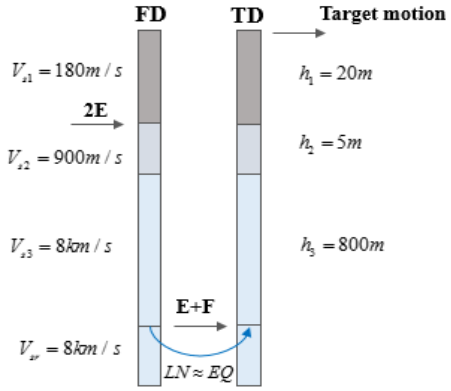
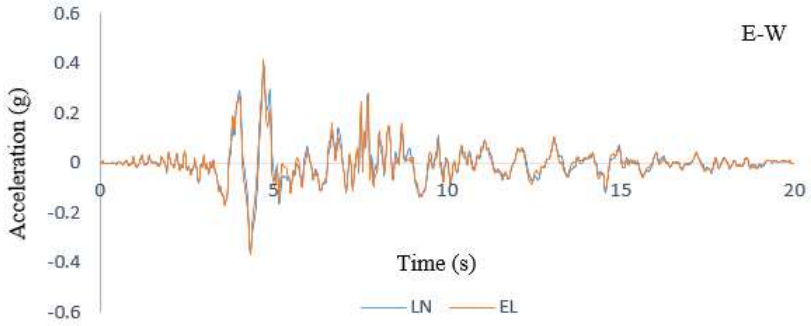


Figure 9. Analytical procedure for input motion in TD



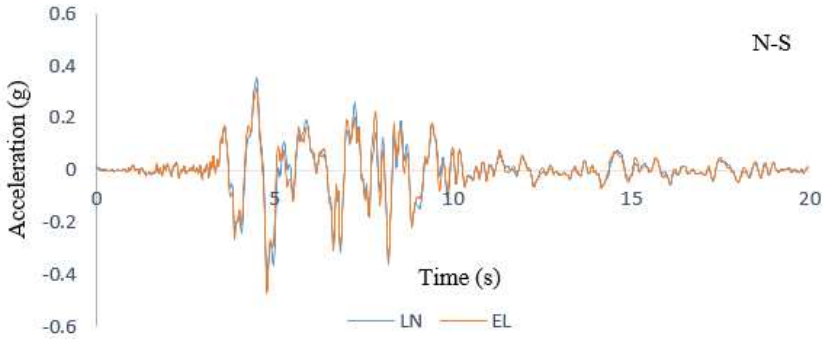


Figure 10. Within input motion (E+F) in TD

As shown in Fig. 10, the within output results ($E+F$) from both analyses showed a good agreement and adequate for input motion for TD analysis. These within outputs ($E+F$) were applied at the same layer and property for TD analysis. The FFGM at the ground surface for both directions and nonlinear response of soil stiffness (G_{NL}) were obtained.

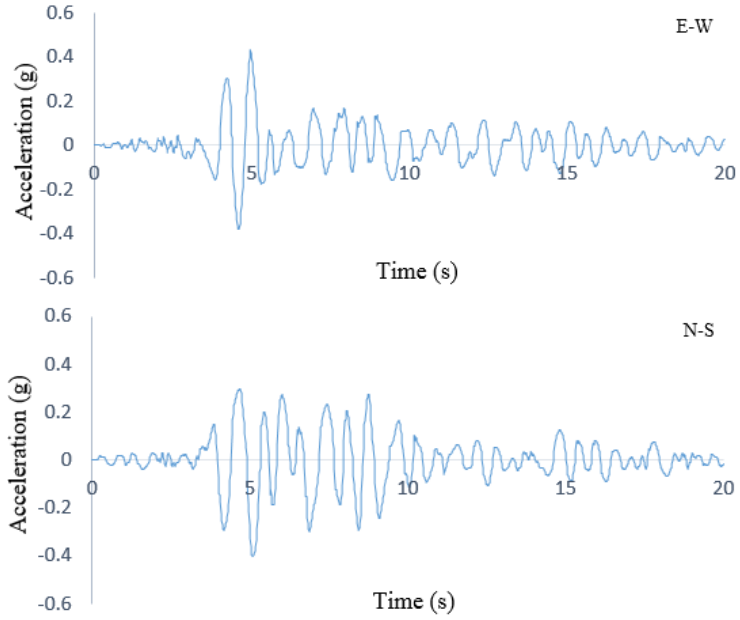
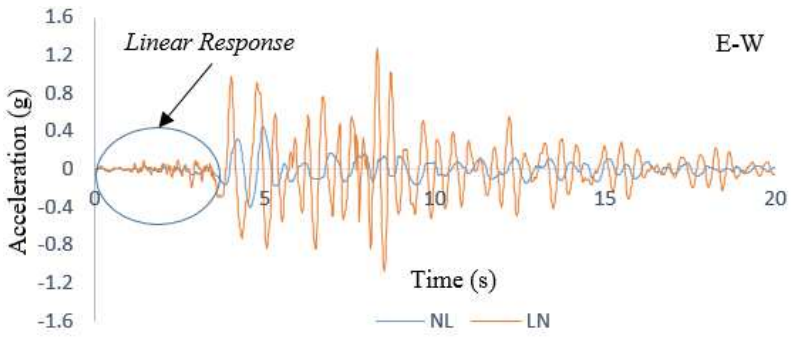


Figure 11. FFGM at the ground surface in TD

However, as described above, the comparison of these nonlinear response motions with linear and equivalent-linear motions at the ground surface in FD was necessary. These comparisons were shown in Fig. 12 and 13.



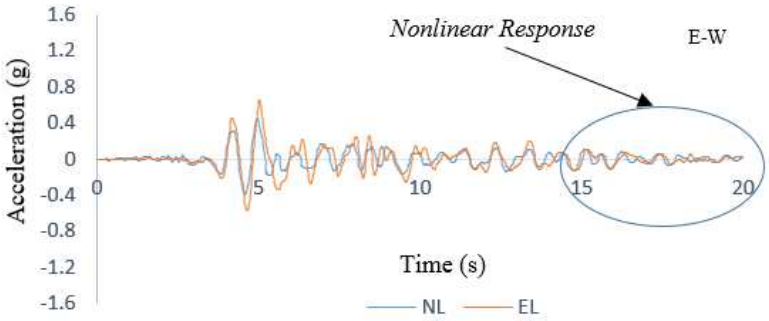


Figure 12. Comparison between LN, EL, and NL motion in E-W direction

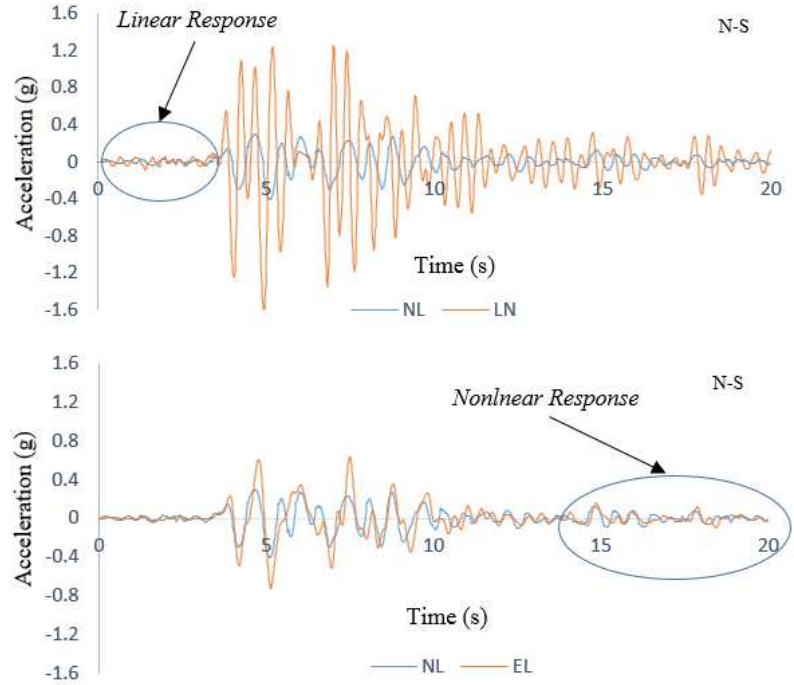
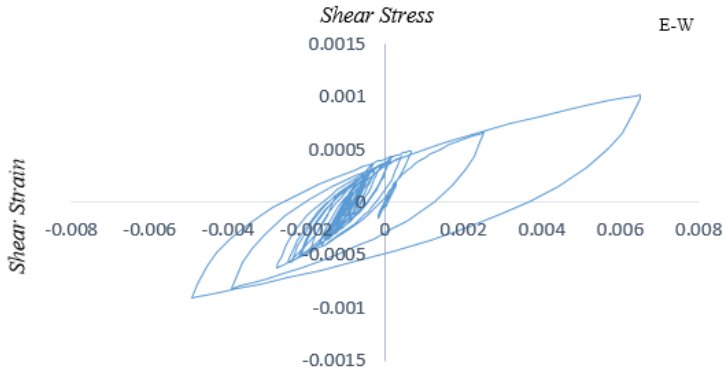


Figure 13. Comparison between LN, EL, and NL motion in N-S direction

As shown in Fig. 12 and 13, the nonlinear response result showed a good agreement with linear motion response for a few seconds from starting point and with equivalent-linear motion response for the last several seconds. These agreements confirmed that the nonlinear motion response at the ground surface in TD started from linear to nonlinear motion response. This confirmation showed about the validation of the proposed analytical model considering the nonlinear response of soil material and motion. The hysteretic curve of nonlinear response of soil medium at the ground surface was also provided, as shown in Fig. 14.



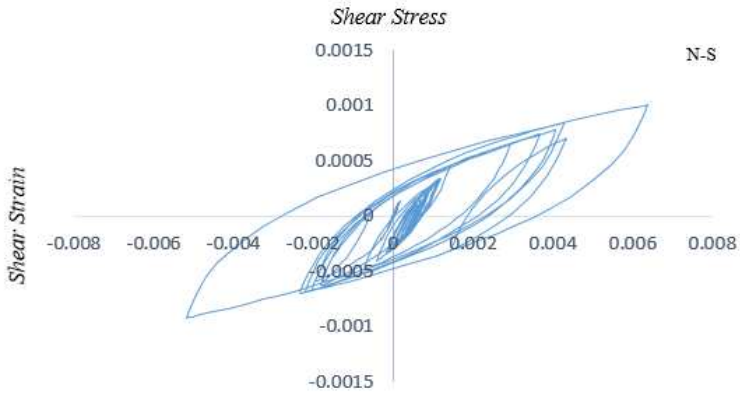


Figure 14. Hysteretic curve of nonlinear response of soil medium

The response of soil stiffness under both analytical models, FD and TD, were shown in Fig.15.

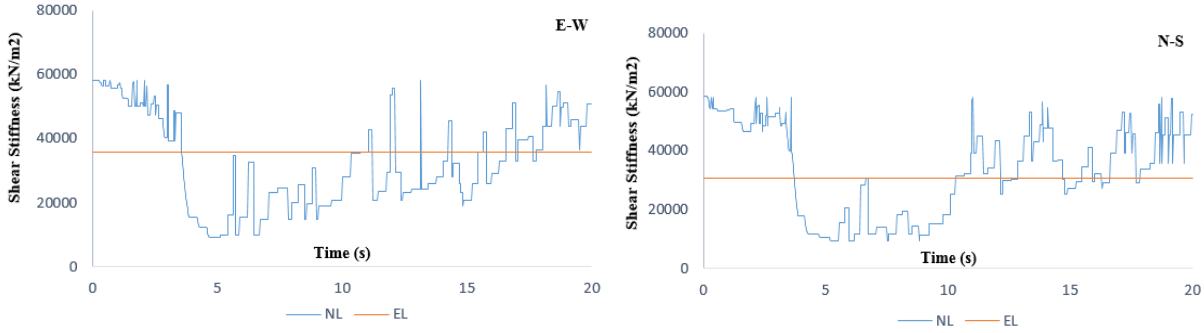


Figure 15. Soil stiffness response analysis in TD and FD

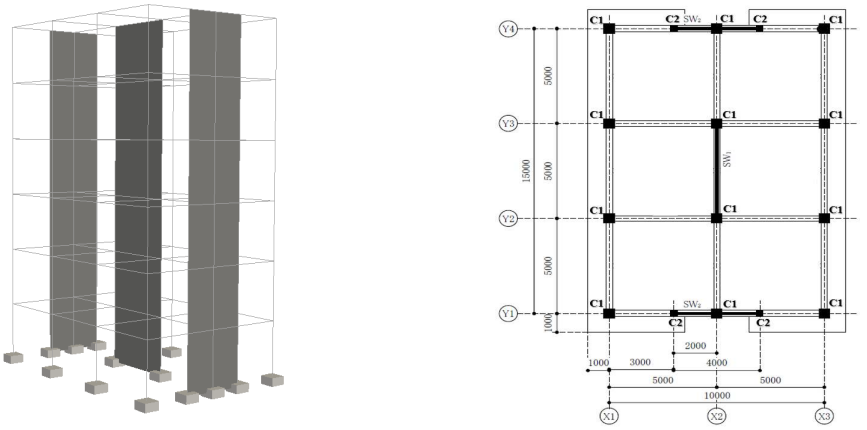
Seismic Response of RC Frame Structure under Nonlinear SSI Effect

After obtaining the nonlinear response of soil materials and FFGM at the ground surface, the linear response of RC frame structure under equivalent-linear and nonlinear SSI effect were conducted.

In order to achieve this objective, the 3D RC frame structural model (from E-defense test [23] [24]) was used in this study. This structure was supported by rigid surface foundation and was assumed to rest on uniform soil deposit subjected to vertically incident S wave. The relevant parameters were shown in the following sections.

Structural Model Assumption

As mentioned above, the model of 3D frame structure was used in this study as shown in Fig. 16. This structural model consisted six stories and 3.5m for height in each story. There were two spans in X-direction and three spans in Y-direction with the same length 5m in each span. In this study, the column C1 section was 0.5mx0.5m with 8-D19 and C2 section was 0.3mx0.3m with 4-D19, beam section was 0.3mx0.5m with 5-D19, and both shear-wall and sidewalls thickness were 0.15m with doubly reinforcing bar D10@300. Furthermore, the shear reinforcing bar of column was D10@100 while beam element was D10@200. The nominal strength of reinforcing bars were SD345 and SD295 for D19 and D10, respectively, and concrete strength was 21MPa for all structural elements. Besides this, the non-structural element load was assumed 3.0 kPa and live load 2.5kPa for each story.



(a) Perspective view (b) Top view [24]

Figure 16. E-Defense test structure model

Soil Property and Input Motion

The uniform soil property and input motions were assumed to be the same as described in the previous sections. The equivalent-linear and nonlinear of soil stiffness values were shown in Fig. 15.

Foundation Input Motion (FIM)

Due to the condition of rigid surface foundation and subjected to the vertically incident S wave, the FIM was the same as the FFGM [4]. The FIM response of both FD and TD were shown in Fig. 17 and 18, respectively.

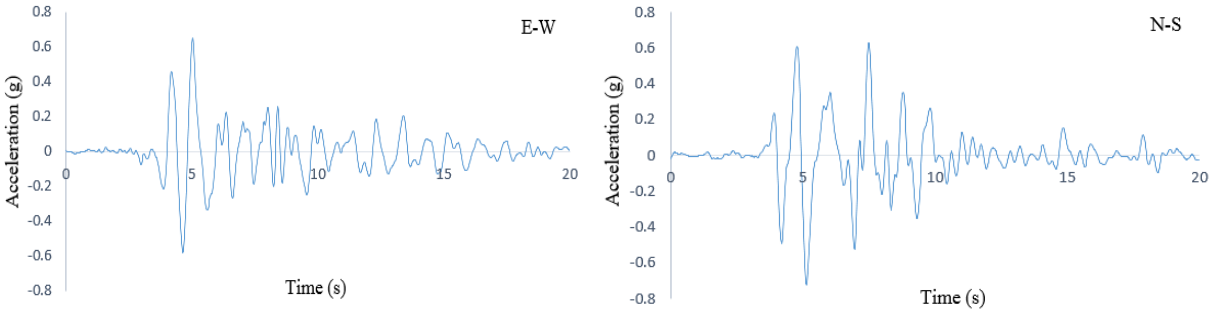


Figure 17. Foundation Input Motion in FD

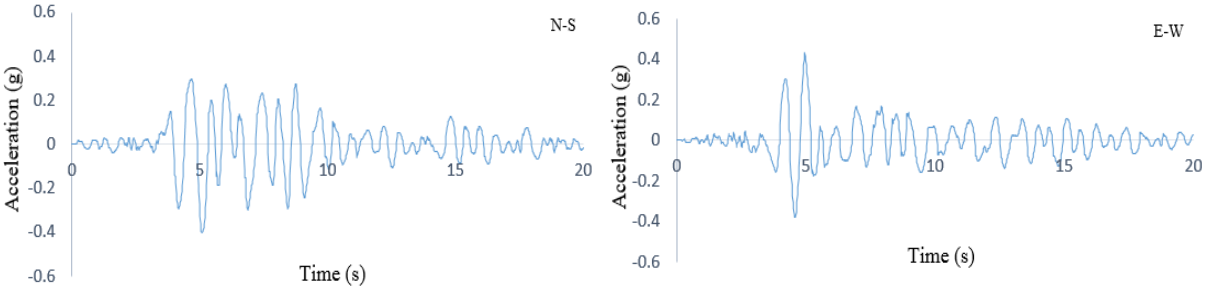


Figure 18. Foundation Input Motion in TD

Dynamic Impedances

The dynamic impedance of rigid surface foundation was expressed in Eq. (3) and (4). The static stiffness, dynamic stiffness modifier and radiation damping can be expressed in the following equations:

1. Static Stiffness

$$K_x = K_y - \frac{0.2}{0.75-\upsilon} GL\left(1-\frac{B}{L}\right) \qquad K_{xx} = \frac{G}{1-\upsilon}\left(I_x\right)^{0.75}\left(\frac{L}{B}\right)^{0.25}\left[2.4+0.5\left(\frac{B}{L}\right)\right] \tag{16}$$

$$K_y = \frac{2GL}{2-\upsilon}\left[2+2.5\left(\frac{B}{L}\right)^{0.85}\right] \qquad K_{yy} = \frac{G}{1-\upsilon}\left(I_y\right)^{0.75}\left[3\left(\frac{L}{B}\right)^{0.15}\right] \tag{17}$$

$$K_z = \frac{2GL}{1-\upsilon}\left[0.73+1.54\left(\frac{B}{L}\right)^{0.75}\right] \qquad K_{zz} = GJ_t^{0.75}\left[4+11\left(1-\frac{B}{L}\right)^{10}\right] \tag{18}$$

2. Dynamic Stiffness Modifier

$$\alpha_x = 1.0 \qquad \alpha_{xx} = 1-\left[\frac{\left(0.55+0.11\sqrt{\frac{L}{B}-1}\right)a_0^2}{\left(2.4-\frac{0.4}{\left(\frac{L}{B}\right)^3}+a_0^2\right)}\right] \tag{19}$$

(19)

$$\alpha_y = 1.0 \qquad \alpha_{yy} = 1-\left[\frac{0.55a_0^2}{\left(0.6+\frac{1.4}{\left(\frac{L}{B}\right)^3}\right)+a_0^2}\right] \tag{20}$$

$$\alpha_z = 1-\left[\frac{\left(0.4+\frac{0.2}{\left(\frac{L}{B}\right)}\right)a_0^2}{\left(\frac{10}{1+3\left(\frac{L}{B}-1\right)}\right)+a_0^2}\right] \qquad \alpha_{zz} = 1-\left[\frac{\left(0.33-0.33\sqrt{\frac{L}{B}-1}\right)a_0^2}{\left(\frac{0.8}{1+0.33\left(\frac{L}{B}-1\right)}\right)+a_0^2}\right] \tag{21}$$

3. Radiation Damping

$$\beta_x = \left[\frac{4\left(\frac{L}{B}\right)}{\frac{K_x}{GB}}\right]\left[\frac{a_0}{2\alpha_x}\right] \qquad \beta_{xx} = \left[\frac{\left(\frac{4\psi}{3}\right)\left(\frac{L}{B}\right)a_0^2}{\left(\frac{K_{xx}}{GB^3}\right)\left[2.2-\frac{0.4}{\left(\frac{L}{B}\right)^3}+a_0^2\right]}\right]\left[\frac{a_0}{2\alpha_{xx}}\right] \tag{22}$$

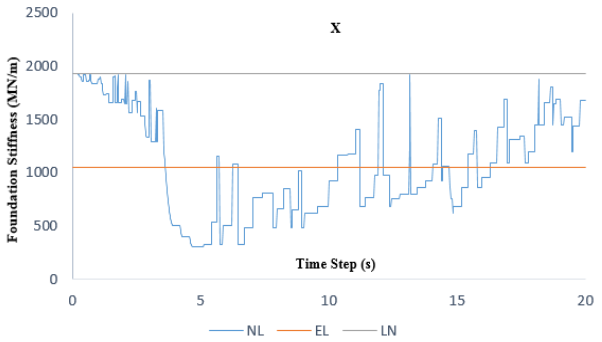
$$\beta_y = \left[\frac{4\left(\frac{L}{B}\right)}{\frac{K_y}{GB}}\right]\left[\frac{a_0}{2\alpha_y}\right] \qquad \beta_{yy} = \left[\frac{\left(\frac{4\psi}{3}\right)\left(\frac{L}{B}\right)^3 a_0^2}{\left(\frac{K_{yy}}{GB^3}\right)\left[\left(\frac{1.8}{1+1.75\left(\frac{L}{B}-1\right)}\right)+a_0^2\right]}\right]\left[\frac{a_0}{2\alpha_{yy}}\right] \tag{23}$$

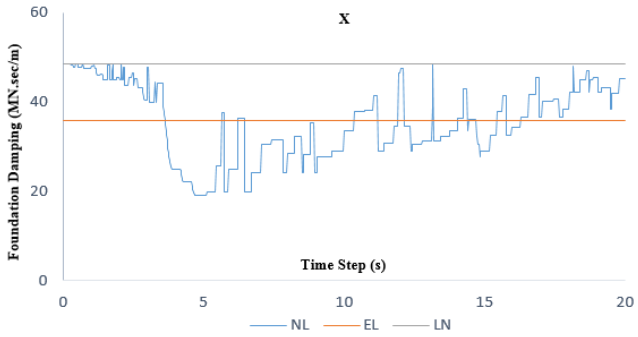
$$\beta_z = \left[\frac{4\psi\left(\frac{L}{B}\right)}{\frac{K_z}{GB}}\right]\left[\frac{a_0}{2\alpha_z}\right] \qquad \beta_{zz} = \left[\frac{\left(\frac{4}{3}\right)\left[\left(\frac{L}{B}\right)^3+\left(\frac{L}{B}\right)\right]a_0^2}{\left(\frac{K_{zz}}{GB^3}\right)\left[\frac{1.4}{1+3\left(\frac{L}{B}-1\right)^{0.7}}+a_0^2\right]}\right]\left[\frac{a_0}{2\alpha_{zz}}\right] \tag{24}$$

Linear, Equivalent-linear and Nonlinear Foundation Stiffness-Damping

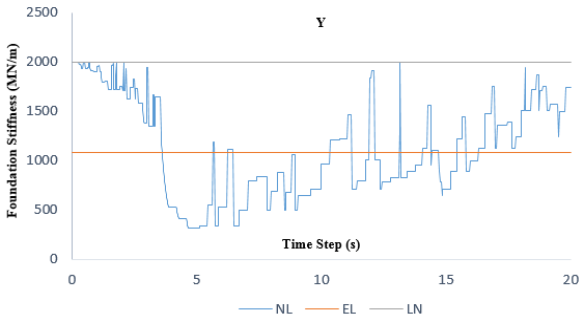
In this comparison, there are six directions for foundation stiffness k_i and damping c_i for rigid surface foundation as shown in Fig. 19.

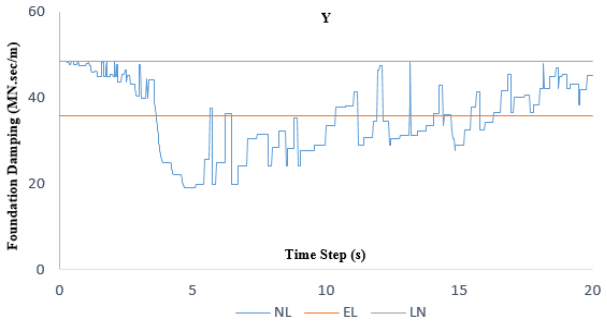
- Direction X:



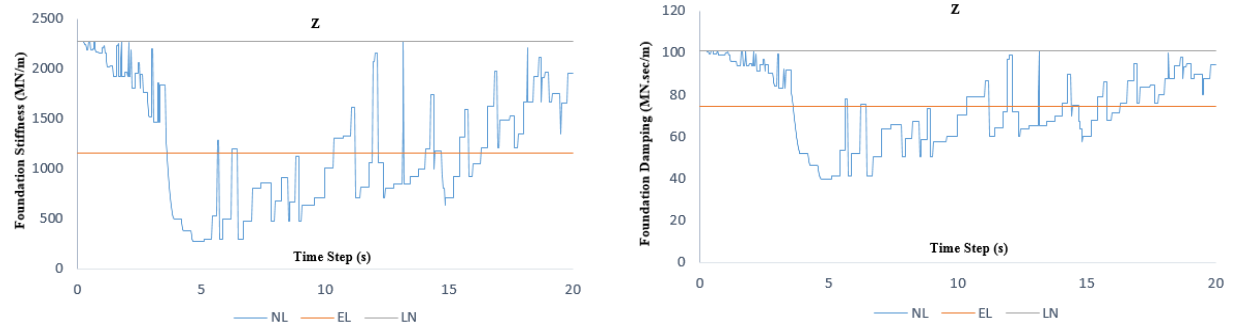


- Direction Y:

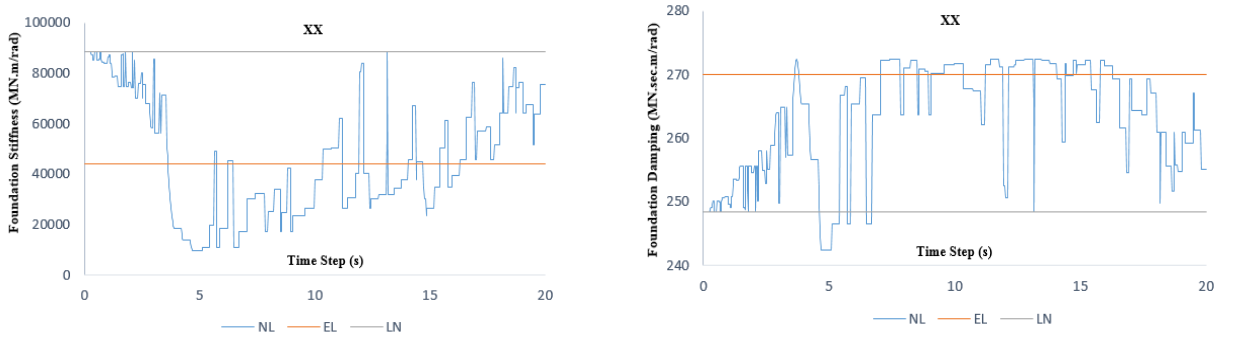




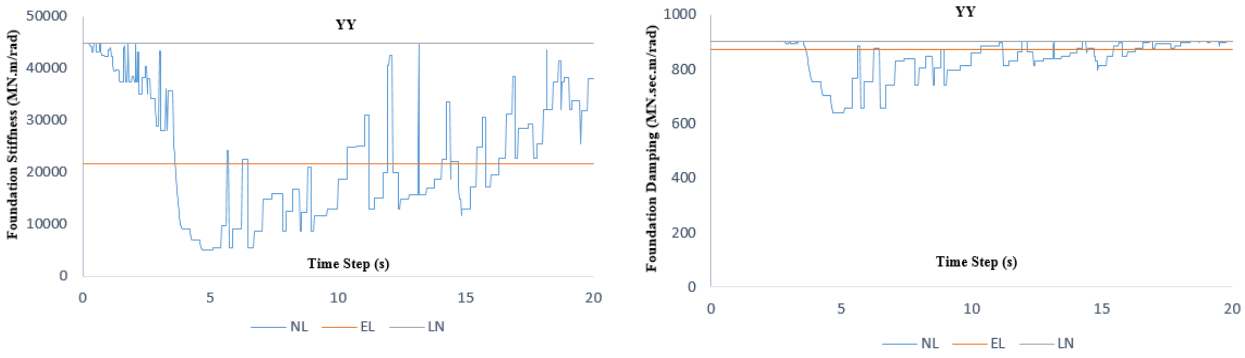
- Direction Z



- Direction XX:



- Direction YY:



- Direction ZZ:

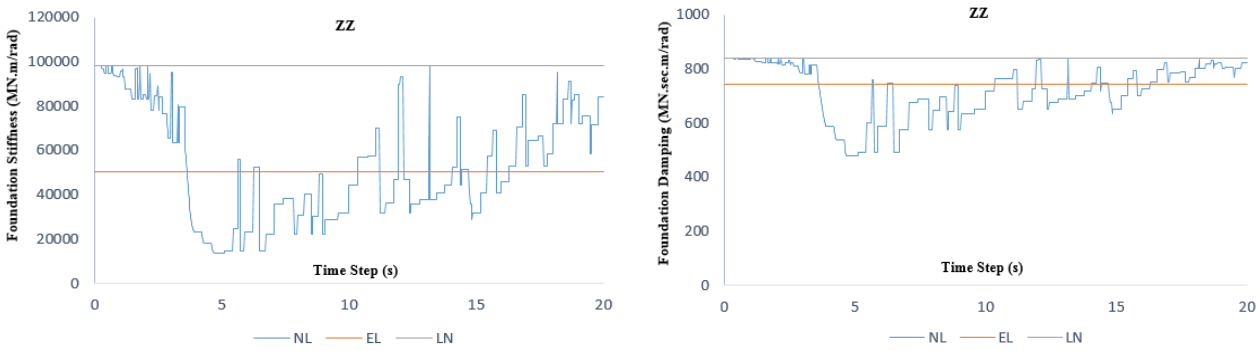
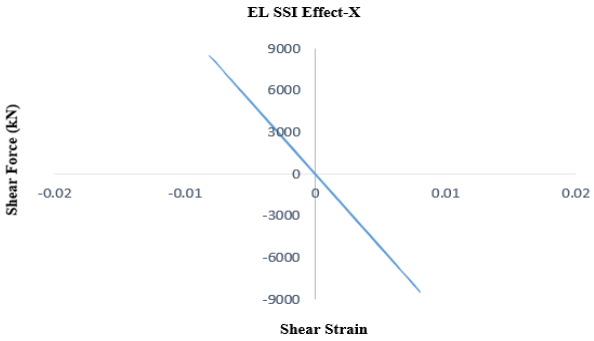
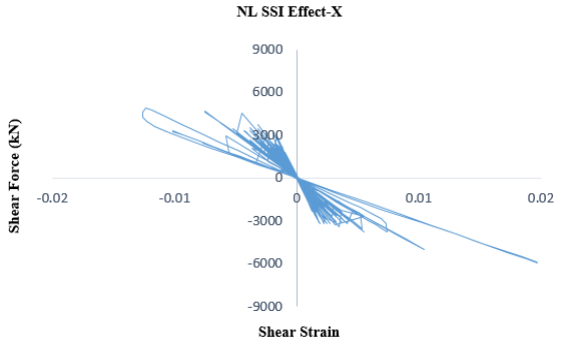
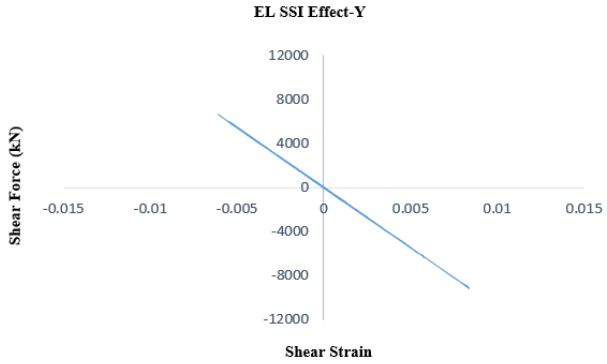


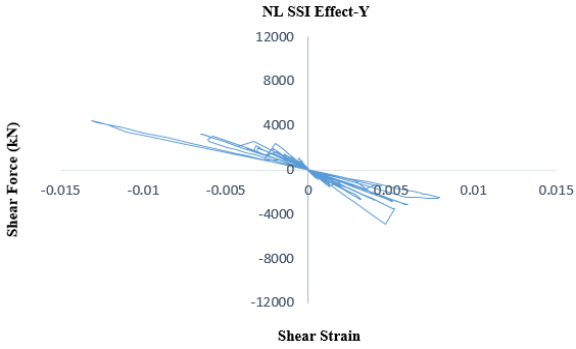
Figure 19. Linear, Equivalent-linear, and Nonlinear Foundation Stiffness

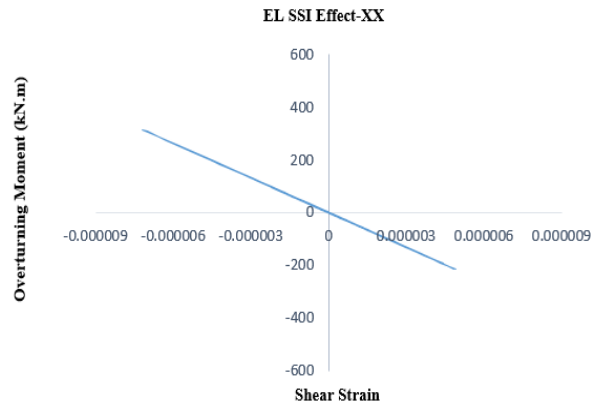
Hysteretic Curve of Foundation-Soil System

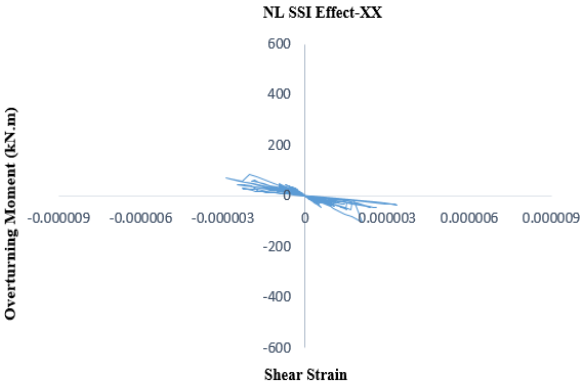


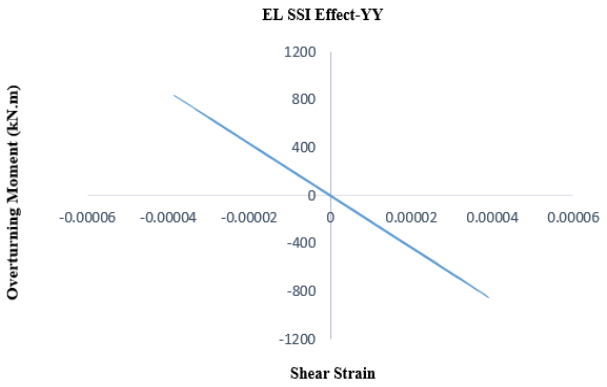


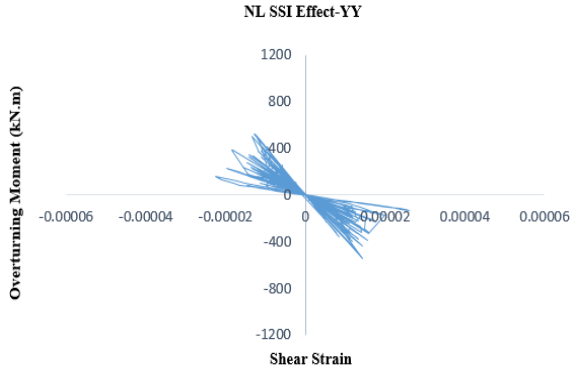












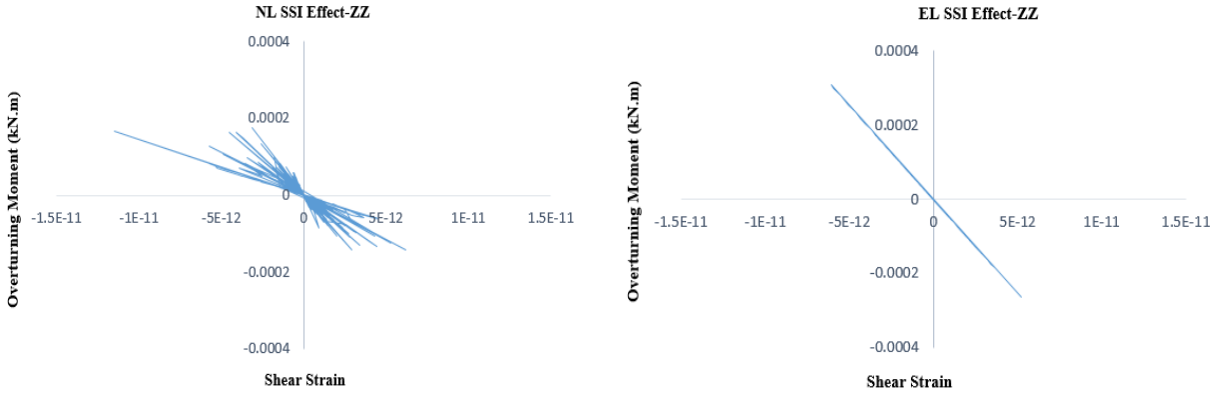


Figure 17 Hysteretic curve of foundation-soil system under both analytical models

Seismic Response of RC frame Structure

In this section, the seismic response of frame structure under both equivalent-linear and nonlinear SSI effect were presented under linear response of base-shear, overturning-moment acceleration, and relative displacement in TD based on Eq. (7), as shown from Fig. 17 to 20, respectively.

- Base-Shear:

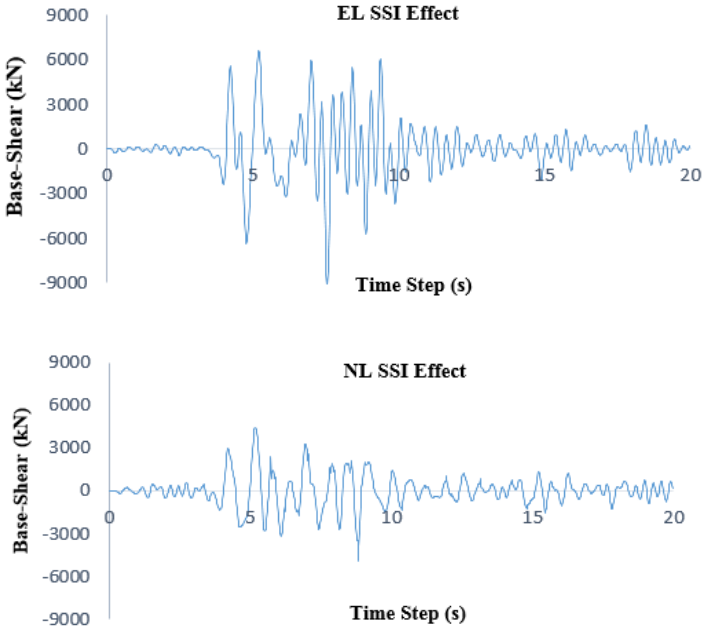
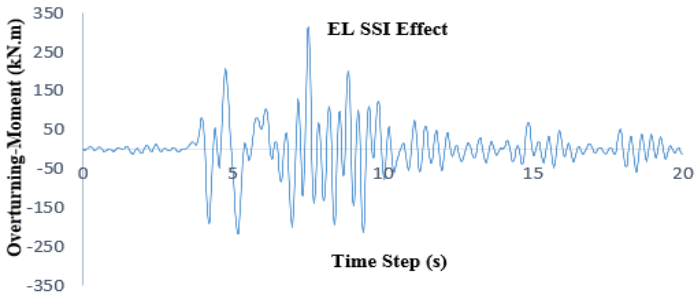


Figure 17. Base-Shear response under equivalent-linear and nonlinear SSI effect

- Overturning-Moment:



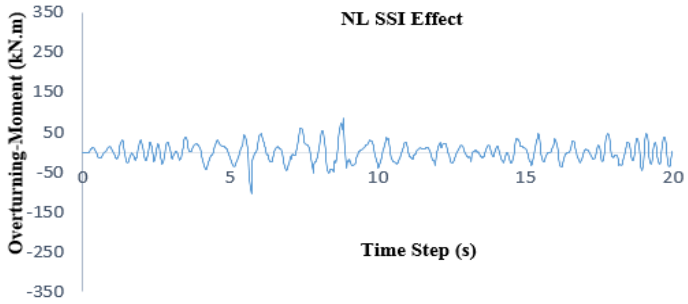


Figure 18. Overturning-Moment under equivalent-linear and nonlinear SSI effect

- Acceleration in each floor:

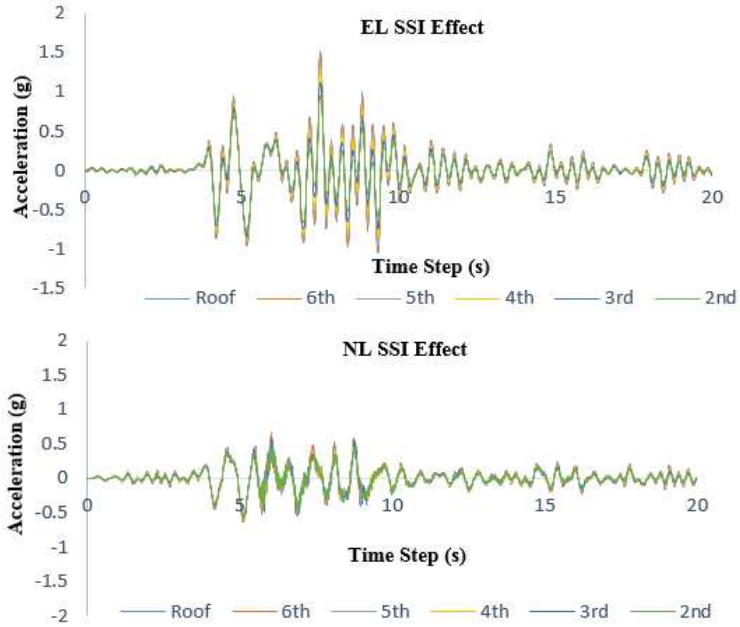


Figure 19. Acceleration under equivalent-linear and nonlinear SSI effect

- Relative Displacement:

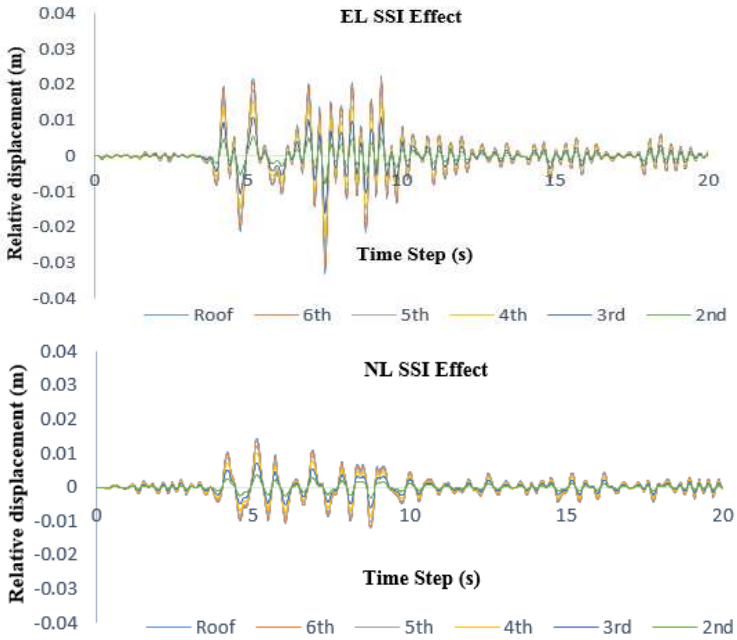


Figure 20. Relative displacement under equivalent-linear and nonlinear SSI effect

Based on the response results above, it was showed that the responses of structure under equivalent-linear SSI effect were larger than the responses under nonlinear SSI effect using substructure approach. These discrepancies showed about overestimated responses of using existing analytical model of SSI problem compared to the actual response of structure under earthquake disaster.

Conclusions

In this paper, the analytical model considering nonlinear response of soil material and motion in TD was proposed. The nonlinear response of motion in TD showed a good agreement with linear response for a few seconds from starting point and with equivalent-linear analysis in FD for the last several seconds. This agreement confirmed about the validation of proposed analytical model.

Furthermore, the seismic response of structure under existing and proposed analytical model were conducted under linear response of base-shear, overturning-moment, acceleration, and relative displacement. The output results showed that the structural response under existing model were larger than the responses under proposed model. These discrepancies showed about the overestimated results of using existing analytical model compared the actual response of structure under earthquake disaster.

In conclusion, the proposed analytical model considering nonlinear SSI effect using substructure approach on the structural response under earthquake loading would be a good candidate for SSI problem and showed about the adequateness of this approach compared to the actual response of structure.

References

[1] FEMA (2009) NEHRP recommended seismic provisions for new buildings and other structures, FEMA P-750/2009 Edition, Building Science for the Federal Emergency Management Agency, Washington, D. C.

[2] Carlo G. Lai, Mario M. (2013) Soil-structure interaction under earthquake loading: Theoretical framework, *ALERT Doctoral School 2013*.

[3] Wolf J. P. (1985) Dynamic soil-structure interaction, Prentice Hall.

[4] Mylonakis G., Sissy N., Gazetas G. (2006) Footings under seismic loading: Analysis and design issues with emphasis on bridge foundations, *Soil Dynamics and Earthquake Engineering* **26**, 824-853.

[5] National Institute of Standards and Technology (2012) Soil-structure interaction for building structures, U.S Department of Commerce, National Institute of Standards and Technology, Engineering Laboratory, Gaithersburg, MD 20899, U.S.

[6] Kramer S. L. (1996) Geotechnical earthquake engineering, Prentice Hall, New Jersey, U.S.

[7] Idriss I. M, Joseph I. S. (1992) User's manual for SHAKE91: A computer program for conducting equivalent linear seismic response analyses for horizontally layered soil deposits, University of California, Davis, California, U.S.

[8] Bardet J. P., Ichi K., Lin C. H. (2000) A computer program for equivalent linear earthquake site response analyses of layered soil deposits, University of Southern California, California, U.S.

[9] Nikolaou S., Mylonakis G., Gazetas G., Tazoh T. (2001) Kinematic pile bending during earthquakes: Analysis and field measurements, *Geotechnique* **51**, 425-440.

[10] Gazatas G. (1991) Formulas and charts for impedances of surface and embedded foundations, *Geotechnical Engineering* **117**, 1363-1381.

[11] Gazatas G. (1983) Analysis of machine foundation vibrations: State of the art, *Soil Dynamic and Earthquake Engineering* **2**, 2-42.

[12] Gazatas G. (1991) Foundation vibration, Foundation Engineering Handbook Chapter 15, 2nd edition, Chapman and Hall, NewYork, U.S.

[13] Pais A., Kausel E. (1988) Approximate formulas for dynamic stiffness of rigid foundations, *Soil Dynamics and Earthquake Engineering* **7**, 213-227.

[14] Newmark N. M. (1959) A method of computation for structural dynamics, *Engineering Mechanics Division*, 67-94.

[15] Youssef M. A. H., Duhee P. (2001) Non-linear one-dimensional seismic ground motion propagation in the Mississippi embayment, *Engineering Geology* **62**, 185-206.

[16] Youssef M. A. H., Duhee P. (2002) Viscous damping formulation and high frequency motion propagation in non-linear site response analysis, *Soil Dynamic and Earthquake Engineering* **22**, 611-624.

[17] Camilo P., Youssef M. A. H. (2009) Damping formulation for nonlinear 1D site response analyses, S*oil Dynamics and Earthquake Engineering* **29**, 1143-1158.
[18] Chang N. Y., Hien M. N. (2010) Viscous damping for time domain finite element analysis, 5th International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics and Symposium in Honor of Professor I. M. Idriss, San Diego, California, U.S.
[19] Tatsuoka Fumio, Fukushima Shinji (1978) Stress-strain relation of sand for irregular cyclic excitation, *Research Bulletin*, 356-359.
[20] Rayleigh J. W. S., Lindsay R. B. (1945) The theory of sound vol. 2(1), New York, U.S.
[21] Ramberg W., Osgood W. R. (1943) Description of stress-strain curves by three parameters, Technical Note 902, National Advisory Committee on Aeronautics.
[22] International Building Code (2006), International Code Council, U.S.
[23] Kabeyasawa T., Matsumori T., Katsumata H., Shirai K. (2005) Design of the full-scale six story reinforced concrete wall-frame building for testing at E-defense, *Proceedings of the first NEES/E-Defense workshop on collapse simulation of reinforced concrete building structures*, Berkeley, California, U.S. 23-45.
[24] Kim Y., Kabeyasawa T., Matsumori T. (2007) Dynamic collapse analysis of the six-story full scale wall-frame tested at E-Defense, Proceeding 8th Pacific Conference on Earthquake Engineering, Singapore, Singapore.
[25] Seismic Response Analysis and Design of Buildings Considering Dynamic Soil-Structure Interaction, Japanese version

Projection-based particle methods - latest achievements and future perspectives

†Abbas Khayyer[1] and Hitoshi Gotoh[2]

[1,2]Department of Civil and Earth Resources Engineering, Kyoto University, Japan.

†Corresponding and presenting author: khayyer@particle.kuciv.kyoto-u.ac.jp

Abstract

The paper presents a concise review on the latest achievements made in the context of projection-based particle methods, including MPS and Incompressible SPH (ISPH) methods. The latest achievements corresponding to stability, accuracy, boundary conditions and energy conservation enhancements as well as advancements related to simulations of multiphase flows, fluid-structure interactions and surface tension are reviewed. The future perspectives for enhancement of applicability and reliability of projection-based particle methods are also highlighted.

Keywords: particle methods, projection method, Moving Particle Semi-implicit (MPS), Incompressible Smoothed Particle Hydrodynamics (ISPH), stability, accuracy, conservation

Introduction

Projection-based particle methods, including MPS [1] and Incompressible SPH (ISPH) [2] methods, are founded on Helmholtz decomposition of an intermediate velocity vector field into a solenoidal (divergence-free) one and an irrotational (curl-free) one. These methods potentially result in accurate solutions to the continuity and Navier-Stokes equations, especially in terms of pressure calculation and volume conservation. In particular, the prediction-correction feature of projection-based particle methods provides the opportunity for numerical error minimization through the application of, for instance, error mitigating functions in the source term of the Poisson Pressure Equation (PPE) [3,4]. This paper aims at illustrating a concise summary of the latest achievements made in the field of projection-based particle methods, as well as some future perspectives.

The latest achievements made in the field of projection-based particle methods correspond to enhancements of stability, accuracy, boundary conditions, energy conservation and enhanced simulations of multiphase flows, fluid-structure interactions, surface tension, etc. In this paper, these achievements will be concisely reviewed.

Latest Achievements

Stability enhancement: A distinct category of methods developed for enhancement of both stability and accuracy for both explicit and semi-implicit projection-based particle methods correspond to particle regularization schemes. For instance, Lind et al. [5] proposed a generalized **P**article **S**hifting (**PS**) technique on the basis of Fick's law of diffusion. Despite its simplicity and effectiveness, the particle shifting scheme may violate the overall conservation properties [5] including conservations of momentum and energy.

To ensure the stability of projection-based particle methods, Tsuruta et al. [6] presented a **D**ynamic **S**tabilization (**DS**) scheme which is aimed at producing exactly adequate repulsive

forces in a momentum-conservative manner. The applicability and effectiveness of this scheme has to be further examined for a wider range of free-surface, internal and multi-phase flows. Recently, the authors have conducted a study on accuracy and conservation properties of particle regularization schemes including PS and DS schemes. Despite providing exact local and thus global momentum conservation, the DS scheme may result in small-scale particle perturbations. This issue can be seen from a simple and well-known numerical benchmark test, namely, the Taylor-Green vortex.

Fig. 1 shows a qualitative comparison in between DS and PS schemes through illustrating calculated normalized pressure and velocity fields at normalized time of tU/L = 1.0 for Re = 10^6 in a Taylor-Green vortex test [5]. In the performed simulations of Taylor-Green vortex, particles are considered to be 5 mm in diameter (d_0 = 0.005 m) resulting in a total number of 40,000 particles. The calculation time step is set based on Courant stability condition and a maximum allowable time step of Δt_{max} = 5.0E-4 s. Without a proper particle regularization scheme, a purely Lagrangian simulation of Taylor-Green vortices will be most likely characterized by unfavourable anisotropic particle distributions along the flow streamlines. Here both DS and PS schemes have been successful in providing stable calculations. Nevertheless, distribution of particles by PS appears to be more regular in comparison to that by DS. As previously stated, at least for this test, the DS scheme has apparently resulted in small-scale particle perturbations. This would indicate the need to revisit the derivation of this scheme and possibly present an enhanced version.

As for the PS scheme, a distinct issue arises for free-surface or multiphase flows. In other words, special care must be taken with application of this scheme to interface particles due to large concentration gradients. Lind et al. [5] proposed a special treatment (Eq. 27 of [5]) for free-surface and its nearby particles to eliminate shifting normal to the free-surface. Theoretically, this treatment is justified for proper implementation of PS to free-surface flows. However, several numerical challenges arise, especially in long term simulations, resulting in unphysical perturbations and/or accumulation of particles at free-surface (e.g. Fig. 17 [5]). In addition, in order to minimize the unphysical perturbations at free-surface, the PS scheme of Lind et al. [5] for free-surface contains two tuning parameters to allow slight diffusion normal to the interface. Recently, the authors proposed a new **OPS** (**O**ptimized **P**article **S**hifting) scheme to enhance the accuracy of PS at the phase interfaces (e.g. free-surface). Unlike PS, the OPS does not contain any tuning parameters. **Fig. 2** illustrates the improved performance of OPS with respect to PS in simulation of a square patch of fluid [7]. **Fig 2**(**c**) and (**d**) present the time histories of mechanical energy dissipation and calculated pressure at the center of the patch. In our performed simulations, the square's length, L, and angular velocity, ω, are considered as 1.0 m and 1.0 m/s, respectively. Particles are considered to be 0.01 m in diameter (d_0 = 0.01 m).

Accuracy enhancement: For both ISPH and MPS methods refined differential operator models have been proposed to enhance the accuracy of pressure calculation [3,8,9,10,11] and particle motion [3,11]. Refined differential operator models have been proposed for discretization of either source term [8,10] or Laplacian of pressure [9,12,13] in the PPE.

Inspired by the excellent work of Kondo and Koshizuka [10], Khayyer and Gotoh [3] proposed a so-called **ECS** (**E**rror **C**ompensating **S**ource) scheme to minimize the projection-related errors. The PPE incorporating the ECS is formulated as [3]:

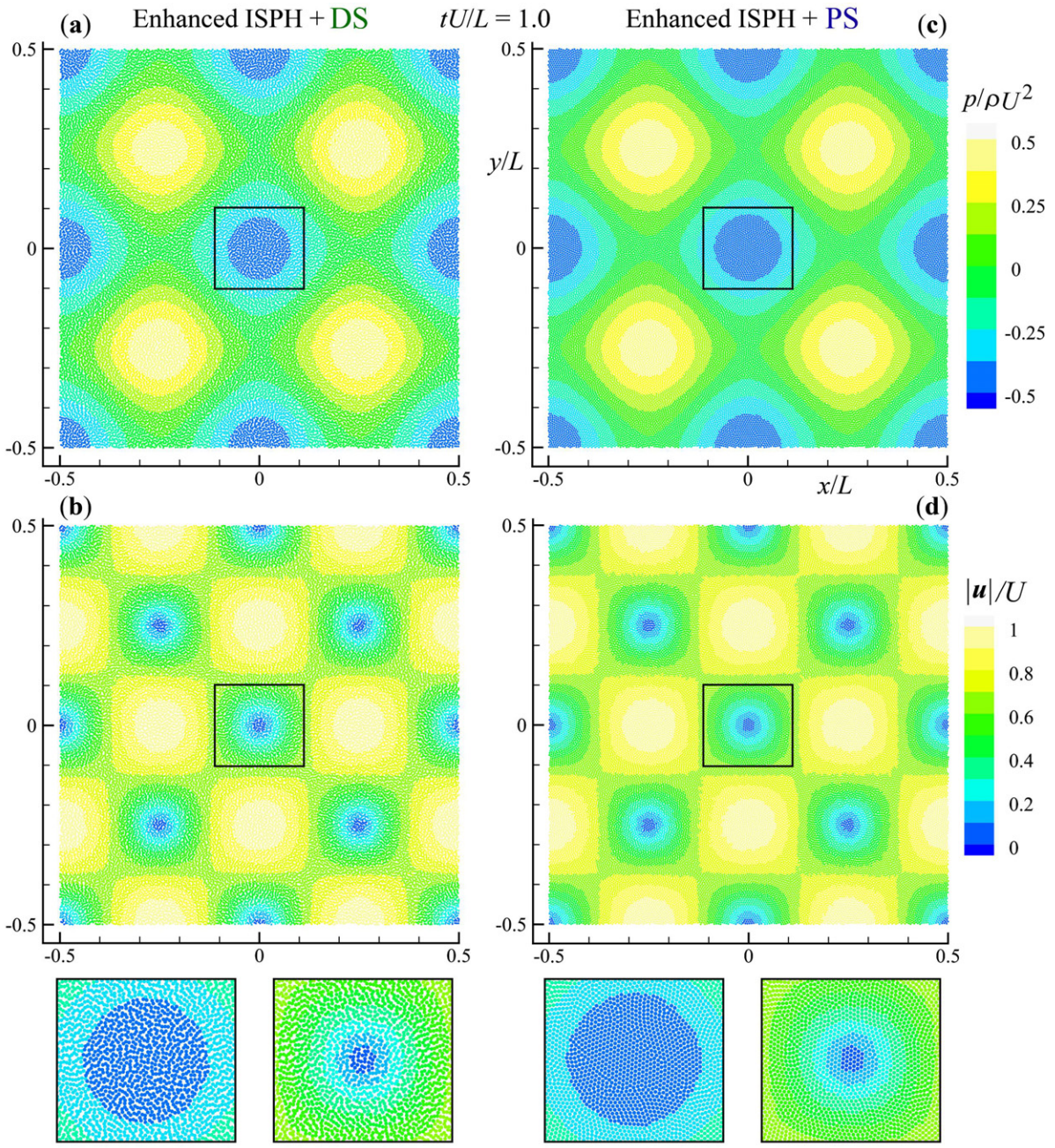


Fig. 1. Calculated normalized pressure (a,c) and velocity (b,d) by Enhanced ISPH + DS [6] (a,b) and Enhanced ISPH + PS [5] (c,d) - Taylor-Green flow (Re =10^6)

$$\left\langle \nabla^2 p_{k+1} \right\rangle_i = \frac{\rho}{n_0 \Delta t}\left(\frac{Dn}{Dt}\right)_i^* + \Lambda_{ECS}$$

$$\Lambda_{ECS} = \frac{\rho}{\Delta t}\left\{\frac{\alpha}{n_0}\left(\frac{Dn}{Dt}\right)_i^k + \frac{\beta}{\Delta t}\frac{n_i^k - n_0}{n_0}\right\} \quad ; \quad \alpha = \left|\frac{n_i^k - n_0}{n_0}\right| \quad ; \quad \beta = \left|\frac{\Delta t}{n_0}\left(\frac{Dn}{Dt}\right)_i^k\right| \tag{1}$$

where p, ρ, n, n_0, t, Δt, i and k represent pressure, density, particle number density, reference particle number density, time, calculation time step, target particle i and calculation step number, respectively. Hence, the source term of PPE is comprised of a main term and two error mitigating terms multiplied by dynamic coefficients (α, β) as functions of instantaneous

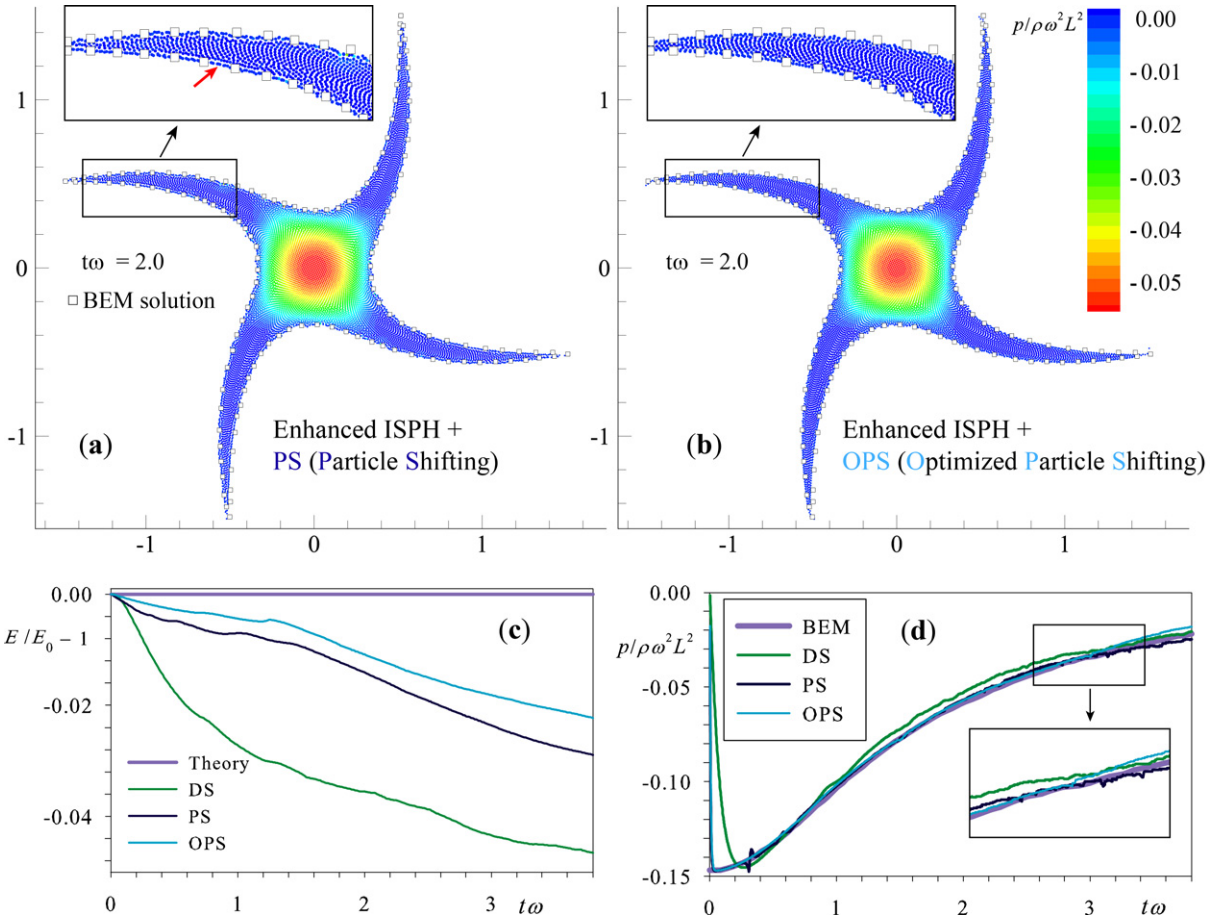


Fig. 2. Qualitative comparison in between PS [5] (a) and newly proposed OPS (b) schemes, elimination of unphysical discontinuity at free-surface (a) by OPS (b) - time histories of energy and normalized pressure at the center of the patch - evolution of a square patch of fluid [7]

flow field. The dynamic coefficients adjust the intensities of error mitigating terms depending on the instantaneous state of flow field. Similar ECS scheme has been formulated and validated for the ISPH [11].

Once an accurate pressure field is obtained, particles should be moved in space according to accurately computed accelerations corresponding to pressure gradient. In this regard, enhanced pressure gradient models with consistency-related corrections (e.g. [3,4,11,14,15,16]) have been proposed.

Improvement of boundary conditions: These improvements correspond to wall, free-surface and inflow/outflow boundary conditions.

Adami et al. [17] proposed a generalized wall boundary condition for SPH which correctly imposes no-slip conditions even for complex geometries. Despite being relatively simple for implementation, application of mirror particles may lead to inaccuracies in the convergence of differential operator models [18]. A more favored and recent approach is related to development of so-called semi-analytical wall boundary conditions. Di Monaco et al. [19] developed a semi-analytic approach for treatment of wall boundaries that can be considered as an integral version of the mirror particles of Adami et al. [17] for fixed boundaries. Similar approaches have been proposed by Ferrand et al. [20] and Mayrhofer et al. [21] that provide accurate and direct modeling of boundary integrals at the frontiers of the fluid domain

resulting in precise pressure forces, wall friction and turbulent conditions. Recently, Leroy et al. [22] extended the unified semi-analytical wall boundary condition of Ferrand et al. [20] for the projection-based particle methods, and more precisely, the ISPH method.

In projection-based particle methods, a challenging issue is to detect free-surface particles accurately to impose the dynamic free-surface boundary condition, i.e. p equal to zero, to them. Khayyer et al. [23] proposed an auxiliary condition based on the non-symmetric distribution of free-surface particles to be used together with the original simple criterion. Ma and Zhou [24] proposed a Mixed Particle Number Density and Auxiliary Function Method (MPAM) for identifying the free surface particles in their Meshless Local Petrov-Galerin method based on Rankine source solution (MLPG-R) method. Park et al. [25] used a so-called Arc Method for an accurate assessment of free-surface particles. Nair and Tomar [26] presented a semi-analytical approach to impose Dirichlet boundary conditions on the free surface and thus eliminating the need for free-surface particle detection. This necessity was also eliminated by proposal of a new free-surface boundary condition referred to as Space Potential Particles (SPP [27]), through introduction of a potential in void space.

There have been a number of researches specifically targeting inlet/outlet boundary conditions in both weakly compressible (e.g. [28]) and incompressible (e.g. [29]) frameworks. In order to enhance the ISPH solution for both pressure and velocity near the boundaries including inlet/outlet ones, Hosseini and Feng [30] presented an approach which utilizes a rotational pressure-correction scheme with a consistent pressure boundary condition.

Energy conservation: Violeau [31] highlighted the compatibility, and more precisely, the skew-adjointness of gradient and divergence operators for energy conservation in calculations by particle methods. In the context of projection-based particle methods, this important property is required for an exact projection [32] which is a necessity for an exact energy conservation. A clear link exists also in between energy conservation and consistency of differential operator models and specifically, pressure gradient model.

Khayyer et al. [33] performed a study on energy conservation properties of projection-based particle methods. Their study highlighted the significance of Taylor-series consistent pressure gradient models and enhancing effect of a consistency-related gradient correction in providing enhanced energy conservation. Both ISPH and MPS were found to provide accurate predictions of physical dissipations in fluid impact problems. **Fig. 3** depicts improved MPS results corresponding to a normal impact of two rectangular fluid patches [34]. The rectangular patches have a length L, width $2H$ and the impact occurs at $t = 0$. The fluid is considered to be inviscid and incompressible, and thus the impact will be associated with a theoretically sudden loss of a fraction of the initial energy [35]. For the performed simulations L = 1.0 m, H = 0.33 m and U = 3.4 m/s. The maximum allowable time step is set as Δt_{max} = 5.0E-5 s and the particles are set to be of 0.01 m in diameter, i.e. d_0 = 0.01 m. A set of typical snapshots illustrating this phenomenon are presented in **Fig**. **3**(**a**-**c**). From **Fig**. **3**(**d**), the improved MPS method has provided an accurate estimation of energy loss corresponding to this impact.

To further illustrate the performance of improved MPS in reproduction of physical dissipation the normal impact of two rectangular fluid patches with different masses is considered. An analytical expression for the energy loss during this specific impact is given by Rogers and Szymczak [36]. A set of snapshots corresponding to this interesting classical fluid mechanics problem are presented in **Fig 4**. The performed simulation is characterized by a Mach number

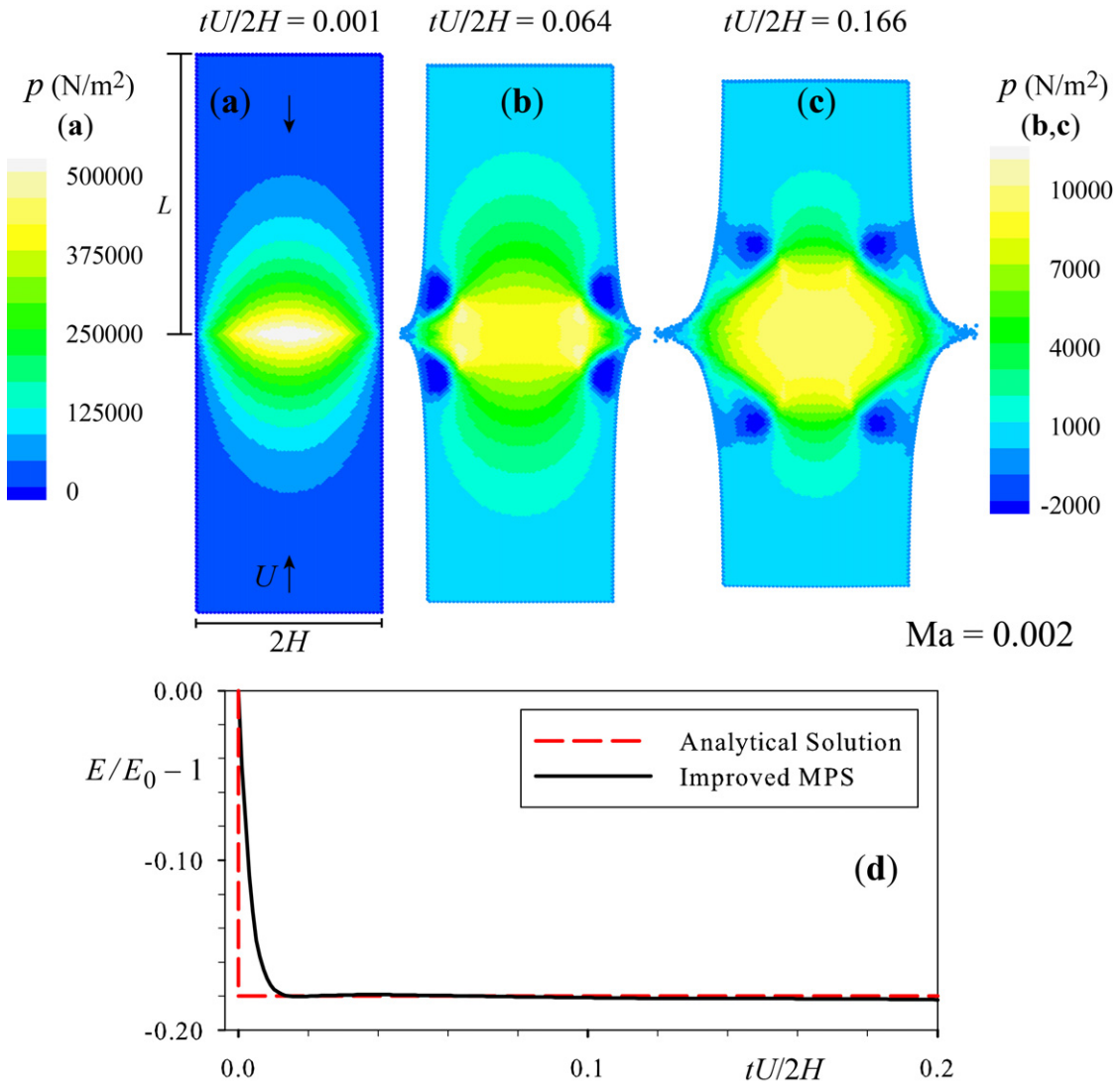


Fig. 3. Snapshots of particles together with pressure field (a-c), analytical [35] and calculated energy loss (d) - results by improved MPS - normal impact of two identical fluid patches [34,37]

of Ma = 0.2. For this simulation, the maximum allowable time step, Δt_{max}, is set as 5.0E-7 s, and particles are set to be of 0.01 m in diameter, i.e. d_0 = 0.01 m. **Fig. 4(e)** shows the excellent performance of improved MPS in providing almost accurate prediction of the energy loss for this impact.

The superior performance of improved MPS in predictions of energy loss in fluid impact problems as well as its excellent capability in shock capturing and propagation can be further pronounced by comparing the achieved results with those of advanced particle methods, including δ-SPH (e.g. Figs 14 and 15 in [37]) and Riemann SPH (e.g. Figs 9 and 10 in [34]). It should be noted in both of the mentioned references [34,37] weakly compressible SPH formulations are adopted.

Enhanced simulations of multiphase flows: Khayyer and Gotoh [4] presented an improved MPS method for multiphase flows characterized by large density ratios. The stability of their calculations was guaranteed through the application of a Taylor-series-based density smoothing scheme, and accuracy enhancement was achieved through the application of a PPE's error mitigating term, i.e. ECS scheme, and refined discretizations of source term and

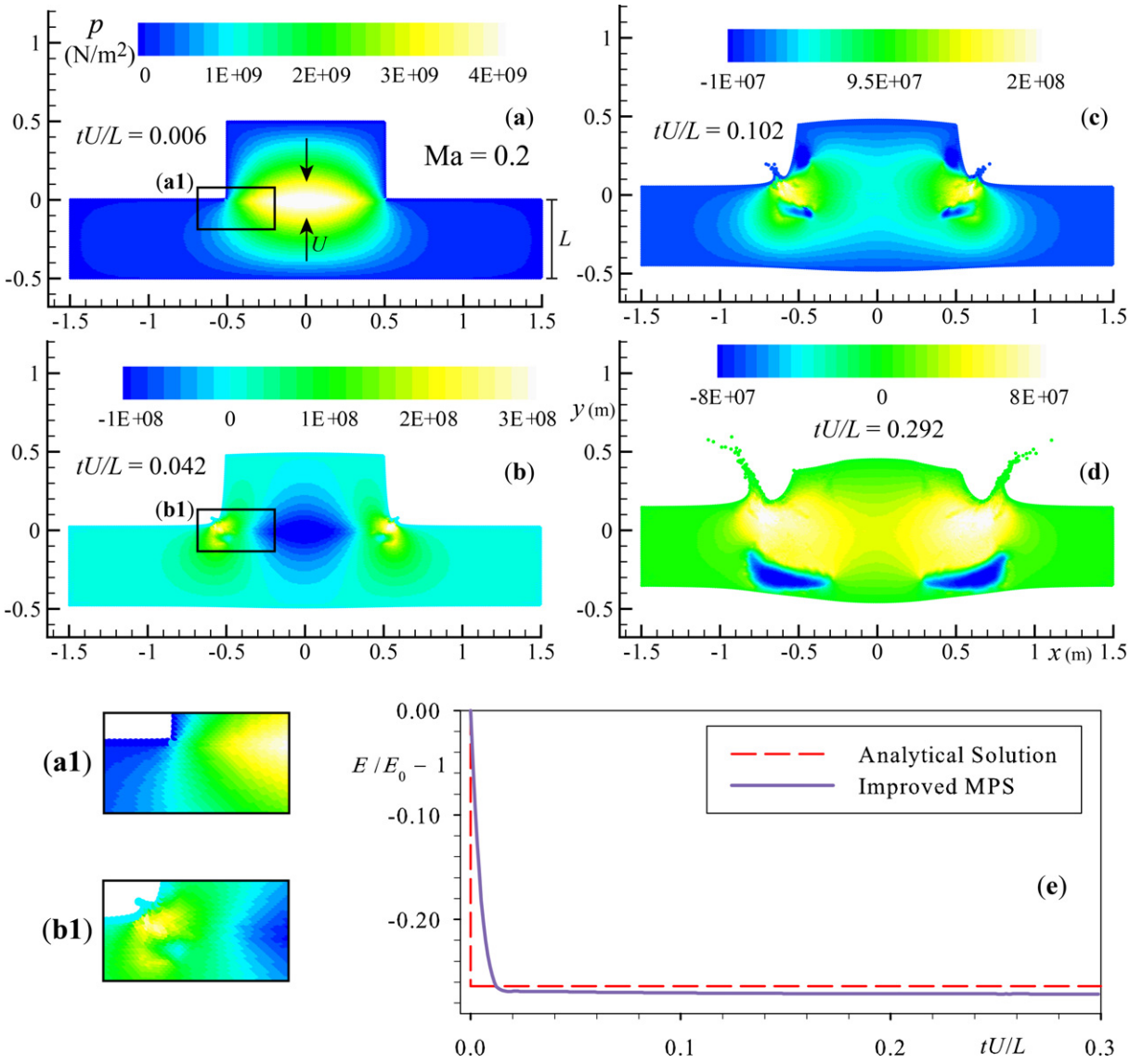


Fig. 4. Snapshots of particles together with pressure field (a-d), analytical [36] and calculated energy loss - results by improved MPS - normal impact of two fluid patches with different masses [34]

Laplacian of pressure. **Fig. 5** presents two typical snapshots corresponding to a multiphase violent sloshing flow characterized by air entrainment/entrapment with a realistic air/water density ratio of 1:1000. Conditions of the performed sloshing simulation corresponded to the experiment by Rognebakke et al. [38]. Sinusoidal excitations with maximum amplitude of 150 mm and frequency of 1.2 Hz were considered. The particles were 5.0 mm in diameter and the calculation time step was set according to the Courant stability condition and a maximum allowable time increment of 4.0E-5 s.

The ECS scheme was extended to minimize the projection-related errors in an incompressible-compressible multiphase calculation of wave slamming where actual speeds of sounds in air and water were implemented [39]. The newly proposed scheme was referred to as **CIECS** (**C**ompressible-**I**ncompressible **ECS**). The effectiveness of CIECS in minimization of projection-related errors in a typical Compressible-Incompressible multiphase flow, namely, slamming with entrapped air was shown through two sets of simulations corresponding to experiments by Lin and Shieh [40] and Verhagen [41].

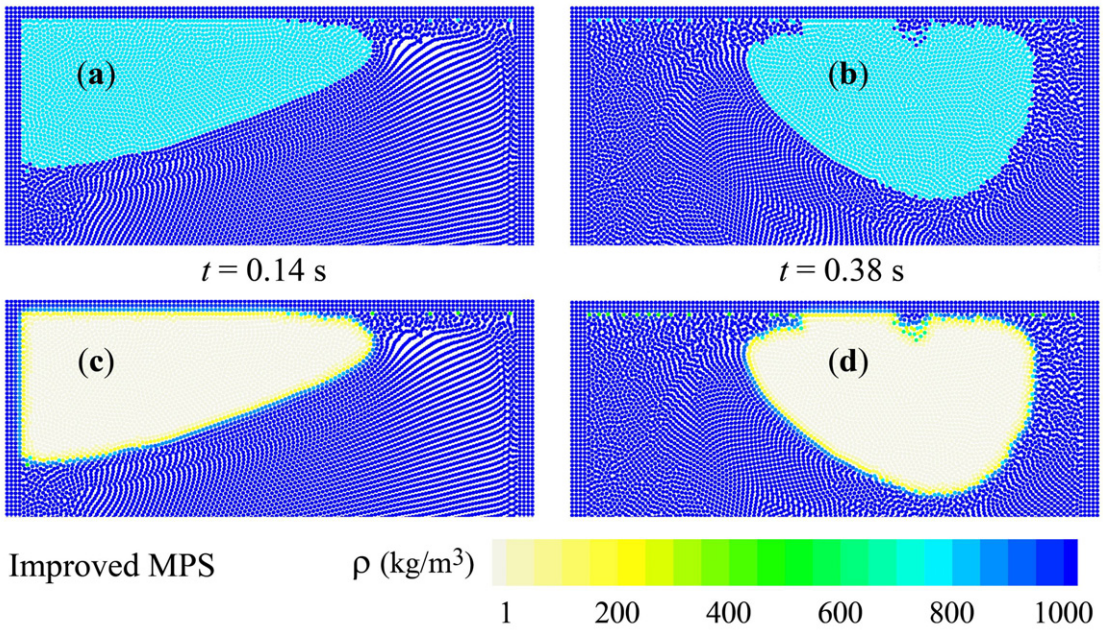


Fig. 5. Snapshots of gas and liquid particles (a,b) and calculated density fields (c,d) - muliphase simulation of a violent sloshing flow [38] by an improved MPS method [4]

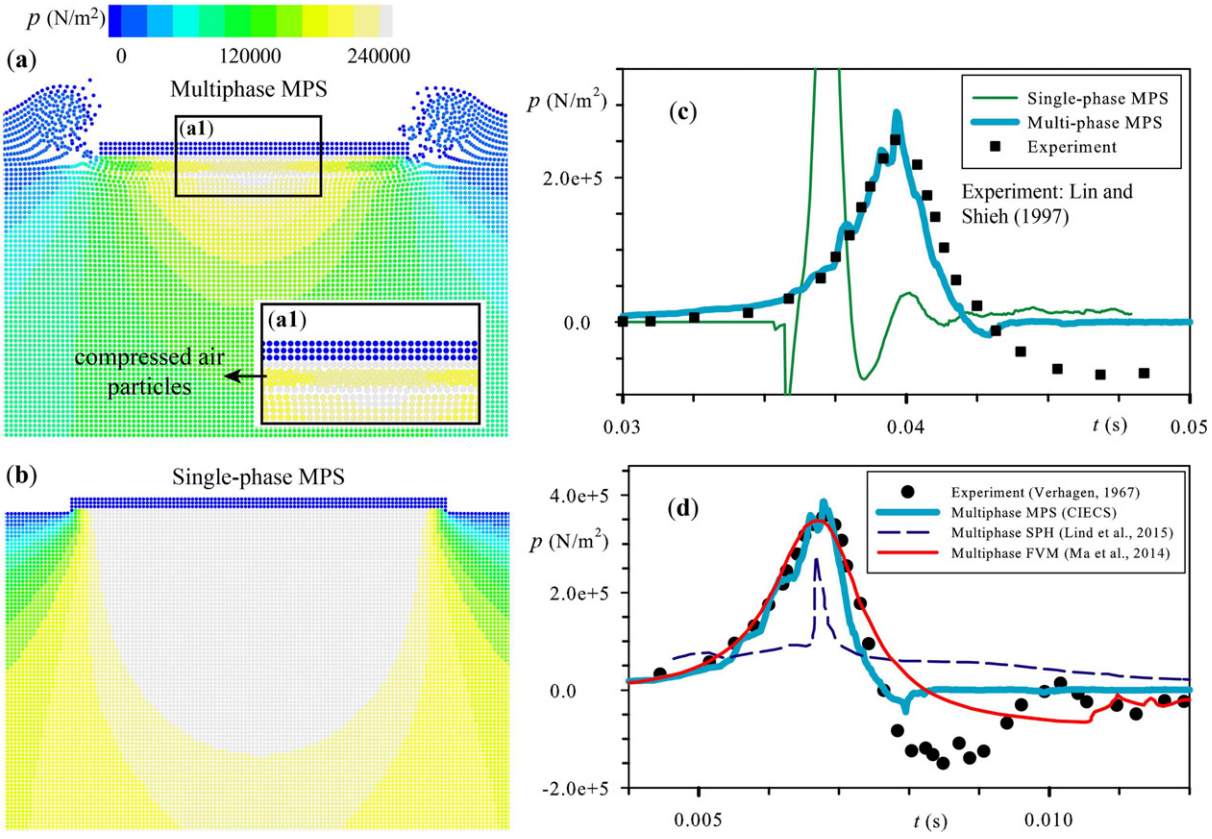


Fig. 6. Multiphase MPS with CIECS scheme applied to water slamming, experiments by Lin and Shieh [40] (a,c) and Verhagen [41] (d) - importance of air cushioning effect in prediction of slam induced pressure (c) and comparisons of multiphase MPS with multiphase SPH [42] and FVM [43] (d)

Fig. **6**(**a**-**c**) depicts the water slamming simulation results related to the experiment by Lin and Shieh [40] by multiphase and single-phase MPS methods. The figure portrays the importance of consideration of air and its cushioning effect for prediction of slamming-induced pressures.

Fig. **6**(**d**) presents a comparison in between the multiphase MPS with CIECS scheme with results by Lind et al. [42] and Ma et al. [43] with respect to the experiment by Verhagen [41]. A common experiment-simulation inconsistency seen in this figure corresponds to inaccurate prediction of post-impact negative pressure. The authors are investigating the probable reasons behind this apparent inconsistency. In the performed water slamming simulations, the diameter of particles was set as 3 mm. Considered viscosities for the water and air phases corresponded to their physical ones, i.e. ν_w = 1.0E-6 m^2/s and ν_a = 1.5E-5 m^2/s. The calculation time step was set based on the Courant stability condition and Δt_{max} = 1.0E-4 s.

Fluid-structure interactions: Particle methods including projection-based ones appear to be suitable computational tools for **FSI** (**F**luid-**S**tructure **I**nteraction) simulations, mainly due to their Lagrangian feature. These methods have been applied to simulate interactions in between fluid flows with either rigid (e.g. [44]) or flexible (e.g. [45]) structures. In the latter case, a proper structural model should be carefully coupled with the fluid solver.

In the context of projection-based particle methods, Lee et al. [46] developed a MPS-FEM coupled method to study incompressible fluid flow interactions with elastic structures. Rafiee and Thiagarajan [45] proposed a fully-Lagrangian SPH-based solver for simulation of incompressible fluid-hypoelastic structure interactions. In their study, the PPE was solved simply using an approximate explicit scheme. Hwang et al. [47] developed a fully-Lagrangian MPS-based FSI analysis method for incompressible fluid-linear elastic structure interactions. The key feature of this solver was absence of any artificial numerical stabilizers commonly applied in particle-based FSI solvers. This feature was achieved by implementation of an appropriate coupling algorithm.

Khayyer et al. [48] presented an enhanced version of Hwang et al.'s method by incorporating several refined schemes for the fluid phase and presenting an improved calculation of fluid force to structure. The achieved enhancements as well as applicability of developed MPS-based FSI solver are portrayed in **Fig**. **7**, corresponding to simulations of an entry of a deformable aluminum beam into an undisturbed water [49] and a dam break flow impacting on an elastic plate [50]. **Fig**. **7**(**a**) presents a representative snapshot of the pressure and stress fields in fluid and beam. A schematic sketch of this beam entry test and time histories of deflection at point C is shown in **Fig. 7**(**b**), where improved results are obtained by the enhanced coupled MPS [48]. For this aluminum beam entry test, the analytical solutions were derived by Scolan [51], on the basis of the hydrodynamic Wagner's model and linear Wan's theory. The material properties of the aluminum beam, namely, its Young's modulus, Poisson ratio and density were considered as 67.5 GPa, 0.34 and 2700 kg/m^3, respectively. Both structural and fluid particles were 0.01 m in size. **Fig**. **7**(**c**) and (**d**) portray two typical snapshots by coupled MPS [47] and enhanced coupled MPS [48] solvers together with their corresponding experimental photo as well as the result by a FDM-FEM solver [50] for the second FSI test. The superior performance of enhanced MPS is clearly illustrated in this figure as this method provides more consistent deflections of the elastic plate.

Surface tension: Surface tension modeling in the context of particle methods have been performed using either potential approach or continuum one. In the so-called potential approach surface tension is modeled by assuming that microscopic cohesive intermolecular forces can be mimicked by macroscopic inter-particle forces. The main advantage of this approach is related to its computational simplicity in that surface tension is modeled via particle-particle interactions explicitly without the necessity of calculating surface normals and curvatures, as required in the continuum approach. The main disadvantage of potential

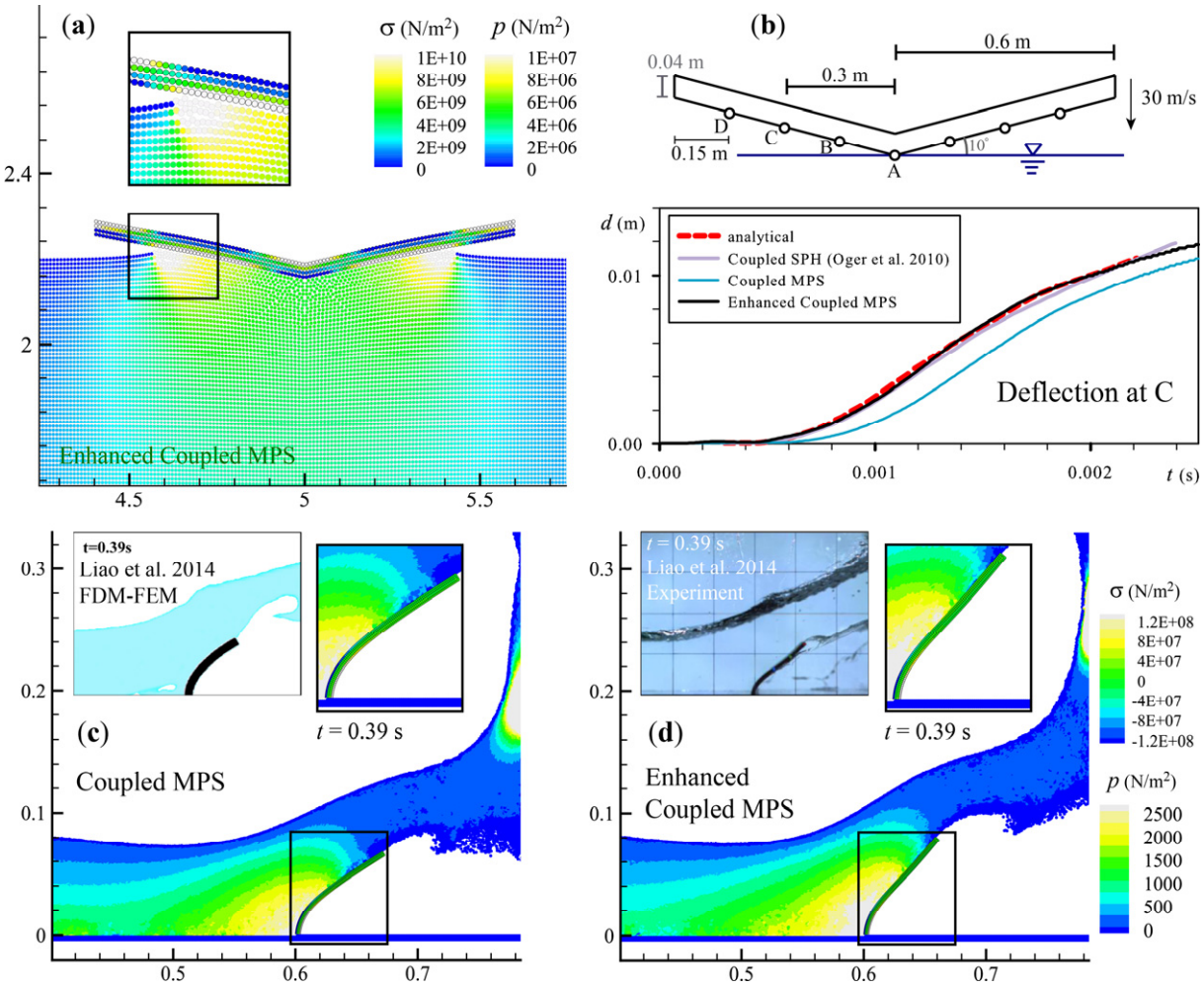


Fig. 7. Entry of an aluminum beam into undisturbed water [49] (a,b) and dam break with elastic plate [50] (c,d), results by an enhanced coupled MPS solver [48] (a,d) and a coupled MPS solver [47] (c)

approach corresponds to the fact that the surface tension forces depend on the intensity of particle-particle interactions. These interactions have to be adjusted numerically by varying the macroscopic input parameters depending on the simulation case to reproduce desired surface tension forces.

The most common approach for incorporation of surface tension in macroscopic particle-based simulations is the continuum approach and specifically those based on the **C**ontinuum **S**urface **F**orce (**CSF**) model introduced by Brackbill et al. [52]. In this approach, the surface tension is treated as a continuous, three-dimensional effect across the interface, derived directly from the Young-Laplace equation. Morris [53] showed several possible implementations of CSF model in SPH and highlighted the challenges in accurate calculations of interface curvature. These challenges are not only limited to difficulties in accurate particle-based calculation of Laplacian of color function for approximation of interface curvature, but also to the fact that a smoothed color function is usually used. The use of a smoothed color function may become problematic for approximation of interface normals near the boundaries and sharp-angled areas.

In MPS-based simulations of surface tension, the CSF based simulations can be categorized into two distinct groups, depending on the computational procedure for calculation of the curvature and the normal vector. These two categories are: arc fitting at interface [54] and

differential approach (e.g. [55]). As the name indicates the arc fitting approach is aimed at approximating the normal vector and curvature by constructing local arcs at the surface particles via specific computational procedures. The accuracy of arc fitting approach is highly dependent upon the instantaneous smoothness of the free-surface. In the differential approach, the continuum surface forces are calculated by applying differential operator models for both gradient and Laplacian so that potentially accurate approximations of the unit normal vector and the curvature can be obtained.

Khayyer et al. [56] proposed a new differential CSF-based model in the context of MPS. Their model benefits from a novel formulation for curvature estimation using direct second order derivatives of color function via a precise discretization. By applying a high-order Laplacian scheme [9] including the approximation of boundary integrals, relatively accurate approximation of interface curvature and thus surface tension could be achieved. Accordingly, the Laplacian of color function, C, at an interface target particle i was calculated as [56]:

$$\left(\nabla^2 C\right)_i = \frac{1}{n_0}\sum_{i\neq j}\left\{\frac{\partial C_{ij}}{\partial r_{ij}}\frac{\partial w_{ij}}{\partial r_{ij}} + C_{ij}\left(\frac{\partial^2 w_{ij}}{\partial r_{ij}^2} + \frac{D_s - 1}{r_{ij}}\frac{\partial w_{ij}}{\partial r_{ij}}\right)\right\} + BI \tag{2}$$

where $C_{ij} = C_j - C_i$, $r_{ij} = r_j - r_i$, w represents kernel function, D_s stands for number of space dimensions and BI denotes the boundary integrals [57] formulated as:

$$BI = \int_{\partial\Omega} \nabla C \cdot \boldsymbol{n}\, w_{ij}\, dS \approx \frac{1}{n_0}\sum_{j\in\partial\Omega} \frac{C_{ij}\,\boldsymbol{r}_{ij}\cdot\boldsymbol{n}_j}{\left|\boldsymbol{r}_{ij}\right|^2} w_{ij}\, S_j \tag{3}$$

where $\boldsymbol{n}$ denotes interface normal, $\boldsymbol{r}$ symbolizes position vector and for 2D simulations S_j signifies the length (diameter) of boundary particle j. Therefore, the surface tension force is evaluated via achieving a direct Laplacian-based approximation of curvature. The enhanced performance of the Laplacian-based surface tension model [56] with respect to the arc fitting one [54] is illustrated in **Fig. 8**, corresponding to simulations of a water drop impact [58] for Froude and Weber numbers of 639 and 395, respectively. The figure portrays the superior performance of Laplacian-based surface tension model in better reproduction of crown development and splash drops.

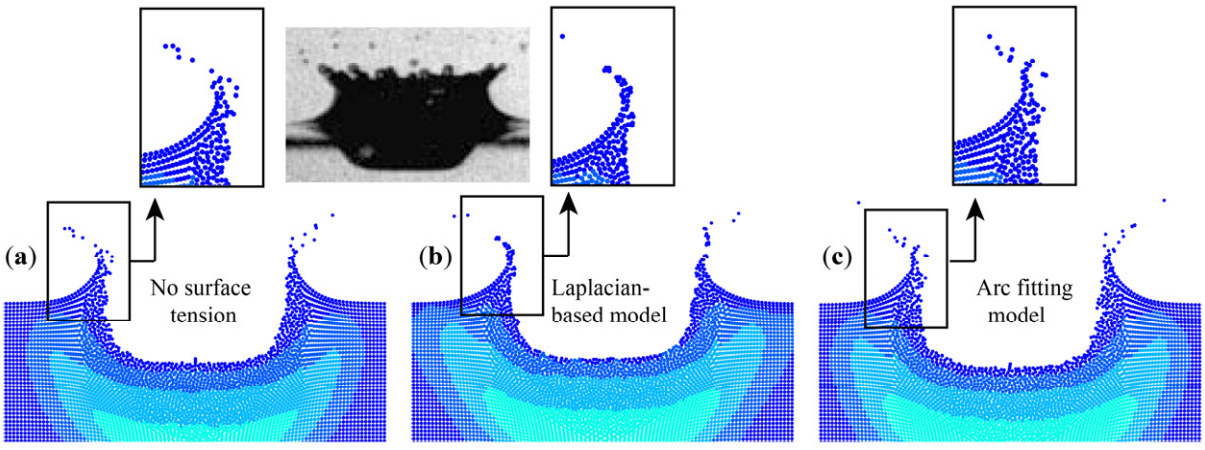


Fig. 8. Improved MPS results of a water drop impact [58], no surface tension model (a), Laplacian-based surface tension model [56] (b) and arc fitting surface tension model [54] (c)

Future Perspectives

In spite of the achieved advancements, rigorous researches should continue to be conducted to further enhance the reliability and accuracy of particle methods for practical engineering and scientific purposes. In particular, important issues of stability, conservation, convergence, boundary conditions, turbulence modeling [59,60], multi-scale and multi-physics simulations [61] will be among the future perspectives corresponding to projection-based particle methods.

For extended engineering and industrial applications, it is important to keep the developed computational methods free of any numerical term with constants that may require calibration. Several key insights on extended engineering and industrial applications of particles methods are highlighted in excellent review papers by Koshizuka [62] and by Violeau and Rogers [63]. Indeed, prior to any practical application, precise verification of particle-based codes must be conducted by consideration of appropriate benchmark tests with analytical solutions in terms of reproduced velocity and pressure together with comprehensive investigations on conservation and convergence properties.

Further advanced multi-scale and multi-physics applications of particle methods are expected to be achieved with forthcoming theoretical and computational enhancements. In particular, rigorous enhancements of stability, accuracy and conservation properties of particle methods along with advancements made in high performance computing as well as developments of accurate variable resolution schemes [64] will enable particle methods, including projection-based ones, to serve as advanced, reliable and efficient computational methods.

References

[1] Koshizuka, S. and Oka, Y. (1996), Moving particle semi-implicit method for fragmentation of incompressible fluid, Nuclear Science and Engineering, **123**, 421-434.

[2] Shao, S. and Lo, E.Y.M. (2003), Incompressible SPH method for simulating Newtonian and non-Newtonian flows with a free surface, Adv. Water Resour., **26**, 787-800.

[3] Khayyer, A. and Gotoh, H. (2011), Enhancement of Stability and Accuracy of the Moving Particle Semi-implicit Method, J. Comp. Phys., **230**, 3093-3118.

[4] Khayyer, A. and Gotoh, H. (2013), Enhancement of performance and stability of MPS meshfree particle method for multiphase flows characterized by high density ratios, J. Comp. Phys., **242**, 211-233.

[5] Lind, S.J., Xu, R., Stansby, P.K. and Rogers, B.D. (2012), Incompressible smoothed particle hydrodynamics for free-surface flows: A generalised diffusion-based algorithm for stability and validations for impulsive flows and propagating waves. J. Comp. Phys., **231**, 1499-1523.

[6] Tsuruta, N., Khayyer, A. and Gotoh, H. (2013), A Short Note on Dynamic Stabilization of Moving Particle Semi-implicit Method, Computers & Fluids, **82**, 158-164.

[7] Colagrossi A. (2003), A meshless Lagrangian Method for Free-Surface and Interface Flows with Fragmentation, PhD Thesis, Universita di Roma, La Sapienza.

[8] Khayyer, A. and Gotoh, H. (2009), Modified Moving Particle Semi-implicit methods for the prediction of 2D wave impact pressure, Coastal Engineering, **56**, 419-440.

[9] Khayyer, A. and Gotoh, H. (2010), A Higher Order Laplacian Model for Enhancement and Stabilization of Pressure Calculation by the MPS Method, Applied Ocean Res., **32**(1), 124-131.

[10] Kondo, M. and Koshizuka, S. (2011), Improvement of Stability in Moving Particle Semi-implicit method, International Journal for Numerical Methods in Fluids, **65**, 638-654.

[11] Gotoh, H., Khayyer, A., Ikari, H., Arikawa, T. and Shimosako K. (2014), On enhancement of Incompressible SPH method for simulation of violent sloshing flows, Applied Ocean Research. **46**, 104-115.

[12] Tamai, T. and Koshizuka, S. (2014), Least squares moving particle semi-implicit method, Computational Particle Mechanics, **1**(3), 277-305.

[13] Tamai, T, Murotani, K. and Koshizuka, S. (2016), On the consistency and convergence of particle-based meshfree discretization schemes for the Laplace operator, Computers and Fluids, in press, http://dx.doi.org/10.1016/j.compfluid.2016.02.012.

[14] Li, S. and Liu, W.K. (2004), Meshfree Particle Methods, Berlin: Springer Verlag. ISBN 3-540-22256-1.

[15] Liu, M.B. and Liu, G.R. (2010), Smoothed Particle Hydrodynamics (SPH): an Overview and Recent Developments, Archives of Computational Methods in Engineering, **17**(1), 25-76.

[16] Oger, G., Doring, M., Alessandrini, B., Ferrant, P. (2007), An improved SPH method: towards higher order convergence, Journal of Computational Physics, **225**(2), 1472-1492.

[17] Adami, S., Hu X.Y. and Adams, N.A. (2012), A generalized wall boundary condition for smoothed particle hydrodynamics, Journal of Computational Physics, **231**(21), 7057-7075.

[18] Macià, F., Antuono, M., Gonzales, L.M., and Colagrossi, A. (2011), Theoretical analysis of the no-slip boundary condition enforcement in SPH methods", Prog. Theor. Phys., **125**(6), 1091-1121.

[19] Di Monaco, A., Manenti, S., Gallati, M., Sibilla, S., Agante G., and Guandalini, R. (2011), SPH modeling of solid boundaries through a semi-analytic approach, Engineering Applications of Computational Fluid Mechanics, **5**(1), 1-15.

[20] Ferrand, M., Laurence, D.R., Rogers, B.D., Violeau, D. and Kassiotis, C. (2013), Unified semi analytical wall boundary conditions for inviscid, laminar or turbulent flows in the meshless SPH method, International Journal for Numerical Methods in Fluids, **71**(4), 446-472.

[21] Mayrhofer, A., Rogers, B.D., Violeau D. and Ferrand, M. (2013), Investigation of wall bounded flows using SPH and the unified semi-analytical wall boundary conditions, Computer Physics Communications **184**(11), 2515-2527.

[22] Leroy, A., Violeau, D., Ferrand, M., Kassiotis, C. (2014), Unified semi-analytical wall boundary conditions applied to 2-D incompressible SPH, Journal of Computational Physics, **261**, 106-129.

[23] Khayyer, A., Gotoh H. and Shao, S.D. (2009), Enhanced predictions of wave impact pressure by improved incompressible SPH methods, Applied Ocean Research, **31**(2), 111-131.

[24] Ma, Q.W. and Zhou, J.T. (2009), MLPG_R Method for Numerical Simulation of 2-D Breaking Waves, Comp Modeling in Eng and Sci, **43**(3), 277-303.

[25] Park, J.I., Park, J.C., Hwang S.C. and Heo, J.K. (2014), Two-Dimensional Particle Simulation for Behaviours of Floating Body near Quaywall during Tsunami, Journal of Ocean Engineering and Technology, **28**(1), 12-19.

[26] Nair, P. and Tomar, G. (2014), An improved free surface modeling for incompressible SPH, Computers & Fluids, **102**, 304-314.

[27] Tsuruta, N., Khayyer, A. and Gotoh, H. (2015), Space potential particles to enhance the stability of projection-based particle methods, International Journal of Computational Fluid Dynamics. **29**, 100-119.

[28] Lastiwka, M., Basa M. and Quinlan, N.J. (2009), Permeable and non-reflecting boundary conditions in SPH, International Journal for Numerical Methods in Fluids, **61**, 709-724.

[29] Khorasanizade, S. and Sousa, J.M.M. (2016), An innovative open boundary treatment for incompressible SPH, International Journal for Numerical Methods in Fluids, **80**, 161-180, 2016.

[30] Hosseini, S.M. and Feng, J.J. (2011), Pressure boundary conditions for computing incompressible flows with SPH", Journal of Computational Physics, **230**, 7473-7487.

[31] Violeau, D. (2012), Fluid Mechanics and the SPH Method, Theory and Applications, Oxford University press, ISBN: 978-0-19-965552-6.

[32] Cummins, S.J. and Rudman, M. (1999), An SPH projection method, Journal of Computational Physics, **152**, 584-607.

[33] Khayyer, A., Gotoh, H., Shimizu, Y. and Gotoh, K. (2015), On Enhancement of Energy Conservation Properties of ISPH and MPS Methods, Proceedings of 10th international SPHERIC workshop, Parma, Italy, 139-146.

[34] Marrone, S., Colagrossi, A., Di Mascio, A. and Le Touzé, D. (2015), Prediction of energy losses in water impacts using incompressible and weakly compressible models, Journal of Fluids and Structures, **54**, 802-822.

[35] Szymczak, W., (1994), Energy losses in non-classical free surface flows, in Bubble Dynamics and Interface Phenomena, ser. Fluid Mechanics and Its Applications, J. Blake, J. Boulton-Stone, and N. Thomas, Eds. Springer Netherlands, **23**, 413-420.

[36] Rogers, J. and Szymczak, W. (1997), Computations of violent surface motions: comparisons with theory and experiment, Philosophical Transactions of the Royal Society of London A355, 649-663.

[37] Antuono, M., Marrone, S., Colagrossi, A. and Bouscasse, B. (2015), Energy balance in the δ-SPH scheme, Computer Methods in Applied Mechanics and Engineering, **289**, 209-226.

[38] Rognebakke, O.F., Hoff, J.R., Allers, J.M., Berget, K., Bergo, B.O., and Zhao, R. (2006), Experimental approaches for determining sloshing loads in LNG tanks, Trans. Soc. Naval Archit. Mar. Eng., **113**, 384-401.

[39] Khayyer, A. and Gotoh, H. (2016), A multiphase compressible-incompressible particle method for water slamming, International Journal of Offshore and Polar Engineering, **26**(1), 20-25.

[40] Lin, M.C. and Shieh, L.D. (1997), Simultaneous measurements of water impact on a two-dimensional body, Fluid Dynamics Research, **19**, 125-148.

[41] Verhagen, J.H.G. (1967), The impact of a flat plate on a water surface, Journal of Ship Research, **11**(4), 211-223, 1967.

[42] Lind, S.J., Stansby, P.K., Rogers, B.D. and Lloyd, P.M. (2015), Numerical predictions of water-air wave slam using incompressible-compressible smoothed particle hydrodynamics, Applied Ocean Research, **49**, 57-71.

[43] Ma, Z.H., Causon, D.M., Qian, L., Mingham, C.G., Gu, H.B. and Martinez Ferrer, P. (2014), A Compressible Multiphase Flow Model for Violent Aerated Wave Impact Problems, Proceedings of the Royal Society A, 470 (2172).

[44] Liu, X., Xu, H., Shao, S.D. and Lin, P. (2013), An improved incompressible SPH model for simulation of wave-structure interaction", Computers and Fluids, **71**, 113-123.

[45] Rafiee, A. and Thiagarajan, K.P. (2009), An SPH projection method for simulating fluid-hypoelastic structure interaction", Computer Methods in Application Mechanics and Engineering, **198**, 2785-2795.

[46] Lee, C.J.K., Noguchi, H. and Koshizuka, S. (2007), Fluid-shell structure interaction analysis by coupled particle and finite element method, Computer and structures, **85**, 668-697.

[47] Hwang, S.C., Khayyer, A., Gotoh, H. and Park, J.C. (2014), Development of a fully Lagrangian MPS-based coupled method for simulation of fluid-structure interaction problems, Journal of Fluids and Structures, **50**, 497-511.

[48] Khayyer, A., Gotoh, H., Park, J.C., Hwang, S.C and Koga, T. (2015), An enhanced fully Lagrangian coupled MPS-based solver for fluid-structure interactions, Journal of JSCE (Coastal Eng.), **71**, 883-888.

[49] Oger, G., Guilcher, P.M., Jacquin, E., Brosset, L., Deuff, J.B. and Le Touzé, D. (2010), Simulations of hydro-elastic impacts using a parallel SPH model, International Journal of Offshore and Polar Engineering, **20**(3), 181-189.

[50] Liao, K., Hu, C. and Sueyoshi, M. (2015), Free surface flow impacting on an elastic structure: Experiment versus numerical simulation, Applied Ocean Research, **50**, 192-208.

[51] Scolan, Y.M. (2004), Hydroelastic behavior of a conical shell impacting on a quiescent-free surface of an incompressible liquid, Journal of Sound and Vibration, **277**, 163-203.

[52] Brackbill, J.U., Kothe, D.B., Zemach, C. (1992), A continuum method for modeling surface tension, Journal of Computational Physics, **100**, 335-354.

[53] Morris, J.P. (2000), Simulating surface tension with smoothed particle hydrodynamics, International Journal for Numerical Methods in Fluids, **33**, 333-353.

[54] Nomura, K., Koshizuka, S., Oka, Y. and Obata, H. (2001), Numerical Analysis of Droplet Breakup Behavior using Particle Method, Journal of Nuclear Science and Technology, **38**(12), 1057-1064.

[55] Ichikawa, H. and Labrosse, S. (2010), Smooth Particle Approach for Surface Tension Calculation in Moving Particle Semi-implicit Method, Fluid Dynamics Research, **42**, 035503.

[56] Khayyer, A., Gotoh H. and Tsuruta, N. (2014), A New Surface Tension for Particle Methods with Enhanced Splash Computation, Journal of Japan Society of Civil Engineers, Ser. B2 (Coastal Engineering), **70**(2), 26-30.

[57] Souto-Iglesias, A., Macià, F., González L.M. and Cercos-Pita, J.L. (2013), On the consistency of MPS, Computer Physics Communications, **184**(3), 732-745.

[58] Liow, J.L. (2001), Splash formation by spherical drops, J. Fluid Mech., **427**, 73-105, 2001.

[59] Gotoh, H. and Sakai, T. (2006), Key issues in the particle method for computation of wave breaking, Coastal Engineering, **53**(2), 171-179.

[60] Leroy, A., Violeau, D., Ferrand, M. and Joly, A. (2015), Buoyancy modelling with incompressible SPH for laminar and turbulent flows, International Journal for Numerical Methods in Fluids, **78**(8), 455-474.

[61] Liu, M.B. and Liu, G.R. (2016), Particle Methods for Multi-Scale and Multi-Physics, World Scientific, 400 pp, ISBN: 978-981-4571-69-2.

[62] Koshizuka, S. (2011), Current achievements and future perspectives on particle simulation technologies for fluid dynamics and heat transfer, Journal of Nuclear Science and Technology, **48**(2), 155-168.

[63] Violeau, D. and Rogers, B.D. (2016), Smoothed particle hydrodynamics (SPH) for free-surface flows: past, present and future, Journal of Hydraulic Research, **54**(1), 1-26.

[64] Vacondio, R., Rogers, B.D., Stansby, P.K. and Mignosa, P. (2016), Variable resolution for SPH in three dimensions: Towards optimal splitting and coalescing for dynamic adaptivity, Comput. Methods Appl. Mech. Engrg., **300**, 442-460.

Free vibration and sound radiation of the rectangular plates based on edge-based smoothed finite element method and application of elemental radiators

†X.Y. You[1], *W. Li[1,2,3], Y.B. Chai[1], and Q.F. Zhang[1]

[1]Department of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, China.

[2]Hubei Key Laboratory of Naval Architecture & Ocean Engineering Hydrodynamics, Wuhan City, China

[3]Collaborative Innovation Center for Advanced Ship and Deep-sea Exploration (CISSE), Shanghai, China

*Presenting author: hhxyyxy@163.com

†Corresponding author: hustliw@hust.edu.cn

Abstract

In this paper, the edge-based smoothed finite element (ES-FEM) method and application of elemental radiators is presented to solve the free vibration and sound radiation problems for the rectangular plates. The edge-based smoothed finite element is utilized for the modeling of plate structure. Three-node triangular elements is used to discretize the three-dimensional (3D) shell, due to its convenience for generating and good adaptability for complicated geometries. The system stiffness is obtained by using the strain smoothing technique over the smoothing domains, such as edge-based domain. Consequently, the employing of the strain smoothing technique can provide a proper softening effect to the FEM model, and cure the "overly-stiff" property existing in the standard FEM. Hence, this implementation can significantly improve the accuracy of the solution for free vibration. The application of elemental radiators can rapidly compute the sound radiation of the rectangular plates without fluid elements.

Keywords: the rectangular plates, free vibration and sound radiation, ES-FEM, elemental radiators.

1. Introduction

Nowadays, the plates have been used widely in many branches of structural engineering, such as aircraft, ships, bridges, buildings, etc. The vibration and sound radiation of plates have attracted engineering's more attention, due to the bad influence to structure's strength and acoustic performance.

Many researchers have carried out the analysis of plates. M. Levinson[1] studied linear elastic theoretical solution to free vibration of the simply-supported plate. Raske, Schlack and Fryba[2][3] researched dynamic response of isotropic rectangular plate under various moving loads. Gbadeyan and Oni[4] also computed dynamic response of rectangular plate under various moving loads based on the improved integral transformation method. The radiation

resistance and efficiency of the plate in frequency domain was computed by using the approximate method, which has been widely applied in many research[5][6] . Williams and Maynard[7] used Rayleigh integral and Fast Fourier Transformation to solve the sound radiation of a plate.

Owing to limitations of the analytical methods, the finite element method (FEM) becomes one of the most popular numerical method to analyze plate structures. In the practical applications, lower-order Reissner-Mindlin shell elements are preferred due to its simplicity and efficiency. However, these low-order shell elements have a defect of the shear locking phenomenon, which has the root of incorrect transverse forces under bending. In order to eliminate shear locking, the discrete shear gap (DSG)[8] was used.

In order to overcome the "overly-stiff" problem in FEM, Liu[9] firstly proposed that the combination of the strain smoothing technique[10] and FEM, so-called the Smoothed Finite Element (S-FEM). In S-FEM models, the finite element mesh is used similarly as in the FEM models, however, the weak form is evaluated based on smoothing domains created from the entities of the element mesh such as cells (CS-FEM), or nodes (NS-FEM), or edges (ES-FEM)[11]. These smoothing domains are linear independent and hence ensure stability and convergence of the S-FEM models.

Due to the easy and automatic generation for complicated domains, the three-node triangular element. In this work, the discrete shear gap technique (DSG) is combined the ES-FEM to give a so-called ES-DSG element for plate analysis. The ES-DSG has a superior property compared to standard FEM. The employing of the strain smoothing technique can provide a proper softening effect to the FEM model, and cure the "overly-stiff" property existing in the standard FEM.

2. Three-node Reissner-Mindlin shell element

The middle surface of plate is defined as the reference plane, and let u, v, w be the displacements of the middle surface in the x, y, z direction, let $\beta_x, \beta_y, \beta_z$ be the rotation in the y, x, z direction, which is shown in Fig. 1.

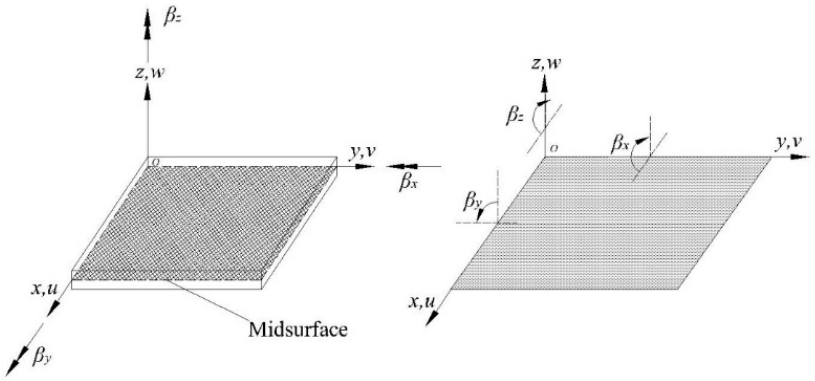


Figure 1. Reissner–Mindlin flat plate

The six independent freedom of three-nodes shell element at any node can be written as below, as is shown in Fig. 2.

$$\mathbf{u} = [u \;\; v \;\; w \;\; \beta_x \;\; \beta_y \;\; \beta_z]^{\mathrm{T}} \tag{1}$$

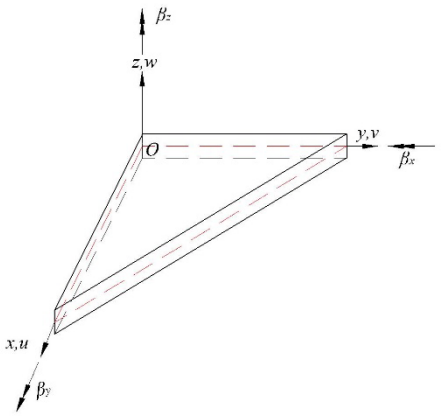


Figure 2. The three-node Reissner–Mindlin shell element

Therefore, the membrane strain ε^{m}, the curvature of the shell element κ and the shear strain γ are constructed as

$$(\boldsymbol{\varepsilon}^m)^{\mathrm{T}} = \left[\frac{\partial u}{\partial x} \;\; \frac{\partial v}{\partial x} \;\; \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right], \;\; \boldsymbol{\kappa}^{\mathrm{T}} = \left[\frac{\partial \beta_x}{\partial x} \;\; \frac{\partial \beta_y}{\partial x} \;\; \frac{\partial \beta_x}{\partial y} + \frac{\partial \beta_y}{\partial x}\right], \;\; \boldsymbol{\gamma} = \begin{bmatrix} \dfrac{\partial w}{\partial x} + \beta_x \\ \dfrac{\partial w}{\partial y} + \beta_y \end{bmatrix}. \tag{2}$$

For the free vibration analysis of Reissner–Mindlin shell, the standard Galerkin weak form can be written as

$$\int_\Omega (\delta\boldsymbol{\varepsilon}^m)^{\mathrm{T}} \mathbf{D}^m \boldsymbol{\varepsilon}^m \, \mathrm{d}\Omega + \int_\Omega \delta\boldsymbol{\kappa}^{\mathrm{T}} \mathbf{D}^b \boldsymbol{\kappa} \, \mathrm{d}\Omega + \int_\Omega \delta\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{D}^s \boldsymbol{\gamma} \, \mathrm{d}\Omega + \int_\Omega \delta\mathbf{u}^{\mathrm{T}} \mathbf{m}\ddot{\mathbf{u}} \, \mathrm{d}\Omega = 0 \tag{3}$$

where $\mathbf{m}$ is the mass matrix containing the density of the material ρ and thickness of the plate t as

$$\mathbf{m} = \mathrm{diag}\left[\rho t, \; \rho t, \; \rho t, \; \rho t^3/12, \; 0\right], \tag{4}$$

$$\mathbf{D}^m = \frac{Et}{1-\nu^2}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix}, \tag{5}$$

$$\mathbf{D}^b = \frac{Et^3}{12(1-\nu^2)}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix}, \tag{6}$$

$$\mathbf{D}^s = \frac{Etk}{2(1+\nu)}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{7}$$

Discretize the problem domain Ω into N_e finite elements, and $\Omega = \bigcup_{e=1}^{Ne} \Omega_e$ and $\Omega_i \bigcap \Omega_j = \varnothing \quad (i \neq j)$. Consequently, the finite element displacement solution $\mathbf{u}^{\mathrm{h}} = \begin{bmatrix} u & \mathrm{v} & \mathrm{w} & \beta_x & \beta_y & \beta_z \end{bmatrix}^{\mathrm{T}}$ of the Reissner–Mindlin shell model is defined as

$$\mathbf{u}^{\mathrm{h}} = \sum_{I=1}^{N_n} N_I(\mathbf{x})\mathbf{I}_6\mathbf{d}_I = \sum_{I=1}^{N_n} \mathbf{N}_I\mathbf{d}_I \tag{8}$$

where $\mathbf{I}_6$ is the 6th rank unit matrix; N_n is the total number of nodes in the problem domain; $N_I(\mathbf{x})$ is the shape function at *I*th node; $\mathbf{d}_I = \begin{bmatrix} u_I & v_I & w_I & \beta_{xI} & \beta_{yI} & \beta_{zI} \end{bmatrix}^{\mathrm{T}}$ is the displacement vector of *I*th node.

In order to eliminate the shear locking, the "Discrete Shear Gap" method is adopted. In each triangular element, the shear strain can be written as

$$\gamma_{yz} = \sum_{I=1}^{3} \frac{\partial N_i(\mathbf{x})}{\partial y}\Delta w_{xi} + \sum_{I=1}^{3} \frac{\partial N_i(\mathbf{x})}{\partial y}\Delta w_{yi} \tag{9}$$

$$\gamma_{xz} = \sum_{I=1}^{3} \frac{\partial N_i(\mathbf{x})}{\partial x}\Delta w_{xi} + \sum_{I=1}^{3} \frac{\partial N_i(\mathbf{x})}{\partial x}\Delta w_{yi} \tag{10}$$

where Δw_{xi} and Δw_{yi} are Discrete Shear Gap at *I*th node given by

$$\begin{aligned} &\Delta w_{x1} = \Delta w_{x3} = \Delta w_{y1} = \Delta w_{y2} = 0, \\ &\Delta w_{x2} = (w_2 - w_1) + \frac{1}{2}a(\beta_{x1} + \beta_{x2}) + \frac{1}{2}b(\beta_{y1} + \beta_{y2}), \\ &\Delta w_{y3} = (w_3 - w_1) + \frac{1}{2}c(\beta_{x1} + \beta_{x3}) + \frac{1}{2}d(\beta_{y1} + \beta_{y3}). \end{aligned} \tag{11}$$

The a, b, c, d in Eq.11 are defined as

$$\begin{aligned} a &= x_2 - x_1, \quad b = y_2 - y_1, \\ c &= x_3 - x_1, \quad d = y_3 - y_1. \end{aligned} \tag{12}$$

where x_i and y_i (i=1-3) are the coordinates of the nodes in a triangular element.

Therefore, the membrane, bending and shear strains can be expressed in the matrix forms as

$$\boldsymbol{\varepsilon}^{\mathrm{m}} = \sum_I \mathbf{R}_I \mathbf{d}_I, \quad \boldsymbol{\kappa} = \sum_I \mathbf{B}_I \mathbf{d}_I, \quad \boldsymbol{\gamma}^{\mathrm{s}} = \sum_I \mathbf{S}_I \mathbf{d}_I \tag{13}$$

where

$$\mathbf{R}_I = \begin{bmatrix} N_{I,x} & 0 & 0 & 0 & 0 & 0 \\ 0 & N_{I,y} & 0 & 0 & 0 & 0 \\ N_{I,y} & N_{I,x} & 0 & 0 & 0 & 0 \end{bmatrix} \tag{14}$$

$$\mathbf{B}_I = \begin{bmatrix} 0 & 0 & N_{I,x} & 0 & 0 & 0 \\ 0 & 0 & 0 & N_{I,y} & 0 & 0 \\ 0 & 0 & N_{I,y} & N_{I,x} & 0 & 0 \end{bmatrix} \tag{15}$$

$$\mathbf{R}_I = \frac{1}{2A_{\mathrm{e}}} \begin{bmatrix} A_{\mathrm{e}} & 0 & b-d & \frac{ad}{2} & \frac{bd}{2} & d & \frac{-bc}{2} & \frac{-bd}{2} & -b \\ 0 & A_{\mathrm{e}} & c-a & \frac{-ac}{2} & \frac{-bc}{2} & -c & \frac{ac}{2} & \frac{ad}{2} & a \end{bmatrix} \tag{16}$$

Thus, the global stiffness matrix $\mathbf{K}$ can be expressed as

$$\mathbf{K} = \int_\Omega \mathbf{R}^{\mathrm{T}} \mathbf{D}^m \mathbf{R} \, \mathrm{d}\Omega + \int_\Omega \mathbf{B}^{\mathrm{T}} \mathbf{D}^b \mathbf{B} \, \mathrm{d}\Omega + \int_\Omega \mathbf{S}^{\mathrm{T}} \mathbf{D}^s \mathbf{S} \, \mathrm{d}\Omega \tag{17}$$

and the global mass matrix $\mathbf{M}$ can be expressed as

$$\mathbf{M} = \int_\Omega \mathbf{N}^{\mathrm{T}} \mathbf{m} \mathbf{N} \, \mathrm{d}\Omega \tag{18}$$

and the load vector $\mathbf{F}$ can be defined as

$$\mathbf{F} = \int_\Omega p\mathbf{N} \, \mathrm{d}\Omega + \mathbf{f}^{\mathrm{b}} \tag{19}$$

For free vibration analysis of the Reissner–Mindlin shell model, we get

$$(\mathbf{K} - \omega^2 \mathbf{M})\mathbf{d} = 0 \tag{20}$$

where ω is the natural frequencies and $\mathbf{d}$ is the mode shape vectors.

3. Edge-based smoothed finite element method

The edge-based strain smoothing technique for shell elements will be implemented in the sub-domain based on edge of triangular elements. The domain is firstly discretized as triangular elements as the standard FEM. However, the numerical integral in Eq. (17) are no longer based on triangular elements, but based on the smoothing domain $\Omega_k \ (k=1,\ 2,\ \ldots,N)$, in which N is the total number of the edge in the problem domain. The smoothing domain of each edge k is constructed by connecting two end-points of the edge and the middle point of its surrounding triangular elements, as is shown in Fig. 3.

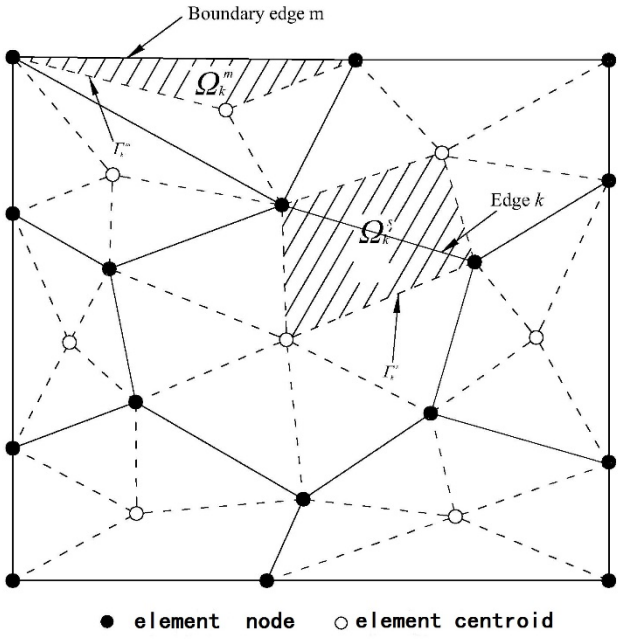


Figure 3. The edge-based smoothing domain

By using the edge-based strain smoothing technique, the integration over the whole triangular elements can be transform to an integral over the whole smoothing domains. Then, the smoothed global stiffness matrix can be rewritten as

$$\bar{\mathbf{K}}=\sum_{k=1}^{N}\left(\int_{\Omega_k}\bar{\mathbf{R}}^{\mathrm{T}}\mathbf{D}^{m}\bar{\mathbf{R}}\,\mathrm{d}\Omega+\int_{\Omega_k}\bar{\mathbf{B}}^{\mathrm{T}}\mathbf{D}^{b}\bar{\mathbf{B}}\,\mathrm{d}\Omega+\int_{\Omega_k}\bar{\mathbf{S}}^{\mathrm{T}}\mathbf{D}^{s}\bar{\mathbf{S}}\,\mathrm{d}\Omega\right) \tag{21}$$

Employing the strain smoothing operation over each smoothing domain on the membrane, bending and shear strains of the shell elements, the smoothing membrane, bending and shear strains over the domain Ω_k can be written as

$$\overline{\boldsymbol{\varepsilon}}^{\mathrm{m}}(\mathbf{x}_k)=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\boldsymbol{\varepsilon}^{\mathrm{m}}(\mathbf{x}_k)\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\mathbf{R}_k\mathbf{d}_k\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Gamma_k}\mathbf{R}_k\mathbf{d}_k\,\mathrm{d}\Gamma \tag{22}$$

$$\overline{\boldsymbol{\kappa}}(\mathbf{x}_k)=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\boldsymbol{\kappa}(\mathbf{x}_k)\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\mathbf{B}_k\mathbf{d}_k\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Gamma_k}\mathbf{B}_k\mathbf{d}_k\,\mathrm{d}\Gamma \tag{23}$$

$$\overline{\boldsymbol{\gamma}}^{\mathrm{s}}(\mathbf{x}_k)=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\boldsymbol{\gamma}(\mathbf{x}_k)\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Omega_k}\mathbf{S}_k\mathbf{d}_k\,\mathrm{d}\Omega=\frac{1}{\mathrm{A}_k}\int_{\Gamma_k}\mathbf{S}_k\mathbf{d}_k\,\mathrm{d}\Gamma \tag{24}$$

where A_k is the area of the smoothing domain Ω_k, and Γ_k is the boundary of the smoothing domain Ω_k.

After performing the integral, the smoothed membrane, bending and shear strains in the smoothing domain Ω_k can then be written in following matrix

$$\overline{\boldsymbol{\varepsilon}}^{\mathrm{m}}(\mathbf{x}_k)=\sum_{i=M_k}\overline{\mathbf{R}}_i(\mathbf{x}_k)\mathbf{d}_i \tag{25}$$

$$\overline{\boldsymbol{\kappa}}(\mathbf{x}_k)=\sum_{i=M_k}\overline{\mathbf{B}}_i(\mathbf{x}_k)\mathbf{d}_i \tag{26}$$

$$\overline{\boldsymbol{\gamma}}^{\mathrm{s}}(\mathbf{x}_k)=\sum_{i=M_k}\overline{\mathbf{S}}_i(\mathbf{x}_k)\mathbf{d}_i \tag{27}$$

where M_k is the total number of the nodes in the smoothing domain Ω_k.

4. The sound radiation analysis of plate

By employing the Rayleigh surface integral, each triangular element on the plate can be treated as a simple point source (elemental radiator) that radiating sound. Therefore, the sound pressure[12] at an arbitrary observation location *Q* of the plate is written as below, as is shown Fig. 4

$$p(Q)=\frac{\mathrm{j}\omega\rho_0}{2\pi}\int_{\mathrm{S}}\frac{\mathrm{e}^{-\mathrm{j}kr}}{r}v(P)\,\mathrm{dS} \tag{28}$$

where k is the wave number, ρ_0 is the air density, S is the area of the plate, r is the distance between the observation location Q and the centroid P of a triangular element.

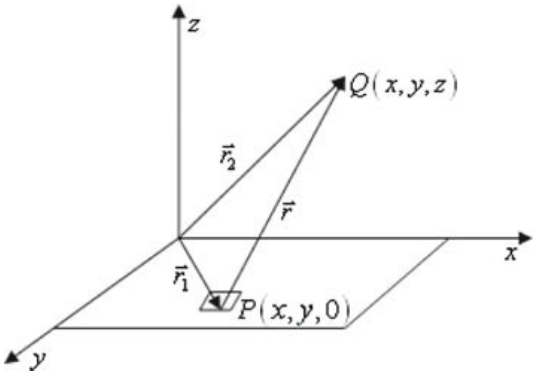


Figure 4. The sound pressure at an arbitrary observation location *Q*

The sound intensity at the observation location Q is defined as

$$I(Q)=\frac{1}{2}\mathrm{Re}\left[p(Q)v^{*}(Q)\right] \tag{29}$$

where $v^{*}(Q)$ is the complex conjugate velocity value at the observation location Q.

The sound power radiating into the semi-infinite space over the plate can be written as

$$W=\int_{S'} I(Q)\,\mathrm{dS}' \tag{30}$$

where S' is an arbitrary surface which cover the plate.

Substituting Eq. 28 and Eq. 29 into Eq. 30, and supposing S' is coincide with S, we can deduce[13]

$$W=\frac{\omega\rho_0}{4\pi}\int_{S'}\left[\int_{S} v(P)\frac{\sin(kr)}{r}v^{*}(Q)\,\mathrm{dS}\right]\mathrm{dS}' \tag{31}$$

where $v(P)$ is the normal velocity at location P, $v^{*}(Q)$ is the normal velocity at location Q.

By discretizing Eq. 31 into a finite form

$$W\approx\frac{\omega\rho_0}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}v(C_i)v^{*}(C_j)\frac{\sin kr}{r}(\Delta\mathrm{S})^2=\mathbf{vZv}^{*} \tag{32}$$

where N is the total number of triangular elements, $v(C_i)$ is the normal centroid velocity

of *i*th triangular element, and ΔS is the area of a triangular element.

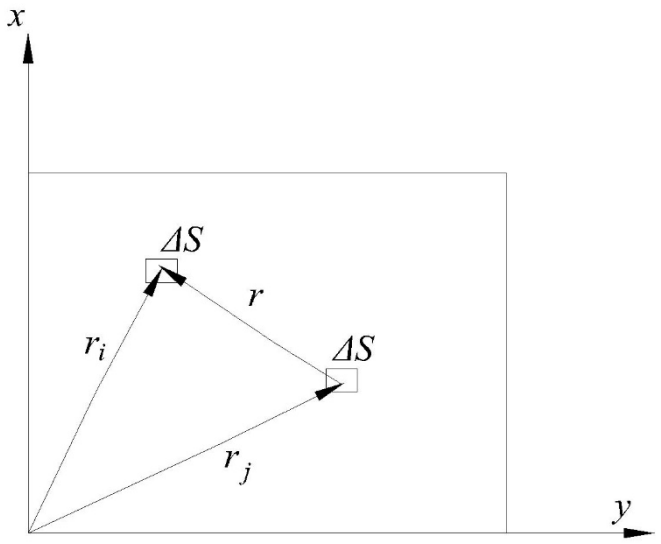


Figure 5. The discretization of a rectangular plate

$\mathbf{Z}$ is the sound resistance matrix defined as below[14]

$$\mathbf{Z}=\frac{\omega^2\rho_0 S^2}{4\pi N c_0}\begin{bmatrix} 1 & \frac{\sin kr_{1,2}}{kr_{1,2}} & \cdots & \frac{\sin kr_{1,N}}{kr_{1,N}} \\ \frac{\sin kr_{2,1}}{kr_{2,1}} & 1 & \cdots & \frac{\sin kr_{2,N}}{kr_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sin kr_{N,1}}{kr_{N,1}} & \frac{\sin kr_{N,2}}{kr_{N,2}} & \cdots & 1 \end{bmatrix} \tag{33}$$

where c_0 is the sound velocity in air.

Finally, the sound power level L_p of the plate can be defined as

$$L_p = 10\log\frac{W}{W_0} \tag{34}$$

where W_0 is the reference sound power defined as 10^{-12} W.

5. Numerical example

Consider two rectangular plates that is simply supported and clamped. The length, width and thickness of the plates are 0.5 m, 0.4 m and 0.005 m, respectively. The material parameters of the plates are given by Young's modulus 210 GPa; Poisson's ratio $v=$ 0.3 and the density

$\rho =$ 7850 kg / m^3. A uniform discretization of 20 × 16 elements is used, as shown in Fig. 6.

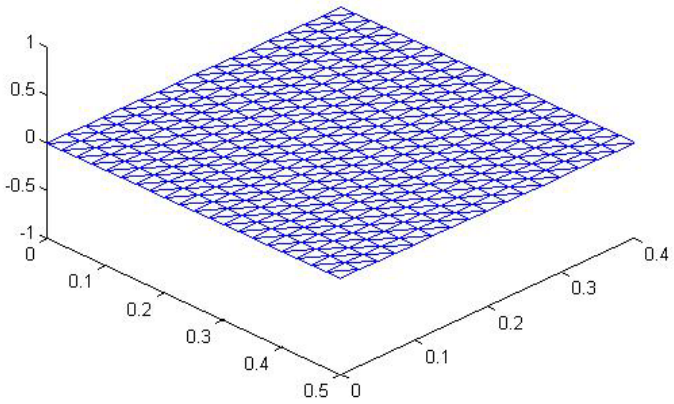


Figure 6. The mesh of a rectangular plate

For free vibration analysis, the eigen frequencies of the plates by the ES-FEM, together with the reference of the commercial software ANSYS are listed in Table 1 below. Fig. 7-9 is the mode shape of the clamped plate computed by ES-FEM and the ANSYS, Fig. 10-12 is the mode shape of the simply supported plate by ES-FEM and the ANSYS.

Table 1. The eigen frequencies results (Hz) of the plates from different methods

	The simply supported plate			The clamped plate		
Mode	ES-FEM (20 × 16)	ANSYS (20× 16)	ANSYS (high quality mesh)	ES-FEM (20 × 16)	ANSYS (20 × 16)	ANSYS (high quality mesh))
1	126.0	127.3	125.4	235.0	238.2	232.3
2	275.6	279.4	272.3	417.6	425.3	407.8
3	360.6	365.6	355.3	543.9	556.6	531.8
4	512.4	523.8	501.0	715.1	729.3	692.5
5	529.9	540.5	517.3	726.9	751.3	693.0

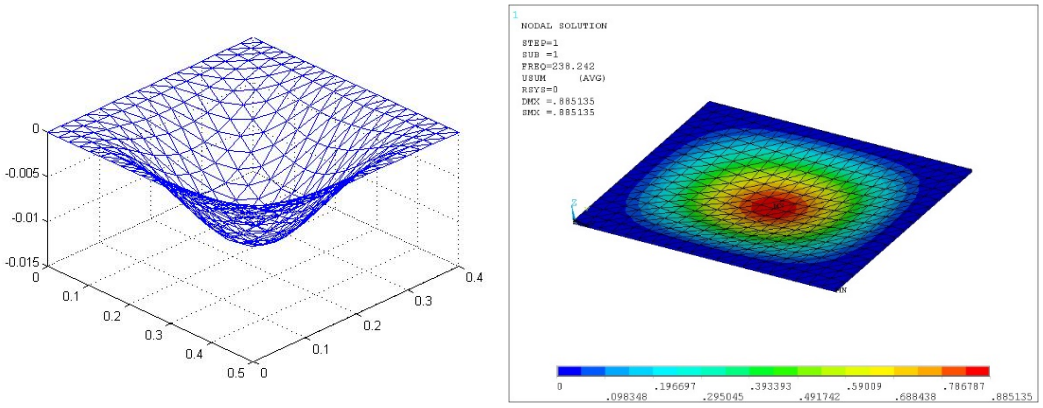


Figure 7. The first mode shape of the clamped plate

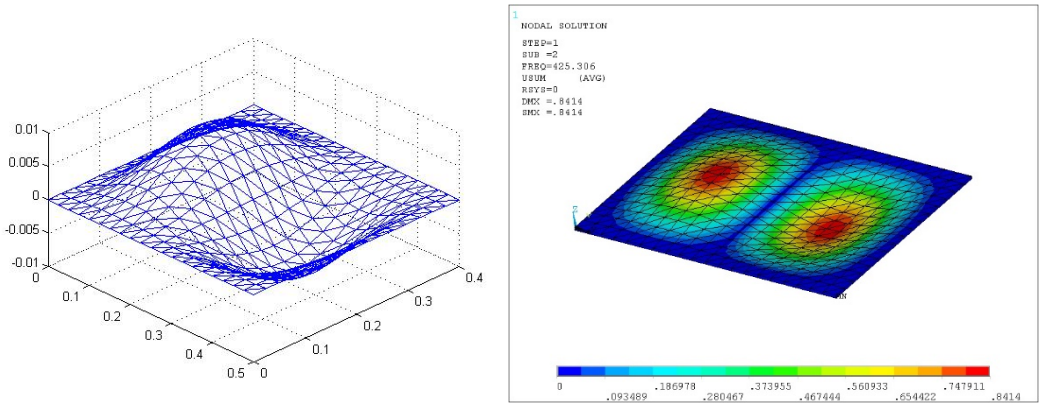


Figure 8. The second mode shape of the clamped plate

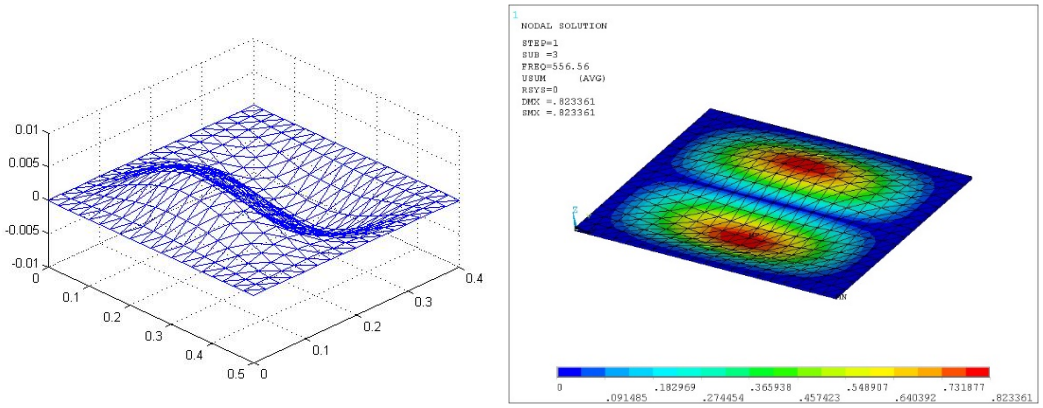


Figure 9. The third mode shape of the clamped plate

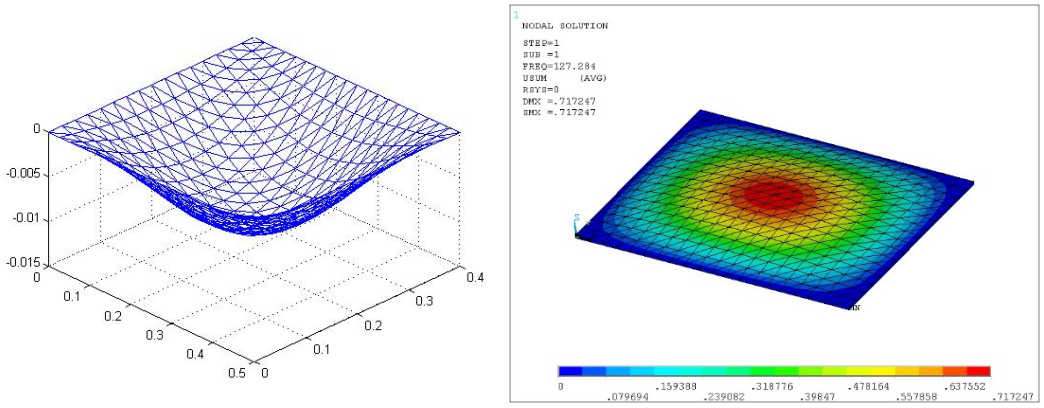


Figure 10. The first mode shape of the simply supported plate

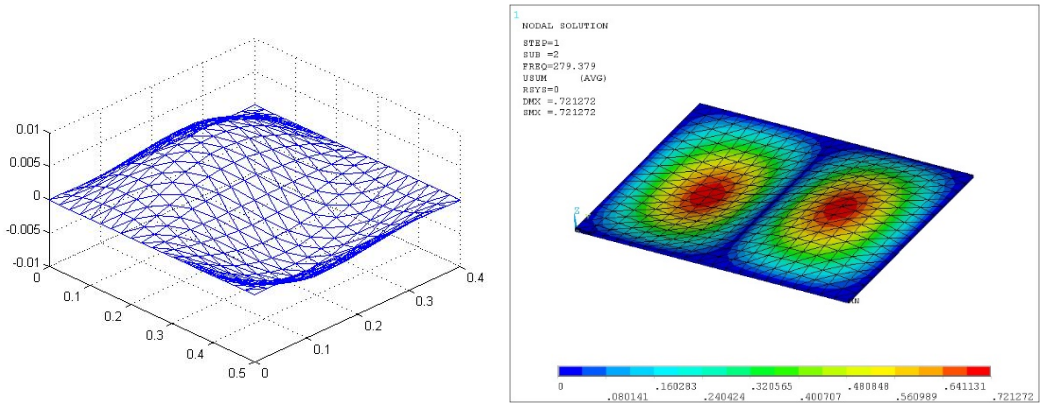


Figure 11. The second mode shape of the simply supported plate

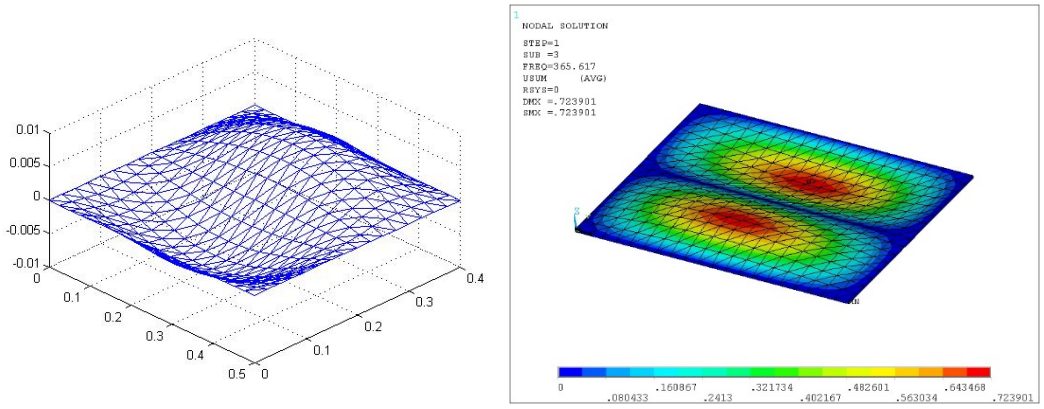


Figure 12. The third mode shape of the simply supported plate

From Table 1, it is observed that the results of the ES-FEM are more accurate than results of the commercial software ANSYS with the same mesh.

As shown in Fig. 13, the rectangular plate is subjected to a normal concentrated force 1 N on centroid of the surface. Then, computing sound power level of the plates from 1-1000 Hz in semi-infinite domain, together with the reference of the commercial software LMS Virtual.Lab are listed in Fig. 14-15. The density and velocity of air are defined as 1.205 kg/m^3 and 340 m/s.

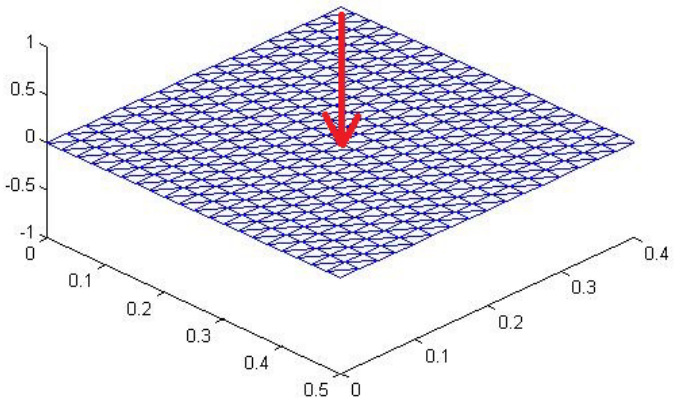


Figure 13. The rectangular plates with a concentrated force

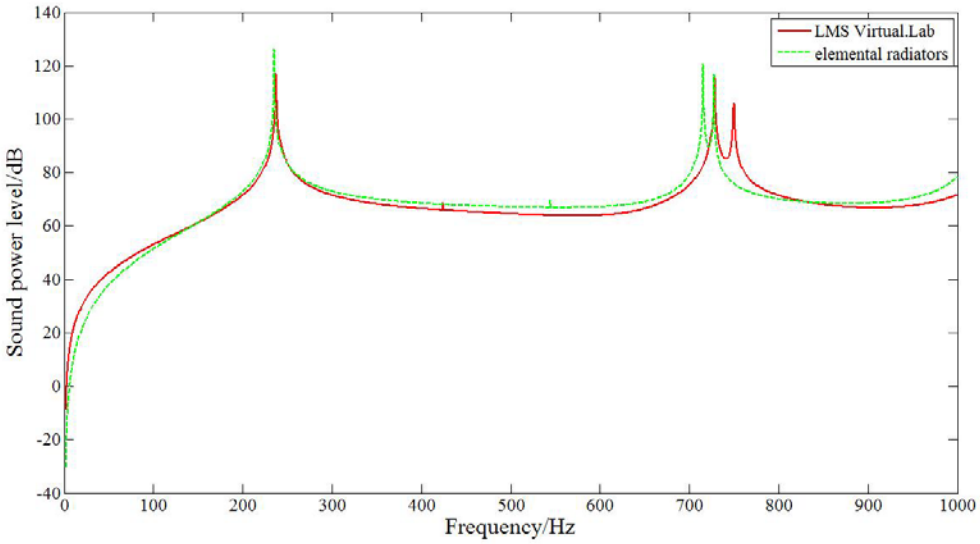


Figure 14. The sound power level of the clamped plates

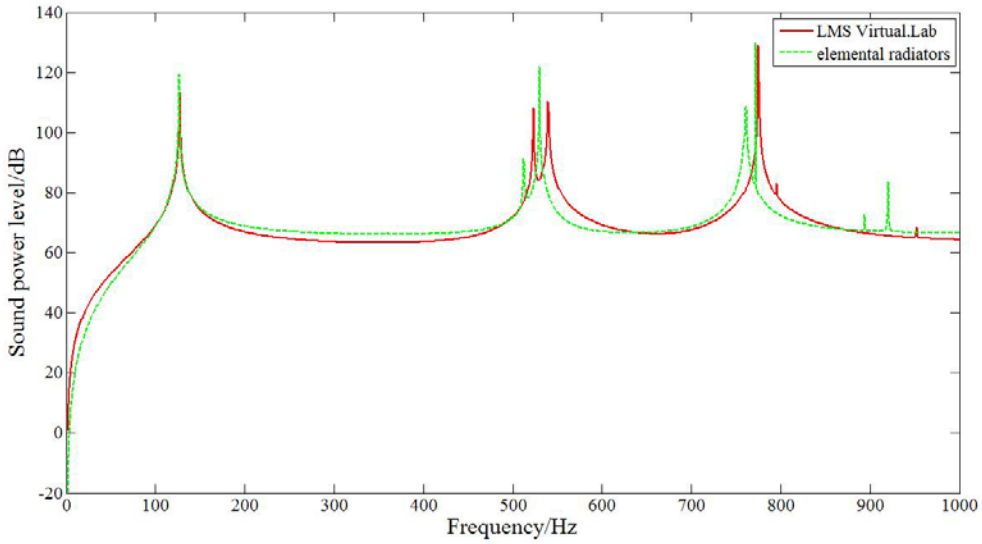


Figure 15. The sound power level of the simply supported plates

From Fig. 14-15, the results from elemental radiators are similar to the results from the Virtual.Lab, especially for the first peak. Due to the difference of two methods above in free vibration analysis, the rear peaks are slightly noncoincidence.

6. Conclusion

In this work, the edge-based smoothed finite element method with the Discrete Shear Gap is used in free vibration analysis of the plates, and the application of elemental radiators is utilized in sound radiation analysis. Through the numerical examples, some conclusion can be drawn below:

(1) The ES-DSG can give better accuracy than standard FEM in free vibration analysis using the same element mesh.

(2) The application of elemental radiators can not only provide a rapid computation, but manifest a desirable accuracy.

References

[1] M.Levinson (1985) Free vibrations of a simply supported rectangular plate, *Journal of Sound and Vibration* **98(2)**, 289–298.
[2] T. F. Raske, A. L. Schlack (1967) Dynamic response of plates due to moving loads, *Journal of the Acoustical Society of America* **42(3)**, 625-635.
[3] L. Fryba (1999) Vibration of solids and structures uder moving loads, Thomas Telford, London, UK.
[4] J. A. Gbadeyan, S. T. Oni. (1995) Dynamic behavior of beams and rectangular plates under moving loads, *Journal of Sound and Vibration* **182(5)**, 677–695.
[5] G. Maidanik (1974) Vibration and radiative classification of modes of a baffled finite panel, *Journal of Sound and Vibration* **34**, 447–455.
[6] C. E.Wallace (1972) Radiation resistance of rectangular panels, *Journal of the Acoustical Society of America* **53**, 946-952.
[7] E. G. Williams, J. D. Maynard (1982) Numerical evaluation of the Rayleigh integral for planar radiators using the FFT, *Journal of the Acoustical Society of America* **72(6)**, 2020-2030.
[8] K. U. Bletzinger, M. Bischoff, E. Ramm (2000) A unified approach for shear-locking free triangular and rectangular shell finite elements, *Comput. Struct* **75**, 321-334.

[9] Liu GR, Dai KY, Nguyen TT (2007) A smoothed finite element method for mechanics problems, *Comput Mech* **39**, 859–877.
[10] J. S. Chen, C. T. Wu, S. Yoon, Y. You (2001) A stabilized conforming nodal integration for Galerkin mesh-free methods, *International Journal for Numerical Methods in Engineering* **50**, 435-466
[11] Liu G. R., Nguyen-Thoi, T., and Lam, K. Y. (2009) An Edge-based Smoothed Finite Element Method (ES-FEM) for static, free and forced vibration analyses of solids, *Journal of Sound and Vibration* **320(4-5)**, 1100-1130.
[12] Rayleigh (1987) *The theory of sound*, 2nd edn, Dover Publications, New York, USA.
[13] Sung C. C., Jan, J. T. (1997) The response of and sound power radiated by a clamped rectangular plate, *Journal of sound and vibration* **207(3)**, 301-317.
[14] Qiao Y., Huang Q. (2007) Fluctuation strength of modulated sound radiated from rectangular plates with mixed boundary conditions, *Archive of Applied Mechanics* **77(10)**, 729-743.
[15] Zheng G., Cui X., Li G., Wu S. (2011) An edge-based smoothed triangle element for non-linear explicit dynamic analysis of shells, *Computational Mechanics*, **48(1)**, 65-80.

Modeling and simulating methods for the desiccation cracking

Sayako Hirobe[1,a)] and Kenji Oguni[1,b)]

[1]Department of System Design Engineering, Keio University, Japan

[a)]Corresponding author: s.hirobe@keio.jp

[b)]oguni@sd.keio.ac.jp

ABSTRACT

The desiccation cracks can be observed on dry-out soil fields or other various materials under desiccation. These cracks have a net-like structure and tessellate the surface of the materials into polygonal cells. The averaged cell sizes change systematically depending on the size of the specimen. In spite of the varieties of the materials, these fundamental features of the cell topology are conserved. This implies the existence of the governing mechanism behind the desiccation crack phenomenon regardless of the material. In this paper, the desiccation crack phenomenon is modeled by the coupling of the desiccation, deformation, and fracture. We perform the simulations for the reproduction of the desiccation cracking based on this coupling model. In the simulation, the finite element analysis for the desiccation problem and the analysis of particle discretization scheme finite element method for the deformation and fracture problems are weakly coupled. The results of the simulation show the satisfactory agreements with the experimental observation in terms of the geometry of the crack pattern, the increase tendency of the averaged cell size depending on the size of the specimen, and the hierarchical sequence of the cell formation. These agreement indicate that the proposed model and method capture the fundamental features and mechanism of the desiccation cracking.

Keywords: Desiccation cracks, Pattern formation, Coupled problem, PDS-FEM.

Introduction

The desiccation cracks can be observed on dry-out soil fields or other various materials under desiccation. These cracks have a net-like structure and tessellate the surface of the materials into polygonal cells with almost constant size and the averaged cell sizes change systematically depending on the size of the specimen. These features of the desiccation cracks are searched on the various materials in previous researches [1]-[8]. In spite of the varieties of the materials, the fundamental features of the cell topology (i.e., the net-like structure of the cracks, the change in averaged cell size depending on the thickness of the specimen) are conserved. This conservation of the features implies the existence of the governing mechanism behind the desiccation crack phenomenon regardless of the choice of the materials. However, the experimental researches cannot explain this governing mechanism because the measurement of the local distribution of the physical quantities near the cracks such as the water content and the stress is still difficult. Thus, the numerical approaches are required for detailed quantitative discussion.

In previous researches, a number of models and analysis methods for the analysis of the desiccation crack phenomenon are proposed. Most of these models assume the homogeneous water distribution and ensuing uniform drying shrinkage [9]-[12]. While this assumption might be sufficient for the thin-layer specimen where the gradient of the water distribution can be neglected, it cannot be applied for the thick-layer specimen where the gradient of the water distribution remarkably appears. On the other hand, some models attempt to embed the inhomogeneous water distribution due to desiccation in the stress analysis [13][14]. However, these models and methods do not introduce the effect of the cracks in the desiccation and deformation problem. Thus, they can be regarded as the pseudo-coupling analysis. This pseudo-coupling analysis can reproduce the crack initiation or the final crack pattern in the limited case, the process of the crack pattern formation cannot be reproduced.

In this paper, the desiccation crack phenomenon is modeled as the coupling of the desiccation, deformation, and fracture. The desiccation problem and the deformation problem are described by the diffusion equation and the equation of force equilibrium respectively and the effect of fracture is embedded in each problem. We perform the simulations for the reproduction of the desiccation cracking based on this coupling model. In the simulation, the finite element analysis for the desiccation problem and the analysis of particle discretization scheme finite element method (PDS-FEM) [15][16] for

the deformation and fracture problems are weakly coupled. The simulation results are compared with the results of drying experiments of calcium carbonate slurry to validate the proposed model and simulation method qualitatively. Throughout this paper, the summation convention is employed for the subscripts in the equation.

Drying Experiment of Calcium Carbonate Slurry

We performed the drying experiments of calcium carbonate slurry to observe the change in cell sizes depending on the thickness of the specimen and the pattern formation process of the desiccation cracking. The change in averaged volumetric water content was measured during desiccation for the determination of the parameters used in the numerical analysis. The saturated calcium carbonate slurry was prepared at the volumetric water content rate 72%. Then, the slurry was poured into the rectangular acrylic container; the size of the container was $100 \times 100 \times 50$ mm. The thickness of the specimen was set as 5 mm, 10 mm, 20 mm, and 30 mm. The slurry was dried at 20 °C temperature and at 50% relative humidity in the air until the entire of the specimen dried out completely.

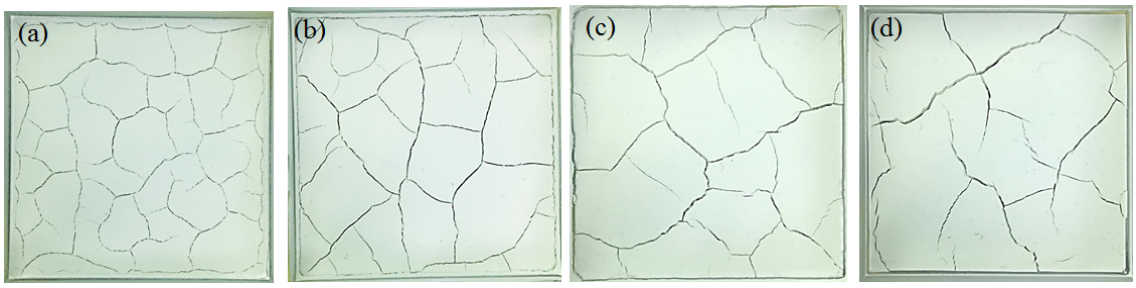


Figure 1. The final crack patterns formed on the top surface of the specimen after the desiccation with different thickness. (a) 5 mm, (b) 10 mm, (c) 20 mm, (d) 30 mm.

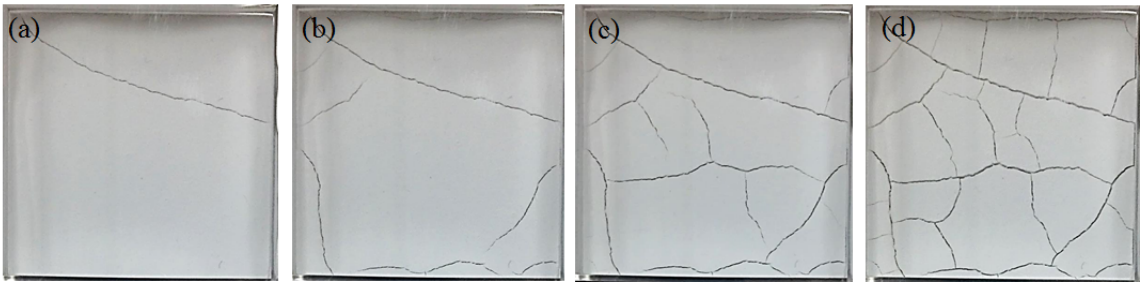


Figure 2. The cell formation process of the drying experiment in the case of 10 mm thickness. (a) the crack initiation, (b) the primary cracks growth, (c) the secondary cracks growth and the tessellation of the lager cells, (d) the final crack pattern.

During the desiccation, the excessive water layer on the top surface of the specimen disappeared at the volumetric water content 56.6% and the cracks initiated on the top surface of the specimen at the volumetric water content 22.4%. The pattern formation of the desiccation cracks terminated before the entire specimen dries out (at the volumetric water content 20.4%). Figure 1 shows the final patterns of cracks formed on the top surface of the specimen with different thickness. The size of the cells framed by the cracks is kept almost constant in each thickness and the averaged cell size increase with the increase of the thickness of the specimen. Figure 2 shows the pattern formation process of the cracks on the top surface of the specimen in the case of 10 mm thickness. On the initial stage of the desiccation cracking, some long and curved cracks initiate on the edge of the specimen and extend to the other edges traversing the specimen (Fig 2 (a) and (b)). These cracks (considered as the primary cracks) do not branch and form the largest structure of the cells. Then, relatively short cracks are formed and tessellate the lager cells (Fig 2 (c) and (d)). These cracks (considered as the secondary cracks) often branch and terminate when they meet the existing cracks. This hierarchical cell tessellation by the secondary cracks continues until the crack initiation terminates. During the desiccation, the volumetric water content reduced almost linearly.

Mathematical Model for Desiccation Cracking

Field Equations for Desiccation Cracking

The desiccation crack phenomenon can be regarded as the coupled problem of the desiccation, deformation, and fracture. For the formulation of this coupled model, we introduce the governing equations for the desiccation problem in fractured medium and the deformation problem in fractured medium.

The desiccation process of the mixture of the powder and the water can be expressed by the Richards' equation:

$$\frac{\partial \theta}{\partial t} = \nabla(D(\theta)\nabla\theta) + \frac{\partial K(\theta)}{\partial z} \tag{1}$$

where θ is a volumetric water content, t is time, $K(\theta)$ is an unsaturated hydraulic conductivity, and $D(\theta)$ is a moisture diffusion coefficient. When we assume the constant moisture diffusion coefficient and neglect the gravitational effect, the Richards' equation is simplified to the linear moisture diffusion equation in terms of the volumetric water content θ:

$$\frac{\partial \theta}{\partial t} = D\nabla^2\theta. \tag{2}$$

Here, the volumetric water content θ is a function of the position $\boldsymbol{x}$ and time t.

Consider a permeable and linearly elastic body Ω with external boundary $\partial\Omega$. When the initial volumetric water content in Ω is set as $\theta^0(\boldsymbol{x})$ and the water evaporates from the external boundary $\partial\Omega$, the desiccation process in Ω is expressed as the next initial boundary value problem:

$$\begin{cases} \dot{\theta} = D\nabla^2\theta & \boldsymbol{x} \in \Omega \quad (3a) \\ \theta(\boldsymbol{x}, 0) = \theta^0(\boldsymbol{x}) & \boldsymbol{x} \in \Omega \quad (3b) \\ D\dfrac{\partial \theta}{\partial \boldsymbol{n}} = -q^{\Omega}(\boldsymbol{x}, t) & \boldsymbol{x} \text{ on } \partial\Omega \quad (3c) \end{cases}$$

where $q^{\Omega}(\boldsymbol{x}, t)$ is a water flux due to evaporation from the external boundary $\partial\Omega$. In the desiccation problem, the crack surfaces Γ can be regarded as the newly created evaporation surfaces. Therefore the effect of cracks on the desiccation process is embedded as the Neumann boundary condition:

$$D\frac{\partial \theta}{\partial \boldsymbol{n}} = -q^{\Gamma}(\boldsymbol{x}, t) \quad \boldsymbol{x} \text{ on } \Gamma \tag{4}$$

where $q^{\Gamma}(\boldsymbol{x}, t)$ is a water flux due to evaporation from the crack surfaces Γ.

On the other hand, the deformation process of an isotropic and elastic body Ω corresponding to the change in the volumetric water content θ given by the initial boundary value problem (3) is governed by the equation of force equilibrium:

$$\begin{cases} \sigma_{ij,j} = 0 & \boldsymbol{x} \in \Omega \quad (5a) \\ \sigma_{ij} = c_{ijkl}(\varepsilon_{kl} - \varepsilon^s_{kl}) & \boldsymbol{x} \in \Omega \quad (5b) \\ \varepsilon_{ij} = \dfrac{1}{2}(u_{i,j} + u_{j,i}) & \boldsymbol{x} \in \Omega \quad (5c) \end{cases}$$

where σ_{ij} is a stress, c_{ijkl} is a elastic modulus, ε_{ij} is a total strain, ε^s_{ij} is a shrinkage strain, and u_i is a displacement. In the case of the drying shrinkage, since the drying shrinkage strain ε^s_{ij} resulting from the volume reduction due to desiccation is inelastic, the drying shrinkage strain does not contribute to the generation of the stress. Therefore, the elastic strain $\varepsilon^e_{ij} = \varepsilon_{ij} - \varepsilon^s_{ij}$ becomes the source of the stress instead of the total strain ε_{ij} as shown in Eq. (5b). This approach can be also seen in Peron et al.[14]. Considering the isotropy of Ω, the shrinkage strain ε^s_{ij} is derived from the volumetric drying shrinkage strain ε^v corresponding to the reduction of the volumetric water content θ as follows:

$$\varepsilon^v(\boldsymbol{x}, t) = \frac{1}{\alpha}\frac{\rho_w}{\rho_d}\{\theta(\boldsymbol{x}, t) - \theta(\boldsymbol{x}, 0)\} \tag{6}$$

$$\varepsilon^s_{ij} = \frac{1}{3}\varepsilon^v\delta_{ij} \tag{7}$$

where ρ_w is the mass density of the water, ρ_d is the dry bulk density of the powder, α is the moisture shrinkage coefficient of the powder and δ_{ij} is the Kronecker's delta.

When the displacement boundary condition $\bar{u}_i(\boldsymbol{x})$ is prescribed on the external boundary $\partial\Omega$, the deformation process of Ω is given by the next boundary value problem:

$$\begin{cases} \sigma_{ij,j} = 0 & \boldsymbol{x} \in \Omega \quad (8a) \\ \sigma_{ij} = c_{ijkl}(\varepsilon_{kl} - \varepsilon^s_{kl}) & \boldsymbol{x} \in \Omega \quad (8b) \\ \varepsilon_{ij} = \dfrac{1}{2}(u_{i,j} + u_{j,i}) & \boldsymbol{x} \in \Omega \quad (8c) \\ u_i = \bar{u}_i & \boldsymbol{x} \text{ on } \partial\Omega^u. \quad (8d) \end{cases}$$

In the deformation problem, the crack surfaces Γ can be regarded as the traction-free surfaces and the effect of crack surfaces Γ is embedded as

$$\sigma_{ij} n_j = 0 \quad \boldsymbol{x} \text{ on } \Gamma \tag{9}$$

where n_i is a unit normal vector of the crack surfaces Γ.

Thus, the problems of desiccation and deformation in fractured medium are coupled by Eq. (6) (the relationship between volumetric water content and volumetric shrinkage strain) and embedding the effect of common crack surfaces in each problem.

Discretized Form of the Field Equations of Desiccation Cracking

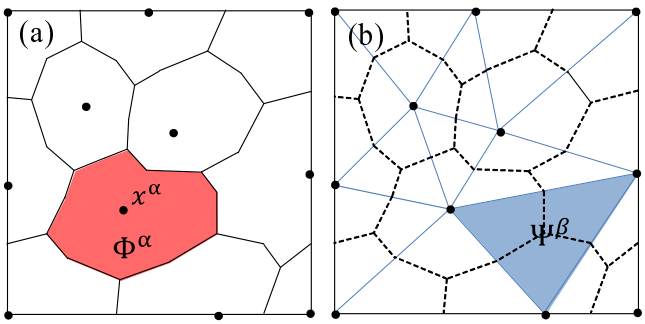


Figure 3. The discretization of the analysis domain Ω in two-dimension. (a) Volonoi tessellations Φ^α, (b) The Delaunay tessellations Ψ^β.

In this research, the analysis of the deformation and fracture are performed by using PDS-FEM. PDS-FEM applies the particle discretization for the variables using a discontinuous and non-overlapping characteristic functions defined on the Voronoi blocks $\{\Phi^\alpha\}$ and the Delaunay tetrahedrons $\{\Psi^\beta\}$; this conjugate pair of geometries are uniquely defined for the set of nodes $\{\boldsymbol{x}^\alpha\}$ as shown in Fig. 3. In two-dimension, the Delaunay block becomes a triangle. The characteristic functions are defined as

$$\phi^\alpha(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \Phi^\alpha) \\ 0 & (\boldsymbol{x} \notin \Phi^\alpha) \end{cases} \tag{10}$$

$$\psi^\beta(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \Psi^\beta) \\ 0 & (\boldsymbol{x} \notin \Psi^\beta). \end{cases} \tag{11}$$

Then, the displacement u_i and the strain ε_{ij} are discretized as

$$u_i(\boldsymbol{x}) = \sum_{\alpha=1}^{N} u_i^\alpha \phi^\alpha(\boldsymbol{x}) \tag{12}$$

$$\varepsilon_{ij}(\boldsymbol{x}) = \sum_{\beta=1}^{M} \varepsilon_{ij}^\beta \psi^\beta(\boldsymbol{x}) \tag{13}$$

where N is a number of Voronoi blocks and M is a number of Delaunay tetrahedrons. Thus, the displacement is discretized by the Voronoi blocks $\{\Phi^\alpha\}$ and the variables related to the spatial gradient of the displacement (i.e., strain and stress) are averaged over the Delaunay tetrahedrons $\{\Psi^\beta\}$.

The boundary value problem (8) for the deformation problem is equivalent to the next variational problem:

$$\begin{aligned} &\mathrm{I}(u_i(\boldsymbol{x})) = \int_\Omega \frac{1}{2}\left(\varepsilon_{ij} - \varepsilon^s_{ij}\right) c_{ijkl}\left(\varepsilon_{kl} - \varepsilon^s_{kl}\right) dV \\ &\text{Minimize } \mathrm{I}(u_i(\boldsymbol{x})) \text{ s.t. } u_i(\boldsymbol{x}) = \bar{u}_i(\boldsymbol{x}) \text{ on } \partial\Omega \end{aligned} \tag{14}$$

Applying the particle discretization scheme to this functional I, the discretized functional $\hat{\mathrm{I}}$ becomes

$$\hat{\mathrm{I}} = \sum_{\beta=1}^{M} \frac{1}{2}\left(\varepsilon^\beta_{ij} - \varepsilon^{s\beta}_{ij}\right) c^\beta_{ijkl}\left(\varepsilon^\beta_{kl} - \varepsilon^{s\beta}_{kl}\right) \Psi^\beta \tag{15}$$

where Ψ^β is the volume of the β-th Delaunay block.

In PDS-FEM, the strain-displacement relation is expressed as

$$\varepsilon^\beta_{ij} = \sum_{\alpha=1}^{N} \frac{1}{2}(B^{\beta\alpha}_j u^\alpha_i + B^{\beta\alpha}_i u^\alpha_j) \tag{16}$$

where

$$\begin{aligned} B^{\beta\alpha}_i &= \frac{1}{\Psi^\beta}\int_{\Psi^\beta} \phi^\alpha_{,i}(\boldsymbol{x})\psi^\beta(\boldsymbol{x}) dV \\ &= \frac{1}{\Psi^\beta}\int_{\partial\Psi^\beta} n^\alpha_i(\boldsymbol{x}) dS \\ &= \frac{1}{\Psi^\beta}\int_{\partial\Phi^\alpha\cap\Psi^\beta} n^\alpha_i(\boldsymbol{x}) dS. \end{aligned} \tag{17}$$

Note that $B^{\beta\alpha}_i$ is identical to the B matrix for the strain field in the ordinary FEM with the linear tetrahedral elements. Applying this strain-displacement relation to the discretized functional $\hat{\mathrm{I}}$ in Eq. (15), the discretized functional $\hat{\mathrm{I}}$ can be expressed in terms of the nodal displacement u^α_i. Then, the stationary condition for $\hat{\mathrm{I}}$ results in the equation of force equilibrium

$$\sum_{\gamma=1}^{N} K^{\alpha\gamma}_{ik} u^\gamma_k = f^\alpha_i, \tag{18}$$

where stiffness matrix $K^{\alpha\gamma}_{ij}$ and external force vector f^α_i is

$$K^{\alpha\gamma}_{ik} = \sum_{\beta=1}^{M} B^{\beta\alpha}_j c^\beta_{ijkl} B^{\beta\gamma}_l \Psi^\beta \tag{19}$$

$$f^\alpha_k = \sum_{\beta=1}^{M} \varepsilon^{s\beta}_{ij}\left(c^\beta_{ijkl} B^{\beta\alpha}_l\right)\Psi^\beta. \tag{20}$$

In PDS-FEM, fracture is expressed as the loss of the interaction between Voronoi blocks and the fracture surfaces are defined on the boundary of Voronoi blocks (i.e., in the Delaunay tetrahedron). The loss of the interaction between Voronoi blocks is expressed as the removal of the contribution of the nodal displacement to the strain averaged over the fractured Delaunay tetrahedron. Thus, $B^{\beta\alpha}_i$ related to the Delaunay tetrahedron and Voronoi blocks composing the fracture surface becomes zero. The effect of this removal of $B^{\beta\alpha}_i$ is finally embedded in the stiffness matrix $K^{\alpha\gamma}_{ij}$. Thus, the Neumann boundary condition (9) on crack surfaces Γ (i.e., traction-free surface) is introduced as the change in the stiffness matrix $K^{\alpha\gamma}_{ij}$ of the equation of force equilibrium (18) in discretized form of the deformation problem.

The analysis of desiccation process is performed by using the ordinary FEM with linear tetrahedral elements corresponding to the Delaunay tetrahedrons used in analysis of deformation and fracture by PDS-FEM. The initial boundary value problem (3) is spatially discretized by using the shape function for the linear tetrahedral elements. Therefore, the Eq. (6) expressing the relation between the volumetric water content and the volumetric drying shrinkage strain is discretized as

$$\varepsilon^{v\beta}(\boldsymbol{x}, t) = \frac{1}{\alpha}\frac{\rho_w}{\rho_d}\{\bar{\theta}^\beta(t) - \bar{\theta}^\beta(0)\} \tag{21}$$

where $\bar{\theta}^\beta$ (function of time t) is an average of the nodal volumetric water content consisting the β-th Delaunay tetrahedron.

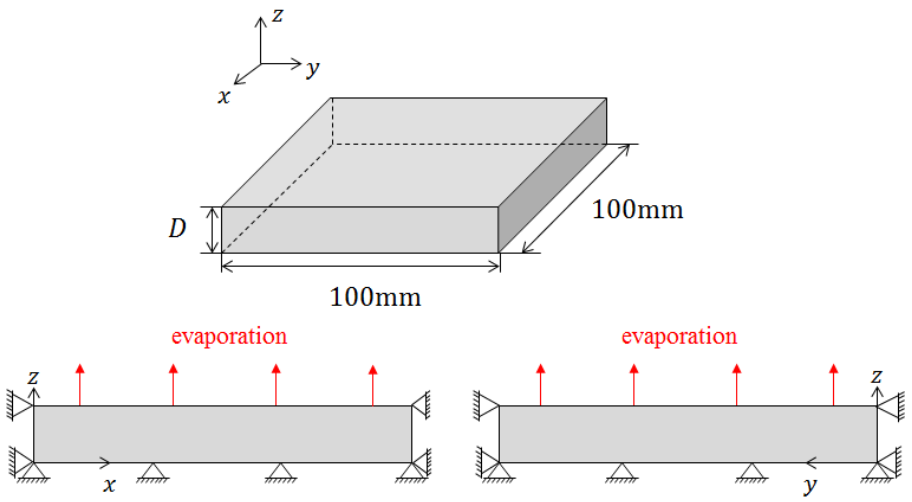


Figure 4. The geometry and the boundary conditions for the numerical analysis of the desiccation cracking.

For the introduction of the Neumann boundary condition (4) to the discretized form of the initial boundary value problem (3), the expression of the crack surfaces on PDS-FEM is applied to the FEM analysis of the desiccation process. Thus, the crack surfaces Γ is defined on the boundary of Voronoi blocks. The Neumann boundary condition (4) can be interpreted as i) elimination of the water flux normal to the crack surfaces Γ and ii) the prescribed water flux q^{Γ} due to evaporation on the crack surfaces Γ. The elimination of the water flux normal to Γ can be introduced as the anisotropic moisture diffusion coefficient. The water flux vector $\boldsymbol{J}$ in the orthonormal coordinate system $\{\boldsymbol{e}_i\}$ is expressed by Darcy's law:

$$\boldsymbol{J} = -D\nabla\theta. \tag{22}$$

We set the orthonormal coordinate system $\{\boldsymbol{e}'_i\}$ with $\boldsymbol{e}'_3$ in the normal direction of the crack surface Γ. The components of the projection of $\boldsymbol{J}$ on Γ (denoted as $\boldsymbol{J}^c$) in the $\{\boldsymbol{e}_i\}$ coordinate system is

$$J^c_i = T_{ji}P_{jk}T_{kl}J_l \tag{23}$$

where coordinate transform matrix T_{ij} and the projection matrix P_{ij} are

$$T_{ij} = \boldsymbol{e}'_i \cdot \boldsymbol{e}_j \tag{24}$$

$$P_{ij} = \begin{cases} 1 & \text{if } i = j = 1, 2 \\ 0 & \text{otherwize.} \end{cases} \tag{25}$$

Thus, the elimination of the water flux normal to Γ (i.e., the replacement of $\boldsymbol{J}$ with $\boldsymbol{J}^c$) corresponds to the introduction of the anisotropic moisture diffusion coefficient ($DT_{ji}P_{jk}T_{kl}$) to the fractured tetrahedral elements.

Since nodes are not placed on the crack surfaces expressed in the analysis of PDS-FEM, the water flux q^{Γ} is prescribed on the nodes placed on the boundary of fractured tetrahedral elements and unfractured tetrahedral elements. This prescription of the water flux corresponds to the assumption of the blunt crack.

Numerical Analysis of Desiccation Cracking

We perform the simulations for the reproduction of the crack patterns and the cell formation process observed in the drying experiments of calcium carbonate slurry. The distribution of the volumetric water content is obtained by the FEM analysis for the initial boundary value problem (3) with a constant time step $\Delta t = 0.1$ hour. Then, the seamless analysis for the deformation and the fracture by PDS-FEM is performed at each time step. Since the time scale for the desiccation and the fracture have a strong contrast, we performed the weak coupling analysis for these problems (i.e., weak coupling of the FEM analysis and the analysis of PDS-FEM). To capture the effect of the fracture surfaces promptly, the time step is reduced to $\Delta t = 0.01$ hour when the maximum traction among all elements reached to the 97% of the tensile strength t_c.

The model sizes and the boundary conditions is set to fit the drying experiments of calcium carbonate slurry; see Fig. 4 The thickness is set as 5 mm, 10 mm, 20 mm, and 30 mm. The nodal displacement of the sides and the bottom surfaces of

Table 1. Mesh sizes for the numerical analysis of the desiccation cracking

model size [mm]	number of elements	number of nodes
$100 \times 100 \times 5$	253,930	50,355
$100 \times 100 \times 10$	278,337	51,726
$100 \times 100 \times 20$	309,509	55,304
$100 \times 100 \times 30$	347,551	61,146

Table 2. The parameters for the numerical analysis of the desiccation cracking

parameters	
Soil dry density ρ_g	$800\,\mathrm{kg/m^3}$
Initial volumetric water content θ^0	0.560
Volumetric water content at the end of the simulation θ^f	0.204
Evaporation speed on the top surface q^Ω	$8.8 \times 10^{-5}\,\mathrm{m/hour}$
Evaporation speed of the crack surfaces q^Γ	$4.4 \times 10^{-5}\,\mathrm{m/hour}$
Moisture shrinkage coefficient α	0.69
Moisture diffusion coefficient D	$3.6 \times 10^{-6}\,\mathrm{m^2/hour}$
Poisson's ratio ν	0.3
Young's modulus E	5.0 MPa
Tensile strength t_c	1.6 MPa

the analysis model are constrained in all directions to express the adhesion between the slurry and the container wall on the drying experiments. The water evaporates from the top surface of the analysis model and crack surfaces. We prepare the finite element models with unstructured mesh for each model; the mesh sizes are shown in Table 1.

The measurable parameters are determined from the drying experiments of calcium carbonate slurry; see Table 2. The initial volumetric water content is set as the volumetric water content at which the excessive water layer disappeared in the drying experiments. The analysis is stopped when the volumetric water content reach to the 20.4% at which the crack pattern formation terminated in the drying experiments. The other parameters which are not measured in the drying experiments of calcium carbonate slurry (Young's modulus, tensile strength, and the moisture diffusion coefficient) are determined from the drying experiments of clayey silt in previous researches[6]. The evaporation speed on the crack surfaces q^Γ can be considered as slower than that on the top surface q^Ω because the opening width of the cracks is narrow. Therefore, the evaporation speed on the crack surfaces q^Γ is set as 50% of q^Ω.

Figure 5 shows the final crack patterns formed on the top surface of the analysis models with different thickness. The cracks have net-like structure and form polygonal cells. The cell sizes are kept almost constant on each thickness and the averaged cell size increases with the increase of the thickness. These geometric features of the crack pattern (i.e., net-like structure and polygonal cells) and the increasing tendency of the cell sizes depending on the thickness of the analysis model can be also observed in the drying experiment of calcium carbonate slurry.

The cell formation process on the top surface of the analysis model in the case of 10 mm thickness is shown in Fig. 6. In the early stage of the desiccation process, some long cracks initiate on the edge of the analysis model and extend traversing the top surface (Fig.4 (a) and (b)). These cracks can be considered as primary cracks observed on the drying experiment of calcium carbonate slurry. Then, relatively short cracks propagate to tessellate the lager cells (Fig.4 (c) and (d)). These cracks often branch and propagate until they meet other cracks; the emergence of the secondary cracks. The features of the shape of the cracks and the hierarchical sequence of the cell formation coincide with the drying experiment of calcium carbonate slurry.

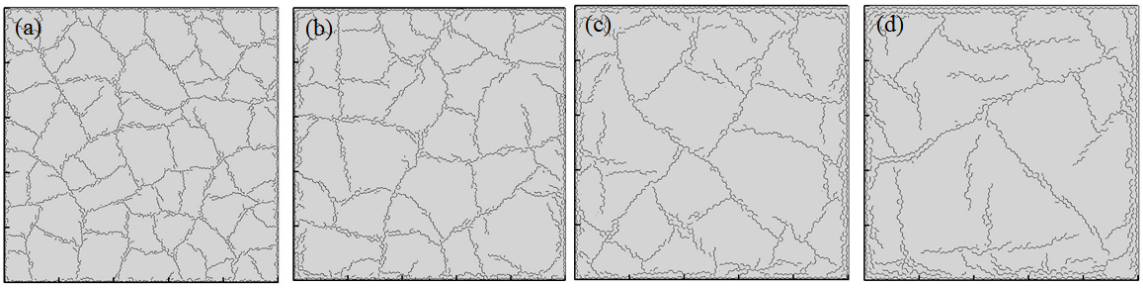


Figure 5. The final crack pattern formed on the top surface for each analysis model. (a) 5 mm, (b) 10 mm, (c) 20 mm, (d) 30 mm.

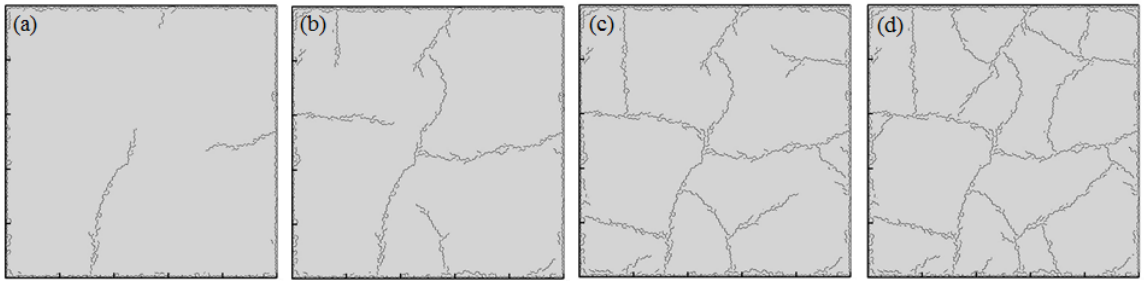


Figure 6. The cell formation process of the numerical analysis in the case of 10 mm thickness. (a) the crack initiation, (b) the primary cracks growth, (c) the secondary cracks growth and the tessellation of the lager cells, (d) the final crack pattern.

Conclusions

In this paper, the problem of the desiccation cracking is modeled by the coupling of desiccation, deformation, and fracture. In the proposed model, the diffusion equation with the anisotropic diffusion coefficient and the equation of the force equilibrium with the stiffness matrix reflecting the loss of the interaction due to fracture are coupled. This coupling analysis is performed by introducing the relation between volumetric drying shrinkage strain and embedding the effect of the common crack surfaces in desiccation and deformation problem.

The simulations with the FEM analysis and the analysis of PDS-FEM are performed to reproduce the crack patterns and the cell formation process observed in the drying test of calcium carbonate slurry. The simulation results show the satisfactory agreements with the experimental observation in terms of the geometry of the crack pattern, the increasing tendency of the averaged cell size depending on the thickness of the specimen, and the hierarchical sequence of the cell formation. These agreement indicate that the proposed model and method capture the fundamental features and mechanism of the desiccation cracking. For more quantitative discussion, we need the parametric study on the parameters which can not be measured in experiments.

References

[1] Corte, A. and Higashi, A. (1960). Experimental research on desiccation cracks in soil. Technical report, U.S. Army Snow Ice and Permafrost Research Establishment, Illinois, USA.

[2] Groisman, A. and Kaplan, E. (2006). An experimental study of cracking induced by desiccation. *Europhysics. Letters*, 25(6):415–420.

[3] Kindle, E. M. (1917). Some factors affecting the development of mud-cracks. *Journal of Geology*, 25:135–144.

[4] Kodikara, J. K., Barbour, S. L., and Fredlund, D. G. (2000). Desiccation cracking of soil layers. In *Unsaturated soils for Asia. Proceedings of the Asian Conference on Unsaturated Soils, UNSAT-ASIA 2000, Singapore, 18-19 May, 2000*, pages 693–698.

[5] Miller, C., Hong, M., and Nazli, Y. (1998). Experimental analysis of desiccation crack propagation in clay liners. *Journal of the American Water Resources Association*, 34(3):677–686.

[6] Peron, H., Hueckel, T., Laloui, L., and Hu, L. B. (2009). Fundamentals of desiccation cracking of fine-grained soils: experimental characterisation and mechanisms identification. *Canadian Geotechnical Journal*, 46(10):1177–1201.

[7] Rodríguez, R., Sánchez, M., Ledesman, A., and Lloret, A. (2007). Experimental and numerical analysis of desiccation of mining waste. *Canadian Geotechnical Journal*, 44(6):644–658.

[8] Vogel, H. J., Hoffmann, H., and Roth, K. (2005b). Studies of crack dynamics in clay soil I. experimental methods, result, and morphological quantification. *Geoderma*, 125:203–211.

[9] Sánchez, M., Manzoli, O. L., and Guimarães, L. J. N. (2014). Modeling 3-D desiccation soil crack networks using a mesh fragmentation technique. *Computers and Geotechnics*, 62:27–39.

[10] Sima, J., Jiang, M., and Zhou, C. (2014). Numerical simulation of desiccation cracking in a thin clay layer using 3D discrate element modeling. *Computers and Geptechnics*, 56:168–180.

[11] Trabelsi, H., Jamei, M., Zenzri, H., and Olivella, S. (2012). Crack patterns in clayey soils: Experiments and modeling. *International journal for numerical and analytical and methods in geomechanics*, 36:1410–1433.

[12] Vogel, H. J., Hoffmann, H., Leopold, A., and Roth, K. (2005a). Studies of crack dynamics in clay soil II. a physically based model for crack formation. *Geoderma*, 125(3-4):213–223.

[13] Musielak, G. and Śliwa, T. (2012). Fracturing of clay during drying: Modelling and numerical simulation. *Transport in porous media*, 95:465–481.

[14] Peron, H., Delenne, J., Laloui, L., and El Youssoufi, M. (2008). Discrete element modeling of drying shrinkage and cracking of soils. *Computers and Geotechnics*, 36:61–69.

[15] Hori, M., Oguni, K., and Sakaguchi, H. (2005). Proposal of FEM implemented with particle discretization for analysis of failure phenomenon. *Journal of the Mechanics and Physics of Solids*, 53:681–703.

[16] Oguni, K., Wijerathne, M. L. L., Okinaka, T., and Hori, M. (2009). Crack propagation analysis using PDS-FEM and comparison with fracture experiment. *Mechanics of materials*, 41(11):1242–1252.

Smoothed Particle Hydrodynamics (SPH) Applications in Some Sediment Dispersion Problems

E. Bertevas[1], T. Tran-Duc[1], B. C. Khoo[2] and N. Phan-Thien[2]†

[1]Keppel-NUS Corporate Laboratory, National University of Singapore, Singapore.
[2]Department of Mechanical Engineering, Faculty of Engineering, National University of Singapore, Singapore.
†Corresponding and presenting author: nhan@nus.edu.sg

Abstract

We present results from an on-going research effort which aims at applying the Smoothed Particle Hydrodynamics (SPH) method to the large-scale particle transport as well as complex flows involving particle-suspension/structure interaction. The main application to this modelling work is related to the estimation of the seabed disturbance created by a moving harvesting device near the seabed for deep-sea applications. Current results are presented for a lab-scale model of sediment disturbance in the vicinity of the harvester as well as ocean-scale sediment transport. The latter includes new developments on SPH formulations of anisotropic diffusion.

Keywords: Smoothed Particle Hydrodynamics, turbulent sediment transport, anisotropic diffusion, sediment/equipment interaction.

Introduction

In view of the environmental impact assessment required prior to harvesting operations in deep-sea environment, predictions of the extent of sediment disturbance and transport need to be provided. The proposed method relies on the description of particle suspensions via a mixture model [1] which was adapted to the Lagrangian framework of SPH [2-3]. Particle transport is modelled through the convection-diffusion of the sediment volume fraction. This accounts for particle sedimentation and particle turbulent diffusion which can be obtained from standard models relating the diffusion coefficient to the turbulent viscosity. In the case of complex flows such as equipment/seabed interactions, the latter can be extracted from the solution of standard turbulence models which are coupled to the momentum equation via Boussinesq's concept of turbulent viscosity. For ocean-scale sediment transport however, the turbulent diffusion is usually specified as directionally dependent and an improved SPH formulation for anisotropic diffusion is presented. The proposed implementation offers the possibility to account for the non-Newtonian nature of the seabed rheology which is mainly composed of clay material. Cohesive particle suspensions may be modelled through volume fraction dependent yield stress fluid models such as Herschel-Bulkley's (for classical viscoplastic fluid models, see [4]) or Papanastasiou's [5] models. The rheological properties of the sediment suspensions may be regarded as functions or functionals of local particle concentration and shear rate. With this perspective in mind, the formulation is applied to a near-field and a far-filed sediment dispersion problem. In the near-field problem, a laboratory-scale setup was designed and comprises a horizontally translating inclined blade partially immersed into a layer of clay sediment. The behavior of the sediment layer was characterised by means of rheological measurements and is modelled as a yield stress fluid. The induced sediment dispersion is investigated for various cases and the SPH predictions are compared with visual observations obtained from the experiments, highlighting the capture of various flow and sediment transport characteristics. In the far-field problem, the spread of the sediment due to a logarithmic steady current from a distributed source is investigated. With a sediment released rate of 3.6 tons per hour for 5 hours (total of 18 tons), the simulation results show that about 93% of the released sediment has been deposited on the floor up to about 11km from the source location, after 3.5 days.

Methodology

The model used is based on the mixture model [5] in which the continuity, momentum conservation and transport of the sediment concentration ϕ are given respectively by

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\boldsymbol{u} = 0, \tag{1}$$

$$\frac{\partial}{\partial t}\boldsymbol{u} + \nabla \cdot \rho \boldsymbol{u}\boldsymbol{u} = -\nabla P + \nabla \cdot \left(\mathbf{T}^{\eta} + \mathbf{T}^{\mathcal{T}}\right) + \rho \boldsymbol{g}, \tag{2}$$

$$\frac{D\phi}{Dt} = -\phi \nabla \cdot \boldsymbol{u} - \nabla \cdot \left(\phi\left(1 - \frac{\phi\rho_s}{\rho}\right)\boldsymbol{u}_s\right) + \nabla \cdot \left(D^{\mathcal{T}} \nabla \phi\right). \tag{3}$$

Here, $\boldsymbol{u}$ is the barycentric velocity of a volume of mixture, D/Dt is the material derivative, and $\rho = \phi\rho_s + (1-\phi)\rho_f$ is the mixture density, $\mathbf{T}^{\eta}$ and $\mathbf{T}^{\mathcal{T}}$ are the viscous and turbulent diffusion stresses, $\boldsymbol{u}_s$ is the sediment settling velocity and $D^{\mathcal{T}}$ is the diffusivity. In addition, the standard $k-\varepsilon$ and $k-\omega$ SST models have been considered to model turbulent viscosity and particle diffusivity. These equations have been transformed in a Lagrangian framework and discretized using the standard SPH derivatives. The turbulent modelling aspects are not in focus here – rather, the diffusion of the sediment is of concern.

Anisotropic diffusion of sediment

Due to stratifications of the water column in oceans, disturbed sediment diffuses mainly in the horizontal directions and is quite limited in the vertical direction. In other words, the sediment diffusion process near the ocean bottom is basically anisotropic. In the transport equation, diffusion coefficient thus is not a scalar but a tensor,

$$\boldsymbol{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}, \tag{4}$$

which is a symmetric and positive definite tensor. Currently, there is a lack of an appropriate SPH formulation for an anisotropic diffusion operator. Thus, a new SPH expression for a general diffusion operator is derived in this study and it is named ASPHAD (Anisotropic SPH approximation for Anisotropic Diffusion). ASPHAD has the form

$$\left(\nabla \cdot (\boldsymbol{D}\nabla c)\right)_i = 2\int_{\Omega_i} \frac{c(\boldsymbol{r}) - c(\boldsymbol{r}')}{|\boldsymbol{r}' - \boldsymbol{r}||\boldsymbol{L}^{-1}\boldsymbol{e}_{\boldsymbol{r}'\boldsymbol{r}}|^2} \frac{\partial W(|\boldsymbol{r}' - \boldsymbol{r}|, h)}{\partial |\boldsymbol{r}' - \boldsymbol{r}|} d\boldsymbol{r}' + \mathcal{O}(h^2) \tag{5}$$

in which $\boldsymbol{e}_{\boldsymbol{r}'\boldsymbol{r}} = (\boldsymbol{r}' - \boldsymbol{r})/|\boldsymbol{r}' - \boldsymbol{r}|$ and $\boldsymbol{L}^{-1}$ is inverse matrix of $\boldsymbol{L}$, which is given by $\boldsymbol{D} = \boldsymbol{L}\boldsymbol{L}^T$. The tensor decomposition could be either singular value decomposition or Cholesky decomposition. Particle, or SPH, discretization form of ASPHAD is

$$\left(\nabla \cdot (\boldsymbol{D}\nabla c)\right)_i = 2\sum_j \frac{V_j c_{ij}}{r_{ij}\left|\boldsymbol{L}^{-1}\boldsymbol{e}_{ij}\right|^2} \frac{\partial W_{ij}}{\partial r_{ij}} \tag{6}$$

For illustration purposes, ASPHAD is used to simulate anisotropic diffusion of a contaminant source in fluid with diffusion coefficients given by

$$\boldsymbol{D} = \begin{bmatrix} 0.12 & 0 \\ 0 & 0.02 \end{bmatrix} (m^2/s) \tag{7}$$

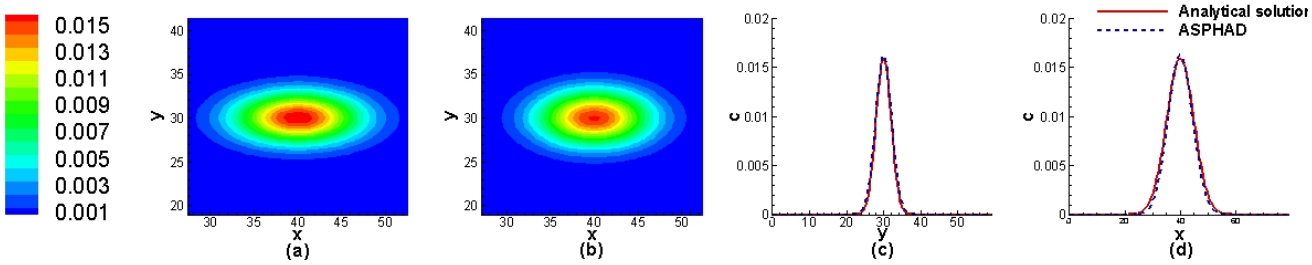


Figure 1: Anisotropic diffusion with diffusing rates of $0.12\ m^2/s$ in x-direction and $0.02\ m^2/s$ in y-direction. (a) Analytical solution, (b) ASPHAD, (c) and (d) Vertical and horizontal concentration distribution through source location $(x_0, y_0) = (40m, 30m)$

Sediment/Equipment Interaction Problems

In order to generate relevant experimental data to validate the proposed model for sediment disturbance induced by a harvester in operation on the seabed, a lab-scale experimental setup was designed in which an inclined blade moves horizontally through a layer of clay suspension mimicking the behavior of the seabed. The experiments take place in a $2m \times 0.5m \times 0.5m$ tank with a blade velocity $O(0.1m.s^{-1})$ and the scale of the experiment is only about one order of magnitude smaller than the real operation scales. Seabed samples recovered from the operation region revealed a behavior similar to suspensions of bentonite clay. Hence, the latter was used as a model for the seabed and the rheological properties of suspensions of various concentrations were measured and fitted to a Papanastasiou model [4] for yield stress fluids. The concentration dependent parameters obtained were fed into our numerical model. The induced seabed disturbance is then compared with SPH numerical simulations and a typical example is shown in Fig. 2, where it can be observed that the numerical simulation captures several of the flow features and sediment disturbance, notably the development of vortex trails in the wake of the blade.

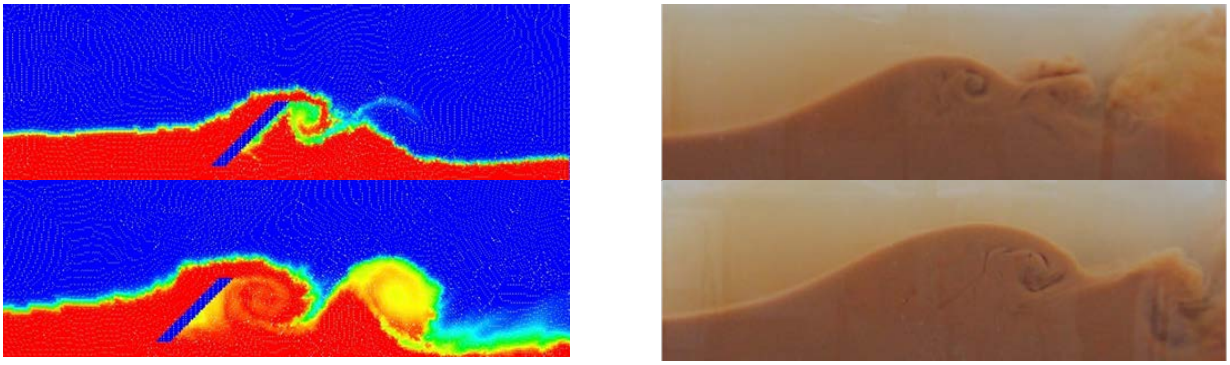
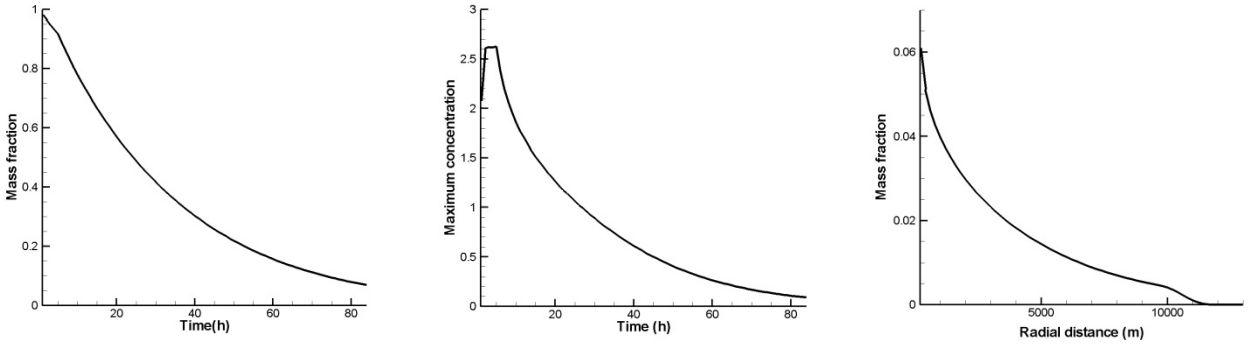


Figure 2. Comparison between SPH simulation results (left) and the experiments (right)

Far-field sediment dispersion simulation



(a) (b) (c)

Figure 3: (a) mass fraction of still suspended sediment (to the total amount of released sediment), (b) maximum sediment concentration (kg/m^3) during the simulation period, (c) mass fraction of sediment deposited on the floor versus radial distance from the source location.

Sediment at seabed disturbed by technical activities has a spectrum of sizes, from a few microns to hundreds of microns. These sediment particles are heavier than water and therefore in the presence of a subsea current, they can be persistent in ocean water and indeed can be carried away to relatively long distances away from the disturbance site. Together with current-related convections, sediment particles also settle downward and deposit on the ocean floor. These deposited sediments

may be re-suspended if the shear stress on the floor is high enough. The distributions of suspended and deposited sediments are the main concern in an environmental impact assessment.

A large-scale sediment dispersion simulation is carried out with SPH in a 150m(H) ×15km(W) domain. Mathematical model basically follows equations (1)-(3). Boundary condition for the sediment concentration at the ocean bottom is generally given by $w_s c_b f_d + M_e f_e$, in which f_d and f_e are probabilities of sediment deposition and bottom erosion, w_s and c_b are sediment settling velocity and sediment concentration right above the bottom. Current flows from left to right and is assumed to follow logarithmic profile with a height-averaged velocity of 0.05m/s. Eddy viscosity and sediment-fluid mixing coefficients are calculated from a mixing length model. Average settling velocity of sediment is $w_s = 10^{-4}\, m/s$ (corresponding to a sediment size of about 10 μm). Sediment source, locating at 2km from the left boundary, follows Gaussian distribution in horizontal direction and exponential distribution in vertical direction and continually releases sediments at a rate of 3.6 tons per hour during 5 hours. The simulation results for suspended and re-deposited sediment after 3.5 days are shown in the figure 3. After 3.5 days, about 93% of the released sediment deposited along the floor up to about 11km from the source location. Maximum concentration reaches to $2.6 g/l$ (or kg/m^3) at 5 hours and gradually reduces to $0.09 g/l$ after 3.5 days.

Conclusions

The motivation behind these works is to provide an assessment of the sediment dispersion problem due to a sediment source disturbance at a certain location. To this end, we find that SPH is a good numerical method that tracks fluid particles and interfaces, including the complexity of the constitutive equations for the fluids and flows with an immersed moving structure. The experimental setup is currently being equipped with sensors in order to measure the force exerted on the blade and to compare it with the one extracted from simulations. The prediction and minimization of this force is also useful in the design of the harvesting equipment. Further work includes the extension to 3D and harvester-scale simulations in order to provide of sediment source estimates for the ocean-scale simulations.

References

[1] Manninen M., Taivassalo V. and Kallio S. (1996) *On the Mixture model for Multiphase Flow*. VTT publications 288, Finland.
[2] Violeau, D. (2012) *Fluid Mechanics and the SPH method*. Oxford University Press, Oxford.
[3] Liu, G. R. and Liu, M. B. (2003) *Smoothed Particle Hydrodynamics — A Meshfree Particle Method*, World Scientific, Singapore.
[4] Huilgol, R. R. (2015) *Fluid Mechanics of Viscoplasticity*, Springer-Verlag, Berlin.
[5] Papanastasiou, T. C. (1987), J. Rheology 31, 385.

Sequential Stochastic Response Surface Method Using Moving Least Squares Based Sparse Grid Scheme for Efficient Reliability Analysis

Amit Kumar Rathi[1], Sudhi Sharma P V[2]and Arunasis Chakraborty[1,a)]

[1]Department of Civil Engineering, Indian Institute of Technology Guwahati, Assam 781039, India

[2]Bentley Systems India Pvt. Ltd., Kolkata, WB 700156, India

[a)]Presenting & corresponding author: arunasis@iitg.ernet.in

ABSTRACT

Present work demonstrates an efficient method for reliability analysis using sequential development of the stochastic response surface in sparse grid framework. Here, stochastic response surface is formed by orthogonal Hermite polynomial basis, whose unknown coefficients are evaluated using moving least squares technique. To construct the response surface, collocation points (as in the conventional stochastic response surface method (SRSM)) are replaced by the sparse grid scheme that reduces the number of function evaluations. Additionally, the sparse grid is populated sequentially based on the optimization process for finding the most probable failure point. After constructing the sequential SRSM, reliability analysis is conducted using importance sampling. Numerical study shows the efficiencies of the proposed sequential SRSM in terms of accuracy and number of time-exhaustive evaluation of the original performance function.

Keywords: Reliability Analysis, Polynomial Chaos Expansion, Moving Least Squares, Hermite Polynomial, Sparse Grid.

Introduction

Surrogate modelling has become an important numerical tool in the recent past for various engineering applications like optimization [1], reliability analysis [2] [3], uncertainty quantification [4]. However, this approximate modelling is not always convergent and may yield significant modelling error [3][5]. One of the method is polynomial chaos expansion (PCE) [2] which can be used for complex scientific and engineering problems. Using PCE, Isukapalli [6] proposed stochastic response surface method (SRSM) to approximate the performance function. Formation of the SRSM is done using the actual function evaluations at Gauss quadrature points (a.k.a. collocation points) to determine the unknown coefficients by regression. The method proves to be robust and stable as it uses regression and orthogonal polynomials (i.e. Hermite polynomial) making it convergent in L^2 sense [2]. Later, Xiu and Karniadakis [4] proposed generalized PCE formulation for solving stochastic differential equations using the Askey polynomial scheme. Application of this method have been studied for various engineering problems like foundation on heterogeneous soil, aircraft joined-wing structure and so on [7][8]. Sudret and Der Kiureghian [7] proposed an application of PCE along with first order reliability method (FORM) and importance sampling technique for solving random field problems. Similar effort has been made by Kameshwar *et al.* [5] to combine PCE with FORM for reliability analysis. They substituted the individual contribution of the random variables in limit state by unidimensional Hermite polynomials which is later solved for reliability index. Gavin and Yau [9] proposed a higher order SRSM (HO-SRSM) using Chebyshev polynomials in which the polynomial order with respect to individual random variable is determined based on the significance of the respective polynomial coefficients.

Apart from using regression analysis which is a widely accepted tool for determining the unknown coefficients in SRSM, advanced techniques like least angle regression (LARS) [10], moving least squares (MLS) [11] and so on, have been adopted for improving the accuracy of the stochastic response surface approximation. However, MLS based SRSM can give inaccurate estimation of failure probability in some cases as shown by Xiong *et al.* [12]. They suggested a double weighted strategy where extra weightage is imposed on the support points near the limit state for better local approximation of SRSM.

Although SRSM is considered to be accurate but it suffers from substantial increase in the number of actual function calls with the increase of number of random variables (i.e. *curse of dimensionality*). This, in turn, makes

the process computationally exhaustive and inefficient. To counter this issue, Blatman and Sudret [10] proposed hyperbolic PCE with LARS. Another technique is adaptive-sparse PCE method [13] where the insignificant terms in the bivariate polynomial expansion are dropped. However, the literatures still lack in limiting the number and the location of support points which affects the overall performance to replicate the original surface.

With this in view, present work suggests an efficient method for reliability assessment for large field problem. A sequential algorithm for SRSM is presented to address the curse of dimensionality. The stochastic response surface is formed using the Hermite polynomial basis function. The unknown coefficients of the response surface are evaluated by MLS technique. For constructing the response surface, collocation points (as in the conventional SRSM) are replaced by the sparse grid scheme which reduces the function evaluations. Additionally, the sparse grid is populated sequentially based on the optimization process for finding the most probable failure point. Once the sequential SRSM is formed, importance sampling is adopted for reliability evaluation.

MLS based SRSM

In stochastic response surface method, the original performance function is replaced by a summation of orthogonal polynomial described in terms of the random variables. Different orthogonal polynomials are described in the literature (e.g. Legendre, Laguerre, Hermite, Chebyshev) for different applications. Off all these polynomials, Hermite polynomial is popular in reliability analysis and stochastic finite element modelling. Mathematically, Hermite polynomial of order o can be expressed as [6]

$$\Gamma_o(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_o}) = e^{\frac{1}{2}\xi^T\xi}(-1)^o \frac{\partial^o e^{-\frac{1}{2}\xi^T\xi}}{\partial\xi_{i_1}\partial\xi_{i_2}\ldots\partial\xi_{i_o}}. \tag{1}$$

where, $\xi = \{\xi_1\ \xi_2\ \ldots\ \xi_n\}^T$ denotes the vector of standard normal random variables. Thus, using polynomial of a predefined order o, the original performance function can be expressed as

$$y(\xi) = \alpha_0 + \sum_{i_1=1}^{n}\alpha_{i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\alpha_{i_1 i_2}\Gamma_2(\xi_{i_1}, \xi_{i_2}) + \ldots + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\cdots\sum_{i_o=1}^{i_{o-1}}\alpha_{i_1 i_2 \ldots i_o}\Gamma_o(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_o}). \tag{2}$$

The unknown coefficients of the aforementioned equation can be denoted by $\mathbf{b} = \{\alpha_0\ \alpha_1\ \ldots\ \alpha_n\ \alpha_{11}\ \alpha_{21}\ \ldots\ \alpha_{nn\ldots n}\}^T$. Thus, rewriting the Eq. 2 in simplified matrix form, one gets

$$y(\xi) = \Xi(\xi)\mathbf{b} \tag{3}$$

where, $\Xi(\xi)$ consists of the Hermite polynomial basis of order $\leq o$. A total of $n_b = \frac{(n+o)!}{n!\,o!}$ coefficients need to be determined in the representation of SRSM. Usually, these unknown coefficients $\mathbf{b}$ are evaluated using regression analysis [6]. However, a global approximation of the original performance function often lead to large error as it fails to capture local variations, if any [14]. To address this problem, moving least square (MLS) based regression is often prescribed in the literature where the unknown coefficients change with locations (i.e. support points). Using this modified regression technique, the unknown coefficients $\mathbf{b}$ can be expressed as follows [15]

$$\mathbf{b} = [\mathbf{\Xi}^{\mathrm{T}}\,\mathbf{W}\,\mathbf{\Xi}]^{-1}\,[\mathbf{\Xi}^{\mathrm{T}}\,\mathbf{W}]\,\mathbf{y} \tag{4}$$

where, the actual values of the performance function evaluated at the support points are expressed by the vector $\mathbf{y}$. Also, each row of the matrix $\mathbf{\Xi}$ represents the polynomial basis corresponding to the location. The weight matrix $\mathbf{W}$ in the above equation consists of weight function w which is given by [15]

$$w(\delta) = \begin{cases} \frac{\bar{w}(\delta)}{\sum_{i=1}^{k}\bar{w}(\delta_i)} & \text{if } \delta \leq r \\ 0 & \text{elsewhere} \end{cases} \tag{5}$$

where,

$$\bar{w}_i(\delta) = \frac{\{(\frac{\delta}{r})^2 + \epsilon\}^{-2} - (1+\epsilon)^{-2}}{\epsilon^{-2} - (1+\epsilon)^{-2}} \tag{6}$$

In the above equation, δ represents the Euclidian distance between the respective support point, r is the influence radius of the weight function and ϵ is adopted as 10^{-5} [15] [14]. Although, this evolving regression technique improves the performance of the meta model significantly, it still suffers computational challenges in problems with large dimensions and multiple optima. Besides this, computational cost for problem with large dimension

has remained a major challenge to the designers. Thus, there is a constant demand for a more efficient technique for reliability estimation that involves less functional evaluations and subsequent computational cost without compromising with the quality of the end results. With this in view, present study aims to demonstrate sequential development of stochastic response method where the support points are generated using sparse grid technique.

Proposed Sequential SRSM with Sparse Grid Scheme

In this section, the details of the proposed sequential stochastic response surface using MLS based PCE in sparse grid scheme is presented.

Sequential SRSM

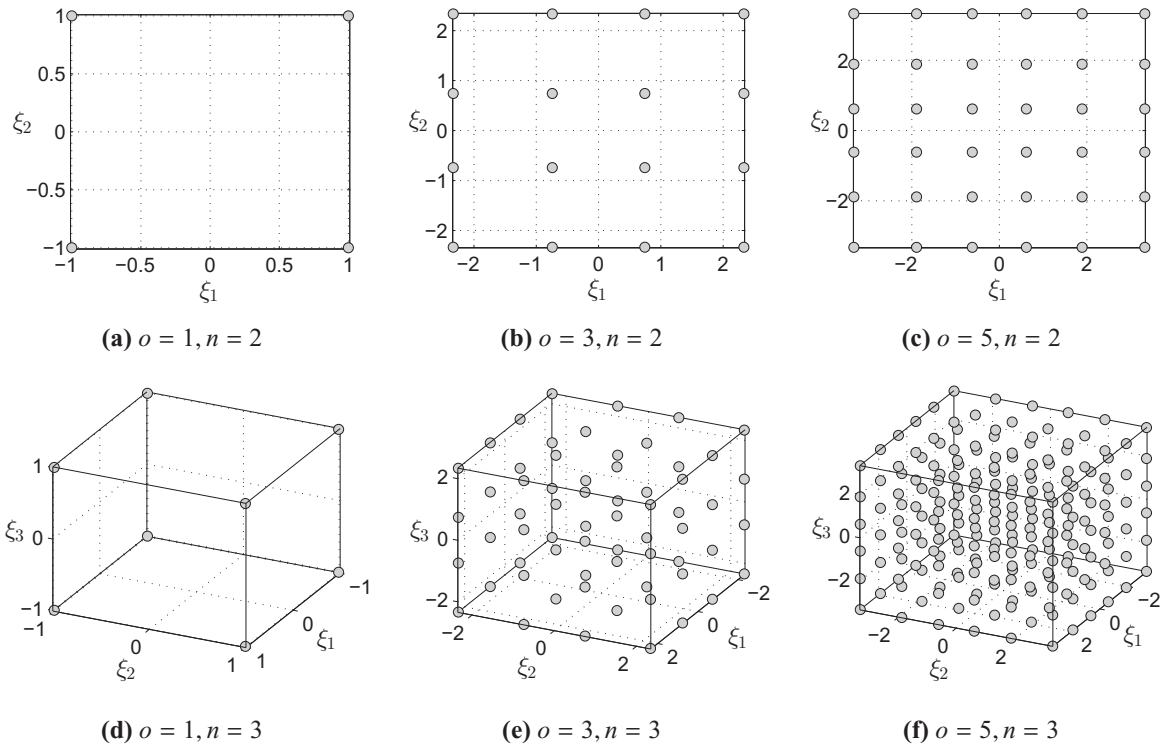
As in the Eq. 2, stochastic response surface is constructed by Hermite polynomial basis with unknown coefficients. To evaluate these coefficients, support points are generated using different techniques namely collocation method, Latin hypercube design, monomial cubature rule among many others [11]. The location of these points are known prior to the determination of the coefficients for constructing the polynomial basis matrix $\boldsymbol{\Xi}$. Hence, the number of support points n_e should be at least equal to the number of unknown coefficient (i.e. n_b). Thus, the support points generation scheme can be dense or sparse depending on various issues like number of random variables n and order of the polynomial o. In case of dense generation, computation cost rises making SRSM less effective whereas the sparse population of support points might lead to inaccurate approximation. Additionally, for an ideal situation in reliability analysis, more support points are required in the vicinity of the limit state (i.e. $y(\mathbf{x}) = 0$). As these points account for better approximation and accuracy in estimation of probability of failure [12]. Therefore, customizing the number of support points n_e in a single go based on o, n etc. becomes a difficult task. To overcome this problem, the present study uses an iterative scheme where the support points are generated only in the vicinity where it is required. This is done by optimizing the Gaussian space of the approximated surface to find the most probable failure point (or design point) as

$$\begin{aligned} &\text{Find :} && \xi \\ &\text{Minimize :} && |\xi|_2 \\ &\text{Subjected to :} && \tilde{y}(\mathbf{x}) = 0. \end{aligned} \tag{7}$$

Above optimization is executed over the approximated surface [i.e. $\tilde{y}(\,.\,)$] which imposes no restriction on the choice of the optimization tool. Hence, different searching tools like gradient based methods, genetic algorithm can be adopted to perform the constrained optimization. Thus, the accuracy of the optimization largely depends on the accuracy of the meta modelling. Although, the accuracy of the meta model increases with number of support points, a tradeoff between the number of points and modelling error is adopted to optimize computational cost. In this study, sequential quadratic programming (SQP) inbuilt in MATLAB® [16] is adopted for solving the Eq. 7 to evaluate the design point ξ^* in the Gaussian space.

From the above discussion, it is evident that the ideal condition requires more points in the failure region and less points elsewhere. This, in turn, improves the efficiency of the reliability method by limiting the number of function calls in order to reduce cost of computation. To explain this phenomenon, Fig. 1 demonstrates the points generated by the full grid formation using the collocation method. These points almost uniformly cover the domain with increase in the order o of PCE, especially the domain with high probability [11]. This might lead to inclusion of points that are insignificant for estimation of probability of failure and lead to excessive computational burden. Whereas the points generated by the sparse grid scheme are selective which is explained later.

Thus, to minimize the function calls (in other words, the number of support points), the proposed method initiates at a predefined point. Without loss of generality, this initial design point is considered as the mean values of the random variables (i.e. $\mu_{\mathbf{X}}$). The support points are allocated around this design point, preferably with the extent to incorporate the failure region (i.e. $y(\mathbf{x}) \leq 0$). PCE is constructed using Eq. 2 of order o for n random variables. In this context, the number of support points must be $\geq n_b$ for solving the PCE based approximation by MLS technique as explained in the Eq. 4. After constructing the approximated surface $\tilde{y}(\mathbf{x})$, constrained optimization problem as explained in Eq. 7 is solved to identify the failure region and the corresponding new optima. This new deign point is further used for generating more support points as explained earlier for the following iteration. In order to generate more support points in the area near the limit state, the spatial extent of the generated support



(a) $o = 1, n = 2$ **(b)** $o = 3, n = 2$ **(c)** $o = 5, n = 2$

(d) $o = 1, n = 3$ **(e)** $o = 3, n = 3$ **(f)** $o = 5, n = 3$

Figure 1: Collocation points with different order and dimension

points in every iteration *it* is reduced by factor λ ($\lambda < 1.0$). The value of this reduction factor depends on the required convergence speed and the accuracy to be attained. After every iteration, convergence of the solution is checked by $|\tilde{y}(\mathbf{x}^*)_{it-1} - \tilde{y}(\mathbf{x}^*)_{it}| \leq \check{e}_1$ and $|\mathbf{x}^*_{it-1} - \mathbf{x}^*_{it}| \leq \check{e}_2$, where $\check{e}_1$ and $\check{e}_2$ are the permissible errors of order, typically in the range of $10^{-2} - 10^{-3}$. Once the convergence is achieved, reliability analysis is conducted using importance sampling method as explained later in this section. The proposed sequential SRSM provides the information of the most probable failure point $\mathbf{x}^*$ and the probability of failure p_f associated with it.

Sparse Grid Scheme

The proposed method employs sparse grid scheme where support points are judiciously selected from the full grid. The sparse grid formation follows Smolyak's algorithm which includes the points from the lower product grids [17] and is controlled by a factor *l* such that

$$SG_l = \sum_{\sum_{i=1}^{m} i_j \leq l+m-1} (\Delta^{i_1} \otimes \Delta^{i_2} \otimes \cdots \otimes \Delta^{i_m})(y) \tag{8}$$

where, this factor *l* is the level of the sparse grid scheme. In Eq. 8, Δ represents the unidimensional difference quadrature term defined in the unit space i.e. [0, 1]. Present study uses an equidistant sparse grid scheme as proposed by Clenshaw and Curtis [18]. The number of points generated by this scheme in unidimensional direction is given by [19]

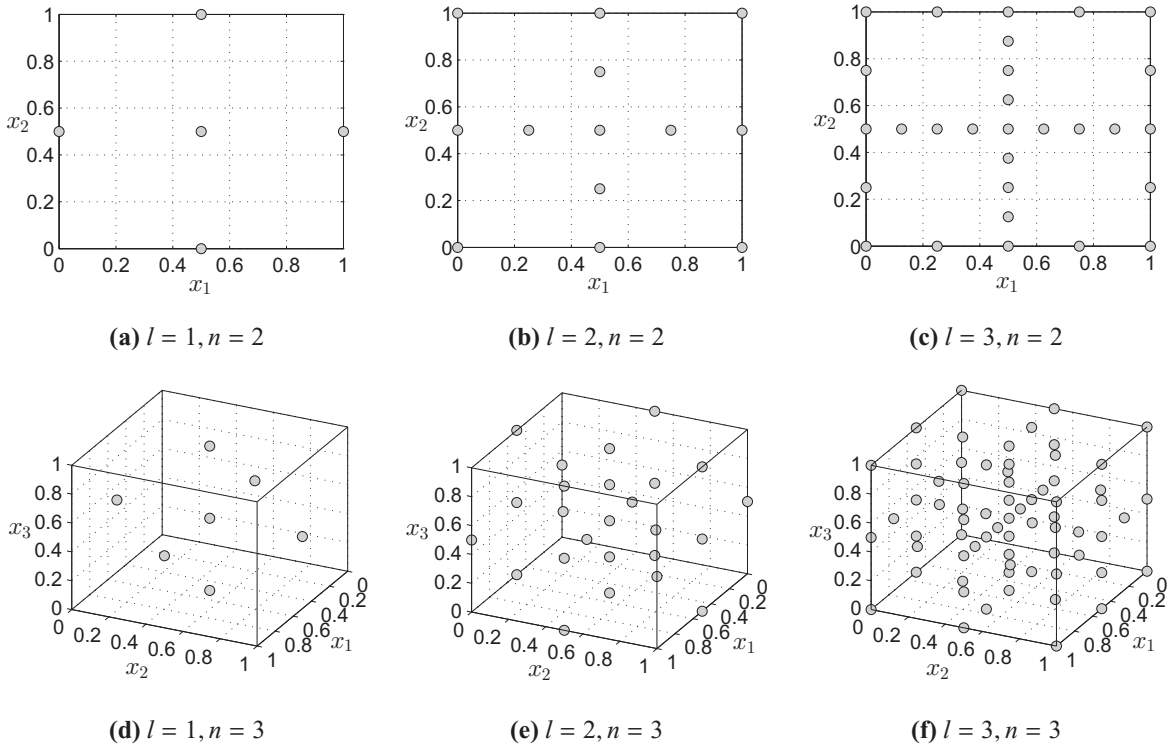
$$n_{c,i} = \begin{cases} 1 & \text{if } i = 1 \\ 2^{i-1} + 1 & \text{if } i > 1 \end{cases} \tag{9}$$

where, *i* denotes a positive integer like 1, 2, 3, …. The coordinates of these points in the unit space is obtained from [19]

$$x_k^{i,j} = \begin{cases} 0.5 & \text{for } j = 1 \text{ if } n_{c,i} = 1 \\ \frac{j-1}{n_{c,i}-1} & \text{for } j = 1, 2, \ldots, n_{c,i} \text{ if } n_{c,i} > 1 \end{cases} \tag{10}$$

where, $x_k^{i,j}$ represents the set of coordinates for the k^{th} random variable. Fig. 2 shows the Clenshaw-Curtis sparse grid generated for various *l* values. In contrary to the collocation points which is a full grid scheme as shown in

Fig. 1, sparse grid scheme fills the domain (in this case it is a unit space) non-uniformly. Thus, creating large voids between adjacent points. The proposed sequential method in this study attempts to fill such voids in the vicinity of critical regions of the limit state only.



(a) $l = 1, n = 2$ **(b)** $l = 2, n = 2$ **(c)** $l = 3, n = 2$

(d) $l = 1, n = 3$ **(e)** $l = 2, n = 3$ **(f)** $l = 3, n = 3$

Figure 2: Support points generated by Clenshaw-Curtis sparse grid scheme

Reliability Assessment

After satisfying convergence criteria, reliability analysis is conducted for estimating the probability of failure p_f. Based on the support points generated sequentially in the iterative manner, approximate surface $\tilde{y}(\mathbf{x})$ is constructed. In this context, it may be noted that support points generated in every iteration along with the corresponding values of the original function are saved to construct the global response surface. Using this global response surface, proposed method estimates the most probable failure point $\mathbf{x}^*$. Here, importance sampling is chosen with sample size (say 10^3 or 10^4) for conducting the reliability assessment in this study. The probability of failure p_f is calculated as [20]

$$p_f \approx \frac{1}{n_s} \sum_{p=1}^{n_s} \frac{\mathscr{S}[\tilde{y}(\mathbf{x}^p) \leq 0] f_{\mathbf{X}}(x^p)}{f^*_{\mathbf{X}}(x^p)} \tag{11}$$

where, n_s is sample size of the simulation. In the above equation, $\mathscr{S}[\,.\,]$ is a discrete indicator function with binary output (i.e. either 0 or 1) based on the satisfaction of the condition stated in the third bracket. Additionally, $f_{\mathbf{X}}(x)$ is the joint probability density function (*pdf*) of the random variables and $f^*_{\mathbf{X}}(x)$ denotes a modified *pdf* applied as the weight function to balance the simulation in the vicinity of the most probable failure point. Readers may refer [20] for further details of this technique. A flowchart of the proposed sequential modelling is shown in Fig. 3 and the application of this method is discussed in the following section.

Numerical Results and Discussion

The proposed sequential SRSM discussed in the previous section is considered here for numerical analysis. Results obtained from this method is compared with other methods (e.g. FORM, SORM, HO-SRSM, SRSM, MCS) to

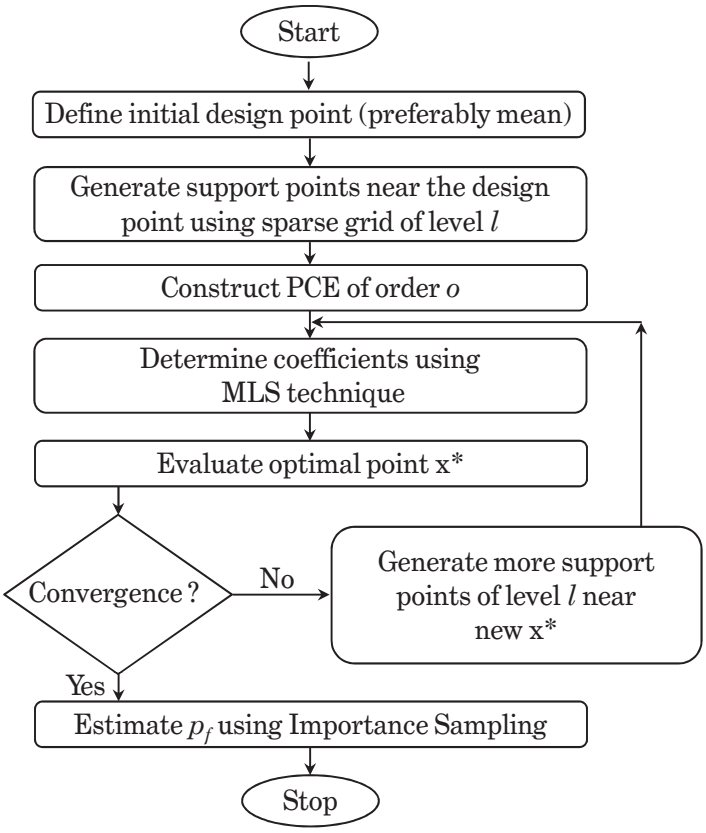
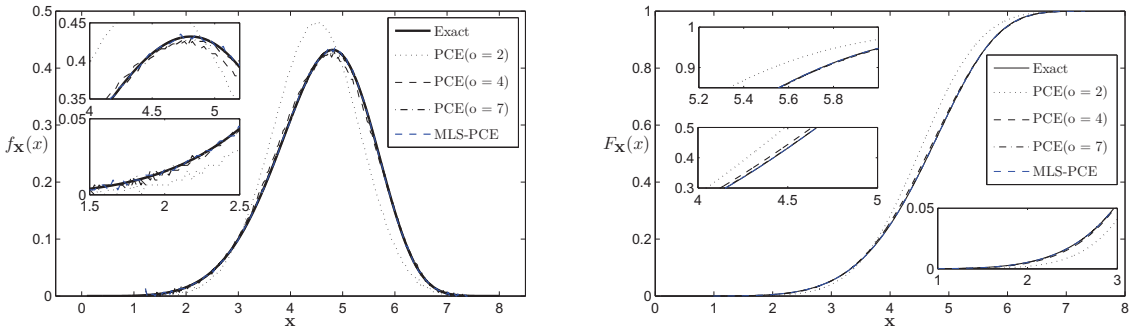


Figure 3: Flowchart of the proposed sequential SRSM



(a) **(b)**

Figure 4: Comparison of (a) *pdf* and (b) CDF of a random variable following Weibull distribution evaluated from different methods

demonstrate its efficiency and accuracy. The order of the proposed MLS based SRSM is fixed at 2 as it is sufficient to accurately capture the nature of the non-normal *pdf*. For this purpose, Weibull distribution is considered to check the adequacy of the order. The random variable is assumed to have mean and variance as 4.60 and 0.85 respectively. For MLS based PCE modelling, the effective range of [0, 8] is subdivided in 8 equidistant segments. Fig 4 shows the *pdf* and CDF of the Weibull distribution using conventional PCE of different order and propose MLS-PCE. As the order increases, conventional PCE matches with the exact value. It is noticed that an order 7 is adequate for

conventional PCE to map the Weibull distribution. However, same accuracy is achieved with MLS-PCE of order 2.

With this performance of the MLS-PCE in hand, proposed sequential SRSM is tested with different benchmark problems. For this purpose, three different problems are considered from literature with different complexities. The performance of the sequential SRSM is tested vis-à-vis with other methods. Finally, a design problem involving nonlinear finite element analysis of a composite plate is presented to demonstrate the superiority of the proposed algorithm for reliability analysis.

Example 1: Franke's Test Surface

In this example, a non-algebraic bivariate performance function is considered which is given by

$$\begin{aligned} y(\mathbf{x}) &= 0.75\exp\{-0.25(9x_1-2)^2-0.25(9x_2-2)^2\}+0.75\exp\left\{-\tfrac{(9x_1-2)^2}{49}-\tfrac{(9x_2-2)^2}{10}\right\} \\ &+0.50\exp\{-0.25(9x_1-7)^2-0.25(9x_2-3)^2\}-0.25\exp\{-(9x_1-4)^2-(9x_2-7)^2\}-0.25 \end{aligned} \tag{12}$$

where, x_1 and x_2 are independent Gaussian random variables with mean $\mu_{\mathbf{X}} = 0.40$ and standard deviation $\sigma_{\mathbf{X}} = 0.10$. The function in Eq. 12 is a modified version of the original Franke's test surface where the limit for failure is set to 0.25 [21]. It is widely considered as a benchmark exercise for testing the interpolation of scattered data. Fig. 5a shows the profile of the original surface which includes two humps (i.e. maxima) and a crater (i.e. minima).

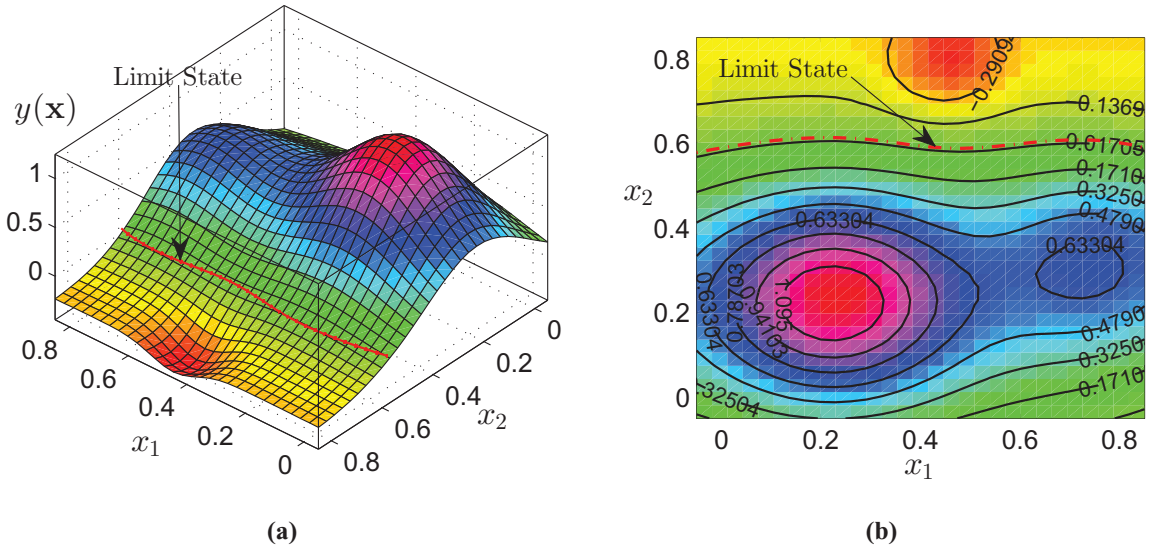


Figure 5: (a) Surface plot and (b) contour plot of the Franke's test surface with limit state

Now, to demonstrate the sequential response surface, Franke's test surface is simplified by cutting the surface using an imaginary plane through the mean value of the random variable x_1 as shown in Fig. 6. This, in fact, reduces the test surface to a 1D function of the random variable x_2. As stated in the flowchart (Fig. 3), the process commences from the mean of the random variable (i.e. μ_{x_2}) which is assumed as the initial design point. The support points using Clenshaw-Curtis sparse grid of level l are generated around this design point. Lower and upper bounds of the support points are adopted using a prior guess which is sufficient enough to accommodate the limit state condition (i.e. $y(\mathbf{x}) = 0$). The sequential SRSM employs Hermite polynomials which are in Gaussian space. Hence, the limits for the random variable x_2 in the standard normal space is considered to be $[-5\ 5]$. Based on the sparse grid level $l = 2$, three equidistant support points are generated in the first iteration such that the points are placed on the limits and the mean as shown in Fig. 6a. In this context, it may be noted that support points for ξ_2 are generated in the standard normal space and converted into its original space defined by x_2. Using these three points, an approximated surface with PCE of order $o = 2$ is constructed using MLS technique as discussed earlier. Eq. 7 is optimized for determining the next optimal location (i.e. new design point) for generating more support points. In

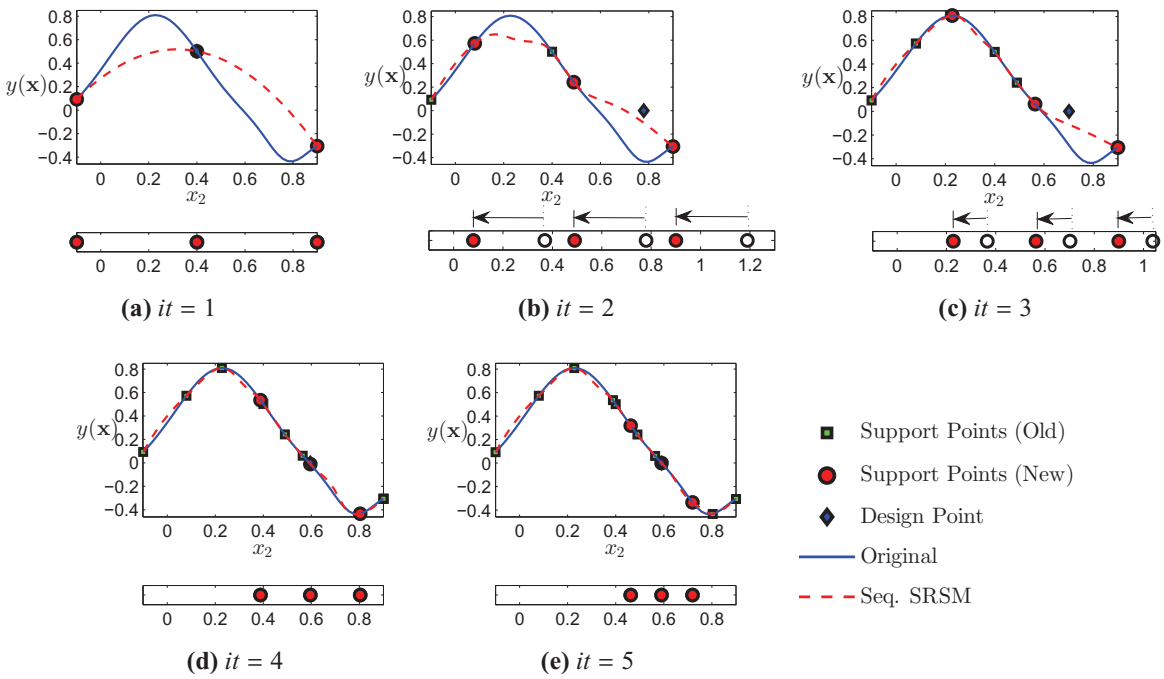


Figure 6: Sequential generation of the support points using the sparse grid scheme

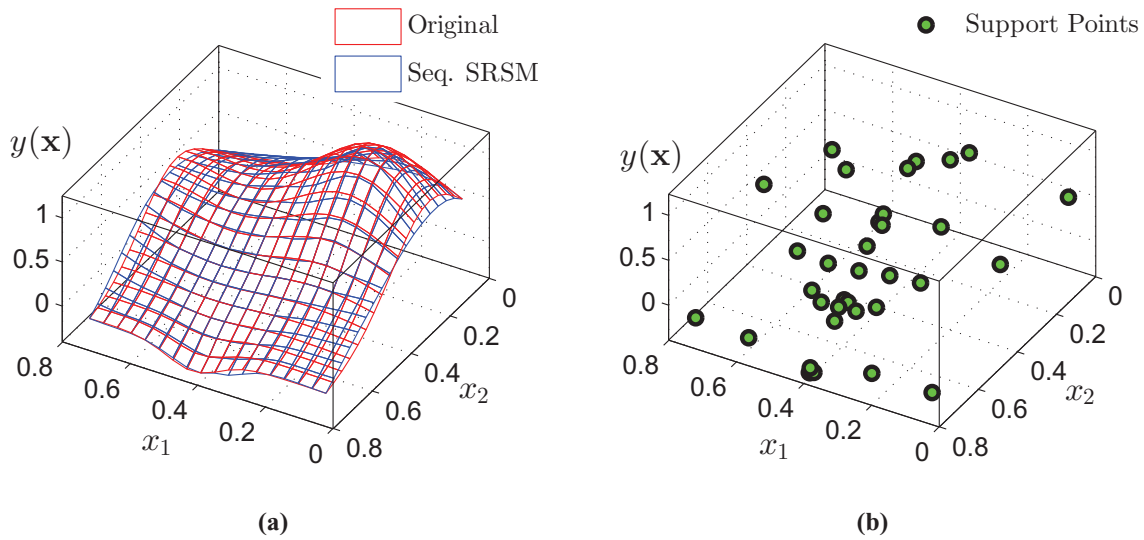


Figure 7: Plot of the approximate surface from (a) sequential SRSM and (b) the support points generated by the sparse grid scheme

the following iteration $it = 2$, additional support points are generated around the new design point using $l = 2$. Here, the extent (i.e. difference between the bounds) of the new support points is reduced from the previous one by a factor $\lambda < 1$ so that more points are segregated near the design point. In this case, $\lambda = 0.82$ is adopted which is observed to give quick convergence. The positions of these new support points are shown in Fig. 6b. It can be observed that the position of one support point lies beyond the domain $[-5\ 5]$. Therefore, the position of all new support points are shifted uniformly to fit within the given limits. This uniform shifting of the positions is done to maintain the symmetry of the new support points. However, if the initial guess on bounds does not contain the

Table 1: Comparison of the probability of failure for the Franke's function

Method	Order	p_f	Function Calls	Error (%)	Remark
MCS	-	0.02459	10^6	-	-
FORM	-	0.02714	36	10.37	-
SORM	-	0.02386	56	2.97	-
SRSM	4	0.03489	25	41.89	-
	6	0.03142	49	27.78	-
	8	0.02968	81	20.70	-
	10	0.02812	121	14.36	-
HO-SRSM	-	0.02647	70	7.65	$o = [6, 7]$
Seq. SRSM	2	0.02446	36	0.53	$l = 2^{(for\ it = 1,2)}$ and $l = 1^{(for\ it = 3,4)}$

limit state [i.e. $y(\mathbf{x}) = 0$], the bounds may be suitably readjusted to incorporate the failure region. Once the position of the new support points are shifted as shown in Fig. 6b, Eq. 4 is solved to determine the unknown coefficients at these positions. The approximated surface is improved using the old as well as the new support points. Again the optimization is executed to determine the next location for $it = 3$. In every successive iteration, new support points are accommodated by further narrowing their extent by reducing the factor λ from $it = 3$ onwards. This eventually helps is concentrating the support points at the failure region for fast convergence. In this case, the convergence is achieved in 5 iterations requiring 15 function calls. However, only 13 actual function calls are executed because in $it = 2$ and 3 have two common points. Fig. 6e shows the performance function (i.e. original and proposed) after convergence.

Using this sequential development of response surface, Franke's function is modelled with two different sparse grid levels. Iteration 1 and 2 used $l = 2$ while all other it used $l = 1$ until convergence. This leads to an total of 4 iterations and 36 actual function calls. Fig. 7 shows the approximation achieved and the support points generated in the proposed sequential SRSM for the bivariate Franke's test surface. The example is also solved by different methods for reliability estimation and the results from these methods are summarized in Table 1. In this study, MCS is assumed to be most accurate estimation of probability of failure which gives $p_f = 0.02459$ with one million samples (i.e. $n_s = 10^6$). The solution using FORM and SORM converges after 36 and 56 function calls and p_f estimated from both these methods are 0.02714 and 0.02386 with 10.37% and 2.97% error, respectively. Additionally, it was observed that the initial choice of design point in FORM and SORM (e.g. [0.2 0.2]) may lead to difficulty in achieving convergence and eventually yields erroneous results. Conventional SRSM is executed with PCE of order 4, 6, 8 and 10 for both the variables. It requires 25 to 121 function calls for estimating p_f with error 14.36–41.89% as shown in the Table 1. HO-SRSM [9] solves the limit state with the order $o = 6$ and 7 for random variables x_1 and x_2, respectively which gives $p_f = 0.02647$ (i.e. 7.65% error) with 70 function calls. Almost half number of function calls are demanded to perform the proposed sequential SRSM to get $p_f = 0.02446$ which is fairly accurate as the error is 0.53%.

Example 2: Fortini's Clutch

Next example in this study is Fortini's clutch assembly which consists of a hub and four roller bearings placed in a cage as shown in Fig. 8. Its application for reliability methods is discussed by Lee and Kwak [22]. The clutch is designed for the overturning based on the contact angle θ as shown in the Fig. 8. This angle θ is formed by a vertical axis passing through center of hub and a line connecting the centers of the opposite roller bearings and the hub center. Thus, the limit state can be defined as

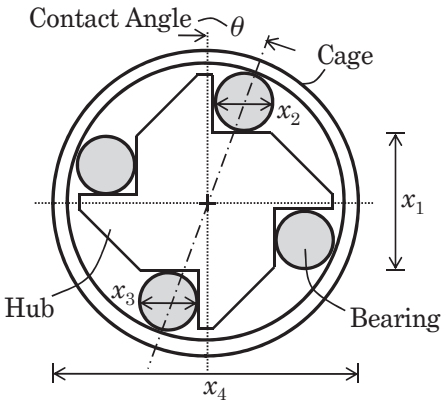


Figure 8: Fortini's clutch assembly

$$y(\mathbf{x}) = \underbrace{\arccos\left[\frac{x_1 + \frac{1}{2}(x_2 + x_3)}{x_4 - \frac{1}{2}(x_2 + x_3)}\right]}_{\theta} - 0.08726 \tag{13}$$

where, x_1, x_2, x_3 and x_4 are independent random variables with statistical properties mentioned in Table 2. In order

Table 2: Statistical properties of the variables in Fortini's clutch assembly

Random variables	Mean (μ_X)	Standard deviation (σ_X)	Distribution
x_1(mm)	55.29	7.93×10^{-2}	Beta
x_2(mm)	22.86	4.30×10^{-3}	Gaussian
x_3(mm)	22.86	4.30×10^{-3}	Gaussian
x_4(mm)	101.60	7.93×10^{-2}	Rayleigh

to avoid overturning of the clutch and operate smoothly, the contact angle θ must lie within 5° (i.e. 0.08726 radian). Table 3 summarizes the probability of failure, error and number of function calls obtained from various methods. The gradient based methods (i.e. FORM and SORM) diverge [22] and are unable to estimate the probability of failure. MCS estimates the $p_f = 0.00130$ with a sample size of 10^5. For this example, conventional SRSM computes probability of failure using orders $o = 3$, 5 and 7 for all random variables which yields 0.00116, 0.00120 and 0.00118, respectively. It suffers from inaccuracies of 7.69% to 10.77% in spite of using significant number of support points (i.e. $n_e = 256$, 1296 and 4096, respectively). The problem is further solved using HO-SRSM where the order of the polynomials in Eq. 13 are determined to be 5, 2, 2 and 5 respectively. This consumes 192 support points to give $p_f = 0.00094$ of error nearly 28%. Finally, p_f is estimated using the proposed method for different values of sparse grid level l as shown in Table 3. A mixed usage of l is demonstrated where the initial few iterations adopt higher value of l as compared to the later iterations. Four cases are shown in the Table 3, including three cases where $l = 2$ is considered for initial few iterations and then, level l is reduced to 1 until convergence is achieved. From this table, it may be noticed that the accuracy decreases with the decrease in level of sparse grid. It is obvious as there are less number of points in lower level over the effective range leading to inaccurate meta modelling. However, as the level is increased (i.e. more support points), algorithm demands more computational cost which may impose serious restrictions for large problems. Thus, there must be a tradeoff between the level and the accuracy that varies from problem to problem and demands an intermittency criteria to choose an optimal level for the specific problem. In this context, $l = 2$ in first two iterations followed by $l = 1$ in successive iterations is found to produce optimal results.

Table 3: Estimation of probability of failure p_f using MCS and the proposed sequential SRSM for different sparse grid levels l

Method	Order	p_f	Error (%)	Function Calls	Remark
MCS	-	0.00130	-	10^5	-
FORM	-	-	-	-	-
SORM	-	-	-	-	-
SRSM	3	0.00116	10.77	256	-
	5	0.00120	7.69	1296	-
	7	0.00118	9.23	4096	-
HO-SRSM	-	0.00094	27.77	192	o = [5, 2 , 2, 5]
Seq. SRSM	2	0.00125	3.85	164	$l = 2^{(for\ it=1,...,4)}$
	2	0.00130	0.00	132	$l = 2^{(for\ it=1,2,3)}$; $l = 1^{(for\ it=4)}$
	2	0.00124	4.62	100	$l = 2^{(for\ it=1,2)}$; $l = 1^{(for\ it=3,4,5)}$
	2	0.00116	10.66	77	$l = 2^{(for\ it=1)}$; $l = 1^{(for\ it=2,...,5)}$

Example 3: Non-differentiable Function

In this example, a hypothetical limit state is examined where it is expressed as [21]

$$y(\mathbf{x}) = 35 - \sum_{i=1}^{2} x_i^2 - \sum_{j=3}^{6} x_j - \frac{x_7 x_8 x_9}{\max(1, x_{10})} \tag{14}$$

In the above limit state, max(.) gives the largest value among the set in the first bracket. This leads to non-

Table 4: Random variables in the non-differentiable function example

Random variables	Mean (μ_X)	Standard deviation (σ_X)	*pdf*
x_1, x_2	-0.200	1.200	Gaussian
x_3	2.500	0.400	Gaussian
x_4, x_5, x_6	2.500	1.400	Gaussian
x_7	1.000	1.000	Gaussian
x_8	1.230	0.350	Gaussian
x_9	0.980	0.023	Gaussian
x_{10}	2.000	1.000	Gaussian

differential nature of the performance function that restricts the use of any gradient based reliability analysis (e.g. FORM, SORM) [21]. Thus, for this case, the random variable x_{10} is responsible for discontinuity in the limit state. Table 4 shows the statistical properties of the five uncorrelated random variables in this case. For checking the accuracy, results from MCS with 10^6 simulations is presented in Table 5. Conventional SRSM is performed with order o = 2 and 3 for all the random variables which demands 59046 and 1048576 support points, respectively. This makes the process very expensive as the later (i.e. SRSM with o = 3) requires more function calls than MCS. Besides constant order of PCE for all the variables as in the fist two cases, SRSM with variable orders are also tried

Table 5: Comparison of the probability of failure for the non-differentiable function example

Method	Order	p_f	Function Calls	Error (%)	Remark
MCS	-	0.000396	10^6	-	-
FORM	-	-	-	-	-
SORM	-	-	-	-	-
SRSM	2	0.000422	59046	6.57	-
	3	0.000417	1048576	5.30	-
	-	0.000400	19440	1.01	$o = [2, 2, 1, 1, 1, 1, 2, 2, 2, 4]$
HO-SRSM	-	0.000430	1804	8.59	$o = [2, \ldots, 2, 4]$
Seq. SRSM	2	0.000387	504	2.27	$l = 2^{(for\ it = 1,2)}$ and $l = 1^{(for\ it = 3,4)}$

and tabulated above. Different orders of these PCE for ten random variables are shown in the third case of SRSM. The result obtained in this case with 19440 function calls has 1.01% error. Although the result is fairly accurate, the computation cost is significantly high. HO-SRSM is further used to calculate the probability of failure which is estimated to be 0.000430 (i.e. 8.59%) with 1804 support points, consuming relatively less computation cost. This results in huge reduction in n_e and enhanced accuracy as compared to previously discussed case of $o = 2$ and 3. The observation clearly indicates that in full grid schemes, the unnecessary support points might reduce its efficiency. Application of sequential SRSM further reduces the computational effort to 504 function calls which is nearly 72% reduction from that in HO-SRSM. Additionally, the accuracy of the proposed method is also well within acceptable limits.

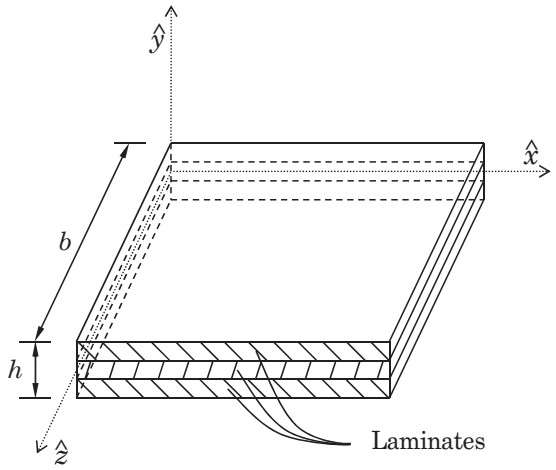


Figure 9: Composite plate

Example 4: Geometrically Nonlinear Composite Plate

Finally, the reliability based design of a carbon-epoxy composite plate is carried out to study the performance of the proposed sequential SRSM. The composite plate, as shown in Fig. 9, consists of laminates stacked in proper

sequence with different orientations. In the present work, geometrically nonlinear composite plate is studied for reliability analysis where the properties of the laminates are adopted to be similar with identical thickness. The laminates in the plate are analyzed using first order shear deformation theory (FSDT) which is based on the assumption that transverse normal is allowed to rotate. This, helps to include transverse shear strains in equilibrium equation and thus, the displacement field is given as [23]

$$\left.\begin{array}{ll} u(\hat{x},\hat{y},\hat{z}) & = u_0(\hat{x},\hat{y}) + \hat{z}\phi_{\hat{x}}(\hat{x},\hat{y}) \\ v(\hat{x},\hat{y},\hat{z}) & = v_0(\hat{x},\hat{y}) + \hat{z}\phi_{\hat{y}}(\hat{x},\hat{y}) \\ w(\hat{x},\hat{y},\hat{z}) & = w_0(\hat{x},\hat{y}) \end{array}\right\}. \tag{15}$$

In the above equation, mid-plane displacements u_0, v_0 and w_0 are associated to $\hat{x}$, $\hat{y}$ and $\hat{z}$ directions, respectively. The rotation with respect to transverse normal denoted by $\phi_{\hat{x}}$ is about $\hat{y}$ and similarly, $\phi_{\hat{y}}$ is about $\hat{x}$. Using thin plate condition where rotations are determined by slopes of the transverse deflection and applying the von-Karman assumptions in Eq. 15, yields [23]

$$\left[\begin{array}{c} \varepsilon_{\hat{x}\hat{x}} \\ \varepsilon_{\hat{y}\hat{y}} \\ \gamma_{\hat{x}\hat{y}} \end{array}\right] = \left[\begin{array}{c} \varepsilon^m_{\hat{x}\hat{x}} \\ \varepsilon^m_{\hat{y}\hat{y}} \\ \gamma^m_{\hat{x}\hat{y}} \end{array}\right] + \hat{z}\left[\begin{array}{c} \varepsilon^f_{\hat{x}\hat{x}} \\ \varepsilon^f_{\hat{y}\hat{y}} \\ \gamma^f_{\hat{x}\hat{y}} \end{array}\right] \tag{16}$$

$$\gamma = \left[\begin{array}{c} \gamma_{\hat{x}\hat{z}} \\ \gamma_{\hat{y}\hat{z}} \end{array}\right] \tag{17}$$

where, $\gamma_{\hat{x}\hat{z}}$ and $\gamma_{\hat{y}\hat{z}}$ represents shear strains. In Eq. 16, the superscripts m and f denotes membrane and flexural components, respectively. Geometric nonlinearity caused by large deformations gives rise to additional higher order terms in the strain field. This modifies the membrane strain to [23]

$$\left.\begin{array}{ll} \varepsilon^m_{\hat{x}\hat{x}} = & u_{0,\hat{x}} + \frac{1}{2}(w_{0,\hat{x}})^2 \\ \varepsilon^m_{\hat{y}\hat{y}} = & v_{0,\hat{y}} + \frac{1}{2}(w_{0,\hat{y}})^2 \\ \varepsilon^m_{\hat{x}\hat{y}} = & u_{0,\hat{y}} + v_{0,\hat{x}} + w_{0,\hat{x}}\, w_{0,\hat{y}} \end{array}\right\}. \tag{18}$$

Thus, the constitutive equation for a laminate is expressed as [23]

$$\left\{\begin{array}{c} \sigma_1 \\ \sigma_2 \\ \tau_{12} \\ \tau_{23} \\ \tau_{13} \end{array}\right\} = \left[\begin{array}{ccccc} Q_{11} & Q_{12} & 0 & 0 & 0 \\ Q_{21} & Q_{22} & 0 & 0 & 0 \\ 0 & 0 & Q_{66} & 0 & 0 \\ 0 & 0 & 0 & Q_{44} & 0 \\ 0 & 0 & 0 & 0 & Q_{55} \end{array}\right] \left\{\begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \gamma_{12} \\ \gamma_{23} \\ \gamma_{13} \end{array}\right\} \tag{19}$$

where, Q_{ij} represents the plane stress reduced stiffness coefficients. In Eq. 19, these coefficients are defined in the material axes of the laminate which can be transformed into the global axes by

$$\breve{Q} = [T][Q][T]' \tag{20}$$

where, $[T]$ is the transformation matrix. The orthotropic laminate, after this transformation, acts as anisotropic which also include coupling terms. Thus, the constitutive equation for a composite plate by adding the laminates to get a equivalent single layer is given as

$$\left[\begin{array}{c} [N] \\ [M] \end{array}\right] = \left[\begin{array}{cc} [A] & [B] \\ [B] & [D] \end{array}\right] \left[\begin{array}{c} \varepsilon^m \\ \varepsilon^f \end{array}\right] \tag{21}$$

where, $[N]$ and $[M]$ are the force and moment resultants, respectively. Also, $[A]$ is the extensional stiffness matrix, $[B]$ represents the bending-extension coupling stiffness matrix and $[D]$ denotes the bending stiffness matrix.

Here, total Lagrangian incremental formulation is adopted for developing the geometric nonlinear equilibrium equation [23]. Hence, the virtual work equation of undeformed plate with volume V and area a is given as

$$\int_V (d\varepsilon^T\, \sigma\, dV) = \int_V (\rho\, du^T\, g\, dV) + \int_A (du^T\, p\, da) \tag{22}$$

where, $d\varepsilon$ is the virtual Green strain vector, du gives the virtual displacement, σ represents Piola-Kirchoff stress vector, ρ is the mass density of the material, g is the body force per unit mass and p denotes the pressure applied on the plate. The displacement field can be discretized using finite elements which is given by

$$\bar{u} = \sum_{i=1}^{n_n} [\Omega_i I] q_i \tag{23}$$

where, I is the identity matrix, Ω_i denotes the shape function for any arbitrary node i, $q_i = [u_i\ v_i\ w_i\ \phi_{xi}\ \phi_{yi}]^T$ gives the nodal displacements. Extending the above equation for the complete thickness yields the elemental nonlinear equilibrium equation as

$$\psi(q) = \sum_{i=1}^{n_e} \left[\int_a ([\check{B}]'\ \bar{\sigma}\ da) - \{\bar{P}\}_e \right] = 0 \tag{24}$$

where, $\check{B}$ denotes the strain-displacement matrix with nonlinear terms, $\psi(q)$ represents the summation of external and internal forces, $\bar{\sigma}$ gives the force and moment resultant and $\{\bar{P}\}_e$ is the total external force at the element level. The nonlinear equilibrium equation can be evaluated by solving Eq. 24 with respect to the displacement vector q. This yields the expression in the terms of displacement, force and stiffness. In the present study, Newton-Raphson iterative technique [23] is used for solving this nonlinear equilibrium equation.

Using this finite element model of a geometrically nonlinear plate, fragility analysis is carried out for the reliability based design. Table 6 shows the statistical properties of the random variables used in this example. In this study, modified Tsai-Hill failure [24] criterion is adopted to define the limit state. This criterion is an extension of the von-Mises distortion energy theory [25]. The limit state based on the modified Tsai-Hill failure index $\mathscr{F}$ is expressed as

$$y(\mathbf{x}) = 1 - \underbrace{\left\{ \left(\frac{\sigma_{11}}{X}\right)^2 + \left(\frac{\sigma_{22}}{Y}\right)^2 - \left(\frac{\sigma_{11}\sigma_{22}}{X^2}\right) + \left(\frac{\sigma_{12}}{T_{12}}\right)^2 \right\}}_{\mathscr{F}} \tag{25}$$

where, σ_{11}, σ_{22} and T_{12} are the laminate stresses along the respective material coordinates whereas X and Y are the tensile and compressive strengths. As shown in the Eq. 25, the limit state reflects the condition when the failure index $\mathscr{F}$ of any laminate exceeds unity [24].

Table 6: Statistical parameters and distribution of the random variables in the composite plate

Random Variables	Units	Mean	cov (%)	*pdf*	Parameters	
$\nu_{12} \rightarrow x_1$	-	0.281	7.5	Lognormal	-	-
$G_{12} \rightarrow x_2$	GPa	4.5	8.8	Lognormal	-	-
$X_t \rightarrow x_3$	GPa	2.409	6.7	Lognormal	-	-
$X_c \rightarrow x_4$	GPa	1.148	18.1	Lognormal	-	-
$T_{12} \rightarrow x_5$	GPa	0.083	5	Lognormal	-	-
$E_1 \rightarrow x_6$	GPa	154.9	5.9	Weibull	158.820	21.6
$E_2 \rightarrow x_7$	GPa	8.7	9.5	Weibull	9.055	12.9
$Y_t \rightarrow x_8$	GPa	0.046	20	Weibull	0.050	5.7
$Y_c \rightarrow x_9$	GPa	0.196	15.3	Weibull	0.209	7.7

A square carbon-epoxy composite plate is considered of dimension 1×1 m and thickness 0.010 m. The plate is simply supported in all four sides with uniformly distributed load (UDL = 0.090 MPa) acting downwards. The laminates are placed with orientation $[0°/90°/0°]$. To perform the finite element analysis, quadrilateral nine noded element with a mesh size 8×8 is used. The random variables in this study are elastic modulus E_1 and E_2, shear modulus G_{12}, Poisson's ratio ν_{12} and strength parameters X_t, X_c, Y_t, Y_c and T_{12}. Their statistical parameters and *pdf* types are adopted from Sasikumar *et al.* [24] and mentioned in Table 6. Using these parameters, the reliability of the plate against modified Tsai-Hill failure criterion is evaluated by all the methods described in previous examples and the results are tabulated in Table 7. As usual, the MCS is conducted with 10^4 samples and is considered as the benchmark for further analysis. Gradient based methods (i.e. FORM and SORM) gives satisfactory results with 220 and 751 function calls respectively. The probability of failure estimated for these two cases have error around 3.1% and 2.99%. A marginal improvement in second order method is noticed. However, the quality of the results largely depends on the initial guess which is known drawback of the gradient based techniques. With these results

Table 7: Comparison of the probability of failure for the nonlinear composite plate

Method	Order	p_f	Function Calls	Error (%)	Remark
MCS	-	0.1906	10^4	-	-
FORM	-	0.1965	220	3.10	-
SORM	-	0.1963	751	2.99	-
SRSM	-	0.2187	1280	14.74	o = [1, 1, 1, 1 1, 1, 1, 4, 1]
	-	0.2632	1536	38.09	o = [1, 1, 1, 1 1, 1, 1, 5, 1]
HO-SRSM	-	0.1900	3324	0.31	o = [2, 2, 2, 2 2, 5, 3, 5, 2]
Seq. SRSM	2	0.1975	456	3.64	$l = 2^{(for\ it = 1,2)}$ and $l = 1^{(for\ it = 3,4,5)}$

in hand, SRSM and its modified versions are tried. First, the conventional SRSM is tried with order 2 and 3 that need 19683 and 262144 function calls. These are well above the function calls required for MCS. As each function call needs to solve nonlinear finite element code to check the failure criterion involving significant computation time, this method for reliability estimation of the nonlinear composite plate is not feasible. Moreover, the results from the SRSM have 14.74% and 38.09% error for two different combination of orders of random variables which are not satisfactory at all. Once the performance of conventional SRSM is studied, HO-SRSM is tried with different order. It is found that it converged with $p_f = 0.19$ that has 0.31% error with 3324 function calls. Finally, proposed sequential SRSM is used to study the failure. It is noticed that the proposed method with $l = 2$ in first two iterations followed by $l = 1$ converges after 5 iterations with error less than 3.84%. It requires 456 function calls which is more than FORM but well below that required for SORM and HO-SRSM. This clearly justifies the superiority of the proposed method for actual design problems involving large finite element models.

Summary

An efficient reliability analysis using sequential development of SRSM is demonstrated here. In this process, the proposed MLS based SRSM is formed with Clenshaw-Curtis sparse grid scheme with equidistant support points in each successive iterations. The order of the polynomials and the level of the sparse grid are adjusted in every iteration that offers significant flexibility to optimize computational cost while compromising with the accuracy of the end result. In this context, different optimization tool may be adopted as the proposed imposes no restriction. Using this sequential SRSM, different benchmark problems are solved to demonstrate its performance. The numerical study presented in this paper clearly demonstrates that level of accuracy and computation cost involved in the proposed method. It may be concluded that the proposed sequential SRSM offers appreciable accuracy at an optimal computational cost. Overall, it proves to be an effective tool for reliability analysis for problems with large dimension and other complexities.

With simple modifications, the proposed method can be adopted for problems with multiple performance function and/or multiple failure points and uncertainty quantification. Authors wish to address these issues in their future work.

References

[1] Myers, R. H., Montgomery, D. C. and Anderson-Cook, C. M. (2009) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd edn. John Wiley & Sons, Hoboken, New Jersey, USA.

[2] Ghanem, R. G. and Spanos, P. D. (1991) *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag New York, USA.

[3] Haldar, A. and Mahadevan, S. (2000) *Reliability Assessment using Stochastic Finite Element Analysis*, John Wiley & Sons, Inc., New York, USA.

[4] Xiu, D. and Karniadakis, G. E. (2003) Modeling uncertainty in flow simulations via generalized polynomial chaos, *Journal of Computational Physics* **187** (1), 137 – 167.

[5] Kameshwar, S., Rathi, A. K. and Chakraborty, A. (2012) A Modified Gradient Based Reliability Analysis for Nonlinear Nonalgebraic Limit States Using Polynomial Chaos Expansion, *Proceedings of* 4^{th} *International Congress on Computational Mechanics and Simulation*, Hyderabad, India, December 10–12.

[6] Isukapalli, S. S. (1999) *Uncertainty Analysis of Transport-Transformation Models*, Ph.D. dissertation. Rutgers, The State University of New Jersey.

[7] Sudret, B. and Der Kiureghian, A. (2002) Comparison of Finite Element Reliability Methods, *Probabilistic Engineering Mechanics* **17** (4), 337 – 348.

[8] Choi, S.-K., Grandhi, R. V. and Canfield, R. A. (2004) Structural Reliability under Non-Gaussian Stochastic Behavior, *Computers & Structures* **82** (13–14), 1113 – 1121.

[9] Gavin, H. P. and Yau, S. C. (2008) High-Order Limit State Functions in the Response Surface Method for Structural Reliability Analysis, *Structural Safety*, **30** (2), 162 – 179.

[10] Blatman, G. and Sudret, B. (2010) Reliability Analysis of a Pressurized Water Reactor Vessel using Sparse Polynomial Chaos Expansions, *Reliability and Optimization of Structural Systems*, 9 – 16, CRC Press.

[11] Xiong, F., Chen, W., Xiong, Y. and Yang, S. (2011) Weighted Stochastic Response Surface Method Considering Sample Weights, *Structural and Multidisciplinary Optimization*, **43** (6), 837 – 849.

[12] Xiong, F., Liu, Y., Xiong, Y. and Yang, S. (2012) A Double Weighted Stochastic Response Surface Method for Reliability Analysis, *Journal of Mechanical Science and Technology*, **26** (8), 2573–2580.

[13] Hu, C. and Youn, B. D. (2011) Adaptive-Sparse Polynomial Chaos Expansion for Reliability Analysis and Design of Complex Engineering Systems, *Structural and Multidisciplinary Optimization*, **43** (3), 419–442.

[14] Rathi, A. K. and Chakraborty, A. (2016) Reliability-Based Performance Optimization of TMD for Vibration Control of Structures with Uncertainty in Parameters and Excitation, *Structural Control and Health Monitoring*.

[15] Most, T. and Bucher, C. (2005) A Moving Least Squares weighting function for the Element-free Galerkin Method which almost fulfills essential boundary conditions, *Structural Engineering and Mechanics* **21** (3), 315 – 332.

[16] MATLAB® version 7.13.0.564 (R2011b), The MathWorks Inc., Natick, Massachusetts, USA.

[17] Holtz, M. (2011) Sparse Grid Quadrature, In *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance* (51–76). Springer Berlin Heidelberg, Berlin, Heidelberg.

[18] Clenshaw, C. W. and Curtis, A. R. (1960) A Method for Numerical Integration on an Automatic Computer, *Numerische Mathematik*, **2** (1), 197 – 205.

[19] Klimke, A. and Wohlmuth, B. (2005) Computing Expensive Multivariate Functions of Fuzzy Numbers using Sparse Grids, *Fuzzy Sets and Systems*, **154** (3), 432 – 453.

[20] Melchers, R. E. (1989) Importance Sampling in Structural Systems, *Structural Safety*, **6** (1), 3 – 10.

[21] Chowdhury, R., Rao, B. N. and Prasad, A. M. (2009) High-Dimensional Model Representation for Structural Reliability Analysis, *Communications in Numerical Methods in Engineering*, **25** (4), 301 – 337.

[22] Lee, S. H. and Kwak, B. M. (2006) Response Surface Augmented Moment Method for Efficient Reliability Analysis, *Structural Safety* **28**, 261 – 272.

[23] Reddy, J. N. (2004) *Mechanics of Laminated Composite Plates and Shells: Theory and Analysis*, CRC Press.

[24] Sasikumar, P., Suresh, R. and Gupta, S. (2014) Analysis of CFRP Laminated Plates with Spatially Varying Non-Gaussian Inhomogeneities using SFEM, *Composite Structures* **112**, 308 – 326.

[25] Mohite, P. M. *Composite Materials and Structures*, National Programme on Technology Enhanced Learning (NPTEL), Accessed on: 16 May 2016. `http://www.nptel.ac.in/courses/101104010/`

Designing photonic crystals with complete band gaps

Fei Meng[1,3], Shuo Li[2], Baohua Jia[2] and Xiaodong Huang[1*]

[1]Centre for Innovative Structures and Materials, School of Engineering, RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia
[2]Centre for Micro-Photonics, Faculty of Science, Engineering and Technology, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia
[3]School of Civil Engineering, Central South University, Changsha 410075, China

*Presenting author: s3459321@student.rmit.edu.au
†Corresponding author: huang.xiaodong@rmit.edu.au

Abstract

In order to design photonic crystals with complete band gaps, a topology optimization algorithm is proposed based on finite element analysis and bi-directional evolutionary structural optimization method. The photonic crystals are assumed to be periodically composed of two materials with different electromagnetic property. By introducing discrete design variables and calculating the sensitivity of each element, the BESO algorithm gradually re-distributes the dielectric materials within the unit cell until the photonic crystal have a complete band gap between specified photonic bands. The proposed optimization algorithm is efficient and some innovative designs have been obtained.

Keywords: Topology optimization; complete band gap; finite element analysis (FEA); bi-directional evolutionary structural optimization (BESO).

Introduction

Photonic crystals are micro optical periodic structures in 1, 2 or 3 dimensions. Due to the periodic arrangement of dielectric materials with different electromagnetic properties, photonic crystals will be able to modulate the propagation of light and generate some special functions like photonic band gap[1], negative refraction[2] and slow light[3]. Photonic band gap refers to the property that the propagation of electromagnetic waves within a certain frequency ranges are totally prohibited in the photonic crystal. It is an important and fundamental feature which lays the ground for many utilities, for example wave guide[4, 5] and resonant cavity[5, 6]. It has significant meaning to design photonic crystals with large band gaps which can modulate light signals in a broader frequency range.

For two dimensional photonic crystals, electromagnetic wave can be decomposed to transverse magnetic waves, whose electric field is perpendicular to crystal plane, called TM mode, and transverse electric waves, whose magnetic field is perpendicular to crystal plane, called TE mode. Design and optimization of photonic band gap structures for an independent polarization have been reported in many literatures. These methods include the traditional trial-and-error method based on physical intuitions[7, 8] and advanced topology optimization methods such as genetic algorithms[9, 10], level set method[11], SIMP[12] and BESO[13].

Compared with band gaps for a single polarization, complete band gaps, which can prevent electromagnetic waves of both polarizations, are apparently more meaningful. However, the calculation of complete band gap takes the photonic bands of both TM and TE mode into consideration, which makes the optimization process very low-efficient and laborious. Therefore, only limited results have been reported so far [14-19], and most of them are designed based on physical intuitions. Furthermore, some of them share a similar pattern, like the results in Refs. [15], [17] and [18].

While these designs are attractive, it is important to attempt new methods and algorithms in order to find wider complete band gaps or different topologies. In this paper, a new approach based on bi-directional evolutionary structural optimization (BESO) method is proposed. BESO is a structural optimization method based on finite element analysis (FEA). Its key concept is gradually removing inefficient materials from and adding high efficient materials into the design domain, until the optimal design is achieved[20]. BESO has been successfully applied to the optimization of materials with periodic micro structures, including photonic band gap crystals[13].

In this paper, the finite element method used to calculate the photonic bands is firstly introduced. An objective function is put forward and the corresponding sensitivity analysis is conducted. Then based on the FEA and the sensitivity analysis, a BESO algorithm is established by introducing discrete design variables. Starting from a simple initial design without band gap, BESO evolves the topology of the unit cell step by step until a desired complete band gap emerges and enlarges to its maximum. Finally, several numerical examples are presented to demonstrate the effectiveness and efficiency of the proposed optimization algorithm.

Finite element analysis of photonic crystals

The propagation of light in a photonic crystal is governed by the Maxwell's equations. For 2D photonic crystals, when there is no point source or sink of electric and magnetic fields, the Maxwell equations can be reduced to two decoupled master equations as

$$-\nabla\cdot\left(\nabla E(\mathbf{k},\mathbf{r})\right)=\varepsilon(\mathbf{r})\left(\frac{\omega}{c}\right)^2 E(\mathbf{k},\mathbf{r}) \quad \text{for TM mode} \tag{1a}$$

$$-\nabla\cdot\left(\frac{1}{\varepsilon(\mathbf{r})}\nabla H(\mathbf{k},\mathbf{r})\right)=\left(\frac{\omega}{c}\right)^2 H(\mathbf{k},\mathbf{r}) \quad \text{for TE mode} \tag{1b}$$

where $\mathbf{k} = (k_x, k_y)$ is the wave vector and $\mathbf{r} = (x, y)$ denotes the coordinates. $\varepsilon(\mathbf{r})$ is the dielectric function. $E(\mathbf{k}, \mathbf{r})$ is the electric field, $H(\mathbf{k}, \mathbf{r})$ is the magnetic field, c is the speed of light, and ω is the corresponding eigenfrequency.

Due to the periodicity of the crystal, $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}+\mathbf{R})$, $E(\mathbf{k}, \mathbf{r}) = E(\mathbf{k}, \mathbf{r}+\mathbf{R})$ and $H(\mathbf{k}, \mathbf{r}) = E(\mathbf{k}, \mathbf{r}+\mathbf{R})$, where $\mathbf{R}$ is the lattice translation vector. Based on the Bloch-Floquet theory [21], $E(\mathbf{k}, \mathbf{r})$ and $H(\mathbf{k}, \mathbf{r})$ can be represented by the product of a periodic function and an exponential factor

$$E(\mathbf{k},\mathbf{r})=E(\mathbf{r})\cdot\exp(i\mathbf{k}\cdot\mathbf{r}) \quad \text{for TM mode} \tag{2a}$$

$$H(\mathbf{k},\mathbf{r})=H(\mathbf{r})\cdot\exp(i\mathbf{k}\cdot\mathbf{r}) \quad \text{for TE mode} \tag{2b}$$

Substitute Eqs. 2a and 2b into Eqs. 1a and 1b, the governing equations can be converted to eigenvalue problems within the representative unit cell.

$$-(\nabla+i\mathbf{k})\cdot\left((\nabla+i\mathbf{k})E(\mathbf{r})\right)=\varepsilon\left(\frac{\omega}{c}\right)^2 E(\mathbf{r}) \quad \text{for TM mode} \tag{3a}$$

$$-(\nabla+i\mathbf{k})\cdot\left(\frac{1}{\varepsilon}(\nabla+i\mathbf{k})H(\mathbf{r})\right)=\left(\frac{\omega}{c}\right)^2 H(\mathbf{r}) \quad \text{for TE mode} \tag{3b}$$

For a given wave vector $\mathbf{k} = (k_x, k_y)$, Eqs. 3a and 3b can be solved by finite element method. The weak expressions $F_E(v, E(\mathbf{r}))$ and $F_H(v, H(\mathbf{r}))$ corresponding to TM modes and TE modes respectively are

$$F_E(v, E(\mathbf{r})) = v\frac{\partial^2 E(\mathbf{r})}{\partial x^2} + v\frac{\partial^2 E(\mathbf{r})}{\partial y^2} + v\frac{\partial}{\partial x}(ik_x E(\mathbf{r})) + v\frac{\partial}{\partial y}(ik_y E(\mathbf{r})) + ik_x v\frac{\partial E(\mathbf{r})}{\partial x} + ik_y v\frac{\partial E(\mathbf{r})}{\partial y} - k^2 vE(\mathbf{r}) - \left(\frac{\omega}{c}\right)^2 v\varepsilon E(\mathbf{r}) \tag{4a}$$

$$F_H(v, H(\mathbf{r})) = v\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H(\mathbf{r})}{\partial x}\right) + v\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H(\mathbf{r})}{\partial y}\right) + v\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}ik_x H(\mathbf{r})\right) + v\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}ik_y H(\mathbf{r})\right) + ik_x\frac{1}{\varepsilon}v\frac{\partial H(\mathbf{r})}{\partial x} + ik_y\frac{1}{\varepsilon}v\frac{\partial H(\mathbf{r})}{\partial y} - k^2\frac{1}{\varepsilon}vH(\mathbf{r}) - \left(\frac{\omega}{c}\right)^2 vH(\mathbf{r}) \tag{4b}$$

where v is the test function. By discretizing the unit cell with 4-node square elements, the above eigenvalue problems can be written in the matrix format as

$$\left[\mathbf{K} - \left(\frac{\omega}{c}\right)^2 \mathbf{M}\right]\mathbf{u} = 0 \tag{5}$$

where $\mathbf{K} = \sum \mathbf{K}_e$, $\mathbf{M} = \sum \varepsilon_e \mathbf{M}_e$ and $\mathbf{u} = \mathbf{E}$ for TM modes, while $\mathbf{K} = \sum \frac{1}{\varepsilon_e}\mathbf{K}_e$, $\mathbf{M} = \sum \mathbf{M}_e$ and $\mathbf{u} = \mathbf{H}$ for TE modes. $\mathbf{K}_e$ and $\mathbf{M}_e$ denote the elemental stiffness and mass matrices, ε_e is the relative permittivity of element e.

BESO process

In this paper, our aim is to design photonic crystals with a large complete band gap. The size of the band gap can be measured by the band gap-midgap ratio, which is independent on the size of the photonic crystal and hence more meaningful than the absolute value of the band gap. The position of the band gap is controlled manually by specify the adjacent TM bands (referred as band ω_i^{TM} and band $\omega_{i+1}^{\mathrm{TM}}$) and TE bands (referred as band ω_j^{TE} and band $\omega_{j+1}^{\mathrm{TE}}$). The upper limit of the band gap is the smaller value of $\omega_{i+1}^{\mathrm{TM}}$ and $\omega_{j+1}^{\mathrm{TE}}$, while the lower limit is the larger value of ω_i^{TM} and ω_j^{TE}. Therefore, the objective function $f(\mathbf{X})$ in can be expressed as

$$f(\mathbf{X}) = \frac{\min(\omega_{i+1}^{\mathrm{TM}}(\mathbf{k}), \omega_{j+1}^{\mathrm{TE}}(\mathbf{k})) - \max(\omega_i^{\mathrm{TM}}(\mathbf{k}), \omega_j^{\mathrm{TE}}(\mathbf{k}))}{(\min(\omega_{i+1}^{\mathrm{TM}}(\mathbf{k}), \omega_{j+1}^{\mathrm{TE}}(\mathbf{k})) + \max(\omega_i^{\mathrm{TM}}(\mathbf{k}), \omega_j^{\mathrm{TE}}(\mathbf{k})))/2} \tag{6}$$

where $\mathbf{X} = \{x_1, x_2 \ldots x_n\}$ is the elemental design variable, n is the total number of elements. Each elemental design variable corresponds to its material property, ε_e,, so $\mathbf{X}$ represents the topology of the unit cell. In the optimization process, the evolution of topology is reflected by the change of design variables.

The design variable is constructed by assuming $x_e = 0$ means element e is consist of material 1 which has a low permittivity ε_1, and $x_e = 1$ denotes material 2 with high permittivity ε_2. According to our numerical experience, in order to have a stable and reliable optimization process, x_e is set to be a discrete value between 0 and 1 with a custom step size. Then, the permittivity of this element is interpolated by following functions.

$$\varepsilon(x_e) = \varepsilon_1(1-x_e) + \varepsilon_2 x_e \quad \text{for TM mode} \tag{7a}$$

$$\varepsilon(x_e) = \frac{1}{(1-x_e)/\varepsilon_1 + x_e/\varepsilon_2} \quad \text{for TE mode} \tag{7b}$$

Finally, the optimization problem can be stated as

$$\begin{aligned} &\text{Maximize}: f(\mathbf{X}) \\ &\text{Subject to}: 0 \le x_{\mathrm{e}} \le 1 \end{aligned} \tag{8}$$

The topology of the unit cell is updated iteratively based on the elemental sensitivity numbers, i.e. the relative ranking of elemental sensitivities, which is the derivative of the objective function with regard to a design variable. Based on the objective function, the sensitivity of element *e* can be expressed as

$$\alpha_e = \frac{\partial f(\mathbf{X})}{\partial x_e} = \frac{\omega_{bot}\dfrac{\partial \omega_{top}}{\partial x_e} - \omega_{top}\dfrac{\partial \omega_{bot}}{\partial x_e}}{(\omega_{top}+\omega_{bot})^2/4} \tag{9}$$

where $\omega_{top} = \min(\omega_{i+1}^{\mathrm{TM}}(\mathbf{k}), \omega_{j+1}^{\mathrm{TE}}(\mathbf{k}))$, $\omega_{bot} = \max(\omega_i^{\mathrm{TM}}(\mathbf{k}), \omega_j^{\mathrm{TE}}(\mathbf{k}))$. Based on the FEA method, for a given frequency $\omega_i(\mathbf{k})$ and its corresponding eigenvector $\mathbf{u}_{\mathrm{i}}$, its derivative to x_{e} can be expressed as

$$\frac{\partial \omega_i(\mathbf{k})}{\partial x_e} = \frac{1}{2\omega_i(\mathbf{k})}\mathbf{u}_i^{\mathrm{T}}\left(\frac{\partial \mathbf{K}}{\partial x_e} - (\omega_i(\mathbf{k}))^2\frac{\partial \mathbf{M}}{\partial x_e}\right)\mathbf{u}_i \tag{10}$$

The derivatives of matrix **K** and **M** can be calculated from the interpolation functions 7a and 7b

$$\frac{\partial \mathbf{K}}{\partial x_e} = 0, \quad \frac{\partial \mathbf{M}}{\partial x_e} = (\varepsilon_2 - \varepsilon_1)\mathbf{M}_e \quad \text{for TM mode} \tag{11a}$$

$$\frac{\partial \mathbf{K}}{\partial x_e} = \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right)\mathbf{K}_e, \quad \frac{\partial \mathbf{M}}{\partial x_e} = 0 \quad \text{for TE mode} \tag{11b}$$

In order to improve the stability and convergence of optimization process, a heuristic filter scheme is integrated into the optimization algorithm and the elemental sensitivity numbers are further averaged with their corresponding values in the previous iteration. The specific procedure of the filter and average scheme can refer to Ref. [13].

After obtaining the elemental sensitivity numbers, BESO will modify the design variables based on the specified proportions of two constitutive materials. The BESO process starts from an initial design filled up with material 2 except a tiny void at the center of the unit cell. The total volume of material 2 in each iteration is evolved gradually[13].

BESO will increase design variables for elements with highest sensitivity numbers and decrease design variables for elements with lowest sensitivity numbers simultaneously. Based on the relative ranking of the elemental sensitivity numbers, a threshold of the sensitivity number, α^*, is determined by using bi-section method so that the volume of material 2 in the next iteration is equal to the target volume. The design variable for each element is modified by comparing its sensitivity number with the threshold. Different from other topology optimization methods with continuous

design variable, BESO method uses discrete design variable. In each iteration, the variation of a design variable is a constant Δx ($\Delta x = 0.1$ is used in this paper). The design variable x_e is updated as

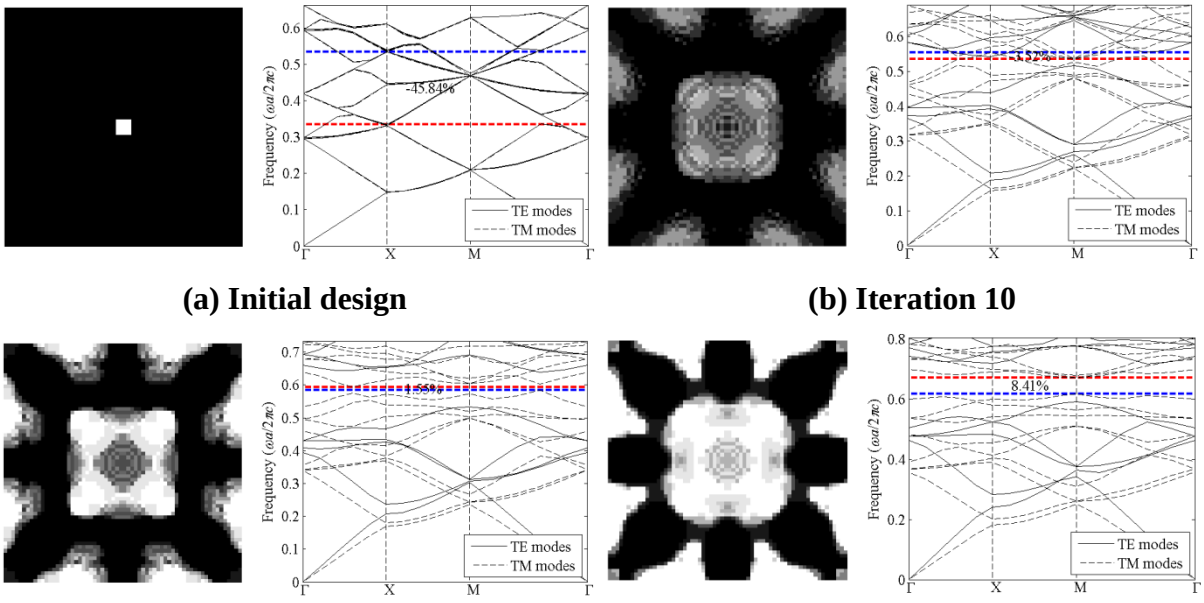
$$x_e = \begin{cases} \min(x_e + \Delta x, 1), & \text{if } \alpha_e > \alpha^* \\ \max(x_e - \Delta x, 0), & \text{if } \alpha_e < \alpha^* \end{cases} \tag{12}$$

Although discrete intermediate design variables are used in the optimization process, the final design tends to be a clear 0/1 design because a larger permittivity contrast leads to a larger band gap[22].

4. Results and discussion

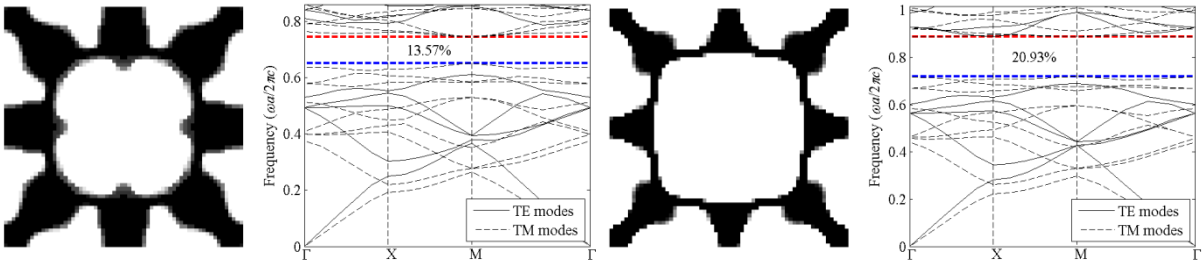
2D photonic crystals with square lattice and C_{4v} symmetry are considered in this paper. The photonic crystals consist of 2 materials: Air, relative permittivity $\varepsilon_1 = 1$ and GaAs, relative permittivity $\varepsilon_2 = 11.4$. In the topology images below, the air is indicated by white color and the GaAs is indicated by black. The model is meshed with 64×64 four-node square elements. The FEA and BESO are programed with MATLAB codes.

To illustrate the optimization process of BESO method, the evolution history of the topology and band diagrams of an example are shown in Fig. 1. The position of the band gap is between the 5th and 6th TM photonic bands and the 9th and 10th TE bands. For the simple initial design, the band gap-midgap ratio is a negative value, which means there is no band gap at all. With the optimization continues, the topology gradually evolves and the band gap-midgap ratio gradually increases. The volume fraction of GaAs gradually decreases from almost 100% to 30.06% at the end of the optimization process, and a complete band gap with a band gap-midgap ratio of 20.93% emerges. The whole optimization process cost 91 iterations which demonstrate the high computational efficiency of the proposed optimization algorithm.



(a) Initial design **(b) Iteration 10**

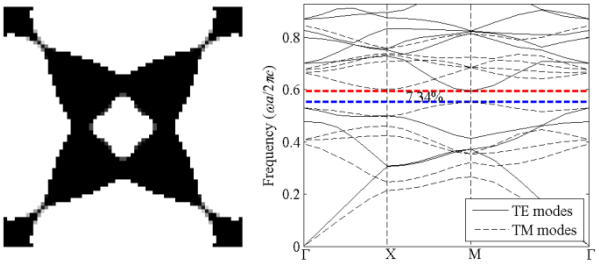
(c) Iteration 20 **(d) Iteration 30**



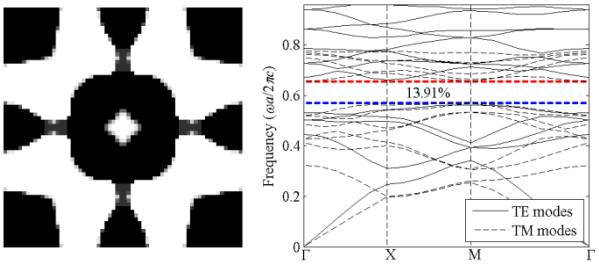
(e) Iteration 40 **(f) Final design**

Figure 1. Evolution history of a complete band gap formed by the overlapping of 9th TM and 5th TE band gap

Although the size of the maximized band gap illustrated in Fig. 1f is slightly larger than the result in Ref. [15], [17] and [18], all of them have a similar structure. However, by appointing different position of the complete band gap, some new designs can be obtained, as illustrated in Fig. 2. These designs are both obtained in less than 100 iterations. The resulting complete band gaps are 7.34% and 13.91%, respectively. It indicates that the solution is highly depended on the specified TE and TM bands for optimization. Therefore, the further study is recommended for finding a maximum complete band gap by appointing appropriate TE and TM bands.



(a) Complete band gap formed by the overlapping of 5th TM and 3rd TE band gap



(b) Complete band gap formed by the overlapping of 8th TM and 5th TE band gap

Figure 2. New designs of photonic crystals with complete band gap

Conclusions

This paper investigates the topology optimization of 2D photonic crystals with complete band gaps. An optimization scheme based on FEA and BESO is proposed to find the optimal design. According to the defined objective function and sensitivity analysis, the initial design gradually evolves to its optimum and a large complete band gap is formed between specified photonic TE and TM bands. The numerical results indicate the high-efficiency of the proposed algorithm and some new topologies with complete band gaps have been obtained. The proposed method can be equally applied for photonic crystals with other lattices.

Acknowledgements

The authors wish to acknowledge the financial support from the Australian Research Council (FT130101094 and DE120100291) and the China Scholarship Council.

References

[1] E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics, Phys Rev Lett, 58 (1987) 2059-2062.
[2] C. Luo, S.G. Johnson, J.D. Joannopoulos, J.B. Pendry, All-angle negative refraction without negative effective index, Physical Review B, 65 (2002) 201104.
[3] T. Baba, Slow light in photonic crystals, Nature Photonics, 2 (2008) 465-473.
[4] N. Skivesen, A. Têtu, M. Kristensen, J. Kjems, L.H. Frandsen, P.I. Borel, Photonic-crystal waveguide biosensor, Optics express, 15 (2007) 3169.
[5] J.D. Joannopoulos, S.G. Johnson, J.N. Winn, R.D. Meade, Photonic crystals: molding the flow of light, Princeton university press, Princeton, 2011.
[6] Y. Akahane, T. Asano, B.-S. Song, S. Noda, High-Q photonic nanocavity in a two-dimensional photonic crystal, Nature, 425 (2003) 944-947.
[7] J.H. Wu, A.Q. Liu, L.K. Ang, Band gap optimization of finite photonic structures using apodization method, Journal of Applied Physics, 100 (2006) 084309.
[8] M.M. Hossain, G. Chen, B. Jia, X.-H. Wang, M. Gu, Optimization of enhanced absorption in 3D-woodpile metallic photonic crystals, Optics express, 18 (2010) 9048-9054.
[9] S.J. Cox, D.C. Dobson, Band Structure Optimization of Two-Dimensional Photonic Crystals in H-Polarization, Journal of Computational Physics, 158 (2000) 214-224.
[10] S. Preble, M. Lipson, H. Lipson, Two-dimensional photonic crystals designed by evolutionary algorithms, Applied Physics Letters, 86 (2005) 061111.
[11] C.Y. Kao, S. Osher, E. Yablonovitch, Maximizing band gaps in two-dimensional photonic crystals by using level set methods, Applied Physics B, 81 (2005) 235-244.
[12] O. Sigmund, K. Hougaard, Geometric properties of optimal photonic crystals, Physical Review Letters, 100 (2008) 153904.
[13] F. Meng, X. Huang, B. Jia, Bi-directional evolutionary optimization for photonic band gap structures, Journal of Computational Physics, 302 (2015) 393-404.
[14] M. Qiu, S. He, Optimal design of a two-dimensional photonic crystal of square lattice with a large complete two-dimensional bandgap, Journal of the Optical Society of America B, 17 (2000) 1027.
[15] L. Shen, Z. Ye, S. He, Design of two-dimensional photonic crystals with large absolute band gaps using a genetic algorithm, Physical Review B, 68 (2003).
[16] H. Li, L. Jiang, W. Jia, H. Qiang, X. Li, Genetic optimization of two-dimensional photonic crystals for large absolute band-gap, Optics Communications, 282 (2009) 3012-3017.
[17] H. Men, N.C. Nguyen, R.M. Freund, K.M. Lim, P.A. Parrilo, J. Peraire, Design of photonic crystals with multiple and combined band gaps, Physical Review E, 83 (2011).
[18] D. Wang, Z. Yu, Y. Liu, P. Lu, L. Han, H. Feng, X. Guo, H. Ye, The optimal structure of two dimensional photonic crystals with the large absolute band gap, Optics express, 19 (2011) 19346-19353.
[19] P. Shi, K. Huang, Y.-p. Li, Photonic crystal with complex unit cell for large complete band gap, Optics Communications, 285 (2012) 3128-3132.
[20] X. Huang, Y.-M. Xie, Evolutionary topology optimization of continuum structures: methods and applications, John Wiley & Sons, Chichester, 2010.
[21] C. Kittel, Introduction to solid state physics, Wiley, 2005.
[22] H. Men, N.C. Nguyen, R.M. Freund, P.A. Parrilo, J. Peraire, Bandgap optimization of two-dimensional photonic crystals using semidefinite programming and subspace methods, Journal of Computational Physics, 229 (2010) 3706-3725.

Propagation properties of elastic waves in the 3D nacreous composite material

***S. Zhang, J. Yin, †H. W. Zhang, and B. S. Chen**

Department of Engineering Mechanics, State Key Laboratory of Structural Analysis of Industrial Equipment, Dalian University of Technology, Dalian, Liaoning, China

*Presenting author: zhangs@dlut.edu.cn
†Corresponding author: zhanghw@dlut.edu.cn

Abstract

Inspired by natural nacreous materials with the excellent performance, a kind of 3D nacreous composite material is designed based on the thought of staggered and combined soft and hard materials. In the 3D band structure analysis, the designed material generates an ultrawide low frequency band gap. Additionally, the influences on the band gap with different material parameters of the model are examined. Furthermore, the numerical tests for the transmission characteristics reveal the significant vibration attenuation effect of the nacreous material which fit remarkably well with the band gap.

Keywords: Nacreous composite material, Phononic crystal, Band gap, Vibration isolation, Multi-level substructure.

Introduction

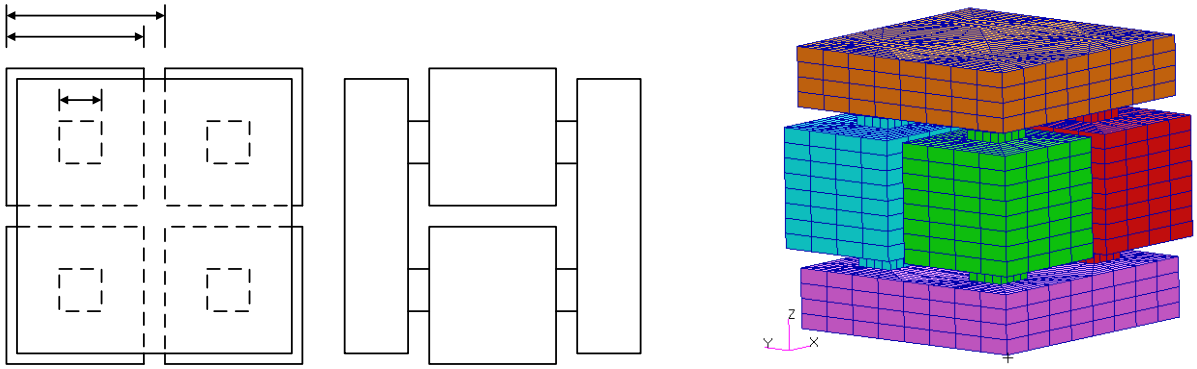
Mother nature is the best designer. This belief enlightens people to set foot on the way to mimic and understand nature. In the exploration of nature, people find out that many biological materials feature excellent mechanical or physical properties compared with engineering materials people usually use [1]. The feet of geckos and insects possess remarkably strong adhesion contact ability; cobwebs boast considerably high toughness and strength; animal bones are shaped in multi-scale and porous structures with light weight and high strength; nacreous composite materials exhibit the strength stronger than any member of single-phase material. All of these have prompted the flourish of biomimetic materials. Nacre is composed of 95% mineral substance (which is relatively hard material) and 5% protein (which is relatively soft biological material). This material, however, is more than several times strength than the single-phase mineral material [2]. Taking this into consideration, Gao *et al.* [3] proposed an explanation that the microstructures of nacreous composite materials are insensitive to flaw, and then built a Brick-and-Mortar (B-and-M) model to describe nacreous materials. Yao's study [4] showed that nacreous composite materials are not only insensitive to flaw under the micro scale, but perform well to restrain the stress concentration.

It should be noticed that nacreous composite materials exhibit periodicity in the soft and hard hierarchical structures, which is the basic characteristic of phononic crystals [5]. They are capable of tailoring the wave propagation through some frequency ranges (band gaps) in which the propagation of sound and elastic waves is forbidden [6]. Inspired by nacreous composite materials with the excellent performance, the elastic wave propagation in the 3D nacreous composite material is studied.

Three-dimensional Band Strucuture

Most researchers spare much less time for the dynamic characteristics of the B-and-M composite structure. As our attention in this paper is mainly paid to the band structure and wave propagation characteristics of the designed 3D B-and-M composite material. The geometric parameters of the 3D finite element model (FEM) are shown in Fig. 1(a). The model is divided into six basic

substructures [Fig. 1(b)] based on a multi-level substructure technique [7] to improve computational efficiency and reduce memory usage significantly which has the same accuracy with the traditional FEM. The material of the Brick is aluminum, and that of the Mortar is silicone rubber (material parameters shown in Table 1).



(a) Geometric parameters (b) Substructure model

Figure 1. The unit cell of a 3D nacreous composite material

Table 1. Two-phase parameters of nacreous material

Material	Density /kg/m^3	Elastic modulus /MPa	Poisson's ratio
Aluminum (Brick)	2700	70000	0.3
Silicone rubber (Mortar)	1300	0.1175	0.4688

The band structure of the designed 3D nacreous material (Fig. 2) illustrates that this material opens up an ultrawide band gap in the low frequency regime (84.4155Hz ~ 467.5685Hz). This kind of nacreous composite material, however, is a typical kind of Bragg phononic crystal rather than the locally resonant phononic crystal, which can be verified from its reduced frequency of the central gap frequency. Owing to its band characteristics, this material is suitable for the engineering application in the vibration isolation in the low frequency range. Foreseeably, this material will have a remarkable effect on the vibration isolation if elastic waves' frequencies locate in the band gap regime.

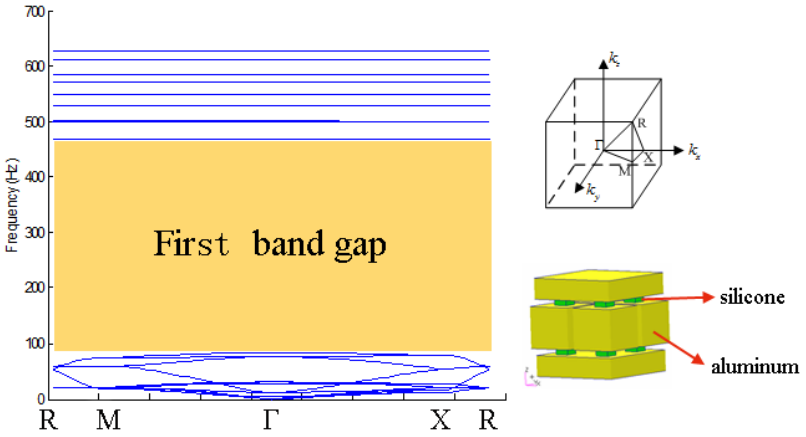


Figure 2. The band structure of the 3D nacreous material

It is known that the scaling law [8] uniformly expanding or shrinking the physical sizes of phononic crystals by a factor β results in the frequency spectrum being scaled by $1/\beta$. This law can also be

explained by a view of finite element method. The generalized eigenvalue problem of the phononic crystal with finite element method can be written as,

$$[\boldsymbol{K}]\{\boldsymbol{u}\} = \omega^2[\boldsymbol{M}]\{\boldsymbol{u}\}. \tag{1}$$

When the physical size of the phononic crystal is expanded or shrinked by a factor β, the new generalized eigenvalue equations with the constant density and elasticity can be written as,

$$\beta[\boldsymbol{K}]\{\boldsymbol{u}\} = \bar{\omega}^2\beta^3[\boldsymbol{M}]\{\boldsymbol{u}\},\ \bar{\omega} = \frac{1}{\beta}\omega. \tag{2}$$

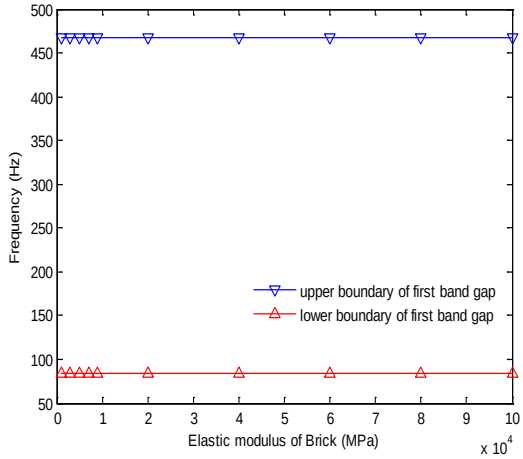
In addition, Chen's study [9] showed that the size-effect becomes more important and the non-classical elastic continuum should be taken into account when a system is in the dimension of several nanometers. The results illustrated that the classical elastic continuum is still applicative above the dimension of ten nanometers, otherwise the size-effect should be considered. Therefore, the analysis of nacreous material with the B-and-M pattern is reasonable in the classical elastic continuum. Moreover, a lot of researches using the classical elastic continuum to discuss phononic crystals in the micron scale and nano scale were found in Ref. [10][11]. The vibration isolation performance for a given B-and-M structure is dependent on the length scale, but the B-and-M structures at macroscale and microscale have similar vibration isolation performance. According to the scaling law, the first band gap of the 3D nacreous composite material (Fig. 1) has a range of 16.88MHz～93.51MHz, if the length of the unit cell is 350nm.

Influences of Material Parameters on the Band Gap

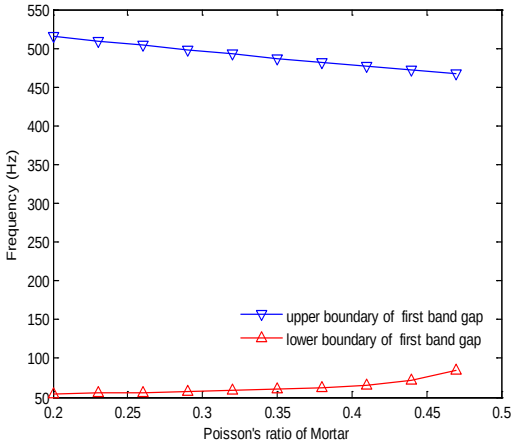
In order to design the nacreous composite material for vibration reduction in the low frequency regime, four material parameters of the 3D B-and-M model are studied to examine their influence on the band gap. The four material parameters are the elastic modulus and density of Brick, the elastic modulus and Poisson's ratio of Mortar (with corresponding ranges shown in Table 2).

Table 2. The range of parameters of four materials with two-phase model

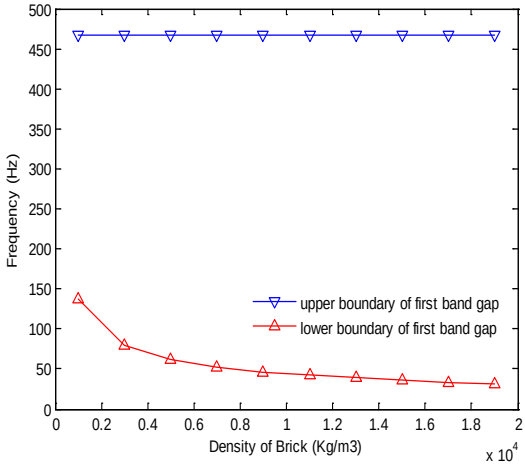
	Brick		Mortar	
	Elastic modulus (GPa)	Density (kg/m^3)	Elastic modulus (MPa)	Poisson's ratio
Lower limit	1	1000	0.1	0.2
Upper limit	100	20000	100	0.49



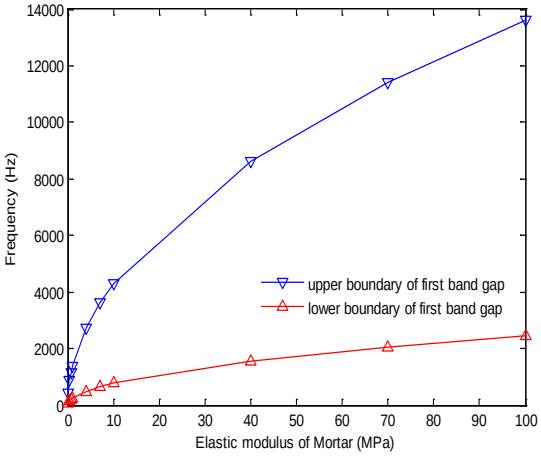
(a) Elastic modulus of Brick



(b) Poisson's ratio of Mortar



(c) Density of Brick



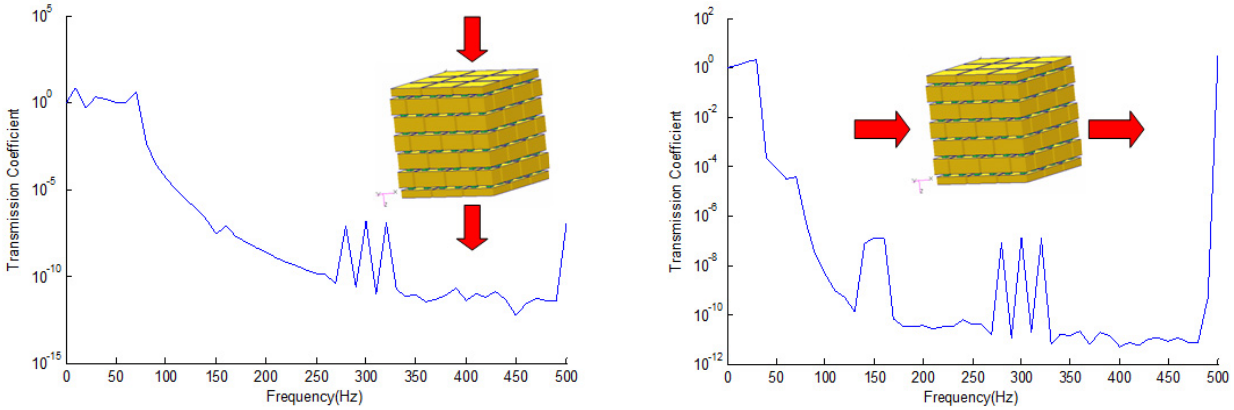
(d) Elastic modulus of Mortar

Figure 3. The influence of the four material parameters on the first band gap

The influence exerted by the four material parameters is shown in Fig. 3. The band gap results demonstrate that the elastic modulus of Brick and the Poisson's ratio of Mortar exert little influence on the first band gap. However, a larger density of Brick, which leads to a larger mass, contributes to a wider first band gap. Moreover, a smaller elasticity modulus of Mortar, which makes the shear stiffness of Mortar smaller, leads to the lower boundaries of the first band gap with a narrower first band gap.

Transmission Characteristics

To verify the vibration isolation effect of the 3D nacreous composite material, the transmission characteristics of the designed material are studied in this paper. A finite periodic structure of 3×3×3 cells is modeled for the numerical analysis with the help of the software MSC.Nastran. We are interested in the *x*-directional (horizontal) and *z*-directional (vertical) displacement transmission characteristics, which are valued by the ratio between the output displacement and the input displacement. The response curves are usually described in the logarithmic form.



(a) Horizontal direction (b) Vertical direction

Figure 4. The transmission characteristics of the designed material

The horizontal and vertical transmission characteristics of the designed material are shown in Fig. 4. According to the results, the 3D nacreous composite material isolates the vibration effectively within an ultrawide frequency regime in both horizontal and vertical directions. Given that the frequency range of the elastic wave attenuation and the band gap regime are consistent with each other, the correctness of the computed results is verified from an additional aspect. The results also show that the 3D nacreous composite material is a kind of Bragg phononic crystals which do not present the Fano-like interference phenomenon [12] found in locally resonant phononic crystals, thus benefiting the application in the engineering vibration reduction.

Conclusions

Enlightened by the nacreous composite material with the excellent performance, a 3D phononic crystal material is designed based on the idea of staggered and combined soft and hard materials. The results demonstrate that the material generates an ultrawide first band gap in the low frequency regime, and that its size makes it suitable to be applied to the engineering vibration reduction and isolation. Moreover, the band gap could be furtherly changed via adjusting material parameters. In the numerical tests for transmission characteristics, this material boasts remarkable effect on vibration reduction and isolation, which is in consistency with the band gap results.

Acknowledgments

The supports of the National Natural Science Foundation of China (11232003, 91315302), the National 111 Project of China (No.B08014, No.B14013), and the Doctoral Fund of Ministry of Education of China (20130041110050) are gratefully acknowledged.

References

[1] Meyers, M., Chen, P., and Lin, A. (2008) Biological materials: Structure and mechanical properties, *Progress in materials science* **53**, 1-206.

[2] Barthelat, F., and Rabiei, R. (2011) Toughness amplification in natural composites, *J. Mech. Phys. Solids* **59**, 829-840.

[3] Gao, H., Ji, B., and Jager, I. *et al.* (2003) Materials become insensitive to flaws at nanoscale: Lessons from nature, *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5597-5600.

[4] Yao, H., Song, Z., and Xu, Z. *et al.* (2013) Cracks fail to intensify stress in nacreous composites, *Composites Science and Technology* **81**, 24-29.

[5] Yin, J., Huang, J., and Zhang, S. *et al.* (2014) Ultrawide low frequency band gap of phononic crystal in nacreous composite material, *Physics Letters A* **378**, 2436-2442.

[6] Swinteck, N., Vasseur, J., and Hladky-Hennion, A. *et al.* (2012) Multifunctional solid/solid phononic crystal, *J. Appl. Phys.* **112**, 024514.

[7] Yin, J., Zhang, S., and Zhang H. *et al.* (2015) Band structure and transmission characteristics of complex phononic crystals by multi-level substructure scheme, *International Journal of Modern Physics B* **29**, 1550013.
[8] Kushwaha, M., Halevi, P., and Martinez, G. *et al.* (1994) Theory of acoustic band-structure of periodic elastic composites, *Phys. Rev. B* **49**, 2313-2322.
[9] Chen, A., and Wang, Y. (2011) Size-effect on band structures of nanoscale phononic crystals, *Physica E* **44**, 317-321.
[10] Veres, I., Berer, T., and Matsuda, O. *et al.* (2012) Focusing and subwavelength imaging of surface acoustic waves in a solid-air phononic crystal, *J. Appl. Phys.* **112**, 053504.
[11] Maldovan, M., and Thomas, E. (2006) Simultaneous localization of photons and phonons in two-dimensional periodic structures, *Appl. Phys. Lett.* **88**, 251907.
[12] Goffaux, C., Sanchez-Dehesa, J., and Yeyati, A. *et al.* (2002) Evidence of Fano-like interference phenomena in locally resonant materials, *Phys. Rev. Lett.* **88**, 225502.

Optimal sensors/actuators placement in smart structure using island model parallel genetic algorithm

†Animesh Nandy[1], †*Debabrata Chakraborty[1], and Mahesh S Shah[2]

[1]Department of Mechanical Engineering, IIT Guwahati, Guwahati 781039, India
[2]Center for Development of Advanced Computing (C-DAC), Pune, India

*Presenting author: chakra@iitg.ernet.in
†Corresponding author: chakra@iitg.ernet.in

Abstract

Determination of optimal placements of sensors/actuators in large structures is a difficult job as large number of possible combinations leads to a very high computational time and storage. Therefore this kind of optimization problem demands a parallel implementation of the optimization schemes. Island model genetic algorithm (GA) being inherently parallel has been used for searching optimal placements of collocated sensors/actuators. Numerical simulations have been done for determination of optimal placements of collocated PZT sensors and actuators in smart fiber reinforced shell structures using island model parallel GA (IMPGA) in conjunction with electro-mechanical finite element analysis with an objective of maximizing the controllability index. It has been observed that the present IMPGA based formulation not only makes it possible to determine optimal sensors/actuators locations for large structures but also leads to a better solution at a much reduced and achievable computational time.

Keywords: Optimal placement, Sensors/Actuators, Island Model Parallel Genetic Algorithms, Smart Structures.

Introduction

Optimal placement of sensors and actuators plays an important role in deciding the efficacy of smart structures in suppressing undesirable disturbances. For active vibration control of large structures requiring a large number of sensors/actuators a very large number of possibilities exist from which the optimal locations of the sensors/actuators need to be chosen to achieve the maximum actuation. Therefore, determination of optimal placements of sensors and actuators has been an important area of research and a number of works have already been reported. Some of the important works are described here. Kang et al [1] has worked on optimal placement of piezoelectric sensor/actuator for active vibration control of laminated beams. Kim and Kim [2] presented optimal distribution of an active damping layer consuming minimum control energy on a flexible plate. Since optimal placement of sensors and actuators is a discrete optimization problem, genetic algorithm (GA) ideally suits as an optimization tool for this kind of problems. Rao et al [3] used GAs to obtain the optimal actuators placement in an actively controlled two-bay truss. Dhuri and Seshu [4] used GA for active vibration control of flexible structure. Roy and Chakraborty [5] presented an improved GA for optimal vibration control of smart fiber reinforced polymer (FRP) composite structure. Multi-objective optimization of hybrid composite laminates using serial genetic algorithm (SGA) and finite element method (FEM) has also been reported by Rahul et al. [6]. Agarwal et al [7] proposed a gene manipulation, multi-objective genetic algorithm to optimize the placement of active devices and sensors in frame structures. Roy and Chakraborty [8] used GA based linear-quadratic regulator (LQR) control scheme for designing an optimal controller to maximize the closed loop.

It has already been reported that GA based placements leads to superior results compared to commonly used mode shape based placement [5]. However necessary requirements of large population size and a large number of generations for convergence to the optimal solution put constraints on computational time and storage. Moreover, for structural applications, the fitness is calculated using FEM whose accuracy is again decided by spatial and time

discretizations. This is more important for large structures where the number of combinations to be searched for converging to the optimal solution is very large. Therefore, IMPGA could be advantageously used to search optimal sensors/actuators placements in such smart structures. Even though there are few works [9] where IMPGA has been used for design of optimal stacking sequence of composite structures, to the best of author's knowledge, no work has been reported in literature to obtain optimal sensors and actuators placement using IMPGA. Therefore the present paper aims at developing an island model parallel GA based methodology to search for optimal placements of collocated sensors and actuators leading to a better solution compared to SGA and at a reduced and achievable computational time.

Problem Formulation

Figure 1 shows the schematics of a smart laminated structure having patches of piezoelectric material bonded on the top and bottom surfaces of the base structure, one as sensor and the other as actuator. Signal from the sensor is used as a feedback in a closed–loop feedback control system. An appropriate control law determines the feedback signal to be given to the actuator. In Fig. 1, F_t is the excited force, ϕ_s is the voltage generated by the sensor and ϕ_a is the voltage input to the actuator in order to control the displacement.

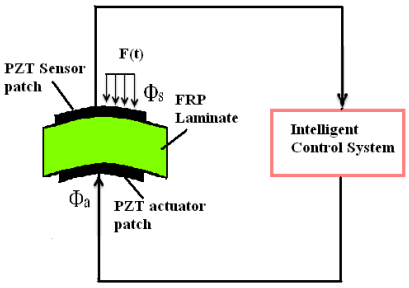


Figure 1. Smart structure

Finite Element Formulation for Controllability Index

An eight noded isoparametric shell elements have been used for finite element electromechanical analysis of the smart FRP shells [10]. The direct and converse piezoelectric equations are given by equations (1) and (2) respectively as

$$\{\boldsymbol{D}\} = [\boldsymbol{e}]\{\boldsymbol{\varepsilon}\} + [\boldsymbol{\epsilon}]\{\boldsymbol{E}\} \tag{1}$$

$$\{\boldsymbol{\sigma}\} = [\boldsymbol{C}]\{\boldsymbol{\varepsilon}\} - [\boldsymbol{e}]^T\{\boldsymbol{E}\} \tag{2}$$

where, $\{\boldsymbol{D}\}$ denotes the electric displacement vector, $\{\boldsymbol{\sigma}\}$ denotes the stress vector, $\{\boldsymbol{\varepsilon}\}$ denotes the strain vector and $\{\boldsymbol{E}\}$ denotes the electric field vector. Further $[\boldsymbol{e}] = [\boldsymbol{d}][\boldsymbol{C}]$, where $[\boldsymbol{e}]$ comprises the piezoelectric coupling constants, $[\boldsymbol{d}]$ denotes the piezoelectric constant matrix and $[\boldsymbol{\epsilon}]$ denotes the dielectric constant matrix. Electrical potential has been assumed to only vary in the thickness direction linearly and the electric field strengths of an element in terms of the electrical potential for the actuators and the sensors patches respectively are expressed as

$$\{-\boldsymbol{E}_a^e\} = [\boldsymbol{B}_a^e]\{\boldsymbol{\phi}_a^e\} = \begin{bmatrix} 0 \\ 0 \\ 1/h_a \end{bmatrix}\{\boldsymbol{\phi}_a^e\} \text{ and } \{-\boldsymbol{E}_s^e\} = [\boldsymbol{B}_s^e]\{\boldsymbol{\phi}_s^e\} = \begin{bmatrix} 0 \\ 0 \\ 1/h_s \end{bmatrix}\{\boldsymbol{\phi}_s^e\} \tag{3}$$

where subscripts *a* and *s* refer to the actuator patch and the sensor patch respectively. $[\boldsymbol{B}_a^e]$ and $[\boldsymbol{B}_s^e]$ are the electric field gradient matrices of the actuator and the sensor elements respectively. The dynamic finite element equations of a piezo-laminated composite shell can be derived from the Hamilton principle and for one-element it is

$$\begin{pmatrix} \left[\boldsymbol{M}_{uu}^{e}\right] & [\boldsymbol{0}] & [\boldsymbol{0}] \\ [\boldsymbol{0}] & [\boldsymbol{0}] & [\boldsymbol{0}] \\ [\boldsymbol{0}] & [\boldsymbol{0}] & [\boldsymbol{0}] \end{pmatrix} \begin{Bmatrix} \{\ddot{\boldsymbol{d}}\} \\ \{\ddot{\boldsymbol{\phi}}_a\} \\ \{\ddot{\boldsymbol{\phi}}_s\} \end{Bmatrix} + \begin{pmatrix} \left[\boldsymbol{K}_{uu}^{e}\right] & \left[\boldsymbol{K}_{ua}^{e}\right] & \left[\boldsymbol{K}_{us}^{e}\right] \\ \left[\boldsymbol{K}_{au}^{e}\right] & \left[\boldsymbol{K}_{aa}^{e}\right] & [\boldsymbol{0}] \\ \left[\boldsymbol{K}_{su}^{e}\right] & [\boldsymbol{0}] & \left[\boldsymbol{K}_{ss}^{e}\right] \end{pmatrix} \begin{Bmatrix} \{\boldsymbol{d}\} \\ \{\boldsymbol{\phi}_a\} \\ \{\boldsymbol{\phi}_s\} \end{Bmatrix} = \begin{Bmatrix} \{\boldsymbol{F}^{e}\} \\ \{\boldsymbol{G}^{e}\} \\ \{\boldsymbol{0}\} \end{Bmatrix} \tag{4}$$

where $\left[\boldsymbol{M}_{uu}^{e}\right]$ is the global mass matrix, $\left[\boldsymbol{K}_{uu}^{e}\right]$ is the global elastic stiffness matrix, $\left[\boldsymbol{K}_{ua}^{e}\right]$ and $\left[\boldsymbol{K}_{us}^{e}\right]$ are the global piezoelectric coupling matrices of actuator and sensor patches respectively. $[\boldsymbol{K}_{aa}]$ and $\left[\boldsymbol{K}_{ss}^{e}\right]$ are the global dielectric stiffness matrices of actuator and sensor patches respectively. $\{\boldsymbol{d}\}$ is displacement vector, $\{\boldsymbol{F}^{e}\}$ is the element external mechanical force vector and $\{\boldsymbol{G}^{e}\}$ is the element external electrical force vector. After assembling the overall dynamic finite element equation is

$$[\boldsymbol{M}_{uu}]\{\ddot{\boldsymbol{d}}\} + \left[[\boldsymbol{K}_{uu}] - [\boldsymbol{K}_{ua}][\boldsymbol{K}_{aa}]^{-1}[\boldsymbol{K}_{au}] - [\boldsymbol{K}_{us}][\boldsymbol{K}_{ss}]^{-1}[\boldsymbol{K}_{su}]\right]\{\boldsymbol{d}\} = \{\boldsymbol{F}\} - [\boldsymbol{K}_{ua}]\{\boldsymbol{\phi}_a\} \tag{5}$$

The decoupled dynamic equations considering modal damping can be written as

$$\{\ddot{\boldsymbol{\eta}}_i(t)\} + 2\xi_{di}\omega_i\{\dot{\boldsymbol{\eta}}_i(t)\} + \omega_i^2\{\boldsymbol{\eta}_i(t)\} = [\boldsymbol{\psi}]^T\{\boldsymbol{F}\} - [\boldsymbol{\psi}]^T[\boldsymbol{K}_{ua}]\{\boldsymbol{\phi}_a\} \tag{6}$$

where ω_i the i^{th} natural frequency and ξ_{di} is the damping ratio, $[\boldsymbol{\psi}] = [\boldsymbol{\psi}_1 \boldsymbol{\psi}_2 \ldots \boldsymbol{\psi}_r]$ is the truncated modal matrix which transforms the generalized coordinates $d(t)$ to the modal coordinates $\eta(t)$ as $\{\boldsymbol{d}(t)\} = [\boldsymbol{\psi}]\{\boldsymbol{\eta}(t)\}$. In state-space form

$$\{\dot{\boldsymbol{X}}\} = [\boldsymbol{A}]\{\boldsymbol{X}\} + [\boldsymbol{B}]\{\boldsymbol{\phi}_a\} + \left[\hat{\boldsymbol{B}}\right]\{\boldsymbol{u}_d\} \tag{7}$$

where $[\boldsymbol{A}] = \begin{bmatrix} [\boldsymbol{0}] & [\boldsymbol{I}] \\ -\left[\omega_i^2\right] & [2\xi_{di}\omega_i] \end{bmatrix}$ is the system matrix, $[\boldsymbol{B}] = \begin{bmatrix} [\boldsymbol{0}] \\ -[\boldsymbol{\psi}]^T[\boldsymbol{K}_{ua}] \end{bmatrix}$ is the control matrix, $\left[\hat{\boldsymbol{B}}\right] = \begin{bmatrix} [\boldsymbol{0}] \\ [\boldsymbol{\psi}]^T\{\boldsymbol{F}\} \end{bmatrix}$ is the disturbance matrix, $\{\boldsymbol{u}_d\}$ is the disturbance input vector, $\{\boldsymbol{\phi}_a\}$ is the control input, and $\{\dot{\boldsymbol{X}}\} = \begin{Bmatrix} \dot{\eta} \\ \ddot{\eta} \end{Bmatrix}$ and $\{\boldsymbol{X}\} = \begin{Bmatrix} \eta \\ \dot{\eta} \end{Bmatrix}$. The sensor output equation can be written as

$$\{\boldsymbol{y}\} = [\boldsymbol{C}_0]\{\boldsymbol{X}\} \tag{8}$$

where $[\boldsymbol{C}_0]$ depends on the modal matrix $[\boldsymbol{\psi}]$ and the sensor coupling matrix $\left[\boldsymbol{K}_{us}\right]$. The modal control force f_c applied to the system can be written as

$$\{\boldsymbol{f}_c\} = [\boldsymbol{B}]\{\boldsymbol{\phi}_a\} \tag{9}$$

It follows from Eq. (9) that

$$\{f_c\}^T\{f_c\}=\{\phi_a\}^T[B]^T[B]\{\phi_a\} \tag{10}$$

Using the singular value analysis, $[B]=[M][S][N]^T$ where $[M]^T[M]=[I]$, $[N]^T[N]=[I]$ and $[S]=\begin{bmatrix}\sigma_1 & \dots & 0\\ 0 & \ddots & \vdots\\ \vdots & \dots & \sigma_{n_a}\\ 0 & \dots & 0\end{bmatrix}$ where n_a is the number of actuator. Eq. (10) can be rewritten as

$$\|\{f_c\}\|^2=\|\{\phi_a\}\|^2\|S\|^2 \tag{11}$$

Thus, maximizing this norm independently on the input voltage $\{\phi_a\}$ induces maximizing $\|S\|^2$. The magnitude of σ_i is a function of location and the size of piezoelectric actuators. Wang and Wang [11] proposed maximizing the controllability index as

$$Maximize\ \Omega=\prod_{i=1}^{n_a}\sigma_i \tag{12}$$

Island Model Parallel Genetic Algorithm for Optimum Sensor/Actuator

In the present problem, the design variables are the positions of the actuators, and are represented in a string of integers specifying the locations of actuators. Referring to Eq. (12), the higher the controllability index, the smaller will be the electrical potential required for control. In modal control, however, one of the important issues is to decide the number of control modes where actuations need to be done. Providing actuations to higher modes (which are residual modes actually not excited) might lead to instability known as control spill over. In the present work therefore the fitness/objective function which needs to be maximized in the GA ensuring optimal actuators locations has been proposed as follows

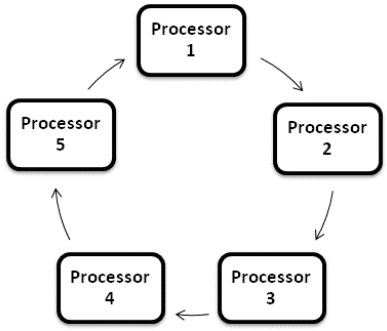


Figure 2. A 5 processor IMPGA

$$\boldsymbol{\Omega}=\begin{Bmatrix}\left(\prod_{i=1}^{n_a}\sigma_i-\gamma'\prod_{i=1}^{n_a}\sigma_i^R\right)\ if\ \left(\prod_{i=1}^{n_a}\sigma_i\right)>\left(\gamma'\prod_{i=1}^{n_a}\sigma_i^R\right)\\ \left(\prod_{i=1}^{n_a}\sigma_i-\gamma'\prod_{i=1}^{n_a}\sigma_i^R\right)\times 10^{-12},\ otherwise\end{Bmatrix} \tag{13}$$

where σ_i^R are the components of $[S^R]$ corresponding to residual modes and γ' is a weight constant. In this objective function, if the contribution of residual modes dominates, fitness of that population is forced to a very low value thereby eliminating the chances of such populations to grow in successive generations.

In IMPGA approach (Fig.2), first the population size is decided as a multiple of number of processors so that the total population of chromosomes is divided into a number of sub-

populations. String length of each population is decided by the number of actuators. For example referring to Fig. 3 if the population size is 60 and there are 5 processors (islands), each processor will handle a sub population size of 12. In each of the processor, for sub populations, the fitness value for each chromosome is obtained using the FEA independently and new sets of chromosomes are generated by applying genetic operators after each generation. After a certain number of generations, the best population of one processor is allowed to migrate only to its neighboring processor, replacing the worst population. For example, the best candidate from processor 1 will replace the worst candidate of processor 2, the best candidate of processor 2 will replace the worst candidate of processor 3 and so on. Thus, migration does not change the size of population. At the end of each generation, a better population results, and is used in successive generations to achieve populations with even better fitness. This is repeated until the solution converges and the optimal locations of actuators are selected corresponding to the chromosome with best controllability index.

Results and Discussions

A parallel code has been developed using MPI libraries as well as migration routines *(Island Model)* for optimization. The parallel code has been run on parallel cluster at IIT Guwahati. The cluster has 5 nodes and each node consists of 8 (1.5 GHz) processors. On one of the nodes of the cluster, the code has been run using SGA corresponding to same genetic parameters and population. The results obtained from IMPGA as well as SGA model for optimal placement of sensors and actuators have been compared to study the efficacy of the present approach.

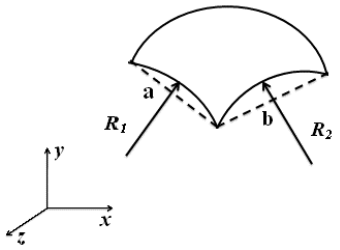


Figure 3. Curved shell

Problem Definition

In this study, a [p/[0/90]$_s$/p] graphite/epoxy (GR/E) doubly curved shell with the four edges simply supported, having a=b=0.02m, R_1=$2R_2$=R=0.06m R/a =3, a/h=10 (Fig. 3) has been considered. Here 'p' stands for piezo-patches one for sensing and the other for actuation. Thickness of each piezoelectric patch has been considered as 0.5 mm and that of each GR/E lamina has been considered as 0.25 mm. A 10×10 finite element mesh has been used to model the shell panel and optimal actuators placements have been calculated considering the first eight modes with first four modes as control modes and others as residual modes. The material properties have been listed in the Table 1.

Table 1. Material properties

Property	Gr/E	PZT
E_1 (GPa)	172.5	63.0
E_2=E_3 (GPa)	6.9	63.0
G_{12}=G_{13} (GPa)	3.45	24.6
G_{23} (GPa)	1.38	24.6
υ_{12}=υ_{13}=υ_{23}	0.25	0.28
ρ (kg m^{-3})	1600	7600
e_{31}=e_{32} (C m^{-2})	0.0	10.62
$\in_{11}$=$\in_{22}$=$\in_{33}$ (F m^{-1})	0.0	0.15 x10^{-7}

Table 2. Input parameters for GA

Initial population	60
Maximum generation	100
Number of actuators/sensors	6
Mutation rate	20%
Crossover rate	90%

Table 2 shows various input parameters considered for SGA as well as IMPGA. The stated genetic parameters used in IMPGA are finalized from the values obtained from multiple runs in the SGA. Thus, an optimized value of 100 generation with an initial population 60 is used to obtain comparative results regarding fitness and time between IMPGA and SGA. Six numbers of actuators are considered leading to a string length of 6.

Table 3. Controllability index for SGA

Initial population	Maximum generation	Fitness	Time (sec)
5	100	0.237	210993
30	100	0.248	1264376
60	100	0.255	2526150

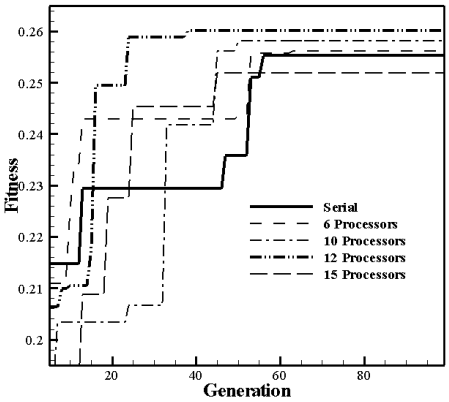


Figure 4. Convergence of fitness

Effect of Population Size on Controllability Index and CPU time

The code has been run up to 100 generations in one processor as SGA with increasing population size and table 3 shows the effect of population size on the controllability index. It is clear from the table that as the population size increases, controllability index increases but as expected the computational time also increases. It is therefore necessary that the optimal placement of sensors and actuators are searched from a larger population. However computational time requirement puts a restriction on the upper limit of the population size when such a problem is run on a serial platform. Therefore a parallel GA provides a feasible solution for such problems and island model GA being inherently parallel has been advantageously used in the present study.

Optimal Placement using Island Model Parallel GA

Five different schemes were used to study the effect of parallelization. The schemes are decided based on two factors viz. a maximum population size which could be run in a serial GA and with different number of processors such that in each case the number of processors is an integer factor of the population size. Different schemes considered here are:

- Scheme **1**:- SGA with one processor having population size of 60.
- Scheme **2**:- IMPGA with 6 processors having a sub population size of 10 in each
- Scheme **3**:- IMPGA with 10 processors having a sub population size of 6 in each
- Scheme **4**:- IMPGA with 12 processors having a sub population size of 5 in each
- Scheme **5**:- IMPGA with 15 processors having a sub population size of 4 in each

Table 4 shows the comparative performances of these 5 schemes up to 100 generations. It could be observed that increasing number of processors leads to increase in fitness up to 12 processors but beyond that the fitness decreases. This indicates that for this particular problem, the maximum number of processors that could be used for a population of 60 is 12. This is due to the fact that depending upon the number of populations increasing the number of processors leads to smaller sub population size in each processor and more communication

overheads. Thus the minimum population size for this problem is 5. Figure 4 shows the convergence of fitness for the optimal placement problem of smart shell structure for these 5 schemes. It is clear that with the increasing number of processors, it is not only that the fitness is higher compared to that in SGA but this fitness is achieved at a much less number of generations. This is due to the fact that the better solutions evolve independently in different processors and those processors (islands) interact (migration) after certain number of generations thereby passing on populations with better fitness only in each processor. Therefore even though the population size is larger, increasing number of processors still reduces the number of generation required for convergence.

Table 4. Controllability index for different schemes

Scheme	Initial population	Maximum generation	Number of processor	Fitness	Time (Sec)
1	60	100	1	0.255	2526150
2	60	100	6	0.256	443612
3	60	100	10	0.258	270812
4	60	100	12	0.260	227651
5	60	100	15	0.252	183234

Comparison Serial GA and IMPGA

Figure 5 shows the variation of computational time with the increasing number of processors

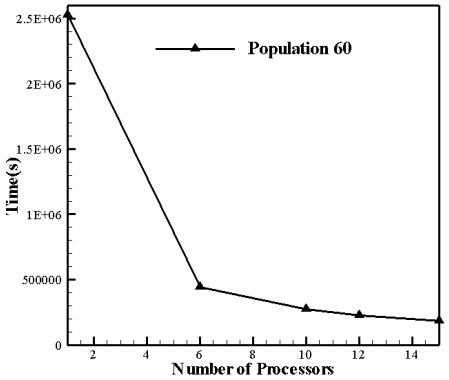


Figure 5. Time Vs no. of processors

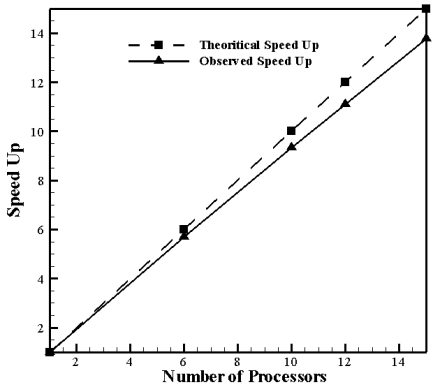


Figure 6. Speedup Vs no. of processors

while using IMPGA. Here, the use of one processor implies SGA executed using a single processor on the parallel platform. It could be observed from Fig. 5 that for a fixed number of initial population and generation, increase in number of processors leads to significant decrease in computational time. In the present problem of optimal placement of collocated actuators/ sensors using IMPGA with 12 numbers of processors takes 2,27,651 seconds while SGA takes 25,26,150 seconds under the same condition. The better computational performance of IMPGA is only because of better mixing of population due to migration which leads to faster convergence to optimal solution. The performance of a parallel code is

evaluated in terms of factors such as speedup, efficiency and scalability. The speedup, $S_N = T_S / T_{par}$ where, T_S and T_{par} represent time taken with a single processor multiple processors respectively. The efficiency of a parallel algorithm, $E_N = S_N / N$ where N is the number of processors. In the present study, a comparison has been made between SGA and IMPGA based on these factors. Table 5 shows the speed up obtained with increasing number of processors for a fixed population size of 60 up to 100 generations. Figure 6 shows the speedup comparison between SGA and IMPGA. It could be observed that with the increase in number of processors, speedup increases effectively. This is also clearly observed that there is a decrease in efficiency with the increase in processors (Fig.7). This is due to the fact that in the IMPGA, overhead increases due to increase in migration as the number of processors increases.

Further, to understand the behavior of the present IMPGA application, with increasing number of population, scalability analysis has been carried out keeping the number of processors fixed and the same is compared with SGA. In the present study, computational time has been noted for 100 generations for SGA and 15 processors IMPGA. Figure 8 shows the CPU time versus population size for both the cases. It could be clearly observed that in the

Table 5: Speedup and efficiency for 100 generation with fixed population size of 60

No. of Processors	Time (Sec)	Speedup Actual	Speedup Theoretical	Efficiency Actual	Efficiency Theoretical
1(Serial)	2526150	1	1	--	--
6	443612	5.69	6	94.9%	100%
10	270812	9.32	10	93.2%	100%
12	227651	11.09	12	92.5%	100%
15	183234	13.78	20	91.9%	100%

case of SGA the magnification in CPU time is equal to the magnification in population size. However, in the case of 15 processors IMPGA, increase in CPU time is much less compared to the magnification in population size. This shows that the proposed IMPGA based model in determination of optimal sensors/actuators location will be more efficient for larger population size and hence for larger structures.

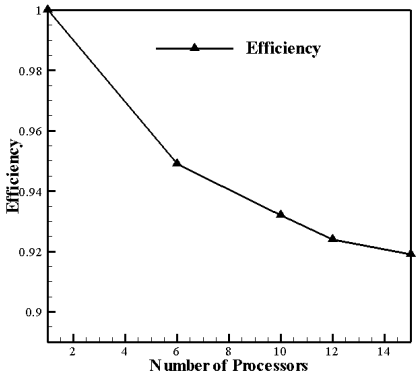


Figure 8 Time Vs no of processors

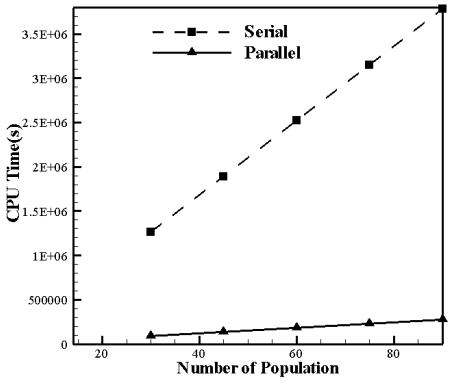


Figure 7. Efficiency Vs no of processors

Conclusions

In the present work an island model parallel genetic algorithm in conjunction with FEA has been developed for evaluation of optimal placements of collocated actuators/ sensors on a smart FRP shell structure. Controllability index determined from finite element analysis has been used as the measure of fitness with the actuators location as the variables. The present method not only leads to a better solution, but also finds that at a much reduced computational time. This method will be especially suitable for large structures where large number of sensor and actuators need to be used requiring larger population size and sequential GA fails due to limitation in population size. It has been observed from the present study that the present IMPGA based method is far superior compared to the sequential GA method in determining the optimal placements of actuators/sensors.

References

[1] Kang, Y.K., Park, H.C, Hwang, W., Han, K.S. (1996) Optimum placement of piezoelectric sensor/actuator for vibration control of laminated beams, *AIAA Journal* **34(9),** 1921–1926.

[2] Kim, T-W, Kim, J-H. (2005) Optimal distribution of an active layer for transient vibration control of a flexible plate, *Smart Materials and Structures* **14**, 904-16.

[3] Rao S. S., Pan T-S, Venkayya, V. B. (1991) Optimal placement of actuators in actively controlled structures using genetic algorithms, *AIAA Journal* **29**, 942–943.

[4] Dhuri, K. D. and Seshu, P. (2009) Multi objective optimization of actuator placement and sizing using genetic algorithms, *Journal of Sound & Vibration* **323**, 495-514.

[5] Roy, T. and Chakraborty, D.(2009) Optimal Vibration Control of Smart Structures using Improved Genetic Algorithms *Journal of Sound & Vibration* **319**, 15-40.

[6] Rahul, Sandeep, G., Chakraborty, D. and Dutta, A. (2006) Multi-objective optimization of hybrid laminates subjected to transverse impact, *Composite Structures* **73**, 360-369.

[7] Cha1, Y-J, Raich, A., Barroso, L and Agrawal, A.(2011) Optimal placement of active control devices and sensors in frame structures using multi-objective genetic algorithms, *Structural Control Health Monitoring* **20**, 16-44.

[8] Roy, T. and Chakraborty, D.(2009) Genetic algorithm based optimal design for vibration control of composite shell structures using piezoelectric sensors and actuators. *International Journal of Mechanics of Materials and Design* **5**, 45-60.

[9] Rahul, Chakraborty, D. and Dutta, A. (2005) Optimization of FRP composites against impact induced failure using island model parallel genetic algorithm, *Composites Science and Technology* **65**, 2003- 2013.

[10] Roy, T., Manikandan, P. and Chakraborty, D.(2010), Improved shell finite element for piezothermoelastic analysis of smart fiber reinforced concrete structures, *Finite Elements in Analysis and Design* **46**, 710-720.

[11] Wang, Q. and Wang, C. (2001) A controllability index for optimal design of piezoelectric actuators in vibration control of beam structures, *Journal of Sound and Vibration* **242(3)**, 507-518.

An examination of multiplicity of steady states for two- and four-sided lid-driven cavity flows through an HOC scheme

†Chitrarth Prasad[1] and *Anoop K. Dass[2]

[1]Department of Mechanical Engineering, Indian Institute of Technology Guwahati, India
[2]Department of Mechanical Engineering, Indian Institute of Technology Guwahati, India

*Presenting author: anoop@iitg.ernet.in
†Corresponding author: chitrarth2009@gmail.com

Abstract

This work is concerned with the computation of two- and four-sided lid-driven cavity flows using a transient higher-order compact (HOC) scheme. The multiplicity of steady states for most of these configurations through the use of non-compact schemes is well known. In this work, these cases are re-examined using a spatially fourth- and a temporally second-order compact scheme. The threshold values of certain parameters such as the cavity aspect ratio (A) and the flow Reynolds number (Re) are also computed beyond which there is multiplicity of solutions. It is observed that for the motion of non-facing walls of a square cavity, multiple solutions can be obtained for $Re = 975$, which is significantly lower than the previously established value. For all the other cases, these critical values of parameters are in good agreement with the existing investigations. Multiple solutions are also obtained for antiparallel wall motion in two-sided square cavities, which do not feature in any of the previous investigations using non-compact schemes.

Keywords: lid-driven, higher-order compact, Reynolds number, aspect ratio

1 Introduction

Over the years the single lid-driven cavity flow has been used as a benchmark problem to test the performance of numerical schemes and algorithms for incompressible flows. The problem has attracted researchers because it contains a wide variety of interesting phenomenon in the simplest of geometric settings. The single lid-driven cavity flow was extended to two- and four-sided cavity flows by various investigators [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], who observed that a plethora of vortex patterns can be generated with different aspect ratios and directions of motion of the walls.

It is well known that many nonlinear problems exhibit multiple steady solutions even though the governing equations and boundary conditions remain the same. As the governing equations for fluid flow are nonlinear in nature, the possibility of multiple solutions exists. Many researchers have found multiple solutions for parallel wall motion of facing walls for both rectangular and square cavities, and for antiparallel wall motion for rectangular cavities [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Albensoeder et al. [2] were among the first to investigate the nonlinear regime and find multiple 2D steady states in rectangular two-sided lid-driven cavities. They have found upto five different flow states for both parallel and antiparallel motion of facing walls. Very recently Lemée et al. [6] addressed the issue of multiple solutions in square cavity with parallel motion of facing walls and found out a critical Reynolds number above which multiple solutions exist, which is consistent with [2]. Similar investigations have been carried out for two-sided cavities with motion of non-facing walls and for four-sided cavities [1, 8]. All these existing

investigations have been carried out using non-compact schemes. It can be observed from these investigations, that the additional solutions, if they exist, always exist in pairs.

In this work, we re-examine these solutions using a higher-order compact (HOC) scheme of spatially fourth- and temporal second-order accuracy. This scheme was developed by Kalita et. al [11] by differentiating the governing equation to obtain compact approximations for the leading truncation error terms. Grid independent results are carefully computed so that the results can be used as means to test other schemes and algorithms. For the two-sided rectangular cavity, computations are carried out at a fixed Reynolds number (Re) of 600 at various aspect ratios (A) for parallel motion of facing walls. It is seen that at $Re = 600$ multiple solutions exist only above a critical aspect ratio of 0.556. This value is very close to the value 0.559 reported in [2]. Computations are also carried out for square cavities having antiparallel and parallel motion of facing walls at various Re's. For parallel wall motion in a square cavity, a threshold value of $Re = 983.5$ is observed below which only stable symmetric solutions exist. This value is in good agreement with previously reported values in [2, 6]. For antiparallel motion of facing walls in a square cavity, till very recently, existence of multiple steady solutions was not experienced. In a recent communication [12], using the same HOC scheme as the one used here, we demonstrate that existence. This shows the accuracy and effectiveness of the present scheme, which we use here to compute multiple solutions for the motion of nonfacing walls in two- and four-sided configurations as well. It is observed that the limiting value of Re for four-sided cavity is very close to previously reported value in [8], however for motion of non-facing walls in two-sided cavity, multiple solutions are seen to exist even for $Re = 975$, which is significantly lower than the previously reported threshold value of 1071 [8].

This paper is organized in five sections. Section 2 describes the HOC scheme formulation and associated descritization. In Section 3 the credibility of the present HOC code is established through a comparison exercise with the results of a previously known benchmark work [13]. Section 4 presents the results and discussion. Concluding remarks are made in Section 5.

2 Scheme formulation

There have been various attempts at developing HOC schemes [14, 11, 15, 16, 17]. The scheme used in this work was developed by Kalita et al. [11]. We present here a brief description of their HOC scheme formulation.

The unsteady 2D transport equation for a general variable ϕ in some continuous domain with suitable boundary conditions can be written as

$$a\frac{\partial \phi}{\partial t} - \nabla^2\phi + c(x,y,t)\frac{\partial \phi}{\partial x} + d(x,y,t)\frac{\partial \phi}{\partial y} = g(x,y,t) \tag{1}$$

where a is const, c and d are convection coefficients and g is forcing function. We take the steady state form of equation (1), which is obtained when ϕ,c,d and g are independent of t.

$$-\nabla^2\phi + c(x,y)\frac{\partial \phi}{\partial x} + d(x,y)\frac{\partial \phi}{\partial y} = g(x,y) \tag{2}$$

Discretization with second-order central differencing on a uniform grid with spacing h and k in the $x-$ and $y-$directions respectively yields

$$-\delta_x^2\phi_{ij} - \delta_y^2\phi_{ij} + c\delta_x\phi_{ij} + d\delta_y\phi_{ij} - \tau_{ij} = g_{ij} \tag{3}$$

where ϕ_{ij} denotes $\phi(x_i, y_j)$; δ_x, δ_x^2 and δ_y, δ_y^2 are the first and second-order central difference operators along $x-$ and $y-$directions respectively. The truncation error τ_{ij} is given by

$$\tau_{ij} = \left[\frac{h^2}{12}\left(2c\frac{\partial^3\phi}{\partial x^3} - \frac{\partial^4\phi}{\partial x^4}\right) + \frac{k^2}{12}\left(2d\frac{\partial^3\phi}{\partial y^3} - \frac{\partial^4\phi}{\partial y^4}\right)\right] + O(h^4, k^4) \tag{4}$$

In order to obtain a fourth-order compact formulation for equation (2), each of the derivatives of the leading term in equation (4) are compactly approximated to $O(h^2, k^2)$. In order to do this the original PDE (2) is differentiated to yield expressions for higher derivatives. After these substitutions (3) yields [11]

$$\begin{aligned} &-\alpha_{ij}\delta_x^2\phi_{ij} - \beta_{ij}\delta_y^2\phi_{ij} + C_{ij}\delta_x\phi_{ij} + D_{ij}\delta_y\phi_{ij} \\ &-\frac{h^2+k^2}{12}\left[\delta_x^2\delta_y^2 - c_{ij}\delta_x\delta_y^2 - d_{ij}\delta_x^2\delta_y - \gamma_{ij}\delta_x\delta_y\right]\phi_{ij} = G_{ij} \end{aligned} \tag{5}$$

where the coefficients $\alpha_{ij}, \beta_{ij}, \gamma_{ij}, C_{ij}, D_{ij}, G_{ij}$ are as follows

$$\alpha_{ij} = 1 + \frac{h^2}{12}\left(c_{ij}^2 - 2\delta_x c_{ij}\right) \tag{6}$$

$$\beta_{ij} = 1 + \frac{k^2}{12}\left(d_{ij}^2 - 2\delta_y d_{ij}\right) \tag{7}$$

$$\gamma_{ij} = \frac{2}{h^2+k^2}\left(h^2\delta_x d_{ij} + k^2\delta_y c_{ij}\right) - c_{ij}d_{ij} \tag{8}$$

$$C_{ij} = \left[1 + \frac{h^2}{12}(\delta_x^2 - c_{ij}\delta_x) + \frac{k^2}{12}(\delta_y^2 - d_{ij}\delta_y)\right]c_{ij} \tag{9}$$

$$D_{ij} = \left[1 + \frac{h^2}{12}(\delta_x^2 - c_{ij}\delta_x) + \frac{k^2}{12}(\delta_y^2 - d_{ij}\delta_y)\right]d_{ij} \tag{10}$$

$$G_{ij} = \left[1 + \frac{h^2}{12}(\delta_x^2 - c_{ij}\delta_x) + \frac{k^2}{12}(\delta_y^2 - d_{ij}\delta_y)\right]g_{ij} \tag{11}$$

For unsteady case (1), the equation with variable coefficients will be similar to (2), but the coefficients c and d are functions of x, y and t; and the expression on the RHS becomes $g(x, y, t) - a[(\partial\phi)/(\partial t)]$. Using this we can obtain the semi-discrete form of the unsteady equation (1) using HOC as

$$\begin{aligned} &a\left[1 + \frac{h^2}{12}(\delta_x^2 - c_{ij}\delta_x) + \frac{k^2}{12}(\delta_y^2 - d_{ij}\delta_y)\right]\delta_t\phi_{ij} = \\ &\alpha_{ij}\delta_x^2\phi_{ij} + \beta_{ij}\delta_y^2\phi_{ij} - C_{ij}\delta_x\phi_{ij} + \left[\delta_x^2\delta_y^2 - c_{ij}\delta_x\delta_y^2 - d_{ij}\delta_x^2\delta_y - \gamma_{ij}\delta_x\delta_y\right]\phi_{ij} + G_{ij} \end{aligned} \tag{12}$$

This scheme can be used to solve any 2D unsteady transport equation using a suitable time integration technique along with proper boundary conditions. In this work Crank Nicholson scheme has been used for time integration. This makes our HOC scheme fourth-order accurate

in space and second-order accurate in time.

$$a\left[1+\frac{h^2}{12}(\delta_x^2-c_{ij}\delta_x)+\frac{k^2}{12}(\delta_y^2-d_{ij}\delta_y)\right]\delta_t^+\phi_{ij}^n=\frac{1}{2}\left(H_{ij}^n+H_{ij}^{n+1}\right) \tag{13}$$

where δ_t^+ denotes the forward difference operator and the superscript n stands for the time level. H_{ij}^n is given as

$$H_{ij}^n=\alpha_{ij}\delta_x^2\phi_{ij}^n+\beta_{ij}\delta_y^2\phi_{ij}^n-C_{ij}\delta_x\phi_{ij}^n+\frac{h^2+k^2}{12}\left[\delta_x^2\delta_y^2-c_{ij}\delta_x\delta_y^2-d_{ij}\delta_x^2\delta_y-\gamma_{ij}\delta_x\delta_y\right]\phi_{ij}^n+G_{ij}^n \tag{14}$$

3 Code Validation

The above HOC formulation (13) can be used to solve streamfunction-vorticity form of the 2D Navior-Stokes given by

$$-\nabla^2\psi=\omega \tag{15}$$

$$\frac{\partial\omega}{\partial t}+u\frac{\partial\omega}{\partial x}+v\frac{\partial\omega}{\partial y}=\frac{1}{Re}\nabla^2\omega \tag{16}$$

where ψ stands for streamfunction and ω for vorticity. The velocities in $x-$ and $y-$directions are given by

$$u=\psi_y \tag{17}$$

$$v=-\psi_x \tag{18}$$

To lend credibility to the present HOC code its results for single lid-driven square cavity flow (Fig. 1(a)) are compared with the results of Ghia et al. [13] at various Re's at grids 51 X 51, 71 X 71 and 101 X 101. Fourth-order compact approximations for velocities obtained from equations (16)-(18) are given by

$$u_{ij}=\delta_y\psi_{ij}+\frac{h^2}{6}\left(\delta_y\omega_{ij}+\delta_x^2\delta_y\psi_{ij}\right)+O(h^4) \tag{19}$$

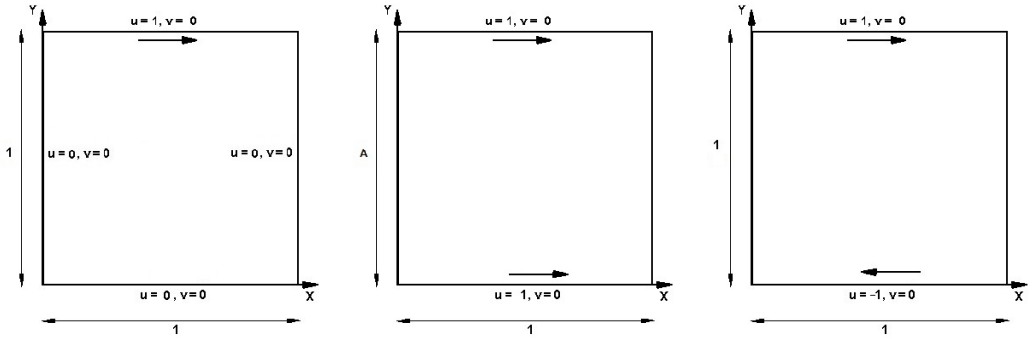
$$v_{ij}=-\delta_x\psi_{ij}-\frac{h^2}{6}\left(\delta_x\omega_{ij}+\delta_x\delta_y^2\psi_{ij}\right)+O(h^4) \tag{20}$$

The value of the streamfunction ψ is taken to be zero on all the boundaries while the Neumann boundary condition for vorticity is derived using a fourth-order compact scheme. For example, on the leftmost wall ($x=0, 0\le y\le 1$), the approximation for ω can be found from the relation $v=-\psi_x$ and equations (15) and (16) to get

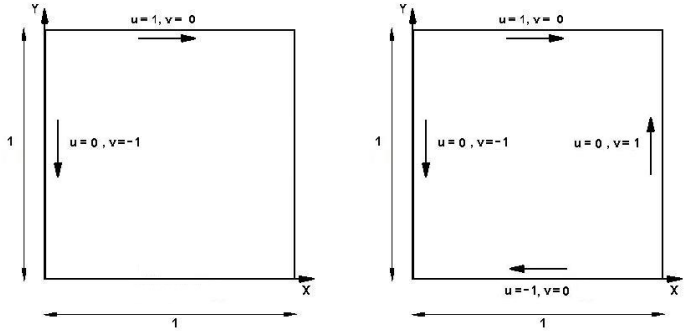
$$-\delta_x^+\psi_{0j}-\left[\frac{h}{2}+\frac{h^2}{6}\delta_x^+-\frac{h^3}{24}\left(Rev_{0j}\delta_y-\delta_y^2\right)\right]\omega_{0j}=v_{0j}-\frac{h^3}{24}\left(\delta_x^+\delta_y v_{0j}-\delta_t\omega_{0j}\right) \tag{21}$$

where the suffixes 0 and j stand for the leftmost wall and the vertical space index respectively. Using the boundary conditions for left wall, i.e., $v_{0j}=0$ and $\psi_{0j}=0$, the vorticity at the left wall can be explicitly written as

$$\omega_{0j}^{n+1}=\frac{24\Delta t}{h^3}\left[-\frac{\psi_{1j}^n}{h}-\frac{h}{2}\omega_{0j}^n-\frac{h}{6}\left(\omega_{1j}^n-\omega_{0j}^n\right)-\frac{h}{24}\left(\omega_{0j+1}^n-2\omega_{0j}^n+\omega_{0j-1}^n\right)\right]+\omega_{0j}^n \tag{22}$$

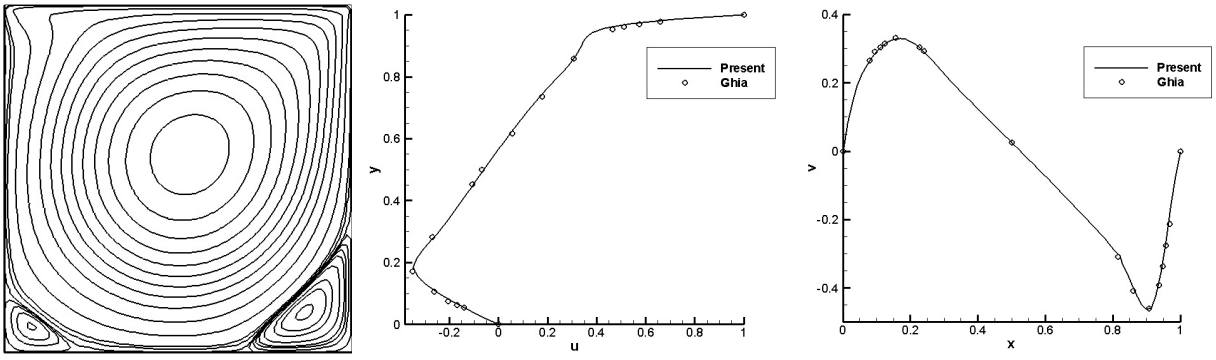


(a) Single lid-driven Cavity (b) Two-sided lid-driven cavity (parallel motion of facing walls) (c) Two-sided lid-driven cavity (antiparallel motion of facing walls)



(d) Two-sided lid-driven cavity (motion of non-facing walls) (e) Four-sided lid-driven cavity

Figure 1: Geometry and boundary conditions for cavities considered in this work.



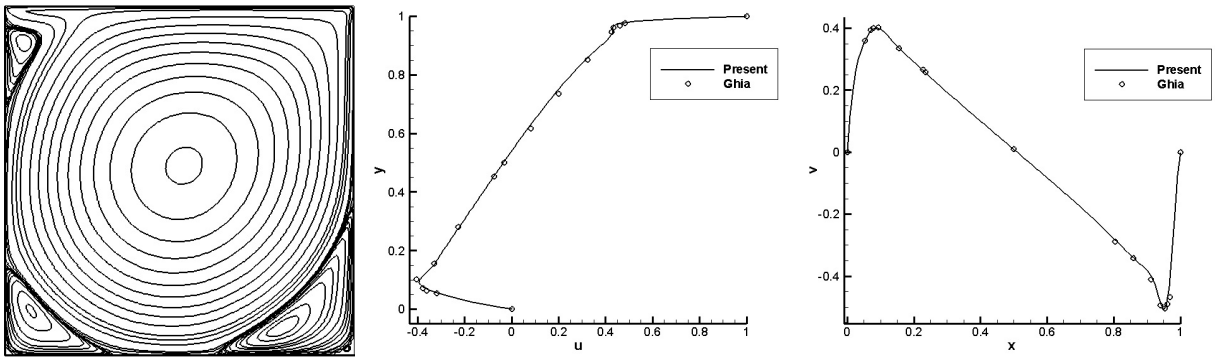
(a) Streamlines (b) u velocity along vertical centreline (c) v velocity along horizontal centreline

Figure 2: Code validation: single lid-driven square-cavity ($Re = 1000$).

The boundary conditions for the other walls can be derived in a similar manner.

For corners it is not possible to obtain fourth-order expressions of vorticity due to geometry constraints. Hence we use a third-order approximation. For example, for the upper left corner (x_0, y_{N-1}), vorticity can be written by approximating equations (17) and (18) in both $x-$ and $y-$directions and approximating the higher-order terms appropriately. This results in

$$\left[\frac{h}{2} + \frac{h^2}{6}\left(\delta_x^+ - \delta_y^-\right)\right]\omega_{0,N-1} = -\left[\delta_x^+ + \delta_y^-\right]\psi_{0,N-1} - u_{0,N-1} - v_{0,N-1}$$



(a) Streamlines (b) u velocity along vertical centreline (c) v velocity along horizontal centreline

Figure 3: Code validation: single lid-driven square-cavity ($Re = 5000$).

$$-\frac{h^2}{6}\left[\delta_x^+\delta_y^- u_{0,N-1} + \delta_x^+\delta_y^- v_{0,N-1}\right] + O(h^3) \tag{23}$$

where the suffix $N-1$ denotes all the points lying on $y = 1$. The boundary conditions for other corners can be written in a similar manner. The procedure for boundary conditions is outlined in [18].

The results presented here are on a 101 X 101 grid for Re's 1000 and 5000 (Figs. 2 and 3). The obtained results are in very good agreement even for a relatively high Reynolds number ($Re = 5000$) on a coarse grid of 101 X 101. This shows the higher-order nature of the scheme. Thus the present HOC code stands validated.

4 Results and Discussion

4.1 Two-sided cavity - Motion of facing walls

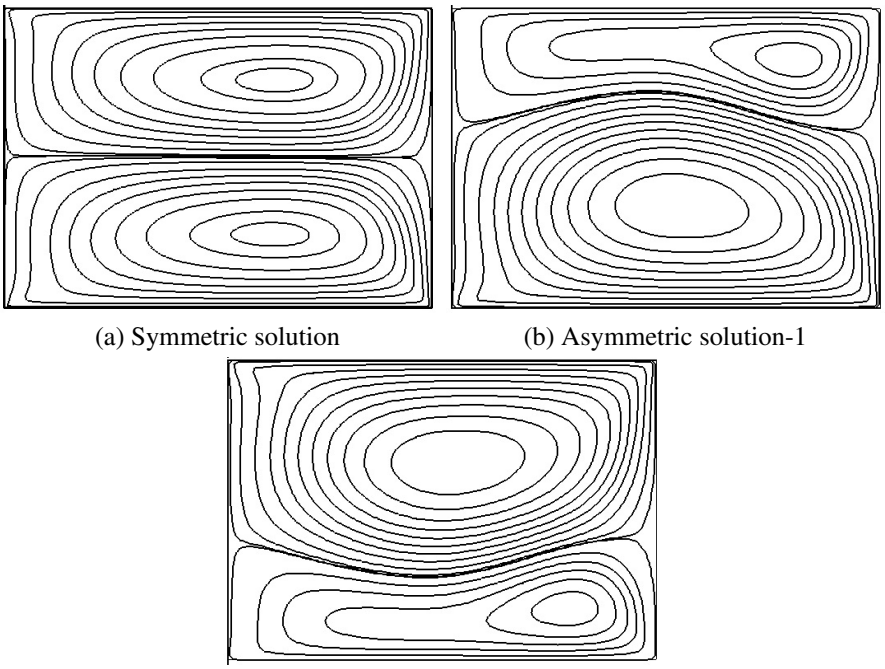
Figs. 1(b) and 1(c) show the two-sided cavity with parallel and antiparallel motion of facing walls respectively. All computations are performed at $Re = 600$ for parallel wall motion of rectangular cavity (Fig. 1(b)). Multiple solutions are obtained when $A \geq 0.556$. This value is in good agreement to the previously reported value of 0.559 [2]. Fig. 4 shows the multiple solutions obtained at $A = 0.65$. It is observed that one symmetric and a pair of asymmetric solutions exist at this aspect ratio. However at $A = 0.875$, an extra pair of weakly asymmetric solutions can also be seen to exist. Thus a total of five solutions exist for $A = 0.875$. These are shown in Fig. 5.

Computations are also carried out at various Re's for parallel motion of a square cavity (Fig. 1(b) with $A = 1$). It is seen that a total of three solutions exist for $Re \geq 983.5$. This threshold value is again in good agreement with previously investigations [2, 6]. Fig. 6 shows multiple solutions at $Re = 3500$ for this configuration.

For antiparallel motion of facing walls in a square cavity (Fig. 1(c)), multiple solutions are seen for $Re \geq 3203$ [12]. Below this Re value, the additional solutions merge to form a single solution. Fig. 7 shows the multiple solutions at $Re = 3300$.

4.2 Two-sided cavity - Motion of non-facing walls

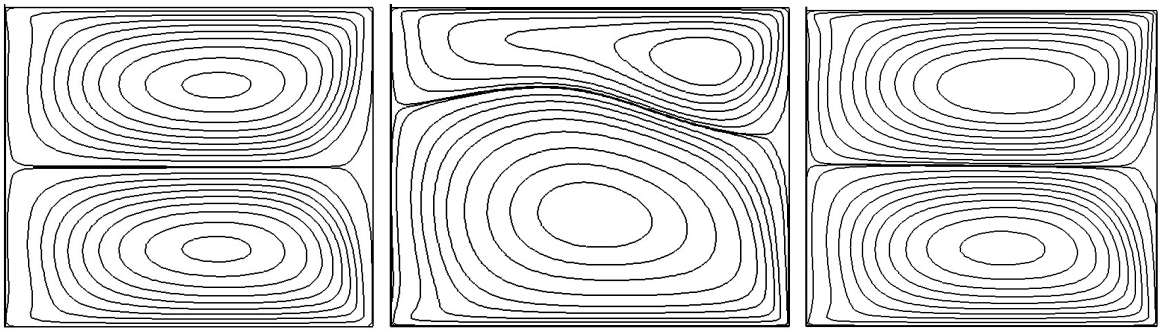
Using the same HOC scheme, an attempt is now made to re-examine the existence of multiple solutions for motion of non-facing walls of a square cavity. Existing literature has shown that a total of three solutions exist above a critical Re value of 1071 [8]. In this work, we show that



(a) Symmetric solution (b) Asymmetric solution-1

(c) Asymmetric solution-2

Figure 4: Multiple solutions of parallel wall motion for $A = 0.65$ at $Re = 600$.



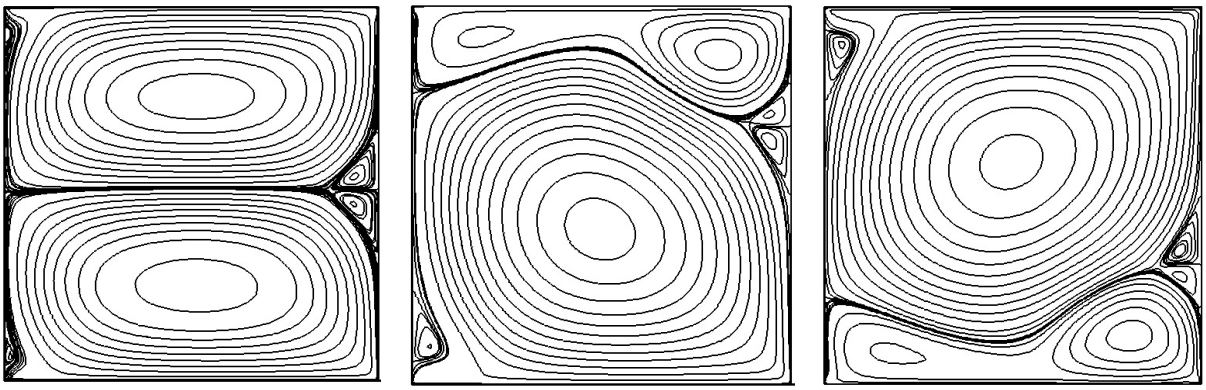
(a) Symmetric solution (b) Asymmetric solution (c) Weakly asymmetric solution

Figure 5: 3 out of 5 multiple solutions of parallel wall motion for $A = 0.875$ at $Re = 600$.

these extra solutions can be obtained at a significantly lower Re value of 975. Figs. 8 and 9 show multiples solutions at $Re = 975$ and 2000 respectively. It may be noted that $\psi = 0$ along the main diagonal for the solutions shown in Figs. 8(a) and 9(a).

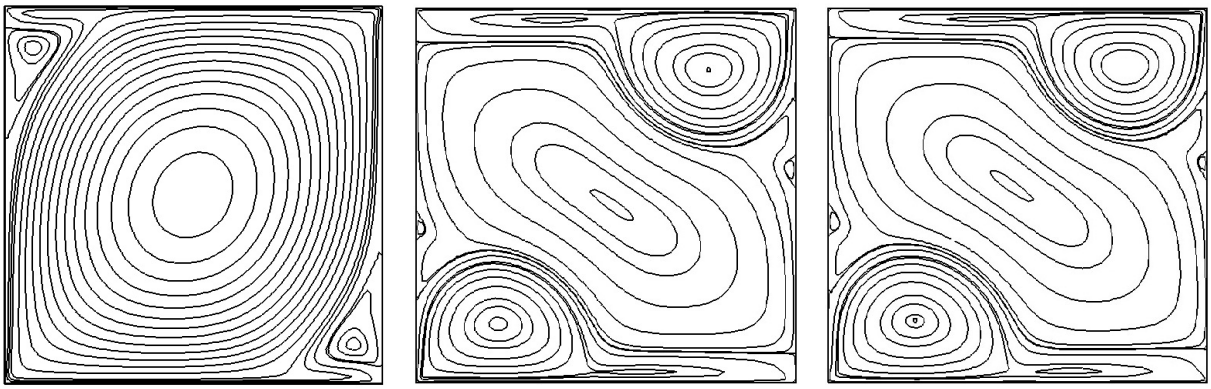
4.3 Four-sided cavity

The geometric configuration for four-sided cavity flow is given in Fig. 1(e). Wahba [8] obtained a threshold value of $Re = 129$ above which multiplicity of solutions was obtained. Using the HOC scheme (13), multiple solutions are seen for $Re \geq 130$. For $Re < 130$, only a single symmetric solution can be obtained. Figs. 10 and 11 show all the multiple solutions obtained at $Re = 150$ and 300 respectively.



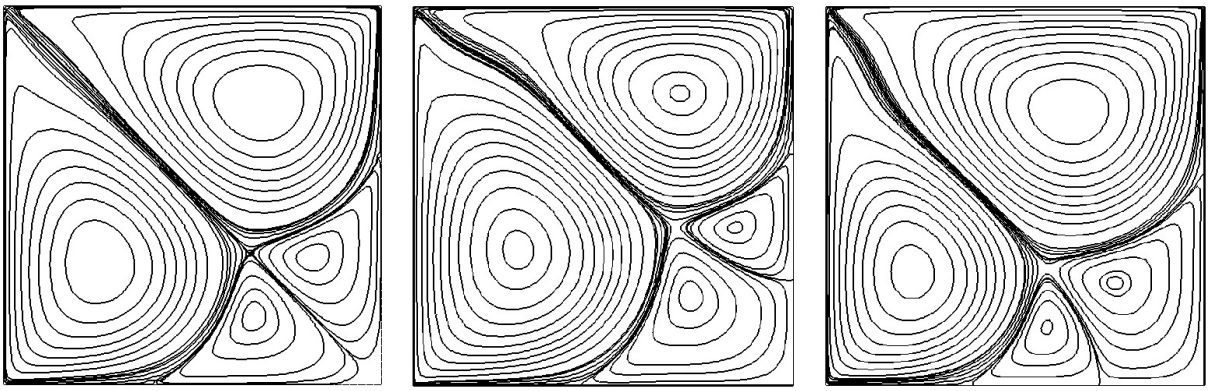
(a) Symmetric solution (b) Asymmetric solution 1 (c) Asymmetric solution 2

Figure 6: All solutions at $Re = 3500$ for parallel lid motion in a square cavity (Grid: 101 X 101).



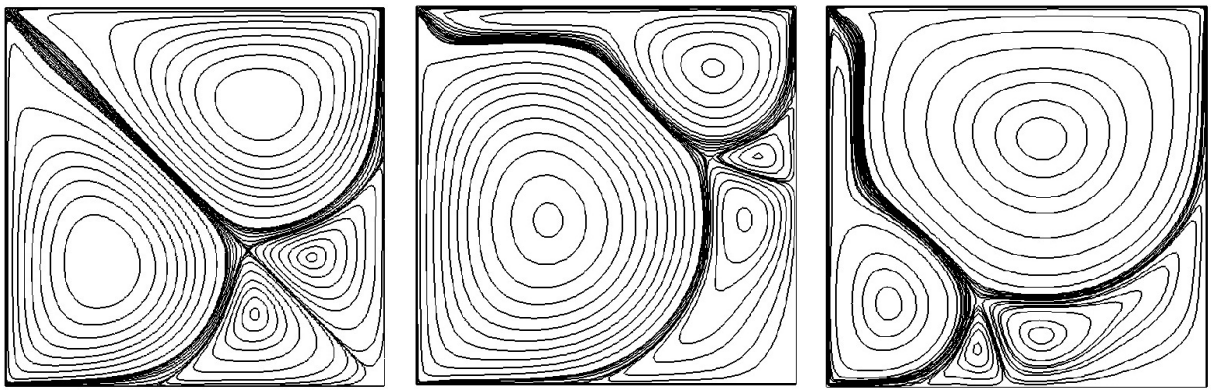
(a) Symmetric (b) Asymmetric 1 (c) Asymmetric 2

Figure 7: All solutions at $Re = 3300$ for antiparallel lid motion in a square cavity (Grid: 101 X 101).



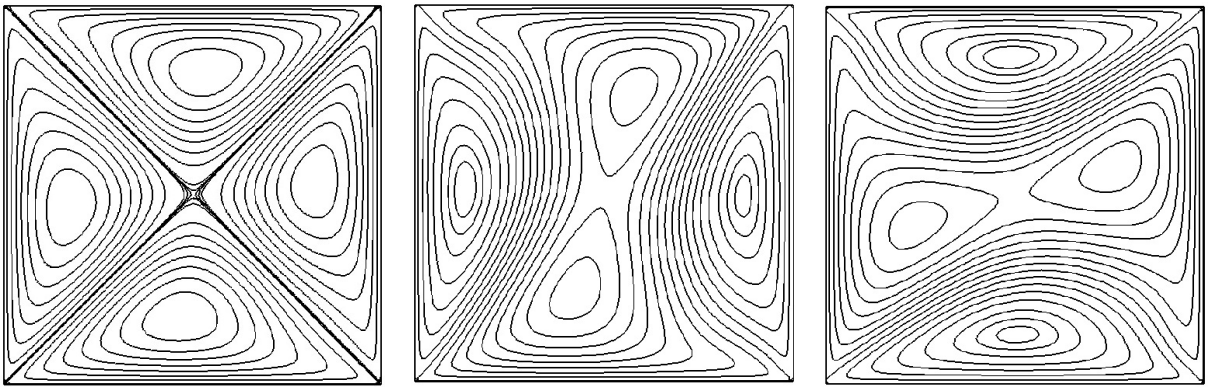
(a) Symmetric solution (b) Asymmetric solution 1 (c) Asymmetric solution 2

Figure 8: All solutions at $Re = 975$ for motion of non-facing walls in a square cavity (Grid: 101 X 101).



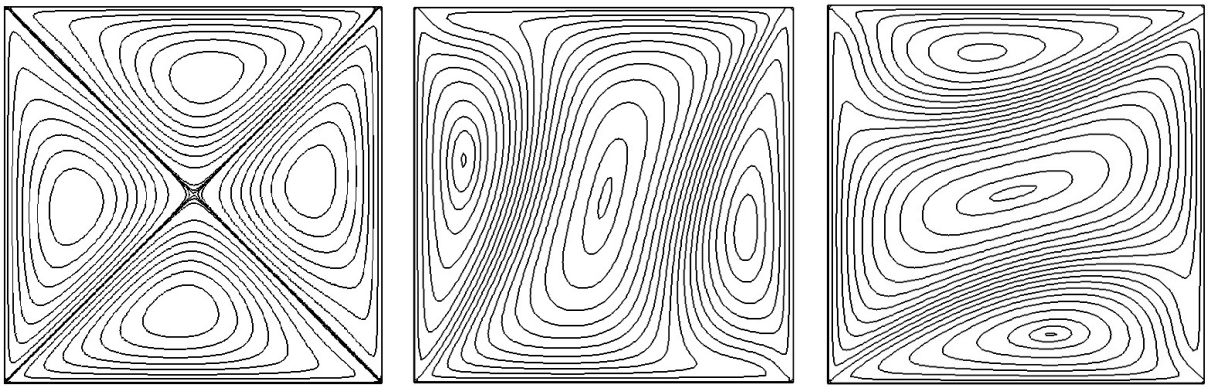
(a) Symmetric solution (b) Asymmetric solution 1 (c) Asymmetric solution 2

Figure 9: All solutions at $Re = 2000$ for motion of non-facing walls in a square cavity (Grid: 101 X 101).



(a) Symmetric solution (b) Asymmetric solution 1 (c) Asymmetric solution 2

Figure 10: All solutions at $Re = 150$ for four-sided square cavity (Grid: 101 X 101).



(a) Symmetric solution (b) Asymmetric solution 1 (c) Asymmetric solution 2

Figure 11: All solutions at $Re = 300$ for four-sided square cavity (Grid: 101 X 101).

Table 1: Comparison of current computation with previous investigations

Cavity Configuration	Threshold Parameter	
	HOC	non-compact
Two-sided rectangular cavity with parallel facing wall motion ($Re = 600$)	$A \geq 0.556$	$A \geq 0.559$ [2]
Two-sided square cavity with parallel facing wall motion	$Re \geq 983.5$	$Re \geq 980$ [6] $Re \geq 990$ [2]
Two-sided square cavity with antiparallel facing wall motion	$Re \geq 3203$	-
Two-sided square cavity with non-facing wall motion	$Re \geq 975$	$Re \geq 1071$ [8]
Four-sided square cavity	$Re \geq 130$	$Re \geq 129$ [8]

5 Conclusion

A higher-order compact scheme of fourth-order spatial and second-order temporal accuracy is used to examine multiple stable steady-state solutions for both two-sided and four-sided cavity flows. For two-sided cavity flows with parallel motion of facing walls at $Re = 600$, three to five multiple steady solutions are obtained depending on the cavity aspect ratio. These solutions consist of a symmetric solution and one or two pairs of asymmetric solutions. It is seen that multiple solutions do not exist and only the symmetric solution exists for the movement of the longer plate below an aspect ratio of 0.556. For parallel lid motion of facing walls in a square cavity (when aspect ratio equals one) only symmetric solutions exist below $Re = 983.5$. The threshold values, of aspect ratio for rectangular and Re for square cavities, are carefully computed and they are close to those obtained in earlier investigations. Multiple solutions are also presented for antiparallel of facing walls in a square cavity. For two-sided cavity flows with motion of non-facing walls, we establish the existence of multiple solutions at $Re = 975$, which is significantly below the previously known threshold value of 1071. An exhaustive grid independence exercise was undertaken in order to obtain this threshold value. Multiple solutions are computed for four-sided square cavity flows as well. The critical Re value above which the multiple solutions exist in this configuration is in good agreement with previous investigations. It is well known that many nonlinear flow problems exhibit multiple solutions and this work demonstrates the ability of HOC schemes to obtain such solutions that violate the desirable mathematical condition of well-posedness.

References

[1] Perumal DA, Dass AK. Multiplicity of steady solutions in two-dimensional lid-driven cavity flows by lattice boltzmann method. *Computers & Mathematics with Applications* 2011; **61**(12):3711–3721.

[2] Albensoeder S, Kuhlmann H, Rath H. Multiplicity of steady two-dimensional flows in two-sided lid-driven cavities. *Theoretical and Computational Fluid Dynamics* 2001; **14**(4):223–241.

[3] Kuhlmann H, Wanschura M, Rath H. Elliptic instability in two-sided lid-driven cavity flow. *European Journal of Mechanics-B/Fluids* 1998; **17**(4):561–569.

[4] Kuhlmann H, Wanschura M, Rath H. Flow in two-sided lid-driven cavities: non-uniqueness, instabilities, and cellular structures. *Journal of Fluid Mechanics* 1997; **336**:267–299.

[5] Bruneau CH, Saad M. The 2d lid-driven cavity problem revisited. *Computers & Fluids* 2006; **35**(3):326–348.

[6] Lemée T, Kasperski G, Labrosse G, Narayanan R. Multiple stable solutions in the 2d symmetrical two-sided square lid-driven cavity. *Computers & Fluids* 2015; **119**:204–212.

[7] Peng YF, Shiau YH, Hwang RR. Transition in a 2-d lid-driven cavity flow. *Computers & Fluids* 2003; **32**(3):337–352.

[8] Wahba E. Multiplicity of states for two-sided and four-sided lid driven cavity flows. *Computers & Fluids* 2009; **38**(2):247–253.

[9] Blohm C, Kuhlmann HC. The two-sided lid-driven cavity: experiments on stationary and time-dependent flows. *Journal of Fluid Mechanics* 2002; **450**:67–95.

[10] Luo WJ, Yang RJ. Multiple fluid flow and heat transfer solutions in a two-sided lid-driven cavity. *International journal of heat and mass transfer* 2007; **50**(11):2394–2405.

[11] Kalita JC, Dalal D, Dass AK. A class of higher order compact schemes for the unsteady two-dimensional convection–diffusion equation with variable convection coefficients. *International Journal for Numerical Methods in Fluids* 2002; **38**(12):1111–1131.

[12] Prasad C, Dass AK. Use of an hoc scheme to determine the existence of multiple steady states in the antiparallel lid driven flow in a two-sided square cavity. *Computers & Fluids* 2016; **submitted for publication**.

[13] Ghia U, Ghia KN, Shin C. High-re solutions for incompressible flow using the navier-stokes equations and a multigrid method. *Journal of computational physics* 1982; **48**(3):387–411.

[14] Spotz WF, Carey G. *High-order compact finite difference methods with applications to viscous flows*. Citeseer, 1994.

[15] Abarbanel S, Kumar A. Compact high-order schemes for the euler equations. *Journal of Scientific Computing* 1988; **3**(3):275–288.

[16] Lele SK. Compact finite difference schemes with spectral-like resolution. *Journal of computational physics* 1992; **103**(1):16–42.

[17] Bassi F, Rebay S. A high-order accurate discontinuous finite element method for the numerical solution of the compressible navier–stokes equations. *Journal of computational physics* 1997; **131**(2):267–279.

[18] Spotz W, Carey G. High-order compact scheme for the steady stream-function vorticity equations. *International Journal for Numerical Methods in Engineering* 1995; **38**(20):3497–3512.

Research on complex hydrodynamic interaction when UUV recovered by submarine

***LUO Yang, †PAN Guang, Yang Zhi-dong, HUANG Qiao-gao, and QIN Deng-hui**

College of Marine, Northwestern Polytechnical University, Xi'an, China
*Presenting author: lawyer0818@hotmail.com
†Corresponding author: panguang601@163.com

Abstract

Hydrodynamic interaction performance between an unmanned underwater vehicle (UUV) and a submarine was presented using Reynolds Average Navier-Stokes (RANS) techniques, when submarine was recovering an UUV. The hydrodynamic characteristics of UUV in different positions relative to submarine was simulated numerically based on RANS techniques, and the variation of UUV's hydrodynamic coefficients interfered by flow around the submarine was analyzed. Then combined with the dynamic grid techniques, unsteady hydrodynamic performance was numerically calculated when an UUV performed parallel movement and vertical movement relative to the longitudinal axis of the submarine, and the changing law of the hydrodynamic coefficients of UUV under corresponding conditions was revealed. The method presented could predict the maneuvering and controlling performance of the UUV retrieved to a submarine.

Keywords: Computational Fluid Dynamics, Hydrodynamic Interaction, Unmanned Underwater Vehicle, Submarine.

Introduction

With the exploitation of marine resources being intensified and more extensive military applications, the Unmanned Underwater Vehicle (UUV) is required to have longer operation time. For an UUV equipped in a submarine, energy refuel and information exchange can be achieved through underwater recovery[1]. Ronald W. Yeung and Wei-Yuan Hwang[2] has predicted nearfield hydrodynamic interactions of ships in shallow water based on the slender-body theory. H. Zhang[3] et al. studied effect of turbulence intensity to the fluid dynamic interference of two cylinders side in side. Zeng Yifei[4] presented equivalent extension-body method to calculate the interaction between two underwater cylinders in relative motion. Y.R. Choi[5] investigated the hydrodynamic interference between floating multi-body system with the boundary element method. Wang Fei[6] developed a program based on the panel method and calculated the hydrodynamic performance of an underwater vehicle in motion near the submarine. B.J. Koo[7] and S.Y. Hong[8] simulated numerically the hydrodynamic interaction and mechanical coupling effects of two floating platforms and vessels connected by elastic lines respectively. Chen Li and Zhang Liang et al.

pointed that the minimal interaction path exists for underwater bodies in approaching process in theory from the computation results of a cylinder near a plane wall according to potential flow theory[9], and according to the combination bodies in periodic heave and pitch motion in unbounded flow field and near a plane wall, they used unsteady theory method to calculated nonlinear unsteady interference force related to the vortex evolution and motion. They, based on which, also revealed the typical characteristics of unsteady hydrodynamic interference. The hydrodynamic characteristics of a 2D oval with length-thickness ratio 7.0 while moving near plane wall were presented through towing-tank tests by them, then, they gave the regressive formula of hydrodynamic coefficients relative to clearances, attack angles and divided three typical interaction regions, defined as Lifting, Mixed and Blocking Region[10]-[11]. HeYuzhi[12] simulated the process that an UUV approached a conical docking device with numeral method and obtained the change law of drag coefficient and lift coefficient of the UUV during which. Leong, Z. [13] investigated the hydrodynamics performance when an AUV moved in various horizontal and vertical positions of a submarine at a series of relative speeds with CFD method based on N-S equation. S.A.T. Randeni P.[14] et al. investigated the hydrodynamic interaction between an AUV operating in close proximity to a submarine, with the development of a CFD model to replicate the pure sway motion of the AUV and figured out that the percentage difference between the CFD and EFD (experimental fluid dynamics) sway forces were generally below 6%.

However, UUV's appendages were not taken into consideration in most of the researches above. Steady and unsteady hydrodynamic performance between an UUV and a submarine was numerically calculated based on RANS techniques and the results of UUV models with appendages were compared with which without attached parts in this paper. The condition in focus is more complex and more applicable to engineering reality, which may provide a reference for prediction and analysis on the hydrodynamic performance during UUV's underwater recovery.

1 Calculated Models

1.1 Geometric models and calculated conditions

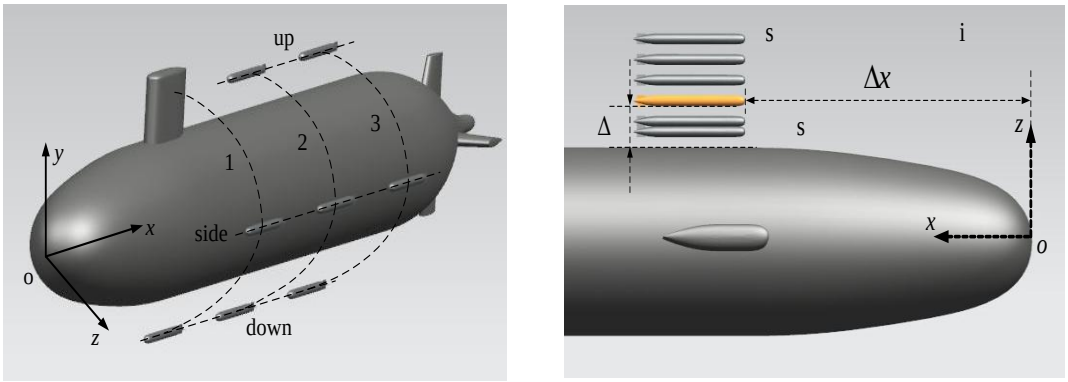
An axisymmetric body with characteristic diameter D = 0.534m and length L = 7m was elected as UUV model and the model without appendages was defined as UUV-1 while the other one with which was identified as UUV-2. The distance from the buoyancy center of the UUV-1 to the head is 3.2446m while which of the UUV-2 is 3.2499m. As for submarine model, the research utilized a 1:18 scaled model of the SUBOFF submarine hullform as the submarine's main body and defined as SUB-1, while the model with complete appendages including the main body, fairwater and caudal fin as UUV-2. The full length L_s of the submarine model is 78.408m, length between perpendiculars L_{pp} is 76.698m and maximum diameter D_s of which

equals 9.144m. In addition, dimensionless fluid force and torque coefficients are defined as

$$C_{xx} = f_{xx} / \left(\frac{1}{2} \rho S_a U^2 \right), \quad M_{xx} = m_{xx} / \left(\frac{1}{2} \rho S_a U^2 \right) \qquad (1)$$

In the formula (1), f_{xx} and m_{xx} represent the component of fluid force and torque along the direction of the coordinate system *xx* respectively, while C_{xx} and M_{xx} correspond to the dimensionless coefficient of fluid force and torque components. The largest cross-sectional area of the UUV model was selected as S_a and UUV's full length was considered as L_a in dimensionless hydrodynamic coefficients. The L_a and S_a of submarine's dimensionless hydrodynamic coefficients were set as L_{pp} and its square accordingly.

In order to facilitate the description of working conditions and the analysis of calculation results, a coordinate system *oxyz* was established and the feature positions of UUV relative to the submarine were defined as shown in Figure 1.



（a）3D **（b）*xoz***

Figure 1 The coordinate system definition and feature positions of UUV relative to the submarine

As shown in Figure 1 (a), the central point of the bow was defined as the coordinate origin, the *ox* axis is along the vertical symmetry axis of the submarine, *oy* axis lies in submarine's longitudinal symmetry plane and vertical to the *ox* axis, and the *oz* axis meets the right-hand rule. The position of the UUV relative to the submarine was determined by its longitudinal position(the coordinate of the UUV along the *ox* axis) and relative direction, and chose three different longitudinal positions along the *ox* axis relative to the submarine and labeled as "1, 2, 3" respectively. When the UUV model situates in the submarine's lateral plane, it is recorded as "side"; in the submarine's vertical plane and the positive *ox* axis, it is denoted as "up", otherwise as

"down". Based on the definition of the above markers, the feature orientation of the UUV relative to the submarine can be represented as "side1" and "down2". As depicted in Figure 1 (b), take "side1" as the example, the distance from the UUV's wall to the submarine's was marked as Δs. Therefore, the feature position of the UUV relative to the submarine can be obtained with the combination of feature orientation and Δs, which was denoted as "side1-Δs". The *ox* coordinate of the UUV's head point was marked as Δx. In order to study the influence of UUV's three different longitudinal position: near the head, tail and central of the submarine, on the hydrodynamic coefficients, the values of Δs corresponding to three different longitudinal position labeled as "1, 2, 3," are shown in Table 1.

Table 1 The values of Δs corresponding to different longitudinal position

ox axis label	1	2	3
Distance Δx (m)	18	31.5	45

1.2 Meshing

A rectangular domain with a size of $5L_s \times 20D_s \times 20D_s$ was chosen as computational domain. The SUBOFF model was arranged in the center of the domain, and the distance from the velocity inlet to the head of the model is $1.5L_s$. Except for the velocity inlet and pressure outlet, the four remaining faces of the rectangular domain were set as slip wall.

Structured grid was utilized for computational domain meshing, and based on based on the concept of block partition, the computational domain of a single submarine was generated firstly, then a portion of grid blocks in the overall calculation domain was excised and embeded in the grid block containing the UUV model inside the overall domain through internal interface to complete the grid generation, which is as shown in Figure 2 and Figure 3.

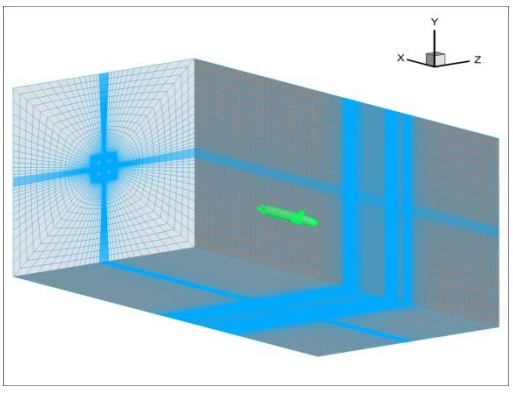


Figure 2 The overall grid generation

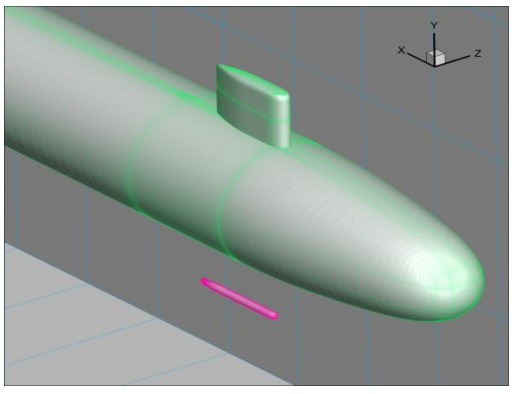


Figure 3 The merged grid generation

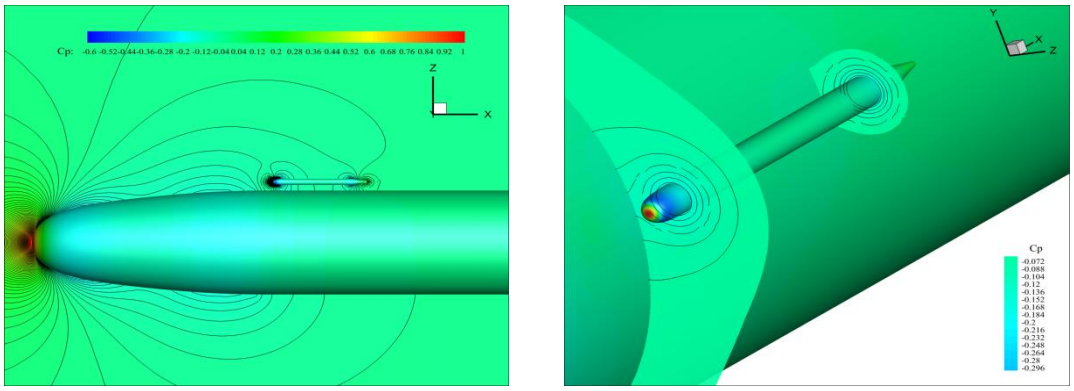
For the unsteady motion, used dynamic layering method to simulate the motion of the interface of the UUV and whole grid in the dynamic area relative the submarine.

2 Analysis of steady hydrodynamic interference between the UUV and submarine

2.1 Results and analysis of hydrodynamic interference of models without appendages

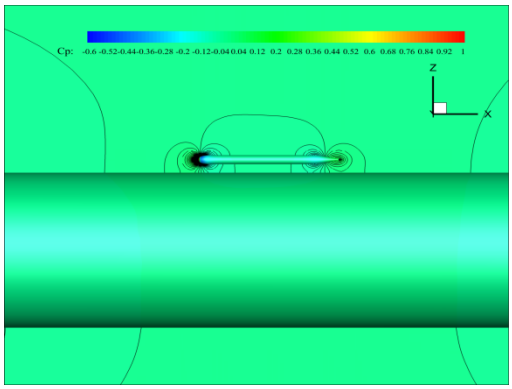
The turbulence model selected was RNG $k-\varepsilon$ and set inlet velocity as 2 kn ignoring the influence of gravity.

Figure 4 to Figure 6 shows the pressure coefficient distribution corresponding to three feature orientations: “side1”, “side2” and “side3” when Δs = 0.5. From the pressure coefficient contours in the lateral *xoz* section it can be seen that: in “side1” feature orientation the head of the UUV-1 model was close to the low pressure area of the SUB-1 head, and influenced by which the pressure drag of the UUV-1 got smaller than that in unbounded flow condition, even resulted in a pressure surplus (a negative value). Therefore, the hydrodynamic interference of the flow around the submarine to the UUV-1 performed as drag reduction. In “side2” feature orientation, the tail of the UUV-1 model was close to the low pressure area of the SUB-1 aft body and influenced by which the pressure drag of the UUV-1 got bigger than that in unbounded flow condition, so the hydrodynamic interference increased the resistance. While in “side2” feature orientation, the UUV-1 was in the stable flow field near the parallel middle part of the SUB-1, and the low pressure area near the SUB-1 tail had little effect on the UUV-1, as a result, the pressure drag of the UUV-1 approximated that in unbounded flow condition. It can be seen from pressure coefficient contours in the local transverse *yoz* section that the isobar shaped as an inverted “C” type, namely a low pressure region formed in the adjacent zone of the UUV-1 and SUB-1, which shows that the SUB-1 acted suction on the UUV-1.

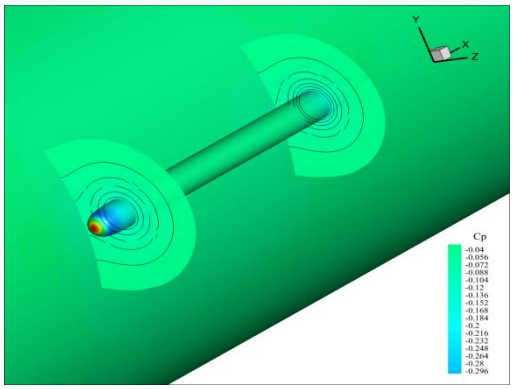


（a）The lateral *xoz* section **（b）The local transverse *yoz* section**

Figure 4 The pressure coefficient distribution in the position “side1-0.5m”

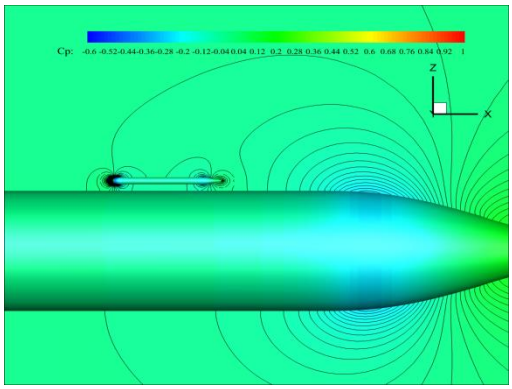


（a）The lateral *xoz* section

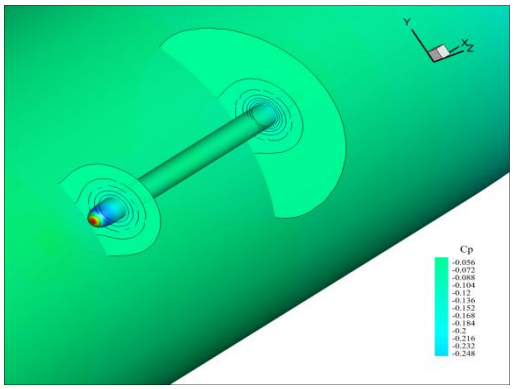


（b）The local transverse *yoz* section

Figure 5 The pressure coefficient distribution in the position "side2-0.5m"



（a）The lateral *xoz* section



（b）The local transverse *yoz* section

Figure 6 The pressure coefficient distribution in the position "side3-0.5m"

Now defined the dimensionless distance from the UUV's wall to the submarine's as $\Delta\bar{s}$ and the UUV's diameter was selected as the feature space. The variation of the drag coefficient C_x with $\Delta\bar{s}$ according to the three feature positions of the UUV-1 above was as shown in Figure 7. Regarded the drag coefficient of the UUV-1 in unbounded flow field as reference value, Figure 8 shows the change percentage of the resistance coefficient with $\Delta\bar{s}$.

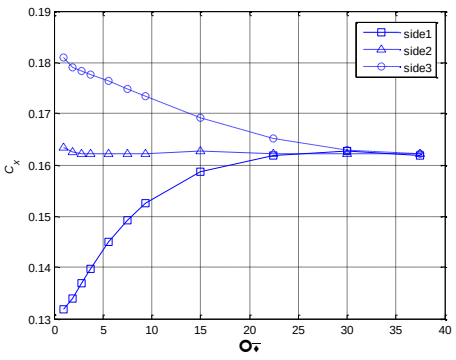


Figure 7 C_x corresponding to $\Delta\bar{s}$

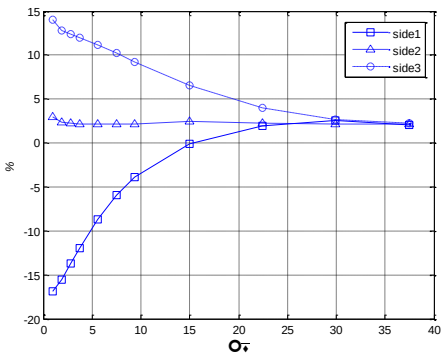


Figure 8 The change percentage of C_x

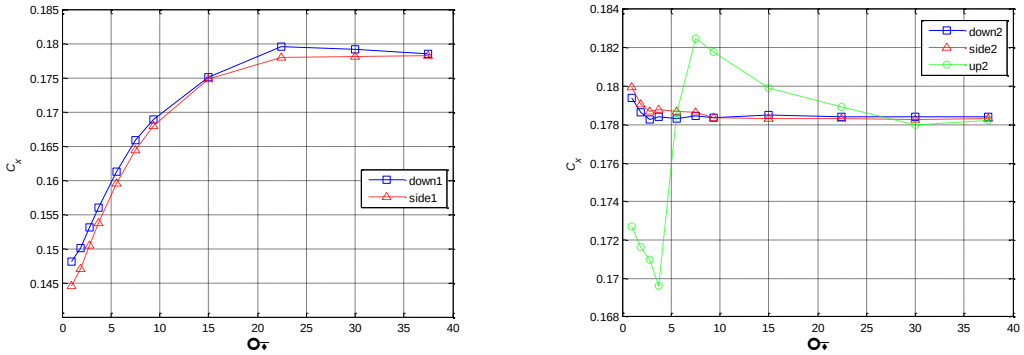
According to Figure 7 and Figure 8 it can be found that when the UUV-1 was located in "side1" near the head of the submarine, C_x of the UUV-1 decreased with the decrease of $\Delta\bar{s}$; while in "side2" where the flow field was stable, C_x varied little; however, C_x increased with the decrease of $\Delta\bar{s}$ in "side3". And C_x approached to the value in unbounded flow field with the increase of $\Delta\bar{s}$ under the three conditions. When $\Delta\bar{s}$ >30, C_x of the UUV-1 corresponding to the three conditions all tended to converge, in another word, the interference function distance of the flow around the SUB-1 on UUV-1 model is about 30 times its diameter.

2.2 Results and analysis of hydrodynamic interference of models with appendages

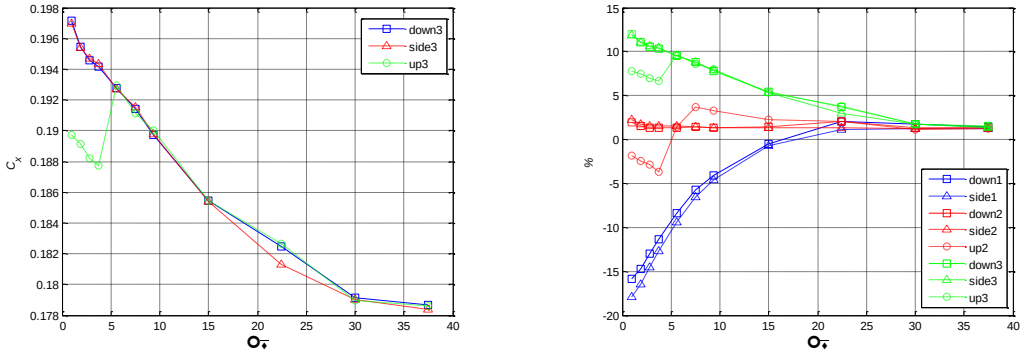
The turbulence model selected was also RNG k-ε and set inlet velocity as 2 kn ignoring the influence of gravity as well as the condition without appendages.

The model with full appendages involved eight different feature orientations. For the convenience to analyze the hydrodynamic interference of the flow field in different feature orientations, defined the plane determined by the longitudinal axes of the SUB-2 and SUB-2 as the main interference plane, based on which, the main interference force coefficient along the vertical direction to the *ox* axis was marked as C_{xx}, and suction was recorded as positive, repulsion as negative; the main interference moment coefficient vertical to the main plane was denoted as M_{xx}, the moment deviating the head of the model UUV-2 from the SUB-2 was denoted by positive, otherwise negative.

The numerical results of C_x and its change percentage compared to the unbounded condition corresponding to different feature orientations are shown in Figure 9.



(a) Result in longitudinal position "1 **（b）Result in longitudinal position "2"**



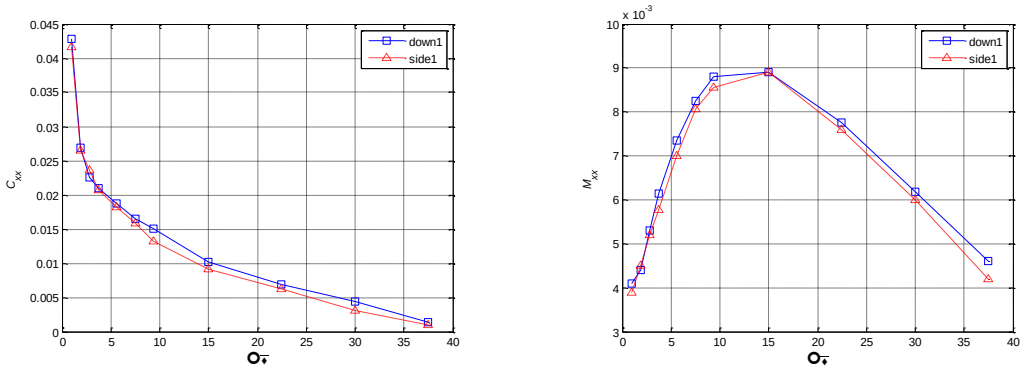
（c）Result in longitudinal position "3"　（d）The change percentage of C_x

Figure 9　Results of C_x and its change percentage compared to that in unbounded condition

It can be seen from Figure 9 that when the UUV was in the same longitudinal position and the relative direction was "side" and "up", the change laws of C_x with $\Delta\bar{s}$ were almost identical. For the condition "up2": when $\Delta\bar{s} \leq 4$, frictional resistance was small, pressure drag was large and the overall was small for the flow field around fairwater; when $\Delta\bar{s} > 5.6$, with the increase of $\Delta\bar{s}$, the model UUV-2 got close to the up edge of the wake flow and the interaction on frictional resistance decreased while the interference to frictional resistance came to an effect gradually; when $\Delta\bar{s} = 7.5$, C_x in the condition "up2" was larger than that in the conditions "side2-4m" and "down2-4m"; when $\Delta\bar{s} > 7.5$, the wake flow of the fairwater had little effect on the velocity field of UUV-2, with the increase of $\Delta\bar{s}$, the influence of local high pressure in the wake flow of the fairwater on the UUV-2 weakened and C_x decreased; when $\Delta\bar{s} = 22.5$, C_x approximated that in "side2" and "down2". For the condition "up3": the interaction of the wake flow of the fairwater on C_x mainly concentrated the range of $\Delta\bar{s} \leq 4$, when $\Delta\bar{s} > 5.6$, C_x approximated that in "side3" and "down3". From Figure (d), the interference distance of the flow around the SUB-2 on UUV-2 model is about 30 times its diameter.

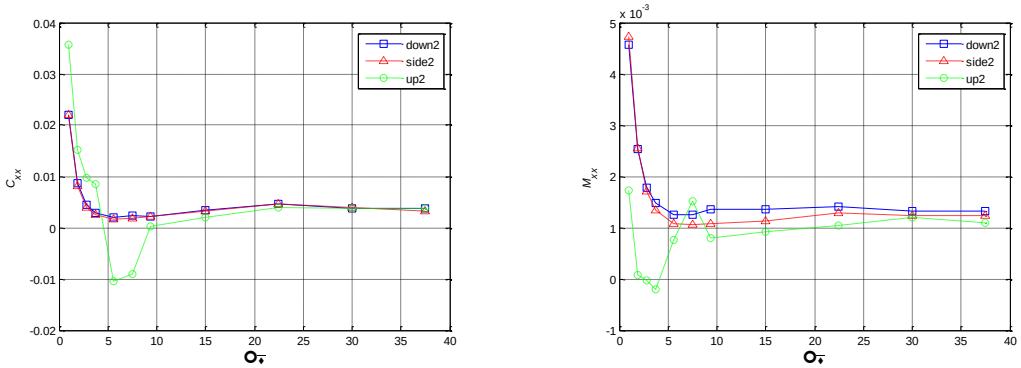
The calculated results of C_{xx} and M_{xx} of the UUV-2 in different longitudinal positions are as shown in Figure 10 to Figure 12. Graphical results show that when the UUV was in the same longitudinal position, and the relative directions were "side" and "up", the change laws of C_{xx} and M_{xx} with $\Delta\bar{s}$ were almost consistent, C_{xx}

both performed as suction; when $\Delta\bar{s}$ =5.6 and $\Delta\bar{s}$ =7.5, for the collective influence of the wake flow of the fairwater and submarine's external flow, C_{xx} corresponding to "up2" was negative, performed as repulsion; C_{xx} corresponding to "up3" was a small positive value close to zero, performed as slight suction; when $\Delta\bar{s} \geq 9.4$, the change laws of C_{xx} and M_{xx} with $\Delta\bar{s}$ in position "up" were almost consistent with the conditions in "side" and "down". When the longitudinal position was "1", M_{xx} were all positive, so the moment deviated the head of the UUV-2 from the SUB-2, and with the increase of $\Delta\bar{s}$, it increased firstly and then decreased, similar to parabola change rules; as the longitudinal position was "2", with the increase of $\Delta\bar{s}$, the change trends of M_{xx} in "side" and "down" conditions were almost identical and they both decreased at first and then remained stable approximately, but due to the influence of the wake flow of the fairwater , when $\Delta\bar{s} < 9.4$, the curve of M_{xx} in "up" condition is similar to an inverted "N" type, then M_{xx} tended to be stable as well; when the longitudinal position was "3", the change of M_{xx} was consistent with the condition in longitudinal position "2".



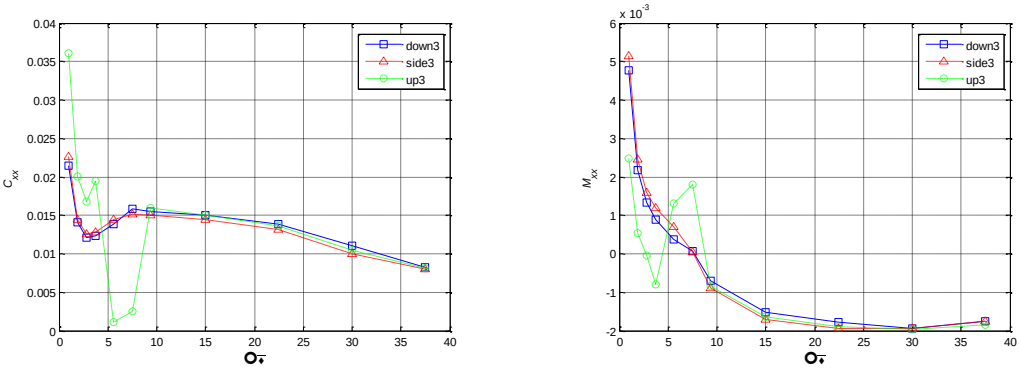
（a）C_{xx} （b）M_{xx}

Figure 10 C_{xx} and M_{xx} in longitudinal position "1"



（a）C_{xx} （b）M_{xx}

Figure 11 C_{xx} and M_{xx} in longitudinal position “2”

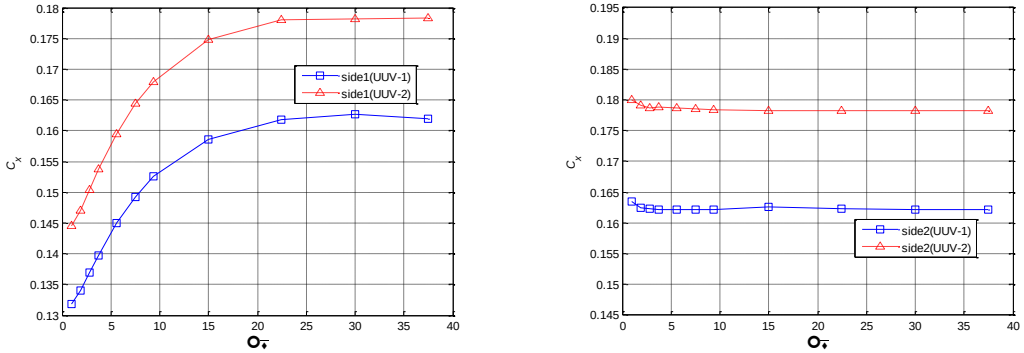


（a）C_{xx} （b）M_{xx}

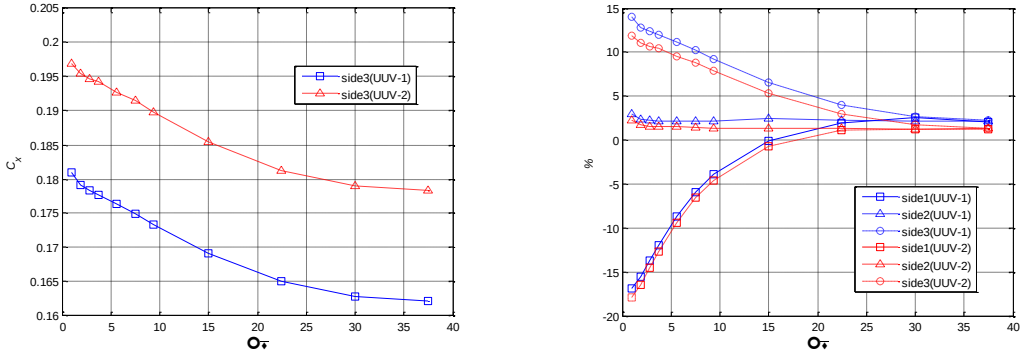
Figure 12 C_{xx} and M_{xx} in longitudinal position “3”

2.3 Comparison of computational results of models with and without appendages

Selected the calculated results of the model without appendages UUV-1 and the model with appendages UUV-2 in “side” position for comparision. Figure 13 shows C_x of the model UUV-1 and UUV-2 corresponding to $\Delta\bar{s}$ in three different “side” conditions. It can be seen from the comparative results that: while in the same feature orientation, for the UUV-1 and UUV-2, the change rules of C_x with $\Delta\bar{s}$ were similar to each other and the difference between C_x of the two models corresponding to the same $\Delta\bar{s}$ basically maintained at a certain range. Combined with Figure (d), the difference was close to that in the unbounded flow field, from which we can see the hydrodynamic interference of the submarine to the UUV mainly acts on its main body and has little effect on the fin.



（a）“side1”　　（b）“side2”

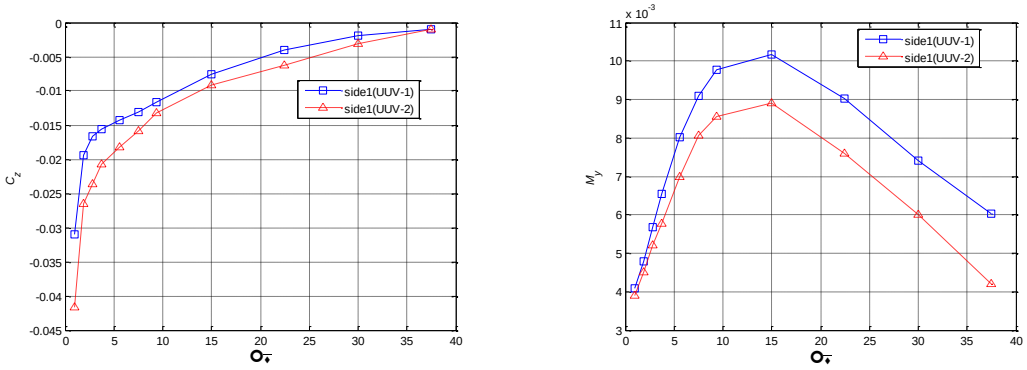


（c）“side3　　（d）The change percentage of C_x

Figure 13 C_x and its change percentage of the UUV-1 and UUV-2 in three “side” conditions

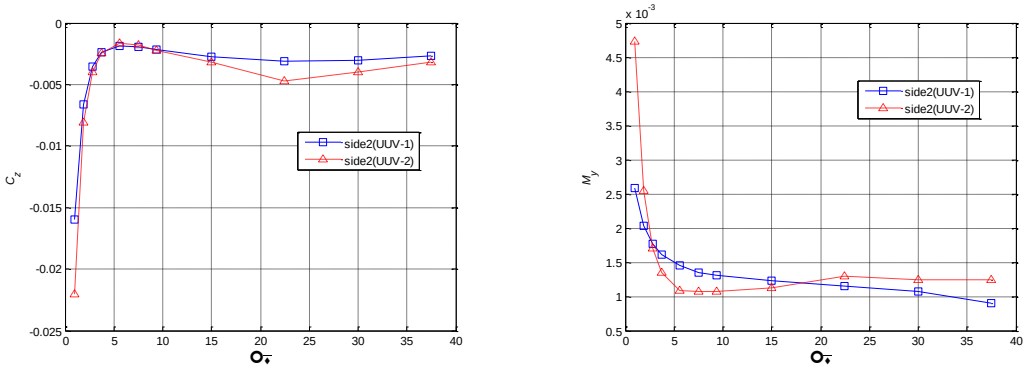
Figure 14 to Figure 16 show the side force coefficient C_z and yawing moment coefficient M_y of the model UUV-1 and UUV-2 corresponding to $\Delta\bar{s}$ in three different “side” conditions. From the comparative results we can find that in the same calculated condition, for the UUV-1 and UUV-2, the change rules of C_z and M_y with $\Delta\bar{s}$ were similar to each other. On the whole, the absolute value of C_z of the UUV-2 was bigger than that of the UUV-1, which illustrates that the side interference to the UUV was greater because of the appendages. For the condition “side1”, M_y of the UUV-2 was smaller than that of the UUV-1 corresponding to the same $\Delta\bar{s}$; for the condition “side2”, when $\Delta\bar{s}$ <2, M_y of the UUV-2 was larger than that of the UUV-1, while $\Delta\bar{s}$ >2, M_y of the two models were almost the same corresponding

to the same $\Delta\bar{s}$; for the condition “side3”, when $\Delta\bar{s}$ <3.7, the absolute value of M_y of the UUV-2 was bigger than that of the UUV-1, while $\Delta\bar{s}$ >3.7, the absolute value of M_y of the UUV-2 was smaller than that of the UUV-1.



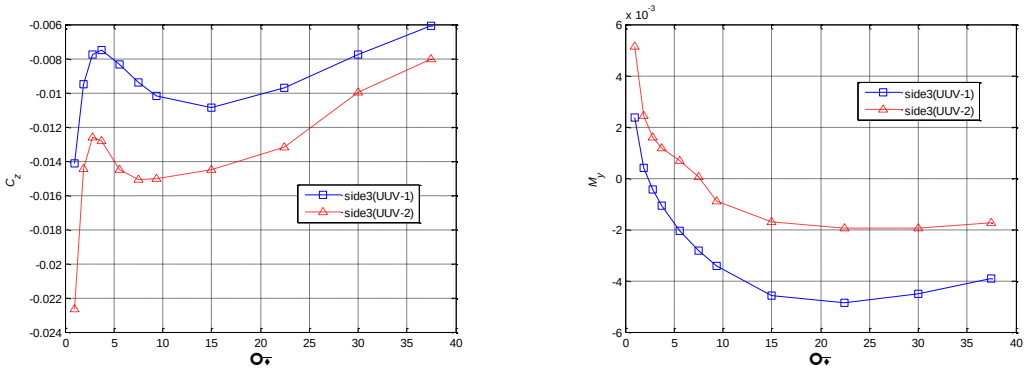
（a）C_z （b）M_y

Figure 14 C_z and M_y of the UUV-1 and UUV-2 in the condition “side1”



（a）C_z （b）M_y

Figure 15 C_z and M_y of the UUV-1 and UUV-2 in the condition “side2”



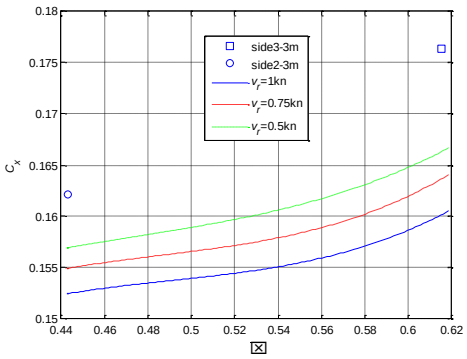
（a）C_z （b）M_y

Figure 16 C_z and M_y of the UUV-1 and UUV-2 in the condition "side3"

3 Analysis of unsteady hydrodynamic interference between the UUV and submarine

3.1 Simulation research on the motion of the UUV parallel to the submarine's longitudinal axis

Selected the light body UUV-1 and SUB-1 as the research objects and the simulation time step was 0.2s. Considering the condition that Δs =3m, the UUV-1 moved from the initial position side3-3 to side3-2 paralleled to the submarine's longitudinal axis at three different speeds 0.4kn, 0.75kn and 1kn, the change rules of the hydrodynamic coefficients of the UUV-1 with $\bar{x}$ were investigated. The specific calculation results are shown in Figure 17 to Figure 20.



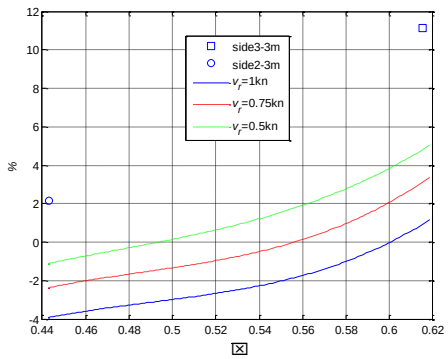


Figure 17 C_x corresponding to $\bar{x}$ **Figure 18 The change percentage of C_x**

Figure 17 and Figure 18 show C_x and its change percentage while the UUV-1 moving paralleled to the submarine's longitudinal axis at various speeds. From which we can figure out that the change laws of C_x with $\bar{x}$ were similar at different rates and ΔC_x between two conditions corresponding to different speeds was about a certain value. The larger the relative velocity was, the smaller C_x was corresponding to the same $\bar{x}$, that is, resistance of the UUV was smaller.

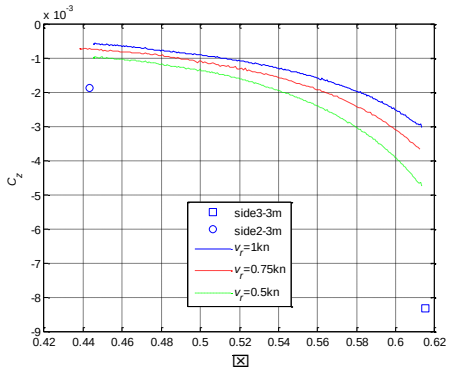


Figure 19 Calculated results of C_z

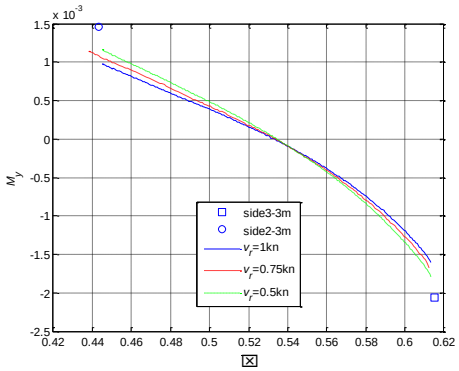
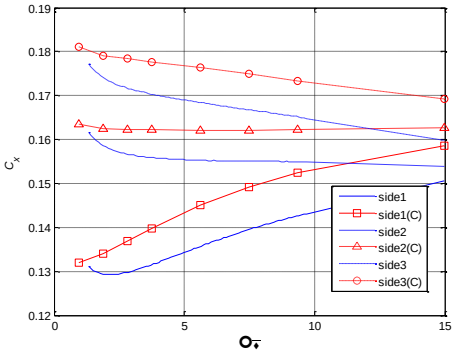


Figure 20 Calculated results of M_y

The change of C_z and M_y with $\bar{x}$ are shown in Figure 19 and Figure 20. From which we can see that the change laws of C_z with $\bar{x}$ were similar at different rates and they were all negative, implying the side force acting on the UUV-1 was suction; the larger the relative velocity was, the smaller the absolute value of C_z was corresponding to the same C_z, but the difference between which was small, in other words, improvement of the local Reynolds number of the UUV-1 can decrease the interference of the submarine on its side force coefficient slightly; the influence of different velocities on M_y was the same as C_z, i.e. improvement of the local Reynolds number of the UUV-1 can also decrease the interference of the submarine on its side force coefficient modestly.

3.2 Simulation research on the motion of the UUV vertical to the submarine's longitudinal axis

We also elected the light body UUV-1 and SUB-1 as the research objects and the simulation time step was 0.2s. Considering the condition that the UUV-1 approached the submarine vertical to the its longitudinal axis from three initial positions side1-8m, side2-8m and side3-8m, the change rules of the hydrodynamic coefficients of the UUV-1 with $\bar{x}$ were investigated. The specific calculation results are as shown in Figure 21 to Figure 24.



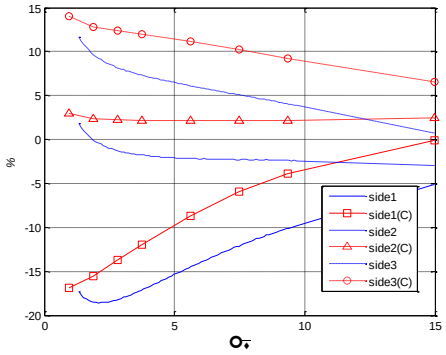


Figure 21 Calculated results of C_x **Figure 22 The change percentage of C_x**

The simulation results of C_x and its change percentage corresponding to $\Delta\overline{s}$ are as shown in Figure 21 and Figure 22. It can be seen that with the UUV-1 approaching SUB-1 laterally, when SUB-1 was located in the side1 feature orientation, $\Delta\overline{s}$ >2.8, C_x showed approximate linear decrease, when $\Delta\overline{s}$ <2.8 it displayed approximate parabolic increase; in side 2 feature orientation, when $\Delta\overline{s}$ >2.8, C_x increased slowly with the decrease of $\Delta\overline{s}$, but once $\Delta\overline{s}$ <2.8, it increased significantly; in side3 feature orientation, when $\Delta\overline{s}$ >2.8, with the decrease of $\Delta\overline{s}$, C_x showed an approximate parabolic increase trend, while $\Delta\overline{s}$ >2.8, it turned out approximate parabolic increase. Compared steady with unsteady numerical results in Figure 22, we can discover that when $\Delta\overline{s}$ >2.8, changes of the two results were similar to each other, and the calculated value in unsteady state was smaller than that in steady state corresponding to the same $\Delta\overline{s}$; when $\Delta\overline{s}$ <2.8, the change gradient of C_x with $\Delta\overline{s}$ was larger in unsteady state.

The simulation results of C_z and M_y corresponding to $\Delta\overline{s}$ when the UUV-1 approaching the SUB-1 from different positions are as shown in Figure 23 and Figure 24. From the graphic results it can be seen that with the UUV-1 getting close to the SUB-1, due to the interference of the submarine, C_z was positive , manifested as repulsion; M_y was also positive, presented as deviating the head of the UUV-1 from the SUB-1.

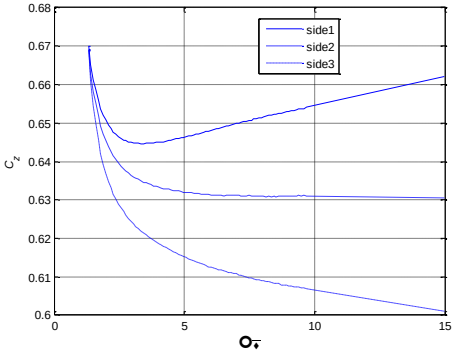


Figure 23 C_z

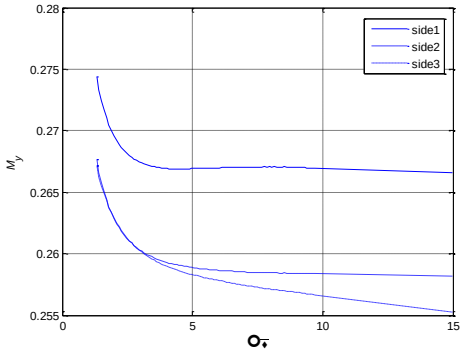


Figure 24 M_y

Conclusions

In this paper RNG $k-\varepsilon$ turbulence model was used to close the RANS equations, and combined with the dynamic grid techniques, unsteady and steady hydrodynamic performance was numerically calculated when the UUV was recovered by the submarine.

When the UUV maintained static relative to the submarine and the distance $\overline{\Delta s}$ between them is small, the hydrodynamic interference of the submarine on the UUV is strong, and with the increase of $\overline{\Delta s}$, it weakens, and the function distance of the flow around the submarine on C_x is about 30 times its diameter; the more rear the longitudinal position is, the larger C_x is. The UUV will also be subjected to the suction of the submarine for the flow around the submarine. When the UUV-2 with appendages is in the same longitudinal position and the relative direction is "side" and "up", the change laws of C_x, C_{xx} and M_{xx} with $\overline{\Delta s}$ are almost identical, while in "up" position, it is more complicated as a result of the wake flow of the fairwater. In the same feature position, for the UUV-1 and UUV-2, the change rules of C_x, C_z and M_y with $\overline{\Delta s}$ are similar accordingly.

When the UUV moves paralleled to the submarine's longitudinal axis at various speeds, the larger the local Re is, the smaller resistance coefficient is; even though the UUV is also subjected to the suction and yawing moment, the velocity has little effect on them. When the UUV approaches the submarine laterally in different feature positions, the change rules of the resistance coefficient of the UUV is similar to that in steady state; C_z is positive , manifested as repulsion; M_y is also positive, presented as deviating the head of the UUV from the submarine.

References

[1] Nicholson JW, Healey AJ.(2008) The Present State of Autonomous Underwater Vehicle (AUV) Applications and Technologies, *Marine Technology Society Journal* **42**, 44-51.
[2] Yeung, R W，Hwang, W. (1977) Nearfield hydrodynamic interactions of ships in shallow water, *Journal of Hydronautics* **11**, 128-135.
[3] H. Zhang, W.H. Melbourne. (1992) Interference between two circular cylinders in tandem in turbulent flow, *Journal of Wind Engineering and Industrial Aerodynamics* **41**, 589-600.
[4] Zeng YF, Zhu JM. (1994) Discussion on the calculation of the hydrodynamic interaction between underwater moving objects, *The Ocean Engineering* **02**, 40-48.
[5] Y.R. Choi, S.Y. Hong. (2002) An analysis of hydrodynamic interaction of floating Multi-Body Using Higher-Order Boundary Element Method, *The Twelfth International Offshore and Polar Engineering Conference*, Kitakyushu, Japan.
[6] Wang F. (2003) Hydrodynamic calculation on an autonomous underwater vehicle in motion near a submarine , Harbin Engineering University.
[7] B.J. Koo, M.H. Kim, (2005) Hydrodynamic interactions and relative motions of two floating platforms with mooring lines in side-by-side offloading operation, *Applied Ocean Research* **27**, 292-310.
[8] S.Y. Hong, J.H. Kim, S.K. Cho, Y.R. Choi, Y.S. Kim. (2005) Numerical and experimental study on hydrodynamic interaction of side-by-side moored multiple vessels, *Ocean Engineering* **32**, 783-801.
[9] Cheng L, Zhang Liang, Wu Deming, Chen Qiang. (2005) Hydrodynamic interactions between two underwater non- lifting bodies, *Journal of Harbin Engineering University* **26**, 1-6.
[10] Zhang L, Cheng L, et al. (2006) Experiment on Hydrodynamic Interaction Between 2D Oval and Wall, *Journal of Ship Mechanics* **10**, 1-10.
[11] Cheng L. (2006) Research on hydrodynamic interactions between two bodies, Harbin Engineering University.
[12] He YZ. (2010) Study on the control of UUV with cable during recovery, Northwestern Polytechnical University.
[13] Leong Z, Saad K, et al. (2013) Investigation into the hydrodynamic interaction effects on an AUV operating close to a submarine, Pacific international maritime conference, 1-11.
[14] S.A.T. Randeni P., Z.Q. Leong, D. Ranmuthugala, A.L. Forrest, J. Duffy. (2015) Numerical investigation of the hydrodynamic interaction between two underwater bodies in relative motion, *Applied Ocean Research* **51**, 14-24.

Modeling Complex Dynamical Systems in MF Range Combining FEM and Energy Methods

†G. Borello[1]

[1]InterAC, France.

*Presenting author: gerard.borello@interac.fr

Abstract

Complex dynamical systems such as car body, aircraft fuselage or train coach are conveniently modeled with Finite Element Method (FEM) in the Low Frequency range (LF). Increasing the frequency range to Mid-Frequencies (MF), typically up to 1000-2000 Hz, requires larger and larger FEM mesh. Presently, MF fluid/structure interaction problems on large structures cannot be solved in decent time at engineering level. Reduction of model size is required especially under random distributed loads. Energy methods like Statistical Energy Analysis (SEA) provide a theoretical framework for building small models based on power-balanced- equations they can be run in High Frequency range (HF). Nevertheless, SEA parameters are derived from analytical solutions of differential operators and submitted to many assumptions and simplifications. They cannot provide robust enough prediction in MF range due to inherent complexity of industrial systems.

To improve predictability of energy models, the relevant parameters are then identified by inverse method from the "statistical" dynamic information contained in side FEM model. The FEM-derived SEA models are called Virtual SEA models (VSEA). They use the same parameters than the classical "analytical" SEA models. VSEA parameters can then be directly compared to their analytical counterparts. VSEA models may be understood as compressed FEM models in which the narrow-band frequency and spaced-varying FEM dynamic is replaced by band-integrated frequency and spaced averaged dynamic applied to a partition of FEM domain into subsystems. This compression leads to very small models while minimizing the information depredation. For example car body-in-white dynamic described by 6 million DoF's in FEM is encapsulated as a real-valued 50x50 matrix relating injected power from impressed forces to energy in each of the subsystems. Problems involving random loads can then be solved by using VSEA models rather than original FEM's. VSEA models can also be complemented by analytical other subsystems such as fluid cavities to solve full vibroacoustic response involving airborne and structure-borne propagation paths. Outputs from VSEA models are also more easily interpreted and provide description of propagation paths in the system.

Keywords: Statistical Energy Analysis, SEA, Virtual SEA, VSEA, Computational Dynamic, Propagation path.

Introduction

SEA [1] has been and is still a very popular method in vibroacoustic engineering to predict random vibrations and fluid –structure interaction over a broadband frequency range. SEA describes the interaction of subdomains of a given dynamical systems in term of energy through a set of power-balanced equations. To build a valid SEA representation of the actual vibrational state, the system needs to be partitioned into subsystems and coupling coefficients calculated with appropriate physical laws. Some restrictive assumptions such as "weak" subsystem coupling must also be fulfilled. The three latter tasks have been for a long time the drawback of the SEA method due to lack of guidance in constructing models of complex systems. The use of analytical dynamical operators for computing SEA parameters was also limiting engineers in their ability to handle the structural complexity.

Nevertheless over years, despite all these obstacles, SEA method has been successful in predicting responses of a large class of industrial systems but generally with lack of control over frequency-band modeling limit and variance. From early 2000th's SEA modeling capabilities have been leveraged up over Mid-Frequency (MF) by relying on Finite Element Method (FEM) to provide more robust SEA representation using FEM derived-parameters. This technology called Virtual SEA (VSEA) has enlightened the general dynamical behavior of complex systems, leading as side effects to improvement of equivalent analytical modeling rules always required for extending VSEA results above the frequency limit of the FEM mesh. There are theoretical connections with the parallel development of stochastic FEM modeling over the MF range [2][17].

The SEA Power Equilibrium

A given dynamical system is partitioned into two subsystems for easier presentation. Between the resulting two subsystems, a set power-balanced equations Eq. (1) traduces the conservation of energy. Power flowing into a subsystem, from either a local source or arising from another coupled subsystem, is, for a fraction, dissipated into the subsystem and for another fraction, sent back to the coupled subsystem.

$$\begin{cases} \pi_1 / \omega_c = \eta_1 E_1 + \eta_{12} E_1 - \eta_{21} E_2 \\ \pi_2 / \omega_c = \eta_2 E_2 + \eta_{21} E_2 - \eta_{12} E_1 \end{cases} \tag{1}$$

$\eta_1 E_1$, $\eta_2 E_2$ are the vibrational energies dissipated by subsystem 1 and 2. $\pi_{12} / \omega_c = \eta_{12} E_1 - \eta_{21} E_2$ is the net power flow between subsystems expressed as the difference of radiated energies from 1 to 2 and 2 to 1 and $\pi_{21} = -\pi_{12}$.

π_1 and π_2 terms represent the injected power by external random loads in resp. subsystem 1 and 2 over a frequency band of width B centered around radian frequency ω_c.

η_1 and η_2 are the mean modal Damping Loss Factors of the subsystems (or DLF) and η_{12}, η_{21} the Coupling Loss Factors between related subsystems (CLF).

Under external random impressed loads, the subsystem vibrational energy is related to the frequency and spaced-averaged mean squared velocity $<\overline{v}^2>$ times the subsystem mass:

$$E = m <\overline{v}^2> \tag{2}$$

Injected power due to random loads is independent of the modal subsystem response. Therefore, knowing DLF and CLF, we can build a loss matrix relating injected power vector to energy vector and calculate the energies in function of injected power in the band *B*:

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} = \begin{pmatrix} \eta_1 + \eta_{12} & -\eta_{21} \\ -\eta_{12} & \eta_2 + \eta_{21} \end{pmatrix}^{-1} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} / \omega_c \tag{3}$$

For Eq. (3) being a valid representation of the actual energy exchange between the two subsystems, impressed powers π_1 and π_2 need to be uncorrelated. For describing the energy exchange between two continuous bounded subsystems (two plates or a plate and an acoustic cavity), the concept of modal energy is introduced. In a band *B*, the continuous subsystem is statistically described by a set of modal oscillators carrying modal energy. If *N* oscillators are resonating in *B*, the statistic modal energy is equal to $\mathfrak{e} = E / N$ where *E* is the sum of the *N* modal energies. Considering two set of modal oscillators resonating in *B* with respectively N_1 and

N_2 number of resonance frequencies in B, SEA states that the net power flow between the set is expressed by Eq. (4).

$$\pi_{12} = N_1 N_2 \beta \left(\varepsilon_1 - \varepsilon_2 \right) \tag{4}$$

It leads the symmetrical SEA power-balanced equations given by Eq. (5).

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} E_1 / N_1 \\ E_2 / N_2 \end{pmatrix} = \left[\begin{pmatrix} \left(\eta_1 + \eta_{12} \right) N_1 & -\eta_{21} N_2 \\ -\eta_{12} N_1 & \left(\eta_2 + \eta_{21} \right) N_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} / \omega_c \tag{5}$$

Off-diagonal terms are equal due to reciprocity relationship, $N_1 \eta_{12} = N_2 \eta_{21}$, making the loss matrix symmetrical as the modal coupling coefficient β is always symmetrical due to linearity of dynamical operators.

The modal formulation of SEA provides an easier interpretation of the "weakly" coupled assumption. Under this assumption, the subsystem modes can be considered as an acceptable orthonormal basis for projecting its responses and the related total energy is then found much closer to the discrete sum of modal energies. If subsystems 1 and 2 are strongly coupled, their modes are hybridized and undistinguishable with non-null cross-correlated energy E_{12}.

The further required assumption of "weak" coupling between subsystems has been discussed for a long time among SEA community and it is only recently that the role of this requirement has been clarified: Eq. (4) is always valid as soon as we restrict the calculation of energy to resonant modes. But on one hand, in case of more than two coupled subsystems, CLF law coupling related oscillators is different from the weak coupling case and on the other hand, all subsystems are found cross-coupled together, independently they are connected or not on a physical boundary.

The degree of physical coupling between thin shells, parts of a complex system, is generally decreasing with frequency as any small discontinuity of mass or stiffness in the system will generate growing reflection coefficient when frequency increases (i.e. wavelength decreases). It leads to progressive confinement of energy in localized subdomains of the dynamical system. Confined energy is thus stored in local modes of the subsystems creating energy gaps between subdomains and SEA representation given by Eq. (4) and Eq. (5) will start to work and calculated CLF can be restricted to near-by coupled subsystems. As a consequence, there is always some cut-off frequency under which SEA scheme will be found defective.

The EDM Power Equilibrium

When subsystems are strongly coupled, power flow is no more discontinuous as in Eq. (4) as the distribution of energy density within the union of subsystems is continuous function of space. The Energy Diffusion Method (EDM) demonstrates [4][5][6] that in a continuous uniform system of extension Ω the conservation of energy between subdomains $\delta\Omega$ is given by the following energy-based equation:

$$-\frac{c_g^2}{\eta\omega} \Delta e + \eta \omega e = \pi_{inj} \tag{6}$$

where c_g is the group velocity of underlying propagating waves and e the energy density in the medium. η is the mean DLF in the medium in which is assumed perfect diffusion of the vibrational field. The local intensity is found proportional to the gradient of energy density:

$$\vec{I} = -\frac{c_g^2}{\eta\omega}\vec{\nabla}e \tag{7}$$

The coefficient $\frac{c_g^2}{\eta\omega}$ plays then the same role than the coefficient of thermal transfer in heat exchange problems.

SEA and EDM Power Flows

A large fluid cavity is now considered containing many acoustic modes and its volume is split into smaller cavities. SEA is assuming weak coupling between subsystems. Therefore such an SEA model cannot represent actual distribution of energy in the volume as obviously the coupling is expected to be very strong between two neighboring sub-volumes as their boundaries are not at all reflective. To illustrate it, an acoustic volume (12m x 1m x 1m) is split into two different partitions. For comparison of calculated local energies, two SEA models are built from the partitions with sub-volumes cross-coupled by the "regular" SEA CLF.

Partition C6 is made of 6 sub-volumes (2m x 1m x 1m each) while partition C12 is made of 12 sub-volumes (1m x 1m x 1m) as sketched in Figure 1.

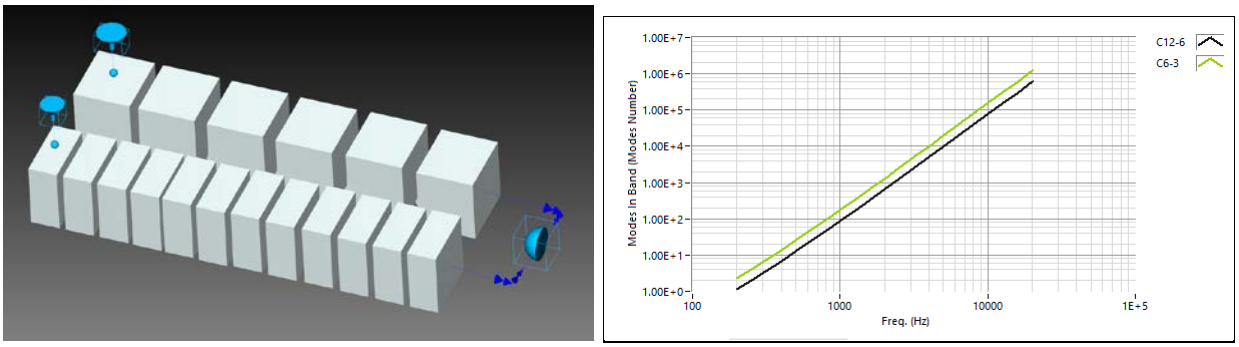


Figure 1. Two SEA partitions of the same acoustic volumes and number of modes per 1/3rd octave band in the two related elementary sub-volumes

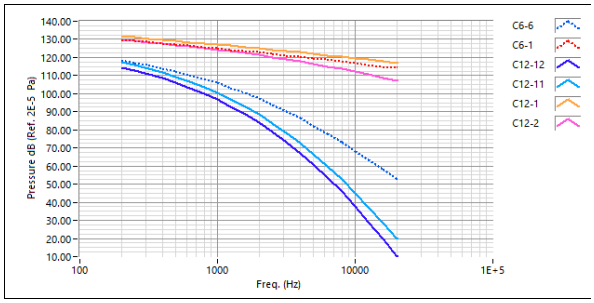


Figure 2. Pressure levels in first and last acoustic sub-volumes in C6 and C12 partitions for unit power injected in resp. C6-1 and C12-1

Elementary sub-volumes are all including at least one acoustic resonance from 200 Hz. When applying a unit source power in the first left volumes, we expect to find similar pressure level at right ends of both partitions. Sub-volumes are numbered C6-1, C6-2… for partition C6 and C12-1, C12-2… for partition C12. Figure 2 shows pressure levels calculated in both models in first and last sub-volumes. For C12 volumes, the two first and the two last are graphed as they are half-sized compared to C6 corresponding volumes. By doubling the number of sub-volumes to mesh the total volume, predicted pressure in C12 drops down from about 25 to 30 dB at 20 kHz and of about 10 dB around 2 kHz. The predicted pressure level is found dependent on mesh size (i.e. to the number of subsystems used to describe the volume).

Same exercise may be done in any SEA software method and will lead to similar result if CLF are computed from wave transmission method.

In this particular example of serial energy transfer within an arbitrary partition of an acoustic volume, weak coupling assumption is not verified. Therefore the volume meshing is not consistent.

EDM provides a more representative model to describe the actual energy transfer based on a different power flow formulation. For computing CLF, classical SEA method, as originally proposed [1] - and still applied in the community of SEA users - relies on wave theory and weak coupling. At subsystem boundary, the power flow from an emitter subsystem to a receiver one, is then given by:

$$\Pi_{1\to 2} = \eta_{12}\omega E_1 = \langle \tau \rangle_\theta \, I_{inc} \Sigma \tag{8}$$

Where $\langle \tau \rangle_\theta$ is the mean random-incidence wave transmission coefficient, I_{inc} is the incident intensity propagating from 1 to 2 and Σ the junction size (area for 3D acoustic volumes). Because $I_{inc} = \frac{c}{4} e$ in the diffuse acoustic field, *e* being the energy density, Wave Transmission (WT) CLF is found equal to:

$$\eta_{12} = \frac{\langle \tau \rangle_\theta \, c\Sigma}{4\omega\Omega} \tag{9}$$

For two identical coupled sub-volumes, the net power flow is finally expressed in function of WT CLF as:

$$\pi_{12} = \frac{c\Sigma}{4\Omega}\left(E_1 - E_2\right) = -\frac{c}{4} \cdot \frac{E_2 - E_1}{L} \tag{10}$$

where is $L=\Omega/\Sigma$ is the characteristic length of the sub-volumes and $\langle \tau \rangle_\theta = 1$ as no refection occurs at sub-volume interface.

Comparing Eq. (10) and Eq. (7), both net power flow expressions are found proportional to the gradient of energy (or energy density with appropriate scaling). But the coefficients of vibrational transfer that relates the gradient of energy to power are following different laws vs. frequency.

Therefore, in any complex dynamical system, the energy distribution over the domain is expected to be continuous over sub-domains with low reflective boundaries between them and submitted to step when crossing a reflective boundary. Because the reflectivity of a boundary is frequency-

dependent when speed of sound slightly varies between two coupled sub-domains, we have a better understanding of Figure 2 result: the discretization of an acoustic volume into sub-volumes coupled by WT CLF implicitly creates artificial loss of energy at each boundary and leads to a larger energy drop in the last cavity. This effect is a consequence of applying WT CLF between strongly-coupled regions. Fluid-structure interaction problems are less entailed by this problem as fluid and structure, at least for air, are always weakly coupled.

SEA method based on WT-CLF is not representative of actual dynamical behavior in analyzing structure borne noise propagation over the MF range when speed of sound between the various parts is not very different as in the case of car body-and-white where the skin is everywhere between 0.7 to 1 mm thick with low reflectivity from boundaries.

Over MF range, for improving robustness of SEA modeling, WT CLF should be only applied for coupling regions separated by steep energy gradient while continuous regions with smooth gradient should be either considered as a single SEA subsystem or split into smaller scale subsystems coupled by EDM modified CLF. Figure 3 shows the prediction with an SEA model with a topology similar to Figure 1 but built following previous recommendations (i.e. with different CLF expression). Sub-volumes in partitions C6 and C12 are cross-coupled by discretized EDM CLF in SEA+ software [15] and again we plot pressure levels in first and last acoustic volumes in Figure 3. In that case, the mesh dependence has nearly been cleared. Pressure in the two last C12 sub-volumes are now close to C6 related sub-volume as expected in the actual physics.

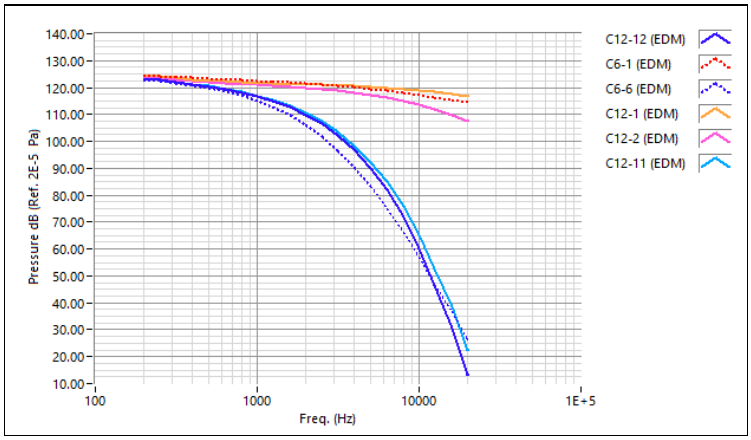


Figure 3. Pressure levels in first and last acoustic sub-volumes in C6 and C12 partitions for unit power injected in resp. C6-1 and C12-1 applying EDM CLF between sub-volumes

Virtual SEA for SEA Prediction over MF Range in Complex Structures

Creation of a representative SEA model requires discriminating subdomains separated by expected energy gap, in function of the vibrational frequency reflectivity of boundaries as previously discussed. Partition into subsystems is then a key feature in designing a “physical” SEA model but is an unknown of the vibration problem.

For complex dynamical system such car body, aircraft, spacecraft or any compact structural component such an electronic equipment, EDM theory cannot be directly applied as constant speed of sound is required among regions of smooth energy distribution. Finding potential energy gaps between various zones is also not always intuitive. Based on expertise developed in

measuring SEA CLF on built-up structures [8][9][10], VSEA method [11][12][13] was introduced for characterizing SEA CLF and weakly coupled regions of car chassis, further extended to spacecraft analysis and full car body or components in operating conditions.

For non-homogeneous systems with complex geometry, VSEA relies on FEM global real modes to get a snapshot of the statistical vibrational behavior in the various targeted frequency bands. VSEA is derived from experimental SEA analysis or ESEA [7] and identifies SEA parameters of the system responses by solving an inverse SEA problem like in ESEA method. Inputs to the inverse problem are the synthesized modal responses at a grid of predefined nodes. VSEA may be viewed as a kind of virtual test where the system dynamics is reduced to responses at a subset of discrete nodes.

Virtual SEA Numerical Process

Real modes are extracted using preferred FEM solver (NX-NASTRAN in next example). Eigenfrequencies and related modal amplitudes, at a set of restitution nodes, are stored and exported to SEA+ VSEA solver. VSEA is synthesizing complex velocity FRF $v_i/f_j = v_{ij}$ at all restitution points M_i in global x, y, z directions due to rain-on-the-roof unitary x, y, z forces applied at each restitution node M_j. Final FE statistical information is reduced to the transfer velocity matrix **V** made of v_{ij} elements

Global DLF for modal synthesis is taken equal to some frequency band-dependent default value for all modes. **V** matrix is compressed into 1/3rd octave band and projected in the direction n_i and n_j of maximal input/output conductances given for at all nodes by Y =Re{Diag(**V**)}n. The final matrix $\bar{\mathbf{V}}^2$ for SEA-parameter identification is then expressed in band-averaged format at center radial frequencyw_c and bandwidth B with elements given by:

$$v_{ij}^2(w_c, B) = \frac{1}{B}\int_B v_{ij}^2(w)dw \tag{11}$$

$\bar{\mathbf{V}}^2$ is finally auto-partitioned by SEA+ peripheral algorithm which groups nodes into a set of weakly coupled subsystems Ω_k leading to SEA rectangular transfer matrix $\langle \bar{\mathbf{V}}^2 \rangle$ of which elements are given by:

$$\langle \bar{v}_{kk'i}^2 \rangle = \frac{1}{N_k}\sum_{j\in\Omega_k} \bar{v}_{kk'ij}^2 \tag{12}$$

SEA parameter identification is performed by solving the SEA inverse problem relating $\langle \bar{\mathbf{V}}^2 \rangle$ to SEA loss matrix **L** through the normalized SEA power balanced equations.

$$\mathbf{I}\cdot Y/w_c = \mathbf{L}\cdot m\langle \bar{\mathbf{V}}^2 \rangle \Rightarrow \mathbf{L} = \left[m\langle \bar{\mathbf{V}}^2 \rangle \right]^{-1^*} \mathbf{I}\cdot Y/w_c \tag{13}$$

m is the subsystem mass vector and * indicates the pseudo-inverse. In practice with

$$m = \mathcal{N}/4Y \tag{14}$$

previous equation is reshaped for direct solve of modal density vector $\mathcal{N}$. It leads to the local modal energy matrix power balanced equations given by:

$$\mathbf{I}/w = \mathbf{L}\mathcal{N}\boldsymbol{\varepsilon} = \mathcal{L}\boldsymbol{\varepsilon} \tag{15}$$

with elements of $\boldsymbol{\varepsilon}$ given by:

$$e_{kk'i} = \frac{1}{N_k} \sum_{j \in \Omega_k} \frac{\overline{v}^2_{kk'ij}}{4 y_i y_j} \tag{16}$$

and

$$\boldsymbol{\mathcal{L}} = \begin{bmatrix} h_{1}\mathcal{N}_1 + \sum_{k'} h_{1k'}\mathcal{N}_{k'} & .. & -h_{N1}\mathcal{N}_N \\ .. & .. & .. \\ -h_{1N}\mathcal{N}_1 & .. & h_{N}\mathcal{N}_N + \sum_{k'} h_{k'N}\mathcal{N}_{k'} \end{bmatrix} \tag{17}$$

$\boldsymbol{\varepsilon}$ has dimension of modal energy and leads to accurate identification of $\boldsymbol{\mathcal{L}}$ thanks to SEA+ algorithm that performs auto-partitioning into weakly coupled regions.

In practice, with lossless junctions, previous system is solved for identifying separately modal density and CLF. Model quality is assessed by comparing $\left\langle \tilde{V}^2 \right\rangle = 4Y^T \boldsymbol{\mathcal{L}}^{-1^*} Y / \omega$ with direct $\left\langle \overline{\mathbf{V}}^2 \right\rangle$ FRF input. Difference $\tilde{V}^2 - V^2$ gives the reconstruction error matrix plot as reconstruction performance index in SEA+ [15].

The MS-VSEA patch method is a variant formulation of the inverse SEA problem where the auto-partition is applied to pre-defined group of nodes instead of nodes. This method is thus providing a partition per frequency band corresponding to a specific grouping of patches into subsystems. Main advantage is to accurately reconstruct FE transfers over the whole frequency range. Figure 1 shows the flow chart of the VSEA process.

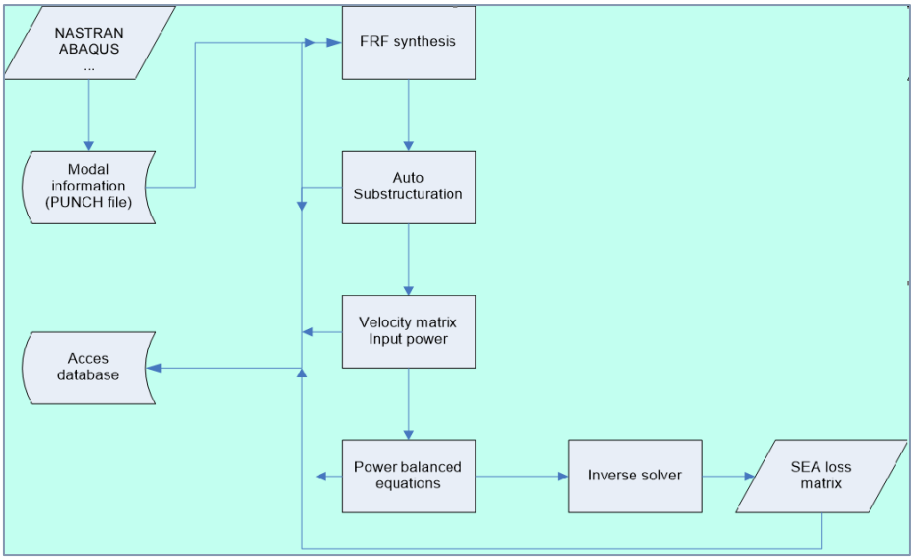


Figure 4. VSEA data flow

VSEA Modeling of a Car Component

A car cockpit is analyzed with VSEA and related subsystems are shown in Figure 5. The cockpit is made of various imbricated plastic parts embedded in a metallic support with various section properties. 500 nodes map the system. Starting from a FEM model of the component, around 2500 eigenvalues are extracted and modal amplitudes at all reference nodes are stored. SEA+ performs the FRF synthesis and numerical MS-VSEA model is generated from it.

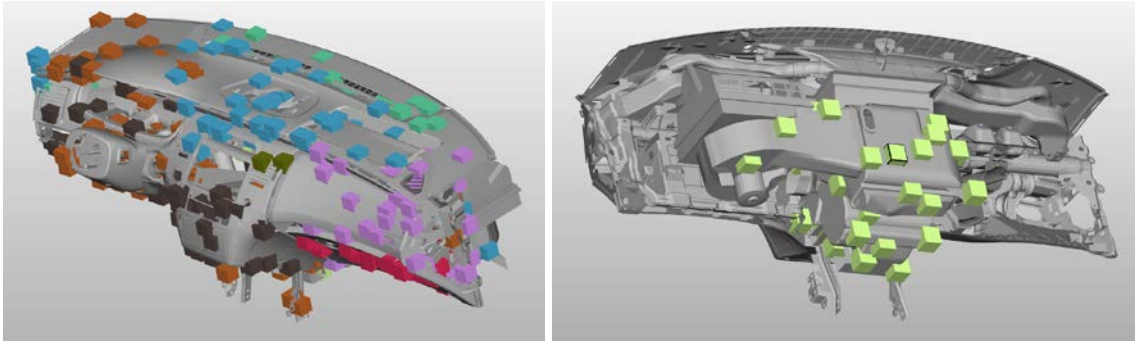


Figure 5. Left-Cockpit with identified nodal VSEA subsystems at 700 Hz and right-air-conditioning block identified as a subsystem

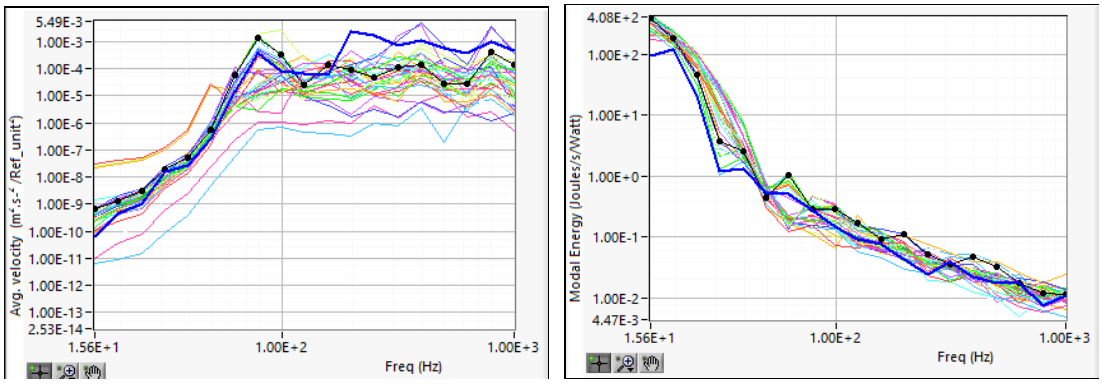


Figure 6. Left-V/F transfer velocity for a given excitation node in air-conditioning block subsystem and right-related local modal energy transfer as given by (16)

Effective subsystem domains are varying with frequency. Only 4 effective subsystems are found at 100 Hz, 7 at 250 Hz, 11 at 500 Hz and 13 at 1000 Hz. Detection of subsystems is performed on the matrix made of elements calculated from (16) that provides much less node-to-node scattering of nodal responses that the classical FRF (V/F) as shown in Figure 6.

In SEA+ GUI, only the finest partition is displayed for easy expansion in HF range as FEM modal extraction was limited to 1000 Hz due to FE size. The performance index of the mode guaranties the MS-VSEA band-averaged transfers to be within 2 dB from related direct FEM calculation. The frequency limit of MS-VSEA model is called the transition frequency, f_t.

Below f_t and unlike standard SEA model, the nodal information is preserved in VSEA model, making possible to predict response in VSEA nodes and not only as a mean over subsystem domain. Correlation with measured data is easier.

An SEA expander is then allocated to each VSEA subsystems. The expander is a “classical” SEA subsystem of which parameters are derived from analytical theory in order to take over the calculation of SEA CLF and modal densities above f_t. Patches may be conveniently chosen by the user for easier modeling of analytical SEA expanders by ascertaining a patch as a group of FE with same section property for example.

Above f_t both junction CLF and subsystem modal densities are analytically expanded to HF.

VSEA and Analytical Fluid/Structure Coupling

VSEA or MS-VSEA subsystems are coupled to analytical SEA cavities through a specific statistical radiation integral calculated by spatial windowing of an elementary infinite structural wave [14]. The related structural VSEA wavenumber is estimated from ratio of rotational over translational nodal principal-direction conductances.

For a subsystem of domain Ω_k,

$$\langle k \rangle = \sqrt{\frac{\langle Y_R \rangle_{\Omega_k}}{\langle Y_T \rangle_{\Omega_k}}} \tag{18}$$

Where $\langle Y_R \rangle$ and $\langle Y_T \rangle$ are respectively maximal local rotational and translational conductances averaged over Ω_k.

Adding Acoustic Trims to the VSEA Bare Model

All soft parts (internal trim panels, acoustic materials) are modeled as additional analytical subsystems or as attenuation spectra that filter the sound radiation. The latter are predicted by Transfer Matrix Method (TMM). Given an acoustic trim made of several layers (porous and/or elastic materials), related transmission (TL) and insertion (IL) losses are predicted by TMM under random incidence and infinite layer dimension. TL and IL are corrected for taking into account finite size of the trim. The base panels modeled as SEA subsystem are modified by trim presence (added mass and added damping).

Conclusions

SEA method provides a general theoretical framework for modeling both airborne and structure borne paths in complex industrial systems in the energy domain. We have brought to the fore some drawbacks of the method which have over years puzzled mechanical engineers in their understanding of SEA limitations. These limitations, mostly related to a priori sub-division of a system into regions satisfying SEA basic theoretical assumptions, have been overcome by introducing new concepts in the SEA modeling approach:

- Automated conversion of statistical dynamical information containing in a FEM model into SEA subsystems compatible with their analytical counterpart,
- Merging Energy Diffusion theory and Wave Transmission for improving coupling loss factor definition in region of strong coupling,
- Coupling Transfer Matrix Method which operates on 2D infinite layers with supporting SEA structures,
- Correcting calculation in infinite domain with the Spatial Windowing Technique.

Connecting all these features in a collaborative Graphical User Interface has helped in going further. Important topics, not covered in previous pages [16] are part of on-going research activities. They are progressively introduced in the SEA solver such as the non-resonant energy propagation in both structure and cavities which improves prediction of mass-controlled transfers of energy under acoustic and turbulent boundary layer loads.

Laminated shell damping and indirect mechanical coupling loss factors predictions are also an issue. Specific theories have been already stated and validation work is in progress.

References

[1] Lyon, R. H. and Maïdanik, G. (1962) Power Flow between Linearly Coupled Oscillators, *JASA* **34**, 623.

[2] Soize, C. (1993) A model and numerical method in the medium frequency range for vibroacoustic predictions using theory of structural fuzzy, *JASA* **94**(2), Pt 1, 849-866.

[3] Wilby, J. F. and Pope, L. D. (1979) Prediction of the acoustic environment in the Space Shuttle payload bay, *AIAA Conference*, Seattle, USA, REF AIAA 79-0643.

[4] Nefske, D.J. (1987) Power Flow finite element analysis of dynamic systems: basic theory and applications, *SAE*.

[5] Le Bot, A. (1994) *Equations énergétiques en mécanique vibratoire, Application au domaine des moyennes et hautes fréquences*, Thèse de l'école centrale de Lyon, France.

[6] Carcaterra, A. and Sestieri, A. (1995) Energy Density Equations and Power Flow in Structures, *JSV* 188(**2**), 269-282.

[7] Lalor, N. (1989) The Experimental Determination of Vibrational Energy Balance in Complex Structures, Paper 108429 Proc. *SIRA Conference on Stress & Vibration*, London, UK.

[8] Borello, G., Alliot and A. Kernilis A. (1991) Identification of SEA Coupling Loss Factors on a Liquid Rocket Engine, *Inter-Noise*, Sydney, Australia.

[9] Borello, G., Geoffroy, P. and Cuny, N. (1995) System Model of a High Speed Train Passenger Coach Using SEA and Prediction in Working Conditions, *Inter-Noise*, Newport Beach, CA, USA.

[10] Audoly, C. and G. Borello (2000) Feasibility of Experimental determination of SEA Damping Loss factors of submarine Structures, *Inter-Noise*, Nice, France.

[11] Gagliardini, L., L. Houillon,. Petrinelli, L. and Borello, G. (2003) Virtual SEA: mid-frequency structure-borne noise modeling based on Finite Element Analysis, *SAE-NVC* 2003-01-1555, Traverse City, MI, USA.

[12] Borello, G., Gagliardini, L., Houillon, L. and Petrinelli, L. (2005) Virtual SEA-FEA-Based Modeling of Structure-Borne Noise, *SVM*, January.

[13] Gagliardini, L., Thenail, D. and Borello, G. (2007) Virtual SEA for noise prediction and structure borne sound modelling, *Rieter Automotive Conference*, Pfaffikon, Switzerland.

[14] Villot, M., Guigou-Carter, C. and Gagliardini, L. (2001) Predicting the acoustical behaviour of finite size multi-layered structures by applying spatial windowing on infinite structures, *JSV* **245**, Number 3, 433-455(23).

[15] *SEA+ User-Guide* version from 2014 (InterAC SARL)

[16] G. Borello, G. (2016) Evolution des méthodes de calcul vibroacoustique des systèmes industriels en fonction de la fréquence, *Congrès Français d'acoustique*, Le Mans, France.

[17] Lyon, R. H. (1995) Statistical Energy Analysis and structural fuzzy, *JASA* **97**, Number 5 Pt1.

Accelerated multi-temporal scale approach to fatigue failure prediction

Rui Zhang[1, 2], Lihua Wen[1], Jinyou Xiao[1] and *†Dong Qian[2]

[1]School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China.
[2]Department of Mechanic Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

*Presenting author: dong.qian@utdallas.edu
†Corresponding author: dong.qian@utdallas.edu

Abstract

In this work, we present a computational approach to high cycle fatigue life prediction with an efficient solver employing time-discontinuous Galerkin (TDG) based space-time finite element method and its enriched version (XTFEM) [1, 2] in three dimensions. While the robustness of TDG based space-time FEM has been extensively demonstrated, a critical barrier for the extensive application is the large computational effort due to the additional temporal dimension and enrichment that are introduced. By formulating a new preconditioner and utilizing the properties of Kronecker product, we developed a generic iterative algorithm for solving the fully-coupled block-structured matrix equations formulated by space-time FEM. This approach reduces the computational cost to the same order of solving the corresponding static FE problems. The established numerical framework is further integrated with a multiscale damage model for the purpose of capturing failure initiation and propagation. The efficiency and robustness of the proposed method are illustrated in numerical examples, in which we show much better performance over direct solution of the original TDG matrix equations using either sparse direct or iterative solvers

Keywords: Space-Time FEM, XTFEM, Parallel Computing, GPU, Fatigue

Introduction

Past studies have shown that space-time finite element based on the time-discontinuous Galerkin (TDG) formulation leads to A-stable, higher-order accurate ODE solvers [3-5]. The TDG-based method has been extended to second-order hyperbolic systems such as elastodynamics [6-9]. It significantly reduces the artificial oscillations that are commonly associated with semi-discrete time integration schemes in capturing sharp gradients. Recently, it has been shown that its predicative capabilities in the temporal domain can be further improved by enriching the standard shape functions with a function that represents the problem physics, such as multi-temporal scale fatigue life prediction problems [2, 10] or coupled atomistic/continuum multiscale problems [1, 11, 12]. The enriched method is termed the extended space-time FEM (XTFEM). However, due to the additional temporal dimension and enrichment that are introduced, space-time FEM and XTFEM lead to systems of coupled equation larger than those emanating from regular semi-discrete methods, which becomes a critical barrier for practical applications in terms of computational cost.

By casting the coupled equations to partly decoupled forms, iterative predictor/multi-corrector algorithms have been developed in past decades [9, 13, 14]. These methods have been proved to be unconditionally stable and widely employed for TDG-based two-field formulation, as the resulting matrix equations are only weakly coupled. However, the single-field formulation employed in current implementation leads to fully coupled matrix systems, thus the algorithms developed for the two-field formulation are not directly applicable. Previously, we proposed a generalized iterative solution approach for both space-time FEM

and XTFEM in two dimensions [10], which significantly reduced the computational effort. In current work, we further extend this approach to three dimensions by developing a new preconditioning technique. Unlike the iterative predictor/multi-corrector algorithms, the new approach reduces the computational cost to the same order of solving the corresponding static finite element equations without explicitly recasting the original block-structured matrix systems. Furthermore, parallel algorithms based on multi-core graphics processing unit (GPU) are established in order to accelerate the solution of nonlinear constitutive model employed in fatigue damage problems. Finally, numerical examples are given to demonstrate the efficiency and robustness of the proposed method.

Space-Time Finite Element Method

Regular Space-Time FEM

The regular space-time FEM in current work follows largely the single-field formulation of TDG for elastodynamics [7]. In TDG formulation, the space-time domain $\Omega \times]0, T[$ is first divided into multiple segments called space-time slabs and the n-th slab given as $Q_n = \Omega \times]t_{n-1}, t_n[$, then Q_n is further discretized into $(n_{el})_n$ space-time elements. We further introduce the jump operators

$$[[\mathbf{u}(t_n)]] = \mathbf{u}(t_n^+) - \mathbf{u}(t_n^-) \tag{1}$$

where $\mathbf{u}(t_n^\pm) = \lim_{\varepsilon \to 0^\pm} \mathbf{u}(t_n \pm \varepsilon)$. By introducing the trial functions $\mathbf{u}^h(\mathbf{x}, t)$ and test functions $\delta\mathbf{u}^h(\mathbf{x}, t)$ to be C^0 continuous within each slab, the weak form of TDG formulation can be expressed as,

$$\begin{aligned} 0 = &\int_{Q_n} \delta\dot{\mathbf{u}}^h \cdot (\rho\ddot{\mathbf{u}} - \mathbf{f})\,\mathrm{d}Q + \int_{Q_n} \delta(\nabla\dot{\mathbf{u}}^h) \cdot \boldsymbol{\sigma}(\nabla\mathbf{u}^h)\,\mathrm{d}Q + \int_{\Gamma_t} \delta\dot{\mathbf{u}}^h \cdot \mathbf{t}\,\mathrm{d}\Gamma \\ &+ \int_\Omega \delta\dot{\mathbf{u}}^h(t_{n-1}^+) \cdot \rho[[\dot{\mathbf{u}}(t_{n-1})]]\,\mathrm{d}\Omega + \int_\Omega \delta(\nabla\mathbf{u}^h(t_{n-1}^+)) : [[\boldsymbol{\sigma}(\nabla\mathbf{u}^h(t_{n-1}))]]\,\mathrm{d}\Omega \end{aligned} \tag{2}$$

for $n = 1, 2, \ldots, N$, where ρ is the mass density, $\boldsymbol{\sigma}$ is the stress, $\mathbf{f}$ is the body force and $\mathbf{t}$ is the prescribed traction on boundary Γ_t. Note that the first line of Eq. (2) represents the regular weak form of linear elastodynamics in Galerkin formulation, while the second line enforces the velocity and displacement continuity in time.

In current work, a multiplicative form of the space-time shape function is adopted as

$$\mathbf{N}(\mathbf{x}, t) = \begin{bmatrix} N_{t_1}\mathbf{N}_\mathbf{x} & \cdots & N_{t_i}\mathbf{N}_\mathbf{x} & \cdots & N_{t_k}\mathbf{N}_\mathbf{x} \end{bmatrix} \tag{3}$$

where $\mathbf{N}_\mathbf{x}$ and N_t are the spatial and temporal shape functions respectively. This form allows us to discretize the spatial and temporal domain independently. Shape functions from the regular finite element can be employed for $\mathbf{N}_\mathbf{x}$. For temporal shape function, a simple 3-node quadratic interpolation scheme has been employed. Three nodes at t_{n-1}, $t_{n-1/2}$ and t_n are equally spaced along the time axis for each space-time slab and

$$\mathbf{N}_t = \frac{1}{\Delta t^2}\begin{bmatrix} 2(t_n - t)(t_{n-1/2} - t) & -4(t_n - t)(t_{n-1} - t) & 2(t_{n-1} - t)(t_{n-1/2} - t) \end{bmatrix} \tag{4}$$

in which Δt is the time step.

After substituting the space-time approximation into the weak form, we arrive at the space-time stiffness equation in the form of $\mathcal{K}\mathbf{d} = \mathcal{F}$, in which the fully-coupled, block-structured linear system matrix is given as

$$\mathcal{K} = \begin{bmatrix} \frac{5\mathbf{M}}{\Delta t^2} + \frac{\mathbf{K}}{2} & -\frac{4\mathbf{M}}{\Delta t^2} - \frac{2\mathbf{K}}{3} & -\frac{\mathbf{M}}{\Delta t^2} + \frac{\mathbf{K}}{6} \\ -\frac{12\mathbf{M}}{\Delta t^2} + \frac{2\mathbf{K}}{3} & \frac{16\mathbf{M}}{\Delta t^2} & -\frac{4\mathbf{M}}{\Delta t^2} - \frac{2\mathbf{K}}{3} \\ \frac{7\mathbf{M}}{\Delta t^2} - \frac{\mathbf{K}}{6} & -\frac{12\mathbf{M}}{\Delta t^2} + \frac{2\mathbf{K}}{3} & \frac{5\mathbf{M}}{\Delta t^2} + \frac{\mathbf{K}}{2} \end{bmatrix} \tag{5}$$

where **K** and **M** are regular spatial stiffness and mass matrix respectively.

Extended Space-Time FEM

The predicative capability of the space-time FEM can be further improved by introducing an enrichment function $\Phi(\mathbf{x}, t)$ into regular space-time shape function. Choice of such an enrichment function depends on the problem physics. The enriched space-time approximation is given as

$$\mathbf{u}(\mathbf{x},t) = \sum_{I=1}^{n_s} \mathbf{N}_I(\mathbf{x},t)\mathbf{d}_I + \sum_{J=1}^{n_e} \tilde{\mathbf{N}}_J(\mathbf{x},t)\mathbf{a}_J \tag{6}$$

where **a** represents the enriched degrees of freedom (DOFs), n_s and n_e are the numbers of standard and enriched DOFs respectively. There resulting formulation is then termed as XTFEM. For the *J*-th node the enriched shape function is

$$\tilde{\mathbf{N}}_J(\mathbf{x},t) = \mathbf{N}_J(\mathbf{x},t)\Phi_J(\mathbf{x},t) \tag{7}$$

in which $\Phi_J(\mathbf{x},t) = \Phi(\mathbf{x},t) - \Phi(\mathbf{x}_J, t_J)$.

Enrichment function adopted in current work has been proposed for high cycle fatigue problems [2, 10] and coupled atomistic/continuum simulations [1, 11, 12]. By employing a time dependent harmonic function, the enrichment function is given as

$$\Phi_J(t) = \Phi(t) - \Phi(t_J) = \sin(\omega t) - \sin(\omega t_J) \tag{8}$$

Similarly, the linear system matrix of XTFEM is obtained as

$$\mathcal{K}_e = \begin{bmatrix} \mathcal{K} & \mathcal{K}_{ea} \\ \mathcal{K}_{eb} & \mathcal{K}_{ee} \end{bmatrix} \tag{9}$$

where $\mathcal{K}$ is the regular space-time system matrix, $\mathcal{K}_{ea}$ and $\mathcal{K}_{eb}$ reflect the coupling between enriched and regular DOFs, $\mathcal{K}_{ee}$ reflects the coupling between enriched DOFs.

An Efficient Iterative Solver

Mathematical Formulation

As shown in Eqs. (5) and (9), linear system matrices formulated by either regular space-time FEM or XTFEM are block-structured and coupled with both conventional FE stiffness matrix **K** and mass matrix **M**, which can be expressed as

$$\mathcal{K}_{rs\times rs} = \mathbf{A}_{r\times r} \otimes \mathbf{K}_{s\times s} + \mathbf{B}_{r\times r} \otimes \mathbf{M}_{s\times s} \tag{10}$$

where **A** and **B** are non-symmetric coefficient matrices obtained by temporal integration, symbol $\otimes$ denotes the Kronecker product. The size of matrices **A** (or **B**) and **K** (or **M**) are denoted by r and s respectively. The value of r is determined by both the order of temporal shape function and the number of enriched DOFs that are introduced to each node, it could be neglected when compared with the value of s for practical problems. For example, the temporal coefficient matrices formulated by regular space-time FEM in Eq. (5) are given as

$$\mathbf{A}_{r\times r}=\frac{1}{6}\begin{bmatrix}3&-4&1\\4&0&-4\\-1&4&3\end{bmatrix}_{3\times 3},\ \mathbf{B}_{r\times r}=\frac{1}{\Delta t^2}\begin{bmatrix}5&-4&-1\\-12&16&-4\\7&-12&5\end{bmatrix}_{3\times 3} \tag{11}$$

where $r = 3$ in this case.

The proposed linear system solver is based on preconditioned iterative methods. By employing the preconditioning techniques, the original linear equation of $\mathcal{K}\mathbf{d} = \mathcal{F}$ is converted to

$$(\mathcal{P}^{-1}\mathcal{K})\mathbf{d}=\mathcal{P}^{-1}\mathcal{F} \tag{12}$$

in which $\mathcal{P}$ is the preconditioner, **d** and $\mathcal{F}$ are unknown and force vectors respectively.

It is well known that the efficiency and robustness of these methods largely depends on the quality of the preconditioners. A good preconditioner $\mathcal{P}$ should be close to $\mathcal{K}$, and makes the resulting system easier to solve. In order to improve the numerical efficiency of the iterative solver, we further exploit and utilize the unique block-structure of the $\mathcal{K}$ matrix as shown in Eq. (10) to propose a new preconditioner. The new preconditioner is obtained as $\mathcal{P}=\mathbf{A}\otimes\mathbf{P}$, where $\mathbf{P}\approx\mathbf{K}$ is a preconditioning matrix obtained by approximating the spatial stiffness matrix **K**. The resulting computational effort is then reduced to the same order of solving the corresponding static finite element stiffness equations.

Numerical implementation

In current work, the Generalized Minimum Residual method (GMRES) [15] is employed as the iterative solver since the system matrix $\mathcal{K}$ is nonsymmetric. Preconditioning matrix **P** = **LU** in which **L** and **U** matrices are obtained by incomplete lower and upper factorization of the spatial **K** matrix with threshold strategy for dropping small terms and column pivoting (ILUTP) [15]. In order to reduce the number of fill-in entries that are introduced to the factor matrices during the factorization, which could lead to very expensive computation, a permutation of the **K** matrix is performed first by employing the Reversed Cuthill-McKee (RCM) reordering algorithm [16]. To overcome the demanding storage efforts, **K** and **M** matrices are stored in Compressed Sparse Row (CSR) format. Note that explicit formulation of the block-structured matrix $\mathcal{K}$ is no longer required in current implementation.

Numerical Example

Prismatic rod subject to cyclic fatigue loading

We consider a prismatic rod as sketched in Figure 1. The rod is fixed at left end and subject to a fully-reversed cyclic fatigue loading $p(t) = P_0\sin(2\pi ft)H(t)$ Pa at right end, where $H(t)$ is the Heaviside function. The amplitude and frequency of the cyclic loading are 10^6 Pa and 10 Hz respectively. The material properties are given as Young's modulus E = 211 GPa, Poisson's ratio ν = 0.3 and mass density ρ = 7850 kg/m^3.

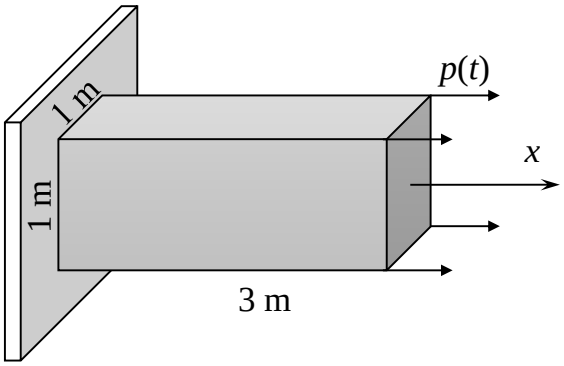


Figure 1. Illustration of the prismatic rod problem

This problem is simulated by XTFEM with a time step $\Delta t = 5T$ where T = 0.1 s is the period of loading cycle. The spatial domain is discretized by 8-node linear cubic elements. The computing environment is a desktop workstation with Intel Xeon CPU E5-2623v3, 16 Gigabytes RAM and NVIDIA TESLA GPU K20c. Displacement response is illustrated in Figure 1 and compared with solutions obtained from both explicit and implicit FEM using ABAQUS. The result obtained by XTFEM agrees well with those from traditional semi-discrete methods which require much smaller time steps. It shows that XTFEM is stable and accurate for the large time steps employed. This advantage of XTFEM would allow fast simulations on high-cycle fatigue loading histories.

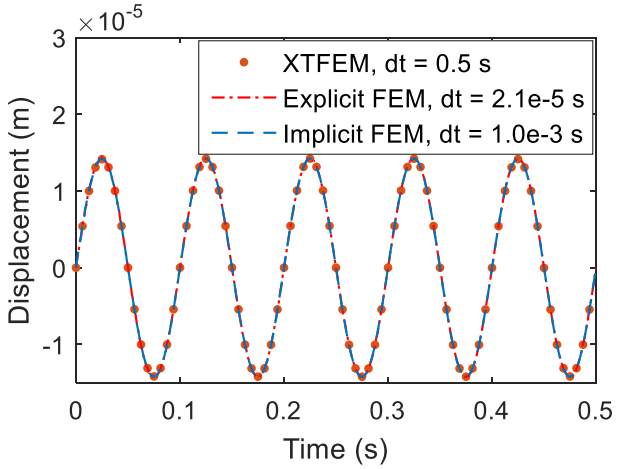


Figure 2. Displacement response at the free end of the rod subject to cyclic fatigue loading

In order to demonstrate the performance of the proposed iterative solver, a comparison study with regarding to both a sparse direct solver (SuiteSparse/UMFPACK) and a regular preconditioned iterative solver is conducted here. Note that the regular preconditioned iterative solver employed here is almost the same with the one developed in current work, except that the preconditioner is obtained directly from the large, block-structured space-time stiffness matrix. For these two iterative solvers, the dropping and pivoting tolerances of ILUTP preconditioner are set to 1.0e-3 and 1.0e-1 respectively, while the GMRES convergence tolerance is 1.0e-8.

By varying the size of spatial elements, N, the number of unknowns in the resulting linear systems formulated by XTFEM, ranges from 5,850 to 3,661,218. The computational performances of different solvers on those linear systems are summarized in Table 1. The memory usage is obtained from the storage of the **L U** factors due to their major contribution,

while time cost is measured by the CPU time for solving the first time step as the LU factorization is only performed at this step. In addition, the number of iterations to converge of the two iterative solvers also provided in Table 1. Symbol "/" indicates no results due to insufficient memory.

Table 1. Performance of different solvers in XTFEM simulations

DOFs	Sparse direct solver		Regular iterative solver			Current solver		
	Mem (MB)	Time (s)	Mem (MB)	Time (s)	Iters	Mem (MB)	Time (s)	Iters
5,850	224	12.5	21.7	4.4	76	1.6	0.04	13
36,450	7,764	3,254	300.5	160	290	20	1.0	47
484,218	/	/	4,865	8,432	1,332	333	41	151
3,661,218	/	/	/	/	/	2,722	680	309

Table 1 clearly demonstrates the advantages of the current solver over the other two and remarkable computational savings are achieved. In terms of computational complexity, the sparse direct solver showed an $O(N^{3.0})$ time cost and $O(N^{1.9})$ memory cost; The regular iterative solver achieved an better performance of $O(N^{1.7})$ and $O(N^{1.2})$ for time and memory costs respectively; Finally, the current solver further reduced the time cost to $O(N^{1.5})$ and memory cost to $O(N)$. In addition, the proposed solver also significantly reduced the number of iterations to converge. Thus, we conclude here that the proposed iterative solver efficiently and robustly accelerated the solution of linear systems formulated by XTFEM.

Conclusion

In summary, an accelerated multi-temporal scale approach is developed in current work for fatigue failure prediction in three dimensions. An efficient iterative solver with a new preconditioning technique is established for the fully-coupled, block-structured matrix equations that are formulated by TDG-based space-time FEM and XTFEM. This solver successfully reduces the computational cost from solving the large space-time matrix equations to the same order of solving the smaller corresponding static finite element equations without explicit matrix recasting. GPU-based parallel algorithms for the nonlinear constitutive fatigue damage model is coupled with XTFEM to predict fatigue failure. Numerical examples with unknowns up to ~3.7 million have been efficiently accelerated by the proposed method using single CPU process on a desktop workstation. The robustness of the solver is also extensively demonstrated. It shows that the computing time and memory of the accelerated implementation scale with the number of DOFs N through $O(N^{1.5})$ and $O(N)$ respectively.

Acknowledgments

The work of R. Zhang is supported by the State Scholarship Fund (China) under grant No. 201406290125 and the University of Texas at Dallas, which are gratefully acknowledged. The authors also would like to thank the National Science Foundation (grant # DMR-0706161, CMMI-1335204, 1334538) for financial support of this research. Any opinions, findings, conclusions, or recommendations expressed are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] Y. Yang, S. Chirputkar, D. N. Alpert, T. Eason, S. Spottswood, and D. Qian, "Enriched space-time finite element method: a new paradigm for multiscaling from elastodynamics to molecular dynamics," *International Journal For Numerical Methods In Engineering,* vol. 92, pp. 115-140, Oct 12 2012.

[2] S. Bhamare, T. Eason, S. Spottswood, S. R. Mannava, V. K. Vasudevan, and D. Qian, "A multi-temporal scale approach to high cycle fatigue simulation," *Computational Mechanics,* vol. 53, pp. 387-400, 2014.

[3] P. Lesaint and P. A. Raviart, "On a Finite Element Method for Solving the Neutron Transport Equation," in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, C. de Boor, Ed., ed New York: Academic press, 1974, pp. 89-123.

[4] M. Delfour, W. Hager, and F. Trochu, "Discontinuous Galerkin Methods for Ordinary Differential Equations," *Mathematics of Computation,* vol. 36, pp. 455-473, 1981.

[5] C. Johnson, "Error Estimates and Adaptive Time-Step Control for a Class of One-Step Methods for Stiff Ordinary Differential Equations," *SIAM Journal on Numerical Analysis,* vol. 25, pp. 908-926, 1988.

[6] T. J. R. Hughes and G. M. Hulbert, "Space-time finite element methods for elastodynamics: formulations and error estimates," *Computer Methods in Applied Mechanics and Engineering,* vol. 66, pp. 339-363, 1988.

[7] G. M. Hulbert and T. J. R. Hughes, "Space-time finite element methods for second-order hyperbolic equations," *Computer Methods in Applied Mechanics and Engineering,* vol. 84, pp. 327-348, 1990.

[8] T. J. R. Hughes and J. R. Stewart, "A space-time formulation for multiscale phenomena," *Journal of Computational and Applied Mathematics,* vol. 74, pp. 217-229, 1996.

[9] X. D. Li and N. E. Wiberg, "Structural dynamic analysis by a time-discontinuous Galerkin finite element method," *International Journal for Numerical Methods in Engineering,* vol. 39, pp. 2131-2152, 1996.

[10] R. Zhang, L. Wen, S. Naboulsi, T. Eason, V. K. Vasudevan, and D. Qian, "Accelerated multiscale space–time finite element simulation and application to high cycle fatigue life prediction," *Computational Mechanics,* pp. 1-21, 2016.

[11] S. Chirputkar and D. Qian, "Coupled atomistic/continuum simulation based on extended space-time finite element method," *Cmes-Computer Modeling in Engineering & Sciences,* vol. 24, pp. 185-202, 2008.

[12] D. Qian and S. Chirputkar, "Bridging scale simulation of lattice fracture using enriched space-time Finite Element Method," *International Journal for Numerical Methods in Engineering,* vol. 97, pp. 819-850, 2014.

[13] C. C. Chien and T. Y. Wu, "An improved predictor/multi-corrector algorithm for a time-discontinuous Galerkin finite element method in structural dynamics," *Computational Mechanics,* vol. 25, pp. 430-437, 2000.

[14] P. Kunthong and L. L. Thompson, "An efficient solver for the high-order accurate time-discontinuous Galerkin (TDG) method for second-order hyperbolic systems," *Finite Elements in Analysis and Design,* vol. 41, pp. 729-762, 2005.

[15] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second ed.: Society for Industrial and Applied Mathematics, 2003.

[16] W. M. Chan and A. George, "A linear time implementation of the reverse Cuthill-McKee algorithm," *BIT Numerical Mathematics,* vol. 20, pp. 8-14, 1980.

Numerical investigation of different tip clearances effect on the performance of Pumpjet Propulsor

***Qin Denghui, †Pan Guang, Huang Qiaogao, Lu Lin and Luo Yang**

College of Marine, Northwestern Polytechnical University, Xi'an, China

*Presenting author: 18729869394@163.com

†Corresponding author: panguang601@163.com

Abstract

Tip clearance loss is a limitation of the improvement of turbomachinery performance. Previous studies show the tip clearance loss is generated by the leakage flow through the tip clearance, and is roughly in close relation with the gap size. In this study, a pumpjet propulsor with different size of tip clearance(δ =0.2mm、0.5mm、1mm、2mm、3mm) has been presented to investigate the influence of the tip clearance on a pumpjet propulsor. This analysis was carried out by solving Reynolds Averaged Navier-Stokes (RANS) method with the commercial Computational Fluid Dynamic (CFD) code CFX14.5, and the SST $k-\omega$ turbulence model is applied. In order to verify the accuracy of numerical simulation method, calculations were carried out with a worldwide employed propeller (the E779A propeller). Simulation results show that the open water efficiency decreases gradually in the same advance coefficient (J) with the increasing of tip clearance. However, the open water efficiency is basically unchanged after the tip clearance is bigger than 2mm. The effects of tip clearance on the tip-separation vortex and the tip-leakage vortex are discussed. the area affected by the tip-separation vortex and the tip-leakage vortex is becoming bigger and bigger as the tip clearance increases. And as the tip clearance increases, the core of tip-separation vortex and tip-leakage vortex are "contact", they transfer to each other. The position of the contact is moved to following edge in the axial direction, and the position is 1/3 of the blade tip away from leading edge in the case δ=3mm. The main effected area of different tip clearance, which is the low pressure area, is mainly focus at the area above 0.9 spanwise of the suction side of rotor blade.

Keywords: Pumpjet Propulsor; Tip clearance; Computational fluid dynamic (CFD). The tip vortex structure.

Introduction

Pumpjet propulsor is a new type of underwater propulsion system, which adopts single-rotor propulsion and decelerating duct. The application of decelerating duct improves the cavitation performance of the propulsion system at a lower velocity.

At present, the research on the characteristics of pump jet propulsion, domestic and international published literature mainly concentrates on the test and numerical calculation of hydrodynamic performance. Ch. Suryanarayana et al[1] make experiment on hydrodynamic performance of the underwater vehicle equipped with pumpjet propulsor. They verify the advantages of the rear stator pumpjet propulsor and indicate that the rear stator can absorb the rotational energy of the rotor and reduce the radial component in the wake, and so as to improve the efficiency of the propulsion. Stefan Ivanell [2] uses computational fluid dynamics method to calculate the hydrodynamic performance of the torpedo with pump jet, and the rationality of the method is verified by comparing with the experimental results. The numerical results show that the stator has contributed about 20% of the thrust. Song Baowei et al.[3] calculate the hydrodynamic performance of a type of pump jet propulsor based on CFD method; using high quality structured grid and using sliding mesh technology. The numerical results and the experimental results are in good agreement. Pan Guang et al [4] carry numerical calculation to the vehicle equipped with a certain type of water pump jet propulsion. The open water performance curve of pump jet propeller is given and it indicates that

the pumpjet propulsor has higher efficiency, and ideal balance performance. The pressure of rotor blades and stator blades in relative height is analyzed. The morphology and the principle of the rotor tip vortex are explained. In the flow of pump jet propulsor gap, Wang Tao et al [5] carry numerical simulation for complex viscous flow field of pumpjet propulsor. By analyzing the local flow field, the influence of the clearance flow on the flow field (including velocity and pressure fields) is revealed. In addition the most of the study about the flow of tip clearance are aimed at the duct propeller or axial flow pump. For example, T. Lee Y. et al [6] study the flow of tip clearance of the duct propeller by solving the three-dimensional RANS equation method. The calculated results are in good agreement with the experimental results. It is shown that the numerical method is feasible for the study of the tip clearance flow. Although the duct propeller and axial flow pump are different with pumpjet propulsor, the results of the duct propeller and axial flow pump research have a good reference to the research of tip clearance flow of pumpjet propulsor.

In this paper, according to the 0.2mm, 0.5mm, 1mm, 2mm, 3mm pumpjet propulsor model, the high quality structured grid is generated based on the block grid coupling technique. By means of numerical simulation, based on the sliding grid technique, the numerical simulation of three-dimensional full channel steady turbulent flow is carried out. The open water performance of the pumpjet propulsor with different tip clearances, the influence of the rotor tip-separation vortex and tip-leakage vortex and the rotor blade surface pressure field is analyzed.

Numerical simulation method

Governing equations

For an incompressible and single phase fluid, the governing equations for Reynolds Averaged Navier-Stokes (RANS) can be written as the mass and momentum conservations in the following tensor form:

$$\frac{\partial \rho U_j}{\partial x_j} = 0 \tag{1}$$

$$\frac{\partial (\rho U_i U_j)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\mu \frac{\partial U_i}{\partial x_j} \right) + \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial P}{\partial x_i} + S_M \tag{2}$$

where i = 1, 2, 3 , j = 1, 2, 3 , ρ is the fluid density, x_i and x_j are the Cartesian coordinate components. S_M , U_i and U_j are different values depending on different situations. For an inertial frame, S_M equals to zero, U_i and U_j represents the absolute velocity component. For a relative rotating frame, S_M is the sum of Coriolis ($2\boldsymbol{\omega}\times\boldsymbol{U}$) and centrifugal forces ($\boldsymbol{\omega}\times(\boldsymbol{\omega}\times r)$), U_i and U_j represent the relative velocity components. μ is the dynamic viscosity, t is the time, τ_{ij} denotes the Reynolds stresses, and P and U represent the pressure and the time averaged velocity, respectively.

Turbulence model

According to the existing study by Ji et al. [7], the $k-\omega$ shear stress transport (SST) turbulence model is applied for closing the numerical simulation in this study. The *SST* $k-\omega$ turbulence model combines the advantages of stability of the near-wall $k-\omega$ turbulence model and independent of the external boundary $k-\varepsilon$ turbulence model. It can adapt to a variety of physical phenomenon caused by the pressure gradient changes, and it can utilize the inner viscous layer combined with the wall function to accurately simulate the phenomenon of the boundary layer without the use of easier distortion viscous-attenuation function.

Verification of numerical simulation method

In order to verify the accuracy of numerical simulation method, the steady flows over a skewed four-bladed marine propeller E779A have been studied. The non-dimensional geometry data of the E779A propeller is taken from Subhas et al. [8] and presented in Tables 1. This propeller has been widely tested for several years and a large number of reliable experimental data are available (see Salvatore et al. [9]). The computational domain and boundary conditions for E779A marine propeller is a 1/4 cylinder passage as shown in Figure 1.

Table 1: Parameters of the E779A propeller

Propeller diameter (D_t)	227.3 *mm*
P/D_t ratio	1.1
Skew angle	$4^{\circ}48^{"}$
Rake	$4^{\circ}3^{"}$
Blade area ratio	0.689
Hub diameter (D_H)	45.53 *mm*

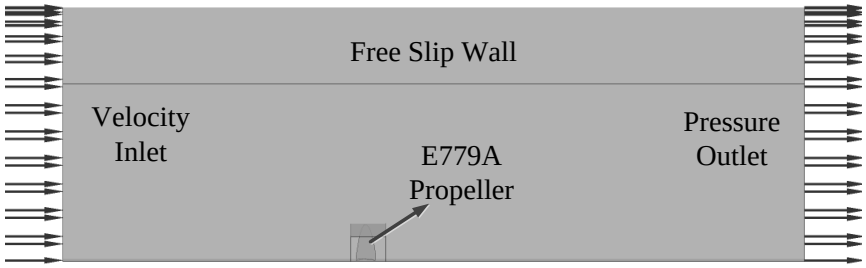


Figure 1: Computational domain and boundary conditions for E779A propeller.

The advance ratio J is defined, respectively, as $J = U_\infty / (nD_t)$, where U_∞ denotes the free stream velocity, n is the blade rotating velocity, p_{out} is the outlet pressure. The thrust coefficient $K_T = Thrust / (\rho_f n^2 D_t^4)$ and torque coefficient $K_Q = Torque / (\rho_f n^2 D_t^5)$ are defined, respectively. The numerical simulations are carried out at three different typical values of advance ratio J =0.71, 0.77 and 0.83. The numerical results of K_T and K_Q at three different advance ratios are compared with the experimental data and summarized in Table 2.

Table 2: Comparison of thrust and torque coefficient with experimental data

J	Numerical Results		Experimental Results		Errors (%)	
	K_T	$10K_Q$	K_T	$10K_Q$	K_T	$10K_Q$
0.71	0.2485	0.4327	0.2474	0.4449	0.44	2.74
0.77	0.2206	0.3913	0.2184	0.4031	1.01	2.93
0.83	0.1913	0.3488	0.1888	0.3590	1.32	2.84

From Table 2 we can see that the numerical prediction results are in good agreement with the experimental results, and the errors of K_T and K_Q are less than 3%. Consequently, it is indicated

that the numerical simulation method with the *SST* $k-\omega$ turbulence model is applicable and reliable for pumpjet propulsor flows.

Steady numerical simulation for pumpjet propulsor

Pumpjet propulsor model

Pumpjet propulsor simulation model in this study is shown in Figure 2: the propeller has 11 rotor blades, 9 stator blades. The rotors are in front of the stators, and the rotors rotate clockwise (seen from the front of the model). The diameter of pumpjet propulsor is D =0.26 *m* and the length of pumpjet propulsor is L =0.17 *m*.

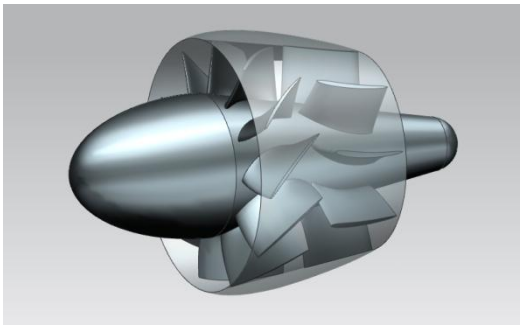


Figure 2: Pumpjet propulsor model

In order to simulate the flow better and get more precise result, two half ellipsoid type flow-guide caps have been added in front and rear of the propulsor model, respectively. In order to study the effects of different tip clearance effect on the performance of pumpjet propulsor, different diameter of the duct has been selected to get different tip clearance. Five models with $0.2mm$, $0.5mm$, $1mm$, $2mm$ and $3mm$ tip clearances have been selected. In order to facilitate the discussion, the δ has been defined to represent the tip clearance.

Computational domain and mesh

The computational domain and boundary conditions are shown as Figure 3 and Figure 4. The computational domain is a length of 10*L*, diameter of 5*D* cylinder surrounding the model, whose axis coincides with the symmetry axis of Propulsor model. The inlet is located 3*L* from the front face of Propulsor model, and the outlet is situated 7*L* from the front face of Propulsor model. According to the structural characteristics of the pumpjet propulsor, the computational domain is divided into three parts: rotor domain, stator domain and external flow field domain. The rotor domain is a rotating domain, and the other two domains are stationary domains. The rotor and stator domains are embedded in the external flow field domain. The interaction between the rotor domain and stator domain and the interaction between the rotor domain and external flow field domain are solved by using the sliding mesh method.

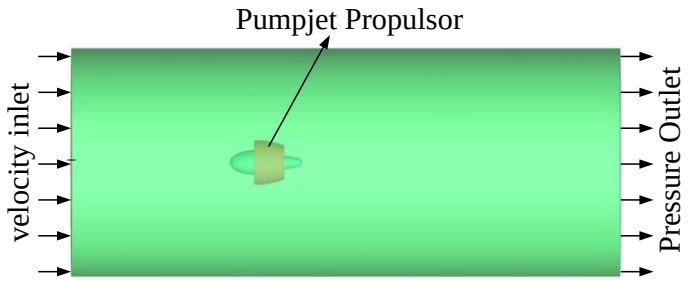


Figure 3 The computational domain and boundary conditions

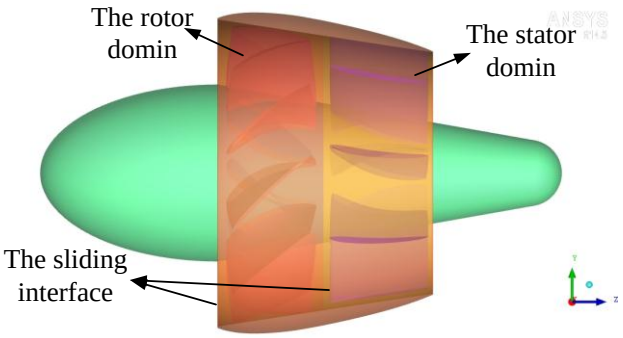
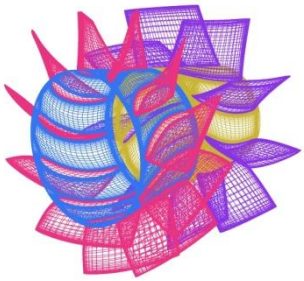


Figure 4 The sliding interface

The computational grid quality directly affects the results of numerical simulations. The structured grid has the advantage of using less memory and is very favorable for the boundary layer calculation. Therefore, the three computational domains are filled with structured grids. Multi-block grid method are used to generate high-quality structured grid by ANSYS ICEM. The grids around PJP adopt H hybrid grids, The PJP surface and propulsor blades are surrounded by O-hexahedral girds. Figure 5 shows the rotor and stator blades surface grids. In addition, in order to accurately capture the phenomenon of tip vortex, the gap is encrypted and the boundary layer is 0.05mm, Figure 6 shows the encrypted mesh between the rotor blades and the duct. The number of entire computational domain grids is approximately 3×10^6.



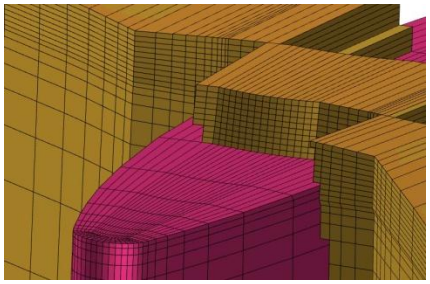


Figure 5 the rotor and stator blades surface grids Figure 6 the encrypted mesh between the rotor blades and the duct

Boundary condition

Software ANSYS CFX is applied in numerical simulation. For computational domain boundary conditions, the inlet boundary is set to normal speed, turbulence intensity is 5% as the default. The no-slip boundary condition is imposed on duct and stator blades. The free-slip wall boundary is imposed on the cylinder surface. The averaged static pressure is 0 Pa at the outlet. The interface between the rotor domain and stator domain is set to frozen rotor. The finite volume method is used to discrete control equations and the turbulence model. The pressure and velocity coupling using the SIMPLEC algorithm and the spatial derivatives are calculated using a second-order upwind algorithm.

Results and discussion

To facilitate the discussion of calculation results, the non-dimensional physical quantities are shown in Table 3.

Table 3. Non-dimensional physical quantities

physical quantities	definition

advance coefficient	$J=\frac{v}{n*D}$
thrust coefficient of rotor	$K_{T_t}=\frac{T_t}{\rho n^2 D^4}$
the torque coefficient of rotor	$K_{M_t}=\frac{M_t}{\rho n^2 D^5}$
thrust coefficient of stator and duct	$K_{T_s}=\frac{T_S}{\rho n^2 D^4}$
the torque coefficient of stator and duct	$K_{M_s}=\frac{M_s}{\rho n^2 D^5}$
Total thrust coefficient	$K_T=K_{T_t}+K_{T_s}$
Total torque coefficient	$K_M=K_{M_s}$
The open water efficiency	$\eta=\frac{J}{2\pi}\frac{K_T}{K_M}\delta$

In the table, υ is the far field flow velocity; n is rotor speed (r/s); D is the diameter of the rotor; ρ is the fluid density; T_t is the thrust of rotor; T_s is the thrust of stator and duct; M_t is the torque of rotor and M_t and M_s is the torque of stator and duct.

Different tip clearances effect on the open water performance of PJP

In the case of δ=3mm, maintain the velocity of inlet equal to $25.72ms^{-1}$ and change n from $2400rps$ to $4200rps$ to obtain different advance ratios. Figure 7 shows the thrust and torque coefficient and open water efficiency curves.

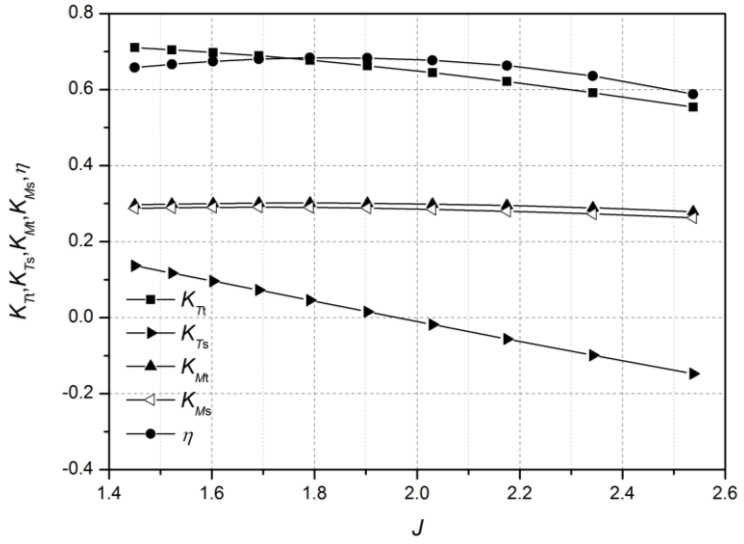


Figure 7 Thrust and torque coefficient and open water efficiency curves

Maintain the velocity of inlet equal to $25.72ms^{-1}$ and calculate models with $0.2mm$, $0.5mm$, $1mm$, $2mm$ and $3mm$ tip clearance. Figure 10 shows the open water efficiency curve of the five models.

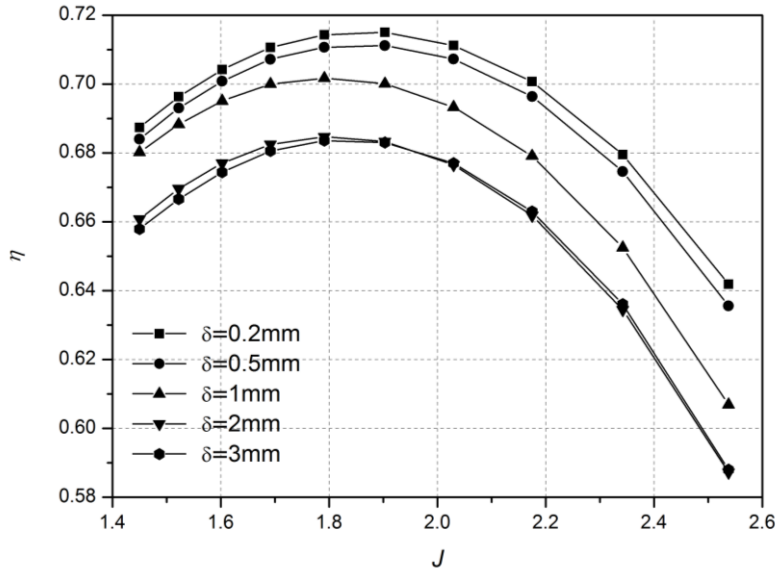
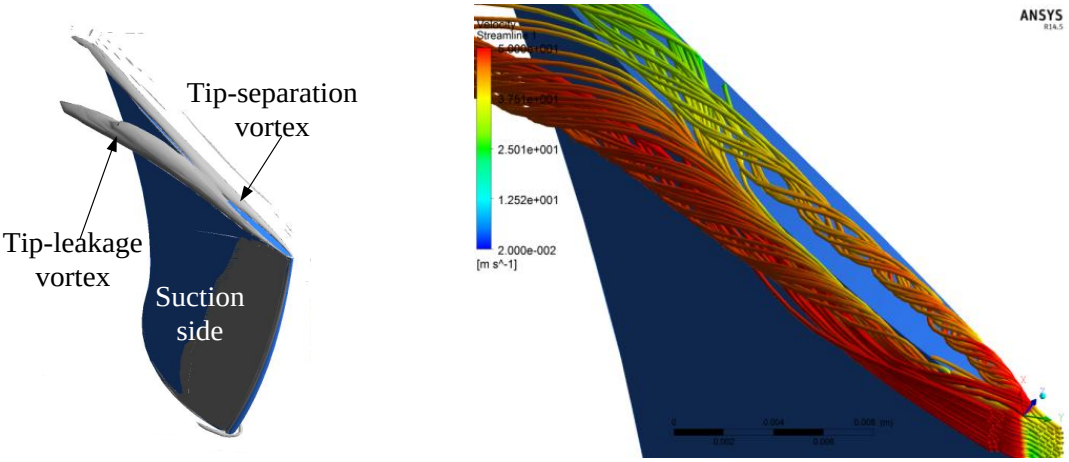


Figure 8 the open water efficiency curve of different tip clearance models

It can be seen from Figure 8: The rotor provides the main thrust because the rotor has much more thrust than stator and duct system; As J increases in the calculation range, the thrust and torque of rotor and the stator and duct system are gradually reduced; As advance coefficient increases in the calculation range, a linear relationship between thrust coefficient of the stator and duct system and advance coefficient. As J increases, the thrust coefficient of the stator and duct system change from trust to resistance; The torque coefficient of rotor and the stator and duct system are close to each other in the calculation range, and the maximum relative error is only 5.86% when advance coefficient is 2.53. It indicates that the PJP used in this study have a ideal balance performance. As J increases in the calculation range, the open water efficiency increased first and then decreased. The PJP has a maximum open water efficiency about 71.5% when J is 1.9 In the case δ=0.2mm . As tip clearance increases, the open water efficiency decreases gradually in the same J . The open water efficiency is basically unchanged after the tip clearance is bigger than 2mm.

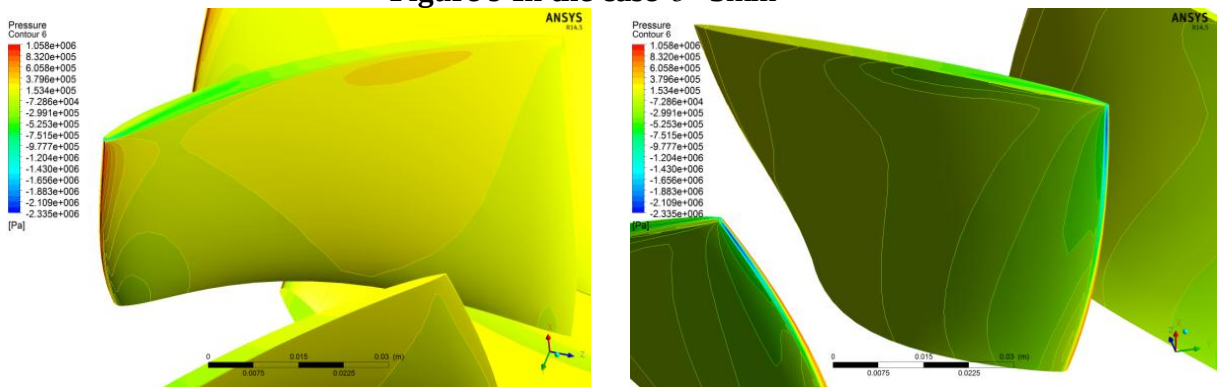
Different tip clearances effect on the tip vortex structure of PJP

You et al. [10] found that the tip vortex structure of ducted propeller is formed by three parts: the tip-separation vortex, the tip-leakage vortex and the induced vortex. The tip-leakage vortex is caused by the pressure different between the pressure and the suction side. The tip-separation vortex is formed due to flow separation underneath the blade tip. The induced vortex is generated by the tip-leakage vortex. Although the tip vortex structure of ducted propeller may be different with PJP, the research conclusion has a great impact on PJP. Because the strength and the influence area of induced vortex are small, the effect of different tip clearance on the tip-separation vortex and the tip-leakage vortex has been mainly analyzed. Figure 11(a) shows the vortex core of rotor blade in the case δ=3mm using the λ_2 vortex-identification (Jeong and Hussain, [11]). Figure 9(b) shows the flow streamlines near the rotor blade tip. The pressure contours of pressure side and rotor suction side of rotor in the case of δ=3mm are illustrated in Figure 10 (a) and (b).



(a) the vortex core of rotor blade **(b) the flow streamlines near the rotor blade tip**

Figure 9 In the case δ=3mm



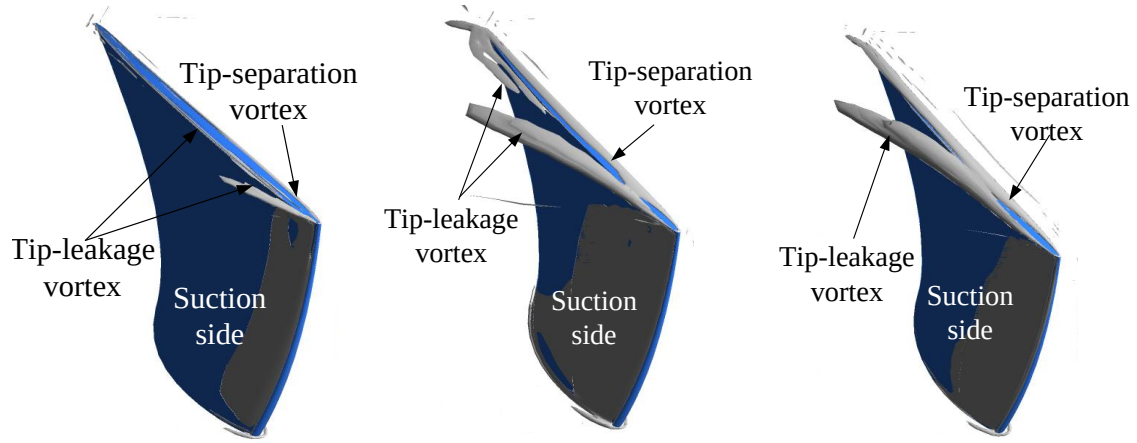
(a) the pressure side **(b) the suction side**

Figure 10 the pressure contours of rotor blade in the case δ=3mm

It can be seen from figure 9(a) that the rotor tip-separation vortex is caused by the flow separation at the leading edge of the rotor blade tip. The rotor tip-separation vortex spreads in the axial direction along the intersecting line of the rotor blade tip and pressure side of the rotor blade, leaves the trailing edge of rotor blade tip, and spreads to the stator passage finally. The tip-separation vortex moves toward to the intersecting line of the rotor blade tip and suction side in the circumferential direction with the spread of the vortex in the axial direction.

As shown in Figure 9(a), the rotor tip-leakage vortex is formed at the blade tip of the suction surface of the rotor blade. It can be seen from Figure 10 that there is obvious area of low pressure on the tip of suction side of the rotor blade near the leading edge. Simultaneously, obvious area of high pressure is formed on the tip of pressure side of the rotor blade near the leading edge. The fluid flow is sucked to the low pressure area of the suction side due to the pressure difference, which causes appearance of the rotor tip-leakage vortex. The tip-separation vortex left the suction side of the rotor blade and moved toward to mid-passage with the spread of the vortex in the axial direction. By Figure 9(b) can be seen that the rotor tip-separation vortex and tip-leakage vortex are not completely independent. A portion of the fluid separates from tip-separation vortex and integrates into tip-leakage vortex. Meanwhile, a portion of the fluid separates from tip-leakage vortex and integrates into tip-separation vortex. The low pressure center of vortex is also called vortex core. Figure 9 shows that the tip-separation vortex core and tip-leakage vortex core are "connected", they transfer to each other.

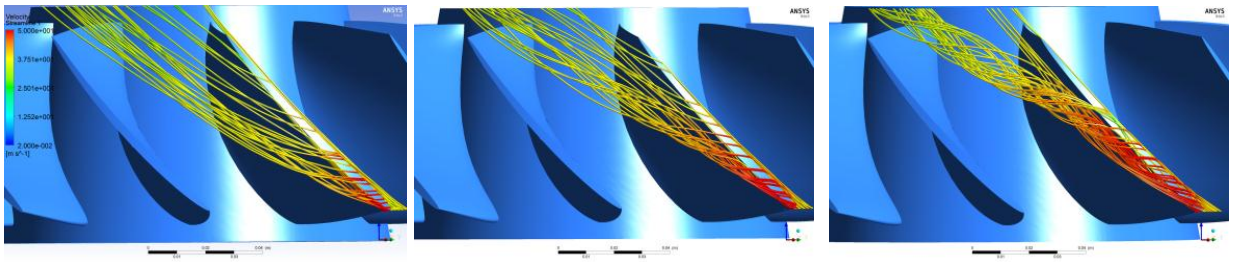
The vortex core of rotor blade in the case δ=0.2mm ; δ=1mm and δ=3mm when J=1.9 are illustrated in Figure 11.



(a) $\delta = 0.2mm$ **(b)** $\delta = 1mm$ **(c)** $\delta = 3mm$

Figure 11 the vortex core of rotor blade of different tip clearance

The flow streamlines near the rotor blade tip in the case δ=0.2mm ; δ=1mm and δ=3mm when J=1.9 are illustrated in Figure 12.



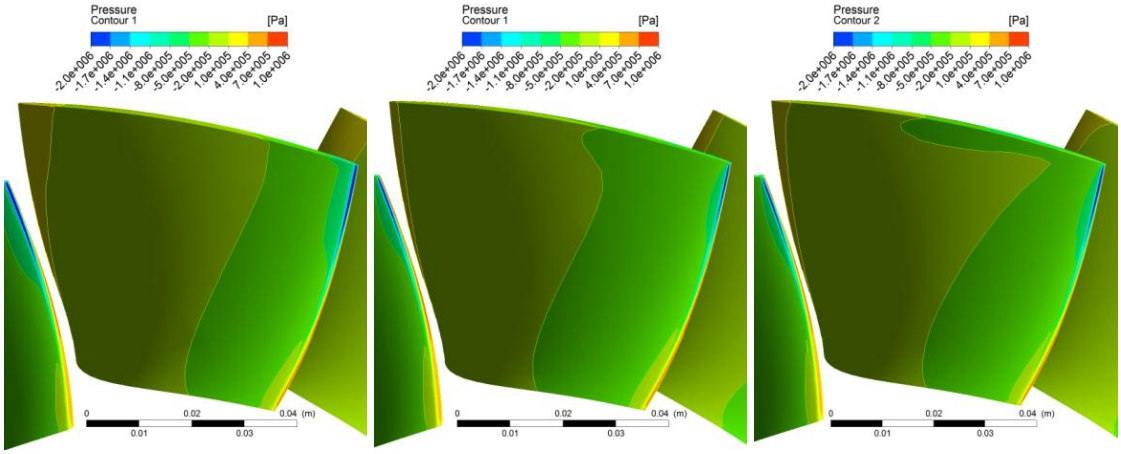
(a) $\delta = 0.2mm$ **(b)** $\delta = 1mm$ **(c)** $\delta = 3mm$

Figure 12 the flow streamlines near the rotor blade tip of different tip clearance

As shown in Figure 11 that the tip-separation vortex spreads toward the suction surface as the tip clearance increases, and the affected area is becoming more and more large. The tip-separation vortex almost covers the whole area of the tip of rotor blade in the case of δ=3mm . As for tip-leakage vortex, it can be seen from figure 12 that as the tip clearance increases, the affected area of tip-leakage vortex is more and more large too. The affected area is only focus on the area near the leading edge of the rotor blade in the case of δ=0.2mm , but the tip-leakage vortex almost affects the entire rotor passage in the case of δ=3mm . Moreover, the distance between the tip-leakage vortex core and the suction side is larger with the increasing of the tip clearance, and the tip-leakage vortex core has moved to about 1/2 in the middle of the passage in the case δ=3mm . Last but not least, as the tip clearance increases, the core of tip-separation vortex and tip-leakage vortex are "connected ", they transfer to each other. The position of the “connected” is moved to trailing edge in the axial direction, and the position is 1/3 of the blade tip away from leading edge.

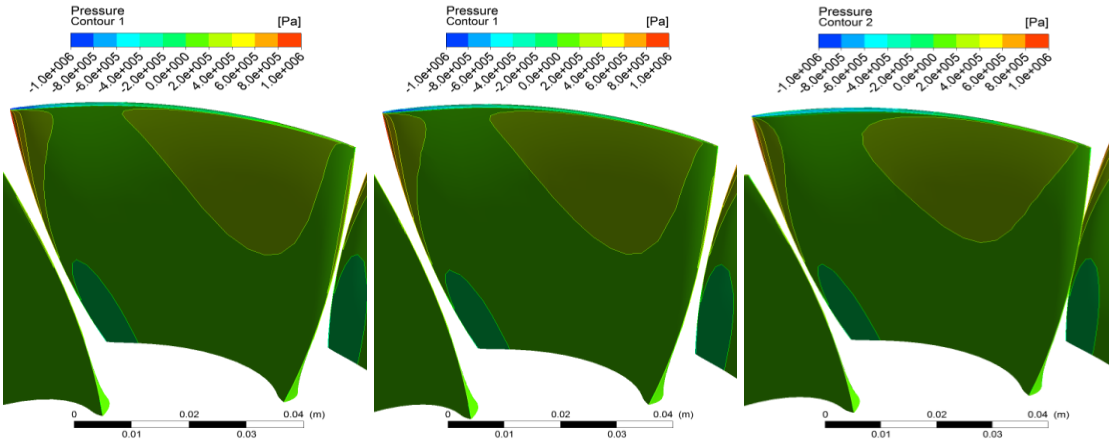
Different tip clearance effect on the pressure field of Rotor blade

Figure 13 and Figure 14 show the pressure contours on the pressure side and suction side of the rotor blade with different tip clearance:



(a) $\delta = 0.2mm$ **(b)** $\delta = 1mm$ **(c)** $\delta = 3mm$

Figure 13 the pressure contours of suction side of rotor blade



(a) $\delta = 0.2mm$ **(b)** $\delta = 1mm$ **(c)** $\delta = 3mm$

Figure 14 the pressure contours of pressure side of rotor blade

By Figure 14 can be seen that the main effected area of different tip clearance is mainly focus at the area above 0.9 spanwise of the suction side of rotor blade, and the effect of the pressure side is not very obvious.

The blade tip loading at constant span of 0.98 is illustrated in Figure 15.

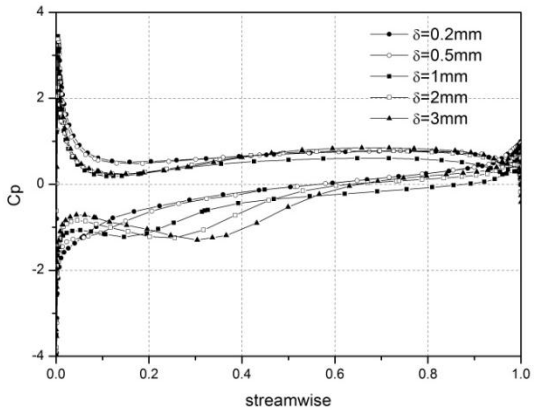


Figure 15 The blade tip loading at constant span of 0.98.

As shown in Figure 15, as the tip clearance increases, the low pressure area appears in the area above 0.9 spanwise of the suction side of rotor blade. The low pressure area gradually moved from the leading edge to the trailing edge in the axial direction, and the effected area gradually increases. The lowest Cp appeares in the axial position at about 30% streamwise, and the low pressure zone affects the area from streamwise 10% to streamwise 50% of rotor blade at constant span of 0.98.

Conclusions

In this study, a pumpjet propulsor with different size of tip clearance(δ =0.2mm、0.5mm、1mm、2mm、3mm) has been presented to investigate the influence of the tip clearance to pumpjet propulsor. This analysis was carried out with RANS method, and the *SST* $k-\omega$ turbulence model is applied. In order to verify the accuracy of numerical simulation method, calculations were carried out with a worldwide employed propeller (the E779A propeller). It is indicated that the numerical simulation method with the *SST* $k-\omega$ turbulence model is applicable and reliable for PJP flows.The influences of the clearance on pumpjet propulsor are reflected in five aspects mainly.

1)As tip clearance increases, the open water efficiency decreases gradually in the same J. The open water efficiency is basically unchanged after the tip clearance is bigger than 2mm.

2) The tip-separation vortex spreads toward the suction surface as the tip clearance increases, and the affected area is becoming bigger and bigger. The tip-separation vortex almost covers the whole area of the tip of rotor blade in the case of δ=3mm.

3) The rotor tip-leakage vortex is formed at the blade tip of the suction surface of the rotor blade and left the suction side of the rotor blade and moved toward to mid-passage with the spread of the vortex in the axial direction. Moreover, the distance between the tip-leakage vortex core and the suction side is larger with the increasing of the tip clearance, and the tip-leakage vortex core has moved to about 1/2 in the middle of the passage in the case δ=3mm.

4) As the tip clearance increases, the core of tip-separation vortex and tip-leakage vortex are "contact", they transfer to each other. The position of the contact is moved to following edge in the axial direction, and the position is 1/3 of the blade tip away from leading edge in the case δ=3mm.

5) The main effected area of different tip clearance, which is the low pressure area, is mainly focus at the area above 0.9 spanwise of the suction side of rotor blade, the effect of the pressure side is not very obvious.

6)As the tip clearance increases, the low pressure area gradually moved from the leading edge to the following edge in the axial direction, and the effected region gradually increases. The lowest point appeared in the axial position at about 30% streamwise, and the low pressure zone affects the area from streamwise 10% to streamwise 50% of rotor blade.

References

[1] Suryanarayana Ch， Satyanarayana B， Ramji K， et al(2010). Experimental evaluation of pumpjet propulsor for an axisymmetric body in wind tunnel. *International Journal of Naval Architechture and Ocean Engineering*， 2：24-33.

[2] Ivanell S. (2001) Hydrodynamic simulation of a torpedo with pump jet propulsion system. Stockholm：*Royal Institute of Technology*.

[3] Liu, Z., Song B., Huang Q., (2010) Simulation method of hydraulic performance of pump jet propulsion system based on CFD Technology28(5)：724-729.

[4] Pan, G., Hu, B., Wang, P., Yang, Z., and Wang, Y. (2013), "Numerical Simulation of Steady Hydrodynamic Performance of a Pump-jet Propulsor," *Journal of Shanghai Jiao Tong University*, Vol. 47, 932-937.
[5] Wang, T., Zhou, L., (2004) ,"Study on numerical simulation and mechanism of interaction with the mainstream flow pump jet clearance" Proceedings of the conference on hydrodynamics of ships, 212-223.
[6] Y. T. Lee， C. Hah and J. Loellbach. (2003) Flow Analyses in a Single-Stage Propulsion Pump.J. Turbomach，118(2)：240-248.
[7] Ji, B., Luo, X., WU, Y., Liu, S., Xu, H., and Oshima, A. (2010), "Numerical Investigation of Unsteady Cavitating Turbulent Flow around a Full Scale Marine Propeller," Proceedings of the 9th International Conference on Hydrodynamics, October 11-15, Shanghai, China.
[8] Subhas, S., Saji, V.F., Ramakrishna, S., and Das, N.H. (2012), "CFD Analysis of a Propeller Flow and Cavitation," International Journal of Computer Applications, Vol. 55, pp 26-33.
[9] Salvatore, F., Testa, C., and Greco, L. (2003), "A Viscous/Inviscid Coupled Formulation for Unsteady Sheet Cavitation Modelling of Marine Propellers," Fifth International Symposium on Cavitation (CAV2003), November 1–4, Osaka, Japan.
Suryanarayana, Ch., Satyanarayana, B., and Ramji, K. (2010), "Performance Evaluation of an Underwater Body and Pumpjet by Model Testing in Cavitation Tunnel," International Journal of Naval Architecture and Ocean Engineering, Vol. 2, pp 57-67.
[10] You, D., Wang, M., Moin, P. and Mittal, R., 2007. Large-eddy simulation analysis of mechanisms for viscous losses in a turbomachinery tip-clearance flow. *Journal of Fluid Mechanics*, 586, .177-204.
[11] Jeong, J. and Hussain, F., 1995. On the identification of a vortex. *Journal of fluid mechanics*, 285, .69-94.

Simple method of approximate calculation of statically indeterminate trusses

***Janusz Rębielak[1]**

[1]Laboratory of Building Structures, Faculty of Architecture, Cracow University of Technology, ul. Warszawska 24, 31-155 Kraków, Poland.

*Presenting author: j.rebielak@wp.pl

Abstract

The paper presents principles of a simple method, which in two stages makes possible the approximate calculations of statically indeterminate truss systems. The proposed two-stage method applies rules of other methods used for calculations of the statically determinate trusses. In each of the both stages there are considered the statically determinate trusses, patterns of which are obtained as results of suitable taking selected members out from pattern of the basic statically indeterminate truss. These intermediate trusses have the same clear span and construction depth like the basic indeterminate truss but they are loaded by forces of the half values applied to nodes of the same positions like in the basic one. The proposed two-stage method uses theorems and features of calculus of vectors as well as principle of the superposition method. Final values of forces acting in particular members of the basic statically indeterminate truss are resultants of forces calculated in each stage in the counterpart members of the statically determinate trusses. There are presented results of calculations carried out for two cases of loading of a selected type of the plane truss. These results are compared with results of forces determined for the same truss by application of computer calculations carried out by method appropriate for the statically indeterminate systems.

Keywords: Truss system, Calculus of vectors, Cremona's method, Superposition method, Statically indeterminate system, Approximate solution.

Introduction

Methods of calculations of statically indeterminate systems have to make possible the exact computation of the force values acting in members of such systems. Results obtained by application of them are the basis for engineers, who are obliged to design the safe and economic structural systems for very various purposes, like for instance roof structures in the building industry. There are numerous methods commonly used to calculate the statically indeterminate systems starting from e.g. the force method, the displacement method, the iterations methods like the method of successive approximations, the finite elements method etc., which were invented in the past and they are still modified and adapted to requirements of needs of the appropriate computer software [1-4]. The force distribution in area of the structural indeterminate trusses depends among others on the ratios of stiffness of members joining in particular nodes. That is why methods of precise calculations of the force values have to be complex what further implies, that computation procedures and computer calculation software have to be equally complex.

Definition of research problem and proposal of method of its solution

In preliminary structural analysis of the statically indeterminate truss it is usually enough to define only the approximate values of forces acting in its members. For these purposes it is not necessary to use a sophisticated method of calculations that is why one can apply some simple ways of the approximate computations. The proposed two-stage method has been invented during the initial statically analyses of a certain group of the spatial tension-strut

structures. It is in detail discussed in papers [5, 6]. These structures consist of cross-braces made in form of struts while other components like vertical members and members of the outer layers are the tension members. Simplified scheme of vertical cross-section of a basic truss system representing this group is shown in Fig. 1a. These types of structural systems have to be suitably pre-stressed. If the tension-strut truss is overloaded by forces F, see Fig. 1b, then certain number of the upper chord members are not able to take the compression forces, because of their big slenderness, what implies that they are excluded from process of the force transmission.

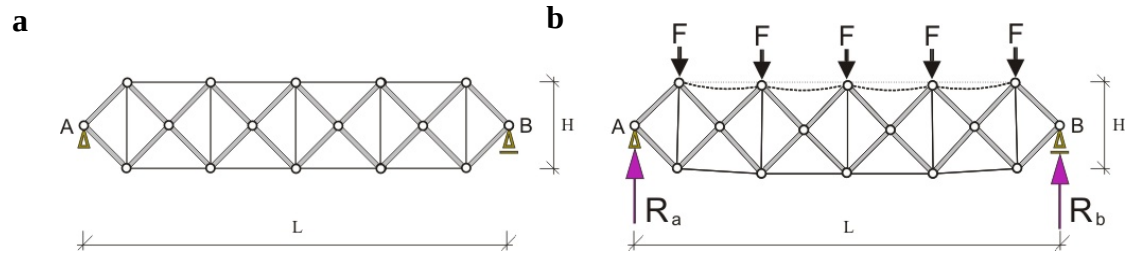


Figure 1. Schemes of plane tension-strut truss systems, a) basic configuration, b) configuration of overloaded structure

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 as:

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

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 created by number of members p = 33 what indicates that the structure is the fourfold statically indeterminate system. From analysis of the scheme shown in Fig. 1b follows that number of the excluded members equals 4, what exactly is equal to the degree of statically indeterminacy of the basic truss system. Thus the overloaded basic plane truss can be considered as the statically determinate system, what directly indicates that it can be calculated by application of one of the simple methods like e.g. Cremona's method, Ritter's method or other methods suitable for this purpose.

The observation brings to mind a following questions: is it possible to apply for instance Cremona's method for computation of statically indeterminate plane trusses? If yes, in what way it has to be done? One should be aware that values of forces determined by means of the sought after method will be of approximate values because stiffness of particular members are not taking into account in methods used for calculations of statically determinate trusses. The considered problem refers to the coplanar force system therefore the three basic conditions of equilibrium have to be fulfilled:

$$\sum_{i=1}^{n} F_{ix} = 0 \tag{3}$$

$$\sum_{i=1}^{n} F_{iy} = 0 \tag{4}$$

$$\sum_{i=1}^{n} M_i = 0 \tag{5}$$

Moreover the basic principles of calculus of forces have to be strictly respected. Taking into consideration all indicated requirements it is proposed to introduce the two-stage procedure of calculations, general scheme of which is shown in Fig. 2.

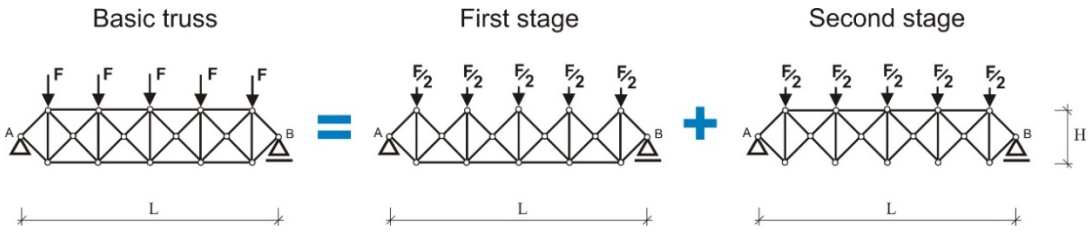


Figure 2. General schemes of two-stage method proposed for approximate calculation of statically indeterminate trusses

The point of proposed method is to carry out static calculations in two independent stages for statically determinate trusses, shapes of which are received through remove the number of members equal to statically indeterminacy from space of the basic truss. The calculated statically determinate truss has in each stage the same geometric parameters like clear span L and construction depth H, but it is loaded by forces of half values applied to the same nodes like in area of the basic truss. Values of the final forces computed in the basic truss will be resultants of forces obtained in each stage for members having the same position in area of considered truss.

Results of calculations and comparative analysis

In order to verify correctness of theoretic assumptions of the two-stage method there were carried out series of computations of simple form of the plane statically indeterminate truss having shape of basic truss shown in Fig. 2, built of steel members, having clear span equals 5.00 meters and of construction depth equal to 1.00 meter. In the basic case the truss is symmetrically loaded in all nodes of the upper chord by concentrated forces, each of value 1.00 kN. In the first stage four members of the upper chord are removed and concentrated forces of value equal to 0.50 kN are applied to all nodes of the upper chord. The own weight of truss is not taken into consideration. After this operation the investigated truss become the statically determinate system what empowers to apply, for instance, the Cremona's method for computation values of forces acting in component members of the truss. Because the basic truss is of symmetric form and it is loaded in the symmetric way that is why the Cremona's method in both the stages can be applied only for half of suitably forms of considered trusses. Results of the first stage of calculations are presented in Fig. 3.

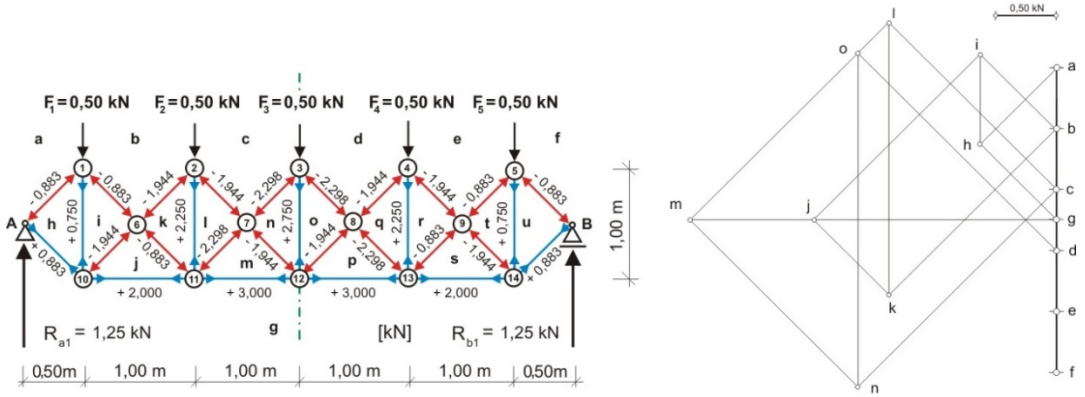


Figure 3. Scheme of distribution of values of forces calculated in the first stage in area of basic truss together with appropriate Cremona's polygon of forces

In the second stage four members, like previously, are rejected but this time from the lower chord of the basic truss and the statically determinate form of truss is loaded by concentrated forces, each of value equal to 0.50 kN and applied to each node of the upper layer. Results of the second stage of calculation are shown in Fig. 4.

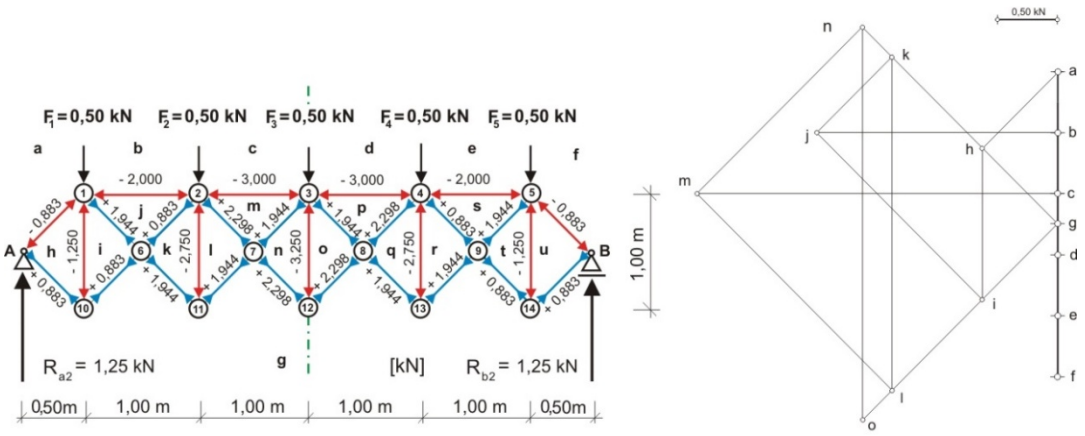
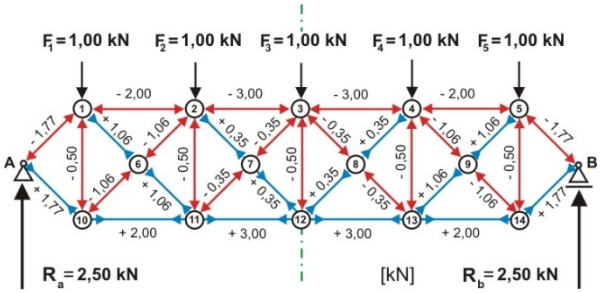


Figure 4. Values of forces defined in the second stage of calculation in area of basic truss together with appropriate Cremona's polygon of forces

Keeping rules of the proposed method the final values of forces acting in particular members are determined as resultants of forces calculated in two independent stages in the counterpart members of trusses considered in each stage, see Fig. 5a. For instance the final force acting in member of the upper chord placed between nodes of numbers 2 and 3 is the resultant of zero value for the not existing member between these nodes in the first stage, see Fig. 3, and the force value equals -3,00 kN acting in the counterpart member, determined in the second stage, see Fig. 4.

a



b

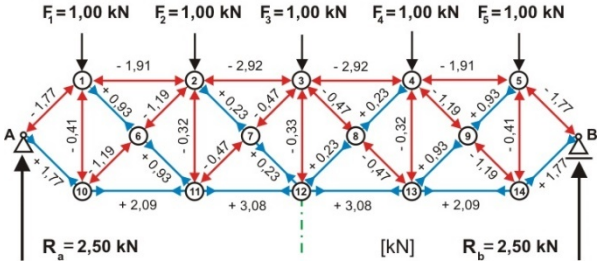


Figure 5. Values of forces calculated in the same members of basic structure by application of, a) proposed two-stage method, b) suitable computer software

The same form of the basic indeterminate truss was 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 presented in Fig. 5b. Value of the force in member placed between nodes 2 and 3 defined in the computer calculation equals -2.92 kN, so the

difference between outcomes received in the two compared methods is only 0.08 kN what is really small relatively difference because it constitutes only ca. 2.6 % regarding to the bigger force. More differences between particular values one can notice in the force values calculated in these two methods carried out in the cross braces. For instance in member placed between nodes 3 and 7 the force value calculated in the two-stage method equals -0.35 kN, while by application of the suitable computer software it is equal to -0.47 kN, what constitutes the differentiation of around 25 % towards the bigger value. In this place one should to point out that the biggest differences of the force values are observed in members, where are acting the really smallest forces. More precise answer for question about degree of approximation of results obtained in the proposed two-stage method in comparison to results defined in the exact method one can receive due to the static calculation of the same basic truss but conducted now e.g. for an asymmetric way of its load. The demanded computations were carried out for selected case, where two concentrated forces of the same value equal to 1.00 kN are applied to nodes of the upper chord and having numbers 4 and 5. Results of the both intermediate calculations are shown in Fig. 6 and in Fig. 7.

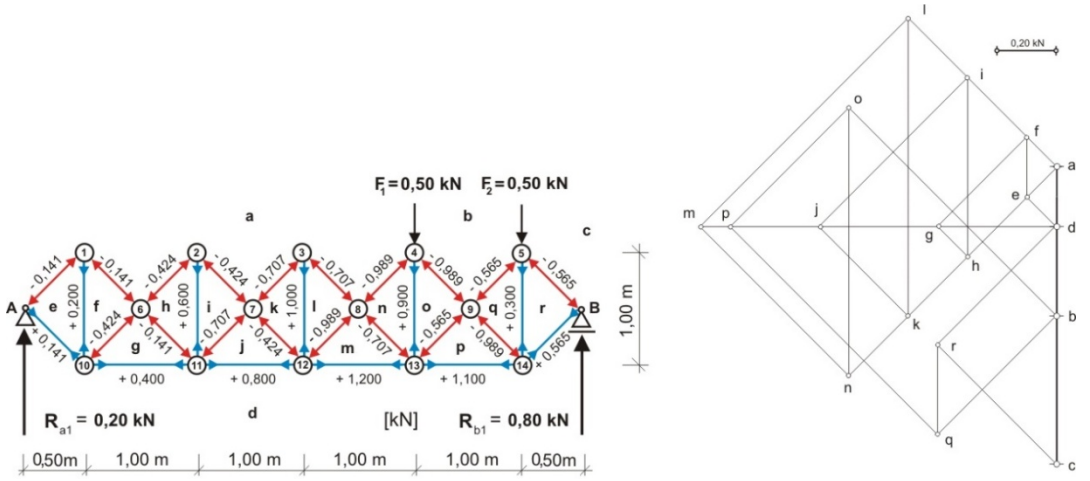


Figure 6. First stage of calculation of basic truss under asymmetric load

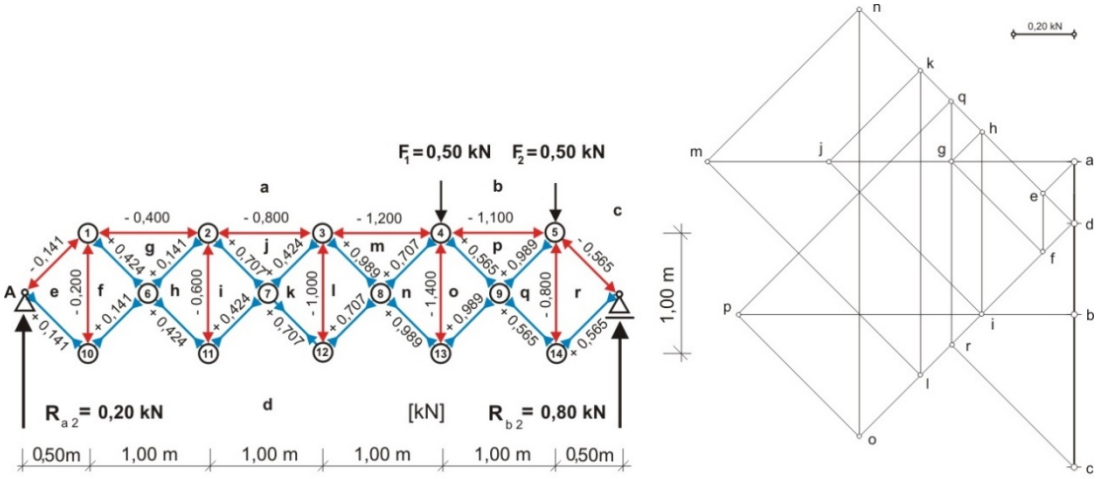


Figure 7. Second stage of calculation of basic truss under asymmetric load

Because of the asymmetric way of application of the load both procedures of computation of the intermediate trusses have to be conducted for the whole structures. From analysis of information overall presented in Fig. 8 follows, that the biggest differences one can notice between the force values calculated by means of both compared methods in certain cross braces. For example the force value in the cross brace located between nodes 5 and 9 defined in the two-stage method equals +0.42 kN, while by means of the computer software its

computed value is equal to +0.28 kN. Then the difference is on level of ca. 33 % towards to the bigger value. Significantly smaller differences one can observe between values of forces calculated in both methods in the most strained members of the outer chords. For instance the force value in member placed between nodes 13 and 14 defined in two-stage method equals +1.10 kN but calculated by application of suitable computer software it is equal to +1.20 kN, what constitutes only around 8.3 % towards the bigger value.

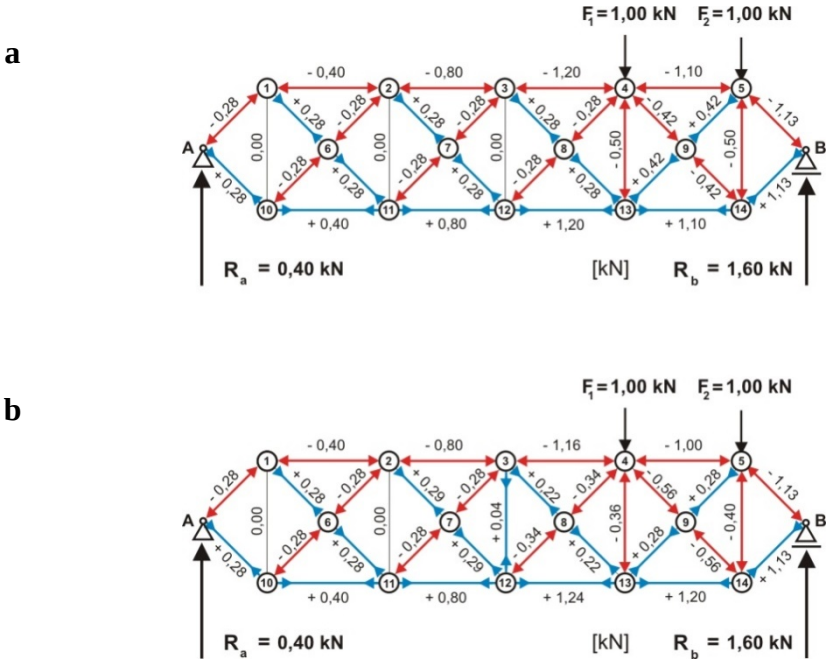


Figure 8. Results of static calculations carried out for the same basic truss by application of a) proposed two-stage method, b) suitable computer software

Conclusions

From comparative analysis of outcomes obtained in the both compared computation ways follows that the two-stage method can be applied as an approximate method of calculation of the plane statically indeterminate trusses. Its accuracy can be improved in the future by taking into consideration the stiffness differences between members connected in particular nodes of the considered structural system.

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] Zienkiewicz, O.C. and Taylor R.L. (2000) *The finite element method*, Oxford Press, UK.
[4] Allen E., Zalewski W. and Boston Structures Group (2010) *Form and forces. Designing efficient, expressive structures*, John Wiley & Sons, Hoboken, New Jersey.
[5] Rębielak, J. (2014) A two-stage method for an approximate calculation of statically indeterminate trusses, *Journal of Civil Engineering and Architecture* **78**, 567-572.
[6] 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.

Chemical Reaction, Heat and Mass Transfer on Unsteady MHD Flow along a Vertical Stretching Sheet with Heat Generation/Absorption and Variable Viscosity

Jatindra Lahkar

Department of Mathematics, Digboi College, Digboi-786171, Assam, India, e-mail:jatindralahkar@gmail.com

Abstract

The effect of chemical reaction on laminar mixed convection flow and heat and mass transfer along a vertical unsteady stretching sheet is investigated, in the presence of heat generation/absorption with variable viscosity and viscous dissipation. The governing non-linear partial differential equations are reduced to ordinary differential equations using similarity transformation and solved numerically using the fourth order Runge-Kutta method along with shooting technique. The effects of various flow parameters on the velocity, temperature and concentration distributions are analyzed and presented graphically. Skin-friction coefficient, Nusselt number and Sherwood number are derived at the sheet.

Introduction

Processes involving magnetohydrodynamics(MHD) heat and mass transfer flow in the boundary layer, induced by a moving surface in a fluid with chemical reaction occur frequently in nature. It occurs not only due to temperature difference but also due to magnetic field or combination of these. In chemical engineering there are many transport processes that are governed by the joint action of the buoyancy forces from both thermal and mass diffusion in the presence of chemical reaction effects. During a chemical reaction between two species heat is also generated. Diffusion and chemical reactions in an isothermal laminar flow along a soluble flat plate were studied by Fairbanks and Wike [1]. Chakrabarti and Gupta [2] investigated hydromagnetic flow and heat transfer over stretching sheet. Apelblat[3] presented mass transfer with a chemical reaction of first order with effects of axial diffusion.The effects of mass transfer on flow past an impulsively started infinite vertical plate with constant heat flux and chemical reaction were studied by Das et al.[4].

Anjalidevi and Kandasamy [5] studied the steady laminar flow along a semi-infinite horizontal plate in the presence of a species concentration and chemical reaction. Fan et al. [6] studied the mixed convective heat and mass transfer over a horizontal moving plate with a chemical-reaction effect. Takhar et al. [7] investigated the flow and mass diffusion of a chemical species with first-order and higher order reactions over a continuously stretching sheet with an applied magnetic field. Muthucumaraswamy [8] studied the effects of a chemical reaction on a moving isothermal vertical infinitely long surface with suction. Anjali Devi and Kandasamy [9] studied effects of chemical reaction, heat and mass transfer on non-linear MHD laminar boundary layer flow over a wedge with suction and injection. Chamkha [10] presented an analytical solutions for heat and mass transfer by laminar flow of a Newtonian, viscous, electrically, conducting and heat generation absorption. The effects of radiation and chemical reactions, in the presence of a transverse magnetic field, on free convective flow and mass transfer of an optically dense viscous, incompressible, and electrically conducting fluid past a vertical isothermal cone surface are investigated by Afify [11]. Kandasamy et al. [12] studied the nonlinear MHD flow with heat and mass transfer characteristics of an incompressible, viscous, electrically conducting and Boussinesq fluid on a vertical stretching surface with chemical reaction and thermal stratification effects.

The combined effects of non-uniform heat source/sink and thermal radiation on heat transfer over an unsteady stretching permeable surface was discussed by Pal [13]. Unsteady mixed convection heat transfer over a vertical stretching surface with variable viscosity and viscous dissipation was studied by Aziz [14]. Radiation and Magnetic field Effects on Unsteady Mixed Convection Flow over a Vertical Stretching/Shrinking surface with suction/injection was discussed by Sandeep et al[15].

The objective of the paper is to investigate the influence of heat and mass and magnetic field on an unsteady flow over a vertical stretching sheet with heat generation/absorption and chemical effects in presence of variable viscosity and viscous dissipation.

Formulation of the problem

An unsteady, two-dimensional, boundary-layer convective flow of an incompressible, viscous and electrically conducting fluid along a vertical stretching sheet embedded in porous media in the presence of heat and mass transfer, chemical reaction is considered. The x-axis is considered along the sheet and y-axis is perpendicular to the sheet. The fluid properties are assumed to be constant except the viscosity, the density term of in buoyancy terms of the momentum equations and the chemical reaction is homogeneous and of first order taking place in the flow. The sheet is stretching in its own plan with velocity

$$u_w = bx/(1-\alpha t) \tag{1}$$

where $b(>0)$ is the stretching parameter and $\alpha(>0)$ is the unsteadiness parameter and both have dimensions of (time)$^{-1}$. The surface temperature T_w and concentration distribution of the sheet C_w, which varies with the distance x along the sheet and time t. The system is influenced by an external transverse magnetic field of strength B defined as

$$B = B_0(1-\alpha t)^{-1/2} \tag{2}$$

The volumetric rate of heat generation/absorption is given as

$$Q = Q_0(1-\alpha t)^{-1} \tag{3}$$

Under above assumptions, the governing equations of continuity, momentum, energy and concentration are given by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{4}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \frac{1}{\rho_\infty}\frac{\partial}{\partial y}\left(\mu\frac{\partial u}{\partial y}\right) + g\beta(T - T_\infty) + g\beta^*(C - C_\infty) - \frac{\sigma B^2}{\rho_\infty}u \tag{5}$$

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{k}{\rho_\infty C_p}\frac{\partial^2 T}{\partial y^2} + \frac{\mu}{\rho_\infty C_p}\left(\frac{\partial u}{\partial y}\right)^2 + \frac{Q}{\rho_\infty C_p}(T - T_\infty) \tag{6}$$

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = D\frac{\partial^2 C}{\partial y^2} - K_1(C - C_\infty) \tag{7}$$

along with the boundary conditions

$$\begin{aligned} &\text{at } y=0:\ u=u_w,\ v=0,\ T=T_w,\ C=C_w \text{ and} \\ &\text{as } y\to\infty:\ u\to 0,\ T\to T_\infty,\ C\to C_\infty \end{aligned} \tag{8}$$

where u and v are the velocity components along the x and y directions respectively, ρ_∞ is the density of the fluid, g is the gravitational acceleration, β is the thermal expansion coefficient, β^* is the concentration expansion coefficient, σ is the electrical conductivity, T is fluid

temperature inside the thermal boundary layer, C is the species concentration in boundary layer, T_∞ is the temperature far away from the sheet, C_∞ is the species concentration far away from the sheet. C_p is the specific heat at constant pressure, k is the thermal conductivity, D is the mass diffusion coefficient. Variation of the viscosity with temperature are assumed to be of the form[16]

$$1/\mu = [1+\tau(T-T_\infty)]/\mu_\infty = a(T-T_r), \quad (9)$$

where $a=\tau/\mu_\infty$ and $T_r=T_\infty - 1/\tau$ (10)

are constants and their values depend on reference state and the thermal property of the fluid τ. Also $T_w=T_\infty+[bx^2/2\nu_\infty](1-\alpha t)^{-2}$ and $C_w=C_\infty+[bx^2/2\nu_\infty](1-\alpha t)^{-2}$, where ν_∞ is the kinematic viscosity of the fluid.

Introducing the similarity variable η and the dimensionless variables f, θ and ϕ as:

$$\eta=(b/\nu_\infty)^{1/2}(1-\alpha t)^{-1/2}y \quad (11)$$

$$\psi=[\nu_\infty b/(1-\alpha t)]^{1/2}x.f(\eta) \quad (12)$$

$$T=T_\infty+[bx^2/2\nu_\infty](1-\alpha t)^{-2}\theta(\eta) \quad (13)$$

$$C=C_\infty+[bx^2/2\nu_\infty](1-\alpha t)^{-2}\phi(\eta) \quad (14)$$

where $\psi(x,y,t)$ is the stream function satisfying the continuity equation (4) with $u=\partial\psi/\partial y$ and $v= -\partial\psi/\partial x$. The components of velocity can be readily expressed as:

$$u= u_w f'(\eta),\ v=-[\nu_\infty b/(1-\alpha t)]^{1/2} f(\eta) \quad (15)$$

Making use of Eqs. (11)-(14), Eqs. (5)-(7) reduce to

$$f''' = \frac{\theta_v-\theta}{\theta_v}[A(f'+0.5\eta f'')+f'^2-ff''-\lambda(\theta+N\phi)+Mf'] - \frac{\theta' f''}{\theta_v-\theta} \quad (16)$$

$$\theta'' = Pr\left[A(2\theta'+0.5\eta\theta')-f\theta'+2f'\theta-S\theta-\frac{\theta_r}{\theta_v-\theta}Ecf''^2\right] \quad (17)$$

$$\theta'' = Sc[A(2\phi+0.5\eta\phi')-f\phi'+2f'\phi+\gamma\phi] \quad (18)$$

The transformed boundary conditions:

at η=0: f=0, f ′ =1, θ=1, ϕ=1 and

at η→∞: f ′→0, θ→0, ϕ →0. (19)

where a prime denotes ordinary differentiation with respect to η, θ = (T-T∞)/(Tw-T∞) is the non-dimensional temperature, $\theta_v = (T_r-T_\infty)/(T_w-T_\infty) = -1/\tau(T_w-T_\infty)$ is the variable viscosity parameter, A= α/b is the unsteadiness parameter, $Ec=2b\nu_\infty/C_p = u_w^2/C_p(T_w-T_\infty)$ is the Eckert number, $M=2\sigma B_0^2(1-\alpha t)/\rho_\infty b$ *is the Magnetic number,* $N=\beta^*(C_w-C_\infty)/\beta(T_w-T_\infty)$ *is the Buoyancy ration parameter,* $Pr=\nu_\infty\rho_\infty C_p/k$ is the Prandtl number, S=Q(1-αt)/aρC_p is the heat generation/absorption Parameter and $\lambda=g\beta x/2b\nu_\infty = Gr_x/Re_x^2$ is the mixed convection parameter with $Gr_x = g\beta(T_w-T_\infty)x^3/\nu_\infty^2$ is the Grashof number. The case in which λ=0 corresponds to the forced convection regime while that in which λ is large corresponds to thefree convection regime.

For practical applications, the physical quantities of major interest are the local friction coefficient C_{fx}

$$C_{fx} = 2\mu(\partial u/\partial y)_{y=0}/\rho_\infty u_w^2 = (2\theta_v/\theta_v-1)Re_x^{-1/2}f''(0) \quad (20)$$

the local Nusselt number Nu_x

$$Nu_x=-x(\partial T/\partial y)_{y=0}=-Re_x^{3/2}\theta'(0)/[2(1-\alpha t)] \quad (21)$$

and local Sherwood number Sh_x

$$Sh_x=-x(\partial C/\partial y)_{y=0} \quad =-Re_x^{3/2}\phi'(0)/[2(1-\alpha t)] \quad (22)$$

where $Re_x = u_w x/\nu_\infty$ is the local Reynolds number based on the sheet velocity u_w.

Results and discussion

The non-linear coupled differential Eqs. (16)-(18) with boundary condition (19) and constitutes a boundary value problem has been solved numerically by fourth order Runge-Kutta Shooting method for different values of the parameters. Effect due to magnetic field and chemical reaction at the wall of the cone over the velocity, temperature and concentration are shown through figures 1-3. Fig. 1 depicts the dimensionless velocity profiles for different values of magnetic field and chemical reaction parameters. It observes that the velocity component of the fluid along the surface of the sheet increase with decrease of the strength of the magnetic field, on the contrary, fig. 2 and 3 shows the dimensionless temperature and concentration of the fluid increase with increase of the strength of the magnetic field. On the other hand the dimensionless velocity and temperature of the fluid reduce with an increase of chemical reaction parameter while the dimensionless concentration has the opposite behavior.

Figures 4-6 present the effects of the unsteadiness parameter A on the velocity, temperature and concentration profiles, respectively. From these figures it observed that increasing value of A results in decreasing the velocity, temperature and concentration keeping other parameters fixed. Figures 7-9 illustrate the influence of the chemical reaction parameter γ and the Schmidt number Sc on the velocity, temperature and concentration profiles in the boundary layer, respectively. Increasing the chemical reaction parameter produces a decrease in the species concentration. In turn, this causes the concentration buoyancy effects to decrease as γ increases. Consequently, less flow is induced along the sheet resulting in decreases in the fluid velocity in the boundary layer. On the other hand, the fluid temperature increases as γ increases. In addition, the concentration boundary layer thickness decreases as γ increases. Moreover, the Schmidt number is an important parameter in heat and mass transfer processes as it characterizes the ratio of thicknesses of the viscous and concentration boundary layers. Its effect on the species concentration has similarities to the Prandtl number effect on the temperature. That is, increases in the values of Sc cause the species concentration and its boundary layer thickness to decrease resulting in less induced flow and higher fluid temperatures. This is depicted in the decreases in the velocity and species concentration and increases in the fluid temperature as Sc increases. These behaviors are clearly evident in Figures 7-9. The influence of heat generation/absorption over velocity, temperature and concentration are elucidated with the help of figures 10-12. It is clear that the velocity of the fluid increases with increase of heat generation parameter S but the temperature and concentration of the fluid increases with increase of S. On the other hand the velocity, temperature and concentration of the fluid decrease with the increasing values chemical reaction parameter γ.

Conclusions

We conclude the following from above results and discussion:

1.The influence of chemical reaction, the fluid flow along the sheet accelerate with increase of chemical reaction parameter, on the other hand, temperature of the fluid increases with increase of chemical reaction parameter but concentration of the fluid reduces with it.

2. For all values of unsteadiness parameter, increasing values of the chemical reaction parameter the boundary layer decreases on the surface of the sheet.

3. The increases in the values of *Sc* cause the species concentration and its boundary layer thickness to decrease resulting in less induced flow and higher fluid temperatures. This is depicted in the decreases in the velocity and species concentration and increases in the fluid temperature as *Sc* increases.
4. Due to heat generation, increases of heat generation parameter accelerate the fluid motion and decelerate the temperature and concentration of the fluid along the sheet.

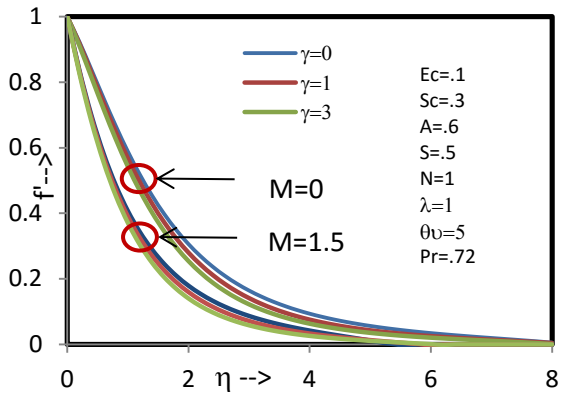


Fig.1 Velocity profiles for various values of chemical reaction γ and magnetic parameter M.

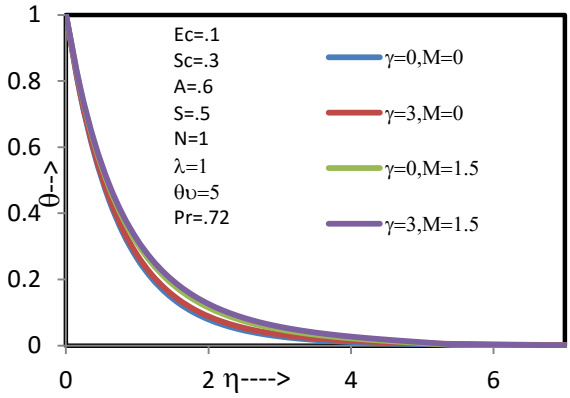


Fig.2 Temperature profiles for various values of chemical reaction γ and magnetic parameter M.

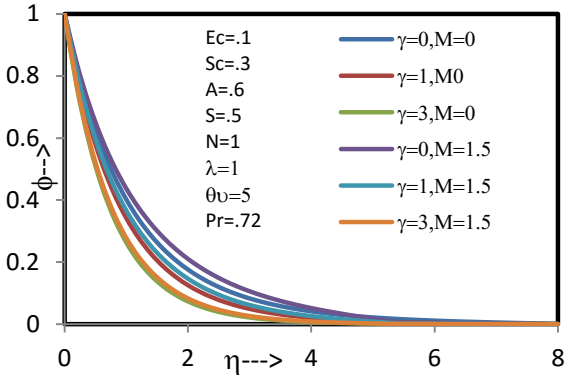


Fig. 3 Concentration profiles for various values of chemical reaction γ and magnetic parameter M.

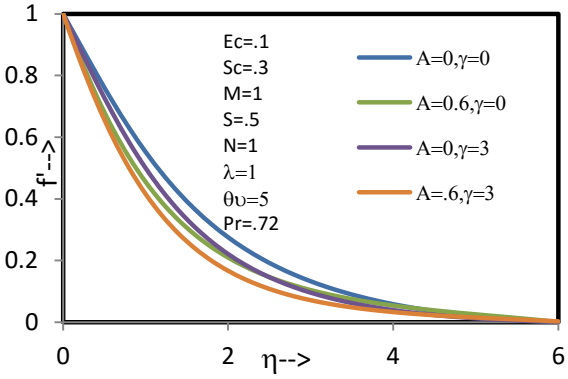


Fig.4 Velocity profiles for various values of chemical reaction γ and unsteady parameter *A*

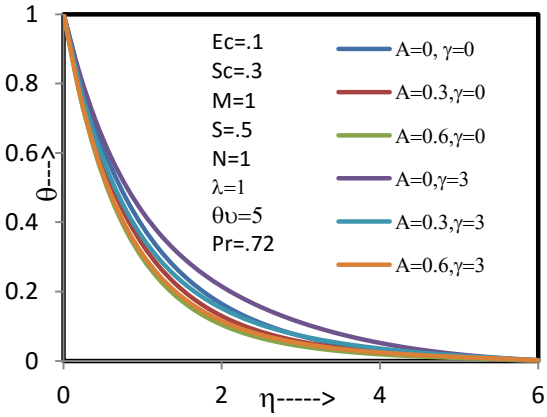


Fig. 5 Temperature profiles for various values of chemical reaction γ and unsteady parameter *A*

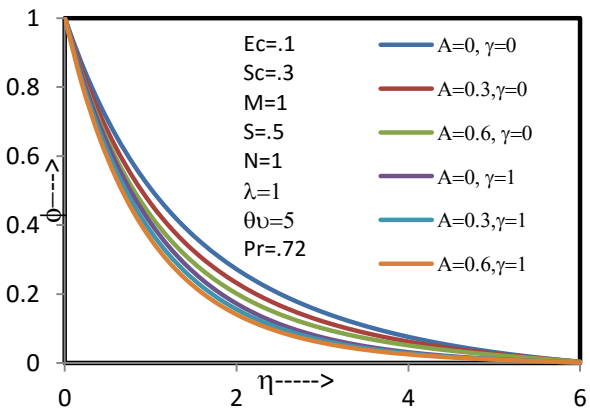


Fig. 6 Concentration profiles for various values of chemical reaction γ and unsteady parameter *A*.

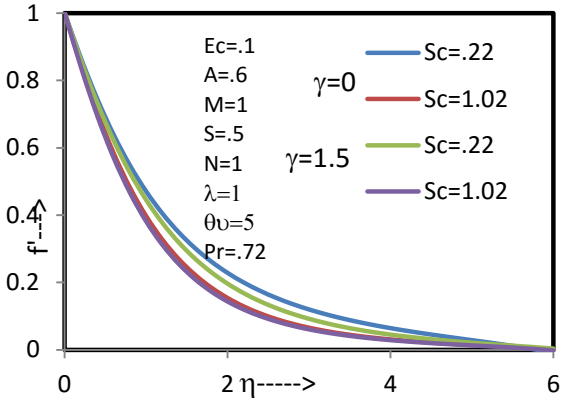


Fig.7 Velocity profiles for various values of chemical reaction γ and Schmidt number Sc.

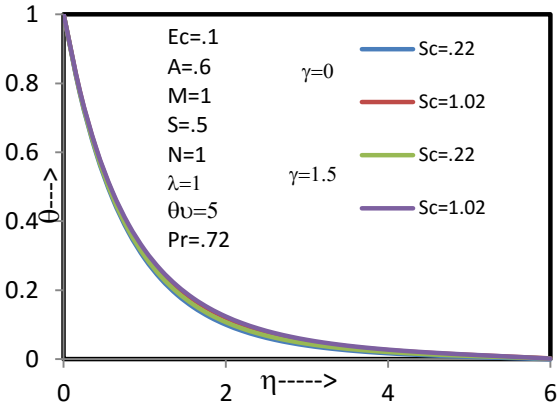


Fig.8 Temperature profiles for various values of chemical reaction γ and Schmidt number Sc.

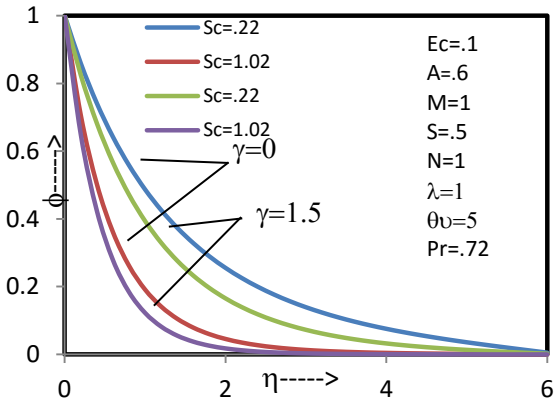


Fig. 9 Concentration profiles for various values of chemical reaction γ and Schmidt number Sc.

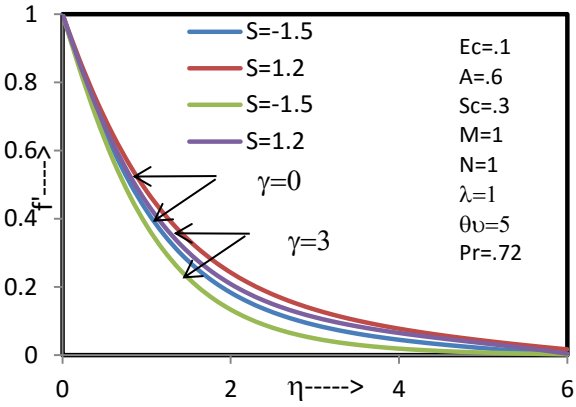


Fig.10 Velocity profiles for various values of chemical reaction γ and heat generation S.

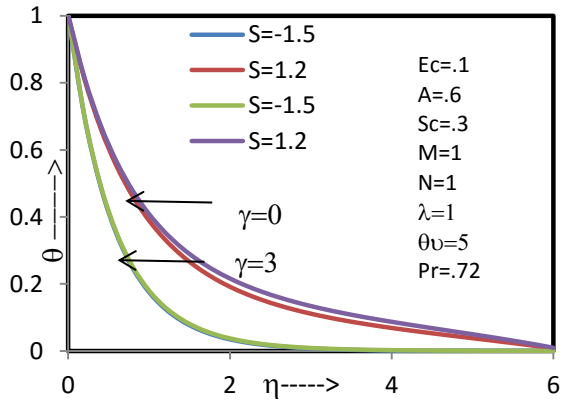


Fig.11 Temperature profiles for various values of chemical reaction γ and heat generation S.

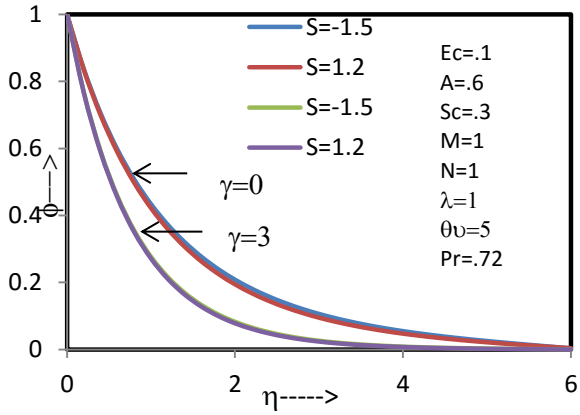


Fig. 12 Concentration profiles for various values of chemical reaction γ and heat generation S.

References

[1] Fairbanks, D.F. and Wike, C.R.(1950) Diffusion and chemical re-action in an isothermal laminar flow along a soluble flat plate, Ind. Eng. Chem. Res., 42, 471- 475.

[2] Chakrabarti, A. and Gupta A. S.(1979) Hydromagnetic flow and heat transfer over stretching sheet, Quarterly Journal of Mechanics and Applied Mathematics, 37, 73-78.

[3] Apelblat A.(1982) Mass transfer with a chemical reaction of the first order: effects of axial diffusion, *The Chemical Engineering Journal*, 23, 193-203.

[4] Das, U.N., Deka, R.A. and Soundalgekar, V.M.(1994) Effect of Mass Transfer on Flow Past an Impulsively Started Infinite Vertical Plate with Constant Heat Flux and Chemical Reaction, Forschung im Ingenieurwesen, 60, 284-287.

[5] Anjalidevi, S. P. and Kandasamy, R. (1999) The steady laminar flow along a semi-infinite horizontal plate in the presence of a species concentration and chemical reaction. *Heat Mass Transfer*, 35 465-467.

[6] J.R. Fan, J.M. Shi, and X.Z. Xu, (1998) The mixed convective heat and mass transfer over a horizontal moving plate with a chemical-reaction effect. *Acta Mech.*, 126, 59-69.

[7] Takhar, H.S., Chamkha, A.J. and Nath, G.(2000) Flow and mass transfer on a stretching sheet with a magnetic field and chemically reactive species, Int. J. Eng. Sci., 38, 1303-1314.

[8] Muthucumaraswamy, R. (2002) Effect of a Chemical Reaction on a Moving Isothermal Vertical Surface with Suction", Acta Mechanica, 155, 65-70.

[9] Anjali Devi S.P. and Kandasamy, R. (2002) Effects of Chemical Reaction, Heat and Mass Transfer on Non-Linear MHD Laminar Boundary Layer Flow over a Wedge with Suction and Injection, Int. Comm. Heat Mass Transfer, 29, 707-716.

[10] Chamkha, A.(2003) MHD Flow of Uniformly Stretched Vertical Permeable Surface in the Presence of Heat Generation/Absorption and a Chemical Reaction, Int. Comm. Heat Mass Transfer, 30, 413-422.

[11] Afify, A.A.(2004) The effect of radiation on free convective flow and mass transfer past a vertical isothermal cone surface with chemical reaction in the presence of a transverse magnetic field, Can. J. Phys., 82, 447-458.

[12] Kandasamy, R., Periasamy, K. and Sivagnana Prabhu, K.K.(2005) Chemical reaction, heat and mass transfer on MHD flow over a vertical stretching surface with heat source and thermal stratification effects, Int. J. of Heat and Mass Transfer, 48, 4557-4561.

[13] Pal D.(2011) Combined effects of non-uniform heat source/sink and thermal radiation on heat transfer over an unsteady stretching permeable surface, *J. Communications in Nonlinear Science and Numerical Simulation*. 16 (4), 1890-1904.

[14] Aziz, M.A.E. (2014). Unsteady mixed convection heat transfer along a vertical stretching surface with variable viscosity and viscous dissipation, J. of the Egyptian Mathematical Society, 22, 529–537.

[15] Sandeep N, Sulochana C., Sugunamma V.(2015) Radiation and magnetic field effects on unsteady mixed convection flow over a vertical stretching/shrinking surface with suction/injection, *Industrial Engineering Letters*, 5(5), 127-136.

[16] Lai, F.C. and Kulacki, F.A. (1990) The effect of variable viscosity on convective heat transfer along a vertical surface in a saturated porous medium, Int. J. Heat Mass Transfer 33(5), 1028-1031.

Development of a cellular automaton for a better consideration of elastic neighborhood effect in polycrystals

***Remy Bretin[1], Philippe Bocher[1] and Martin Levesque[2]**

[1]Mechanical Engineering Department, Ecole de Technologie Superieure (ETS), 1100 rue Notre-Dame Ouest, Montreal, H3C 1K3 Quebec, Canada
[2]Laboratory for Multiscale Mechanics (LM2), Department of Mechanical Engineering, Ecole Polytechnique de Montreal, C.P. 6079, succ. Centre-ville, Montreal, Quebec H3C3A7, Canada
*Presenting and corresponding author: remy.bretin.1@ens.etsmtl.ca

Abstract

This paper presents the development of a cellular automaton (CA) which could take into account the neighborhood effect in the context of polycrystal mechanics. This model aims to have a better estimate of the stress / strain field in polycrystals than conventional analytical models such as the self-consistent model (SCM). As the first step in the consideration of neighborhood effect, the model was developed in the case of a uniaxial loading in linear elasticity. A Kelvin structure is used to represent a polycrystal, considering that all grains have the same size and shape. The primary focus is the influence of crystallographic orientations on the local behavior in the microstructure. The model has been developed based on the hypothesis that the Finite Element Method (FEM) can quantify correctly the influence of a grain's neighborhood on its behavior. FEM, SCM, and the analytical model results are finally compared grain by grain after simulations on 686 grains polycrystalline aggregates in the Kelvin structure. The results show that the developed CA provides an approximation almost three times better than those of the SCM and the importance of taking into account the neighborhood effect. This also gives an opportunity to better understand the parameters that influence the behavior of a grain in a polycristal.

Keywords: Cellular automaton, Homogenization Model, Anisotropy, Eshelby's inclusion, Polycrystal, Neighborhood.

Introduction

The Finite Element Method (FEM) is the most common method used to simulate the micromechanical behavior of polycrystals [1, 2]. It requires a heavy amount of computer resources and the analytic model such as the Self-Consistent Model (SCM) [3] can offer a good approximation of the micromechanical behavior of polycrystals for a much lower calculation time, allowing for the possibility to simulate various microstructure configurations. However, the SCM does not take into account the neighborhood effect as each phase is defined solely by its crystal orientation and volume fraction, and there is no spatial representation of the polycrystal.

This is quite unfortunate as the neighborhood effect is an important criterion to consider if one wants to describe the behavior of the material on a local scale [4]: a grain surrounded by soft grains will not show the same behavior as the same grain surrounded by hard grains. In order to take this into account, the principles of the cellular automaton model have been considered by several authors to address this limitation [5, 6, 7]

A cellular automaton (CA) is a discrete mathematical model where the structure is discretized into several cells. Each cell has a characteristic initial state that characterizes it, and its behavior

Table 1: Elastic constants of the cubic iron crystal ([11])

Fe	$C_{1111} = 226$ GPa	$C_{1122} = 140$ GPa	$C_{1212} = 116$ GPa

depends on the behavior of its neighboring cells. This is described by a transition rule. Such CA can describe and predict complex behaviors in many different fields of research [8, 9, 10].

In this context, the objective of this research is to develop a simple model based on the SCM by adding the principles of CA in order to take into account the neighborhood effect. As a first step, the structure of Kelvin (Fig.1) has been used to annihilate the effects of grain shape and size ratio, and to document only the effect of grain orientation of the neighboring grains. Non cristallographic texture was introduced in the microstructure. They were tested in uniaxial loading with linear elasticity properties. In order to have a better understanding of the influence of the neighborhood of a grain, a full-field numerical study has been proceeded using the FEM. From the hypotheses that FEM simulation gives "the right" results, a model has been developed and its results have been compared with the ones obtained by FEM approach.

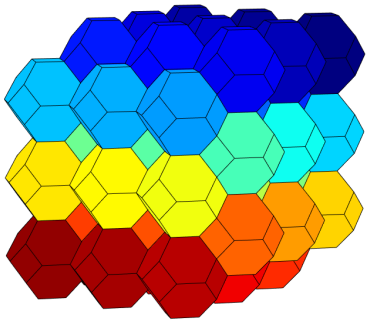


Figure 1: Kevin Structure

The material properties used in the results shown in this paper are the properties of the iron crystal (Tab.1)(Cubic elastic tensor), but the approach was also applied to other materials such as Aluminium, Nickel or Titanium with similar accuracy.

Comparison of the FEM and the SCM

FEM simulations have been performed on a cube of 686 grain, with 20 different crystallographic orientation distributions on a Kelvin structure (Fig. 1) to simplify the study of the neighborhood effect and cancel any size and shape effect (all grains are identical). Periodic boundary conditions have been applied in order to cancel any border effects. Arbitrarily, a uniaxial loading $\underline{\underline{E}}^0$ is applied to the cube, where $E_{ij} = 0$ except for $E_{33} = 0.1\%$. The resulting effective stress is $\Sigma_{1st}^{eff} = 274$MPa. The crystals are purely elastic, and the elastic tensor of the iron crystal (Tab.1) has been used for all the simulations that are presented in the rest of the paper.

On Fig.2 are presented the mean first principal stress in each grain obtained with the FEM and the SCM. For a given grain Young modulus, the FEM shows a dispersion of the stress where as the SCM shows only one possible stress solution to the problem. The SCM is actually an average of the stress observed with the FEM. This dispersion is clearly related to some neighborhood effect. With some neighborhood conditions, that could lead to significant increase of the local stress compared to the SCM. The highest stress observed with the SCM is approximately 108% of Σ_{1st}^{eff} when the highest stress observed with the FEM is 120% of Σ_{1st}^{eff}.

Study of the neighborhood effect

The SCM showed to have a good first approximation of the stress in the grain but it clearly needs to be corrected to consider the neighborhood effect and predict more realistic and statistical results. In order to do that, the influence of one or several grains on a given grain has been studied depending on their spacial and crystal orientation distribution.

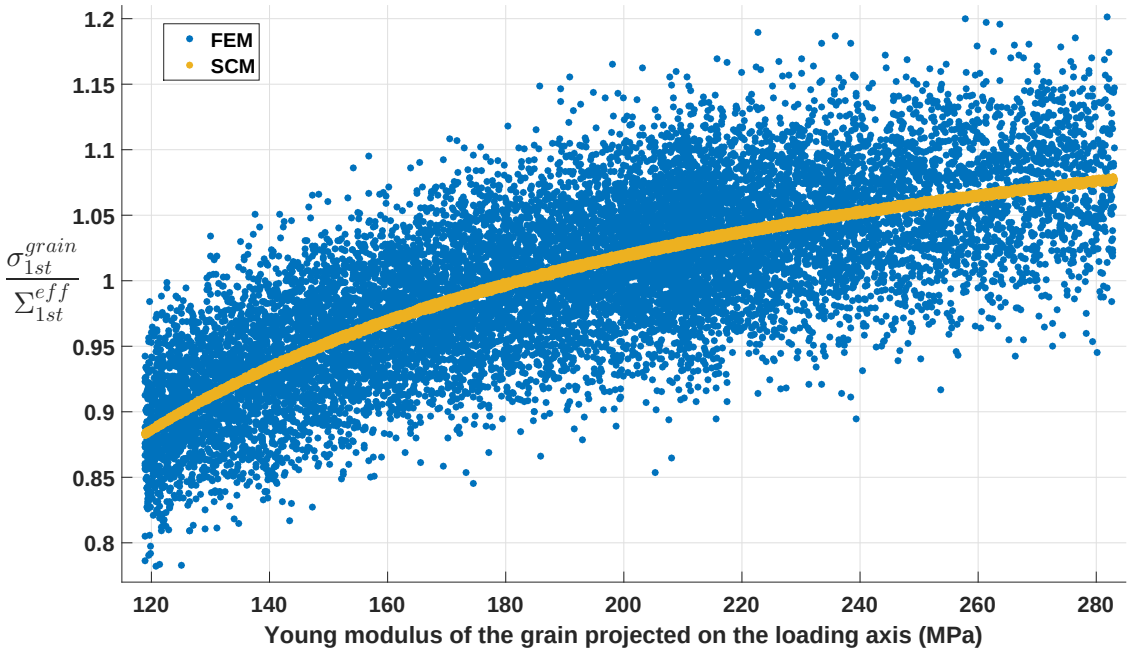


Figure 2: First principal stress in each grain of the polycrystal normalized by the macroscopic first principal stress: Comparison between the FEM and the SCM

Grain interaction

In order to see the influence of one grain on an other grain in the polycrystalline Kelvin structure, the behavior of two grains A and B immersed in an infinite homogeneous matrix has been documented as a function of their crystal orientations and their relative position.

Firstly, the macroscopic properties of the material are attributed to the homogeneous matrix and the central grain A, and crystallographic properties are attributed to grain B.

The first principal stress of grain A (0;0;0) has been observed depending on the position X, Y, Z and the crystal orientation of grain B. An influence factor α_B^A of grain B on grain A is defined such as $\alpha_B^A = \sigma_B^A/\sigma_0^A$, where σ_0^A is the first principal stress of grain A immersed alone in the matrix and σ_B^A is the first principal stress of grain A immersed in the matrix with grain B.

It was found that if the elastic tensor $\mathbb{C}_B$ of grain B is expressed in the local base where the axis Z is parallel to the loading axis and the axis X points toward the projection of grain B on the plan perpendicular to the loading direction (Fig. 3), the components $\mathbb{C}_{3333}$, $\mathbb{C}_{1133}$ or $\mathbb{C}_{3313}$ of this elastic tensor $\mathbb{C}_B$ directly influence the factor α_B^A (Eq.1).

$$\begin{aligned}\alpha_B^A(X,Y,Z) = a_1^{(X,Y,Z)} &\times \mathbb{C}_{3333} \\ + a_2^{(X,Y,Z)} &\times \mathbb{C}_{1133} \\ + a_3^{(X,Y,Z)} &\times \mathbb{C}_{3313} \\ + a_4^{(X,Y,Z)}&\end{aligned} \quad (1)$$

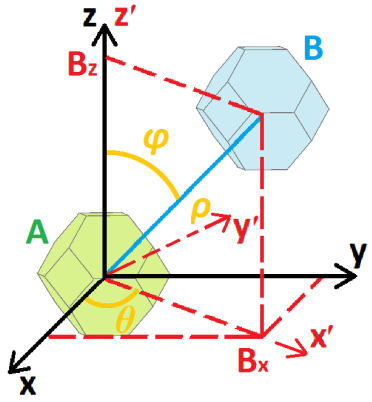


Figure 3: Illustration of the local base formed by grains A and B

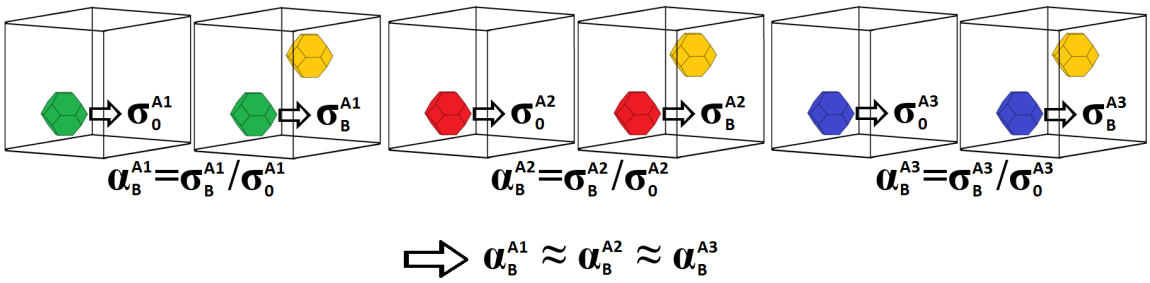


Figure 4: Illustration of the first assumption (each color corresponds to a different crystallographic orientation): the influence of grain B on grain A is independent of the orientation of grain A

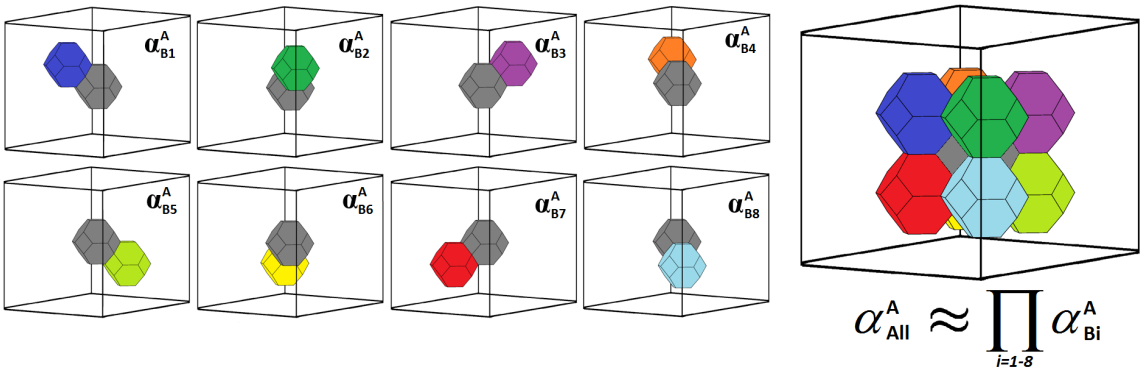


Figure 5: Illustration of the influence of each individual grain compared to the influence of several grains (each color corresponds to a different crystallographic orientation): the influence of several grains is equal to the product of the influence factor of each neighboring grain

The coefficients $a_i^{(X,Y,Z)}$ are calculated from the FEM results for each different relative position (X;Y;Z).

The calculations were run with grain A having the effective properties of the material. In order to generate the equation for any orientation of grain A, the crystal properties are attributed to grain A and calculations are run. It has been observed that the influence of the crystal orientation of grain A is negligible compared to the influence of the crystal orientation of grain B, suggesting a first assumption for the model under development: the influence of grain B on grain A is independent of the orientation of grain A (Fig.4). With that assumption made, equation 1 can be used to calculate the influence factor α_B^A of grain B on grain A for any crystal orientation of grain A.

Influence of several grains on another grain

The influence of several grains on the central grain A has been studied. Knowing the influence factor α_{Bi}^A of each neighboring grain Bi on grain A, it has been observed that the influence of several grains on grain A is equivalent to the product of the influence factor of each neighboring grain (Fig.5). A second assumption can be made: the influence of several grains is equal to the product of the influence factor of each neighboring grain (Eq.2)

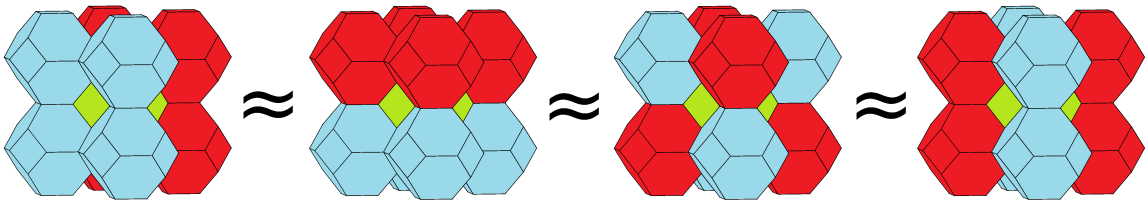


Figure 6: Illustration of the influence of several grains with the same set of influence factors but with different repartitions (the grains in blue correspond to the lowest influence factor α^A_{Bmin} and the grains in red to the highest influence factor α^A_{Bmax}): the influence of a neighboring grain on the central grain is not affected by the other neighboring grains.

$$\alpha^A_{All} = \prod_n^N \alpha^A_n(x, y, z, \varphi_1, \Phi, \varphi_2) \tag{2}$$

A consequence of the latter observation and assumption is that the distribution of the neighborhood doesn't affect the results as long as the α^A_{Bi} values do not change. In figure 6 are illustrated 4 neighboring grains with the lowest influence factor α^A_{Bmin} and 4 other neighboring grains with the highest influence factor α^A_{Bmax} distributed differently. It is observed that no matter how those grains are distributed, the stress in the central grain is not significantly affected (Fig.6). A third assumption is made : the influence of a neighboring grain on the central grain is not affected by the other neighboring grains.

Definition of the Cellular-Automaton

Based on the three assumptions declared in the previous chapter, a cellular-automaton (CA) has been developed using Self Consistent calculations to evaluate the first principal stress σ^A_0 of grain A. The solution σ^A_{SCM} of the SCM for a spherical inclusion with the elastic property of grain A (a Kelvin structure's cell can be considered as spherical) immersed alone in the matrix was used as the base for the calculation. The CA solution consists of applying to the SCM solution the influence factor of each neighboring grain that is considered to have a significant influence:

$$\sigma^A_{AC} = \sigma^A_{SCM} \times \prod_n^N \alpha^A_n(\underline{X}, [\varphi_1; \Phi; \varphi_2]) \tag{3}$$

In Eq.3, N is the number of neighboring grains considered, $\underline{X}$ is the vector form by grain A and the neighboring grain n, $[\varphi_1; \Phi; \varphi_2]$ are the Euler angles representing the orientation of the neighboring grain n, and the influence factor α^A_n of grain n on grain A is calculated with the Eq.1.

Results of CA calculation are shown in Fig.7 and 8, presenting the first principal stress in each grain obtained with the FEM and the CA. Four types of neighborhood are presented:

- 0 neighboring grain is considered. In other words, this is the SCM results.
- The first layer of neighboring grains are considered ($N = 14$ grains): all neighboring grains that have a distance from the central grain $d \leq 2r$, where r is the radius of one grain.
- The second layer of neighboring grains are considered ($N = 64$ grains): $d \leq 4r$.
- The third layer of neighboring grains are considered ($N = 258$ grains): $d \leq 6r$.

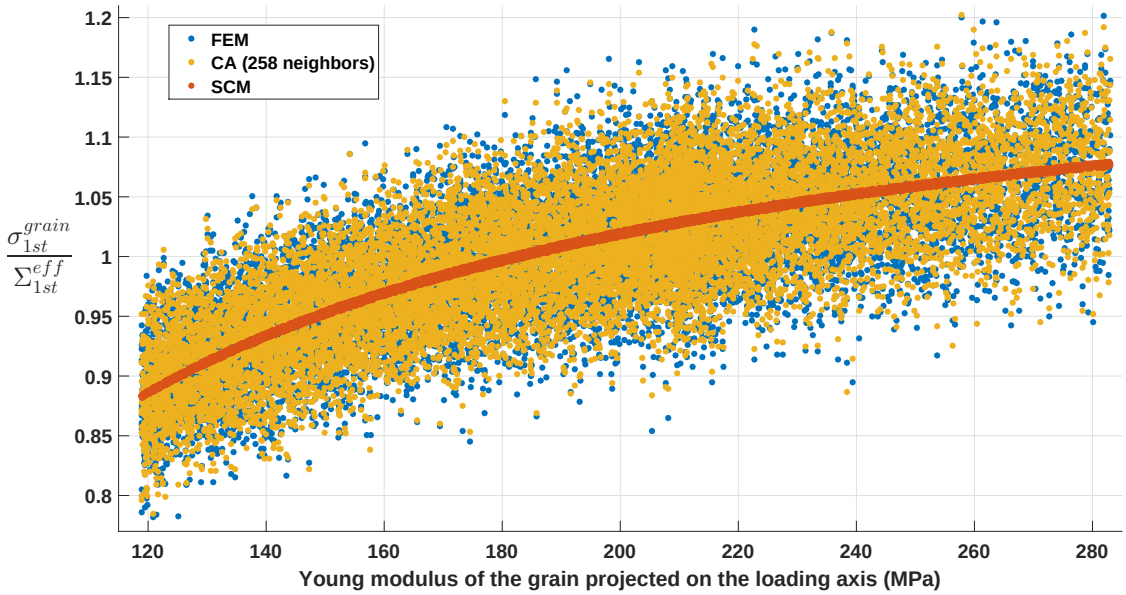


Figure 7: First principal stress in each grain of the polycrystal normalized by the macroscopic first principal stress: Comparison between the FEM, SCM, and CA

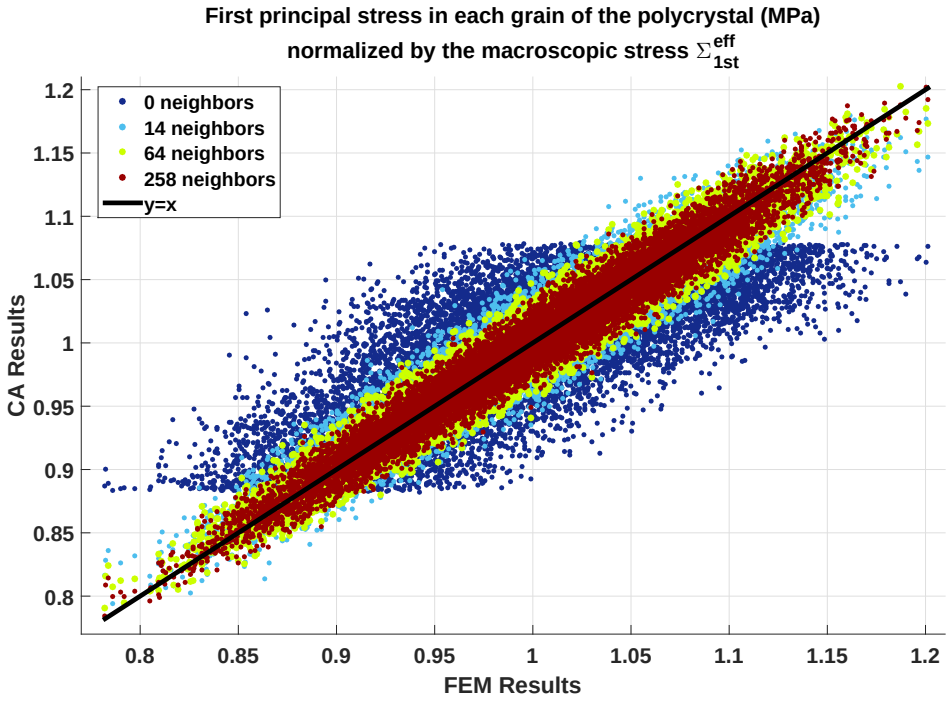


Figure 8: Comparison of the FEM results with the CA results

Table 2: Comparison of the FEM results with the CA results

	$N = 0$	$N = 14$	$N = 64$	$N = 258$
Normalized average difference $\left\langle \frac{\left\lvert \sigma_{EF}^{grain} - \sigma_{SC}^{grain} \right\rvert}{\Sigma_{EF}^{eff}} \right\rangle_{grain}$	3.42%	1.59%	1.22%	1.04%
Normalized maximum difference $max_{grain} \left(\frac{\left\lvert \sigma_{EF}^{grain} - \sigma_{SC}^{grain} \right\rvert}{\Sigma_{EF}^{eff}} \right)$	16.93%	8.80%	5.78%	5.41%

In Fig.7 are presented the FEM, SCM and CA (with 258 neighboring grains considered) results. As we can see, consideration of the neighborhood effect in the model generates a dispersion of the stress similar to the one observed for the FEM results. If we take a closer look at the local behavior on Fig.8 and Tab.2, it is observed that the more neighboring grains are considered, the closer the CA results are to the FEM ones. The accuracy of the CA with $N = 258$ is three times better than a simple SCM, and the extreme values are better captured. For information, using the same computer, the FEM simulation takes approximately 40 minutes to be completed when the CA with $N = 258$ takes 40 seconds.

Conclusions

The present work shows the importance of the neighborhood effect in polycrystal fields. From the observations of the Finite Element simulations, a cellular automaton has been developed taking into account the neighborhood effect. The model is based on the Self-Consistent model to which an influence factor is applied depending on the orientation and distribution of the neighborhood of the grain. Taking the FEM results as a reference, the model showed a significant improvement of the results compared to the original SCM.

References

[1] Forest, S., Cailletaud, G., Jeulin, D., Feyel, F., Galliet, I., Mounoury, V. and Quilici, S. (2002) Introduction au calcul de microstructures. Mecanique et Industries **3**, 439-456.

[2] Roters, F., Eisenlohr, P., Hantcherli, L., Tjahjanto, D. D., Bieler, T. R., Raabe, D. (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. Acta Materialia **58**, 1152-1211.

[3] Kroner, E. (1961) Zur plastischen verformung des vielkristalls. Acta Metallurgic **9**, 155-161.

[4] Kocks, U. F., Tom, C. N. and Wenk, H. R. (1961) Texture and Anisotropy. Preferred Orientations in Polycrystals and Their Effect on Material Properties. Cambridge University Press.

[5] Montheillet, F. and Gilormini, P. (1996) Predicting the mechanical behavior of two-phase materials with cellular automata. International Journal of Plasticity **12**(4), 561-574.

[6] Boutana, N. and Bocher, P. and Jahazi, M. (2008) Discrepancy between fatigue and dwell-fatigue behavior of near alpha titanium alloys simulated by cellular automata. International Journal of Fatigue **51**, 49-56.

[7] Pourian, Meysam H. and Pilvin, Philippe and Bridier, Florent and Bocher, Philippe (2014) Heterogeneous elastic behavior of HCP titanium polycrystalline aggregates simulated by cellular automaton and finite element. Computational Materials Science **92**, 468-475.

[8] Hesselbarth, H.W. and Gbel I.R. (1991) Simulation of recrystallization by cellular automata. Acta Metallurgica et Materialia **39**(9), 2135-2143.

[9] Ding, R., Guo, Z.X. (2002) Microstructural modelling of dynamic recrystallisation using an ex-

tended cellular automaton approach. Computational Materials Science **23**, 209-218.
[10] Solas, D., Thbault, J., Rey, C., Baudin, T. (2010) Dynamic Recrystallization Modeling during Hot Forging of a Nickel Based Superalloy. Materials Science Forum **638**, 2321-2326.
[11] Simmons, G. and Wang, H.F. (1971) Single crystal elastic constants and calculated aggregate properties: a handbook. The M.I.T. Press **2**.

Dislocation Dynamics in polycrystals with atomistic-informed mechanisms of dislocation-grain boundary interactions

***†N.B. Burbery[1], G. Po[2] , R. Das[1], N. Ghoniem[2], W. G. Ferguson[3]**

[1] Department of Mechanical Engineering, 20 Symonds St., University of Auckland, Auckland, New Zealand,
[2] Mechanical and Aerospace Engineering Department, University of California, Los Angeles 1597, USA,
[3]Department of Chemical and Materials Engineering, 20 Symonds St., University of Auckland, Auckland, New Zealand

*† Presenting and corresponding author: nbur049@aucklanduni.ac.nz

Abstract

At the mesoscale, plastic deformation is facilitated by the motion of dislocations and is strongly dependent on the local crystallographic orientation. In polycrystalline materials, the mismatch between adjacent crystals inhibits the inter-granular dislocation mobility, reduces plastic strain homogeneity and significantly influences the hardening and softening stress-strain behavior. Studies have shown that inter-granular slip transmission is possible at high stresses, involving a complex combination of dislocation absorption, junction formation and nucleation interactions with the intrinsic grain boundary dislocations. These effects are thought to contribute significantly to the behavior of dislocation pile-ups and could explain the predominant mechanisms influencing the properties of nanocrystalline materials. Modelling the mesoscale microstructure-property relationships, observed in real materials, would be very useful to guide future developments in the field of grain boundary engineering.

Dislocation dynamics (DD) simulations are a promising framework for computational modelling to provide insights about phenomena that can only be explained from the intermediate scale between atomistic and macro scales. However, a robust framework for modelling dislocation interactions with internal microstructure such as grain boundaries (GBs) has yet to be achieved for 3D models of DD at the meso-scale. Atomistic studies have shown that GBs cannot be assumed to act purely as an inertial damper between two regions with identical crystallography [1], or as an impenetrable barrier [2, 3]. The primary aim of the present study was to establish a sufficiently 'generic' framework to enable the modelling of various GB structures, polycrystal geometries and crystallographic orientations. The framework described is effective for studying GB-dislocation interactions (including inter-granular effects) and the approach for partitioning the DD simulation domain also provides an ideal future basis for modelling precipitate-hardened materials.

To achieve a robust method to differentiate between crystal regions, the present framework utilizes a mesh-based partitioning system. The simulation domain is meshed and "region IDs" are assigned to individual mesh elements. GBs are recognized as internal surfaces separating regions with different "IDs". This flexible construction allows modeling of an arbitrary number of grains and grain orientation. Within each grain, slip systems are determined by the grain orientation, and grain boundary dislocations are created to accommodate the grain misorientation. These special dislocations are either of sessile or glissile character, depending on the grain boundary structure. The glissile structure cases allow for grain boundary sliding. An algorithm was developed to re-position any dislocations which would otherwise cross the mesh-region interface to exactly intersect the GB plane. Dislocations in the GB are constrained to glide in the GB plane. Atomistically informed criteria for "slip transmission" are implemented. In particular, 'Slip transmission' was enabled by simulating dislocation nucleation in the adjacent crystal if the local Peach Koehler force on the secondary slip system exceeds the threshold value (obtained with atomistic studies).

GBs contain intrinsic dislocations (GBDs) which must be considered carefully, particularly when attempting to model inter-granular interactions with mobile lattice dislocations. A dislocation extraction algorithm was used to analyze the atomistic structure of a low angle grain boundary and identify the appropriate spacing of GBDs within the DD simulation bi-crystal model. This work provides a means to study multi-grain deformation processes governed by dislocations that "pile-up" at grain boundaries, in detail beyond feasible limits of experiments.

Keywords: Dislocation dynamics; Molecular dynamics; Slip transmission; Strain burst; Micro-pillar; Coincident-site lattice; Hall-Petch; Bi-crystal.

Introduction

Since the proposal of Taylor's theory of work hardening 1934 [4], the materials research sector has aimed to achieve a physics-based multi-scale model to non-empirically predict the non-linear (plastic) stress-strain behavior and properties of dislocation-hardened metals. Such models need to account for the dynamically evolving dislocation and grain boundary microstructure [5]. Dislocations are well-established to facilitate the bulk of irreversible crystal deformation due to their high mobility along specific crystallographic slip systems [6]. For this reason, the properties of polycrystalline materials are predicated by the orientation of the slip systems with respect to the loading direction, and by the microstructure which inhibits the dislocation mobility. Grain boundaries (GBs) are an intrinsic microstructural component of all metal (excluding single crystals) and contribute both a barrier to dislocation mobility and the transition between different slip-deformation systems [5]. GBs primarily inhibit dislocation motion; however, trans-granular 'slip transmission' can occur via a corresponding nucleation of new, re-orientated dislocations in the adjacent crystal [7]. The GB structure can facilitate dislocation nucleation, annihilation and/or recombination, which may be the rate-limiting effects in nano-crystalline materials [8-10]. For these reasons, the impact of dislocation dynamics on the non-linear stress-strain properties of polycrystalline materials can only be truly understood when interactions with the 3D network of grain boundary microstructures is accounted for. However, GBs remain significantly under-represented within the computational modelling and simulation research sector for studying defect-driven plastic deformation, below the empirical 'macro-scale' crystal plasticity simulations [11].

Dislocation dynamics (DD) simulations are widely acknowledged as a breakthrough meso-scale technique, with the capacity to establish a phenomenological link between fundamental atomistic studies and macro-scale continuum models useful for real-world material design [11-14]. However, DD remains in a development stage and has yet to be implemented in a way that can accommodate dynamic grain boundary interactions in 3D, which is necessary to understand effects of dislocation pile-ups and re-oriented slip transmission [11]. Previous attempts to model polycrystal DD with mesoscale simulations are mostly limited to 2D DD with impenetrable GBs [2, 15-17], which recently have included more complex interactions such as slip transmission through the GB interface [18]. These studies offer valuable insights about the effect of grain boundaries on the unimpeded motion along singular slip systems. However, 2D methods are incapable of modelling the evolution of dislocation density because dislocations are 'pseudo point defects'. Furthermore, the 2D systems are artificially constrained to only 1, 2 or (at best) 3 slip systems [16]. It is unlikely that such models will ever be capable of effectively capturing the complexity of cross-slip, multi-junction formation or more complex long-range dislocation force-field effects. In terms of 3D DD, rudimentary models have been created to evaluate the stress-fields in 'bi-crystals' containing of an array of impenetrable dislocations, akin to a low-angle GB [15]. However, this model did not account for changing crystallography at the interface, and no algorithms were provided to enable dislocation intersection with the GB interface. Hence, this dislocation array study is a good first step but does not provide a realistic representation of a GB interface. A more sophisticated model was established by Kubin et al. in 2009 [2], involving a truly polycrystalline, multi-textured simulation. However, the GBs were modelled with as impenetrable interfaces and the model was incapable of compensating for dislocation interactions with the intrinsic GB dislocations or reproducing inter-granular slip transmission. The present study establishes a 3D DD methodology which is robust for modelling multiple GB character and polycrystal geometries, and applies this for a rudimentary study of a bi-crystal with a 'penetrable GB',

The equilibrium atomistic structure of the GB core and the spacing and Burgers vectors of the intrinsic GB dislocations (GBDs) are entirely dependent on the misorientation angle and interfacial plane of the GB intersecting two adjacent crystals. Low angle GBs can be fully described as an array of 'grain boundary dislocations' (GBDs), and have been observed to occur with misorientation angles less than the 'transition angle' which is approximately between 10-15° [19]. The dislocation structure of higher angle GBs are generally more difficult to classify, however it is commonly believed that in this case, the GB core consists of overlapping dislocations. These are difficult to classify as dislocations, because the overlapped cores cannot be identified by forming a Burgers circuit according to the conventional methodology. Energetically favorable structures of GBs involve a repeated 'structural unit' of equi-spaced clusters of GBDs [20-22]. In situations with high local stress concentration such as near nanoindenters [23] and inside dislocation pile-ups [24], mobile lattice dislocations can 'penetrate' through the GB by interacting with the GBDs. Specifically, lattice dislocations can indirectly 'transmit' across the GB by forming junctions with GBDs, partially annihilating and re-nucleating a new dislocation with different orientation in the adjacent crystal. To establish an initial benchmark for the newly developed simulation methodology, the first case will involve a bi-crystal containing two low angle GBs, which were chosen because of the low GBD density. The bi-crystal was selected as the most simple benchmark geometry for comparison with MD simulations, and to isolate the influence of the GBDs on the mechanical properties [25].

The present study describes a novel modification of 3D DD simulation method, utilizing an array of co-planar intrinsic dislocations to model GB - dislocation interactions at the meso-scale. This will enable future studies of the intrinsically mesoscale effects of dislocation pile-ups and size-strength (Hall-Petch) relationships.

Framework of conventional mesoscale dislocation dynamics simulations

This study utilizes the Mechanics of Defects Evolution Library (MoDEL) code, based on the parametric DD approach described by Ghoniem et al [26] and recently modified to improve the description of the dislocation core by Po et al. [27, 28]. The 'parametric' DD approach is ideal for 3D modelling of multi-defect dynamics to achieve efficient modelling of curved dislocations of arbitrary shape, orientation and length. Although DD remains a 'state-of-the-art' method due to the nature of its ongoing development [13], there is a long history of development since the 1990's [13, 27-33] [34]. At its core, the procedure of evaluating the Peach-Koehler force interactions, discretizing the motion, network configuration and shape is well-established [13, 35]. The present study does not go into detail about the fundamental framework (refer to [13, 28]), but rather describes the novel implementation of polycrystalline effects and GB-dislocation interactions within the established 3D DD framework. However, first it is necessary to describe the elements of the present framework that are modified and which enable the description of grain boundaries in a constitutive linear elastic framework.

In this implementation, DD simulations are coded with object-oriented C++ programming to model the discretized motion of dislocation loops and Frank-Read sources [13]. In its most fundamental form, DD is a meshless-continuum method with 'infinite' dimensions; however a mesh can be utilized for the implementation of fixed surface boundary conditions. This contributes only a surface effect; and retains the single crystal orientation and isotropic elastic properties of the medium without simulation sub-domains. This 'conventional framework' for DD simulations can be decomposed into the following four fundamental elements:

a) Dislocation nodes (1D)

Nodes store discrete positions in the dislocation line at each timestep, within the elastic continuum. *Each node is characterised by a specific ID, mesh tetrahedra and nodal velocity.*

b) Dislocation segments (2D)

Segments are mathematical splines that connect adjacent dislocation nodes in a dislocation loop. The curvature of the spline corresponds with the localised Peach-Koehler force-field at the specific timestep. As such, the discretised positions of dislocation segments are not stored between timesteps as in the case of dislocation nodes. *Segments are defined by the Burgers vector, glide plane normal vector, Peach-Koehler forces and external stress tensor.*

c) Dislocation network (3D)

The network is a container of all the dislocation segments in a 3D ensemble of dislocation loops and dislocation sources. *The network defines the self-interactions of dislocation segments within a single loop and interactions between different dislocations, and asserts the consistency of elastic criteria, such as the Burgers vectors and node-balance.*

d) Finite element mesh (optional – required for certain boundary conditions)

A mesh is not necessary, however must be used to model finite volumes and surface effects. Surface forces are implemented by creating artificial image forces, according to the original description provided by Van der Giessen et al [31]. *The mesh is defined by mesh tetrahedra defined by four positional points (nodes) and four triangular faces. Each mesh tetrahedra, face and node is assigned a unique ID number.*

Computational procedure for modelling polycrystal sub-regions in DD

The distinctive element of the present approach for modelling DD is the concept of ‘region IDs’, which can be assigned to all mesh tetrahedra within a user-specified geometry. Hence, all the mesh tetrahedra within the mesh region geometry (crystal) share the same region ID. The mesh is faceted with faces defined by any three of the four mesh nodes in each of the mesh tetrahedra. Each facet must always share the region IDs of the two adjoining tetrahedra or be a surface with only one region ID, and hence mesh facets must either have one or two region IDs. For the present case with a bi-crystal containing only one GB, this is sufficient to describe the interface. However, the method is also capable of modelling GB junctions of three or more crystals by identifying points lying on a mesh-line adjoining tetrahedra with more than two region IDs.

The mesh is independent of the dynamic behavior of the simulations, and hence within the current framework the region ID is an immutable component of the initial-state crystal geometry. While this inhibits the implementation of GB migration within the current framework, it is valuable to assure efficiency and avoid arbitrary distortion of the interfacial mesh. Hence, this original approach to defining the GB structure provides a robust, efficient and ‘generic’ basis for modelling polycrystals of complexity within a DD simulation.

To establish a polycrystal mesh, a template MATLAB script was developed [36] which could be modified to define the size of the mesh, interface orientation, and crystal region IDs for either a rectangular prism or a cylinder bi-crystal geometry. The mesh itself was generated with tetgen, using a Delauney tetrahedralization constrained by maximum tetrahedron volume to control the coarseness of the mesh [37]. All the mesh tetrahedra within the one of the sub-domains defined by the matlab script are assigned the same integer (region ID), that is unique to the crystal. It was necessary to ensure that dislocation nodes that intersect the mesh faces shared by two region IDs are coincidental with the GB interface. This was achieved by identifying any tetrahedra nodes that were within a nominal floating point distance tolerance interface plane, and modifying the positions of two adjacent mesh nodes so that the mesh faces were correctly aligned. Hence, the mesh-elements of the GB were defined so that any nodes incidental with a face sharing two region IDs would align correctly with both the GB plane and the internal crystallographic lattice.

Utilizing a dislocation array based on atomistic calculations to model GB structure

GBs may be described as a crystallographic structure of repeated atomistic structural units containing intrinsic GB dislocations (GBDs). However, characterization of the GBDs in high-angle GBs (misorientation > 15°) has been difficult to achieve due to the overlapped nature of the dislocation cores within the plane [20, 38]. Low-angle GBs are more readily modelled, due to the greater spacing between GBDs and subsequently greater ease for classifying the distinct atomistic dislocation cores [15, 39]. Three pure-tilt grain boundaries were simulated in full-atomistic from, using bi-crystals obtained with LAMMPs molecular dynamics simulations [40]. The GBD structure of the fully atomistic GB plane were analyzed using Stukowski's dislocation extraction algorithm [41]. The results are shown in Figure 1. The dislocation line-direction is parallel to the tilt axis, which is the same [0 0 1] direction for all three GB structures (as shown in Figure 1.A).

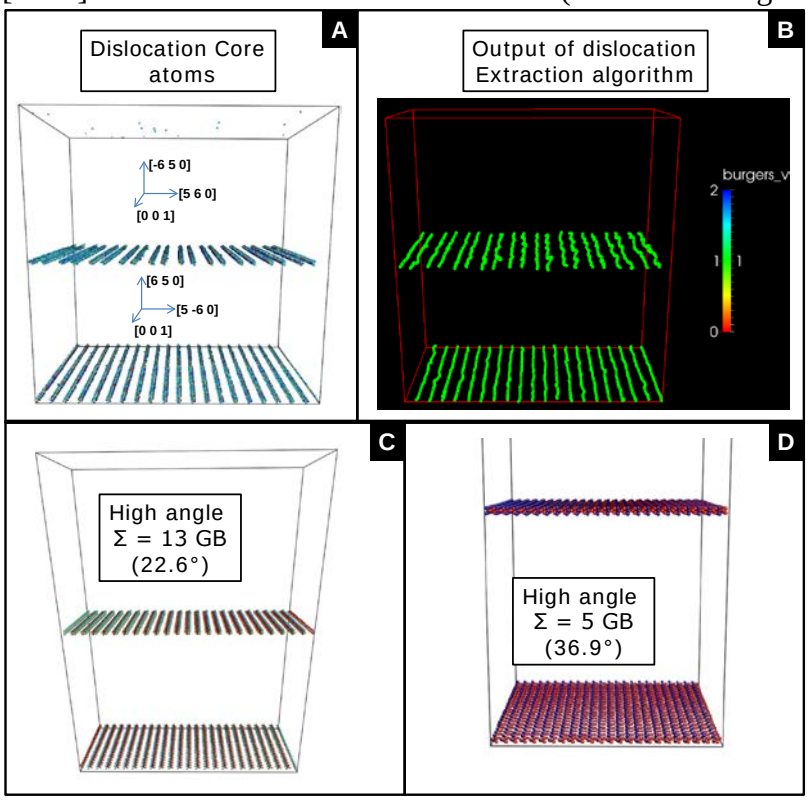


Figure 1: Structure of high and low angle GBs described in two formats. Atomistic structural GB units [20]: A) low angle (8.1°); C) high angle (22.6°); D) high angle (36.9). AND B) low angle GB - array of GBDs

Figure 1.b shows that the dislocation extraction algorithm effectively identifies intrinsic dislocations spaced at intervals equivalent to the atomistic structural GB units, only for the low angle GB case. However, the direct comparison of the atomistic structures of the different GBs provides an invaluable insight for modelling with some of the high angle GBs. This is because the spacing of atomistic structural units can be evaluated despite being unable to extract the dislocation content. For the current crystallographic misorientation, the inter-GBD spacing is 15.7 Å, (i.e., 6b, where b is the Burgers vector). The GBDs can be considered 'perfect' edge dislocations with full Burgers vectors aligned in the direction of the GB normal (the [6 5 0] or the [-6 5 0] directions). This is consistent with the symmetric pure-tilt 'parallel-edge wall' GBs described in ref. [42]. It is noteworthy that the 'nose-to-tail' spacing of the 'C' atomistic structural units (which are also described in detail in ref. [21]) is 6.1 Å for the Σ=13 case. Furthermore, for the case of the Σ=5 GB, which has a higher misorientation angle, the nose-to-tail spacing is 0.0 Å (i.e., there is no inter-GBD gap) between qualitatively identical 'C' atomistic structural units. This suggests that high angle GBs can be modelled in a similar manner as low-angle GBs, however with a reducing spacing between GBDs. The validity of this claim is the subject of future studies.

For the present study, the structure was based on the low-angle atomistic case to provide a first-case benchmark for comparison. As shown in Figure 2, intrinsic GBDs were assigned in DD simulations of a bi-crystal containing a planar GB with a normal vector in the vertical direction, with an equivalent spacing and Burgers vector character as obtained from atomistic analysis.

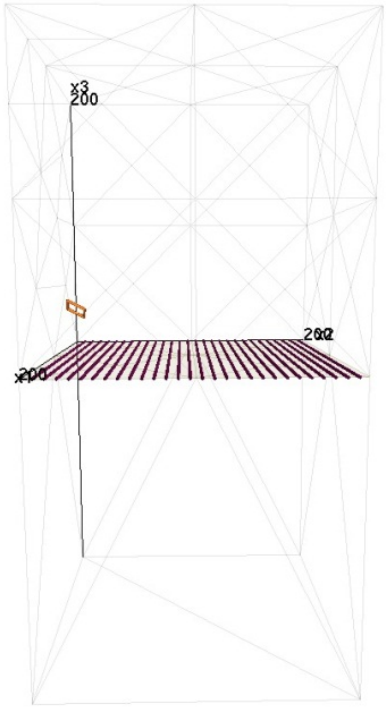


Figure 2: DD simulation of bi-crystal containing interfacial GBDs with identical spacing as an 8.2° GB

The description of GBs as dislocation arrays in DD is ideal for modelling dislocation interactions such as annihilation, junction formation (i.e., absorption and recombination of lattice dislocations); and nucleation that results in slip transmission. The details of modelling dislocation – GB interactions are primarily accommodated by well-established junction formation procedures within the conventional DD framework [13]. However, the computational procedure to obtain 100% confidence that the dislocations would not arbitrarily cross the interface involves extensive modifications to the simulation code. It, it was also necessary to establish a new simulation procedure to enable dislocation nucleation and, hence, slip transmission.

Algorithms to simulate dislocation intersection and absorption into the GB plane

The procedure to inhibit the arbitrary crossing of dislocations across the GB interface involves a necessarily complex and intensive procedure. It was necessary to ensure that a robust system of checks was utilized for all dislocation nodes, including pre-existing mobile nodes, nodes from network re-distribution (annihilation and generation of new nodes) and 'special nodes' in dislocation junctions external surface ledges.

To enforce that dislocations do not artificially cross the GB interface, an algorithm was developed to check the region ID of the mesh position that the node is projected to move towards along its current trajectory at each timestep. When the projected position would lie within a mesh tetrahedron that is assigned a different region ID than the current position's region ID, a complex procedure was triggered to implement dislocation-GB intersection. It is noteworthy that a system of projected positional checking was already a necessary component of fixed-boundary simulation with finite volume. However, the currently described procedure for GB intersection is more complex due to the formation of GBD junctions and ongoing mesh re-distribution of internal (not boundary) segments.

To model intersection with the GB, it was necessary to first identify the interfacial mesh facet shared by tetrahedrons with two different region IDs along the trajectory of the dislocation node. This is achieved by looping over the faces of all tetrahedra in the trajectory between the initial nodal point to the final position. Once the tetrahedron on the path within the original region ID is found that contains this interfacial facet, the dislocation node is then placed at the point along the original dislocation trajectory that intersects this mesh-face.

Once a dislocation node is positioned at the point of intersection on a facet connecting two region IDs, the node is designated as a GB node by assigning a fixed unit vector that defines the GB normal direction (V_{normal}). V_{normal} is thereafter used to eliminate the component of the nodal velocity projected in the direction normal to the GB plane, so that the motion of GB nodes is accurately constrained within the GB plane. In addition to being constrained to glide within the GB plane, GB nodes are also constrained to the original glide plane or by the constraints of a dislocation junction with any intersecting GBDs.

Slip transmission and dislocation nucleation from GBs

Atomistic studies have demonstrated that dislocations rarely penetrate GBs directly at the original point of intersection, but rather that the localized stress concentration activates dislocation 'nucleation' from an adjacent GB lattice site on a new slip system. In the present framework, 'nucleation' involves a recombination reaction between the interfacial GBDs and the trapped lattice dislocation lying on the interfacial 'displacement complete shift' lattice.

Nucleation is only initiated after a check of the resolved Peach-Koehler forces of all segments containing two dislocation nodes within the GB interface is identified to exceed a threshold value. This involves looping over all aforementioned segments, checking the available slip systems of the secondary region ID and computing the forces for all of the 12 FCC slip directions.

When nucleation is triggered, 'slip transmission' is implemented by generating a new dislocation loop comprised of 3 nodes lying in the second crystal and 2 nodes on the GB interface. The GB nodes are placed equidistantly from the midpoint of the lattice dislocation segment lying within the GB, however rotated appropriately so that the dislocation loop is normal to the new glide plane. Subsequently, recombination occurs between the lattice dislocation and the nucleated segment in the GB plane. This segment remains constrained inside the GB normal plane, however the newly

nucleated lattice dislocation nodes in the second crystal are free to move along the new slip system. Unfortunately, because the threshold nucleation stress and/or Peach-Koehler force is strongly dependent on the localized GB structure and/or presence of defects such as GB ledges, there is uncertainty about the most-appropriate nucleation thresholds [25]. Probabilistic modelling may provide an ideal work-around for this limitation in the future, however will need to be tailored to the specific misorientation angle (particularly for high-angle vs low-angle GBs). To avoid unnecessary complexity for low-stress interactions without dislocation pile-ups, nucleation and transmission may be enabled or disabled very easily by modifying one line of code.

Stability testing and robustness of mesh-region barrier

The most-critical requirement of the current computational approach is the assertion that no elements within the dislocation network will ever arbitrarily cross the interface between different mesh regions. If this were to occur at any point in the simulation procedure, the mesh region ID check would fail to correctly identify an intersection event at the subsequent node – motion step. After extensive stability testing, a few challenges were identified and subsequently were resolved to obtain a robust framework which ensures that the GB is never 'arbitrarily penetrated'.

The first challenge observed for artificial interface-crossing, involved an inherent numerical (rounding) error that occurred after dislocation nodes were made to intersect the GB plane. In this case, nodes moving within the GB plane sometimes were incorrectly identified with only one region ID due to truncation and rounding errors, so the node was no longer on a shared mesh face (GB). No solution was identified to completely eliminate this effect, however fortunately this issue has been overcome entirely by asserting that nodes intersecting the GB are constrained to the plane defined by the GB normal vector, without requiring a check for the mesh region ID.

The second issue identified provided a more fundamental challenge for the modelling, caused when lattice dislocations formed a junction that intersected with the GB plane. In this case, in order to assert crystallographic consistency it was necessary that dislocation nodes be placed at the exact point intersecting the glide planes of the lattice dislocations and the GB plane. This was further complicated, when junctions also were formed with GBDs, requiring a point of intersection between four independent planes (mathematically improbable). This meant that in certain cases there was no existing direct solution which effectively merged the dislocation segments into a junction at a point which was also coincidental with the GB plane. The conventional junction-formation protocol assertions would result in a junction that artificially crossed into the second crystal. A temporary fix has been implemented, which disables dislocation junction formation if the projected intersection position that is coincidental with the three (or more) constraining planes does not remain in the original region ID. It is a future aim to establish a new approach to more rigorously model junction formation at the GB by performing sequential junction formation and re-alignment. Due to the crystallographic constraints, it is likely that this will involve a complex procedure of annihilation, nucleation and recombination to maintain the conservation of Burgers vector and glide-plane constraints that are an intrinsic property of dislocation dynamics [43].

Careful checking and testing with a variety of stress conditions, dislocation densities and geometries of the fixed micro-pillar boundaries has demonstrated that the present version of the code is empirically an 'inherently stable' simulation framework. This has very positive implications for 3D DD modelling in the future with dislocation pile-up formations in polycrystals and for modelling the accumulation of very high internal dislocation densities. Examples are provided in Figure 3.A and in Figure 3.B. to demonstrate the efficacy of the presently described method to model arbitrarily complex systems and dislocation pile-ups. It is also noteworthy that this method is also

exceptionally well-suited for simulating defects or precipitate hardened alloys that involve impenetrable inclusions.

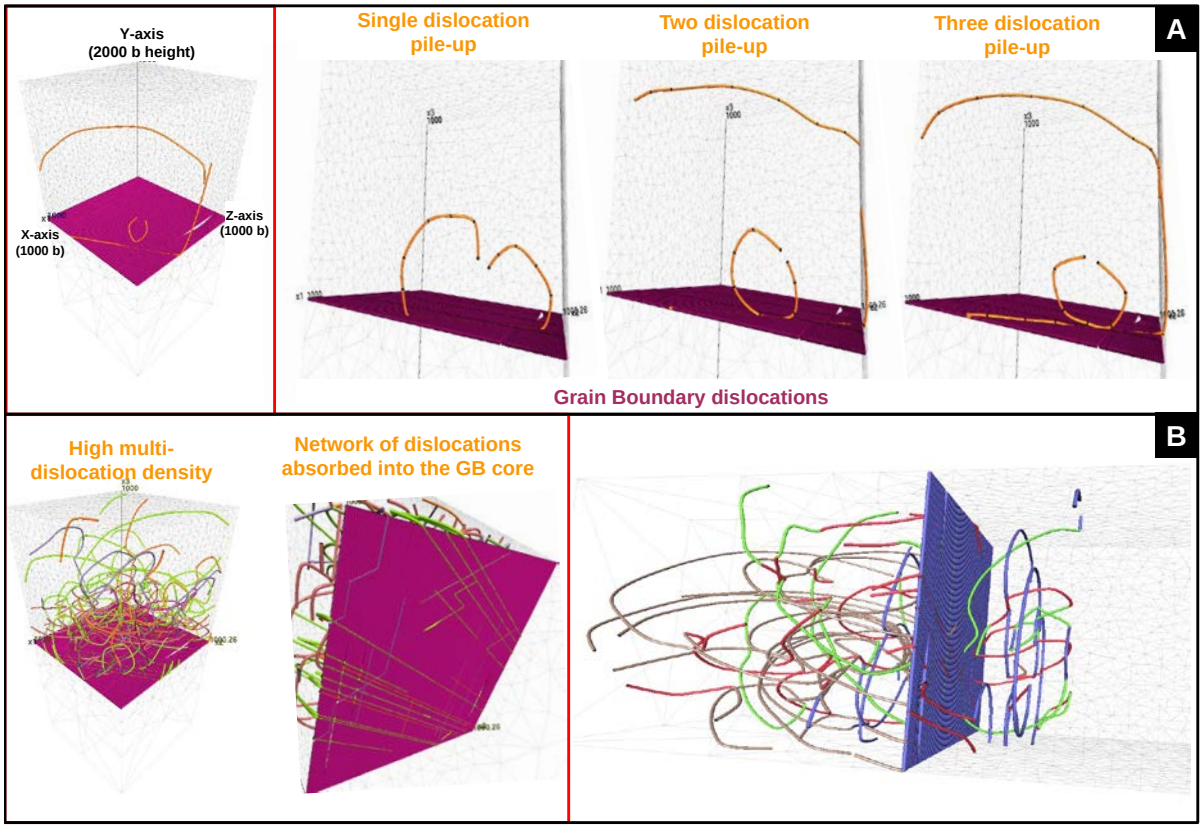


Figure 3: Robust modelling of 'impenetrable' mesh-region interface, up to very high dislocation densities: a) Examples of dislocation pile-up formations at the GB under a singular slip system; b) Examples of multi-junction formation and high dislocation density accumulated at the GB interface.

Dislocation nucleation and slip transmission through the GB

One of the novel elements of the present framework is the capability to model inter-granular plastic deformation, which occurs by slip transmission into the secondary crystal. This has been achieved at both a rudimentary level in terms of a singular set of crystal slip systems, and has also been recently applied to model nucleation along a user-specified selection of secondary crystal slip systems. The definition of crystal-specific slip systems remains in a state of development; however the present study demonstrates that the framework has the capacity in the future. An example of nucleation from a dislocation pile-up located at the GB is demonstrated in Figure 4.

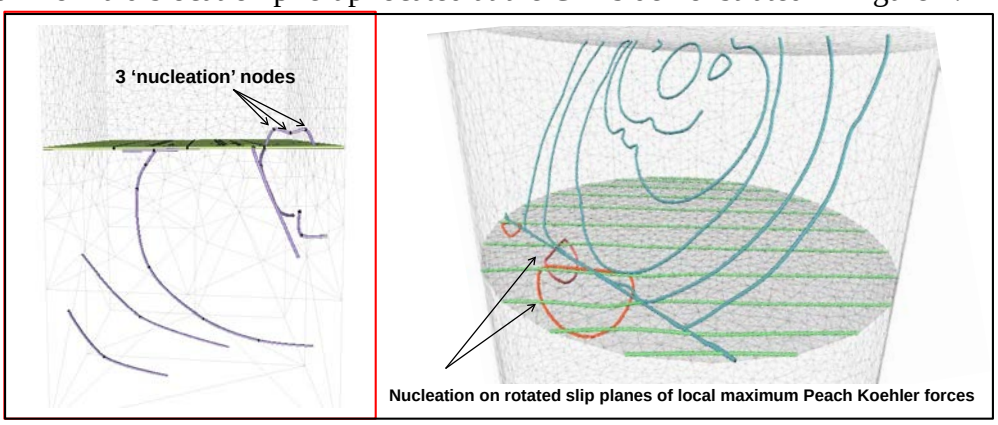


Figure 4: Demonstration of dislocation nucleation. A) Example of nucleation procedure by generating three 'nucleation nodes' from high-density GBDs; B) nucleation on rotated slip systems with low GBD density

Conclusions

This paper has described the implementation of a novel approach to achieve mesoscale dislocation dynamics simulations in polycrystalline materials. The method utilizes a modified mesh, in order to assign a unique 'region ID' to dislocations contained within different crystals. A series of algorithms have been developed to provide a modelling framework that ensures that dislocations do not arbitrarily cross the mesh region-interface. The code also asserts that dislocations which would otherwise cross the grain boundary (GB) interface will instead will exactly intersect the GB plane. An additional modification that remains in a development stage enables slip transmission by initiating dislocation nucleation from the GB into the secondary crystal, which is initiated when the maximum local Peach-Koehler force exceeds the threshold value.

Molecular dynamics simulations coupled with a post-processing method to extract the dislocation content were used to determine the atomistic structure of a low angle GB, and explicitly convert this into a dislocation format (i.e., a planar array of GBDs). On the basis of the atomistic analysis of this interface, replica GB structures were modelled with the modified DD, using the MoDEL library. It has been demonstrated that the code is inherently stable, and will not allow for slip transmission across the mesh-region interfaces unless dislocation nucleation is triggered. This has been used to demonstrate stress concentration within a dislocation pile-up, dislocation absorption into the GB core and the accumulation of high local dislocation density adjacent to the GB. In addition, the complex algorithms used to model slip transmission via the nucleation of dislocations along secondary crystallographic slip systems has been demonstrated to be an effective approach. The future opportunities to discretely evaluate the junction formation; annihilation; recombination and nucleation dislocation reactions between lattice dislocations and GBs can be used to provide significant insights into the defect mechanics of trans-granular plastic deformation.

References

[1] Zhou, C. and R. LeSar, *Dislocation dynamics simulations of plasticity in polycrystalline thin films.* International Journal of Plasticity, 2012. **30–31**: p. 185-201.

[2] Kubin, L.P., B. Devincre, and C. de Sansal, *Grain size strengthening in microcrystalline copper: a three-dimensional dislocation dynamics simulation.* Key Engineering Materials, 2010. **423**: p. 25-32.

[3] Espinosa, H.D., et al., *Discrete dislocation dynamics simulations to interpret plasticity size and surface effects in freestanding FCC thin films.* International Journal of Plasticity, 2006. **22**(11): p. 2091-2117.

[4] Taylor, G.I., *The mechanism of plastic deformation of crystals. Part I. Theoretical.* Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1934: p. 362-387.

[5] Paterson, M., *Deformation Mechanisms: Crystal Plasticity*, in *Materials Science for Structural Geology*. 2013, Springer Netherlands. p. 107-207.

[6] Ghoniem, N.M., et al., *Multiscale modelling of nanomechanics and micromechanics: An overview.* Philosophical Magazine, 2003. **83**(31-34): p. 3475-3528.

[7] Kumar, K.S., H. Van Swygenhoven, and S. Suresh, *Mechanical behavior of nanocrystalline metals and alloys.* Acta Materialia, 2003. **51**(19): p. 5743-5774.

[8] Hans, C., *Plastic deformation kinetics in nanocrystalline FCC metals based on the pile-up of dislocations.* Nanotechnology, 2007. **18**(32): p. 325701.

[9] Burbery N., Das R., and F. WG., *Establishing effective criteria to link atomic and macro-scale simulations of dislocation nucleation in FCC metals*, in *The 6th International Conference on Computational Methods (ICCM2015)*, D. R., Editor 2015, International Journal of Computational Methods (IJCM): Auckland.

[10] Burbery N., Das R., and F. WG., *Molecular dynamics study of the inter-relationships between grain boundary structure and the thermal, mechanical and energy properties.* Acta Materiala, 2015. **A-15-1315R1**: p. A-15-1317.

[11] McDowell, D.L., *A perspective on trends in multiscale plasticity.* International Journal of Plasticity, 2010. **26**(9): p. 1280-1309.

[12] Zbib, H.M. and T. Diaz de la Rubia, *A multiscale model of plasticity.* International Journal of Plasticity, 2002. **18**(9): p. 1133-1163.

[13] Po, G., et al., *Recent Progress in Discrete Dislocation Dynamics and Its Applications to Micro Plasticity.* JOM, 2014. **66**(10): p. 2108-2120.

[14] McDowell, D.L., *Materials design: a useful research focus for inelastic behavior of structural metals.* Theoretical and Applied Fracture Mechanics, 2001. **37**(1–3): p. 245-259.

[15] Schouwenaars, R., M. Seefeldt, and P. Van Houtte, *The stress field of an array of parallel dislocation pile-ups: Implications for grain boundary hardening and excess dislocation distributions.* Acta Materialia, 2010. **58**(13): p. 4344-4353.

[16] Balint, D.S., et al., *Discrete dislocation plasticity analysis of the grain size dependence of the flow strength of polycrystals.* International Journal of Plasticity, 2008. **24**(12): p. 2149-2172.

[17] Nicola, L., et al., *Plastic deformation of freestanding thin films: Experiments and modeling.* Journal of the Mechanics and Physics of Solids, 2006. **54**(10): p. 2089-2110.

[18] Li, Z., et al., *Strengthening mechanism in micro-polycrystals with penetrable grain boundaries by discrete dislocation dynamics simulation and Hall–Petch effect.* Computational Materials Science, 2009. **46**(4): p. 1124-1134.

[19] Winning, M. and A.D. Rollett, *Transition between low and high angle grain boundaries.* Acta Materialia, 2005. **53**(10): p. 2901-2907.

[20] Vitek, V., et al., *Grain Boundary Structure and Kinetics.* ASM, Metals Park, Ohio, 1980: p. 115.

[21] Tschopp, M.A. and D.L. McDowell, *Asymmetric tilt grain boundary structure and energy in copper and aluminium.* Philosophical Magazine, 2007. **87**(25): p. 3871-3892.

[22] Lejček, P. and S. Hofmann, *Thermodynamics and structural aspects of grain boundary segregation.* Critical Reviews in Solid State and Materials Sciences, 1995. **20**(1): p. 1-85.

[23] Soer, W.A., K.E. Aifantis, and J.T.M. De Hosson, *Incipient plasticity during nanoindentation at grain boundaries in body-centered cubic metals.* Acta Materialia, 2005. **53**(17): p. 4665-4676.

[24] Mompiou, F., et al., *Inter- and intragranular plasticity mechanisms in ultrafine-grained Al thin films: An in situ TEM study.* Acta Materialia, 2013. **61**(1): p. 205-216.

[25] Burbery, N.J., R. Das, and W.G. Ferguson, *Modelling with variable atomic structure: Dislocation nucleation from symmetric tilt grain boundaries in aluminium.* Computational Materials Science, 2015. **101**(0): p. 16-28.

[26] Ghoniem, N.M., S.H. Tong, and L.Z. Sun, *Parametric dislocation dynamics: A thermodynamics-based approach to investigations of mesoscopic plastic deformation.* Physical Review B, 2000. **61**(2): p. 913-927.

[27] Po, G. and N. Ghoniem, *A variational formulation of constrained dislocation dynamics coupled with heat and vacancy diffusion.* Journal of the Mechanics and Physics of Solids, 2014. **66**(0): p. 103-116.

[28] Po, G., et al., *Singularity-free dislocation dynamics with strain gradient elasticity.* Journal of the Mechanics and Physics of Solids, 2014. **68**(0): p. 161-178.
[29] Amodeo, R.J. and N.M. Ghoniem, *Dislocation dynamics. I. A proposed methodology for deformation micromechanics.* Physical Review B, 1990. **41**(10): p. 6958-6967.
[30] Canova, G., et al., *3d Simulation of dislocation motion on a lattice: application to the yield surface of single crystals.* Solid State Phenomena, 1993. **35**: p. 101-106.
[31] Van Der Giessen, E. and A. Needleman, *Discrete dislocation plasticity: A simple planar model.* Modelling and Simulation in Materials Science and Engineering, 1995. **3**(5): p. 689-735.
[32] Zbib, H.M., et al., *3D dislocation dynamics: stress–strain behavior and hardening mechanisms in fcc and bcc metals.* Journal of Nuclear Materials, 2000. **276**(1–3): p. 154-165.
[33] Zbib, H., *Advances in discrete dislocations dynamics and multiscale modeling.* Journal of Engineering Materials and Technology, 2009. **131**: p. 041209-1.
[34] Beneš, M., et al., *A parametric simulation method for discrete dislocation dynamics.* The European Physical Journal Special Topics, 2009. **177**(1): p. 177-191.
[35] Zbib, H., *Introduction to Discrete Dislocation Dynamics*, in *Generalized Continua and Dislocation Theory*, C. Sansour and S. Skatulla, Editors. 2012, Springer Vienna. p. 289-317.
[36] *MATLAB version 8.0*, 2012, The MathWorks Inc.: Natick, Massachusetts. p. (computer software).
[37] Si, H. *TetGen: A Quality Tetrahedral Mesh Generator and Three-Dimensional Delaunay Triangulator*. 2016; Available from: http://tetgen.berlios.de/.
[38] Gleiter, H., *The nature of dislocations in high-angle grain boundaries.* Philosophical Magazine, 1977. **36**(5): p. 1109-1120.
[39] Lim, A.T., et al., *Stress-driven migration of simple low-angle mixed grain boundaries.* Acta Materialia, 2012. **60**(3): p. 1395-1407.
[40] Plimpton, S.J., *Fast Parallel Algorithms for Short-Range Molecular Dynamics.* Journal of Computational Physics, 1995. **117**(Refer to: http://lammps.sandia.gov): p. 1-19.
[41] Stukowski, A., *Structure identification methods for atomistic simulations of crystalline materials.* Modelling and Simulation in Materials Science and Engineering, 2012. **20**(4): p. 045021.
[42] Hirth, J., R. Pond, and J. Lothe, *Spacing defects and disconnections in grain boundaries.* Acta Materialia, 2007. **55**(16): p. 5428-5437.
[43] Hirth, J. and J. Lothe, *Theory of Dislocations*. 1982: John Wiley \& Sons.

Design of porous phononic crystals with combined band gaps

***Y.F. Li[1], †X.Huang[1], and S. Zhou[1]**

[1]Centre for Innovative Structures and Materials, School of Engineering, RMIT University, Australia

*Presenting author: s3495356@student.rmit.edu.au
†Corresponding author: huang.xiaodong@rmit.edu.au

Abstract

Phononic crystals are periodic structures known for their abilities to alter the propagation of acoustic or elastic waves, and their characteristics are greatly dependent on the topological configurations of constituent materials within the unit cell. Thus it is possible to engineer a phononic crystal for specific functionality by tailoring its topology. Low manufacturing cost as well as light weight gives porous phononic crystals advantages over other kinds of phononic structures. This paper presented a bi-directional structural optimization (BESO) method in conjunction with homogenization theory for the systematic design of porous phononic crystals. On account of sustaining static loads, a bulk or shear modulus constraint is considered in the design of porous phononic structures. A multi-objective optimization was conducted to simultaneously maximize combined band gap width and bulk or shear modulus with a prescribed volume fraction of consisting solid material. The methodology was briefly introduced and several optimized porous phononic structures with exceptionally large band gaps were presented.

Keywords: Porous phononic crystals, Band gap, BESO, Homogenization, Multi-objective optimization

Introduction

Phononic crystals (PnCs) artificially designed to control the propagation of acoustic and elastic waves are periodic structures consisting of different materials usually with high contrast in their mechanical properties [1]-[5]. The most fundamental feature of PnCs is the existence of band gaps, the frequency ranges within which the propagation of mechanical waves is strictly forbidden. This special property gives rise to mangy applications such as noise and vibration control as well as wave filtering and waveguides, *etc.* [6][7]. Over the past two decades, several classes of PnCs differing in the physical nature of the constituent phases have been studied, including solid/solid, solid/fluid, solid/void, fluid/fluid systems, *etc*. [8].Among them, porous phononic crystals with void or air holes embedded in solid matrix have exceptional advantages over other systems, for they can be very light-weighted while easily fabricated with low manufacturing cost. They hold a promising prospect for applications in noise and vibration control of aircraft, automobile and other industries that have restricted control over weight.

Porous PnCs can be easily engineered by adjusting the spatial distributions of air/vacuum holes in a solid substrate. Initially in analog to studies on composite PnCs, positions, shapes, sizes of air/vacuum holes, as well as the layouts of the unit cell have been carefully investigated to disclose their relations with the phononic band gaps [9][10]. It is apparent that such trial-and-error methods are incapable to get the optimal designs when compared with more systematical means such as topology optimization. Systematic design of phononic band gap crystals was first conducted by Sigmund and Jensen based on finite element method (FEM) in combination with a gradient-based optimization algorithm [11]. Later, genetic algorithm (GA) and another gradient-based topology optimization, in conjunction with FEM or the fast plane wave expansion method (FPWE), are developed to maximize the band gap sizes of phononic band gap crystals [12]-[16]. Most of these works focused on the topological design of the composite PnCs while the porous PnCs are less considered [17]. Previous research on composite and porous PnCs has revealed that the stiffer and heavier material

tends to be isolated by the soft and light counterpart in order to get an optimal band gap size. Such characteristic leads to a tricky problem in the optimization of porous PnCs. Since the transverse/ shear waves are not supported in the air, the discontinuous solid materials would only support the propagation of longitudinal waves, which reduces the problem to sonic crystals. However the initial intention is to find the optimal porous phononic structures that exhibit large band gaps for elastic waves. Therefore, it is necessary to make sure that the optimized phononic band gap structures have continuous distribution of solid material to support all components of elastic waves. Dong et al. conducted a multi-objective optimization of 2D porous PnCs for maximizing band gap width and minimizing mass of structure simultaneously by using non-dominated sorting-based genetic algorithm II (NSGA-II) [18]. In this work, an artificial geometrical constraint was adapted to avoid too narrow connections and guarantee the resulting structure is self-support. It is apparent that current research in this area is insufficient and further systematic investigation into the design of cellular phononic band gap crystals is necessary.

Considering the porous PnCs might sustain certain amount of static loads, it is more meaningful to add extra stiffness constraint than simply adding a geometrical constraint to the optimized structures. To the authors' best knowledge, no work has been reported yet to conduct band gap optimization on the porous phononic crystals with a stiffness constraint and volume constraint simultaneously. In the present paper, the focus is the unit cell topology optimization of porous PnCs by using a specific bi-directional structural optimization algorithm. The objective is to maximize the combined out-of-plane and in-plane band gap size for porous phononic crystals subject to bulk or shear modulus constraint with a given volume fraction. In next Section we introduce the essential governing equations and related theories of topology optimization algorithm used in this paper. This is followed by a number of optimization results and conclusions.

Governing Equations and Topology Optimization Algorithm

Governing Equations

In this paper, we consider two dimensional phononic crystals with square lattice and assume the propagation of elastic waves is restricted to the *x*–*y* plane only. The governing equation for out-of-plane transverse waves is given by:

$$\rho(\mathbf{r})\frac{\partial^2 u_z}{\partial t^2}=\frac{\partial}{\partial x}\left[\mu(\mathbf{r})\frac{\partial u_z}{\partial x}\right]+\frac{\partial}{\partial y}\left[\mu(\mathbf{r})\frac{\partial u_z}{\partial y}\right] \tag{1}$$

while the couple in-plane longitudinal and transverse waves are governed by:

$$\rho(\mathbf{r})\frac{\partial^2 u_x}{\partial t^2}=\frac{\partial}{\partial x}\left[\left(\lambda(\mathbf{r})+2\mu(\mathbf{r})\right)\frac{\partial u_x}{\partial x}+\lambda(\mathbf{r})\frac{\partial u_y}{\partial y}\right]+\frac{\partial}{\partial y}\left[\mu(\mathbf{r})\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right)\right] \tag{2}$$

$$\rho(\mathbf{r})\frac{\partial^2 u_y}{\partial t^2}=\frac{\partial}{\partial x}\left[\mu(\mathbf{r})\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[\lambda(\mathbf{r})\frac{\partial u_x}{\partial x}+\left(\lambda(\mathbf{r})+2\mu(\mathbf{r})\right)\frac{\partial u_y}{\partial y}\right] \tag{3}$$

where λ and μ denote the Lame's coefficients; ρ is the material density; and r = (*x*, *y*) denotes the position vector; $\mathbf{u}=\{u_x,u_y,u_z\}^T$ is the displacement vector, and according to Bloch's theorem when waves propagate in periodic structures it should satisfy the form:

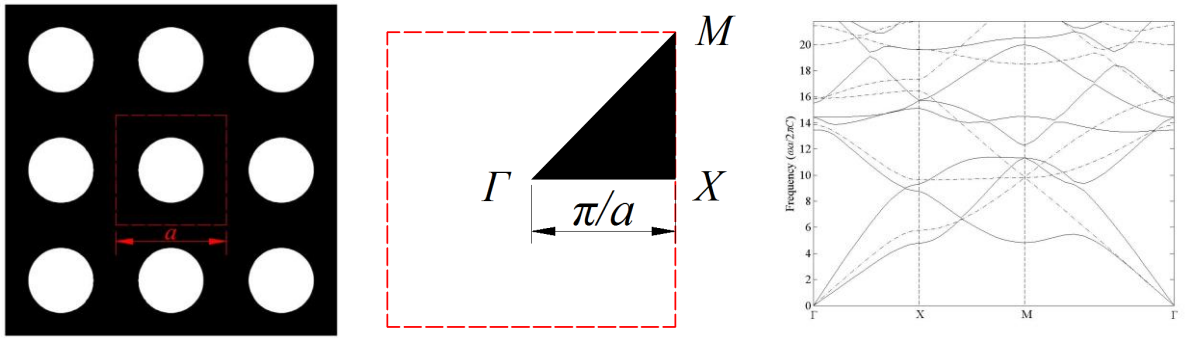
$$\mathbf{u}(\mathbf{r},\mathbf{k})=\mathbf{u}_k(\mathbf{r})e^{i(\mathbf{k}\cdot\mathbf{r})} \tag{4}$$

where $\mathbf{u}_k(\mathbf{r})$ is a periodic function of $\mathbf{r}$ with the same periodicity as the structure. $\mathbf{k} = (k_x, k_y)$ is the Bloch wave vector. With the Bloch boundary conditions, the governing equations can be converted to two eigenvalue problems for in-plane and out-of-plane waves, respectively, which both can be written as the form:

$$\left(\mathbf{K}(\mathbf{k}) - \omega(\mathbf{k})^2\mathbf{M}\right)\mathbf{u} = 0 \tag{5}$$

where eigenvectors $\mathbf{u} = \mathbf{u}_k(\mathbf{r})$. $\mathbf{K}$ and $\mathbf{M}$ are the stiffness matrix and mass matrix, respectively.

We could easily solve the problem using the finite-element method and plot the band structures ($\mathbf{k}$-ω) with eigenvalues obtained from above equations. For a 2D phononic crystal with the square lattice shown in Figure 1a, the sweep scope of Bloch wave vector $\mathbf{k}$ can be reduced to the edges of the irreducible first Brillouin zone, which is the triangle Γ-X-M-Γ shown in Figure 1b. A schematic band diagram is given in Fig.1c. The dashed lines represent eigenfrequencies for out-of-plane waves while the solid lines denote eigenfrequencies for in-plane waves. Apparently there is no any complete band gap between the out-of-plane waves and in-plane waves in the band diagram.



(a) **(b)** **(c)**

Figure 1. (a) Phononic crystals with 3 × 3 unit cells; and (b) irreducible first Brillouin zone (Γ-X-M-Γ); (c) A schematic band diagram without any band gap.

The focus of this paper is to open a complete band gap that exists in both out-of-plane and in-plane mode and gradually enlarge the gap size to obtain an optimal design. For a complete band gap among the n^{th} and $(n+1)^{th}$ dispersion branch of out-of-plane mode and the m^{th} and $(m+1)^{th}$ dispersion branch of in-plane mode, the relative band gap size is computed as the ratio of absolute band gap width and mid-gap value,

$$d_r = 2\frac{\min\left(\omega_{n+1}^{out}(\mathbf{k}),\ \omega_{m+1}^{in}(\mathbf{k})\right) - \max\left(\omega_n^{out}(\mathbf{k}),\ \omega_m^{in}(\mathbf{k})\right)}{\min\left(\omega_{n+1}^{out}(\mathbf{k}),\ \omega_{m+1}^{in}(\mathbf{k})\right) + \max\left(\omega_n^{out}(\mathbf{k}),\ \omega_m^{in}(\mathbf{k})\right)} \tag{6}$$

where ω_n^{out}, ω_{n+1}^{out}, ω_m^{in}, ω_{m+1}^{in} are eigenfrequecies at the bottom and top edges of the target band gap for out-of-plane and in-plane modes, respectively. As a result, the band gap size is a relative value with no length scale.

Objective Function

When the porous PnCs might sustain amount of static loads, ideally we want to design these structures as stiffer as possible and the best way is to maximize bulk or shear modulus and the band gaps simultaneously. However, the optimal directions for two goals are opposite to each other. As mentioned in the introduction, the stiffer and heavier material tends to be isolated by the soft and light counterpart to exhibit an optimal band gap size. In the porous case, the solid material will be isolated by air. Such structures clearly could not sustain any

loads as the solid parts are not connected. Instead of maximizing stiffness and band gaps simultaneously, we add an extra bulk or shear modulus constraint to the optimization of the phononic band gap.

The static effective elasticity tensor of a porous material with periodic microstructures can be found by the homogenization theory [19][20] in terms of the material distribution in the unit cell as,

$$E_{ij}^{H}=\frac{1}{|Y|}\int_{\Omega}\left(\left\{\varepsilon_0^i\right\}-\left\{\varepsilon^i\right\}\right)^{T}[E]\left(\left\{\varepsilon_0^j\right\}-\left\{\varepsilon^j\right\}\right)d\Omega \tag{7}$$

where E_{ij}^{H} is homogenized elasticity tensor, $[E]$ is the constitutive matrix at a given point, $|Y|$ denotes the area of the unit cell Ω , $i, j=1,2,3$ for two dimensional inhomogeneous structures, $\left\{\varepsilon_0^i\right\}$ are three linear independent test strain fields as $\left\{\varepsilon_0^1\right\}=\{1,0,0\}$, $\left\{\varepsilon_0^2\right\}=\{0,1,0\}$, $\left\{\varepsilon_0^3\right\}=\{0,0,1\}$, $\left\{\varepsilon^i\right\}$ are the introduced strain fields, which are the solutions to the standard finite element equation with periodic boundary condition and subjected to the test strain fields $\left\{\varepsilon_0^i\right\}$. Thus effective bulk or shear modulus of a porous material can be expressed as,

$$\kappa^{H}=\frac{1}{4}\left(E_{11}^{H}+E_{12}^{H}+E_{21}^{H}+E_{22}^{H}\right) \tag{8}$$

$$G^{H}=E_{33}^{H} \tag{9}$$

For simplicity, dimensionless stiffness constraints are used instead of effective bulk or shear modulus in the following numerical examples. Specifically, $\kappa = \kappa^{H}/\kappa_0$ and $G = G^{H}/G_0$ are used as effective bulk and shear modulus constraints, where κ_0 and G_0 are the bulk and shear moduli, respectively, of the solid material.

On account of potential weight limitation, a volume share of the solid in the whole design domain should be restricted. Therefore, the optimization problem under consideration can be mathematically formulated with objective and constraint functions as follows:

$$\text{Maxmize: } f\left(x_e\right)=d_r \tag{10}$$

$$\text{Subject to: } V_f^{*}=\sum_{e=1}^{N}x_e V_e \quad x_e=x_{\min} \text{ or } 1 \tag{11}$$

$$\kappa\geq\kappa^{*} \quad or \quad G\geq G^{*} \tag{12}$$

where the objective function $f(x_e)$ denotes the relative band gap size which is defined by the percentage in the following band diagrams; V_f^{*} is the volume constraint; x_e is the artificial design variable, which denotes the material type (air or solid material) for each element. κ^{*} and G^{*} are effective bulk and shear modulus constraints. It should be noted that the bulk or shear modulus constraint should be not greater than the upper limit of the porous structure with the same volume fraction, otherwise the optimization will tend to purely maximize the bulk or shear modulus. The corresponding dimensionless upper limits of bulk or shear modulus are given by Hashin–Shtrikman bounds for two-phase materials [21]. In the following optimizations, the stiffness constraints are set to $\kappa^{*} = \beta^{*}\kappa_{upper}$ or $G^{*} = \beta^{*}G_{upper}$, where β is the ratio of stiffness constraint over its upper bound value at a same volume fraction and is located in the range between 0 and 1.

Bidirectional Evolutionary Structural Optimization (BESO)

Bi-directional evolutionary structural optimization (BESO) method is a gradient-based topology optimization algorithm in optimum material distribution problems for continuum structures , which is a further developed version of evolutionary structural optimization (ESO) [22][23]. The basic concept of BESO is to gradually remove low efficient materials from the structure and meanwhile add materials to the most efficient regions so that the rest part evolves to an optimum [24]-[26]. BESO method has demonstrated its capability in the design of periodic microstructures [27], and already been successfully applied in the design of photonic and phononic band gap crystals [28][29].

To resolve the multi-objective topology optimization problem defined in Eq. (10)-(12), we apply a similar material interpolation scheme with penalization to avoid artificial localized modes as in the studies on the topology optimization of continuum structures for natural frequencies [30]. The interpolation scheme is given as:

$$\rho(x_e) = x_e \rho_0 \tag{13}$$

$$E(x_e) = \left[\frac{x_{\min} - x_{\min}^p}{1 - x_{\min}^p} \left(1 - x_e^p\right) + x_e^p \right] E_0 \qquad (0 < x_{\min} \le x_e \le 1) \tag{14}$$

where ρ_0 and E_0 represent the density and Young's modulus of solid material, respectively; p is the penalty exponent; x_e stands for a design variable, $x_e = x_{min}$ denotes element e is composed of air, and $x_e = 1$ means element e is composed of solid material. To avoid singularity in finite element analysis, x_{min} in the calculation is usually set to be a very small value that is slightly larger than 0. In the following example, the value is chosen as $x_{min} = 1 \times 10^{-6}$.

BESO starts from an initial design and then calculates the elemental sensitivities, i.e. gradients of objective function with respect to the change of design variable x_e. Based on the relative rankings of the elemental sensitivity, it will gradually modify the distribution of solid material in the following iteration steps by changing the value of the design variable of every element until the convergence criterions are satisfied. Details of sensitivity analysis and evolutionary procedure can be found in the literature [28][31][32].

Results and Discussions

We consider silicon as the solid material as an illustration example. The physical properties of silicon are given as $\rho = 2330$ kg/m3, $\lambda = 85.502$ GPa and $\mu = 72.835$ GPa [18]. The following optimizations are conducted with a volume constraint $V_f^* = 50\%$ and constraint ratio β=0.3. A filter scheme has been applied [30]. The unit cell with dimensionless lattice length a=1 is discretised into 64×64 linear four node. The eigenfrequencies (ω) in the band structures are normalized by $\omega a/2\pi C$, where C =340 m/s denotes wave speed in air. By using the aforementioned optimization algorithm, the following topologies with complete band gap have been obtained with silicon/air system.

As shown in Fig.2 and Fig.3, two optimized structures have been found with bulk modulus constraint and shear modulus constraint, respectively. For both cases, the first complete band gap is between the first and second dispersion branch of out-of-plane mode as well as between the third and fourth dispersion branch of in-plane mode, while the second complete band gap is located in the second and third dispersion branch of out-of-plane mode and the sixth and seventh dispersion branch of in-plane mode. All the optimization results again reveal that the solid material is approaching the limiting case of separate columns in air but still keeping connected by slim constructions to support the propagation of shear waves and the prescribed bulk or shear modulus as well.

When maximizing the complete band gap with bulk modulus constraint, the main parts in the resulting topologies are analogous to square inclusions. In comparison, it is interesting to observe that the main parts of the final designs with shear modulus constraint are more like round columns. The other difference between two cases is the position of thin connections. All the complete band gap sizes we obtained in these porous phononic structures have broken

the record value in the literature [17][18]. All designs are amenable to manufacture with appropriate size scaling to the frequency range of interest.

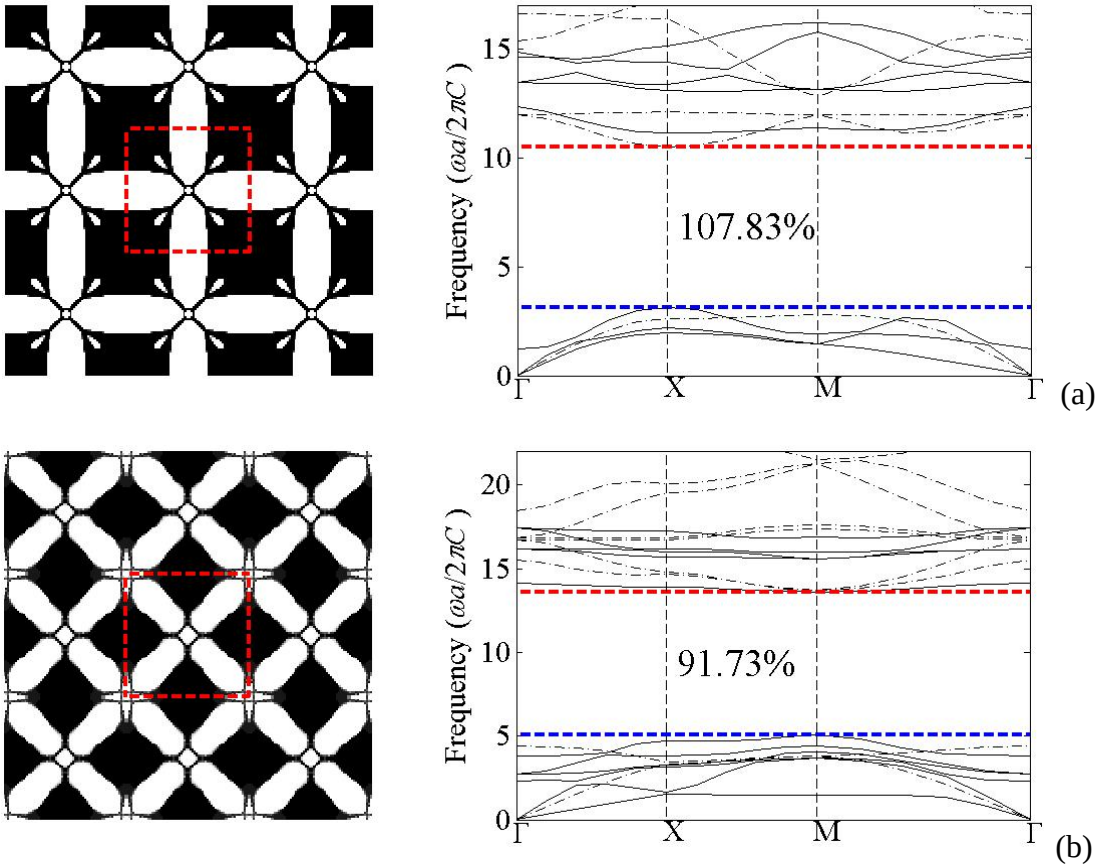
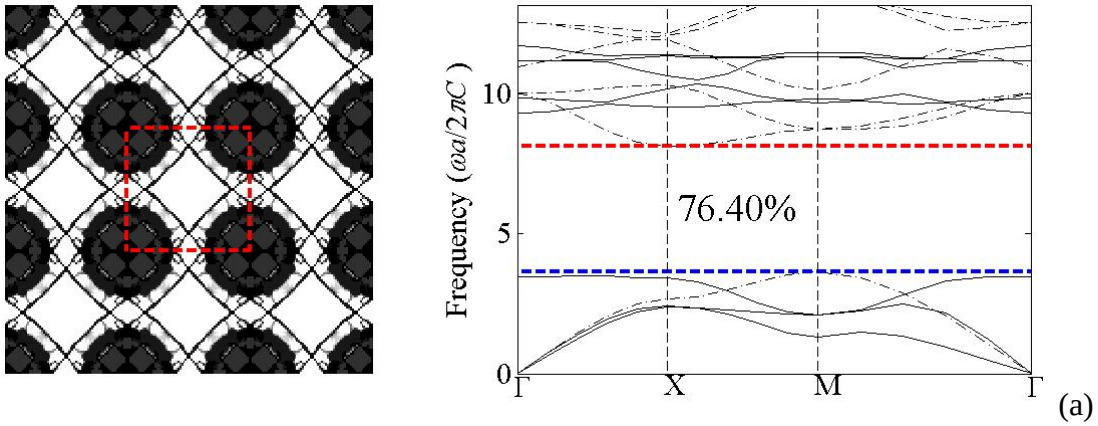


Figure 2. Optimized topologies and corresponding band structures for complete band gap with bulk modulus constraint, (a) between ω_1^{out}, ω_2^{out} and ω_3^{in}, ω_4^{in}; (b)between ω_2^{out}, ω_3^{out} and ω_6^{in}, ω_7^{in}



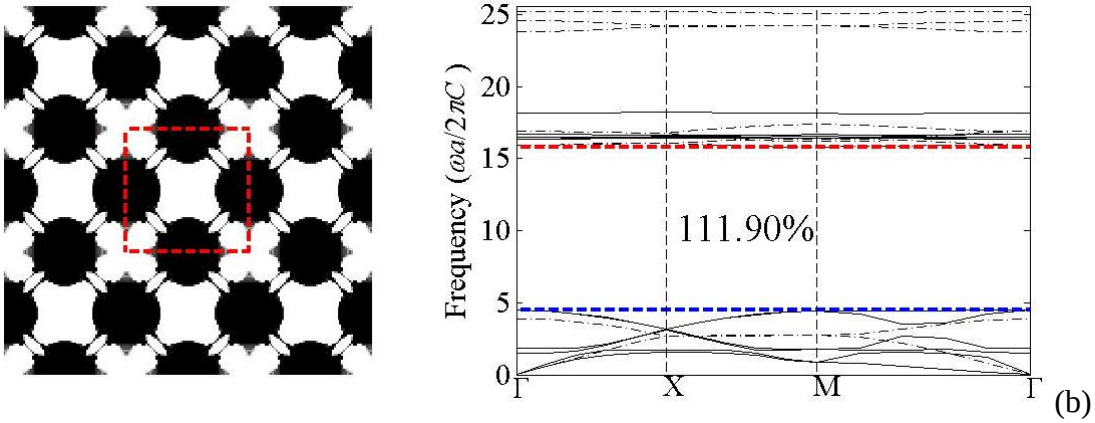


Figure 3. Optimized topologies and corresponding band structures for complete band gap with shear modulus constraint, (a) between ω_1^{out}, ω_2^{out} and ω_3^{in}, ω_4^{in}; (b)between ω_2^{out}, ω_3^{out} and ω_6^{in}, ω_7^{in}

Conclusions

This paper has discussed the topology optimization of porous phononic crystals for maximizing complete band gap between out-of-plane and in-plane mode with a bulk or shear modulus and volume constraint simultaneously. Homogenization theory and BESO algorithm have been adopted to resolve the problem. Several optimization results with bulk and shear modulus constraint were presented. Numerical results showed that there are many slim connections in the optimized topologies of porous phononic crystals. All the presented designs have exceptionally large complete band gaps.

References

[1] Sigalas, M.; Economou, E. (1992) Elastic and acoustic wave band structure. *Journal of Sound and Vibration* **158**, 377-382.

[2] Kushwaha, M.S.; Halevi, P.; Dobrzynski, L.; Djafari-Rouhani, B. (1993) Acoustic band structure of periodic elastic composites. *Physical Review Letters* **71**, 2022-2025.

[3] Sigalas, M.; Economou, E.N. (1993) Band structure of elastic waves in two dimensional systems. *Solid State Communications* **86**, 141-143.

[4] Kushwaha, M.S.; Halevi, P.; Martínez, G.; Dobrzynski, L.; Djafari-Rouhani, B. (1994) Theory of acoustic band structure of periodic elastic composites. *Physical Review B* **49**, 2313-2322.

[5] Kushwaha, M.S. (1996) Classical band structure of periodic elastic composites. *International Journal of Modern Physics B* **10**, 977-1094.

[6] Sigalas, M.; Kushwaha, M.S.; Economou, E.N.; Kafesaki, M.; Psarobas, I.E.; Steurer, W. (2005) Classical vibrational modes in phononic lattices: Theory and experiment. *Zeitschrift für Kristallographie* **220**, 765-809.

[7] Lu, M.H.; Feng, L.; Chen, Y.F. (2009) Phononic crystals and acoustic metamaterials. *Materials Today* **12**, 34-42.

[8] Pennec, Y.; Vasseur, J.O.; Djafari-Rouhani, B.; Dobrzyński, L.; Deymier, P.A. (2010) Two-dimensional phononic crystals: Examples and applications. *Surface Science Reports* **65**, 229-291.

[9] Liu, Y.; Su, J.-y.; Gao, L. (2008) The influence of the micro-topology on the phononic band gaps in 2d porous phononic crystals. *Physics Letters A* **372**, 6784-6789.

[10] Liu, Y.; Su, J.Y.; Xu, Y.L.; Zhang, X.C. (2009) The influence of pore shapes on the band structures in phononic crystals with periodic distributed void pores. *Ultrasonics* **49**, 276-280.

[11] Sigmund, O.; Jensen, J. (2003) Systematic design of phononic band-gap materials and structures by topology optimization. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **361**, 1001-1019.

[12] Gazonas, G.A.; Weile, D.S.; Wildman, R.; Mohan, A. (2006) Genetic algorithm optimization of phononic bandgap structures. *International Journal of Solids and Structures* **43**, 5851-5866.

[13] Rupp, C.J.; Evgrafov, A.; Maute, K.; Dunn, M.L. (2007) Design of phononic materials/structures for surface wave devices using topology optimization. *Structural and Multidisciplinary Optimization* **34**, 111-121.

[14] Dong, H.W.; Su, X.X.; Wang, Y.S.; Zhang, C. (2014) Topological optimization of two-dimensional phononic crystals based on the finite element method and genetic algorithm. *Structural and Multidisciplinary Optimization* **50**, 593-604.

[15] Liu, Z.; Wu, B.; He, C. (2014) Band-gap optimization of two-dimensional phononic crystals based on genetic algorithm and fpwe. *Waves in Random and Complex Media* **24**, 286-305.

[16] Hussein, M.I.; Hamza, K.; Hulbert, G.M.; Saitou, K. (2007) Optimal synthesis of 2d phononic crystals for broadband frequency isolation. *Waves in Random and Complex Media* **17**, 491-510.

[17] Bilal, O.R.; Hussein, M.I. (2011) Ultrawide phononic band gap for combined in-plane and out-of-plane waves. *Physical Review E* **84**, 065701.

[18] Dong, H.-W.; Su, X.-X.; Wang, Y.-S. (2014) Multi-objective optimization of two-dimensional porous phononic crystals. *Journal of Physics D: Applied Physics* **47**, 155301.

[19] Bendsøe, M.P.; Kikuchi, N. (1988) Generating optimal topologies in structural design using a homogenization method. *Computer methods in applied mechanics and engineering* **71**, 197-224.

[20] Sigmund, O. (1995) Tailoring materials with prescribed elastic properties. *Mechanics of Materials* **20**, 351-368.

[21] Hashin, Z.; Shtrikman, S. (1963) A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids* **11**, 127-140.

[22] Xie, Y.; Steven, G.P. (1993) A simple evolutionary procedure for structural optimization. *Computers & structures* **49**, 885-896.

[23] Xie, Y.M.; Steven, G.P. (1997) Basic evolutionary structural optimization. In *Evolutionary structural optimization*, Springer London: pp 12-29.

[24] Huang, X.; Xie, M. (2010) *Evolutionary topology optimization of continuum structures: Methods and applications*. John Wiley & Sons: Chichester.

[25] Huang, X.; Xie, Y. (2007) Convergent and mesh-independent solutions for the bi-directional evolutionary structural optimization method. *Finite Elements in Analysis and Design* **43**, 1039-1049.

[26] Huang, X.; Xie, Y. (2009) Bi-directional evolutionary topology optimization of continuum structures with one or multiple materials. *Computational Mechanics* **43**, 393-401.

[27] Huang, X.; Xie, Y.; Jia, B.; Li, Q.; Zhou, S. (2012) Evolutionary topology optimization of periodic composites for extremal magnetic permeability and electrical permittivity. *Structural and Multidisciplinary Optimization* **46**, 385-398.

[28] Li, Y.F.; Huang, X.; Meng, F.; Zhou, S. (2016) Evolutionary topological design for phononic band gap crystals. *Structural and Multidisciplinary Optimization*, 1-23.

[29] Meng, F.; Huang, X.; Jia, B. (2015) Bi-directional evolutionary optimization for photonic band gap structures. *Journal of Computational Physics* **302**, 393-404.

[30] Huang, X.; Zuo, Z.; Xie, Y. (2010) Evolutionary topological optimization of vibrating continuum structures for natural frequencies. *Computers & structures* **88**, 357-364.

[31] Huang, X.; Xie, Y. (2010) Evolutionary topology optimization of continuum structures with an additional displacement constraint. *Structural and Multidisciplinary Optimization* **40**, 409-416.

[32] Huang, X.; Radman, A.; Xie, Y. (2011) Topological design of microstructures of cellular materials for maximum bulk or shear modulus. *Computational Materials Science* **50**, 1861-1870.

Automatic Programming Via Text Mapping To Expert System Rules

Pedro V. Marcal

Mpact Corp., Oak Park, Ca., USA

pedrovmarcal@gmail.com

Abstract

Starting from semantically parsed text, a program was developed to split up all compound sentences to simple sentences. These sentences were converted to expert systems rules and then processed by a tailored Expert System program. Successful execution of the Expert System program demonstrated a form of Automatic Programming.

Keywords: Automatic Programming, Expert Systems, Knowledge Base, A. I.

Introduction

Computer Programming has always been regarded as a labor intensive process. In the early days of Artificial Intelligence, it was recognized that there was a duality between data (including text) and program and that the two were interchangeable. This was summarized in the mantra, 'Data is Program and Program is Data'. There was therefore many attempts to develop systems for automatic programming. These systems were largely unsuccessful see for example [1]. So this complex task was split up into niche processes. For example programs were written to convert mathematical equations into programs, eg. Matlab and Mathematica and open source Sympy. Another approach was to develop Domain Specific Languages [1]. The main reason for the failure to develop a general system was the computer's inability to understand text. Liu et al. [2] developed a system based on text understanding by requiring the text take the form of a story. With this approach [2] was able to develop a top level schematic programming system while leaving the details to be filled in by the traditional method.

It is obvious that in order to convert text to program, we must first understand the text. Marcal[3] developed a semantic parser in two steps. First, the text was processed by a context free parse in English. The Semantic parse was accomplished by a translation of the text into Simplified Chinese (Mandarin). English and Chinese are orthogonal in meaning. One English word has many meanings because it is based on phonetics. Whereas one Chinese word (set of characters) has only 1 meaning, derived from its ideograph. So there are many Chinese Words with the same meaning. A statistical parsing method was developed in [3] that used Design Of Experiments[4] to reduce the search through the possible combinations of meaning to provide the optimal translation. The statistic was based on ngram of 3 (sequence of 3 words) . This method reduced the computational effort of parsing by two orders of magnitude. This method required good Corpora in both English and Chinese. In addition, using Wordnet and the two Corporas as a basis, the author constructed a Lexical Dictionary that covered most of the English and Chinese Languages. The result of the semantic parse (translation to Chinese) was reflected back to English in the form of an unambiguous sentence. This is the starting point for the current project.

In the early 1970s , Expert systems were regarded as a promising approach to Artificial Intelligence[5]. The writer took part in a project to explain the principles behind a nonlinear Finite Element Program, MARC to the level that enabled interested Engineers to use the program. To this end, the SACON (Stress Analysis Consultant) was developed based on EMYCIN. This project also had important ramifications for the Expert System Developers because for the first time the rules encompassed a

complete Domain. This emphasized the need for completeness in a system. In addition the Domain could be used to develop systems of Computer assisted instructions. In further work, Racz and Marcal[6] extended expert Systems to also call functions and external Programs. In the early days of nonlinear analysis, it was usual to employ Ph. D.s to perform such analysis. The Nuclear Industry had need for a large number of analysis of their components. Two Japanese Companies were able to develop expert systems for the MARC program[7]. It found that Engineering Aides with High School Diplomas were able to perform such nonlinear analysis (I assume under close supervision). The problem with the Expert Systems was that it required specialists to elicit the understanding of the experts and then codify it in the form of rules. It was in fact another form of programming and twice as laborious because the coding was one removed.

Objective

The objective of the current work is to start with the semantic parse of text and split any compound sentences to a sequence of simple sentences which we will call hyper-sentences. These hyper-sentences form simple concepts of the Subject,Verb, Object (SVO) type. We will call these concepts hyper-concepts. It was found that these hyper-concepts are automatically mapped into expert systems rules, called hyper-rules. An expert system was developed to process these hyper-rules in the usual way. The successful development of an expert system constitutes the form of automatic programming that we seek. The hyper-concepts are of interest in their own right and can be classified in the usual hierarchical way, similar to the Wordnet scheme with hypo and hyper relations. In order to assist in this, we adopt the Conceptual Dependency (CD) of Schank[8-9]. Schank and his colleagues showed that the verbs in hyper-concepts could be represented by sixteen hypo-verbs and that these verbs can be classified in three states, viz Property , Physical and Mental respectively. In CD theory these verbs designate actions that are labeled ATRANS, PTRANS and MTRANS respectively.

The following are some examples of the hyper-verbs.

. TRANS transfer (of possession,location, and of ideas)

. MOVE, PROPEL, GRASP movement and forces

.INGEST, EXPEL absorb and its opposite

.ATTEND, SPEAK, RECORD sense, verbal output and write or record

.BUILD construct.

.COMPUTE compute

.STMEM,LTMEM short term remember and long term remember

.TIME passage of time

.LIVE biological

.DO any other verbs that can not be classified above.

Finally, we adopt Schank's construction of a hyper sentence instead of the traditional SVO, we use PP (Picture Painter), ACT (Action), Ppo(Picture Painter Object). The picture Painters are in turn modified by Picture Aiders (Adjectives, prepositions). The ACT are in turn modified by ACT Aiders (AA,adverbs)

Since any of these verbs can be classified in the three states above, we have expanded the original CD to have a total of 48 possible classification. In the original CD theory Schank did not attempt to classify the Subject and Object phrases. In the current work, we expanded CD to also classify these phrases

using a Wordnet type generalization. With such an approach it was found that the hyper concepts took on their own properties. These hyper-concepts were found to follow Zipf's Law with similar types of properties as that found for words which were originally found to obey Zipf's Law.

Theoretical Considerations

There is little additional theory to consider above that already discussed. The difficulties lay in its algorithmic implementation as a computer code. The following is a computer flow of the program. It is conveniently separated into two programs. The first develops the hyper-rules while the second processes the rules in an expert system. The programs are written in Python 2.7. Python's string processing features has made the coding easier and its dictionary with its seamless in-core and out of core storage has proven invaluable.

Rule development for each sentence in sequence.

1. Context-free Parse sentence.
2. Semantic Parse by translation into Chinese.
3. Chunk the phrases and transfer their lexical meaning back to English. (for ease of use).
4. Split the phrases into hyper-sentences with hyper-concepts.
5. Define the cd hypo concept for each hyper sentence.
6. Convert each hyper sentence into a hyper rule.
7. Separate rules into two categories. In the first the verbs are not modified by any adverbial or ACT Aiders. These rules are defined once and for all. The second category contains ACT Aiders and these form conditions in the hyper-rules. We call these eligible rules (for expert system processing)
8. Search for rule conclusion in the hyper-concepts belonging to the same original sentence or following the original sentence. These are marked by prepositions such as then (PP Aiders).

We note that each hyper-rule carries its own words and its conceptual dependency. Because each phrae may be expressed in a myriad of different word combinations, the CD takes on special significance for processing in an expert system. This is in fact the key to converting hyper-concepts into hyper-rules for expert systems. Every paragraph exists for a reason. The paragraph usually answers one of the following questions viz. what, where, when, how? These are then the objective or in expert system parlance the golden rule for each paragraph. In most cases, the first appearance of such a preposition (PPA) denotes a golden rule. If a golden rule is not obvious the expert system prompts the user or asks the user to accept its best estimate.

There is a certain amount of pre processing required to prepare the hyper-rules for efficient processing by the expert system. Lists such as menu items are collected for convenience. It is important that critical eligible rules with multiple conditions be identified, and their transfer locations be identified as system switches. Once transferred to a system switch, the eligible rules are processed in sequence until another switch is encountered. Then control is transferred back to the original switching rule, for execution of the next rule.

Development of Expert system.

1. Pre process rules for efficiency. Each eligible hyper rule is assigned an event. Each event contains a detailed analysis of the rule as to how it affects the expert system.

2. Identify Golden Rules. (noted in event)
3. Identify switching rules. (as compound rules).(noted in event)
4. Identify system switches. (noted in event)
5. Execute rules in sequence. Record each executed rule in a journal.
6. Contact the user for unresolved rules that cannot be processed further. It is at this stage that the dialog is constructed to allow the user to query the actions taken and recorded in the journal. Usually, the actions taken to remedy the situation requires a new text and a repeat of the hyper-rule construction in the first step above.
7. When all the eligible rules have been satisfied, save the conclusions and the journal and exit the program.

Case Study

In this case study, we repeat the problem used by Liu et al [2] to automatically program the solution. The text describes a bar in the following.

This is a bar with a bartender, who makes drinks.
The bar has a menu containing some drinks, which include : a sour apple martini, a margarita, and rum and coke.
When a customer orders a drink, the bartender tries to make it. When the bartender is asked to make a drink, he makes it and gives it to the customer only if the drink is in the menu's drinks.
Otherwise, the bartender says to the customer, ' Sorry I don't know how to make that drink '.
In [2], the solution is neatly encased in the following Class using its internal format.

```
class bartender :
   def make(drink) :
        if (drink in menu.drinks) :
           bartender.make(drink)
           bartender.give(drink, customer)
        else :
           bartender.say( \
                "Sorry I don't know how to make that drink.", customer)
```

The solution here follows the same lines, except that there is less need for collection of concepts that relate to each other. This is implied in the rules and is automatically recognized by the expert system.

The parsed hyper-concepts are listed in Appendix A. These then are turned into hyper-rules which are listed in Appendix B.

The hyper-rules are paraphrased in the following.

Rule (4000,0) defines bar as a structure

Rule (4000,1) defines bartender makes drink.

Rule (4001,0) Defines existence of menu.

Rule (4001,1) Defines menu as containing drinks.

Rule (4001,2) Defines drinks list.

Rule (4002,0) Customer requests drink, Golden Rule.

Rule (4002,1) Bartender makes it, but there is a restriction (either option A or B)

Rule (4003,0) Bartender figures out rule.

Rule (4003,1) Action if option.

Rule (4003,2) Action if option.

Rule (4003,3) Condition for option A, System Switch A.

Rule (4003,4) Give drink to customer as per option A.

Rule (4004,0) Condition for Option B, implied System Switch B.

Rule (4004,1) Bartender speaks an apology.

Rule (4004,2) Excuse is does not know how to make.

Rule (4004,3) That drink.

End of Rules.

The processing of the rules gave the same results as the automatic coding by Liu et al [2].

Hence we conclude that we have achieved our objective of automatic programming of rules for an expert system.

Discussion and further work

The ability to convert text to programs is very important. Mainly because most of our complete history and culture is recorded as text. In this project we have achieved this in a general way. This process may also be extended to obtain summaries from text by systematically asking the questions. What, where, how, why? We would then need to extend this to querying the internet to provide related text. The important action would be to systematically store the texts processed so that the rules can be retrieved as and when they are required. The Watson program[10] does an extensive job in collecting a knowledge base. The difference with the current program is that [10] does not do such a detailed job of parsing and labeling as it is done here..

Another direction that this development can proceed is to set up a mixture of mathematical equations sprinkled with text to control the results of the computing. This process is similar to the current coding in CAE, but here robotized.

Finally for this process to act in real time, it must be accelerated by parellization. It currently takes on average about 33 secs. to process a sentence on a PC.

Conclusions

A program has been developed for generating hyper-rules from text.

These hyper-rules can be processed in a tailored Expert System to achieve automatic programming.

We have achieved the first step in our objective to develop an automatic way of converting text to programs.

References

[1]. C. Rich, R.C. Waters,' Readings in Artificial Intelligence and Software Engineering.' Morgan Kaufman Publishers, 1986.

[2] H. Liu, H. Lieberman,'Metafor: Vizualizing Stories as Code',Proc. IUI'05, San Diego, CA, 2005.

[3]. P.V. Marcal, 'Development of a General Semantic Reader (GPSR)' ,available on Research Gate by following my research, Jan. 1, 2012

[4] P.V. Marcal, and J. Fong, 'A Design-of-Experiments approach to Statistical Parsing of a Natural Language Abstracting Code for Fatigue Data Event Databases', Proc ICCES, June, 2014, Korea.

[5]. P.V. Marcal, "Knowledge Engineering and Artificial Intelligence"; Advanced Topics and New Developments in Finite Element Analysis, ASME WAM, Monterey, California, (July, 1978). Result of work with R.J. Melosh and E.A. Feigenbaum on the development of SACON.

[6].S. Racz, and P.V. Marcal, "SACON2, An Expert System for Automating the FEM Process", Proc. WCCM, 2006

[7]. S. Hiromi, I. Ishihara,H. Kobayashi,M. Kajiwara,'Developm,ent of Expert System MARCAS Supporting the Preparation of MARC Input Deck', Proc.MARC, Users Conference, CA., 1988.

[8] R. Schank, and C.K. Riesbeck, 'Inside Computer Understanding.', editors, Lawrence Erlbaum Associates , 1972

[9]. A.G.Francis, Jnr.'A Pocket Guide To CD Theory', Lecture Notes, College of Computing, Georgia Inst. Technology, GA, 1974,

[10] D. Ferruci, Watson, The Deep Qa Project, IBM Research, Feb, 2011.

Appendix A: details of parsing

```
 *** WriteDict *** hyper_concept
hyper_concept  ('4000', '0')   value
('4000', '0')
['cd_concept', 'object', 'is', 'equal', 'bar', 'ACTION', 'PAo',
'what_alliance', 'human']
[['PP', '0'], ['this', 'PP', 'object', 'ATRANS']]
[['ACT', '1'], ['is', 'ACT', 'equal', 'ATRANS']]
[['PPo', '2'], ['bar', 'PP', 'structure', 'PTRANS']]
[['PAo', '3'], ['with', 'P', 'possess', 'MTRANS']]
[['PAo', '4'], ['bartender', 'PP', 'human', 'PTRANS']]
['PP', 'this', 'object', 'ATRANS', 'ACT', 'is', 'equal', 'ATRANS', 'PPo',
'bar', 'structure', 'PTRANS']
hyper_concept  ('4000', '1')   value
('4000', '1')
['cd_concept', 'human', 'make', 'build', 'material', 'ACTOR']
[['PP', '6'], ['who', 'PP', 'human', 'ATRANS']]
[['ACT', '7'], ['make', 'ACT', 'build', 'PTRANS']]
[['PPo', '8'], ['drink', 'PP', 'food', 'MTRANS']]
['PP', 'who', 'human', 'ATRANS', 'ACT', 'make', 'build', 'PTRANS', 'PPo',
'drink', 'food', 'MTRANS']
hyper_concept  ('4001', '0')   value
('4001', '0')
['cd_concept', 'bar', 'have', 'ingest', 'signal', 'ACTION']
[['PP', '0'], ['bar', 'PP', 'structure', 'PTRANS']]
```

```
[['ACT', '1'], ['have', 'ACT', 'ingest', 'ATRANS']]
[['PPo', '2'], ['menu', 'PP', 'record', 'MTRANS']]
['PP', 'bar', 'structure', 'PTRANS', 'ACT', 'have', 'ingest', 'ATRANS',
'PPo', 'menu', 'record', 'MTRANS']
hyper_concept  ('4001', '1')   value
('4001', '1')
['cd_concept', 'signal', 'contain', 'do', 'material', 'ACTION']
[['ACT', '3'], ['contain', 'ACT', 'do', 'PTRANS']]
[['PPo', '4'], ['some', 'PA', 'quantity', 'ATRANS']]
[['PPo', '4'], ['drink', 'PP', 'food', 'MTRANS']]
['ACT', 'contain', 'do', 'PTRANS', 'PPo', 'drink', 'food', 'MTRANS']
hyper_concept  ('4001', '2')   value
('4001', '2')
['cd_concept', 'object', 'include', 'do', 'material', 'ACTION', 'PAo',
'substance.n.01', 'material']
[['PP', '6'], ['which', 'PP', 'object', 'ATRANS']]
[['ACT', '7'], ['include', 'ACT', 'do', 'PTRANS']]
[['PPo', '9'], ['sour', 'PA', 'sense', 'ATRANS']]
[['PPo', '9'], ['apple', 'PP', 'food', 'PTRANS']]
[['PPo', '9'], ['martini', 'PP', 'food', 'MTRANS']]
[['PAo', '11'], ['margarita', 'PP', 'food', 'MTRANS']]
[['PAo', '14'], ['rum_and_coke', 'PP', 'food', 'MTRANS']]
['PP', 'which', 'object', 'ATRANS', 'ACT', 'include', 'do', 'PTRANS',
'PPo', 'martini', 'food', 'MTRANS']
hyper_concept  ('4002', '0')   value  ### This becomes the golden rule
('4002', '0')
['cd_concept', 'human', 'order', 'plan', 'material', 'THINK']
[['PP', '1'], ['customer', 'PP', 'money_value', 'PTRANS']]
[['AA', '0'], ['when', 'AA', 'age_destination', 'PTRANS']]  ### trigger for
###golden rule. When asked customer replies with A) drink in menu eg. ###
### Margarita or B) drink not in menu eg vodka martini.
[['ACT', '2'], ['order', 'ACT', 'plan', 'MTRANS']]
[['PPo', '3'], ['drink', 'PP', 'food', 'MTRANS']]
['PP', 'customer', 'money_value', 'PTRANS', 'ACT', 'order', 'plan',
'MTRANS', 'PPo', 'drink', 'food', 'MTRANS']
### at this point we activate a switch with value A or B respectively.
hyper_concept  ('4002', '1')   value
('4002', '1')
['cd_concept', 'human', 'make', 'build', 'knowledge', 'ACTOR']
[['PP', '5'], ['bartender', 'PP', 'human', 'PTRANS']]
[['ACT', '6'], ['try', 'ACT', 'compute', 'MTRANS']]
[['ACT', '6'], ['make', 'ACT', 'build', 'PTRANS']]
[['PPo', '7'], ['it', 'PP', 'object', 'MTRANS']]
['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'make', 'build', 'PTRANS',
'PPo', 'it', 'object', 'MTRANS']
hyper_concept  ('4003', '0')   value
('4003', '0')
['cd_concept', 'human', 'ask', 'transmit', 'None', 'ACTOR']
[['PP', '1'], ['bartender', 'PP', 'human', 'PTRANS']]
[['AA', '0'], ['when', 'AA', 'age_destination', 'PTRANS']]
[['ACT', '2'], ['is', 'ACT', 'equal', 'ATRANS']]
[['ACT', '2'], ['ask', 'ACT', 'transmit', 'MTRANS']]
['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'ask', 'transmit', 'MTRANS']
hyper_concept  ('4003', '1')   value ### System Switch A
```

```
('4003', '1')
['cd_concept', 'human', 'make', 'build', 'None', 'ACTOR']
[['ACT', '4'], ['make', 'ACT', 'build', 'PTRANS']]
['ACT', 'make', 'build', 'PTRANS']
hyper_concept  ('4003', '2')   value
('4003', '2')
['cd_concept', 'human', 'make', 'build', 'knowledge', 'ACTOR']
[['PP', '6'], ['he', 'PP', 'human', 'ATRANS']]
[['ACT', '7'], ['make', 'ACT', 'build', 'PTRANS']]
[['PPo', '8'], ['it', 'PP', 'object', 'MTRANS']]
['PP', 'he', 'human', 'ATRANS', 'ACT', 'make', 'build', 'PTRANS', 'PPo',
'it', 'object', 'MTRANS']
hyper_concept  ('4003', '3')   value
('4003', '3')
['cd_concept', 'human', 'give', 'trans', 'knowledge', 'UNDEF', 'PAo',
'source', 'human', 'prepD', '->', 'PAo']
[['CNJ', '9'], ['and', 'CNJ', 'coordinating_object', 'MTRANS']]
[['AA', '14'], ['only', 'AA', 'qualification_value', 'PTRANS']]
[['ACT', '10'], ['give', 'ACT', 'trans', 'PTRANS']]
[['PPo', '11'], ['it', 'PP', 'object', 'MTRANS']]
[['PAo', '12'], ['to', 'P', 'recipient', 'MTRANS']]
[['PAo', '13'], ['customer', 'PP', 'money_value', 'PTRANS']]
['ACT', 'give', 'trans', 'PTRANS', 'PPo', 'it', 'object', 'MTRANS']
hyper_concept  ('4003', '4')   value
('4003', '4')
['cd_concept', 'None', 'is', 'equal', 'signal', 'UNDEF', 'PAo',
'aspect_what_property', 'material']
[['CNJ', '15'], ['if', 'CNJ', 'subordinating_qualification', 'MTRANS']]
[['ACT', '16'], ['is', 'ACT', 'equal', 'ATRANS']]
[['PPo', '17'], ['menu', 'PP', 'record', 'MTRANS']]
[['PAo', '18'], ['of', 'P', 'possess', 'property']]
[['PAo', '19'], ['drink', 'PP', 'food', 'MTRANS']]
['ACT', 'is', 'equal', 'ATRANS', 'PPo', 'menu', 'record', 'MTRANS']
hyper_concept  ('4004', '0')   value ### Swystem Switch B
('4004', '0')
['cd_concept', 'human', 'say', 'speak', 'None', 'ACTOR', 'PAo', 'source',
'human', 'prepD', '->', 'PAo']
[['PP', '2'], ['bartender', 'PP', 'human', 'PTRANS']]
[['AA', '0'], ['otherwise', 'AA', 'property_lest', 'PTRANS']]
[['ACT', '3'], ['say', 'ACT', 'speak', 'MTRANS']]
[['PAo', '4'], ['to', 'P', 'recipient', 'MTRANS']]
[['PAo', '5'], ['customer', 'PP', 'money_value', 'PTRANS']]
['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'say', 'speak', 'MTRANS']
hyper_concept  ('4004', '1')   value
('4004', '1')
['cd_concept', 'language', 'know', 'move', 'None', 'UNDEF']
[['PP', '7'], ['sorry', 'PA', 'transmit', 'ATRANS']]
[['PP', '7'], ['i', 'PP', 'human', 'ATRANS']]
[['PP', '7'], ['do', 'PP', 'None', 'ATRANS']]
[['ACT', '8'], ['know', 'ACT', 'move', 'MTRANS']]
['PP', 'do', 'solfa_syllable.n.01', 'ATRANS', 'ACT', 'know', 'move',
'MTRANS']
hyper_concept  ('4004', '2')   value
('4004', '2')
```

```
['cd_concept', 'human', 'make', 'build', 'None', 'ACTOR']
[['AA', '9'], ['how', 'AA', 'None', 'PTRANS']]
[['ACT', '10'], ['make', 'ACT', 'build', 'PTRANS']]
['ACT', 'make', 'build', 'PTRANS']
hyper_concept  ('4004', '3')   value
('4004', '3')
['cd_concept', 'money_value', 'drink', 'ingest', 'None', 'ACTION']
[['PP', '11'], ['that', 'PP', 'object', 'ATRANS']]
[['ACT', '12'], ['drink', 'ACT', 'ingest', 'ATRANS']]
['PP', 'that', 'object', 'ATRANS', 'ACT', 'drink', 'ingest', 'ATRANS']
```

Appendix B. hyper-rules used by expert system

```
 *** WriteDict *** rule_collector
Here is a brief explanation of the rules.
The first label in the rule '<define>' says its a constant rule. The rest
of the line gives the PP and the PPo phrases. The second line gives the
hyper-concept. The next lines give the prepositional phrases (PA)
In the case of an eligible rule the first label is a tag for which Adverb
influences the processing of the current rule. For example the '<if_age>'
label (hypo-classification of the word 'when') is first encountered as an
eligible rule. The components of the rule follow the same order as before.
It is also labeled as a golden rule. The same label is then used to tag the
following rules which are executed by the system. This continues until a
switch changes the control to '<if_only>' (option A of the drink order.)
Then finally control is switched to '<if_alt>' (option B of the drink
order.)

rule_collector  ('4000', '0')   value
['<define>', [['this', 'PP', 'object', 'ATRANS'], ['bar', 'PP',
'structure', 'PTRANS']]]
['<define>', ['PP', 'this', 'object', 'ATRANS', 'ACT', 'is', 'equal',
'ATRANS', 'PPo', 'bar', 'structure', 'PTRANS']]
['<define>', [['this', 'PP', 'object', 'ATRANS'], ['with', 'P', 'possess',
'MTRANS'], ['bartender', 'PP', 'human', 'PTRANS']]]
rule_collector  ('4000', '1')   value
['<define>', [['who', 'PP', 'human', 'ATRANS'], ['drink', 'PP', 'food',
'MTRANS']]]
['<define>', ['PP', 'who', 'human', 'ATRANS', 'ACT', 'make', 'build',
'PTRANS', 'PPo', 'drink', 'food', 'MTRANS']]
rule_collector  ('4001', '0')   value
['<define>', [['bar', 'PP', 'structure', 'PTRANS'], ['menu', 'PP',
'record', 'MTRANS']]]
['<define>', ['PP', 'bar', 'structure', 'PTRANS', 'ACT', 'have', 'ingest',
'ATRANS', 'PPo', 'menu', 'record', 'MTRANS']]
rule_collector  ('4001', '1')   value
['<define>', ['ACT', 'contain', 'do', 'PTRANS', 'PPo', 'drink', 'food',
'MTRANS']]
['<list>', [['some', 'PA', 'quantity', 'ATRANS'], ['drink', 'PP', 'food',
'MTRANS']]]
['<signal_list>', [('4001', '0'), ['PPo', 'menu', 'record', 'MTRANS'],
[['some', 'PA', 'quantity', 'ATRANS'], ['drink', 'PP', 'food', 'MTRANS']]]]
rule_collector  ('4001', '2')   value
['<define>', [['which', 'PP', 'object', 'ATRANS'], ['martini', 'PP',
```

```
'food', 'MTRANS'], ['margarita', 'PP', 'food', 'MTRANS'], ['rum_and_coke',
'PP', 'food', 'MTRANS']]]
['<define>', ['PP', 'which', 'object', 'ATRANS', 'ACT', 'include', 'do',
'PTRANS', 'PPo', 'martini', 'food', 'MTRANS']]
['<define>', [['sour', 'PA', 'sense', 'ATRANS'], ['apple', 'PP', 'food',
'PTRANS']]]
rule_collector  ('4002', '0')   value
['<if_age>', [['customer', 'PP', 'money_value', 'PTRANS'], ['drink', 'PP',
'food', 'MTRANS']]]
['<if_age>', [['when', 'AA', 'age_destination', 'PTRANS'], ['order', 'ACT',
'plan', 'MTRANS']]]
['<if_age>', ['PP', 'customer', 'money_value', 'PTRANS', 'ACT', 'order',
'plan', 'MTRANS', 'PPo', 'drink', 'food', 'MTRANS']]
['<system_gold>', ['acquire', 'relate_key', '<if_age>', True,
'instantiate', False, 'iterate']]
rule_collector  ('4002', '1')   value
['<if_age>', [['bartender', 'PP', 'human', 'PTRANS'], ['it', 'PP',
'object', 'MTRANS']]]
['<if_age>', ['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'make', 'build',
'PTRANS', 'PPo', 'it', 'object', 'MTRANS']]
rule_collector  ('4003', '0')   value
['<if_age>', [['bartender', 'PP', 'human', 'PTRANS']]]
['<if_age>', [['when', 'AA', 'age_destination', 'PTRANS'], ['is', 'ACT',
'equal', 'ATRANS']]]
['<if_age>', ['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'ask',
'transmit', 'MTRANS']]
rule_collector  ('4003', '1')   value
['<if_age>', ['ACT', 'make', 'build', 'PTRANS']]
rule_collector  ('4003', '2')   value
['<if_age>', [['he', 'PP', 'human', 'ATRANS'], ['it', 'PP', 'object',
'MTRANS']]]
['<if_age>', ['PP', 'he', 'human', 'ATRANS', 'ACT', 'make', 'build',
'PTRANS', 'PPo', 'it', 'object', 'MTRANS']]
rule_collector  ('4003', '3')   value
['<if_only>', [['and', 'CNJ', 'coordinating_object', 'MTRANS']]]
['<if_only>', [['only', 'AA', 'qualification_value', 'PTRANS'], ['give',
'ACT', 'trans', 'PTRANS']]]
['<system_switch>', ['modify', 'call_all', '<if_only>', True, 'then',
False, '<if_alt>']]
['<then>', ['ACT', 'give', 'trans', 'PTRANS', 'PPo', 'it', 'object',
'MTRANS']]
['<then>', [['it', 'PP', 'object', 'MTRANS']]]
['<then>', [['it', 'PP', 'object', 'MTRANS'], ['to', 'P', 'recipient',
'MTRANS'], ['customer', 'PP', 'money_value', 'PTRANS']]]
rule_collector  ('4003', '4')   value
['<if_only>', [['if', 'CNJ', 'subordinating_qualification', 'MTRANS']]]
['<if_only>', ['ACT', 'is', 'equal', 'ATRANS', 'PPo', 'menu', 'record',
'MTRANS']]
['<if_only>', [['menu', 'PP', 'record', 'MTRANS']]]
['<if_only>', [['menu', 'PP', 'record', 'MTRANS'], ['of', 'P', 'possess',
'property'], ['drink', 'PP', 'food', 'MTRANS']]]
rule_collector  ('4004', '0')   value
['<if_alt>', [['bartender', 'PP', 'human', 'PTRANS']]]
['<if_alt>', [['otherwise', 'AA', 'property_lest', 'PTRANS'], ['say',
```

```
'ACT', 'speak', 'MTRANS']]]
['<if_alt>', ['PP', 'bartender', 'human', 'PTRANS', 'ACT', 'say', 'speak',
'MTRANS']]
['<if_alt>', [['bartender', 'PP', 'human', 'PTRANS'], ['to', 'P',
'recipient', 'MTRANS'], ['customer', 'PP', 'money_value', 'PTRANS']]]
rule_collector  ('4004', '1')   value
['<if_alt>', [['sorry', 'PA', 'transmit', 'ATRANS'], ['i', 'PP', 'human',
'ATRANS'], ['do', 'PP', 'None', 'ATRANS']]]
['<if_alt>', ['PP', 'do', 'solfa_syllable.n.01', 'ATRANS', 'ACT', 'know',
'move', 'MTRANS']]
rule_collector  ('4004', '2')   value
['<if_alt>', [['how', 'AA', 'None', 'PTRANS'], ['make', 'ACT', 'build',
'PTRANS']]]
['<if_alt>', ['ACT', 'make', 'build', 'PTRANS']]
rule_collector  ('4004', '3')   value
['<if_alt>', [['that', 'PP', 'object', 'ATRANS']]]
['<if_alt>', ['PP', 'that', 'object', 'ATRANS', 'ACT', 'drink', 'ingest',
'ATRANS']]
```

Transfer and pouring processes of casting by smoothed particle hydrodynamic method

†M. Kazama[1], K. Ogasawara[1], *T. Suwa[1], H. Ito[2], and Y. Maeda[2]

[1]Application development div., Next generation technical computing unit, FUJITSU Limited, Japan
[2]Department of Mechanical Engineering, Daido University, Japan

*Presenting author: suwa.tamon@jp.fujitsu.com
†Corresponding author: kazama.masaki@jp.fujitsu.com

Abstract

The casting process includes the transportation process and the pouring process. The transportation process is to carry the molten metal from the place where it is saved to the mold or the injection molding machine, and the pouring process is to pour and fill molten metal in the mold.

In the transportation process, it is undesirable to expose the surface of the molten metal to air to suppress the defect (for instance, generation of the oxidizing layer). Therefore, it is necessary to move to pour it into the mold as soon as possible after the molten metal is bailed out by the ladle. However, it is also necessary to prevent overflow of the molten metal and intervention of air or gas. So the liquid vibration by the acceleration and deceleration of the transfer machine should be suppressed.

Then, the development of the technique to control the liquid surface oscillation is needed. Computational fluid dynamics is expected to be effective as the means of the verification. The phenomenon handled here should treat the wall in the container to be a moving boundary, and the liquid surface as a free boundary.

In the pouring process, the flow of the molten metal in the mold is the target of interest. Because the quality of a final article of cast is dependent on how the molten metal flows in complex shape of the mold. Also as for this process, the use of numeric fluid analysis on the process including the fission and the fusion on a free surface is expected to be effective.

The particle-based fluid analysis methods are considered as the numerical computation technique which is applicable and useful in the treatment of these moving boundary and free boundary. However, quantitative comparative studies with the data of actual transportation and pouring process are few.

We have applied the smoothed particle hydrodynamics method to the process of transportation and pouring and validated the results with experimental data. We report on the technique and the result because we saw the experimental data and our numerical results are a good agreement.

Keywords: Casting, particle-based method, smoothed particle hydrodynamics, transfer process, pouring process,

Introduction

In computer aided engineering (CAE) of the casting process, it is necessary to simulate the flow behavior (flow analysis that solves the Navier-Stokes equation), the heat transfer phenomenon (coupled analysis of heat conduction equation and the Navier-Stokes equation), and solidification phenomena (the phase change of the liquid and the solid is indispensable) adequately. Moreover, in some cases, it is necessary to treat casting stress analysis, which is an analysis of heat transfer and the thermal stress and strain caused by solidification and is solved with the elastoplasticity and the viscoelasticity equation. Casting CAE software of the

main current is adopting the numerical method for analysis of the Eulerian (grid-based) algorithms. It is not complete in generality and practical use, though these software can suitably simulate some of the casting process. There exists casting processes which are difficult to be treated by the simulators.

On the other hand, the particle-based methods are numerical analysis method of the Lagrangian algorithm, and there is an expectation that it can analyze casting processes which the numerical solution technique of Eulerian algorithms are not suitable for (e.g. Cleary (2010) [1]).

Then, we have performed the reproduction calculation that used the smoothed particle hydrodynamic (SPH) method, which is a particle-based method, simulation about the transportation of the molten metal and the experiment on the mold filling processes, compared the experimental result and the numerical computation result, and examined the possible application to the casting process of the particle method simulation.

Experiments

Transportation of liquid container

The molten metal is frequently transported in the casting process.
When transportation, because the vibration of the liquid surface generates the oxidization layer and causes the defect, it is necessary to execute it promptly with less vibration on the liquid surface. In the tilt type pouring process, the ladle is inclined and molten metal is poured to the ingate of the mold. Although it is hoped that molten metal is filled in the ingate quickly, it is also very important to suppress the vibration of the liquid surface as well as transportation. In order to shorten the lead time, the tilt motion is often begun while moving the transportation container (ladle). That is, there is a case to do transportation and tilting at the same time.

It is thought the numerical analysis of the Eulerian algorithm are not good at this type of process, which include moving container (solid wall). Therefore it is profitable if it is shown to be able to simulate it by the particle-based method.

Then, we do the numerical analysis with the particle-based method on the same condition as the experiment executed by Okatsuka et al. (2011) and Shibuya et al. (2013) [2][3]. A concise explanation of this experiment is as follows: Water is put in the liquid container of 10mm in thickness that installs the level sensor shown in Figure 1 up to the height of 140mm. The container is transported on the x axis and tilted with T shaft center. At this time, the vibration of the liquid surface is controlled by controlling the transportation speed and the tilting speed so that the wave should not occur.

Mold filling processes

Mampaey and Xu (1995) performed experiments of mold filling process in order to directly observe the molten metal flow for the model with different runner shape by arranging the heat-resistant glass in the one side of the mold[4]. The analysis of this behavior that fills the runner and the cavity is an object that cast CAE software of grid-based method analyzes enough, and the part that can be called the indispensable function of cast CAE software. Therefore, we also analyzed to attempt the utility of the cast analysis by the particle method on the same condition as the experiment and it compared it with the result of Mampaey and

Xu (1995)[4]. We reports on the results of two models shown in Figure 2 (called curved gating system and stair like gating system).

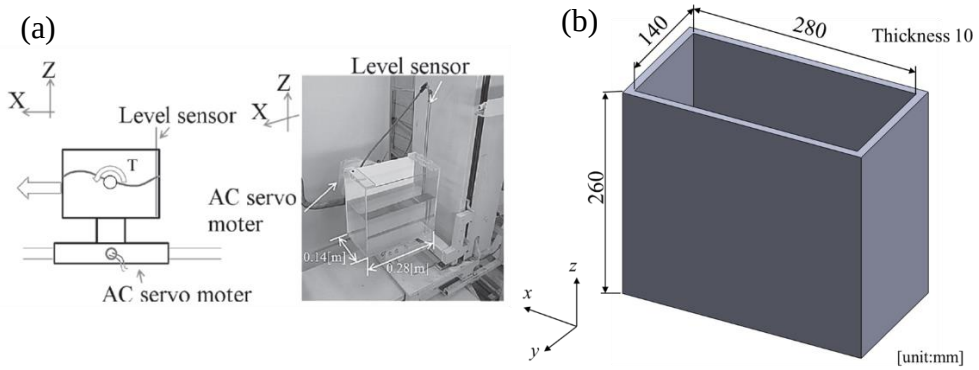


Figure 1. (a) Experimental equipment; (b) size of the container

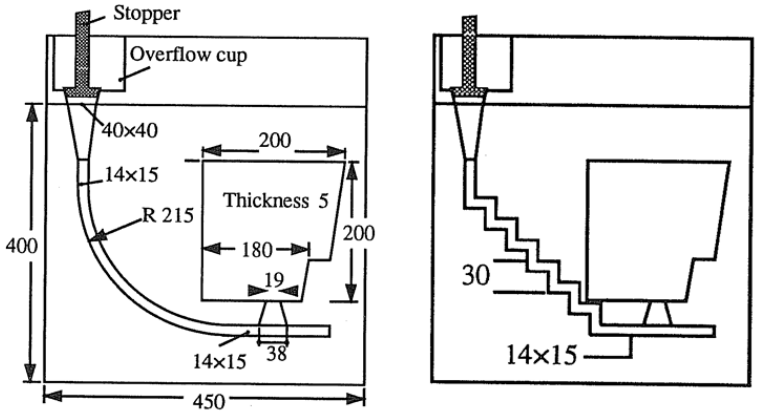


Figure 2. The plane casting systems

Numerical method

We use a form of the equation of continuity and the momentum equation of SPH method as follows (Suwa, Nakagawa, and Murakami (2013)[5]):

$$\frac{D\rho_a}{Dt} = \sum_b m_b (v_a{}^i - v_b{}^i) \frac{\partial W(\mathbf{r}_{ab}, h)}{\partial x^i} \tag{1}$$

$$\frac{D\vec{v}_a}{Dt} = -\sum_b m_b \left(\frac{p_b + p_a}{\rho_b \rho_a} \right) \frac{\partial W(\mathbf{r}_{ab}, h)}{\partial \vec{x}_a} + \vec{g} \tag{2}$$

Here, ρ, t, m, v, p, h, and $\mathbf{r}_{ab}$ are the density, the time, the mass, the velocity, the pressure, the smoothing length, and the relational position vector between the particles a and b, respectively. The constant vector $\vec{g}$ is the gravitational acceleration. The subscripts a and b are indices of the particles, and the sum is over all particles b within a radius $2h$ from the particle a. As a kernel function $W(\mathbf{r}_{ab}, h)$, we employ the quintic spline.

$$W(\mathbf{r}_{ab}, h) = \begin{cases} \alpha_d \left(1 - \frac{q}{2}\right)^4 (2q+1) & (0 \le q \le 2) \\ 0 & (2 < q) \end{cases} \tag{3}$$

Where $q=|\mathbf{r}_{ab}|/h$ and α_d, a normalization constant, is $21/(16\pi h^3)$ in 3-D.

The fluid in our SPH formalism is treated as weakly compressible. The pressure is given by the following equation of state:

$$p = \rho_0 c_s^2 \left(\frac{\rho}{\rho_0} - 1 \right) \tag{4}$$

where ρ_0 and c_s are initial density and sound speed, respectively. In this study, ρ_0 and c_s are set to 1 kg m^{-3} and 50 m s^{-1}, respectively. The sound speed c_s is a numerical parameter, and a value is chosen to be extent where the density of the SPH particle with a typical kinetic energy does not come off from a standard density greatly.

Numerical results

Transportation of liquid container

The transportation of liquid container without the control and that with the control were analyzed as well as the experiment by Okatsuka et al. (2011) and Shibuya et al. (2013) [2][3]. Figure 3 shows transferring input without control. Figure 4(a) shows time series of the level of liquid under this transportation condition. The result of the SPH simulation and the result of the experiment are indicated as the blue solid line and the red broken line, respectively. The measurement point of the water level is a position of level sensor shown in Figure 1. It is shown that the shape of the first wave, which is the most important about transportation, immediately after the completion of the movement (4.1 second) is corresponding to the experiment very well. The wave attenuation occurs in the calculation since the second wave. The attenuation shows the tendency to become small when the resolution rises.

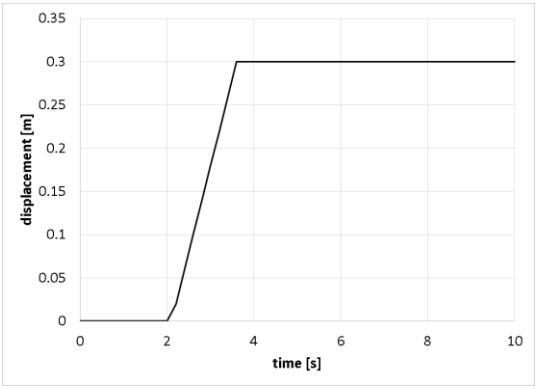


Figure 3. Transferring input without control

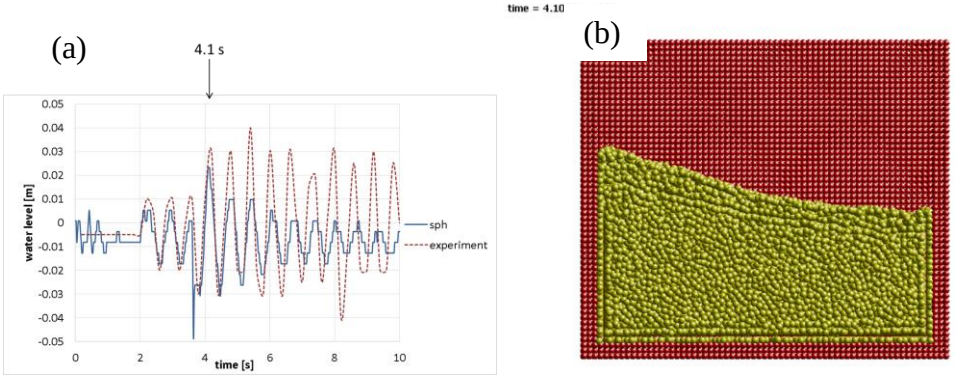


Figure 4. (a) Level of the liquid surface without control
(b) Snapshot of the liquid container without control (4.1 sec.)

Figure 5 shows transferring input with control. Figure 6(a) shows time series of the level of liquid under this transportation condition. As well as Figure 4(a), the result of the SPH simulation and the result of the experiment are indicated as the blue solid line and the red broken line, respectively. We can see that the vibration of the liquid surface becomes gentle by adding the control while the wave falls into disorder without the control of the transportation speed, and the effect is reproduced by the SPH simulation.

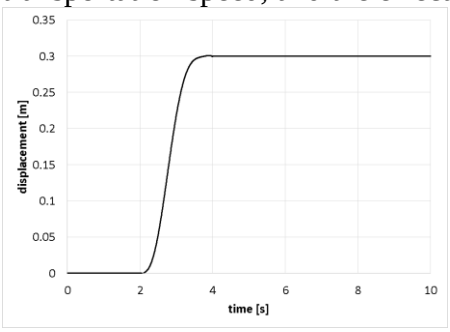


Figure 5. Transferring input with control

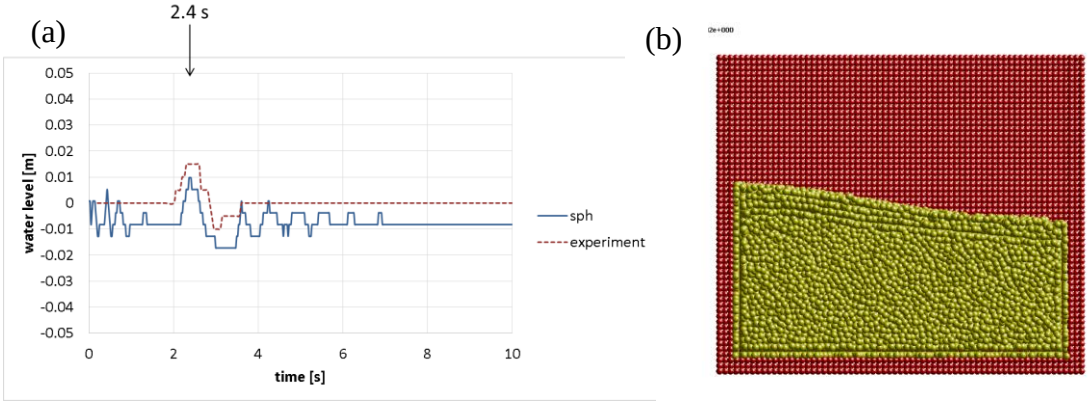


Figure 6. (a) Level of the liquid surface with control
(b) Snapshot of the liquid container with control (2.4 sec.)

Figure 7 (a) and (b) show transferring input and tilt input, respectively. The tilt input is controlled. Figure 8 is a water level history when tilting motion is input with transportation. In the situation in which the energy of the water vibration is supplied as tilting motion, the vibration of the liquid surface is kept being corresponding well since the second wave.

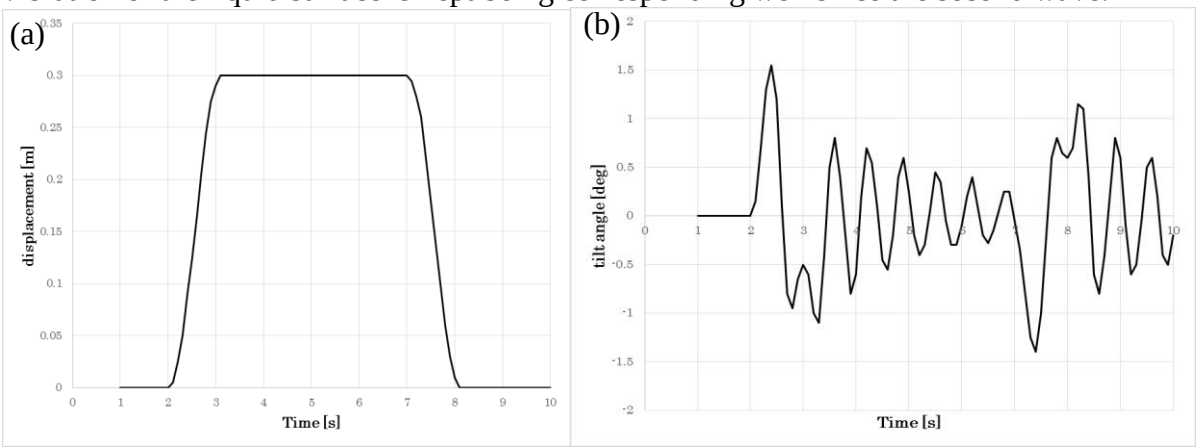


Figure 7. Transferring and tilt input with tilt control

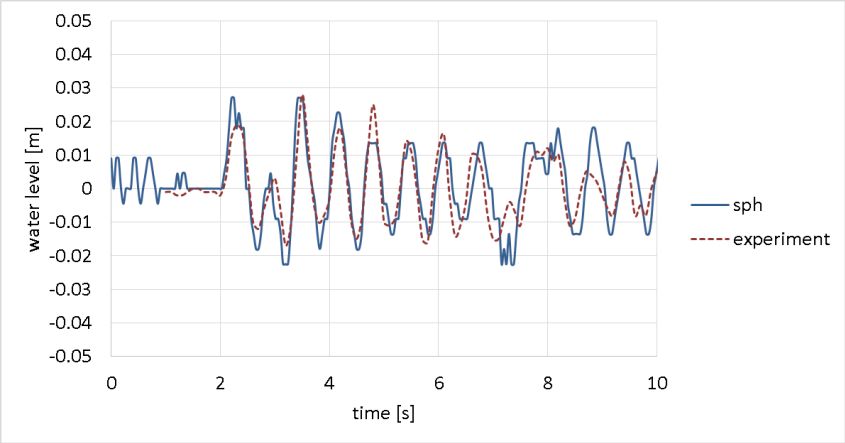


Figure 8. Height of the liquid surface with designed transfer and tilt input

Mold filling processes

The mold filling processes were analyzed with the same condition of the experiments by [4]. We have compared the results of our SPH simulations with the results of experiments and results of a CAE software employing Eulerian algorithm with volume of fluid (VOF) method as surface treatment. In SPH simulation, we put the fluid of rectangular shape (with side-lengths 6cm, 6cm, and 10cm) above the overflow cup, and the fluid has been fallen freely. In grid-based simulation, the inflow condition of the constant rate was given to the inlet. The inlet velocity was calculated to match the filling time of the experimental result (17.707 cm s^{-1} for curved gate model, and 5.646 cm s^{-1} for stair like gate model).

In Figure 8, mold filling sequences of the casting with a curved gating system are shown. The first, second, and third line indicate the results of the experiment, SPH simulation, and grid-based simulation, respectively. Any of the experiment and two simulations show that the velocity of molten metal increase at the time of 0.67s (point that the filling of the runner is completed), and we can see that the molten metal get to the top of the container. It is shown like this that a qualitative tendency is the corresponding between the simulations and the experiment.

In Figure 9, mold filling sequences of the casting with a stair like gating system are shown. As well as Figure 8, the first, second, and third line indicate the results of the experiment, SPH simulation, and grid-based simulation, respectively. It seems that the appearance to which the filling gradually progresses from the lower side is corresponding well by the simulation result and the experimental result.

As the tendency between the models, the velocity of fluid into the cavity of the curved gate model is faster than that of the stair like gate model, in which the energy loss has happened when filling it. This has been achieved by regulating the inlet velocity of the fluid in the grid base simulation. In other words, after the experimental results are known, it is necessary to adjust the input condition. On the other hand, the difference of the behavior of both models appears clearly as a result of analyzing the fluid behavior while giving the same inflow condition in the SPH simulation. This mold filling models are known well as a verification of molten metal flow analysis that has a free surface, and it is very difficult to adjust both the curved gate model and the stair like gate model. The results of the molten metal flow behavior show advantage of SPH method to reproduce a free surface behavior.

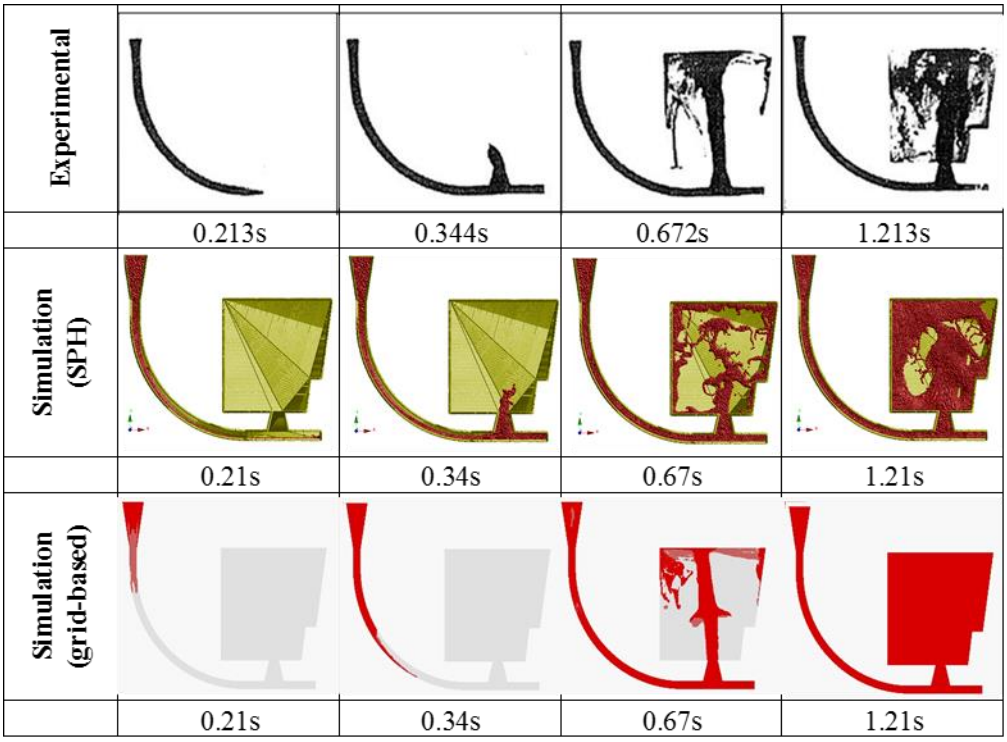


Figure 8. Mold filling sequences of the casting with a curved gating system

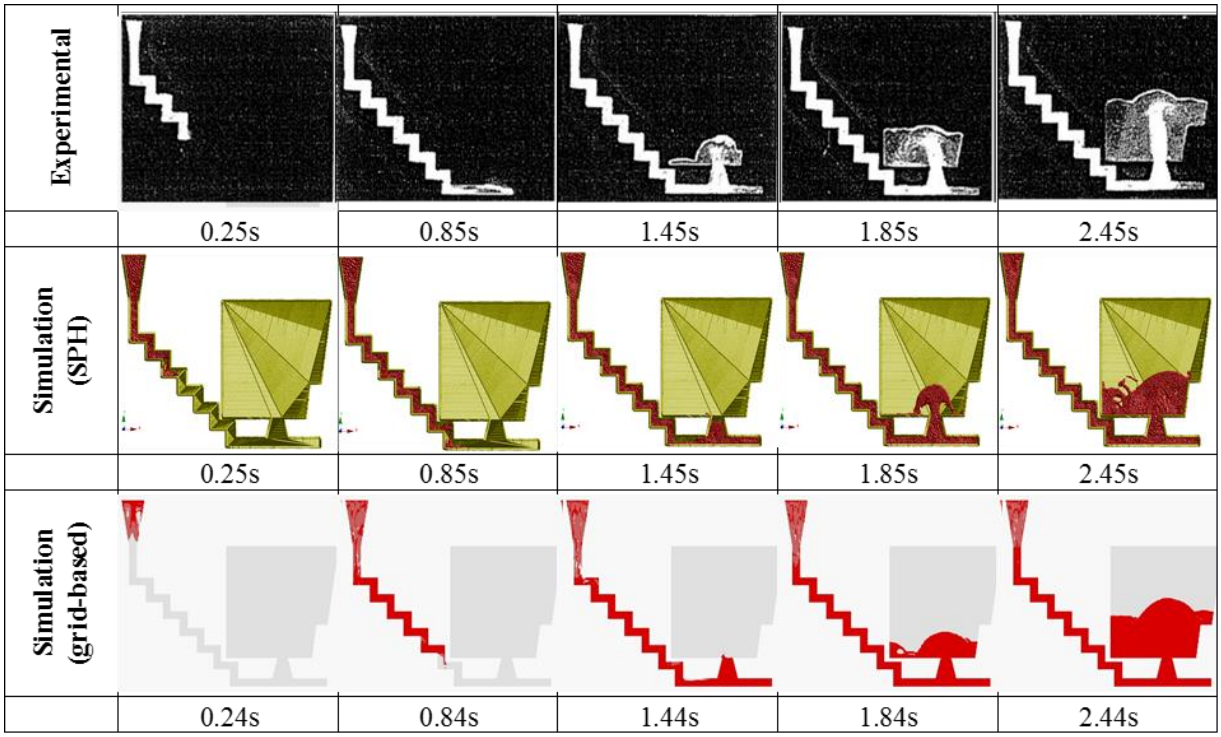


Figure 9. Mold filling sequences of the casting with a stair like gating system

Conclusions

The SPH method was applied to the casting process (transportation of liquid container and mold filling), and we compared them with the experimental results. In the analysis of transportation, it was shown that the result of the particle method analysis reproduced the experiment well as the liquid surface oscillation are suppressed by the presence of the

sloshing control. This depends on the treatment of a solid wall, which is difficult to be taken in Eulerian analysis technics. The particle method analysis can be expected to be applied to the optimization of the carrier control. Moreover, the difference of the tendency to fill the cavity slowly in stair like gate model while spouting under the cavity in great force in curved gate model was reproduced well in the analysis of the mold filling process. It is thought that the SPH method has an enough analytic performance for the region where the solid wall moves and where the dispersion of the molten metal occur.

In the cast field, there are various processes and phenomena, and a lot of objects that should be examined of the propriety of the application of simulations exist. The processes with moving boundaries such as ladle pouring of die casting, sleeve movement, and local pressurizing process are expected to show the advantage to the particle method. We would like to examine the application of the particle method simulation to these processes in the future. It is also needed to show the applicability to heat flow and the solidification behavior that are the indispensable function as cast CAE software.

References

[1] Cleary, P.W. (2010) Extension of SPH to predict feeding, freezing and defect creation in low pressure die casting, *Applied Mathematical Modelling* **34**, 3189–3201

[2] Okatsuka, H., Shibuya, R., Noda, Y., Matsuo, Y., and Terashima, K. (2011) Sloshing Suppression Control by Using Model Predictive Control in Liquid Container Transfer System, *Transactions of the Japan Society of Mechanical Engineers Series C* **77**, 4068-4080

[3] Shibuya, R., Okatsuka, H., Noda, and Y., Terashima, K. (2013) Sloshing Suppression Control with Designed Transfer and Tilt Input by Using Generalized Predictive Method, *Transactions of the Society of Instrument and Control Engineers* **49**, 134-141

[4] Mampaey, F., and Xu, Z.A. (1995) Simulation and experimental validation of mould filling, *Modelling of Casting, Welding and Advanced Solidification Processes VII*, 3-14

[5] Suwa, T., Nakagawa, T., and Murakami, K. (2013) : A Study of the Wave Transformation Passing over an Artificial Reef using SPH Method, *Journal of computational science and technology* **7**, No. 2, 126-133.

Newtonian Gravitational Force for predicting Distribution Centre Location of a Supply Chain Network

***A.A.G. Akanmu[1] and F.Z. Wang[2]**

[1]School of Computing, University of Kent at Medway, UK
[2]School of Computing, University of Kent at Canterbury, UK

*†Presenting/corresponding author: aagakanmu@gmail.com

Abstract
Occasions do arise when researchers and industrialists alike are faced with the decision of where to cite new structures (shops, stores, distribution centers etc) in order to benefit the consumers and the business entity as well. Such decisions might take the importance of vertices and/or edges of a network (e.g. Supply Chain Network) into consideration. In particular, the strength of the vertices and those of the edges play an important role in arriving at such decisions. In this paper, as against the most common and traditional measures of centralities, that is - Degree, Closeness, Betweenness and Eigen-Vector centralities, a new centrality measure, Top Eigen-Vector Weighted Centrality (TEVWC) which takes into consideration the clique structure of a network and the strengths attached to the vertices/edges of the network, was used to predict the location of a distribution center in a supply chain management. The accuracy of prediction on a sample dataset of supply chain network, using the TEVWC was found to be 94.6%, which is 10.6% higher than the result outcome from the method of Newtonian Gravitational Force when driving distances are considered, but with the earth distances the accuracy obtained is 99%.

Keywords: Cliques, Centre of Mass, Link-weights, Node-weights, Network Centrality, Supply Chain Network

Introduction

Any network consists of nodes and links; the nodes are also severally referred to as vertices, actors and points, while the links are also often referred to as edges, arcs, and ties. Different meanings have been adduced to the weighted-ness of a network, so many literature have at instances made references to link-weights as the weights of the entire network, even though any network as described above would at least consist of node(s) and link(s) as the case may be. This therefore implies that there has to be node-weights as separated from link-weights and the combination of the two would thereby emerge as weights of any typical network.
In his work on identifying cohesive subgroups [1] laid emphasis on the link of a graph thus "Further, the definitions based on path length are restrictive in that they specify the nature of the relationship between each pair of actors within a subgroup instead of a general relationship between each actor and all others in the subgroup", thereby leaving out the actors/nodes' strength. According to the definition of the Topological Centrality (TC) of an

edge, the weights of edges are the sum of the weights of its two end nodes [2]. Here, the definitions of the weights of edges and weights of nodes are somehow fuzzy, as it is not clear cut what made up the weights of the end nodes.
[3] defined a weighted network as that in which ties are not just either present or absent, but have some form of weight attached to them, hence the emphasis of his paper on the trade-off between the weight on the tie and the number of ties. This was however silent on the attributes of the node (which in most cases form the weights on the nodes). This viewpoint was partly shared by [8] when he said "Second category of measures (i.e., h-Degree, a-Degree and g-Degree) takes into account the links' weights of a node in a weighted network. Third category of measures (i.e., Hw-Degree, Aw-Degree and Gw-Degree) combines both neighbors' degree and their links' weight."
[4] [5] [6] have also attempted to generalize the traditional network centrality measures (degree, betweenness and closeness) to weighted networks, but they were only able to implement their generalisations as the link-weighted network, thus not putting the node-weights into consideration.

Another emphasis on link-weighted-ness in terms of duration is that by [7] whereby they introduced a time-variant approach to the degree centrality measure, that is, the time scale degree centrality (TSDC), whereby the presence and duration of links between actors are considered while leaving out the node attributes. On hybrid centrality measures, [9] reported having considered a network as having the centrality measures of each node as the attribute of the node, while [10] in their analysis of results for scholars performance and social capital measures also buttressed this view point by submitting that repeated co-authorships are merged by increasing more weight(tie strength) to their link(tie) for each relation, so also [11] whereby they referred to weight of undirected graph as the link-weight. However, all these arguments are again centred on link-weights as against the weights of the network that could have considered a combination or mergers of node-weights and link-weights.
In their new method of constructing co-authorship, [12] used the times of co-authorship to calculate the distance between each pair of authors, and to also evaluate the importance of their cooperation to each other with the law of gravity. This relies again on the use of link weights.
The mixed-mean centrality measure of [13] took into consideration, the number of links, link-weights and node-weights in their study of co-authorship network, while [14] used the clique structure and node-weighted centrality to predict the distribution centre location in a supply chain management, thus clarifying what the link-weights and node-weights actually represent in a weighted network.
It is still largely unknown how newtonian gravitational force of attraction and the top eigen-vector weighted centrality can be applied to predict location of structures in a network. Thus, it is important to still find out whether the attributes of the nodes in any network is of importance or not; one might also want to know how accurate the mergers of node-weights and link-weights can be in terms of prediction of where to cite structures (for example, where to cite a distribution centre); and finally how accurate would the prediction of the location for a DC become, given a new centrality measure, which takes into consideration, the clique structure of a network combined with the node-weights and link-weights of the network.
The nodes of the clique for each of the cities considered are ranked in line with their eigen-vectors, and the representative node (the highest ranking node) for that clique becomes the representative node of that city. The centre of mass for the emergent nodes is thereafter taken into consideration. This method is important in that it only takes the node-weights and link-weights into consideration while trying to achieve the results, thereby saving other resources.

Section II discusses the link-weighted centrality and node-weighted centrality and the third section discusses methods employed in this paper and their implementation, while the fourth lays out the output results from the methodology and the last forms the conclusion.

Weighted Centralities

Link-Weighted Centrality

The equation (1) below represents the weighted degree centrality with respect to the edges or links.

$$S_p = C_D^W(p) = \frac{\sum_q^N w_{pq}}{n-1} \qquad (1)$$

Where C_D^W represents the weighted degree centrality; p is the focal node ; q= adjacent node ; w= weight attached to the edge ; and n= total number of nodes in the graph. This reasoning can be extended to the weighted centrality of the Closeness, Betweenness and the Eigenvector. As an example, the weighted eigenvector centrality could be seen as

$$\lambda \mathbf{x} = \mathbf{A}^{\mathbf{w}}\mathbf{x} \qquad (2)$$

where $\mathbf{A}^{\mathbf{w}}$ is a square matrix of the weights on the edges of **A** and **x** is an eigenvector of **A** .

A tuning parameter α was introduced to determine the relative importance of the number of ties compared to the weights on the ties by [3]. Equation (3) below thereby represents the product of degree of a focal node and the average weight to these nodes as adjusted by the introduced tuning parameter. So, for weighted degree centrality at α we have:

$$c_d^{w\alpha}(p) = k_p \times (\frac{s_p}{k_p})^{\alpha} = k_p^{(1-\alpha)} \times s_p^{\alpha} \quad (3)$$

where k_p = degree of nodes

$S_p = c_D^w(p)$ as defined in (1) above , and α is ≥ 0

This argument could also equally be applied to the closeness centrality; betweenness centrality and eigenvector centrality.

Node-Weighted Centrality

As an extension to equation (3), a tuning parameter β was introduced by [13] to include the weightedness on the nodes, therefore, for weighted degree centrality at α *and* β we shall now have

$$c_d^{wab}(i) = k_i \cdot (\frac{s_i}{k_i})^a = k_i^{(1-a)} \cdot s_i^a \cdot z_i^b \qquad (4)$$

where k_i = degree of nodes

$s_i = c_D^w(s)$ as defined in (1)

z_i = weight of nodes, where $\alpha \geq 0$; $\{\beta \in Z : -1 \leq \beta \leq 1\}$

The value β depends on whether the weight is having positive or negative effect on the centrality measure, if for instance the weight is having a positive effect (e.g. profit) β is +1

else it is -1 (i.e. loss). Values of α *ranging from ¼ to 1¾ is used in order to vary the effect of* α, i.e. values less than 1 and those greater than 1.

Top-Eigen Weighted Vector Centrality and Newtonian Gravitational Force

The node-weights of the sample used for this study is the sales value while the edges are the driving distances between the shops in the sampled area. The sampled shops here are maximally connected as all of them have road links. Hence, we take the advantage of the clique structure by making the most central node (the one with highest centrality) from each clique to be a representative of that clique. By that, we have a representative node each from the two cliques considered for the purpose of the prediction of a proposed DC (see figure 1 below).

In the county of Scotland, two major cities with higher concentration of shops were chosen for our sample, the city of Glasgow and Edinburgh. In each of the cities, the ranking of the nodes(i.e. shops) based on eigen-vector centrality were considered, tested for all the four centralities (degree, closeness, betweeness and eigen-vector), thereafter, the highest ranking node called the top eigen-vector weight was made to be representative of that city (see Table I). The driving distances apart of each of the representative cliques for Glasgow and Edinburgh were obtained from google MAPI. UCINET , tnet and Excel software are used for obtaining the centralities and doing the final calculations (see Figure2).

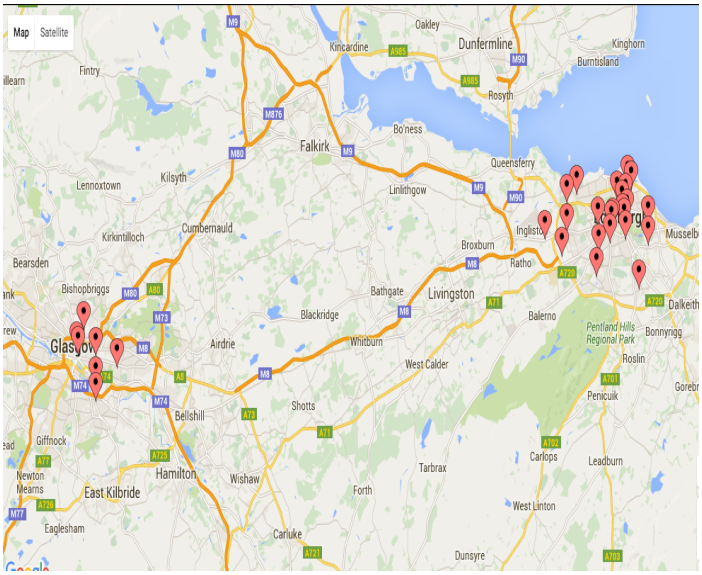


Figure.1. Figure showing the two cliques of Scotland shops (Glasgow on the left and Edinburgh on the right)

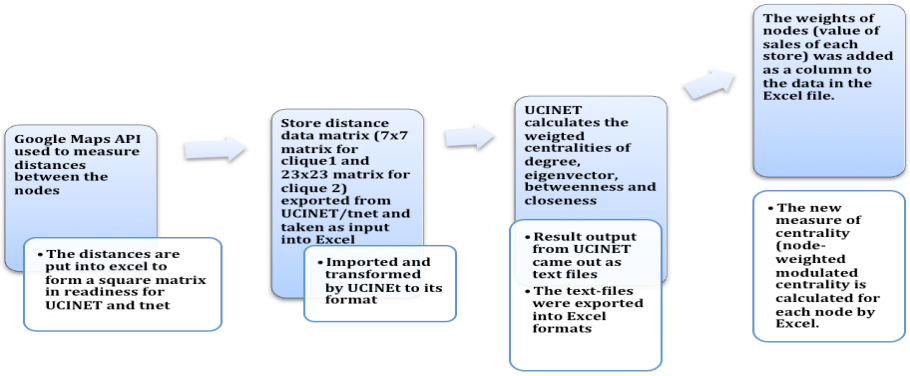


Figure 2. Figure showing the implementation of top-eigen vector weighted centrality measure to the cliques of Scotland

The newton gravitational force was later introduced after the implementation of the Top eigen-vector weighted centrality, and it is later explained with equation (5).

Top-Eigen Weighted Vector Centrality

Node 22 with postcode EH12 7UQ being the highest ranking always, was chosen as the representative of the clique from Edinburgh when the Top eigen-vector weighted centrality is used. Similar procedure was carried out for Glasgow clique and Node 5 with postcode G21 1YL came out being the representative of that clique.(Figure 3 and Figure 4)

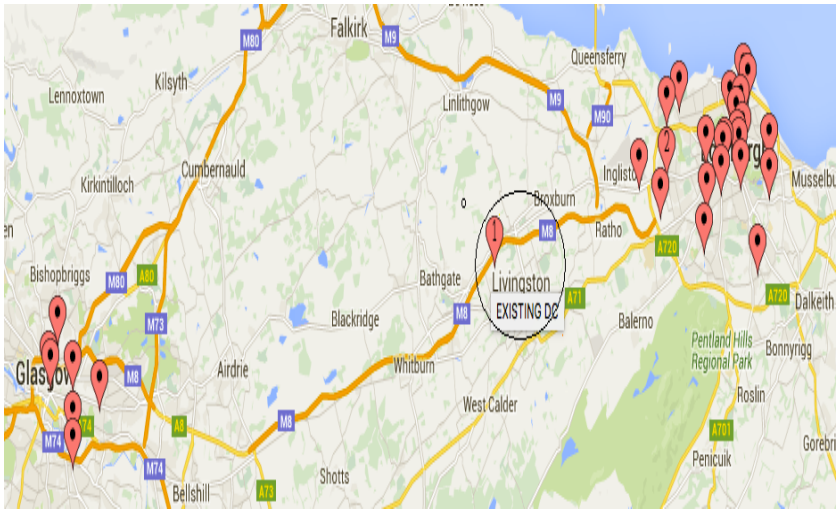


Figure.3 Figure showing the Existing DC at Livingston(encircled) and the clique representative node at Edinburgh marked "2".

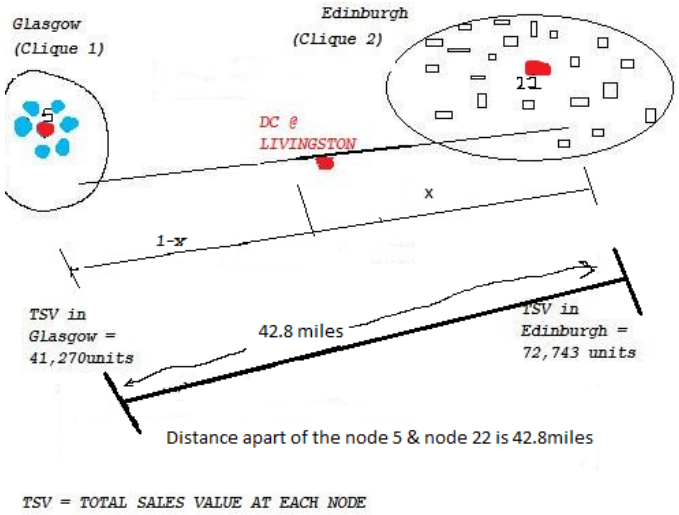


Figure 4. Figure showing the representative cliques at Scotland cities of Glasgowand Edinburgh

From Figure.4 above, let *x* be the proportional distance to the predicted Distribution Centre, and since the driving distance between node 5 (representing Glasgow clique) and node 22 (representing Edinburgh clique) is 42.8miles, by proportion

$$1\text{-}x \,/\, x = 72743/41270,$$
then $x = 0.36$ (i.e. 36% of 42.8) which is 15.4miles

If x is some 15.4miles away from the Edinburgh clique representative, and the existing DC is 13.1 miles away from node 22, the difference of the predicted DC will be 2.3miles away from the existing DC, hence,
the error rate of the predicted DC = (2.3/42.80) x 100 = 5.37% i.e. the percentage accuracy of the prediction = 94.63%

Newtonian Gravitational Force
This method is fashioned after the Newton's gravitational law which ascerts that every object's mass will ascertain some amount of force on any neighboring object, no matter the distance between them. The formula is:

$$F = k * \frac{(m * M)}{R^2} \qquad (5)$$

where
F = Gravitational Force
k = constant
m =the mass of the first object
M=the mass of the second object
R=Distance between the two objects (it can be driving distance or the earth distance)

Table 1. Table showing the clique result according to Top Eigen-Vector Weighted Centrality of selection from Edinburgh and Glasgow

	NODE-WEIGHTED EIGEN-VECTOR CENTRALITY FOR EDINBURGH							NODE-WEIGHTED EIGEN-VECTOR CENTRALITY FOR GLASGOW					
ODE	α= ¼	α= ½	α= ¾	α= 1¼	α= 1½	α= 1¾	NODE	α= ¼	α= ½	α= ¾	α= 1¼	α= 1½	α= 1¾
22	453097.5	510928.4	576140.5	732596.9	826101.5	931540.6	1	51495.5	83776.9	136294.6	360734.7	586870.9	954766.6
23	33870.3	33585.7	33303.6	32746.4	32471.3	32198.5	3	111193.7	183949.4	304310.2	832822.6	1377750.4	2279232.3
24	63165.7	70052.7	77690.6	95555.4	105973.8	117528.2	4	73621.3	130901.3	232747.0	735808.2	1308292.6	2326189.8
25	78587.1	77797.5	77015.8	75475.8	74717.4	73966.6	5	145086.8	264307.3	481493.6	1597913.3	2910948.7	5302930.0
30	152807.8	154952.0	157126.2	161566.5	163833.5	166132.3	11	41543.2	75860.3	138525.5	461912.0	843478.6	1540241.6
31	328540.4	360246.8	395013.1	474935.1	520769.5	571027.4	13	4465.1	7280.7	11871.6	31563.2	51465.8	83918.1
32	297308.6	346383.0	403557.8	547777.5	638195.0	743537.0	15	3662.5	6052.2	10001.1	27309.7	45128.5	74573.5
33	165476.9	173857.7	182662.9	201633.7	211845.6	222574.7							
35	3412.6	3404.2	3395.9	3379.3	3371.0	3362.7							
36	21128.7	21025.6	20923.1	20719.5	20618.5	20518.0							
37	5285.0	6115.7	7076.9	9476.3	10965.7	12689.2							
38	8971.3	9984.7	11112.6	13764.9	15319.7	17050.2							
39	7943.9	7919.5	7895.2	7846.8	7822.7	7798.7							
40	4531.7	5270.7	6130.4	8293.1	9645.7	11218.9							
41	28334.9	28259.3	28183.9	28033.6	27958.7	27884.1							
42	11302.1	11272.0	11242.1	11182.4	11152.7	11123.0							
43	2201.3	2631.5	3145.8	4495.6	5374.3	6424.6							
44	9406.4	9964.9	10556.6	11847.4	12550.8	13296.0							
45	4698.1	5294.3	5966.3	7576.8	8538.3	9622.0							
46	17504.4	17952.4	18411.8	19366.2	19861.8	20370.1							
47	4229.0	4130.8	4034.8	3849.6	3760.1	3672.8							
48	22823.0	23793.4	24805.0	26959.2	28105.4	29300.4							
49	17663.5	17373.2	17087.6	16530.6	16258.9	15991.7							

In case of the objects, which in this case are the 30 shops of Scotland (consisting of seven shops from Glasgow and 23 shops from Edinburgh) as shown in Figure 3 above. The shops have pull effects on the DC at Livingston, as such the vectorial resultant force F of each node(shop) is calculated using the earth distances apart and the driving distances apart.

Earth Distance with 30shops/nodes
When the representative clique (EH12 7UQ i.e. Node 22) was used as origin (leaving 29 shops for consideration) as shown in Figure 5 below, the total force is 314.53units but when the actual DC for Scotland (EH54 8QW) was used as origin (as in Figure 3) for all 30shops the total force was 12.28units.

Figure 5. Figure showing the representative clique of Edinburgh herein marked "1" with other shops in Scotland

In the Figure 5. above, the point marked "1" is the representative clique (node 22) EH12 7UQ. This node is used as the origin for the other 29nodes in the region of Glasgow and Edinburgh, which is, excluding the existing DC (EH54 8QW) at Livingston.

To make things clearer, the figure below shows the existing DC – EH54 8QW (at Livingston) as "1" while "2" represents the predicted DC – EH12 7UQ (at Edinburgh)

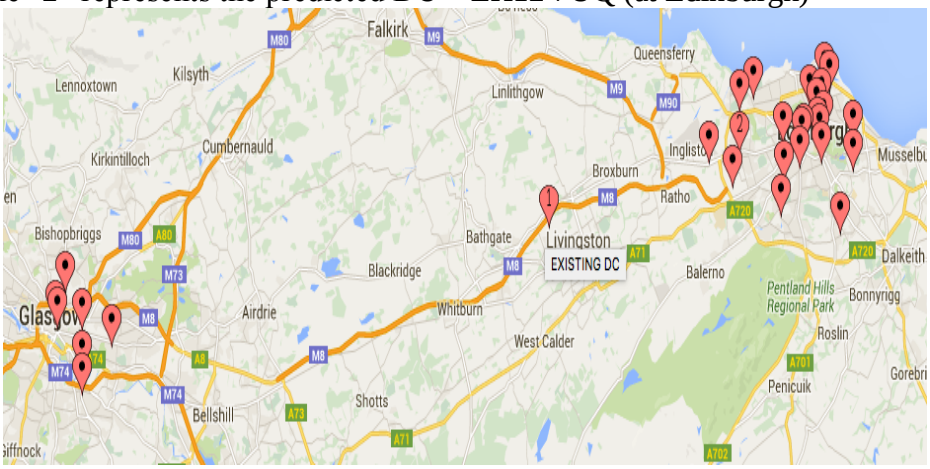


Figure 6. Figure showing the Existing DC at Livingston and the representative clique at Edinburgh

Driving Distance with 30shops/nodes

For the driving distance, the total force for the DC as origin is 1,394,170.15 while the representative clique as origin yielded 29,690,905.18 . The table 2 below summarises the findings of the resultant forces when each of the driving distances and earth distances is used in the calculations.

Table2. Table showing the total force for Earth and Driving forces for Scotland shops

S/No	TYPE OF DISTANCE	EXISTING DC	PREDICTED DC
1	EARTH DISTANCE	1.23 E01	6.0 E01
2	DRIVING DISTANCE	1.39 E06	4.76 E06

Scotland with 7 shops/nodes at Glasgow, 23shops/nodes at Edinburgh and three additional shops

We consider three additional shops which are outliers , that is, not within Glasgow and Edinburgh but within an increased coverage radius of 36miles against the previous 30miles radius. This means we now consider 33 shops as our sample instead of the previous 30 shops, these newly added shops are at South Queenferry, Hardington and Bathgate. With these additional three shops added from within Scotland but outside Glasgow and Edinburgh, we have the results in Figure 7 below:

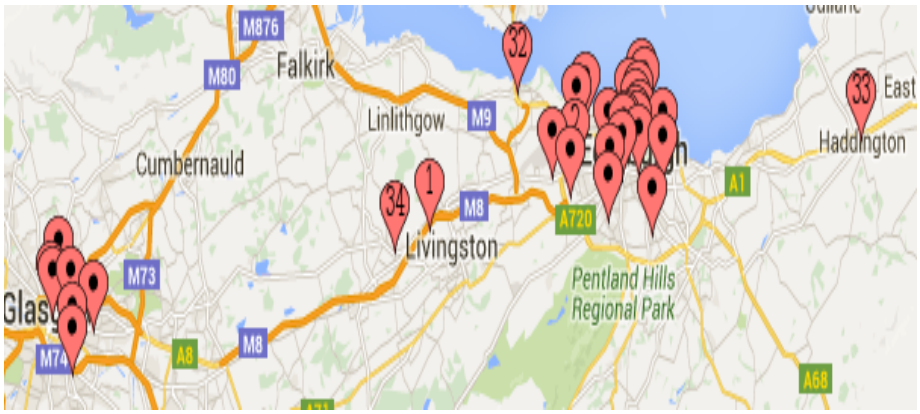


Figure 7. Figure showing newly added nodes 32, 33 & 34 outside Glasgow and Edinburgh

The details of the new shops/nodes are as shown in the table 3 below:

Table 3. Table showing details of the three new nodes added to the existing 30 nodes/shops

S/N	Node	Post Code	Dist to Existing DC	City	Sales Values	Lat	Long
1	32	EH30 9QZ	11.9	SOUTH QUEENSFERRY	7948	55.9828	3.3990
2	33	EH41 3LZ	36.4	HADDINGTON	9358	55.9571	2.7777
3	34	EH48 2ES	3.8	BATHGATE	13746	55.8936	3.6215

With the addition of the three new shops and using each one as the origin to the remaining 32 shops, the table 4 below compares the results with the existing DC and former representative clique node using centrality measures.

Table 4. Table showing the total force for Earth and Driving forces for Scotland with additional three shops as new origins

S/No	TYPE OF DISTANCE	EXISTING DC	PREDICTED DC	New Shop1 (EH30 9QZ) as Origin	New Shop2 (EH41 3LZ) as Origin	New Shop3 (EH48 2ES) as Origin
1	EARTH DISTANCE	1.98 E01	1.3 E02	2.8 E01	6.6 E01	4.4 E01
2	DRIVING DISTANCE	6.1 E06	2.3 E07	1.1 E07	1.3 E07	5.9 E06

Figure 8. Figure shows the newly predicted DC as against the earlier predicted node labeled 2

Newtonian Gravitational Force with 30 shops of Glasgow and Edinburgh
Using the Earth distance between the shops and the Existing Distribution Centre (EDC) as origin, we have the results in Table 5 below:

Table 5. Table showing the forces exerted by highest/lowest valued nodes while considering earth distance

	Glasgow			Edinburgh			
	Node	Post Code	Value of Force	Node	Post Code	Value of Force	Distance Apart of Nodes
Highest Value Nodes	Node5	G21 1YL	1.5890	Node22	EH12 7UQ	1.7703	42.2
Lowest Value Nodes	Node15	G1 1EJ	0.0447	Node43	EH12 9BH	0.0080	41.3

Using the driving distance between the shops and the Existing Distribution Centre (EDC) as origin, we have the results in Table 6 below:

Table 6. Table showing the forces exerted by highest/lowest valued nodes while considering driving distance

	Glasgow			Edinburgh			
	Node	Post Code	Value of Force	Node	Post Code	Value of Force	Distance Apart of Nodes
Highest Value Nodes	Node5	G21 1YL	57,189.34	Node8	EH12 7UQ	404,474.18	2.2
Lowest Value Nodes	Node15	G1 1EJ	1,549.52	Node45	EH8 7NG	1,081.94	53.7

Newtonian Centrifugal Force with 33 shops of Glasgow and Edinburgh
Using the driving distance between the shops and the Existing Distribution Centre (EDC) as origin, we have the results in Table 7 below:

Table 7. Table showing the forces exerted by highest/lowest valued nodes while considering driving distance

	Glasgow			Edinburgh			
	Node	Post Code	Value of Force	Node	Post Code	Value of Force	Distance Apart of Nodes
Highest Value Nodes	Node 5	G21 1YL	66,150.20	Node 52A	EH48 2ES	4,184,638.00	27.7
Lowest Value Nodes	Node 15	G1 1EJ	1,792.31	Node 45	EH8 7NG	1,251.47	53.7

Using the earth distance between the shops and the Existing Distribution Centre (EDC) as origin, we have the results in Table 8 below:

Table 8. Table showing the forces exerted by highest/lowest valued nodes while considering earth distance

	Glasgow			Edinburgh			
	Node	Post Code	Value of Force	Node	Post Code	Value of Force	Distance Apart of Nodes
Highest Value Nodes	Node5	G21 1YL	2.0218	Node22	EH12 7UQ	2.2525	42.2
Lowest Value Nodes	Node15	G1 1EJ	0.0568	Node43	EH12 9BH	0.0102	41.3

SUMMARY OF ACCURACY WITH THE SALES VALUES USED AS NODE-WEIGHTS

Table 9. Accuracy of results obtained for both earth/driving distances for 30 shops and 33 shops

	PERCENTAGE ACCURACY OF THE HIGHEST FORCE NODES FROM GLASGOW TO EDINBURGH	PERCENTAGE ACCURACY OF THE LOWEST FORCE NODES FROM GLASGOW TO EDINBURGH
EARTH DISTANCE WITH 30 SHOPS	64.9%	63.2%
EARTH DISTANCE WITH 33 SHOPS	99.1%	99%
DRIVING DISTANCE WITH 30 SHOPS	64.9%	79.9%
DRIVING DISTANCE WITH 33 SHOPS	63.5%	84%

Conclusions

The Newtonian Gravitational force provides a more accurate percentage of 4.4% more than when the TEVW centrality was applied. The set of input resources for this method are the node-weights and link-weights, even though there are other factors to consider in the citing of a distribution centre, this makes this method a cheaper one with high accuracy of prediction. The assumptions in this study is that the driving distances are taken to be a straight line in the model figures in this paper, whereas in reality this might not necessarily be so.

In future, the range of values for α might transcend the range of ¼ and 1¾ as some interesting outcomes might surface, also, the domain of application could still be further expanded to cover area such as bioinformatics whereby the visualisation and understanding of biology networks will make one to be able to predict the reaction of cells to pharmaceutical drugs due to their positioning in such a network. Healthcare is another area of consideration, as the study of the connections between hospitals, patients, doctors and healthworkers can help a lot in the prediction of where to cite new hospitals and even how to arrest or prevent epidemics. In terms of network security, a more central node is protected and given more attention in order to prevent or repel attacks from any form of intrusion.

It is clear that the node-weights (node attributes) actually count in any network as confirmed in this research whereby it forms the basis of prediction of a distribution centre with a higher accuracy while making use of the newtonian gravitational force as compared with the centrality measure – Top Eigen-Vector Weighted Centrality (TEVWC).

Acknowledgement

We wish to acknowledge Dr. Fred Yamoah and dunnhumby (www.dunnhumby.com) for providing us with the dataset used in this research.

References

[1] K.A. Frank (1995). Identifying cohesive subgroups. Social Networks 17 (1995) 27 – 56. N.H. Elsevier.

[2] H. Zhuge & J. Zhang(2010). Topological Centrality and It's e-Science Applications. Wiley Interscience. arXiv:0902.1911v1

[3]T. Opsahl, F. Agneessens & J. Skvoretz (2010). Node Centrality in Weighted Networks: Generalizing degree and shortest Paths. Social Networks 32(2010) 245-251. Elsevier B.V.

[4] A. Barrat, M. Barthelemy, R. Pastor-Satorras, & A. Vespignani (2004). The Architecture of Complex Weighted Networks. Proceedings of the National Academy of Sciences 101(11), 3747-3752. arXiv:cond-mat/0311416.

[5] U. Brandes (2001). A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology 25, 163-177.

[6] M.E.J. Newman (2001). Scientific Collaboration networks. II. Shortest paths, weighted networks, and centrality. The American Physical Society. Physical review E, Volume 64, 016132

[7] S. Uddin, L.Hossain, & R.T. Wigand(2013). New direction in degree-centrality measure: Towards a time-variant approach.. International Journal of Information Technology & Decision Making. World Scientific Publishing Company.

[8] A. Abbasi(2013). H-Type hybrid centrality measures for weighted networks. Scientometrics (2013) 96:633-640. DOI 10.1007/s11192-013-0959-y

[9] A.Abbasi, & L.Hossain (2013). Hybrid centrality measures for binary and weighted netwroks. In Complex networks (pp.1-7). Springer Berlin Heidelberg.

[10] A. Abbasi, R.T. Wigand, & L.Hossain.(2014). Measuring social capital through network analysis and its influence on

individual performance. Accessed from http://works.bepress.com/alireza_abbasi/21 on 03 Mar, 2015.
[11] P.B. Walker, S.G. Fooshee, & I. Davidson (2015). Complex interactions in social and event network analysis. In N. Agarwal
et al. (Eds.): SBP 2015, LNCS 9021, pp. 440-445. Springer International Publishing Switzerland.
[12] J. Liu et al (2015). A new method to construct co-author networks. Physica A 419 (2015) 29-39. Elsevier.
[13]G.A.A. Akanmu, F.Z. Wang & H. Chen(2012). Introducing weighted nodes to evaluate the cloud computing topology.
Journal of Software Engineering and Applications, 2012, 5, 961-969
[14]A.A.G. Akanmu, F.Z. Wang & A.F. Yamoah(2014). Weighted Marking, Clique Structure and Node-Weighted Centrality
Measures to Predict Distribution Centre's Location in a Supply Chain Management. International Journal of Advanced Computer Science and Applications. Vol.5, No. 12.

Assigning Material Properties to Finite Element Models of Bone: A New Approach Based on Dynamic Behavior

A. Ostadi Moghaddam [1], †M. J. Mahjoob[1], and A. Nazarian [2]

[1] School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2] Center for Advance Orthopaedic Studies, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA
†Corresponding author: mmahjoob@ut.ac.ir

Abstract

Finite element (FE) method is extensively employed to investigate the biomechanical behavior of bone structures. Material and morphological information of bone samples are typically provided by computed tomography (CT) scanning. Assuming that density and elasticity of bone are correlated, many studies have proposed different density-elasticity relationships to determine bone elastic constants. Herein, an innovative method for determining a single mathematical relationship between bone density and elasticity is proposed. Density distribution and morphology of a bovine bone were obtained from CT images, and the natural frequencies were measured using experimental modal analysis. The relationship between density and elasticity has a standard mathematical form with variable constants. Genetic algorithm (GA) was used to obtain the constants by minimizing the discrepancy between experimental and FE results. The relationship was then used in material properties assignment process of FE modeling and proved to be valid by predicting the natural frequencies of bone in different boundary conditions (BCs).

Keywords: Density-elasticity relationship, Modal analysis, Bone tissue, Computed tomography, Finite element method

1. Introduction

Accurate subject-specific finite element models of bone are of great importance in many state of the art research and clinical applications. FE analysis of bone provides valuable information about strain and stress fields within the tissue. Results can be used in fracture risk assessment, designing prosthetic implants and other clinical applications. Dynamic behavior and characteristics of bone such as natural frequencies, mode shapes and response to dynamic loads can also be determined by FE analysis. Obtaining the fundamental frequencies of bone, in particular, has numerous practical applications in medicine and bioengineering. It has been shown that loads with frequencies close to natural frequencies of bone can enhance bone apposition [1]. In fact, patient-specific natural frequencies of targeted bones would help physicians to optimize vibration therapies and exercise regimens and find a solution which suits the patient best [2]. In addition, resonance frequencies and mode shapes of bone provide valuable information about density-elasticity relationships [3] and orthotropic properties of long bones [4].

To generate a subject-specific model, geometry and material properties of bone are usually derived from computed tomography images. The CT images are processed to create three-dimensional (3D) geometry of bone segments. Mechanical properties of bone can also be derived from CT data using mathematical relationships, which relate CT values to material properties [5]–[9]. It has been demonstrated that the relationship between CT numbers and apparent density of bone tissue is approximately linear [10]–[12]. However, obtaining an accurate relationship between density and mechanical properties of bone, particularly elasticity,

is more challenging. Accurate determination of these relationships is important for developing precise FE models.

The relationship between Young's modulus and bone density is described by many different empirical models in the literature [13]–[22]. This relationship is generally reported in power or linear form. The complexity of experimental techniques involved in measuring mechanical properties of an anisotropic and porous material can explain the disparity in predicted values of Young's modulus in different studies. To determine the stiffness, commonly a bone specimen is loaded in a load frame. During the mechanical test, different types of error can arise which makes it difficult to obtain bone stiffness. Methods of measuring bone deformation are widely discussed in the literature [8].

To overcome the difficulties in traditional mechanical testing and improve the accuracy of the density-elasticity relationship, we have developed a new method which determines the model parameters in the general form of the density-elasticity relationship based on the results of experimental modal analysis using GA and FE methods. Unlike many reported models in the literature, this method leads to a single density-elasticity equation which is valid for all ranges of bone density.

2. Materials and methods

2.1 Experimental determination of natural frequencies

Modal analysis is a successful method to validate FE models of bone and to determine bone elastic constants [23] Simple experimental equipment, reasonably short measurement time and accuracy of measurements make modal analysis a potentially useful method for obtaining material properties.

Natural frequencies of a fresh-frozen bovine femur bone were obtained using impact hammer and shaker tests in free-free and clamped-free boundary conditions. To simulate the free-free BCs, soft elastic straps were used to suspend the sample.

The experimental setup of the shaker test is presented in Fig. 2. Computer generated random wave signals containing frequencies from 0 to 5000 Hz were used to excite the bone. Signals were amplified by a signal amplifier. Excitation and response signals were detected by accelerometers (DJB A/120/VT, DJB Co., France).

The experimental setup of the hammer test is presented in Fig. 3. The bone is excited by hitting an impact hammer equipped with a force transducer to five different points normal to the surface to excite different modes of vibration. An accelerometer is used to detect the bone response. The tests were then repeated for different positions of accelerometer. Charge amplifiers are used to condition the force and acceleration signals.

Applying a fast Fourier transform (FFT) algorithm, the frequency response of bone was analyzed considering the excitation and response signals. The resonance frequencies of different vibration modes were obtained using frequency response curves.

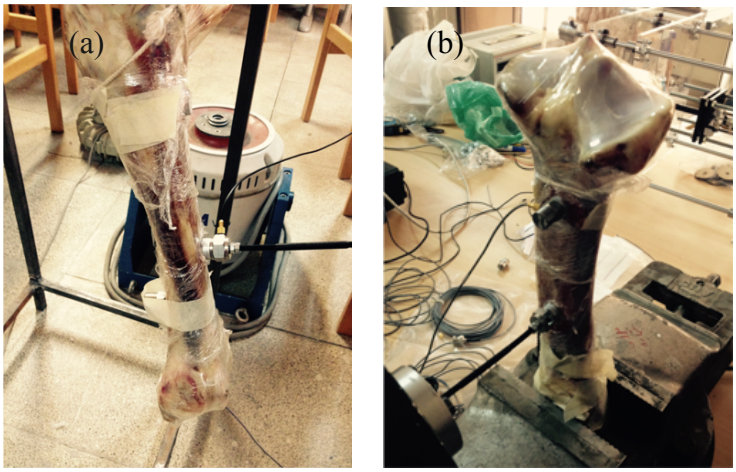


Figure 1. Shaker test setup; (a) free-free (b) clamped-free BCs

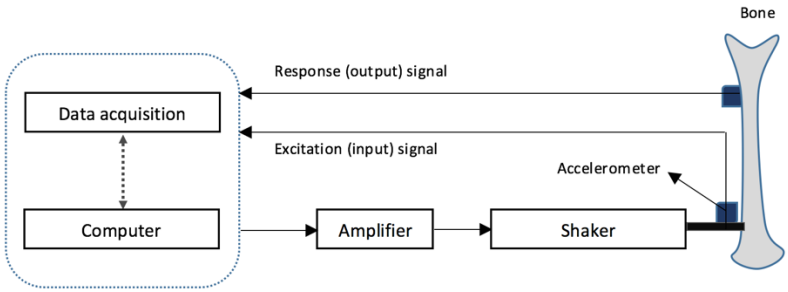


Figure 2. Measuring vibration response of bone using shaker

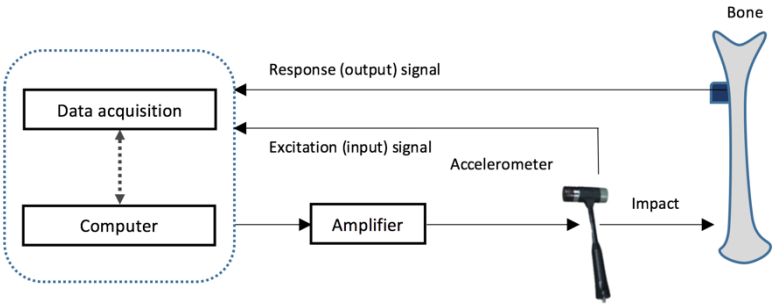


Figure 3. Measuring vibration response of bone using modal hammer

2.2 Finite element modeling

A bovine femur was CT scanned with a slice thickness of 1 mm (16 slice Siemens SOMATON emotion), and a three dimensional model of bone was created using Mimics© v17, MATERIALISE. Exporting the geometry from MIMICS to 3-Matic® v17, tetrahedral volume meshes were generated. A standard procedure (Materialise NV, Leuven, Belgium, 2010) was followed to obtain the three dimensional geometry from DICOM images and mesh the model. The acquired mesh was exported to a commercial FE software for numerical analysis.

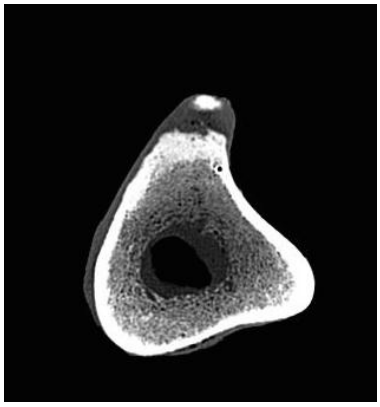


Figure 4. A CT image of the bovine bone

2.3 Material properties assignment

Based on Hounsfield gray values, Mimics can assign material properties to volumetric meshes. After bringing the mesh back to Mimics, an average Hounsfield value is calculated for each element, and the range of gray value is divided into equally sized intervals to represent different material groups. In this study, five material groups were used to model the bone.

The effective bone density and CT numbers are assumed to be linearly correlated [7], [20], [25]. The following equation was used to assign apparent density to the mesh:

$$\rho = 4.64\times10^{-4}\times HU + 1 \tag{1}$$

where ρ is the apparent density (g/cm3) and HU is the CT number (Hounsfield unit).

Considering the literature, the relationship between apparent density and Young's modulus is generally reported in the following form:

$$E = a\rho^{b} + c \tag{2}$$

where E is the Young's modulus, ρ is the apparent density (ash, wet or dry) and a, b and c are the model parameters. A Poisson ratio of 0.3 was considered for all finite elements. Here, experimental results and numerical methods were used to determine the model coefficients.

2.4 Numerical eigenfrequency analysis

The first five natural frequencies and mode shapes of the bone model were calculated using COMSOL Multyphysics v5 without considering the damping effect. The generated mesh together with the material properties were imported to COMSOL. The density-elasticity relationship and model parameters a, b and c were defined in COMSOL according to [20], as a first approximation. LiveLink ™ for MATLAB was used to apply the genetic algorithm and find the optimal coefficients.

2.5 Obtaining coefficients of density-elasticity relationship using GA

The FE model in Matlab was changed to represent a function with the coefficients a, b and c as inputs and the first five natural frequencies as outputs. Assuming that the most exact density-elasticity relationship can result in the most precise values of natural frequencies, we defined an optimization problem to obtain the coefficients in Eq. 2. The following objective function was taken to represent the discrepancy between numerical and experimental results:

$$OF = \sqrt{(f_{E1}-f_{N1})^2 + (f_{E2}-f_{N2})^2 + (f_{E3}-f_{N3})^2} \tag{3}$$

where f_{Ei} is the i[th] natural frequency obtained from experimental modal analysis, and f_{Nj} is the j[th] natural frequency obtained from numerical eigenfrequency analysis. The genetic algorithm toolbar in Matlab® R2014a was used to minimize the objective function. The initial population was chosen to be [a, b, c] = [2, 3, 0], and the boundary for searching the optimal answer was [0-10] for all parameters. Population size and number of generations were set to 40 and 10 respectively.

2.6 Validation

The acquired density-elasticity relationship was used to assign material properties to the FE model of bone with clamped-free BCs. The results of eigenfrequency analysis were compared with the experimental natural frequencies to assess the validity of the relationship in different BCs. Other material assignment strategies were also examined, and the results were compared.

3. Results and discussion

Both hammer and shaker tests were performed to measure natural frequencies of bovine bone in free-free boundary conditions. Accelerometers used in shaker test were only able to measure bending vibrations in the x direction.

Table 1. Natural frequencies of bone in free-free BCs; hammer and shaker tests

mode shape/direction	bending		torsion	bending	
	x	y	-	x	y
natural frequency	1st	2nd	3rd	4th	5th
hammer (Hz)	646	834	1278	1875	2342
shaker (Hz)	645	-	-	1798	-

Table 2 density of different material groups

Material number	1	2	3	4	5
Density (g/cm3)	711.9	1010.5	1309.0	1607.5	1906.1

Table 2 represents the apparent densities of five material groups which are calculated using Eq. 1. Many density-elasticity relationships are proposed in the literature for specific ranges of density which result in different values of elasticity.

In order to obtain more accurate Young's modulus values, genetic algorithm was applied to minimize the objective function defined in Eq. 3. Table 3 represents the results of this optimization process. The natural frequencies were determined using different density-elasticity relationships (initial value, GA and Baca et al), and the results were then compared to experimental findings.

Table 3 results of GA optimization

Study ▸		GA	Initial population	Baca (2008)	experiment
Natural frequency	1	623.3	549.3	572.7	646
	2	825.7	726.2	757.2	834
	3	1286.5	1132.0	1179.0	1278
	4	1877.5	1640.9	1706.2	1875
	5	2267.9	1981.3	2061.3	2342

Objective function	OF	25.65	205.68	145.16	-
Density-elasticity equation coefficients	a	1.26986	2	2.065	-
	b	3.81558	3	3.09	-
	c	2.62971	0	0	-

Considering the values of objective function, it is clear that genetic algorithm can be utilized to find the coefficients of density-elasticity relationship which lead to an accurate FE model. Although the first three natural frequencies were used during the optimization process, results are accurate in all modes of vibration. This fact indicates that the resultant equation predicts the real values of Young's modulus and not those which only minimize the objective function numerically.

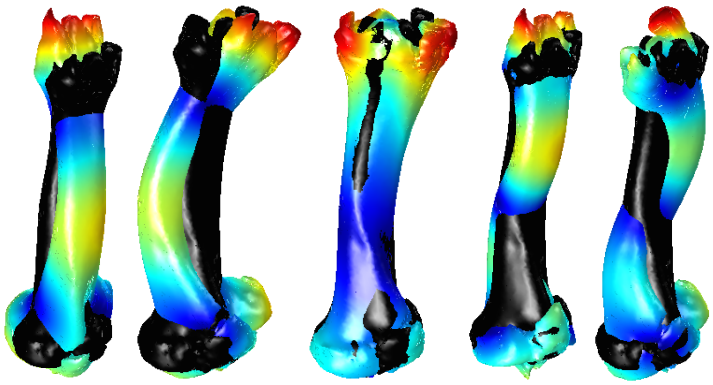


Figure 5. The first five natural modes of vibration; free-free BCs

The first five natural frequencies of bone subjected to clamped-free BCs, based on different density-elasticity relationships, are presented in Table 4. The values are compared with experimental results and the proposed GA method.

Table 4 first five natural frequencies of bone in clamped-free BCs; experimental results vs. FE

		frequency1	frequency2	frequency3	frequency4	frequency5	mean %error
Shaker test		63	-	-	556	-	-
GA method		62.3	80.8	472.4	552.4	656.6	0.879
literature	[13]	59.90	77.00	438.95	510.50	605.60	6.196
	[17]	55.68	72.20	387.25	456.60	546.85	14.417
	[20]	60.40	77.70	448.35	524.70	621.36	4.520

The suggested method results in more accurate natural frequencies not only in free-free BCs (which were used to obtain the model coefficients) but also in clamped-free BCs with totally different values of resonance frequencies. The predicted values of local Young's modulus can therefore be considered as true and reliable.

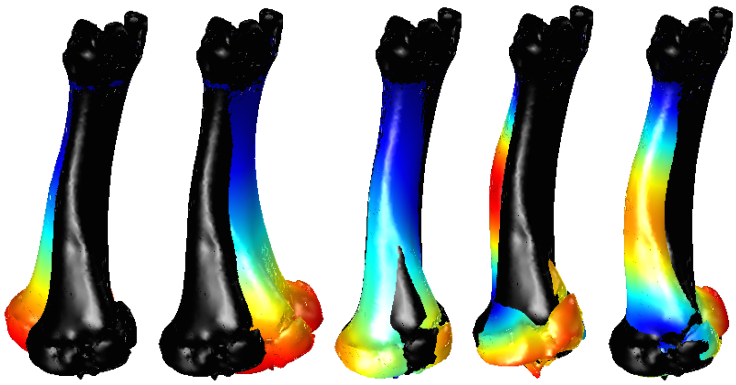


Figure 6. The first five natural modes of vibration; clamped-free BCs

A sensitivity analysis was performed to investigate the effect of changing model parameters in equation (2) on the the first five natural frequencies of the bone. In Figures 7 through 9, two parameters were kept constant while the third parameter changed around a mean value. Variation of the Poisson's ratio did not have a significant effect on the bending natural frequencies. Torsional natural frequency, however, slightly decreased with increasing Poisson's ratio values (Fig. 10).

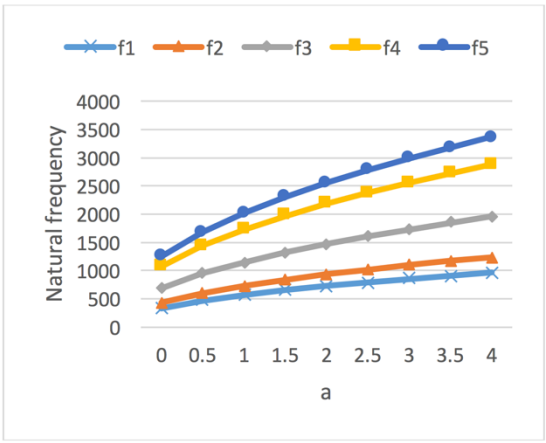


Figure 7. sensitivity analysis; parameter a

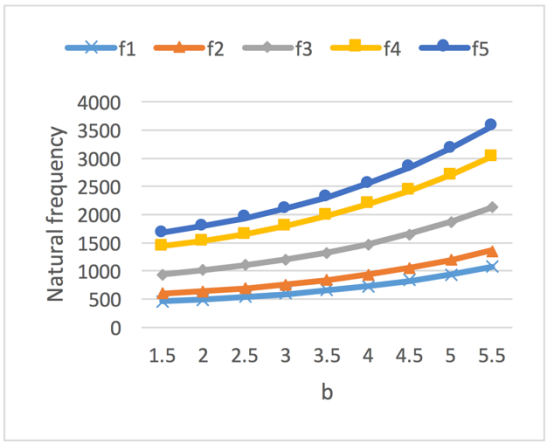


Figure 8. sensitivity analysis; parameter a

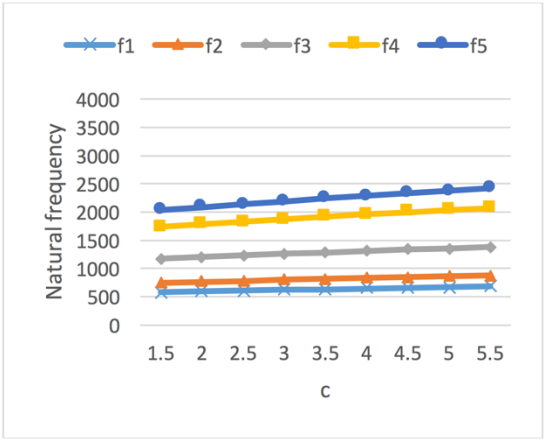


Figure 9. sensitivity analysis; parameter a

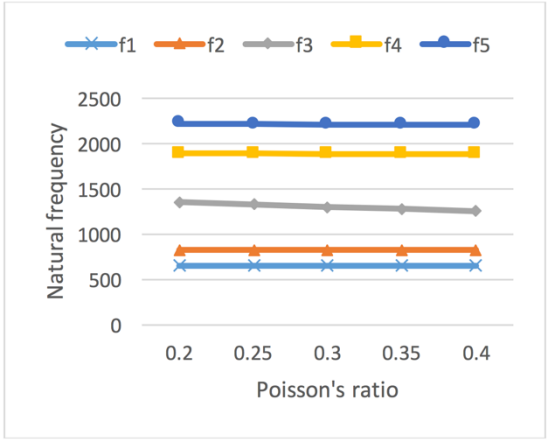


Figure 10 sensitivity analysis Poisson's ratio

Local optimal points were avoided in GA, because of mutations and the final result were closer to the global minimum. However, Genetic algorithm could be time consuming when the number of generations and population increase. To avoid this problem, number of generations and population were limited to 10 and 40, respectively.
There were several limitations associated with the FE model. Five material groups were considered to be enough to represent the distribution of the mechanical properties. Additionally, the effect of marrow on the bone response was presumed negligible, and the material behavior was assumed to be isotropic and linear elastic. A more advanced model may include more groups of materials or a continuous distribution of material properties and consider the effects of nonlinearity, anisotropy and bone marrow in the model.

4. Conclusion

In this study, the density-elasticity relationship of a bovine bone was determined by introducing and solving an optimization problem. Genetic algorithm was used to minimize the difference between natural frequencies obtained from experimental and FE modal analyses. The assumption was that the experimental and numerical results agree, if the material distribution in model approaches the real distribution.
Using the density-elasticity relationship obtained by GA, the numerical resonant frequencies were in good agreement with the experimental results in all modes of vibration with free-free and clamped-free BCs. It can be concluded that the relationship between density and elasticity of bone can be determined with a single mechanical test (experimental modal analysis) and solving an optimization problem based on FE analysis, where the results are valid for all bone density ranges.

References

[1] L. Zhao, T. Dodge, A. Nemani, and H. Yokota, “Resonance in the mouse tibia as a predictor of frequencies and locations of loading-induced bone formation,” *Biomech. Model. Mechanobiol.*, vol. 13, no. 1, pp. 141–151, 2014.

[2] B. A. Wallace and R. G. Cumming, “Systematic review of randomized trials of the effect of exercise on bone mass in pre-and postmenopausal women,” *Calcif. Tissue Int.*, vol. 67, no. 1, pp. 10–18, 2000.

[3] R. Scholz, F. Hoffmann, S. von Sachsen, W.-G. Drossel, C. Klöhn, and C. Voigt, “Validation of density–elasticity relationships for finite element modeling of human pelvic bone by modal analysis,” *J. Biomech.*, vol. 46, no. 15, pp. 2667–2673, 2013.

[4] W. R. R. Taylor, E. Roland, H. Ploeg, D. Hertig, R. Klabunde, M. D. D. Warner, M. C. C. Hobatho, L. Rakotomanana, and S. E. E. Clift, “Determination of orthotropic bone elastic constants using FEA and modal analysis,” *J. Biomech.*, vol. 35, no. 6, pp. 767–773, Jun. 2002.

[5] B. Merz, P. Niederer, R. Mu ller, and P. Ru egsegger, “Automated finite element analysis of excised human femora based on precision-QCT,” *J. Biomech. Eng.*, vol. 118, no. 3, pp. 387–390, 1996.

[6] C. Zannoni, R. Mantovani, and M. Viceconti, “Material properties assignment to finite element models of bone structures: a new method,” *Med. Eng. Phys.*, vol. 20, no. 10, pp. 735–740, 1999.

[7] F. Taddei, E. Schileo, B. Helgason, L. Cristofolini, and M. Viceconti, “The material mapping strategy influences the accuracy of CT-based finite element models of bones: an evaluation against experimental measurements,” *Med. Eng. Phys.*, vol. 29, no. 9, pp. 973–979, 2007.

[8] B. Helgason, E. Perilli, E. Schileo, F. Taddei, S. Brynjlfsson, and M. Viceconti, “Mathematical relationships between bone density and mechanical properties: a

literature review," *Clin. Biomech.*, vol. 23, no. 2, pp. 135–146, 2008.
[9] S. Eberle, M. Gottlinger, and P. Augat, "An investigation to determine if a single validated density-elasticity relationship can be used for subject specific finite element analyses of human long bones," *Med. Eng. Phys.*, vol. 35, no. 7, pp. 875–883, 2013.
[10] J. Y. Rho, M. C. Hobatho, and R. B. Ashman, "Relations of mechanical properties to density and CT numbers in human bone," *Med. Eng. Phys.*, vol. 17, no. 5, pp. 347–355, 1995.
[11] M. J. Ciarelli, S. A. Goldstein, J. L. Kuhn, D. D. Cody, and M. B. Brown, "Evaluation of orthogonal mechanical properties and density of human trabecular bone from the major metaphyseal regions with materials testing and computed tomography," *J. Orthop. Res.*, vol. 9, no. 5, pp. 674–682, 1991.
[12] R. J. McBroom, W. C. Hayes, W. T. Edwards, R. P. Goldberg, and A. A. D. White, "Prediction of vertebral body compressive fracture using quantitative computed tomography.," *J. Bone Jt. Surg.*, vol. 67, no. 8, pp. 1206–1214, 1985.
[13] J. C. Lotz, T. N. Gerhart, and W. C. Hayes, "Mechanical properties of metaphyseal bone in the proximal femur," *J. Biomech.*, vol. 24, no. 5, pp. 317–329, 1991.
[14] S. M. Snyder and E. Schneider, "Estimation of mechanical properties of cortical bone by computed tomography," *J. Orthop. Res.*, vol. 9, no. 3, pp. 422–431, 1991.
[15] J. C. Lotz, T. N. Gerhart, and W. C. Hayes, "Mechanical properties of trabecular bone from the proximal femur: a quantitative CT study.," *J. Comput. Assist. Tomogr.*, vol. 14, no. 1, pp. 107–114, 1990.
[16] F. Linde, I. Hvid, and F. Madsen, "The effect of specimen geometry on the mechanical behaviour of trabecular bone specimens," *J. Biomech.*, vol. 25, no. 4, pp. 359–368, 1992.
[17] M. Dalstra, R. Huiskes, A. Odgaard, and L. Van Erning, "Mechanical and textural properties of pelvic trabecular bone," *J. Biomech.*, vol. 26, no. 4, pp. 523–535, 1993.
[18] B. Li and R. M. Aspden, "Composition and mechanical properties of cancellous bone from the femoral head of patients with osteoporosis or osteoarthritis," *J. Bone Miner. Res.*, vol. 12, no. 4, pp. 641–651, 1997.
[19] E. F. Morgan, H. H. Bayraktar, and T. M. Keaveny, "Trabecular bone modulus-density relationships depend on anatomic site," *J. Biomech.*, vol. 36, no. 7, pp. 897–904, 2003.
[20] V. Baca, Z. Horak, P. Mikulenka, and V. Dzupa, "Comparison of an inhomogeneous orthotropic and isotropic material models used for FE analyses," *Med. Eng. Phys.*, vol. 30, no. 7, pp. 924–930, 2008.
[21] L. Peng, J. Bai, X. Zeng, and Y. Zhou, "Comparison of isotropic and orthotropic material property assignments on femoral finite element models under two loading conditions," *Med. Eng. Phys.*, vol. 28, no. 3, pp. 227–233, 2006.
[22] A. Nazarian, D. von Stechow, D. Zurakowski, R. Müller, and B. D. Snyder, "Bone volume fraction explains the variation in strength and stiffness of cancellous bone affected by metastatic cancer and osteoporosis," *Calcif. Tissue Int.*, vol. 83, no. 6, pp. 368–379, 2008.
[23] B. Couteau, M.-C. Hobatho, R. Darmana, J.-C. Brignola, and J.-Y. Arlaud, "Finite element modelling of the vibrational behaviour of the human femur using CT-based individualized geometrical and material properties," *J. Biomech.*, vol. 31, no. 4, pp. 383–386, Apr. 1998.
[24] Materialise NV, *Mimics Student Edition Course Book*. 2010.
[25] L. C. Kourtis, D. R. Carter, H. Kesari, and G. S. Beaupre, "A new software tool (VA-BATTS) to calculate bending, axial, torsional and transverse shear stresses within bone cross sections having inhomogeneous material properties," *Comput. Methods Biomech. Biomed. Engin.*, vol. 11, no. 5, pp. 463–476, 2008.

An Improved Method of Continuum Topology Optimization Subjected to Frequency Constraints Based on Indenpendent Continuous Topological Variables

H.L. Ye[1], *W.W. Wang[1], Y.K. Sui[1]

[1] College of Mechanical Engineering and Applied Electronics Technology,
Beijing University of Technology, Beijing, China

*Presenting author: yehongl@bjut.edu.cn

Abstract

In this paper, an improved topology optimal model of continuum structures subject to frequency constraints is established based on Independent, Continuous, Mapping (ICM) method. Firstly, two filter functions- Power Function(PF) and Composite Exponential Function(CEF) are selected to recognize the design variables, and to implement the changing process of design variables from "discrete" to "continuous" and back to "discrete". Explicit formulations of frequency constraints are given based on Rayleigh's quotient, filter functions, first -order Taylor series expansion. Then, an improved optimal model is formulated using different filter functions and the explicit frequency constraints. The program based on the dual sequence quadratic programming (DSQP) and global convergent method of moving asymptotes algorithm(GCMMA) for solving the optimal model is developed on the platform of MSC.Patran & Nastran. Finally, numerical examples are given to demonstrate the validity and applicability of the proposed method. By comparison, we find that the results from DSQP method equipped with filter function of composite exponential function are a little better than other methods for the problem of frequency constraints.

Key words： Topological optimization·Continuum·Frequency constraint· Independent Continuous and Mapping(ICM) method·filter function

Keywords: Topological optimization, Continuum, Frequency constraint, Independent Continuous and Mapping(ICM) method, filter function.

Introduction

The essence of topology optimization lies in searching for the optimum path of transferring loads, therefore the computational results of topology optimization are usually more attractive and more challenging than the results of cross-sectional and shape optimization. Although topology optimization is only in conceptual design phase in engineering, the design results significantly impacts the performance of the final structure. Since the landmark paper of Bendsøe and Kikuchi[1], numerical methods for topology optimization of continuum structures have been developed quickly in application[2]. The known are homogenization method[5],6], variable density method(including SIMP and RAMP interpolation model)[7-10], evolutionary structural optimization (ESO) [11-13], level set method [14-16]and so on.

Compared with static topology optimization, the optimization algorithm on dynamic topology optimization is more complicated and the calculation of sensitivity analysis is more enormous. Frequency topology optimization is of great importance in dynamic topology optimization and engineering fields. Topology Optimization with respect to frequencies of structural vibration was first considered by Diaz and Kikuchip[17], who studied the topology optimization of eigenvalues by using the homogenization method where reinforcement of a structure is optimized to maximize eigenvalues. Subsequently, many researches focus on to expand topology optimization in dynamic problems. Ma et al. [18,19], Kosaka and Swan [20]presented different formulations for simultaneous maximization a set of frequencies of free vibration of disk and plate structures. Krog and Olhoff [21],

Jensen and Pedersen[22] utilize a variable bound formulation that facilitates proper treatment of multiple frequencies. Pedersen[23]deals with maximum fundamental frequency design of plates and includes a technique to avoid spurious localized modes. Jensen and Pedersen[22] presents a method to maximize the separation of two adjacent frequencies in structures with two material components. Zhu & Zhang[24] emphasize on layout design which is related to optimization of boundary conditions and it is studied to maximize natural frequency of structures. In 2007， Du and Olhoff[25] introduced SIMP method for maximization of first eigenvalue and frequency gaps. In 2009, Niu et al.[26] proposed a two-scale optimization method and found the optimal figurations of macrostructure- microstructure of cellular material with maximum structural fundamental frequency. Huang et al.[27] investigated the maximization of fundamental frequency of beam, plane and three-dimensional block by applying a new bi-directional evolutionary structural optimization (BESO) method, and dealt with localized modes by modifying the traditional penalization function of SIMP method. Qi et al.[28] presented a level set based shape and topology optimization method for maximizing the simple or repeated first eigenvalues of structure vibration. Further development on frequency topology optimization see references[29-33].

Independent, Continuous and Mapping (ICM) method[34], which is proposed by Sui for skeleton and continuum structural topology optimization in 1996. This method generalizes topological variables abstractly independence of the design variables such as sectional sizes, geometrical shape, density or Young's modulus of material. Filter functions are used to map the changing process of topological design variables from "discrete" to "continuous" and back to "discrete". The smooth model with minimizing structural weight is established and solved by the traditional algorithms in mathematical programming. This model is beneficial to maintain the consistency of objective and constraint in cross-sectional optimization, shape optimization and topology optimization.

In this paper, we extend our previous research[34-36] primarily about Independent, Continuous and Mapping (ICM) method on static topology optimization issues of continuum structures to dynamic topology optimization field. A model of topology optimization for the lightest structures with frequency constraints is investigated. An improved model of continuum topology optimization with Composite Exponential Function(CEF) as filter function instead of Power function is established. Among the methods of mathematic optimization model solving, mathematical programming (MP) method is popular. Because of the nonlinearity of mathematic optimization model in topology optimization of continuum structure， sequential quadratic programming (SQP) in the MP method are widely used. And the dual theory is used to convert the constrained optimization model to one with reduced number of design variables, and the solving efficiency is greatly improved. Therefore, dual sequential quadratic programming (DSQP) algorithm is employed in this paper, and the results is compared with that of the global convergent method of moving asymptotes algorithm (GCMMA)[37,38].

This paper is organized as follows. In section 2, the optimization formulation and description of filter function are introduced. In section 3, an improved frequency topology optimization model based on ICM method is built. Optimal algorithms to solve the mathematical optimization problem are given in section 4. Numerical simulations are presented in section 5. In section 6, conclusions are given.

1 Problem formulation and description of filter function

1.1 Optimization problem formulation

For structural cross-section and shape optimization, natural frequency of structure is often taken as constraint. We denote f_i as the frequency of *i*-th order, and $\underline{f_i}, \overline{f_i}$ are the low and up bound of *i*-th order frequency respectively. They satisfy the following inequality:

(i) $f_1 \geq \underline{f_1}$;

(ii) $f_i \le \bar{f}_i$ and $f_{i+1} \ge \underline{f}_{i+1}$ in non-frequency band constraints.

For elastic structure, the usual relation between frequency f and eigenvalue is $\lambda = (2\pi f)^2$. Therefore, the frequency constraints can apparently be transformed into eigenvalue constraints using the formula. Here we uniformly use $g(\lambda) \le \bar{\lambda}$ to generalize (i) and (ii) based on $\lambda = (2\pi f)^2$.

Thus, the model of continuum topology optimization with frequency constraints can be formulated as follows

$$\begin{cases} find \ \ \mathbf{t} \in E^N \\ make \ W = \sum_{i=1}^{N} w_i \ \to \min \\ s.t. \quad g(\lambda_j) \le \bar{\lambda}_j (j = 1, \cdots, J) \\ \quad 0 < \underline{t} \le t_i \le 1 \ (i = 1, \cdots, N) \end{cases} \tag{1}$$

where $\boldsymbol{t}$ and W denote the topological design variable vector and the weight of structure. i and j are the i-th element and the j-th order frequency respectively, J and N represent the total number of constraints and elements. And $\underline{t}$ is the lower bound of design variables, e.g. $\underline{t} = 0.001$.

1.2 Description of the filter function

In order to develop the model ICM method, we firstly investigate the essential part of ICM—the filter function. Its definition and choosing determine the establishment and solving of optimization model, and further filter funciton will make great impact on the final performance of topology optimization. In order to map the topological variables from "discrete" to "continuous", Sui(1996) studied the filter function $f(t_i)$.

Several types of filter function are suggested in ICM method[34]. Among which, Power Function(PF) is used frequently in application[36] and is as follows

$$f(t_i) = t_i^{\alpha}, \ \alpha \ge 1 \tag{2}$$

Here t_i denotes i-th design variable. α is a positive constant.

We introduce a new filter function -Composite Exponential Function(CEF) to take the place of the old one and it is as follows:

$$f(t_i) = \frac{e^{t_i/\gamma} - 1}{e^{1/\gamma} - 1}, \ \gamma > 0 \tag{3}$$

γ is a given positive constant and it is determined by numerical experiments with different problems. In section 5, we compared the performance of the two types of filter function.

Denote $f_w(t_i)$, $f_k(t_i)$ and $f_m(t_i)$ as filter functions for frequency topology optimization and they are given as follows:

$$w_i = f_w(t_i) w_i^0, \ \boldsymbol{k}_i = f_k(t_i) \boldsymbol{k}_i^0, \ \boldsymbol{m}_i = f_m(t_i) \boldsymbol{m}_i^0 \tag{4}$$

Here w_i^0, $\boldsymbol{k}_i^0$ and $\boldsymbol{m}_i^0$ are the element weight, element stiffness matrix and element mass matrix of original structure before the process of topology optimization, respectively. w_i, $\boldsymbol{k}_i$ and $\boldsymbol{m}_i$ are the ones in the process of topology optimization, respectively.

2 Improved model based on ICM method

2.1 Determination of eigenvalue

In the finite element analysis the dynamic behavior of a continuum structure can be represented by the following general eigenvalue problem

$$(\mathbf{K} - \lambda_j \mathbf{M}) \mathbf{u}_j = 0 \tag{5}$$

where, $\boldsymbol{K}$ is the global stiffness matrix and $\boldsymbol{M}$ is the global mass matrix. λ_j is the *j*th eigenvalue and $\boldsymbol{u}_j$ is the eigenvector corresponding to λ_j. The eigenvalue λ_j and the corresponding eigenvector $\boldsymbol{u}_j$ are related to each other by Rayleigh quotient

$$\lambda_j = \frac{\mathbf{u}_i^T \mathbf{K} \mathbf{u}_i}{\mathbf{u}_i^T \mathbf{M} \mathbf{u}_i} \tag{6}$$

2.2 Sensitivity analysis

Since eigenvalue λ_j is implicitly related with topology variable $\boldsymbol{t}$, we use first-order Taylor series expansion for eigenvalue to express their relationship explicitly. At first, the sensitivity of eigenvalue with respect to design variables should be derived.

Take the reciprocal of stiffness filter function as design variables as follows

$$x_i = \frac{1}{f_k(t_i)} \tag{7}$$

We have

$$t_i = f_k^{-1}(x_i) \tag{8}$$

Therefore, the stiffness matrix filter function, mass matrix filter function and weight filter function are given as follows

$$f_k(t_i) = \frac{1}{x_i} \;\; ; \;\; f_m(t_i) = f_m[f_k^{-1}(\frac{1}{x_i})] \;\; ; \;\; f_w(t_i) = f_w[f_k^{-1}(\frac{1}{x_i})] \tag{9}$$

In view of (6), we have the derivative of λ_j to design variable as follows:

$$\frac{\partial \lambda_j}{\partial x_i} = \boldsymbol{u}_j^T \frac{\partial \boldsymbol{K}}{\partial x_i} \boldsymbol{u}_j - \lambda_j \boldsymbol{u}_j^T \frac{\partial \boldsymbol{M}}{\partial x_i} \boldsymbol{u}_j \tag{10}$$

Considering Eq.(4) and (9), the global stiffness matrix $\boldsymbol{K}$ and the mass matrix $\boldsymbol{M}$ can be calculated by

$$\boldsymbol{K} = \sum_{i=1}^{N} \boldsymbol{k}_i = \sum_{i=1}^{N} f_k(t_i) \boldsymbol{k}_i^0 = \sum_{i=1}^{N} \frac{1}{x_i} \boldsymbol{k}_i^0 \,, \qquad \boldsymbol{M} = \sum_{i=1}^{N} \boldsymbol{m}_i = \sum_{i=1}^{N} f_m(t_i) \boldsymbol{m}_i^0 = \sum_{i=1}^{N} f_m\left[f_k^{-1}\left(\frac{1}{x_i} \right) \right] \boldsymbol{m}_i^0 \tag{11}$$

Substituting Eq.(11) to Eq.(10), we have

$$\frac{\partial \lambda_j}{\partial x_i} = -\mathbf{u}_j^{\mathrm{T}} \frac{2\mathbf{k}_i}{2x_i} \mathbf{u}_j + \beta(x_i) \lambda_j \mathbf{u}_j^{\mathrm{T}} \frac{2\mathbf{m}_i}{2x_i} \mathbf{u}_j = -\frac{2}{x_i} (U_{ij} - \beta(x_i) V_{ij}) \tag{12}$$

where, $\beta(x_i) = \dfrac{f_m'[f_k^{-1}(1/x_i)] f_k(1/x_i)}{f_m[f_k^{-1}(1/x_i)] f_k'(1/x_i)} = \dfrac{f_m'(t_i) f_k(t_i)}{f_m(t_i) f_k'(t_i)}$

In (12), $U_{ij} = \frac{1}{2} \boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{k}_i \boldsymbol{u}_j$ and $V_{ij} = \frac{1}{2} \lambda_j \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{m}_i \boldsymbol{u}_j$ represent the strain energy and the kinetic energy of *i*th element corresponding to the *j*th eigenmode, respectively. At this moment, the derivatives of eigenvalue with respect to all design variables can be obtained by subtracting the strain energy and kinetic energy for element mode from the results of modal analyses.

2.3 Explicit expression of eigenvalue

Using the first-order Taylor series expansion, the approximate expression of eigenvalue $\lambda_j(\boldsymbol{t})$ can be obtained

$$\lambda_j(\boldsymbol{t}) = \lambda_j(\boldsymbol{x}^{(\nu)}) + \sum_{i=1}^{N} \frac{\partial \lambda_j}{\partial x_i} (x_i - x_i^{(\nu)}) \tag{13}$$

where superscript ν is the number at the ν -th iteration.

Substitute Eq.(10) into Eq.(13), we get

$$\lambda_j(\boldsymbol{t}) = \lambda_j(\boldsymbol{x}^{(\nu)}) + \sum_{i=1}^{N} 2(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)}) - \sum_{i=1}^{N} \frac{2}{x_i^{(\nu)}}(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)})x_i \tag{14}$$

The constraint $\underline{\lambda}_j \le \lambda_j(\boldsymbol{x})$ can be rewritten as

$$\sum_{i=1}^{N} \frac{2}{x_i^{(\nu)}}(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)})x_i \le -1 \times [\underline{\lambda}_j - \lambda_j(\boldsymbol{x}^{(\nu)}) - \sum_{i=1}^{N} 2(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)})]$$

For constraint $\lambda_j(\boldsymbol{x}) \le \overline{\lambda}_j$, it can be rewritten as

$$-1 \times \sum_{i=1}^{N} \frac{2}{x_i^{(\nu)}}(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)})x_i \le \overline{\lambda}_j - \lambda_j(\boldsymbol{x}^{(\nu)}) - \sum_{i=1}^{N} 2(U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)})$$

If we define D as

$$D = \begin{cases} 1 \quad for\ \lambda_j \le \overline{\lambda}_j \\ -1\ for\ \lambda_j \ge \underline{\lambda}_j \end{cases} \qquad \tilde{\lambda}_j = \begin{cases} \underline{\lambda}_j & (\lambda_j \ge \underline{\lambda}_j) \\ \overline{\lambda}_j & (\lambda_j \le \overline{\lambda}_j) \end{cases}$$

And further define

$$\begin{aligned} &A_{ij} = U_{ij}^{(\nu)} - \beta(x_i^{(\nu)})V_{ij}^{(\nu)},\ \overline{c}_{ij} = -\frac{2}{x_i^{(\nu)}} A_{ij},\ c_{ij} = -D\overline{c}_{ij} \\ &\overline{d}_j = -\lambda_j(\mathbf{x}^{(\nu)}) - \sum_{i=1}^{N} 2A_{ij} \quad ,\ d_j = D \times (\tilde{\lambda}_j + \overline{d}_j) \end{aligned} \tag{15}$$

Then frequency constraints can be simplified by the following inequality:

$$\sum_{i=1}^{N} c_{ij} x_i \le d_j \tag{16}$$

Thus ends the process of explicitly approximation of the frequency constraints.

2.4 Improved model of frequency topology optimization

Based on the above analysis, the model of topology optimization with frequency constraints by introducing filter function can be transformed as follows:

$$\begin{cases} find\ \ \boldsymbol{t} \in E^N \\ make\ W = \sum_{i=1}^{N} f_w(t_i) w_i^0 \to \min \\ s.t.\quad g(\lambda_j(f_k(t_i), f_m(t_i))) \le \overline{\lambda}_j (j = 1, \cdots, J) \\ \qquad 0 < \underline{t} \le t_i \le 1\ (i = 1, \cdots, N) \end{cases} \tag{17}$$

By using explicitly approximation of the frequency constraints, the model (17) can be written as follows:

$$\begin{cases} \text{find} \quad \boldsymbol{t} \in E^N \\ \text{make}\ \ W = \sum_{i=1}^{N} f_w(t_i) w_i^0 \to \min \\ \text{s.t.}\ \ \sum_{i=1}^{N} c_{ij} x_i \le d_j (j = 1, \cdots, J) \\ \qquad 1 \le x_i \le \overline{x}_i (i = 1, \cdots, N) \end{cases} \tag{18}$$

3. Solution of the improved topology optimization model

3.1 Standardization of objective

Considering model (18) is a programming with nonlinear objective and linear constraints following the explicit process of frequency constraints, the second-order Taylor series expansion is used to approximate the objective function and ignore the constant item. The model is transformed into the following quadratic programming model:

$$\begin{cases} \text{find} \quad \boldsymbol{t} \in E^N \\ \text{make} \ W = \sum_{i=1}^{N} (a_i x_i^2 + b_i x_i) \to \min \\ \text{s.t.} \ \ \sum_{i=1}^{N} c_{ij} x_i \le d_j (j = 1, \cdots, J) \\ \quad 1 \le x_i \le \bar{x}_i (i = 1, \cdots, N) \end{cases} \tag{19}$$

As objective function varies with different filter functions, investigation of the different cases following different types of filter functions is necessary. Here we focus on PF and CEF.

When PF is applied as the filter function, it is given as follows:

$$f_w(t_i) = t_i^{\gamma_w}; \ f_k(t_i) = t_i^{\gamma_k}; \ f_m(t_i) = t_i^{\gamma_m} \tag{22}$$

In view of (8), we have $t_i = \dfrac{1}{x_i^{1/\gamma_k}}$, $t_i^{\gamma_w} = \dfrac{1}{x_i^{\gamma_w/\gamma_k}} = \dfrac{1}{x_i^{\alpha}}$, and $\alpha = \dfrac{\gamma_w}{\gamma_k}$.

Therefore the objective function (19) can be rewritten as:

$$W = \sum_{i=1}^{N} t_i^{\gamma_w} w_i^0 = \sum_{i=1}^{N} \frac{w_i^0}{x_i^{\alpha}} \approx \sum_{i=1}^{N} (a_i x^2 + b_i x) \tag{23}$$

Where $a_i = \dfrac{\alpha(\alpha+1) w_i^0}{2(x_i)^{\alpha+2}}$ and $b_i = \dfrac{-\alpha(\alpha+2) w_i^0}{(x_i)^{\alpha+1}}$ are undetermined parameters.

When CEF is applied as the filter function, it is given as follows:

$$f_w(t_i) = \frac{e^{t_i/\gamma_w} - 1}{e^{1/\gamma_w} - 1}; \ f_k(t_i) = \frac{e^{t_i/\gamma_k} - 1}{e^{1/\gamma_k} - 1}; \ f_m(t_i) = \frac{e^{t_i/\gamma_m} - 1}{e^{1/\gamma_m} - 1} \tag{24}$$

We have $x_i = \dfrac{1}{f_k(t_i)} = \dfrac{e^{1/\gamma_k} - 1}{e^{t_i/\gamma_k} - 1}$, and therefore

$$f_w(t_i) = \frac{\left(\dfrac{e^{1/\gamma_k} - 1}{x_i} + 1 \right)^{\frac{\gamma_k}{\gamma_w}} - 1}{e^{1/\gamma_w} - 1}; \quad f_m(t_i) = \frac{\left(\dfrac{e^{1/\gamma_k} - 1}{x_i} + 1 \right)^{\frac{\gamma_k}{\gamma_m}} - 1}{e^{1/\gamma_m} - 1} \tag{25}$$

Similarly, the objective function in (20) can be expressed as

$$\sum_{i=1}^{N} \frac{\left(\dfrac{e^{1/\gamma_k} - 1}{x_i} + 1 \right)^{\frac{\gamma_k}{\gamma_w}} - 1}{e^{1/\gamma_w} - 1} w_i^0 \approx \sum_{i=1}^{N} (a_i x_i^2 + b_i x_i) \tag{26}$$

Where $a_i = \dfrac{1}{2} \dfrac{\gamma_k}{\gamma_w} \dfrac{w_i^0}{\left(x_i^{(v)}\right)^4} \dfrac{e^{1/\gamma_k} - 1}{e^{1/\gamma_w} - 1} \left(\dfrac{e^{1/\gamma_k} - 1}{x_i^{(v)}} + 1 \right)^{\frac{\gamma_k}{\gamma_w} - 2} \left[\left(\dfrac{\gamma_k}{\gamma_w} + 1 \right) \left(e^{1/\gamma_k} - 1 \right) + 2 x_i^{(v)} \right]$,

$$b_i = -\frac{\gamma_k}{\gamma_w} \frac{w_i^0}{\left(x_i^{(v)}\right)^3} \frac{e^{1/\gamma_k} - 1}{e^{1/\gamma_w} - 1} \left(\frac{e^{1/\gamma_k} - 1}{x_i^{(v)}} + 1 \right)^{\frac{\gamma_k}{\gamma_w} - 2} \left[\left(\frac{\gamma_k}{\gamma_w} + 2 \right) \left(e^{1/\gamma_k} - 1 \right) + 3 x_i^{(v)} \right].$$

They are two undetermined parameters.

3.2 Solving algorithms of optimization model

With the above analysis and solving of (19), DSQP and GCMMA are employed. The optimal topology structure with continuous design variables is obtained. The iterating computation will end until following condition is satisfied

$$\frac{\left\| \boldsymbol{x}^{(v+1)} - \boldsymbol{x}^{(v)} \right\|}{\left\| \boldsymbol{x}^{(v)} \right\|} \le \varepsilon$$

x* obtained at this moment is just the optimal solution of Eq. (19) however. Then t* can be calculated based on Eq. (8). Let $\boldsymbol{t}^{(k+1)} = \boldsymbol{t}^*$ and modify the last structure via immediate iteration

optimizing, and again enter next iteration. Similarly, iterating in this way until the following condition is satisfied

$$\Delta W = \left| \frac{(W^{(v+1)} - W^{(v)})}{W^{(v+1)}} \right| \leq \varepsilon \tag{20}$$

Where $W^{(v)}$ and $W^{(v+1)}$ is the structural weight of previous iteration and current iteration, respectively. ε is a precision of convergence, which is prescribed to be 0.001 herein.

3.3 Discretization of continuous design variables

To map design variables from "continuous" to "discrete" back, filter threshold value is needed. We denote filter threshold value as T_0 to determine whether the element is deleted or not. In order to measure the discreteness degree of topology variables, we use M_{nd}[39] as a criterion and it is given (21).

$$M_{nd} = \frac{\sum_{i=1}^{n} 4T_i(1-T_i)}{n} \times 100\% \tag{21}$$

where T_i is the topological variable value for the i-th element and n is the total number of the elements. Following (21) , if all the topological values are 0 or 1, M_{nd} is 0; if the topological values are 0.5, M_{nd} is 1; the more closer of the topological values to 0 or 1, the more smaller value of M_{nd} and the better of the optimal result.

4. Numerical examples

In this section, we illustrate the proposed method with three examples for the topology optimization with frequency constraints. The first one is a rectangular beam with two frequency constraints. We address the ability of schemes to obtain discrete solutions and compare the solutions obtained using two different filter function. We show how it is possible to formulate and solve optimal problems. The second one is a cylindrical shell structure by second frequency constraint. We aims to compare with the results by using two algorithms combined with two filter functions. For the computation, the initial values of topology variables are all given as unit (t=1), the lowest bounds of topology variables and the convergence precision values are 0.01 and 0.001, respectively.

Example 1 Rectangular beam with two frequency constraints

It is a rectangular beam with two ends clamped and the thickness of beam is assumed as 1mm shown as Fig.1. The design domain is 140mm×20mm, and a concentrated mass (Mc = 50g) is attached at the center of base structure and it has inertia only in Y direction. The Young's modulus E = 100GPa, Poisson's ratio μ=0.3 and mass density ρ = 1000kg/m^3. The structure is divided into 140×20 four-node rectangular elements. We set frequency constraints for the design problem is $f_1 \geqslant$ 8000Hz, $f_2 \geqslant$60000Hz. The topology optimization equation was formulated combine PF and CEF filter functions, respectively.

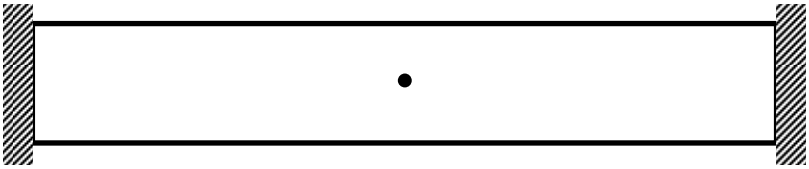


Fig.1 Geometry model of clamped beam

(a)Optimal topology configurations with PF

(b)Optimal topology configurations with CEF

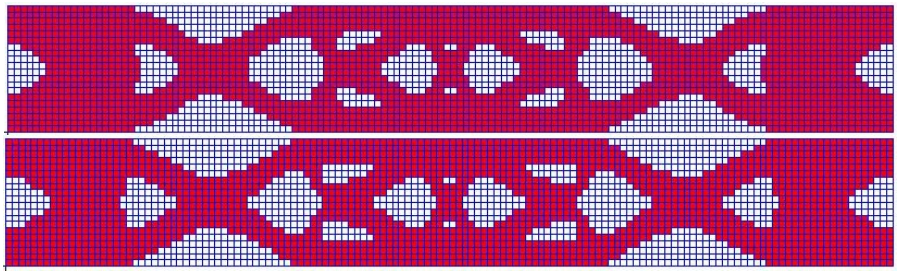


Fig.3 Optimal topology configurations with PF different filter functions

The solving topology configuration of the beam with different filter functions is given in Fig.3. The iterative curve of computation with different filter functions are described in Fig.5-8. To describe the dynamics of optimal structure, the first three modal shapes of optimal structure with two filter functions are computed and displayed in Fig.4-6. The frequency and structural weight changing with time in the optimization process are presented in Fig.7 and Fig.8 with different filter functions. The optimal results with different filter function are shown in Table.1 and the computed distribution of topological design variable values is listed in Table 2.

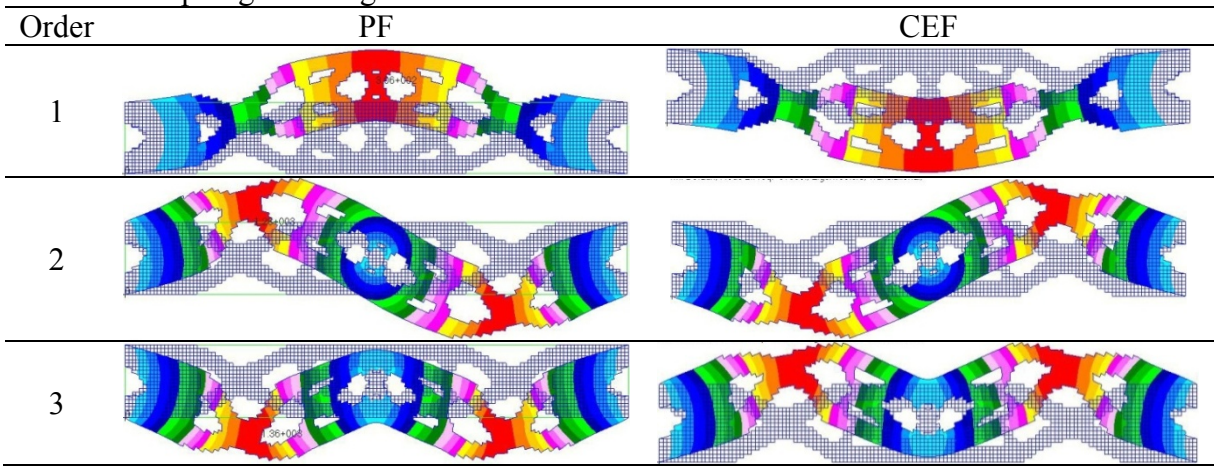


Fig.4 Modal shapes of optimal structure with different filter functions

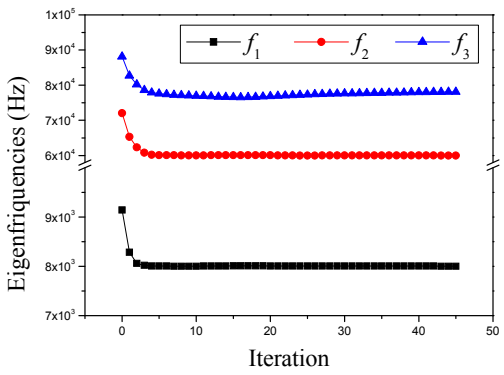


Fig.5 Iteration curves of frequency with PF

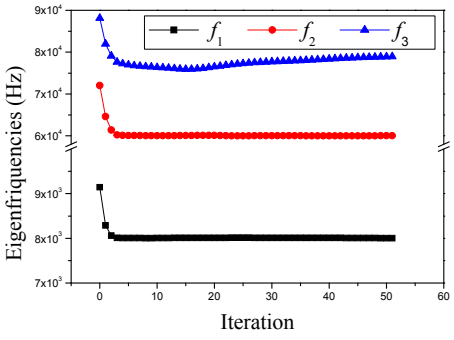


Fig.6 Iteration curves of frequency with CEF

Table1 Optimal results with different filter functions

Filter function	PF	CEF
Iteration	45	51
Mass (g)	2.067093018	1.9778114014
f_1 (Hz)	8003.934082	8003.0073242
f_2 (Hz)	59968.550781	60027.289063

Table2 Distribution of topological value with different filter functions

Distribution of topology value	PF	CEF
(0,0.1]	240	472
(0.1,0.2]	72	60
(0.2,0.3]	52	56
(0.3,0.4]	24	64
(0.4,0.5]	52	48
(0.5,0.6]	108	44
(0.6,0.7]	160	28
(0.7,0.8]	232	84
(0.8,0.9]	216	188
(0.9,1]	1644	1756
Total number of element	2800	2800
M_{nd}	26.74%	16.36%

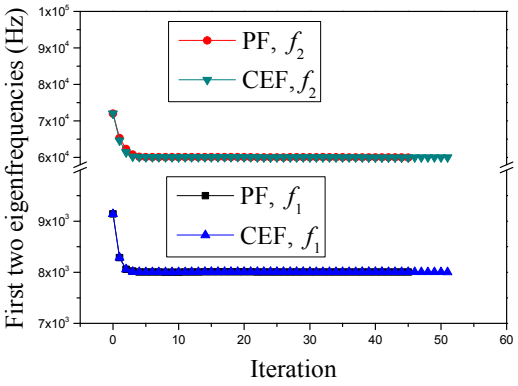


Fig.7 Iteration curves of constrainted frequencies with different filter functions

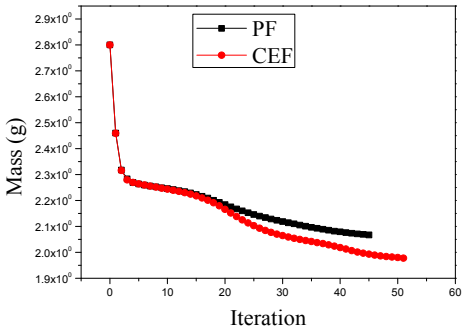


Fig.8 Iteration curves of mass with different filter functions

Example2 Cylindrical shell with the second frequency constraint

A cylindrical shell structure with thickness is 1m, bus-bar a= 20m, arc b=20m, central angle α =0.25 and radius R=80m was shown in Fig.10. In addition, a concentrate mass M=312000kg was attached on the center of cylindrical shell. The Young's modulus E = 100GPa, Poisson's ratio μ=0.3 and mass density ρ = 7800kg/m^3. The structure was divided into 30×30 four-node rectangular elements. The constraint frequency for the design problem is $f_2 \geqslant 28$ Hz. The topology optimization equation was formulated combine PF and CEF filter functions, respectively.

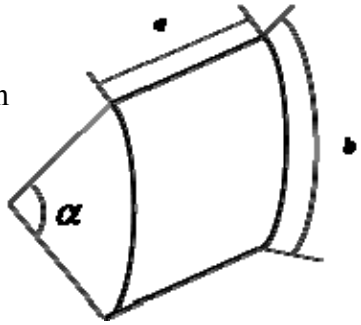
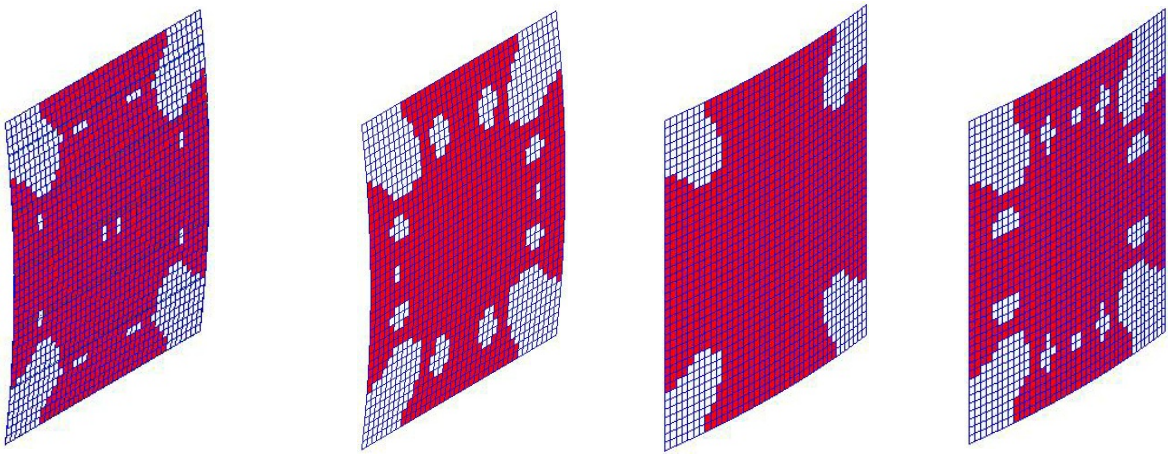


Fig.9 Geometry model of cylindrical shell

Optimal topology configurations after optimization are shown in Fig.10. Iteration curves of first three frequencies with different algorithms and filter functions are given in Figure11. From Fig.12 and Fig.13, we can get the iteration curves of second frequencies and the iteration curves of structural mass for different algorithms and filter functions. Table3 lists the results of optimization for cylinder shell.

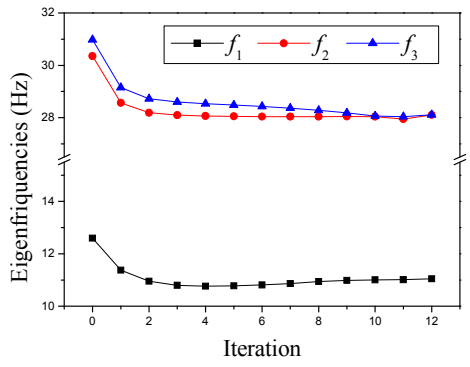
Table3 Results of optimization for cylinder shell

Algorithm and filter function	GCMMA& PF	GCMMA&CEF	DSQP&PF	DSQP&CEF
Iteration	12	29	12	46
Mass (kg)	2634298.85	2309522.20	2633713.28	2146202.33
f_2 (Hz)	28.104261398	28.017398834	28.0874	28.0153
f_3 (Hz)	28.114189148	28.737268448	28.1656	28.7392

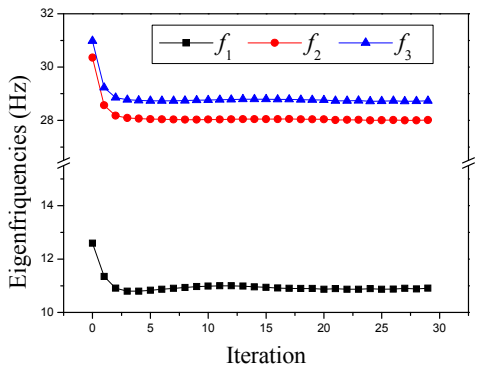


(a) GCMMA& PF (b) GCMMA&CEF (c) DSQP& PF (d) DSQP& CEF

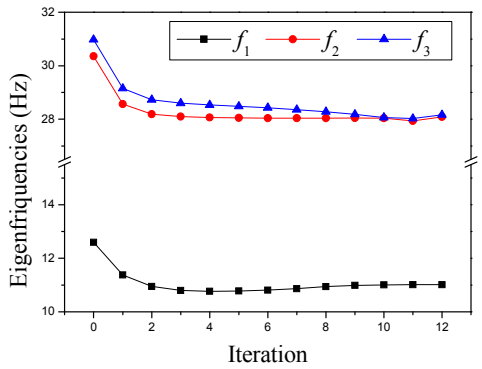
Fig.1 Optimal topology configuration with different algorithms and filter functions.



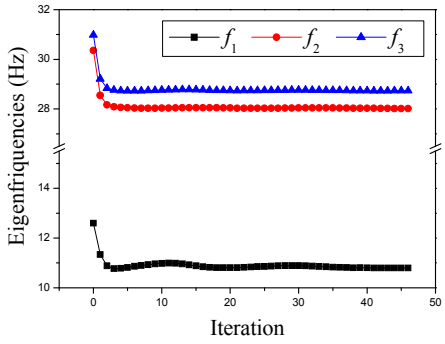
(a) GCMMA& PF



(b) GCMMA&CEF



(c) DSQP& PF



(d) DSQP& CEF

Fig.11 Iteration curves of first three frequencies with different algorithms and filter functions

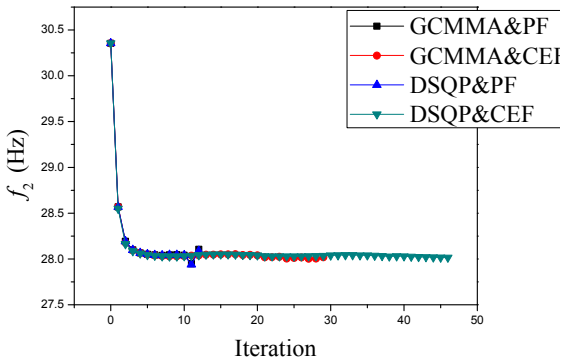


Fig.12 Iteration curves of second frequency

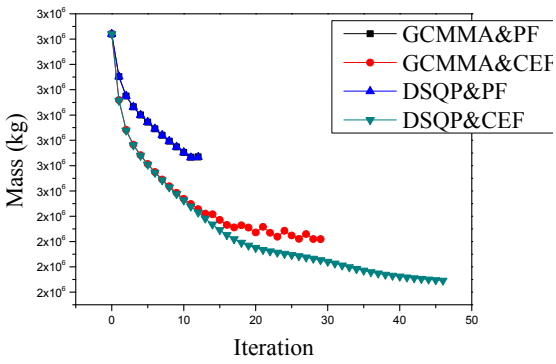


Fig.13 Iteration curves of mass

Conclusion

In this paper, an improved frequency topology optimization model of continuum structure is developed based on ICM method. CEF- a new filter function is selected to recognize the design variables, as well as to implement much better the changing process of design variables from "discrete" to "continuous" and back to "discrete". Explicit formulations of frequency constraints are given by extracting structural strain and structural kinetic energy from the results of structural modal analysis. An improved optimal model is formulated using CEF and the explicit frequency constraints. The program based on DSQP and GCMMA for solving the optimal model is developed on the platform of MSC.Patran & Nastran. Finally, two examples of continuum structure are computed to demonstrate the feasibility of the proposed method.

The performance of the developed program are given in Fig.3,Table1, Table2, Table3, Fig.7, Fig.8, Fig.10, Fig.12, Fig.13. The results from Fig.3 and Fig.10 show that clear and stable configurations can be obtained using different algorithms and filter functions, and we find that configurations computed with DSQP combined PF and DSQP combined CEF, GCMMA combined PF and GCMMA combined CEF are similar between one and the other in Fig.10. From Table 1, we can see that the objective (weight)with CEF is apparent lower than that of PF. However, the iterative step numbers of CEF is larger than that of PF for the convergence. We can also find that DSQP combined CEF has the best performance for the optimization example from the point of view of optimal objective in Fig.13. From the point of the discrete degree, Table2 for the distribution of optimal topological values show that the M_{nd} with PF and CEF are 26.74% and 16.36%, the difference is apparent . CEF has the better performance in the process of optimization.

Although the comparison of DSQP with GCMMA from the recent reference are done, and we have better results coupled with two different filter function, we just give compared results based on ICM method. To improve continuum structure optimal algorithms , it is necessary to investigate the algorithm based on other methods.

Acknowledgment

The project was supported by the National Natural Science Foundation of China (11072009, 111720131). The authors acknowledge Krister Svanberg to offer GCMMA programming.

References

[1] Bendsøe M. P., Kikuchi N(1988). Generating optimal topologies in structural design using a homogenization method[J]. *Computer methods in applied mechanics and engineering*, 71(2): 197-224.

[2] Eschenauer H.A., Olhoff N (2001). Topology optimization of continuum structures: a review[J]. *Appl Mech Rev*, 54(4): 331 –390.

[3] Joshua D. Deaton, Ramana V. Grandhi(2014). A survey of structural and multidisciplinary continuum topology

optimization: post 2000 [J]. *Struct Multidisc Optim* , 49:1-38.
[4] Rozvany GIN (2001). Aims, scope, methods, history and unified terminology of computer-aided topology optimization in structuralmechanics [J]. *Struct Multidiscip Optim* , 21(2): 90–108.
[5] Bendsøe M.P., Sigmund O(2003). Topology optimization: theory, methods and applications [M], 2nd edn. Springer, Berlin.
[6] Hassani B, Hinton E(1998). A review of homogenization and topology optimization —homogenization theory for media with periodic structure[J]. *Computers & Structures*, 69(6): 707-756.
[7] Cao, M.J., Ma, H.T., Wei, P(2015). A novel robust design method for improving stability of optimized structures[J]. *Acta Mechanica Sinica* , 31(1):104-111
[8] Zhang H., Liu S.T., Zhang X(2010). Topology optimization of 3D structures with design-dependent loads[J]. *Acta Mechanica Sinica* , 26:767–775.
[9] Sigmund O (2001). A 99 line topology optimization code written in Matlab [J]. *Struct Multidisc Optim*, 21(2): 120 -127
[10] Bendsøe M.P., Sigmund O(1999). Material interpolation schemes in topology optimization[J]. *Arch Appl Mech*, 69 (9–10): 635–654.
[11] Xie Y.M., Steven G.P(1993). A simple evolutionary procedure for structural optimization. *Comput Struct,* 49(5):885–896.
[12] Xie Y.M., Steven G.P(1997). Evolutionary structural optimization. Springer.
[13] Rozvany GIN, Querin O.M., Gaspar Z, Pomezanski V(2003). Weight increasing effect of topology simplification. *Struct Multidisc Optim* , 25:459-465.
[14] Luo J, Luo Z, Chen S, Tong L, Wang MY(2008). A new level set method for systematic design of hinge-free compliant mechanisms. *Comput Methods Appl Mech Eng* , 198(2): 318 –331.
[15] Wang M.Y., Chen S, Wang X, Mei Y(2005). Design of multimaterial compliant mechanisms using level-set methods. *J Mech Des ,* 127: 941–956.
[16] Wang M.Y., Wang X, Guo D(2003). A level set method for structural topology optimization[J]. *Comput Methods Appl Mech Eng* , 192(1–2): 227–246.
[17] Diaz A, Kikuchi N(1992). Solution to shape and topology eigenvalue optimization problems using a homogenization method[J]. *International Journal for Numerical Methods in Engineering*, 35:1487-1502
[18] Ma Z.D., Cheng H.C., Kikuchi N(1994). Structural Design for Obtaining Desired Frequencies by Using the Topology and Shape Optimization Method[J]. *Computer Systems in Engineering*. 5(1): 77-89
[19] Ma Z.D., Kikuchi N., Cheng H.C(1995). Topological design for vibrating structures[J]. *Computer Methods in Applied Mechanics and Engineering*. 121(1-4): 259-280
[20] Kosaka I., Swan C.C(1999). A summetry reduction method for continuum structural topology optimization[J]. *Comput Struct* 70:47-61
[21]Krog L.A., Olhoff N(1999). Optimum topology and reinforcement design of disk and plate structures with multiple stiffness and eigenfrequency objectives[J]. *Comput Struct* 72:535-563
[22] Jensen J. S., Pedersen N. L (2006). On maximal eigenfrequency separation in two-material structures: the 1D and 2D scalar cases[J]. *Journal of Sound and Vibration*. 289(4): 967-986.
[23] Pedersen N. L. Maximization of eigenvalues using topology optimization[J]. *Struct Multidisc Optim.* 20(1): 2-11. (2000)
[24]Zhu J.H., Zhang W.H (2006). Maximization of structural natural frequency with optimal support layout[J]. *Struct Multidisc Optim* 31:462-469
[25] Du J. B., Olhoff N(2007). Topological design of freely vibrating continuum structures for maximum values of simple and multiple eigenfrequencies and frequency gaps [J]. *Struct Multidisc Optim.* 34: 91-110
[26] Niu B., Yan J., Cheng G.D(2009). Optimum structure with homogeneous optimum cellular material for maximum fundamental frequency[J]. *Struct Multidisc Optim,* 39: 115-132
[27]Huang X., Zuo Z.H., Xie Y.M(2010). Evolutionary topological optimization of vibrating continuum structures for natural frequencies[J]. *Computers & Structures*. 88: 357-364
[28] Qi X., Shi T.L., Wang M.Y(2011). A level set based shape and topology optimization method for maximizing simple or repeated first eigenvalue of structure vibration[J]. *Struct Multidisc Optim,* 43: 473–485
[29] Nandy A. K., Jog C.S (2012). Optimization of vibrating structures to reduce radiated noise[J]. *Struct Multidisc Optim.* 45(5): 717-728
[30] Zheng J., Long S., Li G(2012). Topology optimization of free vibrating continuum structures based on the element free Galerkin method[J]. *Struct Multidisc Optim*.45(1):119-127
[31] Tsai T.D., Cheng C. C (2013). Structural design for desired eigenfrequencies and mode shapes using topology optimization[J]. *Struct Multidisc Optim* 47: 673-686
[32] Park J.Y., Han S.Y (2013). Application of artificial bee colony algorithm to topology optimization for dynamic stiffness problems[J]. *Computers and Mathematics with Applications*, 66: 1879-1891
[33] Joshua D., Deaton, Ramana V. Grandhi(2014). A survey of structural and multidisciplinary continuum topology optimization: post 2000. *Struct Multidisc Optim* 49:1-38

[34] Sui, Y. K(1996). Modeling Transformation and Optimization New Developments of Structural Synthesis Method. *Dalian Univer-sity of Technology Press*,Dalian (in Chinese)
[35] Sui Y.-K., Ye H.-L., Peng X.-R. (2006). Topological Optimization of Continuum Structure with Global Stress Constraints Based on ICM Method. The International Conference on Computational Methods，December 15-17, 2004, Singapore. COMPUTATIONAL METHODS, PTS 1 AND 2 : 1003-1014.
[36] Sui Y.-K., Ye H.-L. (2013). Continuum Topology Optimization Methods ICM [M]. Beijing: Science Press.
[37] Svanberg K (1987) The method of moving asymptotes—a new method for structural optimization. *International journal for numerical methods in engineering* 24(2): 359-373
[38] Svanberg K (1995) A globally convergent version of MMA without linesearch. In: Proceedings of the first world congress of structural and multidisciplinary optimization. Goslar, Germany

Comparisons of Limiters in Discontinuous Galerkin Method

***Su Penghui, Hu Pengju, Zhang Liang**

China Academy of Aerospace Aerodynamics, Beijing, China

*Presenting author: drsubest@163.com

Abstract

The discontinuous Galerkin method (DGM) is an important numerical method in computational fluid dynamics. The characteristics of DGM include its flexibility to construct high order schemes by using high order basis functions, and its compactness regardless of basis function orders. In supersonic simulations, the DGM often perform severe oscillations in regions where strong discontinuity solutions appear, so the slope limiters become necessary. In this study, three slope limiters are considered: TVB limiter, WENO limiter and HWENO limiter. The performance of these limiters are compared and analyzed with two dimensional supersonic cylinder flows. The results of show that all these limiters are able to stabilize the solution procedure, but the solutions show some differences between these limiters. Explanations as well as possible improvements are given.

Keywords: Discontinuous Galerkin Method, Supersonic Flow, Slope Limiter

Introduction

The discontinuous Galerkin method was first proposed by Reed and Hill [1] for neutron transportation problems, since then, the application of this method is widely extended. The applications include fluid simulations, MHD simulations, shallow water simulations and many others. In the area of supersonic flow simulations, the traditional methods include finite volume method and finite difference method, both of these methods have defects in modern supersonic flow simulations, the finite difference method has its weakness of dealing with complex geometric shapes, and the finite volume method has its weakness of constructing high order scheme on unstructured meshes. The DGM could overcome both of the defects of traditional methods. By introducing element-wise polynomial basis functions and inter-cell numerical fluxes, the DGM could have compact stencil on complex geometries. These characteristics make it an ideal candidate of next generation supersonic flow simulations.

When DGM is utilized in supersonic flow simulations with strong shockwaves, numerical instability is a major problem of this scheme, which will cause non-physical oscillations and divergent solutions. Many possible solutions have been proposed to overcome this defect, artificial viscosity and slope limiters are the two main approaches. In this study, slope limiters are utilized to suppress non-physical oscillations. Slope limiters were adopted into DGM by a collective effort of many researchers [2]-[6]. Slope limiters will detect severe oscillations of solutions and smooth them with smoother polynomials. In this study, the performances of TVB limiter [2], WENO limiter [5] and HWENO limiter [6] in shockwave regions are compared.

Governing Equations

The governing equations of two dimensional inviscid supersonic flows are Euler equations are. The conservation forms of these equations are:

$$\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}=0 \tag{1}$$

where *U*, *F* and *G* refer to conservative state vector, *x*-direction inviscid flux and *y*-direction inviscid flux respectively.

$$U=\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ \rho E\end{bmatrix};\ F=\begin{bmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ (\rho E+p)u\end{bmatrix};\ G=\begin{bmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ (\rho E+p)v\end{bmatrix} \tag{2}$$

To enclose the equation system, the equation of state is introduced.

$$p=\rho RT \tag{3}$$

Discontinuous Galerkin Method

The physical domain Ω_h is divided into non-overlapping elements *K*, where $\cup K=\Omega_h$. A reference element K' is introduced to simplify numerical integrations, the reference element and physical element are connected with coordinates mapping.

$$F_K : K' \to K : \xi \mapsto x=\sum_{i=1}^{m} x_i \chi_i(\xi) \tag{4}$$

where *m*, χ_i, and x_i refer to number of element interpolation functions, element shape functions and shape function coefficients respectively.

At any moment *t*, the unknowns $U_h(\xi,t)$ on reference element can be expressed in basis function space $span\{\psi_i(\xi)\}; i=1,..,m$.

$$U_h(\xi,t)=\sum_{i=1}^{m} U_i(t)\psi_i(\xi) \tag{5}$$

In each element, the weak form of governing equations is introduced.

$$\int_{K_j}\left(\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}\right)\phi_i \mathrm{d}\Omega=0 \quad ; \quad (i=1,...,m) \tag{6}$$

with some manipulations, the equations have the following form.

$$\begin{aligned}&\int_{K_j}\frac{\partial U}{\partial t}\phi_i \mathrm{d}\Omega-\int_{K_j}\left(F\frac{\partial \phi_i}{\partial x}+G\frac{\partial \phi_i}{\partial y}\right)\mathrm{d}\Omega\\ &+\int_{\partial K_j}\phi_i\left(Fn_x+Gn_y\right)\mathrm{d}S=0 \quad ; \quad (i=1,...,m)\end{aligned} \tag{7}$$

where *F* and *G* refer to inter-cell fluxes, in DGM, the solution has multiple values on element boundaries, in order to determine the value of inter-cell fluxes, numerical flux functions are introduced, in this study, the Van Leer flux [7] is adopted to calculate fluxes *F* and *G*.

In each time step, a system of ordinary differential equation is formed:

$$\int_{K_j} \frac{\partial U}{\partial t}\phi_i \mathrm{d}\Omega = \mathrm{Res}(K_j;i); \quad (i=1,...,m) \tag{8}$$

where $\mathrm{Res}(K_j;i)$ is residual term.

$$\mathrm{Res}(K_j;i) = \int_{K_j}\left(F\frac{\partial \phi_i}{\partial x}+G\frac{\partial \phi_i}{\partial y}\right)\mathrm{d}\Omega - \int_{\partial K_j}\phi_i\left(Fn_x+Gn_y\right)\mathrm{d}S; \quad (i=1,...,m) \tag{9}$$

so the equations become:

$$M_{K_j}\frac{dU}{dt} = \mathrm{Res}(K_j;i) \quad ; \quad (i=1,...,m) \tag{10}$$

or

$$\frac{dU}{dt} = M_{K_j}^{-1}\mathrm{Res}(K_j;i) \quad ; \quad (i=1,...,m) \tag{11}$$

where M_{K_j} refers to the mass matrix on element K_j.

A third order explicit Runge-Kutta scheme is introduced to solve this ordinary equations system.

$$\begin{aligned} U^{(1)} &= U^n + \Delta t M^{-1}R(U^n), \\ U^{(2)} &= \frac{3}{4}U^n + \frac{1}{4}\left[U^{(1)} + \Delta t M^{-1}R(U^{(1)})\right] \\ U^{(3)} &= \frac{1}{3}U^n + \frac{2}{3}\left[U^{(2)} + \Delta t M^{-1}R(U^{(2)})\right] \\ U^{n+1} &= U^{(3)} \end{aligned} \tag{12}$$

where $U^n = U(t)$, $U^{n+1} = U(t+\Delta t)$.

When high order basis functions are introduced, the dissipation of DGM is not enough to suppress numerical oscillations near strong discontinuity regions. In order to eliminate non-physical oscillations in numerical solutions, slope limiters are adopted. The TVB limiter, WENO limiter and HWENO limiter are commonly used.

The TVB limiter limits the first order components of solutions.

$$\bar{m}\left(a_1,a_2,\cdots,a_m\right) = \begin{cases} a_1, & |a_1| \le M\Delta x^2 \\ \mathrm{m}\left(a_1,a_2,\cdots,a_m\right), & |a_1| > M\Delta x^2 \end{cases} \tag{13}$$

where M refers to a problem dependent constant, and m is minmod function.

$$\mathrm{m}\left(a_1,a_2,\cdots,a_m\right) = \begin{cases} s\min_i|a_i|, & \text{if} \quad s=\mathrm{sign}(a_1)=\mathrm{sign}\left(a_2\right)=\cdots=\mathrm{sign}\left(a_m\right) \\ 0, & \text{else} \end{cases} \tag{14}$$

In WENO limiter, the average solutions of adjacent elements are used to reconstruct smooth solutions. If an element K_0 has three adjacent elements K_1 K_2 and K_3, the reconstruction stencils of polynomial P_1 for this element are $K_0K_1K_2$, $K_0K_1K_3$ and $K_0K_2K_3$.

$$\begin{aligned}
&\frac{1}{|K_0|}\int_{K_0} P_1 d\Omega = q_{K_0} \\
&\frac{1}{|K_m|}\int_{K_m} P_1 d\Omega = q_{K_m} \\
&\frac{1}{|K_n|}\int_{K_n} P_1 d\Omega = q_{K_n}, (m,n) = (1,2),(2,3),(1,3)
\end{aligned} \tag{15}$$

The HWENO limiter takes gradients of solutions into consideration, for an element K_0 with adjacent elements K_1, K_2 and K_3， four additional Hermite polynomials are constructed with stencils K_0K_0, K_0K_1, K_0K_2 and K_0K_3.

$$\begin{aligned}
&\frac{1}{|K_0|}\int_{K_0} P_1 d\Omega = q_{K_0} \\
&\frac{1}{|K_s|}\int_{K_s} \frac{\partial P_1}{\partial x_i} d\Omega = \left.\frac{\partial P_1}{\partial x_i}\right|_{K_s}, s = 0,1,2,3
\end{aligned} \tag{16}$$

the new solution P is reconstructed based on polynomial $P^{(i)}$ and weight w_i.

$$P = \sum_{i=1}^{m} w_i P^{(i)} \tag{17}$$

The WENO and HWENO limiters get a better performance if they are activated only on strong discontinuity regions. In this study, a shock detector [8] is introduced to indicate problem elements, on which the WENO or HWENO limiter is activated.

Numerical Results

Supersonic cylinder flow is chosen as test case for the performance of limiters. There is a strong shockwave in front of the cylinder, which will test the stability of numerical schemes. The radius of cylinder is 0.01, inflow Mach number is 3, and the non-dimensional inflow parameters are: $\rho = 1$, $u = 1$, $v = 0$, $p = 1/\left(\gamma M^2\right)$, Fig.1 shows the sketch of computational mesh, in order to perform large-scale numerical simulations on parallel computers, the mesh is partitioned into 40 sub-domains using Metis software package, 40 processers are utilized to speed up the solution procedure. In order to increase the converge speed to steady state solution, local time stepping method is introduced.

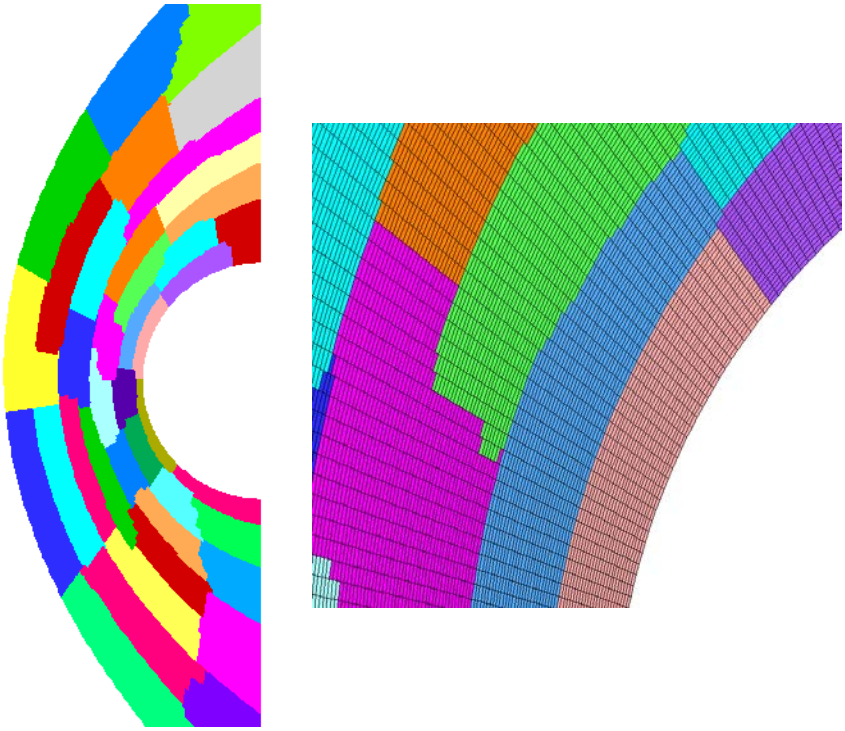


Figure 1. Computational mesh partitions(left) and its details (right)

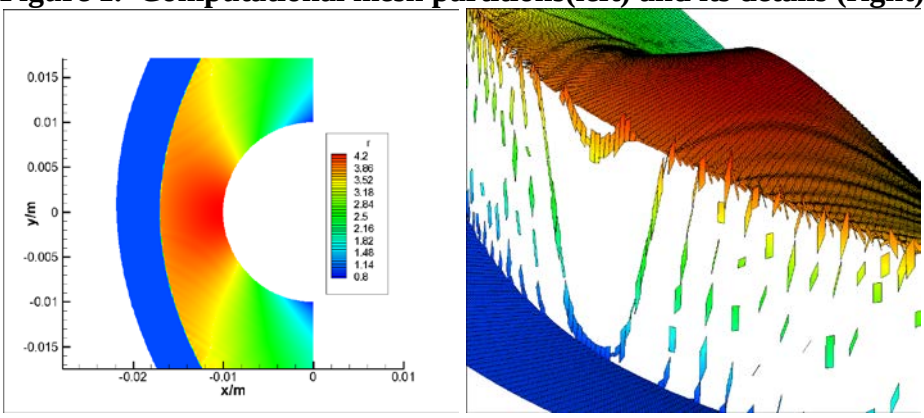


Figure 2. Density contour (left) and 3D view (right), with TVB limiter

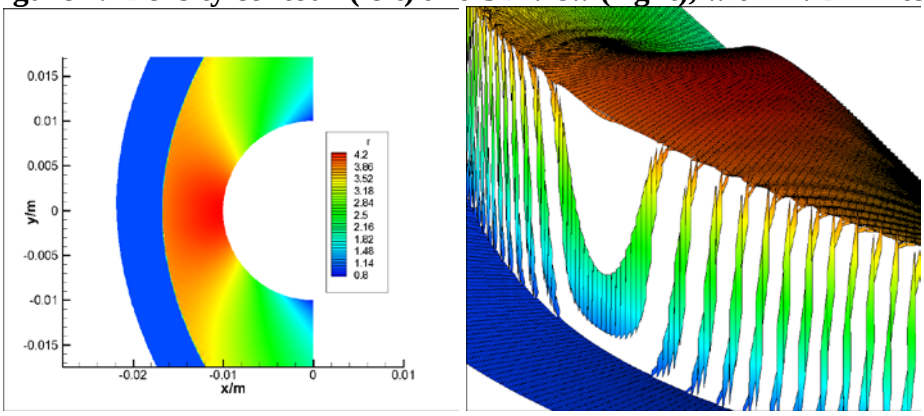


Figure 3. Density contour (left) and 3D view (right), with WENO limiter

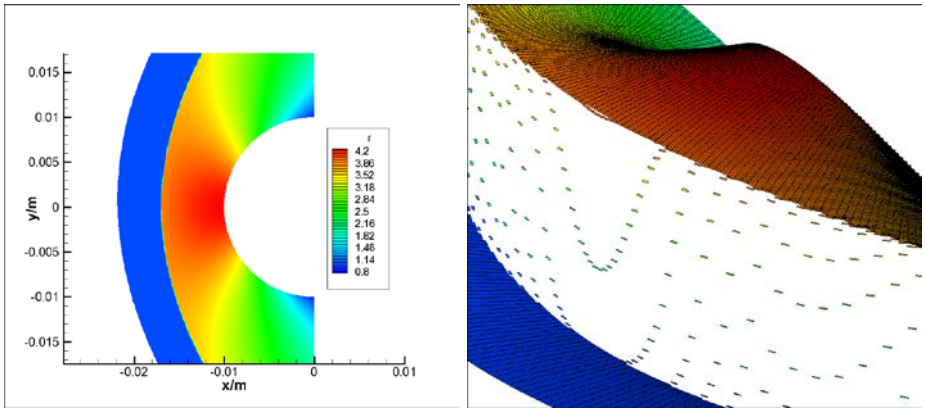


Figure 4. Density contour (left) and 3D view (right), with HWENO limiter

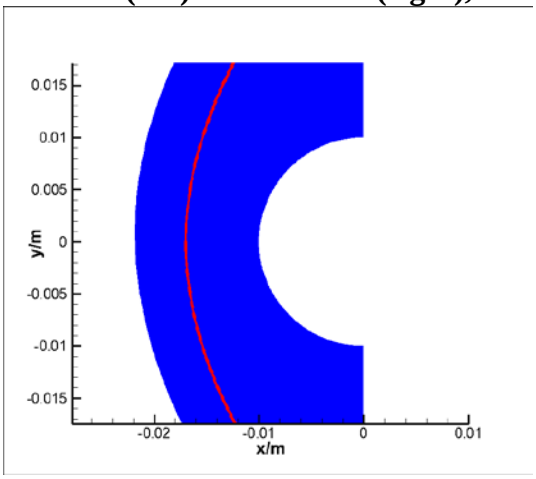


Figure 5. Elements on which limiter are activated

Density distributions and 3D density contours obtained by TVB, WENO and HWENO limiters are shown in Fig.2 to Fig.4. The results show that all these three limiters could stabilize the solution when strong shockwave appears, and capture the shockwave within few elements. In the shockwave regions, the density distribution with TVB limiter shows small overshoot, while the density distribution with WENO and HWENO limiter shows no overshoot. And the HWENO limiter gives more smooth solution than WENO and TVB limiters. Fig.5 shows the shockwave detected by shock detector, the red colored elements indicate that there are shockwaves, limiters are activated only on these elements.

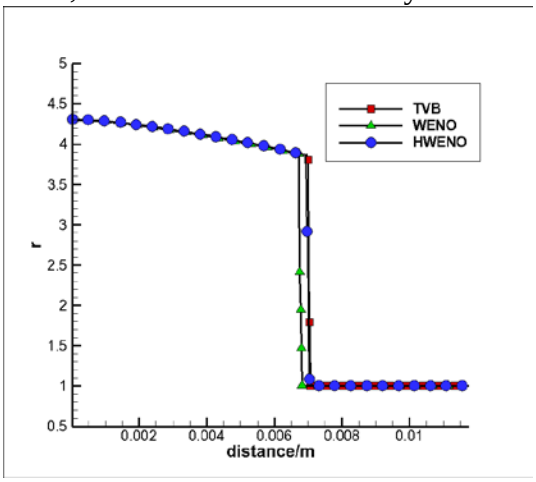


Figure 6. Density distributions along stagnation line.

The density distributions along stagnation line are shown in Fig.6, the density distributions before and after shockwave are identical for all the three limiters. But the shockwave position

predicted with WENO limiter is more closer to the cylinder than the other limiters, the cause of these differences need more investigations.

Conclusions

In this study, the performance of slope limiters in discontinuous Galerkin method are compared and analyzed with two dimensional supersonic cylinder flows. The results show that all these limiters are able to stabilize the solution procedure, in shockwave regions the density fields predicted with WENO and HWENO limiters are smoother than TVB limiter and contain no overshoot. In supersonic simulations, the WENO and HWENO limiters show better performances in suppressing non-physical oscillations and obtaining smooth solutions.

References

[1] Reed, W. H. and Hill, T. R. (1973) Triangular mesh methods for the neutron transport equation, *Los Alamos Scientific Laboratory Report*, LA-UR-73-479.
[2] Cockburn, B. and Shu, C. W. (1989) TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws II: General framework, *Mathematics of Computation* **52**, 411-435.
[3] Cockburn, B. and Shu, C. W. (1998) The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional system, *Journal of Computational Physics* **141**, 199–224.
[4] Kuzmin, D. and Turek, S. (2004) High-resolution FEM-TVD schemes based on a fully multidimensional flux limiter, *Journal of Computational Physics* **198**, 131–158.
[5] Yulong Xing and Chi-Wang, Shu. (2006) High order well-balanced finite volume WENO schemes and discontinuous Galerkin methods for a class of hyperbolic systems with source terms, *Journal of Computational Physics* **214**, 567–598.
[6] Hong, L., Joseph, D. B. and Rainald, L. (2007) A Hermite WENO-based limiter for discontinuous Galerkin method on unstructured grids, *Journal of Computational Physics* **225**, 686–713.
[7] Toro, E. F. (2009) Riemann Solvers and Numerical Methods for Fluid Dynamics –A Practical Introduction, Springer, New York.
[8] L. Krivodonova, et al. (2004) Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws, *Appl. Numer. Math* **48**.

Efficient multi-domain bivariate spectral collocation solution for MHD laminar natural convection flow from a vertical permeable flat plate with uniform surface temperature and thermal radiation

S. Mondal[1], S.P. Goqo†* [1], P. Sibanda[1] and S.S. Motsa[1]

[1]School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa
[1]*Presenting author: spgoqo@gmail.com
[1]†Corresponding author: spgoqo@gmail.com

Abstract

A recently developed numerical method, multidomain quasilinearisation method, is applied on a steady laminar, natural convection boundary layer flow of MHD viscous and incompressible fluid from a vertical permeable flat plate with uniform temperature in this paper. Non-dimensionless variables are used to transform the governing equations to a system of non-dimensional nonlinear partial differential equations. Then the resulting equations are solved numerically by using multidomain quasilinearisation method. The numerical results for tangential velocity, transverse velocity, and temperature, skin friction and Nusselt number are calculated and shown in a table and in various graphs.

Keywords: Natural convection; Magnetohydrodynamics; Multi-domain; Thermal radiation; Boundary layer.

Nomenclature

B	Magnetic induction
C_{fx}	Local skin friction
e	Electronic charge
E	Intensity of electric field
g	Gravitational acceleration
Gr_x	Modified Grashof number
H	Magnetic intensity
J	Electric current density
m	Hall parameter
M	Magnetic parameter
Nu_x	Local Nusselt number
p	Pressure
P_r	Prandtl number
T	Temperature of the fluid
$T\infty$	Free stream temperature
x, y, z	Co-ordinate directions
u, v, w	Velocity components in x, y, z directions
v	Velocity component normal to u
V	Transpiration velocity
x	Axial coordinate
y	Coordinate normal to x
q_r	Thermal radiation
R	Thermal radiation parameter

Greek symbols

α	Thermal diffusivity
β	Volumetric expansion coefficient for temperature
ψ	Stream function
θ	Dimensionless temperature function
ρ	Density
ν	Kinematic viscosity
μ	Dynamic viscosity
ξ	Transpiration parameter
η	Pseudo similarity variable

Subscripts

w	Conditions at wall
∞	Conditions far away from wall

Introduction

In many industrial processes, the study of magnetohydrodynamics natural convection flow and heat transfer has attracted considerable attention during the last decades. This is due to its applications which are found in MHD generator, flight MHD, Plasma studies, nuclear reactors, geothermal extractions, Hall accelerators and boundary layer control in the field of aeronautics and aerodynamics. Another important application of magnetohydrodynamic natural convection boundary layer flow past a semi-infinite vertical permeable flat plate with uniform mass flux is in space flight and in nuclear reactor. This applications normally requires a strong magnetic field and a low density gas and therefore the Hall current and ion slip becomes important.

The natural convection boundary layer flow from a vertical wall with Hall current and heat flux has been discussed by Sato [1], Yamanishi [2], Sherman and Sutton [3], Sing and Cowling [4], Sparrow and Cess [5], Gupta [6]. Free convection flow of a conducting fluid permeated by a transverse magnetic field was studied by Katagiri [7]. It has been observed by Singh and Cowling [4] that regardless of the strength of the applied magnetic field there will always be a region in the neighborhood of the leading edge of the plate where electromagnetic force are unimportant, whilst at large distances from the leading edge this magnetic force dominate. Pop and Watanabe analyzed the free convection flow of a conducting fluid permeated by a transverse magnetic field in the presence of Hall effects and uniform magnetic field.

Numerical solutions of MHD convection and mass transfer flow of viscous incompressible fluid were studied by Wahiduzzaman et al. [9]. They assumed that the induced magnetic field is negligible compared with the imposed magnetic field. Saha et al also studied the effect of Hall current on the steady, laminar, natural convection boundary layer flow of MHD viscous and incompressible fluid from a semi-infinite heated permeable vertical flat plate with an applied magnetic field transverse to it has been investigated, assuming that the induced magnetic field is negligible compared to the imposed magnetic field.

In the design of nuclear plants, gas turbines, propulsion devises for aircraft, missiles, satellites, and space vehicles, radiative heat transfer is a very important factor. This is due to the non-isothermal effects where high temperature is involved. Most studies that involve thermal radiation have been mostly limited to a stretching sheet. Some of the important investigations involving thermal radiation effects can be found in, for example, Englang and Emery [10], Gorla and Pop [11], Raptis [12], Abd El-Aziz [13, 14, 15].

Most of these studies rely on traditional numerical methods which requires the use of many grid points for accurate solutions. This is the result of the presence of local variable ξ which does not give accurate results for, usually, values of $\xi > 1$ [17]. The present study attempts to obtain the accurate solution with the use of few grid points.

It has been demonstrated that the finite difference method gives the solutions for all large values of transpiration parameter ξ. However, nonsimilarity method cannot give solutions for large values of ξ [16]. The aim of this paper is to give an alternative method that will handle solutions for large values of ξ when nonsimilarity transformation methods are used.

Problem Formulation

Consider the steady natural convection boundary layer flow of an electrically conducting and viscous incompressible fluid from a semi-infinite heated permeable vertical flat plate in presence of magnetic field and thermal radiation with the effect of Hall currents.

Applying the Boussinesq approximation, the boundary layer equations governing the flow under the assumption that the fluid is quasi-neutral and ion slip and thermoelectric effect results in the following system of equations:

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

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \nu\frac{\partial^2 u}{\partial y^2} + g\beta(T - T_\infty) - \frac{\sigma B_0^2}{\rho(1+m^2)}(u + mw), \tag{2}$$

$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} = \nu\frac{\partial^2 w}{\partial y^2} + \frac{\sigma B_0^2}{\rho(1+m^2)}(mu - w), \tag{3}$$

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \alpha\frac{\partial^2 T}{\partial y^2} - \frac{1}{\rho c_p}\frac{\partial q_r}{\partial y}, \tag{4}$$

$$\tag{5}$$

where u, v and w are the velocities in the $x-$,$y-$ and $z-$ direction, T is the fluid temperature, $\nu(= \mu/\rho)$ is the kinematic coefficient of viscosity, μ is the fluid viscocity and ρ is the fluid density, $\alpha(= \kappa/\rho c_p)$ is the thermal diffusivity with κ being the fluid thermal conductivity and c_p is the heat capacity of the fluid at constant pressure, q_r is the thermal radiative heat flux, $m(= \omega^2\tau^2)$ is the Hall parameter, with ω as the cyclotron frequency of electron and τ as collision time of electrons with ions.

The radiative heat flux q_r under Rosseland approximation takes the form

$$q_r = -\frac{4\sigma}{3k_1}\frac{\partial T^4}{\partial y}, \tag{6}$$

where σ is the Stefan-Boltzmann constant and k_1 is the mean absorption coefficient. Assuming that the temperature difference within the flow is sufficiently small, T^4 may be approximated in Taylor series form, after ignoring higher order terms, as follows:

$$T^4 = 4T_\infty^3 T - 3T_\infty^4. \tag{7}$$

Applying (6) and (7) to equation (4) we get

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \alpha\frac{\partial^2 T}{\partial y^2} - \frac{16\sigma T_\infty^3}{3k_1\rho c_p}\frac{\partial^2 T}{\partial y^2}. \tag{8}$$

The boundary conditions are:

$$u(x,y)=0,\ \ v(x,y)=-V_0,\ \ w(x,y)=0,\ \ T(x,y)=T_w\ \text{ at } y=0 \tag{9}$$

$$u(x,y)=0,\ \ v(x,y)=0,\ \ w(x,y)=0,\ \ T(x,y)=T_\infty\ \text{ at } y=\infty, \tag{10}$$

where V_0 is the transpiration velocity which is positive for suction and negative for injection.

The set of non-linear partial differential equations are transformed by introduction of dimensionless group of transformations for the dependent and independent variables applicable in natural convection flow from a vertical surface:

$$\psi(x,y)=\nu Gr_x^{1/4}[f(\xi,\eta)+\xi],\ \ \eta=\frac{y}{x}Gr_x^{1/4},\ \ \xi=\frac{V_0x}{\nu}Gr_x^{-1/4}, \tag{11}$$

$$w(x,y)=\frac{\nu}{x}Gr_x^{1/2}g(\xi,\eta),\theta=\frac{T-T_\infty}{T_w-T_\infty} \tag{12}$$

where ψ is the stream function, defined by

$$u=\frac{\partial\psi}{\partial y}\ \text{ and }\ v=-\frac{\partial\psi}{\partial x} \tag{13}$$

which satisfies the continuity condition (1). In the above equation (12) f is the dimensionless stream function, g is the dimensionless velocity and θ is the dimensionless temperature of the fluid. η is the pseudo-similarity variable and ξ is the transpiration parameter depending on the transpiration velocity V_0 and the axial variable x.

Applying these transformations to the system of equations (2) - (4), the resulting governing non-similarity system of partial differential equations are expressed in dimensionless form as [17]:

$$f'''+\frac{3}{4}ff''-\frac{1}{2}f'^2+\theta+\xi f''-\frac{M}{(1-m^2)}(f'+mg)=\frac{1}{4}\xi\left(f'\frac{\partial f'}{\partial\xi}-f''\frac{\partial f}{\partial\xi}\right), \tag{14}$$

$$g''+\frac{3}{4}fg'-\frac{1}{2}f'g+\xi g'-\frac{M}{(1-m^2)}(g-mf')=\frac{1}{4}\xi\left(f'\frac{\partial g}{\partial\xi}-g'\frac{\partial f}{\partial\xi}\right) \tag{15}$$

$$\frac{1}{Pr}\left(1+\frac{3}{4}R\right)\theta''+\frac{3}{4}f\theta'+\xi\theta'=\frac{1}{4}\xi\left(f'\frac{\partial\theta}{\partial\xi}-\theta'\frac{\partial f}{\partial\xi}\right) \tag{16}$$

where the local Grashof number, magnetic field number and thermal radiation parameter are, respectively, given by

$$Gr_x=\frac{g\beta\delta T}{\nu^2}x^3,\qquad M=\frac{\sigma B_0^2x^2}{\rho Gr_x^{\frac{1}{2}}},\qquad R=\frac{4\sigma T_\infty^3}{kk_1} \tag{17}$$

The primes in the above equations denoted differentiation with respect to η and the correspond-

ing boundary conditions are given by

$$f(0,\xi) = f'(0,\eta) = 0, \quad g(0,\xi) = \theta(0,\xi) = 1, \tag{18}$$

$$f'(\infty,\xi) = g(\infty,\xi) = \theta(\infty,\xi) = 0. \tag{19}$$

The physical quantities of interest in this case are the skin-friction, Nusselt and Sherwood numbers which are defined in [23] as

$$C_{fx}Gr_x^{-3/4} = f''(0,\xi), \quad Nu_xGr_x^{-1/4} = -\theta'(0,\xi), \tag{20}$$

respectively.

Bivariate Spectral Quasilinearisation Method (BSQLM)

In this section we first describe the standard bivariate spectral quasilinearisation method for solving coupled non-linear partial differential equations. The quasi-linearisation method is based on Taylor series expansion of system of equations about some previous approximation of the solution. The assumption used is that the difference between the current and previous solution is small. To illustrate the idea of the BSQLM we first write equations as

$$\Omega_k[H_1, H_2, H_3] = 0, \quad \text{for } \; k = 1, 2, 3, \tag{21}$$

where H_1, H_2 and H_3 represents equations (14), (15) and (16) respectively. The quasilinearisation scheme applied in equations (14) - (16) results in

$$a_{0r}f'''_{r+1} + a_{1r}f''_{r+1} + a_{2r}f'_{r+1} + a_{3r}f_{r+1} + a_{4r}\frac{\partial f_{r+1}}{\partial \xi} + a_{5r}\frac{\partial f'_{r+1}}{\partial \xi} + a_{6r}g_{r+1} + a_{7r}\theta_{r+1} = R_{1r}, \tag{22}$$

$$b_{0r}f'_{r+1} + b_{1r}f_{r+1} + b_{2r}\frac{\partial f_{r+1}}{\partial \xi} + b_{3r}g''_{r+1} + b_{4r}g'_{r+1} + b_{5r}g_{r+1} + b_{6r}\frac{\partial g_{r+1}}{\partial \xi} + b_{7r}\theta_{r+1} = R_{2r}, \tag{23}$$

$$c_{0r}f'_{r+1} + c_{1r}f_{r+1} + c_{2r}\frac{\partial f_{r+1}}{\partial \xi} + c_{3r}g_{r+1} + c_{4r}\theta''_{r+1} + c_{5r}\theta'_{r+1} + c_{6r}\theta_{r+1} + c_{7r}\frac{\partial \theta_{r+1}}{\partial \xi} = R_{3r}, \tag{24}$$

where

$$a_{0r}=1,\ \ a_{1r}=\frac{3}{4}f_r+\xi+\frac{1}{4}\xi\frac{\partial f_r}{\partial\xi},\ \ a_{2r}=-f_r'-\frac{M}{1-m^2}-\frac{1}{4}\xi\frac{\partial f_r'}{\partial\xi},$$
$$a_{3r}=\frac{3}{4}f_r'',\ \ a_{4r}=\frac{1}{4}\xi f_r'',\ \ a_{5r}=-\frac{1}{4}\xi f_r'',\ \ a_{6r}=-\frac{Mm}{1+m^2},\ \ a_{7r}=1,$$
$$b_{0r}=-\frac{1}{2}g_r+\frac{Mm}{1+m^2}-\frac{1}{4}\xi\frac{\partial g_r}{\partial\xi},\ \ b_{1r}=\frac{3}{4}g_r',\ \ b_{2r}=\frac{1}{4}\xi g_r',$$
$$b_{3r}=1,\ \ b_{4r}=\frac{3}{4}f_r+\xi+\frac{1}{4}\xi\frac{\partial f_r}{\partial\xi},\ \ b_{5r}=-\frac{1}{2}f_r'-\frac{M}{1+m^2},\ \ b_{6r}=-\frac{1}{4}\xi f_r',\ \ b_{7r}=0,$$
$$c_{0r}=-\frac{1}{4}\xi\frac{\partial\theta_r}{\partial\xi},\ \ c_{1r}=\frac{3}{4}\theta_r',\ \ c_{2r}=\frac{1}{4}\xi\theta_r',\ \ c_{3r}=0,$$
$$c_{4r}=\frac{1}{Pr}(1+\frac{4}{3}R),\ \ c_{5r}=\frac{3}{4}f_r+\xi+\frac{1}{4}\xi f_r'.$$
$$R_{1r}=a_{0r}f_r'''+a_{1r}f_r''+a_{2r}f_r'+a_{3r}f_r+a_{4r}\frac{\partial f_r}{\partial\xi}+a_{5r}\frac{\partial f_r'}{\partial\xi}+a_{6r}g_{r+1}+a_{7r}\theta_{r+1}-H_1,$$
$$R_{2r}=b_{0r}f_r'+b_{1r}f_r+b_{2r}\frac{\partial f_r}{\partial\xi}+b_{3r}g_r''+b_{4r}g_r'+b_{5r}g_r+b_{6r}\frac{\partial g_r}{\partial\xi}+b_{7r}\theta_r-H_2,$$
$$R_{3r}=c_{0r}f_r'+c_{1r}f_r+c_{2r}\frac{\partial f_r}{\partial\xi}+c_{3r}g_r+c_{4r}\theta_r''+c_{5r}\theta_r'+c_{6r}\theta_r+c_{7r}\frac{\partial\theta_r}{\partial\xi}-H_3.$$

Applying spectral collocation on (14) - (16) gives

$$A_{11}\boldsymbol{F}_i+a_{4r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{F}_j+a_{5r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{D}\boldsymbol{F}_j+A_{12}\boldsymbol{G}_i+A_{13}\boldsymbol{\theta}_i=\boldsymbol{R}_{1,i},\tag{25}$$
$$A_{21}\boldsymbol{F}_i+b_{2r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{F}_j+A_{22}\boldsymbol{G}_i+b_{6r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{G}_j+A_{23}\boldsymbol{\theta}_i=\boldsymbol{R}_{2,i},\tag{26}$$
$$A_{31}\boldsymbol{F}_i+c_{2r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{F}_j+A_{32}\boldsymbol{G}_i+A_{33}\boldsymbol{\theta}_i+c_{7r}\sum_{j=0}^{Nt}d_{ij}\boldsymbol{\theta}_j=\boldsymbol{R}_{3,i},\tag{27}$$

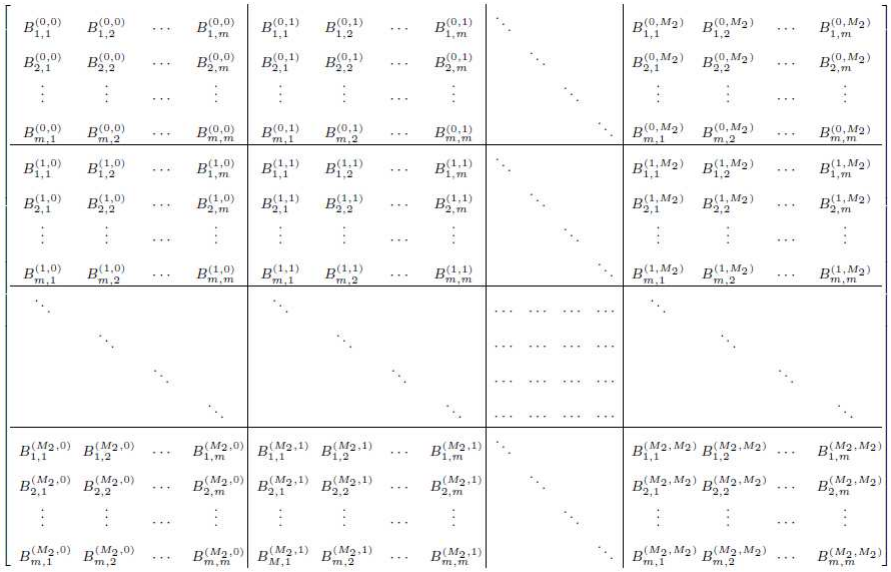
where

$$A_{11}^i=\boldsymbol{a}_{0r}\boldsymbol{D}^3+\boldsymbol{a}_{1r}\boldsymbol{D}^2+\boldsymbol{a}_{2r}\boldsymbol{D}+\boldsymbol{a}_{3r}\boldsymbol{I},\ \ A_{12}^i=\boldsymbol{a}_{6r}\boldsymbol{I},\ \ A_{13}^i=\boldsymbol{a}_{7r}\boldsymbol{I},$$
$$A_{21}^i=\boldsymbol{b}_{0r}\boldsymbol{D}+\boldsymbol{b}_{1r}\boldsymbol{I},\ \ A_{22}=\boldsymbol{b}_{3r}\boldsymbol{D}^2+\boldsymbol{b}_{4r}\boldsymbol{D}+\boldsymbol{b}_{5r}\boldsymbol{I},\ \ A_{23}=\boldsymbol{b}_{7r}\boldsymbol{I},$$
$$A_{11}^i=\boldsymbol{c}_{0r}\boldsymbol{D}+\boldsymbol{c}_{1r}\boldsymbol{I},\ \ A_{32}=\boldsymbol{c}_{3r}\boldsymbol{I},\ \ A_{33}=\boldsymbol{c}_{4r}\boldsymbol{D}^2+\boldsymbol{c}_{5r}\boldsymbol{D}+\boldsymbol{c}_{6r}\boldsymbol{I}.$$

For convenience, equations (25), (26) and (27) are expanded for $i=0,\ldots,M_2$ and rearranged to obtain the following matrix form

$$\mathbf{B}_r\boldsymbol{X}_{r+1}=\mathbf{R}_r\tag{28}$$

where the coefficient matrix $\mathbf{B}_r$ is defined as



where

$$
\begin{aligned}
B_{11}^{ii} &= A_{11}^{i} + \boldsymbol{a}_{4r} d_{ii} \boldsymbol{I} + \boldsymbol{a}_{5r} d_{ii} \boldsymbol{D}, \quad B_{12}^{ii} = A_{12}^{i}, B_{13}^{ii} = A_{13}^{i}, \\
B_{11}^{ij} &= \boldsymbol{a}_{4r} d_{ij} \boldsymbol{I} + a_{5r} d_{ij} \boldsymbol{D}, \quad B_{12}^{ij} = 0, \quad B_{13}^{ij} = 0, \\
B_{21}^{ii} &= A_{21}^{i} + \boldsymbol{b}_{2r} d_{ii} \boldsymbol{I} \quad B_{22}^{ii} = A_{22}^{i} + \boldsymbol{b}_{6r} d_{ii} \boldsymbol{I}, B_{23}^{ii} = A_{23}^{i}, \\
B_{21}^{ij} &= \boldsymbol{b}_{2r} d_{ij} \boldsymbol{I}, \quad B_{22}^{ij} = \boldsymbol{6r} d_{ij} \boldsymbol{I}, \quad B_{23}^{ij} = 0, \\
B_{31}^{ii} &= A_{31}^{i} + \boldsymbol{c}_{2r} d_{ii} \boldsymbol{I}, \quad B_{32}^{ii} = A_{32}^{i}, B_{33}^{ii} = A_{33}^{i} + \boldsymbol{c}_{7r} d_{ii} \boldsymbol{I}, \\
B_{31}^{ij} &= \boldsymbol{c}_{2r} d_{ij} \boldsymbol{I}, \quad B_{32}^{ij} = 0, \quad B_{33}^{ij} = \boldsymbol{c}_{7r} d_{ij} \boldsymbol{I},
\end{aligned}
\tag{29}
$$

The vectors $\boldsymbol{X}_{r+1}$ and $\mathbf{R}_r$ are defined as

$$
\begin{aligned}
\boldsymbol{X}_{r+1} &= \left[\mathbf{F}_{1,r+1}^{(0)} \mathbf{G}_{2,r+1}^{(0)} \cdots \boldsymbol{\theta}_{m,r+1}^{(0)} \left| \mathbf{F}_{1,r+1}^{(1)} \mathbf{G}_{2,r+1}^{(1)} \cdots \boldsymbol{\theta}_{m,r+1}^{(1)} \right| \cdots\cdots\cdots\cdots \left| \mathbf{F}_{1,r+1}^{(M_2)} \mathbf{G}_{2,r+1}^{(M_2)} \cdots \boldsymbol{\theta}_{m,r+1}^{(M_2)} \right.\right]^T \\
\mathbf{R}_r &= \left[\mathbf{R}_1^{(0)} \ \mathbf{R}_2^{(0)} \ \mathbf{R}_3^{(0)} \cdots \mathbf{R}_m^{(0)} \left| \mathbf{R}_1^{(1)} \ \mathbf{R}_2^{(1)} \ \mathbf{R}_3^{(1)} \cdots \mathbf{R}_m^{(1)} \right| \cdots\cdots\cdots\cdots \left| \mathbf{R}_1^{(M_2)} \ \mathbf{R}_2^{(M_2)} \ \cdots \ \mathbf{R}_m^{(M_2)} \right.\right]^T
\end{aligned}
$$

The approximate solutions are obtained by solving (28) iteratively for $r = 0, 1, 2, \ldots$. The inclusion of boundary conditions and multi-domain solution approach is discussed in the next section through a specific example.

Multi-domain bivariate spectral collocation method for systems of PDEs

It is well-known that the standard form of the bivariate spectral quasi-linearisation method described in [24] works well for problem defined over small domains. Large domains require proportionally larger number of nodes to yield accurate results. For the BSQLM, increasing the number of nodes increases the computational effort required to solve the matrix equations almost exponentially. A simple way of ensuring that accurate solutions are obtained efficiently over large domains is to seek to limit the size of the matrix equations. As can be noted from matrix equation (28), the size of the coefficient matrix for a system of m PDEs in m unknowns is $m(M_1+1)(M_2+1)$ by $m(M_1+1)(M_2+1)$, where M_1, M_2 give the number of nodes in the x_1 and x_2 domains, respectively. Below, we introduce a strategy that seeks to reduce the size of the matrix equations by ensuring that the value of M_2 is kept to be as low as possible. For problems where the largest order of the derivative with respect to x_2 is one this can be achieved by evalu-

ating the solution in a sequence of equal intervals, which are subject to continuity conditions at the end points of each interval.

To apply the multi-domain bivariate spectral quasi-linearisation method (MD-BSQLM) we divide the interval $\xi \in [0, \xi_P]$ into P sub-intervals $\Omega_e = [\xi_{e-1}, \xi_e]$ for $e = 1, 2, \ldots, P$ as shown in the illustration 1 below.

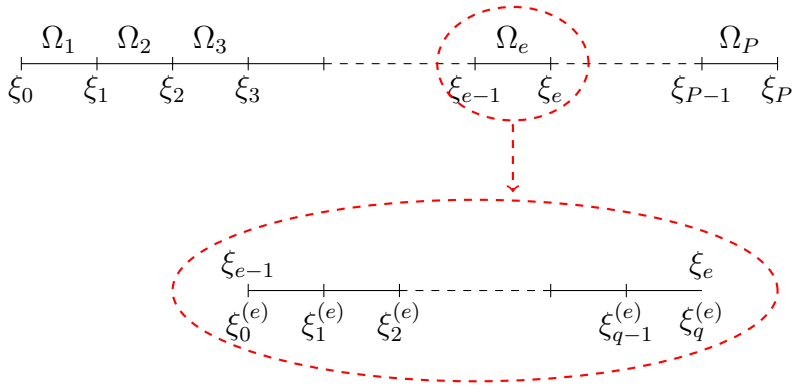


Figure 1: Multi-domain grid

Each interval Ω_e is further divided into q divisions which are not necessarily of equal spacing. The non-linear equations (25), (26) and (27) are solved in each subinterval $[\xi_{e-1}, \xi_e]$ with the solution denoted by $\overset{e}{f}(\eta, \xi)$ in this interval. In the first interval $[\xi_{e-1}, \xi_e]$, the solution is $\overset{1}{f}(\eta, \xi)$ is obtained subject to the "initial" condition $\overset{1}{f}(\eta, 0)$. For each $e \geq 2$, at each interval $[\xi_{e-1}, \xi_e]$, the continuity condition

$$\overset{e}{f}(\eta, \xi_{e-1}) = \overset{e-1}{f}(\eta, \xi_{e-1}) \tag{30}$$

is used to implement the BSQLM over the interval $[\xi_{e-1}, \xi_e]$. This process is repeated to generate a sequence of solutions $\overset{e}{f}(\eta, \xi)$ for $e = 1, 2, \ldots, P$

In our system, the number of equations and unknowns is $m = 3$ and the orders of the highest derivatives that are required as limits in the definition of the coefficient parameters and matrices are

$$n_{1,1} = 3, \quad n_{1,2} = 0, \quad n_{1,3} = 0, \quad n_{2,1} = 1, \quad n_{2,2} = 2 \quad n_{3,1} = 1, \quad n_{3,3} = 2$$

With these values, the coefficient parameters and matrices are obtained using the formulas given in the previous section and are defined in the appendix. Applying the spectral collocation gives

$$A_{11}\overset{e}{\boldsymbol{F}}_{i,r+1} + a_{4r}\sum_{j=0}^{M_2} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + a_{5r}\sum_{j=0}^{M_2} d_{ij}\boldsymbol{D}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{12}\overset{e}{\boldsymbol{G}}_{i,r+1} + A_{13}\overset{e}{\boldsymbol{\theta}}_{i,r+1} = \overset{e}{\boldsymbol{R}}_{1,i}, \tag{31}$$

$$A_{21}\overset{e}{\boldsymbol{F}}_{i,r+1} + b_{2r}\sum_{j=0}^{M_2} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{22}\overset{e}{\boldsymbol{G}}_{i,r+1} + b_{6r}\sum_{j=0}^{M_2} d_{ij}\overset{e}{\boldsymbol{G}}_{j,r+1} + A_{23}\overset{e}{\boldsymbol{\theta}}_{i,r+1} = \overset{e}{\boldsymbol{R}}_{2,i}, \tag{32}$$

$$A_{31}\overset{e}{\boldsymbol{F}}_{i,r+1} + c_{2r}\sum_{j=0}^{M_2} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{32}\overset{e}{\boldsymbol{G}}_{i,r+1} + A_{33}\overset{e}{\boldsymbol{\theta}}_{i,r+1} + c_{7r}\sum_{j=0}^{M_2} d_{ij}\overset{e}{\boldsymbol{\theta}}_{j,r+1} = \overset{e}{\boldsymbol{R}}_{3,i}, \tag{33}$$

where

$$\mathbf{F}_{i,r+1} = [f_{r+1}(\hat{\xi}_i, \hat{\eta}_0), f_{r+1}(\hat{\xi}_i, \hat{\eta}_1), f_{r+1}(\hat{\xi}_i, \hat{\eta}_2), \ldots, f_{r+1}(\hat{\xi}_i, \hat{\eta}_{M_1})]^T,$$
$$\mathbf{G}_{i,r+1} = [g_{r+1}(\hat{\xi}_i, \hat{\eta}_0), g_{r+1}(\hat{\xi}_i, \hat{\eta}_1), g_{r+1}(\hat{\xi}_i, \hat{\eta}_2), \ldots, g_{r+1}(\hat{\xi}_i, \hat{\eta}_{M_1})]^T,$$
$$\boldsymbol{\theta}_{i,r+1} = [\theta_{r+1}(\hat{\xi}_i, \hat{\eta}_0), \theta_{r+1}(\hat{\xi}_i, \hat{\eta}_1), \theta_{r+1}(\hat{\xi}_i, \hat{\eta}_2), \ldots, \theta_{r+1}(\hat{\xi}_i, \hat{\eta}_{M_1})]^T.$$

The boundary conditions for solving equations (39) - (41) are

$$\overset{e}{f}(\xi_i, \eta_{M_1}) = 0, \qquad \sum_{p=0}^{M_1} \mathbf{D}^{(1,0)}_{M_1,p} \overset{e}{f}(\xi_i, \eta_p) = 0, \qquad \overset{e}{g}(\xi_i, \eta_{M_1}) = \overset{e}{\theta}(\xi_i, \eta_{M_1}) = 1, \tag{34}$$

$$\sum_{p=0}^{M_1} \mathbf{D}^{(1,0)}_{0,p} \overset{e}{f}(\xi_i, \eta_p) = 0, \qquad \overset{e}{g}(\xi_i, \eta_0) = \overset{e}{\theta}(\xi_i, \eta_0) = 0. \tag{35}$$

The "*initial*" conditions at $\xi = 0$ ($\hat{\xi} = \hat{\xi}_{M_2} = -1$) are obtained by solving the following ODE set

$$f''' + \frac{3}{4} f f'' - \frac{1}{2} f'^2 + \theta - \frac{M}{(1-m^2)}(f' + mg) = 0, \tag{36}$$

$$g'' + \frac{3}{4} f g' - \frac{1}{2} f' g - \frac{M}{(1-m^2)}(g - mf') = 0 \tag{37}$$

$$\frac{1}{Pr}\left(1 + \frac{3}{4}R\right)\theta'' + \frac{3}{4} f \theta' = 0 \tag{38}$$

The solution of equation (36) - (38), in the first interval, are denoted by $\overset{1}{\mathbf{F}}_{M_2,r+1}$, $\overset{1}{\boldsymbol{G}}_{M_2,r+1}$ and $\overset{1}{\boldsymbol{\theta}}_{M_2,r+1}$. In the next intervals we solve the following equations

$$A_{11}\overset{e}{\boldsymbol{F}}_{i,r+1} + a_{4r}\sum_{j=0}^{M_2-1} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + a_{5r}\sum_{j=0}^{M_2-1} d_{ij}\boldsymbol{D}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{12}\overset{e}{\boldsymbol{G}}_{i,r+1} + A_{13}\overset{e}{\boldsymbol{\theta}}_{i,r+1} = \overset{e}{\boldsymbol{K}}_{1,i}, \tag{39}$$

$$A_{21}\overset{e}{\boldsymbol{F}}_{i,r+1} + b_{2r}\sum_{j=0}^{M_2-1} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{22}\overset{e}{\boldsymbol{G}}_{i,r+1} + b_{6r}\sum_{j=0}^{M_2-1} d_{ij}\overset{e}{\boldsymbol{G}}_{j,r+1} + A_{23}\overset{e}{\boldsymbol{\theta}}_{i,r+1} = \overset{e}{\boldsymbol{K}}_{2,i}, \tag{40}$$

$$A_{31}\overset{e}{\boldsymbol{F}}_{i,r+1} + c_{2r}\sum_{j=0}^{M_2-1} d_{ij}\overset{e}{\boldsymbol{F}}_{j,r+1} + A_{32}\overset{e}{\boldsymbol{G}}_{i,r+1} + A_{33}\overset{e}{\boldsymbol{\theta}}_{i,r+1} + c_{7r}\sum_{j=0}^{M_2-1} d_{ij}\overset{e}{\boldsymbol{\theta}}_{j,r+1} = \overset{e}{\boldsymbol{K}}_{3,i}, \tag{41}$$

where

$$\overset{e}{\mathbf{K}}_{1,i} = \overset{e}{\mathbf{R}}_{1,i} - a_{4r} d_{iM_2}\overset{e}{\boldsymbol{F}}_{M_2,r+1} - a_{5r} d_{iM_2}\boldsymbol{D}\overset{e}{\boldsymbol{F}}_{M_2,r+1}, \tag{42}$$

$$\overset{e}{\mathbf{K}}_{2,i} = \overset{e}{\mathbf{R}}_{2,i} - b_{2r} d_{iM_2}\overset{e}{\boldsymbol{F}}_{M_2,r+1} - b_{6r} d_{iM_2}\overset{e}{\boldsymbol{G}}_{M_2,r+1}, \tag{43}$$

$$\overset{e}{\mathbf{K}}_{3,i} = \overset{e}{\mathbf{R}}_{3,i} - c_{2r} d_{iM_2}\overset{e}{\boldsymbol{F}}_{M_2,r+1} - c_{7r} d_{iM_2}\overset{e}{\boldsymbol{\theta}}_{M_2,r+1}. \tag{44}$$

The continuity conditions in this example are given by

$$\begin{aligned} \overset{e}{f}(\eta, \xi_{e-1}) &= \overset{e-1}{f}(\eta, \xi_{e-1}), \\ \overset{e}{g}(\eta, \xi_{e-1}) &= \overset{e-1}{g}(\eta, \xi_{e-1}), \\ \overset{e}{\theta}(\eta, \xi_{e-1}) &= \overset{e-1}{\theta}(\eta, \xi_{e-1}), \end{aligned} \tag{45}$$

Applying the continuity conditions on (42) - (44) gives

$$\overset{e}{\mathbf{K}}_{1,i} = \overset{e}{\mathbf{R}}_{1,i} - a_{4r} d_{iM_2} \overset{e-1}{\boldsymbol{F}}_{M_2,r+1} - a_{5r} d_{iM_2} \boldsymbol{D} \overset{e-1}{\boldsymbol{F}}_{M_2,r+1}, \tag{46}$$

$$\overset{e}{\mathbf{K}}_{2,i} = \overset{e}{\mathbf{R}}_{2,i} - b_{2r} d_{iM_2} \overset{e-1}{\boldsymbol{F}}_{M_2,r+1} - b_{6r} d_{iM_2} \overset{e-1}{\boldsymbol{G}}_{M_2,r+1}, \tag{47}$$

$$\overset{e}{\mathbf{K}}_{3,i} = \overset{e}{\mathbf{R}}_{3,i} - c_{2r} d_{iM_2} \overset{e-1}{\boldsymbol{F}}_{M_2,r+1} - c_{7r} d_{iM_2} \overset{e-1}{\boldsymbol{\theta}}_{M_2,r+1}. \tag{48}$$

Results and Discussion

The natural convection flow from a vertical permeable equations are derived and solved using multi-domain bivariate spectral collocation method. This is done taking into account the normal magnetic field to the surface of the plates. Also, thermal radiation and the Hall current effects are taken into consideration.

	Saha et al. [17]		MBQLM	
ξ	$f''(0,\xi)$	$\theta'(0,\xi)$	$f''(0,\xi)$	$\theta'(0,\xi)$
2	0.706	1.4028	0.7088928	1.4026916
10	–	–	0.1428570	7.0000000
20	0.0714	13.9995	0.0714227	14.0000000
40	0.0357	27.9985	0.0340195	27.9997087
50	0.0285	349981	0.0247535	34.9964783
60	0.0238	41.9977	0.0182251	41.9806430
70	0.0204	48.9974	0.0140161	48.9328200
80	0.0178	55.9971	0.0115996	55.8259543

Table 2: Comparison of Multi-domain solution local skin friction and the Nusselt number against the transpiration parameter ξ while $Pr = 0.7$, $M = 0.5$, $m = 100$ against the Saha et al results [17]

Table 1 shows the comparison results between the current results and the literature results [17]. The table displays the local skin friction and the Nusselt number with respect to the transpiration parameter ξ ranging from 0 to 80 while $Pr = 0.7$, $M = 0.5$, $m = 100$. It is observed that for the increasing value of the transpiration parameter xi the value of the local skin friction coefficient turn to increase near the leading edge, and then diminished slowly. The local Nusselt number coefficient increases rapidly. This observation validates that the solutions of large transpiration number are in agreement with the literature [17].

We also look at the residual error results in order to ensure that our numerical scheme is accurate. The convergence error results are shown in Figures 2, 3 and 4 for velocity, temperature and temperature profiles.

Figures 5 to 7 shows the tangential velocity, transverse velocity and temperature profiles, respectively, for $M = 0.5$, $m = 2$, $R = 1$ and $Pr = 0.01$ for different values of ξ. The tangential

velocity profile, in Figure 5, decreases as the transpiration parameter ξ is increased. This shows that the local maximum values of the velocity profile occurs at the area of the boundary layer. The same observation is shown in Figure 6 on the transverse velocity. Figure 7 shows that the temperature profiles decreases as the transpiration parameter is increased. The momentum and thermal boundary layer thickness decreases with the increasing values of ξ due to suction effects of the surface mass transfer.

Magnetic field parameters effects are presented in Figures 8 and 9. Tangential velocity profiles decrease with increase in magnetic parameter but the transverse velocity increases with an increase in the magnetic field parameter.

The effect of thermal radiation parameter is presented in Figures 10 to 12. The thermal radiation parameter increases both the tangential and transverse velocity profiles. It also increases the temperature profiles of the fluid. This is due to the decrease in values of R leading to a decrease in Rosselenda radiation absorptivity k_1. Also, an increase in temperature has a direct effect on the buoyancy force which in turn iniduces more flow causing the tangential and transverse velocities to increase.

Conclusion

This paper has presented a recently developed multidomain quasilinearisation method for solving general non-linear differential equations. The multidomain quasilinearisation method is developed based on bivariate spectral quasilinearisation method (BSQLM). The main goal of the current study is to apply this method in a natural convection flow from a vertical plate with uniform surface temperature. The method proves to be efficient especially for large transpiration parameter. Velocity, temperature and temperature profiles are also analysed here. From these investigations we can conclude:

- MD-SQLM overcomes the similarity transformation barrier of not capturing solutions at large transpiration parameter values.
- Increase in the transpiration parameter decreases the momentum and thermal boundary layer
- Thermal radiation parameter increases the tangential velocity, transverse velocity and temperature profiles.

References

[1] H. Sato, The Hall Effect in the Viscous Flow of Ionized Gas between Two Parallel Plates under Transverse Magnetic Field. Journal of the Physical Society of Japan, 16,(1961) 1427-1433.

[2] T. Yamanishi, Hall Effect in the Viscous Flow of Ionized Gas through Straight Channels. 17th Annual Meeting, Physical Society of Japan, 5,(1962) 29.

[3] A. Sherman, and G.W. Sutton, Magnetohydrodynamics.Evanston, 173-175.

[4] K.R. Sing, T.G. Cowling, Thermal convection in magnetohydrodynamic boundary layers, J. Mech. Appl. Math. 16(1963) 1–5.

[5] E.M. Sparrow, R.D. Cess, The effect of magnetic field on free convection heat transfer, Int. J. Heat Mass Transfer 3 (1961) 267–274.

[6] A.S. Gupta, Flow of an electrically conducting fluid past a porous flat plate in the presence of a transverse magnetic field, J. Appl. Math. Phys. (ZAMP) 11 (1960) 43–50.

[7] Katagiri, M. The Effect of Hall Currents on the Viscous Flow Magnetohydrodynamic Boundary Layer Flow Past a Semi-Infinite Flat Plate. Journal of the Physical Society of Japan, 27, 1051–1059, 1969.

[8] I. Pop, and T. Watanabe, Hall Effect on Magnetohydrodynamic Free Convection about a Semiinfinite Vertical Flat Plate. International Journal of Engineering Science, 32,(1994) 1903-1911.

[9] M. Wahiduzzaman, R. Biswas, M.D. Eaqub Ali, M.D. S. Khan, Numerical solution of MHD convection and ass transfer flow of viscous incompressible fluid about an inclined plate with Hall current and constant heat flux, Journal of Applied Mathematics and Physics, 3 1688–1709, 2015.

[10] W. G. England and A. F. Emery, Thermal radiation effects on laminar free convection boundary layer of an absorbing gas, Journal of Heat Transfer, vol. 31, pp. 37-44, 1969..

[11] R. S. R. Gorla and I. Pop, Conjugate heat transfer with radiation from a vertical circular pin in a non-newtonian ambient medium, Warme- und Stoffubertragung, vol. 28, no. 1-2, pp. 11-15, 1993.

[12] A. Raptis, Radiation and free convection flow through a porous medium, International Communications in Heat and Mass Transfer, vol. 25, no. 2, pp. 289-295, 1998.

[13] M. Abd El-Aziz, Thermal radiation effects on magnetohydrodynamic mixed convection flow of a micropolar fluid past a continuously moving semi–infinite plate for high temperature differences, Acta Mechanica, vol. 187, no. 1-4, pp. 113-127, 2006.

[14] M. Abd El-Aziz, Thermal–diffusion and diffusion–thermo effects on combined heat and mass transfer by hydromagnetic three-dimensional free convection over a permeable stretching surface with radiation, Physics Letters A, vol. 372, no. 3, pp. 263-272, 2008.

[15] M. Abd El-Aziz, Radiation effect on the flow and heat transfer over an unsteady stretching sheet, International Communications in Heat and Mass Transfer, vol. 36, no. 5, pp. 521-524, 2009.

[16] L. K. Saha, S. Saddiqa, M. A. Hossain. "Effect of Hall current on MHD natural convection flow from vertical permeable flat plate with uniform surface heat flux." Applied Mathematics and Mechanics (English edition) 32. Vol 9(2011): 1127–1146.

[17] L. K. Saha, M. A. Hossain, R.S.R. Gorla. "Effect of Hall current on MHD natural convection flow from vertical permeable flat plate with uniform surface temperature." International Journal of Thermal Sciences 46.(2007): 1790–801.

[18] L. N. Trefethen, Spectral Methods in MATLAB, SIAM (2000).

[19] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, Spectral Methods in Fluid Dynamics, Springer-Verlag, Berlin (1988).

[20] M.A. Hossain, and S.C. Paul, Free convection from a vertical permeable circular cone with non-uniform surface temperature. Acta Mechanica, 151(1–2) 103–114 (2001) doi:10.1007/BF01272528.

[21] S. S. Motsa, V. M. Magagula, and P. Sibanda, "A Bivariate Chebyshev Spectral Collocation Quasilinearization Method for Nonlinear Evolution Parabolic Equations," The Scientific World Journal, vol. 2014, Article ID 581987, 13 pages, 2014. doi:10.1155/2014/581987

[22] K.A. Yih, MHD forced convection flow adjacent to a non-isothermal wedge, Int. Comm. Heat Mass Transfer, (26) (1999) 819-827

[23] S. Hussain, M.A. Hossain, Natural convection flow from a vertical permeable flat plate with variable surface temperature and species temperature, Engineering Computations, Vol. 17 No. 7, 2000, pp. 789-812.

[24] S. S. Motsa, V. M. Magagula, and P. Sibanda. "A Bivariate Chebyshev Spectral Collocation Quasi-linearization Method for Nonlinear Evolution Parabolic Equations." Hindawi Publishing Corporation (2014).

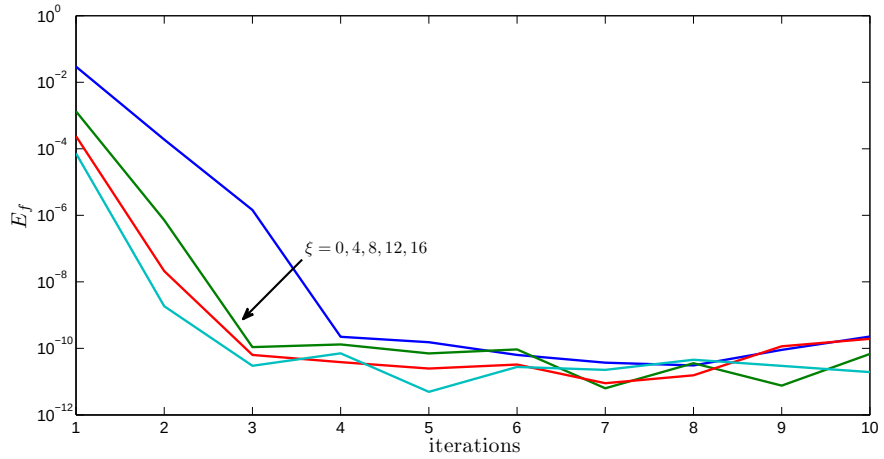


Figure 2: Convergence error in the tangential velocity profile at different values of ξ

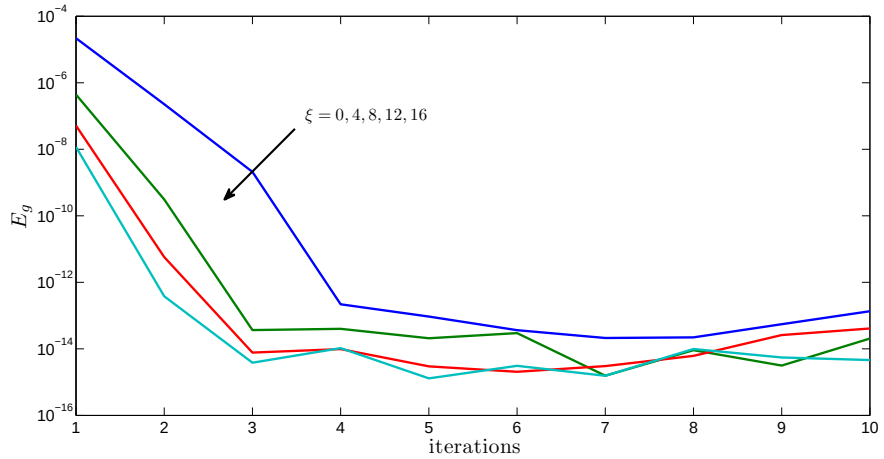


Figure 3: Convergence error in the transverse velocity profile at different values of ξ

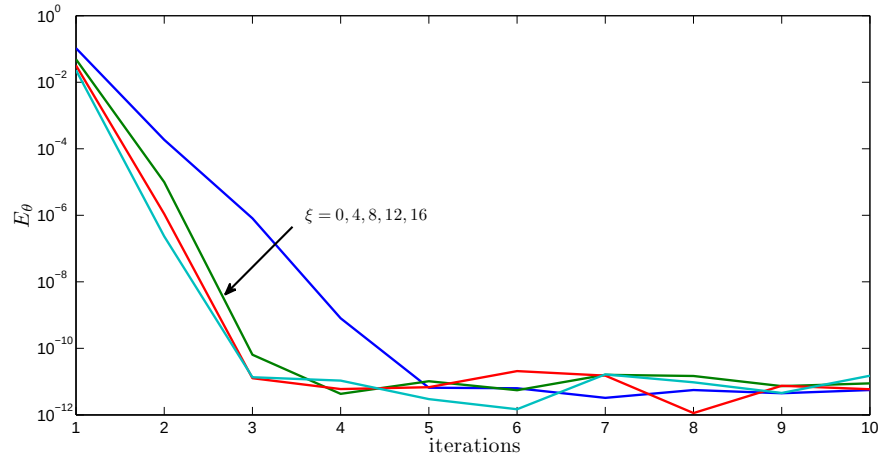


Figure 4: Convergence error in the temperature profile at different values of ξ

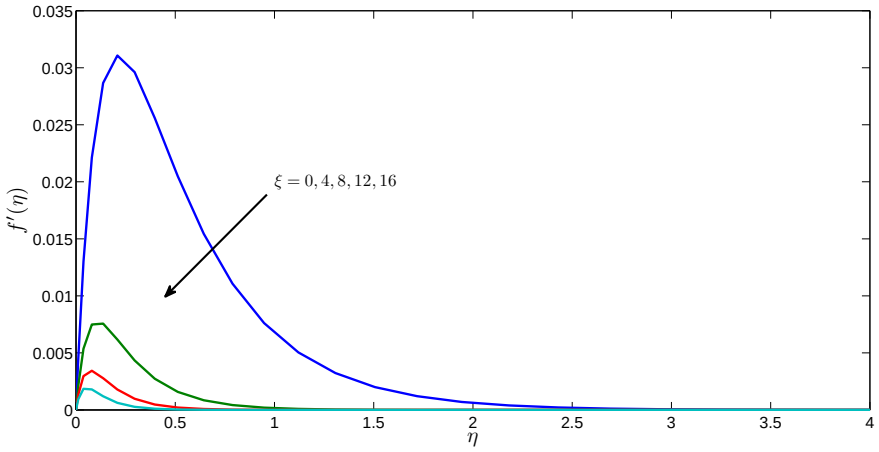


Figure 5: Tangential velocity profile at different values of ξ

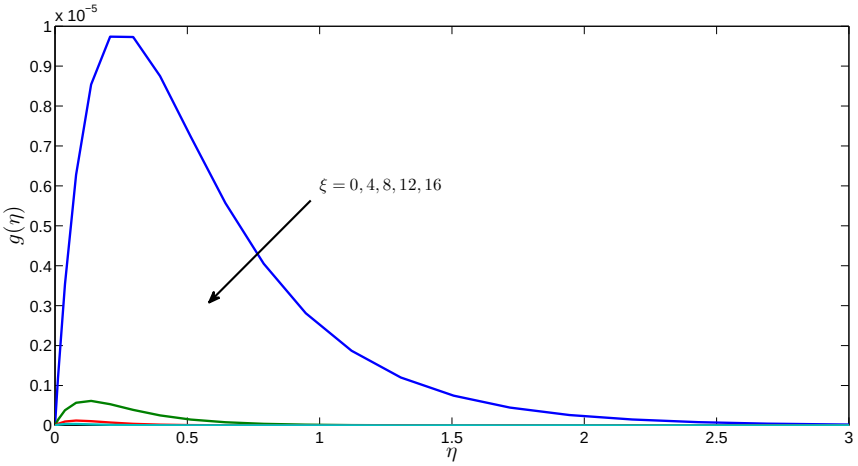


Figure 6: Transverse velocity profile at different values of ξ

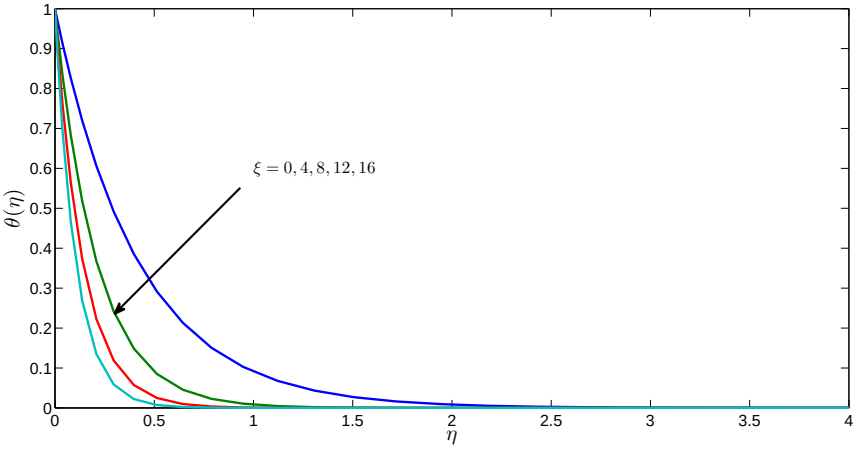


Figure 7: Temperature profile at different values of ξ

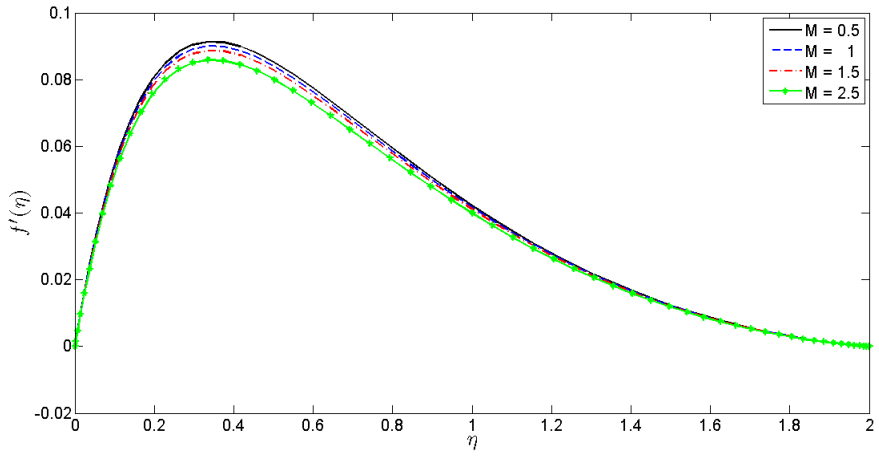


Figure 8: Tangential velocity profile for $R = 3$, $m = 2$, $Pr = 0.7$ at $M = 0.5, 1, 1.5, 2.5$

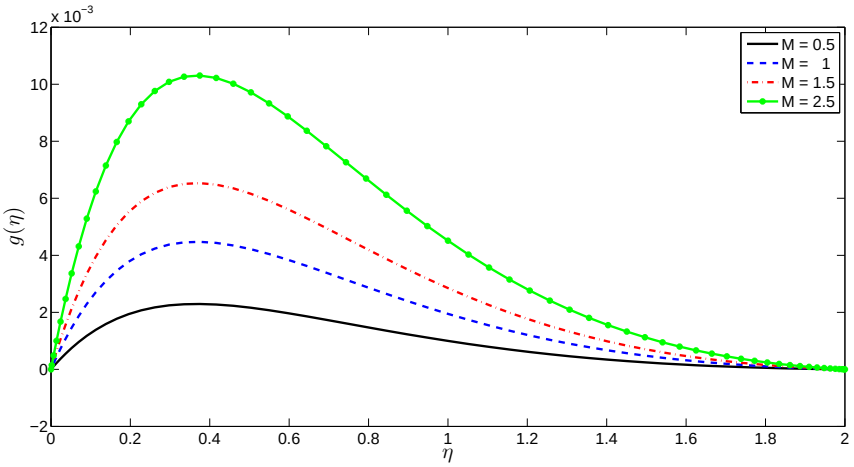


Figure 9: Transverse velocity profile for $R = 3$, $m = 2$, $Pr = 0.7$ at $M = 0.5, 1, 1.5, 2.5$

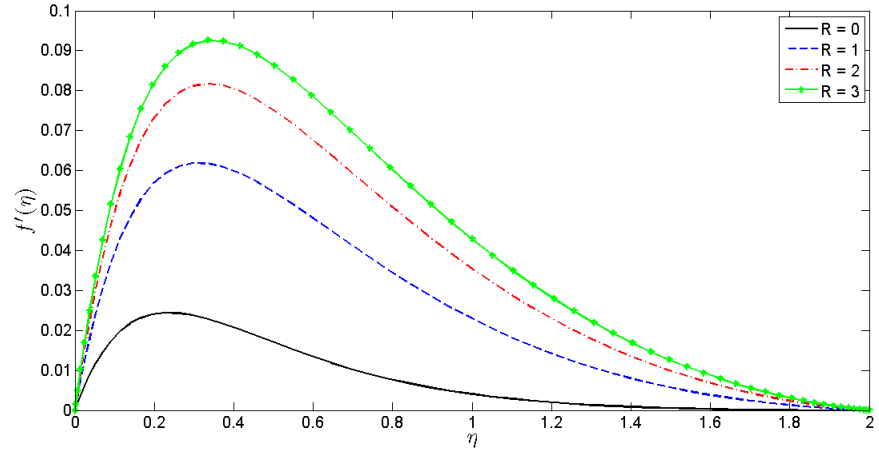


Figure 10: Tangential velocity profile for $M = 1/2$, $m = 100$, $Pr = 0.7$ at $R = 0, 1, 2, 3$

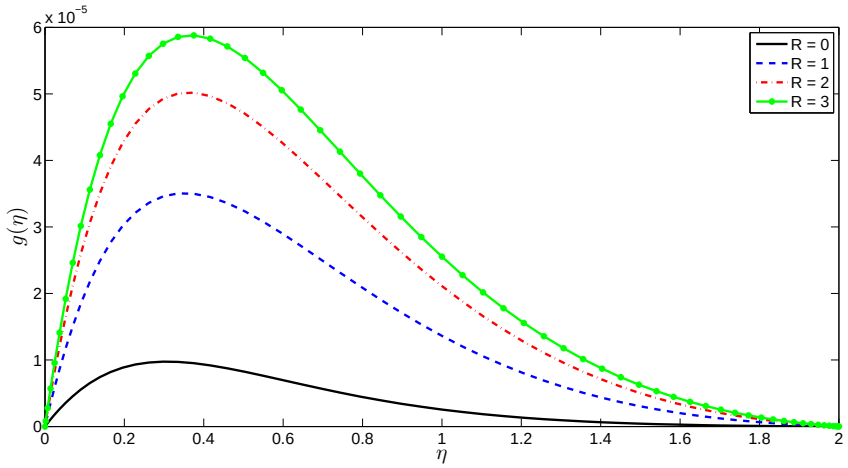


Figure 11: Transverse velocity profile for $M = 1/2$, $m = 100$, $Pr = 0.7$ at $R = 0, 1, 2, 3$

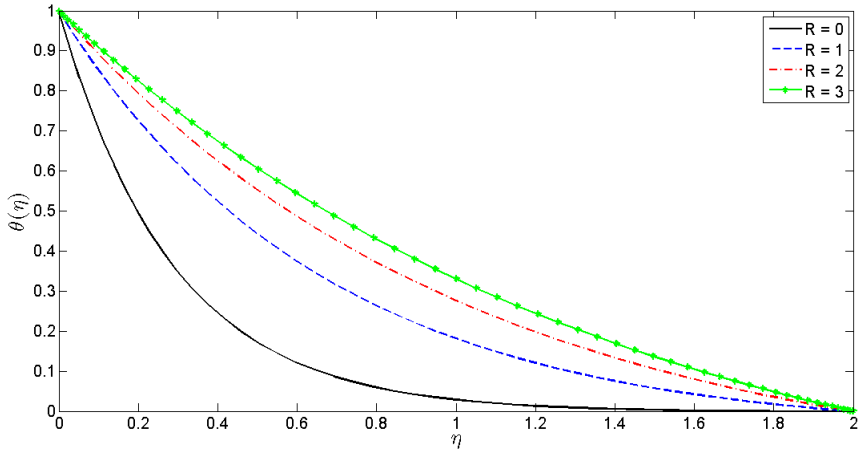


Figure 12: Concentration profile for $R = 3$, $m = 2$, $Pr = 0.7$ at $M = 0.5, 1, 1.5, 2.5$

On a numerical DEM-based approach for assessing thermoelastic properties of composite materials

W. Leclerc[1,a)], H. Haddad[1], C. Machado[1]and M. Guessasma[1]

[1]Laboratoire des Technologies Innovantes (LTI), EA3899, Université de Picardie Jules Verne, France

[a)]Corresponding and presenting author: willy.leclerc@u-picardie.fr

ABSTRACT

The present contribution is dedicated to a Discrete Element Method (DEM)-based approach aiming at assessing the thermomechanical behavior of composite materials. Such an approach presents several advantages in comparison to other classical methods as the Finite Element (FE) one. This enables a better description of the multi-scale behavior of the material with the inherent variability related to the microscopic scale. It also gives the possibility to directly access information such strain and stress fields and heat flux density at the scale of the discrete element. In the current work, a focus is done on the thermoelastic properties of a heterogeneous medium composed of a single inclusion. A 2D representative pattern is generated and discretized using a granular packing composed of cylindrical particles in contact point. This is generated using a process based on the Lubachevsky-Stillinger Algorithm (LSA) coupled to a DEM approach based on a smooth formulation. A hybrid-particulate model is considered to model the mechanical behavior of the material. In this approach, the contact between two particles is described by a beam element which models the cohesive link at the microscopic scale. Heat transfer is simulated using an iterative time-dependent scheme based on the Fourier's law and Voronoï's mosaics generated from granular packings. A full range of thermoelastic properties are considered in order to investigate several configurations of material from an insulative fibre less resilient than the surrounding matrix to a conductive fibre more resilient than the matrix. Estimated properties are compared to those obtained from other numerical methods such as FE and Fast Fourier Transform (FFT)-based calculations and analytical models. Results highlight the ability of the proposed approach to estimate effective thermoelastic properties. These first results pave the way of interesting insights since taking into account non-linear behaviors, interfacial effects and damaging in the proposed approach can be envisaged in a next future.

Keywords: Discrete element method, Multi-scale approach, Composite material, Thermoelastic properties, Equivalent continuous domain.

Introduction

Composite materials arouse the interest of many industrial sectors such as aeronautic, aerospace, automotive, building and marine. These are indeed characterized by excellent stiffness-to-weight and thermal conductivity-to-weight ratios which make them adaptable to different situations and make them able to serve specific purposes and exhibit desirable thermomechanical properties. Besides, the development of biocomposites composed of natural fibres as flax or hemp show their ability to respond to current environnement issues as the reduction of gas emissions. Research to increase performance and safety of composites pieces in many fields requires the development of means of investigation concerning the behavior in service and durability of materials. Durability characterizes the ability of the material to resist to degradation of the thermomechanical properties over time under various types of sollicitations. The scientific challenge therefore consists in developing reliable numerical methods for achieving a better extrapolation of the multi-scale thermomechanical behavior of the composite as well as a better description of various phenomena arising in the material such as crack initiations, debonding effects, local variability and heterogeneity.

Considered as an alternative to the classical FE method, the DEM is an ideal tool for solving mechanical problems in which multiple scales and discontinuities arise. Indeed, DEM is characterized by a good description of microscopic phenomena, an easy treatment of complex structures and a very fine time scale which enables to describe the local behavior of a large number of particles. Among the early studies, DEM was used to explore and gain new insights into various physical applications from geomechanics applications [1, 2] to tribological simulation approaches [3, 4] and heat transfert simulation in multi-contact systems [5, 6]. More recently, André et al. [7] and Haddad et al. [8] considered a hybrid

particulate-lattice model in which particles are linked using cohesive beam elements. Thus, the DEM was made able to quantitatively model the mechanical behavior of homogeneous and heterogeneous materials as well as fracture phenomena as crack formation and propagation.

The present work is dedicated to an extension of the hybrid particulate-lattice model to the characterization of thermoelastic behavior of composite materials. The main objective is to highlight the ability of a DEM-based approach to the assessment of thermoelastic properties such as the thermal conductivity and the Young's modulus. For this purpose, a focus is done on a heterogeneous medium composed of a single inclusion. A 2D square-shaped representative pattern is modeled and discretized by a granular packing composed of cylindrical particles in contact point generated with the help of an efficient process based on the LSA [9] coupled to a DEM approach using a smooth formulation. In order to take into account in the same time the elastic behavior and the heat transfer within the material, the initial set of contacts is densified by a Delaunay triangulation process performed from this initial cloud of particle's centers. It leads to a better description of the heterogeneous medium and more accurate results by the hybrid-particulate model. Besides, a Voronoï mosaic is associated to the Delaunay triangulation which provides in the same time a representative volume and transmission contact surfaces to each particle. Thus, the heat transfer by conduction can be simulated using an iterative time-dependent scheme based on the Fourier's law where representative volumes and surfaces come from the Voronoï mosaic.

This paper is organized as follows. First, we describe the heat transfer scheme and the hybrid-particulate approach for simulating the thermoelastic behavior of the material. Second, the numerical model is validated in the context of a homogeneous material. Thermal and boundary conditions are imposed to the 2D square pattern in order to reproduce simple tests as tensile and shear ones leading to thermoelastic properties. Finally, the DEM-based approach is applied to the case of a single circular inclusion embedded in a matrix. For validation purposes, a large range of material configurations are investigated from an insulative fibre less resilient than the surrounding matrix to a conductive fibre more resilient than the matrix. Comparisons are carried out with several numerical methods, namely FE and FFT-based calculations and analytical models.

Numerical model

Equivalent Continuous Domain

The first step of the proposed DEM-based approach consists in discretizing the continuous domain at the macroscopic scale by a granular packing composed of cylindrical particles in 2D. The generation of the granular packing is done by the efficient LSA coupled to the DEM using a smooth formulation. The idea is that the early stages of the LSA are dominated by the densification of the system and consequently more efficiently performed than the last steps where the number of contacts dramatically increases. In the coupled approach, the last steps are performed by the DEM using a smooth formulation which is more suited to control the multiplicity of contacts than the LSA. Under several assumptions of polydispersity, orientation and size, the granular domain can be considered as an Equivalent Continuous Domain (ECD) in that this is enough representative of the continuous medium. First, the compacity of the granular domain has to be closed to 0.85 which corresponds to the Random Close Packing (RCP) for a random granular packing composed of cylindrical particles in 2D. Second, the coordination number which represents the average number of particles in contact with one given particle has to be close to 4.5. Third, a slight polydispersity of particle size must be introduced in order to avoid undesirable directional effects. Typically, the particle's radius follows a Gaussian distribution law and the dispersion is characterized by the coefficient of variation which is the ratio between the standard deviation and the average radius. For information purposes, this is set to 0.3 in the present work. These three first parameters ensure the randomness of the granular packing and consequently the isotropy of the ECD. In other words, this ensures that thermoelastic properties are independent of the direction. At last, the number of particles represents the fineness of the discretized medium in a similar way to a FE Mesh. As done by previous authors, the network of contacts is finally densified using a Delaunay triangulation process applied from this initial cloud of particle's centers. Thus, the coordination number comes from about 4.5 to about 5.9 and about 10% of new contacts are generated. A Voronoï tessellation is finally associated to the Delaunay triangulation. This provides in the same time an area of representation for each particle and its contacts. Such a process turns out to be not costly in computational time as long as dynamic effects are not considered since the remeshing process is then not required.

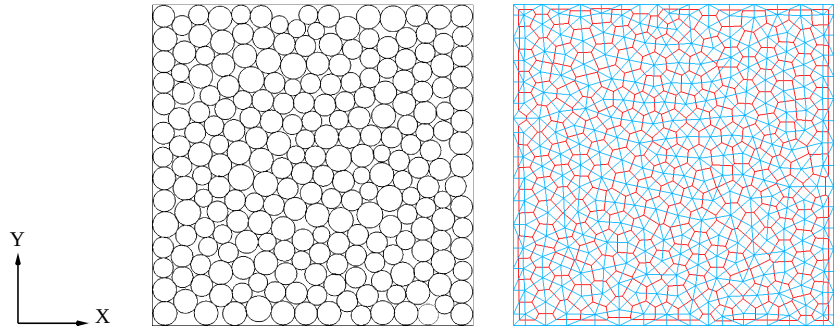


Figure 1. Example of a typical 2D Voronoï construction based on a granular packing constituted of 200 particles: granular packing (a) and corresponding Voronoï tessellation (b)

Heat transfer by conduction

In the present model, each particle i is related to a Voronoï cell considered as its representative element (Fig. 2). This polygon has a number of sides equal to the number of particles j in contact with the particle i. The heat flux transmitted by the contact surface between two particles i, j is defined as follows:

$$W_{ij} = H_c^{i,j}(T_j - T_i) \tag{1}$$

where T_i, T_j are the temperatures of particles i, j and $H_c^{i,j}$ is the coefficient of thermal conductance: $H_c^{i,j} = \frac{S_{ij}^t k}{d_{ij}}$, with λ the conductivity of material, d_{ij} the distance between the centers of particles i, j and S_{ij}^t the area of heat transmission surface related to the corresponding polygon side.

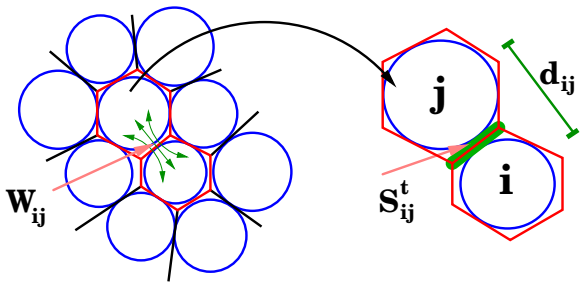


Figure 2. Definition of the heat transmission surface S_{ij}^t

The corresponding equation of heat transfer is expressed for each particle i by:

$$C_i^d \frac{dT_i}{dt} + \sum_{j=1}^{n_i} W_{ij} = Q_i \tag{2}$$

where Q_i represents the external heat flux associated to the particle i and n_i is the number of neighbors of particle i. C_i^d is the heat capacity of the particle given by:

$$C_i^d = c_p \rho_d V_i \tag{3}$$

with V_i and ρ_d are the volume and the density of the particle respectively and c_p is the specific heat of constitutive material. For the purpose of conservation mass, the discrete element mass is adjusted to the polygon one. To satisfy this assumption, we consider ρ_c as the constitutive material density, ρ_d is then connected to ρ_c through the following relationship:

$$\rho_d = \frac{V_{poly}}{V_i} \rho_c \tag{4}$$

where V_{poly} is the polygon's volume. The discretization of equation for heat transfer (2) in time leads to:

$$T_i^{t+\Delta t} = T_i^t + \frac{\Delta t}{c_p \rho_d V_i} \underbrace{[Q_i + \sum_{j=1}^{n_i} \frac{S_{ij}^t \lambda}{d_{ij}} (T_j^t - T_i^t)]}_{Q_i^{tot}} \tag{5}$$

Elastic behavior

We consider a hybrid particulate-lattice model in which the interaction between two cylindrical particles in contact is modeled by a beam of length L_μ, Young's modulus E_μ, cross-section A_μ and quadratic moment I_μ (Fig.3). Therefore, the cohesive contacts are maintained by a vector of three-component generalized forces acting as internal forces. The normal component acts as an attractive force, the tangential component allows to resist to the tangential relative displacement and the moment component counteracts the bending motion [7].

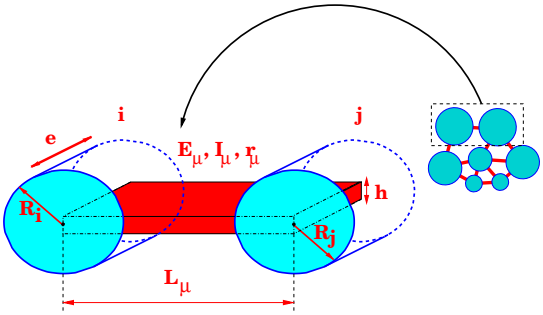


Figure 3. Hybrid particulate-lattice model

The cross-section A_μ is rectangular with sides e and h, where e is the thickness of the granular medium and h is the height of the cross section defined by:

$$h = r_\mu \frac{R_i + R_j}{2} \tag{6}$$

where $r_\mu \in]0, 1]$ is a dimensionless radius. R_i and R_j are respectively the radius of particles i and j in contact. The internal cohesive forces between two particles i and j are given by the following system:

$$\begin{bmatrix} F_n^{j\to i} \\ F_t^{j\to i} \\ M^{j\to i} \end{bmatrix} = \begin{bmatrix} \frac{E_\mu A_\mu}{L_\mu} & 0 & 0 & 0 \\ 0 & \frac{12E_\mu I_\mu}{L_\mu^3} & \frac{6E_\mu I_\mu}{L_\mu^2} & \frac{6E_\mu I_\mu}{L_\mu^2} \\ 0 & \frac{6E_\mu I_\mu}{L_\mu^2} & \frac{4E_\mu I_\mu}{L_\mu} & \frac{2E_\mu I_\mu}{L_\mu} \end{bmatrix} \begin{bmatrix} u_n^i - u_n^j \\ u_t^i - u_t^j \\ \theta_i \\ \theta_j \end{bmatrix} \tag{7}$$

where θ_i and θ_j are respectively the rotations of particles i and j. $u_n^{i,j}$ and $u_t^{i,j}$ are respectively the normal and tangential displacements. The linear system of equations shows the micro-macro relations applied to determine the contact forces between two particles i and j. These relations stem from the classical stiffness matrix of the beam element model. The translational and rotational equations of motion for a particle i are written as follows:

$$m_i \ddot{u}_i = F_i^{ext} + \sum_j F^{j\to i} \tag{8}$$

$$I_i \ddot{\theta}_i = M_i^{ext} + \sum_j M^{j\to i} \tag{9}$$

where m_i is the elementary mass of the particle i and I_i is the quadratic moment of inertia of the particle i. $F^{j\to i}$ et $M^{j\to i}$ are respectively the force and the moment of interaction of the particle j on the particle i. F_i^{ext} et M_i^{ext} are respectively the external force and moment acting on particle i. The numerical resolution is based on an explicit time integration with a formulation based on a Verlet scheme.

Thermoelastic properties

The present section is dedicated to the description of methodologies leading to the assessment of thermoelastic properties, namely the Effective Thermal Conductivity (ETC), the Effective Young's Modulus (EYM) and the Effective Shear Modulus (ESM). For validation purposes, a homogeneous medium with known properties is considered and effective thermoelastic properties are evaluated and finally compared to the expected values. From now on, the continuous domain is a square and flat plate of side L=3.5 cm and the corresponding ECD is a granular packing composed of about 5000 polydisperse cylindrical particles.

ETC

The ETC is estimated by the following approach. A temperature difference (ΔT) is imposed between two opposite edges of the square domain (in the present case $y = 0$ and y =L). The heat transfer within the homogeneous medium is described by the time-dependent methodology described in subsection a). The heat flux density (ϕ) is then numerically estimated at stationary state and the ETC λ deduced from the following Equation :

$$\lambda = \frac{\phi \mathrm{L}}{\Delta T} \tag{10}$$

In the present test, the plate is subjected to thermal and initial conditions defined as follows :

$$\begin{cases} T_1 : T(y = 0) = 25°C \\ T_2 : T(y = \mathrm{L}) = 100°C \\ t = 0 : T(y) = T_0 = 25°C \quad 0 < y < \mathrm{L} \end{cases} \tag{11}$$

Lateral boundaries are under adiabatic conditions and material parameters are listed in Tab. 1 :

Table 1. Thermal properties of the continuous domain

Density	ρ_c	2600	kg/m^3
Thermal conductivity	λ	30	W/mK
Specific heat	c_p	900	J/kgK

The variation of temperatures obtained by an analytic solution [10] and the DEM-based approach at times 3 s, 30 s and 150 s are graphically shown in Fig. 4a. Both models present identical temperature profiles which exhibits the ability of the DEM-based approach to model heat transfer in a continuous domain.

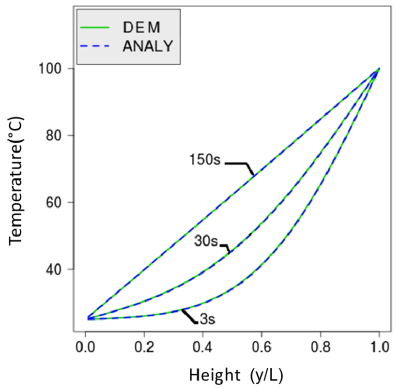


Figure 4. Comparison between analytic and discrete model solutions at several times (a) and field of heat flux density (b)

The heat flux density is estimated at stationary state at the scale of the particle using the following Equation which is analogous to the Love-Weber formulation.

$$\phi_i = \frac{1}{V_i} \sum_j \Phi^{ext,j} x_{ij} \tag{12}$$

where ϕ_i is the heat flux density related to the particle i, V_i is the volume of the particle i, x_{ij} is the length of the contact between particles i and j, and $\Phi^{ext,j}$ is the external flux applied to the particle i by the particle j. The heat flux density ϕ is estimated after averaging heat flux densities over the volume of the plate. In the present example, a value of 64303 W/m^2 is obtained which leads to an ETC λ=30.008 W/(m.K) which is very close to the expected value of 30 W/(m.K). This highlights the ability of the present DEM-based approach to estimate ETC of homogeneous materials.

EYM and ESM

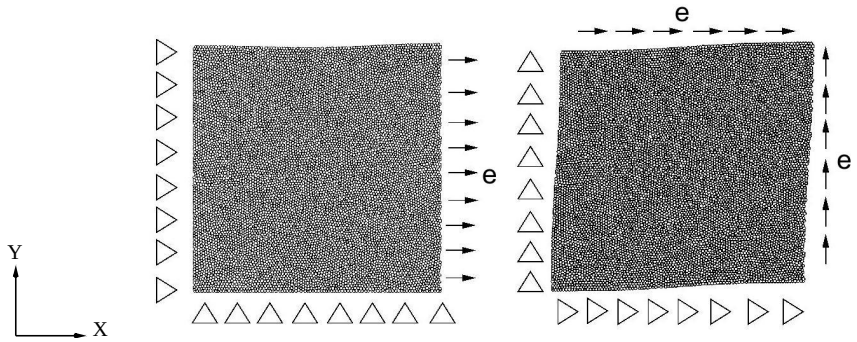


Figure 5. Quasi-static tensile (a) and shear (b) tests

EYM and ESM are estimated via quasi-static tensile and shear tests performed under a plane stress state using the boundary conditions described in Fig. 5. Symmetry boundary conditions are considered and a displacement e is imposed on the right edge of the square in the case of the tensile test on the one hand, on the other hand anti-symmetry boundary conditions are considered and a displacement e is imposed on top and right edges of the plate in the case of the shear test. The main issue of such an approach is that on the contrary of FE calculations for which local properties at the scale of the element are identical to the macroscopic properties in the case of a homogeneous material, microscopic properties of the beam element (E_μ, r_μ) can only be correlated to EYM and ESM as previously done in previous works [7, 8].

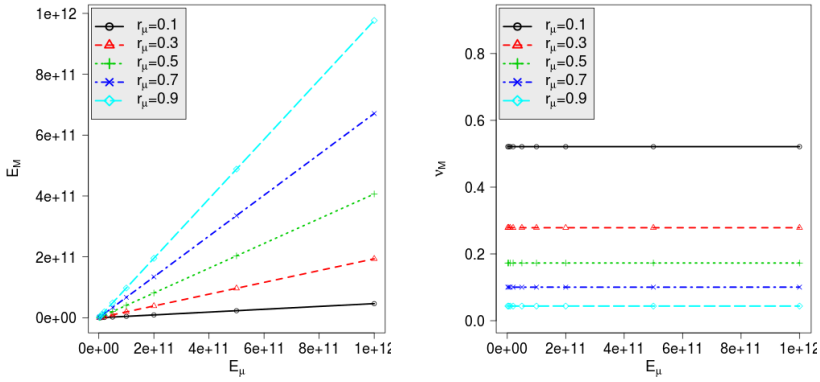


Figure 6. Influence of the microscopic parameters E_μ and r_μ on the EYM (a) and the Poisson's ratio (b)

The calibration process consists in determining the relation between microscopic and macroscopic parameters via a full range of investigated configurations so that the evolution of microscopic properties allows us to choose the desired macroscopic ones. In the present work, we consider a microscopic Young's modulus in the interval [2GPa, 1000GPa], and a r_μ parameter in the interval $[0.1, 0.9]$. Evolution curves are plotted in Figures 6-a and -b. We notice that the macroscopic Poisson's ratio ν_M does not depend on E_μ but quadractically depends on the dimensionless radius r_μ. EYM E_M linearly

depends on r_μ and quadratically depends on E_μ. These conclusions are in good agreement with those obtained by André et al. [7] in the context of spheres in 3D.

Case of a heterogeneous continuous medium with a single inclusion

The section is dedicated to the investigation of the thermoelastic behavior of a heterogeneous continuous medium with a single inclusion. For this purpose a 2D square-shaped representative pattern of the composite material is generated and numerical approaches described in the previous section are considered. The representative pattern consists of a centred circular inclusion which represents the unidirectional fibre and has a radius equal to 0.25 times the length L (Fig. 7). The square pattern is discretized by the same random granular packing constituted of 5000 polydisperse particles than the previous one used for a homogeneous material.

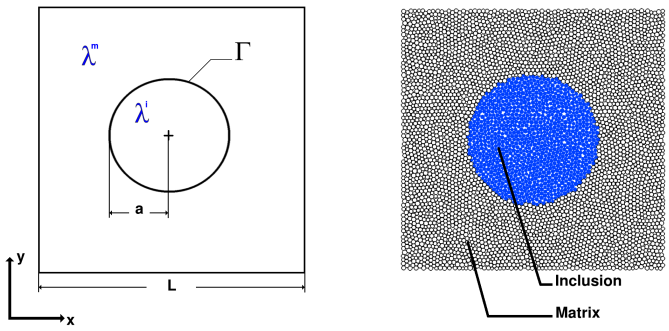


Figure 7. Single inclusion problem: continuous (a) and discrete (b) models

Thermal properties

Our objective is to assess the ETC λ^e of the heterogeneous continuous medium with a single inclusion via the proposed DEM-based approach. Both inclusion and matrix phases are supposed isotropic with thermal conductivities respectively denoted by λ^i and λ^m where superscripts i and m designate the inclusion and matrix phase respectively. λ^m is set to 30 W/(m.K) and λ^i is varied according to the expected contrast of properties $c_\lambda = \frac{\lambda^i}{\lambda^m}$ which can be chosen greater or lower than 1. In other words, the inclusion can be considered more conductive or more insulative than the matrix phase. The specific heat capacity is supposed set to 900 J/(K.kg) for both phases but this is of little importance since we are only interested by results at stationary state in the present section. The evaluation of the ETC is performed considering the methodology described in subsection a). Results are compared to two numerical homogenization techniques. The first technique is the FFT-based method which consists in solving the Lippmann-Schwinger's equation in Fourier space using an iterative algorithm [11, 12]. In the present work, calculations are performed using the Eyre-Milton scheme and a digitized map of the representative pattern consisted of 1048576 (1024^2) pixels [13]. The second one is the double-scale homogenization method (2SFEM) [14]. This approach is based on variational considerations and uses the FEM with periodic boundary conditions. Results are also compared with the classical FEM for which thermal conditions are the same as those considered in the DEM-based approach, and a theoretical estimate, namely the Hashin's model (HM) [15]. For information purposes, all FEM calculations are carried out using a structured mesh composed of 980000 (2×700^2) 3-node triangular elements.

Table 2. Influence of the contrast on the normalized ETC for several numerical and theoretical approaches

c_λ		0.01	0.02	0.05	0.1	0.2	0.5	1	2	5	10	20	50	100
λ^*	DEM	0.672	0.678	0.694	0.722	0.768	0.877	1.000	1.139	1.300	1.382	1.432	1.466	1.478
	FEM	0.677	0.682	0.698	0.723	0.769	0.878	1.000	1.140	1.302	1.384	1.433	1.467	1.478
	2SFEM	0.676	0.682	0.698	0.723	0.768	0.877	1.000	1.140	1.301	1.383	1.433	1.466	1.477
	FFT	0.677	0.682	0.698	0.723	0.768	0.877	1.000	1.140	1.302	1.384	1.433	1.467	1.472
	HM	0.677	0.683	0.698	0.723	0.769	0.877	1.000	1.140	1.301	1.383	1.432	1.465	1.477

Calculations are carried out for a range of c_λ from 0.01 to 100. Thus, two main configurations are considered, namely the case of an inclusion more insulative than the matrix ($c_\lambda < 1$) and the reverse case for which the inclusion is more conductive than the matrix ($c_\lambda > 1$). Table 2 illustrates the influence of the contrast on the assessed normalized ETC (λ^*) which is obtained by dividing the ETC λ^e by the thermal conductivity of the matrix. Results are compared with those obtained using other numerical and theoretical approaches. Whatever the contrast, less or more than 1, predictions given by the DEM are very close to other assessments with a maximum relative difference of 0.6%. This highlights the ability of the DEM to estimate the ETC of a heterogeneous continuous medium with a single inclusion.

Elastic properties

Tensile and shear tests are carried out using the boundary conditions already seen in Fig. 5 in order to assess EYM and ESM. The macroscopic Young's modulus E^m of the matrix is set to 65 GPa. Different values of macroscopic Young's modulus E^i of the inclusion are considered so that the contrast of properties $c_r = \frac{E^i}{E^m}$ varied from 0.01 to 100. Poisson's ratios of both phases are set to 0.3 and we suppose a plane stress state. DEM-based results are compared to those obtained using the same numerical approaches than previously seen for evaluating the ETC, namely the FFT-based method, the double-scale homogenization method (2SFEM), the classical FEM for which boundary conditions are identical to those considered in the DEM approach. Comparisons are also performed with the theoretical estimate given by Mori and Tanaka (MT) [16]. For information purposes, all FE and FFT-based calculations are carried out considering the same discretizations than previously used for evaluating the ETC.

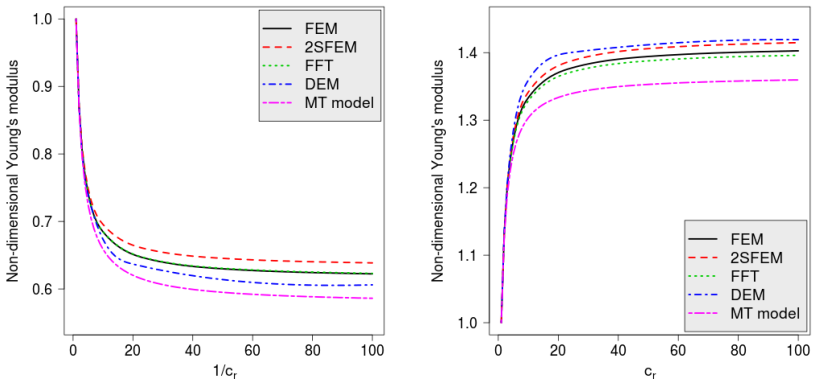


Figure 8. Non-dimensional Young's modulus as a function of the contrast of properties, case $c_r \leq 1$ (a) case $c_r \geq 1$ (b)

Two configurations are investigated. The first problem corresponds to the case of an inclusion less stiff than the matrix with a Young's modulus less than that of the matrix. The second one corresponds to the case of an inclusion stiffer than the matrix with a Young's modulus higher than that of the matrix. Fig. 8 illustrates the influence of the contrast of properties c_r on the non-dimensional Young's modulus which is obtained by dividing E by E^m. Results exhibit a good agreement between DEM, FEM, numerical homogenization techniques and the theoretical estimate whatever c_r. For example, for a contrast of 100, the relative differences with respect to the value given by the FEM is 5.39% for the Young's moduli in the case where $c_r < 1$, and the relative difference is only 0.06% when $c_r > 1$. Globally, relative differences do not exceed 5% whatever the considered contrast of properties. This highlights the ability of the DEM approach to estimate elastic properties of a heterogeneous continuous medium with a single inclusion.

Conclusion

The present paper dealt with a DEM-based approach for characterizing the thermoelastic behavior of composite materials. A focus was done on a 2D plate structure with a single inclusion embedded in a matrix. Comparisons with other numerical and theoretical approaches highlight the suitability of the proposed approach to estimate ETC, EYM and ESM. These results are encouraging and pave the way to interesting prospects. In a next future, we expect to extend the present approach to model the thermomechanical behavior of complex heterogeneous media where fracture phenomena and interfacial effects arise.

References

[1] Cleary, P. W. and Campbell, C. S. (1993) Self-lubrication for long run-out landslides: examination by computer simulation. *Journal of Geophysical Research Solid Earth* **98**, 21911-24.

[2] Campbell, C. S., Cleary, P. W. and Hopkins, M. A. (1995) Large-scale landslide simulations: global deformation, velocities, and basal friction. *Journal of Geophysical Research Solid Earth* **100**, 8267-83.

[3] Sève, B., Iordanoff, I. and Berthier, Y. (2001) A discrete solid third body model: Influence of the intergranular forces on the macroscopic behaviour. Elsevier.

[4] Fillot, N., Iordanoff, I. and Berthier, Y. (2007) Modelling third body flows with a discrete element method -a tool for understanding wear with adhesive particles. *Tribology International* **40**, 973-981.

[5] Vargas, W. L. and McCarthy, J. (2001) Heat conduction in granular materials. *American Institute of Chemical Engineers Journal* **47**, 1052-1059.

[6] Haddad, H., Guessasma, M. and Fortin, J. (2014) Heat transfer by conduction using DEM-FEM coupling method. *Computational Materials Science* **81**, 339-347.

[7] André, D., Iordanoff, I., Charles, J. L. and Néauport, J. (2012) Discrete element method to simulate continuous material by using the cohesive beam model. *Computer Methods in Applied Mechanics and Engineering* **213-216**, 113-125.

[8] Haddad, H., Leclerc, W., Guessasma, M., Pélegris, C., Ferguen, N. and Bellenger E. (2015) Application of DEM to predict the elastic behavior of particulate composite materials. *Granular Matter* **21**, 537-554.

[9] Lubachevsky, B. D. and Stillinger, F. H. (1990) Geometric Properties of Random Disk Packings. *Journal of Statistical Physics* **60**, 561-583.

[10] Weigand, B. (2004) Analytical Methods for Heat Transfer and Fluid Flow Problems. Springer Verlag.

[11] Moulinec, H. and Suquet, P. (1998) A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering* **157**, 69-94.

[12] Michel, J-C., Moulinec, H. and Suquet, P. (1999) Effective properties of composite materials with periodic microstructures: a computational approach. *Methods in Applied Mechanics and Engineering* **172**, 109-143.

[13] Eyre, D. and Milton, G. (1999) A fast numerical scheme for computing the response of composites using grid refinement. *European Physical Journal. Applied Physics.* **6**, 41-47.

[14] Sanchez-Palencia, E. (1980) Non-homogeneous Media and Vibration Theory. Springer-Verlag.

[15] Hashin, Z. (1965) On elastic behaviour of fibre reinforced materials of arbitrary transverse phase geometry. *Journal of the Mechanics and Physics of Solids* **13**, 119-134.

[16] Mori, A. and Tanaka, K. (1973) Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica* **21**, 571-574.

Large-Eddy Simulation of Porous-Like Canopy Forest Flows Using Real Field Measurement Data for Wind Energy Application

†*Zeinab Ahmadi Zeleti[1], Antti Hellsten[1,2], Ashvinkumar Chaudhari[1], Heikki Haario[1]

[1]School of Engineering Science, Lappeenranta University of Technology, Lappeenranta, Finland.
[2]Finnish Meteorological Institute, Helsinki, Finland

*Presenting author: zeinab.ahmadi.zeleti@lut.fi
†Corresponding author: zeinab.ahmadi.zeleti@lut.fi

Abstract

Forests are an integral part of the world's landscape and are often characterized as regions with considerable potential on wind power. In order to take into account the existence of the forest for wind energy assessment, most of previous researches have implemented the drag force or roughness length approaches. However, the goal of this study is to model the forest with porous medium approach and investigate the mean wind and turbulence around porous-like-forest by means of Large Eddy Simulation (LES). For this purpose, *in situ* wind measurements, is obtained at Skinnarila forest, near the campus of Lappeenranta University of Technology, Finland.

Keywords: Wind Energy, LES, Canopy Flows, CFD, Porous Media, Experimental Validation.

Introduction

Predicting turbulence over the wind farms is highly important for wind energy assessment. Many sites with high wind speeds may not be a good candidate for wind energy production due to their high degree of turbulence. Forested terrain is an example to be given here. Nowadays, many onshore wind projects are being planned in or very close to forested landscapes due to its vast availability of wind and less number of inhabitants. Nevertheless, these regions are recognized with complex flow due to large variability of vertical and horizontal foliage distribution which induces the amount of turbulence within, above and around forested trees. To better understand the behavior of wind flow motions around forested area, extensive studies from site and laboratory experiments to numerical simulations have been carried out for many years [1,2,3]. Many researchers have examined how best they can evaluate and impose the effect of forest into numerical predictions of canopy flows. In Fabian et al. 2012 [4], the detailed representation of canopy derived from terrestrial laser scanning was used for LES to observe the aerodynamic influence of small scale plant distribution on clearing inside forest. The turbulent structures developed by a pine forest was numerically studied and validated with field data [5]. Here, the authors have utilized the measured mean vertical distribution of frontal leaf area density (LAD) for LES simulation and detected wakes behind the trunks. Similarly, LES were carried out using the drag force induced by trees under three different atmospheric conditions, namely stable, unstable and neutral [6]. In this work, the canopy model was implemented by considering a homogeneous forest with leaf area Index (LAI) very close to the measured *in situ* value.

Nevertheless for wind park simulations, the effect of canopy is often done by specifying a relationship between frontal LAD, local wind speed and drag coefficient which is added to the right hand side of momentum equations and turbulence models to account for turbulence length within forest [7,8,9]. It is important to note that measuring LAD is very costly in terms

of technology and time. Also, it is highly case dependent and may vary quite much from different forests.

In this work, we will model canopy forest with porous medium approach. In order to validate the porous LES results over forested and non/forested terrain, field measurements have been conducted at Skinnarila forest, near the campus of Lappeenranta University of Technology, Finland.

Methodology

Study Site:

The field measurement obtained at Skinnarila (near Lappeenranta University of Technology, about 7 km north-west of city of Lappeenranta in Finland) recorded wind continuously at 11 different heights from 24th of May till 6th of June, 2013. Part of the forest in which the experimental measurement took place was classified as a non-uniform plantation of pine trees. The average tree height was about 20 m at the forest edge and within.

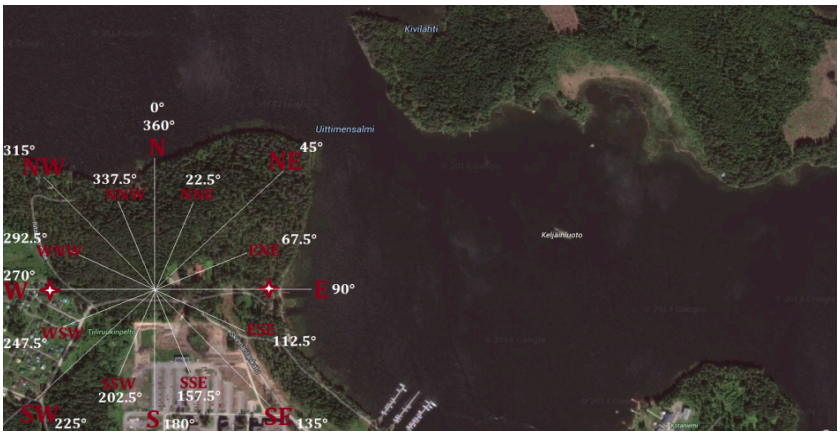


Fig.1: Aerial view of Skinnarila forest with the two lidar devices (little diamonds) positioned at east and west.

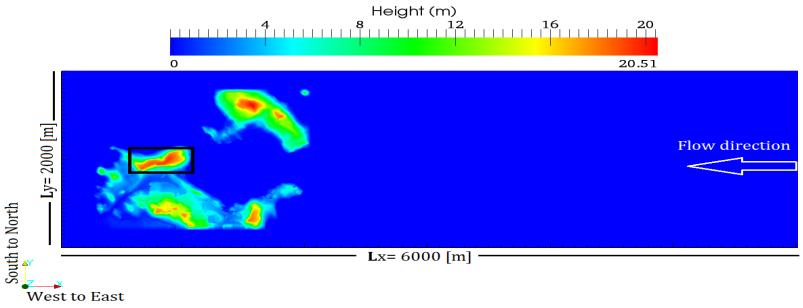


Fig.2: Photo of terrain surface utilized for simulation. The black solid frame indicates the region of interest.

Fig.1 displays the aerial view of the investigated domain together with the two lidar devices positioned at 6770864N, 559016E (before forest) and 6770848 N, 558604 E (after forest)

which are marked as white-red diamond. As seen from the picture, these two devices are aligned according to the predominant wind direction which flows from Lake Saimaa through forest (east to west) with much less disturbances. Moreover, the terrain elevation above sea level is demonstrated in Fig. 2.

Data Selection:

In order to constrain our study under neutral meteorological conditions, the harmony weather forecast of every 3 hours data, including temperature and pressure for heights up to 3 km, provided by Finnish Meteorological Institute (FMI), has been used. Potential temperature, the most important and frequently used quantity in atmospheric science, together with velocity profiles limited to wind directions of around 90° have been plotted during the diurnal cycle for the period of measurement. As a result of boundary evolution presented by harmony potential temperature and velocity profiles, the dates satisfying the neutral atmospheric regime with lowest boundary layer thickness (less than 500 m) are identified and utilized over the real site measurement data. Again, based on availability of data at all 11 different heights, the decision has been made for 2nd day of June 2013 between 21:00 to 23:10 o'clock.

Numerical Descriptions:

In order to represent the forest effect into computational fluid dynamic (CFD) simulation, a porous media model is used. This is by additional sink/source term added to the right hand side of the LES equations in the form $S = -\left(\frac{\mu}{k}U + C\frac{1}{2}\rho|U|U\right)$. This is the general form of porous model composed of two parts: viscous and inertial drag loss terms, respectively. Here the ability of the medium to permit flow is denoted as k and the canopy inertial resistance coefficient as C. In the previous work [10] where porous parameterization study on flow through obstacles representing trees was investigated, we concluded the insignificant effect of permeability and porosity for high Reynolds number. By following this finding, the above sink term reduces to inertial drag loss term.

To solve the flow equations, the entire computational domain ($6\times2\times0.5$ km^3) is discretized into 11625000 of hexagonal grid cells with resolution of about 8 m in all three directions. The finite-volume method based un-structured code OpenFOAM is used in this study. In particular, the simulations are being carried out using our own in-house LES solver called "rk4ProjectionFoam" [11] recently implemented into OpenFOAM. For the numerical computations, the inflow boundary condition is defined according to the selected 2 hours and 10 minutes averaged horizontal velocity and wind directions recorded at 11 different heights of *in situ* measurement before the Skinnarila forest. To fully develop the turbulence structure, the so-called recycling technique [11] is employed at the inlet. The pressure is fixed to zero at outflow boundary and periodic boundary condition is assigned in the lateral sides. The symmetry boundary condition is set at the top surface. The logarithmic wall-function based on roughness-length is used to account the roughness effect.

Before employing LES, a series of Reynolds-averaged Navier-Stokes (RANS) simulations are carried out to parameterize the inertial resistance coefficient of porous-like forest. Afterwards, the most suitable coefficient is implemented into the porous model for the LES calculations. However, it is a good practice to perform LES without forest in order to better observe the turbulence induced by the forest.

Results

Field Experiments:

The 10 minutes averaged wind directions during 2 weeks of field measurement are shown in Fig. 3 for both locations: before (left) and after (right) forest. It can be seen that majority of wind is blowing into the forest approximately from east (close to east-north-east) at all 11 heights. However, the wind has turned its direction at lower heights (especially at 15 m) right after forest edge. This is shown more visibly in Fig. 4. Here the 2 weeks averaged data are plotted with height. Also, it is observed that the forest resistance causes the wind speed to slow down within forest.

Moreover, the 2 hours and 10 minutes averaged wind data over neutral atmospheric condition are plotted (see Fig. 5) which indeed depicts the drop of wind speed and a slight change in wind directions after vegetated area.

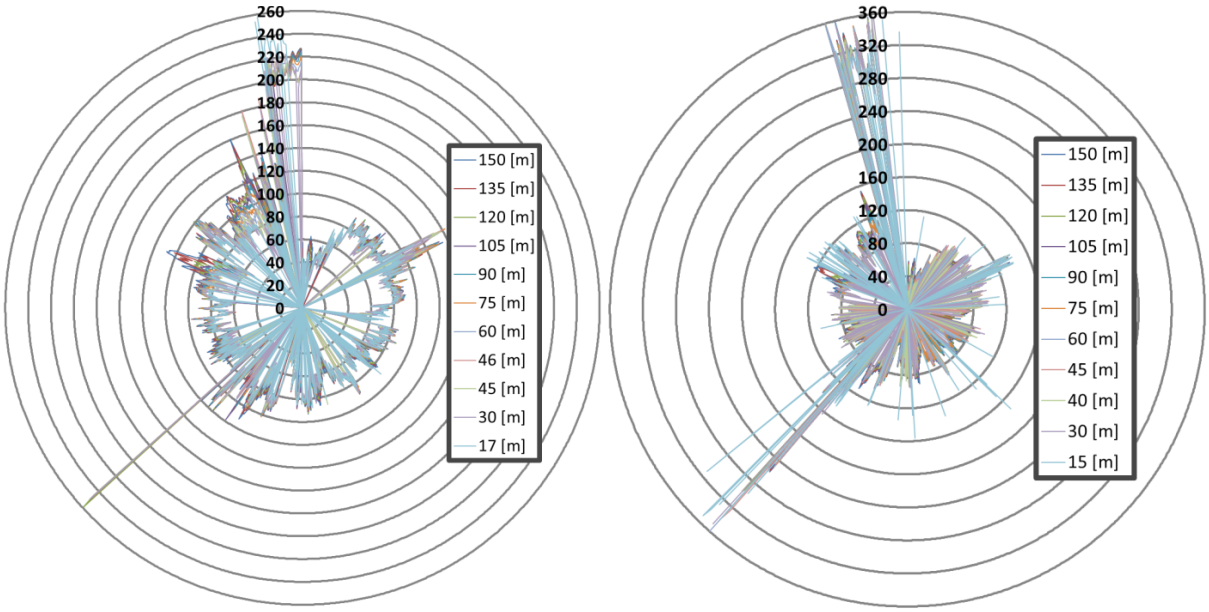


Fig. 3: Comparison of 2-weeks measured wind directions before (left) and after (right) forest at 11 heights.

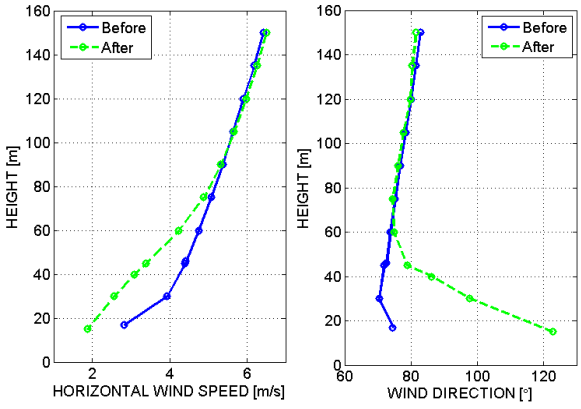


Fig. 4: Comparison of two-weeks-averaged horizontal wind speed (left) and wind direction (right) at two positions: before (blue-line) and after (green-dashed line) forest.

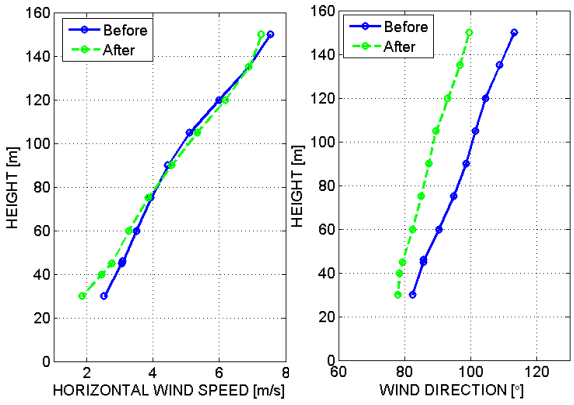


Fig. 5: Comparison of about two hours averaged horizontal wind speed (left) and wind direction (right) at two positions: before (blue-line) and after (green-dashed line) forest.

Numerical Simulations:

Here, we report the preliminary results obtained from the first LES over the site shown in Fig. 2. In the following Figs. 6 and 7 the instantaneous and the 30-min time-averaged horizontal flow fields on the middle planes in stream-wise and span-wise directions are shown, respectively.

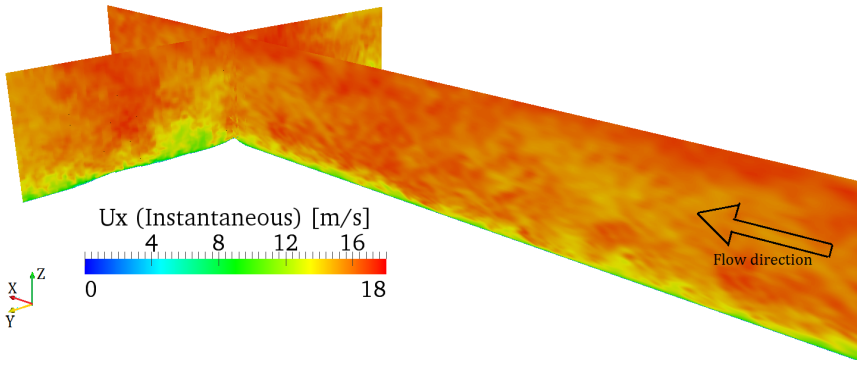


Fig. 6: Instantaneous horizontal flow fields on the stream-wise and span-wise planes.

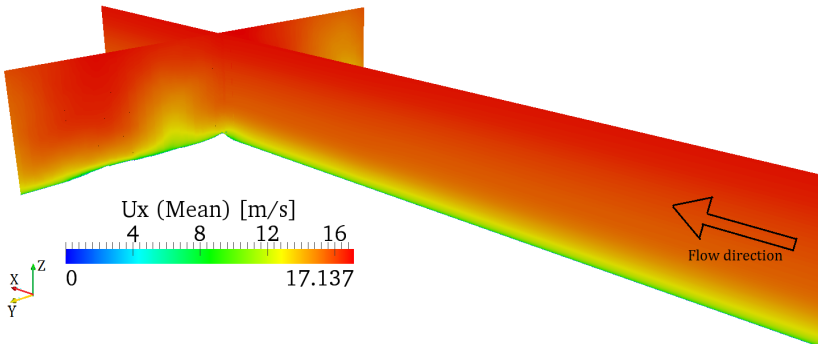


Fig. 7: 30-min time-averaged horizontal flow fields on the stream-wise and span-wise planes.

Conclusions

Wind flow near and after forest edge was investigated in a field experiment using two light detection and ranging devices which were capable to record wind up to 150 m height from the ground. After forest, flow recirculation at lower heights was observed. Also, the wind

decelerated which indicates the existence of high resistance within forest due to distribution of vegetation.

Moreover, this paper has presented a procedure for post-processing a real field wind data in order to classify the involved atmospheric boundary condition based on harmony forecast data, during 2 weeks of measurement. As a result, the neutral weather condition was selected to be on 2nd day of June 2013 between 21:00 to 23:10 useful for our numerical simulations.

Here, turbulent flow over Skinnarila terrain which has a small hill is studied by means of LES. Due to high amount of CPU time required to run these simulation, the LES results over porous-like forest together with smooth-non-forested terrain will be compared with field data in near future.

Acknowledgements

High-performance computing resources provided by CSC-IT Center for Science Ltd was used for all CFD simulations. In addition, we would like to acknowledge the National Land Survey of Finland for providing the terrain elevation dataset.

References

[1] Kunlun Bai, Joseph Katz and Charles Meneveau. (2015) Turbulent Flow Structure Inside a Canopy with Complex Multi-Scale Elements, *Boundary-Layer Meteorol* **155**, 435-457.
[2] Wenxin Huai, Wanyun Xue and Zhongdong Qian. (2015) Large-Eddy Simulation of Turbulent Rectangular Open-Channel Flow with an Emergent Rigid Vegetation Patch, *Advances in Water Resources* **80**, 30-42.
[3] Ebba Dellwik, Ferhat Bingol and Jakob Mann. (2014) Flow Distortion at a Dense Forest Edge, *Quarterly Journal of the Royal Meteorological Society* **140**, 676-686.
[4] Fabian Schlegel, Jörg Stiller, Anne Bienert, Hans-Gerd Maas, Ronald Queck and Christian Bernhofer. (2012) Large-Eddy Simulation of Inhomogeneous Canopy Flows Using High Resolution Terrestrial Laser Scanning Data, *Boundary-Layer Meteorol* **142**, 223-243.
[5] Sylvain Dupont, Mark R. Irvine, Jean-Marc Bonnefond, Eric Lamaud and Yves Brunet. (2012) Turbulent Structures in a Pine Forest with a Deep and Sparce Trunk Space: Stand and Edge Regions, *Boundary-Layer Meteorol.***143**, 309-336.
[6] Pierre Aumond, Valery Masson, Christine Lac, Benoit Gauvreau, Sylvain Dupont and Michel Berengier. (2013) Including the Drag Effects of Canopies: Real Case Large-Eddy Simulation Studies, *Boundary-Layer Meteorol* **146**, 65-80.
[7] Lopes Da Costa, J., Castro, F., Palma, J. and Stuart, P. (2006) Computer Simulation of Atmospheric Flows Over Real Forests for Wind Energy Resource Evaluation, *Journal of Wind Engineering and Industrial Aerodynamics* **94**, 603-620.
[8] Andrey Sogachev. (2009) A Note on Two-Equation Closure Modelling of Canopy Flow, *Boundary-Layer Meteorol* **130**, 423-435.
[9] Louis-Étienne Boudreault, Andreas Bechmann, Lasse Tarvainen, Leif Klemedtsson, Iurii Shendryk and Ebba Dellwik. (2015) A LiDAR Method of Canopy Structure Retrieval for Wind Modeling of Heterogeneous Forests, *Agricultural and Forest Meteorology* **201**, 86-97.
[10] Zeinab Ahmadi Zeleti, Heikki Haario and Jari Hämäläinen. (2014) Experimental validation of a porous medium modeling for air flow through forest canopy, *Europe's Premier Wind Energy Conference*, Barcelona, Spain.
[11] Ashvinkumar Chaudhari. (2014) Large-Eddy Simulation of Wind Flows Over Complex Terrains for Wind Energy Applications, *Doctoral Dissertation, Lappeenranta University of Technology.* 1-110.

Gust effect factors and natural sway frequencies of trees for wind load estimation

***Seunghoon Shin[1], Ilmin Kang[1], Seonggeun Park[1], Yuhyun Lee[1], Kyungjae Shin[1], Whajung Kim[1], and †Hongjin Kim[1]**

[1]Department of Architectural Engineering, Kyungpook National University, Korea

*Presenting author: ssh10004ok@knu.ac.kr
†Corresponding author: hjk@knu.ac.kr

Abstract

Strong winds have caused increasing wind damages for fruit trees such as uprooting and fruit drop in orchards worldwide. In order to prevent these wind damages, the various prop systems or support systems have been introduced for fruit trees. When a prop system is designed against strong winds, it is essential to calculate the wind load acting on each tree for accurate evaluation of wind resistance of prop system.

It is often to treat the applied wind load acting on a tree as a static load and to use beam theory to determine the maximum bending moment at the base of the tree. However, the response of a tree is frequency dependent and is affected mostly by wind gusts at frequencies close to its resonant frequency. In this situation, the dynamic effects are likely to increase the bending of stems and hence the maximum bending moment at the base of the tree. These dynamic effects are likely significant and cannot be ignored when the natural sway frequency of a tree is relatively small, that is, the tree is flexible.

There are two approaches to quantifying the response of a tree to a given fluctuating wind load. First, the wind load and tree response spectra are experimentally measured and a transfer function from the wind load to tree response is developed. Alternatively, if the information on the dynamic properties such as natural sway frequency and damping ratio of trees are available, then it is possible to characterize their response to any fluctuating wind load by employing a wind engineering theory. In many design codes or standards, this dynamic effect is considered adopting the gust effect factor and empirical formulae for the factor are given as functions of the natural sway frequency and damping ratio. The threshold natural sway frequency in most design codes that the dynamic effect against fluctuating wind load needs to be considered carefully is 1.0 Hz.

This paper presents the system identification method to measure the natural sway frequencies and damping ratios of fruit trees for the evaluation of the wind load acting on the trees. Both the ambient vibration test and free vibration test are performed and the identified dynamic properties are compared. It is found the average natural frequency of fruit trees is less than 1.0 Hz, and thereby the dynamic effect against fluctuating wind load needs cannot be ignored. Further, it is found that the damping ratios of fruit trees are quite larger than those of civil and building structures due to the soil-structure interaction. Therefore, a special care is required when the prop systems for fruit trees are designed against strong winds.

Keywords: Tree Supporting system, Wind load, Gust effect factor, Ambient vibration test, System identification.

Introduction

Typhoon has caused increasing wind damages for fruit trees at the orchard such as uprooting and fruit drop in Korea recently. The resistance to uprooting moment of the tree is typically the weakest mechanical link for shallow-rooted trees subjected to strong winds (Lundström et al. 2007). In order to prevent these wind damages and to enhance the uprooting moment

capacities of trees in orchards, the wind break forest and the various prop systems, or supporting systems, have been introduced to fruit orchards (He and Hoyano, 2010). Three most typical types of apple tree prop system used in Korea are 1) individual prop system, 2) steel pipe fence prop system, and 3) concrete column fence prop system.

However, most of these prop systems were originated from the regions where the strong tropical storm like a typhoon does not occur (Lespinasse and Delort, 1986). Further, the most of studies on the apple tree prop system has focused on the annual yield and profits (Robison et al., 2007, Palmer et al., 1992). Therefore, it is required to evaluate the wind resisting performance of fruit tree prop systems which are frequently used in Korea.

When a prop system is designed against strong winds, it is essential to calculate the wind load acting on each tree for accurate evaluation of wind resistance of the prop system. It is often to treat the applied wind load acting on a tree as a static load and to use beam theory to determine the maximum bending moment at the base of the tree. However, the response of a tree is frequency dependent and is affected mostly by wind gusts at frequencies close to its resonant frequency (Hu et al., 2009). In this situation, the dynamic effects are likely to increase the bending of stems and hence the maximum bending moment at the base of the tree. These dynamic effects are significant and cannot be ignored when the natural frequency of a tree is relatively small, that is, the tree is flexible.

There are two approaches to quantifying the response of a tree to a given fluctuating wind load (Moore and Maguire, 2004). First, the wind load and tree response spectra are experimentally measured and a transfer function from the wind load to tree response is developed. Alternatively, if the information on the dynamic properties such as natural frequency and damping ratio of trees are available, then it is possible to characterize their response to any fluctuating wind load by employing a wind engineering theory. In many design codes or standards, this dynamic effect is considered adopting the gust effect factor and empirical formulae for the factor are given as functions of the natural frequency and damping ratio. The threshold natural frequency in most design codes that the dynamic effect against fluctuating wind load needs to be considered carefully is 1.0 Hz (AIJ, 2009, ASCE, 2010).

This paper presents the system identification to measure the natural frequencies and damping ratios of fruit trees for the evaluation of the wind load acting on the trees. Both the ambient vibration test and free vibration test are performed and the identified dynamic properties are compared. The dynamic properties obtained using the previously reported empirical formulae are also compared to experimentally identified ones. Next, the gust effect factors for each tree are evaluated using the formula given in Korean Building Code, which is termed as KBC2009 hereafter (AIK, 2009).

Wind Load on the tree supporting system

Wind load on a tree

Figure 1 shows the steel pipe fence type prop system, which is most commonly used for apple orchards in Korea. Three or more trees are planted between two vertical pipes that are spaced 6 m, and the wind load acting on trees are transferred to vertical supports by horizontal wires installed at every 80 cm. Since the stiffness in the longitudinal direction is much larger than that in the normal direction, uprooting damages generally occurs in the normal direction and most trees connected to a fence are damaged simultaneously.

Figure 1. Steel pipe fence type prop system

The wind load acting on a tree, P, is calculated as (Simiu and Scanlan, 1996)

$$P = q_w A \tag{1}$$

where q_w is the wind pressure (N/m^2). The wind pressure q_w is given by

$$q_w = 0.5\rho C_D G_f V_z^2 \tag{2}$$

where ρ is the air density (kg/m^3), C_D is the drag coefficient (dimensionless), G_f is the gust effect factor (dimensionless), and V_z is the design wind velocity at height z (m).

The drag coefficients C_D of trees in Eq. (2) are generally obtained experimentally using a wind tunnel and some typical values are given for various tree types (Mayhead, 1973, Vollsinger et al., 2005). On the contrary, only limited studies have been performed on the gust effect factor G_f of trees since it is affected by many features such as tree species, age, height, stem diameter, and spacing (Gardiner et al., 2000). In this study, the gust effect factor is obtained and analyzed applying empirical formulae provided in literatures and design codes that are obtained based on a wind engineering theory.

Gust effect factor

The gust effect factor is defined as a ratio of the maximum response to mean response of a structure and is given as (Simiu and Scanlan, 1996)

$$G_f = \frac{X_{\max}}{\overline{X}} = 1 + g_f \frac{\sigma_X}{\overline{X}} \tag{3}$$

where X_{max} is the maximum response, $\overline{X}$ is the mean response, g_f is a peak factor, and σ_x is the standard deviation of the response.

Gardiner et al. (2000) proposed the following empirical formula obtained from a wind tunnel test using scaled tree models.

$$G_{\max} = \left(2.7193\frac{s}{H} - 0.061\right) + \left(-1.273\frac{s}{H} + 0.9701\right)\left(1.1127\frac{s}{H} + 0.0311\right)^{x/H}$$

$$G_{mean} = \left(0.68\frac{s}{H} - 0.0385\right) + \left(-0.68\frac{s}{H} + 0.4875\right)\left(1.7239\frac{s}{H} + 0.0316\right)^{x/H}$$

$$G_f = \frac{G_{\max}}{G_{mean}} \qquad \text{(3.a, b, and c)}$$

where s is the tree spacing (m), H is the tree height, and x is the distance from the forest edge (m).
Davenport and Surray (1990) defined the gust effect factor for low rise structures as

$$G_f = 1 + \psi\varphi\sqrt{k_1 + k_2} \qquad (4)$$

where ψ is the peak factor (dimensionless), φ is the exposure factor (dimensionless), k_1 is the background turbulence factor (dimensionless), and k_2 is the gust resonant factor.

Peak factor ψ depends on the natural frequency of the structure, that is, it increases as a logarithmic function of natural frequency of the structure increases. Further, the gust resonant factor k_2 also is a function of the natural frequency of the structure. The damping ratio of the structure affects the gust resonant factor as well. Consequently, the accurate evaluation of the natural frequency and damping ratio is critical for the gust factor calculation.

Eq. (4) is adopted in many design codes including KBC2009. The peak factor ψ and the exposure factor φ in KBC2009 are given as

$$\psi = \sqrt{2\ln(600\nu_f) + 1.2} \qquad (5)$$

$$\varphi = \left(\frac{3 + 3\alpha}{2 + \alpha}\right) I_z \qquad (6)$$

where α is the power law exponent of mean wind speed profile for a given terrain roughness category, υ_f and I_z are , respectively, the level crossing number and turbulence density at the reference height and given as

$$\nu_f = n_0\sqrt{\frac{k_2}{k_1 + k_2}} \qquad (7)$$

$$I_z = 0.1\left(\frac{z}{Z_g}\right)^{-\alpha-0.05} \tag{8}$$

where n_0 is the natural frequency of the structure (Hz) and Z_g is the nominal height of the atmospheric boundary layer.
The background turbulence factor k_1 and the gust resonant factor k_2 in KBC2009 are defined as

$$k_1 = 1 - \left[\frac{1}{\left\{1 + 5.1\left(L_H/\sqrt{HB}\right)^{1.3}\left(B/H\right)^{0.33}\right\}^{1/3}}\right] \tag{9}$$

$$k_2 = \frac{\pi}{4\varsigma_f} S_f F_s \tag{10}$$

where B is the width of the structure, ς_f is the damping ratio, L_H is turbulence density at the reference height, and S_f and F_s are, respectively, the size reduction factor and the spectral energy factor given as

$$S_f = \frac{0.84}{\left\{1 + 2.1(n_0 H/V_H)\right\}\left\{1 + 2.1(n_0 B/V_H)\right\}} \tag{11}$$

$$F_s = \frac{4(n_0 L_H/V_H)}{\left\{1 + 71(n_0 L_H/V_H)^2\right\}^{5/6}} \tag{12}$$

where V_H is the design wind speed at the top of the structure.
For the structure with a natural frequency of less than 1.0 Hz, the structures is classified as a rigid structure and its gust effect factor is simply given as Eq. (13) omitting the gust resonant factor k_2 and letting the value of and the exposure factor φ to be 4 from Eq. (4)

$$G_f = 1 + 4\varphi\sqrt{k_1} \tag{13}$$

Natural frequency and damping ratio of trees

From Eqs. (7), (11), and (13), it can be noticed that the natural frequency is required for the gust effect factor calculation. Further, it can be noted from Eq. (10) that the damping ratio needs to be known as well.

Moore and Maguire (2004) investigated previously reported natural frequency measurement from 602 trees, which belong to eight different species, and showed that natural frequency is strongly and linearly related to the ratio of diameter at breast height to total height squared. They presented the following empirical formula based on a regression analysis.

$$n_0 = 0.0766 + 3.1219\frac{D_{bh}}{H^2} \tag{14}$$

where D_{bh} is the diameter at breast height (cm).

They proposed another empirical formula to consider the species difference given by Eq. (15) where I_p is an indicator variable.

$$n_0 = 0.0948 + 3.4317 \frac{D_{bh}}{H^2} - 0.7765 I_p \frac{D_{bh}}{H^2} \tag{15}$$

The value of I_p is 1.0 if the genus is Pinus and 0.0 otherwise.
Moore and Maguire also investigated the damping ratio of trees from the previous researches and classified it into two categories; 1) internal damping is due to the friction of the root-soil connection, structural damping resulting from the movement of branches and the internal friction of the wood, and 2) external damping due to the aerodynamic drag of the crown and also to collisions between crowns of neighboring trees. They concluded that the internal damping ratios are generally less than 0.05 and do not appear to be related to tree size, while the external damping is wind velocity dependent and much larger than the internal one.

Field measurement of natural frequencies and damping ratio

Test specimens and methods

A field vibration test was performed to measure the natural frequencies and damping ratios of orchard trees. The apple trees were used for the test. Both the ambient vibration test and free vibration test were performed and the identified dynamic properties were compared.

The trees were supported by the steel pipe fence type prop system and four to five trees were planted between two vertical steel pipes. The test was performed when trees were heavy with clusters of apples since the typhoon damages occur mostly before and during harvest season. Total of 20 trees were used in the test.

In order to analyze the effect of the prop on the dynamic properties of trees, a half of specimens were tested after cutting all horizontal wires connected to the trees while the prop for the rest of specimens remained intact.

Two piezoelectric accelerometers were installed at 1.5 m high, one in the longitudinal direction (x-direction) and another in the normal direction (y-direction) to measure the accelerations of trees without a steel pipe prop. On the contrary, only one accelerometer was used in the y-direction for trees with a prop because the frequency in the x-direction is considerably affected by the prop due to large stiffness.

The ambient vibration test was carried out for 10 minutes with a sampling frequency of 360 Hz. The free vibration test was performed by simply pushing trees about 30 cm slowly by human and letting trees vibrate freely. Five human-induced free vibrations were performed continuously for both x- and y-directions for trees without a steel pipe prop, while those were performed for the y-direction only for trees with a prop. The only acceleration measured in the same direction to the free vibration direction is utilized for the identification of dynamic properties for the free vibration test.

Identified natural frequencies and damping ratio

The power spectrum densities (PSDs) of measured accelerations from two test methods were obtained to identify the natural frequencies of trees. Then the half-power band-width method was applied to the obtained PSDs for damping ratio estimation (Clough, R. W. and Penzien, J, 1995, Xiong et al., 2011).

The peaks of PSDs are considerably noticeable at the fundamental natural frequencies in both ambient and free vibration tests, while the values of PSDs in the ambient vibration test contains the higher modes and DC contents. The distinction of PSDs near the fundamental frequency in the free vibration test is mainly due to the fact that trees oscillate at their fundamental frequency under a free vibration.

The identified natural frequencies of the test specimens are summarized in Tables 1 and 2. Note that the only natural frequencies in the y-direction are identified in Table 2 for trees with a prop since the acceleration were measured in that direction only.

It can be seen from Table 1 that the natural frequencies of trees in the x- and y-directions are almost same except the test specimen T3, T8, and T9. It can be also seen that the identified natural frequencies from the free vibration tests are generally smaller than those from the ambient vibration tests. This is because the natural frequency of a structure is generally inversely proportional to its response amplitude and the amplitudes of measured accelerations in the free vibration test are significantly larger than those in the ambient vibration test. In average, the natural frequencies obtained from the free vibration test are 3.90 % and 6.06% smaller in the x- and y-directions, respectively, than those from the ambient vibration test.

The natural frequencies of the trees with a prop are found to be increased compared to those without a prop. Those with a prop are 15.73 % and 13.70 % larger than those without a prop in average (Table 2). This concludes that the stiffness of the steel pipe fence prop helps to increase the stiffness of trees in the y-direction. That is the overall uprooting moment resistance capacities of trees are increased due to the installation of the prop.

Table 1. Identified natural frequencies of trees without a prop

Specimen	Ambient vibration test		Free vibration test	
	x-dir. (Hz)	y-dir. (Hz)	x-dir. (Hz)	y-dir. (Hz)
T1	0.807	0.807	0.779	0.791
T2	0.907	0.907	0.908	0.870
T3	0.807	0.630	0.756	0.655
T4	0.857	0.882	0.807	0.857
T5	0.907	0.958	0.907	0.907
T6	0.958	1.058	0.958	1.008
T7	1.134	1.046	1.008	1.008
T8	1.210	1.411	1.159	1.109
T9	1.084	0.958	1.008	0.907
T10	1.109	1.159	1.109	1.109
Average	0.978	0.982	0.922	0.922

Table 2. Identified natural frequencies of trees with a prop in the y-direction

Specimen	Ambient vibration test (Hz)	Free vibration test (Hz)
T11	0.958	0.907
T12	0.857	0.756
T13	0.807	0.756
T14	1.512	1.411
T15	1.445	1.336
T16	0.907	0.832
T17	1.498	1.210
T18	1.033	1.008
T19	1.159	1.109
T20	1.184	1.159
Average	1.136	1.048

In Table 3, the calculated natural frequencies of trees using the empirical formulae provided in Eqs. (14) and (15) are presented for comparison with experimentally identified ones. Compared with identified natural frequencies provided in Tables 1 and 2, the empirical formulae proposed by Moore and Maguire overestimate the natural frequencies up to 234 %

in average. Therefore, it can be concluded that the empirical formulae do not cover every genus of trees even though they are obtained from more than 600 experimental data.

Table 3. Natural frequencies of trees from empirical formulae

Specimen	Eq. (14) (Hz)	Eq. (15) (Hz)	Specimen	Eq. (14) (Hz)	Eq. (15) (Hz)
T1	4.169	4.593	T11	2.612	2.881
T2	2.497	2.755	T12	2.375	2.622
T3	2.358	2.603	T13	2.301	2.540
T4	2.351	2.595	T14	1.545	1.709
T5	2.892	3.190	T15	3.046	3.359
T6	3.193	3.521	T16	2.455	2.709
T7	2.723	3.048	T17	2.443	2.696
T8	4.063	4.477	T18	2.825	3.116
T9	2.618	2.888	T19	2.682	2.959
T10	2.083	3.092	T20	2.945	3.248
Average	2.971	3.276	Average	2.523	2.784

Table 4. Identified damping ratios of trees without a prop

Specimen	Ambient vibration test		Free vibration test	
	x-dir. (%)	y-dir. (%)	x-dir. (%)	y-dir. (%)
T1	6.53	7.68	4.63	3.05
T2	6.56	7.08	11.10	16.41
T3	7.35	6.10	7.42	8.90
T4	7.18	3.79	8.62	7.03
T5	6.86	7.21	9.73	7.03
T6	6.48	7.05	7.82	11.10
T7	3.54	5.60	9.52	9.61
T8	7.92	6.62	6.98	14.59
T9	3.49	6.36	9.84	7.78
T10	4.15	5.74	5.95	5.06
Average	6.01	6.32	8.16	9.06

Table 5. Identified damping ratios of trees with a prop in the y-direction

Specimen	Ambient vibration test (%)	Free vibration test (%)
T11	5.31	10.91
T12	2.64	17.27
T13	8.97	8.72
T14	2.48	14.19
T15	8.30	12.41
T16	6.30	10.55
T17	6.73	20.61
T18	19.43	21.79
T19	6.98	11.15
T20	9.29	8.17
Average	7.64	13.58

Tables 4 and 5 present the identified damping ratios of trees with and without a prop from the ambient and free vibration test. It can be seen from Table 4 that the identified damping ratios of trees without a prop from the free vibration test were significantly larger than those from the ambient vibration test. They are 35.88% larger in the x-direction and 43.22 % larger in the y-direction in average. This is because the external damping as well as internal damping plays a role when the trees are oscillating with large magnitudes as Moore and Maguire reported. The average damping values of 6.01 % and 6.32 % obtained from the ambient vibration test match well to the internal damping of 5 % reported by Moore and Maguire.

Compared to the damping ratio of trees without a prop, those with a prop in Table 5 are 20.88 % and 49.92 % larger in the ambient and free vibration tests, respectively. Consequently, it can be concluded that the wires attached to the trees in the steel pipe fence prop increase not only stiffness but also damping ratios of trees.

Gust effect factor evaluations

The gust effect factors are calculated and summarized in Tables 6 and 7. Both formulae for non-rigid structures in Eq. (3) and rigid structures in Eq. (4) are utilized since the identified natural frequencies of trees are almost 1 Hz. For comparison, the results of the empirical formula in Eq. (13) proposed by Gardiner et al. are also presented in Tables 6 and 7. For Eqs. (3) and (4), the identified natural frequencies and damping ratio from the ambient vibration test were used because the smaller damping ratios produce more conservative wind load estimation. For Eq. (13), the tree spacing is set to be 1.5 m, and the distance from the forest edge is assumed to be zero for conservative condition.

Table 6. Gust effect factors of trees without a prop

Specimen	Eq. (3)	Eq. (4)	Eq. (13)
T1	2.925	2.493	3.992
T2	2.938	2.494	3.985
T3	3.101	2.468	3.827
T4	3.312	2.473	3.859
T5	2.865	2.469	3.827
T6	2.922	2.508	4.054
T7	2.962	2.469	3.840
T8	2.875	2.514	4.095
T9	2.963	2.485	3.920
T10	2.961	2.491	3.955
Average	2.982	2.486	3.935

Table 7. Gust effect factors of trees with a prop

Specimen	Eq. (3)	Eq. (4)	Eq. (13)
T11	2.783	2.362	3.330
T12	3.490	2.422	3.588
T13	2.746	2.435	3.631
T14	3.207	2.425	3.565
T15	2.629	2.433	3.652
T16	2.910	2.450	3.726
T17	2.685	2.426	3.597
T18	2.506	2.467	3.802
T19	2.740	2.428	3.616
T20	2.602	2.413	3.547
Average	2.830	2.426	3.606

It can be noticed that the gust effect factors obtained using Eq. (4) are 14.27 % to 16.63 % smaller than those obtained using Eq. (3). Therefore, if the flexible nature of trees is neglected, the total wind load can be underestimated noticeably.

The empirical formula proposed by Gardiner et al. yields 27.41 % to 31.96 % lager gust effect factors compared to those by the formula for non-rigid structures, and 48.62 % to 58.29 % lager ones compared to those by the formula for rigid structures Therefore, it can be concluded that the empirical formula that does not require the exact values of natural frequency and damping ratio overestimates the gust effect factor considerably.

Conclusions

The gust effect factors of trees are analyzed for the wind load estimation of the tree supporting system. Since the value of gust effect factor depends on the natural frequency and damping ratio, the field experiment was performed to identify the accurate dynamic properties of the trees.

The 20 apple trees were used for the field test, in which a half of them were tested after cutting all horizontal wires connected to the trees while the prop for the rest of specimens remained intact. Both the ambient vibration test and free vibration test were performed and the identified dynamic properties were compared.

It was found that the average natural frequency of fruit trees is about 1.0 Hz, and thereby the dynamic effect against fluctuating wind load needs cannot be ignored. Further, it is found that the damping ratios of fruit trees are quite larger than those of civil and building structures due to the external damping effect. The wires attached to the trees in the steel pipe fence prop increase both stiffness and damping ratios of trees.

The gust effect factor analysis results indicate that the total wind load can be underestimated noticeably if the flexible nature of trees is neglected. If the empirical formula that does not require the exact values of natural frequency and damping ratio is used, the gust effect factor was overestimated considerably.

Acknowledgement

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries(IPET) through Advanced Production Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(315092)

References

[1] American Society of Civil Engineers (2010) ASCE 7-10 Minimum Design Loads for Buildings and Other Structures.
[2] Architectural Institute of Japan (2004) Recommendations for Loads on Buildings.
[3] Architectural Institute of Korea (2009) Korea Building Code.
[4] Clough, R. W. and Penzien, J (1995), Dynamics of structures, 3rd ed. Computers & Structures, Inc., Berkeley, CA.
[5] Crook, M. J., and Ennos, A. R. (1996) The anchorage mechanics of deep rooted larch, Larix europea x L japonica, Journal of Experimental Botany, 47(10), 1509-1517.
[6] Davenport, A. C. and Surry, D. J. (1974) The pressure on low rise structures in turbulent wind, Canadian Structural Engineering Conference.
[7] Gardiner, B., Peltola, H., and Kellomaki, S., (2000) Comparison of two models for predicting the critical wind speeds required to damage coniferous trees, Ecological Modeling, 129, 1-23.
[8] He, J. and Hoyano, A. (2009) The effects of windbreak forests on the summer thermal environment in a residence, Journal of Asian Architecture and Building Engineering, 8(1), 291-298.

[9] Hu, S., Fujimoto, T., and Chiba, N. (2009) Pseudo- dynamics model of a cantilever beam for animating flexible leaves and branches in wind field. Computer Animation and Virtual Worlds, 20(2-3), 279-287.
[10] Lespinasse, J.M. and Delort, J.F. (1986) Apple tree management in vertical axis: appraisal after ten years of experiments. Acta Horticulture, 160, 139-155.
[11] Lundström, T., Jonas, T., Stöckli, V., and Ammann, W. (2007) Anchorage of mature conifers: resistive turning moment, root–soil plate geometry and root growth orientation, Tree Physiology, 27(9), 1217-1227.
[12] Mayhead, G. J. (1973) Some drag coefficients for British forest trees derived from wind tunnel studies, Agricultural Meteorology, 12, 123-130.
[13] Moore, J. R., and Maguire, D. A. (2004) Natural sway frequencies and damping ratios of trees: concepts, review and synthesis of previous studies. Trees, 18(2), 195-203.
[14] Palmer, J.W., Avery, D.J., and Wertheim, S.J. (1992) Effect of apple tree spacing and summer pruning on leaf area distribution and light interception, Scientia Horticulturae, 52, 303-312.
[15] Robinson, T., Hoying, S. A., DeMaree, A., Iungerman, K, and Fargione, M. (2007) The evolution towards more competitive apple orchard systems in New York, New York Fruit Quarterly. 15(1), 3-9.
[16] Simiu, E. and Scanlan, R. H. (1996), Winds Effects on Structures: Fundamentals and Applications to Design, Wiley, New York.
[17] Vollsinger, S. Mitchell, S. J., Byrne, K. E., Noval, M. D., and Rudnicki, M. (2005) Wind tunnel measurements of crown streamlining and drag relationships for several hardwood species, Canadian Journal of Forest Research, 35 (5), 1238-1249.
[18] Xiong, H., Kang, J., and Lu, X. (2011), Field Testing and Investigation of the Dynamic Performance and Comfort of Timber Floors, Journal of Asian Architecture and Building Engineering, 10(2), 407-412

Numerical simulation of the grains growth on titanium alloy electron beam welding process

†Xiaogang Liu[1,2], Haiding Guo[1,2], and M. M. Yu[1]

[1]Nanjing University of Aeronautics and Astronautics, Jiangsu Province Key Laboratory of Aerospace Power System, Nanjing 210016, China
[2] Collaborative Innovation Center of Advanced Aero-Engine, China
†Corresponding author: liuxg03@nuaa.edu.cn

Abstract

In the paper, the grains growth of TC4-DT alloy joint during EBW (electron beam welding) process was simulated by using Cellular Automaton method. In order to consider the effects of the growth of neighborhood cellular on the centre cells in the model, the solid fraction and solute distribution algorithms of classical CA model was improved. The growth of equiaxed grains and columnar crystals under uniform and non-uniform temperature field were simulated successfully by applying the modified model respectively. The temperature distribution near the fusion line of TC4-DT EBW joint was also calculated by using double ellipsoid heat source model. Then coupling the CA model with the temperature field, the grains growth process of the cross section of the welded zone was simulated. The simulation result fits well with experimental ones on the morphology and the size of the columnar crystals.

Keywords: Grain growth, Cellular Automata, Electron Beam Welding, Columnar Crystal, Titanium alloy

1. Introduction

TC4-DT (Damage Tolerance) alloy is a kind of $\alpha+\beta$ dual phase titanium alloy, its chemical composition approximate to Ti6Al4V. Compared with other medium strength titanium alloy TC4-DT alloy has higher fatigue resistance and damage tolerance properties (lower fatigue crack propagation rate and high fatigue crack propagation threshold), P. F. Fu (2014)[1], L. Tong (2010)[2]. In addition, with excellent weld-ability, TC4-DT alloy is suitable for EBW (Electron Beam Welding) process well. Recently, this alloy has been widely used in industry of aviation and aerospace for its superior mechanical properties.

The final mechanical properties of welded joints primary controlled by physical behavior and microstructure of weld fusion zone during solidification. Therefore, more and more investigations on microstructure simulation of weld pool during the solidification process have been performed to predict the properties of weld joints, T. Zacharial and J. M. Goldakt (1995)[3].

Rapid development in computer technology in recent decades have allowed the use of numerical simulation as powerful tools for developing our understanding of grain growth during solidification. Numerous investigations have been performed to develop various computation models, such as Monte Carlo (MC) models, D. J. Srolovitz (1983)[4], P. P. Zhu (1992)[5], Cellular Automata (CA) models, M. A. Zaeem (2012)[6], A. Choudhury (2012)[7], Phase Field (PF) models, G. J. Fix (1983) [8], R. Kobayashi (1993) [9], C. Beckermann (2001)[10], and so on. Among of these, CA models are the most promising methods for

description the growth of equiaxed and columnar grains in two or three dimensions. S.Wolfarm (1983)[11] firstly discussed the self-organizing behavior in cellular automata as a computational process. In this investigation, formal language is used to extend dynamical systems theory descriptions of cellular automata.J. D. Hunt (1984)[12] presented a CA model for the growth of equiaxed grains ahead of the columnar front during directional solidification. The model considers both single-phase and eutectic equiaxed growth. A simple expression is obtained which can predicts when fully equiaxed structures should occur. M. Rappaz , Thevoz and J. L. Desbiolles(1989)[13] proposed a FEM coupling with CA approach to model equiaxed microstructure formation in casting. In this CA model takes into account nucleation of new grains within the undercooled melt, and the kinetics of the dendrite tips of the eutectic front in the case of dendritic alloys. Subsequently, lots of intensively research on the CA method had been conducted by M. Rappz and C. A. Gandin (1993-1997)[14]-[16], they established the overall framework of this approach, and the application range of the model was developed from two-dimensional to three-dimensional, from the uniform temperature field to the non uniform temperature field. Later, some advanced and modified model have been proposed based on M. Rappaz's work. O. Zinovieva (2015)[17] proposed a improvement two-dimensional CA by introducing two new corrections to eliminate the artificial anisotropy , which based on a combination of the CA and FD methods developed by Rappaz and Gandin. The improvement CA model can be applied to simulate the complex grain morphologies during solidification. Baichen Liu and Q. Y. Xu (2015)[18] presented a three-dimensional CA model to prediction of single dendrite and polycrystalline dendrite growth of ternary alloys. In their model, introduces a modified decentered octahedron algorithm for neighborhood tracking to eliminate the effects of mesh dependency on dendrite growth.

In this paper, considering the TC4-DT alloy EBW process the microstructural evolution of weld pool during solidification was simulated by using a improved CA model. The morphology and size of columnar grains in weld pool were predicted.

2. Model theory

A modified CA model coupled with finite element (FE) method was developed to simulate the grains growth of EBW molten pool during solidification. The nodes temperature calculated with software MSC. Marc were conversion into cells of CA model by applying linear interpolation method.

2.1 Heat Source Model

In order to accurately calculate the temperature distribution of weld zone, the most important is to establish a reasonable heat source model. Taking into account that the EBW has the characteristics of energy input intensively, small heating area, fast moving speed, non-uniform energy density distribution and so on, a double ellipsoid heat source model was employed to simulate the temperature field of the welding process. The double ellipsoid heat source mode, as shown in Fig.1, composed of two quarter ellipsoid of front and rear with different parameter.

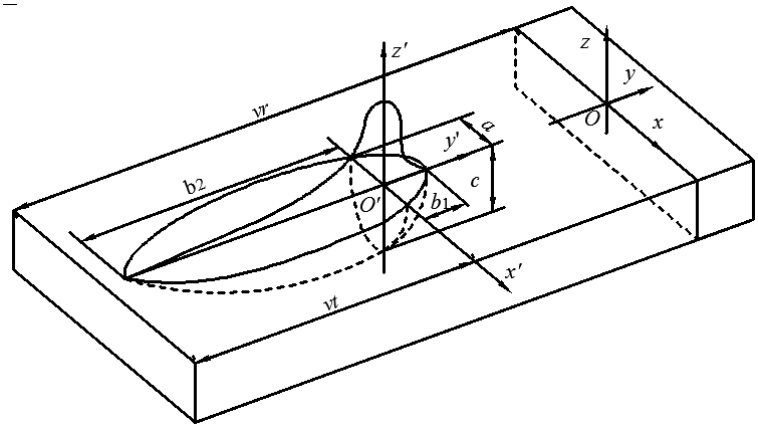


Figure 1. Schematic drawing of double ellipsoid heat source model

The heat flux within any *xy* sectional plane fits Gaussian distribution. Moreover, the total energy density reduced gradually along the depth direction of the weld pool. The heat flux $q(x,y,z)$ can be expressed as the following, P. Lacki (2011)[19]:

$$q_1(x,y,z)=\frac{6\sqrt{3}f_1Q}{\pi ab_1c}\exp(-3\frac{x^2}{a^2})\exp(-3\frac{y^2}{b_1^2})\exp(-3\frac{z^2}{c^2}) \quad \text{(front model)} \quad (1)$$

$$q_2(x,y,z)=\frac{6\sqrt{3}f_2Q}{\pi ab_2c}\exp(-3\frac{x^2}{a^2})\exp(-3\frac{y^2}{b_2^2})\exp(-3\frac{z^2}{c^2}) \quad \text{(rear model)} \quad (2)$$

Where Q is the overall input power given by $Q=\eta U_0 I_0$, η is thermal efficiency, η and η are welding voltage and current, respectively, a, b_1, b_2, c the ellipsoid semi-axes, f_1 and f_2 are the fraction power assigned to ellipsoid quarter, and $f_1+f_2=2$.

2.2 Description of CA Model

In CA models, the simulated area is discretized to be finite cells and time is discretized as time steps. Each time step is called 1CAS, which is defined as the time interval for all cells to undergo a variable calculation. The CA model includes four important parts, such as cellular state, cellular space, cellular neighborhood and transition rule. During the simulation, the state of each cell is determined by the states of its nearest neighbors through a transition rule. The solidification process can be simulated by the transition of the microcosmic cells from liquid to solid, i.e. the change of the solid fraction in each cell from 0 to 1.

The mesh of two dimension CA can be regular triangle or square in most cases. Two types of neighborhood, Von Neumann and Moore, are mostly used in square mesh. The Moore neighborhood model, as shown in Fig.2 , was employed in this paper. There are eight neighbors to the central cell. The traditional CA model thought that all eight neighbors around the central have the same possibility of being capture to transit its stage during solidification, it ignore the difference of distance to different neighborhood. In this paper, the neighborhood cells were divided into two types depend on the location to the central cell. The cells orthogonal to the central cells called type I neighborhood, such as (i, j-1), (i, j+1), (i-1, j), (i+1, j) show in Fig.3 and the cells located on the diagonal of the central cells called type II neighborhood, such as (i-1, j-1), (i-1, j+1), (i+1, j-1), (i+1, j-1). The probability that type I cells were captured is $\sqrt{2}$ times as much as that of type II cells.

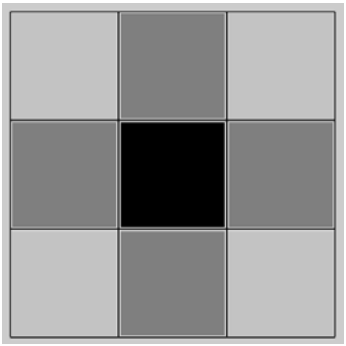


Figure 2. Moore neighborhood model

***i*-1, *j*-1 (II)**	***i*-1, *j* (I)**	***i*-1, *j*+1 (II)**
***i*, *j*-1 (I)**	***i*, *j***	***i*, *j*+1 (I)**
***i*+1, *j*-1 (II)**	***i*+1, *j* (I)**	***i*+1, *j*+1 (II)**

Figure 3. illustration of the type I & type II neighborhood

2.3 Nucleation Model

According to the solidification feature of weld pool, the paper considers only the heterogeneous nucleation. A continuous nucleation distribution, dn/d(ΔT), can be used to describe the grains density increase dn with the increase in undercooling by d(ΔT). The total density of nuclei for a given ΔT is determined by:

$$n(\Delta T) = \int_0^{\Delta T} \frac{dn}{d(\Delta T)} d(\Delta T) \tag{3}$$

The change rate of nucleation density varies with degree of undercooling can be expressed by a Gaussian function, then the formula is as follows:

$$\frac{\mathrm{d}n}{\mathrm{d}(\Delta T)} = \frac{n_{\max}}{\sqrt{2\pi}(\Delta T_\sigma)} \exp\left[-\frac{1}{2}\left(\frac{\Delta T - \Delta T_N}{\Delta T_\sigma} \right)^2 \right] \tag{4}$$

Where $n_{\max}$ is the maximum nucleation density, ΔT_σ is the standard deviation of undercooling and ΔT_N is the mean nucleation undercooling.

2.4 Grain growth

KGT (Kurz, Giovanola and Trivedi, 1986) [20] model was applied to calculate the growth process of dendrite tip. The total undercooling at the dendrite tip can be expressed as the sum of four contributions accounting for the solute ΔT_C, curvature ΔT_r, thermal ΔT_T, and kinetic ΔT_K effects. Thus:

$$\Delta T = \Delta T_C + \Delta T_r + \Delta T_T + \Delta T_K \tag{5}$$

In this paper, the last two terms are neglected for their minor contributions to the total undercooling. Then the remaining terms can be expressed as follows.

$$\Delta T_C = (C_L - C_0) m_L \tag{6}$$

$$\Delta T_r = -\Gamma K \tag{7}$$

Where m_L is liquidus slope, C_L and C_0 are the liquid concentration in interface cells and initial liquid concentration, respectively. Γ is the Gibbs-Thomson coefficient, and K is curvature.

At a certain time, the liquid concentration in interface cells C_L^0 can be expressed by the previous step concentration C_L^0 and solid fraction f_S^0, solid fraction increment Δf_S, equilibrium partition coefficient k_0, as follow.

$$C_L = \frac{C_L^0(1-f_S^0)-k_0 C_L^0 \Delta f_S}{1-f_S^0-\Delta f} \tag{8}$$

The interface cell is not only the solute absorption, there will be the rejection of solute at the same time. The solute concentration at interface cell will keep constant when the solute absorption equal to rejection. Then the equilibrium solute concentrations (C_L^* and C_S^*) are given by:

$$C_L^* = C_0 - \frac{1}{m_L}[\Delta T - \Gamma K] \tag{9}$$

and $$C_S^* = k_0 C_L^* \tag{10}$$

It is assumed that the solute is distributed evenly to the neighbor cell in most current model, which ignore the difference of solute concentration in cells and the distance to cells. A improved solute partition algorithm was proposed in the paper. A weighting coefficient Φ_i ,considering the cells concentration difference and distance to central cells by division type I and II neighborhood mention above (section 2.2) , was introduced to the solute partition equation. The formula can be expressed as follows.

$$\Delta C = \sum_1^n \Delta C_i + \sum_1^m \Delta C_i \tag{11}$$

$$\Delta C_i = \Phi_i \Delta C \tag{12}$$

$$\Phi_i = \begin{cases} \dfrac{\sqrt{2}(C_L^* - \Delta C_i)}{\sum_{j=1}^n \sqrt{2}(C_L^* - \Delta C_i) + \sum_{k=1}^m (C_L^* - \Delta C_k)} & \text{(type I cells)} \\ \dfrac{C_L^* - \Delta C_i}{\sum_{j=1}^n \sqrt{2}(C_L^* - \Delta C_i) + \sum_{k=1}^m (C_L^* - \Delta C_k)} & \text{(type II cells)} \end{cases} \tag{13}$$

Where *n* and *m* are the number of type I cells and type II cells, respectively.

The state of a cell depends on its solid fraction during solidification. The cells are allowed to be one of three states: all liquid, all solid, or a mixture (an interface cell). The solid fractions

in liquid and solid cells are zero and unity, respectively, while interface cells have $0 < f_S < 1$. The solid fraction increment can be obtained by the following formula.

$$\Delta f_S = \frac{\Delta t}{\Delta x}(V_x + V_y - V_x V_y \frac{\Delta t}{\Delta x}) \tag{14}$$

Where V_x and V_x are the moving velocity of solid-liquid interface on *x* and *y* direction, respectively. The Δt is time step and Δx is grid size.
The solid fraction of a captured cell is usually increased uniformly with interface moving velocity, ignoring the increment direction and the influence of neighbor cell around it, which is disadvantage for square cell. An improved calculation method of solid fraction increment was proposed in this paper.

Three cases of solid fraction increased way were discussed according to the relative position of the captured cell (i.e. interface cell) and solidified neighborhood cells, as shown in Fig.4 (a1) (b1) (c1). The interface growth angle, φ , was employed, which has three candidate values, 0° , 45° , 63.4° (arctan(2)), as shown in Fig.4 (a2)(b2) (c2).
Three solid fraction incremental models with different growth angles are illustrated in Fig.4 (a)~(c) respectively.

- φ=0°
  Traditional approach with interface moving from 0 to Vt_1.
- φ=45°
  Two-stage model: (1) interface moving from 0 to Vt'_1; (2) interface moving from Vt'_1 to (Vt'_1+ Vt'_2).
- φ=63.4°
  Three-stage model: (1) interface moving from 0 to Vt_1; (2) interface moving from Vt''_1 to (Vt''_1+ Vt''_2); (3) interface moving from (Vt''_1+ Vt''_2) to (Vt''_1+ Vt''_2+ (Vt''_3).

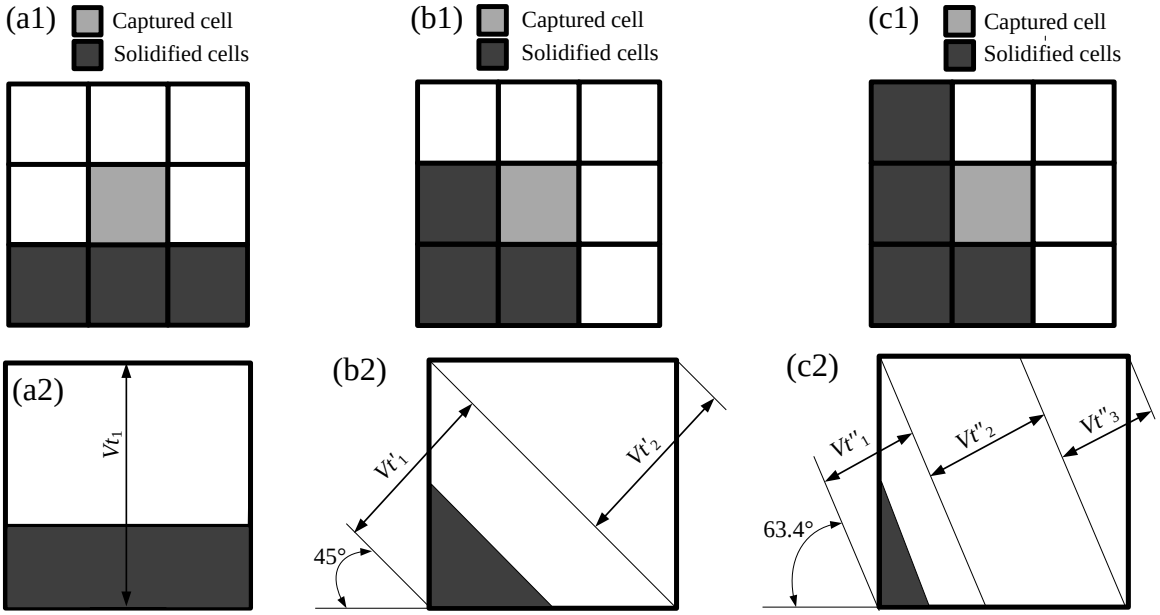


Figure 4. Different relative position of the captured cell and solidified neighborhood cells with (a1) $\varphi = 0^\circ$; (b1) $\varphi = 45^\circ$; (c1) $\varphi = 63.4^\circ$, and interface moving way in a captured cell of (a2) $\varphi = 0^\circ$; (b2) $\varphi = 45^\circ$; (c2) $\varphi = 63.4^\circ$

3. Results and discussion

Considering a titanium alloy, TC4-DT, numerical simulations of grains growth during solidification were conducted. The material property parameters used in this paper were listed in table 1.

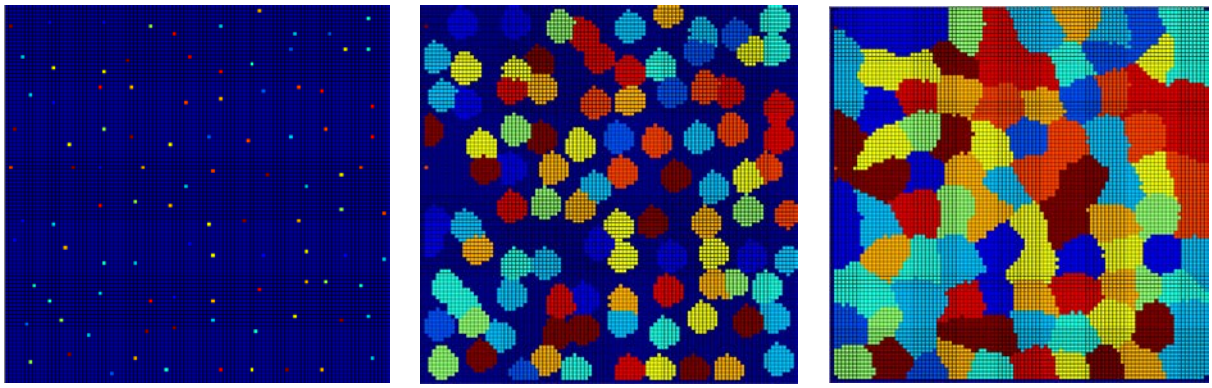
Table 1. Material properties parameters used in the simulation

Property	Variable	Value
Liquidus temperature	T_L	1703℃
Solidus temperature	T_S	1678℃
Partition coefficient	k_0	0.95
Diffusion coefficient in liquid	D_L	$5\times10^{-9}\,\mathrm{m}^2/\mathrm{s}$
Diffusion coefficient in solid	D_S	$5\times10^{-13}\,\mathrm{m}^2/\mathrm{s}$
Liquidus slope	m_L	-1.4
Maximum nucleation density	$n_{\max}$	$4\times10^{9}/\mathrm{m}^3$
Standard deviation of undercooling	ΔT_σ	0.5℃
Maximum undercooling	$\Delta T_{\max}$	2℃
Gibbs-Thomson coefficient	Γ	$3.66\times10^{-7}\,\mathrm{m\cdot K}$
Initial concentration	C_0	10.26 wt%

3.1 Growth of equiaxed grains under isothermal conditions

In order to test the validity of the model and program, the numerical simulation of equiaxed grains growth under hypothetical isothermal conditions were performed. The simulation region was divided into 100 ×100 square cells with the size of 0.01mm, and constant cooling rate was applied.

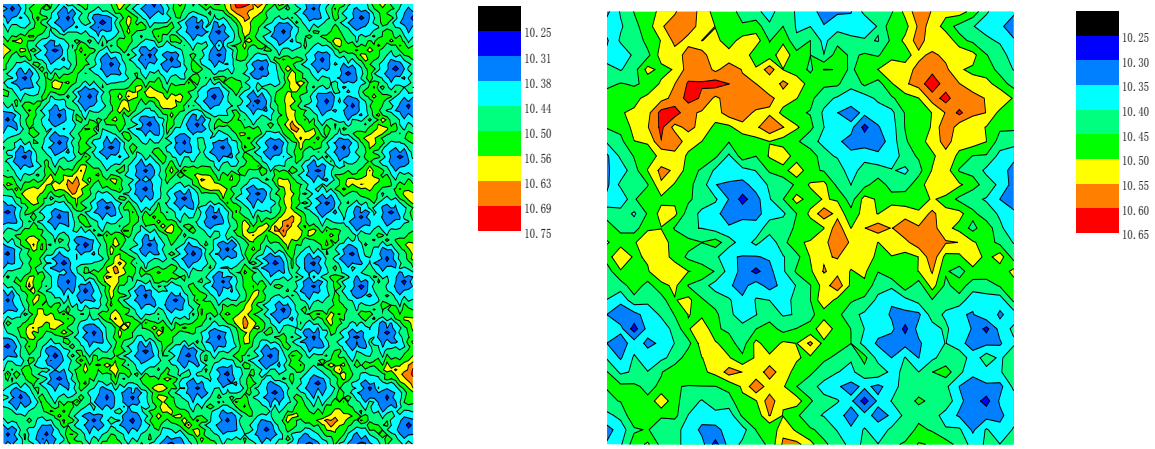
The simulation results was shown in Fig. 5. It can be observed that at the beginning of the solidification process, a large number of nuclei appeared randomly from liquid due to the undercooling. As the time step increases, the grains grow up gradually with relatively normal shape. And the grains which have the same characteristic value will merge into large one when they contact with each other. At the end of the solidification, the whole region filled with comparatively uniform grains distinguished by different colors as Fig5(c).



(a) *t*=200CAs (b) *t*=240CAs (c) *t*=300CAs

Figure 5. Growth of equiaxed grains during solidification

Fig. 6 shows the solute distribution in solidified grains at 300CAs. The results revealed that the solute concentration of the grain boundary is higher than that of the grain interior, closer to the center of grains the lower concentration is. During solidification process, the solidified cells will reject solute to the liquid phase due to solute redistribution which will lead to the solute enrichment at the grain boundary, namely grain boundary segregation. Where the multiple grain boundary segregation is more significant.



(a) (b)

Figure 6. The solute distribution at 300CAs during solidification: (a) the whole simulation region; (b) magnification of the local solute concentration region

3.2 Growth of columnar crystals of TC4-DT alloy EBW molten pools

The actual TC4-DT alloy EBW process was simulated by employed CA method mention above. Only the width of 4mm simulation region was selected from the cross section, as shown in Fig. 7, considering the narrow of EBW heat affected zone, usually less than 3mm. The simulation region was divided into 80 ×400 square cells with the size of 0.05mm.

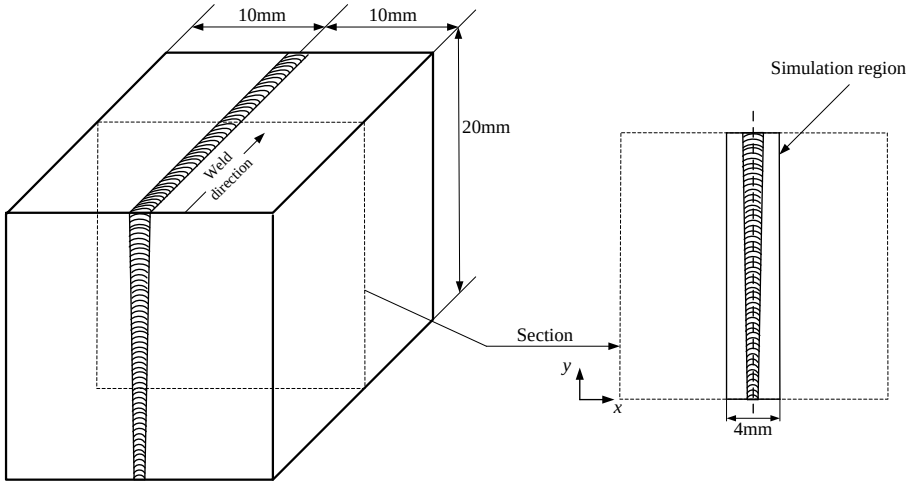


Figure 7. Schematic of TC4-DT EBW joints and selected simulation region

Fig. 8 shows the simulation results of grains nucleation and growth process in the weld pool at different CA time steps. Because of the symmetrical of the weld poor, only half of the model was considered. It can be found that at the beginning of the solidification, the nucleation firstly happened at the lower part of the weld pool near to the fusion line, due to the higher cooling rate and greater undercooling of these area, as shown in Fig. 8(a). As the solidification process, the nucleus at lower part grew gradually towards weld pool center for temperature gradient. In addition, the large number of new nucleus appeared along fusion line

from lower to upper, due to the temperature decrease, as shown in Fig. 8(b). At the time of 6000CAs, as Fig.8 (c), the early nucleus grew over and formed slender columnar grans, while the columnar at top area was just beginning to grow. It is mainly due to the great depth to width ratio of EBW pool and non-uniform temperature distribution from lower to upper of the cross section. Fig. 8(d) shows the final morphologies of grains with different colors. When the weld pool solidification completely, the columnar grains can be seen with irregular shape and size. General speaking, the size of the lower part among firstly solidified grains is evidently less than that the upper part among later solidified grains. What's more, due to the competitive growth among the neighborhood grains only several nucleus can grow up and form complete columnar crystal finally.

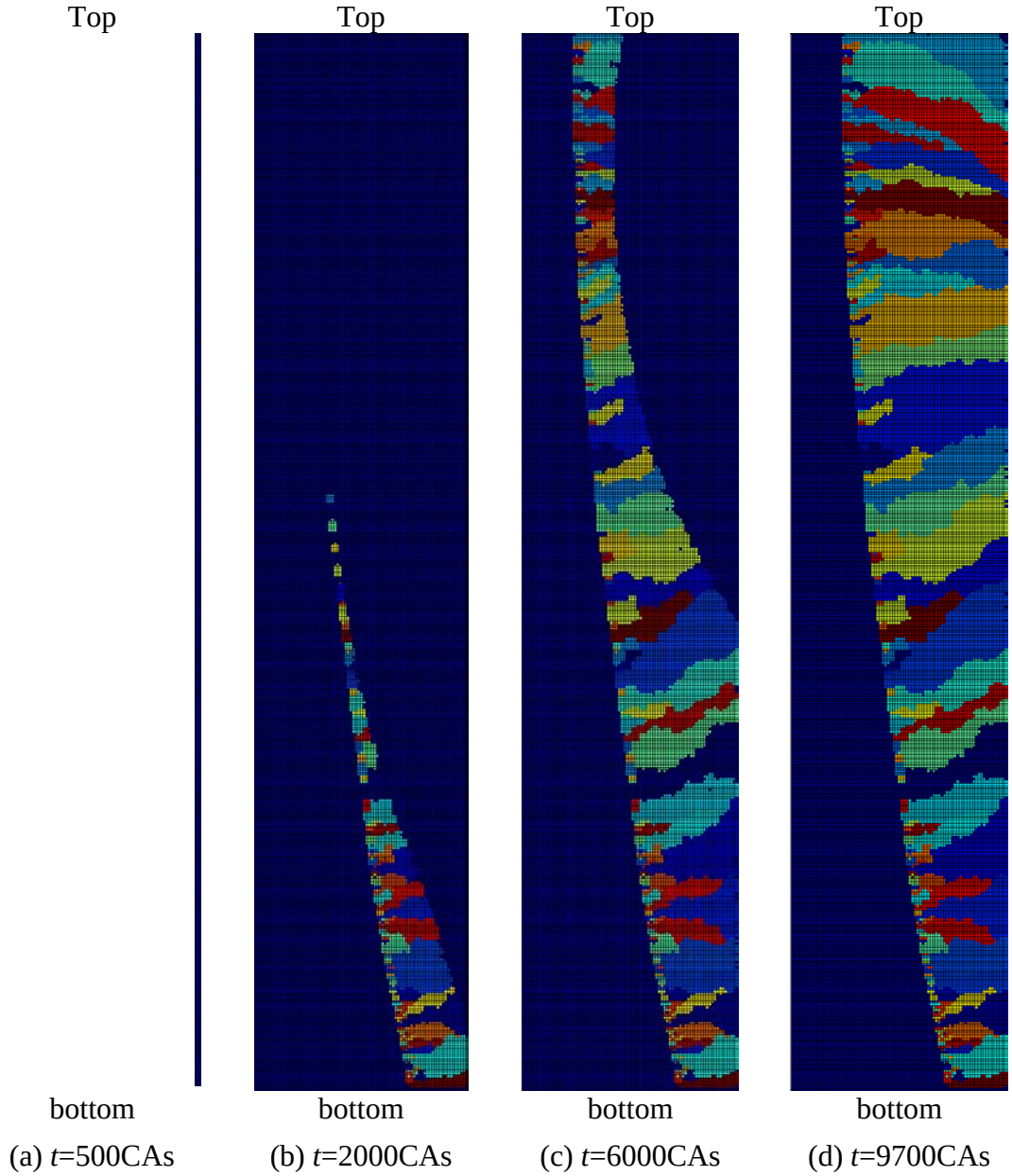
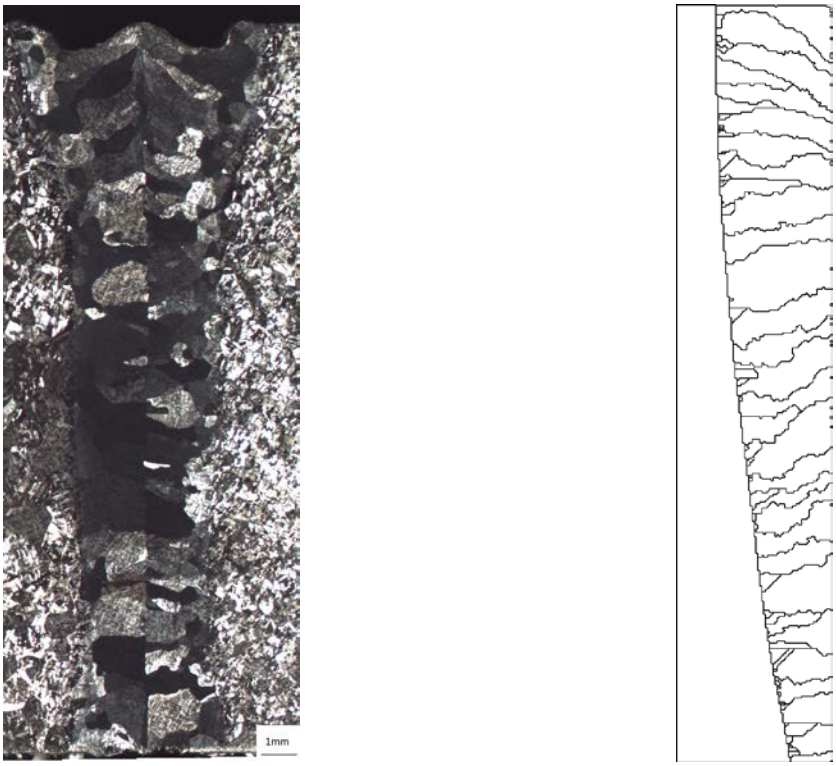


Figure 8. Simulation the growth process of columnar grains of TC4-DT alloy EBW weld pool (1/2 model)

Fig. 9 revealed the comparison of simulation results with experimental results. It can be observed that the morphology of simulated columnar crystal is very close to actual EBW results. The number and size of columnar crystals obtained from experiments and simulations

were measured respectively, as listed in the table 2. It is evidently that simulated result shows good agreement with experimental.



(a) (b)

Figure 9. The comparison of experimental and simulation: (a) The metallograph of TC4-DT EBW fusion zone; (b) The Topology morphology of simulation result

Table 2. The columnar crystal number and size of simulation and experimental

	Number of columnar	Maximum length /mm	Minimum length /mm	Average length /mm
Simulation results	27	1.475	0.202	0.712
Experimental results	28	1.550	0.250	0.714

4. conclusions

In this paper, a CA model was developed which can be used for the numerical simulation of grain growth during weld process. Considering the influence of the neighborhood on interface cell, the algorithm of solid fraction and solute redistribution was improvement. According to the state of neighborhood, the calculation of solid fraction was distinguished into three cases, i.e. the interface growth angle φ taken 0°, 45°, 63.4° respectively. In each case, the solid fraction increment is represented by different piecewise function. In addition, the calculation of solute redistribution was modified more reasonably by taking into account the distance and the concentration difference between the interface cell and its neighborhood.

The proposed model was applied to simulate the equiaxed grains growth under isothermal condition successfully. Finally, the developed model was employed to simulation the

solidification process of TC4-DT alloy EBW weld pool and predict the microstructure of the weld zone. The prediction of columnar crystal showed good agreement with experimental results on grain morphology and size.

Acknowledgement

Financial support by the Aeronautical Science Fund of China (No. 2015ZB52023) and the Fundamental Research Funds for the Central Universities (No. NS2016022) are gratefully acknowledged.

References

[1] P. F. Fu, Zh. Y. Mao, C. J. Zuo, Y. J. Wang and Ch. M. Wang (2014) Microstructures and fatigue properties of electron beam welds with beam oscillation for heavy section TC4-DT alloy, *Chinese Journal of Aeronautics* **27(4)**, 1015–1021.

[2] L. Tong, Z. S. Zhu, H. Q. Yu, Z. L. Wang (2010) Influence factors of microstructure and property of TC4-DT titanium alloy free forgings, *Chinese Journal of Nonferrous Material* **20**, 87–90.

[3] T. Zacharial, J. M. Vitek, J. A. Goldakt (1995) Modeling of fundamental phenomena in welds, *Modeling Simulation* 3(2), 265-288.

[4] D. J. Srolovitz (1983) Grain growth in two-dimensions, *Scr. Metall.* 17(2), 241-252.

[5] P. P. Zhu (1992) Dynamic simulation of crystal growth by Monte Carlo method, *Acta Metal* 40(12), 3369-3379.

[6] M. A. Zaeem, H. Yin, S. D. Felicelli (2012) Comparison of Cellular Automaton and phase field models to simulate dendrite growth in hexagonal crystals, *Journal of Materials Science & technology* 28(2), 137-146.

[7] A. Choudhury, K. Reuther (2012) Comparison of phase-field and cellular automaton models for dendritic solidification in Al–Cu alloy, *Computational Materials Science* 55, 263-268.

[8] G. J. Fix (1983) Phase field for free boundary problemsin: Fasono A, PrimicerioM, Pitman eds. Free Boundary Problems, *Theory and Applications*. 580~589

[9] R. Kobayashi (1993) Modeling and numerical simulations of dendritic crystal growth. *Physica D* 63(10), 410-423.

[10] X. Tong, C. Beckermann, A. Karma, et al (2001) Phase-field simulations of dendritic crystal growth in a forced flow, *Phys Rev E* 63(6), 1-16.

[11] S. Wolfram (1984) Computation theory of cellular automata, *Communications in Mathematical Physics*, **96(1)**, 15-57.

[12] J. D. Hunt (1984) Steady state columnar and equiaxed growth of dendrites and eutectic, *Materials Science & Engineering B*, 65(1), 75-83.

[13] Thevoz, J. L. Desbiolies, M. Rappaz (1989) Modeling of Equiaxed Microstructure Formation in Casting, *Metal. Trans. A* 20, 311-321.

[14] C. A. Gandin, M. Rappaz (1993) Probabilistic modeling of microstructure formation in solidification processes, *Acta Metallurgical Et Material* **41(2)**, 345-360.

[15] C. A. Gandin, M. Rappaz (1994) A coupled finite element-cellular automaton model for the prediction of dendritic grain structure in solidification processes, *Acta Metallurgical Et Material* **42(7)**, 2233-2246.

[16] C. A. Gandin, M. Rappaz (1997) A 3d cellular automaton algorithm for the prediction of dendritic grain growth, *Acta Material*, **45(5)**, 2187-2195.

[17] O. Zinovieva, A. Zinoviev, V. Ploshikhin, V. Romanova, R. Balokhonov (2015) A solution to the problem of the mesh anisotropy in cellular automata simulations of grain growth, *Computational Materials Science* 108, 168-176.

[18] R. Chen, Q. Y. Xu , B. C. Liu (2015) Cellular automaton simulation of three-dimensional dendrite growth in Al–7Si–Mg ternary aluminum alloys, *Computational Materials Science* 105, 90-100.

[19] P. Lacki, K. Adamus (2011) Numerical simulation of the electron beam welding process, *Computers and Structures* 89, 977-985.

[20] W. Kurz, B. Giovanola, R. Trivedi(1986) Theory of microstructure development during rapid solidification, *Acta Metallurgica* **34(5)**, 823-830.

Extending a 3D Parallel Particle-In-Cell Code For Heterogeneous Hardware

***Grischa Jacobs[1], Thomas Weiland[2] and Christian Bischof[3]**

[1]Graduate School of Computational Engineering, TU Darmstadt, Darmstadt 64293, Germany
[2]Computational Electromagnetics Laboratory (TEMF) , TU Darmstadt, Darmstadt 64293, Germany
[3]Scientific Computing, TU Darmstadt, Darmstadt 64293, Germany

*Presenting author: jacobs@gsc.tu-darmstadt.de

Abstract

An evaluation for a parallel Particle-In-Cell code leveraging heterogeneous hardware is presented. Hybrid parallelization is implemented to support optional workload offloading to 40 Intel® Xeon Phi™ coprocessors. A performance model is applied to load balance the particle data for this heterogeneous setup. Performance measurements of a benchmark show the speedups for the balanced and unbalanced cases and the execution without the coprocessor. The code computes particle-field interactions in the time domain, typically used in plasma or particle physics. A multi beam gun is chosen as a benchmark. The gun uses an electrostatic field to accelerate the particles and a magnetostatic field, generated by a current driven coil to focus the particle beam. Calculated results are compared with CST Particle Studio [11]. For solving the electrodynamic and electrostatic fields, described by the coupled MAXWELL equations, a 3D solver has been implemented, facilitating the Finite Integration Technique (FIT) [1].
Keywords: High Performance Computing, Intel Xeon Phi, Particle-In-Cell

Introduction

Modern HPC systems provide diverse processor architectures, making efficient parallel computing a difficult task. Keeping the physical limitations with high clock speed rates and energy consumptions of processors in mind, the attractiveness of modern multicore processors becomes obvious. To leverage their benefits, hybrid parallelization strategies become necessary. As the variety of heterogeneous computing systems will increase in the future, this motivates investigations for realistic performance and scalability models to explore potentials for code optimizations and load balancing strategies. Typically used in computational accelerator and plasma physics, Particle-In-Cell (PIC) simulations calculate the movement of free charges in electromagnetic fields. Solving those physics requires a solution of the coupled MAXWELL equations

$$\begin{aligned} \nabla\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t}, & \nabla\cdot\vec{B}=0,\\ \nabla\times\vec{H} &= \frac{\partial\vec{D}}{\partial t}+\vec{J}, & \nabla\cdot\vec{D}=\rho, \end{aligned} \tag{1}$$

and the relativistic NEWTON-LORENTZ equation

$$\begin{aligned} \frac{\partial\vec{u}}{\partial t} &= \frac{q}{m_0 c}\left(\vec{E}+\vec{v}\times\vec{B}\right), & \frac{\partial\vec{r}}{\partial t}=\vec{v},\\ \vec{u} &= \gamma\frac{\vec{v}}{c}, \end{aligned} \tag{2}$$

where $\vec{u}$ is the normalized momentum and $q, m_0, \vec{r}, \vec{v}$ represent charge, rest mass, position and velocity of particles. As equations 1 and 2 lead to separate computations within this approach those are referred as computational kernels.

3D Particle-In-Cell Simulation

As moving charges describe a current in eq. (1), a cyclic dependency needs to be solved for every time step. This is shown in figure 1. To solve the fields numerically, the Finite Integration Technique (FIT) [1] is implemented. For further information about FIT the reader is referred to [1] for the general theory and to [4] for a setting with PIC. For the time integration of the fields a leap-frog scheme is chosen. For the integration of eq. (2) the well known Boris scheme is used. Charge conservation is ensured by using an algorithm described in [5]. The following subchapters will provide a coarse overview. The basic kernels of the PIC method are: (1) Calculating the dynamic electromagnetic fields in time domain with eq. 1 (”field” kernel). (2) Gather all static and dynamic field values at particle positions (”gather fields” kernel). (3) Integrate particle trajectories for one time step (”push” kernel). (4) Calculate currents introduced by the charge movement and scatter those (”scatter current” kernel). The costs of these kernels depend on the problem setting in terms of particles per cell, particle distribution and the sizes of the computational mesh.

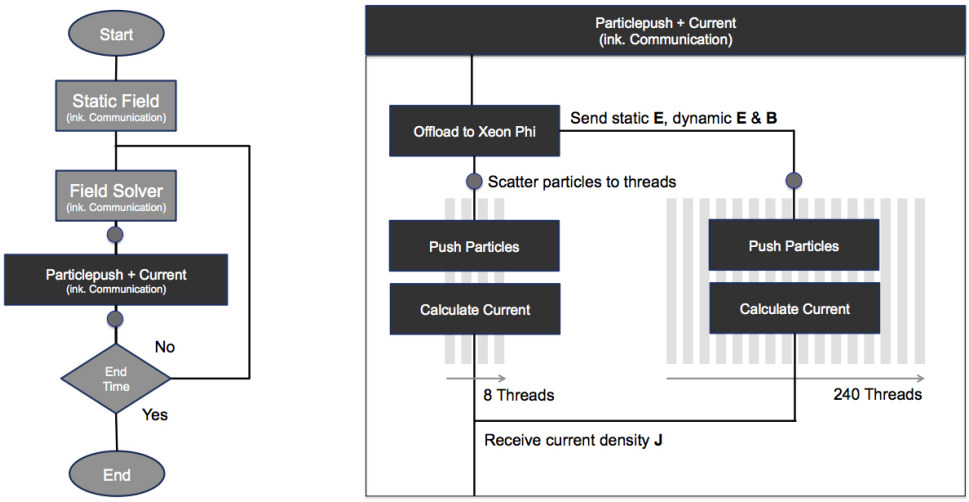


Figure 1: The left figure shows the sequential steps of the Particle-In-Cell algorithm. Each loop calculates one physical time step. The right figure explains the offloading to the Intel® Xeon Phi™ accelerator card. Only the particle trajectory and current calculations are offloaded to the card.

Field Solver

The field solver facilitates FIT. The FI discretization scheme is related to the well known Yee scheme and based on a dual grid-doublet $\{G, \widetilde{G}\}$, which decomposes the computation domain into two sets of dual cells. Integral quantities $\widehat{\mathbf{q}}$, $\widehat{\mathbf{e}}$ and $\widehat{\widehat{\mathbf{b}}}$ are defined on the grid G, corresponding to the total charge in the cell volumes, to the electric voltage along the cell edges and to the magnetic induction flux on the cell facets, respectively. Electric voltage $\widehat{\mathbf{e}}$ is defined by

$$\int_{L_v(i,j,k)} \vec{\mathbf{E}}(\vec{\mathbf{r}}, t) \cdot \vec{\mathbf{e}}_v \, dv \;=\; \widehat{\mathbf{e}}_v(i,j,k). \qquad v \in \{x, y, z\} \tag{3}$$

The integral quantities $\hat{\tilde{\mathbf{j}}}$, $\hat{\tilde{\mathbf{d}}}$ and $\hat{\mathbf{h}}$ are the vectors of charge current, electric displacement flux and magnetic voltage defined on the facets and edges of the dual grid $\widetilde{G}$. Fig. 3 illustrates the allocation of the electric voltage in the case of rectangular dual grids G and $\widetilde{G}$. Using these integral quantities, Maxwell's equations in discrete form, the so-called Maxwell-Grid-Equations are obtained:

$$\mathbf{C}\,\hat{\mathbf{e}} \;=\; -\frac{d}{dt}\hat{\tilde{\mathbf{b}}}, \tag{4a}$$

$$\widetilde{\mathbf{C}}\,\hat{\mathbf{h}} \;=\; \hat{\tilde{\mathbf{j}}} \;+\; \frac{d}{dt}\hat{\tilde{\mathbf{b}}}, \tag{4b}$$

$$\widetilde{\mathbf{S}}\,\hat{\tilde{\mathbf{d}}} \;=\; \mathbf{q}. \tag{4c}$$

The support matrix operators $\{C, S\}$ and $\{\widetilde{C}, \widetilde{S}\}$ defined on G and $\widetilde{G}$ are discrete mappings of the differential "curl" and "div". The operators $C, S, \widetilde{C}$ and S fulfill the identities $SC = C\widetilde{S}^T = 0$ and $\widetilde{S}\widetilde{C} = \widetilde{C}S^T$. This corresponds to the continuum relations $div\ curl = 0$ and $curl\ grad = 0$. The discretization approximation enters FIT through the constitutive material equations

$$\hat{\tilde{\mathbf{d}}} \;=\; \mathbf{M}_{\epsilon}\,\hat{\mathbf{e}} \quad , \quad \hat{\mathbf{h}} \;=\; \mathbf{M}_{\mu^{-1}}\,\hat{\tilde{\mathbf{b}}} \quad \text{and} \quad \hat{\tilde{\mathbf{j}}} \;=\; \mathbf{M}_{\sigma}\,\hat{\mathbf{e}}. \tag{5}$$

Particle Solver

The particle solver models groups of particles "macroparticles" using a ballistic approach by solving eq. (2). Solving it requires a three step process: 1. interpolating $\mathbf{E}$ and $\mathbf{B}$ fields in the center of the macroparticle within one cell by choosing an interpolation with at least order one. 2. Solving eq. (2) with the a method proposed by Boris [5] using the interpolated field values of step (1). Note that solving this equation is not trivial as of the term $\vec{v} \times \vec{B}$. 3. Calculating current densities induces by the particle movement, with the equations 6 for the 2D case as shown in figure 2. Bunemann et. al. [2] describe how to solve this with rigorous charge conservation.

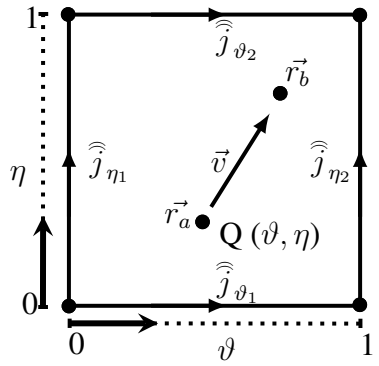


Figure 2: Macroparticle moving in a 2D cell. As the particle is moving it creates currents on the edges of the cell.

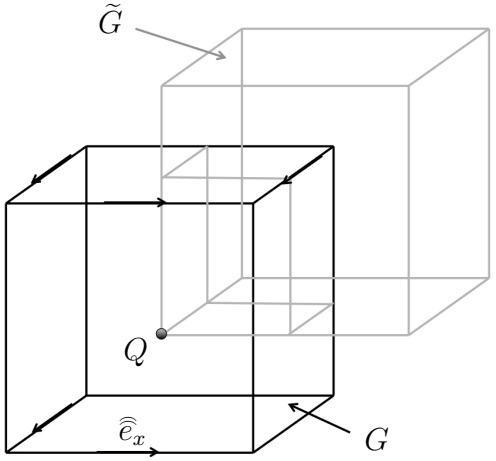


Figure 3: Illustrates the allocation of fluxes and voltages in the case of rectangular dual grids G and $\widetilde{G}$.

The currents in figure 2 for the 2D case are calculated by:

$$\hat{\tilde{j}}_{\vartheta_1} = Q \cdot \frac{\vartheta_2 - \vartheta_1}{\triangle t} \cdot \left(1 - \frac{\eta_1 + \eta_2}{2}\right) \qquad \hat{\tilde{j}}_{\vartheta_2} = Q \cdot \frac{\vartheta_2 - \vartheta_1}{\triangle t} \cdot \frac{\eta_1 + \eta_2}{2} \tag{6a}$$

$$\hat{\tilde{j}}_{\eta_1} = Q \cdot \frac{\eta_2 - \eta_1}{\triangle t} \cdot \left(1 - \frac{\vartheta_1 + \vartheta_2}{2}\right) \qquad \hat{\tilde{j}}_{\eta_2} = Q \cdot \frac{\eta_2 - \eta_1}{\triangle t} \cdot \frac{\vartheta_1 + \vartheta_2}{2} \tag{6b}$$

Integration in the Time-Domain

The time domain equivalent to the FI Method is the well known FDTD scheme of leapfrog integration. Applied to the time dependent equations (4) this procedure is restricted by the Courant–Friedrichs–Lewy stability criterion on the time step length

$$\triangle t \leq \triangle t_{maxCFL}. \tag{7}$$

For the time integration an explicit forward time difference scheme is used. The corresponding update relation is

$$\hat{\mathbf{h}}^{(m+1)} = \hat{\mathbf{h}}^{(m)} - \triangle t\, \mathbf{M}_{\mu^{-1}}\, \tilde{\mathbf{C}}\, \hat{\mathbf{e}}^{(m+\frac{1}{2})} + O(\triangle t^2)\,, \tag{8a}$$

$$\hat{\mathbf{e}}^{(m+\frac{3}{2})} = \hat{\mathbf{e}}^{(m+\frac{1}{2})} + \triangle t\, \mathbf{M}_{\epsilon}{}^{-1} \left(\tilde{\mathbf{C}}\, \hat{\mathbf{h}}^{(m+1)} - \hat{\tilde{\mathbf{j}}}^{(m+1)}\right) + O(\triangle t^2). \tag{8b}$$

Integrating the particle trajectories is straight forward. Replacing the differential operator in eq. (2) with a central differential quotient will lead to a numeric representation. Again a leapfrog scheme for the integration of $\vec{r}$ and $\vec{u}$ is used.

$$\mathbf{u}^{n+1/2} = \mathbf{u}^{n-1/2} + \triangle t \cdot \frac{Q}{m_0 \cdot c} \cdot (\mathbf{E}^n + \mathbf{v}^n \times \mathbf{B}^n)\,, \tag{9a}$$

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \triangle t \cdot \frac{c}{\gamma^{n+1/2}} \cdot \mathbf{u}^{n+1/2} \tag{9b}$$

Eq. (9a) cannot be solved explicit as of $\vec{v}$. Changes in the velocity $\vec{v}$ will also effect the normalized momentum $\vec{u}$. For this reason eq. (9a) is split into three steps as suggested by Boris [5]. First the momentum gets calculated over half a time step by

$$\mathbf{u}_- = \mathbf{u}^n + \frac{\triangle t}{2} \cdot \frac{Q}{m_0 \cdot c} \cdot \mathbf{E}^n, \tag{10}$$

second a rotation is calculated according to the LORENTZ force over a full time step

$$\mathbf{u}^* = \mathbf{u}_- + \mathbf{u}_- \times \mathbf{T}, \tag{11a}$$

$$\mathbf{u}_+ = \mathbf{u}_- + \mathbf{u}^* \times \frac{2 \cdot \mathbf{T}}{1 + |\mathbf{T}|^2}, \tag{11b}$$

$$\mathbf{T} = \triangle t \cdot \frac{Q \cdot \mathbf{B}^n}{2 \cdot m_0 \cdot \sqrt{1 + |\mathbf{u}_-|^2}}. \tag{11c}$$

Finally the momentum gets calculated over the lasting half time step

$$\mathbf{u}^{n+1} = \mathbf{u}_+ + \frac{\triangle t}{2} \cdot \frac{Q}{m_0 \cdot c} \cdot \mathbf{E}^n. \tag{12}$$

Parallel Particle-In-Cell

In this work Intel® Xeon Phi™ coprocessors are evaluated for the parallelization of the PIC method. The existing parallel PIC code facilitates distributed and shared memory parallelization using MPI and OpenMP. The code is build on top oh the PETSC framework [10] supporting sophisticated MPI data structures to be used. The Intel® Xeon Phi™ card has been chosen instead of a GPU card, as the the existing OpenMP code may be easy to offload.

Strategies for Parallelization

To minimize the overall runtime, a suitable parallelization strategy needs to be chosen. Such a strategy may be influenced by application specific properties, e.g. different particle distributions or geometry resolutions and by hardware specific properties such as vectorization in CPU's, multicore systems and coprocessors. Two strategies for distributed memory PIC parallelization have been investigated in the context of accelerator physics (e.g. beam simulation) [6], [7], [8] and [9]:
1.) The whole computational domain is decomposed by the number of computing nodes available. Every node calculates the DOF's for the fields and the trajectories for the particles, that are moving within the domain assigned to the node. Hence only uniform particle distributions, where every node calculates an equal number of particles, benefit from this strategy.
2.) Only the field DOF's are spatially distributed to the nodes, whereas the particle calculations are equally distributed independent from their position. This guarantees an equal workload for every node, with the drawback of additional communication costs. This strategy is characterized by a satisfying weak scaling behaviour, but may not be the fastest solution.

Shared Memory Parallelization / Offloading to Coprocessors

Using one Intel® Xeon Phi™ coprocessor with 60 effective cores, each with four hardware threads, the "field" kernel can make use of the (theoretical) high memory bandwidth and the "push" and "current" kernels can leverage the highly concurrent SIMD nature of the particle calculations using up to 240 hardware threads available on the card. The coprocessor can be used in two different modes. A "native mode" where the executable gets compiled to run on the coprocessor as a standalone MPI process and a "offload mode" that enables offloading selected kernels that benefit from the multicore architecture. Due to the memory limitation of 8 GB main memory for the smaller Intel® Xeon Phi™ 5110P card and the fact that not every computational kernel can benefit from the shared memory scalability of the coprocessor (e.g field solver), the "native mode" is not evaluated in this work. In the benchmark used for evaluation, the computations of the particle solver takes up to 80% of the overall time to compute one physical time step, suggesting that particle integrations and current density calculations are offloaded to the Xeon Phi™ coprocessor, whereas the field computations and all MPI communications are exclusively performed on the host. As the communication to the coprocessor over PCIe is one bottleneck, this work evaluates only the second parallelization strategy mentioned above making benefit of the fact that particle data will stay on the coprocessor across all time step calculations, thus PCIe traffic is reduced. In order to calculate on the coprocessor in parallel with the host, an asynchronous offload with OpenMP 4.0 LEO is implemented. This way only one thread of the host executes the offloading procedure to the coprocessor, whereas the remaining $n-1$ threads of the host are facilitated to compute the particle movement and current calculations.

Performance

In this work performance is defined as "time-to-solution". From a performance bottleneck perspective computational kernels can be classified as memory bounded and CPU bounded.

Evaluating the code showed, that both the "field" kernel and the "push" kernel tend to be to memory (bandwidth) bounded.

Performance Modeling

In some cases it might be inefficient to offload computations to a coprocessor, as the time for sending and receiving data from the coprocessor makes the speedup for calculation neglectable. Therefore a performance model, on the basis of a model proposed by Kredel et.al. [3], is introduced predicting performance achieved by the PIC code. Further performance prediction creates space for robust load balance strategies. By counting all floating point operations $\#op_j$ as well as the number of network bytes exchanged $\#x_j$ by kernel j, taking the communication bandwidth b^k into account and measuring floating point operation per second l_j^k for each node k the performance is estimated by

$$t_k \leq \sum_{j=0}^{Kernel} \frac{\#op_j}{l_j^k} + \sum_{j=0}^{Kernel} \frac{\#x_j}{b^k}. \tag{13}$$

The parameters $\#op_j$ and $\#x_j$ describe the software performance where as b^k and l_j^k are hardware representatives. Software parameters can be derived from the code by hand or with measurements by sweeping all parameters (e.g. mesh and particle size). As it is intended to model a system with one host and one Intel® Xeon Phi™ coprocessor the effective bandwidth for the communication between those needs to be measured as shown in figure 7. For large data sizes ($>$ 30MB) the bandwidth for sending and receiving may differ by up to one dimension. As the offloaded kernels are running in parallel on the host and on the coprocessor, the performance is estimated by

$$t_k \leq \max \left[\frac{w_1 op_1}{l_1^1}, \frac{w_2 * op_1}{l_1^2} + \frac{w_2 * x_{send}}{b_{send}^2} + \frac{w_2 * x_{recv}}{b_{recv}^2} \right] \tag{14}$$

where the *max* operator describes the parallel execution, as the slower system will degrade the performance. The sum of the performance of the "push" and "current" kernels, executed by the host, is denoted by $l_{1,2}^1$ and the operation count by $op_{1,2}$. For the coprocessor those are denoted by $l_{1,2}^2$ and $op_{1,2}$, respectively. As the bandwidth for sending and receiving data from the coprocessor differs, two terms modeling the data transfer are added. As it is intended to perform load balancing, two scalar weights are added, w_1 and w_2. The sum of both weights must be equal to one.

Optimization for Accelerator

In order to get the code running efficiently on the coprocessor some optimizations are carried out. When calculating current density values by the "current scatter" kernel after the movement of the particles, those values are stored in a hash map where the key is the face index in the computational mesh, of the face the particle has crossed. This is done by every thread for all particles those threads are responsible for and merged in the end. Reading and writing to 240 hash maps on the coprocessor, each controlled by one thread, is decreasing the performance in an order of one dimension compared to the performance of the host with 16 threads. Using the concurrent unordered map provided by Intel Thread Building Blocks library [13] the performance is improved to be competitive with the host's performance. The data of particles and fields transmitted from the host to the coprocessor and current data transmitted from the coprocessor to the host is communicated over PCIe 2, having a peak bandwidth of 6 GB/s as shown in figure 7. To achieve high communication speeds the environment variable MIC_USE_2M_BUFFERS of the coprocessor is set to 2 MB.

Architectural Testbed

The Lichtenberg cluster [12] located at Technische Universität Darmstadt, has 647 computing nodes available for various applications and provides an accelerator section that with 24 nodes, configured to be used with 48 Intel® Xeon Phi™ coprocessors (two cards each node). Every host node has two sockets with one Intel® Xeon® Processor E5-2670 having 8 cores, hyperthreading disabled and 32 GByte main memory. Each core runs on 2.6 GHz. Nodes in this section are connected with 1x FDR-10 InfiniBand. Two nodes provide Intel® Xeon Phi™ 7120P coprocessors whereas the remaining 22 nodes provide Intel® Xeon Phi™ 5110P coprocessors. The Intel® Xeon Phi™ coprocessor 5110P has 8GB main memory, 59 effective cores each with four hardware threads, 1.05 GHz clock speed and a theoretical peak memory bandwidth of 320 GB/s. This system is used to evaluate the speedup achieved by PIC code when incorporating Intel® Xeon Phi™ coprocessors.

Results

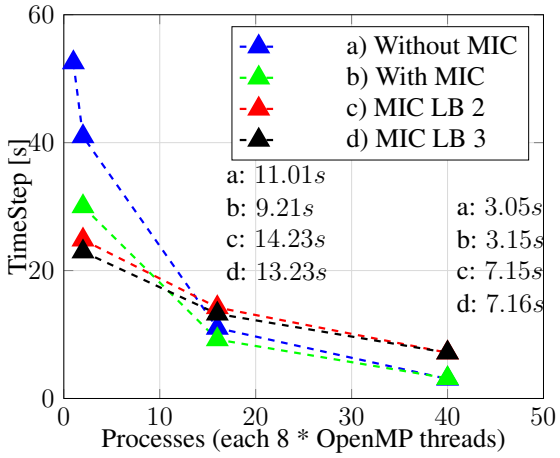


Figure 4: Execution times to compute one physical time step. The overall problem size is fixed (10^6 DOF and 10^7 particle), whereas the number of MPI processes is increased. Each computing node executes two MPI processes, as each MPI process has one Intel® Xeon Phi™ card. The blue line shows measured times without the support of coprocessors. Green line shows measured times with support of the coprocessors. Red and black line show execution times with particle load balancing, using the ratios of 1:2 and 1:3 between the host and the coprocessor.

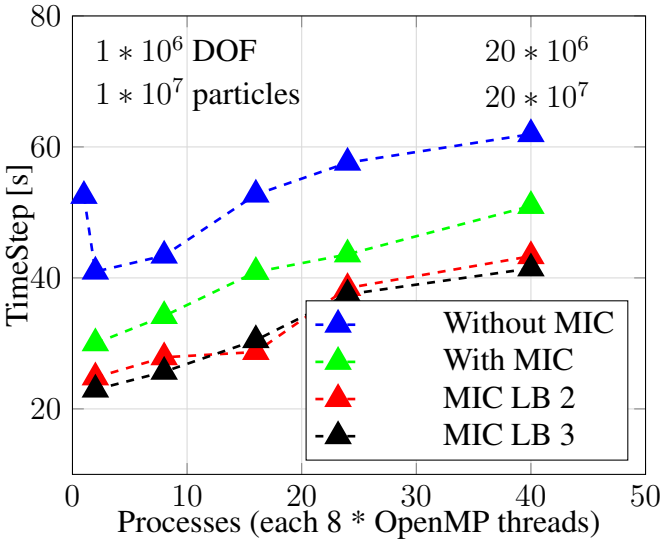


Figure 5: Execution times to compute one time step. The problem size per MPI process is constant (10^6 DOF and 10^7 particle), whereas the number of MPI processes is increased. Each computing node executes two MPI processes. The blue line shows measured times without the support of coprocessors. Green line shows measured times with support of the coprocessors. Red and black line show execution times with particle load balancing, using the ratios of 1:2 and 1:3 between the host and the coprocessor.

As a benchmark problem, a multi beam particle source of a particle accelerator is chosen. The benchmark is provided by CST [11]. It simulates multiple electron beams with free movement in a constant electric and magnetic field and perfect electric conducting boundaries. The magnetic field is generated by a current driven coil to focus the beam. The problem size is designed

to meet the main memory limitations of one computing node in the accelerator supported section of the Lichtenberg cluster, calculating 10 million DOF for the mesh and 100 million particles on one node. Scaling up to 20 nodes a problem with 200 million DOF and 2 billion particles is solved.

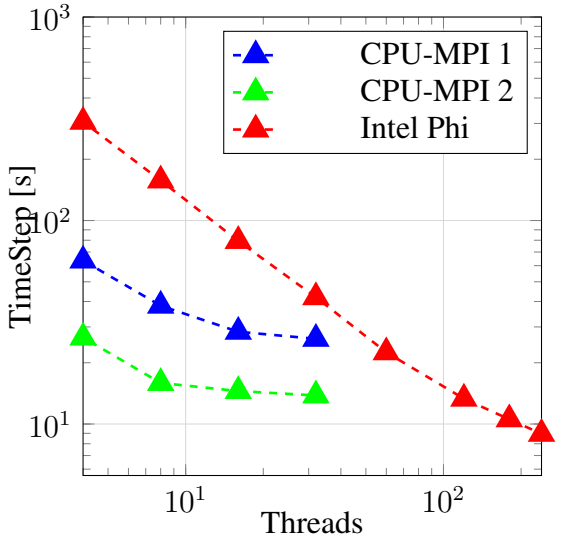


Figure 6: Execution time for "push" and "current" kernels calculating $5 * 10^6$ particles with a varying number of threads. Red line shows the execution times of the offloaded kernels on Intel® Xeon Phi™ 5110P. A speedup of 2-3 against the hosts execution is measured. Blue line shows the execution on the host with one MPI process and varying threads. Green line shows the execution on the host with two MPI processes and varying threads.

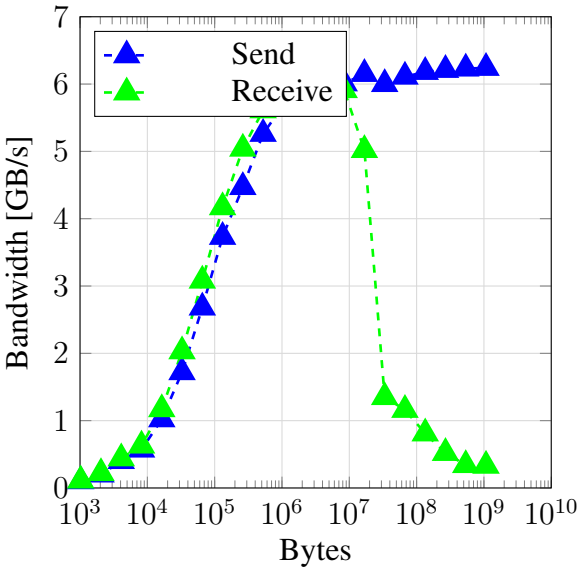


Figure 7: Showing measured bandwidth values when communicating with an Intel® Xeon Phi™ 5110P coprocessor. The blue line shows performance for received data, whereas the green line shows the bandwidth measured when data was sent. Both measurements can be fittet with spline functions.

Speedup and Execution Time

Three studies are carried out to evaluate the speedup achieved with the coprocessors. A study evaluating strong scalability shown in figure 4, where the problem size is constant and the number of parallel units is increased, a weak scalability study, where the problem size increases linearly with the number of parallel units, shown in figure 5 and a study measuring the shared memory scalability shown in figure 6. All measurements shown are executed with two MPI processes per node, each running 8 threads in parallel. This way each MPI process makes use of one Intel® Xeon Phi™ coprocessor, as one node holds two cards. Also the performance for one MPI process running on one node with 16 threads is measured, but only for the setup where no coprocessors were used. Each figure shows time values in seconds measured when executing one physical time step. This physical time step is calculated by the sequential execution of each computational kernel in parallel by all MPI processes. The values are mean values of ten physical time steps measured. Four setups are configured and shown in both figures 4 and 5. The blue line ("w/o MIC") shows measurements were the code is executed without a coprocessor, whereas the green line ("w MIC") incorporates the coprocessors. The red and black

lines are setups, where the number of particles calculated by the host and the ones calculated by the coprocessor are load balanced by the ratio 1:2 and 1:3, respectively. From figure 4 one can infer that using two coprocessors reduces the time to compute one physical time step by up to 56% compared to a host only execution with one MPI process, and a reduction of 23% is achieved compared to a host only execution with two MPI processes. One can also derive that the support of the coprocessor gets insignificant as the number of particles per MPI process gets to small, as measured with 16 and 40 MPI processes in figure 4. Having a constant problem size on each node, as shown in figure 5, the benefit of the coprocessors becomes noticeable, as a mean runtime reduction of 39% for the load balanced setup is measured. Figure 6 shows the scalability of the ”push” and ”current” kernels, when increasing the number of threads, having the problem size kept fixed. The blue line plots a setup where the code is executed without the coprocessor using one MPI process and a varying number of threads, whereas the green line shows measurements for a setup with two MPI processes executed in parallel on one node. The red line plots the time the coprocessor (Intel® Xeon Phi™ 5110P) needs to execute both kernels. Calculating the same number of particles on both accelerator cards, the performance difference between the smaller and the bigger accelerator card is negligible. In figure 6 it is also shown that calculating the same number of particles with the same number of threads on the host and on the coprocessor, the host system exceeds the coprocessors. This may be caused by the different core cache architectures, bigger caches sizes and the existence of L3 caches on the host. As the coprocessor can scale up to 240 threads (the Intel® Xeon Phi™ 5110P only to 236 threads) a speedup between 2 and 3 can be measured compared to the host with 16 threads running. The speedup of the host saturates at 8 threads.

According to Intel, the pinning of threads onto the cores can have a major performance impact. Table 1 shows measurements for various thread affinity setups. Scatter, balanced and compact affinity settings are evaluated when 59, 118, 177 and 236 threads are used. Balanced and scattered thread distributions lead to similar execution times, whereas compact distribution tends to be slower. Using 236 threads all affinity settings show similar results.

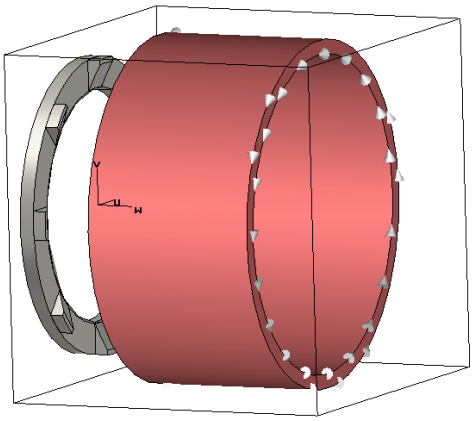


Figure 8: CAD model of the multibeam particle source created in CST Particle Studio [11]. The red structure is a current driven coil generating a magnetostatic field. The ring structure has 8 particle sources.

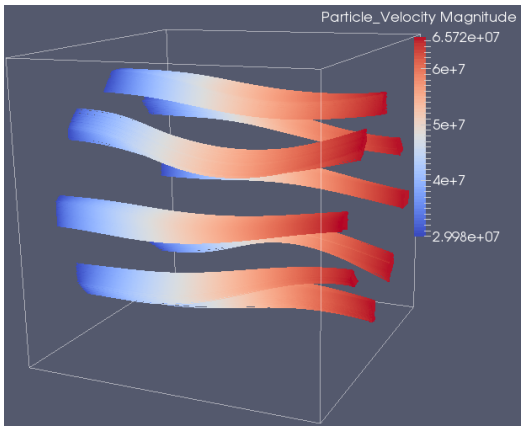


Figure 9: Particle-In-Cell multi beam benchmark simulating electron beams with free movement in a constant magnetic field and perfect electric conducting boundaries. The benchmark is calculating 10 million DOF for the mesh and 100 million particles on one node.

Table 1: Thread affinity on xeon Phi 5110P

KMPAFFINITY	59	118	177	236
granularity=fine, scatter	22.74s	12.75s	9.85s	8.29 s
granularity=fine, balanced	22.82s	12.81s	9.89s	8.16s
granularity=fine, compact	29.40s	15.06s	10.39s	8.16s

Conclusion

Performance measurements are presented evaluating "time-to-solution" with 40 Intel® Xeon Phi™coprocessors incorporated executing a parallel Particle-In-Cell code. It is shown that the on-node performance is improved by 56% for realistic problem sizes, when controlling the balance of data that is computed on the host and on the coprocessor with a load balancing strategy. Therefore an analytical performance model is used and evaluated for the host and the coprocessor.

Acknowledgment

The work of Grischa Jacobs is supported by the 'Excellence Initiative' of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt.

References

[1] Thomas Weiland, A discretization method for the solution of Maxwell's equations for six-component fields, *Electronics and Communication*, Vol.31, p116-121, 1977.

[2] John Villasenor and Oscar Buneman, Rigorous charge conservation for local electromagnetic field solvers, *Computer Physics Communications*, 69:306-316, 1992.

[3] Heinz Kredel, Sabine Richling, Jan Philipp Kruse, Erich Strohmaier, Hans-Günther Kruse, A simple concept for the performance analysis of cluster-computing, *Supercomputing*, 165-180, 2013.

[4] U. Becker, T. Weiland, Particle-in-Cell simulations within the FI-Method, *Surveys on Mathematics in Industry*, Vol.8, No.3-4, pp.233-242, 1999.

[5] Boris, J.P., Relativistic plasma simulation-optimization of a hybrid code, Proceeding of Fourth Conference on Numerical Simulations of Plasmas, November 1970

[6] F. Wolfheimer, E. Gjonaj, T. Weiland: A parallel 3D Particle-In-Cell (PIC) with dynamic load balancing. *Nuclear Instruments and Methods in Physics Research (NIM)*, Vol. 558, pp. 202-204, 2006

[7] E. A. Carmona, L. J. Chandler, On parallel PIC versatility and the structure of parallel PIC approaches, *Concurrency: Practice and Experience*, Vol.9(12), pp.1377-1405, 1997.

[8] Ji Qiang, Xiaoye Li, Particle-field decomposition and domain decomposition in parallel particle-in-cell beam dynamics simulation, *Computer Physics Communications*, Vol. 181, Issue 12, 2010.

[9] A. C. Elster, Parallelization issues and particle-in-cell codes, PhD Book, 1994.

[10] S. Balay et. al., Efficient Management of Parallelism in Object Oriented Numerical Software Libraries, *Modern Software Tools in Scientific Computing*, Pages 163-202, 1997.

[11] "CST Software", https://www.cst.com/

[12] ”Lichtenberg cluster”, http://www.hhlr.tu-darmstadt.de/hhlr/index.en.jsp

[13] ”Intel Thread Building Block”, https://www.threadingbuildingblocks.org/

Discrete Particle Methods for Simulating High-Velocity Impact Phenomena

***M.O. Steinhauser**[1,2]

[1]Fraunhofer Institute for High-Speed Dynamics, Ernst-Mach-Institute, EMI, Freiburg, Germany
[2]Department of Chemistry, University of Basel, Switzerland
*Presenting author: martin.steinhauser@emi.fraunhofer.de

Abstract

In this paper we introduce a mesh–free computational model for the simulation of high-speed impact phenomena. Within the framework of particle dynamics simulations we model a macroscopic solid ceramic tile as a network of overlapping discrete particles of microscopic size. Using potentials of the Lennard-Jones type we integrate the classical Newtonian equations of motion and perform uni-axial, quasi-static load simulations to customize our three model parameters to the typical tensile strength, Young's modulus and the compressive strength of a ceramic. Subsequently we perform shock load simulations in a standard experimental set-up, the edge-on impact (EOI) configuration. Our obtained results concerning crack initiation and propagation through the material agree well with corresponding high-speed EOI experiments with Aluminum Oxinitride (AlON), Aluminum Oxide (Al_2O_3) and Silicon Carbide (SiC), performed at the Fraunhofer Ernst-Mach-Institute (EMI). Additionally, we present initial simulation results where we use our particle–based model to simulate a second type of high-speed impact experiments where an accelerated sphere strikes a thin aluminum plate. Such experiments are done at our institute to investigate the debris clouds arising from such impacts, which constitute a miniature model version of a generic satellite structure that is hit by debris in the earth's orbit. Our findings are that a discrete particle based method leads to very stable, energy–conserving simulations of high–speed impact scenarios. Our chosen interaction model seems to work particularly well in the velocity range where the local stresses caused by impact shock waves markedly exceed the ultimate material strength.

Keywords: Computer Simulation, Discrete particle model, Multiscale modeling, High-speed impact, Molecular Dynamics, Hypervelocity.

Introduction

Understanding the mechanisms of failure in materials on various length- and time scales is a prerequisite for the design of new materials with desired superior properties such as high toughness or strength. On the macroscopic scale, many materials such as concrete or ceramics may be viewed as being homogeneous; however, on the scale of a few microns these materials often exhibit an inhomogeneous polyhedral granular structure which is known to influence its macroscopic mechanical and/or optical properties [1]. Whether a material under load displays a ductile, metal-like behavior or ultimately breaks irreversibly, in essence depends on the atomic crystal structure and on the propagation of defects in the material. Broken atomic bonds (cracks) and dislocations are the two major defects determining mechanical properties on an atomic scale. Due to the ever increasing computing power of modern hardware, many-particle molecular dynamics (MD) simulations taking into account the degrees of freedom of several billion atoms are nowadays feasible [2][3]. Molecular dynamics investigations of this type using generic models of the solid state have lead to a basic understanding of the processes that govern failure and crack behavior, such as the dynamical instability of crack tips [4][5], the limiting speed of crack propagation[6]-[8], the dynamics of dislocations [9][10], or the universal features of energy dissipation in fracture [11]. However, investigations of materials which

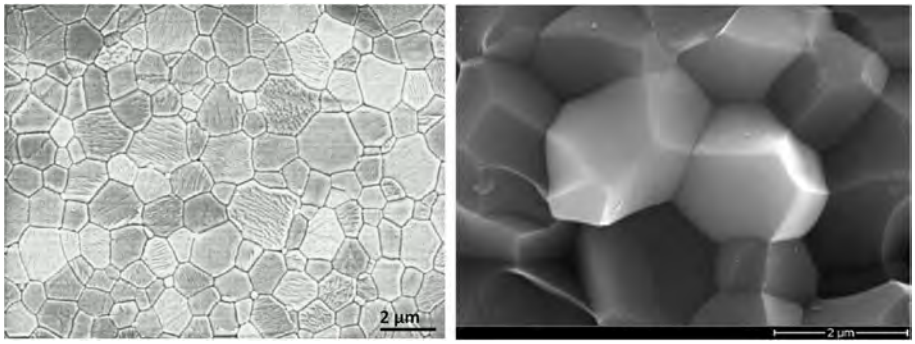


Figure 1: Left: Photomicrograph of an Al_2O_3 ceramic tile. Right: 3D view of the polyhedral granular surface structure of Al_2O_3.

involve multiple structure levels, such as polycrystalline solids, require very large ensembles of atoms to accurately reflect the structures on the microscopic and mesoscopic levels [12]. For systems of reasonable size, atomistic simulations are still limited to following the dynamics of the considered systems only on time scales of nanoseconds. Such scales are much shorter than what is needed to follow many dynamic phenomena that are of experimental interest [13][14]. On the microscale, the typical structure of many brittle materials is composed of convex polyhedra, as seen in two-dimensional (2D) photomicrographs of polycrystalline ceramics (Fig. 1).

High-Performance Ceramics (HPC)s

With ceramics, the specific shape and size of their polycrystalline grain structures is formed in a sintering process where atomic diffusion plays a dominant role. Usually the sintering process results in a porous microstructure with grain sizes of several hundred micrometers. Using a nano-sized fine-grained granulate as a green body along with an adequate process control, it is possible to minimize both, the porosity (which is smaller than 0.05% in volume), as well as the generated average grain size (smaller than 1 m). It is known that both leads to a dramatic increase in hardness which outperforms most metal alloys at considerably lower weight and thus yields a HPC such as AlON, Al2O3, SiC or Boron Carbide (B4C). Characteristic for HPCs are an extremely low porosity (less than 0.1% in volume), high purity, and high hardness of the final macroscopic structure. An additional beneficial property of HPCs is the fact that, depending on the final grain size, the ceramics exhibit translucency or even complete transparency which renders these materials the prime source for future engineering applications [15][16]. Typical applications of HPCs that benefit from high hardness at low weight are e.g. wear resistant brake discs, protection shields in the automobile industry or bone substitutes in medical devices.

The use of extremely small grain sizes below 100 nm in the making of HPCs results again in decreasing hardness [16]. Hence, there is no simple connection between grain size and hardness of a polycrystalline material. As a consequence, today, one is compelled to search for the optimal micro structure for a specific application by intricate and expensive experimental trial-and-error studies. Some of the mechanical properties of HPCs are measured at EMI by means of ballistic high-speed impact experiments in the experimental standard set-up of the EOI configuration, where a fast impactor hits the edge of a ceramic tile of typical dimension $(10 \times 10 \times 2)\,\mathrm{cm}^3$, cf. Fig. 2.

Here, the ceramic specimens are placed at a distance of 1cm in front of the muzzle of a gas gun in order to achieve reproducible impact conditions. In this set-up the rear of the projectile is still guided by the barrel gun when the front hits the target.

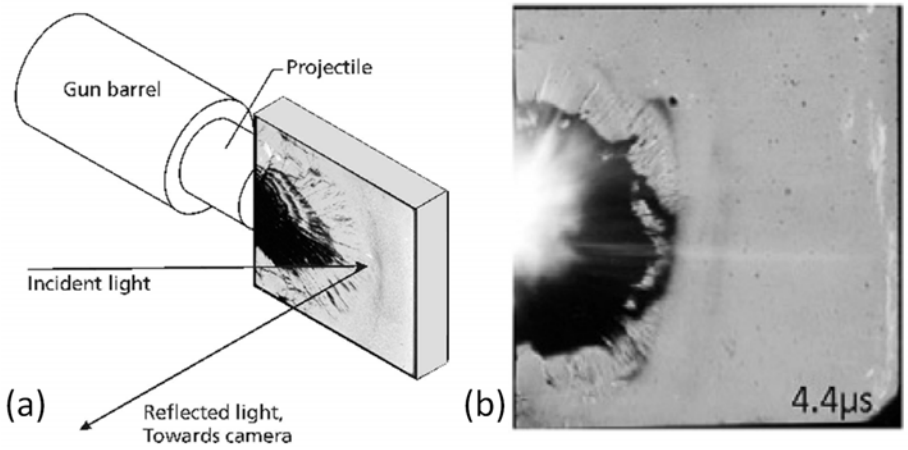


Figure 2: (a) The edge-on impact (EOI) configuration for the reflected light set-up. (b) Sample high-speed photograph of an EOI experiment with SiC impacted at striking velocity 1040 m/s displaying the propagating shock wave through the material.

Modeling and Simulation of Granular Microstructures

With numerical investigations taking explicitly into account the microstructural details, one can expect to achieve a considerably enhanced understanding of the structure-property relationships of such materials [17][18]. For simulations of macroscopic material behavior, techniques based on a continuum approximation, such as the Finite Element Method (FEM) or Smooth Particle Hydrodynamics (SPH), are almost exclusively used. In a continuum approach the considered grain structure of the material is typically subdivided into smaller (finite) elements, e.g. triangles in 2D or tetrahedra in 3D. Upon failure, the elements are separated according to some predefined failure modes, often including a heuristic Weibull distribution [19], which is artificially imposed upon the system. Results using these methods are usually strongly influenced by mesh resolution and mesh quality [20]. On the other hand, classical molecular dynamics (MD) simulations based on Newtonian dynamics of particles have been shown to capture the occurring physical phenomena of shock waves in solids correctly, but are usually limited to the nanoscale and to timescales which are too small to allow for a direct comparison with experiments.

Particle Simulations of Failure in Polycrystalline Materials

In our approach to modeling impact failure of polycrystalline, brittle materials such as ceramics, we use the framework of classical particle dynamics simulations.

Particle Model

Using Occams Razor, instead of trying to directly reproduce the geometrical shape of grains of ceramics as seen in photomicrographs, cf. Fig. 1, we model the solid state as an unordered network of monodisperse soft particles with radii , connected by non-linear elements (springs) which are allowed to overlap in the initial random configuration, see Fig. 3.

The initial random degree of overlap between each particle pair and determines the force needed to detach these particles from each other. The force is imposed on the particles by elastic springs. This simple model can easily be extended to incorporate irreversible changes of state such as plastic flow. However, for brittle materials, where catastrophic failure occurs after a short elastic strain, plastic flow behavior can be neglected. The initial disc overlap and thus the overall density of the model solid can be adjusted by a compactness parameter as dimensionless input parameter of the simulation model. In the example in Fig. 3, . The same overall system configuration can then be visualized as a network of links that connects the centers of overlap-

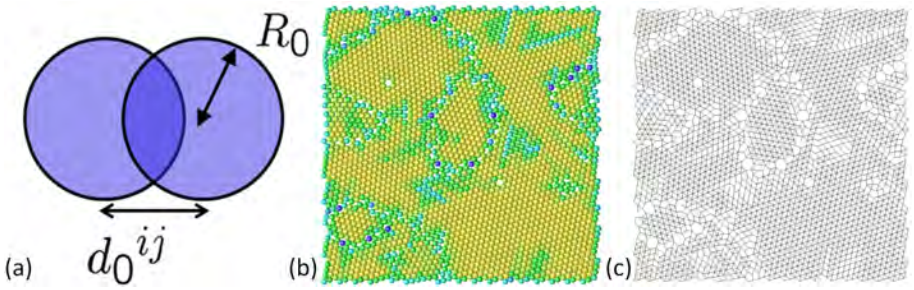


Figure 3: (a) Scheme of overlapping soft particles. (b) Realization of an initial particle configuration in the simulation with $N = 2500$ and $\Theta = 0.9$. The color code displays the structure in terms of nearest next neighbors (coordination number): blue: 0, green: 2, and yellow: 4. (c) The corresponding illustration of the system as unordered network of links.

ping particles. This way of modeling a solid composed of particles was originally used for the simulation of granular flow problems in geophysical models by Cundall and Strack [21], often referred to as the Discrete Element Method (DEM). Though DEM is very closely related to the MD method, it is generally distinguished by its inclusion of rotational degrees-of-freedom, of complicated contact forces and often complicated geometries (including polyhedra).

Potentials and Scaling Properties

As a fundamental requirement for our particle model we demand to have very few parameters which model the basic material properties of a brittle ceramic material; in essence, these are first, the resistance to pressure, second, the cohesive forces that keep the material constituents together, and then the microscopic failure. A material resistance against pressure is introduced by a Lennard-Jones-type repulsive potential

$$V_R{}^{ij}(d^{ij}) = \alpha {R_0}^3 \left[\left(\frac{{d_0}^{ij}}{d^{ij}} \right)^{12} - 2 \left(\frac{{d_0}^{ij}}{d^{ij}} \right)^{6} + 1 \right] \tag{1}$$

which acts on every pair $\{ij\}$ of particles for $0 < d^{ij} \leq {d_0}^{ij}$ and which vanishes for $d^{ij} > {d_0}^{ij}$, i. e. for particle pairs which do not overlap. Factor α (which relates to the compressive strength) in Eqn. (1) scales the energy density and the pre-factor ${R_0}^3$ ensures the correct scaling behavior of the calculated total stress $\sigma_{ij}\sigma^{ij} = \sum_{ij} F^{ij}/A$ in the system (with A being the area where the force F^{ij} is applied), independent of the number of particles N. The cohesive potential is modeled by a harmonic function $V_C{}^{ij}(d^{ij})$, given that there are no irreversible changes of state when the material is submitted to small external forces. Each pair of particles $\{ij\}$ can be visualized as being connected by a spring, the equilibrium length of which equals the initial distance ${d_0}^{ij}$, cf. Fig. 3. Thus, for $d^{ij} > 0$ we have:

$$V_C{}^{ij}(d^{ij}) = \beta R_0 \left(d^{ij} - {d_0}^{ij} \right)^2 . \tag{2}$$

In Eqn. (2) parameter β (which has the dimension [energy/length] and relates to the tensile strength of the material) determines the strength of the potential and the pre-factor R_0 again ensures proper scaling behavior of the macroscopic physical material response, e.g. the stress-strain curve upon external load, for details, see Steinhauser [12]. We demonstrate this model property in Fig. 4 which displays the stress-strain relation obtained for different realizations of solids with a different number of particles. The idea of this particular scaling of the potential

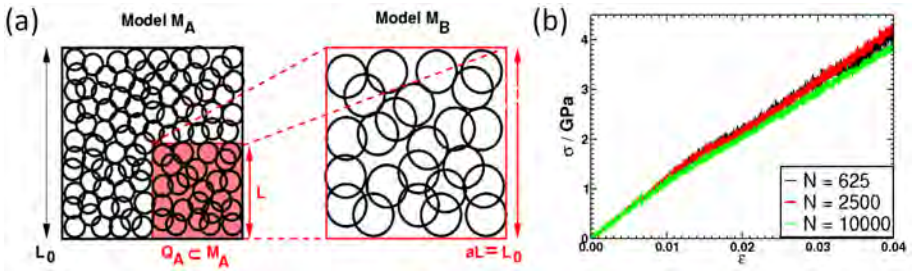


Figure 4: (a) Illustration of the scaling property of our particle model. A number of particles of a small subset $Q_A \subset M_A$ of the original model M_A are enlarged, until they are of the same size as the original system. The resulting system M_B contains fewer particles than M_A, but the macroscopic physical properties, e.g. Youngs modulus stay the same upon external load. (b) Averaged stress-strain curves obtained from simulations of systems with different number of particles. $N = 625, 2500, 10^4$, $\alpha = 20$, $\beta = 350$. The slope of the different realizations is in essence independent of N.

is that one should always obtain the same macroscopic physical properties of a system, e.g. the same stress-strain relation, independent of the arbitrarily chosen number of particles which represent the solid. As a result of this, when up- or downscaling our system, i.e. when changing the number of particles, it is not necessary to re-adjust the pre-factors α and β in the potentials of Eqn. (1) and (2).

Finally we consider failure in our model by introducing two breaking thresholds for the springs with respect to compressive and to tensile failure. If either of these thresholds is exceeded, the respective spring is defined as broken and is removed from the system. A simple tensile criterion is reached when the overlap between two particles vanishes, i.e. when the distance between two particle centers exceeds the sum of their constant radii:

$$d^{ij} > 2R_0\,. \tag{3}$$

Failure under pressure load occurs when the actual mutual particle distance is less by a factor γ than the initial mutual distance, i.e. when

$$d^{ij} < \gamma/, {d_0}^{ij}\,, \tag{4}$$

where $(0 < \gamma < 1)$. Parameter γ is later fitted to reproduce Young's modulus of the real material. We note that the repulsive potential is independent from the failure criteria of Eqn. (3) and (4), i.e. even if bonds described by Eqn. (2) are broken in the system due to pressure or tensile failure, the involved particles still interact via the repulsive potential of Eqn. (1) and cannot artificially move through each other.

Initial Configurations

For our numerical analysis, we simulate directly the experimental geometry of the edge-on impact configuration (EOI) as shown in Fig. 2. We use as initial configuration a random distribution of particles in a cubic simulation box. Generally, we observe an increase in the number of pronounced peaks in the distribution of initial particle distances ${d_0}^{ij}$ when parameter Θ is increased, as shown in Fig. 5. This clearly indicates a change of the initial structure, i.e. of the arrangement and packing density of particles in the system. In Fig. 5 we display the coordination numbers for two initial realization of a brittle, granular solid at two different densities. Thus, by fine-tuning parameter Θ, one can fix the density ρ of the model material according to

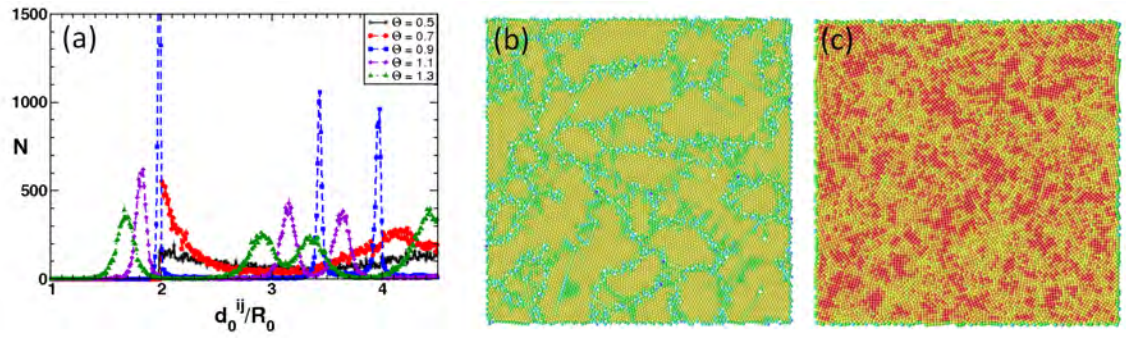


Figure 5: (a) Distribution of mutual initial distances for different values of in a system with particles. (b) Sample initial configurations with and with a preferred hexagonal arrangement of soft particles. The color code displays the coordination number within the range 4-6 (yellow to green). (c) Another realization with the same and , coordination numbers 6-8 (green to red) and with a predominantly quadratic packing.

the one obtained in sintered ceramics of interest, e.g. in the case of Al_2O_3, the experimental density ρ is typically is larger than 98% in volume.

High-Speed Impact Simulations

By adjusting the three free model parameters α, β, γ to experimental values typical for HPC materials, one is able to obtain the correct stress–strain relationship of a specific material as observed in (macroscopic) biaxial loading experiments. After this fixing of parameters the model is applied to other types of external loading, e.g. ballistic high–speed impact in the EOI configuration or a direct impact which can be used as a model system for investigating the situation of a satellite being hit by space debris. This is done with no further model adjustments, and the results are compared with experimental findings. In Fig. 6 we present non–equilibrium molecular dynamics simulation (NEMD) results for a SiC system with impact velocity $v = 150$m/s using $N = 10^5$ particles in a direct comparison with corresponding high-speed experimental results. In general, one can conclude that the physics of shock wave propagation is captured rather well in the simulations, opening a route to a detailed quantitative investigation of observed shock wave and failure phenomena in brittle materials, for example by investigating the number of broken bonds in the system as displayed in the bottom row of part (a) in Fig. 6. Part (b) of Fig. 6 analyzes the ratio of broken bonds to the total number of initial bonds for SiC and Al_2O_3 for different impact velocities and system sizes. The percentage of failed bonds which can be considered as a simple measure for the degree of failure in the material, is consistently larger for SiC, which agrees well with experimental findings.

Finally, in Fig. 7 we show a 3D series of simulation snapshots of the developing debris cloud resulting from a direct ballistic impact of a spherical particle onto a plate, directly after the impact occurred. Debris from the impactor is colored in red and gray particles represent the target. The long–term purpose of this type of impact simulations is to develop a computational model that reproduces the debris cloud distribution which is observed in corresponding high-speed ballistic experiments. This is important for evaluating satellite safety in the earth's orbit, where a continuously increasing number of debris particles increase the risk of collisions with satellites.

Figure 8 displays a typical experimental high-speed photograph of a ballistic impact experiment, shortly after the impact occurred. One can clearly see the debris cloud forming. For comparison we have displayed the debris cloud obtained from a simulation snapshot.

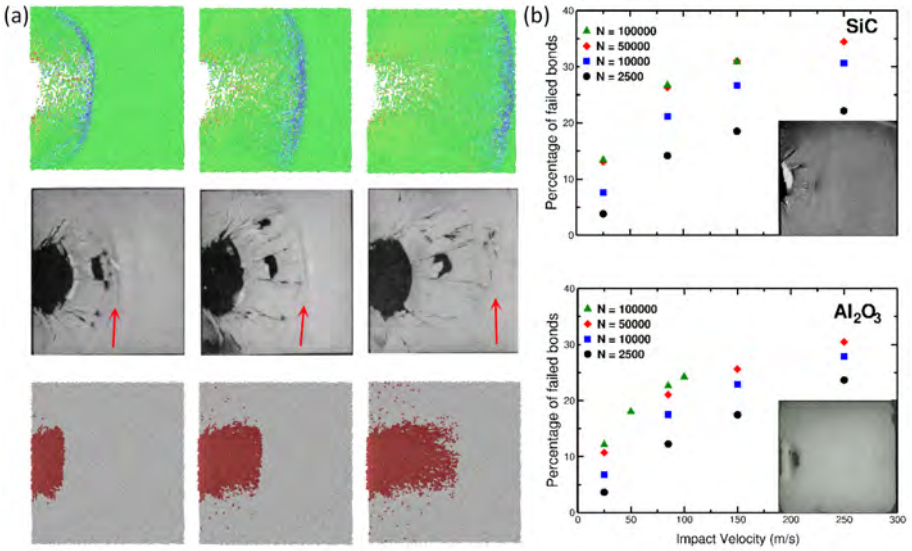


Figure 6: (a) Top row: Simulation results of a SiC EOI simulation at $v = 150$m/s. The material is hit at the left edge. A shock wave (color-coded in blue) propagates through the system. The time interval between the individual snapshots from left to right is 2μs. Middle row: The same experiment with a real SiC specimen. The time interval between the photomicrographs is comparable with the ones in the top row. Arrows indicate the location of the shock wave front. Bottom row: The same computer simulation, this time displaying the occurring damage in the material with respect to broken bonds. (b) Degree of damage at 3μs after impact for SiC and Al_2O_3 for different N and impact velocities. The insets show high-speed camera snapshots indicating the corresponding degree of damage in the materials at striking velocity $v = 85$m/s.

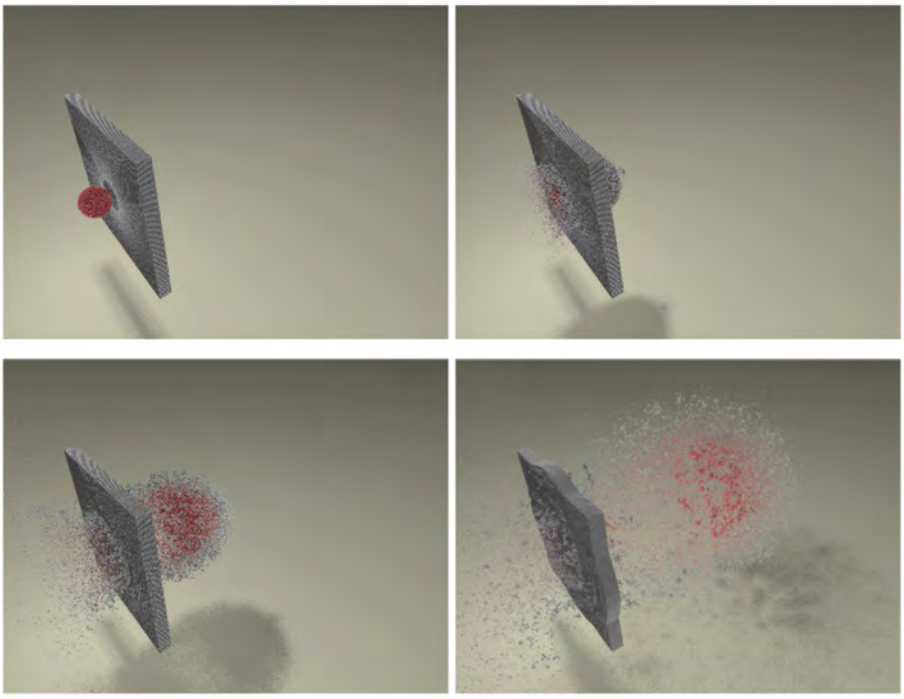


Figure 7: A series of 3D simulation snapshots showing a sphere impacting a plate at very high speed ($v = 6.7$km/s) in a ballistic impact simulation. Red indicates impactor particles and gray indicates target plate particles.

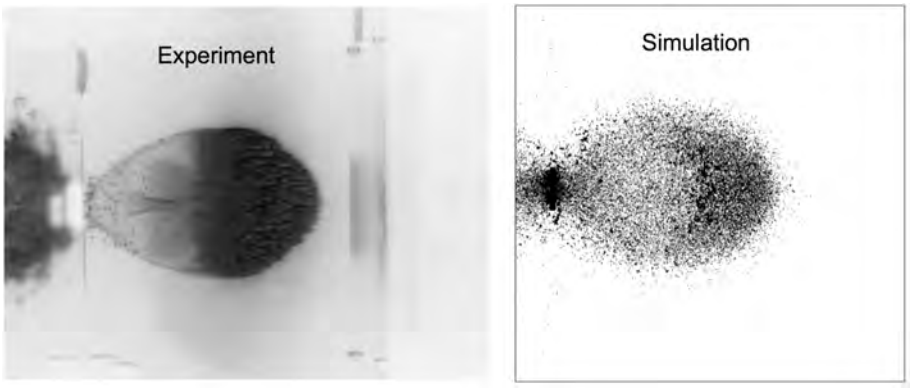


Figure 8: Left: Experimental snapshot of the debris cloud. Right: Realization of this ballistic impact experiment in our particle–based computer simulation.

Conclusions

The proposed simulation scheme in this paper, which uses discrete particles to model the basic properties of a solid, has proven to be stable and convergent. It allows for studying in detail the fracture and failure mechanisms of brittle materials. The simulated failure dynamics, shock wave propagation and the degree of damage with the proposed three-parameter model are in good agreement with experimental findings, albeit in the presented study for the edge-on-impact configuration, only moderate velocities are used to impact the material, as here, we want to demonstrate the principal usefulness of the proposed model by fitting its parameters to a specific, brittle ceramic material. In a first attempt to go to very high impact velocities (larger than 6km/s, which is sometimes called *hypervelocity impact*) we have presented a ballistic impact simulation which results in the formation of a debris cloud. Here, we also find good numerical stability of the proposed particle model and reasonable agreement with corresponding ballistic impact experiments.

In future investigations, it is planned to extend the proposed modeling approach to the simulation of yet larger impact velocities and to other, more complex materials, such as compounds, e.g. fiber-reinforced SMCs (Sheet Moulding Compounds) or to structures typically encountered in soft matter, e.g. biological bilayer membranes, which exhibit much more complex microstructural features.

Acknowledgements

We acknowledge financial support by the German Aerospace Center (DLR) under grant number 50LZ1502 ”DEM-O”.

References

[1] Steinhauser, M. O and Grass, K. (2005) Failure and plasticity models of ceramics: A numerical study. In: Khan A, Kohei S, Amir R., editors, Dislocations, Plasticity, Damage and Metal Forming, Materials Response and Multiscale Modeling. Neat Press, Fulton, ML USA, pp. 370–373.

[2] Abraham, F. F, Walkup R., Gao, H., Duchaineau, M., De La Rubia, T. D., Seager, M. (2002) Simulating materilas failure by using up to one billion atoms and the worlds fastest computer: Brittle fracture. *Proc. Natl. Acad. Sci.* **99**, 5777–5782.

[3] Kadau, K., Germann, T. C., Lomdahl, P.S. (2006) Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/L. *Int J. Mod. Phys. C* **18**, 1755–1761.

[4] Fineberg, J. Materials science: Close-up on cracks. (2003) *Nature* **426**, 131–132.

[5] Buehler, M. J., Gao, H. Dynamical fracture instabilities due to local hyperelasticity at crack tips. (2006) *Nature* **439**, 307–310.

[6] Abraham, F. F., Brodbeck, D., Rafey, R. A., Rudge, W. E. Instability dynamics of fracture: A computer simulation investigation. (1994) *Phys. Rev. Lett.* **73**, 272–275.

[7] Abraham, F. F., Gao, H. How fast can cracks propagate? (2000) *Phys. Rev. Lett.* **84**, 3113–3116.

[8] Rosakis, A. J., Samudrala, O., Coker, D. (1999) Cracks faster than the shear wave speed. *Science* **284**, 1337–1340.

[9] Abraham, F. F., Schneider, D., Land, B., Lifka, B., Skovira, J., Gerner, J., Rosenkrantz, M. (1997) Instability dynamics in three-dimensional fracture: An atomistic simulation. *J. Mech. Phys. Solids* **45**, 1461–1471.

[10] Bulatov, V., Abraham, F. F., Kubin, L., Devrince, B., Yip, S. (1998) Connecting atomistic and mesoscale simulations of crystal plasticity. *Nature* **391**, 669–672.

[11] Gross, S. P., Fineberg, J., Marder, M., McCormick, W. D., Swinney, H. L. (1991) Acoustic emissions from rapidly moving cracks. *Phys. Rev. Lett.* **71**, 3162–3165.

[12] Steinhauser M. O. (2008) *Computational Multiscale Modeling of Fluids and Solids Theory and Applications*, Springer, Heidelberg, Berlin , New York.

[13] Zhou, S. J., Beazley, D. M., Lomdahl, P. S., Holian, B. L. (1997) Large-scale molecular dynamics simulations of three-dimensional ductile failure. *Phys. Rev. Lett.* **78**, 479–482.

[14] Zhou, S. J., Preston, D. L., Lomdahl, P. S., Beazley, D. M. (1998) Large-Scale molecular dynamics simulations of dislocation intersection in copper. *Science* **279**, 1525–1527.

[15] Krell, A., Blank, P., Ma, H., Hutzler, T., van Bruggen, M., Apetz, R. (2003) Transparent sintered corundum with high hardness and strength. *J. Am. Chem. Soc.* **86**, 12–18.

[16] Krell, A., Blanck, P., Ma, H. W., Hutzler, T., Nebelung, M. (2003) Processing of high-density submicrometer Al_2O_3 for new applications. *J. Am. Ceram. Soc.* **86**, 546–553.

[17] Chen, M., McCauley, J. W., Hemker, K.J. (2003) Shock-induced localized amorphization in boron carbide. *Science* **299**, 1563–1566.

[18] Bringa, E. M., Caro, A., Wang, Y., Victoria, M., McNaney, J. M., Remington, B. A., Smith, R. F., Torralva, B. R., Van Swygenhoven, H. (2005) Ultrahigh strength in nanocrystalline materials under shock loading. *Science* **309**, 1838–1841.

[19] Weibull, W. A statistical distribution function of wide applicability. (1951) *J. Appl. Mech.* **18**, 293–297.

[20] Steinhauser, M. O., Hiermaier, S. (2009) A review of computational methods in materials science: Examples from shock-wave and polymer physics. *Int. J. Mol. Sci.* **10**, 5135–5216.

[21] Cundall, P. A., Strack, O. D. L. (1979) A discrete numerical model for granular assemblies. *Gotechnique* **29** 47–65.

Heat flux identification using reduced model and the adjoint method. Application to a brake disk rotating at variable velocity

S. Carmona, Y. Rouizi, O. Quéméner, F. Joly*

Laboratoire de Mécanique et d'Energétique d'Evry, Université d'Evry Val d'Essonne
40 rue du Pelvoux CE1455, Courcouronnes, 91020 Evry Cédex, France
*Corresponding author : Frédéric JOLY (f.joly@iut.univ-evry.fr)

Abstract

In previous works [1], reduced models have been used for solving inverse problems, characterized by a complex geometry requiring a large number of nodes and / or an objective of online identification. The treated application was a brake disc in two-dimensional representation, in rotation at variable speed. The dissipated heat flux at the pad-disk interface had been identified by Beck's method. We present here a similar application using the adjoint method. The modal reduction is done by using special bases (called branch bases) that offer the advantage of dealing with nonlinear problems and / or unsteady parameters. Adjoint method provides particularly accurate results in this configuration.

Keywords : Reduced model, Modal method, Inverse problem, Advection-diffusion equation, Adjoint method

Nomenclature

c	Heat capacity [$J.m^{-3}.K^{-1}$]
e	Disk thickness [m]
k	Thermal conductivity [$W.m^{-1}.K^{1}$]
h	Heat exchange coefficient [$W.m^{-2}.K^{-1}$]
U	Disk velocity [$m.s^{-1}$]
T	Temperature [°C]
z_i	Eigenvalue [s^{-1}]
x	Modal temporal amplitude
V	Eigenvector [K]
N_t	Number of measurement steps

Greek symbols

ϕ	Heat flux [W]
ω	Rotation velocity [$rad.s^{-1}$]
ζ	Steklov number [$kg.s^{-2}.K^{-1}$]

subscript

u	dimensionless quantity
m	Maximum Value
~	Reduced quantity
^	Estimate

Introduction

In the domain of heat conduction, inverse problems are generally ill-posed in the sense of Hadamard and then require complex procedures to obtain satisfactory results. Two techniques are used, the future time step method (Beck) [2] which has the particularity of being a sequential method, and the adjoint method [3] which is an iterative method based on successive computation of descent directions to minimize a criterion taking into account all the data.

In these inverse problems, mathematical complexity of the technique limits the size of the characteristic matrices of the thermal problem and the different studied geometries are often reduced to a simple, two-dimensional appearance. This problem is even more blatant when it comes to conduct an online identification, which involves fast calculations [4]. Under these conditions, the use of modal models [5], which allows a significant decrease in the number of unknowns while maintaining a satisfactory accuracy over the entire domain, allows the extension of the inverse techniques to geometries characterized by mesh of large size. Already developed to a diffusion-transport problem, this identification technique using low order

models involved the identification by Beck's method of the heat flux dissipated by friction during braking phases of a brake disk [1]. A similar configuration is studied in order to extend the use of reduced models to the adjoint method. A comparison between the two techniques is then presented.

Position of the problem

A brake disk (Fig. 1) rotating at variable speed is considered. It receives during the braking phase a time dependant heat flux on the friction zone with the brake pads (domain Ω_1). In the studied case, the surrounding temperature is set to T_{ext} = 0°C and the uniform initial temperature field is T_0 = 0°C. The different time dependant parameters, the radial velocity $\omega(t)$, the heat exchange coefficient $h(t)$ and flux dissipated by friction $\phi(t)$ are expressed in terms of their maximum values and are therefore dimensionless:

$$\omega(t) = \omega_u(t)\,\omega_m, \tag{1}$$

$$h(t) = h_u(t)\,h_m, \tag{2}$$

$$\phi(t) = \phi_u(t)\,\phi_m = \phi_u(t) \int_{\Omega_1} \varphi_m(r)\,d\Omega \tag{3}$$

The heat flux dissipated by friction φ_m is not uniform on Ω_1 but varies linearly with velocity, so with the radius. The temporal evolution of $\omega_u(t)$, $h_u(t)$, and $\phi_u(t)$ are shown in Figure 2, and their maximum values are $\omega_m = 2\pi$ rad/s, h_m = 110W.m^{-2}.K^{-1}, and ϕ_m = 600 W.

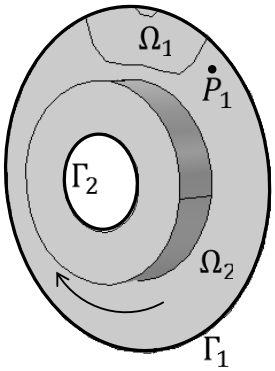


Figure 1 : Computational domain.

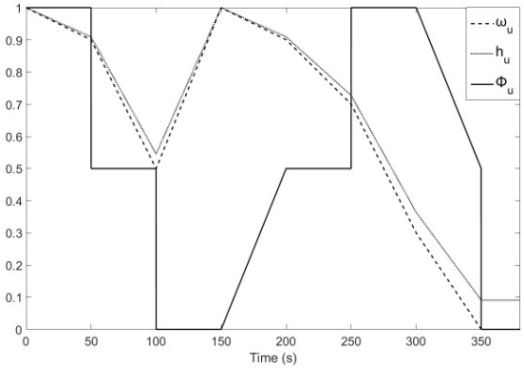


Figure 2 :Temporal evolution of thermal loads

Numerical solution : the detailed model

Given the characteristic dimensions of the disk (k = 50W.m^{-1}.K^{-1}, c = 3.66.10^{6}J.m^{-3}.K^{-1}, e = 8 mm), the Biot number corresponding to the worst case (h_m = 110W.m^{-2}.K^{-1}) has a value Bi = 0.018 << 1. It is then possible to neglect the thermal gradient in the thickness e of the disc. By setting (η, ζ) local coordinates Ω in the plane perpendicular to this thickness, temperature is expressed as $T(x, y, z) = T(\eta, \zeta)$. This produces a thermal problem of shell type whose variational formulation is :

$$\begin{aligned} \int_{\Omega} ec\frac{\partial T}{\partial t} g\,d\Omega = &-\int_{\Omega} ek\,\vec{\nabla}T\cdot\vec{\nabla}g\,d\Omega - \omega_u(t)\int_{\Omega} ec\,\vec{U}_m.\vec{\nabla}T\;g\,d\Omega \\ &- h_u(t)\left(\int_{\Omega_1} h_m T\,g\,d\Omega + \int_{\Gamma_1} h_m T\,g\,d\Gamma\right) + \phi_u(t)\int_{\Omega_2}\varphi_m\,g\,d\Omega \end{aligned} \tag{4}$$

with $g \in H_1(\Omega)$ a test function, and $\Omega = \Omega_1 \cup \Omega_2$.

The discretization of this problem by linear finite element reveals a matrix system of dimension *N* (number of nodes) which is written in the order of previous terms:

$$\mathbf{C}\dot{\mathbf{T}} = \left[\mathbf{K} + \omega_u(t)\mathbf{U} + h_u(t)\mathbf{H}\right]\mathbf{T} + \phi_u(t)\mathbf{\Pi} \tag{5}$$

After a sensitivity analysis, the mesh consists of 9,860 nodes forming 19,362 triangle elements. For a direct problem, the temporal heat flux evolution is known and the evolution of the discrete temperature field **T** is done by solving Eq. (5). Figure 3 represents the evolution of temperature at point A, placed 10mm downstream from the friction area (see Fig. 1). The analysis of the temperature field shows that the local friction on the Ω_1 area leads to the appearance of a sharp temperature front conveyed at the rotational speed. A fixed sensor detects a very rapid temperature variation, which as will be seen later makes the inverse problem difficult to solve.

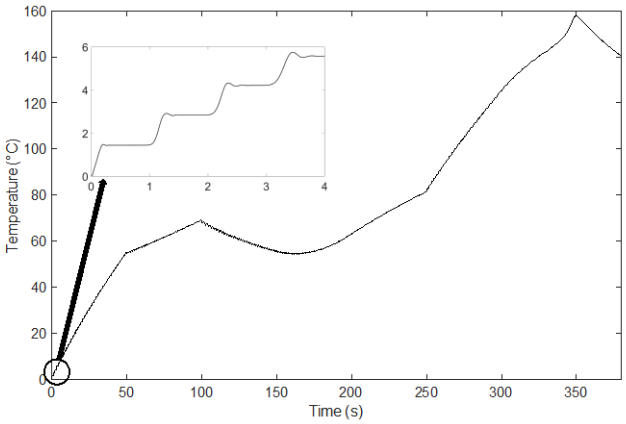


Figure 3 :Temperature evolution at point A

Modal reduction

The branch problem

The modal decomposition supposes the existence of a base such that the following decomposition is unique:

$$T(M,t) = \sum_{i=1}^{N} x_i(t) V_i(M) \tag{6}$$

where $V_i(M)$ are eigenvectors (or modes), and $x_i(t)$ are the unknown coefficients named hereafter modal amplitudes. The modes can be seen as elementary thermal fields.

The branch problem associates to the previous physical problem an eigenvalue problem defined by equations (7) and (8):

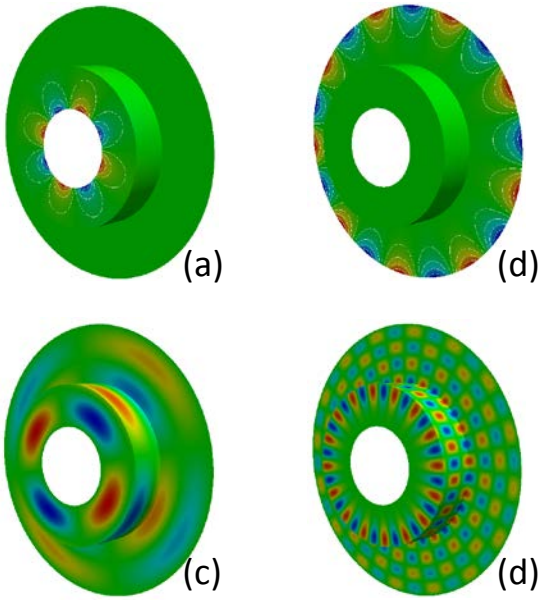
$$\forall M \in \Omega_1 \cup \Omega_2, \quad k\nabla^2 V_i = z_i\, c V_i \tag{7}$$

$$\forall M \in \Gamma_1 \cup \Gamma_2, \quad k\,\vec{\nabla} V_i \cdot \vec{n} = -z_i\,\zeta\, V_i \tag{8}$$

where z_i is the eigenvalue associated with the eigenvector V_i.

The boundary condition (Eq. (8)) is a non physical condition that involves the eigenvalue of the mode. The number of Steklov ζ ensures dimensional homogeneity of the boundary condition and prevents degeneration of the modal problem, *i.e.* to balance Eqs. (7) and (8).

This special boundary condition reveals two types of modes. The first type is constituted of modes quasi null on the boundary but not on the domain (domain modes), and the second one formed of modes quasi null on the domain but not on the boundary (boundary modes). This second type of modes allows to link the temperature fields on the interface. Examples of such modes are given in Fig. 4. The existence of boundary modes allows one to rebuild temperature and thermal flux density for all convective coefficient. This basis is then adapted to nonstationary and nonlinear thermal problems.



Figures 4 : Examples of branch modes : boundary modes ((a) and (b)) and domain modes ((c) and (d))

Reduction method

The modal formulation only shifts the problem : instead of being temperature values at the nodes of a mesh, the unknowns are the amplitudes of the modes $x_i(t)$. The number of modes needed to approach correctly the solution needs to be reduced. This is done by the amalgam method [5] [6]. In this method, the most influential eigenmodes are kept (they are called major eigenmodes), and the remaining eigenmodes (called minor) are added to them, weighted by a factor $\alpha_{i,p}$. This results in new amalgamated eigenmodes $\tilde{V}_i$, which are a linear combination of eigenvectors of the original branch basis.

$$\tilde{V}_i = \sum_{p=0}^{n_r} \alpha_{i,p} V_{i,p} \tag{9}$$

The determination of factors $\alpha_{i,p}$ is performed by minimizing the deviation of energy between a reference model and the reduced model. Note that in our case the reference problem used is constructed independently of the temporal evolution $\phi_u(t)$ to be identified. With these amalgamated modes, the modal decomposition of temperature is given by :

$$T(M,t) \cong \sum_{i=1}^{\tilde{N}} \tilde{x}_i(t) \tilde{V}_i(M) \tag{10}$$

The amplitude equation is obtained by replacing the temperature by its modal decomposition (Eq. (10)) in the physical problem (Eq. (4)), while the test functions are the modes. It replaces

the problem on temperatures at the nodes of the mesh size by a problem on the temporal amplitudes of the modes. In discrete form, Eq. (4) becomes:

$$\begin{aligned}\mathbf{L}\dot{\tilde{\mathbf{X}}} &= \left[\mathbf{M_K} + \omega_u(t)\mathbf{M_U} + h_u(t)\mathbf{M_H}\right]\tilde{\mathbf{X}} + \phi_u(t)\mathbf{N} \\ &= \mathbf{M}(t)\,\tilde{\mathbf{X}} + \phi_u(t)\mathbf{N}\end{aligned} \tag{11}$$

where $\mathbf{L} = \tilde{\mathbf{V}}^{\mathbf{t}}\mathbf{C}\tilde{\mathbf{V}}$, $\mathbf{M_K} = \tilde{\mathbf{V}}^{\mathbf{t}}\mathbf{K}\tilde{\mathbf{V}}$, $\mathbf{M_U} = \tilde{\mathbf{V}}^{\mathbf{t}}\mathbf{U}\tilde{\mathbf{V}}$, $\mathbf{M_H} = \tilde{\mathbf{V}}^{\mathbf{t}}\mathbf{H}\tilde{\mathbf{V}}$ et $\mathbf{N} = \tilde{\mathbf{V}}^{\mathbf{t}}\mathbf{\Pi}$, $\tilde{\mathbf{V}}$ being the matrix containing the $\tilde{N}$ amalgamated eigenvectors, and vector $\tilde{\mathbf{X}}$ contains the $\tilde{N}$ temporal amplitude $\tilde{x}(t)$.

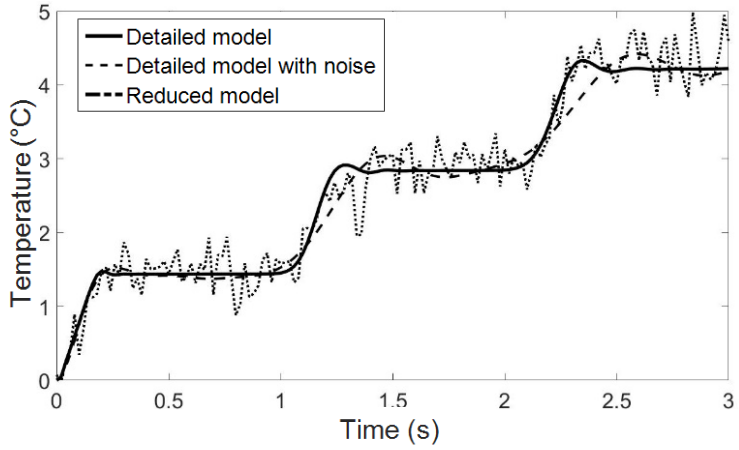


Figure 5 :Temperature evolution at point A obtained with different models for the first 3 seconds of the simulation

A reduced base with 50 modes is used. In the case of the direct problem (Eq. (11)), the modal model recovers the evolution of the thermal field with an average error compared to the detailed model of 0.046°C and a maximum error of 6.18°C, the temperature range being of 139°C, showing the good agreement between reduced and detailed models. At point A, the error averaged over time is 0.105°C. Figure 5 shows the temperature difference between these two models at the measurement point for the first seconds of the simulation.

Inverse problem

The temporal evolution of the heat flux received by friction by the rotating disc is identified from an observable vector **Y**, consisting here of a single measurement point located at A. Given the size of the discrete problem, modal formulation is used to reduce the size of the inverse problem. The relationship between the output vector **Y** and the modal amplitude **X** has to be added to the direct problem defined by equation (11) :

$$\mathbf{Y} = \mathbf{E}\mathbf{T} = \mathbf{E}\,\tilde{\mathbf{V}}\tilde{\mathbf{X}} \tag{12}$$

Two inversion techniques are used, Beck and adjoint method.

Beck's method

Beck's method consists in determining the amplitude of flux at each time step so that the temperature difference between the measurement and the simulation is the smallest possible.
An implicit time discretization (at a fixed time-step Δt = 0.02 s) of Eq. (11) yields the amplitude of each mode:

$$\tilde{\mathbf{X}}^{k+1} = \left[\mathbf{L} - \Delta t\,\mathbf{M}(t)\right]^{-1}\left[\mathbf{L}\tilde{\mathbf{X}}^{k} + \Delta t\;\phi_u^{k+1}\,\mathbf{\Pi}\right] \tag{13}$$

A least squares minimization between measurement and temperature computed from the estimate at the previous time-step brings the estimation of the searched solicitation :

$$\phi_u = \left[\mathbf{\Theta}^{\mathbf{t}}\mathbf{\Theta}\right]^{-1}\mathbf{\Theta}^{t}\;\mathbf{Z}^{k+1} \tag{14}$$

with $\mathbf{\Theta}$ and $\mathbf{Z}$ defined by :

$$\mathbf{\Theta} = \mathbf{E}\left[\mathbf{L} - \Delta t\,\mathbf{M}\right]^{-1}\left[\Delta t\,\mathbf{\Pi}\right] \tag{15}$$

$$\mathbf{Z}^{k+1} = \mathbf{Y}^{k+1} - \mathbf{E}\left[\mathbf{L} - \Delta t\,\mathbf{M}\right]^{-1}\left[\mathbf{L}\,\hat{\tilde{\mathbf{X}}}^{k}\right] \tag{16}$$

This technique is first used in an ideal case, wherein the temperature variation at point A used for identification comes directly from the simulation performed by the reduced model. There is then no error in this situation between the measurement and the direct model. The accuracy of the identification carried out is characterized by global error on the flux (σ_{ϕ_u}) and the temperature (σ_T), which are defined by equations (17) and (18)

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{i=Nt}\left(Y(i) - \hat{Y}(i)\right)^2}{Nt}} \tag{17}$$

$$\sigma_{\phi_u} = \sqrt{\frac{\sum_{i=1}^{i=Nt}\left(\phi_u(i) - \hat{\phi}_u(i)\right)^2}{Nt}} \tag{18}$$

These simulations were performed for a time-step equal to 0.02s. The choice of this reduced time-step is explained by the influence of the transport term that creates sudden temperature changes that need to be taken into account for identification. The choice of the time-step then does not depend on the simple diffusion time between the source and the sensor, but also on the time of transport and the ability of the model to detect sudden changes in temperature. In an ideal case, results are satisfying as $\sigma_{\phi_u} = 0.039$ and $\sigma_T = 0.962°C$.

In a real case, the temperature of a probe is simulated by the full thermal model (Eq. (5)), to which is added a Gaussian white noise characterized by a quadratic error σ_b= 0.3 °C. In this case the identification results are directly unusable: the error on the identified flux is σ_{ϕ_u} =1.59. Indeed, as shown in Fig. 5, the bias to both the use of reduced model in the inverse procedure and measurement error strongly modifies the rapid changes in temperature. The various regularization attempts (increasing the number of measurement points, using a growing number of future time-steps) do not improve significantly the results. This problem had already been shown in previous work [1], and the recommended solution was the use of a low-frequency filter on the identified flux by Fourier transform, assuming that any variation

frequency greater than the frequency rotation could only be a numerical distortion. The application of this technique to our configuration is shown in Figure 6. The results are satisfactory since the use of a 0.4 Hz cutoff frequency results in an error on the identified flow equal to σ_{ϕ_u} = 0.038 and an error on the temperature σ_T = 0.832 ° C.

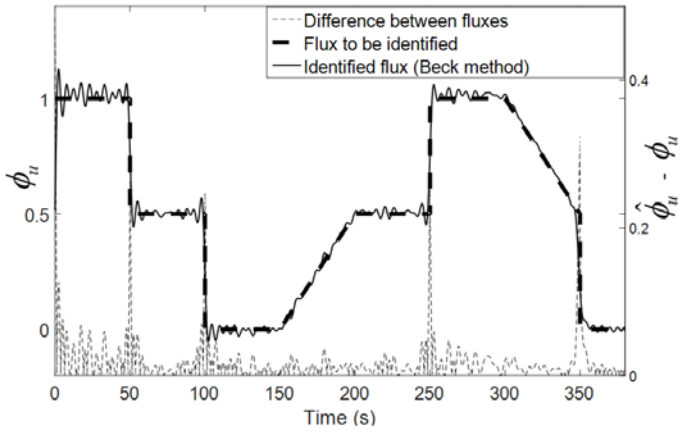


Figure 6 :Heat flux identification obtained with Beck's method after filtering

Adjoint method

The second inversion technique is the adjoint method. It is a global method in which a quadratic functional built on the differences between the measured temperatures and those computed with the identified heat flux is minimized. This function can also be penalized by a regularization term ε :

$$J(\phi_u) = \frac{1}{2}\left[\int_0^\tau \left\|\mathbf{Y}(t) - \hat{\mathbf{Y}}(t)\right\|^2 dt + \varepsilon \left\|\phi_u(t)\right\|^2\right] \tag{19}$$

The identification process consists in finding optimum solicitations $\bar{\phi}_u$ such that J is minimum.

$$\bar{\phi}_u = arg[min\, J(\phi_u)] \tag{20}$$

This problem is solved using a descent method. These methods require the estimation of the functional gradient with respect to the solicitations. The amplitude equation of the model can be seen as a constraint between the thermal loads and temperatures. It involves the Lagrangian L_a associated with the minimization problem under the constraint of the state equation. This term is constructed by summing the functional and the state equation weighted by a Lagrange multiplier (λ):

$$L_a(\phi_u, T, \lambda) = J(\phi_u) + \int_0^\tau \boldsymbol{\lambda}(t)\left(-\mathbf{L}\frac{d\tilde{\mathbf{X}}}{dt} + \mathbf{M}\tilde{\mathbf{X}} + \mathbf{N}\phi_u\right)dt \tag{21}$$

At the point where the functional is minimal, the derivatives of the Lagrangian with respect to these three variables are null :

$$\frac{\partial L_a}{\partial \lambda} = 0 \tag{22}$$

$$\frac{\partial L_a}{\partial \phi_u} = 0 \tag{23}$$

$$\frac{\partial L_a}{\partial T} = 0 \tag{24}$$

The computation of derivative defined by Eq. (22) retrieves the amplitude equation (Eq. (11)). The two last derivatives (Eqs. (23) and (24)) bring two new relations, called gradient equation and adjoint equation:

$$\nabla \mathbf{J} = \varepsilon \mathbf{U} - \mathbf{\Pi}^{\mathbf{t}} \mathbf{V} \boldsymbol{\lambda} \tag{25}$$

$$-\mathbf{L}\dot{\boldsymbol{\lambda}} = \mathbf{M}^{*}\boldsymbol{\lambda} + \mathbf{V}^{\mathbf{t}}\mathbf{E}\left(\mathbf{Y}(t) - \hat{\mathbf{Y}}(t)\right) \tag{26}$$

where **M*** is the adjoint matrix of **M**.

Thus the interest of this formulation is to compute the gradient $\nabla \mathbf{J}$ (Eq. (25)) from the resolution of the single equation (26). The iterative calculation of the thermal load φ_u^k is done using this gradient $\nabla \mathbf{J}$. Many descent patterns exist. We present here the conjugate gradient method, which combines the flux value at a previous iteration with a descent direction (noted $\mathbf{w}^{\mathbf{k}}$ here):

$$\phi_u^{k+1} = \phi_u^k + \rho^k \mathbf{w}^k \tag{27}$$

This iterative calculation is finished when one of the following criteria is met. The first is based on the evolution of the functional *J* (Eq. (28)). The second compares the difference between the estimated temperature and the measurements, which should be of the same order of magnitude as the level of uncertainty of the measurement (principle of Morozov Eq. (29)).

$$\frac{J\left(\phi_u^k\right) - J\left(\phi_u^{k-50}\right)}{J\left(\phi_u^k\right)} < 1\% \tag{28}$$

$$\sigma_T \approx \sigma_b \tag{29}$$

The direction of descent $\mathbf{w}^{\mathbf{k}}$ is a combination between the current and previous descent directions weighted by a coefficient γ^k called Fletcher-Reeves conjugation parameter:

$$\mathbf{w}^{\mathbf{k}} = \gamma^k \mathbf{w}^{\mathbf{k-1}} - \nabla \mathbf{J}^k \tag{30}$$

$$\gamma^k = \frac{\left\| \nabla \mathbf{J}^{\mathbf{k}} \right\|^2}{\left\| \nabla \mathbf{J}^{\mathbf{k-1}} \right\|^2} \tag{31}$$

and ρ^k is the optimal descent step, computed by the secant method (α is a small non null random number)

$$\rho^k = -\alpha \frac{\left\langle \nabla \mathbf{J}(\phi_u^k), \mathbf{w}^k \right\rangle}{\left\langle \nabla \mathbf{J}(\phi_u^k + \alpha \mathbf{w}^k), \mathbf{w}^k \right\rangle - \left\langle \nabla \mathbf{J}(\phi_u^k), \mathbf{w}^k \right\rangle} \tag{32}$$

The treated case corresponds to noisy temperatures (σ_b = 0.3 °C) issued from the detailed model. In the inverse procedure, the penalization term is null (ε = 0). The above presented algorithm converges to the imposed flux in 379 iterations. As shown in Figure 7, this method does not need additional filtering to recover properly the temporal flux variations. Flux deviation is σ_{ϕ_u} = 0.051, which is very slightly greater than the deviations obtained by Beck's method with low frequency filtering, and in terms of temperatures σ_T = 0.389 ° C, which is less than the error obtained by Beck's method with filtering.

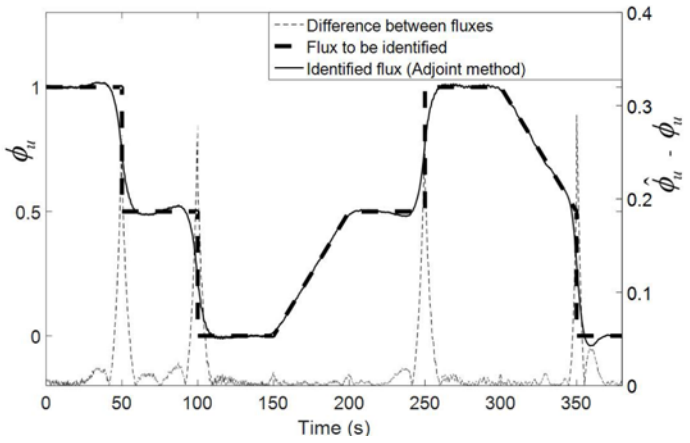


Figure 7 : Heat flux identification with adjoint method

Obtaining such satisfying results without filtering can be explained by the fact that the estimated flux is issued from a minimization including the entire temperature variation. In contrast, in Beck's method only the next time step is used to estimate the flux at a given time, which makes this technique much more sensitive to sudden changes and noise measurements.

Conclusion

The study first of all showed the interest of using low order models in inverse problems, as the loss of information generated by the reduction remains below the noise measurements. Regarding the comparison of the two inverse techniques used in this paper, results showed the difficulty in obtaining correct results with Beck's method. In fact, the sequential aspect of this method does not filter the errors directly obtained from the measurement which are amplified significantly during the flux identification process. A solution is possible, however, but at the cost of additional low frequency filtering, which eliminates the sequential aspect of this technique. The effectiveness of the adjoint method was shown, since very satisfactory results were obtained, with no obligation to use any additional filtering or penalty term functional. This method, more comprehensive, naturally filters the noise during the functional minimization process. The price to pay is that the adjoint method requires more computation time (750s) that Beck's method (115s). These are very encouraging results, paving the way for online identification, both by a search for the minimum acceptable reduction of the modal model, and by the development of a more appropriate adjoint technique (order 2 descent method, temporal sliding window).

Références

[1] O. Quéméner, F. Joly, A. Neveu (2009), On-line heat flux identification from a rotating disk at variable speed, *Int. J. of Heat and Mass Transfer* 53, 1529-1541

[2] S. Chantasiriwan (2001), An algorithm for solving multidimensional inverse heat conduction problem, *Int. J. of Heat and Mass Transfer* 44, 3823-3832

[3] R.A. Khachfe, Y. Jarny (2001), Determination of heat sources and heat transfer coefficient for two-dimensional heat flow - numerical and experimental study. *Int. J. of Heat and Mass Transfer* 1309-1322

[4] A. Vergnaud, L. Perez, L. Autrique (2016), Quasi-online parametric identification of moving heating devices in a 2D geometry, *Int. J. of Thermal Sciences*, 102, 47–61

[5] O. Quéméner, A. Neveu, E. Videcoq (2006), A specific reduction method for branch modal formulation : Application to highly non linear configuration, *Int. J. of Thermal Sciences* 46, 890-907.

[6] O. Quéméner, F. Joly, and A. Neveu (2012), The generalized amalgam method for modal reduction, Int. J. Heat and Mass Transfer, 55, 1197–1207

A computational method for the identification of plastic zones and residual stress in elastoplastic structures.

Thouraya Nouri Baranger[1,a)] and Stephane Andrieux[2,b)]

[1]Université de Lyon, CNRS, LMC2 - Université Lyon1, 69622 Villeurbanne, France

[2]ONERA, Chemin de la Hunière, BP 80100, 91123 Palaiseau, France.

[a)]Corresponding author: thouraya.baranger@univ-lyon1.fr

[b)]Presenting author: stephane.andrieux@onera.fr

ABSTRACT

Two inverse problems arising in the context of isothermal elastoplasticity with small strains are dealt with in this paper. Both of them use as input data fullfield displacement measurements on a stress-free part of the boundary at various time increments during the loading and unloading steps of a structure. In the first one, the recovery of the plastic strain fields (and then of the plastic zones) during the process of loading is addressed, whereas in the second one the residual stress field after complete unloading is looked for. The computational method derived here is grounded on the minimization of an error in constitutive equations. An illustration of the performance and accuracy of the fields recovery is given for each type of inverse problem on a L-shaped structure.

Keywords: Inverse problem, plasticity, residual stress, identification, computational method.

Introduction

The problem of exploiting (measured) boundary data on a part of a solid (displacement and stress vector fields) in order to extend the mechanical fields within the solid, or to identify missing or unknown boundary conditions is still partially open but potential applications are extremely numerous in mechanical and material sciences and in industry as well. One promising approach dealing with this problem is first to reformulate it within the continuous framework, taking advantage of the fact that the amount and spatial density of information gained allows to consider that the complete displacement field is available on a part of the boundary. And secondly to reformulate it then as a Cauchy problem taking into account the fact that an overspecified data pair is given on a part of the boundary. Cauchy problems belong to the class of inverse problems and are usually ill-posed in the sense of [1].

In this paper, advantage is taken of the information available on a part of the boundary of a structure in order to set up two inverse or identification problems. Both of them use as input data full-field displacement measurements on a stress-free part of the boundary at various time increments during the loading and unloading steps of a structure. In the first one, the recovery of the plastic strain fields (and then of the plastic zones) during the process of loading is addressed, whereas in the second one the residual stress field after complete unloading is looked for. The computational method derived here is grounded on the minimization of an error in constitutive equations, and extends previous methods dedicated to linear and non linear elasticity ([7][5]).

The identification problems

Let be given a regular domain Ω, the boundary of which is decomposed into three non overlapping parts Γ_m, Γ_b, and Γ_u. On Γ_b the stress vector $\mathbf{b}$ is prescribed. Γ_m (the subscript m stands for "measurements") is the part where, thanks to DIC's acquisition for example, both displacements U^m and stress vectors F^m (usually zero for the latest) are available, and make up an overspecified boundary data pair. The remaining part Γ_u of the boundary where not any boundary data is known is generally non connected and can possibly contain internal surfaces such as cracks or boundaries of cavities and inclusions.

$$\partial\Omega = \Gamma_b \cup \Gamma_m \cup \Gamma_u \quad \Gamma_i \cap \Gamma_j = \varnothing \quad i, j = m, b, u \tag{1}$$

If the material is elastoplastic and can be described within the framework of isothermal small strains and Generalized Standard Materials [9], the constitutive equation is written in the following incremental format, when choosing the Euler implicit scheme for the time discretization:

$$\left\{\begin{array}{l} \sigma+\Delta\sigma=\frac{\partial W}{\partial \varepsilon}(\epsilon+\Delta\varepsilon-\varepsilon^p-\Delta\varepsilon^p,\alpha+\Delta\alpha), \\ \sigma+\Delta\sigma\in\partial_{\dot{\varepsilon}^p}\Psi(\Delta\varepsilon^p,\Delta\alpha;\varepsilon^p+\Delta\epsilon^p,\alpha+\Delta\alpha) \\ A+\Delta A=-\frac{\partial W}{\partial \alpha}(\varepsilon+\Delta\varepsilon-\varepsilon^p-\Delta\varepsilon^p,\alpha+\Delta\alpha), \\ A+\Delta A\in\partial_{\dot{\alpha}}\Psi(\Delta\varepsilon^p,\Delta\alpha;\varepsilon^p+\Delta\varepsilon^p,\alpha+\Delta\alpha) \end{array}\right. \tag{2}$$

where $\sigma,\varepsilon,\varepsilon^p,\alpha$ are respectively the Cauchy stress tensor, the strain tensor, the plastic strain, and the supplementary internal variables associated with thermodynamic forces A. The potential W is the free energy and Ψ is the positively 1-homogeneous pseudo-potential of dissipation, both are convex functions. $\partial_x\Psi$ stands for the sub-differential of Ψ with respect to x. For the sake of simplicity we shall drop the arguments of the potentials.The two inverse or identification problems can be formulated as follows :

(IP1) *Provided the data* (U^m,F^m) *on* Γ_m*, and* $\boldsymbol{b}$ *on* Γ_b *at various time instants during the loading and unloading of the solid* ($t\in[0,D]$) *are given, to determine the plastic strain field* $\varepsilon^p(\boldsymbol{x},t)$ *within the solid along the process.*

(IP2) *Provided the data* (U^m,F^m) *on* Γ_m*, and* $\boldsymbol{b}$ *on* Γ_b *at various time instants during the loading and unloading of the solid* ($t\in[0,D]$) *are given, to determine the residual stress field* $\sigma^{res}(\boldsymbol{x})=\sigma(\boldsymbol{x},D)$ *within the solid at the end of the process.*

Reformulation as a Cauchy problem

In order to solve problems (IP1) and (IP2), a formulation can be done to recast them into the framework of a Cauchy problem just by considering the conditions that the time increments $(\Delta u,\Delta\sigma,\Delta\varepsilon^p,\Delta A)$ have to fullfil,namely the the incremental evolution equations within the solid. The Cauchy problem is in looking directly for these increments. Then the plastic strain fields at various time instants and the residual stress field can be simply recovered. The Cauchy Problem for incremental plasticity is the following :

(CP) *Provided the data* (U^m,F^m) *on* Γ_m*, and* $\boldsymbol{b}$ *on* Γ_b *at various time instants during the loading and unloading of the solid* ($t\in[0,D]$) *are given, to determine the incremental fields* $(\Delta u,\Delta\sigma,\Delta\varepsilon^p,\Delta\alpha,\Delta A)$ *fulfilling*

$$\left\{\begin{array}{l} div\,[\sigma+\Delta\sigma]=0\ ,\quad \varepsilon(u+\Delta u)=[\nabla(u+\Delta u)]^{sym} \\ \sigma+\Delta\sigma=\frac{\partial W}{\partial \varepsilon}\ ,\ A+\Delta A=-\frac{\partial W}{\partial \alpha} \\ \sigma+\Delta\sigma\in\partial_{\dot{\varepsilon}^p}\Psi(\Delta\varepsilon^p,\Delta\alpha)\,,\ A+\Delta A\in\partial_{\dot{\alpha}}\Psi(\Delta\varepsilon^p,\Delta\alpha) \\ \Delta u=\Delta U^m,\ \Delta\sigma.n=\Delta F^m\ on\ \Gamma_m,\quad \Delta\sigma.n=\Delta b\ on\ \Gamma_b \end{array}\right. \tag{3}$$

Cauchy Problems solution is extensively studied in the literature but mainly for linear operators (Lamé operator for linear elasticity, Laplace equation for conductivity problems,Stokes equation for fluids etc.). Existence of solution for non linear Cauchy problem have been studied also by Leitao *et al.* ([2] [3]) by a constructive method using a fixed point algorithm similar to the one designed by Kozlov *et al.* [4]. Here an extension of the variational method previously designed by the authors for linear and nonlinear elasticity is developed for dissipative solids governed by an elastoplastic constitutive relation described in the Generalized Standard Materials format.

A computational method for solving the Cauchy problem in plasticity

The general method for solving this problems relies on two steps. First, two families of auxiliary usual incremental problems $\Delta\mathcal{P}_1$ and $\Delta\mathcal{P}_2$ are defined, each one using one only of the overspecified boundary data on Γ_m and a given normal stress vector field η over $[0,D]$ on Γ_u :

$$\left\{\begin{array}{l} div\,[\sigma+\Delta\sigma_i]=0\ ,\quad \varepsilon(u+\Delta u_i)=[\nabla(u+\Delta u_i)]^{sym} \\ \sigma+\Delta\sigma_i=\frac{\partial W}{\partial \varepsilon}\ ,\ A+\Delta A_i=-\frac{\partial W}{\partial \alpha} \\ \sigma+\Delta\sigma_i\in\partial_{\dot{\varepsilon}^p}\Psi(\Delta\varepsilon_i^p,\Delta\alpha_i) \\ \Delta\sigma_i.n=\Delta b\ on\ \Gamma_b \end{array}\right. \qquad for\quad i=1,2 \tag{4}$$

and respectively for $\Delta\mathcal{P}_1$ and $\Delta\mathcal{P}_2$:

$$(\Delta\mathcal{P}_1)\left\{\begin{array}{l} \Delta u_1=\Delta U^m\ on\ \Gamma_m \\ \Delta\sigma_1.n=\Delta\eta\ on\ \Gamma_u \end{array}\right. \qquad (\Delta\mathcal{P}_2)\left\{\begin{array}{l} \Delta\sigma_2.n=\Delta F^m\ on\ \Gamma_m \\ \Delta\sigma_2.\,n=\Delta\eta\ on\ \Gamma_u \end{array}\right. \tag{5}$$

If a an incremental surface traction field $\Delta\eta_{opt}$ on Γ_u is such that $\Delta u_1 = \Delta u_2 + \textit{Rigid Body Motion}$, the two problems $\Delta\mathcal{P}_1$ and $\Delta\mathcal{P}_2$ will have the same solution $(\Delta\sigma, \Delta\varepsilon^p, \Delta\alpha)$. Therefore the Cauchy Problem is solved with $(\Delta u_1, \Delta\sigma, \Delta\varepsilon^p, \Delta\alpha, \Delta A)$. A general variational solution method can thus be derived by a second step consisting in building an error function $\mathcal{E}$ between the state increments $\Delta e_1 = (\Delta u_1, \Delta\sigma_1, \Delta\varepsilon_1^p, \Delta A_1)$ and $\Delta e_2 = \Delta u_2, \Delta\sigma_2, \Delta\varepsilon_2^p, \Delta A_2)$ as a functional of $\Delta\eta$ and by minimizing it over all the possible surface traction fields increments defined on Γ_u.

Owing to the general form of the constitutive equation and taking advantage of the convexity of the functions W and Ψ, two errors can be derived with suitable properties ([6][7][8]). They are positive quantities and whenever they vanish then the distance between the two state variable increments vanishes together with the distance of their dual counterparts.

$$\begin{cases} \mathcal{E}_W(\Delta\sigma_1, \Delta\varepsilon_1; \Delta\sigma_2, \Delta\varepsilon_2) = (\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\varepsilon_1^e - \Delta\varepsilon_2^e) - (A_1 - A_2).(\Delta\alpha_1 - \Delta\alpha_2) \\ \mathcal{E}_\Psi(\Delta\sigma_1, \Delta\varepsilon_1^p; \Delta\sigma_2, \Delta\varepsilon_2^p) = (\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\varepsilon_1^p - \Delta\varepsilon_2^p) + (A_1 - A_2).(\Delta\alpha_1 - \Delta\alpha_2) \end{cases} \tag{6}$$

A parametrization enables to put different weights on the errors in stored energy and dissipated one, but outstandingly the value of the parameter, that balances exactly between free energy error and dissipated one, leads to what can be called the Drücker error [**?**]. It involves only the stress and strain tensors, and is then an error in mechanical energy.

$$\mathcal{E} = \frac{1}{2}(\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\epsilon_1 - \Delta\epsilon_2). \tag{7}$$

We can then define the general error functional to be minimized in order to get the solution of the Cauchy problem, and then to design the solution method for (IP1) and (IP2):

$$\Delta\eta_{opt} = ArgMin\left[\mathcal{J}_\chi(\Delta\eta)\right] \quad with \quad \mathcal{J}_\chi(\Delta\eta) = \int_\Omega \mathcal{E}_\chi(\Delta e_1(\Delta\eta)), \Delta e_2(\Delta\eta)))d\Omega \tag{8}$$

The Drücker error can be computed by boundary integration on the whole external surface of the body, thanks to the virtual power principle. This feature has been largely exploited previously to improve the global performance of the solution algorithm for linear Cauchy problems, see [5].

Illustration

The computational method was implemented for both problems on a L-shaped structure submitted to an increasing then a decreasing loading. The overspecified Cauchy data were taken on a part of the right side boundary, whereas the unknown data are located on the top boundary and the left side one. The identification of plastic strain field (IP1) was carried out at various steps of the loading. Figure 1 shows the identified equivalent plastic strain compared to reference values. The identification of the residual stress field (IP2) is carried out at the unloading step. Figure 2 show the identified Von Mises Stress compared to reference one. Let us point out that this result derives from the very good identification of the plastic strain field at the onset of unloading (IP1). Indeed, the residual stress field results directly from the geometric incompatibility of the residual plastic strains within the solid. Because the unloading phase is totally elastic, the residual plastic strain field is exactly the same than at the onset of unloading.

Conclusion

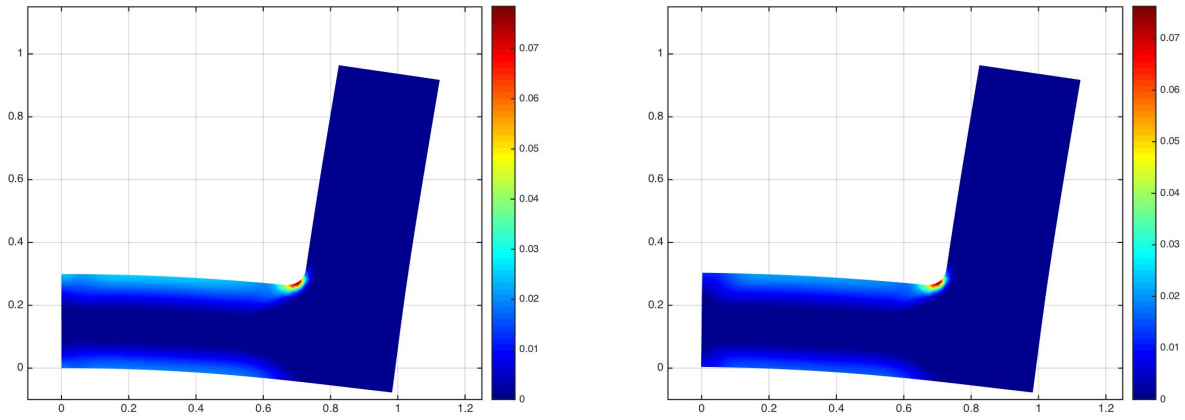
We presented in this paper a computational method for the identification of plastic strains fields and plastic zones during the loading process of a structure, and residual stress field after unolading. The method relies on the solution of a nonlinear Cauchy Problem solved by using a specially designed error in constitutive equation between the solutions of two well-posed problems and minimizing it.

Some improvements have still to be made in the computation of the gradient of the error functional for the case of non twice differentiable potentials for which the general adjoint method can not be directly applied. It is generically the case in elastoplasticity (for the pseudo-potential of dissipation).

References

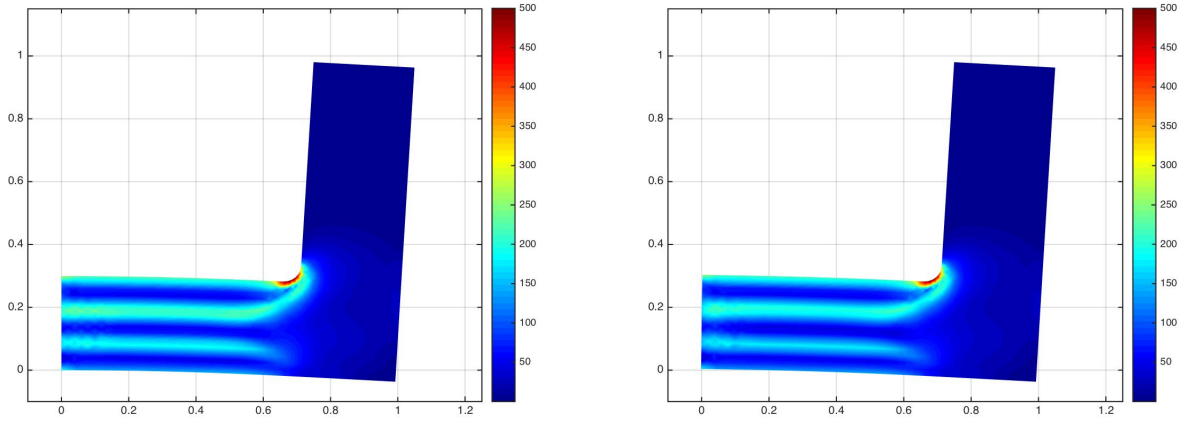
[1] J. Hadamard (1953) *Lectures on Cauchy's problem in Linear Partial Differential Equation.* New York: Dover.

[2] Egger, H., Leitão, A., (2009). Nonlinear regularization methods for ill-posed problems with piecewise constant or strongly varying solutions. Inverse Problems 25 (11), 115014.



(a) Reference value. (b) Identified value.

Figure 1: Equivalent plastic strain at the end of loading step.



(a) Reference value. (b) Identified value.

Figure 2: Von Mises stress at the end of unloading step.

[3] Klüger, P., Leitao, A., 2003. Mean value iterations for nonlinear elliptic Cauchy problems. Numer. Math. 96 (2), 269–293.

[4] Kozlov, V. A., Maz'ya, V. G., Fomin, A. F., January 1992. An iterative method for solving the Cauchy problem for elliptic equations. Comput. Math. Math. Phys. 31, 45–52.

[5] Baranger, T. N., Andrieux, S., 2011. Constitutive law gap functionals for solving the Cauchy problem for linear elliptic PDE. Applied Mathematics and Computation 218 (5), 1970 – 1989.

[6] Andrieux, S. Baranger, T. N. (2016) On the determination of missing boundary data for solids with nonlinear material behaviors, using displacement fields measured on a part of their boundaries. *Journal of the Mechanics and Physics of Solids* **50**, 937-951.

[7] Andrieux, S., Baranger, T. N., 2015. Solution of nonlinear Cauchy problem for hyperelastic solids. Inverse Problems 31 (11), 115003–115022.

[8] Baranger T. N., Andrieux S., and Dang T. B. T. The incremental Cauchy problem in elastoplasticity: General solution method and semi-analytic formulae for the pressurised hollow sphere. *Comptes Rendus Mécanique*, 343(56):331 – 343, 2015.

[9] Halphen B. and Nguyen Q. S. Sur les matériaux standard généralisées. *Journal de Mecanique*, 14:3963, 1975

Multi-model finite element approach for stress analysis of composite laminates

U.N. Band[1] and *†Y.M. Desai[2]

[1,2]Civil Engineering Department, Indian Institute of Technology Bombay, Mumbai, India.

*Presenting author: desai@civil.iitb.ac.in

†Corresponding author: desai@civil.iitb.ac.in

Abstract

A multi-model global-local approach to study free edge effects in laminated composites subjected to uniaxial in-plane loads is presented in this paper. Mixed layer-wise (LW) finite element (FE) model is used in critical free edge zone. Remaining part of plate is modelled by using higher order equivalent single layer (ESL) theory. A transition element is developed to ensure a compatibility between differently modelled subdomains. This combined model possesses traits of both ESL and LW mixed models. Higher order ESL predicts global parameters efficiently, on the other hand, mixed LW model captures the interlaminar stresses at local zones. Mixed LW model includes the transverse stresses as nodal degrees of freedom (DOF) ensuring continuity of the transverse stresses over layer interfaces without using any additional stress functions. Both, ESL and LW mixed models are developed by using three dimensional (3D) elasticity relationships and principle of minimum potential energy. The present combined model is a good blend of computational efficiency and accuracy in prediction of local transverse stresses. Plates with different stacking sequences are investigated for free edge stresses developed in the transverse direction under uniaxial in-plane load conditions.

Key Words: Mixed Finite Element; Free edge stresses; Higher order theory; Principle of minimum potential energy; transition element; global-local analysis.

1.0 Introduction

Laminated composites having several layers with uni-directional fibres are utilized as structural members for variety of applications. Advantageously, these exhibit good strength, stiffness, environmental resistance and are light in weight as compared to homogeneous materials. Depending on configuration of loading, strength parameters can be altered by using appropriate stacking sequence of layers. Evaluation of laminate response to applied load becomes complex due to heterogeneous properties of different layers in a laminate.

Apart from elasticity approach, various displacement based and hybrid models have been proposed for analysis of laminates. These models are implemented using analytical or FE formulations. A three-dimensional (3D) elasticity solution by Pipes and Pagano (1970) [1] has shown that in a laminate under simple uniaxial loading there is a "*boundary layer*" region along the free edges where a three-dimensional state of stress exists, and that the boundary layer thickness is roughly equal to laminate thickness. Wang and Crossman (1977) [2] presented a displacement based FE model to study edge effects for symmetrically stacked laminates. It has been shown that steep stress gradients of the transverse normal and shear stresses prevail near free edges. These high magnitudes of multi-axial stresses in vicinity of free edges may lead to delamination of a laminate. A state of plane stress is seen to prevail towards the interior of plate. Moreover, delamination failure is most common mode of failure in laminated composites, which initiates at geometrical discontinuities like free edges, notches and holes.

Evidently, a correct evaluation of complete 3D state of stress at free edges is important for assessment of strength and durability of a laminate under a certain load configuration. Effect of stacking sequence on laminate strength was investigated by Pagano and Pipes (1971) [3]. Rybicki (1971) [4], Wu and Hsu (1993) [5], Flesher and Herakovich (2006) [6] presented different approaches for evaluation of the transverse stresses and prediction of onset of delamination. Shi and Chen (1992) [7]presented a mixed FE model by using a hybrid stress element at free edges and conventional displacement based FE's at other locations. Chorng-Fuh and Horng-Shian (1993) [8] also presented a mixed FE model to predict the transverse stresses developed at free edges of a laminate subjected to uniform in-plane strain

A displacement model depicting the kinematics of a particle in a laminate must encompass rigid body, extension, bending and warping modes of deformation to correctly predict response in a realistic manner. Many ESL models are seen in literature for analysis of laminated composites. Kant and Swaminathan (2002) [9] presented a comprehensive ESL higher order theory which incorporates all these deformation modes and predicts all global responses effectively. Laminate is considered as a single smeared plate with the properties averaged over thickness. However, evaluation of interlaminar transverse stresses is done by using 3D stress equilibrium equations. On the other hand, a better mathematical representation of laminate behaviour is portrayed by LW models which incorporate discrete individual properties of all layers in a laminate. Displacement based LW models also need either some additional stress function or integration of stress equilibrium equations to estimate magnitudes and through thickness variation of the transverse stresses in a laminate. Ramtekkar, Desai (2002) [10] presented a FE mixed LW formulation having the transverse stresses invoked as nodal DOF along with displacements. Continuity of the transverse stresses over layer interfaces is inherently satisfied. ESL's demand less computational effort as compared to LW models as they map the domain involving less DOF. Computational efficiency is achieved by using ESL but accuracy of solution is sacrificed. LW models exhibit accuracy of solution but demand high computational effort. Application of LW models on a laminate domain involve high DOF in the solution and face restrictions due to limitation of computational resources in cases where fine discretization of domain becomes essential for accuracy of solution.

In this paper a multi-model meshing methodology is presented which advantageously uses both higher order ESL and mixed LW models simultaneously over the domain of a laminate. A transition element is developed to establish compatibility between two models. Presence of ESL in non-critical zones in a laminate ensures accurate assessment of global parameters and reduction of computational cost. At the same time, mixed LW model used in critical free edge region accurately predicts the transverse stresses. Efficacy of present multi-model approach is illustrated by using it on examples of laminates subjected to uniaxial in-plane loading.

2.0 Theoretical formulation

Three models have been formulated for analysis of laminated composite plates consisting of several orthotropic laminae.

(a) **Model 1**: This model adopts a cubic displacement field in the thickness direction for displacements (U,V,W) and has 12 DOF. The theory has been identified as HOST12. The model is based on the three dimensional state of stresses and strains.

(b) **Model 2**: In this model, mixed finite element LWT, which has three displacements (U,V,W) and the transverse stresses ($\tau_{xz}, \tau_{yz}, \sigma_z$) as the nodal DOF, is used. The theory

is based on elasticity relationships. Therefore, introduction of any additional parameters/stress variation functions are advantageously avoided.

(c) **Model 3**: This model is based on a global-local finite element procedure to take advantage of computational efficiency of the higher order ESL theory and accuracy of the 3D mixed model.

2.1 Model 1 : Development of ESL theory based model (HOST 12)

Displacements in three principal directions of the laminate as a fully cubic function of the thickness co-ordinate are

$$\begin{aligned} u(x,y,z) &= u_0(x,y) + z\theta_x(x,y) + z^2 u_0^*(x,y) + z^3\theta_x^*(x,y) \\ v(x,y,z) &= v_0(x,y) + z\theta_y(x,y) + z^2 v_0^*(x,y) + z^3\theta_y^*(x,y) \\ w(x,y,z) &= w_0(x,y) + z\theta_z(x,y) + z^2 w_0^*(x,y) + z^3\theta_z^*(x,y) \end{aligned} \quad (1)$$

The above displacement field eliminates any requirement of shear correction factor and chances of shear locking. Here u_0, v_0 and w_0 are the deformations in the x,y,z (laminate co-ordinate) directions respectively at the mid-plane. θ_x, θ_y and θ_z, on the other hand, are the rotations at mid-plane about the principal directions of laminate. $u_0^*, \theta_x^*, v_0^*, \theta_y^*, w_0^*$ and θ_z^* are higher order terms stemming from the Taylor's series. By using material property, the strain displacement relationship and the principle of minimum potential energy, the stiffness matrix for laminate is developed. By using shape functions similar to the stiffness evaluation, the mass matrix is also developed. Detailed formulation can be seen in the work presented by Kant and Swaminathan (2002) [9]. A nine node Lagrangian isoparametric element has been used to discretize a laminate.

Numerical integration is performed by employing 3 X 3 Gauss quadrature rule for the extension, bending, mass component, whereas, 2 X 2 Gauss rule for the shear part.

2.2 Model 2: Development of mixed LW model

An 18-node three-dimensional element based on mixed formulation is used by considering displacement fields $u(x,y,z)$, $v(x,y,z)$ and $w(x,y,z)$ having quadratic variation along the plane of plate and cubic variation in the transverse direction. The cubic variation of field has been adopted to invoke the transverse stresses as the nodal parameters in addition to the nodal deformations. The displacement field is expressed as

$$u_k(x,y,z) = \sum_{i=1}^{3}\sum_{j=1}^{3} g_i h_j a_{0ijk} + z\sum_{i=1}^{3}\sum_{j=1}^{3} g_i h_j a_{1ijk} + z^2\sum_{i=1}^{3}\sum_{J=1}^{3} g_i h_j a_{2ijk} + z^3\sum_{i=1}^{3}\sum_{j=1}^{3} g_i h_j a_{3ijk} \quad (2)$$

where

$$g_1 = \frac{\xi}{2}(\xi - 1),\ g_2 = 1-\xi^2,\ g_3 = \frac{\xi}{2}(1+\xi), \quad \xi = x/L_x$$

$$h_1 = \frac{\delta}{2}(\delta - 1),\ h_2 = 1-\delta^2,\ h_3 = \frac{\delta}{2}(1+\delta), \quad \delta = y/L_y$$

$k = 1, 2, 3$ and $u_1 = u; \quad u_2 = v; \quad u_3 = w;$

Further, a_{mijk} ($m = 0, 1, 2, 3$; $i, j, k = 1, 2, 3$) are the generalized coordinates.

Variation of displacement fields has been assumed to be cubic through the thickness of element, although there are only two nodes along 'z' axis of an element. Derivative of displacement with

respect to the thickness coordinate has also been included in the displacement field. Such a variation is required for invoking transverse stress components σ_z, τ_{xz} and τ_{yz} as nodal DOF in the present formulation. Further, it also ensures quadratic variation of the transverse stresses through the thickness of an element.

By making use of the elasticity relationship and introducing derivative of displacements, displacement field $u_k(x,y,z)$) in Eq. (2) becomes

$$u_k(x,y,z) = \sum_{n=1}^{18} g_i h_j (f_q u_{kn} + f_p \widehat{u}_{kn}) \tag{3}$$

Here, $i = 1, 2, 3$ for the nodes with $\xi = -1$, $\xi = 0$ and $\xi = 1$, respectively;
$j = 1, 2, 3$ for the nodes with $\delta = -1$, $\delta = 0$ and $\delta = 1$, respectively;

$q = 1,2$ and $p = 3,4$ for the nodes with $\eta = -1$ and $\eta = 1$, respectively for node numbers 1 to 18 and

$$f_1 = \frac{1}{4}(2-3\eta+\eta^3);\ f_2 = \frac{1}{4}(2+3\eta-\eta^3);\ f_3 = \frac{L_z}{4}(1-\eta-\eta^2+\eta^3);\ f_4 = \frac{L_z}{4}(-1-\eta+\eta^2+\eta^3).$$

Here, f_3 and f_4 correspond to derivative of displacements with respect to thickness co-ordinate whereas f_1 and f_2 correspond to displacement DOF , u_{kn} ($k = 1, 2, 3$ and $n = 1,2,3,...18$) are nodal displacement variables, whereas $\widehat{u}_{kn}$ ($= \frac{\partial u_{kn}}{\partial z}$) contains the nodal transverse stress variables. Principle of minimum potential energy is used to develop the element property matrix. Detailed formulation can be seen in the work presented by Ramtekkar, Desai (2002) [10].

Numerical integration of system matrices has been performed by using Gauss quadrature rule with 3 X 3 integration scheme in plane of plate and a 5 X 5 integration scheme in the thickness direction.

2.3 Model 3 - Development of transition element between 2D ESL (HOST12) and 3D mixed LW model

Compatibility between two differently modelled sub-domains (by using Model 1 and Model 2) is enforced by degenerating a continuum 3D element through kinematic constraints compatible with deformations predicted by 2D element.

A 3D-to-2D transition element has one or two faces of a 3D element that are kinematically restrained to enforce compatibility with adjacent 2D elements. Such a face is denoted as a transition face in the sequel. The 3D element on the transition face needs to be conditioned for compatibility with DOF of the ESL (HOST12) element to ensure continuity of the combined model. Such an element acts as a transition element to connect two independently modelled sub-domains. Transition is achieved by placing a stack of such transition elements used in different layers of a laminate at the transition face.

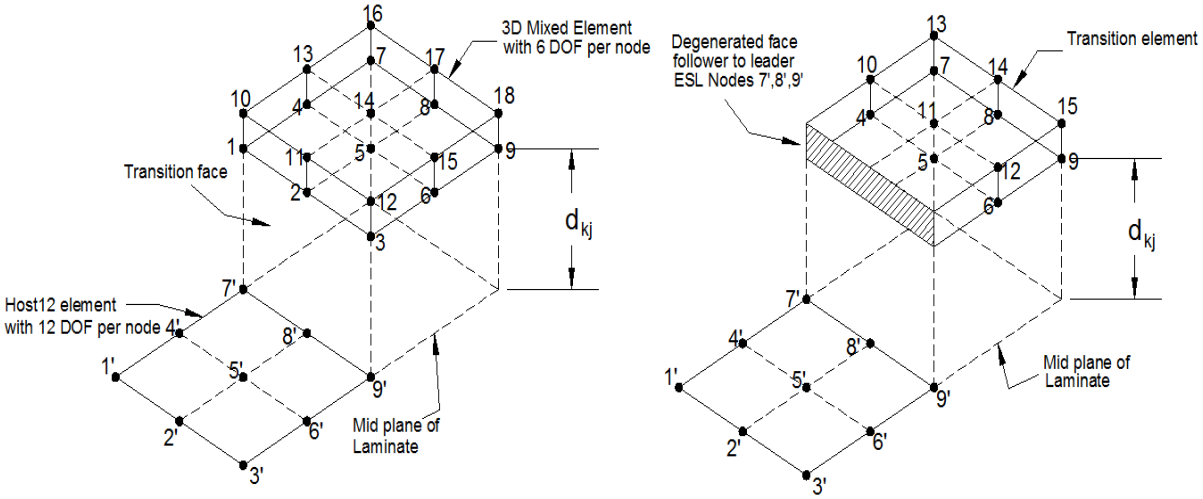


Fig. 1 (a) Configuration of connection between 3D elements and HOST12 elements, and (b) Illustration of degenerated face of the 3D element

A pair of incompatible mesh formulations is shown in Fig. 1(a) wherein a nine node ESL element with twelve DOF per node (node numbers denoted with a prime) is connected to a stack of 3D mixed elements with six DOF per node (three translations and three transverse stresses). Fig. 1(b) shows diagrammatic representation of the transition element with the degenerated transition face. Differently modelled meshes meeting at the transition face represent the same laminate configuration and thickness.

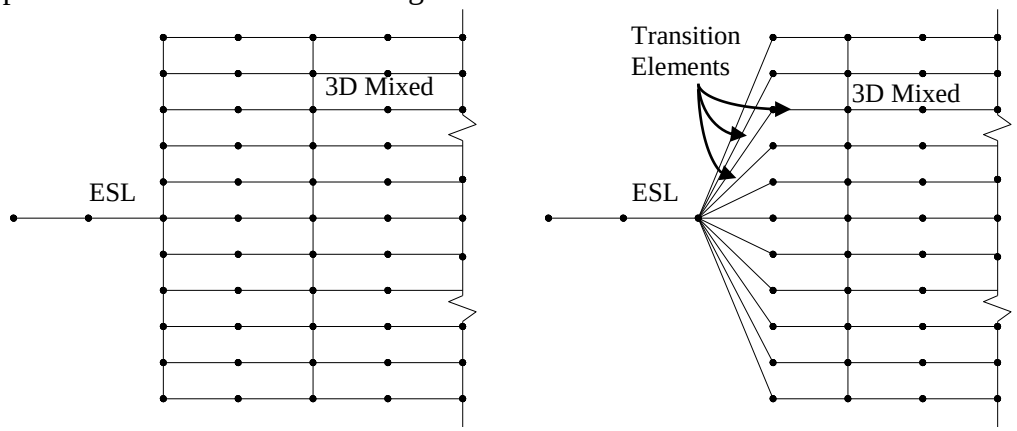


Fig. 2 An indicative impression of unidirectional transition (a) before implementation of restraint; and (b) after implementation of restraint

Kinematics of any point at a distance 'd_{kj}' from the reference plane of the laminate on the transition face is completely described by displacement field for the ESL. Because 2D elements and stack of 3D elements represent the same laminate, motion of the corner 3D node (node 1) (refer Fig. 1(a)) is entirely prescribed by the three translations, three rotations and the higher order terms of its corresponding ESL node (node 7'). Consequently, the DOF associated with nodes 1,2,3,10,11 and 12 are followers to the DOF associated with ESL leader nodes 7', 8' and 9', and hence must be restrained. A transition element is shown in Fig.1(b), where kinematic restraint is imposed on the hatched surface. Three nodes of the ESL on the transition face form 3D elements' transition edge. This edge represents transition face of 3D element stack. An indicative impression of the change in configuration of the 3D element on imposition of the restraint is shown in Fig. 2.

By using displacement field of HOST12 in Eq. (1), kinematics $\{\hat{q}\}_{3D}^{k}$ of any 'k[th]' node of 3D element on the transition face and corresponding to the ESL leader 'j[th]' node can be completely prescribed as

$$\{\hat{q}\}_{3D}^{k}=\begin{Bmatrix}u\\ \hat{v}\\ w\end{Bmatrix}_{3D}^{k}=\begin{bmatrix}1&0&0&d_{kj}&0&0&d_{kj}^{2}&0&0&d_{kj}^{3}&0&0\\ 0&1&0&0&d_{kj}&0&0&d_{kj}^{2}&0&0&d_{kj}^{3}&0\\ 0&0&1&0&0&d_{kj}&0&0&d_{kj}^{2}&0&0&d_{kj}^{3}\end{bmatrix}\{q\}_{2D}^{j}$$

$$where\ \{q\}_{2D}=\begin{bmatrix}u_0 & v_0 & w_0 & \theta_x & \theta_y & \theta_z & u_0^{*} & v_0^{*} & w_0^{*} & \theta_x^{*} & \theta_y^{*} & \theta_z^{*}\end{bmatrix}^{T} \tag{4}$$

or

$$\{\hat{q}\}_{3D}^{k}=[R]_{kj}\{q\}_{2D}^{j} \tag{5}$$

By developing the restraint sub-matrices $[R]_{kj}$ for all pairs of 2D and 3D nodes, the transformation matrix [R] for the entire element can be formulated by appropriately populating sub-matrices $[R]_{kj}$ corresponding to every pair. Finite element stiffness property, mass/inertia property matrices and internal force/influence vector for the transition element are obtained by matrix transformations using the constructed corresponding matrices of 3D element and associated transformation matrix as follows,

$$\left.\begin{aligned}[K_e]_{Tr}&=[R]^{T}[K_e]_{3D}[R]\\ \{F_e\}_{Tr}&=[R]^{T}\{F\}_{3D}\\ [M_e]_{Tr}&=[R]^{T}[M_e]_{3D}[R]\end{aligned}\right\} \tag{6}$$

The transformation in Eq. (6) degenerates the transition face of the 3D element which becomes follower to the corresponding HOST12 leader nodes. All elements in the interior of the local transition face are 18 node elements with all nodes modelled using mixed formulation. Stress DOF at the 3D nodes on the transition face are condensed prior to imposition of the restraint. By considering stiffness and mass matrices of the ESL elements, transition elements and the interior LW mixed elements, the global matrices are obtained in the following form after assembly.

$$\begin{aligned}[K]^{G}&=\sum_{i=1}^{m}\left[K_e^{i}\right]+\sum_{j=1}^{n}\left[K_e^{j}\right]_{Tr}+\sum_{l=1}^{k}\left[K_e^{l}\right]\\ [M]^{G}&=\sum_{i=1}^{m}\left[M_e^{i}\right]+\sum_{j=1}^{n}\left[M_e^{j}\right]_{Tr}+\sum_{l=1}^{k}\left[M_e^{l}\right]\\ \{F\}^{G}&=\sum_{i=1}^{m}\{F_e^{i}\}+\sum_{j=1}^{n}\{F_e^{j}\}_{Tr}+\sum_{l=1}^{k}\{F_e^{l}\}\end{aligned} \tag{7}$$

Here

$[K]^{G}$, $[M]^{G}$ and $\{F\}^{G}$ are the global stiffness property matrix, inertia property matrix and nodal influence vector, respectively;

$\left[K_e^{i}\right]$, $\left[M_e^{i}\right]$ and $\{F_e^{i}\}$ are the element property matrix, inertia property matrix and the element influence vector of i^{th} element, respectively, formed by using mixed LWT;

$\left[K_e^j\right]_{Tr}$, $\left[M_e^j\right]_{Tr}$ and $\left\{F_e^j\right\}_{Tr}$ are the element property matrix, inertia property matrix and element influence vector of j^{th} transition element, respectively; and

$\left[K_e^l\right]$, $\left[M_e^l\right]$ and $\left\{F_e^l\right\}$ are the element stiffness matrix, mass matrix and element nodal load vector of l^{th} ESL element, respectively.

m, *n* and *k* in Eq. (7) represent number of mixed LW, transition and ESL elements.

The displacement vector $\left\{\hat{q}\right\}_{Tr}$ of a transition element is composed of DOF of ESL nodes on the transition edge, and DOF of 3D nodes on the other faces.

The transition element developed by the application of the restraints consists of 108 DOF and 15 nodes for unidirectional transition and a corner element with two adjacent transition edges has 13 nodes and 108 DOF. Such a corner element is developed by applying the kinematic restraint on two adjacent faces.

3.0 Numerical examples

To study 3D state of stresses in the free edge regions, a laminate is modelled by using 3D mixed LW elements at free edge and higher order ESL in remaining part to reduce computational effort. Both models are implemented simultaneously and compatibility between subdomains is established by introducing transition elements. Examples of symmetrical cross ply laminates under in-plane unidirectional strain and transverse doubly sinusoidal load are considered for illustration. Plate under transverse load is considered to be simply supported on all four edges. Substantial reduction in computational effort is achieved as compared to a complete LW mixed FE solution.

3.1 Example 1: Free edge stress analysis of a symmetric cross ply laminate

A symmetric (0/90/90/0) cross ply laminate is considered for free edge stress analysis under action of uniform uniaxial in-plane strain. Width of laminate '2*b*' is considered as '4*h*' and length of laminate '*l*' is taken as '10*h*', where '*h*' is thickness of laminate. Material of laminae is assumed to possess following properties.
E_1 = 138.00GPa; $E_2 = E_3$ = 9.66 GPa; $G_{12} = G_{13}$ = 5.52 GPa;
G_{23} = 4.14 GPa; $\nu_{12} = \nu_{13} = \nu_{23}$= 0.21;

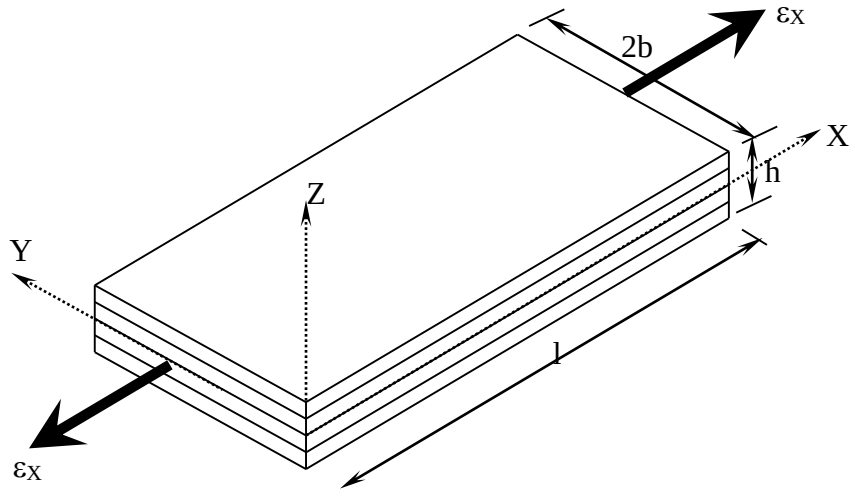


Fig 3 Typical laminate and coordinate axes

Uniform in-plane strain ($\varepsilon_x = 1X10^{-6}$) is introduced along the length of laminate. Implementation of this novel multi-model finite element mesh is done on a quarter part of the laminate. Advantage of symmetry in configuration is taken in implementation of the multi-model FE scheme for a finer discretization. A typical laminate with coordinate axes is shown in Fig 3. Laminate is restrained against deformations along *X* axis at *X=l/2*. A uniform strain is introduced at *X=0*.

Zone in vicinity of *X=0* over entire half width is modelled using a stack of 3D mixed LW elements and in remaining part, higher order ESL (HOST12) is used. Length of local zone (3D mixed LW zone) is taken equal to thickness of laminate. This amounts to about 10% of entire domain of plate. Laminate is discretized using 7 elements along the length and 8 elements along width. A strip of 1 element at free edge along the width is modelled by 3D mixed LW elements. Each layer of laminate is subdivided in 4 sub-layers to accommodate 16 LW mixed 3D elements over the thickness at local free zone. Hence, a total of 176 elements are employed over the domain of laminate. Composition of these elements comprises of 56 ESL and 128 3D mixed LW elements.

Variation of the transverse normal stress at free edge (*X=0*), along the half width of plate is obtained by present multi-model approach. Variation of the transverse normal stress at mid-plane (90-90 interface) and at (90-0) interface are presented in Fig 4 and Fig 5, respectively. Variation of the transverse shear stress (τ_{yz}) at (90-0) interface along the width at free edge is shown in Fig 6.

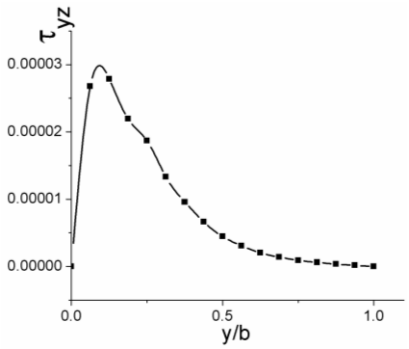


Fig 4 Width wise variation of σ_z on free edge at *X=0* at mid-plane

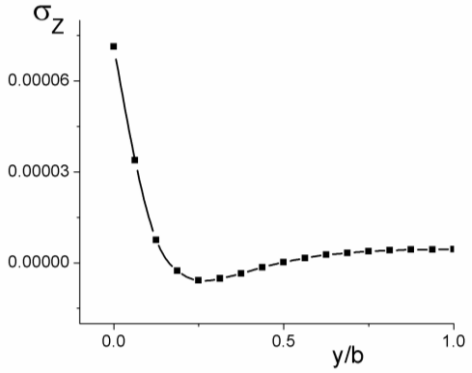


Fig 5 Width wise variation of $\overline{\tau}_{yz}$ at free edge at *X=0* at mid-plane

It is observed that the transverse stresses at the free edge are correctly estimated by the present multi-model approach. Steep stress gradient is predicted at free edge. At the same time, a substantial reduction in computational effort is also achieved. Saving in computational effort as compared to complete 3D mixed LW model can be appreciated. For a complete 3D solution with same mesh discretization, a total of 896 elements would have been required. Reduction in number of elements required to map the domain leads to reduction of DOF and therefore, the computational effort.

3.2 Example 2: Complete stress analysis of a square simply supported sandwich plate under bi-directional sinusoidal transverse load (Core=0.8h)

A (0^o/core/0^o) square sandwich plate (*l*=2b) under bi-directional sinusoidal transverse load is considered for in-plane as well as inter-laminar stresses. The plate is simply supported on all

four edges. The thickness of each face sheet is one tenth of total thickness of sandwich plate. Determination of in-plane and the transverse stresses is accomplished using present combined model. To capture (τ_{yz}), a stack of 3D mixed LW elements are placed at and in vicinity of $(\frac{l}{2},0)$ and remaining laminate is modelled using HOST12 elements. To capture (τ_{xz}), a stack of 3D mixed LW elements are placed at and in vicinity of $(0,b)$ and remaining laminate is modelled using HOST12 elements. For obtaining (σ_z), a stack of 3D mixed LW elements are placed at and in vicinity of $(\frac{l}{2},b)$ and remaining laminate is modelled using HOST12 elements. Material properties and normalization factors used for the analysis are mentioned alongside Table 1. Results for aspect ratios *S*=*l*/*h*= 2, 4, 10, and 20 have been compared in Table 1 with elasticity solution given by Pagano (1970) [11], FE solution by Ramtekkar, Desai (2003) [12] as well as the analytical and finite element solutions presented by various authors. Through thickness variations of the normalized transverse shear stress components and transverse normal stress for the plate with aspect ratio *S* = 4 have been presented in Fig. 7(a-c). Results are in close proximity of exact elasticity solution obtained by Pagano (1970) [11], FE solution by Ramtekkar, Desai (2003) [12]. The agreement of the results with the elasticity solution and 3D fully mixed formulation clearly suggests that such problems can be analyzed with good accuracy by using the present formulation. A substantial reduction in DOF and effort as compared to complete mixed LW solution is observed.

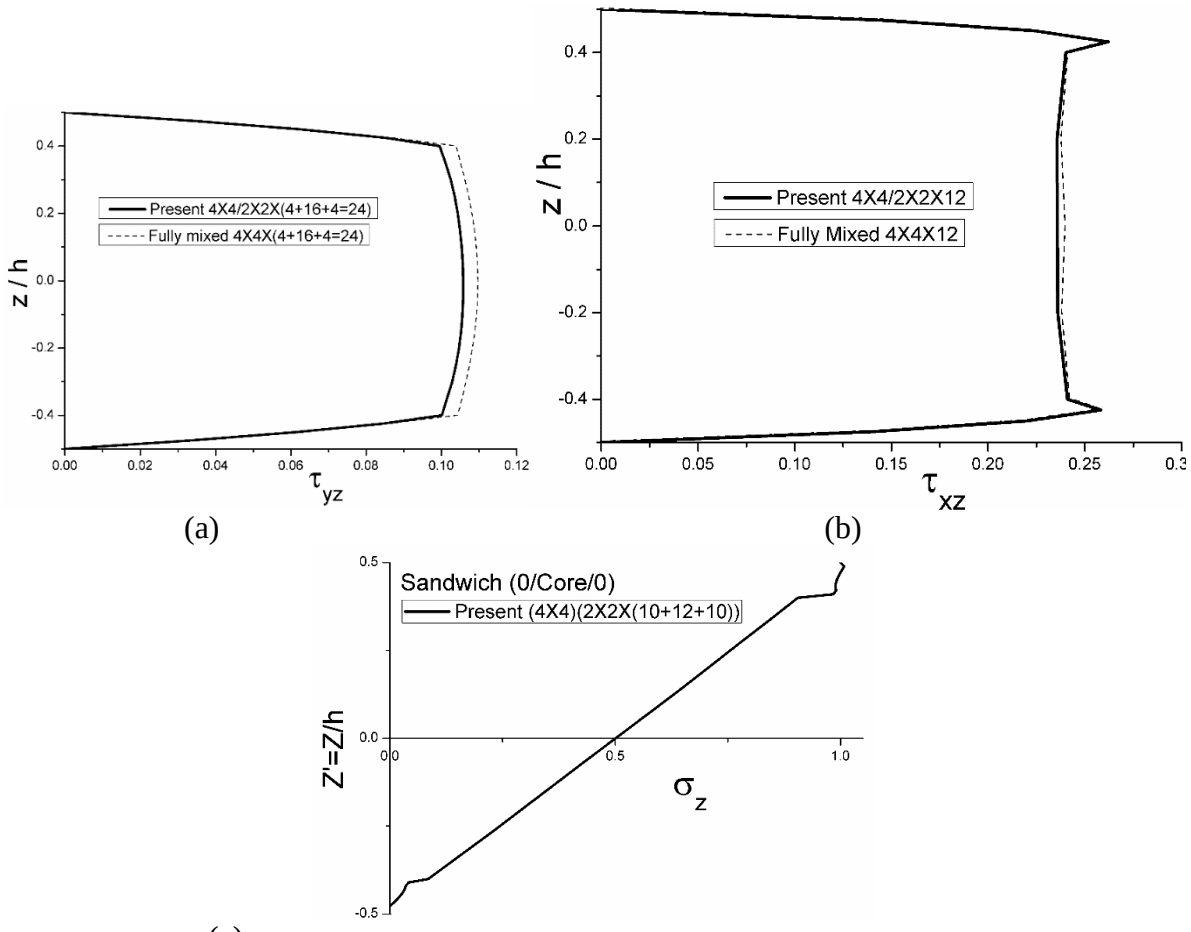


Fig 7 Through thickness variation of a) $\bar{\tau}_{yz}$, b) $\bar{\tau}_{xz}$ and c) σ_z

Table 1 Maximum stresses in square sandwich plate under bi-directional sinusoidal transverse load (Core=0.8h)

($FaceSheet: E_1 = 172.4GPa, E_2 = E_3 = 6.89GPa, G_{12} = G_{13} = 3.45GPa, G_{23} = 1.378GPa, \nu_{12} = \nu_{13} = \nu_{23} = 0.25$;
$Core: E_1 = E_2 = 0.276GPa, E_3 = 3.45GPa, G_{12} = 0.1104GPa, G_{13} = G_{23} = 0.414GPa, \nu_{12} = \nu_{31} = \nu_{32} = 0.25$)

$$\overline{W} = w\,\frac{100E_2h^3}{q_0a^4},\ \left(\bar{\sigma}_X, \bar{\sigma}_Y, \bar{\tau}_{XY}\right) = \frac{\left(\sigma_X, \sigma_Y, \tau_{XY}\right)h^2}{q_0a^2},\ \left(\bar{\tau}_{XZ}, \bar{\tau}_{YZ}\right) = \frac{\left(\tau_{XZ}, \tau_{YZ}\right)h}{q_0a}$$

S		Source	$\bar{\sigma}_X\left(\frac{a}{2}, \frac{b}{2}, \pm\frac{h}{2}\right)$		$\bar{\sigma}_Y\left(\frac{a}{2}, \frac{b}{2}, \pm\frac{h}{2}\right)$		$\bar{\tau}_{XZ}\left(0, \frac{b}{2}, 0\right)$	$\bar{\tau}_{YZ}\left(\frac{a}{2}, 0, 0\right)$	$\bar{\tau}_{XY}\left(0, 0, \pm\frac{h}{2}\right)$	
2	i.	Pagano (1970b) [11]	3.278	-2.653	0.452	0.392	0.185	0.142	-0.240	0.234
	ii.	Present	3.1225	-2.516	0.468	-0.417	0.183	0.136	-0.2328	0.2295
4	i.	Pagano (1970b) [11]	1.556	-1.512	0.259	-0.253	0.239	0.107	-0.144	0.148
	ii.	Present	1.501	-1.460	0.267	-0.263	0.2388	0.1055	-0.1424	0.1474
	iii.	Pandya and Kant (1988) [13]	1.523	-	0.241	-	0.275	-	-0.142	-
	iv.	Reddy and Chao (1981) [14]	0.865	-	0.151	-	0.099	-	-0.088	-
	v.	Wu and Lin (1993) [15]	1.548	-	0.241	-	0.249	-	-0.134	-
	vi.	Ramtekkar, Desai (2003) [12]	1.570	-1.524	0.260	-0.255	0.240	0.108	-0.145	0.149
10	i.	Pagano (1970b) [11]	1.153	-1.152	0.110	-0.110	0.300	0.053	-0.071	0.072
	ii.	Present	1.146	-1.145	0.113	-0.113	**0.306**	**0.058**	-.0707	0.0718
	iii.	Pandya and Kant (1988) [13]	1.166	-	0.105	-	0.340	-	-0.069	-
	iv.	Reddy and Chao (1981) [14]	1.015	-	0.077	-	0.111	-	-0.053	-
	v.	Wu and Lin (1993) [15]	1.210	-	0.111	-	0.324	-	-0.071	-
	vi.	Ramtekkar, Desai (2003) [12]	1.159	-1.158	0.111	-0.110	0.303	0.055	-0.071	0.072
20	i.	Pagano (1970b) [11]	±1.110		±0.070		0.317	0.036	∓ 0.051	
	ii.	Present	±1.115		±0.0729		**0.335**	**0.048**	-0.0512	0.0515
	iii.	Present (8X8)/(2X2X16)	±1.106		±0.0713		**0.3216**	**0.0393**	-0.0506	0.0503
	iv.	Wu and Lin (1993) [15]	1.173		0.072		0.353	-	0.052	
	v.	Ramtekkar, Desai (2003) [12]	±1.115		±0.070		0.317	0.036	∓0.051	

4.0 Conclusions

A multi-model FE approach is developed for stress analysis of composite laminates. An unique transition element is developed for appropriate compatibility between higher order ESL and 3D mixed LW formulation. The present multi-model approach has been tested over a laminate under uniaxial strain. Results for a transversely loaded simply supported sandwich are also presented. Results obtained through This approach enables mapping of the domain of a laminate with reduced numbers of DOF as compared to any 3D solution. At the same time, accuracy in prediction of inter-laminar stresses at critical zones is also achieved. Reduction in number of DOF renders the methodology a computationally economical.

References

[1] Pipes, R.B. and N. Pagano, Interlaminar stresses in composite laminates under uniform axial extension. Journal of Composite Materials, 1970. 4(4): p. 538-548.
[2] Wang, A. and F.W. Crossman, Some new results on edge effect in symmetric composite laminates. Journal of Composite Materials, 1977. 11(1): p. 92-106.
[3] Pagano, N. and R.B. Pipes, The influence of stacking sequence on laminate strength. Journal of Composite Materials, 1971. 5(1): p. 50-57.
[4] Rybicki, E., Approximate Three-Dimensional Solutions for Symmetric Laminates Under Inplane Loading*. Journal of Composite Materials, 1971. 5(3): p. 354-360.
[5] Wu, C.-P. and C.-S. Hsu, A new local high-order laminate theory. Composite Structures, 1993. 25(1): p. 439-448.
[6] Flesher, N.D. and C.T. Herakovich, Predicting delamination in composite structures. Composites science and technology, 2006. 66(6): p. 745-754.
[7] Shi, Y.-B. and H.-R. Chen, A mixed finite element for interlaminar stress computation. Composite structures, 1992. 20(3): p. 127-136.
[8] Chorng-Fuh, L. and J. Horng-Shian, A new finite element formulation for interlaminar stress analysis. Computers & structures, 1993. 48(1): p. 135-139.
[9] Kant, T. and K. Swaminathan, Analytical solutions for the static analysis of laminated composite and sandwich plates based on a higher order refined theory. Composite structures, 2002. 56(4): p. 329-344.
[10] Ramtekkar, G., Y. Desai, and A. Shah, Mixed finite-element model for thick composite laminated plates. Mechanics of Advanced Materials and Structures, 2002. 9(2): p. 133-156.
[11] Pagano, N., Exact solutions for rectangular bidirectional composites and sandwich plates. Journal of composite materials, 1970b. 4(1): p. 20-34.
[12] Ramtekkar, G., Y. Desai, and A. Shah, Application of a three-dimensional mixed finite element model to the flexure of sandwich plate. Computers & structures, 2003. 81(22): p. 2183-2198.
[13] Pandya, B. and T. Kant, Higher-order shear deformable theories for flexure of sandwich plates-finite element evaluations. International Journal of Solids and Structures, 1988. 24(12): p. 1267-1286.
[14] Reddy, J. and W. Chao, A comparison of closed-form and finite-element solutions of thick laminated anisotropic rectangular plates. Nuclear Engineering and Design, 1981. 64(2): p. 153-167.
[15] Wu, C.-P. and C.-C. Lin, Analysis of sandwich plates using a mixed finite element. Composite Structures, 1993. 25(1): p. 397-405.

Numerical Simulation of Raceway Formation in Blast Furnace

Tyamo Okosun, Guangwu Tang, Dong Fu, Armin K. Silaen, Bin Wu, and †*Chenn Q. Zhou

Center for Innovation through Visualization and Simulation
Purdue University Calumet
2200 169th Street
Hammond, IN 46323

*Presenting author: czhou@purduecal.edu
†Corresponding author: czhou@purduecal.edu

Abstract
Pulverized Coal Injection (PCI)/ Natural Gas (NG) co-injection has a significant impact on the size and shape of the raceway inside a Blast Furnace. Consequently, this affects the gas distributions, as well as iron ore reduction, and the furnace pressure drop. The raceway size and shape are influenced by incoming gas momentum entering the raceway envelope from the tuyere jet, the combustion of coke and injected fuels, coke particle size, fuel injection rates, slag volume, and other complex factors. A 3-D CFD mathematical model has been established for estimating the raceway geometry and combustion inside the blast furnace. This model considers the effects of coke combustion and injection fuels on the raceway geometry and the raceway gas flow patterns. The combustion effects are treated as additional source terms of mass and momentum in the gas phase because the combustion converts solid fuels into the gaseous phase and releases heat. In this paper, the raceway geometry and raceway combustion models are presented, along with the methodologies for raceway simulation, and some selected model applications (including parametric study analyses of blast furnace operation under a variety of fuel injection conditions).

Keywords: Ironmaking, blast furnace, Raceway, PCI, CFD

Introduction

The injection of pulverized coal into the blast furnace is a crucial technology for lowering hot metal production costs and reducing coke consumption. In order to achieve high injection rates and coke replacement ratios, the combustion process of injection fuel and the fluid dynamics inside the raceway must be well understood and documented.

The thermodynamics and kinetics of coal and coke combustions under laboratory conditions are well published, and, based on this fundamental knowledge, mathematical modeling can be conducted to simulate injected fuel and coke combustion inside the blast furnace raceway. Additionally, numerical modeling can be utilized to optimize furnace operation conditions in an effort to achieve higher fuel injection rates and optimal coke replacement ratios. Early in published model development, some simulations of the raceway regions assumed one-dimensional plug flow for the tuyere and the raceway. However, these early models, in many cases, failed to properly predict the de-volatilization process and combustion of released volatiles [1, 2]. In certain two-

dimensional models [3], the combustion of coal and coke in the raceway were considered, however the turbulent features of the gas phase were either ignored or simplified [4]. Additionally, these two-dimensional models were mostly applied to simplified raceway shapes, such as cylinders or spheres. For more realistic scenarios, three-dimensional modeling is needed to simulate the full raceway geometry, the combustion of injected fuels and coke, and the fluid flow phenomena inside the raceway envelope.

Accurate modeling of the raceway is important, because it contains the critical parameters that control gas distribution in the blast furnace. Kawabata et al [5] developed a one-dimensional raceway mathematical model designed to predict the gas temperature and composition distributions inside the raceway region. In this model, coke particles inside the raceway were treated as a continuous phase. Additionally, the raceway depth and the void fraction inside raceway were assumed to be constant. Hatano[6] and Nogami [7] developed a two-dimensional model with a similar approach. However, it was also reported that raceway sizes obtained in pseudo two-dimensional and real two-dimensional apparatuses could be different. In the pseudo two-dimensional model, a jet of air is injected through a tuyere placed in the longitudinal central plane of the model domain. The jet can expand in all directions after leaving the tuyere tip, however, it is assumed that any impacts on flow due to jet expansion in the perpendicular direction to the tuyere axis are negligible. Moreover, the combustion of injection fuel has significant effects on the size and shape of raceway of a blast furnace, a phenomena which has not been well examined by any of the aforementioned modeling techniques.

In this paper, an established three dimensional (3-D) computation fluid dynamics (CFD) mathematical model for simulating the raceway shape and combustion in a blast furnace is examined. This model was developed through the efforts from the Global R&D-East Chicago of ArcelorMittal and Purdue University Calumet. In this model, the effects of coke and injected fuel combustion on the raceway geometry and the raceway gas flows are considered. The combustion of injected fuels and coke convert solid mass into a gaseous phase and generate heat in the raceway. The increase in the gaseous mass will increase the gas volume, and the increase in the temperature will expand the gas volume and/or increase the pressure. Therefore, in the modified 3-D CFD raceway model, source terms are added accordingly to present the effects due to these increases.

Mathematical Model

The simulation is divided into two major portions: (a) simulation of NG/PC combustion inside the tuyere, and (b) simulation raceway formation and combustion. The simulation starts with the NG/PCI combustion simulation. The flow profiles at the tuyere outlet obtained from this simulation are used as inlet boundary conditions in the raceway model. The commercial CFD solver ANSYS Fluent is used to model flow inside the blowpipe/tuyere geometry, as well as the initial stages of NG/PC combustion inside the tuyere region as shown in Figure 1.

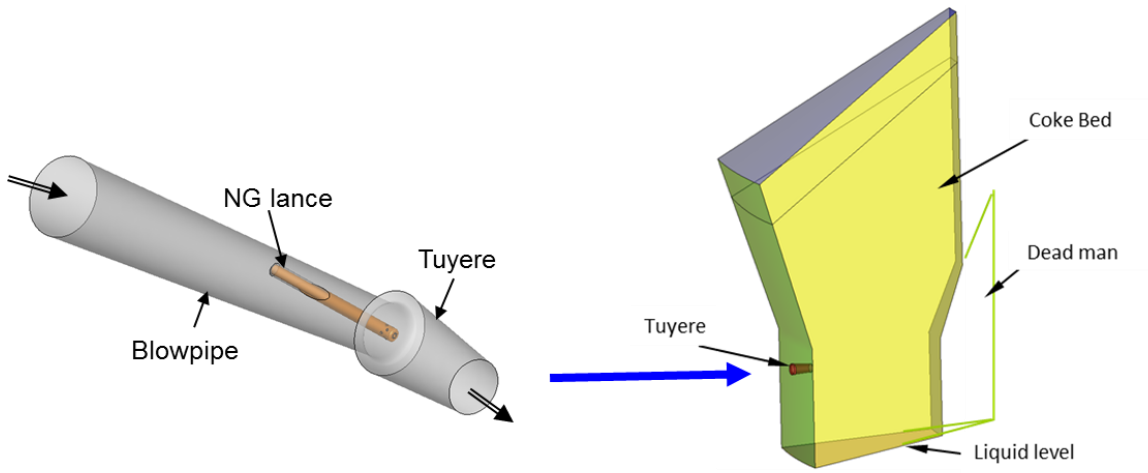


Figure 1. Schematic of CFD models of the tuyere and the raceway

The raceway model is divided into two sub-models, (a) the raceway formation model and (b) the raceway combustion model. A methodology has been developed to predict the effects of the combustion on the raceway geometry (sizes and shapes).

The raceway model estimates the raceway geometry and the distributions of coke injection fuels in the raceway. The model determines the raceway size and shape based on the porosities of the active coke zone at the front of a tuyere. The outputs of the raceway formation model are then used as boundary conditions for the raceway combustion model to prescribe a flow domain and porosity map.

The raceway combustion model employs a Eulerian approach to describe the gas-particle flow and combustion of the injected fuel and coke inside the raceway envelope. Conservation equations are utilized to describe mass transfer, heat transfer, and the motion of gas and particles.

The combustion of injected fuels and coke have an impact on the raceway geometry in two ways, (a) mass transfer from solid to gas, and (b) gas volume expansion due the temperature increase. To account for these phenomena and their effects on the raceway shape and size, mass and volume are added to each local cell accordingly. The mass and gas volume increments determined by the combustion are fed cell by cell to the formation model. The formation model will then determine a new raceway shape and size with the increments taken into account. Table 1 lists the detailed procedure of this methodology. The commercial CFD solver ANSYS Fluent and a proprietary CFD code are used in this study.

Table 1. Iteration between raceway formation and combustion models

Step	Model	Simulator	Raceway geometry and combustion simulating status
1	Raceway Formation	ANSYS Fluent®	To predict raceway without combustion
2	Raceway Combustion	Proprietary Code	Combustion simulating with raceway geometry from Step 1
3	Raceway Formation	ANSYS Fluent®	To predict raceway with combustion source term from Step 2
Repeat Steps 2 and 3 till the shape and size converges.			

Raceway Formation Model

A transient 3D Eulerian approach was employed to predict the raceway geometry. In the Eulerian approach, the different phases are treated as interpenetrating continua. The Eulerian approach uses the concept of phasic volume fraction, α. The volume fractions are assumed to be the continuous functions of space and time, and the sum of the fractions is equal to one. The conservation equation for each phase is given below,

$$\frac{\partial(\alpha_i \rho_i \phi)}{\partial t} + div(\alpha_i \rho_i \phi u) = div(\Gamma grad\phi) + S_\phi \tag{1}$$

where S_ϕ is the mass transfer between the two phases due to chemical reactions of coal and coke. The flow behavior of the fluid-solid mixture is described using a multi-fluid granular flow model. The granular multi-fluid model consists of granular phase conservation equations and fluid phase conservation equations. The inter-phase momentum exchange is modeled using the Syamlal-O'Brien model [8], where the fluid-solid interaction coefficient is defined using an empirical relationship based on terminal velocity measurements in fluidized beds and settling beds. The coefficient is a function of volume fraction and Reynolds number. The solid-solid interaction is based on the assumption that there is instantaneous binary collisions between particles and that the energy dissipation is due to inelasticity of collisions.

Raceway Combustion Model

The proprietary in-house model uses an Eulerian system to describe the gas-particle flow and coal combustion. Conservation equations of mass, energy, and momentum are used to describe the mass transfer, heat transfer, and the motion of gas-particle, respectively. The k- ε two-phase turbulence model is used to simulate gas and particle turbulence. The eddy break-up-Arrhenius combustion model is used for gas combustion. Two-competing reaction model is used for the coal devolatilization rate [9]; and the diffusion-kinetic model is used for the overall reaction rate of char reaction.

Governing equations

Gas-particle phase continuity, momentum, species mass fraction, and energy equations, as well as the equations of the turbulence momentum and its dissipation rate at steady state are described below.

$$\frac{\partial}{\partial x}(\rho u\phi) + \frac{\partial}{\partial y}(\rho v\phi) + \frac{\partial}{\partial z}(\rho w\phi) = \frac{\partial}{\partial x}(\Gamma_\varphi \frac{\partial \phi}{\partial x}) + \frac{\partial}{\partial y}(\Gamma_\varphi \frac{\partial \phi}{\partial y}) + \frac{\partial}{\partial z}(\Gamma_\varphi \frac{\partial \phi}{\partial z}) + S_\varphi + S_{\varphi Pg} \tag{2}$$

$$\begin{aligned} &\frac{\partial}{\partial x}(\rho_P u_P \phi_P) + \frac{\partial}{\partial y}(\rho_P v_P \phi_P) + \frac{\partial}{\partial z}(\rho_P w_P \phi_P) \\ &= \frac{\partial}{\partial x}(\Gamma_{\varphi P} \frac{\partial \phi_P}{\partial x}) + \frac{\partial}{\partial y}(\Gamma_{\varphi P} \frac{\partial \phi_P}{\partial y}) + \frac{\partial}{\partial z}(\Gamma_{\varphi P} \frac{\partial \phi_P}{\partial z}) + S_{\varphi P} + S_{\varphi Pg} \end{aligned} \tag{3}$$

where ϕ and ϕ_P are the general independent variable, Γ_φ and $\Gamma_{\varphi P}$ are the effective transport coefficient; S_φ, $S_{\varphi P}$ and $S_{\varphi Pg}$ are source terms.

Interphase momentum exchange

The particle exchanges momentum with the gas through the drag force. When the void fraction is greater than or equal to 0.8, the momentum exchange coefficient is expressed as,

$$\beta_{\mathrm{k}} = \frac{3}{4} C_{\mathrm{D}} \rho \frac{|\mathrm{u} - \mathrm{u}_k|(1-\varepsilon_{\mathrm{k}})^2}{d_{\mathrm{k}}} f(\varepsilon_{\mathrm{k}}) \tag{4}$$

$f(\varepsilon_k)$ accounts for the effect of the presence of other particles and is a correction to the Stokes law for free fall of a single particle. The following equation is used in this work.

$$f(\varepsilon_{\mathrm{k}}) = (1-\varepsilon_{\mathrm{k}})^{-3.8} \tag{5}$$

The drag coefficient is estimated as a function of the Reynolds number and is described as follows.

$$C_D = \begin{cases} \dfrac{24}{Re_k}\left(1 + \dfrac{Re_k^{0.667}}{6}\right) & Re_k < 1000 \\ 0.44 & Re_k \geq 1000 \end{cases} \tag{6}$$

where the Reynolds number is given by

$$Re_{\mathrm{k}} = \frac{\rho d_k |\mathrm{u} - \mathrm{u}_k|}{\mu} \tag{7}$$

When the void fraction is less than 0.8, the momentum exchange coefficient is calculated by the Ergun's equation below.

$$S_u = (150\frac{(1-\varepsilon_g)^2 \mu_g}{\varepsilon_g \psi_c^2 d_b^2} + 1.75\frac{|U_g - U_b|(1-\varepsilon_g)\rho_g}{\psi_c d_b}) \times \varepsilon_g (U_g - U_b) \tag{8}$$

Gas combustion

The eddy break-up turbulent combustion model is used to quantify the effect of turbulence on the combustion rates of volatiles, carbon monoxide, and hydrogen. The reaction rate is determined as

$$W_s = \min(W_{\mathrm{s,\ EBU}},\ W_{\mathrm{s,Arr}}) \tag{9}$$

where

$$W_{\mathrm{s,Arr}} = B_{\mathrm{s}} \rho^2 Y_{\mathrm{F}} Y_{\mathrm{ox}} \exp(-\frac{E_s}{\mathrm{R}T}) \tag{10}$$

$$W_{\mathrm{s,EBU}} = C_{\mathrm{R}} \rho \frac{k}{\varepsilon} \min(Y_{\mathrm{F}}, \frac{Y_{\mathrm{ox}}}{\beta}) \tag{11}$$

Interphase heat transfer

The heat transfer between a single reacting particle and the gas phase is calculated based on the stagnant film theory.

$$Q_{\mathrm{k}} = \pi d_{\mathrm{k}} Nu_{\mathrm{k}} \lambda_{\mathrm{s}} (T - T_{\mathrm{k}}) \frac{B_{\mathrm{k}}}{\exp(B_{\mathrm{k}}) - 1} \tag{12}$$

$$B_{\mathrm{k}} = -\frac{\dot{m}_{\mathrm{K}}\, C_{\mathrm{ps}}}{\pi d_{\mathrm{k}} Nu_{\mathrm{k}} \lambda_{\mathrm{s}}} \tag{13}$$

$$Nu_{\mathrm{k}} = 2 + 0.5\,\mathrm{Re}_{\mathrm{k}}^{0.5} \tag{14}$$

where the so-called 1/3 Law is used to calculate the thermal conductivity and around the coal particles.

Moisture evaporation rate

A diffusion model is used to calculate the moisture evaporation rate. The moisture in a coal particle is assumed to diffuse to the surface of the particle to form a liquid film. This film is treated as a surface layer of a water droplet with the same diameter. The moisture evaporation rate is calculated as follows.

$$\dot{m}_{wk} = \begin{cases} -\pi d_{\mathrm{k}} Nu_{\mathrm{k}} D_{\mathrm{s}} \rho_{\mathrm{s}} \ln\left(1 + \dfrac{Y_{\mathrm{H_2O,s}} - Y_{\mathrm{H_2O,g}}}{1 - Y_{\mathrm{H_2O,s}}}\right) & T_{\mathrm{k}} < T_{\mathrm{b}} \\ -\pi d_{\mathrm{k}} Nu_{\mathrm{k}} \dfrac{\lambda_{\mathrm{s}}}{C_{\mathrm{ps}}} \ln\left(1 + \dfrac{C_{\mathrm{ps}}(T - T_{\mathrm{k}})}{1 - L_{\mathrm{w}}}\right) & T_{\mathrm{k}} \geq T_{\mathrm{b}} \end{cases} \tag{15}$$

$$Y_{\mathrm{H_2O,s}} = B_{\mathrm{w}} \exp(-E_{\mathrm{w}} / RT_{\mathrm{k}}) \tag{16}$$

where Y_{H2O} is mass fraction of vapor at the surface of the particle. The Nusselt number, Nu_k, is calculated as

$$Nu_{k} = 2 + 0.5\,\mathrm{Re}_{k}^{0.5} \tag{17}$$

Coal devolatilization rate

Coal is assumed to decompose to form char and combustible volatiles. The combustible volatile is assumed to consist of hydrocarbons (C_dH_d) and carbon monoxide.

$$\text{Coal} = [\text{Volatiles}] + \text{Char} \tag{18}$$

$$C_aH_bO_c = [C_dH_b + cCO] + eC \quad (\text{Volatiles} = C_dH_b + cCO) \tag{19}$$

The constants a to e are determined from the coal ultimate analysis.

The coal devolatilization rate is proportional to the mass of the dry ash free (daf) coal. The devolatilization is modeled by two simultaneous competing first-order irreversible reactions.

$$daf\ coal \begin{cases} (1-\alpha_1)\, Ch_1 + \alpha_1 V_1 & (Reaction1) \\ (1-\alpha_2) Ch_2 + \alpha_2 V_2 & (Reaction2) \end{cases} \tag{20}$$

The devolatilization rate is calculated as

$$\dot{m}_{vk} = -\alpha_1 m_{dk} B_{v1} exp(-\frac{E_{v1}}{RT_k}) - \alpha_2 m_{dk} B_{v2} exp(-\frac{E_{v2}}{RT_k}) \tag{21}$$

where α_1 is obtained from the volatiles matter percentage in coal proximate analysis, and α_2 is equal to $2\alpha_2$.

The reduction rate of the daf coal mass due to the devolatilization is calculated as

$$\dot{m}_{dk} = -m_{dk} B_{v1} exp(-\frac{E_{v1}}{RT_k}) - m_{dk} B_{v2} exp(-\frac{E_{v2}}{RT_k}) \quad (22)$$

The volatile matters released into the gas phase undergo the following homogeneous combustion.

$$C_d H_b + \frac{d}{2} O_2 = d\ CO + \frac{b}{2} H_2 \quad (23)$$

$$2CO + O_2 \rightarrow 2CO_2 \quad (24)$$

$$2H_2 + O_2 \rightarrow 2H_2O \quad (25)$$

Char reaction rate

The following heterogeneous char reactions are included in the model.

$$C + O_2 \rightarrow CO_2 \quad (26)$$

$$2C + O_2 \rightarrow 2CO \quad (27)$$

$$C + CO_2 \rightarrow 2CO \quad (28)$$

$$C + H_2O \rightarrow CO + H_2 \quad (30)$$

All the char reactions are assumed to be of first-order with respect to O_2, CO_2, and H_2O. The reaction rates for char reactions in equations (26) to (30) in terms of the gas consumption rates are given below.

$$\dot{m}_{ck,A} = -\pi d_k^2 \rho_s Y_{O_2,s} B_A \exp(-\frac{E_A}{RT_k}) \quad (31)$$

$$\dot{m}_{ck,B} = -\pi d_k^2 \rho_s Y_{O_2,s} B_B \exp(-\frac{E_B}{RT_k}) \quad (32)$$

$$\dot{m}_{ck,C} = -\pi d_k^2 \rho_s Y_{CO_2,s} B_C \exp(-\frac{E_C}{RT_k}) \quad (33)$$

$$\dot{m}_{ck,D} = -\pi d_k^2 \rho_s Y_{H_2O,s} B_D \exp(-\frac{E_D}{RT_k}) \quad (34)$$

Applications

The existing raceway models have been applied to analyze a variety of blast furnaces in a broad range of operating conditions. In this paper, two recent analysis projects utilizing the computational raceway simulation model are examined. Both simulation projects detailed herein were previously published and presented at AISTech 2015 [10][11]. The first furnace examined was the No. 1 blast furnace at United States Steel Canada Lake Erie Works. The project was

undertaken to study injected natural gas (NG) combustion performance and the impact of various lance designs on blast furnace operation and stability.

Research was conducted on three lance designs. The first, referred to as the 'fast lance' had a large number of small holes for gas egress. The design of this lance was expected to improve gas dispersion and enhance combustion inside the tuyere region. The second lance was a simple straight pipe, and the third design was a modification of the fast lance, created by boring out the lance tip. The three lances can be observed in Figure 2.

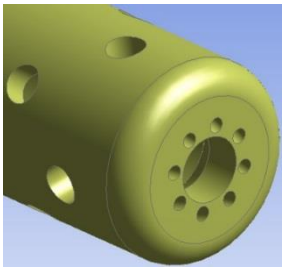
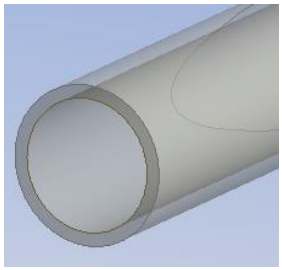
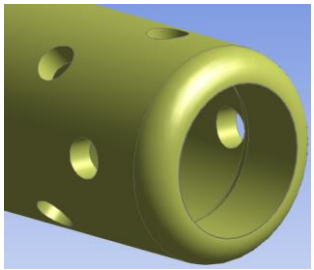
(a)  (b)  (c) 

Figure 2. (a) Fast Lance, (b) Straight Lance and (c) Bored Lance

Identical operating conditions were maintained in all three lance design simulations so that combustion characteristics could be accurately compared. After calibrating the kinetics of natural gas combustion (to accurately model flame liftoff and blowout), combustion of injected natural gas inside the tuyere was modeled [12]. Combustion characteristics and temperature profiles inside the tuyere were examined to compare combustion speed between the three lance designs. A comparison between temperature profiles inside the tuyere is shown in Figure 3.

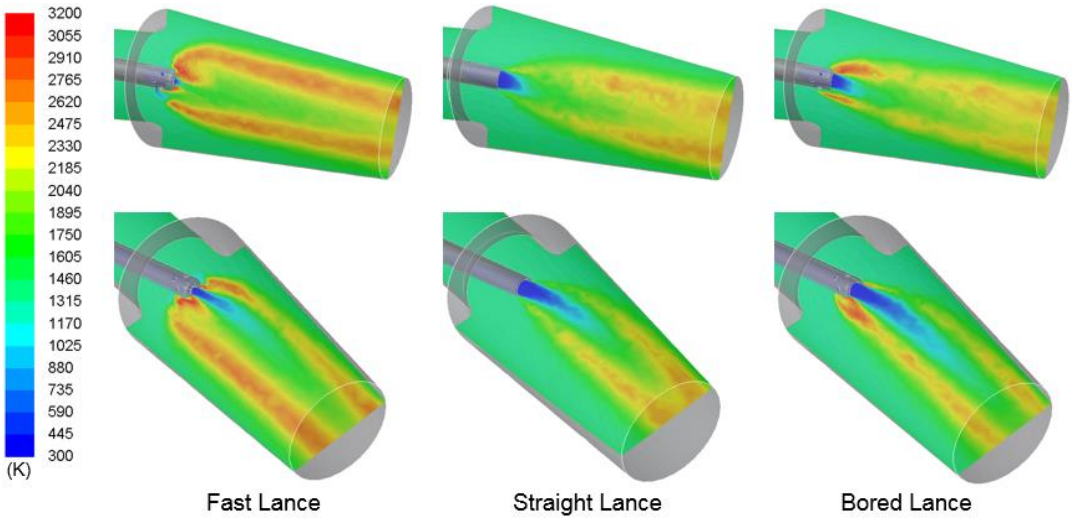


Figure 3. Temperature distributions in the tuyere

It is obvious from an examination of the temperature profiles that the additional holes in the lance tip in bored lance and fast lance designs contribute to the enhancement of combustion inside the

tuyere. The increased mixing and combustion inside the tuyere also leads to the release of additional CO and CO_2 in the gas flow that enters the raceway. The fast lance leads significantly more reactions than the other two designs, with the difference between the bored and straight lances being relatively minor. Figure 4 details the CO_2 distribution inside the tuyere for each lance design.

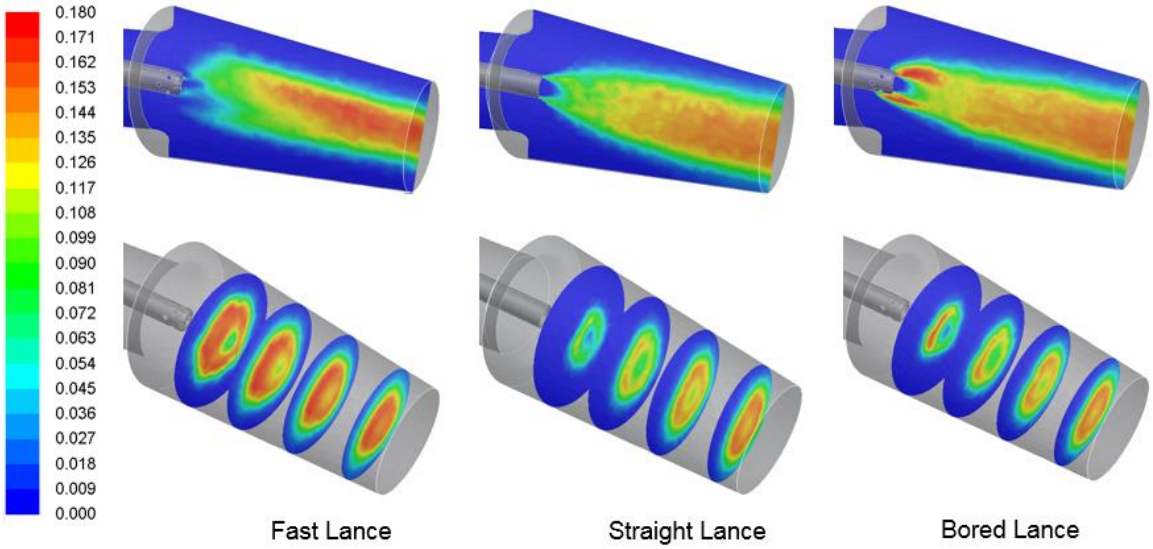


Figure 1. CO_2 mass fraction distribution inside the tuyere

The additional combustion observed in the fast lance case results in higher pressure drops over the tuyere region. Plant operators observed that higher NG injection rates also resulted in increased pressure drops, which can lead to limitation on furnace wind and production rates. Due to this, the increased pressure drop in the fast lance case, when utilized at high production rates, can result in poor stability due inability to supply enough wind to the furnace.

The raceway shape is not significantly impacted by the lance design, however gas species and temperature distributions are heavily altered. As visible in Figure 5, high gas temperatures are present on the side of the raceway, due to the angle at which NG is injected into the tuyere and the resulting consumption of oxygen. With more oxygen available on one side of the raceway, the gas temperatures increase due to coke combustion.

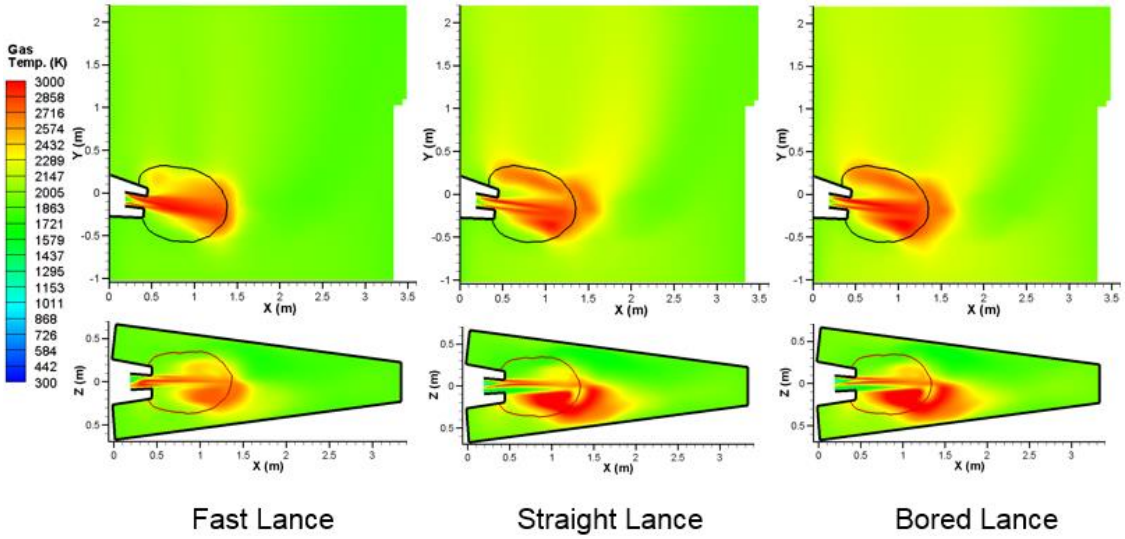


Figure 5. Temperature distribution in the raceway, side view (top) and top view (bottom)

Based on the raceway model outlet temperatures and gas distributions, data was then exported to a second in-house CFD code known as the blast furnace shaft simulator. This additional model allowed for the examination of blast furnace operation in the stack. Minor variations in gas and temperature distribution, as well as overall gas utilization in the furnace provided the basis for selecting improved operating conditions for the Lake Erie Works blast furnace.

The second furnace examined was located at AK Steel Dearborn Works in Dearborn, MI. This project was performed to examine the impact of co-injecting pulverized coal (PC) and NG in a variety of operating conditions on combustion and furnace performance [11]. One of the key factors examined in this project was the use of NG as a carrier gas for PC. The hope was that the replacement of nitrogen with NG could possibly improve combustion performance, in an effort to avoid some of the difficulties in the use of high rate pulverized coal injection, such as reduced permeability. Initially, a baseline case was modeled at standard operating conditions provided by AK Steel Dearborn Works. It was quickly discovered that the baseline design of the tuyere region resulted in poor heat transfer between NG combustion and the PC particles. Additionally, as can be observed in Figure 6, the method of NG injection (a tuyere port) leads to increased thermal wear on the tuyere surface.

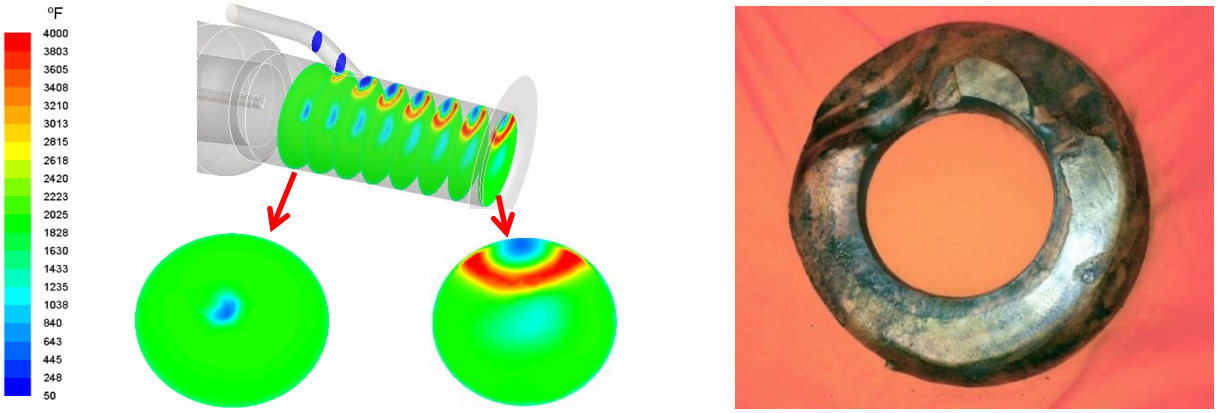


Figure 6. Temperature contours located on planes through the tuyere (left) and sectioned nose piece of tuyere from AK Steel Dearborn Works with wear/ablation zones visible (right)

The gas flow from the tuyere produces a standard jet into the raceway, with gas velocities lower at the center due to the momentum required to accelerate the pulverized coal. Areas of recirculation are easily visible in the raceway envelope, with regions of high temperature present in the recirculation areas as shown in Figure 7.

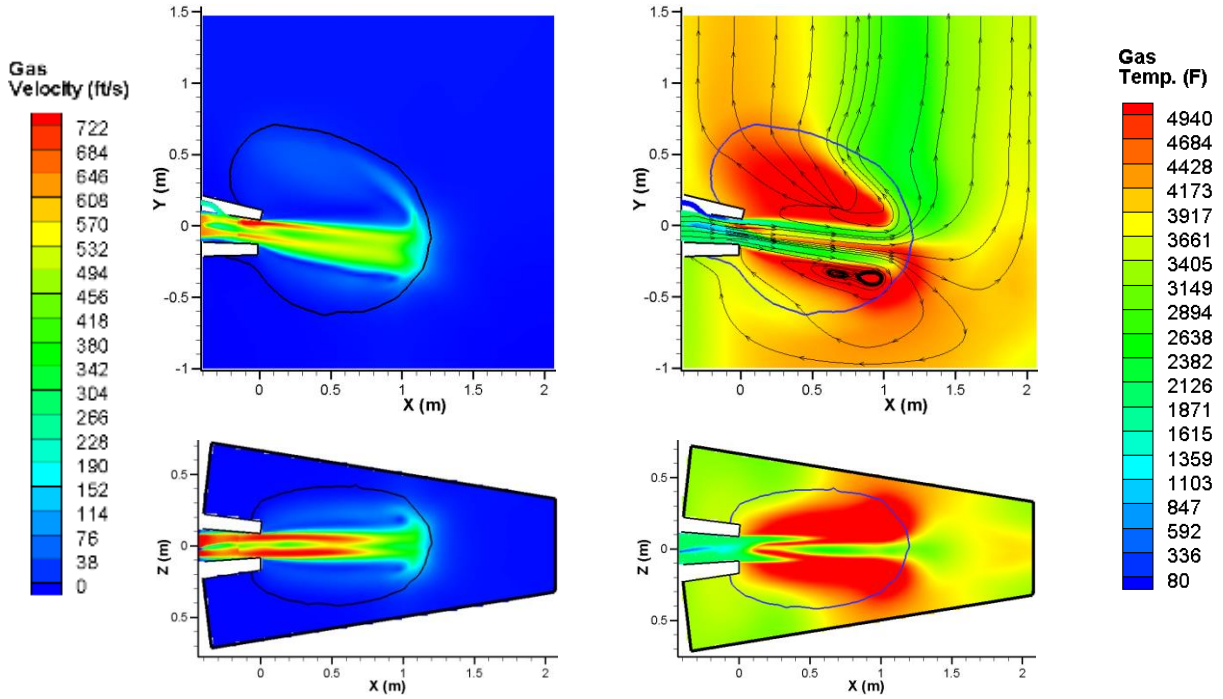


Figure 7. Contours of gas velocity (left) and gas temperature with streamlines (right) in the raceway region. Upper contours are located on section A-A, lower contours located on section B-B.

A key finding of this study was the discovery of incomplete PC burnout in the raceway. As seen in Figure 8, in some regions, the burnout reaches only 60% before passing out of the raceway envelope and into the coke bed. This phenomena can also result in performance and stability issues in furnace operation to unburnt coal buildup in the blast furnace coke bed.

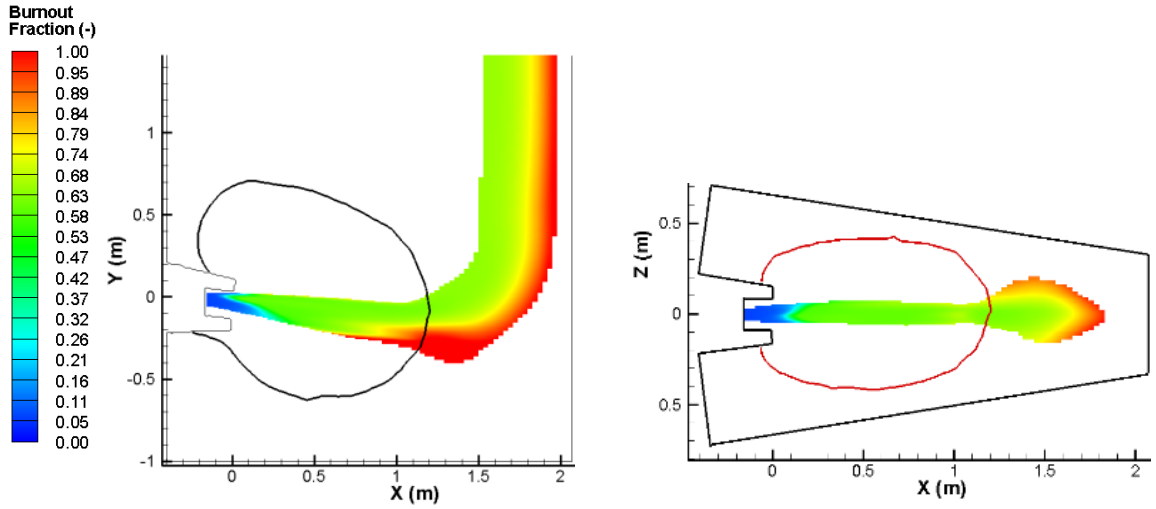


Figure 8. Contours of pulverized coal burnout fraction through the raceway viewed from the side (left) and top (right).

A variety of simulation cases were modeled in the study, utilizing two different gas injection designs, as well as the modification of the PCI carrier gas from nitrogen to natural gas. It was determined that the total burnout rate was improved for all cases that utilized NG to convey PC into the furnace. With the addition of a secondary lance for NGI, the thermal wear problem could be easily resolved. Additionally, when utilizing both the secondary NGI lance and NG for the conveyance of PC, the total fuel burnout neared 96%, a significant increase over standard operating condition values.

These two projects provide a representative examination of the broad applications of the blast furnace raceway modeling capabilities developed. Ranging from design modification and troubleshooting, to operational capabilities and combustion performance analysis, the existing model has been well validated and applied across a number of analyses for a variety of steel producers and their blast furnaces.

Conclusions

A 3-D CFD mathematical model has been established for estimating the raceway geometry and combustion in the raceway, in which the effects of coke and injected fuel combustion on the raceway geometry and gas flow/species distribution through the raceway envelope are considered. The combustion effects are treated as additional source terms of mass and momentum in the gas phase, because combustion serves to convert solid mass into gaseous mass and contribute to the generation of heat, resulting in additional momentum. The simulation results indicate that combustion has significant effects on the raceway shape and size.

The raceway model can be utilized to complete parametric studies analyzing natural gas and coal combustion in the raceway, raceway geometry, raceway gas flow, raceway temperature, raceway gas compositions, and other parameters. The parametric studies examined using this model thus far have helped to optimize to tuyere operation, coal and coke properties to achieve high fuel injection rates, improve fuel replacement ratios, and enhance PCI performance in a variety of ironmaking facilities and operating conditions.

Acknowledgement

The author would like to all the funding agencies and companies including AK Steel, AISI, AIST, ArcelorMittal – America, ArcelorMittal – Canada, DOE, Indiana 21st Century Technology and Development Fund, US Steel - America, US Steel – Canada, and Union Gas. The efforts and support by industrial collaborators as well as CIVS staff and students are also greatly appreciated.

Reference

1 Suzuki T., Smoot L.D., Fletcher T.H., Smith P.J. (1986). Prediction of high-intensity Pulverized Coal Combustion, Combustion Science and Technology, 45, 167-183.

2 Jamaluddin A.S., Wall T.F., Truelove J.S. (1986). Mathematical Modeling of Combustion in Blast Furnace Raceways, including injection of Pulverized Coal. Ironmaking and Steelmaking, 13, 91-99.

3 Nogami H., Miura T., Furukawa T. (1992). Simulation of Transport Phenomena around Raceway Zone in the lower part of Blast Furnace. Tetsu-to-Hagane, 78, 1222-1229.

4 Takeda K., Lockwood F.C.(1997). Integrated Mathematical Model of Pulverized Coal Combustion in a Blast Furnace. ISIJ International, 37, 432-440.

5 H. Kawabata, Z. G. Liu, F. Fujita and T. (2005). Characteristics of Liquid Hold-ups in a Soaked and Unsoaked Fixed Bed. : ISIJ Int., 45, 1466-1473.

6 Hatano, M., Kurita, K. and Tanaka, T. (1981). Aerodynamics study on raceway in blast furnace, International Blast Furnace Hearth and Raceway Symposium, Newcastle, Australia, Symposium Series, Australasian Institute of Mining and Metallurgy, 26,

7 Nogami, H., Miura, T. and Furukawa, T. (1992) Simulation of transport phenomena around raceway zone in the lower part of blast furnace, Tetsu to Hagane, 78, 1222-1229.

8 Syamlal, M. and O'Brien. T. (1987) The Derivation of a Drag Coefficient Formula from Velocity-Voidage Correlations.

9 Kobayashi, H, Howard, J.B., and Sarofim, A.F. (1977) *16th Symp. (Int.) Comb.*, Combustion Institute, **411**.

10 Silaen, A.K., Okosun, T., Chen, Y., Wu, B., Zhao, J., Zhao, Y., D'Alessio, J., Capo, J.C, Zhou, C.Q. (2015), Investigation of High Rate Natural Gas Injection through Various Lance Designs in a Blast Furnace, Proceedings of AISTech 2015, Cleveland, Ohio, United States.

11 Okosun, T., Street S.J., Chen, Y., Zhao, J., Wu, B., Zhou, C.Q. (2015), Investigation of Co-Injection of Natural Gas and Pulverized Coal in a Blast Furnace, Proceedings of AISTech 2015, Cleveland, Ohio, United States.

12 Chen, Y. (2015), Simulation of Natural Gas Combustion Liftoff and Blowout Phenomenon in Blast Furnace, *2015 TMS Annual Meeting & Exhibition*.

A Domain Language for Constructive Block Topology for Hexa Mesh Generation

***R. Rainsberger[1], Pedro V Marcal[2]**

[1]XYZ Scientific Applications, Inc.,2255 Morello Ave. Suite 220, Pleasant Hill, CA. 94523

[2]MPACT Corp.,5297 Oak bend Lane, Suite 105, Oak Park, CA. 91377

*Presenting and Corresponding author : r.rainsberger@yahoo.com

Abstract

A Topological Domain Language has been developed to aid the **True*Grid*®** beginning user to model a restricted set of problems by simplifying the formation of the hexa block topology. A computer program has been developed to convert scripts in this language to the journal format of **True*Grid*®**.

Keywords : topology, hexa mesh generation, domain language

Introduction

The element type has always been important in determining the accuracy of the results in a Finite Element Analysis. Recent results, Marcal et al[1], Fong et al [2]and Marcal et al [3] have demonstrated the superiority of the hexa 27 fully quadratic element[4]. In order to exploit such an element, we need to be able to generate hexa meshes over a wide spectrum of shapes and sizes. Ideally it would be preferable to generate the hexa meshes automatically. The only such method known to the authors is to generate the mesh via a subdivision of a tetra mesh. Such a procedure results in hexa meshes with poor element quality and an unnecessarily large number of degrees of freedom. The **True*Grid*®** Pre-processor [5] has been developed for hexa mesh generation. This program has many advanced features for the generation of complex meshes. It is not our intention to exercise such advanced features. Instead, we wish to explain the basis of this mesh generator and explore the possibility of developing an expert system which could help a new user generate some useful meshes. The **True*Grid*®** program uses a surface projection method. This method in its simplest form defines a block in 3D space and in a unit topological space, respectively. The block in 3D space is in turn mapped to an enclosing surface. Here we restrict the enclosing surface to those resulting from simple geometric solids. This allows us to develop a simple Domain Language.

We introduce the projection method by creating the simplest 3 X 3 X 3 unit topology and using this to create a mesh for a cylinder. The result is shown in Fig. 1. The **True*Grid*®** Journal for these problems is given in Appendix A.

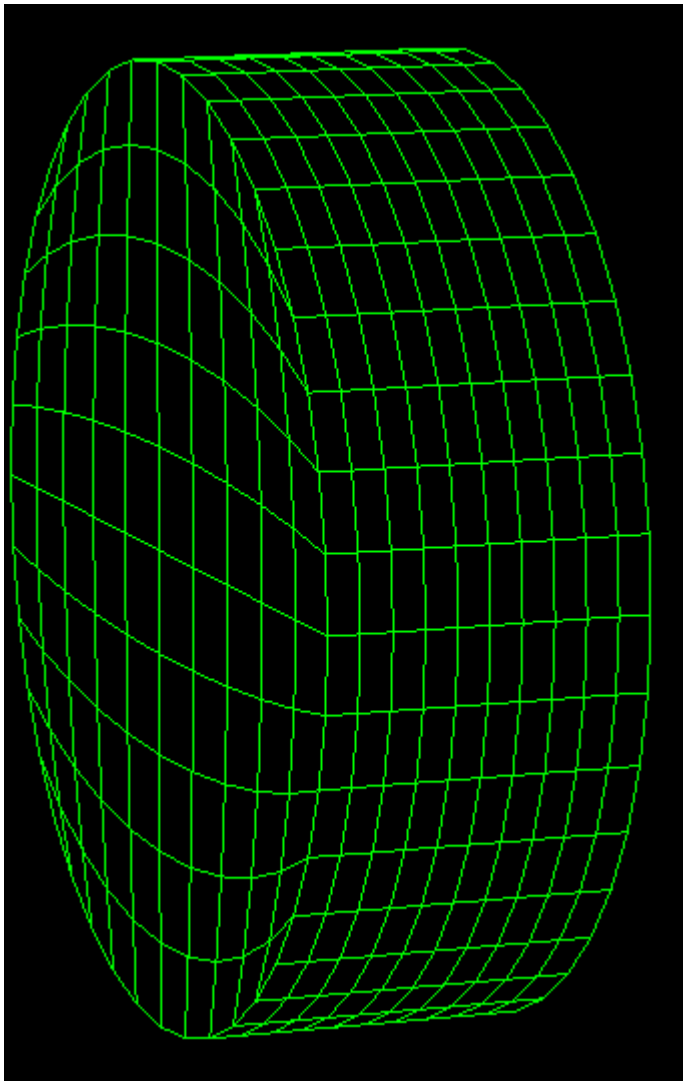


Fig. 1, Hexa mesh for cylinder, simple topology

The problem with such a mesh is that the quality of the elements at the edges of the topological blocks is poor.

We can improve this by introducing a cross pattern for the topology and requiring that the adjacent faces of the cross be joined for the projected model as shown in Fig. 2.

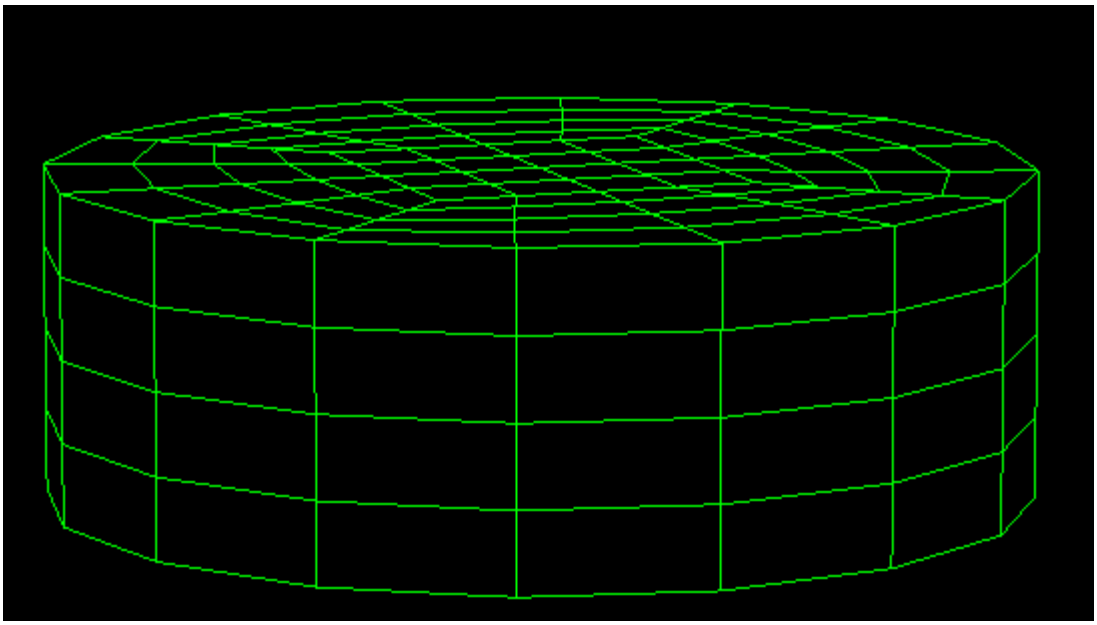
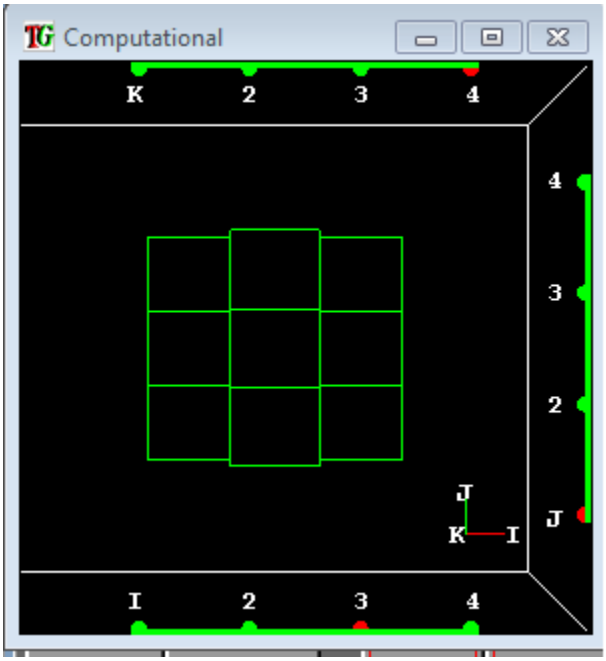


Fig. 2 a , Hexa mesh with cross topology



Frig. 2 b, Cross pattern in topological space

We can now see that the problem with the element quality has disappeared.

Finally we generalize this concept by introducing the cruciform topology in three directions. This is used to project to sphere and sphere like geometries.

This is shown in Fig. 3

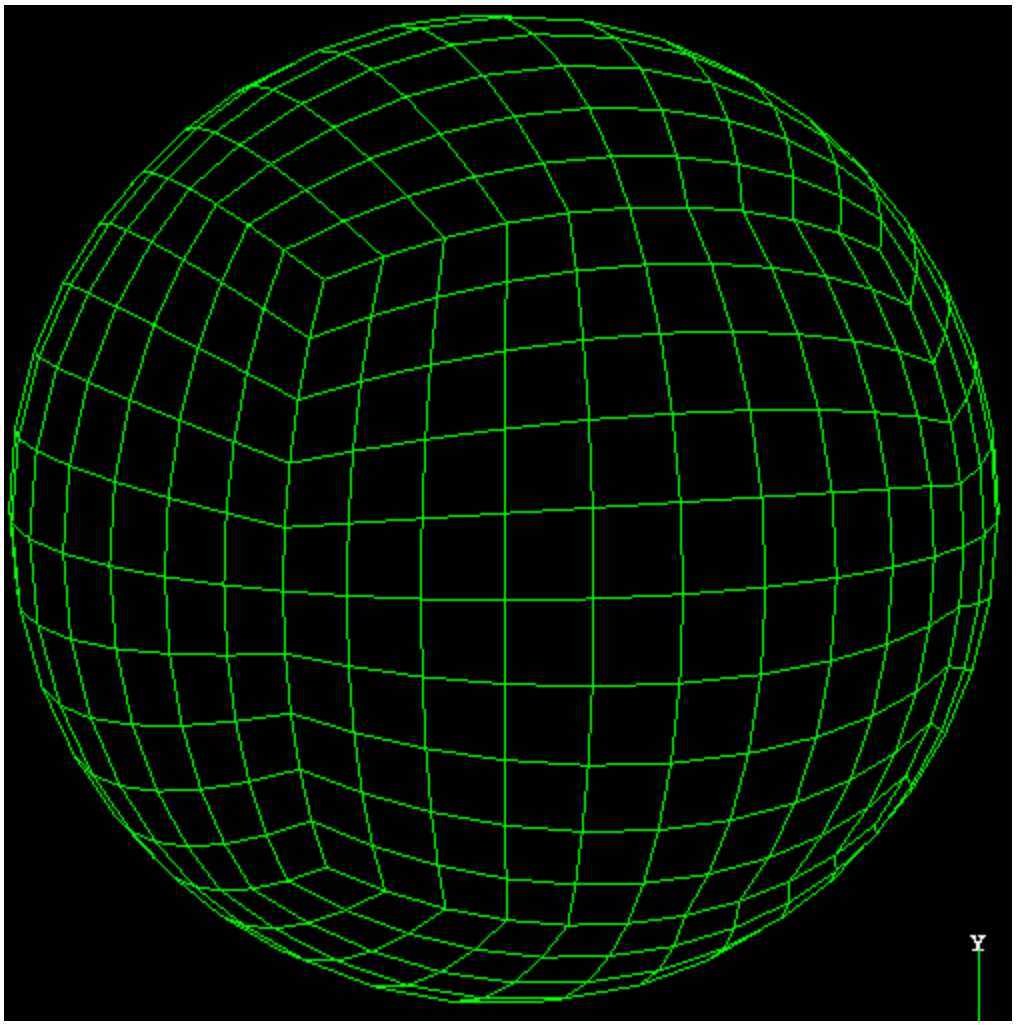


Fig. 3 a, Hexa Mesh for sphere with Cruciform topology

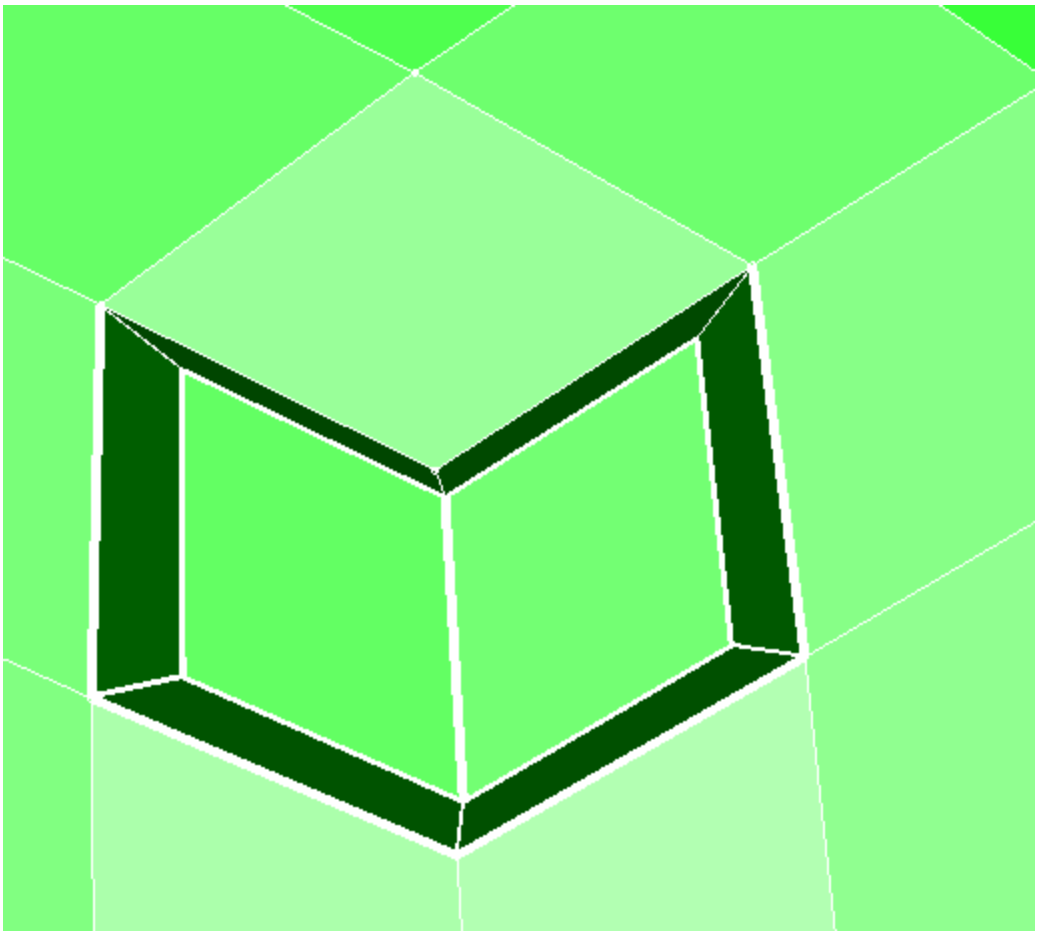


Fig. 3 b , Projection of the corner with three adjoining elements (two hidden)

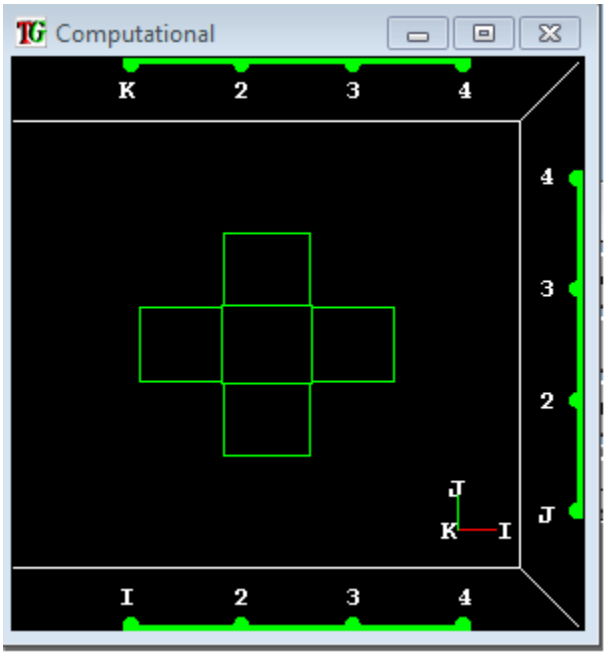


Fig 3 c, Cruciform topology

We have now established the basic requirements for the creation of a topology that can be used to project to the boundary surfaces to create hexa meshes via the projection method.

Theoretical Considerations

In the projection method, we first develop the topology in a three dimensional integer space. Our topology is then described here as a sequence of blocks. The blocks are joined together in a topologically constructive process. Because blocks are the only basic topology admitted in our process, the constructive topology may be restricted to a few simple operations. Here we adopt the concepts and the procedures used by **True*Grid*®**. However, to simplify our task, we will create a simple Topological Domain Language (TDL).

Naturally, in order to take advantage of the existing **True*Grid*®** software, we must also define a method to map the defined TDL operations to a language that the **True*Grid*®** can understand. This we achieve by mapping our operations into the parametric journal language currently used by **True*Grid*®**. The TDL is defined so that it can be easily translated from the TDL to the **True*Grid*®** Scripting Language.

In order to create a general purpose TDL, we have identified the following functions for the language. The functions that are preceded by a required name are the functions that will perform the translation to equivalent terms in the **True*Grid*®** journal.

Finite Element Format

name=FemFormat(mpact)

FemOutput(mpact)

or

Mesh

Coordinate System

Crd(3D||TP,(list i),(list j),(list k))

Crdr(3D||TP, (range i),(range j),(range k))

Crdf(3D,(min i,min j,min k),(max i,max j,max k))

Crdfi(3D,(min i,max i),(min j,max j),(min k,max k)) # see directions below in art. Translation to **True*Grid*®** Script.

CrdfPos(3D,(pos min i,pos max i),(pos min j,pos max j),(pos min k,pos max k)) # the values corresponds to count in the coordinates of the block definition list

SubPos(i or j or k,(min0,max0),(min1,max1)) # use node numbers

CrdfsPerm(SubPos0,SubPos1,SubPos2) # up to three SubPos can be included, permute the subpos between the two of i j k directions. This allows us to change the number of nodes in a block.

CrdfhPerm(SubPos) # is used as an Insert Sub Block (insprt) to add subdivisions to a direction of a block in preparation for creating a part. A negative value indicates insertion to the left of the min index. The max units are in elements (ie nodes+1)

CrdfPermute(TP,(i-min,i-max),(j-min,j-max),(k-min,k_max)) # when more pairs are used in a dir. Add 0 as a separator. Use (:) to indicate the full range in a particular direction.

Block Creation

The basic building block is created by defining a rectangular block in 3D space and its equivalent in topological space.

name=Block(Crd in 3D,Crd in TP)

name=SubBlock(Block, Crd of Sub_block)

name=Cross(Block start, CrdfPermute)

name=Cruciform(Block start,CrdfPermute,CrdfPermute) # two orthogonal CrdfPermutes

name=TieFace(Crdf, tie-#)

name=MapBlock(Sub_block , Block)

name=EndBlock(Block) # end operations for the named Block

Utilities and other useful quantities

Vec((i0,j0,k0),(i1,j1,k1))

VecAdd(vec1,vec2)

VecSub(vec1,vec2)

Dir(0 or 1, 0 or 1,0 or 1)

Pos(i, j,k)

DisplayWindow(x,y)

Block Editing

BlockTranslate(Block,Vec)

BlockRotate(Block,VecDir, num rt angles)

Block Boolean Operations

Union(Block 1, Block 2)

Subtract(Block 1, Block 2)

Intersection(Block 1, Block 2)

InsertSubBlock(SubBlock)

Slice(Block,VecPos,normal to VecDir) # result is two blocks _p and _m added to block name

DeleteBlock(Block)

Topological Properties for Surface Projection

Faces(Crdf) # list corresponds to position on list of block definition

name=Project(Faces,Geom)

Geometric Surfaces For Blocks

Sph(center,rad)

Cyl(center,dir vector)

Tor(center,torus rad, Center,Major rad)

Con(center,dir vector,end rad, other end radius)

Projection Operations to Surfaces

name=BlockSurface(Crd,geometric surfaces) # from block to surface

Insertion of a Sub-Block

The operation of inserting a block into a current block is about the most complex process that we will address with our TDL. Because of the number of operations involved, we have to be systematic about the procedure. We assume that the current block has been defined in the usual way. Because we need to increase the number of elements around our point of insertion, we need to do so consistently because of the block nature of our elements. To be specific, we will use the example of a cylinder joined to a hemisphere and insert another cylinder with its axis in the y direction. This is actually the example we use in our case 2 discussions.

It is listed here before the insertion

prog=FemFormat(mpact)

coord_1=Crdl(3D,(1 6 11 16),(1 6 11 16),(1 6 11 16))

coordt=Crd(TP,(0 12 18 18),(-6 -6 6 6),(-6 -6 6 6))

blk1=Block(coord_1,coordt)

create holes in topology for better elements about spheres and cylinders.

crdfp5=CrdfPermute(TP,(3 ,4),(1, 2 ,0 ,3, 4),(2,3))

crdfp6=CrdfPermute(TP,(:),(1, 2 ,0 ,3, 4),(1 ,2, 0 ,3 ,4))

```
crdfp7=CrdfPermute(TP,( 3,4 ),(2,3),(1 ,2, 0 ,3 ,4))
cruci=Cruciform(blk1,crdfp6,crdfp5,crdfp7)
crdfp8=CrdfPermute(TP,(1,3),(2,3),(2,3))
cross5=Cross(blk1,crdfp8)
# Projection of block face to geometry faces.
orig9=Pos(12,0,0)
rad9=Rad(5)
sph1=Sph(orig9,rad9)
faces8=Crdfi(3D,(2,-3),(-2,-3),(-2,-3))
proj1=Project(faces8,sph1)
orig6=Pos(12,0,0)
rad6=Rad(6)
sph2=Sph(orig6,rad6)
faces7=Crdfi(3D,(2,-4),(-1,-4),(-1,-4))
proj2=Project(faces7,sph2)
orig=Pos(0,0,0)
dir=Dir(1,0,0)
rad=Rad(5)
cyl1=Cyl(orig,dir,rad)
faces1=Crdfi(3D,(1,2),(-2,-3),(-2,-3))
proj=Project(faces1,cyl1)
orig1=Pos(0,0,0)
dir1=Dir(1,0,0)
rad1=Rad(6)
cyl2=Cyl(orig1,dir1,rad1)
faces=Crdfi(3D,(1,2),(-1,-4),(-1,-4))
proj3=Project(faces,cyl2)
```

```
# insert Crdfh here
sub5=SubPos(k,(2,4))
crdfh1=CrdfhPerm(sub5)
sub6=SubPos(k,(3,7))
crdfh2=CrdfhPerm(sub6)
sub7=SubPos(i,(1,5))
crdfh3=CrdfhPerm(sub7)
sub8=SubPos(i,(2,5))
crdfh4=CrdfhPerm(sub8)
crdfp9=CrdfPermute(TP,(2,3),(3,4),(3,4))
cross6=Cross(blk1,crdfp9)
# insert Cfrds here
orig3=Pos(6,0,0)
dir3=Dir(0,1,0)
rad3=Rad(2)
cyl3=Cyl(orig3,dir3,rad3)
faces3=Crdfi(3D,(-2,-3)(3,4),(-3,-4))
proj5=Project(faces3,cyl3)
# insert hole and map 4 faces to the defined surface
crdf1=Crdf(3D,(2, 3, 3),( 2, 4, 4 ))
crdf2=Crdf(3D,(3, 3, 3),( 3, 4, 4 ))
crdf3=Crdf(3D,(2, 3,3),( 3, 4, 3 ))
crdf4=Crdf(3D,(2, 3, 4),( 3, 4, 4 ))
# insert tie definitions
tie1=TieFace(crdf1,1)
tie2=TieFace(crdf3,2)
tie3=TieFace(crdf2,3)
tie4=TieFace(crdf4,4)
Pause
```

```
endb1=EndBlock(blk1)
```

1. We now discuss the coding in the original block that implements the hole that allows the insert to be joined.

We now expand the number of elements. We make use of the CrdfsPerm function. This allows us to change the number of nodes in a block. We insert the following after the Block definition.

```
sub1=SubPos(i,(1,5))

crdfs=CrdfsPerm(sub1)

sub2=SubPos(j,(2,10))

sub3=SubPos(k,(2,10))

sub4=SubPos(i,(1,5))

crdfs1=CrdfsPerm(sub2,sub3,sub4)
```

2. Create nodes and topology for new elements created above. Use function CrdfhPerm. The new nodes and topology are updated and printed to the screen, so that they can help in defining the insert block later. These appear as shown below, respectively.

```
*** PrintList *** new_block_3d

 [1, 6, 11, 16, 21, 26]   [1, 6, 21, 26]   [1, 6, 10, 17, 21, 26]

 *** PrintList *** new_block_tp

 [0, 1, 2, 12, 18, 18]   [-6, -6, 6, 6]   [-6, -6, -5, -4, 6, 6]


sub5=SubPos(k,(2,4))
crdfh1=CrdfhPerm(sub5)
sub6=SubPos(k,(3,7))
crdfh2=CrdfhPerm(sub6)
sub7=SubPos(i,(1,5))
crdfh3=CrdfhPerm(sub7)
sub8=SubPos(i,(2,5))
crdfh4=CrdfhPerm(sub8)
```

3. Make a hole to allow the new block to be inserted. The following creates a hole and maps it to the surface of the shape of the insert. We note from the cylindrical surface being used that its origin is (6,0,0) and its radius is 2.

```
crdfp9=CrdfPermute(TP,(2,3),(3,4),(3,4))

cross6=Cross(blk1,crdfp9)
```

```
orig3=Pos(6,0,0)

dir3=Dir(0,1,0)

rad3=Rad(2)

cyl3=Cyl(orig3,dir3,rad3)

faces3=Crdfi(3D,(-2,-3)(3,4),(-3,-4))

proj5=Project(faces3,cyl3)
```

4. Label the 4 surfaces to be tied to the corresponding surfaces in the Sub-Block to be created.

```
crdf1=Crdf(3D,(2, 3, 3),( 2, 4, 4 ))

crdf2=Crdf(3D,(3, 3, 3),( 3, 4, 4 ))

crdf3=Crdf(3D,(2, 3,3),( 3, 4, 3 ))

crdf4=Crdf(3D,(2, 3, 4),( 3, 4, 4 ))

tie1=TieFace(crdf1,1)

tie2=TieFace(crdf3,2)

tie3=TieFace(crdf2,3)

tie4=TieFace(crdf4,4)
```

5. Now we are in a position to create the Sub-Block**.** We first terminate the specification for the first block. Then we define the block that will become the cylindrical insert. This block must have its nodal coordinates coincide with those of the cylindrical hole. (**True*Grid*®** only requires this to be close. It has eight different ways of making the hole and insert equal exactly.) The nodes (4,9),(1,6),(4,11) are chosen to coincide exactly with the nodes (6,11),(21,26),(10,17) of the first block respectively. With reference to the topology of the insert, we are free to choose a new reference system as long as its core coordinates are consistent with its nodal coordinates and its positioning on the original block, blk1. For the x direction we choose (5,5,7,7) as required by the symmetry for the cylinder and with the cylinder origin at 6. For the y direction we choose (5,6,10) because the first two indices coincide with the topology of the hole and cylinder in the first block. Finally in the z direction we choose (-1,-1,1,1) to give the required symmetry for the cylinder. Note that **True*Grid*®** looks for key points in the insert to match the two blocks. When it does not find it, it defaults to using the nodal coordinates. That is why only the symmetry is defined here since the hole origin and depth to be filled are already defined by the x and y topology, respectively.

```
endb1=EndBlock(blk1)
```

```
coord_2=Crdl(3D,(1 4 9 12),(1 6 16),( 1 4 11 14 ))
coordt2=Crd(TP,(5 5 7 7) ,(5 6 10),(-1 -1 1 1))
blk2=Block(coord_2,coordt2)
```

6. Now we define the topology holes and the surfaces to be tied.

```
crdfp10=CrdfPermute(TP,(1, 2 ,0 ,3, 4),( : ),(1 ,2, 0 ,3 ,4))
cross7=Cross(blk2,crdfp10)
crdfp11=CrdfPermute(TP,(2,3),( : ),(2,3))
cross8=Cross(blk2,crdfp11)
crdf5=Crdf(3D,( 1, 1, 2),( 1, 2, 3))
crdf6=Crdf(3D,(2, 1, 1),( 3, 2, 1))
crdf7=Crdf(3D,( 4, 1, 2),( 4, 2, 3))
crdf8=Crdf(3D,( 2, 1, 4),( 3, 2, 4))
tie5=TieFace(crdf5,1)
tie7=TieFace(crdf7,3)
tie6=TieFace(crdf6,2)
tie8=TieFace(crdf8,4)
orig4=Pos(6,0,0)
dir4=Dir(0,1,0)
rad4=Rad(2)
cyl4=Cyl(orig4,dir4,rad4)
faces4=Crdfi(3D,(-1,-4),(2, 3),(-1,-4))
proj6=Project(faces4,cyl4)
```

7. Finally we create the internal surface of the insert, namely another cylinder. Then end the part and merge the two blocks, concluding the project to insert a sub-block. The diagram for the insert block is shown as Fig. 6, when we discuss the case study for the cylinder with insert.

```
orig5=Pos(6,0,0)
dir5=Dir(0,1,0)
rad5=Rad(1)
cyl5=Cyl(orig5,dir5,rad5)
```

```
faces5=Crdfi(3D,(-2,-3),( : ),(-2,-3))

proj7=Project(faces5,cyl5)

DisplayWindow(10,20)

Pause

endb1=EndBlock(blk2)

Merge

Mesh
```

Translation to True*Grid*® Script

The objective of the project was to develop a Domain Language with simple concepts that may subsequently be used in an expert system. In the above we have used the concept of Constructive Solid Geometry as our guide in defining a constructive Topological model for the Projection method in **True*Grid*®**. In the process we have simplified our model to first define a mapping from the Unit Topological Space to a Block indexed cube in 3D space. This block indexed cube is in turn projected to an enclosing surface space defined by simple engineering solids. By this process we have restricted the full capabilities of the **True*Grid*®** program. However, we are of the opinion that the domain covers a significant spectrum of mesh generation problems that it may be of interest to most analysts wanting to generate hexa meshes. A python program (TDL2TG.py) was written to translate the TDL script to the **True*Grid*®** journal script.

This program was first used to translate scripts used to generate the basic cases discussed for cylinders and spheres with simple topology and also with cross and cruciform topology respectively. These examples are listed in Appendix A.

The surfaces or faces of each block are defined by six faces which in our case are projected to the geometric figure specified by the sf index.

Each face is specified by sf defined by the indices
sf = i-min j-min k-min; i-max j-max k-max; sd* # =Crdf in our notation
Instead of specifying 6 faces, we introduce a short-hand indicial notation. We note that the back and front face is specified by -ve and +ve prefixes respectively. However we will visit each axis direction in turn and allow two index values *_min *_max. A - in front of both indices mean a range with either a back or front face. A single negative index selects only that face. Meanwhile the negative signs are ignored in the other two axis directions. The values there define the range to be combined with the active axis direction being processed.
We have the face indicial notation,
sfi = -//+ i-min -//+ i-max, -//+ j-min -//+ j-max, -//+ k-min -//+ k-max; sd * # Crdfi our notation
As an example of the sphere in SC2i.tg we have
block 1 6 16 21; 1 6 16 21; 1 6 16 21; -1 -1 1 1; -1 -1 1 1; -1 -1 1 1;
sfi = -1 -4; -1 -4; -1 -4; sd 1
To deal with a hemisphere, we only specify 5 faces, and to have same size mesh along the 3 axis,
block 1 6 16 21; 1 6 11; 1 6 16 21; -1 -1 1 1; 0 1 1; -1 -1 1 1;

sfi = -1 -4; 1 -3; -1 -4; sd 1
The top face in the j direction is selected by the -3 in the j-direction.

Case Studies

1. Case Study: Hexa mesh for Cylindrical Pressure Vessel with Hemispherical Closure

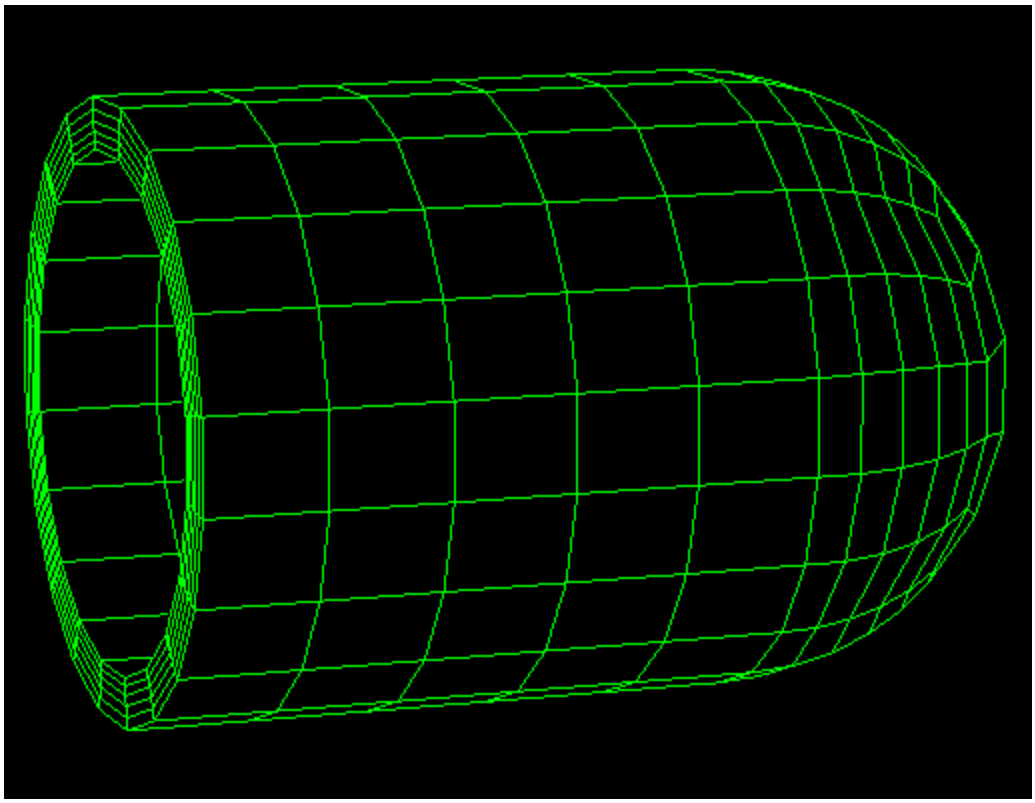


Fig. 4 a, Hexa Mesh for Pressure vessel

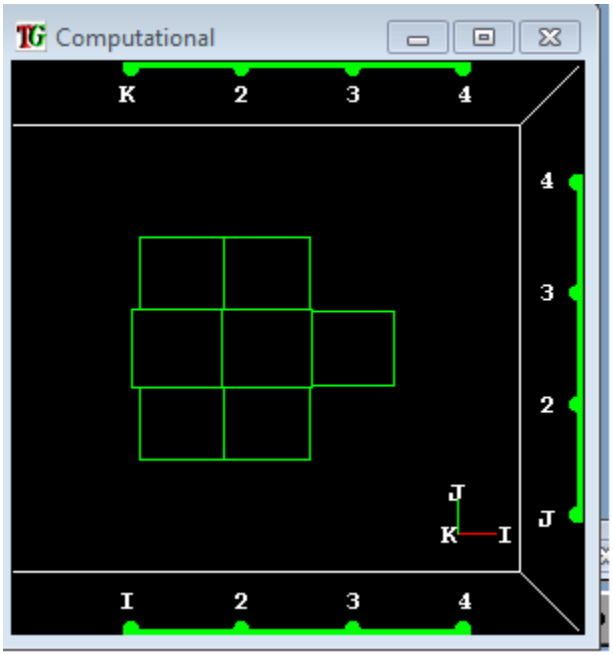


Fig. 4 b, Combined Topology for Cylinder and Sphere (half)

2. Case Study: Insertion of a Cylindrical Nozzle into Pressure Vessel.

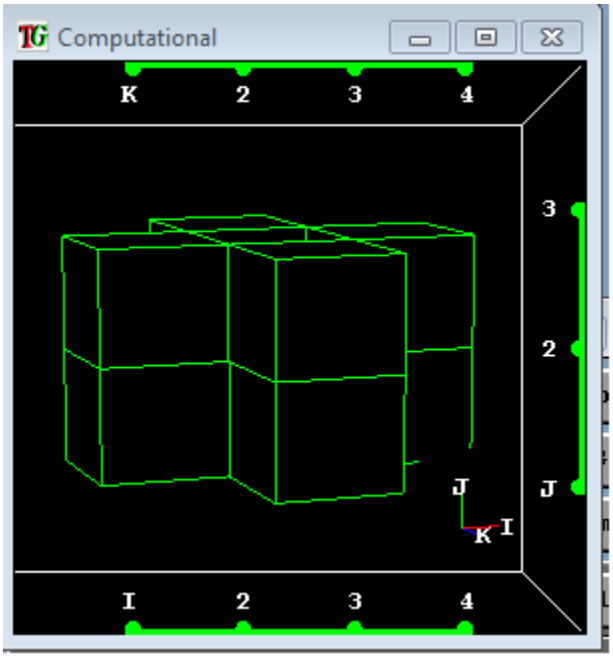


Fig. 5 a, Topology for Cylindrical Part before Insert

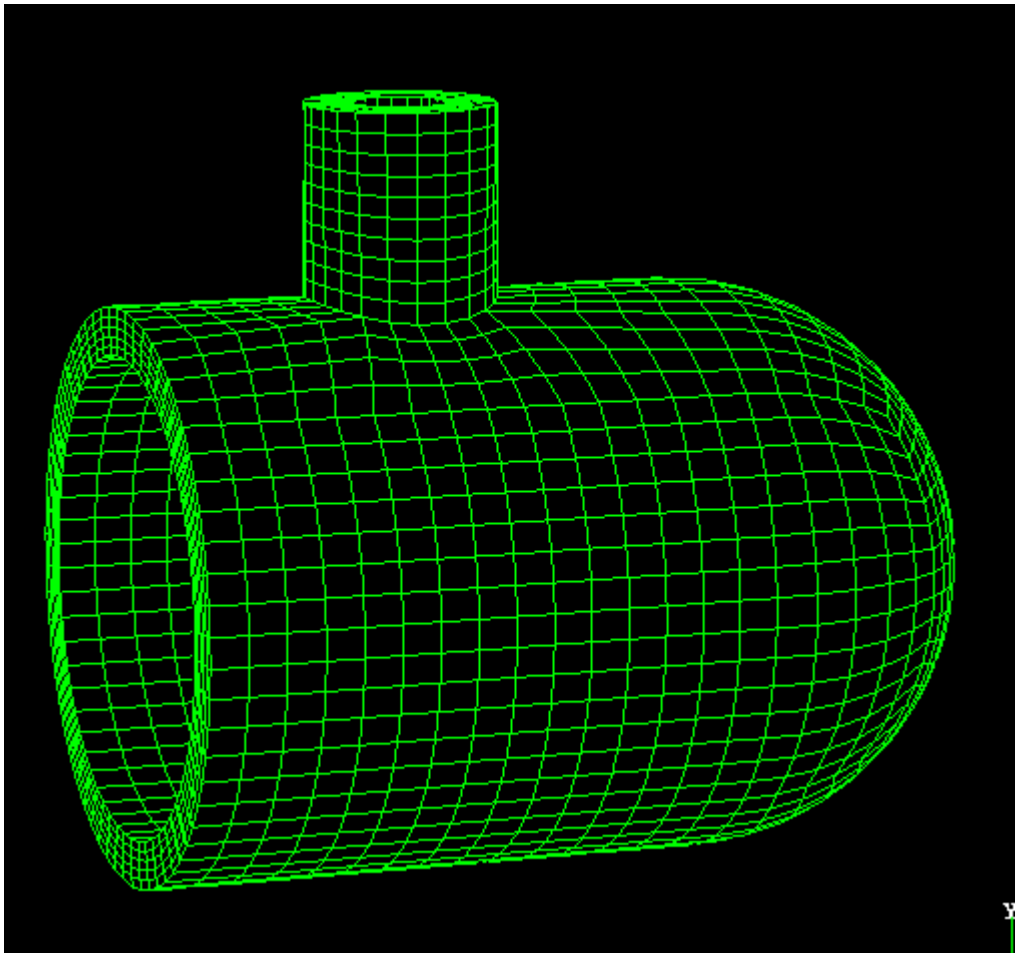


Fig. 5 b , Hexa Mesh for Pressure Vessel With Nozzle Insert

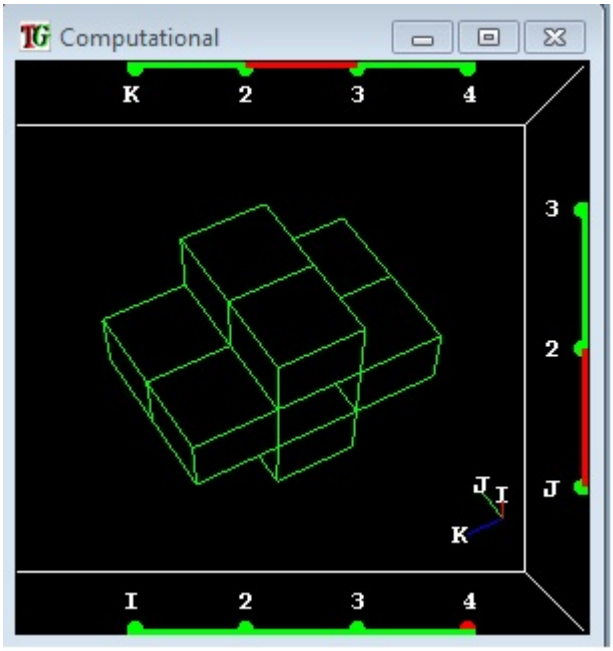


Fig. 6a Topology for insert. Same as fig. 5A but with different axis orientation.

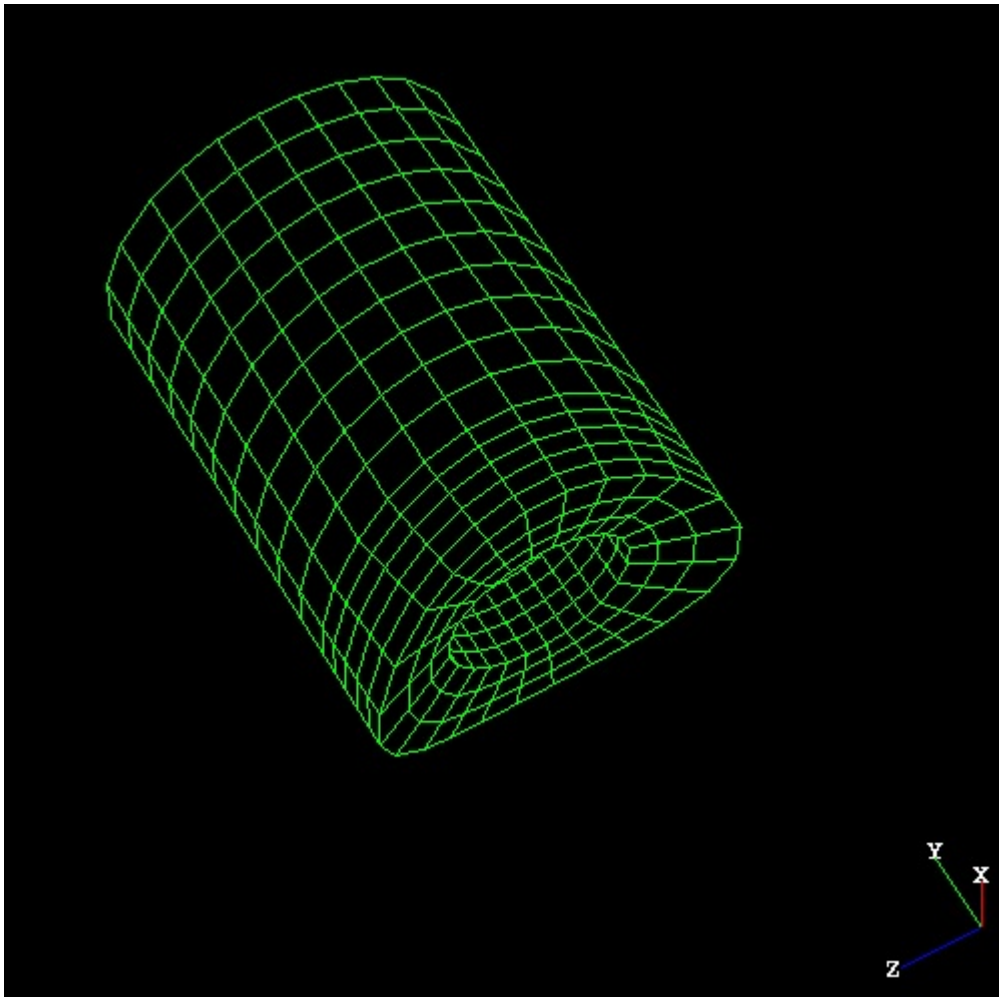


Fig. 6b Mesh for cylindrical insert.

Conclusions

A topological Domain Language has been developed to assist in the use of the **True*Grid*®** Hexa mesh generator. A Python program has been developed to convert scripts in the Domain Language to the journal format in **True*Grid*®**. The Language is useful in cases where the CAD model is created with Constructive Solid Modeling.

References

[1] P.V. Marcal, J.T. Fong, R. Rainsberger, L. Ma, Finite Element Analysis of a Pipe Elbow Weldment Creep-Fracture Problem Using an Extremely Accurate 27-node Tri-Quadratic Shell and Solid Element Formulation, Proc. 14th International Conf. on Pressure Vessel Technology, ICPVT-14, Sep. 23-26, 2015, Shanghai, China, 2015.

[2] J.T. Fong, J.J. Filliben, N.A. Heckert, P.V. Marcal, R. Rainsberger, L. Ma, Uncertainty Quantification and Extrapolation of Finite Element Method-estimated Stresses in a Cracked Pipe Elbow Weldment using a Logistic

Function Fit and a Nonlinear Least Square Algorithm, Proc. International Conf. on Pressure Vessel Technology, ICPVT-14, Sep. 23-26, Shanghai, China, 2015.

[3]P.V. Marcal, J.T. Fong, R. Rainsberger, L. Ma, High Accuracy Approach To Finite Element Analysis Using the Hexa 27-node Element, Proc. PVP-2016, July 17-21, Vancouver B.C., Canada, 2016

[4] P.V. Marcal, MPACT User Manual, Mpact Corp. , Oak Park, CA, 2001.

[5] R. Rainsberger, **True*Grid*®** User's Manual: A Guide and a Reference, Volumes I, II, and III, Version 3.0.0. Published by XYZ Scientific Applications, Inc., Pleasant Hill, CA 94523 , 2014, www.truegrid.com/pub/TGMAN300.1.pdf

APPENDIX A: EXAMPLE SCRIPTS

SC1.tdl

```
prog=FemFormat(mpact)
coord_1=Crdr(3D,11,11,11)
coordt=Crd(TP,(-1,1),(-1,1),(-1,1))
blk1=Block(coord_1,coordt)
orig=Pos(0,0,0)
dir=Dir(1,0,0)
cyl=Cyl(orig,dir)
faces=Crdf(3D,(1,2),(-1,-2),(-1,-2))
proj=Project(faces,cyl)
Pause
DisplayWindow(10,20)
Pause
endb1=EndBlock(blk1)
Merge
Mesh
```

SC1.tg, output from TDL2TG.py, see Fig. 1

```
mpact
block 1 11; 1 11; 1 11;-1 1 -1 1 -1 1
sd 1 cy 0 0 0 1 0 0 3
sfi 1 2; -1 -2; -1 -2; sd 1
interrupt
ry 20 rx 10
```

```
 center
 disp
 interrupt
endpart
merge
write
```

spcy.tdl for Case Study of Cylindrical Pressure Vessel with Hemispherical Closure

```
prog=FemFormat(mpact)
coord_1=Crdl(3D,(1 6 16 21),(1 6 16 21),(1 6 16 21))
coordt=Crd(TP,(-1 -1 1 1),(-2 0 1 1),(-1 -1 1 1))
blk1=Block(coord_1,coordt)
crdfp5=CrdfPermute(TP,(1, 2 ,0 ,3, 4),(3,4),( : ))
crdfp6=CrdfPermute(TP,(1, 2 ,0 ,3, 4),( : ),(1 ,2, 0 ,3 ,4))
crdfp7=CrdfPermute(TP,( : ),(3,4),(1 ,2, 0 ,3 ,4))
cruci=Cruciform(blk1,crdfp5,crdfp6,crdfp7)
crdfp8=CrdfPermute(TP,(2,3),(1,3),(2,3))
cross5=Cross(blk1,crdfp8)
orig5=Pos(0,0,0)
rad5=Rad(3)
sph1=Sph(orig5,rad5)
faces5=Crdfi(3D,(-1,-4),(2,-4),(-1,-4))
proj1=Project(faces5,sph1)
orig6=Pos(0,0,0)
rad6=Rad(2)
sph2=Sph(orig6,rad6)
faces6=Crdfi(3D,(-2,-3),(2,-3),(-2,-3))
proj2=Project(faces6,sph2)
orig=Pos(0,0,0)
dir=Dir(0,1,0)
rad=Rad(3)
```

```
cyl1=Cyl(orig,dir,rad)
faces=Crdfi(3D,(-1,-4),(1,2),(-1,-4))
proj=Project(faces,cyl1)
orig1=Pos(0,0,0)
dir1=Dir(0,1,0)
rad1=Rad(2)
cyl2=Cyl(orig1,dir1,rad1)
faces1=Crdfi(3D,(-2,-3),(1,2),(-2,-3))
proj3=Project(faces1,cyl2)
Pause
DisplayWindow(10,20)
Pause
endb1=EndBlock(blk1)
Merge
Mesh
```

spcy_td.tg output for Case Study of Pressure Vessel, see Fig. 4

```
mpact
block 1 6 16 21 ; 1 6 16 21 ; 1 6 16 21 ; -1 -1 1 1 ;-2 0 1 1 ;-1 -1 1 1 ;
dei 1 2 0 3 4 ;3 4 ; ;
dei 1 2 0 3 4 ; ;1 2 0 3 4 ;
dei  ;3 4 ;1 2 0 3 4 ;
dei 2 3 ;1 3 ;2 3 ;
sd 1 sp 0 0 0 3
sfi -1 -4; 2 -4; -1 -4;  sd 1
sd 2 sp 0 0 0 2
sfi -2 -3; 2 -3; -2 -3;  sd 2
sd 3 cy 0 0 0 0 1 0 3
sfi -1 -4; 1 2; -1 -4;  sd 3
sd 4 cy 0 0 0 0 1 0 2
sfi -2 -3; 1 2; -2 -3;  sd 4
interrupt
```

ry 20 rx 10

center

disp

interrupt

endpart

merge

write

spcys2.tdl for Case Study with insert.

```
prog=FemFormat(mpact)
coord_1=Crdl(3D,(1 6 11 16),(1 6 11 16),(1 6 11 16))
coordt=Crd(TP,(0 12 18 18),(-6 -6 6 6),(-6 -6 6 6))
blk1=Block(coord_1,coordt)
sub1=SubPos(i,(1,5))
crdfs=CrdfsPerm(sub1)
sub2=SubPos(j,(2,10))
sub3=SubPos(k,(2,10))
sub4=SubPos(i,(1,5))
crdfs1=CrdfsPerm(sub2,sub3,sub4)
crdfp5=CrdfPermute(TP,(3 ,4),(1, 2 ,0 ,3, 4),(2,3))
crdfp6=CrdfPermute(TP,( : ),(1, 2 ,0 ,3, 4),(1 ,2, 0 ,3 ,4))
crdfp7=CrdfPermute(TP,( 3,4 ),(2,3),(1 ,2, 0 ,3 ,4))
cruci=Cruciform(blk1,crdfp6,crdfp5,crdfp7)
crdfp8=CrdfPermute(TP,(1,3),(2,3),(2,3))
cross5=Cross(blk1,crdfp8)
orig9=Pos(12,0,0)
rad9=Rad(5)
sph1=Sph(orig9,rad9)
faces8=Crdfi(3D,(2,-3),(-2,-3),(-2,-3))
```

```
proj1=Project(faces8,sph1)
orig6=Pos(12,0,0)
rad6=Rad(6)
sph2=Sph(orig6,rad6)
faces7=Crdfi(3D,(2,-4),(-1,-4),(-1,-4))
proj2=Project(faces7,sph2)
orig=Pos(0,0,0)
dir=Dir(1,0,0)
rad=Rad(5)
cyl1=Cyl(orig,dir,rad)
faces1=Crdfi(3D,(1,2),(-2,-3),(-2,-3))
proj=Project(faces1,cyl1)
orig1=Pos(0,0,0)
dir1=Dir(1,0,0)
rad1=Rad(6)
cyl2=Cyl(orig1,dir1,rad1)
faces=Crdfi(3D,(1,2),(-1,-4),(-1,-4))
proj3=Project(faces,cyl2)
sub5=SubPos(k,(2,4))
crdfh1=CrdfhPerm(sub5)
sub6=SubPos(k,(3,7))
crdfh2=CrdfhPerm(sub6)
sub7=SubPos(i,(1,5))
crdfh3=CrdfhPerm(sub7)
sub8=SubPos(i,(2,5))
crdfh4=CrdfhPerm(sub8)
crdfp9=CrdfPermute(TP,(2,3),(3,4),(3,4))
cross6=Cross(blk1,crdfp9)
orig3=Pos(6,0,0)
```

```
dir3=Dir(0,1,0)
rad3=Rad(2)
cyl3=Cyl(orig3,dir3,rad3)
faces3=Crdfi(3D,(-2,-3)(3,4),(-3,-4))
proj5=Project(faces3,cyl3)
crdf1=Crdf(3D,(2, 3, 3),( 2, 4, 4 ))
crdf2=Crdf(3D,(3, 3, 3),( 3, 4, 4 ))
crdf3=Crdf(3D,(2, 3,3),( 3, 4, 3 ))
crdf4=Crdf(3D,(2, 3, 4),( 3, 4, 4 ))
tie1=TieFace(crdf1,1)
tie2=TieFace(crdf3,2)
tie3=TieFace(crdf2,3)
tie4=TieFace(crdf4,4)
Pause
endb1=EndBlock(blk1)
coord_2=Crdl(3D,(1 4 9 12),(1 6 16),( 1 4 11 14 ))
coordt2=Crd(TP,(5 5 7 7) ,(5 6 10),(-1 -1 1 1))
blk2=Block(coord_2,coordt2)
crdfp10=CrdfPermute(TP,(1, 2 ,0 ,3, 4),( : ),(1 ,2, 0 ,3 ,4))
cross7=Cross(blk2,crdfp10)
crdfp11=CrdfPermute(TP,(2,3),( : ),(2,3))
cross8=Cross(blk2,crdfp11)
crdf5=Crdf(3D,( 1, 1, 2),( 1, 2, 3))
crdf6=Crdf(3D,(2, 1, 1),( 3, 2, 1))
crdf7=Crdf(3D,( 4, 1, 2),( 4, 2, 3))
crdf8=Crdf(3D,( 2, 1, 4),( 3, 2, 4))
tie5=TieFace(crdf5,1)
tie7=TieFace(crdf7,3)
tie6=TieFace(crdf6,2)
```

```
tie8=TieFace(crdf8,4)
orig4=Pos(6,0,0)
dir4=Dir(0,1,0)
rad4=Rad(2)
cyl4=Cyl(orig4,dir4,rad4)
faces4=Crdfi(3D,(-1,-4),(2, 3),(-1,-4))
proj6=Project(faces4,cyl4)
orig5=Pos(6,0,0)
dir5=Dir(0,1,0)
rad5=Rad(1)
cyl5=Cyl(orig5,dir5,rad5)
faces5=Crdfi(3D,(-2,-3),( : ),(-2,-3))
proj7=Project(faces5,cyl5)
DisplayWindow(10,20)
Pause
endb1=EndBlock(blk2)
Merge
Mesh
```

spcys2_td.tg output from Case Study, see figs. 5-6.

```
mpact
block 1 6 11 16 ; 1 6 11 16 ; 1 6 11 16 ; 0 12 18 18 ;-6 -6 6 6 ;-6 -6 6 6 ;
lmseq i  1 5
lmseq j  2 10  lmseq k  2 10  lmseq i  1 5
dei  ;1 2 0 3 4 ;1 2 0 3 4 ;
dei 3 4 ;1 2 0 3 4 ;2 3 ;
dei 3 4 ;2 3 ;1 2 0 3 4 ;
dei 1 3 ;2 3 ;2 3 ;
sd 1 sp 12 0 0 5
sfi 2 -3; -2 -3; -2 -3;  sd 1
```

sd 2 sp 12 0 0 6

sfi 2 -4; -1 -4; -1 -4; sd 2

sd 3 cy 0 0 0 1 0 0 5

sfi 1 2; -2 -3; -2 -3; sd 3

sd 4 cy 0 0 0 1 0 0 6

sfi 1 2; -1 -4; -1 -4; sd 4

insprt 1 6 2 4

insprt 1 6 3 7

insprt 1 2 1 5

insprt 1 2 2 5

dei 2 3 ;3 4 ;3 4 ;

sd 5 cy 6 0 0 0 1 0 2

sfi -2 -3; 3 4; -3 -4; sd 5

bb 2 3 3 2 4 4 1;

bb 2 3 3 3 4 3 2;

bb 3 3 3 3 4 4 3;

bb 2 3 4 3 4 4 4;

interrupt

endpart

block 1 4 9 12 ; 1 6 16 ; 1 4 11 14 ; 5 5 7 7 ;5 6 10 ;-1 -1 1 1 ;

dei 1 2 0 3 4 ; ;1 2 0 3 4 ;

dei 2 3 ; ;2 3 ;

bb 1 1 2 1 2 3 1;

bb 4 1 2 4 2 3 3;

bb 2 1 1 3 2 1 2;

bb 2 1 4 3 2 4 4;

sd 6 cy 6 0 0 0 1 0 2

sfi -1 -4; 2 3; -1 -4; sd 6

sd 7 cy 6 0 0 0 1 0 1

sfi -2 -3; ; -2 -3; sd 7

ry 20 rx 10

center

disp

interrupt

endpart

merge

write

Numerical study of the effects of strain rate on the behaviour of dynamically penetrating anchors in clay

***H. Sabetamal[1], J. P. Carter[1], M. Nazem[2] and S.W. Sloan[1]**

[1] ARC Centre of Excellence for Geotechnical Science and Engineering, The University of Newcastle, NSW, Australia.
[2] School of Engineering, RMIT University, Melbourne, Australia

*Presenting author: Hassan.sabetamal@uon.edu.au
*Corresponding author: Hassan.sabetamal@uon.edu.au

Abstract

The installation of torpedo anchors at high impact velocities imposes high strain rates in the surrounding soil. The high strain rates enhance the mobilised undrained shear strength compared to that measured statically by laboratory or in situ tests. To illustrate the implications of such high strain rates for the behaviour of dynamic anchors, large deformation Finite Element (FE) analyses were carried out. The numerical FE scheme is based on a dynamic coupled effective stress framework with the Modified Cam Clay constitutive model. The soil constitutive model is adapted to incorporate the dependence of clay behaviour on strain rata. In order to assess the validity of the numerical scheme, some laboratory tests on model free falling penetrometers have been simulated. The results indicate that overall the agreement between computations and measurements is good. It is seen that the generation of excess pore pressure around dynamically installed anchors and the frictional resistance at the soil-anchor interface are significantly affected by the strain rate. Moreover, increased strain rate dependency of the soil leads to a marked reduction in the embedment depth, reflecting a noticeable increase in the soil penetration resistance.

Keywords: Torpedo anchors, Strain rate dependency, Dynamic coupled analysis, Large deformations.

Introduction

Deepwater oil and gas reserves have become an important component of global energy supply, and the recovery of hydrocarbons from these regions has resulted in a broad range of relatively new engineering practices. The scale of the foundation and anchoring elements, along with their novel construction and installation techniques, are key aspects of offshore geotechnical engineering. Depending on the depth of the seabed, offshore structures may be divided into two main types: fixed and floating structures. All floating systems used in deep waters require moorings and ultimately some form of anchor on the seabed, which typically include surface (gravity) and embedded anchors. Dynamically installed anchors (*i.e.*, torpedo anchors and deep penetrating anchors) are promising embedment systems used in ultra-deep waters, mainly due to their installation cost advantage compared to other systems such as drag embedment anchors and suction embedded plate anchors. A torpedo anchor is embedded using the kinetic energy attained by gravity free fall through the water column, so that its installation cost is largely independent of water depth. This anchoring system also has a relatively lower fabrication cost which often makes it more attractive than suction caissons.

Despite the economic advantages afforded by dynamically installed anchors, there remain significant uncertainties in the prediction of the embedment depth and the anchor holding capacity. The prediction of the embedment depth is complicated by the very high strain rate adjacent to the soil-anchor interface (resulting from high penetration velocities) and hydrodynamic aspects which can involve inertial and viscous drag forces.

It is well known that the mechanical behaviour of clayey soil is affected by the rate of induced strains. Typically, the undrained strength increases with increasing shear strain rate (e.g., [1]-[5]). Therefore, for high velocity penetrations, the soil resistance under fully undrained conditions might be expected to vary as a function of the strain rate. Numerical studies have actually shown that the effect of the strain rate on the shear strength of the soil should not necessarily be ignored in problems involving the fast penetration of objects into soil layers (*e.g.*, [6][7]). However, there is a lack of knowledge on how the excess pore pressures and frictional forces at the anchor-soil interface might be affected by strain rate effects.

Sabetamal *et al.* [9][10] presented rigorous coupled analyses for a few free falling torpedo anchors. These initial studies reported successful simulations of the installation process and reconsolidation stage of torpedo anchors, and revealed the pattern in which excess pore pressures are generated and dissipated. In this paper, we extend our earlier study to capture the effects of strain rate on the behaviour of torpedo anchors. Accordingly, some numerical findings are reported on the performance of this anchoring system during the installation phase, taking both the strain rate and inertial drag forces into account.

Numerical Scheme

Problems in offshore geomechanics are typically characterized by the existence of hydrodynamic and cyclic loadings, large deformations, extreme soil-structure interactions and soil disturbance typically due to installation effects. The installation of offshore structures, such as a dynamically embedded anchor, is usually an undrained process during which excess pore pressures are generated. The time scale of consolidation is also important for predicting the holding capacity of these anchors under different loading events. A fully coupled analysis is then required to incorporate pore-fluid pressure development and its subsequent dissipation.

Governing equations

A continuum approach based on the theory of mixtures [11] and the concept of volume fractions [12] is employed to derive the governing equations. Sabetamal [13] provided a detailed account of the governing differential equations and the corresponding weak statements that form the basis of our finite element (FE) modelling. A mixed formulation were selected to describe both incompressible and compressible fluids, in which the resulting formulation predicts all field variables, including the solid matrix displacements **U**, pore-fluid pressure **P**, and Darcy velocity of the pore fluid $\mathbf{V}_r$. The resulting equation system governing the behaviour of the soil-water mixture may be written in matrix form as

$$\begin{bmatrix} \mathbf{M}_{ss} & 0 & \mathbf{M}_{sr} \\ 0 & 0 & 0 \\ \mathbf{M}_{rs} & 0 & \mathbf{M}_{rr} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{U}} \\ \ddot{\mathbf{P}} \\ \dot{\mathbf{V}}_r \end{bmatrix} + \begin{bmatrix} \mathbf{C}_s & 0 & 0 \\ \mathbf{C}_{ps} & -\mathbf{C}_{pr} & \mathbf{C}_{pp} \\ 0 & \mathbf{C}_{rr} & 0 \end{bmatrix} \begin{bmatrix} \dot{\mathbf{U}} \\ \dot{\mathbf{P}} \\ \mathbf{V}_r \end{bmatrix} + \begin{bmatrix} \mathbf{K}_\sigma & \mathbf{K}_{sp} & 0 \\ 0 & \mathbf{K}_{rp} & 0 \\ 0 & \mathbf{K}_{pp} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{P} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^s \\ \mathbf{F}^p \\ \mathbf{F}^r \end{bmatrix} \quad (1)$$

where $\mathbf{M}_{ss}$, $\mathbf{M}_{rr}$, $\mathbf{M}_{rs}=\mathbf{M}_{sr}^{T}$ and $\mathbf{C}_{\alpha\beta}$ are the solid mass, fluid mass, coupled fluid mass and damping matrices, respectively. $\mathbf{K}_{\sigma}$ and $\mathbf{K}_{pp}$ are, respectively, the stiffness and compressibility matrices while $\mathbf{K}_{\alpha\beta}$ represent coupling matrices and $\mathbf{F}^{s}$, $\mathbf{F}^{p}$ and $\mathbf{F}^{r}$ are the vectors of external nodal forces.

Large deformation and mesh refinement

To consider large deformation phenomena and avoid possible mesh distortions, the traditional numerical methods established within a Lagrangian framework are typically replaced by those based on the framework of the Arbitrary Lagrangian Eulerian (ALE) method. ALE approaches for geotechnical applications can be divided into two groups: mesh based methods [14]-[17] and particle based schemes such as the material point method [18][19]. The mesh based ALE schemes used in geotechnical engineering may be divided into three categories: the Remeshing and Interpolation Technique involving Small Strains (RITSS) [15], the ALE scheme [20], and the Coupled Eulerian-Lagrangian (CEL) approach. Wang *et al.* [21] compared the performances of the three approaches for some benchmark problems covering static, consolidation and dynamic geotechnical applications. It was concluded that the RITTS and ALE schemes predict close results whereas, for dynamic problems, the results obtained from the CEL approach differ from those predicted with the RITTS and ALE methods. The ALE scheme [17] is incorporated in this study to handle large deformations. In this approach, some special care should be taken for the solution of the advection equations, where transport of the material and the current solution state through the mesh is considered along the streamlines of the advective flow, provided by the convective velocity. In an ALE framework, this corresponds to a relocation of the FE nodes by the mesh motion scheme, while the material is held fixed in space. Most advection schemes, especially the classical first-order methods, show highly numerical diffusive properties. This appears to be crucial for the cases that hardening/softening is involved in the solution by some constitutive models such as the Modified Cam Clay (MCC) model. The transport step has to be then split into multiple advection steps, based on intermediate mesh configurations, and an advection scheme with only a small amount of diffusion is necessary to retain the special shape of the solution variables properly [22].

Interface modelling

The so called one pass node-to-segment (NTS) discretisation method is commonly used to analyse large sliding and large deformation problems of contact mechanics [23][24]. Sabetamal *et al.* [10] applied the NTS scheme to analyse some coupled dynamic problems and observed that smooth discretisation of the contact interface between soil and structure is a crucial factor to avoid severe oscillations in the predicted dynamic forces and pore fluid pressures. It is also noted that a consistent coupling of the NTS contact with elements of a higher order is not possible because contact constraints are only fulfilled locally at a number of finite connection points. In contrast, the mortar segment-to-segment approach [24][25] considers the enforcement of contact constraints in a weak integral form so that high-order approximation functions can be used to interpolate different field variables. The use of high order elements also provides the possibility to explicitly incorporate smooth continuous geometries in the FE model, thus avoiding the numerical oscillations encountered in NTS approach. Sabetamal *et al.* [26] developed and applied a frictionless mortar scheme to model some dynamic problems of two phase saturated problems. In this paper, we use an extended form of the scheme which can also model frictional interfaces embedded within two phase saturated porous media [13].

Strain rate effect

The dependence of undrained shear strength of soil on applied rate of strain has long been recognized [27] and studied extensively both in triaxial compression tests (e.g., [28][3]) and vane shear tests (e.g., [29][4]). The dependence of shear strength s_u on strain rate $\dot{\gamma}$ maybe characterised in terms of a semi-logarithmic relation [1]

$$s_u = s_{u_{ref}} \left[1 + \eta \log \left(\frac{\dot{\gamma}}{\dot{\gamma}_{ref}} \right) \right] \tag{2}$$

where $s_{u_{ref}}$ is the reference undrained shear strength measured at the reference strain rate $\dot{\gamma}_{ref}$ and η denotes the rate of increase per decade with a suggested range of 0.05 to 0.20. In this study, the nonlinear behaviour of the solid constituent in the two phase saturated mixture is captured by the MCC soil model. Typical undrained strain rates in standard laboratory tests measure around 0.01/h (3×10^{-6} s^{-1}). Assuming this rate as the reference strain rate, the initial undrained shear strength predicted by the constitutive model parameters will correspond to $s_{u_{ref}}$. Fig. 1 depicts the locus of normal consolidation line (*NCL*) and overconsolidation line in v-ln(p') space for an overconsolidated soil (v_i, p'_i), where v_i denotes specific volume, p' is mean effective stress, N is the value of specific volume at unit pressure, λ is the slope of the *NCL*, κ represents the slope of unloading-reloading line and q is deivatoric stress.

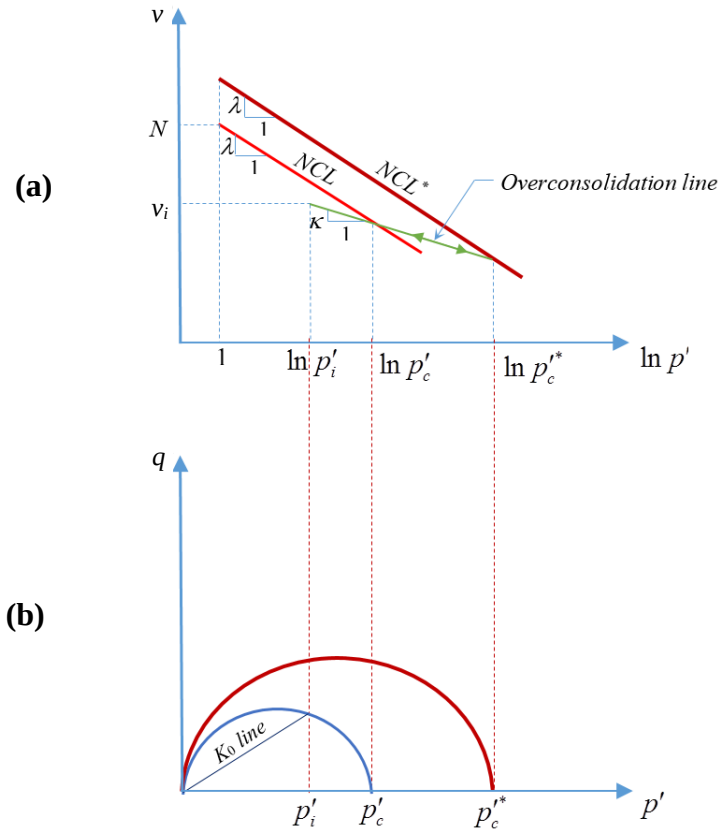


Figure 1. (a) Locus of NCL line; and (b) *q*-*p*' plot

It should be noted that the critical state friction angle is expected to be unaffected by the strain rate, as suggested by numerous experimental studies (e.g., [30]), whereas the normal consolidation line (*NCL*) of clays in the *v*-ln (p') space moves upward with an increase in strain rate (*NCL**). This shift of the *NCL* with increasing strain rate has been observed by many researchers (e.g., [31]-[33]). For undrained conditions, no volume change is allowed so that the specific volume v_i should be constant and must lie on the same unloading-reloading line. Therefore, the updated preconsolidation pressure $p_c'^*$ due to strain rate increase must be at the intersection of *NCL** and the overconsolidation line. Consequently, the upward shift of the *NCL* as a function of strain rate corresponds to an increase in the preconsolidation pressure [34] or overconsolidation ratio (*OCR*) [35], and implies that the soil becomes more dilatant and exhibits larger stiffness and peak undrained shear strength, as observed in reality. The increase in *OCR* adds to the increase in stiffness through the constitutive equations for the plastic modulus and elastic moduli.

To relate the increase in *OCR* and the corresponding preconsolidation pressure to the strain rate increase, Eq. (2) is utilised in this study, along with the theoretical formula that predicts the undrained shear strength based on the MCC model parameters [36]. Consequently, rate-independent plasticity theory is employed to simulate the rate dependent behaviour, avoiding the need to adopt numerically expensive viscoplastic stress-strain integration schemes. Therefore, the adopted model assumes that soil elements at the same initial stress conditions will show different responses if subjected to different strain rates. This is reflected by *OCR* changes and the corresponding enlargements of the yield surface.

Inertial drag force

It seems rational to assume that an inertial drag force exists during penetration of objects into very soft viscous clay, analogous to the hydrodynamic drag experienced by an object passing through water. To show the effect of the drag force on the velocity profile, an inertial drag force is incorporated in the analysis using the following relation

$$F_d = \frac{1}{2} C_d \rho_s A_p V^2 \tag{3}$$

where C_d is the drag coefficient, ρ_s is the density of the soil, A_p is the projected frontal area of the anchor, and V is the current anchor velocity. An approximation of the average drag coefficient equal to 0.7 was suggested by True [37] for a variety of penetrometer geometries and velocities. However, hydrodynamic studies have indicated considerably smaller drag coefficients. Numerical analysis presented by Richardson [38] showed that the drag coefficient C_d decreases with increasing aspect ratio of the penetrometer and ultimately approaches a constant value, which for finless torpedo anchors decreases from 0.35 to a constant value of 0.23 for $L/D \geq 4$, where L and D denote anchor length and diameter, respectively.

Numerical Examples

The numerical framework described previously has been implemented into an in-house FE code, SNAC. This code is employed here to carry out some coupled simulation of dynamic anchors. First, simulation of a model penetrometer is conducted and the analysis results are compared with the corresponding centrifuge data. Then, a series of analyses are performed to study the effect of strain rate on the behaviour of torpedo anchors.

Validation against centrifuge test data

Chow *et al*. [39] reported data from a centrifuge test carried out on a model penetrometer free falling into kaolin clay. The penetrometer had a 60° cone tip and a prototype shaft diameter and length of 1.0 m and 12 m, respectively, and a mass of 28130 kg. The penetrometer achieved impact velocities ranging between 4.7 and 15.6 m/s with corresponding final embedment depths in the range 10.2-16.7 m at prototype scale. The undrained shear strength of the soil $s_u = 1.13z$ kPa was deduced from T-bar penetration tests where z denotes the soil depth in metres. The soil properties are listed in Table 1.

Table 1. Soil parameters

Parameter	Value
Friction angle	$\phi' = 23°$
Slope of normally consolidated line in *e-ln(p')* space	$\lambda = 0.205$
Slope of unloading-reloading line in *e-ln(p')* space	$\kappa = 0.044$
Initial void ratio	$e_0 = 2.14$
Over consolidation ratio	$OCR = 1$
Poisson's ratio	$\upsilon' = 0.3$
Saturated bulk unit weight	$\gamma_{sat} = 17$ kN/m^3
Unit weight of water	$\gamma_w = 10$ kN/m^3
Permeability of soil	$k = 5\times10^{-9}$ m/s

Note: p' is the mean effective stress

Fig. 2 depicts the axisymmetric FE mesh and the corresponding boundary conditions adopted for the numerical simulation. The mesh comprises 3,416 triangular elements and 7,028 nodes.

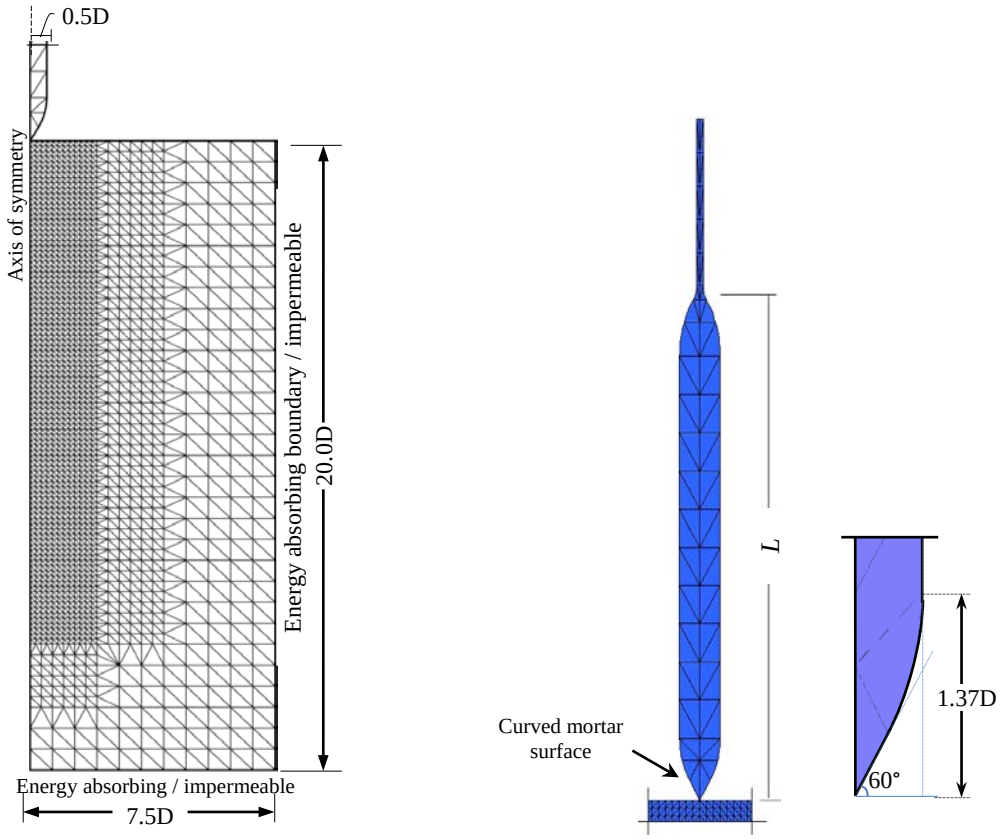


Figure 2. (a) FE model mesh; (b) anchor geometry

The radial thickness of the soil elements underneath the penetrometer is equal to one-third of its shaft radius. Discretisation of the geometry of the penetrometer with quadratic mortar elements facilitates curved surfaces at the cone and top of the anchor (Fig. 2b). Two impact velocities of 4.7 m/s and 6.1m/s were considered in the numerical simulations. The strain rate parameter, the drag coefficient and the friction coefficient at the interface were assumed to be $\eta = 0.2$, $C_d = 0.23$ and $\mu = 0.25$, respectively.

Fig. 3 shows the penetration profile predicted by the numerical analyses and the ultimate penetration depths as measured in the centrifuge test. Good agreement of the ultimate penetration can be observed for the two analyses. The computed anchor tip embedment depths for the impact velocities 4.7m/s and 6.1m/s are, respectively, 10.45m and 12.23m which are only 2.1% and 3% greater than the measured values, providing some experimental validation of the proposed numerical approach and its predictions.

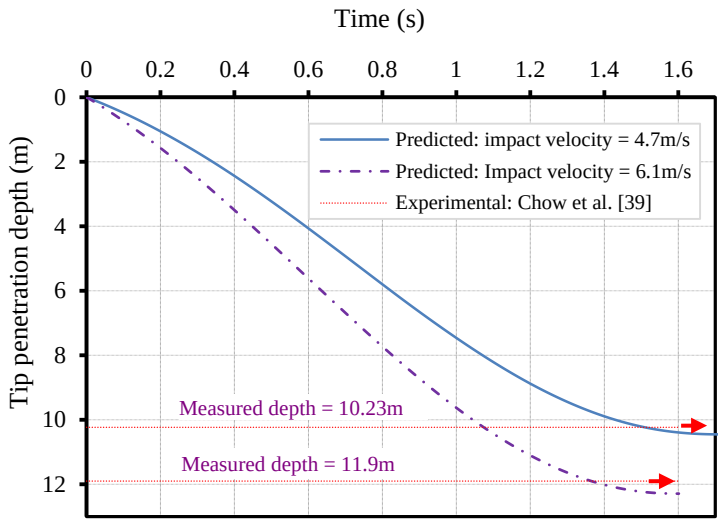


Figure 3. Comparison between numerical prediction and centrifuge test data

Strain rate effect on the behaviour of torpedo anchor

A rigid finless torpedo anchor falling freely into a normally consolidated kaolin clay is analysed in this section. The effect of strain rate on the behaviour of torpedo anchor is then studied in terms of penetration depth, pore pressure generation and frictional resistance. The boundary conditions and geometry of the mesh and torpedo anchor are similar to those adopted in the previous section (Fig. 2), except that the buoyant weight of the anchor is now 150 kN. In order to provide a rather detailed overview of anchor behaviour, two sets of analyses are presented. The first set of simulations assumes a frictionless interface between the soil and anchor so as to study the effects of strain rate only. The second set of analyses incorporates a frictional interface and reveals some practical and important aspects of dynamic anchor behaviour.

Frictionless interface

Fig. 4 depicts the change in the equivalent (apparent) *OCR* value at a penetration depth of 5D for rate parameters of $\eta = 0.15$, and 0.20. The apparent *OCR* value generally increases during penetration and for the rate parameter of $\eta = 0.15$ it reaches a maximum value of 2.7 at some Gauss points, noting that the initial value of *OCR* was 1.0 (Table. 1). The soil elements

within a zone around the cone of the advancing torpedo undergo very high strain rates so that the equivalent value of *OCR* is noticeably increased for that zone. During anchor penetration the shear strain rate varies throughout the soil body in which for soil elements displaced from the tip zone to the anchor shaft the strain rate is alleviated, resulting in decreased magnitudes of the apparent *OCR* along the shaft. However, the final value is still larger than the initial *OCR*. Increasing the strain rate parameter to 0.20, increases the maximum value of apparent *OCR* to 4.1 (Fig. 4b).

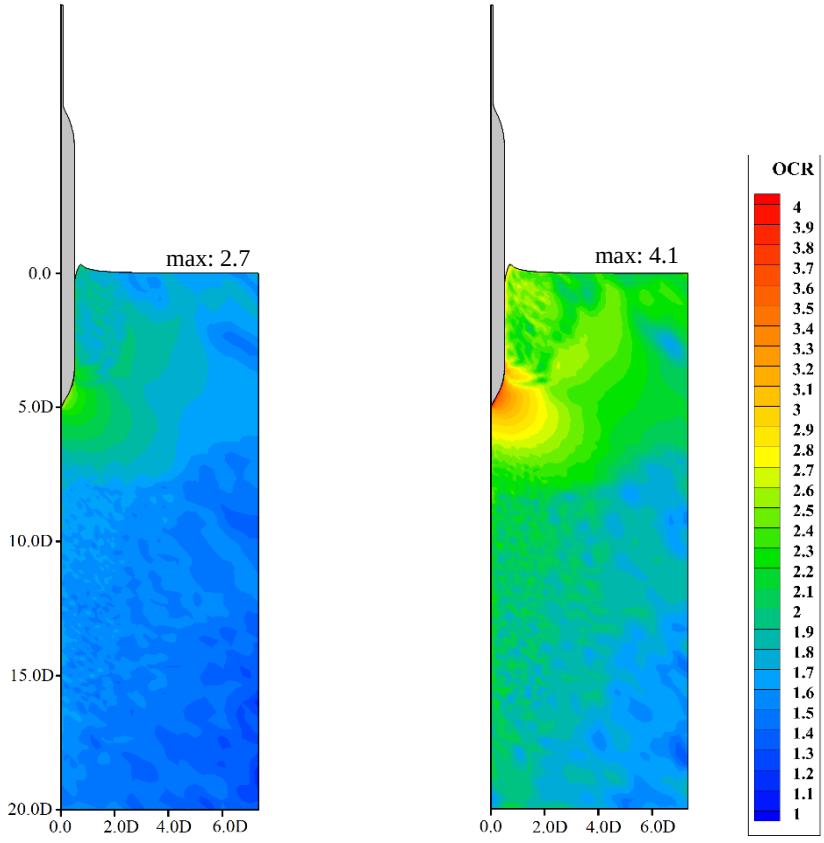


Figure 4. Apparent OCR value evaluated based on the strain rate within the soil body at a penetration depth of 5.0D: (a) η = 0.15; (b) η = 0.20

The soil resistance profile is depicted in Fig. 5 for two values of the rate parameter. It is observed that the total penetration resistance increases for the rate dependent case and the embedment depth is decreased, accordingly. The soil resistance at the end of installation is about 65% larger for the rate dependent case (η = 0.2) compared with the rate-independent one at the same penetration depth.

The embedment depth for the rate independent case is 13.9D whereas it decreases to 8.7D when the rate parameters is 0.20. Therefore, it can be seen that the increases in soil resistance due to strain rate effect is a key factor in the analysis of dynamically penetrating anchors. Although the most of experimental results on free falling anchors have identified the strain rate effect on the ultimate embedment depth, they have not described how strain rate may influence the generation of excess pore pressures and sleeve frictional force. These are explained as follows.

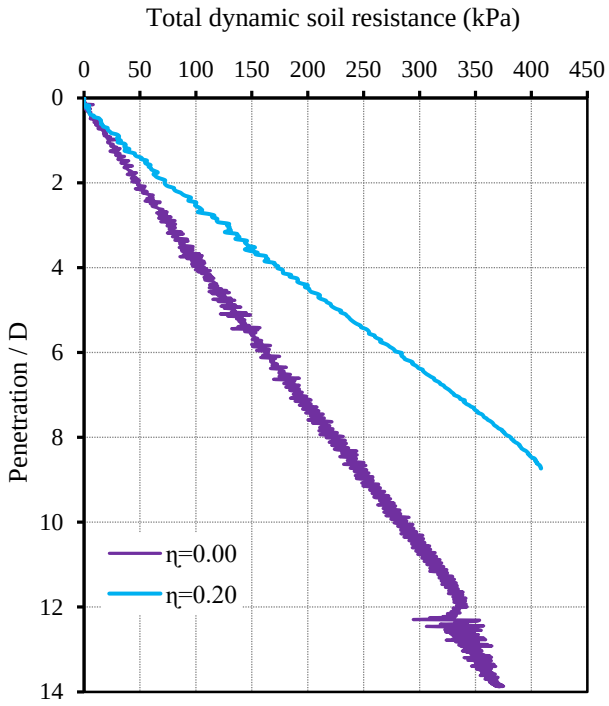


Figure 5. Total dynamic soil resistance profile for different values of rate parameter (η)

Fig. 6 depicts two excess pore pressure contour plots with rate parameters of $\eta = 0.0$ and $\eta = 0.20$. It is seen that, for the rate independent normally consolidated clay (Fig. 6a), a compressive excess pore pressure bulb is typically formed around the anchor tip and shaft. This bulb extends a distance of approximately 4D in the radial direction and about 1D in the vertical direction, as measured from the anchor tip. The maximum compressive values are developed at the anchor tip (~210 kPa) and extend to its shoulder (~160 kPa). Moreover, a tensile region (~-40kPa) is located at a distance of about 2D vertically underneath the anchor tip. This is due to development of plastic expansion (softening) region beneath the pile tip after the compression zone.

A similar plot for the rate dependent case is presented in Fig. 6b. It is observed that a region of suction has been locally created around the cone and also within a thin layer of soil along and adjacent to the anchor shaft. The creation of this suction zone (due to elasto-plastic expansion of soil) is merely a consequence of the high strain rate and the corresponding increase of the apparent *OCR* value. As observed in Fig. 4, soil elements around the conical section experience the highest strain rates and correspondingly much larger values of suction pore pressures (~ -600 kPa) are detected (Fig. 6b). This situation of high strain rates is also combined with the vertical stress relief that happens near the cone shoulder and leads to a more pronounced dilative behaviour of the soil. The normal stress relief may occur at a specific location depending on the geometry of anchor tip implying that the geometry of anchor tip may considerably influence the generation of excess pore pressures. It is also emphasised that the developed suction pore pressures can cause desaturation of the pore pressure measuring systems in experimental tests, and that reliable pore pressure data may not be consistently obtained. Therefore, the finding of a thin zone of suction around the anchor may have important consequences for pore pressure measurements made and interpreted using a conventional cone penetrometer (CPT).

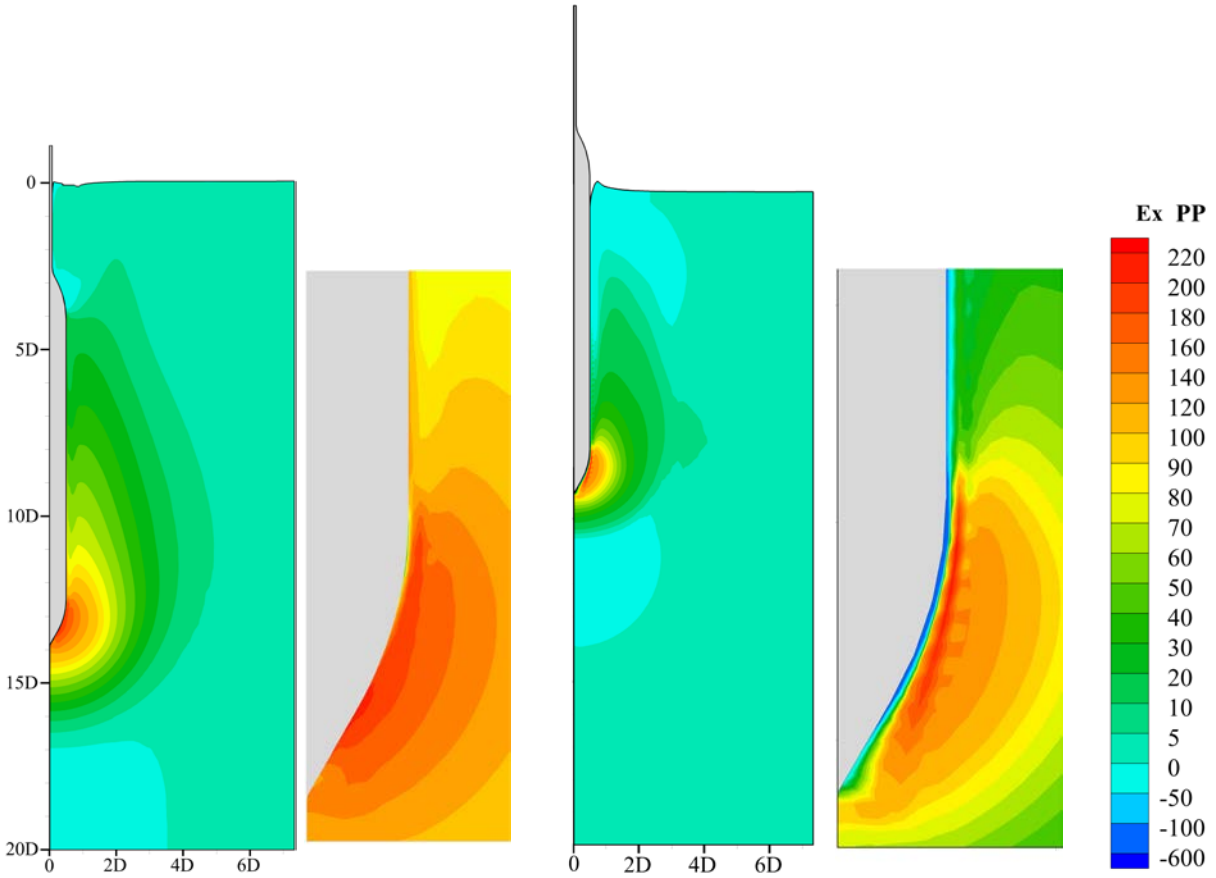


Figure 6. Excess pore pressure (kPa) contour plots: (a) η = 0 and (b) η = 0.2

Frictional interface

It might reasonably be deduced that the tangential frictional force developed at the anchor-soil interface would not be significant because of the undrained behaviour of the soil (*i.e.*, due to the expected lower effective stresses at the interface). However, the numerical results from the previous section revealed that a thin layer of tensile excess pore pressure is actually created along almost the entire length of the torpedo shaft during the installation process. This will increase the effective stresses at the soil-anchor interface and lead to higher frictional forces.

Fig. 7 depicts the soil resistance profile for a rate parameter η = 0.2 with friction coefficients μ = 0 and 0.2. The embedment depth decreases when the friction coefficient is 0.2, as expected. For the frictionless case, the penetration depth is around 9.4D, while it decreases to ~ 7.2D for the frictional case. It is also seen that the frictional soil resistance starts to diverge from the frictionless case at the embedment depth of ~2.7D which is due to the separation of soil and anchor at shallower depths.

Therefore, it is observed that frictional resistance is generated during the fast penetration of dynamic anchors and its effects cannot be ignored.

Conclusions

Numerical analyses have been conducted to evaluate the effect of strain rate on the behaviour of dynamically penetrating anchors. The implications of the strain rate effects on the generation of excess pore pressure and the frictional resistance were specifically studied. It was shown that when the effect of strain rate is taken into account, a zone of suction is typically created around the anchor tip and also within a thin layer of soil along and adjacent to the anchor shaft.

Despite the undrained conditions in the soil, frictional resistance is generated during the fast penetration of dynamically installed anchors. This is largely because of the generation of suction pore pressures and the corresponding increase of the effective stress at the interface between the soil and the anchor. Therefore, it can now be concluded that the strain-rate effects not only increase the bearing resistance, but considerably increase the frictional resistance.

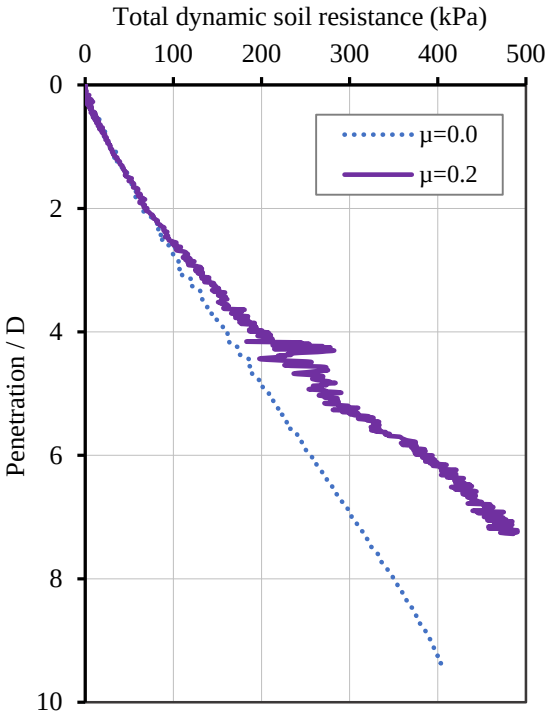


Figure 7. Total dynamic soil resistance profile: μ = 0 & 0.2

References

[1] Graham, J., Crooks, J. H. A. and Bell, L. (1983) Time effects on the stress–strain behaviour of natural soft clays, *Ge'otechnique* **33**(3):327-340

[2] Lefebvre, G. and LeBoeuf, D. (1987) Rate effects and cyclic loading of sensitive clays, *Journal of Geotechnical Engineering* **113**(5):476-489.

[3] Sheahan, T.C., Ladd, C. and Germaine J.T. (1996) Rate dependent undrained shear behaviour of saturated clay, *Journal of Geotechnial and Geoenvironmental Engineering, ASCE* **122**(2):99-108.

[4] Biscontin,G. and Pestana , J. M. (2001) Influence of peripheral velocity on vane shear strength of an artificial clay, *ASTM Geotechnical Testing Journal* **24**(4):423-430.

[5] Lunne, T. and Andersen, K. H. (2007) Soft clay shear strength parameters for deep water geotechnical design, *In: KeynoteAddress, Proc. Sixth Int. Offshore Site Investigation and Geotechnics Conf*, London, UK 51-176.

[6] Nazem, M., Carter, J.P., Airey, D.W. and Chow, S.H. (2012) Dynamic analysis of a smooth penetrometer free-falling into uniform clay, *Geotechnique* **62**(10): 893-905.

[7] Carter, J. P., Nazem, M., Airey, D.W. and Chow, S. H. (2010) Dynamic analysis of free falling penetrometers in soil deposits, *Geotechnical Special Pub. 199. GeoFlorida 2010-Advances in Analysis, Modelling and design, ASCE*, West Palm Beach, Florida 53-68.
[8] Kim, Y. H., Hossain, M. S. aand Wang, D. (2015) Effect of strain rate and strain softening on embedment depth of a torpedo anchor in clay, *Ocean Engineering* **108**(1):704-715.
[9] Sabetamal, H., Nazem, M. and Carter, J. P. (2013) Numerical analysis of Torpedo anchors, *In the proceedings of the 3rd International Symposium on Computational Geomechanics, ComGeo III*, Krakow, Poland 621-632.
[10] Sabetamal, H., Nazem, M., Carter, J. P. and Sloan, S. W. (2014) Large deformation dynamic analysis of saturated porous media with applications to penetration problems, *Computers and Geotechnics* **55**:117–13.
[11] Truesdell, C. and Toupin, R. (1960) The classical field theories, *In Handbuch der Physik, Flugge S (ed.). Springer* Vol. 3.
[12] Morland L. W. (1972) A simple constitutive theory for a fluid-saturated porous solid, *Journal of Geophysical Research* **77**, 890–900.
[13] Sabetamal, H. (2015) Finite Element Algorithms for Dynamic Analysis of Geotechnical Problems, *PhD thesis, The University of Newcastle*, Australia.
[14] van den Berg, P., deBorst, R. and Huetink, H. (1996) An Eulerian finite element model for penetration in layered soil, *International Journal for Numerical and Analytical Methods in Geomechanics* **20**(12) 865-886.
[15] Hu, Y. and Randolph, M.F. (1998) A practical numerical approach for large deformation problems in soil, *International Journal for Numerical and Analytical Methods in Geomechanics* **22**(5) 327-350.
[16] Susila, E. and Hryciw, R.D. (2003) Large displacement FEM modelling of the cone penetration test (CPT) in normally consolidated sand, *International Journal for Numerical and Analytical Methods in Geomechanics* **27**(7) 585-602.
[17] Nazem, M., Sheng, D. and Carter, J.P. (2006) Stress integration and mesh refinement in numerical solutions to large deformations in geomechanics, *International Journal for Numerical Methods in Engineering* **65**,1002**–**1027.
[18] Sulsky, D., Zhou, S. and Schreyer, H.L. (1995) Application of a particle-in-cell method to solid mechanics, *Computer Physics Communications* **87**, 235-252
[19] Beuth, L., Więckowski, Z. and Vermeer, P.A. (2011) Solution of quasi-static large-strain problems by the material point method, *International Journal for Numerical and Analytical Methods in Geomechanics* **35**(13) 1451-1465.
[20] Benson, D.J. (1989) An efficient, accurate, simple ALE method for nonlinear finite element programs, *Computer Methods in Applied Mechanics and engineering* **72**, 305–350.
[21] Wang, D., Bienen, B., Nazem, M., Tian, Y., Zheng, J., Pucker, T. and Randolph, M.F. (2015) Large deformation finite element analyses in geotechnical engineering, *Computers and Geotechnics* **65**, 104-114.
[22] Fressmann, D. and Wriggers. P. (2007) Advection approaches for single- and multi-material arbitrary Lagrangian–Eulerian finite element procedures, *Computational Mechanics* **39**, 153-190.
[23] Hallquist, J. O., Goudreau, G. L. and Benson, D. J. (1985) Sliding interfaces with contact impact in large-scale Lagrangian computations, *Computer Methods in Applied Mechanics and Engineering* 107-137.
[24] Wriggers, P. and Simo, J. C. (1985) A note on tangent stiffness for fully nonlinear contact problems, *Communications in Applied Numerical Methods* **1**,199-203.
[25] Fischer, K. A. and Wriggers. P. (2005) Frictionless 2D contact formulations for finite deformations based on the mortar method, *Computational Mechanics* 36, 226-244.
[26] Sabetamal, H., Nazem, M., Sloan, S. W. and Carter, J. P. (2016) Frictionless contact formulation for dynamic analysis of nonlinear saturated porous media based on the mortar method, *International Journal for Numerical and Analytical Methods in Geomechanics* **40**(1) 25-61
[27] Casagrande, A. and Wilson, S.D. (1951) Effect of Rate of Loading on the Strength of Clays and Shales at Constant Water Contents, *Geotechnique*, **2**(3)251-263.
[28] Bjerrum, L., Simons, N. and Torblaa, I. (1958) The Effect of Time on the Shear Strength of a Soft Marine Clay, *Proceedings of the Brussels Conference on Earth Pressure Problems*, Vol I, 148-158.
[29] Skempton, A. W. (1948) Vane tests in alluvial plain of the River Forth near Grangemouth, *Geotechnique*, **1**, 111-124.
[30] Mitchell, J. K. and Soga, K. (2005) Fundamentals of soil behaviour, *3rd ed. New York: Wiley Inter-Science.*
[31] Bjerrum, L. (1967). Engineering geology of Norwegian normally consolidated marine clays as related to settlements of buildings, *Géotechnique* **17**(2):81-118.
[32] Leroueil, S., Kabbaj, M., Tavenas, F. and Bouchard, R. (1985) Stress-strain rate relation for the compressibility of sensitive natural clays, *Géotechnique* **35**(2), 159-180.
[33] Sheahan, T. C. (2005) A soil structure index to predict rate dependence of stress–strain behavior, *In: Testing, modeling and simulation in geomechanics, ASCE, Geotechnical Special Publication* **143**, 81-97.
[34] Silvestri, V., Yong, R. N., Soulie, M. and Gabriel, F. (1986) Controlledstrain, controlled-gradient and standard consolidation testing of sensitive clays, *In: Yong RN, Townsend FC (eds) Proceedings of consolidation of soils: testing and evaluation: a symposium, issue 892, ASTM Committee D-18 on Soil and Rock* , 433-450.
[35] Katti, D. R., Tang, J. P. and Yazdani, S. (2003) Undrained Response of Clays to Varying Strain Rate, *Journal of Geotechnical and Geoenvirontal Engineering* **129**(3) 278-282.
[36] Potts, D. M. and Zdravkovic, L. (1999) Finite element analysis in geotechnical engineering: theory. *Thomas Telford, London.*
[37] True, D. G. (1976) Undrained Vertical Penetration into Ocean Bottom Soils, *PhD Dissertation, University of California, Berkeley, California.*

[38] Richardson, M. D. (2008) Dynamically installed anchors for floating offshore structures, *Ph.D. thesis, University of Western Australia.*
[39] Chow, S. H., O'Loughlin, C.D. and Randolph, M. F. (2014) Soil strength estimation and pore pressure dissipation for free-fall piezocone in clay, *Géotechnique* **64**(10):817–824.

Stability Investigation of Direct Integration Algorithms Using Lyapunov-Based Approaches

***Xiao. Liang[1], †Khalid M. Mosalam[1]**

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, USA.

*Presenting author: benliangxiao@berkeley.edu
†Corresponding author: mosalam@berkeley.edu

Abstract

In structural dynamics, direct explicit and implicit integration algorithms are commonly used to solve the temporally discretized differential equations of motion for linear and nonlinear structures. The stability of different integration algorithms for linear elastic structures has been extensively studied for several decades. However, investigations of the stability applied to nonlinear structures are relatively limited and rather challenging. Recently, the authors proposed two systematic approaches using Lyapunov stability theory to investigate the stability property of direct integration algorithms of nonlinear dynamical systems. The first approach is a numerical one that transforms the stability analysis to a problem of convex optimization. The second approach investigates the Lyapunov stability of explicit algorithms considering the strictly positive real lemma. This paper reviews and compares these two Lyapunov-based approaches in terms of their merits and limitations.

Keywords: Convex optimization, Direct integration algorithm, Lyapunov stability, Nonlinear, Strictly positive real lemma, Structural dynamics.

Introduction

In structural dynamics, direct integration algorithms are commonly used to solve the differential equations of motion after they are temporally discretized to estimate dynamic responses of structures, e.g., seismic responses of bridges [1]. Integration algorithms are categorized into either implicit or explicit. An integration algorithm is explicit when the responses of the next time step depend on the responses of previous and current time steps only. Otherwise, it is implicit. Numerous implicit and explicit direct integration methods have been developed, including the Newmark family of algorithms [2], the TRBDF2 algorithm [3], and the Operator-Splitting (OS) algorithms [4]. Liang et al. [5,6] investigated the suitability of the OS algorithms for efficient nonlinear seismic response of multi-degree of freedom (MDOF) reinforced concrete highway bridge systems and promising results in terms of accuracy and numerical stability were obtained. The stability of different integration algorithms for linear elastic structures has been studied extensively for several decades, e.g., [7]. Studies related to the stability properties of these integration algorithms applied to nonlinear dynamic analysis are relatively limited and, unlike linear ones, are rather complicated and challenging. This is attributed to specific properties of the nonlinear systems. For example, initial conditions affect the stability of nonlinear systems and the principle of superposition does not hold.

Lyapunov stability theory [8,9], developed by the Russian mathematician Aleksandr Lyapunov in [10], is the most complete framework of stability analysis for dynamical systems. It is based on constructing a function of the system state coordinates that serves as a generalized norm of the solution of the dynamical system. The most important property of Lyapunov stability theory is the fact that conclusions about the stability behavior of the dynamical system can be obtained without

actually computing the system solution trajectories. As such, Lyapunov stability theory has become one of the most fundamental and standard tools of dynamical systems and control theory.
Generally speaking, constructing the above-mentioned energy function for the nonlinear system is not readily available. To address this difficulty, the authors proposed two approaches. In the first, a numerical approach is proposed to transform the problem of seeking a Lyapunov function to a convex optimization problem [11,12], which can solve the problem in a simple and clear manner. Convex optimization minimizes convex functions over convex sets, in which a wide range of problems can be formulated in this way. In this optimization, any local minimum must be a global one, which is an important property leading to reliable and efficient solutions using, e.g., interior-point methods, which are suitable for computer-aided design or analysis tools [13]. The second approach proposed by the authors is based on formulating a generic explicit integration algorithm into a nonlinear system governed by a nonlinear function of the basic forces. This enables investigating the Lyapunov stability of explicit algorithms by means of the strictly positive real lemma [11,14]. The study for nonlinear single degree of freedom (SDOF) systems in [14] was extended to MDOF ones in [15]. This approach transforms the stability analysis of the formulated nonlinear system to investigating the strictly positive realness of its corresponding transfer function matrix. This is further equivalent to a problem of convex optimization that can be solved numerically.

This paper reviews and compares these previously discussed two Lyapunov-based approaches in terms of their merits and limitations. The first numerical approach is shown to be generally applicable to implicit and explicit direct integration algorithms for various nonlinear force-deformation relationships. Moreover, this approach can potentially be extended to nonlinear MDOF systems but may involve extensive computations. The second approach is applicable to explicit algorithms without adopting any approximation and is computationally efficient even for MDOF systems.

Integration Algorithm

The discretized equations of motion of a MDOF system under an external dynamic force excitation is expressed as follows:

$$\mathbf{m}\ddot{\mathbf{u}}_{i+1} + \mathbf{c}\dot{\mathbf{u}}_{i+1} + \mathbf{f}(\mathbf{u}_{i+1}) = \mathbf{p}_{i+1} \tag{1}$$

where $\mathbf{m}$ and $\mathbf{c}$ are the mass and damping matrices, and $\ddot{\mathbf{u}}_{i+1}$, $\dot{\mathbf{u}}_{i+1}$, $\mathbf{f}_{i+1}$, and $\mathbf{p}_{i+1}$ are respectively the acceleration, velocity, restoring force, and external force vectors at time step $i+1$. The restoring force $\mathbf{f}(\mathbf{u})$ is generally defined as a function of the displacement vector $\mathbf{u}$. It is to be noted that bold-faced symbols indicate arrays, either vectors or matrices.

A single-step direct integration algorithms (explicit or implicit) are collectively defined in this paper using the following difference equations:

$$\mathbf{u}_{i+1} = \mathbf{u}_i + (\Delta t)\dot{\mathbf{u}}_i + \eta_1(\Delta t)^2\ddot{\mathbf{u}}_i + \eta_2(\Delta t)^2\ddot{\mathbf{u}}_{i+1} \tag{2}$$

$$\dot{\mathbf{u}}_{i+1} = \dot{\mathbf{u}}_i + \eta_3(\Delta t)\ddot{\mathbf{u}}_i + \eta_4(\Delta t)\ddot{\mathbf{u}}_{i+1} \tag{3}$$

In general, Eqs. (1)–(3) require an iterative solution, which forms the basis of the implicit algorithms. On the other hand, these algorithms become explicit when $\eta_2 = 0$. For example, $[\eta_1,\eta_2,\eta_3,\eta_4] = [1/4,1/4,1/2,1/2]$ leads to implicit Newmark with constant average acceleration, $[\eta_1,\eta_2,\eta_3,\eta_4] = [1/2,0,1/2,1/2]$ transforms the integration to the explicit Newmark algorithm [2].

Lyapunov-Based Numerical Approach

For each direct integration algorithm of SDOF systems, the relationship between the kinematic quantities at time steps $i+1$ and i can be established as follows:

$$\boldsymbol{x}_{i+1} = \mathbf{A}_i \boldsymbol{x}_i + \mathbf{L}_i \tag{4}$$

where $\boldsymbol{x}_i = \left[(\Delta t)^2 \ddot{u}_i \;\; (\Delta t)\dot{u}_i \;\; u_i\right]^T$. It is noted that $\mathbf{A}_i$ and $\mathbf{L}_i$ are the approximation operator and the loading vector at the time step i, respectively. The loading vector, $\mathbf{L}$, is generally bounded and independent of the vector of kinematic quantities, $\boldsymbol{x}$, and does not affect the Lyapunov stability of the direct integration algorithms. Therefore, $\mathbf{L}$ can be set to zero in the sequel of this paper.

For linear structures, the approximation operator, $\mathbf{A}$, remains constant. The stability criterion of linear systems is obvious and well-known, namely the spectral radius of the approximation operator $\rho(\mathbf{A})$ must be less than or equal to 1.0. In contrast, for nonlinear structures, methods that are applicable to linear ones generally do not work. For example, the spectral radius and frequency domain methods basically convey nothing about the stability properties of algorithms. Instead, we turned to Lyapunov stability theory, based on which a numerical approach was proposed. This approach transforms the stability analysis to a problem of convex optimization, which is applicable to direct integration algorithms used to solve nonlinear problems.

As discussed above, we are investigating the system in Eq. (4) with the loading vector $\mathbf{L} = \mathbf{0}$, i.e.,

$$\boldsymbol{x}_{i+1} = \mathbf{A}_i \boldsymbol{x}_i \tag{5}$$

where $\mathbf{A}_i$ is a function of δ_{i+1} which is the tangent stiffness at time step $i+1$ normalized by the initial stiffness. Detailed derivations of $\mathbf{A}_i$ for different algorithms are given in [11,12].

One standard Lyapunov function v_{i+1} at the time step $i+1$ is defined in [16] as follows:

$$v_{i+1} = \boldsymbol{x}_{i+1}^T \mathbf{M}_{i+1} \boldsymbol{x}_{i+1} \tag{6}$$

where the positive definite matrix $\mathbf{M}_{i+1} = \mathbf{M}_{i+1}^T$ is a function of δ_{i+1}. A sufficient condition for the system and thus the direct integration algorithm to be stable is as follows:

$$\begin{aligned} \Delta v_{i+1} &= v_{i+1} - r_t v_i \\ &= \boldsymbol{x}_{i+1}^T \mathbf{M}_{i+1} \boldsymbol{x}_{i+1} - r_t \boldsymbol{x}_i^T \mathbf{M}_i \boldsymbol{x}_i \\ &= \boldsymbol{x}_i^T \left(\mathbf{A}_i^T \mathbf{M}_{i+1} \mathbf{A}_i - r_t \mathbf{M}_i \right) \boldsymbol{x}_i \\ &= \boldsymbol{x}_i^T \mathbf{P}_{i+1} \boldsymbol{x}_i \le 0 \end{aligned} \tag{7}$$

where $0 < r_t \le 1$ controls the rate of convergence, i.e., the smaller the r_t, the faster the convergence. Eq. (7) lead to the negative semi-definiteness of $\mathbf{P}_{i+1}$, i.e., $\mathbf{P}_{i+1} \preceq \mathbf{0}$. For a direct integration algorithm, $\mathbf{M}_{i+1}$ can be expressed as:

$$\mathbf{M}_{i+1} = \sum_{j=1}^{B} \alpha_j \mathbf{\Phi}_j(\delta_{i+1}) \tag{8}$$

where α_j and $\mathbf{\Phi}_j(\delta_{i+1})$ are the *j*–th constant coefficient and base function, respectively, and *B* is the total number of base functions. One example set of base functions is given in [11] where the set

of base functions of $\mathbf{\Phi}_1$ to $\mathbf{\Phi}_6$ represent constant $\mathbf{M}_{i+1}$, $\mathbf{\Phi}_7$ to $\mathbf{\Phi}_{12}$ constitute the base functions that treat $\mathbf{M}_{i+1}$ as a linear function of δ_{i+1}, and nonlinear relationship between $\mathbf{M}_{i+1}$ and δ_{i+1} are considered by base functions $\mathbf{\Phi}_{13}$ to $\mathbf{\Phi}_{18}$.

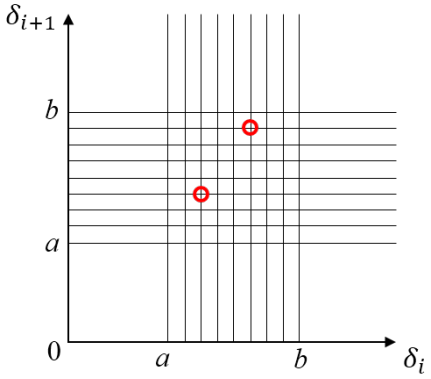


Figure 1. Schematic illustration of discretization process

With the range of δ_i and δ_{i+1} given, e.g., $\delta_i, \delta_{i+1} \in [a,b]$, points can be discretized within this range (Figure 1), e.g., sampling $p+1$ points in $[a,b]$ with interval $\Delta\delta = (b-a)/p$. This yields $(p+1)^2$ possible pairs of (δ_i, δ_{i+1}). Accordingly, the stability analysis becomes a problem of convex optimization that seeks the determination of the coefficients α_j by minimizing their norm for the selected base functions $\mathbf{\Phi}_j(\delta_{i+1})$ where $j: 1 \to B$, subjected to the following conditions on the $(p+1)^2$ possible pairs of (δ_i, δ_{i+1}):

$$\begin{aligned} &\delta_i, \delta_{i+1} \in [a,b], \quad \Delta\delta = (b-a)/p \\ &\mathbf{A}_i^T \mathbf{M}_{i+1} \mathbf{A}_i - r_t \mathbf{M}_i = \mathbf{A}_i^T(\delta_{i+1}) \left(\sum_{j=1}^{B} \alpha_j \mathbf{\Phi}_j(\delta_{i+1}) \right) \mathbf{A}_i(\delta_{i+1}) - r_t \sum_{j=1}^{B} \alpha_j \mathbf{\Phi}_j(\delta_i) \prec = \mathbf{0} \\ &\mathbf{M}_i = \sum_{j=1}^{B} \alpha_j \mathbf{\Phi}_j(\delta_i) \succ \mathbf{0}, \quad \mathbf{M}_{i+1} = \sum_{j=1}^{B} \alpha_j \mathbf{\Phi}_j(\delta_{i+1}) \succ \mathbf{0} \end{aligned} \tag{9}$$

Moreover, with prior knowledge about the variation of δ_{i+1}, the range of $|\delta_{i+1} - \delta_i|$ can be specified, e.g., $|\delta_{i+1} - \delta_i| < \varepsilon$, where ε is an optional parameter that is not necessarily small. For example, suppose we are interested in investigating the stability of a certain algorithm in the range of $\delta_i, \delta_{i+1} \in [1,2]$, and $\delta_i = 1.5$ at the *i*-th time step. If prior knowledge is known such that $\varepsilon = 0.3$, i.e., $\delta_{i+1} \in (1.2, 1.8)$, fewer possible pairs of (δ_i, δ_{i+1}) that require less computational effort can be considered. The problem of convex optimization can be solved numerically by CVX, a software package for specifying and solving convex programs [17].

Two examples of the softening and the stiffening cases for the implicit Newmark algorithm with constant average acceleration are presented to illustrate this approach. The following conditions are considered in these examples:

$$\zeta = 0.05 \quad \mu = 0.05/(2\pi) \quad n = 20 \quad \varepsilon = 0.05 \quad r_t = 1.0 \tag{10}$$

where $\zeta = c/(2m\omega_n)$, $\omega_n^2 = k_I/m$, $\mu = \Delta t/T_n$, $T_n = 2\pi/\omega_n = 2\pi\sqrt{m/k_I}$. The set of base functions $\mathbf{\Phi}_1$ to $\mathbf{\Phi}_{12}$ in [12] is used.

Softening Example

Suppose we are interested in investigating the stability of the implicit Newmark algorithm in the range of $\delta_i, \delta_{i+1} \in [0.9, 1.0]$, therefore $\Delta\delta = (b-a)/p = 0.005$. The coefficients α_j, $j:1\to 12$, are:

$$\begin{aligned}
&\alpha_1 = 1.90\times10^{-8}, \quad \alpha_2 = 2.46\times10^{-9}, \quad \alpha_3 = 1.70\times10^{-10}, \quad \alpha_4 = -2.25\times10^{-9},\\
&\alpha_5 = -2.70\times10^{-10}, \quad \alpha_6 = -4.60\times10^{-10}, \quad \alpha_7 = 1.76\times10^{-8}, \quad \alpha_8 = 1.05\times10^{-9},\\
&\alpha_9 = 6.00\times10^{-11}, \quad \alpha_{10} = -3.35\times10^{-9}, \quad \alpha_{11} = 4.30\times10^{-10}, \quad \alpha_{12} = -2.00\times10^{-10}
\end{aligned} \tag{11}$$

Stiffening Example

Analogous to the procedure of the previous softening example, suppose the range of interest for the stiffening case is $\delta_i, \delta_{i+1} \in [1.0, 1.1]$, the obtained coefficients α_j, $j:1\to 12$, are:

$$\begin{aligned}
&\alpha_1 = 9.81\times10^{-7}, \quad \alpha_2 = 2.28\times10^{-10}, \quad \alpha_3 = 1.93\times10^{-8}, \quad \alpha_4 = -2.25\times10^{-8},\\
&\alpha_5 = -2.30\times10^{-9}, \quad \alpha_6 = -1.53\times10^{-9}, \quad \alpha_7 = 1.01\times10^{-6}, \quad \alpha_8 = 3.39\times10^{-11},\\
&\alpha_9 = 7.03\times10^{-9}, \quad \alpha_{10} = -8.94\times10^{-8}, \quad \alpha_{11} = 7.20\times10^{-9}, \quad \alpha_{12} = -5.01\times10^{-10}
\end{aligned} \tag{12}$$

The set of α_j in Eqs. (11) and (12) from many determined sets has the minimum 2-norm α, i.e. $\min\sqrt{\sum_{j=1}^{12}\alpha_j^2}$, explaining the listed small values of α_j. The existence of such set of α_j implies the existence of $\mathbf{M}_{i+1}$ in Eq. (8) that satisfies the inequality in Eq. (7), which signifies that the implicit Newmark algorithm is stable for the conditions in Eq. (10) in the range of $\delta_i, \delta_{i+1} \in [0.9, 1.0]$ based on Eq. (11) or in the range of $\delta_i, \delta_{i+1} \in [1.0, 1.1]$ based on Eq. (12). Several other examples are provided in [9,10].

The approach discussed above can be applied to investigate the stability of different direct integration algorithms considering various nonlinear effects, e.g., stiffening ($\delta_{i+1} > 1$) and softening ($\delta_{i+1} < 1$) force-deformation relationships. Thus, this approach is generally applicable to direct integration algorithms as long as they can be expressed as given by Eq. (5). Moreover, this approach can potentially be extended to MDOF systems. For m DOF systems, the $3m \times 3m$ approximation operator is a function of δ_{i+1}^j, where $j:1\to m$ denotes the j-th DOF, and thus $(m+1)(9m^2+3m)/2$ selected base functions and corresponding coefficients are needed if $\mathbf{M}_{i+1}$ is expressed as an affine function of δ_{i+1}^j, $j:1\to m$. Thus, this approach involves extensive computations for MDOF systems.

Lyapunov-Based Approach Considering Strictly Positive Real Lemma

This approach was proposed to deal with stability issues of explicit direct integration algorithms, i.e., $\eta_2 = 0$ in Eq. (2). As mentioned previously in the introduction, it transforms the stability analysis of the formulated MDOF nonlinear system to investigating the strictly positive realness of its corresponding transfer function matrix.

For a MDOF system with n DOFs, the j–th term of the restoring force vector, $f^j, j \in [1, n]$, can be expressed as a linear combination of N basic resisting forces of the system, $q^l, l \in [1, N]$, i.e.,

$$f^j = \sum_{l=1}^{N} \alpha_l^j q^l = \boldsymbol{\alpha}^j \mathbf{q} \tag{13}$$

where $\mathbf{q}^T = \left[q^1, q^2, \ldots, q^N\right]$ and $\boldsymbol{\alpha}^j = \left[\alpha_1^j, \alpha_2^j, \ldots, \alpha_N^j\right]$. Therefore,

$$\mathbf{f} = \left[f^1, f^2, \ldots, f^n\right]^T = \boldsymbol{\alpha}\mathbf{q} \tag{14}$$

where $\boldsymbol{\alpha} = \left[\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^n\right]^T$ is an $n \times N$ matrix. In general, N is the summation of the number of the basic resisting forces from each element that contribute to the n DOFs of the system. For the special case of a shear building, $N = n$ because of its assumed shear mode behavior. The *l*–th basic resisting force, q^l, is here defined as a function of $\bar{u}^l$, which is in itself a linear combination of the displacement of each DOF, u^j, $j \in [1, n]$, i.e.,

$$\bar{u}^l = \sum_{j=1}^{n} \beta_j^l u^j = \boldsymbol{\beta}^l \mathbf{u} \tag{15}$$

where $\mathbf{u} = [u^1, u^2, \ldots, u^n]$ and $\boldsymbol{\beta}^l = [\beta_1^l, \beta_2^l, \ldots, \beta_n^l]$. Therefore,

$$\bar{\mathbf{u}} = \left[\bar{u}^1, \bar{u}^2, \ldots, \bar{u}^N\right]^T = \boldsymbol{\beta}\mathbf{u} \tag{16}$$

where $\boldsymbol{\beta} = \left[\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^N\right]^T$ is an $N \times n$ matrix. Detailed explanation of N defining the number of columns and rows of the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, are available in [15]. Moreover, the *l*–th basic resisting force, q^l, is a sector-bounded nonlinearity and is restricted to the following range:

$$\bar{k}_{Min}^l \left(\bar{u}^l\right)^2 \le q^l \bar{u}^l \le \bar{k}_{Max}^l \left(\bar{u}^l\right)^2 \tag{17}$$

where $\bar{k}_{Min}^l \bar{u}^l$ and $\bar{k}_{Max}^l \bar{u}^l$ are the minimum and maximum bounds of q^l, respectively. Define

$$\bar{\mathbf{k}}_{Min} = \mathrm{diag}\left[\bar{k}_{Min}^1, \bar{k}_{Min}^2, \ldots, \bar{k}_{Min}^N\right] \tag{18a}$$

$$\bar{\mathbf{k}}_{Max} = \mathrm{diag}\left[\bar{k}_{Max}^1, \bar{k}_{Max}^2, \ldots, \bar{k}_{Max}^N\right] \tag{18b}$$

$$\bar{\mathbf{k}} = \bar{\mathbf{k}}_{Max} - \bar{\mathbf{k}}_{Min} = \mathrm{diag}[\bar{k}^1, \bar{k}^2, \ldots, \bar{k}^N] \tag{18c}$$

After some manipulation [15], both stiffening and softening systems can be expressed in Eq. (19) with coefficients $\mathbf{A}_e, \mathbf{B}_e$ and $\mathbf{q}_e$ summarized in Table 1.

$$\boldsymbol{x}_{i+1} = \mathbf{A}_e \boldsymbol{x}_i - \mathbf{B}_e \mathbf{q}_e \tag{19}$$

Table 1. Coefficients of MDOF stiffening and softening systems

Matrix	Stiffening Systems	Softening Systems
$\mathbf{A}_e$	$\mathbf{A}_{e1} = \mathbf{A} - \mathbf{B}_1 \boldsymbol{\alpha} \bar{\mathbf{k}}_{Min} \mathbf{C}$	$\mathbf{A}_{e2} = \mathbf{A} + \mathbf{B}_2 \boldsymbol{\alpha} \bar{\mathbf{k}}_{Max} \mathbf{C}$
$\mathbf{B}_e$	$\mathbf{B}_1 \boldsymbol{\alpha}$	$\mathbf{B}_2 \boldsymbol{\alpha}$
$\mathbf{q}_e$	$\mathbf{q}_{e1} = \mathbf{q}_{i+1} - \bar{\mathbf{k}}_{Min} \mathbf{C} \boldsymbol{x}_i$	$\mathbf{q}_{e2} = \bar{\mathbf{k}}_{Max} \mathbf{C} \boldsymbol{x}_i - \mathbf{q}_{i+1}$

where the terms in Table 1 are expressed as follows:

$$\mathbf{B}_1 = -\mathbf{B}_2 = \left[(\Delta t)^2 \mathbf{m}_{eff}^{-1} \quad \eta_3 (\Delta t)^2 \mathbf{m}_{eff}^{-1} \quad \mathbf{0}\right]^T, \quad \mathbf{m}_{eff} = \mathbf{m} + \eta_3 (\Delta t)\mathbf{c} \tag{20a}$$

$$\mathbf{C} = \boldsymbol{\beta}\tilde{\mathbf{C}}, \quad \tilde{\mathbf{C}} = \left[\eta_1 \mathbf{I} \quad \eta_0 \mathbf{I} \quad \mathbf{I}\right], \quad \mathbf{I} = \text{Identity matrix} \tag{20b}$$

Similar to the first numerical approach, the Lyapunov function v_{i+1} at the time step $i+1$ is chosen as:

$$v_{i+1} = \boldsymbol{x}_{i+1}^T \mathbf{M} \boldsymbol{x}_{i+1} \tag{21}$$

The constraints that the basic forces are sector-bounded lead to

$$\Delta v_{i+1} = v_{i+1} - v_i \le -\left(\mathbf{W}\mathbf{q}_e - \mathbf{L}\boldsymbol{x}_i\right)^T \left(\mathbf{W}\mathbf{q}_e - \mathbf{L}\boldsymbol{x}_i\right) \le 0 \tag{22}$$

where there exist matrices $\mathbf{M}$, $\mathbf{L}$ and $\mathbf{W}$ such that

$$\mathbf{M} = \mathbf{A}_e^T \mathbf{M} \mathbf{A}_e + \mathbf{L}^T \mathbf{L} \tag{23a}$$

$$\mathbf{0} = \mathbf{B}_e^T \mathbf{M} \mathbf{A}_e - \boldsymbol{\lambda}\bar{\mathbf{k}}\mathbf{C} + \mathbf{W}^T \mathbf{L} \tag{23b}$$

$$\mathbf{0} = \boldsymbol{\lambda} + \boldsymbol{\lambda}^T - \mathbf{B}_e^T \mathbf{M} \mathbf{B}_e - \mathbf{W}^T \mathbf{W} \tag{23c}$$

where $\boldsymbol{\lambda}$ is a constant diagonal matrix of arbitrary positive coefficients. Derivations from Eq. (21) to Eqs. (23) can be found in [15]. Based on the generalized strictly positive real lemma [18], the stability analysis reduces to seeking $\bar{\mathbf{k}}$ such that the transfer function matrix $\mathbf{G}(z)$ in Eq. (24) is strictly positive real.

$$\mathbf{G}(z) = \boldsymbol{\lambda} + \boldsymbol{\lambda}\bar{\mathbf{k}}\mathbf{C}\left(\mathbf{I}z - \mathbf{A}_e\right)^{-1}\mathbf{B}_e \tag{24}$$

For SDOF systems, the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ become 1, based on [11,14], Eq. (24) reduces to

$$G(z) = 1 + \bar{k}\mathbf{C}\left(\mathbf{I}z - \mathbf{A}_e\right)^{-1}\mathbf{B}_e \tag{25}$$

The strictly positive realness of $G(z)$ can be guaranteed by the asymptotical stability of $\mathbf{A}_e$ and

$$\mathrm{Re}\left[G(z)\right] > 0 \tag{26}$$

which leads to

$$\mathrm{Re}\left[H(z)\right] > -1/\bar{k} \tag{27}$$

where

$$H(z) = \mathbf{C}\left(\mathbf{I}z - \mathbf{A}_e\right)^{-1}\mathbf{B}_e \tag{28}$$

The Nyquist plot [16] can be used to plot $H\left(e^{j\theta}\right) \ \forall \ \theta \in [0, 2\pi]$. From this plot, the minimum value of $\mathrm{Re}\left[H(z)\right]$ that is corresponding to the $-1/\bar{k}$ can be obtained.

For MDOF systems, based on [19], the strictly positive realness of $\mathbf{G}(z)$ in Eq. (24) becomes equivalent to Eq. (29) with $\mathbf{P} = \mathbf{P}^T \succ \mathbf{0}$:

$$\begin{bmatrix} \mathbf{A}_e^T \mathbf{P} \mathbf{A}_e - \mathbf{P} & \mathbf{A}_e^T \mathbf{P} \mathbf{B}_e - \left(\boldsymbol{\lambda}\bar{\mathbf{k}}\mathbf{C}\right)^T \\ \left[\mathbf{A}_e^T \mathbf{P} \mathbf{B}_e - \left(\boldsymbol{\lambda}\bar{\mathbf{k}}\mathbf{C}\right)^T\right]^T & -\left(\boldsymbol{\lambda}^T + \boldsymbol{\lambda}\right) + \mathbf{B}_e^T \mathbf{P} \mathbf{B}_e \end{bmatrix} \prec \mathbf{0} \tag{29}$$

Eq. (29) is a linear matrix inequality (LMI) over variables $\mathbf{P}$ and $\bar{\mathbf{k}}$ [20]. This problem of convex optimization, which seeks $\bar{\mathbf{k}}$ and the corresponding $\mathbf{P}$ by minimizing certain convex cost function, subjected to the constraints of $\mathbf{P}=\mathbf{P}^T \succ \mathbf{0}$ and $\bar{\mathbf{k}} \succ \mathbf{0}$, can be solved numerically by CVX [17].

Multi-story shear buildings with stiffening and softening structural behaviors are used as examples to illustrate this approach. A general multi-story shear building structure is depicted in Figure 2. The detailed derivation of $\mathbf{q}$ and $\bar{\mathbf{u}}$ as well as the corresponding matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for this shear building is given in [15]. Accordingly, the maximum, $\bar{k}_{Max}^{j}$, and minimum, $\bar{k}_{Min}^{j}$, stiffness values of the *j*–th story, where $j:1\rightarrow n$ and the number of stories is n, for stable (in the sense of Lyapunov) stiffening and softening multi-story shear building systems, respectively, are to be determined.

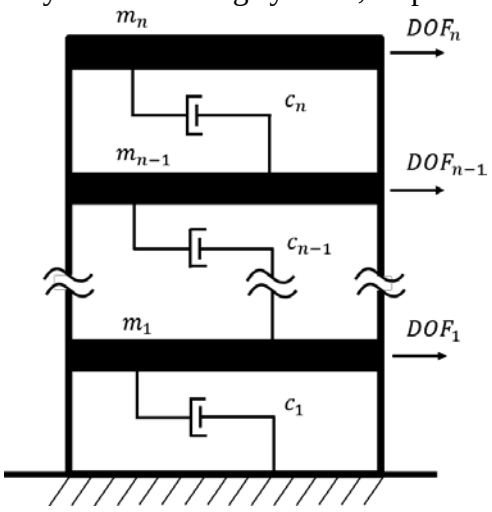


Figure 2. A general multi-story shear building structure

The stability analysis is conducted for the following numerical values:

$$m_j = 0.5, \quad \zeta = 0.05, \quad \bar{k}_I^j = 1000.0 \tag{30a}$$

$$\lambda_j = \omega_j^2 \Big/ \left(\sum_{j=1}^{n} \omega_j^2 \right), \quad \omega_j = 2\pi / T_j, \quad \mu = (\Delta t)/T_n = 0.01 \tag{30b}$$

where T_j is the period of the *j*–th mode of vibration of the analyzed structure. The initial bound matrix is $\bar{\mathbf{k}}_I = \text{diag}[\bar{k}_I^1, \bar{k}_I^2, \ldots, \bar{k}_I^n]$, i.e., $\bar{\mathbf{k}}_{Min}$ and $\bar{\mathbf{k}}_{Max}$ for stiffening and softening systems, respectively. A 20-story (Figure 2 with $n=20$) shear building is used to investigate the Lyapunov stability analysis of the explicit Newmark algorithm, i.e. $[\eta_1, \eta_2, \eta_3, \eta_4] = [1/2, 0, 1/2, 1/2]$. Lyapunov stability analysis following the approach previously discussed in this section is conducted for the analyzed this 20-story shear building with stiffening or softening behavior. The cost function for this building structure is selected as $\min(-\sum_{j=1}^{20} \bar{k}^j)$, which is equivalent to $\max(\sum_{j=1}^{20} \bar{k}^j)$. In this cost function, $\bar{k}^j = \bar{k}_{Max}^j - \bar{k}_{Min}^j$ is the difference of the upper and lower bounds of the basic resisting force q^j associated with the j–th story, where $j:1\rightarrow n$. Table 2 shows that the difference of the upper and lower bounds, $\bar{k} = \bar{k}_{Max} - \bar{k}_{Min}$, of each resisting force for the explicit Newmark algorithm to be stable (in the sense of Lyapunov) for both stiffening, $\bar{\mathbf{u}}^T\mathbf{q} \in [\bar{\mathbf{u}}^T\bar{\mathbf{k}}_I\bar{\mathbf{u}}, \bar{\mathbf{u}}^T(\bar{\mathbf{k}}_I + \bar{\mathbf{k}})\bar{\mathbf{u}}]$, and softening, $\bar{\mathbf{u}}^T\mathbf{q} \in [\bar{\mathbf{u}}^T(\bar{\mathbf{k}}_I - \bar{\mathbf{k}})\bar{\mathbf{u}}, \bar{\mathbf{u}}^T\bar{\mathbf{k}}_I\bar{\mathbf{u}}]$, systems.

Table 2. The $\bar{k}$ of each basic resisting force for the 20-story shear building

Story Number	Stiffening systems	Softening systems	Story Number	Stiffening systems	Softening systems
1	716.1	203.7	11	31.3	35.9
2	125.1	149.6	12	25.3	30.8
3	98.8	150.4	13	21.8	28.5
4	133.0	166.5	14	19.6	26.7
5	163.4	140.0	15	17.3	24.0
6	119.7	97.7	16	15.0	21.2
7	76.9	74.8	17	14.2	20.4
8	56.5	64.1	18	16.9	23.7
9	46.7	55.3	19	29.6	37.3
10	39.0	44.8	20	116.1	106.7

More Examples are given in [10,14,15] to illustrate this approach for different direct explicit integration algorithms applied to different structures (buildings and bridges) with stiffening and softening force-deformation relationships.

Summary and Concluding Remarks

This paper reviewed and compared two recently proposed Lyapunov-based approaches of stability analysis in terms of their merits and limitations. Interested readers should consult references [11,12,14,15] for detailed derivations and examples.

The first approach transforms the stability analysis to a problem of existence, that can be solved via convex optimization, over the discretized domain of interest of the restoring force. As such, this approach is a numerical one with certain approximations. It is shown to be generally applicable to both implicit and explicit direct integration algorithms for various nonlinear force-deformation relationships, including stiffening and softening ones. References [11,12] considered nonlinear SDOF systems. This approach can potentially be extended to nonlinear MDOF systems but extensive computations are involved and can be overcome by some methods, e.g., parallel computing [21].

The second approach is specifically applicable to explicit algorithms for nonlinear SDOF and MDOF systems considering strictly positive real lemma. In this approach, a generic explicit algorithm was formulated for a nonlinear system governed by a nonlinear function of the basic force without adopting any approximations. Starting from this formulation and based on Lyapunov stability theory, the stability analysis of the formulated nonlinear system is transformed to investigating the strictly positive realness of its corresponding transfer function matrix. Furthermore, this is equivalent to a problem of convex optimization that can be solved numerically. The basic force in this study was limited to the sector-bounded nonlinearity, including stiffening, softening and even hysteretic force-deformation relationships as long as they are within the sector bounds. Moreover, this approach is more computationally efficient than the first numerical one, especially for MDOF systems. Comparisons between these two approaches are listed in Table 3. It should be emphasized that Eqs. (7) and (22) are sufficient conditions for dynamical systems to be stable. Therefore, both approaches provide a sufficient condition for the direct integration algorithm to be stable. In other words, neither of these two approaches can indicate the condition of instability of the investigated algorithms. For example, having some basic resisting force vector $\mathbf{q}$ that may

fall outside the range in Table 2 does not indicate the instability of the explicit Newmark algorithm for the analyzed 20-story shear building.

Table 3. Comparisons between the two approaches

Property	First approach	Second approach
Algorithm	Implicit & Explicit	Explicit
Nonlinearity	No restriction	Sector-bounded
Condition	Sufficient	Sufficient
Approximation	Yes	No
MDOF	Potentially	Yes
Computational effort	Extensive	Efficient

References

[1] Liang, X., Günay, S. and Mosalam, K.M. (2016) Chapter 12: Seismic Response of Bridges Considering Different Ground Motion Selection Methods, in *Developments in International Bridge Engineering*, Springer Tracts on Transportation and Traffic 9, Springer Int. Publishing, Switzerland.

[2] Newmark, N. M. (1959) A method of computation for structural dynamics, *ASCE J. Eng. Mech. Div.*, **85**(3), 67–94.

[3] Bathe, K. J. (2007) Conserving Energy and Momentum in Nonlinear Dynamics: A Simple Implicit Time Integration Scheme, *Comput. Struct.*, **85**(8-7), 437–445.

[4] Hughes, T. J. R., Pister, K. S. and Taylor, R. L. (1979) Implicit-Explicit Finite Elements in Nonlinear Transient Analysis, *Comput. Methods in Appl. Mech. Eng.*, **17**(18), 159–182.

[5] Liang, X., Günay, S. and Mosalam, K.M. (2014) Integrators for Nonlinear Response History Analysis: Revisited, *Istanbul Bridge Conference*, Istanbul, Turkey.

[6] Liang, X., Mosalam, K. M. and Günay, S. (2016) Direct Integration Algorithms for Efficient Nonlinear Seismic Response of Reinforced Concrete Highway Bridges, *ASCE J. Bridge Eng.*, **21**(7), 04016041.

[7] Bathe, K. J. and Wilson, E. L. (1972) Stability and accuracy analysis of direct integration methods *Earthquake Eng. Struct. Dyn.*, **1**(3), 283–291.

[8] Khalil, H. K. (2002) *Nonlinear Systems*, Pearson Prentice Hall, 3rd Edition, Upper Saddle River, N.J.

[9] Haddad, W. M. and Chellaboina V. (2008) *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*, Princeton University Press, Princeton, N.J.

[10] Lyapunov, A. M. (1892) The General Problem of the Stability of Motion (In Russian), *Doctoral dissertation*, Kharkov National University, Ukraine.

[11] Liang, X. and Mosalam, K. M. (2015) *Lyapunov Stability and Accuracy of Direct Integration Algorithms in Nonlinear Dynamic Problems and Considering the Strictly Positive Real Lemma*, SEMM Technical Report UCB/SEMM-2015/01, April.

[12] Liang, X. and Mosalam, K. M. (2016) Lyapunov Stability and Accuracy of Direct Integration Algorithms Applied to Nonlinear Dynamic Problems, *ASCE J. Eng. Mech.*, **142**(5), 04016022.

[13] Boyd, S. and Vandenberghe, L. (2004) *Convex Optimization*, Cambridge University Press, Cambridge, UK.

[14] Liang, X. and Mosalam, K. M. (2016) Lyapunov Stability Analysis of Explicit Direct Integration Algorithms Considering Strictly Positive Real Lemma, *ASCE J. Eng. Mech.*, 10.1061/(ASCE)EM.1943-7889.0001143, 04016079.

[15] Liang, X. and Mosalam, K. M. (2016) Lyapunov Stability Analysis of Explicit Direct Integration Algorithms Applied to Multi-Degree of Freedom Nonlinear Dynamic Problems, *ASCE J. Eng. Mech.*, in press and available online.

[16] Franklin, G. F., Powell J. D. and Emami-Naeini A. (2015) *Feedback Control of Dynamic Systems*, 7th Edition, Pearson Higher Education Inc., Upper Saddle River, N.J.

[17] CVX Research, Inc. (2011) *CVX: Matlab software for disciplined convex programming*, version 2.0. http://cvxr.com/cvx.

[18] Xiao, C. and Hill, D. J. (1999) Generalizations and New Proof of the Discrete-Time Positive Real Lemma and Bounded Real Lemma, *IEEE Trans. Circuits Syst. I*, **46**(6), 740–743.
[19] Lee L. and Chen J. (2003) Strictly Positive Real Lemma and Absolutely Stability for Discrete-Time Descriptor Systems, *IEEE Trans. Circuits Syst. I*, **50**(6), 788–794.
[20] Boyd, S., El Ghaoui L., Feron E. and Balakrishnan V. (1994) *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA.
[21] Mosalam, K.M., Liang, X., Günay, S. and Schellenberg, A. (2013), Alternative Integrators and Parallel Computing for Efficient Nonlinear Response History Analyses, *International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering* (*COMPDYN 2013*), Kos Island, Greece.

An Interpolative Particle Level Set Method

L. Crowl Erickson[1,a)], K.V. Morris[1] and J.A. Templeton[1]

[1]Sandia National Laboratories, P.O. Box 969, Livermore, California 94550, USA

[a)]Presenting and Corresponding author: lcerick@sandia.gov

Abstract

There exist a wide range of applications for solutions to multiphase flow problems with moving interfacial dynamics. These include engineering, fluid mechanics, melting metals, geophysical, medical, computer graphics and image processing. Over the years there have been a large effort in the numerical method community to solve these types of problems. Capturing topological changes with physical accuracy remains a challenge. The two main computational approaches for simulating moving interfaces can be categorized as interface capturing (most notably volume of fluid (VOF) and the level set method) and interface tracking methods. The advantages of both kinds of methods can be combined using hybrid methods, such as the particle level set method [1]. In this paper we propose a new particle level set method which uses an interpolation scheme to update the radii of the interface particles. Preliminary results show that this method can outperform the original particle level set method using fewer particles.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Keywords: Particle methods, level set method, interface methods, multiphase modeling

INTRODUCTION

Advantages of the level set method include natural merging and pinch-off behavior as well as straightforward calculation for the interface normal vector and the radius of curvature. However, mass conservation due to numerical diffusion is a problem that plagues this approach. Reinitialization of the signed distance function is typically necessary for the level set to retain its signed distance property and to limit mass loss. Reinitialization procedures are also prone to numerical diffusion and without careful implementation have the tendency to move the zero level set interface, which is not desired. Another downside of the level set method is that it is limited by the grid size - finer features of the interface or regions of high curvature cannot be resolved if they are thinner than the local grid width.

Lagrangian particle methods conserve mass by nature and are excellent at resolving fine scales and curvatures of the interface in flow regimes that do not cause major deformation or stretching of the interface. The downside is that a large number of points are needed to create the interface and a special approach must be in place to back out the surface geometry (e.g. the surface normal and curvature) since there is no connectivity between particles. These methods fail the shrinking square test [5] and cases with merging fronts, but this is due to how the velocity gets interpolated from a background mesh [2]. Reseeding is also necessary as the interface gets stretched, since the particles can get spread out and fine scale resolution gets lost. A self organizing particle method [6] has been developed, where particles move to adapt to local resolution requirements. As holes and particle clustering form, particles get essentially remeshed using pseudo-forces and dynamic insertion and removal. In addition, it is worth noting that topological changes must be specially handled in Lagrangian particle methods, again since there is no measure of connectivity between particles. A Lagrangian particle level set method was developed by Hieber *et al.* [4] using techniques from vortex methods and particles as essentially quadrature points. This paper develops an approach to cutting and reconnecting the interface.

The hybrid particle level set method was developed by Enright *et al.* [1]. Lagrangian particles are placed near the interface and are used to correct the level set function for mass loss (in addition to a traditional reinitialization approach) when "escaped" particles are detected. Adjacent to both sides of the interface defined by the level set equation, massless marker particles of randomly varying size are initially placed. They are given a sign (positive or negative) and move with same velocity field used for advection of the signed distance function. When these particles end up on the wrong side of the interface due to numerical error, the particles are used to correct the signed distance field using the radius of the marker

particle as a measure of the local level set. In this method, a 5th order WENO scheme for the computation of the spatial term $\nabla\phi$ is combined with a 3rd order TVD Runge Kutta procedure for time integration [3]. In a following paper [2], they show that this correction procedure makes high order integration schemes for the level set function unnecessary. Instead, a semi-Lagrangian method [8] coupled with a first order fast marching method [7] for reinitialization is used as a faster alternative (and the resulting numerical diffusion is effectively mitigated with the incorporation of the particle correction procedure.

Most particle-level set hybrid methods methods use a large number of particles to preform their calculations (64 per cell in 2D in [1]), and most of these particles do not even contribute to the correction procedure since only the escaped particles contribute to updating the level set function. In this work we suggest a different approach, using all particles adjacent to the level set and within one grid spacing, and are able to get a smooth and accurate method with only 12 particles per cell close to the interface. We are able to accomplish this by instead using an interpolation scheme to update grid points near the interface using the distances of nearby particles (escaped or not). Using this approach we do not have to check the escaped status of a particle or calculate the projection of the distance between particle and grid point to see if it is normal or tangent to the interface. In our approach, the radius of an interface particle is the signed distance from the zero level set and we use bilinear interpolation at each grid point to up date the "coarse" grid level set function with the information from the "finer" set of particles near the interface. Our Lagrangian particles do not get reinitialized since they reside near the zero level set (which, within Δx, remains fixed during a reininitialization event).

Results

In order to compare methods, we test this method against the original particle level set method [1]. We look at a pseudo one dimensional test case in which the particles and the level set defined on the (2D) grid were given a linear profile ($\phi(\mathbf{x}) = x - 0.5$). Then the level set field was given a constant error by shifting it by $\Delta x/2$, so the particles and the level set field differ by $\Delta x/2$. We assume that the particles are "correct" and that there is error in the level set field, and use each of the three methods to attempt to correct the zero level set. The results are shown in Figures 1 and 2. Other particle correction methods require a large number of particles per cell since only the escaped particles get used to update the level set. This new approach alternatively uses all neighboring interfacial particle information.

References

[1] Enright, D., Fedkiw, R., Ferziger J., Mitchell, I. (2002) A hybrid particle level set method for improved interface capturing Journal of Computational Physics **183**, 83–116.

[2] Enright, D., Losasso, F. and Fedkiw, R. (2005) A fast and accurate semi-Lagrangian particle level set method Computers & structures **83**(6), 479–490.

[3] Jiang, G. S. and Peng, D. (2000) Weighted ENO schemes for Hamilton–Jacobi equations SIAM Journal on Scientific Computing **21**(6), 2126–2143.

[4] Hieber, S. E. and Koumoutsakos, P. (2005) A Lagrangian particle level set method **210**, 342–367.

[5] Osher, S. and Fedkiw, R. (2007) Level Set Methods and Dynamic Implicit Surfaces

[6] Reboux, Sylvain and Schrader, Birte and Sbalzarini, Ivo F (2012) Journal of Computational Physics **231**(9), 3623–3646.

[7] Sethian, J. A. (1996) A fast marching level set method for monotonically advancing fronts Proceedings of the National Academy of Sciences **93**(4), 1591–1595.

[8] Strain, J. (1999) Semi-Lagrangian methods for level set equations Journal of Computational Physics **151**(2), 498–533.

[9] Wang, Z., Yang, J. and Stern, F. (2009) An improved particle correction procedure for the particle level set method, Journal of Computational Physics **228**(16), 5819–5837.

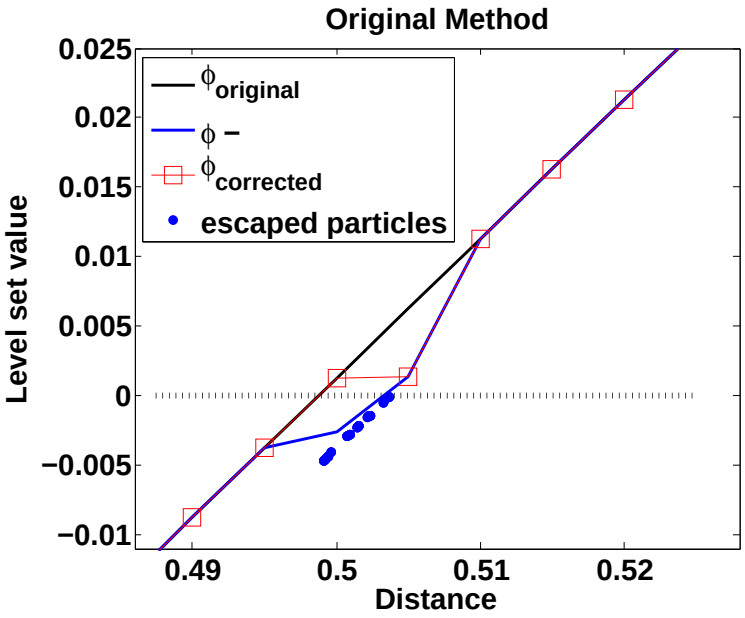


Figure 1: Illustrating particle correction for a psuedo 1D problem in which the level set field is off by $\Delta x/2$ and there is no error in the particle positions and radii. In the original method [1] the location of the zero level set does not get updated.

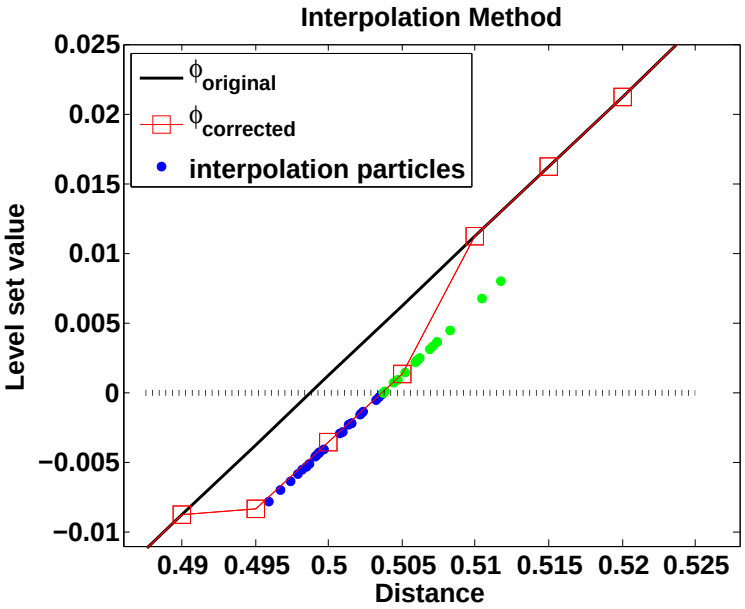


Figure 2: **Illustrating particle correction for a psuedo 1D problem in which the level set field is off by $\Delta x/2$ and there is no error in the particle positions and radii for the new interpolation method proposed in this paper.**

A new BEM for solving multi-medium transient heat conduction

***Wei-Zhe Feng[1], Kai Yang[1], Hai-Feng Peng[1], †Xiao-Wei Gao[1, 2]**

[1]School of Aeronautics and Astronautics, Dalian University of Technology,
Dalian 116024, P.R. China
[2]State Key Laboratory of Structural Analysis for Industrial Equipment,

*Presenting author: fengwz@mail.dlut.edu.cn
†Corresponding author: xwgao@dlut.edu.cn

Abstract

In this paper, a new single interface integral equation method is presented for solving transient heat conduction problems consisting of multi-medium materials with variable thermal properties. Firstly, adopting the fundamental solution for the Laplace equation, the boundary-domain integral equation for transient heat conduction in single medium is established. Then from the established integral equation, a new single interface integral equation is derived for transient heat conduction in general multi-medium functionally graded materials, by making use of the variation feature of the material properties. The derived formulation, which makes up for the lack of boundary integral equation in solving multi-medium problems, has the feature that only a single boundary integral equation is used to solve multi-medium transient heat conduction problems. Compared with conventional multi-domain boundary element method, the newly proposed method is more efficient in data preparing, program coding and computational cost. Based on the implicit backward differentiation scheme, an unconditionally stable and non-oscillatory time marching solution scheme is developed for solving the time-dependent system of differential equations. Numerical examples are given to verify the correctness of the presented method.

Keywords*:* Transient heat conduction, Multi-medium problems, Non-homogeneous problem, Interface integral equation.

1. Introduction

With the advantages of semi-analytical feature and dimensional reduction characteristic, the boundary element method (BEM) has been successfully applied to solve transient heat conduction problems [1-4]. According to the differences of solution procedures, most of the existing approaches can be classified into two broad categories: the transformed space approach (Rizzo and Shippy [5]; Sutradhar et al.[6]; Sutradhar and Paulino [7]; Simoes[8]; Guo et al. [9]), and the time domain approach (Wrobel and Brebbia [10]; Ochiai et al.[11]; Tanaka et al.[12]; Yang and Gao[13]; Al-Jawary el al. [14]; Yu et al. [15]). In the transformed space approach, the time dependent derivative is removed by applying an algebraic transform variable, and the system of equations is solved in the transform space, then inverse transform is employed to reconstitute the solution in time domain. The other kind is the time domain approach, by which the solutions are found directly in the time domain. One implementation

of the time domain approach is the use of time-dependent fundamental solution [10, 11], that can result in a pure boundary integral equation algorithm. However, numerically evaluating the boundary integrals requires both space and time discretization. More details about time-dependent fundamental solution approaches can be found in the works of Wrobel and Brebbia [10] and Ochiai and Sladek [11]. Another implementation of the time domain approach is to employ the fundamental solution for the Laplace equation, and transform the volume integrals associated with time dependent derivative into equivalent boundary integrals. Among the transforming techniques, the dual reciprocity method (DRM) [16, 17], Multiple reciprocity method (MRM) [18], and radial integration method (RIM) [19] are most widely used.

Transient heat conduction BEM has been broadened to a wide range of engineering problems, including non-homogeneous [21], anisotropic [20], and non-linear problems [33]. But most studies mainly focus on single medium. However, most engineering problems involve objects composed of different materials. Therefore, it is important to develop the multi-medium BEM. The conventional widely used technique for solving multi-medium problems is the multi-domain boundary element method (MDBEM) [25-29]. The basic idea of this method is that the whole domain of concern is broken up into a number of separate sub-domains, then a boundary integral equation is written for each sub-domain, and the final system of equations is formed by assembling all contributions of the discretized integral equations for each sub-domain based on the compatibility condition and equilibrium relationship. In the transient heat conduction field, Erhart et al. [31] developed a parallel domain decomposition Laplace transform BEM algorithm for solving the large-scale transient heat conduction problems. Recently, Gao et al. [25, 32] proposed a three-step multi-domain BEM for solving multi-medium non-homogeneous problems.

Although MDBEM is flexible in solving multi-medium problems, it has disadvantages in data preparation and computational time, since twice the element information over the same interface needs to be defined for the adjacent two sub-domains, and twice integrations need to be carried out over interface elements. Moreover, the variable condensation and assembling processes require a higher coding skill to develop a universal program, which heavily influences the computational efficiency. Tracing the issue to its source, the existing boundary integral equations were established on a single medium assumption, therefore it is awkward to solve multi-medium problems through using MDBEM, which involves tedious domain decomposing and assembling processes.

Recently, Gao and his coworkers proposed a single integral equation method, named interface integral BEM (IIBEM), for solving multi-medium problems [34-37]. Through a degeneration method from domain to interface integrals, the integral equation for solving single medium problems can be extended to interface integral equation capable of solving multi-medium steady heat conduction [34], elasticity [35, 36] and elastoplasticity [37] problems. Comparing with the conventional boundary integral equation, an additional interface integral appears in the basic integral equation, embodying the difference of material properties between two adjacent media. The derived formulations make up for the lack of a boundary integral equation in solving multi-medium problems. Compared with MDBEM, the derived integral

equation is very simple in form and only requires integration once over the interface elements. Attributed to the feature of being single integral equation, it is easy to adopt the fast multi-pole method to solve large-scale problems [41].

In this paper, a new single integral equation method is developed for solving general multi-medium transient heat conduction problems. Firstly, the boundary-domain integral equation for single medium non-homogeneous transient heat conduction is established. Then from the established integral equation, the interface integral equation for multi-medium transient heat conduction problems is derived, by a degeneration technique from a domain integral to an interface integral. The new formulation allows the thermal material properties (i.e., thermal conductivity, specific heat and mass density) varying spatially within each medium, and jump across the interfaces between every two adjacent different media. For the first time, a single integral equation method is employed to solve multi-medium transient heat conduction problems with variable material properties.

To solve the time-dependent system of differential equations, the finite difference method (FDM) is used in the discretization of time to approximate the time evolution of temperatures. Based on an implicit backward differentiation scheme, an unconditionally stable and non-oscillatory time marching solution scheme is developed for solving the normal time-dependent system of equations, in which only temperature is involved as the time-dependent unknown variable. Numerical examples are given to verify the correctness of the presented method. The results show that, the presented formulations are robust in solving transient heat conduction in multi-medium functionally graded materials.

2. Review of boundary-domain integral equation for transient heat conduction in single non-homogeneous medium

In this paper, the thermal conductivity *k*, specific heat c_p and mass density ρ are assumed to be functions of spatial coordinates $\mathbf{x}$, i.e. $k(\mathbf{x})$, $c_p(\mathbf{x})$, $\rho(\mathbf{x})$. In this case, the governing equation for transient heat conduction problems can be written as follows:

$$\nabla\left[k(\mathbf{x})\nabla T(\mathbf{x},t)\right]+Q(\mathbf{x})=\rho(\mathbf{x})c_p(\mathbf{x})\frac{\partial T(\mathbf{x},t)}{\partial t} \qquad (t>t_0,\ \mathbf{x}\in\Omega) \tag{1}$$

where, $T(\mathbf{x},t)$ is the temperature at location $\mathbf{x}$ at time *t*; $Q(\mathbf{x})$ is the heat generation; t_0 is the initial time, and Ω represents the computational domain.

The initial condition is

$$T(\mathbf{x},0)=T_0(\mathbf{x}) \tag{2}$$

where, $T_0(\mathbf{x})$ is the initial temperature. On the boundary, Dirichlet and Neumann boundary conditions are prescribed as follows:

$$T(\mathbf{x},t) = \overline{T}(\mathbf{x},t), \qquad \mathbf{x} \in \Gamma_T \tag{3}$$

$$q(\mathbf{x},t) = -k(\mathbf{x})\frac{\partial T(\mathbf{x},t)}{\partial n} = \overline{q}(\mathbf{x},t), \qquad \mathbf{x} \in \Gamma_q \tag{4}$$

where, $q(\mathbf{x},t)$ is the normal heat flux on the boundary Γ of the computational domain Ω; n is the unit outward normal to Γ; and $\Gamma = C(\Gamma_T \bigcup \Gamma_q) = \partial\Omega$, $\Gamma_T \bigcap \Gamma_q = \varnothing$. In Eqs. (3) and (4), $\overline{T}(\mathbf{x},t)$, $\overline{q}(\mathbf{x},t)$ are the given temperature and heat flux on the boundary, usually prescribed as given functions.

Taking the fundamental solution for the Laplace equation as the weight function, applying the weighted residual technique to Eq.(1), and using the Gauss' divergence theorem, the boundary-domain integral equation for solving single medium transient heat conduction problems can be established [13]:

$$\begin{aligned} c\widetilde{T}(\mathbf{y},t) = &-\int_\Gamma G(\mathbf{x},\mathbf{y})\, q(\mathbf{x},t)\, d\,\Gamma(\mathbf{x}) - \int_\Gamma \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n}\widetilde{T}(\mathbf{x},t)\, d\,\Gamma(\mathbf{x}) \\ &+ \int_\Omega G(\mathbf{x},\mathbf{y})\, Q(\mathbf{x})\, d\,\Omega(\mathbf{x}) + \int_\Omega V(\mathbf{x},\mathbf{y})\,\widetilde{T}(\mathbf{x},t)\, d\,\Omega(\mathbf{x}) \\ &- \int_\Omega \frac{\rho(\mathbf{x})c_p(\mathbf{x})}{k(\mathbf{x})} G(\mathbf{x},\mathbf{y})\frac{\partial \widetilde{T}(\mathbf{x},t)}{\partial t} d\,\Omega(\mathbf{x}) \end{aligned} \tag{5}$$

where c=1 for internal points and 1/2 for smooth boundary points; $\mathbf{y}$ represents the source point, and $\mathbf{x}$ the field point; $G(\mathbf{x},\mathbf{y})$ is the fundamental solution for Laplace equation, $\partial G(\mathbf{x},\mathbf{y})/\partial n$ and $V(\mathbf{x},\mathbf{y})$ are the derived kernels. These quantities can be expressed as follows

$$G(\mathbf{x},\mathbf{y}) = \begin{cases} \dfrac{1}{2\pi}\ln(\dfrac{1}{r}) & \text{for 2D problem} \\ \dfrac{1}{4\pi r} & \text{for 3D problem} \end{cases} \tag{6}$$

$$\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n} = \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial x_i} n_i(\mathbf{x}) = \frac{-1}{2\alpha\, r^\alpha}\frac{\partial r}{\partial x_i} n_i \tag{7}$$

$$\begin{aligned} V(\mathbf{x},\mathbf{y}) &= \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial x_i}\frac{\partial \widetilde{k}(\mathbf{x})}{\partial x_i} \\ &= \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial x_i}\frac{1}{k(\mathbf{x})}\frac{\partial k(\mathbf{x})}{\partial x_i} \end{aligned} \tag{8}$$

where, $\alpha = \beta - 1$ ($\beta = 2$ for 2D problems and 3 for 3D problems); r is the distance between source point $\mathbf{y}$ and field point $\mathbf{x}$; $\partial r / \partial x_i$ is the partial derivative of r with respect to coordinate x_i; n_i is the i-th component of n. In Eqs. (7) and (8) and through the paper, the repeated subscripts represent summation.

In Eq. (5) normalized temperature and thermal conductivity are utilized, by considering the product of temperature and thermal conductivity as the unknown variable [13, 38]

$$\tilde{T}(\mathbf{x}) = k(\mathbf{x})T(\mathbf{x}) \tag{9}$$

$$\tilde{k}(\mathbf{x}) = \ln k(\mathbf{x}) \tag{10}$$

Integral equation (5) is the boundary-domain integral equation for solving general single medium transient heat conduction problems. And through the radial integration method (RIM) transforming the involved domain integrals in Eq.(5) to the boundary, a pure boundary element algorithm without internal cells for single medium transient heat conduction can be developed [13].

From Eq.(10) we can see that the kernel function $V(\mathbf{x},\mathbf{y})$ involves the spatial derivative of the thermal conductivity $\partial k(\mathbf{x}) / \partial x_i$, which indicates that $k(\mathbf{x})$ should vary continuously without jump in the domain Ω. However, for a problem consisting of multiple media, the thermal conductivity jumps across the interfaces between two adjacent materials, the derivative $\partial k(\mathbf{x}) / \partial x_i$ will lead to an infinity. Therefore, Eq.(5) is not valid for multi-medium problems. However, the singular kernel is in fact integrable as shown in section 3. In section 3, we will deal with multi-medium problems in which the conductivity is not continuous across the interfaces of media. In this case, the domain integral involved in Eq. (5) is degenerated into an interface integral between two adjacent materials.

3. Interface integral equation for multi-medium transient heat conduction

For the sake of convenience and not losing generality, a problem consisting of two media characterized by conductivities $k_1(\mathbf{x})$ and $k_2(\mathbf{x})$ is considered as shown in Fig. 1, in which Γ is the outer boundary of the problem, Γ_I is the interface between media $k_1(\mathbf{x})$ and $k_2(\mathbf{x})$, and n' is the outward normal to Γ_I. Since the thermal conductivity jumps across the interface Γ_I, we separate a narrow domain Ω_3 around Γ_I, which has a constant infinitesimal thickness Δh along the interface (see Fig.1).

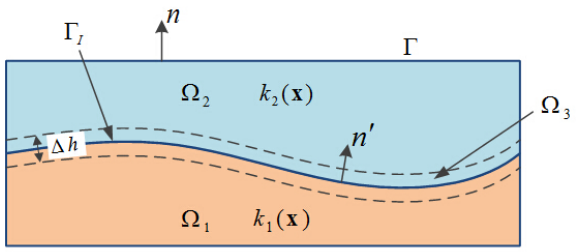


Figure 1. A narrow domain separated around interface of two media

Referring to Fig. 1, the domain integral involving kernel $V(\mathbf{x},\mathbf{y})$ in Eq. (4) can be written as

$$\begin{aligned}\int_\Omega V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)d\Omega &= \lim_{\Delta h\to 0}\left(\int_{\Omega_1+\Omega_2} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\Omega\right)+\lim_{\Delta h\to 0}\left(\int_{\Omega_3} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\Omega\right)\\ &= \int_{\overline{\Omega}} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\Omega+\lim_{\Delta h\to 0}\left(\Delta h\int_{\Gamma_I} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\Gamma\right)\end{aligned} \tag{11}$$

where $\overline{\Omega}$ represents the whole integration domain consisting of all media with an infinite narrow domain isolated out, and in a specific medium $V(\mathbf{x},\mathbf{y})$ is determined by Eq.(8). From Eq.(8), we can see that the kernel $V(\mathbf{x},\mathbf{y})$ involved in the above equation is related to the gradient of the normalized conductivity $\partial\tilde{k}(\mathbf{x})/x_i$. With the existence of a jump effect across the interface Γ_I, the second integral item on the right hand side of Eq.(8) can be manipulated as follows [34, 37]:

$$\begin{aligned}\lim_{\Delta h\to 0}\left(\Delta h\int_{\Gamma_I} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\Gamma\right) &= \int_{\Gamma_I}\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial x_i}n_i'\frac{1}{k(\mathbf{x})}\Delta k(\mathbf{x})\,\tilde{T}(\mathbf{x},t)\,d\Gamma\\ &= \int_{\Gamma_I}\frac{\partial G(x,y)}{\partial n_i'}\Delta k(\mathbf{x})\,T(\mathbf{x},t)\,d\Gamma\end{aligned} \tag{12}$$

Substituting Eq.(12) into Eq.(11), and the result into Eq.(5), the final temperature integral equation is derived as follows:

$$\begin{aligned}c\,\tilde{T}(\mathbf{y},t) = &-\int_\Gamma G(\mathbf{x},\mathbf{y})\,q(\mathbf{x},t)\,d\,\Gamma(\mathbf{x})-\int_\Gamma\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n}\tilde{T}(\mathbf{x},t)\,d\,\Gamma(\mathbf{x})\\ &+\int_{\Gamma_I}\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n'}\Delta k(\mathbf{x})\,T(\mathbf{x},t)\,d\,\Gamma(\mathbf{x})+\int_\Omega G(\mathbf{x},\mathbf{y})\,Q(\mathbf{x})\,d\,\Omega(\mathbf{x})\\ &+\int_{\overline{\Omega}} V(\mathbf{x},\mathbf{y})\,\tilde{T}(\mathbf{x},t)\,d\,\Omega(\mathbf{x})-\int_\Omega\frac{\rho(\mathbf{x})c_p(\mathbf{x})}{k(\mathbf{x})}G(\mathbf{x},\mathbf{y})\frac{\partial\tilde{T}(\mathbf{x},t)}{\partial t}\,d\,\Omega(\mathbf{x})\end{aligned} \tag{13}$$

Eq.(13) is the established interface integral equation for solving multi-medium transient heat conduction problems. The time-dependent effect is embodied by the domain integral involving the time derivative of temperature $\partial\tilde{T}(\mathbf{x},t)/\partial t$. The jump effect of thermal conductivities across the interfaces between every two adjacent media is embodied by the

interface integral item carried out on Γ_I; The non-homogeneous effect of material properties is embodied by the domain integral item involving kernel $V(\mathbf{x},\mathbf{y})$.

In numerical implementation, three types of points are introduced in discretization: outer boundary points on Γ, interface points on Γ_I, and internal points in Ω. Eq.(13) is only suitable for the outer boundary points and internal points by setting $c=1/2$ for smooth outer boundary and $c=1$ for internal points, respectively. When the source point $\mathbf{y}$ is located on the interface points, a similar integral equation can be obtained by letting $\mathbf{y}\to\Gamma_I$ [34]:

$$\begin{aligned} c^I\,\tilde{T}(\mathbf{y}^I,t) = &-\int_\Gamma G(\mathbf{x},\mathbf{y}^I)\,q(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) - \int_\Gamma \frac{\partial G(\mathbf{x},\mathbf{y}^I)}{\partial n}\,\tilde{T}(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) \\ &+ \int_{\Gamma_I} \frac{\partial G(\mathbf{x},\mathbf{y}^I)}{\partial n'}\,\Delta k(\mathbf{x})\,T(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) + \int_\Omega G(\mathbf{x},\mathbf{y}^I)\,Q(\mathbf{x})\,d\,\Omega(\mathbf{x}) \\ &+ \int_{\overline{\Omega}} V(\mathbf{x},\mathbf{y}^I)\,\tilde{T}(\mathbf{x},t)\,d\,\Omega(\mathbf{x}) - \int_\Omega \frac{\rho(\mathbf{x})c_p(\mathbf{x})}{k(\mathbf{x})}\,G(\mathbf{x},\mathbf{y}^I)\,\frac{\partial\tilde{T}(\mathbf{x},t)}{\partial t}\,d\,\Omega(\mathbf{x}) \end{aligned} \tag{14}$$

where, $\mathbf{y}^I$ represents the source points located on the interface; c^I is the free term coefficient, and for smooth interface, the expression of c^I is

$$c^I = \frac{1}{2}[k_1(\mathbf{y}^I) + k_2(\mathbf{y}^I)] \tag{15}$$

where, $k_1(\mathbf{y}^I)$ and $k_2(\mathbf{y}^I)$ are the thermal conductivities for the adjacent two different materials on the location of $\mathbf{y}^I$.

For the convenience, taking into account Eqs. (9) and (10), we can rewrite Eqs. (13) and (14) in an uniform form

$$\begin{aligned} \hat{k}(\mathbf{y})\,T(\mathbf{y},t) = &-\int_\Gamma G(\mathbf{x},\mathbf{y})\,q(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) - \int_\Gamma \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n}\,k(\mathbf{x})\,T(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) \\ &+ \int_{\Gamma_I} \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n'}\,\Delta k(\mathbf{x})\,T(\mathbf{x},t)\,d\,\Gamma(\mathbf{x}) + \int_\Omega G(\mathbf{x},\mathbf{y})\,Q(\mathbf{x})\,d\,\Omega(\mathbf{x}) \\ &+ \int_{\overline{\Omega}} \hat{V}(\mathbf{x},\mathbf{y})\,T(\mathbf{x},t)\,d\,\Omega(\mathbf{x}) - \int_\Omega \rho(\mathbf{x})c_p(\mathbf{x})\,G(\mathbf{x},\mathbf{y})\,\frac{\partial T(\mathbf{x},t)}{\partial t}\,d\,\Omega(\mathbf{x}) \end{aligned} \tag{16}$$

where,

$$\hat{k}(\mathbf{y})=\begin{cases}\dfrac{1}{2}k(\mathbf{y}) & \text{for smooth outer boundary points on } \Gamma \\ \dfrac{1}{2}[k_1(\mathbf{y})+k_2(\mathbf{y})] & \text{for smooth interface points on } \Gamma_I \\ k(\mathbf{y}) & \text{for internal points in } \overline{\Omega}\end{cases} \tag{17}$$

$$\hat{V}(\mathbf{x},\mathbf{y})=\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial x_i}\frac{\partial k(\mathbf{x})}{\partial x_i} \tag{18}$$

To avoid discretizing the domain Ω into internal cells for evaluating domain integrals involved in the above integral equations using the conventional cell-integration technique [39], a robust transformation technique from domain integrals into equivalent boundary integrals is described in reference [19]. In the paper, the three domain integrals involved in Eq.(16) are transformed into equivalent boundary integrals by the radial integration method (RIM) [13].

4. Numerical implementation

Eq. (16) is the boundary-interface-domain integral equation for solving multi-medium transient heat conduction problems with variable material properties, and by employing RIM transforming the involved domain integrals into equivalent boundary integrals, a pure boundary element method without internal cells can be developed.

4.1 System of differential equations

After discretizing the outer boundary Γ and interface Γ_I into a series of boundary elements and collocating the source point $\mathbf{y}$ through all boundary, interface, and internal nodes, we can form the system of differential equations for Eq.(16). Assuming that the BEM model involves N_b boundary nodes, N_c interface nodes, and N_i internal nodes, the total number of nodes is $N_A = N_b + N_c + N_i$. The discrete form of integral equation (16) is as follows:

$$\begin{bmatrix}\mathbf{H}_{bb} & \mathbf{H}_{bc} & \mathbf{0} \\ \mathbf{H}_{cb} & \mathbf{H}_{cc} & \mathbf{0} \\ \mathbf{H}_{ib} & \mathbf{H}_{ic} & \mathbf{H}_{ii}\end{bmatrix}_{N_A\times N_A}\begin{Bmatrix}\mathbf{T}_b \\ \mathbf{T}_c \\ \mathbf{T}_i\end{Bmatrix}_{N_A}=\begin{bmatrix}\mathbf{G}_{bb} \\ \mathbf{G}_{cb} \\ \mathbf{G}_{ib}\end{bmatrix}_{N_A\times N_b}\mathbf{q}_b+\begin{Bmatrix}\mathbf{f}_b \\ \mathbf{f}_c \\ \mathbf{f}_i\end{Bmatrix}_{N_A}+\mathbf{V}_{N_A\times N_A}\mathbf{T}-\mathbf{C}_{N_A\times N_A}\dot{\mathbf{T}}_A \tag{19}$$

where, $\mathbf{H}_{bc}$, $\mathbf{H}_{cc}$ and $\mathbf{H}_{ic}$ correspond to the coefficients of the interface integrals; $\mathbf{H}_{bb}$, $\mathbf{H}_{cb}$, $\mathbf{H}_{ib}$ and $\mathbf{G}_{bb}$, $\mathbf{G}_{cb}$ $\mathbf{G}_{ib}$ correspond to the outer boundary integrals; $\mathbf{H}_{ii}$ is diagonal matrix consisting of free term coefficients for internal points. $\mathbf{V}$ and $\mathbf{C}$ (both with dimensions of $N_A \times N_A$) correspond to the last two domain integrals in Eq.(16). And $\mathbf{f}_b$, $\mathbf{f}_c$

and $\mathbf{f}_i$ are the domain integration results for heat sources. $\mathbf{T}_b$ and $\mathbf{q}_b$ are the temperatures and heat fluxes for the boundary nodes respectively, and

$$\mathbf{T}_b = \begin{Bmatrix} \overline{\mathbf{T}}_1 \\ \mathbf{T}_2 \end{Bmatrix}, \quad \mathbf{q}_b = \begin{Bmatrix} \mathbf{q}_1 \\ \overline{\mathbf{q}}_2 \end{Bmatrix} \tag{20}$$

In which, $\overline{\mathbf{T}}_1$ and $\overline{\mathbf{q}}_2$ are the given temperatures on the Dirichlet boundaries and and heat fluxes on the Neumann boundaries, respectively.

Rearrange the system of equations Eq.(19) by transposing columns of [**H**], [**G**] and [**V**] from one side to the other, gathering all unknowns to the left-hand side, then we can rewrite Eq.(19) as

$$\mathbf{A}\,\mathbf{x} = \mathbf{y} - \mathbf{C}\,\dot{\mathbf{T}} \tag{21}$$

where,

$$\mathbf{x} = \begin{Bmatrix} \mathbf{q}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_c \\ \mathbf{T}_i \end{Bmatrix} = \begin{Bmatrix} \mathbf{q}_1 \\ \mathbf{T}_x \end{Bmatrix} \tag{22}$$

In which, $\mathbf{T}_x$ consists of unknown temperatures on the Neumann boundary conditional nodes, the interface nodes and internal nodes.

By writing the coefficient matrices $\mathbf{A}$, $\mathbf{C}$ in block form, we can reconstitute Eq.(21) as follows:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{1x} \\ \mathbf{A}_{x1} & \mathbf{A}_{xx} \end{bmatrix} \begin{Bmatrix} \mathbf{q}_1 \\ \mathbf{T}_x \end{Bmatrix} = \begin{Bmatrix} \mathbf{y}_1 \\ \mathbf{y}_x \end{Bmatrix} - \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{1x} \\ \mathbf{C}_{x1} & \mathbf{C}_{xx} \end{bmatrix} \begin{Bmatrix} \dot{\mathbf{T}}_1 \\ \dot{\mathbf{T}}_x \end{Bmatrix} \tag{23}$$

In Eq.(21), the unknown heat fluxes at nodes on Dirichlet boundary can be expressed by the unknown temperatures as following

$$\mathbf{q}_1 = [\mathbf{A}_{11}]^{-1}(\mathbf{y}_1 - \mathbf{A}_{1x}\mathbf{T}_x - \mathbf{C}_{11}\dot{\mathbf{T}}_1 - \mathbf{C}_{1x}\dot{\mathbf{T}}_x) \tag{24}$$

Given that $\mathbf{T}_1$ are known temperatures on Dirichlet boundary, which do not vary with time, therefore $\dot{\mathbf{T}}_1 = \mathbf{0}$, substituting back into Eq.(24) yields the following equation:

$$\mathbf{q}_1 = \overline{\mathbf{E}}_{1x}\mathbf{T}_x + \overline{\mathbf{D}}_{1x}\dot{\mathbf{T}}_x + \overline{\mathbf{F}}_1 \tag{25}$$

where,

$$\overline{\mathbf{E}}_{1x} = -[\mathbf{A}_{11}]^{-1}\mathbf{A}_{1x} \tag{26}$$

$$\overline{\mathbf{D}}_{1x} = -[\mathbf{A}_{11}]^{-1}\mathbf{C}_{1x} \tag{27}$$

$$\overline{\mathbf{F}}_{1} = [\mathbf{A}_{11}]^{-1}\mathbf{y}_{1} \tag{28}$$

Substituting Eq.(25) back to Eq.(23), $\mathbf{q}_1$ can be eliminated from the system of differential equations, and the regularized form of differential equations that is only concerned with temperature can be derived:

$$\overline{\mathbf{A}}_{xx}\dot{\mathbf{T}}_{x} = \overline{\mathbf{B}}_{xx}\mathbf{T}_{x} + \overline{\mathbf{C}}_{xx}\dot{\mathbf{T}}_{1} + \overline{\mathbf{Y}}_{x} \tag{29}$$

Similarly, $\dot{\mathbf{T}}_1 = \mathbf{0}$ with the assumption that the temperature boundary conditions do not vary with time, Eq.(29) can be changed into the following form

$$\overline{\mathbf{A}}_{xx}\dot{\mathbf{T}}_{x} = \overline{\mathbf{B}}_{xx}\mathbf{T}_{x} + \overline{\mathbf{Y}}_{x} \tag{30}$$

where,

$$\overline{\mathbf{A}}_{xx} = \mathbf{C}_{xx} - \mathbf{A}_{x1}[\mathbf{A}_{11}]^{-1}\mathbf{C}_{1x} \tag{31}$$

$$\overline{\mathbf{B}}_{xx} = \mathbf{A}_{x1}[\mathbf{A}_{11}]^{-1}\mathbf{A}_{1x} - \mathbf{A}_{xx} \tag{32}$$

$$\overline{\mathbf{Y}}_{x} = \mathbf{y}_{x} - \mathbf{A}_{x1}[\mathbf{A}_{11}]^{-1}\mathbf{y}_{1} \tag{33}$$

Now, Eq.(30) is the normalized system of differential equations only concerned with unknown temperatures. To solve the time-dependent system of equations Eq. (30), the finite difference method (FDM) or precise integration method (PIM) [15] can be used to approximate the time evolution of temperatures. In this paper, we adopt the backward differentiation scheme [42], which is unconditionally stable and non-oscillatory in solving system of ordinary differential equations, to solve Eqs.(30) and (25).

4.2 Time marching scheme

To solve the equation set (30) and (25), we adopt the finite difference method to approximate the time derivative term:

$$\dot{\mathbf{T}}_{x} = \frac{\mathbf{T}_{x}^{n+1} - \mathbf{T}_{x}^{n}}{\Delta t} \tag{34}$$

$$\mathbf{T}_{x} = \theta\,\mathbf{T}_{x}^{n+1} + (1-\theta)\mathbf{T}_{x}^{n} \tag{35}$$

where, $\mathbf{T}_x^n$ represents the temperature at the *n*-th time step, and θ is the Euler parameter which usually takes a value between 0.5 and 1 [40]. In this study, we take $\theta = 1$. Substituting Eqs.(34) and (35) into Eq.(30), yields

$$\mathbf{T}_{x}^{n+1} = \mathbf{M}\mathbf{T}_{x}^{n} + \mathbf{N} \tag{36}$$

where,

$$\mathbf{M} = [\overline{\mathbf{A}}_{xx} / \Delta t - \theta \overline{\mathbf{B}}_{xx}]^{-1} [\overline{\mathbf{A}}_{xx} / \Delta t + (1-\theta)\overline{\mathbf{B}}_{xx}] \tag{37}$$

$$\mathbf{N} = [\overline{\mathbf{A}}_{xx} / \Delta t - \theta \overline{\mathbf{B}}_{xx}]^{-1} \overline{\mathbf{Y}}_{x} \tag{38}$$

where $\overline{\mathbf{A}}_{xx}$, $\overline{\mathbf{B}}_{xx}$ and $\overline{\mathbf{Y}}_{x}$ are defined by Eqs.(31)-(33).

With a similar process, substituting Eqs.(34) and (35) into Eq.(25), the heat fluxes $\mathbf{q}_1$ at nodes on Dirichlet boundary can be evaluated at each time step:

$$\mathbf{q}_1^{n+1} = \mathbf{J}\, \mathbf{T}_x^{n+1} + \mathbf{K}\, \mathbf{T}_x^{n} + \overline{\mathbf{F}}_1 \tag{39}$$

where,

$$\mathbf{J} = \theta \overline{\mathbf{E}}_{1x} + \overline{\mathbf{D}}_{1x} / \Delta t \tag{40}$$

$$\mathbf{K} = (1-\theta) \overline{\mathbf{E}}_{1x} - \overline{\mathbf{D}}_{1x} / \Delta t \tag{41}$$

In Eqs.(39)-(41), $\overline{\mathbf{E}}_{1x}$, $\overline{\mathbf{D}}_{1x}$ and $\overline{\mathbf{F}}_1$ are determined by Eqs.(26)-(28). Now, Eq.(36) and Eq.(39) can be employed to trace the time evolution of temperature and heat flux.

5. Numerical example

A Fortran code, named SIEBEM (single interface integral equation boundary element method) using the presented interface integral formulations in this paper has been developed.

5.1 Transient heat conduction in a two-media composed square flange

This example focuses on a square flange with four reinforced mounting holes, which are equally distributed along a circle with radius of $\Phi = 5.374\,\mathrm{cm}$, as shown in Fig. 2. The flange and the mounting holes are made of different materials, marked with different colors in Fig.2. The initial temperature is assumed to be $T_0 = 0$℃. The temperature at the inner circular side suddenly changes to 800℃, while the temperature at the outer side of the square keeps 0℃. Inner sides of the mounting holes are temperature insulated.

Due to symmetry of the flange, only a quarter is analyzed. The geometry and boundary conditions are shown in Fig. 3, where point O(x=0, y=0) is the spatial origin, point C (x =3.8, y =3.8) represents the center of the mounting hole. Symbols Ω_1 and Ω_2 are the computational domains for two different media, respectively.

The material properties for media Ω_1 and Ω_2 are listed in Table 1, where k represents the thermal conductivity, c_p represents the specific heat, and ρ the mass density.

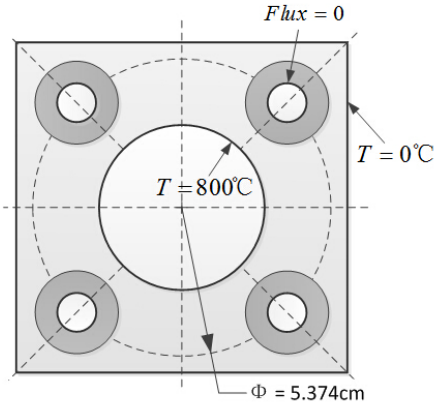


Figure 2. Square flange

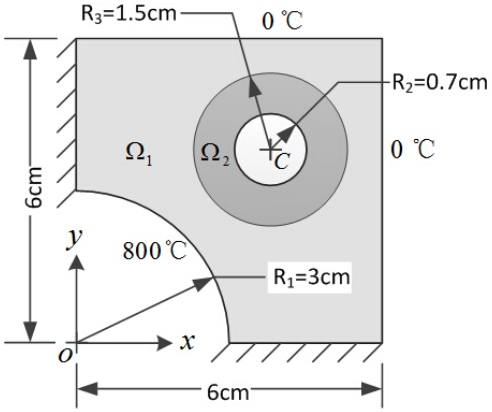


Figure 3. Quarter of the flange

Table 1. Material properties for each medium

Medium	k (W/m·K)	c_p (J/kg·K)	ρ (kg/m^3)
Ω_1	200	490	8.9×10^3
Ω_2	40	900	6.6×10^3

The inner circular side of the flange is discretized into 30 equally-spaced linear boundary elements, and each of the two straight outer boundary lines is discretized into 35 equally-spaced linear elements. The whole BEM model employs 998 nodes, in which 180 are outer boundary nodes, 40 are interface nodes, and 778 are internal nodes distributed within the domain. Fig. 4 shows the BEM model for computation. For comparison, this model is also analyzed using the conventional multi-domain boundary element method (MDBEM) reported in [25]. By using the same scale of mesh discretization, the MDBEM shown in Fig. 5 employs 998 nodes and 260 boundary elements. Since the interface marked with 'F' shown in Fig.5 has to be discretized into elements in each medium, the number of elements used in the MDBEM model is bigger than that used in the SIEBEM model. Therefore the computation scale for the MDBEM model is bigger than that of the SIEBEM model.

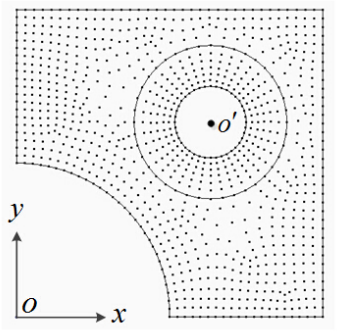


Figure 4. SIEBEM model

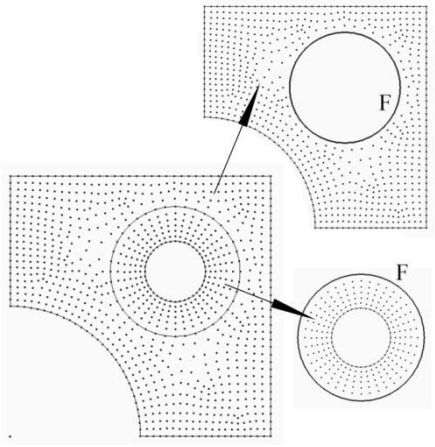


Figure 5. MDBEM model

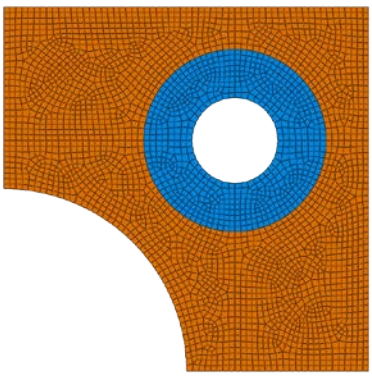


Figure 6. FEM mesh

A 10s time period is analyzed with 100 equally discretized time steps, and the length of each time step is $\Delta t = 0.1\,\mathrm{s}$. To provide a reference solution to compare with the BEM results, the solution of this problem is computed using the commercial software ABAQUS. Fig.6 shows the FEM mesh.

Around the inner circle of the mounting hole with radius of R_2=0.7cm, the temperature distribution at different times calculated by SIBEM, MDBEM and FEM software are shown in Fig.7. And Fig.8 shows the temperature distribution along x direction at the y =0 symmetric straight edge. Fig. 9 compares the BEM results with FEM results for the temperature with time around the interface circle with the radius R_3=1.5cm in Fig.3. From Figs. 7 - 9, we can see that the results of SIEBEM coincide well with the results of MDBEM and FEM software, which validates the correctness of the presented method.

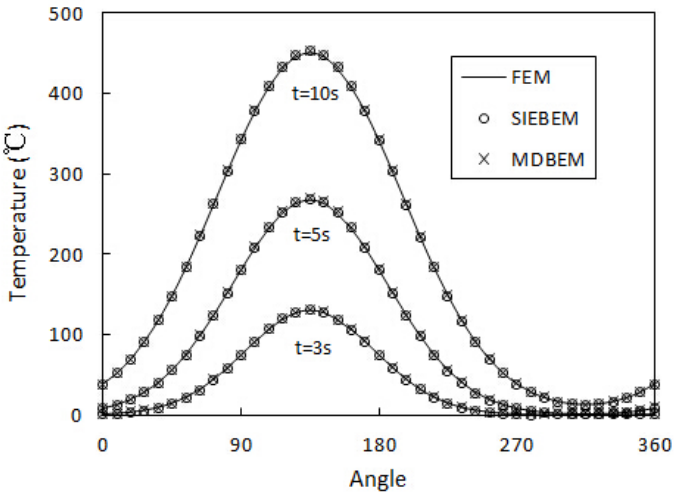


Figure 7. Temperature distribution along inner circle R2=0.7cm

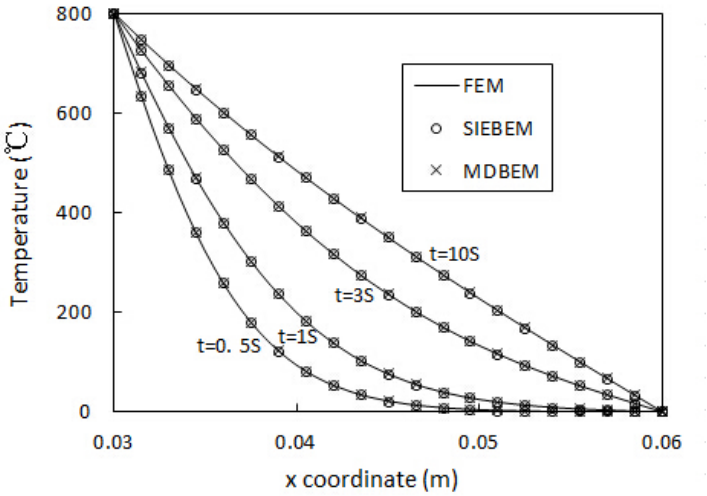


Figure 8. Temperature distribution along the y = 0 straight edge

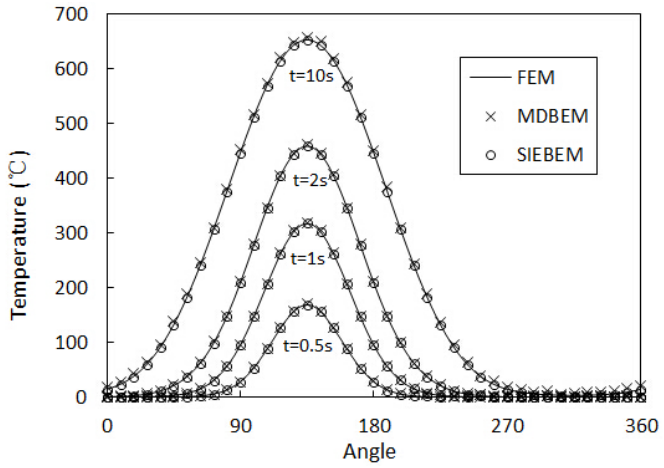
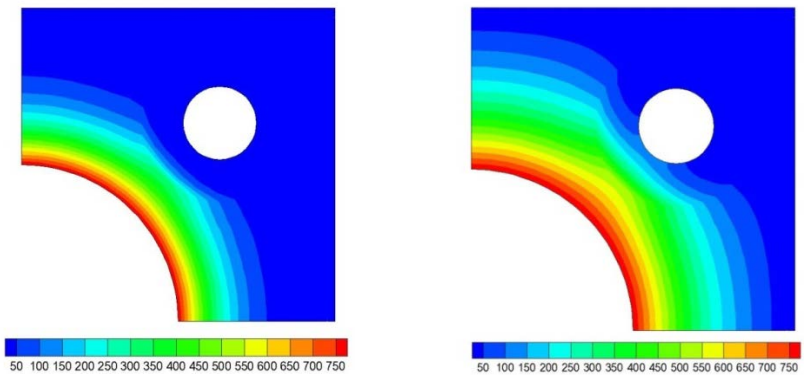
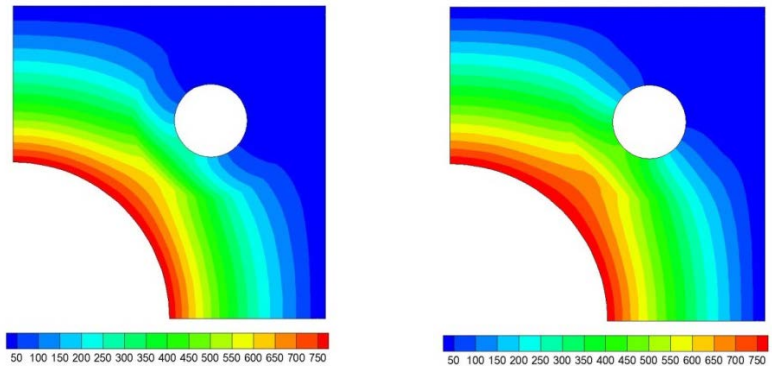


Figure 9. Temperature distribution along the interfacial circle R_3=1.5cm

Fig.10 shows the contour plots of the temperature distribution at different time. From Fig.10, we can easily find the discontinuous effect of temperature distribution when crossing the interfacial circle between the body of the flange and the mounting hole.



a ***b***



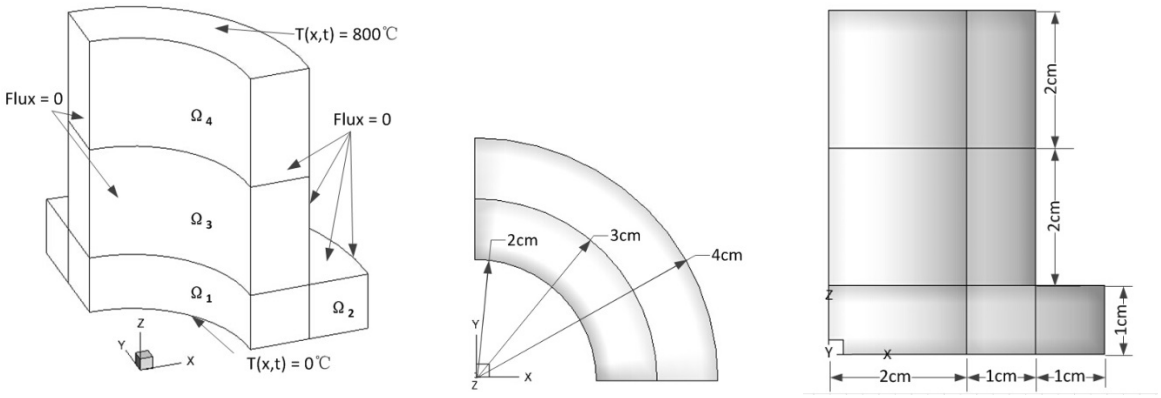
c ***d***

Figure 10. Counter plot of the temperature at different times:

(*a*) *t* = 1s ; (*b*) *t* = 2s; (*c*) *t* = 5s; (*d*) *t* = 10s

5.2 3D transient heat conduction in a four-media composed hollow cylinder

The third example to be considered is a hollow cylinder with a reinforcing stair, which is composed of four different media denoted by Ω_1, Ω_2, Ω_3 and Ω_4, as shown in Fig.11 (*a*) . The initial temperature is assumed to be T_0=0℃. Then the temperature at the top surface changes to 800℃, while the temperature at the bottom surface stays as 0℃. The other sides are thermally insulated. Due to symmetry of the problem, only a quarter of the hollow cylinder is modeled. Figs. 11(*b*) and 11(*c*) shows detailed dimensions and boundary conditions for the geometrical model.



a ***b*** ***c***

Figure 11. Four-media composed hollow cylinder: (*a*) 3D global view; (*b*) top view;

(*c*) right-side view

The material properties of the four media are prescribed as functions of spatial coordinates, and Table 2 gives these specific functions of coordinates for each medium. In order to show the variation of material properties with respect to the spatial coordinates more vividly, the profiles of thermal conductivity k and specific heat c_p are illustrated in Fig. 12. From Fig.12 we can see that the material properties vary in space continuously within each medium but

jump across the interfaces between different media.

Table 2. Material properties for each medium

	k (W/m·K)	c_p (J/kg·K)	ρ (kg/m^3)
Ω_1	$200\times e^{50z}$	$500\times e^{30z}$	8900
Ω_2	$400+10^4(\sqrt{x^2+y^2}-0.03)$	$900-2\times10^4(\sqrt{x^2+y^2}-0.03)$	2700
Ω_3	$200+10^4(z-0.01)$	$500-10^4(z-0.01)$	7900
Ω_4	$600-10^6(z-0.03)^2$	$700-5\times10^5(z-0.03)^2$	6900

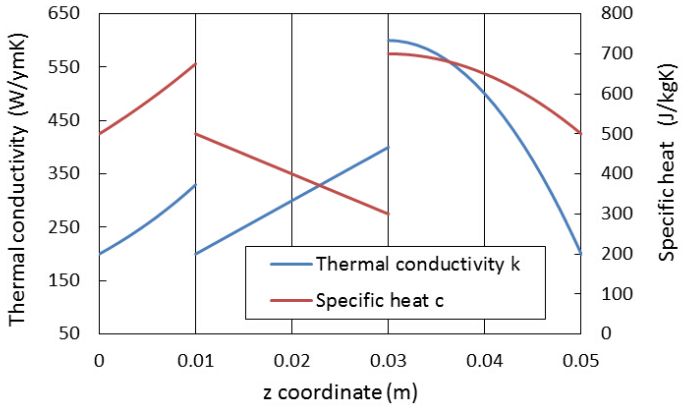


Figure 12. Profiles of thermal conductivity and specific heat along *z*-direction

The BEM mesh employs 880 4-node linear elements, in which 144 are interface elements distributed on the three interfaces between every two different media. Discontinuous elements are used at the intersection points between the interface and outer boundary, ensuring that a collocation point is either used by an outer boundary element or an interface element, see Fig. 13. The total number of nodes is 1546, among which 823 are boundary nodes, 195 are interface nodes, and 528 are internal nodes. Fig. 13 shows the BEM model for computation, in which different media are marked with different colors.

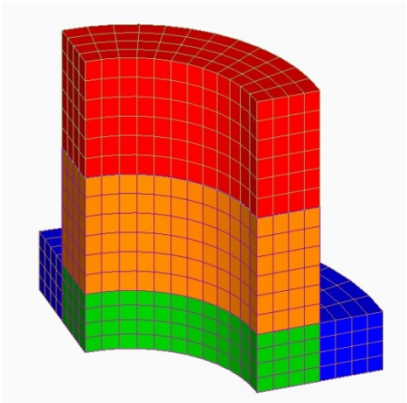


Figure 13. BEM mesh for the hollow cylinder

A 10 second time period is analyzed with 100 equally discretized time steps, and the length of each time step is Δt = 0.1s. For comparison, this model is also analyzed with ABAQUS by using the UMATHT subroutine [43]. Fig. 14 shows the distribution of temperature along z direction over the inner side vertical line of x = 2 cm and y = 0 cm. Fig. 15 shows the temperature distribution along x direction over the spatial straight line of y = 0 cm and z = 1 cm. From Figs. 14 and 15 we can see that the BEM results coincide well with the FEM results, demonstrating the correctness of the proposed method. From Fig. 14, we can easily find that three segment of curves compose the profile of temperature at each time step. And in Fig.15, the profile is composed by two segments. This effect is caused by the jump effect of material properties in multi-medium problems.

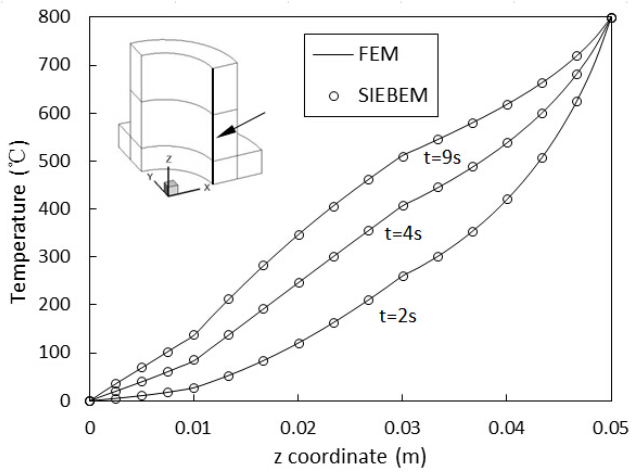


Figure 14. Temperature distribution along the z coordinate direction

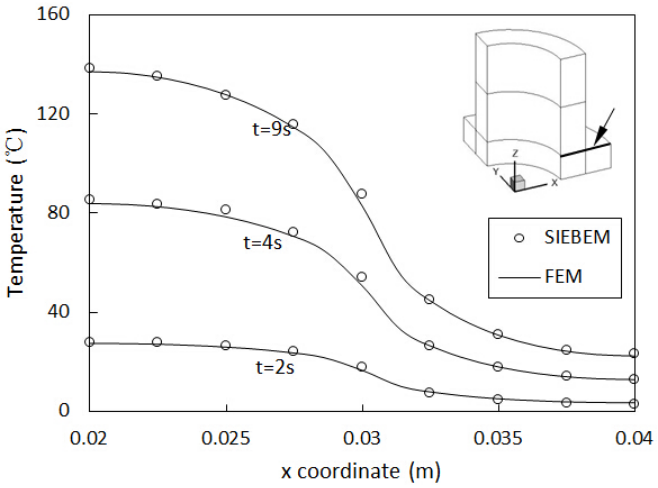


Figure 15. Temperature distribution along the *x* coordinate direction

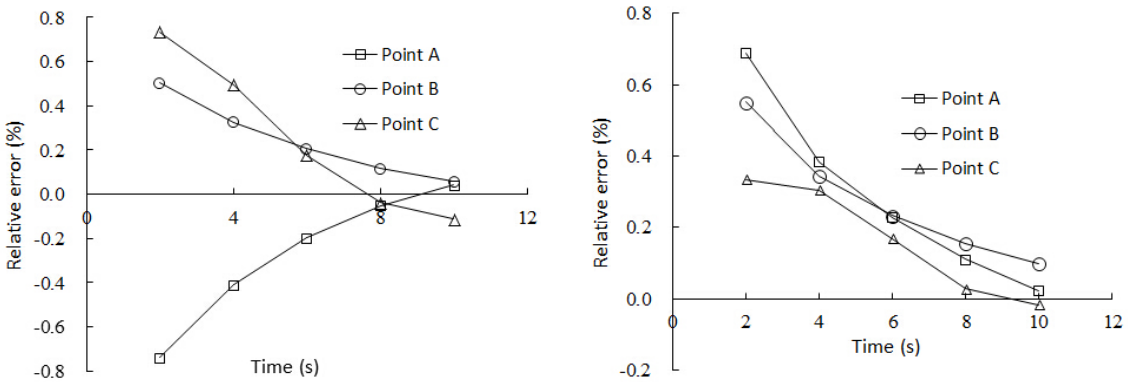
To examine the time evolution of temperature, three points *A* (1.7678, 1.7678, 3), *B* (1.4142, 1.4142, 4) and *C* (1.4142, 1.4142, 2), are investigated. Table 3 shows the comparison of the temperature results at each time step between BEM and FEM method. Relative errors are also calculated, taking the ABAQUS results as standard values. From Table 3 we can see that the relative errors converge to zero with time evolution, indicating that the presented method is

stable with time.

Table 3. Computed temperatures at points *A*, *B* and *C* with Δt = 0.1s

	A			*B*			*C*		
t(s)	BEM	Abaqus	Error(%)	BEM	Abaqus	Error(%)	BEM	Abaqus	Error(%)
1	125.116	123.891	0.989	292.614	290.528	0.718	35.345	35.523	-0.500
2	262.106	260.178	0.741	421.561	419.452	0.503	121.269	120.420	0.705
3	350.276	348.363	0.549	493.677	491.700	0.402	194.533	193.266	0.656
4	408.304	406.638	0.410	539.235	537.480	0.326	247.504	246.302	0.488
5	447.202	445.886	0.295	569.295	567.807	0.262	284.317	283.406	0.321
6	473.492	472.552	0.199	589.482	588.271	0.206	309.561	309.012	0.178
7	491.323	490.739	0.119	603.138	602.187	0.158	326.784	326.588	0.060
8	503.434	503.161	0.054	612.403	611.681	0.118	338.511	338.627	-0.034
9	511.665	511.653	0.002	618.696	618.167	0.086	346.688	346.865	-0.051
10	517.470	517.458	0.002	622.974	622.601	0.060	352.483	352.500	-0.005

To examine the influence of the length of each time step Δt on the computed results, temperatures at points *A*, *B* and *C* are also computed by using different values of Δt. Fig. 16 (*a*) shows the change of relative errors using the time step Δt = 2s. In Fig.16 (*a*), both SIEBEM and ABAQUS results are calculated on Δt = 2s, and the ABAQUS results are utilized as the standard values. Meantime, Fig. 16 (*b*) shows the change of relative errors using Δt = 0.04s, equally the ABAQUS results on Δt = 0.04s are also given as the standard values. By comparing Figs. 16 (*a*) and 16 (*b*) we can see that, even Δt = 2s is 50 times the length of Δt = 0.04s, the results calculated by SIEBEM coincide well with ABAQUS results, and their relative errors converge to zero, indicating that the presented method is stable and highly precise.



a ***b***

Figure 16. Relative errors of temperature along with time: (*a*) Δt = 2s; (*b*) Δt = 0.04s

6. Conclusions

In this paper, based on a newly derived interface integral equation, a new and simple BEM characterized as interface integral equation method is developed for solving transient heat conduction in multi-medium materials with variable material properties. To solve the time-dependent system of differential equations, firstly the unknown heat fluxes are eliminated from the system of differential equations, then based on an implicit backward differentiation scheme, an unconditionally stable and non-oscillatory time marching solution scheme is developed for solving the normal time-dependent system of equations.

Acknowledgement

The authors gratefully acknowledge the National Natural Science Foundation of China (11172055, 11202045).

References

[1] Wrobel, L. C. and Brebbia, C. A. (1992) *Boundary Element Methods in Heat Transfer*, Computational Mechanics Publications, Boston.
[2] Divo, E. and Kassab, A. J. (2003) *Boundary Element Method For Heat Conduction: With Applications in Non-Homogeneous Media*, WIT Press, Southampton, UK.
[3] Gao, X. W., Peng, H. F., Yang, K. and Wang, J. (2015) *Higher Boundary Element Method - Theory and Programming* (Chinese edition), Science Press, Beijing, 2015. ISBN: 978-7-03-042689-5
[4] Liu, Y. J., Mukherjee, S., Nishimura, N., Schanz, M., Ye, W., Sutradhar, A., Pan, E., Dumont, N. A., Frangi, A. and Saez, A. (2011) Recent advances and emerging applications of the boundary element method, *ASME Applied Mechanics Reviews* **64**, 617-636.
[5] Rizzo, F. J. and Shippy, D. J. (1970) A method of solution for certain problems of transient heat conduction, *AIAA Journal* **8**, 2004-2009.
[6] Sutradhar, A., Paulino, G. H. and Gray, L. J. (2002) Transient heat conduction in homogeneous and non-homogeneous materials by the Laplace transform Galerkin boundary element method, *Engineering Analysis with Boundary Elements* **26**, 119-132.
[7] Sutradhar, A., and Paulino, G. H. (2004) The simple boundary element method for transient heat conduction in functionally graded materials, *Computer Methods in Applied Mechanics and Engineering* **193**, 4511-4539.
[8] Simoes, N., Tadeu, A., Antonio, J. and Mansur, W. (2012) Transient heat conduction under nonzero initial conditions: A solution using the boundary element method in the frequency domain, *Engineering Analysis with Boundary Elements* **36**, 562-567.
[9] Guo, S. P., Zhang, J. M., Li, G. Y. and Zhou, F. L. (2013) Three-dimensional transient heat conduction analysis by Laplace transformation and multiple reciprocity boundary face method, *Engineering Analysis with Boundary Elements* **37**, 15-22.
[10] Wrobel, L. C. and Brebbia, C. A. (1981) A formulation of the boundary element method for axisymmetric transient heat conduction, *International Journal of Heat and Mass Transfer* **24**, 843-850.
[11] Ochiai, Y., Sladek, V. and Sladek, J. (2006) Transient heat conduction analysis by triple-reciprocity boundary element method, *Engineering Analysis with Boundary Elements* **30**, 194-204.
[12] Tanaka, M., Matsumoto, T. and Takakuwa, S. (2006) Dual reciprocity BEM for time-stepping approach to the transient heat conduction problem in nonlinear materials, *Computer Methods in Applied Mechanics and Engineering* **195**, 4953-4961.
[13] Yang, K. and Gao, X. W. (2010) Radial integration BEM for transient heat conduction problems, *Engineering Analysis with Boundary Elements* **34**, 557-563.
[14] Al-Jawary, M. A., Ravnik, J., Wrobel, L. C. and Skerget, L. (2012) Boundary element formulations for the numerical solution of two-dimensional diffusion problems with variable coefficients, *Computers and Mathematics with Applications* **64**, 2695-2711.
[15] Yu, B., Yao, W. A. and Gao, Q. (2014) A precise integration boundary-element method for solving transient heat conduction problems with variable thermal conductivity, *Numerical Heat Transfer Part B: Fundamentals* **45**, 472-493.
[16] Nardinia, D. and Brebbia, C. A., A new approach to free vibration analysis using boundary elements, *Applied Mathematical Modelling* **7,** 157-162.
[17] Partridge, P. W., Brebbia, C. A. and Wrobel, L. C. (1992) *The Dual Reciprocity Boundary Element Method*, Computational Mechanics Publications, Boston.

[18] Nowak, A. J. and Brebbia, C. A. (1989) The multiple-reciprocity method – a new approach for transforming BEM domain integrals to the boundary, *Engineering Analysis with Boundary Elements* **6**, 164-167.
[19] Gao, X. W. (2002) The radial integration method for evaluation of domain integrals with boundary-only discretization, *Engineering Analysis with Boundary Elements* **26**, 905-916.
[20] Divo, E. and Kassab, A. J. (1997) A generalized boundary-element method for steady-state heat conduction in heterogeneous anisotropic media, *Numerical Heat Transfer Part B-Fundamentals* **32**, 37-61.
[21] Kassab, A. J. and Divo, E. (1996) A generalized boundary integral equation for isotropic heat conduction with spatially varying thermal conductivity, *Engineering Analysis with Boundary Elements* **18**, 273-286.
[22] Divo, E. and Kassab, A. J. (1998) Generalized boundary integral equation for transient heat conduction in heterogeneous media, *Journal of Thermophysics and Heat Transfer* **12**, 364-373.
[23] Sladek, J., Sladek, V., Krivacek, J. and Zhang, Ch. (2003) Local BIEM for transient heat conduction analysis in 3-D axisymmetric functionally graded solids, *Computational Mechanics* **32**, 169-176.
[24] Sladek, J., Sladek, V. and Zhang, Ch. (2003) Transient heat conduction analysis in functionally graded materials by the meshless local boundary integral equation method, *Computational Materials Science* **28**, 494-504.
[25] Gao, X. W., Guo, L. and Zhang, Ch. (2007) Three-step multi-domain BEM solver for nonhomogeneous material problems, *Engineering Analysis with Boundary Elements* **31**, 965-973.
[26] Dong, C. Y. and de Pater, C. J. (2000) A boundary-domain integral equation for a coated plane problem, *Mechanics Research Communications* **27**, 643-652.
[27] Wang, C. B., Chatterjee, J. and P.K. Banerjee (2007) An efficient implementation of BEM for two- and three-dimensional multi-region elastoplastic analyses, *Computer Methods in Applied Mechanics and Engineering* **196**, 829-842.
[28] Zhou, A., Hui, K. and Lai, Y. (2008) Simulating deformation of objects with multi-materials using boundary element method, *International Journal for Numerical Methods in Engineering* **74**, 1088-1108.
[29] Giannopoulos, G. I. and Anifantis, N. K. (2007) A BEM analysis for thermomechanical closure of interfacial cracks incorporating friction and thermal resistance, *Computer Methods in Applied Mechanics and Engineering* **196**, 1018-1029.
[30] Divo, E., Kassab, A. J. and Rodriguez, F. (2003) Parallel domain decomposition approach for large-scale three-dimensional boundary-element models in linear and nonlinear heat conduction, *Numerical Heat Transfer Part B-Fundamentals* **44**, 417-437.
[31] Erhart, K., Divo, E. and Kassab, A. J. (2006) A parallel domain decomposition boundary element method approach for the solution of large-scale transient heat conduction problems, *Engineering Analysis with Boundary Elements* **30**, 553-563.
[32] Peng, H. F., Bai, Y. G., Yang, K. and Gao, X. W. (2013) Three-step multi-domain BEM for solving transient multi-media heat conduction problems, *Engineering Analysis with Boundary Elements* **37**, 1545-1555.
[33] Tanaka, M., T. Matsumoto and Takakuwa, S. (2006) Dual reciprocity BEM for time-steeping approach to the transient heat conduction problem in nonlinear materials, *Computer Methods in Applied Mechanics and Engineering* **195**, 4953-4961.
[34] Gao, X. W. and Wang, J. (2009) Interface integral BEM for solving multi-medium heat conduction problems, *Engineering analysis with boundary elements* **33**, 539-546.
[35] Gao, X. W. and Yang, K. (2009) Interface integral BEM for solving multi-medium elasticity problems, *Computer Methods in Applied Mechanics and Engineering* **198**, 1429-1436.
[36] Yang, K., Feng, W. Z. and Gao, X. W. (2015) A new approach for computing hyper-singular interface stresses in IIBEM for solving multi-medium elasticity problems, *Computer Methods in Applied Mechanics and Engineering* **287**, 54-68.
[37] Feng, W. Z., Gao, X. W., Liu, J. and Yang, K. (2015) A new BEM for solving 2D and 3D elastoplastic problems without initial stresses/strains, *Engineering Analysis with Boundary Elements* **61**, 134-144.
[38] Gao, X. W. (2006) A meshless BEM for isotropic heat conduction problems with heat generation and spatially varying conductivity, *International Journal for Numerical Methods in Engineering* **66**, 1411-1431.
[39] Gao, X. W. and Davies, T. G. (2002) *Boundary Element Programming in Mechanics*, Cambridge University Press, Cambridge, UK.
[40] Cui, M., Duan, W. W. and Gao, X. W. (2015) A new inverse analysis method based on a relaxation factor optimization technique for solving transient nonlinear inverse heat conduction problems, *International Journal of Heat and Mass Transfer* **90,** 491-498.
[41] Wang, H. T. and Yao, Z. H. (2013) Large-scale thermal analysis of fiber composites using a line-inclusion model by the fast boundary element method, *Engineering Analysis with Boundary Elements* **37,** 319-326.
[42] Zienkiewicz, O. C. and Taylor, K. L. (2000) *The Finite Element Method*, 5th edn . Butterworth Heinemann.
[43] ABAQUS Version 6.13 (2013) *Abaqus Analysis User's Guide*, Dassault System Simulia Corp., Providence, RI, USA.

Modelling of Hydrogen Assisted Stress Corrosion Crack Extension along Centerline of Austenitic Stainless Steel Welds

***Ishwar Londhe[1], † S. K. Maiti[2]**

[1,2] Department of Mechanical Engineering, IIT Bombay, Powai, Mumbai - 400076, India

†Presenting author:skmaiti@iitb.ac.in

*Corresponding author: ishwar.londhe1571@gmail.com

Abstract

This paper presents a numerical scheme for modelling hydrogen assisted stress corrosion cracking (HASCC) along centerline of gas tungsten arc (GTA) welds of austenitic stainless steel 21Cr-6Ni-9Mn (21-6-9). FEM based cohesive zone modelling (CZM) is used to examine the crack extension through the weld fusion zone (FZ). Diffusion of hydrogen through the lattice is analyzed by finite difference method incorporating effects of hydrostatic stress σ_h. J versus crack extension curves are obtained. Results are presented by considering both constant diffusivity and its variation with hydrogen concentration. The results based on the later case compare well with published experimental data. Temporal variations of hydrogen concentration and σ_h along the crack line ahead of the tip at various stages of crack extension are included.

Keywords: Hydrogen assisted stress corrosion cracking, cohesive zone modelling, J - Δa variation,FZ, HAZ, fracture initiation toughness.

1. Introduction

Austenitic stainless steels consists of 16-26% Cr, 8-24% Ni + Mn, up to 0.40% C and small amounts of a few other elements such as Mo, Ti, Nb and Ta. The steel contains about 90-100% of austenitic microstructure which is made possible by adjusting the amount of Cr and Ni + Mn. These alloys provide good strength and high toughness over a wide temperature range and oxidation resistance to little over 1000°F. Due to such excellent properties, they are mostly employed in machines, pipelines and structures subjected to hydrogen and other corrosive environments. During welding of such steels Cr content in base metal is generally kept high in filler wire as Cr is ferrite stabilizer whereas Ni is austenite stabilizer. After welding of austenitic steel and during solidification, melting of certain low melting point constituents like sulfur, phosphorous, manganese and silicon cause shrinkage induced strain. δ-ferrite has capacity to dissolve such harmful elements. Hence, residual amount of stable δ-ferrite is always preferred in steel microstructure to prevent hot cracking [1]. But, it is also reported that δ-ferrites are the dominating sites for microcrack formation and its propagation under load [2]. Weld joints of the austenitic stainless steels therefore becomes weaker against HASCC due to retained δ-ferrite. A common source of hydrogen during welding is the flux used which has ingredients containing chemically bonded water (H_2O) in their microstructure. This water dissociates as hydrogen and oxygen at high temperatures. Ingress of hydrogen is facilitated further by increase of hydrogen solubility in steel with increasing temperature. If the cooling is slow, some of the dissolved hydrogen may escape to the atmosphere; if the cooling is fast then there is no such possibility [3].

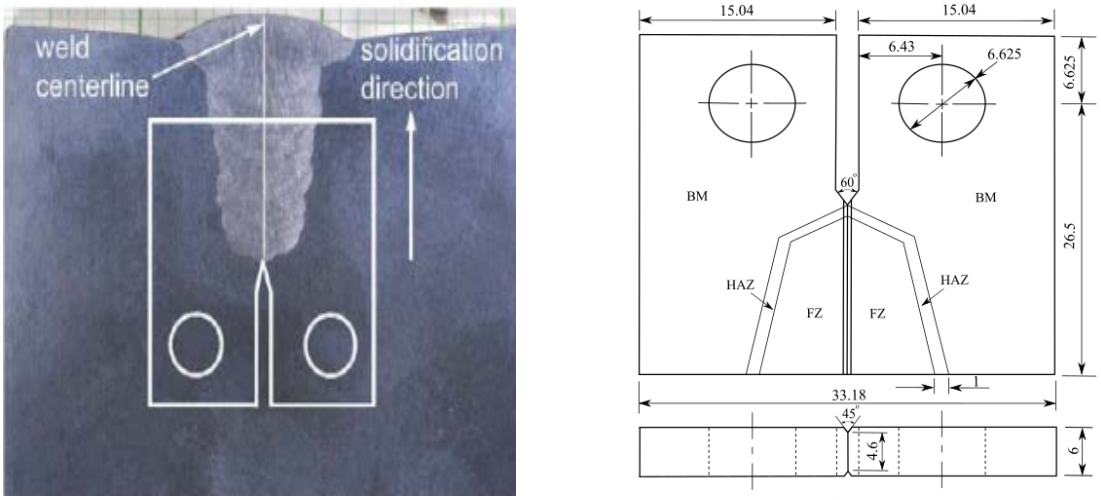
Several experiments [2][4][5] have shown that the dominant sites for initiation of micro-cracks are the ferrite and ferrite-austenite boundaries in the weld microstructure. The micro-cracks gradually develop into macro-cracks, which grow subsequently both along, and

perpendicular to, the initial crack line/plane. A 2-D analysis of such a crack growth only along the weld centerline, which is a FZ, is the objective of this study.

Analysis of the problem is difficult because of existence of three distinct material zones, i.e. FZ, HAZ and the base metal (BM). Fracture in such steels is a complex phenomenon involving ferrite, austenite-ferrite boundary, micro-crack formation, shear linkage between micro-cracks [2]. Both tensile and H_2 diffusion properties also differ from one zone to another. There is not much published data on the properties except some experimental results on variation of J with crack extension [2][4]. The study of the problem is further complicated by the fact that the corrosion affects the crack extension and the later, in turn, affects the diffusion and corrosion. The two phenomena are therefore coupled. The analysis of such a problem through homogenous material in the presence of HASCC has been reported earlier by several investigators. Both sequential [6][7] and coupled analysis have been reported [8]. A sequential analysis of a crack propagation along the weld centerline is considered in this paper. Due to non-availability of all required exhaustive material properties/data, e.g., tensile strength, % elongation, diffusivity parameters, reduction of cohesive strength with hydrogen concentration, etc., the appropriate data are iteratively adjusted to get the best predictions for J vs. Δa variations. In the modelling, variation of yield strength across the HAZ has been interpolated linearly from σ_Y = 485 MPa at the interface of HAZ and BM to σ_Y = 675 MPa at the interface of FZ and HAZ. The case studies presented here concerns internal hydrogen assisted corrosion (IHAC) in CT specimen with crack along the centerline of the weld. The complex failure mechanism is modelled using a hydrogen concentration dependent cohesive zone modelling technique (HCD-CZM).

2. Experimental Details

The experimental results of Somerday et al. [4] provide the basis for the present analysis. Similar studies were also carried by Jackson et al. [5] and Nibur et al. [9] for 304L/308L and 21Cr-6Ni-9Mn/308L austenitic stainless steel welds respectively. The base metal for the present analysis is 21Cr-6Ni-9Mn (21-6-9) steel, which was available in the form of rectangular bar stock of size 75×75 mm.



(a) (b)

Figure 1. a) Macrograph of 21-6-9/21-6-9 GTA weld [4] and (b) CT specimen considered for modelling the weld

The details of preparation of specimen and testing is given in [4]. It suffices to state here that, in order to prepare the weld, a tapered "U" groove was made at the centre of a rectangular bar stock (Fig. 1(a)). This groove was filled with 21-6-9 filler wire by GTA welding operation. A standard CT specimen was then cut out of the bar stock. The machined specimen was provided with a 45^0 side groove. Based on the overall dimensions provided in [4], dimensions of a typical specimen are: width (W) = 26.5 mm, nominal thickness (B) = 6 mm, reduced thickness near the weld (B_C) = 4.6 mm and pre-crack length to specimen width ratio (a/W) = 0.50.

Before the actual testing pre-cracking was appropriately done ahead of the machined notch. Specimens were then kept in hydrogen bath for charging for 29 days to reach a uniform hydrogen concentration of 230 ppm (by weight) and tested at loading rates of 0.4 and 0.04 mm/min [4]. The J integral (J) vs crack extension (Δa) curves are reproduced in Fig. 2. These clearly shows that hydrogen reduces the fracture initiation toughness as well as the slope of crack growth resistance curve significantly. Fracture initiation toughness (J_Q) dropped by more than 53 % (Fig. 2) for the specified pre-charging in comparison with charge-free specimens [4].

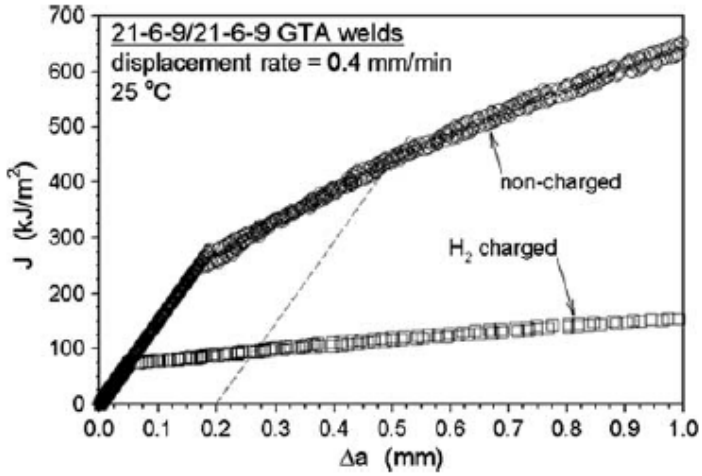


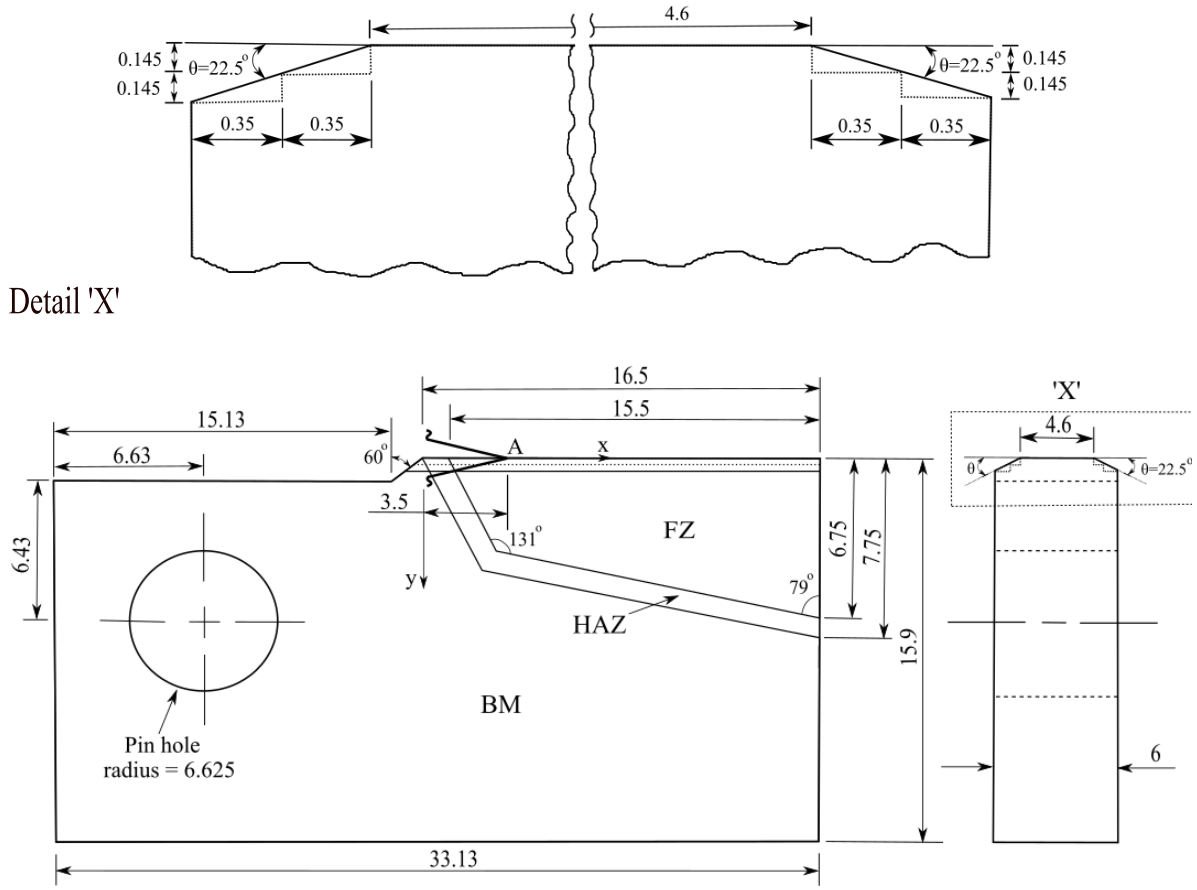
Figure 2. J vs. Δa curves for 21-6-9/21-6-9 GTA weld [4]

3. Finite Element (FE) Model

3.1 CT specimen geometry and model

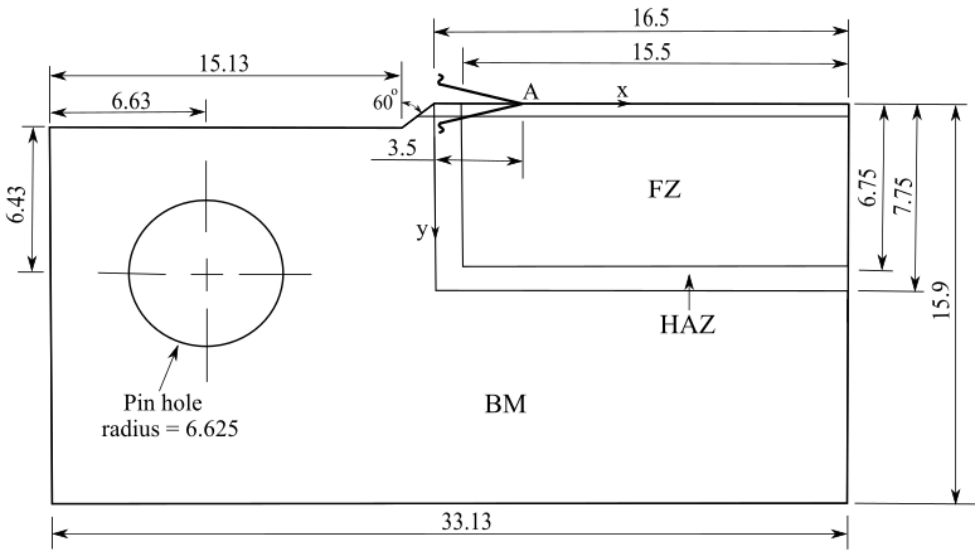
Only one half of the specimen is considered for analysis. Figs. 3, 4 and 5 shows the details of three geometries of FZ, HAZ and BM considered for the simulation. Fig. 3 considers a quadrilateral FZ where as Figs. 4 and 5 consider respectively a rectangular and triangular fusion zone. The dimensions of the three zones were approximated from the photograph of the specimen (Fig. 1(a)) using a plot digitizer software. The fusion and heat affected zones exhibit very different mechanical properties than that of the base metal. For example, the fusion zone exhibits a typical cast structure while heat affected zone exhibits a heat-treated structure involving phase transformation, recrystallization and grain growth. The BM and FZ have yields strength (σ_Y) of 485 MPa and 675 MPa respectively [12]. Young's modulus (E) of the two materials is 196.6 GPa and the Poisson's ratio is 0.3 [4]. Over the HAZ, as mentioned

earlier, the material properties are assumed to vary linearly from σ_Y = 485 MPa at the interface of HAZ and BM to σ_Y = 675 MPa at the interface of FZ and HAZ.



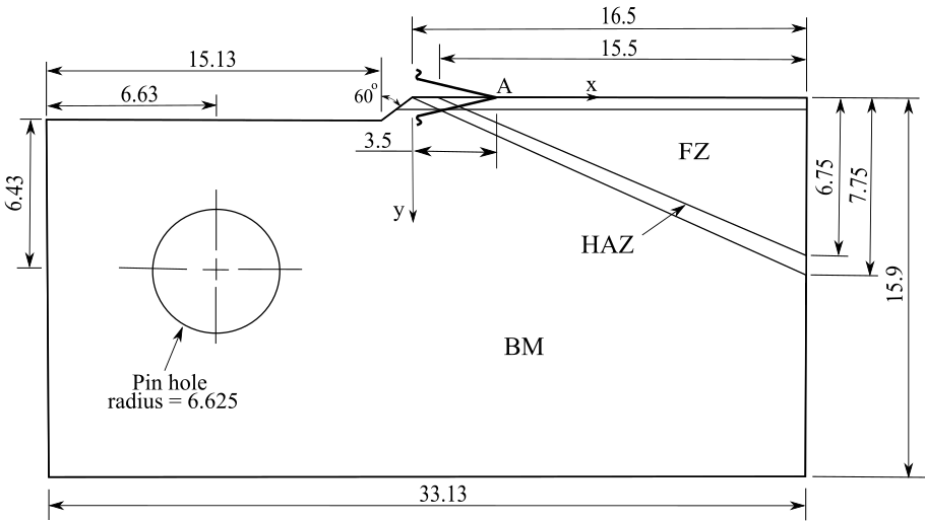
All dimensions are in mm

Figure 3. Quadrilateral FZ dimensions



All dimensions are in mm

Figure 4. Rectangular FZ dimensions



All dimensions are in mm

Figure 5. Triangular FZ dimensions

3.2 TSL parameters (T_0, δ_0)

In the present analysis the TSL employed is similar to the one used by Raykar et al. [13] and Scheider et al. [14]. This type of TSL introduces flexibility as the TSL shape can be varied easily by changing parameters δ_1 and δ_2 (Fig. 6). The two important TSL parameters, traction (T_0) and critical separation displacement (δ_0), were settled by analyzing the case of crack propagation through the quadrilateral FZ without any hydrogen charging and comparing the predicted J *vs.* Δa variations with the corresponding experimental data (Fig. 2). The analysis under IHAC condition was done by replacing T_0 by $T_0(1-\mu C)$. Actual traction separation variation for uncharged and charged cases are shown schematically in Fig. 6. The exact form of TSL is given below.

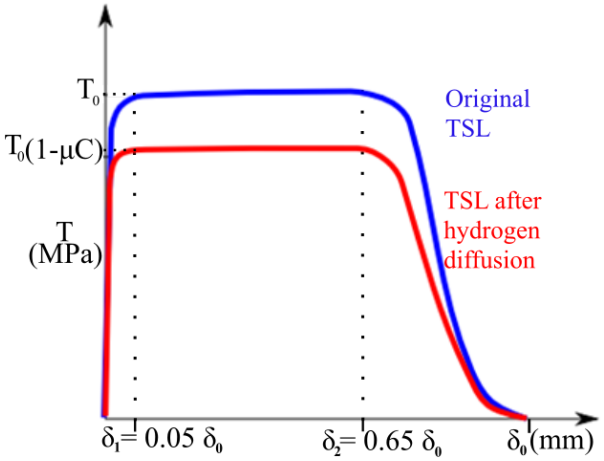


Figure 6. Traction separation law for cohesive element

$$T = T_0(1-\mu C)\begin{cases} 2\left(\dfrac{\delta}{\delta_0}\right)-\left(\dfrac{\delta}{\delta_0}\right)^2, & \delta < \delta_1 \\ 1, & \delta_1 < \delta \le \delta_2 \\ 2\left(\dfrac{\delta-\delta_0}{\delta_0-\delta_2}\right)^3 - 3\left(\dfrac{\delta-\delta_0}{\delta_0-\delta_2}\right)^2, & \delta_2 < \delta \le \delta_0 \end{cases} \quad (1)$$

As can be seen from this equation, the term μC affects the cohesive strength (T). The reduction factor μ in the cohesive strength is considered constant for a particular loading rate. It is settled by comparing the predicted J vs. Δavariations with the corresponding experimental values for hydrogen charged specimens. μ=0 corresponds to testing under charge-free conditions. The hydrogen concentration (C) in Eqn. (1) is not constant for a given loading rate; it varies with time at any node of a cohesive element.

3.3 Mesh size determination

The specimen was discretized in such a way that top layer (Fig. 7) consists of cohesive elements of zero thickness. These elements are placed along the crack propagation direction as shown. Just below these elements, there are few layers of refined continuum elements; rest of the specimen have comparatively coarser mesh. The mesh size near the crack tip was fine enough to capture the stress distribution accurately around the crack tip.

The side groove was accommodated by considering a 3-step variation of thickness (Fig. 3) of normal elements immediately below the cohesive element. The depth of the top two layers is 0.145 mm each and their widths are 4.6 mm and 5.3 mm respectively. The cohesive element width is therefore 4.6 mm. The size of continuum and cohesive elements along the crack propagation direction were arrived at by trial and error by comparison of predicted and experimental J vs. Δadiagrams for charge free specimens for loading rate of 0.4 mm/min (Fig. 7).

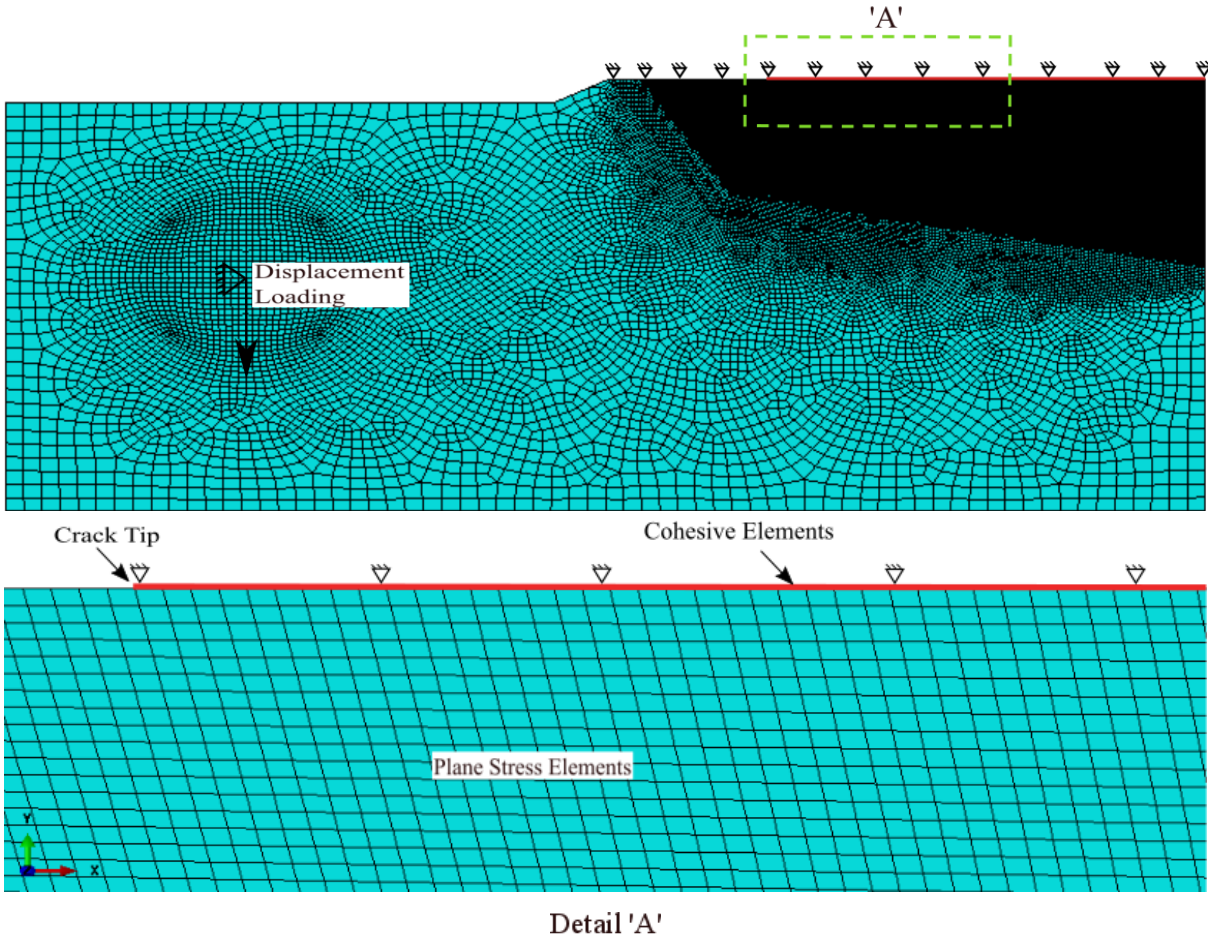


Figure 7. Mesh discretization of CT specimen

Case studies were performed by considering various combination of sizes for cohesive and continuum elements (Figs. 8, 9). Continuum elements are 4 noded quadrilateral plane stress elements (CPS4 of ABAQUS® software). Cohesive elements are 4 noded linear elements (COH2D4 of ABAQUS® software). From Figs. 8 and 9, it is observed that the optimum size of continuum and cohesive elements are 0.1 mm and 0.02 mm respectively.

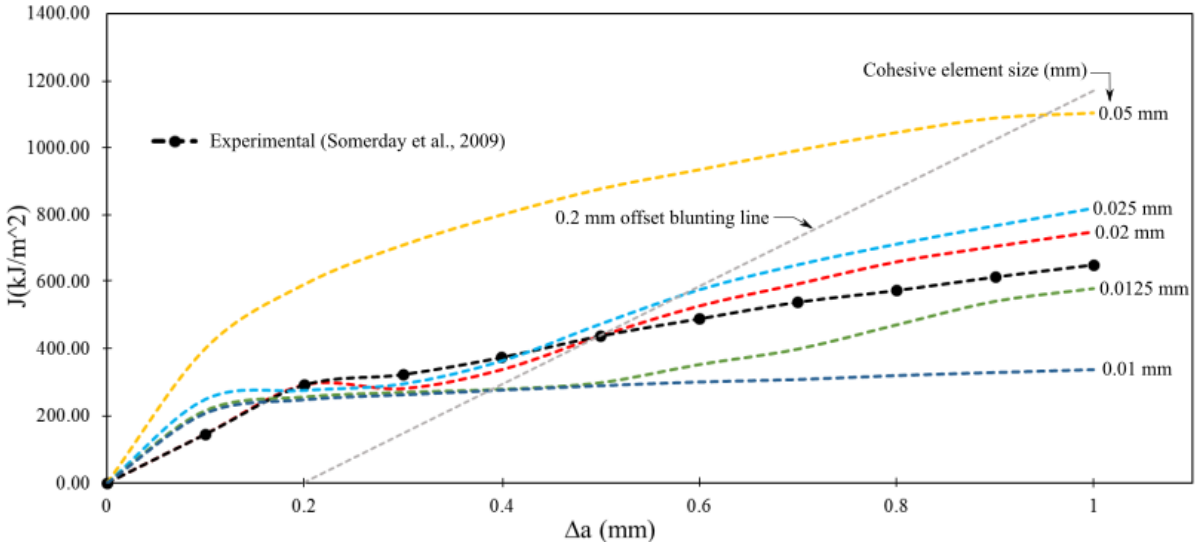


Figure 8. *J vs.* Δ*a* variation for continuum element size=0.1 mm and variable cohesive element sizes

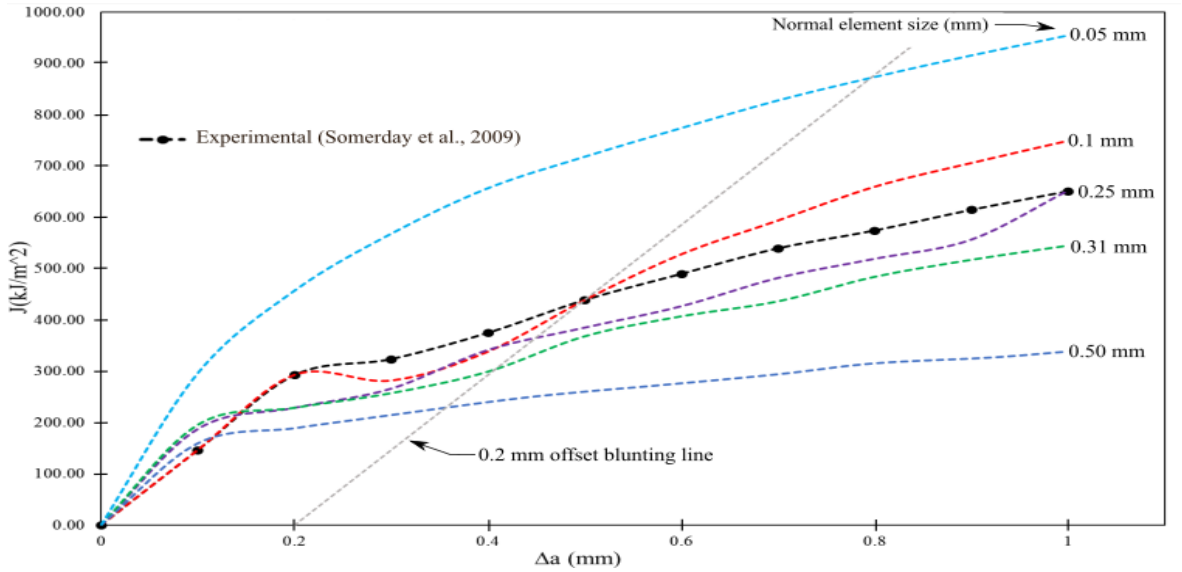


Figure 9. *J vs.* Δ*a* variation for cohesive element size=0.02 mm and variable continuum element size

4. Analysis of hydrogen charged specimens

When a charged specimen is tested, hydrogen concentration keeps changing near the crack tip because there is mobility of hydrogen atoms towards the location of high stress concentration. The equation governing this movement was first given by Sofronis and McMeeking [15] in the form given below.

$$\frac{\partial C_L}{\partial t} = D_{eff}\frac{\partial^2 C_L}{\partial x^2} - D_{eff}\frac{V_H}{RT}\frac{\partial C_L}{\partial x}\frac{\partial \sigma_h}{\partial x} - D_{eff}\frac{V_H C_L}{RT}\frac{\partial^2 \sigma_h}{\partial x^2} = 0, \qquad (2)$$

where x is the distance from the crack tip along the crack path, σ_h is hydrostatic stress. R, T and D_{eff} are the universal gas constant, absolute temperature, effective diffusivity of hydrogen respectively. V_H (= 2×10^3 mm^3/mol) is partial molar volume of hydrogen in the metal at temperature (T) of 300K. C_L and C_T are the number of hydrogen atoms per unit volume present in the lattice and trap sites respectively.

$C_L = \theta_L \beta N_L$. β denotes number of normal interstitial lattice sites (NILS) per solvent atom, θ_L denotes the fraction of the NILS occupied by lattice hydrogen atom and N_L is the number of solvent lattice atoms per unit volume. Parameters β and N_L are constant for a given lattice; β is taken as 1. Similarly C_T can be expressed as $\theta_T \alpha N_T$. α signifies number of hydrogen atoms per trap, θ_T is trap occupancy and N_T is number of traps per unit lattice volume [4]. α is taken as 10. C_T is related to C_L by Oriani's law [16] as follows.

$$C_T = \frac{K_{tr}\left(\frac{\alpha.N_T}{\beta.N_L}\right)C_L}{1+\left(\frac{K_{tr}}{\beta.N_L}\right)C_L}, \qquad (3)$$

where $K_{tr} = e^{\frac{E_B}{RT}}$. K_{tr} is trap equilibrium constant which depends upon the trap binding energy (E_B) for hydrogen atoms and absolute temperature (T). When E_B is small, the trap is called as reversible trap; this type of trap sites releases hydrogen and causes more damage. E_B = 60 kJ is considered as the upper limit for binding energy for reversible trap sites. E_B and T for 21-6-9 austenitic stainless steel are given as 9.65 kJ/mol and 298K respectively [3].

N_T is dependent upon plastic strain (ε_p). The relation between N_T and ε_p is given by McMeeking [15] and Krom et al. [17] as follows.

$$\log N_T = 23.26 - 2.33e^{-5.5\varepsilon_p}, \qquad (4)$$

$$N_L = \frac{N_A}{V_M}, \qquad (5)$$

where N_A is the Avogadro's number (6.023×10^{23}) and V_M is the molar volume of the host lattice (7.116×10^{-6} m^3/mol). The effective diffusivity (D_{eff}) is given by Sofronis and McMeeking [15].

$$D_{eff} = D_L \frac{C_L}{C_L + C_T(1-\theta_T)}, \tag{6}$$

where D_L is the lattice diffusivity of hydrogen atoms and $\theta_T = \frac{C_{TO}}{\alpha N_T}$. In the present analysis total hydrogen concentration is normalized as follows.

$$C = \frac{C_L + C_T}{C_{LO} + C_{TO}}, \tag{7}$$

where C_{L0} and C_{T0} are stress free equilibrium solubility of hydrogen in iron at 300K. Value of C_{L0} is taken as 2.0845×10^{21} atoms/m^3 [15]. C_{T0} is obtained through Eqn. (3) as 2.203×10^{17} atoms/m^3. Further, the initial hydrogen concentration has been taken to be equal to specified concentration C_{L0} throughout the domain.

Based on the observation that, for a two dimensional problem of a homogenous material, 1-D analysis of diffusion along the crack line is quite sufficient [13][14], 1-D form of diffusion Eqn. (2) was solved by finite difference method. In the present study results are obtained by considering the diffusivity to be constant in one case and variable in the other.

4.1 Solution of 1-D diffusion considering constant diffusivity (D_{eff})

One dimensional form of hydrogen diffusion equation with the inclusion of hydrostatic stress σ_h is obtained from Eqn. (2) as follows [13].

$$\frac{\partial C_L}{\partial t} = D_{eff}\frac{\partial^2 C_L}{\partial x^2} - E_H\frac{\partial C_L}{\partial x}\frac{\partial \sigma_x}{\partial x} - E_H C_L\frac{\partial^2 \sigma_h}{\partial x^2}, \tag{8}$$

where x is measured from crack tip and $E_H = \frac{D_{eff}V_H}{RT}$.

This equation was solved numerically using Crank-Nicholson scheme (central difference method) along the line of Raykar et al. [13]. Let $(C_L)_j^n$ be the magnitude of C_L at time step n; j = 1, 2, 3… are the grid points; Δt is the time interval between $(n+1)^{\text{th}}$ and n^{th} step. Δx is equal to width (0.02 mm) of cohesive elements. Total number of grid points considered is 131, i.e., maximum value of *j*=131. The lattice concentration of hydrogen at a given location $\left((C_L)_j^n\right)$ and at the two time steps n and $n+1$ are related.

$$\begin{aligned}\frac{(C_L)_j^{n+1} - (C_L)_j^n}{\Delta t} = {} & D_{eff}\left\{\frac{\left[(C_L)_{j-1}^{n+1} - 2(C_L)_j^{n+1} + (C_L)_{j+1}^{n+1}\right] + \left[(C_L)_{j-1}^{n} - 2(C_L)_j^{n} + (C_L)_{j+1}^{n}\right]}{2\Delta x^2}\right\} \\ & - E_H\left\{\frac{\left[(C_L)_{j+1}^{n+1} - (C_L)_{j-1}^{n+1}\right] + \left[(C_L)_{j+1}^{n} - (C_L)_{j-1}^{n}\right]}{4\Delta x}\right\}\left\{\frac{(\sigma_h)_{j+1}^{n} - (\sigma_h)_{j-1}^{n}}{2\Delta x}\right\} \\ & - E_H\left\{\frac{(C_L)_j^{n+1} + (C_L)_j^{n}}{2}\right\}\left\{\frac{(\sigma_h)_{j+1}^{n} - 2(\sigma_h)_j^{n} + (\sigma_h)_{j-1}^{n}}{(\Delta x)^2}\right\}\end{aligned} \tag{9}$$

This relation is obtained through the Crank-Nicholson scheme.

With the following substitutions,

$$\alpha_1 = \frac{D_{eff}\Delta t}{2\Delta x^2} \tag{10}$$

$$\beta_1 = \frac{E_H \Delta t}{8\Delta x^2}((\sigma_h)^n_{j+1} - (\sigma_h)^n_{j-1}) \tag{11}$$

$$\gamma_1 = \frac{E_H \Delta t}{2\Delta x^2}((\sigma_h)^n_{j-1} - 2(\sigma_h)^n_j + (\sigma_h)^n_{j+1}) \tag{12}$$

the following simplified form is obtained from Eqn. (9).

$$\begin{aligned}(-\alpha_1 - \beta_1)(C_L)^{n+1}_{j+1} + (1 + 2\alpha_1 - \gamma_1)(C_L)^{n+1}_j + (-\alpha_1 + \beta_1)(C_L)^{n+1}_{j-1} \\ = (\alpha_1 + \beta_1)(C_L)^n_{j+1} + (1 - 2\alpha_1 + \gamma_1)(C_L)^n_j + (\alpha_1 - \beta_1)(C_L)^n_{j-1}\end{aligned} \tag{13}$$

Eqn. (13) can be applied at a particular time step at all the grid points to get their hydrogen concentration at the time step (n+1). This process is repeated as many times as required in a case study. These hydrogen concentrations were utilized to amend the reduction in strength of material due to variations in hydrogen concentration and hydrostatic stress. This was adopted in the crack propagation analysis through ABAQUS® subroutine USDFLD.

4.2 Solution of 1-D diffusion equation considering variable diffusivity (D_{eff})

In this case the term D_{eff} is not constant but varies with change in level of C_L and C_T with time. This effect is introduced in the model by substituting D_{eff} in Eqn. (9) as follows.

$$D_{eff} = \frac{D_L.(C_L)^{n+1}_j}{(C_L)^{n+1}_j + \dfrac{\dfrac{K_{tr}\alpha_{tr}N_T}{\beta N_L}.(C_L)^{n+1}_j}{1 + \dfrac{K_{tr}}{\beta N_T}.(C_L)^{n+1}_j}.(1-\theta_T)} \tag{14}$$

After substitution of Eqn. (14), Eqn. (9) gives rise to following non-linear relation in $(C_L)^{n+1}_j$. This was solved by Newton-Rhapson method following Kaiser et al. [18].

$$\frac{(C_L)_j^{n+1}-(C_L)_j^n}{\Delta t}=\left\{\frac{D_L.(C_L)_j^{n+1}}{(C_L)_j^{n+1}+\dfrac{\dfrac{K_{tr}\alpha_{tr}N_T}{\beta N_L}.(C_L)_j^{n+1}}{1+\dfrac{K_{tr}}{\beta N_T}.(C_L)_j^{n+1}}.(1-\theta_T)}\right\}.$$
$$\left\{\begin{array}{l}\left\{\dfrac{\left[(C_L)_{j-1}^{n+1}-2(C_L)_j^{n+1}+(C_L)_{j+1}^{n+1}\right]+\left[(C_L)_{j-1}^{n}-2(C_L)_j^{n}+(C_L)_{j+1}^{n}\right]}{2\Delta x^2}\right\}\\ -\dfrac{V_H}{RT}\left\{\dfrac{\left[(C_L)_{j+1}^{n+1}-(C_L)_{j-1}^{n+1}\right]+\left[(C_L)_{j+1}^{n}-(C_L)_{j-1}^{n}\right]}{4\Delta x}\right\}\left\{\dfrac{(\sigma_h)_{j+1}^{n}-(\sigma_h)_{j-1}^{n}}{2\Delta x}\right\}\\ -\dfrac{V_H}{RT}\left\{\dfrac{(C_L)_j^{n+1}+(C_L)_j^{n}}{2}\right\}\left\{\dfrac{(\sigma_h)_{j+1}^{n}-2(\sigma_h)_j^{n}+(\sigma_h)_{j-1}^{n}}{(\Delta x)^2}\right\}\end{array}\right\} \quad (15)$$

4.3 Boundary and initial conditions for diffusion analysis

For the case of internal hydrogen assisted cracking (IHAC), it is assumed that the hydrogen is not allowed to diffuse out of the material (i.e. hydrogen flux is zero at crack tip and end of ligament). As the temperature of the specimen increases, the tendency of hydrogen to diffuse out of the specimen increases. To prevent hydrogen egress from the CT specimen, the temperature was maintained at 250 K after hydrogen pre-charging [4]. In the present model the boundary conditions employed are as follows. Both at the crack tip and ligament end, flux $\left(J_H\right)=0.$ That is,

$$D_L\nabla C_L-\frac{D_L C_L V_H}{RT}\nabla\sigma_h=0, \quad (16)$$

at both the locations.

For Eqn. (16) to be zero it is necessary that,at the crack tip and the end of ligament,

$$\left(\frac{\partial C_L}{\partial x}\right)=0, (17)$$

$$\left(\frac{\partial \sigma_h}{\partial x}\right)=0, (18)$$

These conditions are enforced by ensuring $C_L=C_{L0}$ at the first three (i.e., j = 1, 2, 3) and last three consecutive grid points (i.e., j = 129, 130, 131) (Fig. 10). The end conditions on hydrostatic stresses (Eqn. (18)) are similarly introduced. This small adjustment is implemented in ABAQUS® software through a user subroutine.

During the initial state, the specimen is fully charged with hydrogen, i.e., at $t = 0$, $C_L = C_{L0}$ throughout the crack plane. For incorporating the displacement boundary conditions the respective top and bottom nodes are constrained such that they have same horizontal displacements and can exhibit separation only in the vertical direction [19]. Displacement loading (at the rate 0.4 mm/min) is applied at the node representing the load point.

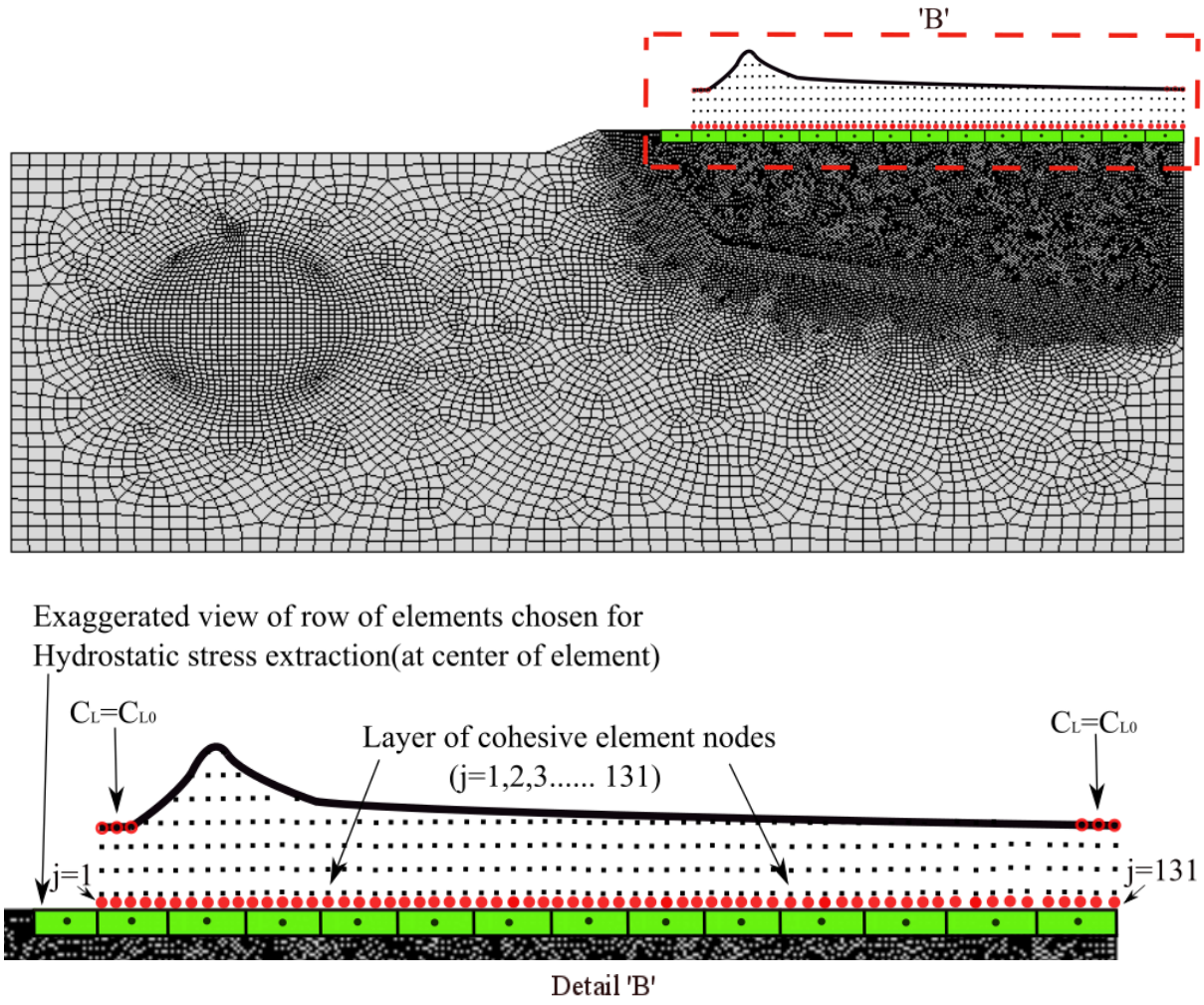


Figure 10. Diagrammatic representation of boundary conditions

5. Comparison of simulation and experimental results

As per ASTM E1820 J is given as follows.

$$J = J_{elastic} + J_{plastic}, \tag{19}$$

where $J_{elastic}$ is represented by following equation.

$$J_{elastic} = (1-\nu^2)\frac{K^2}{E}, \tag{20}$$

where E is Young's modulus and ν is the Poisson's ratio. K is the stress intensity factor which depends upon load (E) corresponding to a particular instant of crack extension (a), width of specimen (W), nominal thickness (B) and minimum thickness (B_c).

$$K = \frac{P}{\sqrt{BB_cW}} \frac{\left(2+\frac{a}{W}\right)}{\left(1-\frac{a}{W}\right)^{1.5}\left(0.886 + 4.64\left(\frac{a}{W}\right) - 13.32\left(\frac{a}{W}\right)^2 + 14.72\left(\frac{a}{W}\right)^3 - 5.6\left(\frac{a}{W}\right)^4\right)} \quad (21)$$

$J_{plastic}$ is given by following equation.

$$J_{plastic} = \frac{n_p \times U_p}{B_c \times (W - a)}, \quad (22)$$

where U_p is the total plastic energy calculated from load displacement plot, n_p is a factor given by Clarke and Landes [20]. It depends on specimen type and varies with crack size (a) and width of specimen (W) as follows.

$$n_p = 2 + 0.522\left(1 - \frac{a}{W}\right) \quad (23)$$

5.1 Charge - free specimen

Initially a number of iterations were performed considering various shapes of FZ. The analysis was done considering the loading rate of 0.4 mm/min. Fig. 11 shows comparison of experimental and predicted J vs. Δa variations for three types of FZ shapes, i.e. rectangular, triangular and quadrilateral. The comparison is done for crack extension up to 1mm. The cohesive parameters are obtained comparing the predicted J vs. Δavariations with the corresponding experimental data for quadrilateral FZ are as follows: T_0 = 1241 MPa and $\delta_0 / 2 = 0.0489$ mm. Similarly for rectangular fusion zone the cohesive parameters giving best results are: $T_0 = 1231$ MPa and $\delta_0 / 2 = 0.0489$ mm; for triangular fusion zone $T_0 = 1234$ MPa and $\delta_0 / 2 = 0.0489$ mm.

The fracture initiation toughness (J_Q) is defined by intersection of J resistance curve with 0.2 mm blunting line. For charge-free specimens and loading rate of 0.4 mm/min the experimental fracture initiation toughness is 439 kJ/m^2 [4].The corresponding predicted results by numerical modelling analysis are 438.31 kJ/m^2, 308.03 kJ/m^2 and 330.63 kJ/m^2 for quadrilateral, rectangular and triangular FZs. This indicates errors of -0.16 %, - 29.83 % and – 24.68 % with respect to J_Q respectively. The maximum difference in the predicted results in the case of quadrilateral FZ is just +14.97 % at crack extension of 1 mm (Fig. 11). In the case of rectangular and triangular FZs, the maximum differences are – 45.07 % and – 40.08 % at crack extension of 1 mm. On the whole, the quadrilateral FZ gives better results. It is selected for the analysis of charged case.

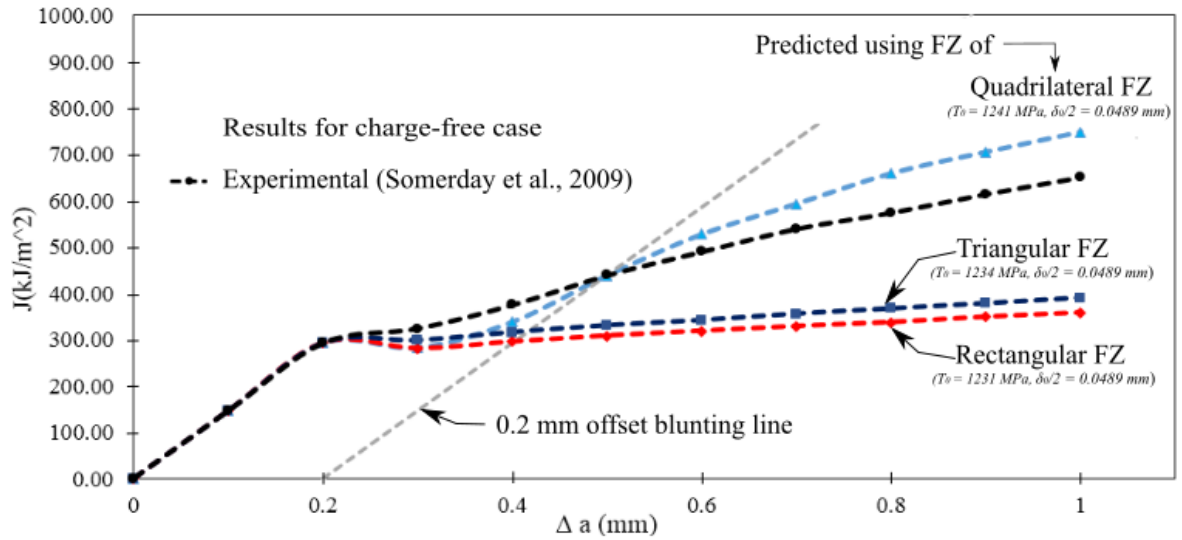


Figure 11. Comparison of experimental and simulation results for charge free specimens

5.2 Charged specimen

The specimens with internal hydrogen assisted cracking (IHAC) were analyzed considering both constant diffusivity and variable diffusivity.

5.2.1 Constant Diffusivity

As indicated earlier, during loading, hydrogen atoms move towards the crack tip, where there is high stress concentration. These movements were studied. Set of Eqns. (2-22) are considered in the analysis. The loading rate is 0.4 mm/min as before. Three D_{eff} values were considered. They were calculated from lattice diffusivity (D_L) using,

$$\varepsilon_p = 0,\ N_T = 1.82\times10^{22},$$

$$C_T = C_{T0} = 2.2\times10^{17},$$

$$C_L = C_{L0} = 2.08\times10^{21},$$

$$\theta_T = \frac{C_{T0}}{\alpha N_T} = \frac{2.20\times10^{17}}{10\times1.82\times10^{22}} = 2.20\times10^{-6},$$

$$D_{eff} = D_L \frac{C_L}{C_L + C_T(1-\theta_T)} = D_L \frac{2.08\times10^{21}}{2.08\times10^{21} + 2.2\times10^{17}(1-1.2\times10^{-6})},$$

$$= 0.9999\times D_L \approx D_L,$$

This indicates that D_{eff} is almost the same as D_L. Analysis has been done for three trial values for D_L, 1.2×10^{-2} mm^2/min, 1.2×10^{-3} mm^2/minand 1.2×10^{-5} mm^2/min[21].

μ is varied in the range 0.2 to 0.6. The best value, based on comparison of predicted and experimental J vs. $\Delta a$is obtained as 0.28. The comparison of predicted and experimental J vs. Δa for three D_{eff} is presented in Fig. 12. This shows that $D_{eff} = 1.2\times10^{-3}$ mm^2/min gives the best comparison with experimental data over the later stages of crack extension.

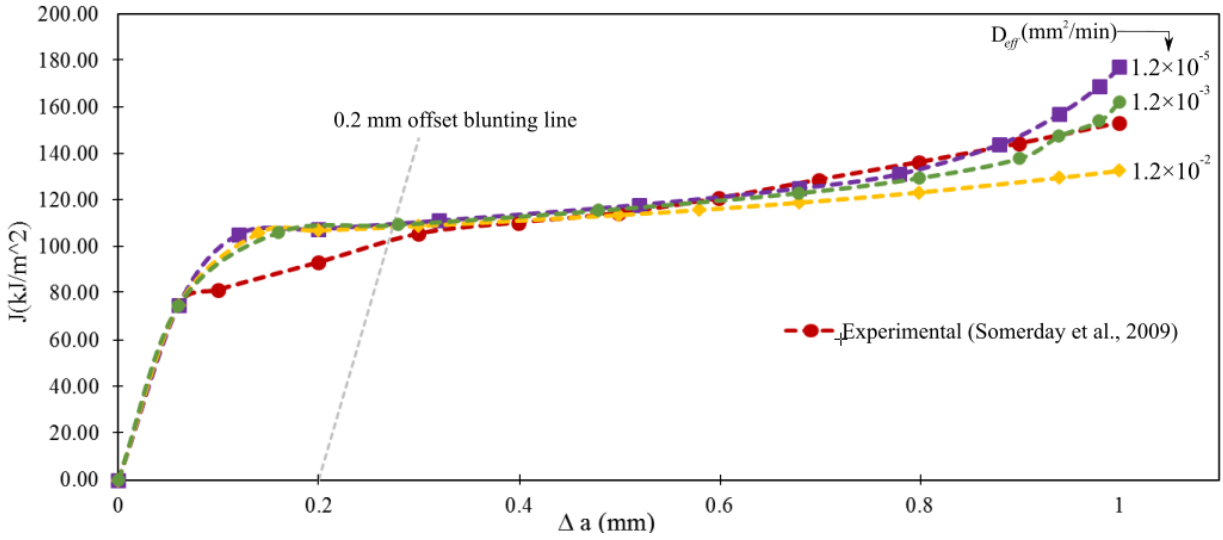


Fig. 12 Comparison of experimental and simulation results with different D_{eff} (mm^2/min)

5.2.2 Variable Diffusivity

In this case diffusivity is considered to be varying with the level of local hydrogen concentration. This method has the advantage that it eliminates the need for iterations to determine D_{eff} [18]. The iterations start with initial value of $D_{eff} = D_L$. Three D_L values are again considered, 1.2×10^{-2} mm^2/min, 1.2×10^{-3} mm^2/min, 1.2×10^{-5} mm^2/min. The finite difference formulation of the resulting diffusion equation leads to a set of non-linear simultaneous equations. These equations are solved with appropriate initial and boundary conditions and are linked to the crack propagation analysis through ABAQUS® (version 6.11) user subroutine USDFLD. Table 1 gives the comparison of J values calculated by constant and variable diffusivity respectively.

By a suitable adjustment of the material data associated with diffusion and crack extension, better correlation is obtained for $D_L = 1.2\times10^{-5}$ mm^2/min. A maximum difference of +19.85% in J is observed at $\Delta a = 0.16$ mm and the difference reduces to +1.46 % at the later stages (Table 1). The results further shows that, with the variable diffusivity, there is an overall improvement in comparison between experimental and predicted J vs. Δa variations for $D_L = 1.2\times10^{-5}$ mm^2/min.

For charged specimens the experimental fracture initiation toughness is 100.75 kJ/m^2. The simulation yielded 109.27 kJ/m^2 for constant diffusivity ($D_{eff} = 1.2\times10^{-3}$mm^2/min) and 109.21 kJ/m^2 for variable diffusivity ($D_L = 1.2\times10^{-5}$mm^2/min). The crack extension (Δa) corresponding to fracture initiation toughness was taken as 0.27 mm [4].

Table 1. Comparison of J at various stages of crack extension considering constant and variable diffusivity

Constant diffusivity (D_{eff} = 1.2×10^{-3} mm^2/min)				Variable diffusivity											
				(D_L = 1.2×10^{-5} mm^2/min)				(D_L = 1.2×10^{-2} mm^2/min)				(D_L = 1.2×10^{-3} mm^2/min)			
Δa mm	Simulation J (kJ/m^2)	Experimental J (kJ/m^2)	%Error	Δa mm	Simulation J (kJ/m^2)	Experimental J (kJ/m^2)	%Error	Δa mm	Simulation J (kJ/m^2)	Experimental J (kJ/m^2)	%Error	Δa mm	Simulation J (kJ/m^2)	Experimental J (kJ/m^2)	%Error
0.16	106.30	88.66	19.89	0.16	106.26	88.66	19.85	0.16	106.20	88.66	19.78	0.16	106.29	88.66	19.88
S0.28	109.54	103.21	6.14	0.30	110.01	103.21	6.59	0.22	107.33	95.83	12.00	0.3	110.09	103.21	6.67
0.48	115.65	113.69	1.73	0.48	115.35	113.69	1.46	0.30	108.71	103.21	5.33	0.38	112.46	109.50	0.35
0.68	123.22	127.27	-3.18	0.64	121.26	124.08	-2.27	0.40	110.69	110.46	0.21	0.54	117.82	117.06	0.66
0.80	129.65	136.34	-4.91	0.84	133.25	139.54	-4.51	0.62	115.97	130.45	-11.11	0.68	123.09	127.27	-3.29
0.90	138.05	144.35	-4.36	0.92	142.39	146.12	-2.55	0.78	121.03	134.34	-9.91	0.8	129.51	136.34	-5.01
0.94	147.67	147.90	-0.15	0.94	144.57	147.90	-2.25	0.90	125.65	144.35	-12.95	0.9	140.02	144.35	-3.00
1	162.39	153.23	5.98	1	161.09	153.23	5.13	1	131.88	153.23	-19.93	1	159.55	153.23	4.12

The distribution of hydrogen concentration and hydrostatic stress (σ_h) ahead of the crack tip over a span along the crack line is presented (Fig. 13) at different stages of crack extension. The concentration of hydrogen (Fig. 13) reaches the highest value at a small distance from the instantaneous crack tip. This is very similar to a case reported earlier for hydrogen environment assisted cracking for a homogenous material [7]. The hydrostatic stress has the highest value close to the point of maximum C_L. This is because near the point of maximum hydrostatic stress, the lattice opens up the highest and has the maximum room for accommodation of hydrogen atoms. Thereby the hydrogen concentration becomes the maximum at this location [22].

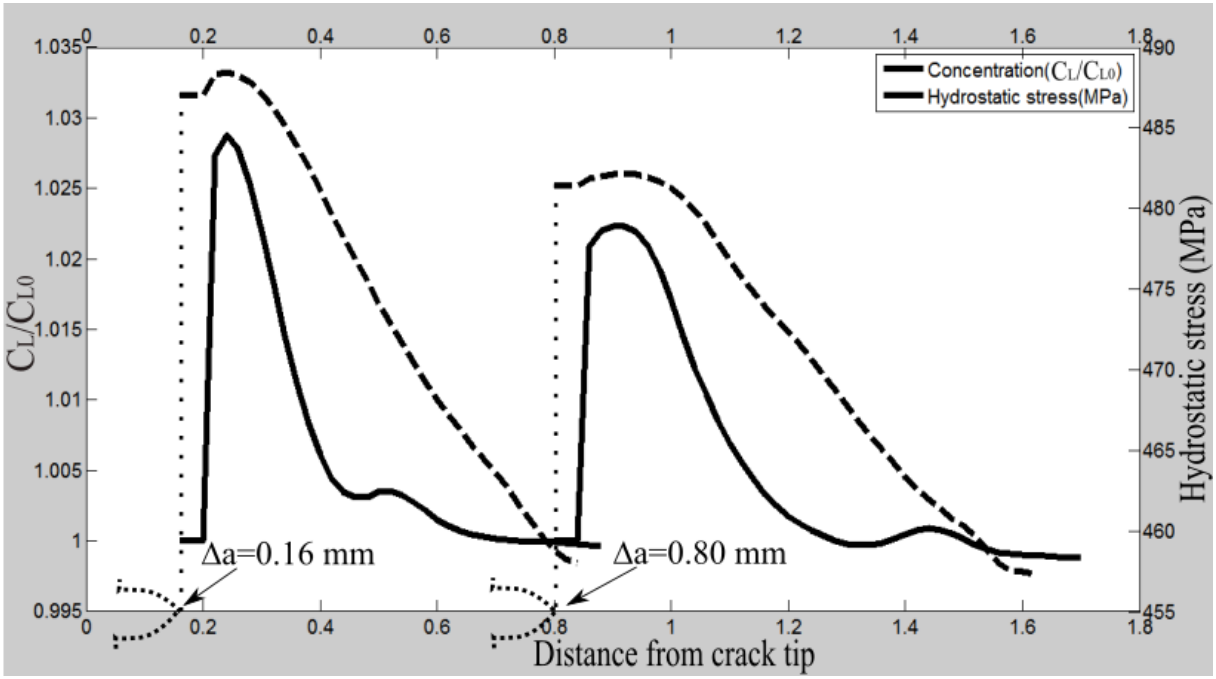


Figure 13. Variation of hydrogen concentrationand hydrostatic stress ahead of crack tip for constant D_{eff} (1.2×10^{-3} mm^2/min)

5.3 Results for loading rate 0.04 mm/min

Similar analysis was also carried out for loading rate of 0.04 mm/min. The quadrilateral fusion zone was again considered and iterations were performed to fix cohesive strength reduction factor (μ). μ is obtained as 0.36. The simulation was carried out by considering variable diffusivity D_L =1.2×10^{-5} mm^2/min. The analysis yielded fracture initiation toughness 78.92 kJ/m^2 compared with the experimental value 77.4 kJ/m^{2}at Δa = 0.25 mm.

6. Conclusions

In this study, an attempt has been made to examine the applicability of cohesive zone modelling to a heterogeneous specimen consisting of weld joint. The comparison (Table 1) indicates that a good prediction for *J vs.* Δa variation for the case of IHAC is possible with the help of CZM technique. The CZM parameters T_0 and $\delta_0$can be settled through combined numerical-experimental study. The same parameters can be employed for situations with

hydrogen charging. However, the cohesive strength reduction factor μ is required to be adjusted. This can be done through the combined numerical-experimental study.

For charged specimens, analysis has been carried out considering both constant and variable diffusivity. There is an overall reduction in error when the analysis is done considering the variable diffusivity. Further, there is better agreement between the experimental and predicted fracture initiation toughnesses for the two loading rates.

Out of the three shapes of weld fusion zones examined, i.e. triangular, rectangular and quadrilateral, the quadrilateral fusion zone gives the best comparison with experimental results.

References

[1] Lippold, J. C., and Savage, W. F. (1982) Solidification of austenitic stainless steel weldments: Part III-the effect of solidification behavior on hot cracking susceptibility, *Welding J*., **61(12)**, 388.

[2] Brooks and West, A. J. (1981) Hydrogen Induced Ductility Losses in Austenitic Stainless Steel Welds, *Metallurgical Transactions A*, **12**A, 213-223

[3] Kumar, Padhy Girish, and K.O.M.I.Z.O. Yu-ichi. (2013) Diffusible hydrogen in steel weldments, *Transactions of JWRI*, **42**, 39-62.

[4] Somerday, B.P., Dadfarnia, M., Balch, D.K., Nibur, K.A., Cadden, C.H. and Sofronis, P. (2009) Hydrogen-Assisted crack propagation in austenitic stainless steel fusion welds. *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, **40**, pp.2350–2362.

[5] Jackson, H.F., Marchi, C.S., Balch, D.K. (2013) Effect of low temperature on hydrogen-assisted crack propagation in 304L/308L austenitic stainless steel fusion welds, *Corrosion Science***77**, pp.210–221.

[6] Scheider, I., Pfuff, M. and Dietzel, W. (2008) Simulation of hydrogen assisted stress corrosion cracking using the cohesive model. *Engineering Fracture Mechanics*, **75**, pp.4283–4291.

[7] Raykar, N.R., Maiti, S.K. and Singh Raman, R.K. (2011) Modelling of mode-I stable crack growth under hydrogen assisted stress corrosion cracking, *Engineering Fracture Mechanics*, **78(18)**, pp.3153–3165.

[8] Brocks, Wolfgang, Rainer Falkenberg, and Ingo Scheider. (2012) Coupling aspects in the simulation of hydrogen-induced stress-corrosion cracking. *Procedia IUTAM 3*, 11-24.

[9] Nibur, K.A., Somerday, B.P., Balch, D.K., and SanMarchi, C. (2009) The role of localized deformation in hydrogen-assisted crack propagation in 21Cr–6Ni–9Mn stainless steel, *Acta Mater.*, vol. **57**, pp. 3795–3809.

[10] Fassel, V. (1959) Spectrographic Determination of Oxygen, Nitrogen and Hydrogen in Metals. *Bunseki kagaku*, **8(5)**, pp.324–335.

[11] Gangloff, R. P. (2003) Hydrogen assisted cracking in high strength alloys in Comprehensive Structural Integrity, Environmentally-Assisted Fracture. *Elsevier, Oxford.* Vol. 6.

[12] Alexander, D. J. and G. M. Goodwin. (1992) Thick-section weldments in 21-6-9 and 316LN stainless steel for fusion energy applications, *Materials. Springer US*, 101-107.

[13] Raykar, N. R., Maiti, S. K. and Singh, R.K. (2011) Modelling of mode-I stable crack growth under hydrogen assisted stress corrosion cracking, *Engineering Fracture Mechanics* **78(18)**, 3153-3165**.**

[14] Scheider, I., Pfuff. M., and Dietzel W. (2008) Simulation of hydrogen assisted stress corrosion cracking using the cohesive model. *Engineering Fracture Mechanics* **75(15)**, 4283-4291.

[15] Sofronis, P., and McMeeking, R. M. (1989) Numerical analysis of hydrogen transport near a blunting crack tip. *Journal of the Mechanics and Physics of Solids* **37(3)** 317-350.

[16] Oriani, R. A. (1970) The diffusion and trapping of hydrogen in steel. *Acta Metallurgica***18**, 147–157.

[17] Krom, A. H. M., Koers, W. J., and Bakker A. (1999) Hydrogen transport near a blunting crack tip. *Journal of the Mechanics and Physics of Solids* **47(4)**, 971-992.

[18] Kaiser, K., Raykar, N.R. (2015)Modelling of Hydrogen Assisted Stress Corrosion Cracking with Hydrogen Concentration Dependent Diffusivity, M. Tech dissertation, department of mechanical engineering, SP COE, India.

[19] Brocks, Wolfgang, Diya Arafah, and Mauro Madia. (2013) Exploiting Symmetries of FE Models and Application to Cohesive Elements. Milano/Kiel.

[20] Clarke, G. A., and Landes J. D. (1979) Evaluation of the J Integral for the Compact Specimen. *Journal*

of Testing and Evaluation **7(5)**, 264-269.
[21] F.R. Coe. (1973) Welding Steels Without Hydrogen Cracking, *Welding Institute, Cambridge, England*
[22] Geneon, Steven A. (1988) Hydrogen assisted cracking of high strength steel welds. *No, MTL-TR-88-12. Army Lab Command Watertown Ma Materials Technology Lab*.

Flow-excited vibration of a large-scale Axial-flow pump station with steel flow passageway based on FSI

†*H.Y. Zhang[1], L.J. Zhang[1], and L.J. Zhao[1, 2]

[1] College of water conservancy and hydropower engineering, Hohai University, China.

[2] Zhejiang design institute of water conservancy & hydroelectric power, China.

*Presenting author: zhanghanyun@hhu.edu.cn

†Corresponding author: zhanghanyun@hhu.edu.cn

Abstract:

Instead of the traditional concrete passageway, a new type structure of pump station with steel passageway is proposed for rapid construction and elimination temperature cracks. However, flow-excited vibration in the operation process of the pump station is still a crucial issue in design. A numerical model considering the interactions of the 3-dimensional (3D) unsteady turbulent flow with the concrete structure and steel passageway was established. Vibration characteristics and transient vibration for a pump station were analyzed based on fluid-structure interaction (FSI) method to predict the vibration responses of the concrete structure and steel passageway, and assess the vibration safety of the pump station structure system.

Keywords: Axial-flow pump houses; Steel flow passageways; 3D unsteady turbulent flow; Fluid structure interaction; Flow-excited vibration

Introduction

Vibration is a common problem in the operation process of the pump station. This long-term vibration has influence on the durability of equipment and the health of staff. Serious vibration could affect the safety and reliability of the pump station [1]. As a result, for the large pump stations with steel passageways, a new type structure of pump station, it is very crucial to predict and assess the vibration response and safety of the pump station structure including steel passageways.

Although a lot of researches on the flow-excited vibration analysis of the pump were carried, fluid-structure interaction was usually not taken into account, in particular, the steel passageway. In this study, in order to obtain the vibration responses occurred in the large axial-flow pump station, a numerical model considering the interactions of the 3D unsteady turbulent flow with the concrete structure and steel passageway was established based on ADINA. The impressible continuity equation, reynolds average Navier-Stokes equation and $k-\omega$ turbulent equations were used to simulate the 3D whole passageway unsteady turbulent flow of the axial-flow pump of steel passageway pump station [2]. The FSI boundary in the interface of fluid and structure is used for the energy transition between them [3]. The vibration characteristics and transient vibration for a pump station were analyzed based on fluid-structure interaction (FSI) method to predict the vibration responses of the

concrete structure and steel passageway, and assess the vibration safety of the pump station structure system.

Description of the numerical model

A designing pump station in East China was investigated. The typical section of the pump station is shown in Fig. 1. The main parameters are given in Table 1. A FSI model of pump station structure-steel passageway-fluid was established. The structural features and design proposal of the pump station (including concrete water inlet and outlet, concrete pump house structure, concrete supports, steel passageway and stiffening ribs) were simulated. The whole finite element model is shown in Fig. 2, where the *x*-axis of the coordinate system was vertical to the main stream, the *y*-axis was along the main stream, and the *z*-axis was upward vertically. The original point was in the impeller center. The concrete structure was discretized into 3D-solid elements; the steel passageway was discretized into shell elements; the stiffening ribs in the steel passageway was discretized into beam elements; the water in passageway was discretized into 3D-fluid elements; the upper structure and pump equipment were considered as added masses. The whole FEM model was totally discretized into 234120 elements, including 43488 3D-solid elements, 2730 shell elements, 2430 beam elements and 185472 3D- fluid elements. The normal constraint boundary was applied in the bottom and the wall vertical to the main stream of the concrete structure. In vibration analysis, single unit was investigated to improve the computational efficiency, as shown in Fig. 3 and 4. The finite element meshes of the steel passageway, stiffening ribs and water in passageway are shown in Fig. 5-7, respectively. The material parameters of the concrete structure, steel passageway and stiffening ribs are shown in Table 2. The dynamic elastic modulus of concrete was increased by 30%, and the Rayleigh damping was adopted with the damping ratio of 5% in the transient analysis.

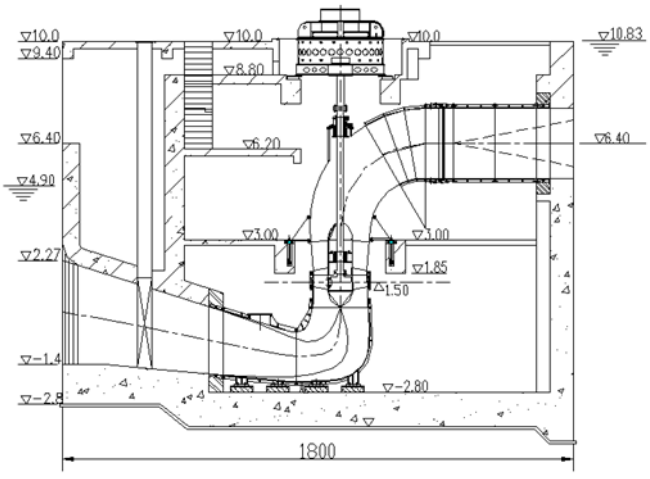


Figure 1. The typical section of the pump station

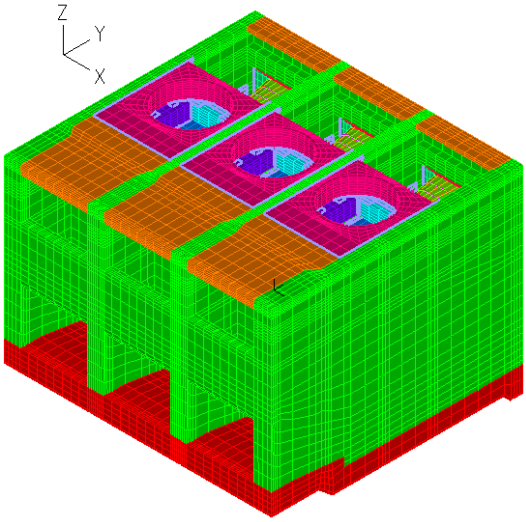


Figure 2. Whole finite element model

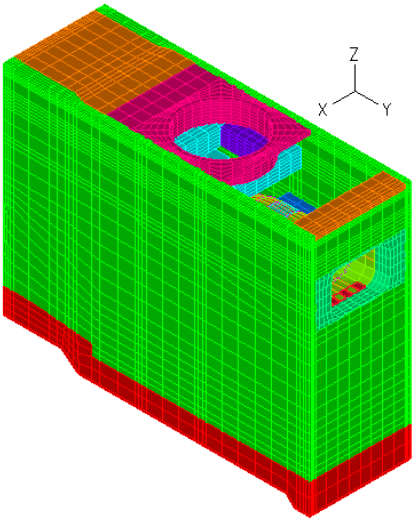


Figure 3. Finite element model of single unit

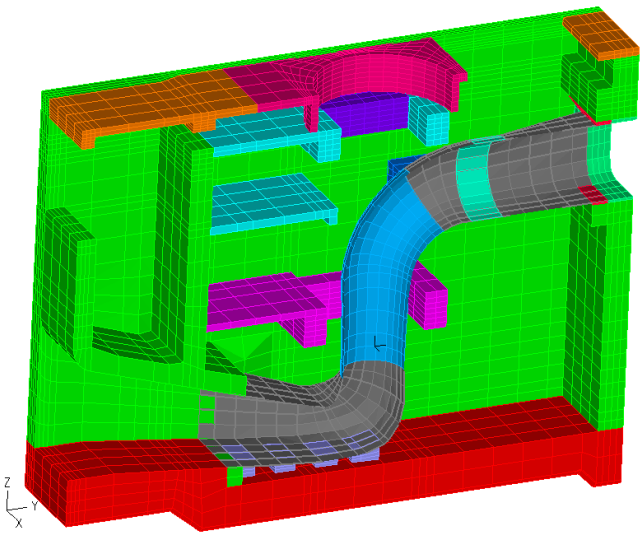


Figure 4. Cross-section of the mesh

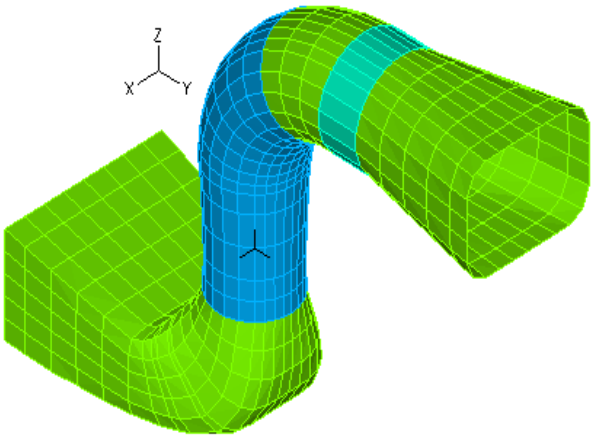


Figure 5. Finite element mesh of the steel passageway

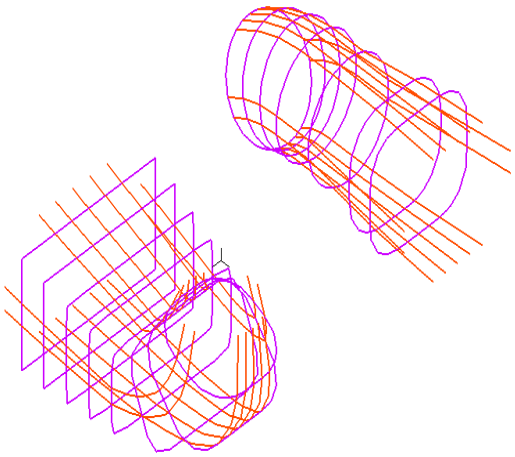


Figure 6. Finite element mesh of the stiffening ribs

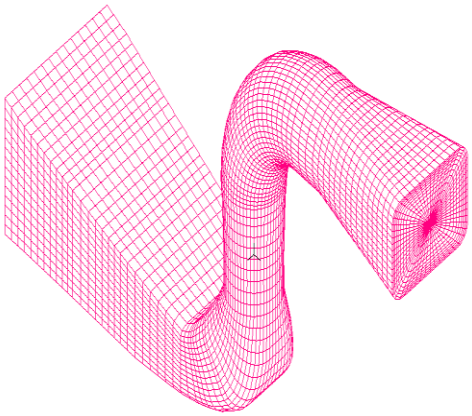


Figure 7. Mesh of the fluid

Table 1. Main parameters of the investigated pump station

Parameters	Q_d	H_d	n_d	Inlet width	Inlet height	Outlet width	Outlet height
Value	86 m^3/s	5.93 m	214.3 rpm	4.505 m	3.678 m	5.5 m	2.5 m

Table 2. Material parameters

	Density (kg/m^3)	Elastic modulus (GPa)	Poisson's ratio
Concrete structure	2400	28	0.167
Steel passageway	7800	210	0.3
Stiffening ribs	7800	210	0.3
Cast iron casing of pump	7800	210	0.3

Vibration characteristics analysis

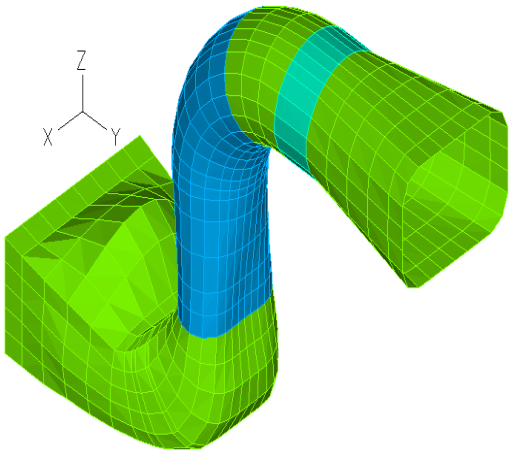
The vibration characteristics of the whole structure system in completion and operating conditions were analyzed. In operating condition, the fluid in the steel passageway was simulated by potential-based fluid elements. The first 15 order frequencies and the different main vibration positions are listed in Table 3.

Comparison of two conditions

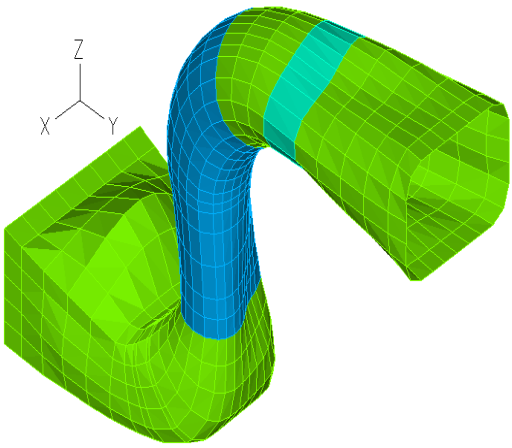
The results indicate that the water in the passageway has obvious effects on the vibration characteristics of the whole pump station system. All-order frequencies of the whole pump station system in operating condition are smaller than that in completion condition. Taking the fundamental frequency as an example, the fundamental frequency is 7.17 Hz in the operating condition, which is decreasing by 56.8 % compared with that in completion condition. The vibration modes of two conditions are different. For instance, the primary vibration position is the concrete structure in the first mode of vibration in completion condition, whereas the primary vibration position is the steel passageway in the first mode of vibration in operating condition. For the first mode of vibration, in operating condition the whole steel passageway vibrates along the *y*-axis, and inlet of the passageway vibrates along the z-axis upward vertically, whereas in completion condition the whole concrete structure vibrates along the y-axis. The first several modes of vibration are shown in Fig.8

Table 3. The first 15 frequencies in two conditions

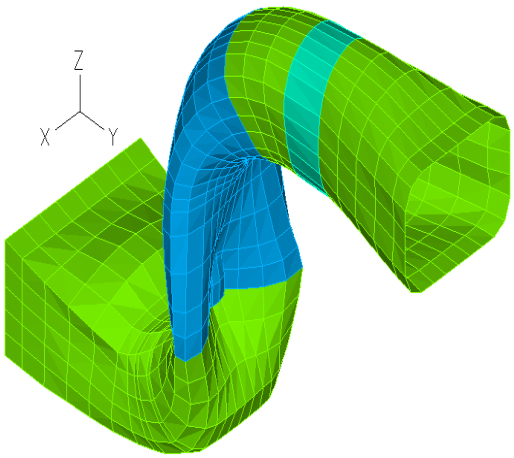
No.	Completion state		Operating state	
	Frequency (Hz)	Primary vibration position	Frequency (Hz)	Primary vibration position
1	16.60	Concrete structure	7.17	Steel passageway
2	32.72	Steel passageway	15.21	Concrete structure
3	36.97	Steel passageway	17.06	Steel passageway
4	38.30	Steel passageway	18.61	Steel passageway
5	42.93	Concrete structure	20.46	Steel passageway
6	44.97	Steel passageway	22.41	Steel passageway
7	46.25	Steel passageway	23.75	Steel passageway
8	49.82	Concrete structure	24.10	Concrete structure
9	50.92	Steel passageway	25.09	Steel passageway
10	56.25	Concrete structure	27.00	Steel passageway
11	56.80	Concrete structure	27.39	Steel passageway
12	57.49	Steel passageway	30.00	Steel passageway
13	58.10	Steel passageway	31.45	Concrete structure
14	59.98	Concrete structure	33.61	Steel passageway
15	62.59	Concrete structure	35.11	Concrete structure



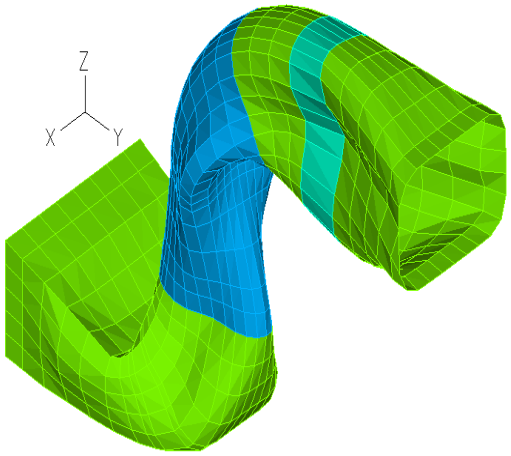
(a) The 1st mode of vibration



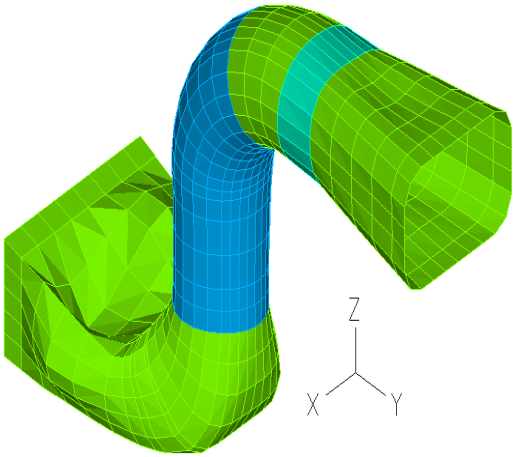
(b) The 3rd mode of vibration



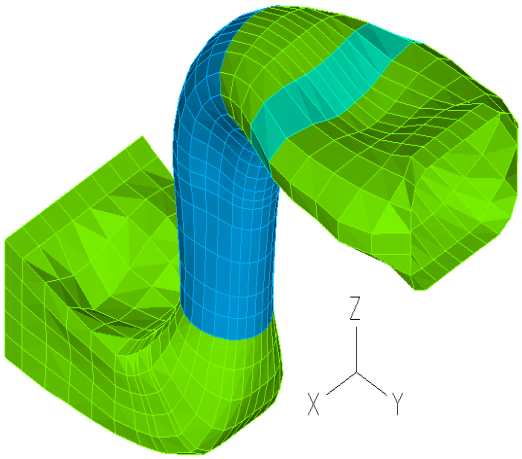
(c) The 4th mode of vibration



(d) The 5th mode of vibration



(e) The 6th mode of vibration



(f) The 7th mode of vibration

Figure 8. Vibration modes of the steel passageway

Resonance check

The resonance of equipment and structures must be avoided in operating condition. The interval of the frequency between the structure and exciting vibration frequency should be greater than 20~30% in operating condition according to *pump station design specification (GB50265-2010).*The expression is as follow [4]:

$$\frac{(f_i - f_{0i})}{f_i} \times 100\% > 20\% \sim 30\% \quad \text{or} \quad \frac{(f_{0i} - f_i)}{f_{0i}} \times 100\% \quad 20\% \sim \tag{1}$$

Where, f_{0i} is the i[th] order vibration frequency of the structure, f_i is the frequencies of vibration sources of various equipment.

The vibration of pump station structures mainly results from machines, electromagnetism and hydraulic vibration which have relation to rotation frequency of the pump and close to its

rotational frequency. In this project, the rotational frequency of the pump is 3.575 Hz. The ±20% of the rotational frequency ranges between 2.86 and 4.29 Hz. The frequency of the whole structure system is beyond the range both in completion and operating conditions.

Transient vibration analysis

Numerical method

The unsteady flow in the passageway was simulated using the RNG $k-\varepsilon$ model [5]. The pressure-velocity coupling was performed using the SIMPLEC algorithm. Second-order format was used for pressure term [6]. 5000 time steps were picked out, with time step as 0.01s in transient analysis.

There were three combinations of boundary condition used in pump flow analysis. (1) Inlet: according to the previous research, the predominant frequency of the pulsating pressure in passageway is close to the rotational frequency and unevenness of the pulsating pressure is in the range of 16%. Therefore, a simple harmonic velocity, $v=\bar{v}(1\pm 0.1\sin wt)$, with $\bar{v}$ =0.865 m/s, pulsation amplitude of 10% and pulsation frequency being equal to rotational frequency, was specified at the inlet[7] [8][9][10]. A averaged velocity, $\bar{v}$ =0.865m/s, is obtained based on designed single unit flow 14.3 m^3/s and area of the inlet of the pump station. (2) The vent was set as outflow boundary condition [11] [12]. (3) The SFI boundary was set in interface between the steel passageway and fluid.

Concrete structure vibration

The vibration amplitude of the displacement and stress of key points in the concrete structure are listed in Table 4 and 5. The positions of the key points were shown in Fig. 9.

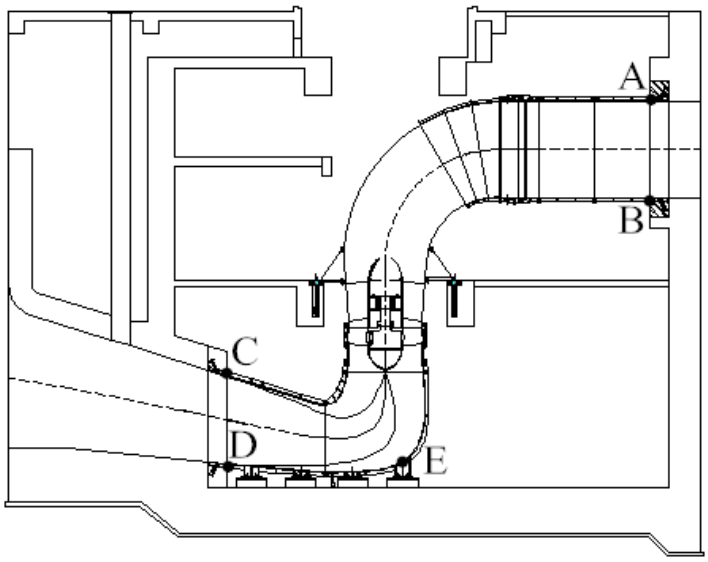


Figure 9. Key points in the concrete structure

The vibration displacement of key points is tiny. The maximum X-direction vibration amplitude of displacement is 0.019 um in the concrete supports in the elbow of the passageway. The maximum Y-direction and Z-direction vibration amplitude of displacement is 2.311 and 3.046 um respectively, and both in the joint between the concrete inlet and steel passageway. The vibration amplitude of stress of all key points is much smaller than strength of the concrete, which is not the controlling factor in the concrete structure design.

Table 4. The amplitude of the vibration displacement of the key points in the concrete structure in operating condition（μm）

Key point	Amplitude of vibration displacement		
	X-direction	Y-direction	Z-direction
A	0.004	1.390	0.176
B	0.002	1.064	0.199
C	0.002	2.311	3.046
D	0.004	0.945	0.038
E	0.019	0.864	0.004

Table 5. The amplitude of the vibration stress of the key points in the concrete structure in operating condition（kPa）

Key points	Normal stress amplitude			Amplitude of the first principal stress	Amplitude of the third principal stress
	σ_x	σ_y	σ_z		
A	1.327	0.385	3.245	2.409	1.326
B	1.388	0.349	4.094	1.388	4.064
C	4.391	0.521	10.492	1.921	11.991
D	0.176	0.200	0.033	0.598	0.448
E	1.919	0.851	0.226	1.946	0.703

Steel passageway vibration

The vibration amplitude of the displacement and stress of key points in the steel passageway were list in Table 6 -8. The positions of the key points were shown in Fig. 10.

The vibration amplitude of displacements of key points in the steel passageway is also tiny.

The maximum X-direction vibration amplitude of displacement is 31.520 um in Point 2 in the side wall of the inlet segment of the passageway. The maximum Y-direction and Z-direction vibration amplitude of displacement are 53.215 and 176.435 um respectively, and both in Point 1 in the top surface of the inlet segment of the passageway. The vibration amplitude of stress of all key points is much smaller than strength of the steel, which is not the controlling factor in the steel passage design.

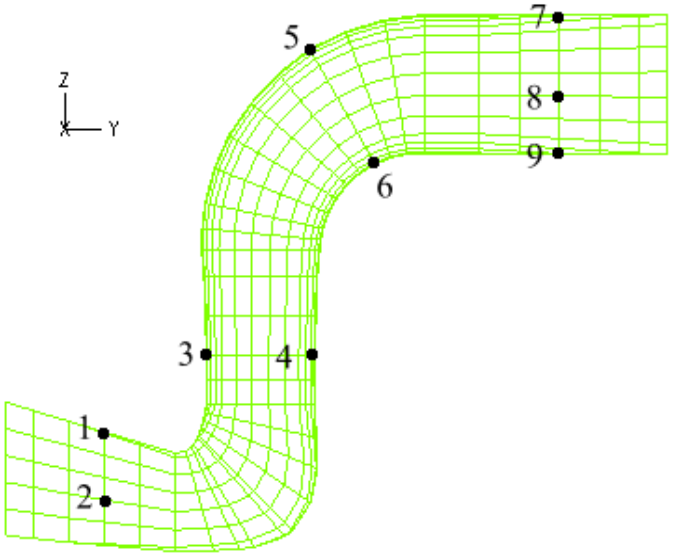


Figure 10. Key points in the steel passageway

Table 6. The amplitude of the vibration displacement of the key points in steel passageway in operating condition（μm）

Key point	Amplitude of vibration displacement		
	X-direction	Y-direction	Z-direction
1	0.008	53.215	176.435
2	31.520	3.879	0.606
3	0.013	12.045	1.282
4	0.014	4.382	0.755
5	0.040	1.192	0.371
6	0.024	1.654	2.103
7	0.012	1.389	6.322
8	4.518	1.394	0.388
9	0.014	1.106	7.791

Table 7. The amplitude of the normal stress of the key points in steel passageway in operating condition（kPa)

Key point	Normal stress amplitude		
	σ_x	σ_y	σ_z
1	298.000	282.660	19.909
2	1.958	301.050	398.900
3	358.400	2.268	33.386
4	767.900	4.967	46.750
5	17.750	18.332	6.875
6	69.215	37.165	9.418
7	120.634	58.260	0.005
8	1.598	21.142	57.662
9	140.634	75.242	0.009

Table 8. The amplitude of the shear stress and principal stress of the key points in steel passageway in operating condition（kPa)

Key point	Shear stress amplitude			Amplitude of the first principal stress	Amplitude of the third principal stress
	τ_{xy}	τ_{xz}	τ_{yz}		
1	0.486	0.165	75.480	302.850	0.230
2	24.329	1.816	47.720	418.350	0.019
3	0.066	0.969	0.566	33.317	358.400
4	0.014	0.197	2.738	46.910	767.900
5	0.190	0.091	11.331	25.247	17.750
6	0.478	0.288	18.067	0.710	69.220
7	0.337	0.172	0.594	58.287	0.095
8	3.472	4.582	15.371	46.237	39.900
9	0.238	0.016	0.424	0.429	140.631

Vibration safety assessment

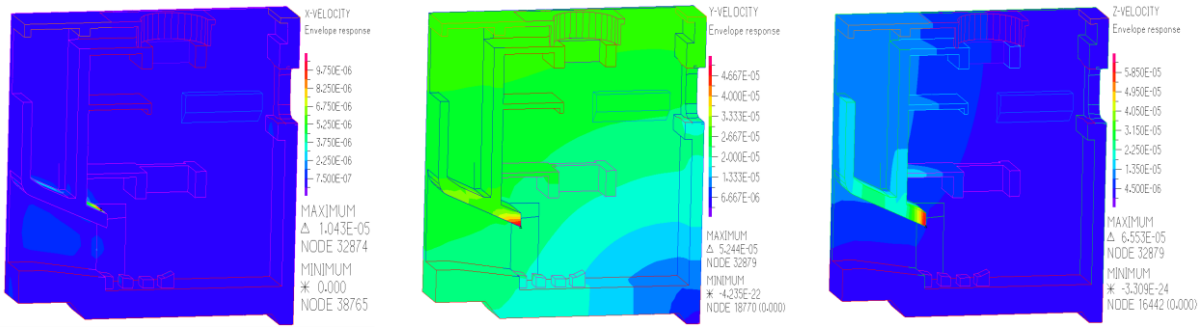
At present, there is no explicit control standard for the vibration checking of pump station structures in China. The vibration of pump station structures is long-term continuous forced vibration, which is similar to the vibration of the hydropower house. Major research focused on the vibration control standard for hydropower house. The suggested vibration control standard for hydropower house is proposed, as shown in Table 9 [5]. The vibration of this pump station is assessed based on the standard for hydropower house.

Table 9. Suggested vibration control standard for hydropower house

Structure member		Vibration displacement (mm)	Vibration velocity (mm/s)		Acceleration (m/s^2)	
			Vertical	Horizontal	Vertical	Horizontal
Floor	As general structure	0.2	5.0		1.0	
	As instrument foundation	0.01	1.5			
	Human health	0.2	3.2	5.0	0.4	1.0
Solid wall		0.2	10.0		1.0	
Generator pier		0.2	5.0		1.0	

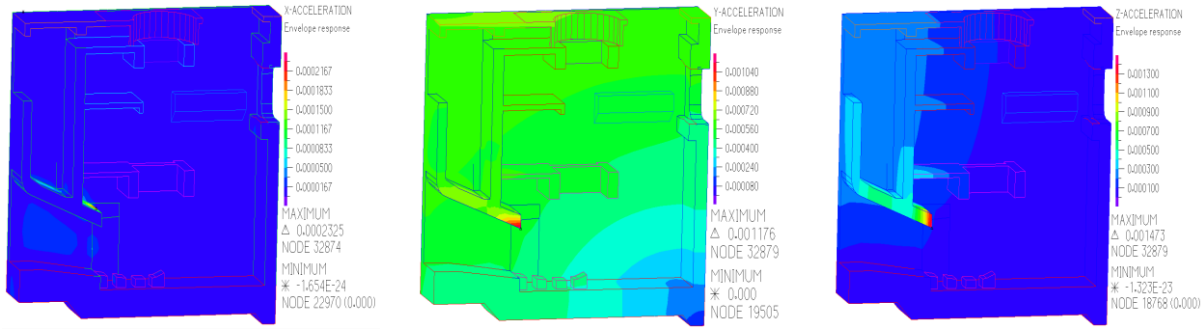
The maximum vibration amplitude of the concrete structure and steel passageway including displacement, velocity and acceleration are list in Table 10-12. The Envelope diagrams of vibration responses are shown in Fig. 11 to 14.

Compared the data in Table 10-12 with the suggested standard in Table 9, the amplitude of vibration displacement of whole pump station system is not large. The amplitude of vibration displacement of concrete structure belongs to the allowed value listed in Table 9. The amplitude of vibration displacement in top surface of the inlet section of the steel passageway is 0.176mm. It should be relieved by strengthening stiffening ribs. The vibration velocity of the concrete structure belongs to the allowed value listed in Table 9. The maximum Z-direction vibration velocity of the steel passageway is 3.765mm/s, exceeding the index of human health (3.2mm/s). There is no office area near the passageway, so it is available. The vibration acceleration of the concrete structure and steel passageway belong to the allowed value listed in Table 9.



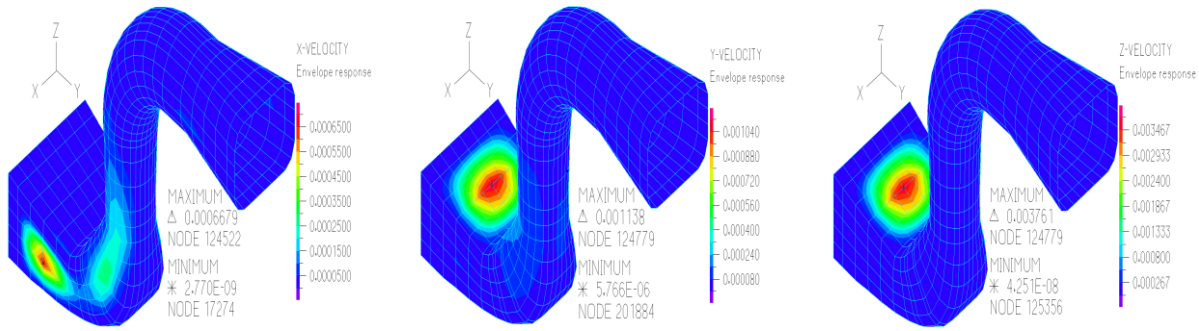
(a) X-direction **(b) Y-direction** **(c) Z-direction**

Figure 11. Envelope diagram of the vibration velocity of the concrete structure (m/s)



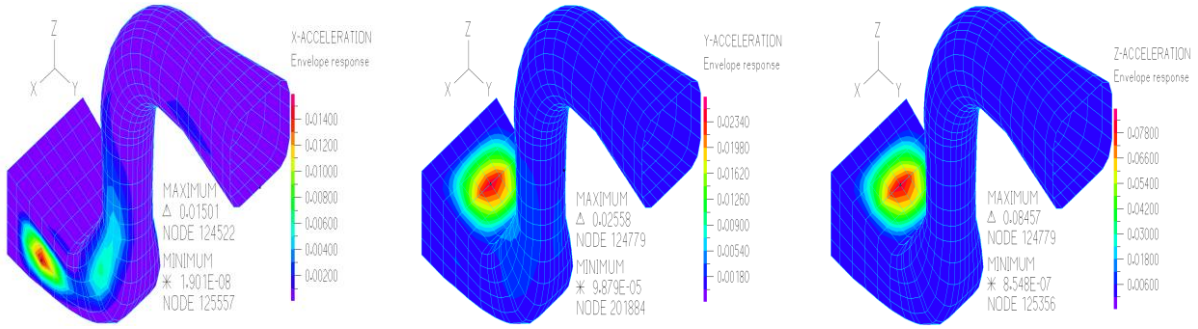
(a) X-direction **(b) Y-direction** **(c) Z-direction**

Figure 12. Envelope diagram of the vibration acceleration of the concrete structure (m/s^2)



(a) X-direction **(b) Y-direction** **(c) Z-direction**

Figure 13. Envelope diagram of the vibration velocity of the steel passageway (m/s)



(a) X-direction **(b) Y-direction** **(c) Z-direction**

Figure 14. Envelope diagram of the vibration acceleration of the steel passageway (m/s^2)

Table. 10 The maximum vibration amplitude of displacement（μm）

	X-direction	Y-direction	Z-direction
Concrete structure	0.019	2.311	3.046
Steel passageway	31.520	53.215	176.435

Table. 11 The maximum vibration velocity（mm/s）

	X-direction	Y-direction	Z-direction
Concrete structure	0.010	0.052	0.066
Steel passageway	0.067	1.139	3.765

Table. 12 The maximum vibration acceleration（m/s^2）

	X-direction	Y-direction	Z-direction
Concrete structure	0.0002	0.001	0.001
Steel passageway	0.015	0.025	0.085

Conclusion

(1) The water filling in the passageway has obvious effects on the vibration characteristics of the whole pump station system. The fluid-structure interaction is essential factor in resonance check of the pump station structure system.

(2) The flow-excited vibration of the pump station is tiny and high frequency. The joint between the concrete inlet and steel passageway and the top surface of the inlet of the steel passageway are weakness positions where should be strengthened in design.

(3) A FSI model considering the interactions of the concrete structure with steel passageway, fluid and pump machinery should be established and investigated.

Acknowledgement

Special thanks are given to financial supports provided by CRSRI Open Research Program (CKWV2015215/KY), Fundamental Research Funds for the Central Universities (2014B03214) and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Wang, X., and Li T. C. (2008) Vibration analysis of pump house of one large-scale bulb tubular pumping station. Proceeding of first international conference of modeling and simulation, Vol. V: 426-431.

[2] Wei, S. H. (2010) Simulation of transmission path and research on structural vibration characteristics of hydropower house. Hohai University

[3] ADINA R&D, Inc. theory and modeling guide volume Ⅲ: ADINA CFD & FSI 2009

[4] Zhao, L. J., Zhang L. J., Zhang, H. Y., Wang, J. Q., Cheng, J. (2016)Vibration characteristics analysis and safety evaluation of steel-made flow passage in vertical axial-flow pumping station, *Water Resources and Power* 34(01) ,146-149.

[5] Mompean, G. (1998) Numerical simulation of a turbulent flow near a right-angled corner using the special non-linear model with RNG k-e Equations, *Computers and Fluids* 27(7), 847-859.

[6]Shuai, Z. L., Li, W. Y., Zhang, X. Y., Jiang, C. X., and Li, F. C. (2014) Numerical study on the character of pressure fluctuation in an axial-flow water pump, *Advanced in Mechanical Engineering*.

[7] Ma, Z. Y., Zhang, Y. L., Cheng, J. (2013) *Theory and application on coupling dynamic of hydropower house and equipment* , 1st edn, China Water & Power Press, Beijing, China

[8] Yang, S. S., Kong, F. Y., Qu, X. Y., Jiang, W. M. (2012) Influence of blade number on the performance and pressure pulsations in a pump used as a turbine, *Journal of Fluids Engineering* 134, 1-10.

[9] Yang, S. S., Kong F., Y., and Chen B. (2011) Research on pump volute design method using CFD, International Journal of Rotating Machinery, 13780.

[10]Ohashi, H. etc. (1991) *Vibration and oscillation of hydraulic machinery*, Cambridge University Press, London, U.K.

[11]Wang, F. J., Qu, L. X., He, L. Y. and Gao, J. Y. (2013) Evaluation of flow-induced Dynamic stress and vibration of volute casing for a large-scale double-suction centrifugal pump, *Mathematical Problems in Engineering* ID764812.

[12] Majidi, K. (2005) Numerical study of unsteady flow in a centrifugal pump, *Journal of turbo machinery* 125(02), 363-371.

A 3-D Meshfree Numerical Model to Analyze Cellular Scale Shrinkage of Different Categories of Fruits and Vegetables during Drying

†*C.M. Rathnayaka Mudiyanselage [1, 2], H.C.P. Karunasena [3], Y.T. Gu [1], L. Guan [1], J. Banks [1] and W. Senadeera [1]

[1]Queensland University of Technology (QUT), Science and Engineering Faculty, School of Chemistry Physics and Mechanical Engineering, 2-George Street, Brisbane, QLD 4001, Australia.
[2]Department of Chemical and Process Engineering, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka.
[3]Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, University of Ruhuna, Hapugala, Galle, Sri Lanka.

*Presenting author: charith.rathnayaka@hdr.qut.edu.au
†Corresponding author: charith.rathnayaka@hdr.qut.edu.au

Abstract

In order to optimize food drying operations, a good understanding on the related transport phenomena in food cellular structure is necessary. With that intention, a three-dimensional (3-D) numerical model was developed to better investigate the morphological changes and related solid and fluid dynamics of single parenchyma cells of apple, carrot and grape during drying. This numerical model was developed by coupling a meshfree particle based method: Smoothed Particle Hydrodynamics (SPH) with a Discrete Element Method (DEM). Compared to conventional grid-based numerical modelling techniques such as Finite Element Methods (FEM) and Finite Difference Methods (FDM), the proposed model can better simulate deformations and cellular shrinkage within a wide range of moisture content reduction. The model consists of two main components: cell fluid and cell wall. The cell fluid model is based on SPH and represents the cell protoplasm as a homogeneous Newtonian liquid. The cell wall model is based on a DEM and approximates the cell wall to an incompressible Neo-Hookean solid material. A series of simulations were conducted to mimic the gradual shrinkage during drying as a function of moisture content.

Keywords: Food drying; Meshfree methods; Plant cell modelling; Smoothed Particle Hydrodynamics (SPH); Three-dimensional (3-D) model

Introduction

Drying is one of the most common and cost effective techniques for extending the shelf life of food materials and also is used for the production of numerous traditional and innovative food products [1]. It is employed to preserve approximately 20% of the planet's fruits and vegetables [2]. During drying, the moisture is removed out of food material in order to slow down biological activities. With the removal of moisture, the food cellular structure undergoes major structural deformations which influence the drying process performance, food quality and the final market value. Therefore, to develop effective and efficient food drying operations, it is important that these cellular structural deformations are well understood and optimised. In doing so, a thorough understanding of the underlying solid and fluid dynamics is pivotal. The key driving forces for the transport phenomena are the moisture content [3-8] and the drying temperature [9]. The moisture content has a strong relationship with the cell turgor pressure [10] and the drying temperature links with the relative humidity and the rate at which moisture is removed from the cellular structure during drying. To derive appropriate

relationships among such driving forces, cellular morphogenesis and underlying transport phenomena, various microscale theoretical [11, 12] and empirical [3, 5, 13] models have been developed.

Numerical modelling has been utilised as an efficient tool in the studies of deformational analysis of various materials. Until recent times, this had not been used for comprehensive analysis of micro-structural deformations of food materials during drying. However, numerical modelling has recently attracted much attention as a viable tool for this purpose [14-16]. It is believed that through an accomplished numerical model, vast benefits could be obtained in food engineering in terms of drying process performance and predicting the final quality of the dried food product. With this background, this study aimed to develop a three-dimensional (3-D) numerical model in order to investigate the morphological changes and related solid and fluid dynamics of parenchyma cells of apple, carrot and grape during drying. For the implementation, more versatile and novel meshfree particle methods have been chosen over the classical grid-based methods. A series of simulations were conducted to predict the shrinkage of each food tissue variety as a function of the moisture content.

Methodology

3-D Representation of a Single Cell

For this study, a single cell of a cortex (parenchyma) tissue is considered, which is the fundamental building block of most bulk plant tissue structures. This type of a cell could be physically regarded as a stiff and thin-walled vessel containing a viscous fluid. Therefore, the developed numerical model is composed of two main components: cell wall and cell fluid. Based on the literature, the basic geometrical shape of a single cell was assumed to be spherical (see Figure 1) [17]. In the cell fluid model, the fluid volume was approximated to a sphere and the cell wall was approximated to a hollow 3-D spherical shell, enclosing the fluid sphere. The cell fluid hydrodynamic pressure is counterbalanced by the tension of the cell wall. Cell fluid was assumed to be incompressible and the system as a whole was treated as an isothermal unit.

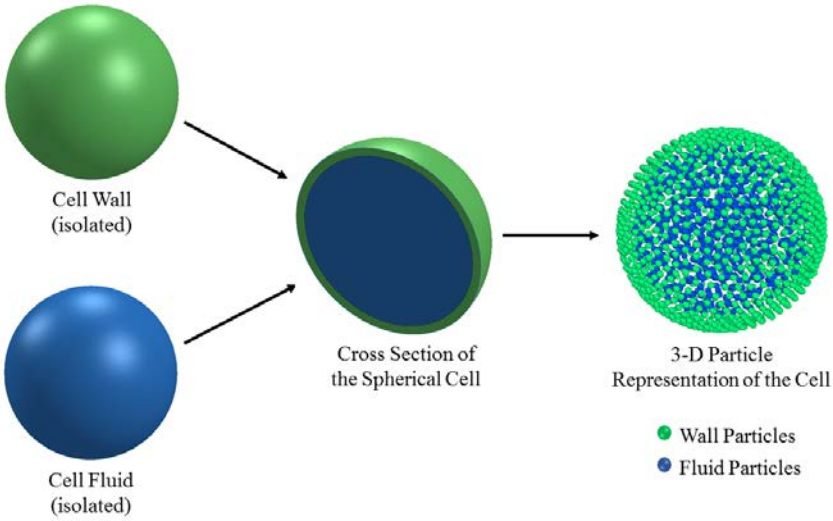


Figure 1. 3-D particle representation of the cell model, which is composed of two sub-models: cell fluid model and cell wall model

After establishing these fundamental assumptions, the cell fluid and cell wall were separately discretised into particle schemes. The intention of this discretisation was to represent the whole system using a large number of non-interconnected particles in order to satisfy the fundamentals of Meshfree Particle Methods (e.g. Smoothed Particle Hydrodynamics (SPH) [18]). Due to the adaptivity and flexibility of the adopted particle framework, it could be easily extended up to multiple cell systems by aggregating more cells together [19, 20]. At the same time, it has the capability to analyse different types of cells (apple, carrot etc.) without significant changes in the main modelling and simulation framework [21]. Furthermore, this particular modelling technique also ensures the ability to incorporate the mechanisms at the subcellular level.

Cell Fluid Model

protoplasm, which can be about 80-90% by volume [22], the cell fluid was approximated to an incompressible homogeneous Newtonian fluid equivalent to water with an elevated viscosity. This can be effectively modelled with Smoothed Particle Hydrodynamics (SPH) considering low Reynolds number flow characteristics [19, 23-25]. Accordingly, in order to model the cell fluid, four different types of forces were used: pressure forces (F^p), viscous forces (F^v), wall-fluid repulsion forces (F^{rw}) and wall-fluid attraction forces (F^a) as presented in Figure 2 [14, 26, 27]. The cumulative effect of these forces is used to define the total force (F_i) on any fluid particle i as,

$$\boldsymbol{F}_i = \boldsymbol{F}^{p}_{ii'} + \boldsymbol{F}^{v}_{ii'} + \boldsymbol{F}^{rw}_{ik} + \boldsymbol{F}^{a}_{ik} . \tag{1}$$

Where *i'* represents the neighboring fluid particles and *k* the interacting wall particles

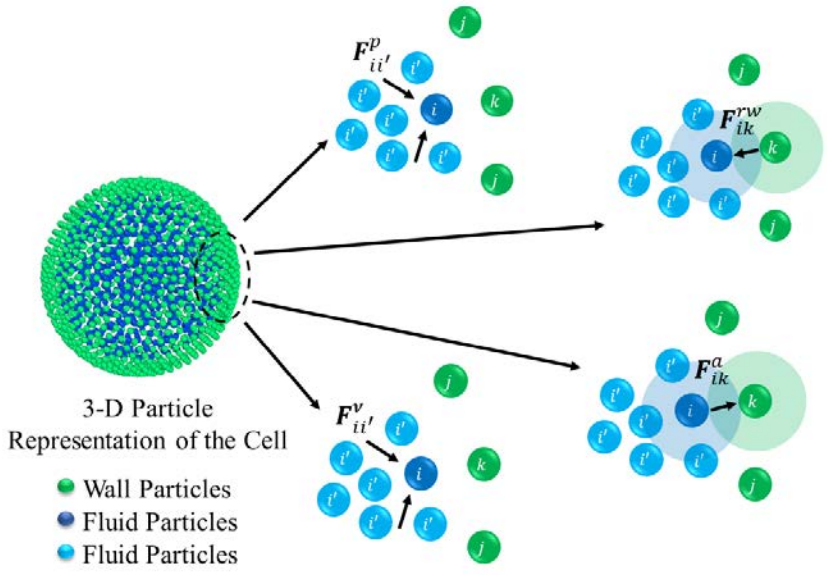


Figure 2. Force fields on the 3-D fluid particle domain: pressure forces (F^p), viscous forces (F^v), wall-fluid repulsion forces (F^{rw}) and wall-fluid attraction forces (F^a)

Cell Wall Model

The cell wall was approximated to a Neo-Hookean solid material. It has been treated as a particle scheme composed of interconnected discrete elements connected to each other via a network, such that each element carries properties of the corresponding cell wall element. The cell wall deformations are represented by the displacement of respective particles under four types of force interactions: stiff forces (F^e), damping forces (F^d), wall-fluid repulsion forces (F^{rf}) and wall-fluid attraction forces (F^a), as illustrated in Figure 3. [14, 15, 26]. Accordingly, the total force (F_k) on any wall particle k is derived as,

$$F_k = F^e_{kj} + F^d_{kj} + F^{rf}_{ki} + F^a_{ki}\,. \tag{2}$$

Where, for each wall particle *k*, *i* indicate the neighboring fluid particles, *j* the bonded wall particles and *l* the non-bonded wall particles

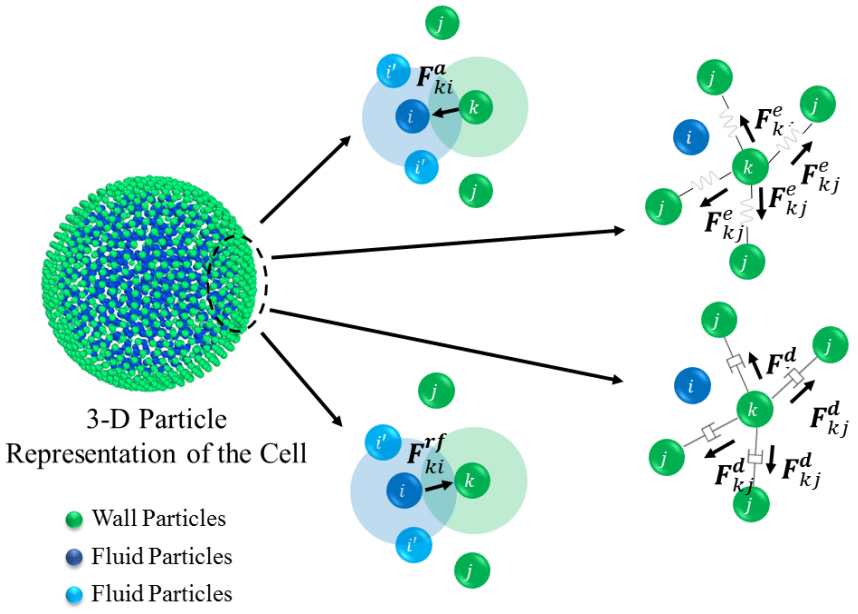


Figure 3. Force fields on the 3-D wall particle domain: stiff forces (F^e), damping forces (F^d), wall-fluid repulsion forces (F^{rf}) and wall-fluid attraction forces (F^a)

Modelling of Different Categories of Fruits and Vegetables

In this study, individual cells of apple, carrot and grape have been modelled. Each food plant material is modelled via customized model parameters obtained from microscopic experimental observations and other numerical models available in the literature. The physical properties necessary for modelling apple, carrot, and grape cells were directly extracted from sources in literature. There were a few properties which had to be calculated and assumed in the process. For instance, shear moduli (G) of the cell wall for carrot and grape were set so that the Young's modulus (E) was 100 MPa which resulted in comparable values for cell wall stiffness at relevant cell wall thickness values. Turgor pressure of grape cells were set equal to the value of apple cells due to the lack of necessary information in literature. This approach has been proven to be successful in similar studies [21]. These model

parameters are outlined in Table 1. The parameters which were used in common for all four plant food categories are shown in Table 2.

Modelling Different Dryness States

The previously mentioned model features were numerically set up with the physical properties of the cells as given in Table 1 and 2. The software tool, COMSOL Multiphysics (COMSOL) was used to generate the initial 3-D particle arrangement in a 3-D sphere corresponding to both the cell fluid and cell wall. There is the possibility to define and fine-tune the initial particle gap and the cell geometrical characteristics using COMSOL in order to obtain the initial particle positions with the preferred and effective particle resolution. The fluid particle scheme was placed without any interconnections among particles, adhering to the SPH fundamentals. In the cell wall, a series of spring networks joining the cell wall particles were used according to the fundamentals of DEM [14, 15, 20] .

It should be noted here that the simulations were carried out mainly based on the moisture content domain, similar to the recent 2-D meshfree based dried plant cell and tissue models [14]. It has also been assumed that the cell turgor pressure stays positive during the entire drying process and it would reduce at a regular rate with the moisture content variation. At the same time, the osmotic potential values corresponding to each dryness states were set equal to the minus value of the relevant turgor pressure and in the meantime the the magnitudes were kept constant in order to assure the stability of the numerical scheme [21].

Table 1. Values of the physical parameters adopted in the model

Parameter	Food variety used for modelling		
	Apple	Carrot	Grape
Initial cell diameter (D_0) (μm)	150	100	150
Cell wall thickness (T_0) (nm)	126	126	62
Wall shear modulus (G) (MPa)	18	33	33
Fresh cell turgor pressure (P_T) (kPa)	200	400	200
Fresh cell osmotic potential (π_T) (kPa)	-200	-400	-200

As the model evolves with time according to the difference between the cell turgor pressure and the osmotic potential, the mass of the cell fluid tends to change and causing slight density variations [14, 15]. Such changes of density cause significant changes in cell turgor pressure as governed by an equation of state (EOS). These turgor pressure variations tend to push the cell wall inwards (shrinkage) or outwards (inflation) causing the cell volume, equivalent diameter and surface area to change. Based on such cell volume changes, the cell turgor pressure varies since it has to be counterbalanced by the cell wall tension. The changes in cell turgor pressure leads to the cell fluid mass gains or losses, which is governed by a mass transfer equation in the cell fluid model [14, 15]. This is repeated as a cycle until the cell

reaches a steady state condition where the cell size and the physical properties tend to reach steady values. For each cell type, this whole process was implemented and simulations were conducted.

Table 2. Values of the physical parameters adopted in the model

Parameter	Value	Reference
Initial cell fluid mass	1.767×10^{-9} kg	[14]
Initial cell wall mass	1.767×10^{-10} kg	[14]
Wall damping ratio (γ)	5×10^{-6} $Nm^{-1}s$	[14]
Cell fluid viscosity (μ)	0.1 Pas	[25]
Cell wall permeability (L_P)	2.5×10^{-6} $m^2N^{-1}s$	[28]
Fluid compression modulus (K)	20 MPa	[14]

The model was developed as a C++ source code and it was executed in a High Performance Computer (HPC). Algorithms in an existing SPH source code based on FORTRAN [18] were referred in developing the C++ source code. For the visualisations, Open Visualization Tool (OVITO) [29] was used [21]

Results and Discussion

Experimental data on drying of plant cellular materials indicate that there is an acceptable linear relationship between the removed moisture content and the bulk volumetric shrinkage [4, 30-33]. Further, the reductions of the cell area, diameter and perimeter are proportional to the overall volumetric shrinkage as well as the removed moisture content [3, 5]. All these experimental findings indicate that there is a strong connection between the moisture content of a plant food material and its shrinkage characteristics. In Figure 4, the model predictions have been visualized for apple, carrot and grape cells. Next, in order to quantify shrinkage characteristics of the cells, a set of geometrical parameters were used. Here, the moisture content (X) of the cell at a given dryness state is a critical parameter and the normalized moisture content (X/X_0) was used here to assist comparison of the model behavior over different dryness states (see Eq. 3). Similarly, the geometrical parameters used to quantify the cellular shrinkage characteristics were also normalized (see Eq. 4, 5 and 6) [14, 15]. The model predictions for these parameters were then compared with the corresponding experimental results in literature [3]. In Figure 5, normalized cell area variation for apple cells are presented and in Figure 6 and 7, normalized cell diameter variation and the normalized cell perimeter variation are presented, respectively. The parameter variation comparison for

the other types of food categories (i.e. carrot and grape) follow a similar trend and were not included here.

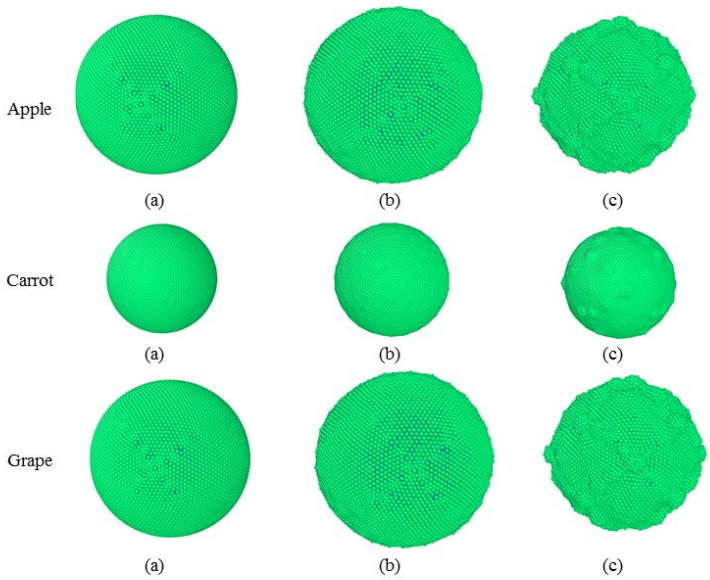


Figure 4. Visualization of the numerical results of the 3-D SPH-DEM model (a) initial conditions before simulations (b) the inflated fresh apple cell ($X/X_0 = 1$) (c) dried state ($X/X_0 = 0.1$)

In the course of this study, results of our work (see the comparisons in the Figures 5, 6 and 7) has shown that there is the possibility to successfully develop a 3-D numerical model for the simulation of single parenchyma cells of apple, carrot and grape during the process of drying using a meshfree approach. There is a reasonably good agreement between the SPH-DEM model predictions and the experimental results [3, 5]. As it could be observed in Figures 5, 6 and 7, this agreement is more positive in the higher moisture content values (i.e. $X/X_0 \geq 0.4$). When it comes to extremely low moisture contents (i.e. $X/X_0 \leq 0.25$), the model predictions tend to deviate from the realistic values considerably.

$$normalised\ moisture\ content = \frac{X}{X_0} = \frac{steady\ state\ cell\ fluid\ mass}{fresh\ cell\ fluid\ mass}. \quad (3)$$

$$normalised\ surface\ area = \frac{A}{A_0} = \frac{steady\ state\ cell\ surface\ area}{fresh\ cell\ surface\ area}. \quad (4)$$

$$normalised\ diameter = \frac{D}{D_0} = \frac{steady\ state\ cell\ Diameter}{fresh\ cell\ Diameter}. \quad (5)$$

$$normalised\ perimeter = \frac{P}{P_0} = \frac{steady\ state\ cell\ Perimeter}{fresh\ cell\ Perimeter}. \quad (6)$$

Therefore, it could be deduced that the developed 3-D SPH-DEM plant cell model has got the ability to approximate the true cellular scale drying behavior much quantitatively and

qualitatively. When comparing with the most recent 2-D numerical models for plant tissue drying [34, 35], it could be seen that this model shows great competency and potential to closely describe the true plant food tissue drying scenario, particularly in 3-D. Therefore credentials are there in these modelling schemes to be further developed and utilized in the field of food engineering. Furthermore, it should be emphasized that there is room for further improvements in the model especially at extremely dried stages ($X/X_0 \leq 0.25$). These improvements would add more details into the true deforming behavior of the cellular system.

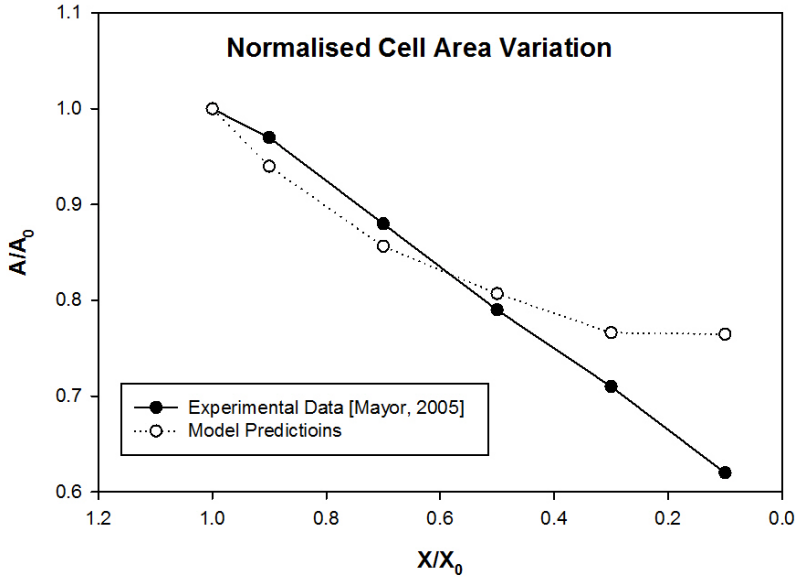


Figure 5. Comparison of model predictions and experimental results [3] for normalised cell area of a single apple cell during drying

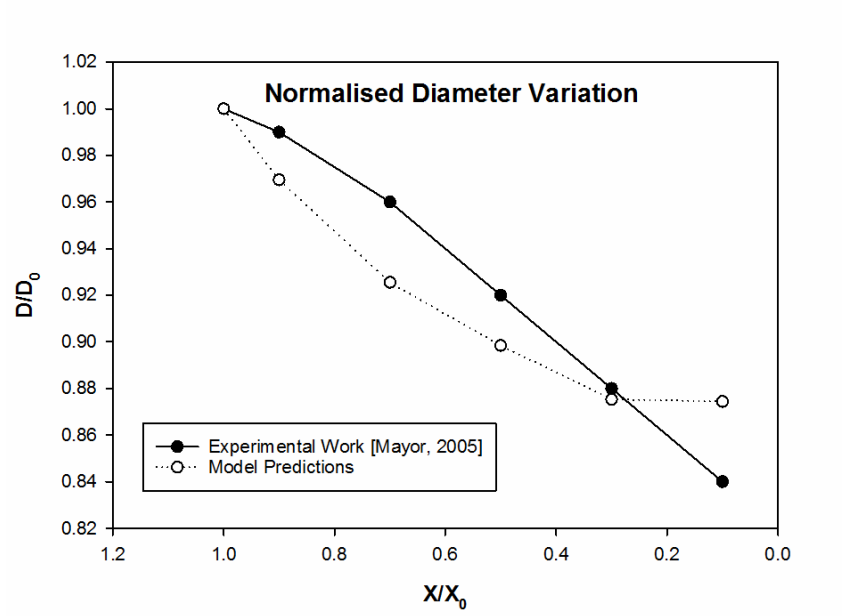


Figure 6. Comparison of model predictions and experimental results [3] for normalised cell diameter of a single apple cell during drying

Conclusion and Outlook

A 3-D plant cell model has been developed using a coupled SPH-DEM numerical method in order to predict the shrinkage characteristics during drying. The model composed of two major parts: cell fluid model and cell wall model. The cell fluid model is based on SPH and approximates the cell protoplasm to a homogeneous Newtonian liquid. The cell wall model is based on a DEM and approximates the real cell wall to an incompressible Neo-Hookean solid material. The drying of single cells of apple, carrot and grape were modelled and simulated for the drying related deformations. Cell shape parameters such as surface area, diameter and perimeter were used to quantify the cell shape alterations. The quantitative shrinkage characteristics were compared with the results from relevant experiments on similar type of plant food materials. Comparisons show that there are similar trends in experimental results and the model predictions, even though there are deviations particularly at very low moisture contents values (extremely dried states of the cells). The reasoning behind such differences have been discussed.

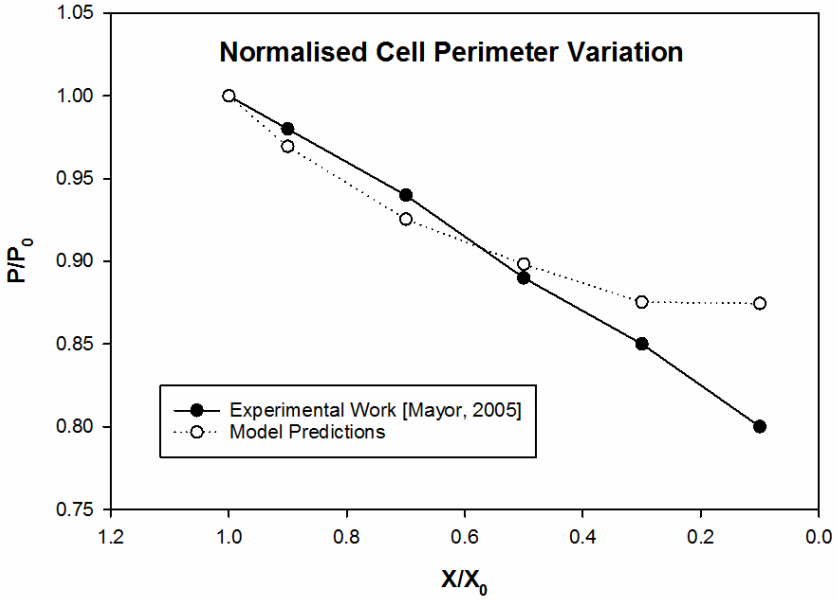


Figure 7. Comparison of model predictions and experimental results [3] for normalised cell perimeter of a single apple cell during drying

It could be noticed that the fundamental capabilities of the adopted numerical modelling technique can effectively handle large deformations of a multiphase cellular system in a comprehensive manner, particularly addressing the 3-D details. Moreover, it has been discussed that there is room for improvements, which can make the model predictions more realistic. This study has the potential to be extended to the level of multi-cell systems by using the developed single cell 3-D model as a fundamental building block.

Acknowledgements

The authors of this study kindly acknowledge the High Performance Computing (HPC) facilities of Queensland University of Technology (QUT), Brisbane, Queensland, Australia; the financial support provided by the Chemistry, Physics and Mechanical Engineering (CPME) scholarship provided by the Science and Engineering Faculty (SEF), QUT; and the

first author specifically extends the sincere support provided by University of Moratuwa, Sri Lanka.

References

1. Jangam, S.V., *An overview of recent developments and some R&D challenges related to drying of foods.* Drying Technology, 2011. **29**(12): p. 1343-1357.
2. Stefan, G., S.R. Hosahalli, and M. Michele, *Drying of Fruits, Vegetables, and Spices*, in *Handbook of Postharvest Technology*. 2003, CRC Press. p. 653-695.
3. Mayor, L., M. Silva, and A. Sereno, *Microstructural changes during drying of apple slices.* Drying technology, 2005. **23**(9-11): p. 2261-2276.
4. Lozano, J.E., E. Rotstein, and M.J. Urbicain, *TOTAL POROSITY AND OPEN-PORE POROSITY IN THE DRYING OF FRUITS.* Journal of Food Science, 1980. **45**(5): p. 1403-1407.
5. Ramos, I.N., et al., *Quantification of microstructural changes during first stage air drying of grape tissue.* Journal of Food Engineering, 2004. **62**(2): p. 159-164.
6. Hills, B.P. and B. Remigereau, *NMR studies of changes in subcellular water compartmentation in parenchyma apple tissue during drying and freezing.* International journal of food science & technology, 1997. **32**(1): p. 51-61.
7. Lee, C.Y., D.K. Salunkhe, and F.S. Nury, *Some chemical and histological changes in dehydrated apple.* Journal of the Science of Food and Agriculture, 1967. **18**(3): p. 89-93.
8. Lewicki, P.P. and G. Pawlak, *Effect of Drying on Microstructure of Plant Tissue.* Drying Technology, 2003. **21**(4): p. 657-683.
9. Ratti, C., *Hot air and freeze-drying of high-value foods: a review.* Journal of food engineering, 2001. **49**(4): p. 311-319.
10. Bartlett, M.K., C. Scoffoni, and L. Sack, *The determinants of leaf turgor loss point and prediction of drought tolerance of species and biomes: a global meta-analysis.* Ecology Letters, 2012. **15**(5): p. 393-405.
11. Crapiste, G.H., S. Whitaker, and E. Rotstein, *Drying of cellular material—I. A mass transfer theory.* Chemical Engineering Science, 1988. **43**(11): p. 2919-2928.
12. Zogzas, N., Z. Maroulis, and D. Marinos-Kouris, *Densities, shrinkage and porosity of some vegetables during air drying.* Drying Technology, 1994. **12**(7): p. 1653-1666.
13. Karathanos, V., G. Villalobos, and G. Saravacos, *Comparison of two methods of estimation of the effective moisture diffusivity from drying data.* Journal of Food Science, 1990. **55**(1): p. 218-223.
14. Karunasena, H.C.P., et al., *A coupled SPH-DEM model for micro-scale structural deformations of plant cells during drying.* Applied Mathematical Modelling, 2014. **38**(15–16): p. 3781-3801.
15. Karunasena, H.C.P., et al., *A particle based model to simulate microscale morphological changes of plant tissues during drying.* Soft Matter, 2014.
16. Karunasena, H.C.P., et al., *Numerical investigation of plant tissue porosity and its influence on cellular level shrinkage during drying.* Biosystems Engineering, 2015. **132**(0): p. 71-87.
17. Nilsson, S.B., C.H. Hertz, and S. Falk, *On the Relation between Turgor Pressure and Tissue Rigidity. II.* Physiologia Plantarum, 1958. **11**(4): p. 818-837.
18. Liu, G.-R. and M. Liu, *Smoothed particle hydrodynamics: a meshfree particle method*. 2003: World Scientific.
19. Van Liedekerke, P., et al., *Particle-based model to simulate the micromechanics of biological cells.* Physical Review E, 2010. **81**(6): p. 061906.
20. Van Liedekerke, P., et al., *A particle-based model to simulate the micromechanics of single-plant parenchyma cells and aggregates.* Physical biology, 2010. **7**(2): p. 026006.
21. Karunasena, H.C.P., et al., *Application of meshfree methods to numerically simulate microscale deformations of different plant food materials during drying.* Journal of Food Engineering, 2015. **146**(0): p. 209-226.
22. Karunasena, H.C.P., et al., *A particle based model to simulate microscale morphological changes of plant tissues during drying.* Soft Matter, 2014. **10**(29): p. 5249-5268.
23. Liedekerke, P.V., et al., *A particle based model to simulate plant cells dynamics*, in *4th international SPHERIC workshop*. 2009: Nantes, France.
24. Liedekerke, P.V., et al., *A particle-based model to simulate the micromechanics of single-plant parenchyma cells and aggregates.* Physical Biology, 2010. **7**(2): p. 026006.

25. Van Liedekerke, P., et al., *Mechanisms of soft cellular tissue bruising. A particle based simulation approach.* Soft Matter, 2011. **7**(7): p. 3580-3591.
26. Karunasena, H.C.P., et al., *Simulation of plant cell shrinkage during drying – A SPH–DEM approach.* Engineering Analysis with Boundary Elements, 2014. **44**(0): p. 1-18.
27. Helambage, C.P.K., et al., *A meshfree model for plant tissue deformations during drying.* ANZIAM Journal, 2014. **55**: p. C110-C137.
28. Taiz, L. and E. Zeiger, *Plant physiology*. New York: Sinauer, 2002.
29. Alexander, S., *Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool.* Modelling and Simulation in Materials Science and Engineering, 2010. **18**(1): p. 015012.
30. Moreira, R., A. Figueiredo, and A. Sereno, *Shrinkage of apple disks during drying by warm air convection and freeze drying.* Drying Technology, 2000. **18**(1-2): p. 279-294.
31. Mayor, L. and A.M. Sereno, *Modelling shrinkage during convective drying of food materials: a review.* Journal of Food Engineering, 2004. **61**(3): p. 373-386.
32. Ratti, C., *Shrinkage during drying of foodstuffs.* Journal of Food Engineering, 1994. **23**(1): p. 91-105.
33. SUZUKI, K., et al., *Shrinkage in dehydration of root vegetables.* Journal of Food Science, 1976. **41**(5): p. 1189-1193.
34. Karunasena, H., et al., *Numerical Investigation of Case Hardening of Plant Tissue During Drying and Its Influence on the Cellular-Level Shrinkage.* Drying Technology, 2015. **33**(6): p. 713-734.
35. Fanta, S.W., et al., *Microscale modeling of coupled water transport and mechanical deformation of fruit tissue during dehydration.* Journal of Food Engineering, 2014. **124**: p. 86-96.

F-bar aided edge-based smoothed finite element methods with 4-node tetrahedral elements for static large deformation hyperelastic and elastoplastic problems

Yuki Onishi[1,a)], Ryoya Iida[1] and Kenji Amaya[1]

[1] Department of Systems and Control Engineering, Tokyo Institute of Technology, Japan

[a)]Corresponding and Presenting author: onishi.y.ad@m.titech.ac.jp

ABSTRACT

A new type of smoothed finite element method, F-barES-FEM-T4, is demonstrated in static large deformation hyperelastic and elastoplastic cases. F-barES-FEM-T4 combines NS-FEM-T4 and ES-FEM-T4 with the aid of F-bar method in order to resolve all the major issues of Selective ES/NS-FEM-T4: limitation of material models, pressure oscillation, and corner locking. As well as other S-FEMs, F-barES-FEM-T4 inherits displacement-based formulation and thus has no increase in DOF. Moreover, the cyclic smoothing procedure introduced in F-barES-FEM-T4 is effective to adjust the smoothing level so that pressure oscillation is suppressed reasonably. A few examples of analyses for rubber-like hyperelastic and elastoplastic materials proof the excellent performance of F-barES-FEM-T4 in contrast to the conventional hybrid elements.

Keywords: Smoothed finite element method, F-bar method, Large deformation, Cyclic smoothing, Pressure oscillation, Locking-free.

Introduction

In the practical use of the finite element method (FEM) for complex shapes, analyses with tetrahedral meshes are indispensable. However, the standard 4-node linear (constant strain) tetrahedral (T4) element has many accuracy issues such as shear locking. Especially when the incompressibility arises in rubber-like or plastic materials, it also suffers from volumetric locking and pressure oscillation issues. Due to the poor performance of the standard T4 element, there have been many researches on the advanced FE formulations of tetrahedral elements.

The hybrid 10-node quadratic (2nd-order) tetrahedral (T10) elements [1] generally represent good results; however, they have accuracy and convergence problems in severe large deformation analysis or contact analysis because of the presence of intermediate nodes. The hybrid T4 element [1] is also used late years but has accuracy issues [5, 6] and brings significant increase in the degree of freedom (DOF) as well. An alternative approach to this problem is the smoothed finite elements methods (S-FEMs) [3]. Selective ES/NS-FEM-T4 [3, 4] would be one of the current best S-FEM-T4 formulations; yet, it still has three major issues: limitation of material models, pressure oscillation, and corner locking [5]. Recently, we proposed a new type of S-FEM-T4 formulation called F-bar aided edge-based smoothed finite element method (F-barES-FEM-T4) [6]. As the adoption of the F-bar method [2] to combine NS-FEM-T4 and ES-FEM-T4 [3], F-barES-FEM-T4 is able to resolve all the major issues of Selective ES/NS-FEM-T4.

In this study, the effectiveness of F-barES-FEM-T4 in static large deformation analyses is demonstrated not only in rubber-like hyperelastic cases but also in elastoplastic cases. Plastic deformation in progress generally decreases the shear modulus drastically and thus presents near incompressibility, thereby inducing volumetric locking and pressure oscillation frequently. A few examples of analyses show that F-barES-FEM-T4 is locking-free and pressure oscillation-free in elastoplastic analyses as well as in nearly incompressible hyperelastic analyses.

Methods

The presenting method, F-barES-FEM-T4, takes advantages of ES-FEM-T4 and NS-FEM-T4 by combining them with F-bar method [2]. The conceptual illustration of F-barES-FEM-T4 is shown in Fig. 1. In F-barES-FEM-T4, the isovolumetric part of the deformation gradient ($\boldsymbol{F}^{\mathrm{iso}}$) is evaluated by using ES-FEM-T4, whereas the volumetric part ($\boldsymbol{F}^{\mathrm{vol}}$) is evaluated

by using NS-FEM-T4 multiply. Combining $\boldsymbol{F}^{\mathrm{iso}}$ and $\boldsymbol{F}^{\mathrm{vol}}$ with F-bar method, the final deformation gradient $\boldsymbol{F}$ is given at edges in the same manner as ES-FEM-T4.

A brief explanation of F-barES-FEM-T4 is described later in this section. See reference [6] for the detail.

Calculation of ${}^{Edge}\widetilde{\boldsymbol{F}}^{iso}$

The isovolumetric part of the deformation gradient at each edge, ${}^{\mathrm{Edge}}\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$, is given in the same manner as ES-FEM-T4.

$$ {}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}}^{\mathrm{iso}} = \left(\frac{1}{{}^{\mathrm{Edge}}_{h}\widetilde{J}} \right)^{1/3} {}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}}; \tag{1} $$

$$ {}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}} = \frac{1}{{}^{\mathrm{Edge}}_{h}V^{\mathrm{ini}}} \sum_{e \in {}^{\mathrm{Edge}}_{h}\widetilde{\mathbb{E}}} {}^{\mathrm{Elem}}_{e}\boldsymbol{F}\ {}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}/6, \tag{2} $$

$$ {}^{\mathrm{Edge}}_{h}\widetilde{J} = \det({}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}}), \tag{3} $$

where ${}^{\mathrm{Edge}}_{h}\widetilde{\mathbb{E}}$ is the set of elements attached to edge h, ${}^{\mathrm{Edge}}_{h}V^{\mathrm{ini}}$ and ${}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}$ are the initial corresponding volume of edge h and element e, respectively.

Calculation of ${}^{Edge}\overline{\boldsymbol{F}}^{vol}$

On the other hand, the volumetric part of the deformation gradient at each edge, ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{vol}}$, is given by the cyclic smoothing procedure as follows.

i. Calculate ${}^{\mathrm{Elem}}\boldsymbol{F}$ and ${}^{\mathrm{Elem}}J$ at each element in the same manner as the standard FEM-T4:

$$ {}^{\mathrm{Elem}}_{e}F_{ij} = {}^{\mathrm{Elem}}_{e}N^{\mathrm{ini}}_{P,j}\ x_{P:i}, \tag{4} $$

$$ {}^{\mathrm{Elem}}_{e}J = \det({}^{\mathrm{Elem}}_{e}\boldsymbol{F}), \tag{5} $$

where ${}^{\mathrm{Elem}}_{e}N^{\mathrm{ini}}_{P,j}$ is the spatial derivative of the shape function ${}^{\mathrm{Elem}}_{e}N^{\mathrm{ini}}_{P}$ in the x_j direction and $x_{P:i}$ is the coordinate of node P in the x_i direction.

ii. Calculate the smoothed J at each node, ${}^{\mathrm{Node}}\widetilde{J}$, in the same manner as NS-FEM-T4:

$$ {}^{\mathrm{Node}}_{n}\widetilde{J} = \frac{1}{{}^{\mathrm{Node}}_{n}V^{\mathrm{ini}}} \sum_{e \in {}^{\mathrm{Node}}_{n}\mathbb{E}} {}^{\mathrm{Elem}}_{e}J\ {}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}/4, \tag{6} $$

where ${}^{\mathrm{Node}}_{n}\mathbb{E}$ is the set of elements attached to node n, ${}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}$ is the initial volume of element e, and ${}^{\mathrm{Node}}_{n}V^{\mathrm{ini}}$ is the initial corresponding volume of node n given by $\sum_{e \in {}^{\mathrm{Node}}_{n}\mathbb{E}} {}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}/4$.

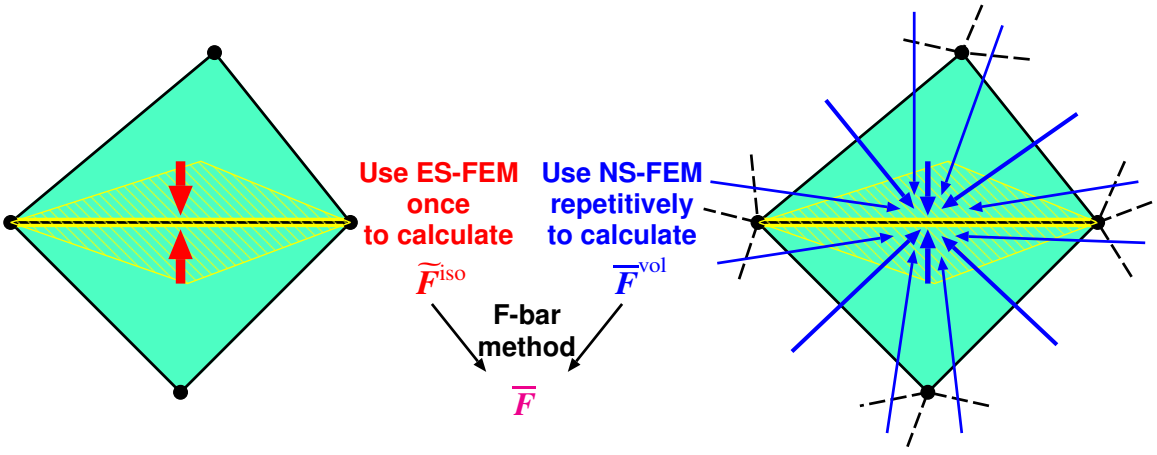


Figure 1. Conceptual illustration of F-barES-FEM.

iii. Calculate the smoothed J at each element, ${}^{\mathrm{Elem}}\widetilde{J}$, as follows:

$$ {}^{\mathrm{Elem}}_{e}\widetilde{J} = \frac{1}{4} \sum_{n \in {}^{\mathrm{Elem}}_{e}\mathbb{N}} {}^{\mathrm{Node}}_{n}\widetilde{J}, \tag{7} $$

where ${}^{\mathrm{Elem}}_{e}\mathbb{N}$ is the set of four nodes comprising element e.

iv. Repeat ii. and iii. c times and obtain the multiply smoothed J at each element, ${}^{\mathrm{Elem}}\overline{J}$. Note that ${}^{\mathrm{Elem}}_{e}J$ is regarded as ${}^{\mathrm{Elem}}_{e}\widetilde{J}$ in the second or later evaluation of Eq. (6). Also, ${}^{\mathrm{Elem}}_{e}\widetilde{J}$ is regarded as ${}^{\mathrm{Elem}}_{e}\overline{J}$ in the last evaluation of Eq. (7).

v. Calculate the multiply smoothed J at each edge, ${}^{\mathrm{Edge}}\overline{J}$, in a similar fashion as ES-FEM-T4:

$$ {}^{\mathrm{Edge}}_{h}\overline{J} = \frac{1}{{}^{\mathrm{Edge}}_{h}V^{\mathrm{ini}}} \sum_{e \in {}^{\mathrm{Edge}}_{h}\mathbb{E}} {}^{\mathrm{Elem}}_{e}\overline{J}\, {}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}/6, \tag{8} $$

where ${}^{\mathrm{Edge}}_{h}\mathbb{E}$ is the set of elements attached to edge h and ${}^{\mathrm{Edge}}_{h}V^{\mathrm{ini}}$ is the initial corresponding volume of edge h given by $\sum_{e \in {}^{\mathrm{Edge}}_{h}\mathbb{E}} {}^{\mathrm{Elem}}_{e}V^{\mathrm{ini}}/6$.

vi. Calculate the multiply smoothed $\boldsymbol{F}^{\mathrm{vol}}$ at each edge, ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{vol}}$:

$$ {}^{\mathrm{Edge}}_{h}\overline{\boldsymbol{F}}^{\mathrm{vol}} = {}^{\mathrm{Edge}}_{h}\overline{J}^{\,1/3}\, \boldsymbol{I}. \tag{9} $$

where $\boldsymbol{I}$ is the unit tensor.

Note that Eq. (6), (7) and (8) satisfy the partition of unity condition and thus the near incompressibility of rubber-like materials is satisfied at the multi-smoothing domain of each edge.

The number of cyclic smoothing, c, is the tuning parameter of F-barES-FEM-T4. F-barES-FEM-T4 with c-time cyclic smoothing is referred to as "F-barES-FEM-T4(c)" hereafter in this paper.

Calculation of ${}^{Edge}\overline{\boldsymbol{F}}$

The final deformation gradient at each edge, ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$, is obtained by combining ${}^{\mathrm{Edge}}\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$ of Eq. (1) and ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{vol}}$ of Eq. (9) with F-bar method.

$$ {}^{\mathrm{Edge}}_{h}\overline{\boldsymbol{F}} = {}^{\mathrm{Edge}}_{h}\overline{\boldsymbol{F}}^{\mathrm{vol}} \cdot {}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}}^{\mathrm{iso}}. \tag{10} $$

Calculation of ${}^{Edge}\boldsymbol{T}$

The Cauchy stress at each edge, ${}^{\mathrm{Edge}}\boldsymbol{T}$, is then derived in the standard way with ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$. In case of history-dependent materials such as elastoplastic materials, ${}^{\mathrm{Edge}}\boldsymbol{T}$ is derived with the history of ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$.

Calculation of ${}^{Edge}\boldsymbol{f}^{int}$

The contribution of each edge to the nodal internal force, ${}^{\mathrm{Edge}}\boldsymbol{f}^{\mathrm{int}}$, is calculated in manner of the F-bar method as

$$ {}^{\mathrm{Edge}}_{h}f^{\mathrm{int}}_{P:p} = \frac{\partial {}^{\mathrm{Edge}}_{h}\widetilde{D}_{ij}}{\partial \dot{u}_{P:p}}\, {}^{\mathrm{Edge}}_{h}T_{pl}\, {}^{\mathrm{Edge}}_{h}V. \tag{11} $$

Note that the stretching tensor in this equation, ${}^{\mathrm{Edge}}\widetilde{\boldsymbol{D}}$, is not the deformation rate of ${}^{\mathrm{Edge}}_{h}\overline{\boldsymbol{F}}$ in Eq. (10) but that of ${}^{\mathrm{Edge}}_{h}\widetilde{\boldsymbol{F}}$ in Eq. (2).

Results

Barreling of Hyperelastic Cylinder

A hyperelastic large deformation analysis of a 1/8 cylinder with enforced displacements is performed. Figure 2 illustrates the outline of the analysis. Barreling deformation grows as the enforced displacement progresses, and then the lateral

surface is squeezed out. The material constitutive model of the cylinder is the neo-Hookean hyperelastic model, $\boldsymbol{T} = 2C_{10}\frac{\mathrm{Dev}(\bar{\boldsymbol{B}})}{J} + \frac{2}{D_1}(J-1)\boldsymbol{I}$, where $C_{10} = 4\times10^7$ Pa and $D_1 = 5\times10^{-11}$ Pa^{-1} and thus the initial Poisson's ratio is 0.499. The mesh seed size is 0.05(= 1/20) m constant for 1st-order elements and is 0.1(= 1/10) m constant for 2nd-order elements.

Firstly, results of 4-node hybrid tetrahedral element of ABAQUS/Standard (ABAQUS C3D4H), 10-node quadratic modified hybrid tetrahedral element (ABAQUS C3D10MH), and 8-node hybrid hexahedral element (ABAQUS C3D8H) are shown in Figs. 3–5. ABAQUS C3D4H is free from shear and volumetric locking; however, it has two major issues: pressure oscillation and corner locking [5]. The corner locking is a type of locking that brings a strangely hard deformation around corners in large deformation cases. ABAQUS C3D10MH is free from shear, volumetric, and corner locking; how-

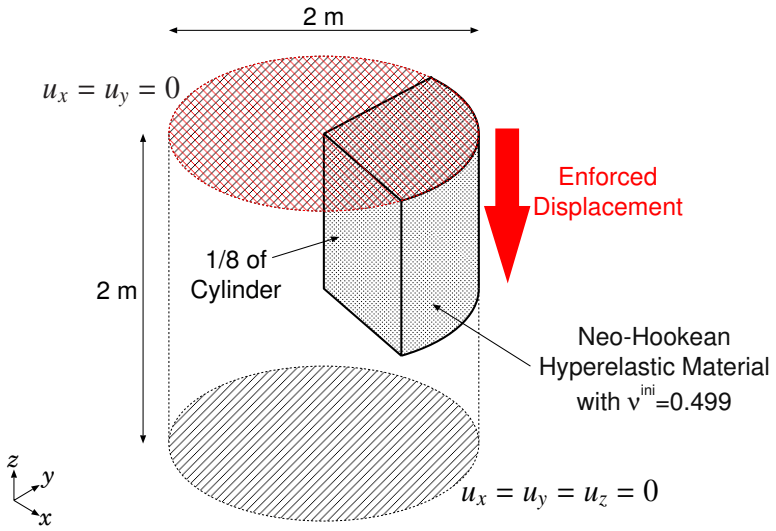


Figure 2. Outline of the hyperelastic barreling analysis.

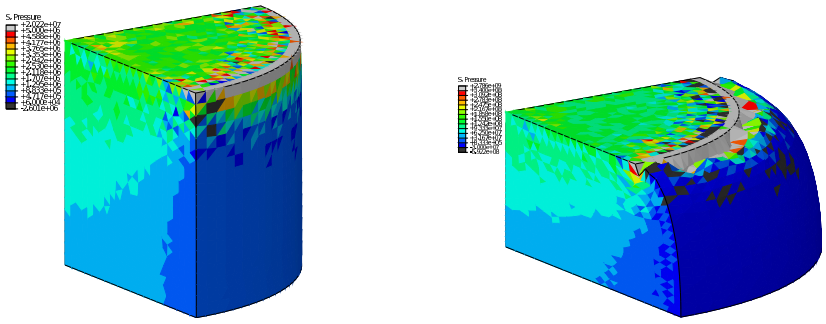


Figure 3. Pressure distributions of ABAQUS C3D4H results. Left: $u_z = 0.01$ m. Right: $u_z = 0.40$ m.

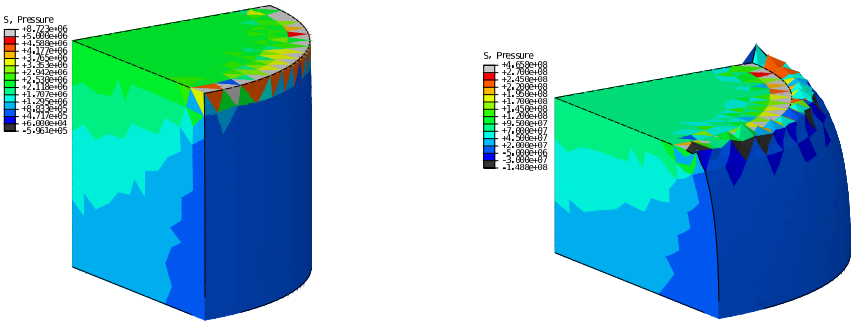


Figure 4. Pressure distributions of ABAQUS C3D10MH results. Left: $u_z = 0.01$ m. Right: $u_z = 0.32$ m.

ever, it suffers from convergence failure in a relatively earlier stage. Moreover, the presence of intermediate nodes causes accuracy loss of interpolation in large deformation cases. ABAQUS C3D8H is also free from shear, volumetric, and corner locking; however, it suffers from pressure oscillation.

Secondly, results of Selective ES/NS-FEM-T4, F-barES-FEM-T4(1), (2), (3) and (4) are shown in Figs. 6–10. Selective ES/NS-FEM-T4 and all F-barES-FEM-T4s are free from shear and volumetric locking and have no convergence problem. Selective ES/NS-FEM-T4 and F-barES-FEM-T4(1) have pressure oscillation and corner locking issues, whereas F-barES-FEM-T4(2) or later suppresses these issues. It should be noted that F-barES-FEM-T4(2) or later are not much different each other and thus c is not much sensitive to the result. Therefore, F-barES-FEM-T4 with a sufficient cycles of smoothing c resolves all the accuracy issues of conventional methods.

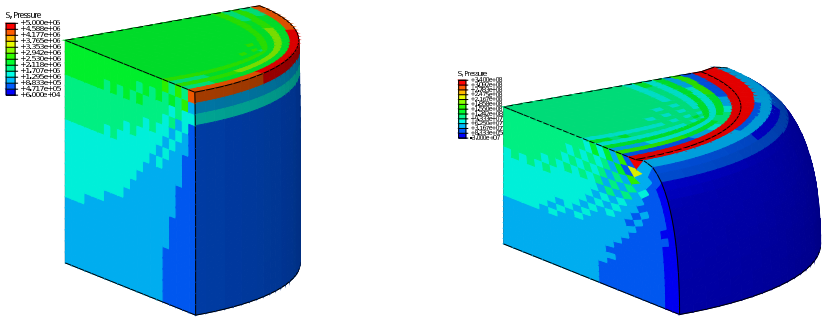


Figure 5. Pressure distributions of ABAQUS C3D8H results. Left: $u_z = 0.01$ m. Right: $u_z = 0.40$ m.

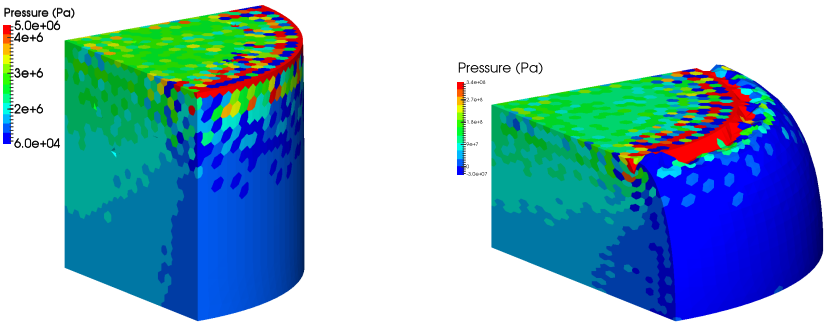


Figure 6. Pressure distributions of Selective ES/NS-FEM-T4 results. Left: $u_z = 0.01$ m. Right: $u_z = 0.40$ m.

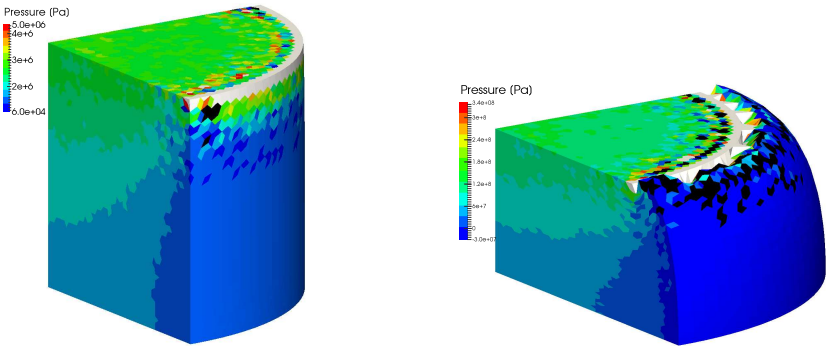


Figure 7. Pressure distributions of F-barES-FEM-T4(1) results. Left: $u_z = 0.01$ m. Right: $u_z = 0.40$ m.

Shear-Tensioning of Elastoplastic Bar

An elastoplastic large deformation analysis of a bar with enforced displacements is performed. Figure 11 illustrates the outline of the analysis. Shear deformation dominates at the middle part of the bar in the early stage of the analysis, whereas stretch deformation dominates in the later stage. The material constitutive model of the bar is the elastoplastic model with Hencky's strain measure, von Mises yield criterion, and the isotropic hardening flow rule. The material properties are 1 GPa Young's modulus, 0.3 Poisson's ratio, 1 MPa yield stress, and 0.1 GPa constant work hardening rate. Hence, the Poisson's ratio under large plastic deformation in progress is greater than 0.48. The mesh seed size is 0.2(= 1/5) m constant.

Results of ABAQUS C3D4H and F-barES-FEM-T4(2) are shown in Fig. 12 and 13. Figure 12 compares the deformations and distributions of the equivalent plastic strain, while Figure 13 compares those of the pressure. The results of ABAQUS

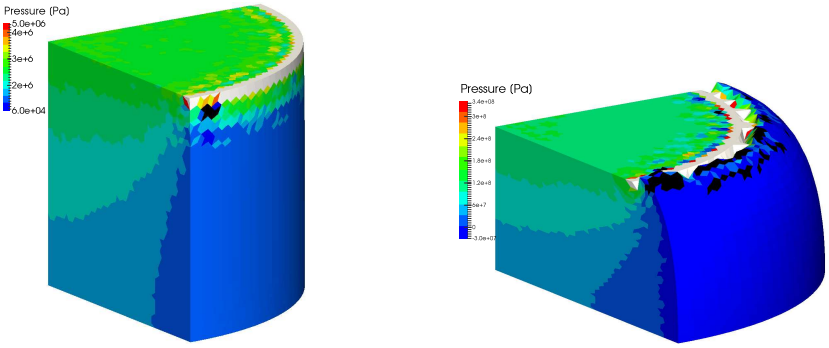


Figure 8. Pressure distributions of F-barES-FEM-T4(2) results. Left: $u_z = 0.01$ **m. Right:** $u_z = 0.40$ **m.**

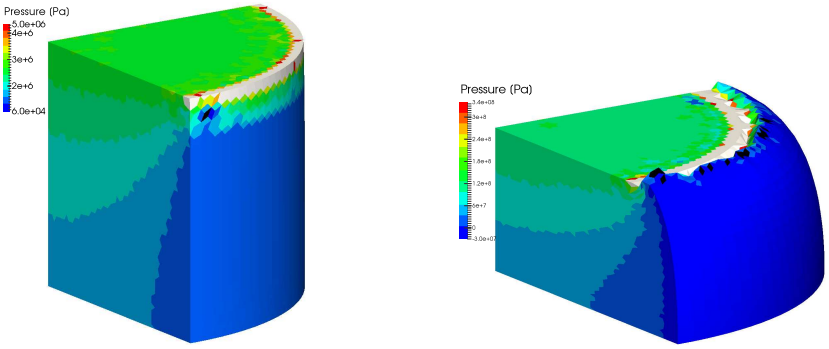


Figure 9. Pressure distributions of F-barES-FEM-T4(3) results. Left: $u_z = 0.01$ **m. Right:** $u_z = 0.40$ **m.**

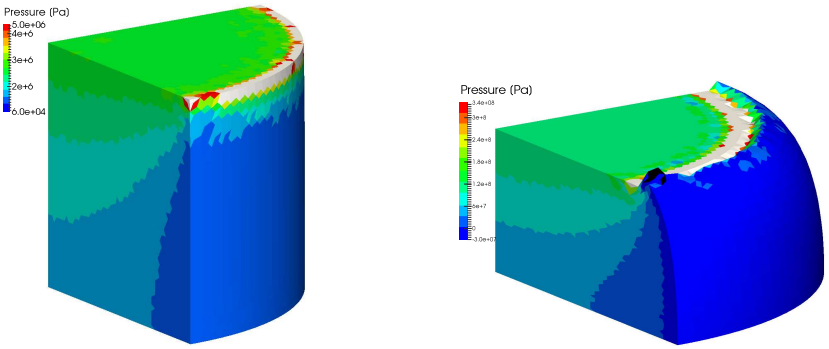


Figure 10. Pressure distributions of F-barES-FEM-T4(4) results. Left: $u_z = 0.01$ **m. Right:** $u_z = 0.40$ **m.**

C3D4H represent strange spatial oscillation on both the equivalent plastic strain and pressure distributions. On the other hand, the results of F-barES-FEM-T4(2) are smooth in the both distributions and thus seem valid. F-barES-FEM-T4 is considered effective not only for rubber-like materials but also for elastoplastic materials.

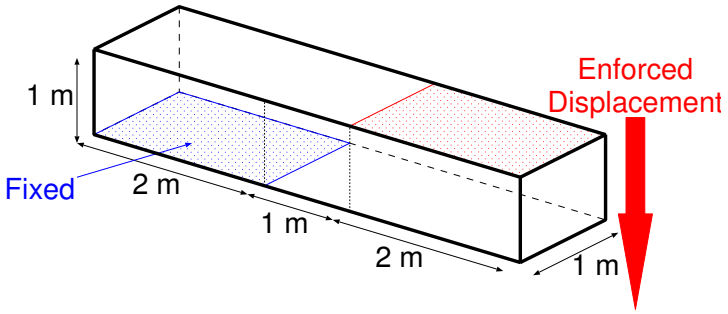
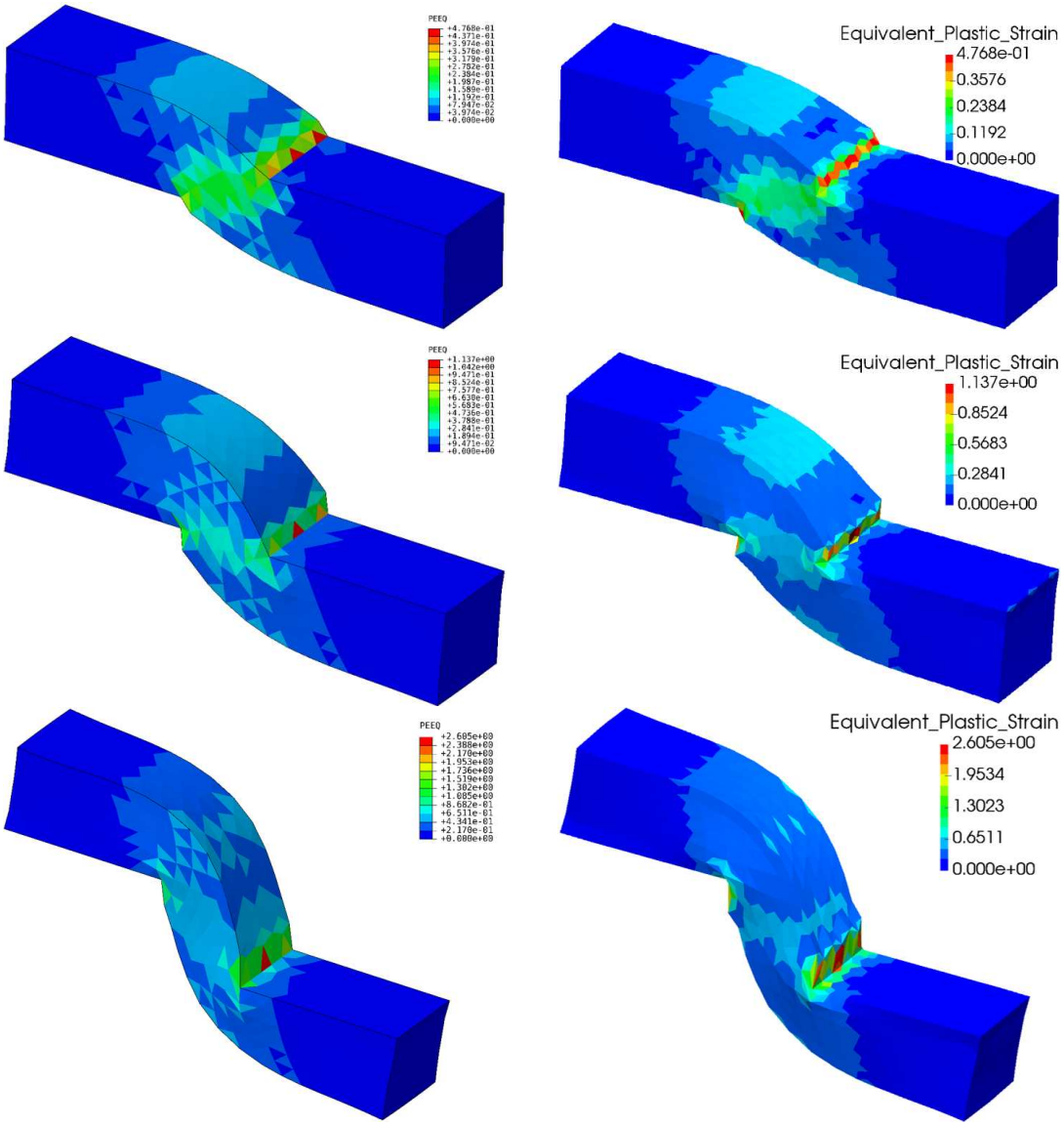


Figure 11. Outline of the elastoplastic shear-tensioning analysis.



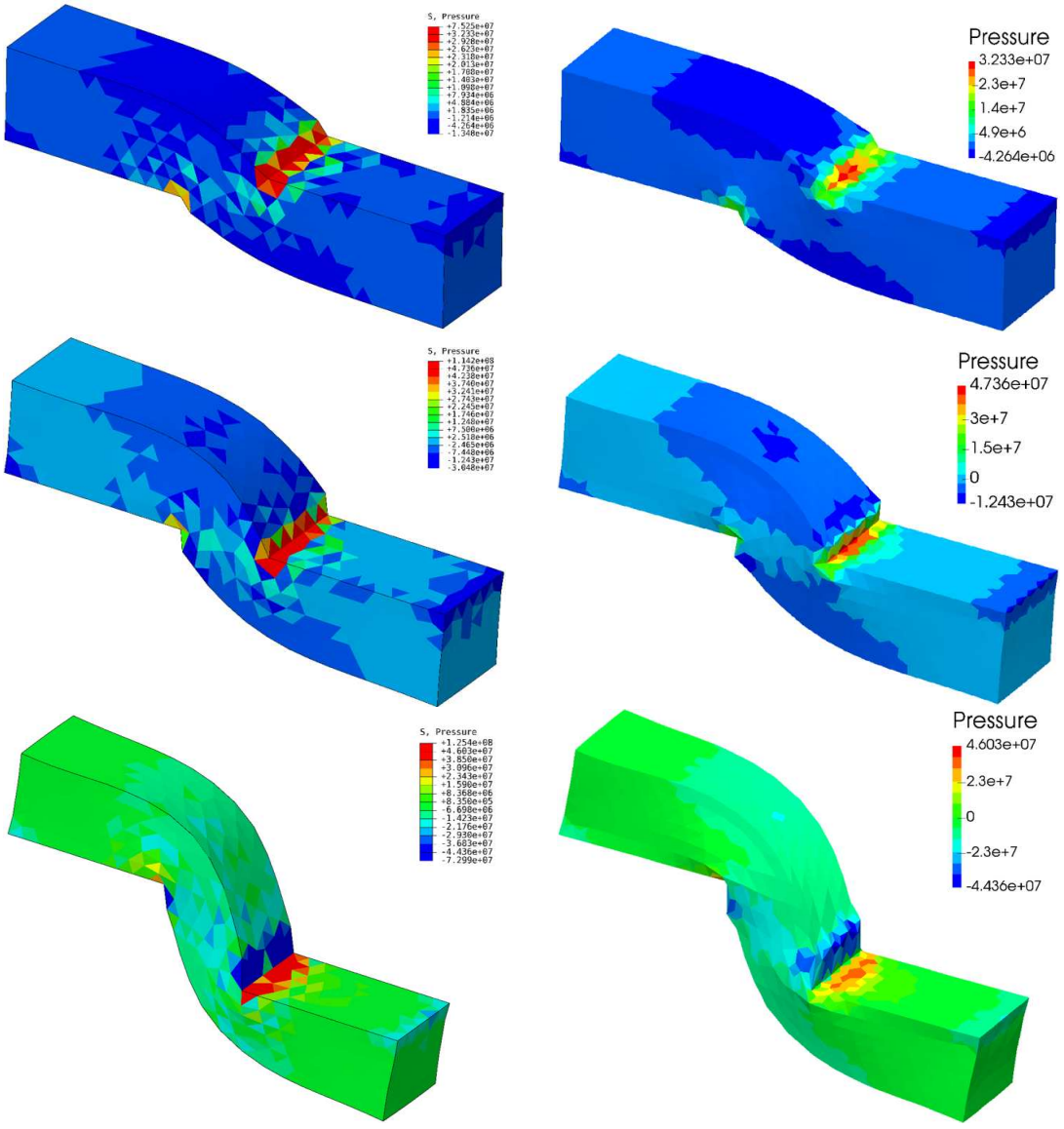
(a) ABAQUS C3D4H (b) F-barES-FEM-T4(2)

Figure 12. Comparison of equivalent plastic strain distributions on the elastoplastic shear-tensioning analysis.

Conclusion

A new type of smoothed finite element method, F-barES-FEM-T4, is demonstrated in static large deformation hyperelastic and elastoplastic problems. The characteristics of F-barES-FEM-T4 are summarized as follows.

- No increase in DOF.
- No limitation of material models.
- No convergence problem in large deformation.
- Free from shear, volumetric, and corner locking.
- Suppress pressure oscillation in rubber-like/elastoplastic materials.
- Adjustable smoothing level with the number of cyclic smoothings (c).



(a) ABAQUS C3D4H (b) F-barES-FEM-T4(2)

Figure 13. Comparison of pressure distributions on the elastoplastic shear-tensioning analysis.

References

[1] Dassault Systèmes Simulia Corp. (2013). *ABAQUS 6.13 Theory Guide*. Dassault Systèmes Simulia Corp., Providence, RI, USA.

[2] de Souza Neto, E., Peric, D., Dutko, M., and Owen, D. (1996). Design of simple low order finite elements for large strain analysis of nearly incompressible solids. *International Journal of Solids and Structures*, 33(20-22):3277–3296.

[3] Liu, G. R. and Nguyen-Thoi, T. (2010). *Smoothed Finite Element Methods*. CRC Press, Boca Raton, FL, USA.

[4] Onishi, Y. and Amaya, K. (2014). A locking-free selective smoothed finite element method using tetrahedral and triangular elements with adaptive mesh rezoning for large deformation problems. *International Journal for Numerical Methods in Engineering*, 99(5):354–371.

[5] Onishi, Y. and Amaya, K. (2015). Performance evaluation of the selective smoothed finite element methods using tetrahedral elements with deviatoric/hydrostatic split in large deformation analysis. *Theoretical and Applied Mechanics Japan*, 63:55–65.

[6] Onishi, Y., Iida, R., and Amaya, K. (under review). F-bar aided edge-based smoothed finite element method using tetrahedral elements for finite deformation analysis of nearly incompressible solids. *International Journal for Numerical Methods in Engineering*.

Finite Element Simulation of the Device CAR1 on Braced Frames

***M.D. Titirla[1]**

[1]Department of Civil Engineering, Aristotle University of Thessaloniki, Greece.

*Presenting and corresponding author: mtitirla@civil.auth.gr

Abstract

The developed device, has the codename CAR1, belongs to the passive energy dissipation systems, as it doesn't require external power to generate system control forces. It can be used on new or existing structures and can be easily adapted to the particular demands of structures. It can be installed in a variety of ways such as in single or X diagonal bracing in building frames. Moreover the use of this device may result in improving (i) the increase of stiffness (ii) the absorption of seismic energy, (iii) as well as control of the axial forces that are developed at the diagonal steel braces. The main part of CAR1 device is the groups of superimposed blades, which absorb seismic energy through simultaneous friction and yield. Firstly this paper discusses the experimental and numerical evaluation of the effectiveness of this steel device. Full scale CAR1 device was experimentally investigated under cyclic loading in Laboratory for Strength of Materials and Structures of Aristotle University of Thessaloniki. Finite Element Models of CAR1 device were developed and analyzed using the software ABAQUS, checking the credible documentation of the device. In addition, a numerically robust finite element model of a whole one storey structure is described, for high-fidelity simulations of inelastic responses of device CAR1 on braced frame. Aim of this study is to compare the response of one storey structure with and without the existence of device CAR1 on diagonal braces.

Keywords: Experimental validation, Finite Element verification, Absorption Seismic Energy, Friction, Dynamic Explicit analysis.

Introduction

The safety of construction (existing or new) is one of the major priorities of engineering globally, because structures often subject to large and often devastating, for their viability, loadings. So, great interest is in the study of the innovations of the design and materials of construction that minimize the probability of failure of the structure in any charging.

Steel concentrically braced frames have been used widely in high-seismic regions due to their efficiency in meeting lateral-load resisting requirements. Based on extensive research since the 1970's, it is well known that the cyclic loading performance of steel braces depend on their slenderness ratio and on the width-to-thickness ratio of their cross sectional elements, and that adequate detailing of the bracing connection is critical to avoid premature fracture at the end of the brace. Braced frame systems are presently being designed to satisfy performance-based seismic design criteria [1, 2]. In terms of analysis capabilities, researchers have proposed methodologies to predict the occurrence of fracture from cumulative damage and to exhibit significant ductility in life safety and collapse prevention limit states, which are governed by inelastic post-buckling and tensile yielding behaviors of the brace elements.

There are essentially two main orientations in order to protect braced elements, first one is to provide the structural system in order to avoid unexpected premature failure modes (mid-

length or connections) and the second is to be incorporated in braces a passive energy dissipation devices.
Buckling restrained braced frames (BRBFs) for seismic load resistance have been widely used in recent years because it yields under both tension and compression without significant buckling [3, 4, 5]. Others researchers create numerical models to approach brace elements. Numerical models can be classified in three categories, the phenomenological models [6, 7, 8], the beam-column Finite elements models [9, 10, 11] and the 3-D Finite elements models [12, 13, 14, 15, 16, 17].
On the other hand, passive energy dissipation devices such as visco-elastic dampers, metallic dampers and friction dampers have widely been used to reduce the dynamic response of civil engineering structures subjected to seismic loads [18, 19, 20] and can easily replaced or repaired. Their effectiveness for seismic design of building structures is attributed to minimizing structural damages by absorbing the structural vibratory energy and by dissipating it through their inherent hysteresis behavior. So, several of these devices have been selected for seismic strengthening of existing or new buildings in the US, Canada and Japan [21, 22, 23].
In order to demonstrate the effectiveness of the devices, many passive energy dissipation systems were studied in experimental research [24, 25, 26, 27] or in numerical research [28, 29, 30]. The Finite Element Method (FEM) has become the most popular method in both research and industrial numerical simulations, as it takes into consideration material laws, contact interface conditions and others parameters, which lead to the exact response of the device. Several algorithms, with different computational costs, are implemented in the finite codes, such as ABAQUS [31], which is commonly used software for finite element analysis. Comparison of numerical results with the same experimental one is very useful and necessary as it provides the possibility to researchers to study the behavior of their devices more widely [32, 33]. The calibrated FEM models are used to conduct a series of simulations to study the effect of different parameters. In this way, results come out that are harder to obtain experimentally.
In the present paper, a numerically robust finite element model is described, which is based on explicit time-stepping, for high-fidelity simulations of inelastic responses of device CAR1 on braced frame. The effectiveness of the investigated device was recently developed at the Laboratory of Strength of Materials and Structures of Aristotle University of Thessaloniki. Aim of this study is to compare the response of one storey structure with and without the existence of device CAR1 on diagonal braces.

Study of the individual device CAR1

A short description:

The developed device has the codename CAR1 and belongs to the passive energy dissipation system, as it doesn't require external power to generate system control forces. This device proposed by Papadopoulos et al. [34] and it consists of 4 main elements, as illustrated in Figure 1. Device CAR1 has the advantage to (i) provide additional stiffness as well as (ii) absorption of seismic energy, through yield and friction, (iii) provision of control of the axial forces that are developed at the diagonal steel rods and last but not least the ability to retain the plastic displacements to a desired level, due to the restrain bolt. Energy dissipation is provided by inelastic bending of superimposed blades.

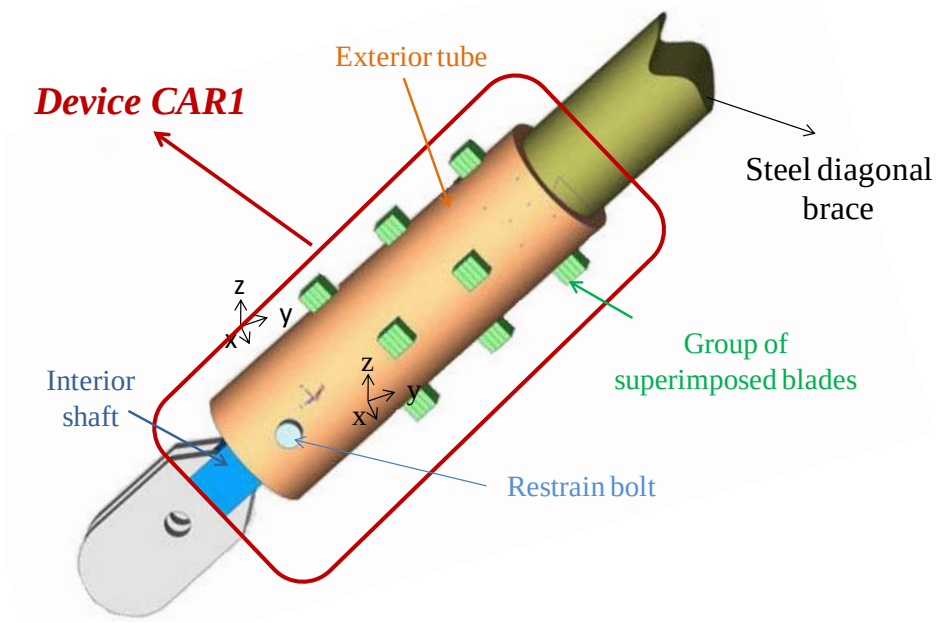


Figure 1: The investigated device CAR1.

Moreover, device CAR1 can be used on new or existing structures and can easily be adapted to the particular demands of structures. However, it can be installed in a variety of ways which include using them in single diagonal braces or in X braces (Figure 2) and in accordance with the requirements of each construction, it can be used one or more devices.

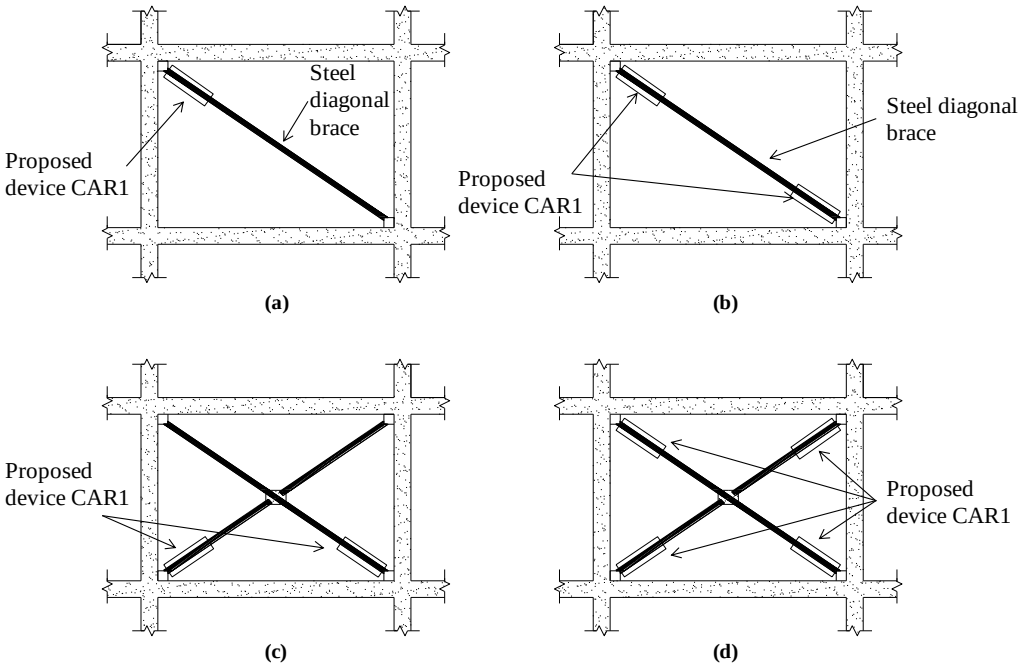


Figure 2: Possible positions of the device CAR1, incorporated in steel diagonal braces.

Experimental set up:

A standard test has been carried out in order to establish the basic material properties of the superimposed blades (Figure 3). These experimentally derived material properties were utilized in the subsequent numerical study.

Young Modulus E (Gpa)	203.6
Yield Strength F_y (Mpa)	220
Maximum Strength (MPa)	290

Figure 3: A standard test in order to establish the basic properties of the superimposed blades.

Full scale CAR1 device was experimentally investigated under cyclic loading. The experimental sequences have been conducted at the Laboratory of Strength of Materials and Structures of Aristotle University of Thessaloniki. The specimen details of the experiment are depicted in Figure 4. The load was controlled with a 100kN capacity load cell under deflection control. Two LVDT's were positioned at each side of the longitudinal axis of the device CAR1, which measure the relative movement of the interior shaft to the exterior tube. All data were recorded and were stored in a digital data system via a computer. We notice that only two group of superimposed blades were tested. Every group consists of five steel blades, each 4mm thick. Quasi-static cyclic tests were carried out in order to ascertain device's CAR1 behavior to absorbed seismic energy. The experimental sequence is 17 cycles displacement control with values starting from 4.5 mm up to 10 mm with rate 3mm/minute.

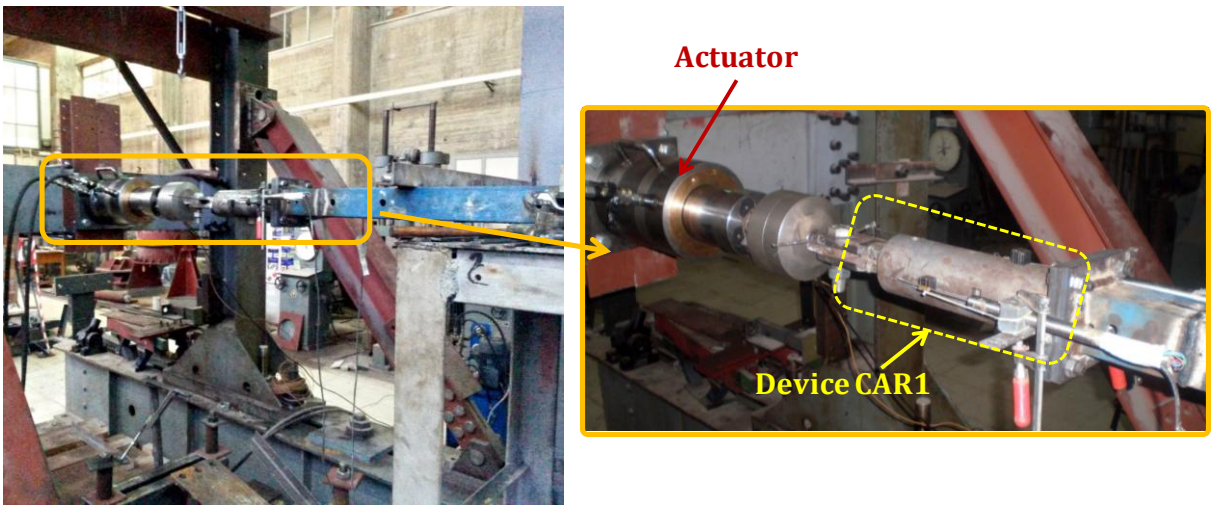


Figure 4: Specimen details.

Finite Element Modeling:

The general purpose FE software ABAQUS was employed to generate FE models to simulate numerically the behavior of the device CAR1. It was selected to use an explicit dynamic solver because this allows the definition of very general contact conditions for complicated contact problems, without generating numerical difficulties. The explicit dynamics analysis procedure is based upon the implementation of an explicit integration rule together with the use of diagonal ("lumped") element mass matrices.

To the comparison with the Standard, the explicit dynamic solver is computationally inefficient for quasi-static problems if real time is used, because the time needed to finish an analysis is proportional to its duration. However, it is often possible to scale the real time to a very small time period if the response of the structure remains basically static. According to classical dynamic theory, when a dynamic system is subjected to a linearly rising load, its response can be approximately treated as static if the duration of the loading stage is large compared to the natural period of the system. For solving this problem, check the ratio of kinetic to internal energy can be used to check if the structure has failed and the analysis is continuing simply as dynamic motion. It is stated in the ABAQUS/Explicit manual [31] that the procedure is quasi-static if the ratio of the kinetic energy to the internal energy is less than 2%. Any responses which have an energy ratio larger than this should be treated as dynamic and removed from the results.

The FEM model geometry reproduced the actual geometry of the tests set up of the device CAR1 to characterize the behavior of the device. The geometry of FE model was reproduced in full detail (Figure 5).

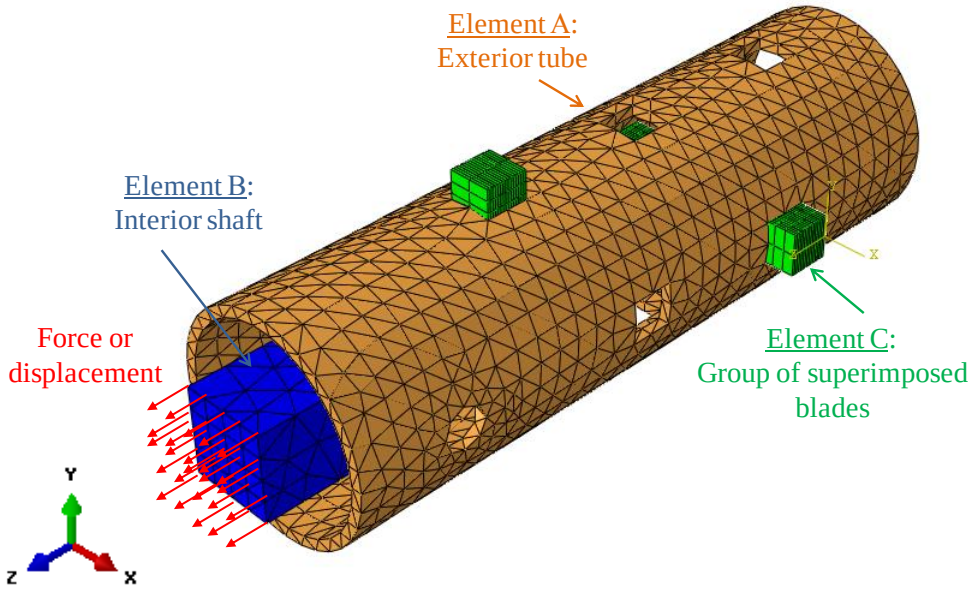


Figure 5: The FEM model used for the device CAR1 in software ABAQUS.

Several simulations were conducted to identify the best meshing. For the explicit method, blades and interior shaft are meshed using 3D reduced integration solid element **C3D8R** (eight-node bricks), while exterior tube is meshed using 3D solid element **C3D4** (four-node tetrahedron) available in ABAQUS. Normally, a higher mesh density provides for higher accuracy but also increases the computational time without improving substantially the accuracy of the results, therefore, a trade-off between time and accuracy becomes crucial [35].

The final mesh has 8126 elements and it resulted in a solution that correlated with the experimental results.
The uniaxial stress–strain relation of the blades, exterior tube and interior shaft are modeled as elastic with Young's modulus (Es) and Poisson's ratio (v) of which typical values are 200 GPa and 0.3, respectively. Plastic behavior are defined in a tabular form, including yield stress and corresponding plastic strain. The experimentally obtained stress (σ_{nom})-strain (ε_{nom}) curves for the blades was converted into the true stress (or Cauchy) (σtrue)-logarithic plastic strain (ε_{ln}^{pl}) format according to Eq. 1 and 2 and utilized to define the material response.

$$\sigma_{true} = \sigma_{eng}\ (1 + \varepsilon_{eng}) \quad (1)$$

$$\varepsilon_{pl} = \varepsilon_{true} - \frac{\sigma_{true}}{\mathrm{E}} \quad (2)$$

The surface-to-surface contact formulation technique with small sliding between the contacting surfaces was chosen. The contact definition includes the specification of two surfaces, one acting as the ‘‘master’’ surface and the other as the ‘‘slave’’ surface. The contact algorithm searches whether the nodes of the slave surface are in contact with the nodes of the master surface and enforces contact conditions in an average sense over a region of slave nodes using a Lagrange multiplier formulation [31]. A friction coefficient equal to 0.2 [36] was assumed between the contacting surfaces. A flowchart for carrying out the FEM analysis procedure is presented in Figure 6.

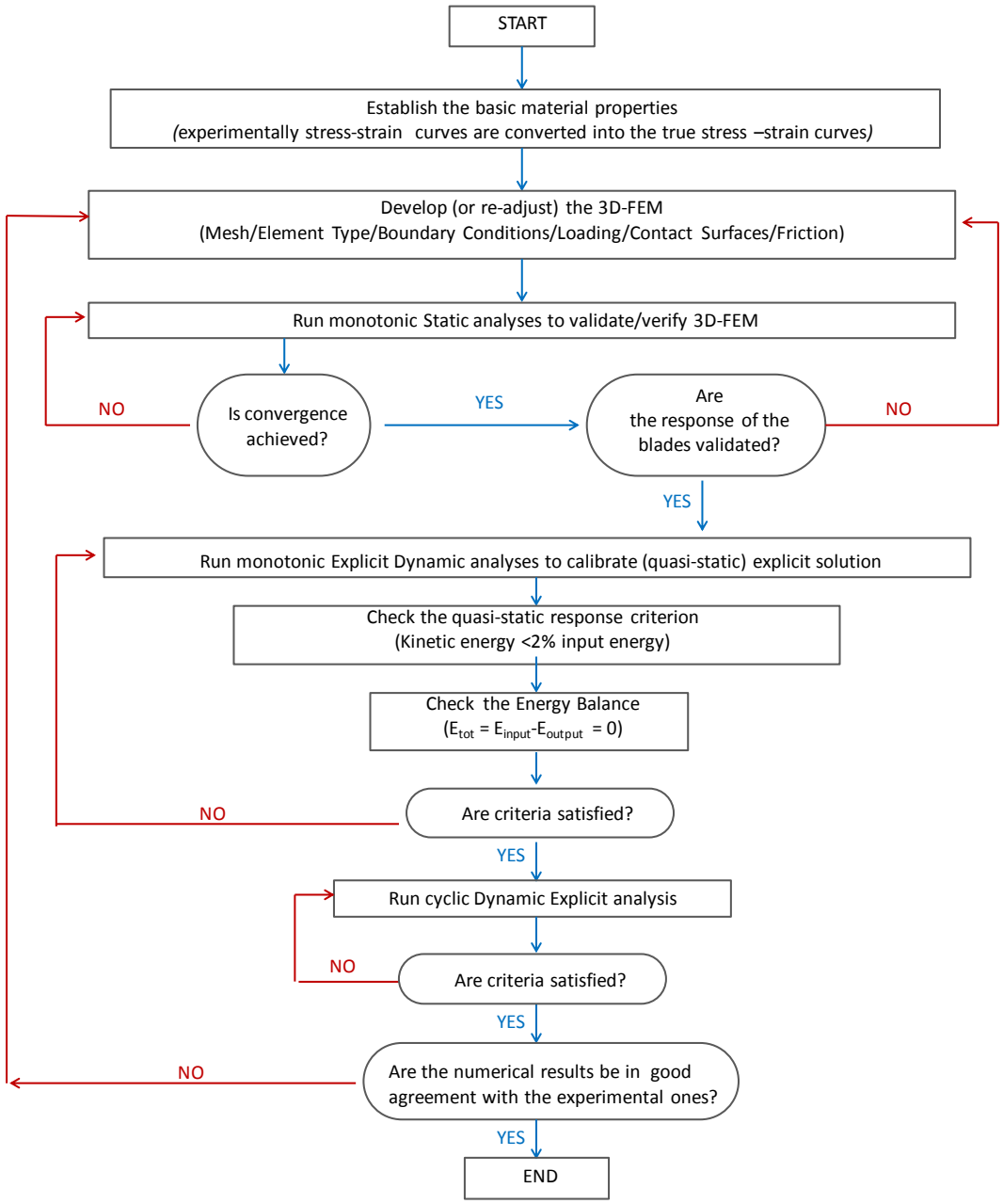


Figure 6: Flowchart in order to develop the Finite Element Model of Device CAR1 in ABAQUS.

Figure 7 plots the force versus relevant displacement from FEM analyses along with the experimental hysteresis. Blue lines illustrate hysteresis loops of experiments, while green lines shows hysteresis loops of Finite Element Models. The predicted values for the load and displacement are in very good agreement with the corresponding experimental ones. The comparisons between the FEM analyses and experiments show that the proposed FEM model is capable of reproducing the inelastic response of the device CAR1. Therefore, it is a reliable tool for the simulation of the hysteretic behavior of the device CAR1 and can be used to contact further studies to investigate the effect of various parameters. In addition, the area within a hysteresis loop is equivalent to the amount of seismic energy that the device is dissipating. Since the shape and consistency of the hysteresis loops, observe the device's ability to absorb CAR1 seismic energy, whereas will not break during the cyclic loading.

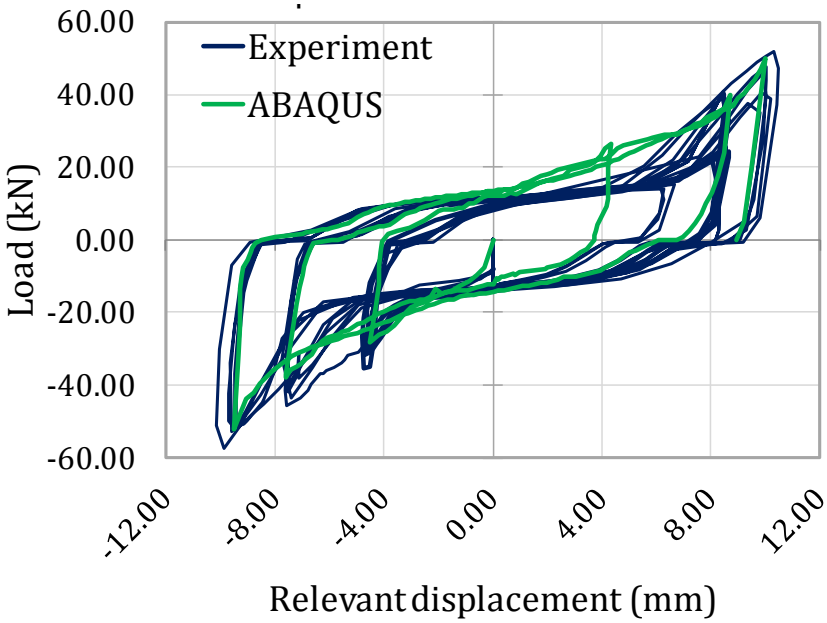


Figure 7: Comparison of the experimental and the numerical force–displacement hysteresis of the device CAR1

In addition, the numerical deformed shapes are compared with the corresponding experimental ones for relevant movement Uz=±5mm in Figure 8.

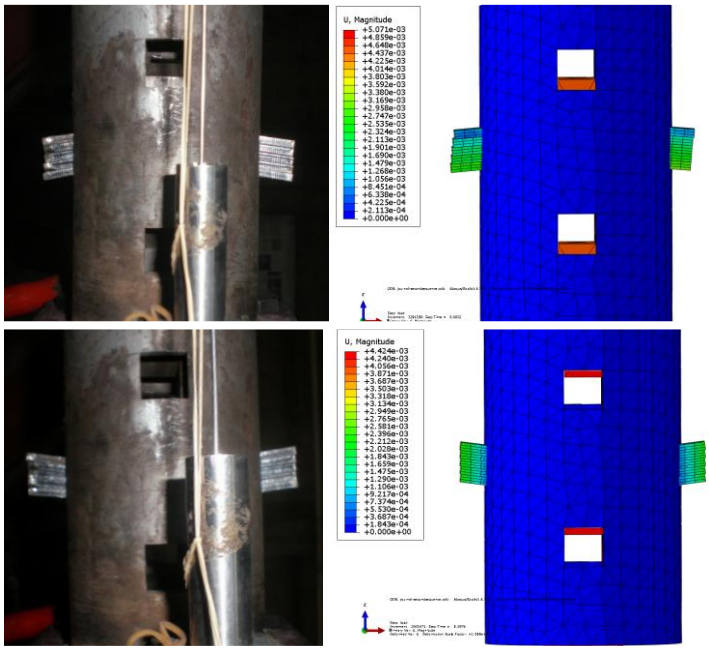


Figure 8: Distribution of deformed shapes (values in m)

Study of one storey structure with and without the existence of device CAR1on diagonal braces

A one storey reinforced concrete structure (Figure 9) was chosen to be studied. It has a height of 3m and length equal to 4.5m. The horizontal elements are beams with dimensions in plan 25x50cm and the vertical are columns with dimensions in plan 35x35cm.

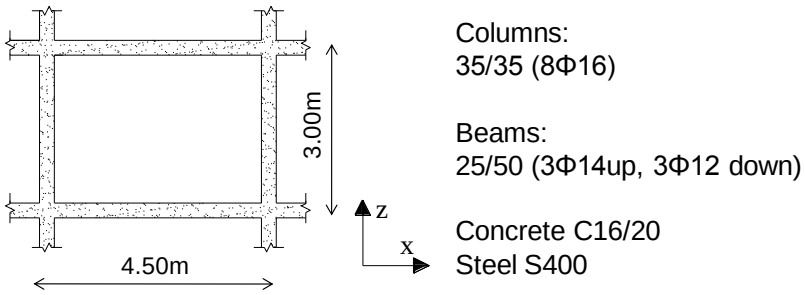


Figure 9: The longitudinal section of structure.

Structure was modeled and analyzed in SAP 2000 ver. 11.0.3 [37] in order to define floor's displacement drifts for seismic performance "Life Safety" (drift=1.6%) and "Collapse" (drift=2.1%). Columns and beams were modeled by frame elements. As it is drawn, the maximum horizontal displacement was chosen equal to 6cm (drift=2.0%), smaller than the collapse displacement (6.3cm). Also, the 2004 NEHRP provisions [38] allow the design of buildings with passive damping systems to experience controlled inelastic deformations associated with typical design drifts limits, e.g. a 2% drift limit. For this horizontal displacement, both braced structures ((i) with diagonal brace and (ii) with diagonal brace and CAR1 device) will compare in software ABAQUS.

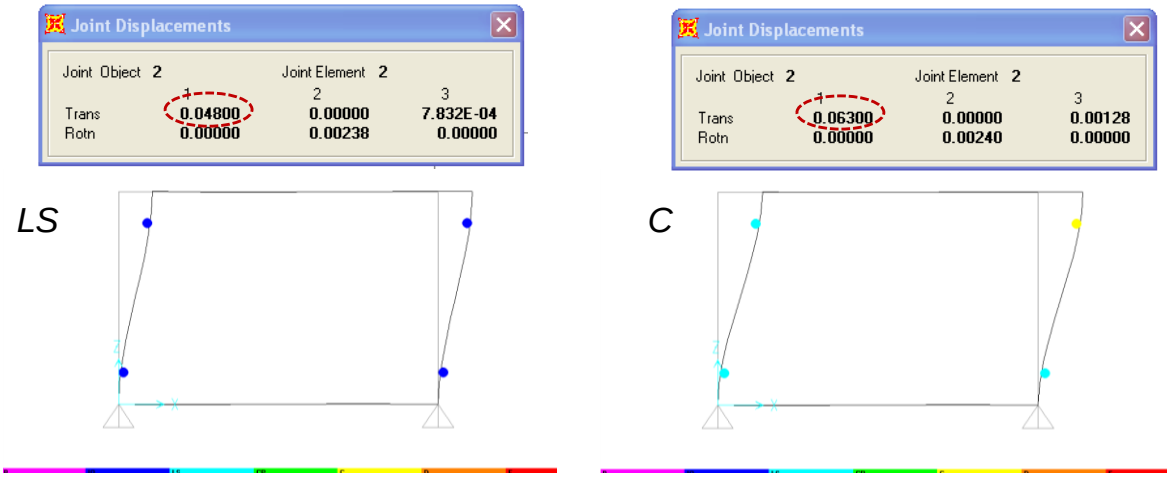
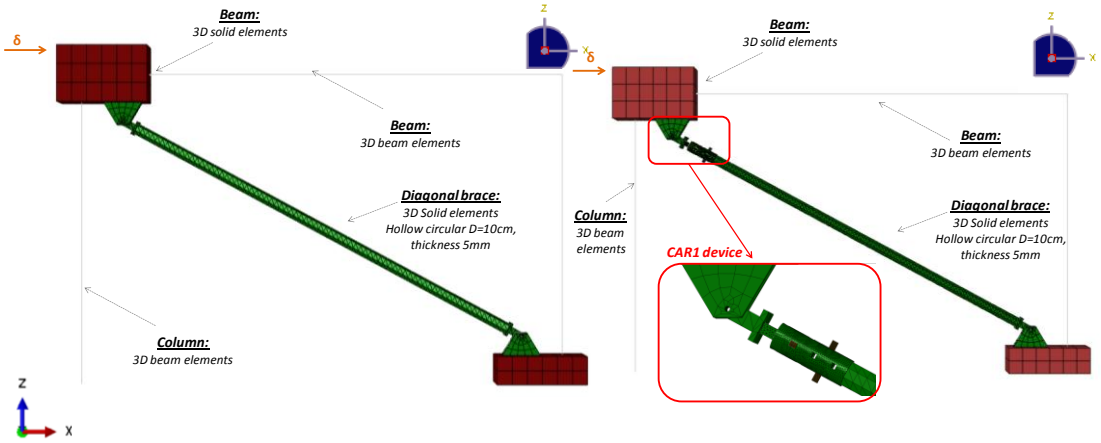


Figure 10: Deformed shape in SAP2000

Both braced structures were model and analyses in software ABAQUS, as it is illustrated in Figure 11. Columns were modeled with 3D beam elements while part of beam, diagonal brace and device CAR1 were modeled with 3D solid elements. The main parameters of modeling are mentioned in section 2.3 and it is not considered necessary to re-commented. Horizontal displacement (δ) imposed at the top of the floor increased step by step until the maximum

displacement of 6cm. Dynamic Explicit analysis were contacted and useful results were observed.



(i) (ii)

Figure 11: (i) Braced Structure, (ii) CAR1-Braced Structure in ABAQUS.

Figure 12 shows the distribution of horizontal displacement around x-x axis at the end of analysis. In Braced Structure, diagonal brace fracture especially at the middle length of diagonal, while in CAR1-Braced Structure the brace remains un-deformed without plastic hinges. Figures 13 and 14 show the peak plastic strains at the end of the analysis. In braced structure, maximum plastic strain observed at the middle length of diagonal brace (fracture point), while in CAR1-Braced Structure at the superimposed blades, which are easily be replaced with minimum cost. As a result, using CAR1 device on diagonal brace, the fracture life of brace is increased. The system exhibited uniform energy absorption with more stability, as strength and maximum deformation of the system increased considerably.

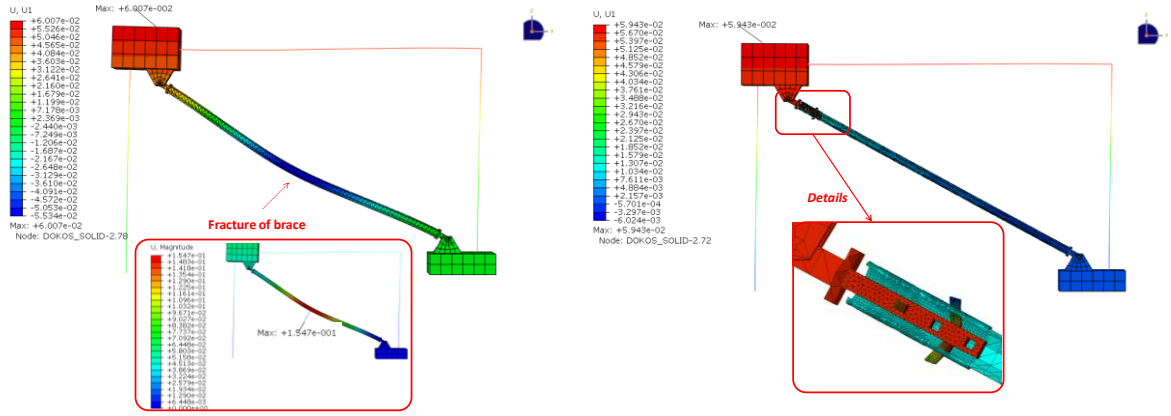


Figure 12: Deformed model at end of analysis, (i) Braced Structure, (ii) CAR1-Braced Structure (values in m).

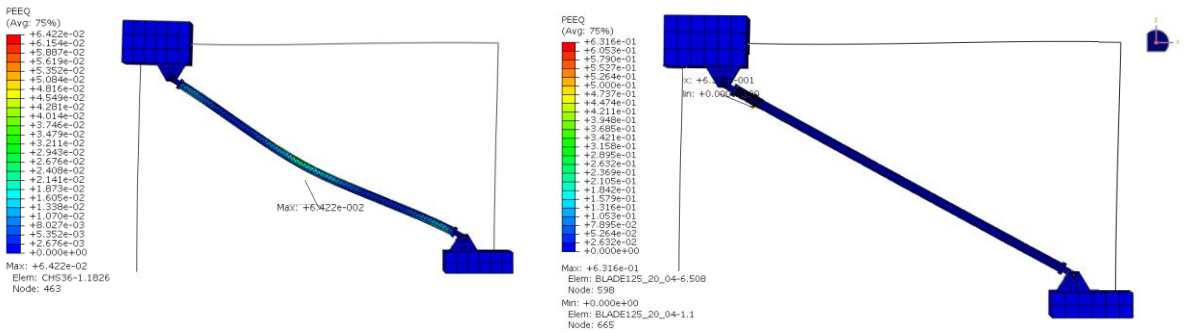


Figure 13: Deformed shape and maximum plastic strain at the end of analysis.

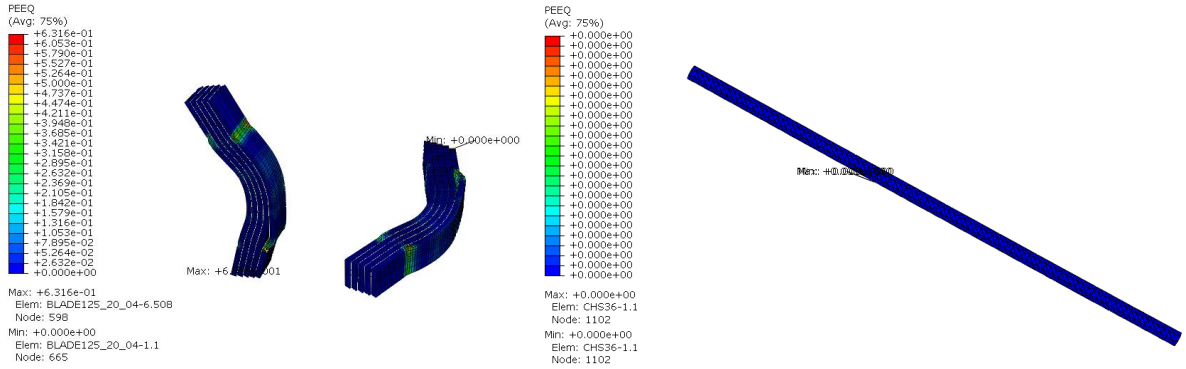


Figure 14: Maximum plastic strain in superimposed blades and diagonal brace at the end of analysis for CAR1-Braced Structure.

Conclusions

In the present paper, an anti-seismic steel device (with code name CAR1) for seismic strengthening of existing or new buildings, which was recently developed at the Laboratory of Strength of Materials and Structures of Aristotle University of Thessaloniki, is studied experimental. A detailed nonlinear finite element model (FEM) was also developed. This model was calibrated against experimental results and used to explain the response of the device CAR1. In addition, a numerically robust finite element model of a whole one storey structure is described, for high-fidelity simulations of inelastic responses of device CAR1 on braced frame.

Based on the findings of this paper, the following conclusions are drawn:

1. Device CAR1 is a reliable energy dissipated device, which can be used on new or existing structures and minimize the probability of failure of the structure in any charging.
2. The developed nonlinear FEM models can be reliable used to access the behavior of the proposed anti-seismic steel device CAR1 as they are capable to trace the hysteretic behavior and predict the deformed shape of the device with good accuracy.
3. Based on the shape and consistency of the hysteresis loops, it is recommended seismic energy, whereas will not break during repeated cyclic loading.

4. The calibrated FEM model permits a thorough investigation of the stress state in the blades and helps to identify all possible local failures
5. Using CAR1 device on diagonal brace, the fracture life of brace is increased. The system exhibited uniform energy absorption with more stability, as strength and maximum deformation of the system increased considerably.

Acknowledgments

I would like to sincerely thank my supervisor, Papadopoulo Paniko (Assistant Professor at Aristotle University of Thessaloniki), for his valuable support.

References

[1] Federal Emergency Management Agency (FEMA). FEMA 356: Pre-standard and commentary for the seismic rehabilitation of buildings, Washington, DC; 2000.
[2] American Institute of Steel Construction (AISC). AISC 341-10: Seismic provisions for structural steel buildings. Chicago, IL; 2010.
[3] Wada A, Huang YH, Iwata M. (1999) Passive damping technology for buildings in Japan. Prog Struct Engng Mater; **2(3)**:1–15.
[4] Xie Q, (2005). State of arte of buckling-restrained braces in Asia, Journal of Constructional Steel Research, **61**, 727-748.
[5] Hoveidae, N and Rafezy, B.. (2012). Overall buckling behavior of all-steel buckling restrained braces. Journal of Constructional Steel Research **79**,151–158.
[6] Zayas VA, Popov EP, Mahin SA. Cyclic inelastic buckling of tubular steel braces. Earthq Eng Rsrch Ctr, Report No. UCB/EERC 80/16 1980; Univ. of California Berkeley, CA
[7] Ikeda K, Mahin SA, Dermitzakis SN. Phenomenological modeling of steel braces under cyclic loading. Earthq Eng Rsrch Ctr, Report No. UCB/EERC 84/09 1984; Univ. of California Berkeley, CA
[8] Khatib F, Mahin SA, Pister KS. Seismic behavior of concentrically braced steel frames. Earthq Eng Rsrch Ctr, Report No. UCB/EERC 88/01 1988; Univ. of California Berkeley, CA
[9] Ikeda K, Mahin SA. Cyclic response of steel braces. J Struct Eng ASCE 1986; **112(2)**:342-361.
[10] Jin J, El-Tawil S. Inelastic cyclic model for steel braces. J Eng Mech ASCE 2003; **129(5):** 548-557.
[11] Lee PS, Noh HC.(2010). Inelastic buckling behavior of steel members under reversed cyclic loading. Eng Struct. **32(9):** 2579–2595.
[12] Lotfollahi M, Alinia MM, Taciroglu E.(2011). Inelastic buckling simulation of steel braces through explicit dynamic analyses. Num Anal & Appl Math, ICNAAM, AIP Conf Proc 2011a; 1389: 2012-2015.
[13] Lotfollahi M, Alinia MM, Taciroglu E. (2011). A validated finite element procedure for buckling simulation of diagonally braced moment resisting frames. Num Anal & Appl Math, ICNAAM, AIP Conf Proc 2011b; 1389: 2016-2019.
[14] Lumpkin EJ, Hsiao PC, Roeder CW, Lehman DE, Tsai CY, Wu AC, Wei CY, Tsai KC. (2012). Investigation of the seismic response of three-story special concentrically braced frames. J Constr Steel Res; **77**: 131-144.
[15] Nip KH, Gardner L, Elghazouli AY.(2010). Cyclic testing and numerical modelling of carbon steel and stainless steel tubular bracing members. Eng Struct; **32(2):** 424-441.
[16] Yoo JH, Roeder CW, Lehman DE.(2008) Analytical performance simulation of special concentrically braced frames. J Struct Eng ASCE; **134(6)**: 881-889.
[17] Yoo JH, Roeder CW, Lehman DE. (2008). Influence of connection design parameters on the seismic performance of braced frames. J Constr Steel Res; **64(6):** 607-623.
[18] Aiken, I. D., Nims, D. K., & Kelly, J. M. (1992). Comparative study of four passive energy dissipation systems, Bulletin of the New Zealand National Society for Earthquake Engineering, **25(3)**, 175-192.
[19] Soong, T.T. & Jr, B.F.Spencer. (2002). Supplemental energy dissipation: state-of-the-art and state-of-the-practice. Engineering Structures, **24**, 243-259.
[20] Symans, M.D., Charney, F.A., Whittaker,A.S., Constantinou, M.C., Kircher, C.A., Johnson, M.W., and McNamara, R.J. 2008. Energy dissipation systems for seismic applications: Current practice and recent developments. J.Struct. Engrg., **134**(3-1).
[21] Pall, A.S., Verganelakis, V., March, C., (1987). Friction-Dampers for seismic control of Concordia University Library Building. Proc 5th Canadian Conference on Earthquake Engineering, Ottawa, pp 191-200.

[22] Makris, N. and Constantinou, M. C., (1992), .Spring-Viscous Damper Systems for Combined Seismic and Vibration Isolation,. Earthquake Engineering and Structural Dynamics, **21**, pp. 649-664.
[23] Martinez-Romero, E., (1993), .Experiences on the Use of Supplemental Energy Dissipators on Building Structures,. Earthquake Spectra, **9**, pp. 581-624.
[24] Whittaker, A. S., Bertero, V. V., Alonso, J. L. and Thompson, C. L. (1989). "Earthquake Simulator Testing of Steel Plate Added Damping and Stiffness Elements", Report No. UCB/EERC 89/02, University of California, Berkley.
[25] Anagnostides G., Hargreaves A.C., Wyatt T.A., (1989). Development and applications of energy absortion devices based on friction. J. Construct. Steel Research, 13, 317-336.
[26] Aiken, I. D., Nims, D. K., Whittaker, A. S., & Kelly, J. M. (1993). Testing of passive energy dissipation systems. Earthquake spectra, **9(3)**, 335-370.
[27] Papadopoulos, P.K., Salonikios, Th., Dimitrakis, S., Papadopoulos, A. (2013). Experimental investigation of a new steel friction device with link element foe seismic strengthening of structures. Structural Engineering & Mechanics, **46(4).**
[28] Pall, A.S., and March, C., (1982). Seismic response of friction damped braced frames. ASCE, Journal of Structural Division, **108 (9)**, 1313-1323.
[29] Papadopoulos, P. (2012). New nonlinear anti-seismic steel device for the increasing the seismic capacity of multi-storey reinforced concrete frames. The structural design of tall and special buildings, **21**, 750-763.
[30] Ramirez, J. D. M., & Tirca, L. (2012). Numerical Simulation and Design of Friction- Damped Steel Frame Structures damped. 15th World Conference in Earthquake Engineering.
[31] Abaqus Simulia, (2012). Analysis User's Manual Volume IV. Analysis User's Manual Volume IV . Providence: Dassault Systèmes.
[32] Vasdravellis, G., Karavasilis, Th., Uy, Br. (2013). Finite element models and cyclic behavior of self-centering steel post-tensioned connections with web hourglass pins. Engineering Structures, **52**, 1-16.
[33] Manos, G.C, Theofanous, M., Katakalos, K. (2014). Numerical simulation of the shear behaviour of reinforced concrete rectangular beam specimens with or without FRP-strip shear reinforcement. Advances in Engineering Software, **67**, 47-56.
[34] Papadopoulos P.K., Titirla M.D., & Papadopoulos A.P. (2014). A new seismic energy absorption device through simultaneously yield and friction used for the protection of structures. 2nd European Conference on Earthquake Engineering and Seismology, Istanbul.
[35] Doudoumis, I.N., (2007). Finite element modelling and investigation of the behaviour of elastic infilled frames under monotonic loading. Engineering Structures **29**, 1004–1024.
[36] Eurocode 3, (2003). Design of steel structures. Part 1-8: design of joints. prEN 1993-1-8:2003. European Committee for standardization: Brussels.
[37] Computers and Structures Inc (2007). SAP 2000 nonlinear version 11.0.3 User's Reference Manual. Berkeley, California
[38] BSSC (2004) NEHRP recommended provisions for seismic regulations for new buildings and other structures. Report FEMA 450, FEMA, Washington, DC

Particle Method Simulation of Wave Impact on Structures

†M. Luo, *C.G. Koh and W. Bai

Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576

†Corresponding author: ceelm@nus.edu.sg
*Presenting author: cgkoh@nus.edu.sg

Abstract

In this paper, a Consistent Particle Method (CPM) is presented to model violent wave impact with compressible air pockets. The novelty of this method lies in four key aspects: (1) accurate computation of spatial derivatives for Laplacian and gradient operators (and hence better pressure prediction) without the use of kernel function unlike some other particle method, (2) rational treatment of density discontinuity at the water-air interface without any smoothing or smearing scheme, (3) a thermodynamics-based compressible solver for modelling compressible air that eliminates the need of determining the artificial sound speed, and (4) two-phase coupling of compressible air solver and incompressible water solver without iteration between the two solvers. An experimental study of sloshing impact with entrapped air pocket is conducted to validate the numerical model.

Keywords: Particle Method, Wave Impact, Two-phase Flow, Air Compressibility.

Introduction

Modelling of wave impact on structures is of great practical interest in offshore and marine engineering e.g. for design of seawalls against tsunami waves in terms of the required height and strength. With the rapid advances of computer power, many numerical methods have been developed to predict the wave profile and impact forces. However, most of these studies[1, 2] do not consider the presence of entrapped air pockets, or treat the air pockets as incompressible. While incompressibility is a reasonable assumption in some water-air flow scenarios[3], air entrapment or entrainment may be generated in some other problems such as violent wave impact on structures[4]. The compressibility of entrapped air pockets can play an important role in the water-air interaction in terms of influencing the pressure peak and impact duration in a wave impact process[5]. Therefore, it is necessary to include air compressibility to better simulate such water-air flow problems.

The numerical difficulties to model wave impact problems with entrapped air pockets include the large and discontinuous deformation of fluid and the abrupt discontinuity of fluid properties (density and viscosity) at the interface between water and air. A greater challenge is to have an integrated solution for water and air that behave very differently, the former being practically incompressible and the latter highly compressible. To address these issues, many mesh-based methods (such as Finite Difference Method and Finite Volume Method) and particle methods have been developed. Due to the meshless and Lagrangian nature, particle methods possess three inherent advantages over mesh-based methods: (1) better capability in modelling large and discontinuous fluid motion such as breaking waves, (2) better tracking of moving interface of different fluids, and (3) no numerical diffusion induced by the convection term in the Navier-Stokes equation. Therefore, a particle method was selected as the underlying tool of the present study. The most commonly used particle methods include SPH, ISPH, MPS and CPM. The primary difference between them lies in the computation of spatial derivatives. Compared to the other three, CPM computes the gradient and Laplacian operators in a more fundamental way by using Taylor series

expansion. Eliminating the use of a kernel function, the spatial derivatives can be approximated much more accurately and hence no artificial schemes are required[6].

The main difficulties of using CPM to simulate violent waves with air entrapment is the approximation of spatial derivatives with sharp density change across fluid interface and the consistent modelling of incompressible water and compressible air. To address these two issues, an improvement of the derivative-approximation scheme in the original CPM was recently proposed to deal with the sharp density discontinuity[6]. In addition, a thermodynamically-consistent compressible solver that not only can be integrated with the developed incompressible solver seamlessly but also can overcome some issues encountered by other compressible solvers is developed[7]. In this paper, the main features and advantages of CPM are presented systematically. Using this method, water sloshing with entrapped air pocket in a specially designed oscillating tank is studied with our own experimental validation.

Governing equations and CPM formulations

The governing equations for viscous Newtonian fluids (both incompressible and compressible) in a two-fluid system are the Navier-Stokes equations as follows[8]:

$$\frac{1}{\rho}\frac{D\rho}{Dt}+\nabla\cdot\mathbf{v}=0 \tag{1}$$

$$\frac{D\mathbf{v}}{Dt}=-\frac{1}{\rho}\nabla p+\frac{1}{\rho}\nabla\cdot\left[\mu\left(\nabla\mathbf{v}+\left(\nabla\mathbf{v}\right)^{T}\right)\right]+\mathbf{g} \tag{2}$$

where ρ is the density of fluid, $\mathbf{v}$ the particle velocity vector, p the fluid pressure, μ the dynamic viscosity of fluid and $\mathbf{g}$ the gravitational acceleration.

For both incompressible and compressible fluids, the governing equations are solved by a predictor-corrector scheme[9, 10]. In the predictor step, the temporary particle velocities and positions are computed by neglecting the pressure gradient term. In the corrector step, a pressure Poisson equation (PPE) can be derived as follows

$$\nabla\cdot\left(\frac{1}{\rho^{*}}\nabla p^{(k+1)}\right)=\frac{1}{\Delta t^{2}}\frac{\rho^{(k+1)}-\rho^{*}}{\rho^{(k+1)}} \tag{3}$$

For incompressible fluids, the incompressibility condition is enforced by setting the fluid density at the current time step ($\rho^{(k+1)}$) to the initial value (ρ_0). The intermediate fluid density (ρ^{*}) is evaluated in the same way introduced in Luo et al. [6]. For compressible fluids, although a similar approach is used to evaluate fluid density, a slow-slope weighting functions whose value at r = 0 is smaller is adopted to allow more compressibility of fluid (more details can be referred to Luo et al. [7]). Another distinct feature in the simulation of compressible flows is that, without the incompressibility condition, the fluid density $\rho^{(k+1)}$ in Equation (3) should be treated as unknown (more details will be presented later).

Gradient and Laplace operators involving density discontinuity

The derivative computation scheme in CPM is derived based on Taylor series expansion. This scheme has been demonstrated to work well for 1-phase flows[11, 12]. In two-phase flows, the pressure function is continuous at the fluid interface but its gradient changes drastically because of the large density difference between two fluids (e.g. water and air densities differ by three orders of magnitude)[6]. Hence, when applied to pressure, the scheme introduced in the previous section does

not give good approximation of gradient and Laplacian terms near the fluid interface. This problem, nevertheless, can be resolved by observing that the pressure gradient normalized with respect to density, i.e. $\nabla p / \rho$, is of the same order of magnitude in the two fluids of a general dynamic problem and, in the hydrostatic case, is in fact constant. By addressing the normalized pressure gradient term, the formulation to compute the gradient and Laplacian operators with abrupt density discontinuity can be derived to be (more details can be referred to Luo et al. [6])

$$\left(\frac{1}{\rho}\frac{\partial p}{\partial x}\right)_i = \sum_{j\neq i}\left[\frac{1}{0.5(\rho_i+\rho_j)}C_{1j}\left(p_j - p_i\right)\right] \tag{4}$$

and

$$\left(\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial p}{\partial x}\right)\right)_i = \sum_{j\neq i}\left[\frac{1}{0.5(\rho_i+\rho_j)}C_{3j}\left(p_j - p_i\right)\right] \tag{5}$$

The coefficients C_{1j} and C_{3j} are the same as those in 1-phase CPM[11]. The above reformulation retains the consistency with Taylor series expansion in computing the required gradient and Laplace terms with abrupt density discontinuity. Since no density smoothing or smearing scheme is needed, this scheme is able to model sharp fluid interface (e.g. water and air whose density difference is about three orders of magnitude) with good accuracy.

Compressible solver based on thermodynamics

For compressible flows, $\rho^{(k+1)}$ in Equation (3) is unknown and hence a closure condition is needed to solve the PPE. The polytropic gas law as shown in Equation (4) is selected to be the closure relation since it does not require the input of speed of sound (c_s), which is dependent on the composition and temperature of a fluid. This avoids the need to determine the actual or numerical sound speed, unlike in the c_s dependent EOS.

$$\frac{p}{\rho^{\gamma}} = \text{constant} \tag{4}$$

where γ is the ratio of specific heats at constant pressure and constant volume. Its value for air is about 1.4.

Incorporating the closure condition of Equation (4) to Equation (3), the PPE accounting for fluid compressibility can be obtained as (more details can be referred to Luo et al. [7])

$$-\nabla\cdot\left(\frac{1}{\rho_i^*}\nabla p_i^{(k+1)}\right) + \frac{1}{\Delta t^2\rho_i^*}\frac{\rho_{a0}}{p_{a0}}\frac{1}{\gamma}p_i^{(k+1)} = -\frac{1}{\Delta t^2}\frac{\rho_{a0}-\rho_i^*}{\rho_i^*} + \frac{1}{\Delta t^2\rho_i^*}\frac{\rho_{a0}}{\gamma} \tag{5}$$

Since the speed of sound c_s is not involved in Equation (5), the issue of how to determine the actual or numerical value of c_s is avoided. This is a significant benefit of the present compressible solver. More importantly, this thermodynamically-consistent compressible solver and the previously proposed incompressible solver[6] both use the predictor-corrector scheme to solve the same governing equations and thus can be easily integrated, leading to the complete two-phase model. Named 2-phase CPM, it is capable of simultaneously and consistently simulating two-phase incompressible and compressible flows with large density difference.

Numerical examples

Sloshing impact with entrapped air pocket

To study wave impact scenario with entrapped air pocket, a new experiment is designed and conducted as shown in Figure 1. The water container comprises a big (left) tank connected by a short channel to a small (right) tank. It is designed such that when water in the left tank sloshes to the right (or left), some water will move through the connecting channel and compress (or expand) the air in the right tank. The same tank as shown in Figure 16 of Luo et al. [7] is used. Air pressure at the middle of the top wall of the right tank, i.e. P_{A1}, is measured by an absolute pressure sensor. Water pressures at 60 mm from the bottom on the right wall of the right tank (P_{W1}) and 30 mm from the bottom on the left wall of the left tank (P_{W3}) are measured by gauge pressure sensors.

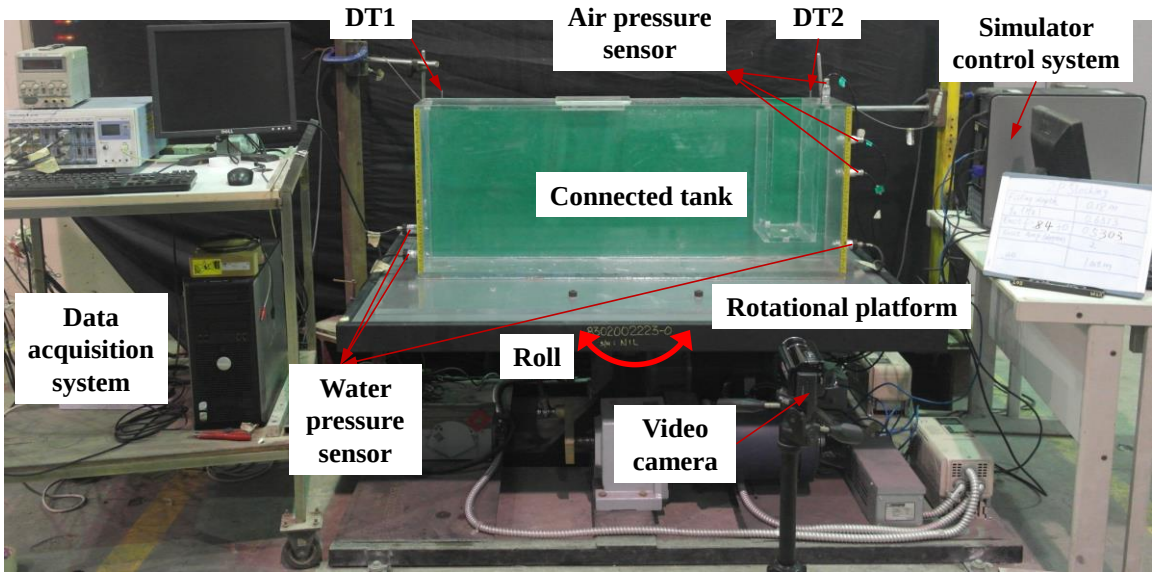
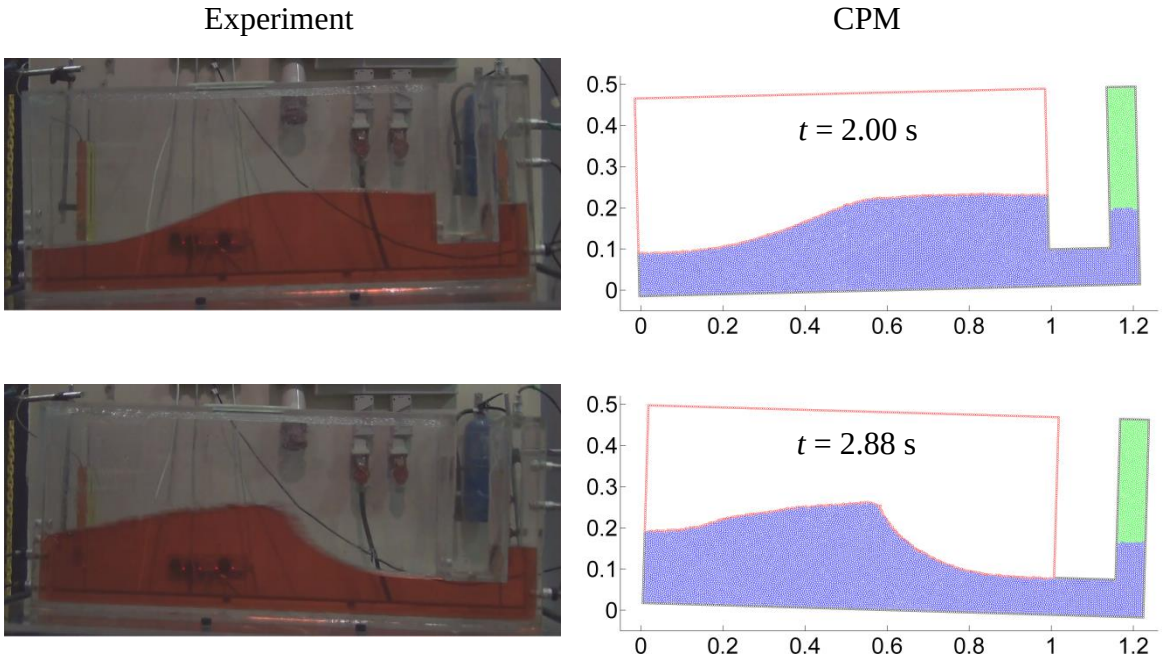


Figure 1. Setup of water-air sloshing experiments in a connected container under rotational excitation



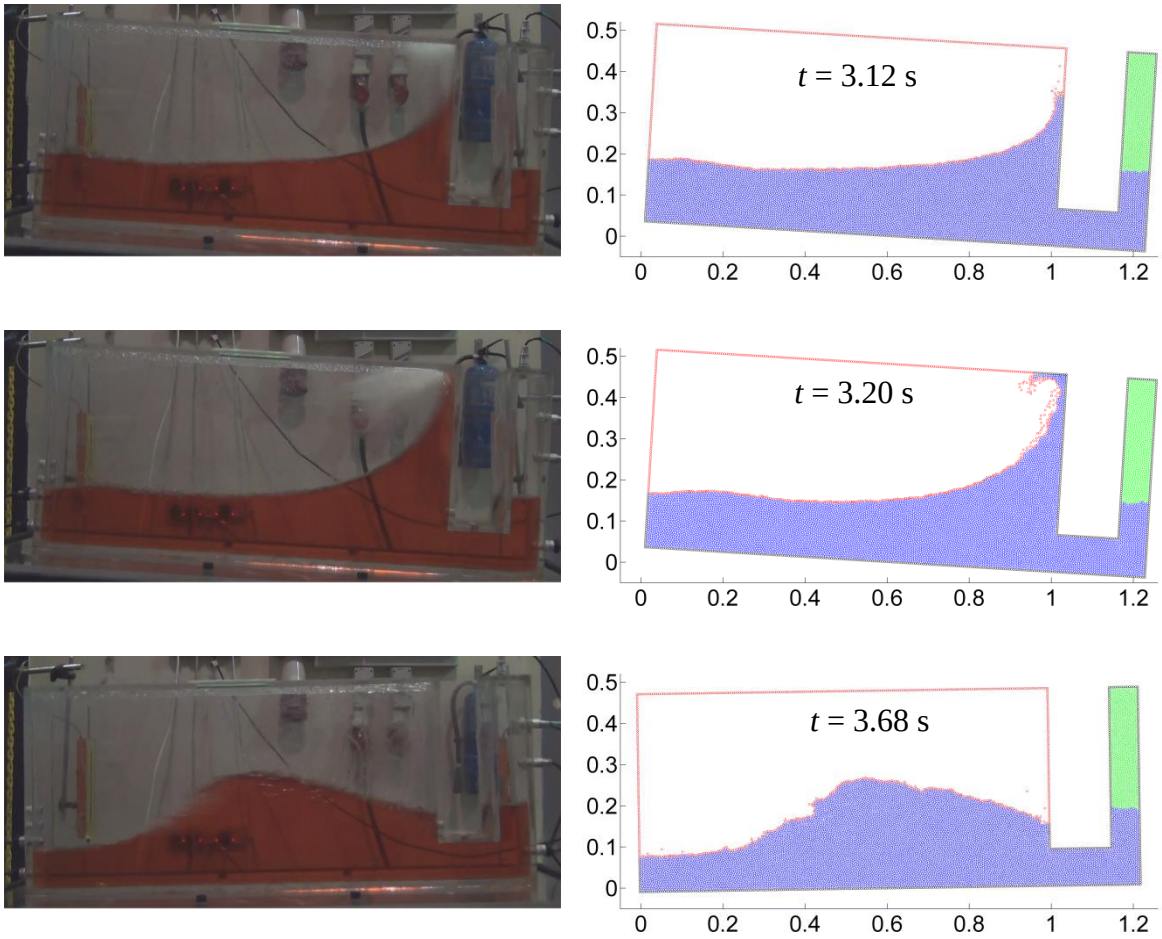


Figure 2. Wave profiles of sloshing in a connected tank with closed air pocket under rotational excitation: experimental result and CPM simulation

In the case presented in this section, the filling depth is adopted to be 0.18 m. The excitation frequency of $0.92\omega_0$ (= 3.6493 rad/s) is found to generate a relatively large variation of air pressure in the right tank, where ω_0 is the reference frequency (not the natural frequency of the sloshing system but only a reference value) computed based on the linear wave theory with water depth (d_L) and length (L_L) in the left tank. In numerical simulation, an initial particle distance of 0.005 m and fixed time step 0.0005 s are adopted on the tradeoff between accuracy and efficiency. The water and air densities at the NTP (Normal Temperature and Pressure) condition are adopted. The dynamic viscosities of water and air are selected to be 10^{-3} Pa s and 1.983×10^{-5} Pa s respectively.

The wave profiles and pressure histories at points A_1, W_1 and W_3 are presented in Figure 2 and Figure 3. Generally good agreement between numerical simulation and experimental result is obtained. The water moves like a bore (because of the relatively low filling depth) which develops over time (see t = 2.00 s and 2.88 s in Figure 2). At t = 3.12 s, violent wave impact occurs near the connecting channel, generating large compression force to the air pocket in the right tank. This can be clearly seen in Figure 3a, which shows a large peak for the air pressure at point A_1. As the water in the left tank runs up along the right wall of the left tank (t = 3.20 s in Figure 2), the compression force continues to exert on the air pocket in the right tank. At t = 3.68 s, the run-up water falls back to the water body and begins to move towards left. It is noted that the air pressure in the right tank shows vibration during the impact process. The air

pressure also influences the water pressure near the air pocket (see the water pressure at Pw1 as shown in Figure 3b).

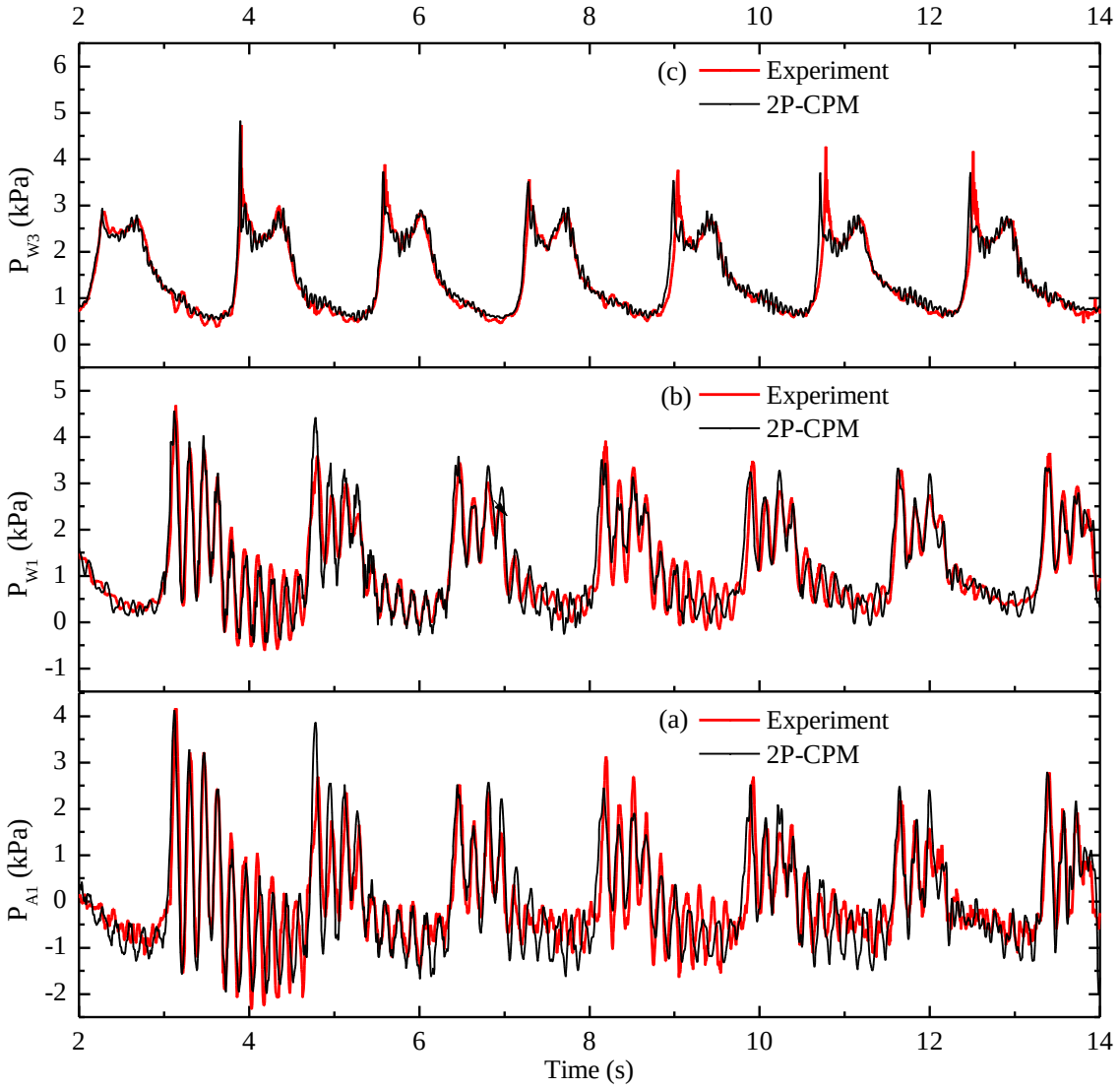


Figure 3. Simulated air pressure at Point P_{A1} and water pressures at Point P_{W1} and P_{W3} in comparison with experimental results

The pressure vibration in the air pocket is further investigated through a power spectral analysis using the Fast Fourier Transform (FFT). It is interesting to note that there is only one peak value, i.e. 6.120 Hz, in the frequency-power curve. It means that the air pressure vibrates with one distinctive frequency. To verify that this pressure vibration is real and not spurious due to the numerical algorithm, the natural frequency of the air tube (under the compression of water) is derived. Following Ramkema [13] who addressed the problem of wave impact on coastal structures, the air-pocket-water system is represented by a mass-spring system as shown in Figure 4, in which the spring is the air pocket and the mass is the water effectively contributing to the impact. The upper bound of the effective water mass is the water in the connecting channel and the right tank, while the lower bound is the upper bound excluding the water in the rectangular region at the right bottom corner of the container (the region within the dash-dot line in Figure 4). Since water at the right bottom corner (dark shaded region in Figure 4) is almost stationary relatively to the tank (theoretically the right bottom point of the container is a stagnation point), the effective mass of the present problem (light shaded region in Figure 4) is

approximated to be water in the connecting channel and the right tank excluding the right bottom corner.

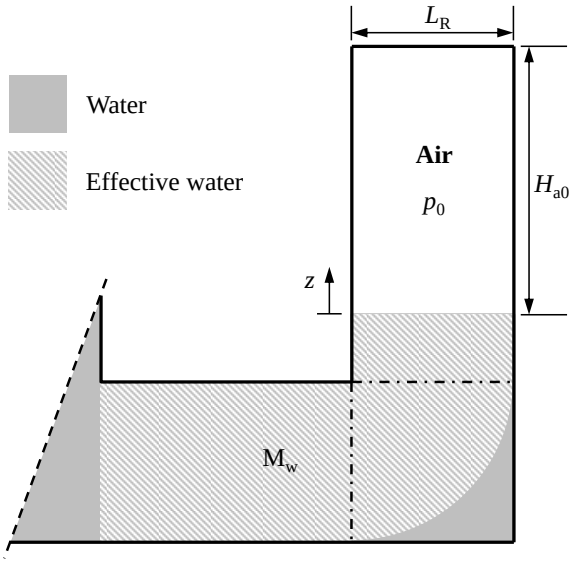


Figure 4. Schematic view of water impact on an air pocket (not to scale)

Assuming the water level in the right tank to be horizontal and giving it a small perturbation z, the force (per unit width) applied on the effective water mass is as follows

$$\begin{aligned} F = (p_{a0} - p)L_{\mathrm{R}} &= \left[1 - \left(1 + \frac{z}{H_{\mathrm{a0}} - z}\right)^{\gamma}\right] p_{a0} L_{\mathrm{R}} \\ &\approx \left[1 - \left(1 + \gamma \frac{z}{H_{\mathrm{a0}} - z}\right)\right] p_{a0} L_{\mathrm{R}} = -\frac{\gamma p_{a0} L_{\mathrm{R}}}{H_{\mathrm{a0}} - z} z \approx -\frac{\gamma p_{a0} L_{\mathrm{R}}}{H_{\mathrm{a0}}} z \end{aligned} \tag{6}$$

where p_{a0} is the initial air pressure in the right tank, L_{R} the length of the right tank and H_{a0} the initial height of the air tube. Ignoring the friction forces from the tank walls, the dynamic equation for the effective water mass is as follows

$$M_{\mathrm{w}} \frac{d^2 z}{dt^2} + \frac{\gamma p_{a0} L_{\mathrm{R}}}{H_{\mathrm{a0}}} z = 0 \tag{7}$$

where M_{w} is the effective water mass (per unit width). Then the natural frequency of the dynamic system can be obtained as

$$f = \frac{1}{2\pi} \sqrt{\frac{\gamma p_{a0} L_{\mathrm{R}}}{H_{\mathrm{a0}} M_{\mathrm{w}}}} \tag{8}$$

the form of which is similar to that derived by Cuomo *et al.* [14] who analytically studied wave impingement entrapping an air pocket against vertical wall. Substituting the upper and lower bounds of M_{w} into Equation (8), the lower and upper bounds of the natural frequency of the entrapped air pocket can be obtained to be 5.668 Hz and 6.507 Hz, whereas the natural frequency corresponding to the adopted value of M_{w} is 6.296 Hz. Compared to the observed frequency of pressure vibration (i.e. 6.120 Hz) in the experimental result, the relative differences are only 7.3 %, 6.3 % and 2.8 %, respectively, for the lower and upper bounds and the adopted value of M_{w}. Therefore, the accuracy of this simplified model is acceptable. The study on the natural frequency

of the air pocket further substantiates that the pressure oscillations observed in the experiment and CPM simulation are real and due to the natural vibration of the entrapped air pocket (air cushion effect).

Conclusions

In this paper, the novel CPM is presented with three features: (1) Accurate computation of first- and second-order derivatives in a way consistent with Taylor series expansion even in two-phase cases with abrupt density change to about 1000; (2) A thermodynamically-consistent compressible solver by employing the polytropic gas law; (3) Seamless integration of the incompressible and compressible solvers such that wave impact problems with entrapped air pocket can be simulated in a simultaneous way.

An experimental study of water sloshing in a specially designed tank is conducted to measure the pressure change of a closed air pocket under wave impact. Numerical results including wave profiles, wave impact pressures and particularly the pressure vibration in the air pocket predicted by CPM agree generally well with the experimental results.

Acknowledgement

The authors appreciate the research grant provided by the Singapore Maritime Institute (Project SMI-2014-OF-02) as well as the funding and technical support of Sembcorp Marine Technology Pte Ltd.

References

[1] Liu D, Lin P. A numerical study of three-dimensional liquid sloshing in tanks. Journal of Computational Physics 2008;227:3921-39.
[2] Wang Y, Shu C, Huang HB, Teo CJ. Multiphase lattice Boltzmann flux solver for incompressible multiphase flows with large density ratio. Journal of Computational Physics 2015;280:404-23.
[3] Heyns JA, Malan AG, Harms TM, Oxtoby OF. A weakly compressible free-surface flow solver for liquid–gas systems using the volume-of-fluid approach. Journal of Computational Physics 2013;240:145-57.
[4] Peregrine DH. Water-wave impact on walls. Annual Review of Fluid Mechanics 2003;35:23-43.
[5] Abrahamsen BC, Faltinsen OM. The effect of air leakage and heat exchange on the decay of entrapped air pocket slamming oscillations. Physics of Fluids 2011;23:102107.
[6] Luo M, Koh CG, Gao M, Bai W. A particle method for two-phase flows with large density difference. International Journal for Numerical Methods in Engineering 2015;103:235-55.
[7] Luo M, Koh CG, Bai W, Gao M. A particle method for two-phase flows with compressible air pocket. International Journal for Numerical Methods in Engineering 2016:n/a-n/a.
[8] Tofighi N, Yildiz M. Numerical simulation of single droplet dynamics in three-phase flows using ISPH. Computers & Mathematics with Applications 2013;66:525-36.
[9] Shao S, Lo EYM. Incompressible SPH method for simulating Newtonian and non-Newtonian flows with a free surface. Advances in Water Resources 2003;26:787-800.
[10] Koshizuka S, Nobe A, Oka Y. Numerical analysis of breaking waves using the moving particle semi-implicit method. International Journal for Numerical Methods in Fluids 1998;26:751-69.
[11] Koh CG, Gao M, Luo C. A new particle method for simulation of incompressible free surface flow problems. International Journal for Numerical Methods in Engineering 2012;89:1582–604.
[12] Koh CG, Luo M, Gao M, Bai W. Modelling of liquid sloshing with constrained floating baffle. Computers & Structures 2013;122:270-9.
[13] Ramkema C. A model law for wave impacts on coastal structures. Coastal Engineering Proceedings1978.
[14] Cuomo G, Allsop W, Takahashi S. Scaling wave impact pressures on vertical walls. Coastal Engineering 2010;57:604-9.

Consistency-driven Pairwise Comparisons Approach to Abandoned Mines Hazard Rating

Waldemar Koczkodaj[1] **and Michael Soltys**[2]

[1]Department of Mathematics and Computer Science, Laurentian University, Sudbury, Ontario, Canada P3E 2C6
[2]Department of Department of Computer Science, California State University at Channel Islands, One University Drive, Camarillo, CA 93012, USA
*Presenting author: michael.soltys@csuci.edu
†Corresponding author: wkoczkodaj@cs.laurentian.ca

Abstract

The pairwise comparisons method, together with inconsistency analysis, are used to assess the hazard level for abandoned mines. Weights, reflecting the relative importance of the objectives concerned are one of the most commonly used solutions for this type of data. Subjective assessments involve inaccuracy (which is difficult to manage) and inconsistency in assessments (which can be measured and may influence the accuracy). The pairwise comparisons method allows us to define a consistency measure and use it as a validation technique. A consistency-driven knowledge acquisition, supported by a properly designed software, contributes to the improvement of quality of knowledge-based systems.

Keywords: pairwise comparison, knowledge management, multicriteria evaluation, inconsistency, hazard rating.

1 Introduction

The first (somewhat documented but never formally published) use of pairwise comparisons (PC) is attributed to Ramon Llull, a 13th-century mystic and philosopher (see [5]). Thurstone applied pairwise comparisons in the form of "the law of comparative judgment" in [18]. There is a variation of this law known as the BTL (Bradley-Terry-Luce) model (cf. [2]). A number of customized methods of pairwise comparisons followed in numerous (some of them controversial) studies. We do not intend to endorse any such customization here. However, Saaty's seminal work [17] had a considerable impact on the pairwise comparisons (PC) research and should be acknowledged despite serious controversies generated by it.

The technical issues of acquiring this knowledge, representing it, and using it appropriately to construct and explain lines-of-reasoning, are important problems in the design of knowledge-based systems. Knowledge acquisition involves extracting knowledge from human experts, books, documents, sensors, or computer files. In the knowledge validation stage this knowledge is validated and verified until its quality is considered acceptable according to some pre-established standards.

Knowledge acquisition is the extraction of knowledge from sources of expertise and its transfer to the knowledge base. Acquisition is actually done throughout the entire expert system development process. Knowledge is a collection of specialized facts, procedures, and assessment rules and may be collected from many sources. These sources can be divided into two types: documented and undocumented. The latter resides in people's minds. Knowledge can be identified and collected by using any of the human senses. It can also be identified and collected by machines.

The knowledge engineer elicits knowledge from the expert, refines it with the expert, and represents it in the knowledge base. The elicitation of knowledge from the expert can be done

manually or with the aid of computers. The main purpose of computerized support to the expert is to reduce or eliminate the potential problems mentioned earlier, especially those of indeterminate bias and ambiguity. These problems dominate the gathering of information for the initial knowledge base and the interactive refinements of this knowledge. A smart knowledge acquisition tool needs to be able to add knowledge incrementally to the knowledge base and refine, or even correct, existing knowledge. Visual modeling techniques are very important in constructing the initial domain model. The objective of the visual modeling approach is to give the user the ability to visualize real-world problems and to manipulate elements of it through the use of graphics.

The expert's knowledge may be, for example, expressed in assessing the number of preferences, relevant criteria or factors, or possible alternatives. When devising methods for formulating and assessing preferences, a knowledge engineer has to take into account the limitations in human capabilities for undertaking such endeavor. One possible technique of extracting the expert's knowledge and preferences is based on the pairwise comparisons method.

2 Pairwise Comparisons Preliminaries

The pairwise comparisons method utilizes statements about expert's preferences and assessments. These statements are expressed by examination of pairs of criteria or objectives. The presented methodology utilizes mapping of inconsistent evaluations by an expert into a numerical scale (see Table 1) that closely approximate his/her assessments. Ordinal numbers are used to express relative preferences. In particular the numbers do not represent "absolute" measure of the mapped criteria, as such may simply not exist (for example, it is hard to define a global measure of public safety but it is still practical to compare it, in relative terms, with the degree of environmental pollution).

Intensity	definition	explanation
1	equal importance	equal contribution
2	weak importance of one over another	slightly favor one criterion over another
3	essential or strong importance	strongly favor one criterion over another
4	demonstrated importance	strong dominance
5	absolute importance	the highest preference
1.2, 2.3, ...,etc.	Intermediate values	when compromise is needed

Table 1: Scale used for pairwise comparisons

The traditional matrix representation of pairwise comparisons (PC) is by using a PC matrix M of the following format:

$$M = \begin{bmatrix} 1 & m_{1,2} & \cdots & m_{1,n} \\ \frac{1}{m_{1,2}} & 1 & \cdots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{m_{1,n}} & \frac{1}{m_{2,n}} & \cdots & 1 \end{bmatrix}.$$

PC matrix elements represent the intensities of an expert's preference between individual pairs

of entities (or criteria) expressed as ratios chosen from an assumed scale for subjective data and transformed by the recently published formula in [10]. Note the criteria E_1, E_2,..., E_n (where n is the number of criteria to be compared). The entry m_{ij} in the i-th row and j-th column of the PC matrix M, denotes the relative importance of entity (or criterion) E_i compared with objective E_j, as expressed by an expert. This PC matrix M has all positive elements and has the following reciprocal property:

$$\forall i, j, \quad 1 \leq i, j \leq n, \quad m_{ij} = \frac{1}{m_{ji}}.$$

The PC matrix M is called consistent if $\forall i, j, k, 1 \leq i < j < k \leq n$, it is the case that $m_{ij} * m_{jk} = m_{ik}$. The vectors consisting of the three values $[m_{ij}, m_{ik}, m_{jk}]$ are called "triads." By the reciprocity condition, triads have a mirror image below the diagonal, and so it is sufficient to concentrate on the values above the diagonal.

Let w_i denote the unknown weight of the criterion i. How can the vector $w = [w_1, w_2, ..., w_n]$ be estimated on the basis of the PC matrix M? One possible solution can be the following. If the expert's assessments are completely consistent, one would have:

$$\forall i, j, \quad 1 \leq i, j \leq n, \quad a_{ij} = \frac{w_i}{w_j}.$$

The following heuristic:

$$w_i = (\prod_{j=1}^{n} m_{ij})^{1/n}$$

was proposed in [19] for finding vector w for inconsistent PC matrices. In fact it trivially works also for consistent PC matrices.

A definition of consistency proposed in [9] allows us to locate the most inconsistent assessments and reexamine them. New and more consistent assessments may be expressed in an interactive way. They may contribute to the overall reduction of the inconsistency.

3 Abandoned mines hazard rating

The knowledge engineer usually has to cope with a large number of criteria, factors or alternatives during the data acquisition process. Our model is presented visually[1] in Fig. 1, and is used by a tool called "Concluder."[2]

The model was the result of a team effort involving mining experts from the Ministry of Northern Ontario and Mines, with expertise based on years of experience. One episode that was in everyone's mind was the collapse of a school yard (fortunately, at a time when the children were attending classes in the school building). The yard caved in as it was built on a forgotten abandoned mine. Based on the expertise of the mining professionals, and data from historical reports, pairwise comparisons were gathered into a large matrix. Needless to say, with such a large number of experts and data, the matrix that was created was inconsistent.

4 Inconsistency in pairwise comparisons

For a single triad $[x, y, z]$, the inconsistency indicator is given by the following formula:

[1]The graphic has been produced with Prefuse, a set of software tools for creating rich interactive data visualizations [8].

[2]Which we make available on Sourceforge [3].

$$ii = 1 - \min\left(\frac{y}{x * z}, \frac{x * z}{y}\right).$$

The new definition was proposed in [9], formally generalized to the entire matrix by the use of the $\max$ function for all triads (defined by the consistency condition), and simplified in [14]. Making comparative assessments of intangible criteria (e.g., the degree of an environmental hazard or pollution factors) involves not only imprecise or inexact knowledge but also inconsistency in our own assessments. The improvement of knowledge elicitation by controlling the inconsistency of experts' assessments is not only desirable but absolutely necessary.

Checking the consistency in the pairwise comparisons method could be compared to checking that the divisor is not equal to 0. It does not make sense to divide anything by 0. The proposed solution of the pairwise comparisons method is based on the assumption that the given reciprocal matrix is consistent. However, expecting that all subjective assessments are consistent is not realistic especially if they are subjective. We know that most assessments are subjective, inaccurate, and nearly always contain some kind of bias, and therefore the total consistency is not to be expected.

To have inconsistent assessments we must have at least three criteria to be compared. Consequently we may assume, that all indexes i, j, k must be pairwise different. We may calculate inconsistencies only for triads with indexes holding the property $1 \leq i < j < k \leq n$.

The inconsistency indicator of a PC matrix is the indicator of the quality of the knowledge. The "improvement" process of the quality of the knowledge begins with computing the inconsistency of the assessments. The triad with the largest inconsistency is displayed for the experts to have an opportunity to revise their preferences.

In our case, Concluder highlights the worst triad as illustrated by Fig. 2.

The inconsistency of 0.44 is regarded as too high (the threshold value is assumed 1/3 for similar applications) so experts need to reconsider their assessments. By changing 1.5 in the highlighted triad into 1.3, we can decrease inconsistency indicator to 0.32 which is assumed to be acceptable so weights w (automatically computed and illustrated by Fig. 3) can be used for decision making.

5 Conclusions

The consistency-driven approach presented in this paper was tested in a research project related to the decision process of rehabilitation of abandoned mines in Ontario by the Provincial Ministry of Northern Development and Mines. The implemented system assists middle-level management in making semistructured decisions. The main goal of the system is to provide management with the most comprehensive and most updated information necessary to make responsible decisions (for details see [1]).

The consistency-driven pairwise comparisons refocused the attention from the race of finding better and better approximation of weights for inconsistent matrices to devising heuristics to influencing assessments to be more consistent (but by no means totally consistent). Finding an ideal vector of weights for inconsistent (or very inconsistent) matrices is a mirage. It is a theoretically challenging and exciting task but does not have much practicality. It could be compared to an attempt at finding lengths of objects using a ruler which randomly changes (by, for example, extreme temperature) its length for each of them. The truth is that no "ideal" solution exists and understanding the true source of our problem, that is inconsistency of assessments, is absolutely necessary for decreasing the inaccuracy.

Reducing the inconsistency is not easy unless we know its location (not only its value). The presented definition of inconsistency locates it. The expert is given the feedback and opportunity of reconsideration of his/her assessments by using various approaches (e.g., Delphi method). It may not be advisable to allow the expert the full flexibility since his/her subjective assessment may change due to an unsubstantiated race for consistency of assessments instead of non-biased subjective opinions. We may, for example, allow the referee to change only a fixed number of opinions by a factor of a fixed total. For example, in case of a matrix of order 4 when we have 6 assessments we may allow to modify a maximum of three modifications on condition that the total of all changes does not exceed say 3 (so three assessments may be modified by one up or down, or one assessment may be modified by 3 up or down).

Acknowledgments

This project has been supported in part by the Euro Research grant "Human Capital." The authors are grateful to Grant O. Duncan (Team Lead, Business Intelligence and Software Integration, Health Sciences North, Sudbury, Ontario) for his help with proofreading this text. The authors also acknowledges involvement of William O. Mackasey (the retired expert of abandoned mines, formerly employed by the Ministry of Northern Development and Mines). Numerous researchers on four continents (Australia, Asia, Europe, and North America) have been extremely supportive through this project and we would like to thank all of them.

References

[1] Bolger P.M., Duszak Z., Koczkodaj W.W., Mackasey W.O. 1993, Ontario Abandoned Mine Hazards Prioritizing - an Expert System Approach. In: Proceedings of the 15th Annual Abandoned Mine Land Conference, Jackson, Wyoming, September 13-15, 1993, pp. 370-388.

[2] Colonius, H., Representation and uniqueness of the Bradley-Terry-Luce model for pair comparisons, British Journal of Mathematical & Statistical Psychology, 33: 99–103, 1980.

[3] "sourceforge.net/directory/os:windows/?q=concluder",
retrieved 2016-03-10

[4] Duszak, Z.; Koczkodaj, W.W.; Generalization of a New Definition of Consistency for Pairwise Comparisons, Information Processing Letters, 52(5): 273-276, 1994.

[5] Faliszewski, Piotr; Hemaspaandra, Edith; Hemaspaandra, Lane A.; Rothe, J., Llull and Copeland Voting Computationally Resist Bribery and Constructive Control, Conference: 2nd International Workshop on Computational Social Choice Location: Liverpool, England, Journal of Artificial Intelligence Research, 35: 275-341, 2009.

[6] Fülöp, J.; A method for approximating pairwise comparisons matrices by consistent matrices *J. Global Optimization* **42**, 423-442 (2008)

[7] . Fülöp, W. W. Koczkodaj, S. J. Szarek, A different perspective on a scale for pairwise comparisons, Transactions of Computational Collective Intelligence I, LNCS 6220: 71-84, 2010.

[8] Heer, j.; Card, S.K.; Landay, J.A., "prefuse: a toolkit for interactive information visualization" in: Proceedings of the SIGCHI conference on Human factors in computing systems: 421-430, Portland, Oregon, USA: ACM, 2005.

[9] Koczkodaj, W.W., *A New Definition of Consistency of Pairwise Comparisons*, Mathematical and Computer Modelling, 18(7): 79-84, 1993.

[10] Koczkodaj, W.W., Pairwise Comparisons Rating Scale Paradox, Transactions on Computational Collective Intelligence XXII: 1-9, 2016.

[11] Koczkodaj, W.W.; Kulakowski, K.; Ligeza, A., On the quality evaluation of scientific entities in Poland Supported by consistency-driven pairwise comparisons method Scientometrics, 99(3): 911-926, 2014.

[12] Koczkodaj, W.W.; Szarek, S.J. , On distance-based inconsistency reduction algorithms for pairwise comparisons, Log. J. IGPL, 18(6): 859-869, 2010.

[13] Koczkodaj, W.W.; Kosiek, M.; Szybowski, J.; Xu, D., Fast convergence of distance-based inconsistency in pairwise comparisons, Fundamenta Informatice, 137: 355-367, 2015.

[14] Koczkodaj, W.W.; Szwarc, R.; Axiomatization of Inconsistency Indicators for Pairwise Comparisons, Fundamenta Informaticae, .132(4): 485-500, 2014.

[15] Koczkodaj, W.W.; Szybowski, J., Pairwise Comparisons Simplified, Applied Mathematics and Computation 253:387?394, 2015.

[16] Koczkodaj, W.W.; Szybowski, J.; Wajch, E.; Inconsistency indicator maps on groups for pairwise comparisons, *International Journal of Approximate Reasoning* **69**, no2, 81-90 (2016)

[17] Saaty, T.L., *A Scaling Methods for Priorities in Hierarchical Structure*, Journal of Mathematical Psychology, Vol. 15, 234-281, 1927.

[18] Thurstone, L.L., *A Law of Comparative assessments*, Psychological Reviews, 34, 273-286, 1927.

[19] Williams, C.; Crawford, G., Analysis of subjective judgment matrices, The Rand Corporation Report R-2572-AF, 1980, pp. 1–59.

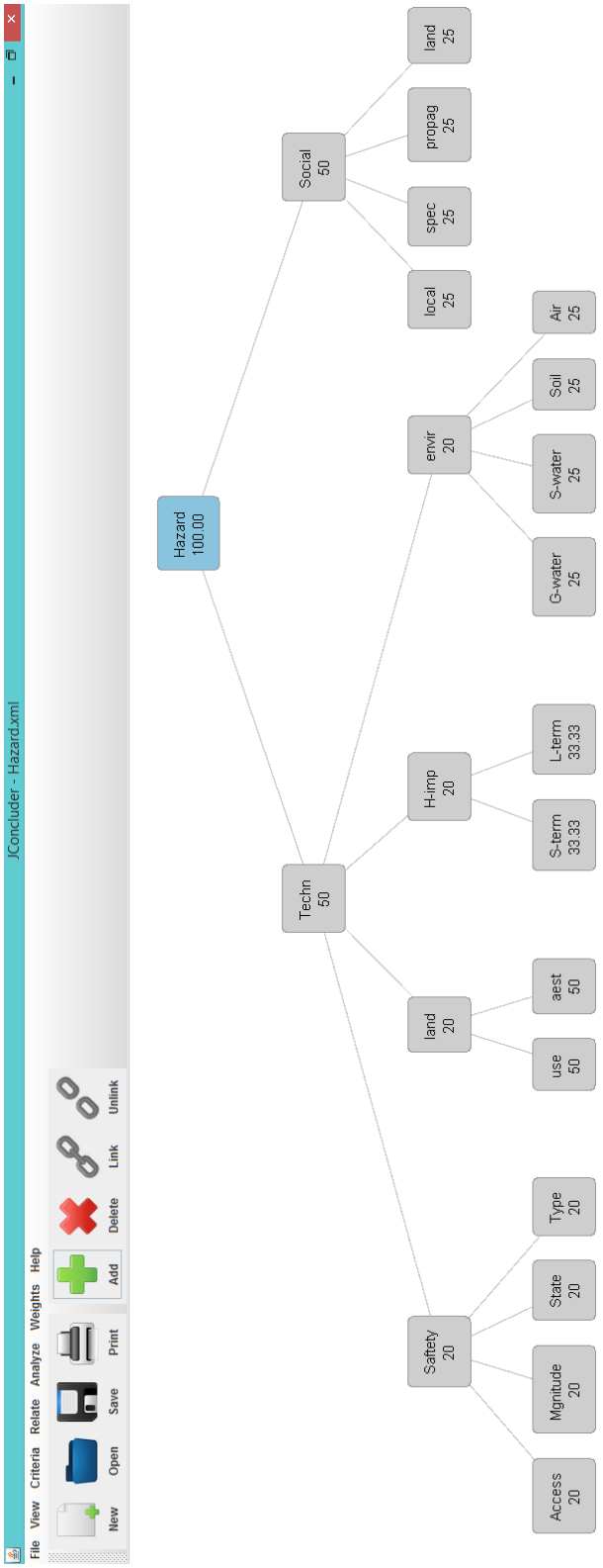


Figure 1: PC model for abandoned mines hazard rating

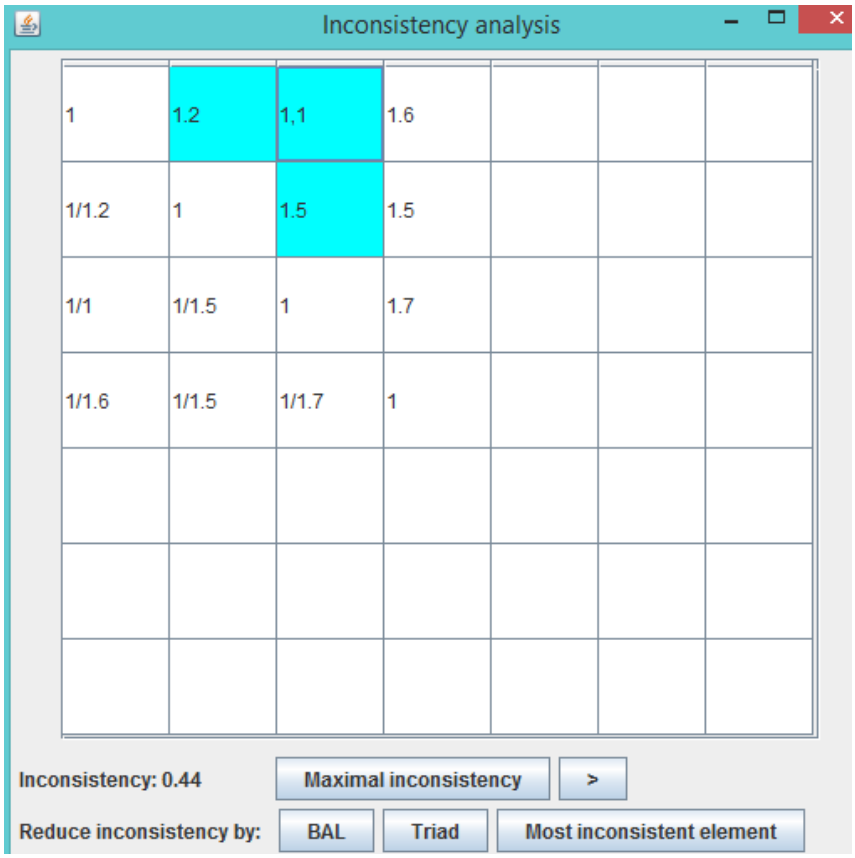


Figure 2: Inconsistency analysis

2.50%	Access
2.50%	State
2.50%	Mgnitude
2.50%	Type
0.83%	S-term
0.83%	L-term
0.21%	G-water
0.21%	S-water
0.21%	Soil
0.21%	Air
5.00%	use
5.00%	aest
12.50%	local
12.50%	spec
12.50%	propag
12.50%	land

Figure 3: The final weights

Computational models for design of concrete segments with symmetrical reinforcement bars under the action of bending moments and axial forces

† Li Shouju [1]*, Shangguan Zichang[2], and Feng Ying [1]

[1] State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China.

[2] Institute of Marine and Civil Engineering, Dalian Ocean University, Dalian, 116023, China

*Presenting author: lishouju@dlut.edu.cn

†Corresponding author: lishouju@dlut.edu.cn

Abstract

Shield-driven tunnels are widely adopted in the development of underground spaces for transportation and utility networks in soft soils. Numerical modeling has now become an important element in the design of underground excavations in soils and rocks. Numerical analysis can provide realistic representation of the field conditions taking into account key elements of the excavation such as the geomechanical characteristics of the ground and the in situ stress condition. In order to solve the problems of section design and verification of concrete segments, the computational models are proposed for designs of concrete segments with symmetrical reinforcement bars under the action of bending moments and axial forces. Based on the constitutive model of steel bars and similarity criterions of strains in beam section, the analytical expression of stress on reinforcement bars located in compressive region is derived. Influences of axial forces on the ultimate bearing bending moment of segments and the area of reinforcement bars in tension region are discussed through analyzing two practical underground tunnels with concrete segment linings. The investigation shows that the depth of compressive region increases ith increasing axial force on segment. The ultimate bearing bending moment of concrete segment increases with increasing axial force on segment when area of reinforcement bars is constant.

Keywords: Concrete segment, Section design and verification, Bending moment, Axial force, Computational model.

Introduction

Reinforced concrete segments are widely used in Metro tunnels, hydraulic tunnels and mining tunnels. The optimal design and analysis of ultimate bearing performance for concrete segments refer to safety and economic problems of underground structures. Especially with commonly using shield machine in underground engineering, the investigations about ultimate bearing performance for concrete segments have received general attentions in domestic and overseas. Jiang studied influences of hybrid tendons, load locations and joint numbers to flexural strength of fully segmental beams. For comparison purpose, a monolithic beam with hybrid tendons was also tested. The deflections, ultimate loads, stresses of prestressing strands and failure modes were investigated. At the ultimate stage, the stresses of all tendons are greater than 1500 MPa[1]. Caratelli performed full-scale tests on both traditional reinforced concrete and fiber reinforced elements. In particular, bending tests were carried out in order to compare the behaviour of the segments under flexural actions, while point load tests were developed with the aim of simulating the thrustforce induced by the Tunnel Boring Machine, and then the effect of load concentration and splitting phenomena[2]. Yan presented a comprehensive experimental study on the comparative behaviour of the reinforced concrete and the hybrid fibre reinforced concrete shield TBM tunnel lining segments exposed to fire. The tests were conducted using a newly developed test facility, which is capable of accommodating different mechanical loading and boundary conditions under different fire scenarios[3]. Shalabi proposed lining structure which was made of bolted and double gasketed precast concrete segment lining with convex to convex longitudinal joint surfaces. Lining evaluation included the sealant performance of different gasket materials under water pressure less than 90 psi. Testing program

was designed to evaluate the longitudinal joint and T-joint sealant behavior under static and dynamic loading using large scale concrete segments[4]. Nehdi investigated the mechanical performance of Ultra-high performance fiber-reinforced concrete tunnel lining segments. Flexural and edge-point load tests were conducted on 1/3-scale tunnel lining segments to evaluate its bending and thrust load resistance[5]. Zhang proposed a method based on the moment-force interaction and the effect of bolt pockets. The method considered that the load corresponding to the appearance of the first crack is the load of bond cracking, and assumed that the K-segment is a column which is subjected to axial loading and biaxial bending. Analytical results were compared with experimental values obtained from four reinforced-concrete K-segments[6].Amau studied the phenomena associated to coupling effects, determines the main involved parameters and analyzes their influence on a real lining structural response by means of a 3D numerical model. The comparison with the usual plane models currently employed in linings designs provide significant conclusions about the coupling effects implications and the conditions in which become more relevant[7].Analysis from Ye on the effective ratio of the transverse bending rigidity values under different load levels with different bolt pre-tightening forces and different assembly modes shows that value of the stagger-jointed segmental ring is obviously lager than that of the straight-jointed segmental ring, and that difference decrease gradually with the load increasing[8].Analysis from Moller shown that installation procedures are most important to be considered in order to arrive at proper predictions for tunneling settlements, horizontal deformations and lining forces. For the installation of closed face shield tunneling a novel simulation method is presented, named the grout pressure method. It is shown that the grout pressure method yields the best predictions for both ground movements and structural forces[9]. Do proposed the influence of joint rotational stiffness, the reduction in joint rotation stiffness under the negative bending moment, the lateral earth pressure factor and Young modulus of ground surrounding the tunnel should not be neglected. On the other hand, the results have also shown an insignificant influence of the axial and radial stiffness of the joints on segmental tunnel lining behavior[10]. The aim of the paper is to propose the relationship between the ultimate bending moment of concrete segment and axial force, analyze stress state of reinforced bars in compressive zone, investigate the worst loading combination between bending moment and axial force, and further develop computing models for evaluating ultimate bearing performances of concrete segments.

Current computing models for ultimate bearing performances of concrete segments

The concrete segments are idealized as column with loading of eccentric force N. Based on Code for design of concrete structures, assume that the deformed bars on compressive zone and tensile zone are yielded. An equivalent rectangular stress distribution is simplified with little loss in accuracy, as shown in Fig. 1

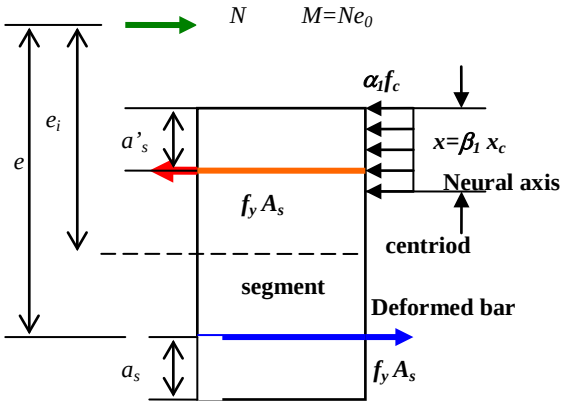


Figure 1. Idealized computational models for concrete segments with symmetrical reinforcement bars

It is assumed that the axial force, concrete grade and area of deformed bar are known. Based on the balance of forces acting on the section, as shown in Fig. 1, it is given by

$$N = \alpha_1 f_c b x \tag{1}$$

Where N is axial force,a_1 is stress coefficient, x is depth to neural axis, b is width of segment, f_c is compressive strength of concrete. The depth of neural axis is expressed as follows

$$x = \frac{N}{\alpha_1 f_c b} \tag{2}$$

The ultimate bearing bending moment of concrete segment is derived as

$$\begin{aligned} M_u &= N e_0 \\ &= \alpha_1 f_c \frac{N}{\alpha_1 f_c b} b(h_0 - 0.5\frac{N}{\alpha_1 f_c b}) \\ &+ (h_0 - a_s) f_y A_s - N(h/2 - a_s + e_a) \end{aligned} \tag{3}$$

Where e_0 is a distance (original eccentricity) from the centroid of deformed bar to axial force, M_u is ultimate bending moment, a_s is vertical distance from the joint point of all longitudinal tension bars to the cross section of the cross section, h is section height, h_0 is section effective height, h_0=h-a_s, A_s is reinforced area, f_y is tensile strength of reinforcement.

$$e_0 = e_i - e_a \tag{4}$$

Where e_i is a distance from the centroid of section to axial force accounting for adding eccentricity, as shown in Fig. 1. e_a is a adding eccentricity.

$$e_i = e - h/2 + a_s \tag{5}$$

$$e = \frac{\alpha_1 f_c x b(h_0 - 0.5x) + (h_0 - a_s) f_y A_s}{N} \tag{6}$$

Where e is a distance from the centroid of deformed bar in tensile zone to axial force.
Based on Code for design of concrete structures, evaluate the segment is in a state of small eccentricity or large eccentricity according to following formulas

$$N_{ub} = \alpha_1 f_c b x_b \tag{7}$$

Where N_{ub} is ultimate compressive force of segment under boundary condition, x_b is boundary depth to neural axis. If $N<N_{ub}$, then segment is in a state of large eccentricity; otherwise in a state of small eccentricity.

$$x_b = \xi_b h_0 \tag{8}$$

Where ζ_b is relative boundary depth to neural axis.

$$\xi_b = \frac{\beta_1}{1+\frac{f_y}{E_s \varepsilon_{cu}}} \tag{9}$$

Where E_s is elastic modulus of reinforcement, ε_{cu} is ultimate strain of concrete, ε_{cu}=0.0033. The calculating steps for ultimate bearing performance of concrete segment are listed as: Step 1) Evaluate the segment is in a state of small eccentricity or large eccentricity according to equation (7); Step 2) Calculate depth to neural axis according to equation (2); Step 3) Calculate a distance (eccentricity) from the centroid of deformed bar to axial force according to equation (4), (5) and (6); Step4) Calculate ultimate bearing bending moment of concrete segment according to equation (3).

New computing models for ultimate bearing performances of concrete segments

Accounting for the specifics of concrete segments in Metro tunnels, such as higher concrete grade, and section height of segments far less than section width of segments, the depth to neural axis is smaller and the stress of deformed bars in compressive zone is less than yield limit, and even the stress of deformed bars in compressive zone is in tensile state. So, based on current computing models, the practical stress of deformed bars in compressive zone is different from model solutions.

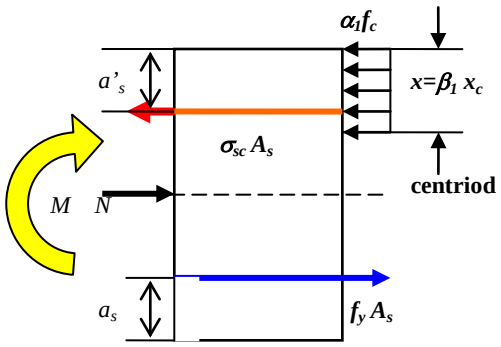


Fig. 2 Force and bending moment balances for column with eccentric compressive loading

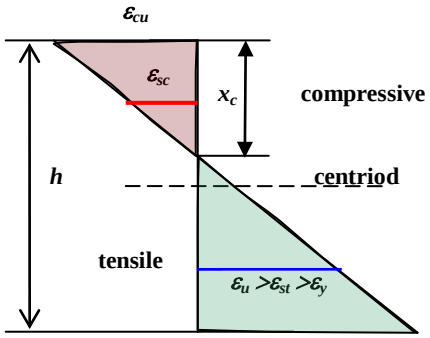


Fig. 3 Simplified strain distributions on concrete segment

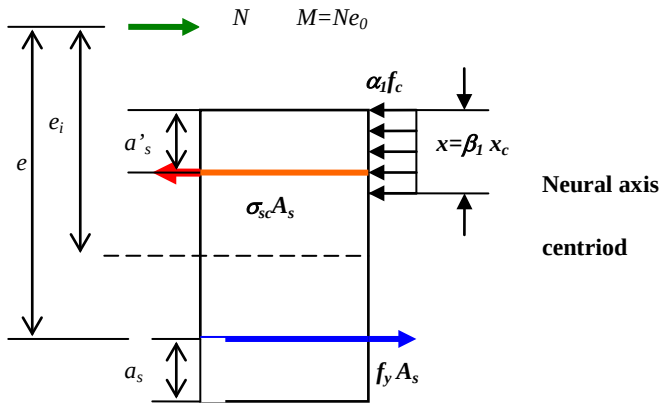


Fig. 4 Idealized computational models for concrete segments with symmetrical reinforcement bars and eccentric compressive loading

Based on force balance, as shown in Fig. 2 and 4, it is derived as

$$N = \alpha_1 f_c b x + \sigma_{sc} A_s - f_y A_s \tag{10}$$

Where σ_{sc} is compressive stress of reinforcement. The depth to neural axis is expressed as

$$x = \frac{N - \sigma_{sc} A_s + f_y A_s}{\alpha_1 f_c b} \tag{11}$$

Assume that the stress of deformed bar in compressive zone is less than yield limit, the relation between stress and strain for deformed bars is given by

$$\sigma_{sc} = \varepsilon_{sc} E_s \tag{12}$$

Under the action of axial force with an eccentricity e, and based on plane deformation assumption, as shown in Fig. 3, the relation between strain of deformed bars in compressive zone and ultimate strain of concrete is given by

$$\frac{\varepsilon_{sc}}{\varepsilon_{cu}} = \frac{x_c - a_s}{x_c} = (1 - \frac{\beta_1 a_s}{x}) \tag{13}$$

Were β_1 is a factor that is a function of the strength of the concrete, a_s is the distance from the center of tensile bars to inter surface of segment, as shown in Fig.4. x_c is distance from the outer compressive fiber to neural axis, and x is depth of neural axis for simplified equivalent rectangular. ε_{cu} is ultimate strain of concrete, ε_{cu}=0.0033. If ε_{sc}>0, then the stress of deformed bars in compressive zone is in compressive state; otherwise in tensile state.

$$\varepsilon_{sc} = \varepsilon_{cu} (1 - \frac{\beta_1 a_s}{x}) \tag{14}$$

Substitute equation (14) into equation(12), it is obtained

$$\sigma_{sc} = E_s \varepsilon_{cu} (\frac{x - \beta_1 a_s}{x}) \tag{15}$$

Substitute equation (15) into equation (11), it is obtained

$$x = \frac{N - E_s \varepsilon_{cu} (\frac{x - \beta_1 a_s}{x}) A_s + f_y A_s}{\alpha_1 f_c b} \tag{16}$$

$$\alpha_1 f_c b x^2 - (N + f_y A_s + E_s \varepsilon_{cu} A_s) x - E_s \varepsilon_{cu} A_s \beta_1 a_s = 0 \tag{17}$$

The depth to neural axis is solved by equation (17), and then the stress of deformed bars in compressive zone is obtained by equation (15). Based on the balance principle of force moment, as shown in Fig.3, it is obtained

$$Ne = \alpha_1 f_c x b (h_0 - 0.5x) + (h_0 - a_s) \sigma_{sc} A_s \tag{18}$$

$$e = \frac{\alpha_1 f_c x b (h_0 - 0.5x) + (h_0 - a_s) \sigma_{sc} A_s}{N} \tag{19}$$

$$\begin{aligned} e_0 &= e - h/2 + a_s - e_a \\ &= \frac{\alpha_1 f_c x b (h_0 - 0.5x) + (h_0 - a_s) \sigma_{sc} A_s}{N} - (h/2 - a_s + e_a) \end{aligned} \tag{20}$$

The ultimate bearing bending moment of concrete segment is derived as

$$\begin{aligned} M_u &= Ne_0 \\ &= \alpha_1 f_c x b (h_0 - 0.5x) + (h_0 - a_s) \sigma_{sc} A_s - N(h/2 - a_s) \end{aligned} \tag{21}$$

The calculating steps of proposed new models for ultimate bearing performance of concrete segment are listed as: Step 1) Evaluate the segment is in a state of small eccentricity or large eccentricity according to equation (7); Step 2) Calculate depth to neural axis according to equation (17); Step 3) Calculate the stress of deformed bars in compressive zone according to equation (15); Step 4) Calculate a distance (eccentricity) from the centroid of deformed bar to axial force according to equation (20); Step5) Calculate ultimate bearing bending moment of concrete segment according to equation (21).

Case study for two Metro tunnels

In order to investigate the differences between current models and new proposed models, two practical Metro tunnels with concrete segment lining are studied. The stress distribution in deformed bars and ultimate bending moment are calculated respectively. The drawbacks of current models are discussed in detail. The first practical engineering example is Beijing Metro tunnel[11]. The maximum embedded depth of tunnel is 10. 31m. The outer diameter of tunnel segment is 6.0m. The height of segment is 300mm. The width of segment is 1.2m. Concrete Grade is C50 with symmetrical reinforcement bars. The yield limit of bars is 300MPa. The area of deformed bars is

2514mm2 both for compressive zone and tensile zone, respectively. The distance from the center of tensile bars to inter surface of segment is as=40mm.

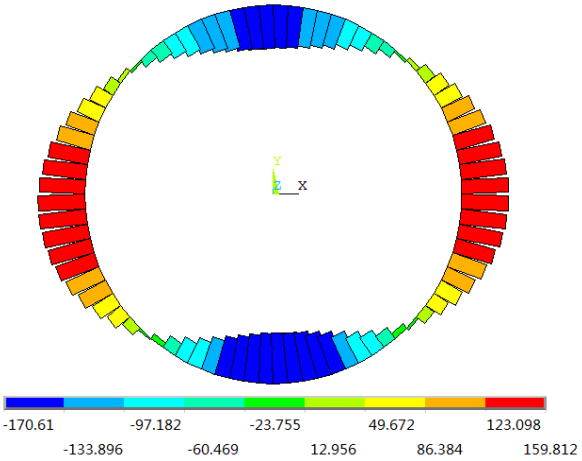


Fig. 5 Bending moment distributions on concrete segments (Unit: kNm, Beijing Metro tunnel)

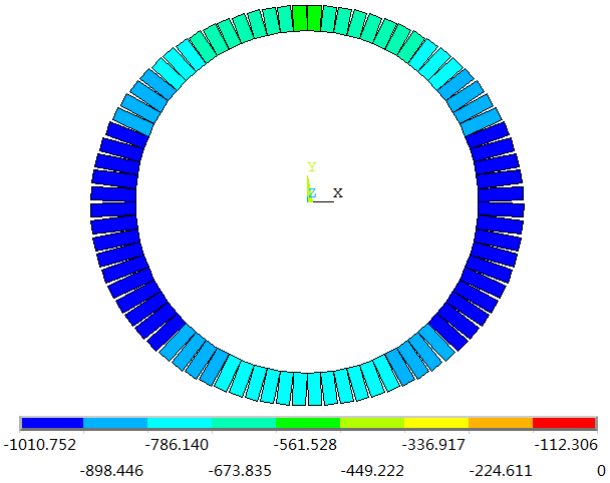


Fig. 6 Axial force distributions on concrete segments (Unit: kN, Beijing Metro tunnel)

Finite element method is used to compute the internal force distributions on segments for Beijing Metro tunnel , as shown in Fig. 6 and 7. It is obtained from Fig.6 and 7 that The maximum bending moment on segments is 160 kNm. The axial force on segments is in compressive state and varied from 400kN to1010kN. Variations of ultimate bending moment of segments versus axial force are listed in Table 1 and 2.

Table 1 Variation of ultimate bending moment of segments versus axial force (Current model, Beijing Metro tunnel)

Axial force /kN	400	500	600	700	800	900	1000
x/mm	14.4	18.1	21.6	25.25	28.8	32.5	36.1
M_u/*kN·m*	215	226	237	248	258	268	278
σ_{sc} /*MPa*	300	300	300	300	300	300	300

Table 2 Variation of ultimate bending moment of segments versus axial force (New proposed model, Beijing Metro tunnel)

Axial force /kN	400	500	600	700	800	900	1000
x/mm	35.6	37.0	38.6	40.2	41.9	43.6	45.5

M_u /kN·m	224	233	242	251	260	269	278
σ_{sc} /MPa	66	90	113	134	156	176	196

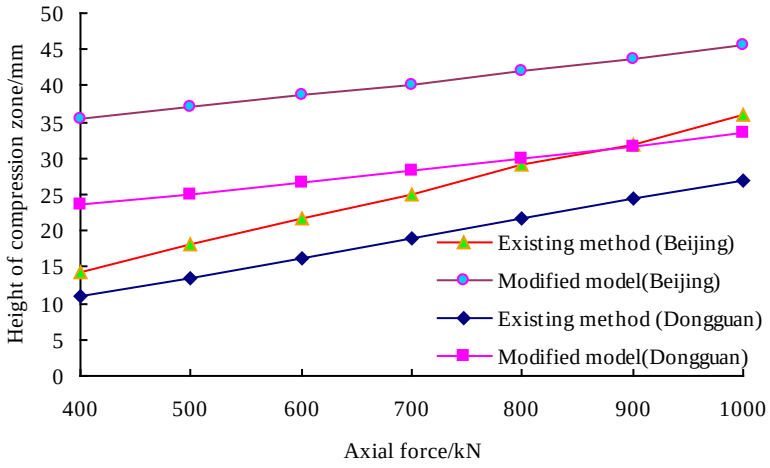


Fig. 7 Variation of compressive depth versus axial forces

It is found from Fig.7 that the compressive depth in concrete segment increases with increasing axial force, and the compressive depth in new proposed model is larger than one in current model.

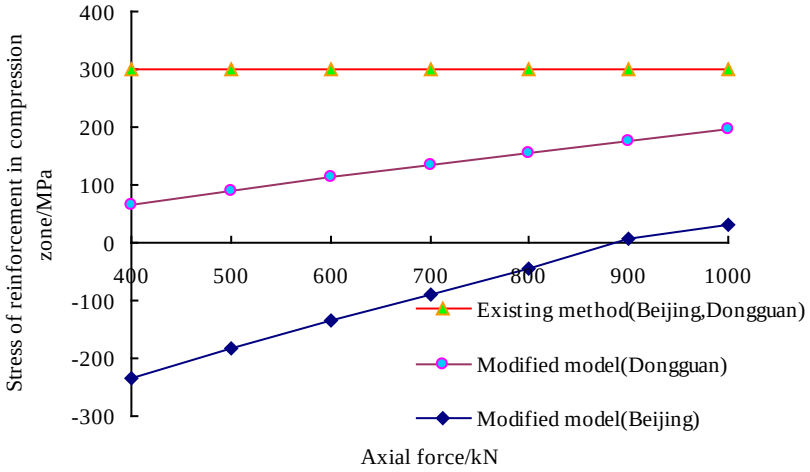


Fig. 8 Variation of stress of reinforced bars in compressive zone versus axial forces

It is found from Fig.8 that the stress of reinforced bars in compressive zone increases with increasing axial force. The stress of reinforced bars in compressive zone is less than yield strength, especially for Dongguan tunnel, the stress of reinforced bars in compressive zone is in tensile state.

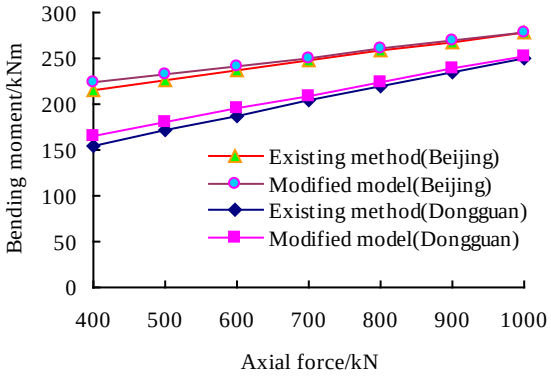


Fig. 9 Variation of ultimate bending moment of segments versus axial force

It is observed from Fig. 9 that the ultimate bending moments of segments for both current and new proposed models are nearly same even if the compressive depths and stresses of reinforced bars in compressive zone for two models are different. The second practical engineering example is Dongguan-Huizhou tunnel under water[12]. The maximum embedded depth of tunnel is 16. 2m. The maximum water depth is 16. 2m. The outer diameter of tunnel segment is 8.5m. The height of segments is 400mm. The width of segments is 1.6m. Concrete Grade is C50 with symmetrical reinforcement bars. The yield limit of bars is 300MPa. The area of deformed bars is 882mm2 both for compressive zone and tensile zone, respectively. The distance from the center of tensile bars to inter surface of segment is as=40mm. The maximum bending moment acting on segments is 90kNm. The maximum axial force is 100kN. Variations of ultimate bending moments of segments versus axial force are listed in Table 3 and 4, respectively.

Table 3 Variation of ultimate bending moment of segments versus axial force (Current model, Dongguan tunnel)

Axial force /kN	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
x/mm	10.8	13.5	16.2	18.9	21.6	24.4	27.0
M_u*//kN·m*	154	171	188	204	220	235	253
σ_{sc} */MPa*	300	300	300	300	300	300	300

Table 4 Variation of ultimate bending moment of segments versus axial force (New proposed model, Dongguan tunnel)

Axial force /kN	400	500	600	700	800	900	1000
x/mm	23.6	25.0	26.6	28.2	30.0	31.6	33.5
M_u */kN·m*	165	180	195	209	224	239	251
σ_{sc} */MPa*	-235	-182	-134	-88	-46	7	30

Note: Negative represents deformed bars in tensile state.

Conclusions

1) The investigation validates that the stress of reinforced bars in compressive zone increases with increasing axial force. The compressive depth in concrete segment is far less than 2as, so the stress of reinforced bars in compressive zone is less than yield strength. Especially for Dongguan tunnel, the stress of reinforced bars in compressive zone is in tensile state.

2) Two practical Metro tunnels with concrete segment lining are computed by using two different models. The results show that the compressive depth in concrete segment increases with increasing axial force, and the compressive depth in new proposed model is larger than one in current model. The worst loading combination is maximum bending moment with minimum axial force.
3) It is observed that the ultimate bending moments of segments for both current and new proposed models increase with increasing axial force, and are nearly same even if the compressive depths and stresses of reinforced bars in compressive zone for two models are obviously different. The proposed computing model can precisely calculate the stresses of reinforced bars in compressive zone.

Acknowledgments

The research described in this paper was financially supported by the National Key Basic Research and Development Program of China (Grant No. 2015CB057804), the National Natural Science Foundation of China (Grant No. 11572079) and Opening Foundation of State Key Laboratory of Structural Analysis for Industrial Equipment (Grant No. S14206).

References

[1] Jiang, H. B., Cao Q., Liu, A. R.. (2016) Flexural behavior of precast concrete segmental beams with hybrid tendons and dry joints. *Construction and Building Materials* **110**, 1-7.

[2] Caratelli, A., Meda, A., Rinaldi, Z.. (2011) Structural behaviour of precast tunnel segments in fiber reinforced concrete. *Tunneling and Underground Space Technology* **26**, 284-291.

[3] Yan, Z. G., Yi, S., He, H. Z.. (2015) Experimental investigation of reinforced concrete and hybrid fibre reinforced concrete shield tunnel segments subjected to elevated temperature. *Fire Safety Journal* **71**, 86-99.

[4] Shalabi, F., Cording, E., Paul, S.. (2012) Concrete segment tunnel lining sealant performance under earthquake loading. *Tunnelling and Underground Space Technology* **31**, 51-60.

[5] Nehdi, M., Abbas, S., Soliman, A.. (2015) Exploratory study of ultra-high performance fiber reinforced concrete tunnel lining segments with varying steel fiber lengths and dosages. *Engineering Structures* **101**, 733-742.

[6] Zhang, W. J., Koizumi, A..(2007) A study of the localized bearing capacity of reinforced concrete K-segment. *Tunnelling and Underground Space Technology* **22**, 467-473.

[7] Arnau, O., Molins, C.. (2013) Three dimensional structural response of segmental tunnel linings, *Engineering Structures* **44** , 210 -221.

[8] Ye, F., Gou, C. F., Sun, H. D.. (2014) Model test study on effective ratio of segment transverse bending rigidity of shield tunnel, *Tunnelling and Underground Space Technology* **41**, 193-205.

[9] Moller, S. C., Vermeer, P. A.. (2008) On numerical simulation of tunnel installation, *Tunneling and Underground Space Technology* **23**, 461-475.

[10] Anh, N., Do, D. D., Oreste, P.. (2013) 2D numerical investigation of segmental tunnel lining behavior. *Tunnelling and Underground Space Technology* **37** ,115-127.

[11] Chen, D..(2009) . Comparative analysis of designing methods of tunnel segments for Beijing Metro. *Railway Standard Design* **10**, 60-64.

[12] Hu, H., Zhang, L., Qiu, W. G.. (2012) Comparative analysis of three kinds of designing methods of tunnel segments and field experiment investigation. *Hydrogeology and Engineering Geology* **39**, 72-76.

Complex normal form method for nonlinear free vibration of a cantilever nan-obeam with surface effects

Demin Zhao†

Department of Engineering Mechanics, College of Pipeline and Civil Engineering, China University of Petroleum (East China), Qingdao 266580, China
†Corresponding author: zhaodemin@upc.edu.cn

Abstract

Nano-beams and nanowires are widely used as building blocks in the rapid development of Nano/Micro-electro-mechanical system (N/MEMS), micro-sensors, energy harvesting and storage devices, etc., and their vibration behaviors have aroused great concerns in both pure science and engineering applications. In this study, we investigate the nonlinear free vibration of a nano-beam considering its surface effects, including the surface elasticity and the residual surface stress. Firstly, a mechanics model on the transverse vibration of a cantilever nano-beam is developed according to Hamilton's principle. In use of the Galerkin and complex normal form methods, the approximate analytical solution of the nonlinear equation is obtained, which has been confirmed by the numerical simulation. The present work can provide theoretical basis for the precise design of nanowires or nanofibers in atomic force microscopy, generators and nano-sensors in electronic devices.

Keywords: Surface elasticity，Residual surface stress，Complex normal form method，Quasi-periodic motion，Chaos

1 Introduction

Nano-beams or nanowires, due to their perfect advantages of greatly magnified sensitivity, increased reliability and reduced sizes, have been highly advanced in Nano/Micro-electro-mechanical systems (N/MEMS), biotechnology, sensors, actuators, resonators and atomic force microscopy [1-3]. Owing to the extremely high surface to volume ratio, surface effects of nanowires have become more important factors than the volumetric forces, which are crucial to their mechanical performance. Based on a number of results from experiments and atomic simulations, the residual surface stress and surface elasticity have proved to be the key origins of the size-dependent properties for most nanomaterials or nanostructures [4].

It has already been known that surface effects can significantly affect the static properties of nanowires, such as the internal force diagram, modulus, deflection and buckling of nano-beams [5–8], and these phenomena have attracted growing interest of many scholars. However, the more interesting issue for a nano-beam is its dynamic response, because it is an essential techinique to measure the dynamic parameters in vibration, in order to characterize the bending stiffness [1]. As a consequence, a great deal of work has been performed to investigate the natural frequency and amplitude of a nano-beam in linear vibration [9–13]. For example, Wang and Feng [9] deduced the natural frequency of a simply supported nano-beam, in consideration of both the surface elasticity and residual surface stress, which indicates that the frequency may be enhanced with a positive residual surface stress and reduced by a negative one. Based on the same model, He and Lilley [10] studied the influence of surface effects on the first-mode natural frequency for a nano-beam with different boundary

conditions. Their theoretical solutions show that the positive surface stress can alter the natural frequency for a cantilever, a simply supported beam and a fixed-fixed nano-beam. From the different viewpoint, Ansari and Hosseini *et al*. [11] explored the impact of surface effects on the natural frequency of a nano-beam in use of the compact finite difference method. They claimed that the surface effects on the natural frequency are dependent on the aspect ratio and thickness of the beam. It should be stressed that for a nano-beam with small aspect ratio, the Timoshenko beam model is normally utilized to analyze its transverse vibration, which is more accurate than the Euler-Bernoulli beam model mentioned previously [12, 13].

Moreover, it is necessary to consider the nonlinear vibration of nanowires in many applications [14, 15]. For example, it is desirable to reduce the size of N/MEMS and achieve high-output energy, but this requires the nano-beam or nanowire in N/MEMS come into operation near the nonlinear working regime [14]. For instance, Chen and Hu *et al*. [16] developed a periodicity-ratio approach to calculate the response of a piezoelectric laminated micro-beam system actuated by AC and DC voltages, and the periodic and chaotic region diagrams were plotted. Miandoab and Yousefi-Koma *et al*. [17] studied the chaotic behaviors of a nano-resonator in MEMS/NEMS subjected to electrostatic forces, and they found that the system undergoes homoclinic and heteroclinic bifurcations in the appearance of chaos. Yet in the real situations, besides such normal nonlinear factors as curvature, geometry nonlinearities and the coupling of multi-fields [18–21], surface effects play an important role in the nonlinear vibration of nano-beams. One example is that, taking the von-Karman geometric nonlinear strain into account, Gheshlaghi and Hasheminejad [22] studied the influence of residual surface stress on the free nonlinear vibration of a nano-beam and got the exact expressions of the natural frequency and vibration amplitude. In addition, Moeenfard and Mojahedi *et al*. [23] used the Homotopy Perturbation Method to analyze the nonlinear free vibration of a clamped-clamped or a clamped–free nano-beam, and the effects of axial loads, rotary inertia, shear deformation and slenderness ratio on the natural frequency have been discussed.

It should be stressed that, to fully probe the dynamics of a nano-beam, it is imperative to consider its semi-analytical solution, and more importantly, surface effects are necessary to be analyzed in this model. To the best of our knowledge, there is hitherto a lack of systematic exploration on this issue, for there is a very strong coupling between the nonlinear factors and surface effects. Therefore, we concentrate on the nonlinear free vibration of a nano-beam, towards extending our understandings on the solution, which is helpful to better design N/MEMS, micro-sensors, energy harvesting and storage devices.

The present paper is organized as follows. In Section 2, the dynamics equation of the free vibration of a cantilever beam including surface effects is derived based on Hamilton Principle. The Galerkin method is then adopted to discrete the partial differential equation into ordinary differential equations in Section 3. Next, in Section 4, we analyze the solution and its stability on the nonlinear vibration by using the complex normal form method (CNFM), and the numerical simulation is followed. Finally, conclusion are given in detail.

2 Kinematics

2.1 Surface effects model

Generally speaking, “surface effects” of nano-materials are mainly attributed to the surface energy or surface stresses on the solid surface [24, 25]. The body and surface layer of a nanowire can be abstracted as a composite beam with a core-shell structure, which includes a

solid core with a Young's modulus and a surface layer with a surface modulus [9], as schematized in Fig. 1. The thickness of the surface layer is normally negligible.

In light of the generalized Young–Laplace equation [9], the residual surface stress in the surface layer of the nano-beam can induce a jump of the normal stress across the interface between the bulk and the surface, and this leads to a transversely distributed pressure $q_s(s)$ along the axial direction of the beam, namely

$$q_s(s) = H\kappa, \tag{1}$$

where H is a constant parameter correlated with the residual surface stress and the cross-sectional shape. The parameter H for a nano-beam with a circular cross section is normally expressed as $H = 2\tau_0 D$ [9], where D is the diameter of the cross section.

For the nano-beam with a circular cross section, the effective bending stiffness $(EI)^*$ can be further modified according to the composite beam model, which is expressed as [6, 9, 10]

$$(EI)^* = \frac{\pi}{64} ED^4 + \frac{\pi}{8} E^s D^3, \tag{2}$$

and the effective tensile stiffness $(EA)^*$ is given by

$$(EA)^* = \frac{\pi}{4} ED^2 + E^s \pi D, \tag{3}$$

where E is the Young's modulus of the bulk material, E^s the surface elastic modulus, I the moment of inertia on the cross section, and A is the area of the cross section.

2.2 Vibration equation

We consider a cantilever nano-beam with surface effects, as shown in Fig. 1. The length of the beam is L, and the mass per unit length is m. Refer to a Cartesian coordinate system (O-xy). The model is assumed to be an Euler-Bernoulli beam, whose axis is initially along the x direction and then it can oscillate in the (x, y) plane [26]. As schematized in Fig. 1, the location of an arbitrary material point B in the beam axis transfers to the position of point B_1 after deformation. Let ϕ be the slope angle between the tangential line of the beam axis and the horizontal line at any point in the axis. We also introduce the curvilinear coordinate, i.e. the arc length s along the axis of the beam, starting from the origin.

To analyze the nonlinear vibration of the beam, the nonlinear effects on the deformation must be incorporated, so the infinitesimal deformation model can not be adopted here. In fact, at any point in the beam axis, there are the following geometric relations:

$$\cos\phi = x', \ \sin\phi = y'. \tag{4}$$

Taking derivative with respect to the arc length s on both sides of the second equation in Eq. (4), one can get the expression of the planar curvature

$$\kappa = \phi' = \frac{1}{\cos\phi} y'' = \frac{y''}{\sqrt{1-y'^2}}, \tag{5}$$

where $(\ )' = \frac{\mathrm{d}(\ )}{\mathrm{d}s}$ and $(\ )'' = \frac{\mathrm{d}^2(\ )}{\mathrm{d}s^2}$.

In use of the Hamilton's principle, one has

$$\int_0^T \delta L(y,T)\mathrm{d}T + \int_0^T \delta W(y,T)\mathrm{d}T = 0, \tag{6}$$

where time is denoted by T and the Lagrangian function is L=U–K.

The strain energy, kinetic energy and external work of the beam can be respectively given by

$$U=\frac{(EI)^*}{2}\int_0^L \kappa^2 \mathrm{d}s=\frac{(EI)^*}{2}\int_0^L \frac{y''^2}{1-y'^2}\mathrm{d}s,\ K=\frac{m}{2}\int_0^L\left(\frac{\mathrm{d}y}{\mathrm{d}T}\right)^2\mathrm{d}s,\ W=\int_0^L q_s y\mathrm{d}s. \tag{7}$$

By virtue of the variational principle, the governing equation of the free vibration for the nano-beam can be deduced as

$$m\ddot{y}+(EI)^*\left(y''''+y''''y'^2+4y'y''y'''+y''^3\right)-H\left(y''+\frac{1}{2}y'^2y''\right)=0. \tag{8}$$

The initial conditions and fixed boundary conditions of the beam are $y(0,T)=0$, $y'(0,T)=0$, $y''(L,T)=0$, $y'''(L,T)=0$.

Introducing the following non-dimensional quantities $w=\frac{y}{L}$, $\xi=\frac{s}{L}$, $\omega^*=\frac{1}{L^2}\sqrt{\frac{(EI)^*}{m}}$, $t=\omega^*T$, $\beta=\frac{HL^2}{(EI)^*}$, the governing equation can be recast as

$$\ddot{w}+w^{(4)}-\beta\left(w''+\frac{1}{2}w'^2w''\right)+\left(w''''w'^2+4w'w''w'''+w''^3\right)=0, \tag{9}$$

where $\dot{w}=\frac{\mathrm{d}w}{\mathrm{d}t}$, $\ddot{w}=\frac{\mathrm{d}^2w}{\mathrm{d}t^2}$, $w'=\frac{\mathrm{d}w}{\mathrm{d}\xi}$, $w''=\frac{\mathrm{d}^2w}{\mathrm{d}\xi^2}$, $w'''=\frac{\mathrm{d}^3w}{\mathrm{d}\xi^3}$, and $w^{(4)}=\frac{\mathrm{d}^4w}{\mathrm{d}\xi^4}$.

3 Galerkin Method

It is noticed that Eq. (9) is a high order and nonlinear partial differential equation (PDE), and it is nearly impossible to find the close-formed solution at hand. Herein, we use the Garlerkin discretization method to transform the PDE to the ordinary differential equations (ODEs).

We select the two-mode approximation on the solution, which is of adequately accuracy to analyze the nonlinear vibration. Assumes $w(\xi,t)$ can be approximated by the two-mode solution as below:

$$w(\xi,t)=q_1(t)\phi_1(\xi)+q_2(t)\phi_2(\xi), \tag{10}$$

where $q_1(t)$ and $q_2(t)$ represent the amplitudes of the first and second principal modes, respectively. The first and the second mode shape functions $\phi_1(\xi)$ and $\phi_2(\xi)$ can be described as:

$$\phi_i(\xi)=\cosh r_i\xi-\cos r_i\xi+\frac{\sin r_i-\sinh r_i}{\cos r_i+\cosh r_i}(\sinh r_i\xi-\sin r_i\xi)\ (i=1,2), \tag{11}$$

where r_i is governed by the frequency equation on a cantilever beam: $\cos(r_i)\cosh(r_i)=-1$. We first substitute Eqs. (10) and (11) into (9), and the obtained equation is multiplied by $\phi_1(\xi)$ or $\phi_2(\xi)$, then the integration of the product from 0 to 1 yields a second-order differential equation, where the orthogonal property of trigonometric functions is used. As a result, the ODEs on $(q_1,q_2)^{\mathrm{T}}$ are presented as:

$$\begin{cases}\ddot{q}_1+a_{11}q_1+a_{12}q_2+\left(a_{13}q_1^3+a_{14}q_1^2q_2+a_{15}q_1q_2^2+a_{16}q_2^3\right)=0\\ \ddot{q}_2+a_{21}q_1+a_{22}q_2+\left(a_{23}q_1^3+a_{24}q_1^2q_2+a_{25}q_1q_2^2+a_{26}q_2^3\right)=0\end{cases}, \tag{12}$$

where the values of the parameters a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, a_{21}, a_{22}, a_{23}, a_{24}, a_{25}, a_{26} are shown in Appendix A.

4 Solution analysis: CNFM

Next, we use the CNFM to solve the nonlinear ODEs in Eq. (12), where the two quantities q_1 and q_2 are coupled together [27]. Firstly, we assume its solutions can be formulated as

$$\begin{cases} q_1 = \zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2 \\ q_2 = \Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right) \end{cases}, \tag{13}$$

where $\bar{\zeta}_1$, $\bar{\zeta}_2$ are the complex conjugates of ζ_1 and ζ_2, respectively, and their normal expressions are

$$\zeta_i = A_i e^{j\omega_i t}\ ,\ \bar{\zeta}_i = \bar{A}_i e^{-j\omega_i t}\ (i=1,2), \tag{14}$$

where ω_1 and ω_2 are two frequencies, and $j^2 = -1$.

The first step is to insert Eq. (13) into the linear part of Eq. (12), and then we can derive the frequency equation on ω_i

$$\omega^4 - \left(a_{11} + a_{22}\right)\omega^2 + a_{11}a_{22} - a_{12}a_{21} = 0. \tag{15}$$

Moreover, the parameters Δ_1 and Δ_2 are given as

$$\Delta_i = -\frac{a_{11} - \omega_i^2}{a_{12}} = -\frac{a_{21}}{a_{22} - \omega_i^2}\ (i=1,2), \tag{16}$$

which are both real numbers.

In combination with Eq. (13) and its derivatives with respect to *t*, it can reach an equation group including four equations. Consequently, the complex variable equations on the parameters ζ_1 and ζ_2 are listed as

$$\dot{\zeta}_1 = j\omega_1\zeta_1 - j\begin{Bmatrix} b_{11}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)^3 \\ +b_{12}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)^2\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right) \\ +b_{23}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right)^2 \\ +b_{14}\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right)^3 \end{Bmatrix}, \tag{17}$$

$$\dot{\zeta}_2 = j\omega_2\zeta_2 - j\begin{Bmatrix} b_{21}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)^3 \\ +b_{22}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)^2\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right) \\ +b_{23}\left(\zeta_1 + \bar{\zeta}_1 + \zeta_2 + \bar{\zeta}_2\right)\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right)^2 \\ +b_{24}\left(\Delta_1\left(\zeta_1 + \bar{\zeta}_1\right) + \Delta_2\left(\zeta_2 + \bar{\zeta}_2\right)\right)^3 \end{Bmatrix}, \tag{18}$$

where the expressions of b_{11}, b_{12}, b_{13}, b_{14}, b_{21}, b_{22}, b_{23}, b_{24} are all shown in Appendix B.

In order to simplify the above equation, we introduce the following near-identity coordinate transformations

$$\begin{cases} \zeta_1 = \eta_1 + h_1\left(\eta_1, \bar{\eta}_1, \eta_2, \bar{\eta}_2\right) \\ \zeta_2 = \eta_2 + h_2\left(\eta_1, \bar{\eta}_1, \eta_2, \bar{\eta}_2\right) \end{cases}, \tag{19}$$

where the functions h_1 and h_2 are expressed in Appendix C.
Inserting Eq. (19) into Eqs. (17) and (18), and if the values of $\Lambda_{1,i}$, $\Lambda_{2,i}$ (i=1, …, 20) in Eq. (19) are properly chosen (shown in Appendix C), one can get the simplest normal forms of Eqs. (17) and (18):

$$\begin{cases} \dot{\eta}_1 = j\omega_1\eta_1 - \varepsilon j\left(c_{11}\eta_1\eta_2\bar{\eta}_2 + c_{12}\eta_1^2\bar{\eta}_1\right) \\ \dot{\eta}_2 = j\omega_2\eta_2 - \varepsilon j\left(c_{21}\eta_1\bar{\eta}_1\eta_2 + c_{22}\eta_2^2\bar{\eta}_2\right) \end{cases}, \tag{20}$$

where ε is small perturbation parameter, and the expressions of the symbols c_{11}, c_{12}, c_{21}, c_{22} are all given in Appendix D.

If the quasi-periodic solutions of the equation exist, the expression of η_1 and η_2 can be written in the polar form

$$\begin{cases} \eta_1 = \dfrac{1}{2}a_1 \exp j\left(\omega_1 t + \theta_1\right) \\ \eta_2 = \dfrac{1}{2}a_2 \exp j\left(\omega_2 t + \theta_2\right) \end{cases}, \tag{21}$$

where a_1, a_2 are amplitudes and θ_1 and θ_2 are phase angles, which are all real numbers.

Substituting Eq. (21) into Eq. (20), and separating the real and imaginary parts yields

$$\begin{cases} \dot{a}_1 = 0 \\ \dot{a}_2 = 0 \\ a_1\dot{\theta}_1 = -\dfrac{1}{4}a_1a_2^2c_{11} - \dfrac{1}{4}a_1^3c_{12} \\ a_2\dot{\theta}_2 = -\dfrac{1}{4}a_1^2a_2c_{21} - \dfrac{1}{4}a_2^3c_{22} \end{cases}, \tag{22}$$

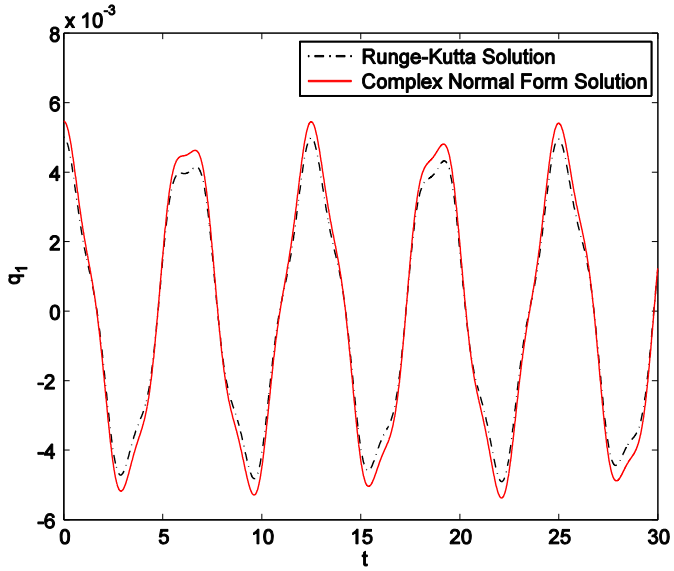
The above equations tells us that, a_1 and a_2 must be constants, and θ_1 and θ_2 can be solved when the a_1 and a_2 are given.

Up to now, the solutions on q_1 and q_2 can be acquired, and their expressions are shown in Appendix E. Therefore, the embryo conclusion is that, since a_1 and a_2 can be taken as arbitrary values, the considered system must have multiple quasi-periodic trajectories; and with different initial conditions, the system will converge to different quasi-periodic trajectories.

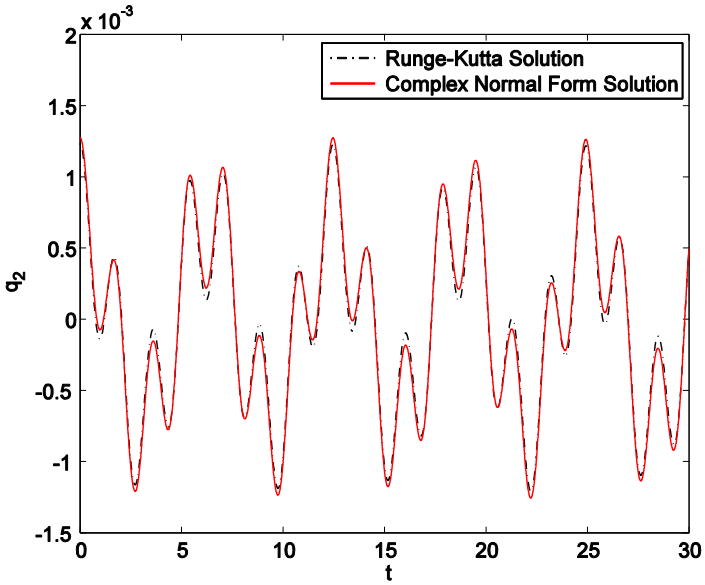
5 Validation of CNFM

To verify the validity of the CNFM, the numerical simulation based on the Runge-Kutta method is performed, where the time span is selected from 0 to 40 seconds. In the simulation process, the physical parameters of the nano-beam are chosen as [5, 8]: D=50 nm, L=500 nm, the mass per meter m=3.7876×10^{-11} kg/m, the bending stiffness $\left(EI\right)^*$ =2.33×10^{-20} Pa$\cdot$m^4 and tensile stiffness $\left(EA\right)^*$ =1.49×10^{-4} Pa$\cdot$m^4. The residual surface stress τ_0, which can be either positive or negative depending on the crystallographic structure for different nanomaterials, is selected in the regime from –2 to 2 N/m [8]. The simulation program is realized in MATLAB using the 4$^{\text{th}}$-order Runge-Kutta method, where the time step is set as 0.02 seconds.

We only study the trivial singular point near the center $(a_2, \theta_1, a_2, \theta_2)$=(0, 0, 0, 0), where multiple stable quasi-periodical solutions exist. For the CNFM, the parameters are selected as: a_1=0.005, a_2=0.0005. Correspondingly, the initial parameters in the numerical simulation are selected as $(q_1, \dot{q}_1, q_2, \dot{q}_2)\big|_{t=0}$ =(0.0055, 0, 0.00129, 0) to ensure they have the same initial values as those of the CNFM. The time history diagrams on q_1 and q_2 are shown in Fig. 2(a) and (b), respectively. From the figure it is clearly seen that the two curves from the Runge-Kutta method and CNFM nearly overlap with each other. This manifests that the CNFM is efficient to analyze the vibration of this system with small magnitude $|q_1| < 0.01$ and $|q_2| < 0.01$, as this method is applicable in weak nonlinear systems.



(a)



(b)

Fig. 2 Comparison between the complex normal form solution and the Runge-Kutta solution, (a) q_1, (b) q_2.

6 Conclusion

In conclusion, the nonlinear free vibration of a cantilever nano-beam has been systematically investigated, and the surface effects are considered. The CNFM and numerical simulation are applied to obtain the solution of the system. The numerical simulation demonstrates that the

complex normal form can be accurate enough to analysis the vibration with small magnitude The study provides insight into the mechanism of the nonlinear dynamics of nanowires, and given theoretical basement for design of the element structures in N/MEMS, sensors, actuators, and resonators, etc.

Acknowledgements

This project was supported by the National Natural Science Foundation of China (11272357), the Natural Science Foundation of Shandong Province of China (ZR2013AL017).

Appendix A

In Section 3, the parameters of Eq. (12) are shown as follows:

$a_{11}=0.0082-0.8588\beta_1$,

$a_{12}=-0.0312+11.7440\beta_1$,

$a_{13}=40.4229-2.1570\beta_1$,

$a_{14}=-306.8797+41.2253\beta_1$,

$a_{15}=2484.0513-147.2018\beta_1$ $a_{15}=2484.0513-147.2018\beta_1$,

$a_{16}=-2178.2983+192.7033\beta_1$,

$a_{22}=0.1131+13.2933\beta_1$,

$a_{21}=-0.02992-1.8738\beta_1$,

$a_{23}=-102.2965-3.4639\beta_1$,

$a_{24}=2484.1298+32.0625\beta_1$,

$a_{25}=-6530.1573-44.4370\beta_1$,

$a_{26}=13416.8369+53.2163\beta_1$.

Appendix B

In Section 4, the parameters of Eqs. (17) and (18) are shown as follows:

$$b_{11}=\frac{a_{23}-\Delta_2 a_{13}}{2\omega_1(\Delta_2-\Delta_1)},\ b_{12}=\frac{a_{24}-\Delta_2 a_{14}}{2\omega_1(\Delta_2-\Delta_1)},\ b_{13}=\frac{a_{25}-\Delta_2 a_{15}}{2\omega_1(\Delta_2-\Delta_1)},\ b_{14}=\frac{a_{26}-\Delta_2 a_{16}}{2\omega_1(\Delta_2-\Delta_1)},$$

$$b_{21}=\frac{\Delta_1 a_{13}-a_{23}}{2\omega_2(\Delta_2-\Delta_1)},\ b_{22}=\frac{\Delta_1 a_{14}-a_{24}}{2\omega_2(\Delta_2-\Delta_1)},\ b_{23}=\frac{\Delta_1 a_{15}-a_{25}}{2\omega_2(\Delta_2-\Delta_1)},\ b_{24}=\frac{\Delta_1 a_{16}-a_{26}}{2\omega_2(\Delta_2-\Delta_1)}.$$

Appendix C

$$\begin{aligned}h_2(\eta_1,\eta_2,\bar{\eta}_1,\bar{\eta}_2)&=\Lambda_{2,1}\eta_1^3+\Lambda_{2,2}\bar{\eta}_1^3+\Lambda_{2,3}\eta_2^3+\Lambda_{2,4}\bar{\eta}_2^3+\Lambda_{2,5}\eta_1^2\bar{\eta}_1\\&+\Lambda_{2,6}\eta_1^2\eta_2+\Lambda_{2,7}\eta_1^2\bar{\eta}_2+\Lambda_{2,8}\eta_1\bar{\eta}_1^2+\Lambda_{2,9}\eta_1\eta_2^2+\Lambda_{2,10}\eta_1\bar{\eta}_2^2+\Lambda_{2,11}\bar{\eta}_1^2\eta_2\\&+\Lambda_{2,12}\bar{\eta}_1^2\bar{\eta}_2+\Lambda_{2,13}\bar{\eta}_1\eta_2^2+\Lambda_{2,14}\bar{\eta}_1\bar{\eta}_2^2+\Lambda_{2,15}\eta_2^2\bar{\eta}_2+\Lambda_{2,16}\eta_2\bar{\eta}_2^2\\&+\Lambda_{2,17}\bar{\eta}_1\eta_2\bar{\eta}_2+\Lambda_{2,18}\eta_1\eta_2\bar{\eta}_2+\Lambda_{2,19}\eta_1\bar{\eta}_1\bar{\eta}_2+\Lambda_{2,20}\eta_1\bar{\eta}_1\eta_2.\end{aligned}$$

$$\begin{aligned}h_1(\eta_1,\eta_2,\bar{\eta}_1,\bar{\eta}_2)&=\Lambda_{1,1}\eta_1^3+\Lambda_{1,2}\bar{\eta}_1^3+\Lambda_{1,3}\eta_2^3+\Lambda_{1,4}\bar{\eta}_2^3+\Lambda_{1,5}\eta_1^2\bar{\eta}_1\\&+\Lambda_{1,6}\eta_1^2\eta_2+\Lambda_{1,7}\eta_1^2\bar{\eta}_2+\Lambda_{1,8}\eta_1\bar{\eta}_1^2+\Lambda_{1,9}\eta_1\eta_2^2+\Lambda_{1,10}\eta_1\bar{\eta}_2^2+\Lambda_{1,11}\bar{\eta}_1^2\eta_2\\&+\Lambda_{1,12}\bar{\eta}_1^2\bar{\eta}_2+\Lambda_{1,13}\bar{\eta}_1\eta_2^2+\Lambda_{1,14}\bar{\eta}_1\bar{\eta}_2^2+\Lambda_{1,15}\eta_2^2\bar{\eta}_2+\Lambda_{1,16}\eta_2\bar{\eta}_2^2\\&+\Lambda_{1,17}\bar{\eta}_1\eta_2\bar{\eta}_2+\Lambda_{1,18}\eta_1\eta_2\bar{\eta}_2+\Lambda_{1,19}\eta_1\bar{\eta}_1\bar{\eta}_2+\Lambda_{1,20}\eta_1\bar{\eta}_1\eta_2;\end{aligned}$$

$$\Lambda_{1,1}=-\frac{b_{11}+b_{12}\Delta_1+b_{13}\Delta_1^2+b_{14}\Delta_1^3}{2\omega_1},\ \Lambda_{1,2}=-\frac{\Lambda_{1,1}}{2},\ \Lambda_{1,3}=-\frac{b_{11}+b_{12}\Delta_2+b_{13}\Delta_2^2+b_{14}\Delta_2^3}{\omega_1-3\omega_2},$$

$$\Lambda_{1,4}=\frac{b_{11}+b_{12}\Delta_2+b_{13}\Delta_2^2+b_{14}\Delta_2^3}{\omega_1+3\omega_2},\ \Lambda_{1,5}=-\Lambda_{1,8},$$

$$\Lambda_{1,8}=\frac{3b_{11}+3\Delta_1b_{12}+3\Delta_1^2b_{13}+3\Delta_1^3b_{14}}{2\omega_1},\ \Lambda_{1,6}=-\frac{3b_{11}+b_{12}\left(2\Delta_1+\Delta_2\right)+b_{13}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{14}}{\omega_1+\omega_2},$$

$$\Lambda_{1,7}=\frac{3b_{11}+b_{12}\left(2\Delta_1+\Delta_2\right)+b_{13}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{14}}{\omega_2-\omega_1},$$

$$\Lambda_{1,9}=-\frac{3b_{11}+b_{12}\left(\Delta_1+2\Delta_2\right)+b_{13}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{14}}{2\omega_2},$$

$$\Lambda_{1,10}=-\Lambda_{1,9},$$

$$\Lambda_{1,11}=\frac{3b_{11}+b_{12}\left(2\Delta_1+\Delta_2\right)+b_{13}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{14}}{3\omega_1-\omega_2},$$

$$\Lambda_{1,12}=\frac{3b_{11}+b_{12}\left(2\Delta_1+\Delta_2\right)+b_{13}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{14}}{3\omega_1+\omega_2},$$

$$\Lambda_{1,13}=-\frac{3b_{11}+b_{12}\left(\Delta_1+2\Delta_2\right)+b_{13}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{14}}{2\omega_1-2\omega_2},$$

$$\Lambda_{1,14}=-\frac{3b_{11}+b_{12}\left(\Delta_1+2\Delta_2\right)+b_{13}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{14}}{2\omega_1+2\omega_2},$$

$$\Lambda_{1,15}=\frac{3b_{11}+3\Delta_2b_{12}+3\Delta_2^2b_{13}+3\Delta_2^3b_{14}}{\omega_1-\omega_2},$$

$$\Lambda_{1,16}=\frac{3b_{11}+3\Delta_2b_{12}+3\Delta_2^2b_{13}+3\Delta_2^3b_{14}}{\omega_1+\omega_2},\ \Lambda_{1,17}=\frac{3b_{11}+b_{12}\left(\Delta_1+2\Delta_2\right)+b_{13}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{14}}{\omega_1},$$

$$\Lambda_{1,18}=-\Lambda_{1,17},$$

$$\Lambda_{1,19}=\frac{6b_{11}+b_{12}\left(4\Delta_1+2\Delta_2\right)+b_{13}\left(2\Delta_1^2+4\Delta_1\Delta_2\right)+6\Delta_1^2\Delta_2b_{14}}{\omega_1+\omega_2},$$

$$\Lambda_{1,20}=\frac{6b_{11}+b_{12}\left(4\Delta_1+2\Delta_2\right)+b_{13}\left(2\Delta_1^2+4\Delta_1\Delta_2\right)+6\Delta_1^2\Delta_2b_{14}}{\omega_1-\omega_2},$$

$$\Lambda_{2,1}=\frac{b_{21}+b_{22}\Delta_1+b_{23}\Delta_1^2+b_{24}\Delta_1^3}{\omega_2-3\omega_1},\ \Lambda_{2,2}=\frac{b_{21}+b_{22}\Delta_1+b_{23}\Delta_1^2+b_{24}\Delta_1^3}{\omega_2+3\omega_1},$$

$$\Lambda_{2,3}=-\frac{b_{21}+b_{22}\Delta_2+b_{23}\Delta_2^2+b_{24}\Delta_2^3}{2\omega_2},\ \Lambda_{2,4}=\frac{b_{21}+b_{22}\Delta_2+b_{23}\Delta_2^2+b_{24}\Delta_2^3}{4\omega_2},$$

$$\Lambda_{2,5}=\frac{3b_{21}+3\Delta_1b_{22}+3\Delta_1^2b_{23}+3\Delta_1^3b_{24}}{\omega_2-\omega_1},\ \Lambda_{2,8}=\frac{3b_{21}+3\Delta_1b_{22}+3\Delta_1^2b_{23}+3\Delta_1^3b_{24}}{\omega_1+\omega_2},$$

$$\Lambda_{2,6}=-\frac{3b_{21}+b_{22}\left(2\Delta_1+\Delta_2\right)+b_{23}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{24}}{2\omega_1},$$

$$\Lambda_{2,7}=\frac{3b_{21}+b_{22}\left(2\Delta_1+\Delta_2\right)+b_{23}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{24}}{2\omega_2-\omega_1},$$

$$\Lambda_{2,9}=-\frac{3b_{21}+b_{22}\left(\Delta_1+2\Delta_2\right)+b_{23}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{24}}{\omega_1+\omega_2},$$

$$\Lambda_{2,10}=\frac{3b_{21}+b_{22}\left(\Delta_1+2\Delta_2\right)+b_{23}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{24}}{3\omega_2-\omega_1},$$

$$\Lambda_{2,11}=\frac{3b_{21}+b_{22}\left(2\Delta_1+\Delta_2\right)+b_{23}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{24}}{2\omega_1},$$

$$\Lambda_{2,12}=\frac{3b_{21}+b_{22}\left(2\Delta_1+\Delta_2\right)+b_{23}\left(\Delta_1^2+2\Delta_1\Delta_2\right)+3\Delta_1^2\Delta_2b_{24}}{2\omega_1+2\omega_2},$$

$$\Lambda_{2,13}=\frac{3b_{21}+b_{22}\left(\Delta_1+2\Delta_2\right)+b_{23}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{24}}{\omega_1-\omega_2},$$

$$\Lambda_{2,14}=\frac{3b_{21}+b_{22}\left(\Delta_1+2\Delta_2\right)+b_{23}\left(\Delta_2^2+2\Delta_1\Delta_2\right)+3\Delta_1\Delta_2^2b_{24}}{\omega_1+3\omega_2},$$

$$\Lambda_{2,15}=-\Lambda_{2,16}\quad;\Lambda_{2,16}=\frac{3b_{21}+3\Delta_2b_{22}+3\Delta_2^2b_{23}+3\Delta_2^3b_{24}}{2\omega_2},$$

$$\Lambda_{2,17}=\frac{6b_{21}+b_{22}\left(2\Delta_1+4\Delta_2\right)+b_{23}\left(2\Delta_2^2+4\Delta_1\Delta_2\right)+6\Delta_1\Delta_2^2b_{24}}{\omega_1+\omega_2},$$

$$\Lambda_{2,18}=\frac{6b_{21}+b_{22}\left(2\Delta_1+4\Delta_2\right)+b_{23}\left(2\Delta_2^2+4\Delta_1\Delta_2\right)+6\Delta_1\Delta_2^2b_{24}}{\omega_2-\omega_1},$$

$$\Lambda_{2,19}=\frac{6b_{21}+b_{22}\left(2\Delta_1+4\Delta_2\right)+b_{23}\left(2\Delta_2^2+4\Delta_1\Delta_2\right)+6\Delta_1\Delta_2^2b_{24}}{2\omega_2},$$

$\Lambda_{2,20}=-\Lambda_{2,19}$.

Appendix D

$c_{11}=6b_{11}+\left(4\Delta_2+2\Delta_1\right)b_{12}+\left(2\Delta_2^2+4\Delta_1\Delta_2\right)b_{13}+6\Delta_1\Delta_2^2b_{14}$,

$c_{12}=3b_{11}+3\Delta_1b_{12}+3\Delta_1^2b_{13}+3\Delta_1^3b_{14}$,

$c_{22}=3b_{21}+3\Delta_2b_{22}+3\Delta_2^2b_{23}+3\Delta_2^3b_{24}$,

$c_{13}=3b_{11}+\left(2\Delta_1+\Delta_2\right)b_{12}+\left(\Delta_1^2+2\Delta_1\Delta_2\right)b_{13}+3\Delta_1^2\Delta_2b_{14}$,

$c_{21}=6b_{21}+\left(4\Delta_1+2\Delta_2\right)b_{22}+\left(2\Delta_1^2+4\Delta_1\Delta_2\right)b_{23}+6\Delta_1^2\Delta_2b_{24}$.

Appendix E

$$q_1 = a_1\cos\left(\omega_1 t+\theta_1\right)+a_2\cos\left(\omega_2 t+\theta_2\right)+\left(\Lambda_{1,1}+\Lambda_{1,2}+\Lambda_{2,1}+\Lambda_{2,2}\right)a_1^3\cos\left(3\omega_1 t+3\theta_1\right)$$
$$+\left(\Lambda_{1,3}+\Lambda_{1,4}+\Lambda_{2,3}+\Lambda_{2,4}\right)a_2^3\cos\left(3\omega_2 t+3\theta_2\right)+\left(\Lambda_{1,5}+\Lambda_{1,8}+\Lambda_{2,5}+\Lambda_{2,8}\right)$$
$$a_1^3\cos\left(\omega_1 t+\theta_1\right)+\left(\Lambda_{1,6}+\Lambda_{1,12}+\Lambda_{2,6}+\Lambda_{2,12}\right)a_1^2a_2\cos\left(2\omega_1 t+\omega_2 t+2\theta_1+\theta_2\right)$$
$$+\left(\Lambda_{1,6}+\Lambda_{1,12}+\Lambda_{2,6}+\Lambda_{2,12}\right)a_1^2a_2\cos\left(2\omega_1 t+\omega_2 t+2\theta_1+\theta_2\right)$$
$$+\left(\Lambda_{1,7}+\Lambda_{1,11}+\Lambda_{2,7}+\Lambda_{2,11}\right)a_1^2a_2\cos\left(2\omega_1 t-\omega_2 t+2\theta_1-\theta_2\right) \quad ,$$
$$+\left(\Lambda_{1,9}+\Lambda_{1,14}+\Lambda_{2,9}+\Lambda_{2,14}\right)a_2^2a_1\cos\left(\omega_1 t+2\omega_2 t+\theta_1+2\theta_2\right)$$
$$+\left(\Lambda_{1,10}+\Lambda_{1,13}+\Lambda_{2,10}+\Lambda_{2,13}\right)a_2^2a_1\cos\left(\omega_1 t-2\omega_2 t+\theta_1-2\theta_2\right)$$
$$+\left(\Lambda_{1,15}+\Lambda_{1,16}+\Lambda_{2,15}+\Lambda_{2,16}\right)a_2^3\cos\left(\omega_2 t+\theta_2\right)+\left(\Lambda_{1,17}+\Lambda_{1,18}+\Lambda_{2,17}+\Lambda_{2,18}\right)$$
$$a_1a_2^2\cos\left(\omega_1 t+\theta_1\right)+\left(\Lambda_{1,19}+\Lambda_{1,20}+\Lambda_{2,19}+\Lambda_{2,20}\right)a_2a_1^2\cos\left(\omega_2 t+\theta_2\right).$$

$$q_2 = \Delta_1 a_1\cos\left(\omega_1 t+\theta_1\right)+\Delta_2 a_2\cos\left(\omega_2 t+\theta_2\right)+\left(\Delta_1\left(\Lambda_{1,1}+\Lambda_{1,2}\right)+\Delta_2\left(\Lambda_{2,1}+\Lambda_{2,2}\right)\right)$$
$$a_1^3\cos\left(3\omega_1 t+3\theta_1\right)+\left(\Delta_1\left(\Lambda_{1,3}+\Lambda_{1,4}\right)+\Delta_2\left(\Lambda_{2,3}+\Lambda_{2,4}\right)\right)a_2^3\cos\left(3\omega_2 t+3\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,5}+\Lambda_{1,8}\right)+\Delta_2\left(\Lambda_{2,5}+\Lambda_{2,8}\right)\right)a_1^3\cos\left(\omega_1 t+\theta_1\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,6}+\Lambda_{1,12}\right)+\Delta_2\left(\Lambda_{2,6}+\Lambda_{2,12}\right)\right)a_1^2a_2\cos\left(2\omega_1 t+\omega_2 t+2\theta_1+\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,6}+\Lambda_{1,12}\right)+\Delta_2\left(\Lambda_{2,6}+\Lambda_{2,12}\right)\right)a_1^2a_2\cos\left(2\omega_1 t+\omega_2 t+2\theta_1+\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,7}+\Lambda_{1,11}\right)+\Delta_2\left(\Lambda_{2,7}+\Lambda_{2,11}\right)\right)a_1^2a_2\cos\left(2\omega_1 t-\omega_2 t+2\theta_1-\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,9}+\Lambda_{1,14}\right)+\Delta_2\left(\Lambda_{2,9}+\Lambda_{2,14}\right)\right)a_2^2a_1\cos\left(\omega_1 t+2\omega_2 t+\theta_1+2\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,10}+\Lambda_{1,13}\right)+\Delta_2\left(\Lambda_{2,10}+\Lambda_{2,13}\right)\right)a_2^2a_1\cos\left(\omega_1 t-2\omega_2 t+\theta_1-2\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,15}+\Lambda_{1,16}\right)+\Delta_2\left(\Lambda_{2,15}+\Lambda_{2,16}\right)\right)a_2^3\cos\left(\omega_2 t+\theta_2\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,17}+\Lambda_{1,18}\right)+\Delta_2\left(\Lambda_{2,17}+\Lambda_{2,18}\right)\right)a_1a_2^2\cos\left(\omega_1 t+\theta_1\right)$$
$$+\left(\Delta_1\left(\Lambda_{1,19}+\Lambda_{1,20}\right)+\Delta_2\left(\Lambda_{2,19}+\Lambda_{2,20}\right)\right)a_2a_1^2\cos\left(\omega_2 t+\theta_2\right). \quad .$$

References

[1] Z. L.Wang, Nanomaterials for Nanoscience and Nanotechnology. Charaterization of Nanophase Materials, Wiley-VCH Verlag GmbH, 2000.
[2] G.Wang, X. Feng, Effect of surface stresses on the vibration and buckling of piezoelectric nanowires, EPL–Europhys Lett. 91 (2010) 56007.
[3] H. Moeenfard, A. Darvishian, M. T. Ahmaidan, Static behavior of nano/micromirrors under the effect of Casimir force, an analytical approach, J. Mech. Sci.Techno. 26 (2012) 537–543.
[4] S. Cuenot, C. F retigny, S. Demoustier-Champagne, B. Nysten, Surface tension effect on the mechanical properties of nanomaterials measured by atomic force microscopy, Phys. Rev. B. 69 (2004) 165410.
[5] D. M. Zhao, J. L. liu, J. Sun, R. N. Wu, R. Xia, A Revisit of Internal Force Diagrams on Nanobeams with Surface Effects. Curr. Nanosci. 11 (2015) 388–393.
[6] L. Y. Jiang, Z. Yan, Timoshenko beam model for static bending of nanowires with surface effect, Physica E 42 (2010) 2274–2279.
[7] H. Li, Z. Yang, Y. M. Zhang, B. C. Wen, Deflections of Nanowires with consideration of Surface Effects, Chin. Phys. Lett. 27 (2010) 126201.
[8] J. He, C. M. Lilley, Surface effect on the elastic behavior of static bending nanowires. Nano Lett. 8 (2008) 1798–1802.
[9] G.Wang, X. Feng, Effects of surface elasticity and residual surface tension on the natural frequency of microbeams, Appl. Phys. Lett. 90 (2007) 231904.
[10] J. He, C. M. Lilley, Surface stress effect on the bending resonance of nanowire with different boundary conditions, Appl. Phys. Lett. 93 (2008) 263108.
[11] R. Ansari, K. Hosseini, A. Darvizeh, B. Daneshian, A sixth-order compact finite difference method for non-classical vibration analysis of nanobeams including surface stress effects, Appl. Math. Comput. 219 (2013) 4977–4991.
[12] G.Wang, X. Feng, Timoshenko beam model for buckling and vibration of nanowires with surface effects, J. Phys. D: Appl. Phys. 42 (2009) 155411.
[13] A. T. Samaei, B. Gheshlaghi, G. Wang, Frequency analysis of piezoelectric nanowires with surface effects, Curr. Appl. Phys. 13 (2013) 2098–2102.
[14] J. Han, Q. Zhang , W. Wang, Design considerations on large amplitude vibration of a doubly clamped microresonator with two symmetrically located electrodes, Commun. Nonlinear Sci. 22 (2015), 492–510.
[15] A. Yao, T. Hikihara, Counter operation in nonlinear micro-electro-mechanical resonators, Phys. Lett. A. 377 (2013) 2551–2555.
[16] C. Chen, H. Hu, L. Dai, Nonlinear behavior and characterization of a piezoelectric laminated microbeam system. Commun. Nonlinear Sci. **18** (2013) 1304–1315.
[17] E. M. Miandoab, A. Yousefi-Koma, H. N. Pishkenari, F. Tajaddodianfar, Study of nonlinear dynamics and chaos in MEMS/NEMS resonators. Commun. Nonlinear Sci. 22 (2015) 611–622.
[18] M. I. Younis, A. H. Nayfeh, A study of the nonlinear response of a resonant microbeam to an electric actuation, Nonlinear Dyn. 31 (2003) 91–117.
[19] E. M. Abdel-Rahman, A. H. Nayfeh, Secondary resonances of electrically actuated resonant microsensors, J. Micromech. Microeng. 13 (2003) 491–501.
[20] W. Zhang, G. Meng, Nonlinear dynamical system of micro-cantilever under combined parametric and forcing excitations in MEMS, Sensor. Actuat. A–Phys. 119 (2005) 291–299.
[21] S. Gutschmidt, O. Gottlieb, Nonlinear dynamic behavior of a microbeam array subject to parametric actuation at low, medium and large DC-voltages, Nonlinear Dyn. 67 (2012) 1-36.
[22] B. Gheshlaghi, S. M. Hasheminejad, Surface effects on the nonlinear free vibration of nanobeams, Compos. Part B–Eng. 42 (2011) 934–937.
[23] H. Moeenfard, M. Mojahedi, M. T. Ahmadian, A homotopy perturbation analysis of nonlinear free vibration of Timoshenko microbeams, J. Mech. Sci. Technol. 25 (2011) 557–565.
[24] G. Ouyang, C. X. Wang , G. W. Yang, Surface energy of nanostructural materials with negative curvature and related size effects, Chem. Rev. 109 (2009) 4221–4247.
[25] M. E. Gurtin, J. Weissmüller, F. Larche, A general theory of curved deformable interfaces in solids at equilibrium, Phil. Mag. A. 78 (1998) 1093–1109.
[26] C. Semler, G. X. Li , M. P. Paidoussis, The non-linear equations of motion of pipes conveying fluid, J. Sound Vib. 169 (1994) 577-599.
[27] A. H. Nayfeh,The Method of Normal Forms, Wiley, Singapore, 2011.

Damage and failure prediction in Alumina Tri-Hydrate/Epoxy core composite sandwich panels subjected to impact loads

G. Morada[1], R. Ouadday[1], †A. Marouene[1], *A. Vadean[1], and R. Boukhili[1]

[1]Department of Mechanical engineering, Polytechnique Montreal, Montreal, Quebec, Canada.

*Presenting author: aurelian.vadean@polymtl.ca

†Corresponding author: aymen.marouene@polymtl.ca

Abstract

This paper reports an experimental and numerical analysis of the impact behavior of composite sandwich panels. An innovative sandwich construction with an ATH/Epoxy core (i.e. epoxy resin filled with alumina tri-hydrate (ATH) particles) and non-crimp glass fabric fibre-reinforced epoxy face-sheets was subjected to impact loads. Explicit nonlinear finite elements model was developed to predict the damage characteristics in both the face-sheets and core. The obtained numerical results were compared with the test data to assess the effectiveness of the proposed model. A good correlation with respect to the contact force and energy-time relationships, permanent deformation, and impact-induced damage was achieved. The contribution of each component of the sandwich structure to its energy absorption capabilities was also evaluated. It was found, for an impact energy of 21J, that the energy dissipated in the ATH/Epoxy core is almost two times more than that dissipated in the face-sheets. The important role of the core material for reducing face-sheet damage was identified.

Keywords: Impact behaviour, Composite sandwich panel, Alumina trihydrate (ATH) particles, damage mechanisms.

Introduction

Composite sandwich structures are finding increasing utilization in many engineering applications such as the aerospace, automotive, building, and water turbine industries, because of their relative benefits over other structural materials [1]. For instance, conventional structures in hydraulic turbine are nowadays replaced with composite sandwich structures to improve energy production and to facilitate in-site manufacturing. However, in such application, it has been found that the river flow can provoke huge amount of waterborne debris and the waterborne debris impact was highlighted as a major source of damage for the composite hydraulic turbine blades. Therefore, impact resistance is an important topic in engineering communities.

Impact resistance of composite sandwich depend on the mechanical and geometrical properties of its constituents such as the face-sheet material, core material, and the adhesive interface properties. Core crushing was identified as the major failure mechanism under an impact event [2]. Meanwhile, one major drawback of sandwich structures is its poor transverse stiffness [3]. Therefore, the core material properties are the main parameters to improve impact resistance of composite sandwich panels. A wide variety of material can be used as core in sandwich

constructions such as synthetic foam, honeycomb, balsa wood, and corrugated cores among others [1]. The main functions of the core materials are to absorb impact energies and provide the overall bending resistance. However, the problem with light-weight cores is that they are not enough resistant to withstand high impact loads.

Mines et al. [2] reported that the core density affects the failure progression. Furthermore, it has been shown that absorbing impact energy via the plastic deformations of the core can improve the damage tolerance of sandwich structures [4]. Torre and Kenny [5] used an innovative sandwich construction made of glass/phenolic composite skins and a rigid polymer foam core with fibre reinforced plastic to enhance crush resistance for civil engineering structures. The sandwich addressed herein is a high density core made of epoxy resin filled with Alumina tri-hydrate particles. This sandwich construction was designed to increase the core crushing resistance and hence improve damage tolerance of sandwich panels at high impact loads.

In light of the aforementioned considerations and the existence of some limitations for performing experimental tests, there is a strong need to develop a numerical model that can be used to predict the structural impact response and the damage process and locations under impact conditions.

There are several numerical approaches reported in the open literature for prediction of the response of sandwich structures under impact loads. In order to reduce the computational time, some researchers [6-8] have used 2D shell elements to model the face-sheets. Among them, Zhou et al. [6] studied the perforation resistance of foam-based sandwich panels using 2D elements for the face-sheets, however, it should be noted that these elements are not accurate for failure analysis since the stress distribution in the face-sheets is a 3D problem. Feng et al. [9] used a progressive damage model to simulate the damage scenarios in foam-based sandwich composites subjected to impact loads. In their proposed model, a 3D damage model was used to track the intra-laminar damages in face-sheets and cohesive elements were used to simulate interface delaminations.

The objective of this work is to investigate the impact response of a particular composite sandwich panel designed to the water turbine industries. This sandwich is made of a high-density core (ATH/Epoxy: epoxy resin filled with alumina trihydrate particles) and Non-Crimp Fabrics glass/epoxy skins. To the best of the authors' knowledge, there is no published studies deal with this sandwich construction. A numerical 3D continuum model was implemented in LS-DYNA/Explicit code to simulate the intra-laminar damage initiation and development within the face-sheets. This model included an enhanced non-linear shear model and a mixed-matrix damage initiation and propagation law. The cohesive elements approach is also used to simulate the inter-laminar delamination. Furthermore, a specific continuum damage model is developed to simulate the behaviour of the ATH/Epoxy core. This model accounts for the damage initiation and propagation as well as the residual strength after final failure. The numerical results were compared with the test data and a good correlation was obtained. The numerical model was also used to assess the contribution of each component of the sandwich structure to its energy absorption capacity.

Compression test on ATH/Epoxy core

Flatwise compressive characteristic of ATH/Epoxy core with 50 wt% ATH was studied. Note that the ATH amount was selected on the basis of a preliminary experimental study (not reported herein), which was conducted earlier to identify the optimum ATH amount that can be used to minimize the heat generated during the epoxy curing reaction. The square cross-section specimens of 51 × 51 mm dimensions with thickness of 25.4 mm were prepared according to the ASTM D1621-10 standard procedure [10]. Testing was carried out on the MTS testing machine with displacement rate of 2.5 mm/min. The uniform distributed load was applied on specimens by two flat and parallel plates (Fig. 1).

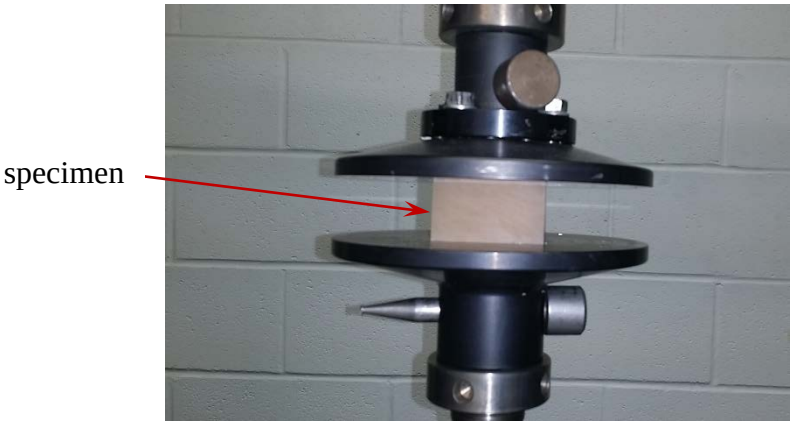


Fig. 1 Flatwise compression test setup

Fig. 2 depicts the load-displacement curves from compression testing experiments which served us to calculate the compressive Young's modulus and crush strength values.

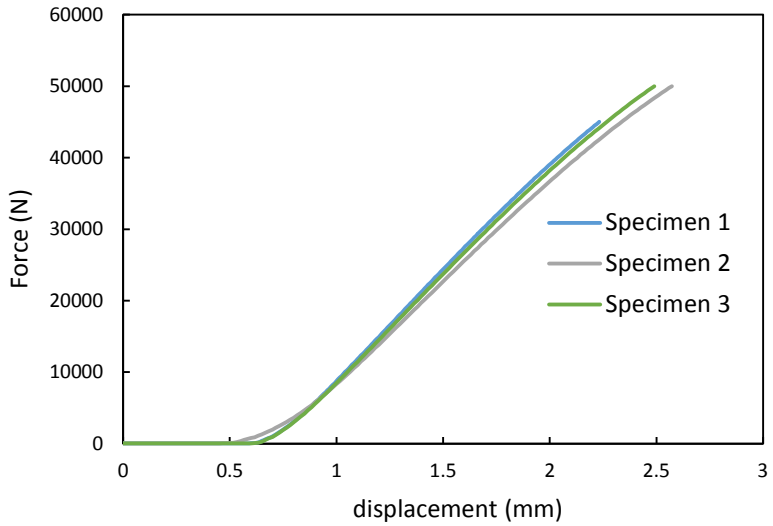


Fig. 2. Compressive force-displacement response of ATH/Epoxy

Face-sheets damage model

Material constitutive model and nonlinear shear response

For a better definition of the material constitutive model of composite laminates, both the non-linear behaviour due to the plastic deformation and the damage in the laminate must be considered [11]. These two phenomena can be simulated using plasticity and continuum damage theories, respectively. Thus, in the present work, elastic Hooke's law for linear orthotropic materials is adopted to contemplate the non-linear shear behavior.

The material constitutive model can be expressed as follows:

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\end{bmatrix} = \frac{1}{\Omega}\begin{bmatrix} E_{11}(1-v_{23}v_{32}) & E_{22}(v_{12}-v_{32}v_{13}) & E_{33}(v_{13}-v_{12}v_{23}) \\ E_{11}(v_{21}-v_{31}v_{23}) & E_{22}(1-v_{13}v_{31}) & E_{33}(v_{23}-v_{21}v_{13}) \\ E_{11}(v_{31}-v_{21}v_{31}) & E_{22}(v_{32}-v_{12}v_{31}) & E_{33}(1-v_{12}v_{21}) \end{bmatrix}\begin{bmatrix}\varepsilon_{11}\\ \varepsilon_{22}\\ \varepsilon_{33}\end{bmatrix} \quad (1)$$

$$\Omega = 1 - v_{12}v_{21} - v_{23}v_{32} - v_{31}v_{13} - 2v_{21}v_{32}v_{13}$$

The nonlinear shear stress-strain part of the constitutive model is assigned as follows:

$$\tau_{ij} = G_{ij}\left(\gamma_{ij} - \gamma_{ij}^{in}\right)\left(1 - \alpha\gamma_{ij}\right) \ where \ \ ij = 1,2,3 \quad (2)$$

where γ_{ij} is total shear strain that can be decomposed into elastic γ_{ij}^{e} and inelastic components γ_{ij}^{in}:

$$\gamma_{ij} = \gamma_{ij}^{e} + \gamma_{ij}^{in} \quad (3)$$

Before damage initiation, inelastic component of the strain can be obtained by:

$$\gamma_{ij}^{in} = \gamma_{ij} - \frac{\tau_{ij}}{G_{ij}^{0}} - \frac{\tau_{ij}}{G_{ij}^{0}(1-\alpha\gamma_{ij})} \quad (4)$$

where G_{ij}^{0} is initial shear modulus, α is a material constant expressing the gradual shear modulus which can be found experimentally. To depict the nonlinear shear behaviour, a polynomial cubic stress-strain as follow was used:

$$\tau_{ij}(\gamma_{ij}) = c_1\gamma_{ij} + (\gamma_{ij})c_2\gamma_{ij}^{2} + c_3\gamma_{ij}^{3} \quad (5)$$

where c_1, c_2, and c_3 are the coefficients obtained by curve fitting to experimental shear stress-strain response.

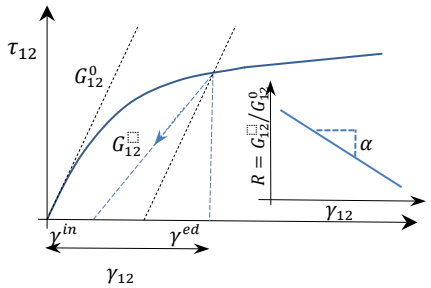


Fig. 3 Typical shear stress-strain response

Damage initiation and propagation in material constitutive was taken into account through the continuum damage mechanic model (CDM). Therefore, a physically-based CDM model was developed in the FE software. The continuous damage evaluation in each ply of laminate was described by a damage matrix *D*, which defined by three internal damage variables d_{ij} correspond to the different damage modes. Each of the damage variables reduces a component of the undamaged stress tensor σ to simulate the stiffness degradation.

$$\sigma^d = D\sigma \tag{6}$$

Intra-laminar damage model

Fibre failure modes

Two strain-based failure criteria, F_{11}^T and F_{11}^C, were used to detect fibre damage initiation under tensile and compressive loading, respectively:

$$\begin{aligned} F_{11}^T &= \left(\frac{\varepsilon_{11}}{\varepsilon_{11}^{ot}}\right)^2 - 1 \geq 0 \\ F_{11}^C &= \left(\frac{\varepsilon_{11}}{\varepsilon_{11}^{oc}}\right)^2 - 1 \geq 0 \end{aligned} \tag{7}$$

where ε_{11}^{ot} and ε_{11}^{oc} are the damage initiation strain in tension and compression, respectively.

Once the damage initiates, material starts to gradually lose its stiffness up to the final failure as sketched in Fig. 4. Here, the damage variables for tensile (d_{11}^t) and compressive (d_{11}^c) fibre failures are defined as follows:

$$\begin{aligned} d_{11}^t &= \frac{\varepsilon_{11}^{ft}}{\varepsilon_{11}^{ft} - \varepsilon_{11}^{ot}}\left(1 - \frac{\varepsilon_{11}^{ot}}{\varepsilon_{11}}\right)^2 \\ d_{11}^c &= \frac{\varepsilon_{11}^{fc}}{\varepsilon_{11}^{fc} - \varepsilon_{11}^{oc}}\left(1 - \frac{\varepsilon_{11}^{oc}}{\varepsilon_{11}}\right)^2 \end{aligned} \tag{8}$$

where ε_{11}^{ft} and ε_{11}^{fc} are the maximum strain at failure which are calculated as a function of the critical energy release rates (G_{11}^t and G_{11}^c), maximum longitudinal stresses (X^t, X^c) and the characteristic length, l^* as follows:

$$\varepsilon_{11}^{ft} = \frac{2G_{11}^t}{X^t\, l^*}\ ;\ \ \varepsilon_{11}^{fc} = \frac{2G_{11}^c}{X^c\, l^*} \tag{9}$$

One coupled tension-compression damage variable, d_{1f}, was used to simulate fibre degradation in the longitudinal direction:

$$d_{1f} = d_{11}^c + d_{11}^t - d_{11}^t d_{11}^c \tag{10}$$

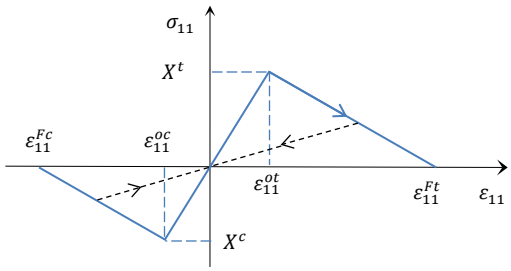


Fig. 4. Intra-laminar damage model behaviour for fiber failure

Matrix failure modes

Matrix damage initiation: Failure criterion proposed by Catalanotti et al. [12] was used to detect matrix cracking, F_{22}^T, and Puck failure criterion [13] was used to identify matrix crushing, F_{22}^C.

These criteria were defined as:

$$F_{22}^T = \left(\frac{\sigma_{\mathrm{nn}}}{S_{\mathrm{t}}^{is}}\right)^2 + \left(\frac{\tau_{\mathrm{nl}}}{S_{\mathrm{l}}^{is}}\right)^2 + \left(\frac{\tau_{\mathrm{nt}}}{S_{\mathrm{t}}^{is}}\right)^2 + \lambda\left(\frac{\sigma_{\mathrm{nn}}}{S_{\mathrm{t}}^{is}}\right)^2\left(\frac{\tau_{\mathrm{nl}}}{S_{\mathrm{l}}^{is}}\right)^2 + \kappa\left(\frac{\sigma_{\mathrm{nn}}}{S_{\mathrm{t}}^{is}}\right)^2 - 1 \geq 0$$

$$F_{22}^C = \left(\frac{\tau_{\mathrm{nl}}}{S_{\mathrm{l}}^{is} - \mu_{\mathrm{nt}}\sigma_{\mathrm{nn}}}\right)^2 + \left(\frac{\tau_{\mathrm{nt}}}{S_{\mathrm{t}}^{is} - \mu_{\mathrm{nl}}\sigma_{\mathrm{nn}}}\right)^2 - 1 \geq 0 \quad (11)$$

where Y_t , S_{t}^{is}, and S_{l}^{is} are the matrix tensile strength and the *in situ* shear strength in transverse and longitudinal directions, respectively; κ and λ are defined as $\kappa = (S_l^2 - Y_t)/S_t\,Y_t$ and $\lambda = 2\mu_{nl}S_t/S_l - \kappa$; μ_{nt} and μ_{nl} are friction coefficients defined as $\mu_{nt} = -1/\tan(2\theta_f)$ and $\mu_{nl} = \mu_{nt}\,S_{12}/S_t$ where $S_t = Y_c/2\tan(\theta_f)$ and Y_c is the matrix compressive strength. The angle of fracture plane, θ_f, is approximately 53° for unidirectional laminate under pure compressive loading.

The two previous criteria depend on the stresses in the potential fracture plane (Fig. 5) which can be calculated using the standard transformation matrix $T(\theta)$:

$$\sigma_{nlt} = [T(\theta)]\sigma_{123}[T(\theta)]^T \quad (12)$$

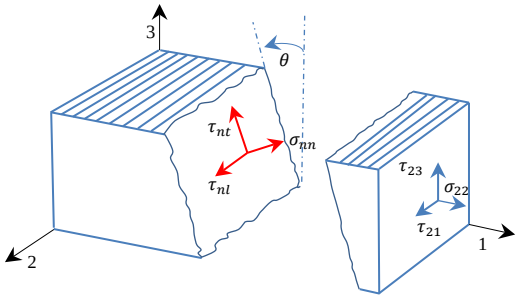


Fig. 5. Fracture plane in compression loading

Matrix damage propagation: when the matrix failure initiates under combined loading, the resulted stress, σ_r, and the corresponding strain, ε_r, on the potential fracture plan should be recorded as follows:

$$\begin{aligned}\sigma_r &= \sqrt{\langle\sigma_{nn}\rangle^2 + (\tau_{nt})^2 + (\tau_{nl})^2}\\ \varepsilon_r &= \sqrt{\langle\varepsilon_{nn}\rangle^2 + (\gamma_{nt})^2 + (\gamma_{nl})^2}\\ \varepsilon^0_{\mathrm{r,in}} &= \sqrt{(\gamma^{in}_{nt})^2 + (\gamma^{in}_{nl})^2}\end{aligned} \tag{13}$$

Here, $\varepsilon^0_{\mathrm{r,in}}$, is the inelastic component of the strain at the moment of failure initiation.

The matrix damage parameter, d_m, is defined as:

$$d_m = \frac{\varepsilon^f_{\mathrm{r}} - \varepsilon^0_{\mathrm{r,in}}}{\varepsilon^f_{\mathrm{r}} - \varepsilon^o_{\mathrm{r}}}\left(\frac{\varepsilon^0_{\mathrm{r}} - \varepsilon_r}{\varepsilon_r - \varepsilon^0_{\mathrm{r,in}}}\right) \tag{14}$$

The shear and tensile stresses on the fracture plane are reduced by the following relations and then they are transformed to the original plane.

$$\begin{aligned}\sigma_{nl} &= (1 - d_m)\sigma_{nl}\\ \sigma_{nt} &= (1 - d_m)\sigma_{nt}\\ \sigma_{nn} &= \sigma_{nn} - d_m\sigma_{nn}\end{aligned} \tag{15}$$

The fracture energy of the matrix, G_m, under combined stresses can be calculated as follows:

$$G_m = G_{IC}\left(\frac{\sigma_{nn}}{\sigma_r}\right)^2 + G_{IIc}\left(\frac{\tau_{nt}}{\sigma_r}\right)^2 + G_{IIC}\left(\frac{\tau_{nl}}{\sigma_r}\right)^2 \tag{16}$$

where G_{IC} and G_{IIC} are the critical strain energy release rates for modes I and II, respectively.

The final failure strain, ε^f_r, which is governed by the critical strain energy release rate, G_m, and characteristic length, l, is defined as follows:

$$\varepsilon_r^f = \frac{2G_m}{\sigma_r\, l} \tag{17}$$

Inter-laminar damage model

Cohesive elements —defined by a linear traction-separation model— are frequently used for simulating the delamination between two successive plies with different fiber orientations. This cohesive model is composed of an elastic behaviour until the damage initiation according to a stress-based quadratic interaction criterion, followed by decohesion of the two plies as a result of the damage propagation.

The quadratic stress-based criterion adopted herein to detect delamination initiation was defined as follows:

$$\left(\frac{\sigma_1}{T}\right)^2 + \left(\frac{\tau_2}{S}\right)^2 + \left(\frac{\tau_3}{S}\right)^2 = 1 \tag{18}$$

where σ_1, τ_2, τ_3 are the interface tangential and normal stresses and T, S are the maximum traction stresses in normal and tangential directions.

The delamination propagation was modeled using the Benzeggagh-Kenane rule [14] for mixed-mode loading:

$$\delta^F = \frac{1+\beta^2}{A_{TSLC}(T+\beta^2 S)}\left[G_{IC} + (G_{IIC} - G_{IC})\left(\frac{\beta^2 S}{T+\beta^2 S}\right)^{XMU}\right] \tag{19}$$

where β is the mixed mode ratio, XMU is exponent of the mixed mode criterion, A_{TSLC} is the area under the load-displacement curve, and G_{IC}, G_{IIC} are the inter-laminar fracture toughness in mode I, II.

ATH/Epoxy core damage model

In order to model the core damage behavior, some numerical approaches have been proposed in the open literature. Some authors [15, 16] applied a yield criterion that considers the transvers normal and shear stresses to predict the initiation of plasticity. Atkay et al. [8] proposed a removing failed element technique to simulate the damage propagation in honeycomb and foam cores. Nevertheless, this approach can not represent the residual strength of material after compressive failure. In this work, a damage model based on the continuum damage mechanic was proposed to simulate the damage initiation and propagation in ATH/Epoxy core. This model takes into account the residual strength after compression failure as sketched in Fig. 6.

The Besant's failure criterion [15] was adopted to detect the core failure initiation under combined shear and compression loads

$$\left(\frac{\sigma_{zz}}{\sigma_{cu}}\right)^2 + \left(\frac{\tau_{xz}}{\tau_{lu}}\right)^2 + \left(\frac{\tau_{yz}}{\tau_{tu}}\right)^2 \geq 1 \tag{20}$$

where σ_{cu}, τ_{lu}, and τ_{tu} are the corresponding yields stresses.

After damage initiation, the stresses (σ_{zz}, τ_{xz}, and τ_{yz}) are gradually reduced using a damage variable, d_c, defined as follows:

$$d_c = \frac{\varepsilon_c^f}{\varepsilon_c^f - \varepsilon_c^o}\left(1 - \frac{\varepsilon_c^o}{\varepsilon}\right)^2 \tag{21}$$

where ε_c^o is the strain at the failure initiation and ε_c^f is the strain at the final failure.

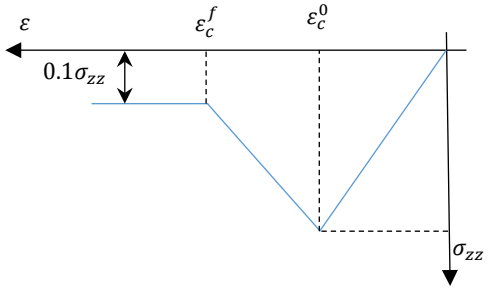


Fig. 6. Stress-strain response of the ATH/Epoxy core

Experimental details

In this investigation, non-crimp fabric (NCF) glass reinforced composite laminates are used as skins for sandwich panels. The composite skins were composed of six layers of E-glass/epoxy reinforcement. Each NCF lamina consists of three plies of [90°/0°/90°] tied together using polyester yarn. At first, the composite skins were manufactured using the vacuum infusion (VI) process. Meantime, sandwich core was prepared by mixing the resin epoxy with 50 wt% of ATH particles. The polymerization mixture was poured into a wood mould where the skins are earlier positioned at its both ends as sketched in Fig. 7. The nominal thickness of sandwich core is 34 mm. After the casting process was completed, the curing of the plastic core (ATH/Epoxy) was achieved at room temperature for 24h. Following the curing process, the sandwich panels were cut into specimens with 100 mm × 100 mm in dimension.

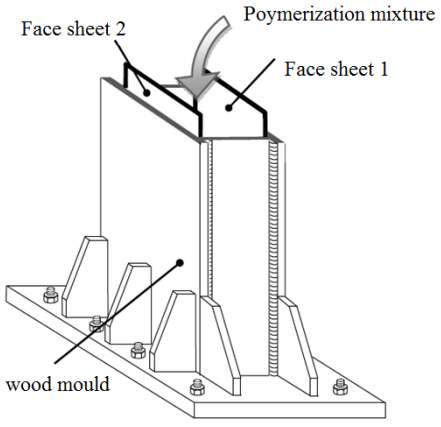


Fig. 7. Wood mould for fabrication of sandwich panel

Impact tests were performed using a drop weight machine following the guideline given in the ASTM standard D3763 [17]. The impactor had a mass of 22 kg and a diameter of 25.4 mm. During impact test, the specimen was constrained between two parallel rigid supports with a hole of 75 mm diameter in the center (see Fig. 8). A sufficient clamping pressure was applied to prevent slippage of the specimen during experiments.

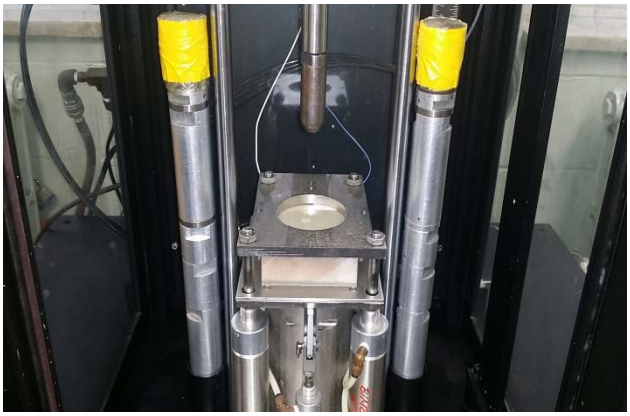


Fig. 8 Specimen fixture apparatus

Finite element model

A 3D finite element model was implemented in LS-DYNA/Explicit code to predict the structural behavior of the whole sandwich panels as well as the damage characteristics for the core and face-sheets during impact loading. To decrease the computational time, only one quarter of the sandwich panel with symmetric boundary conditions was modelled as illustrated in Fig. 9.

Both the plastic core and face-sheets were modelled using eight-node solid elements with reduced integration and hourglass control. Zero-thickness cohesive elements were used to simulate delamination between adjacent plies with different fiber orientations. The impactor and support plate are defined to be rigid bodies. A surface-to-surface type contact element was defined between the upper face-sheet and the impactor surface.

Since no damage was observed in the bottom face-sheet following the experimental testing, the face-sheets damage model was only defined for the upper face-sheet. The ATH/Epoxy core behaviour was simulated through the core material model described in the previous section.

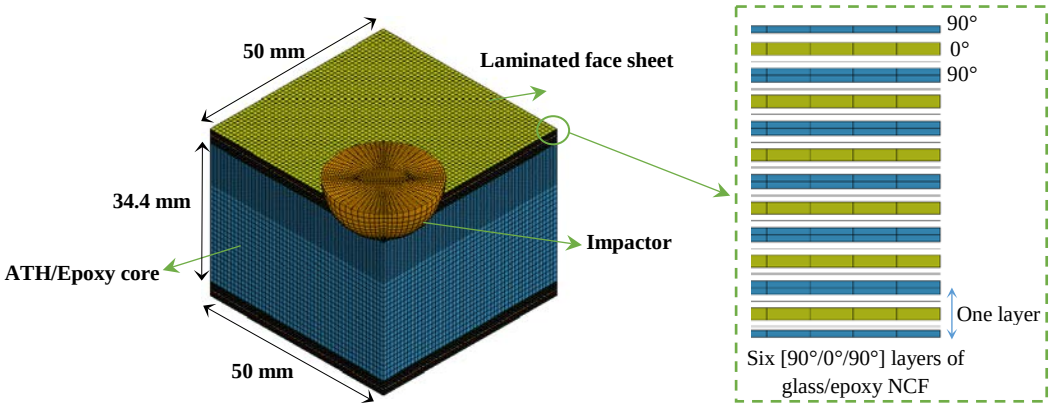


Fig. 9. Finite element model for impact simulations

Results and discussions

Impact response of ATH/Epoxy core

In order to validate the damage model proposed to simulate the AHT/Epoxy damage behavior, impact tests on the ATH/Epoxy specimens were performed for an impact energy of 21J. The choice of this energy level was made to avoid damaging of the used cell load since no data are available in the open literature regarding the impact resistance of the studied sandwich construction.

Figs. 10a and b present a comparison between numerical and experimental force-time curves and energy-time curves, respectively. In general, close correlation is achieved between the numerical prediction and the experimental data. The maximum recorded contact force is about 22.5 kN which can be considered as a high impact load.

Moreover, with regards to impact energy, the experimental results show that about 9.5J energy was absorbed through plastic deformations and matrix damage in the ATH/Epoxy core. Numerical model tends to underestimate the value of absorbed energy as is evident in Fig. 10b. The difference between numerical predictions and experimental data seems *a priori* due to an underestimation of the plastic deformation that the ATH/Epoxy material suffered during the test.

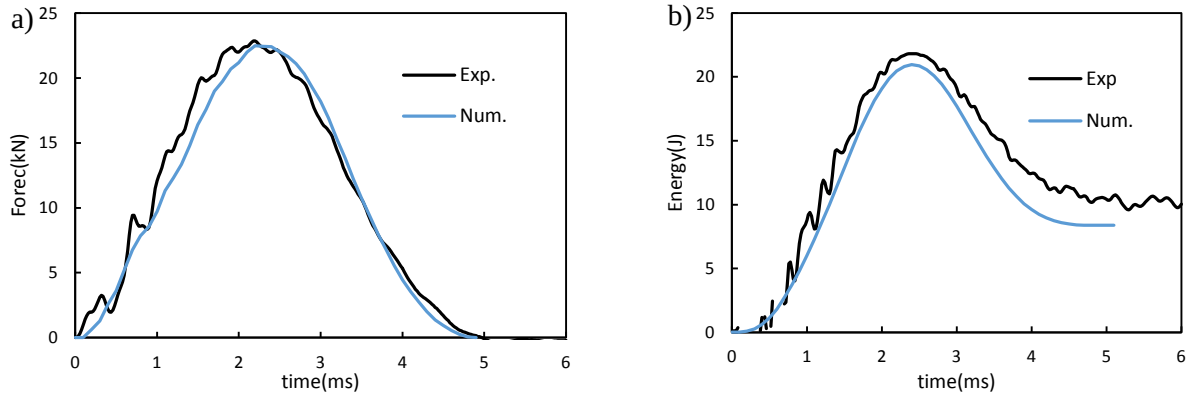
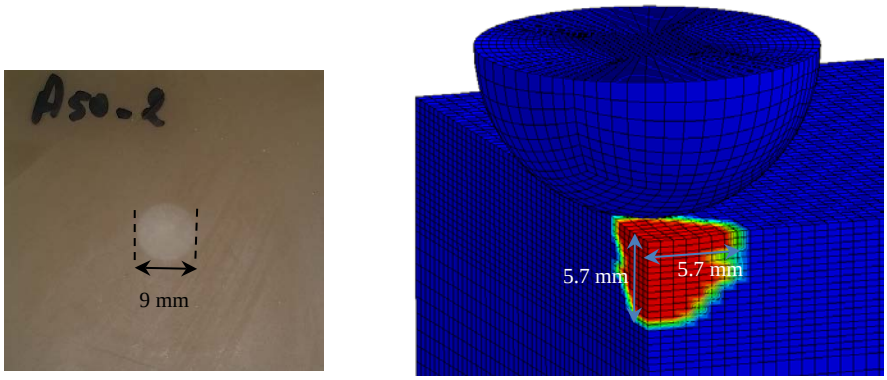


Fig. 9. Impact response of ATH/Epoxy core for an impact energy of 21J

A comparison between the experimental and predicted damage area at impact energy of 21J is presented in Fig. 11. The damage area reported herein represents the projected damage area towards the impacted surface. At first sight, it can be noticed that the numerical model is able to capture the shape (circular shape) and size of the damage area. This pointed out the appropriateness of the proposed core material model to simulate the damage pattern in the ATH/Epoxy plastic core.

From the numerical results, it can be noticed that the predicted damage depth is equal to almost one-half of the predicted damage diameter. Thus, it can be assumed that the experimental damage depth is about 4.5 mm. Microscopic observations will be needed to confirm this hypothesis.

On the other hand, the numerical results show that the compressive stresses in the ATH/Epoxy core are highly intense in the localized contact area. One can therefore draws the conclusion that the damage in ATH/Epoxy core resulted from high compressive stresses under the impactor. Moreover, numerical results show the presence of an irreversible deformation of the ATH/Epoxy core close to the impact zone. This residual deformation is manifested as a permanent indentation of 0.3 mm depth.



a) experimental damage zone

b) numerical damage zone

Fig. 10. Damage zone in ATH/Epoxy specimen

Impact response of NCF laminated face sheet

In order to investigate the influence of sandwich core on the damage evolution in NCF glass/epoxy laminates face-sheets, the impact response of the face-sheets laminates was simulated herein under the same boundary conditions.

Fig. 12a and b illustrate the contact force and energy as a function of time for an impact energy of 21J. The predicted maximum contact force and absorbed energy are about 8.5 kN and 7.5J, respectively. The NCF composite laminates absorb energy through matrix damage and interface delamination mechanisms.

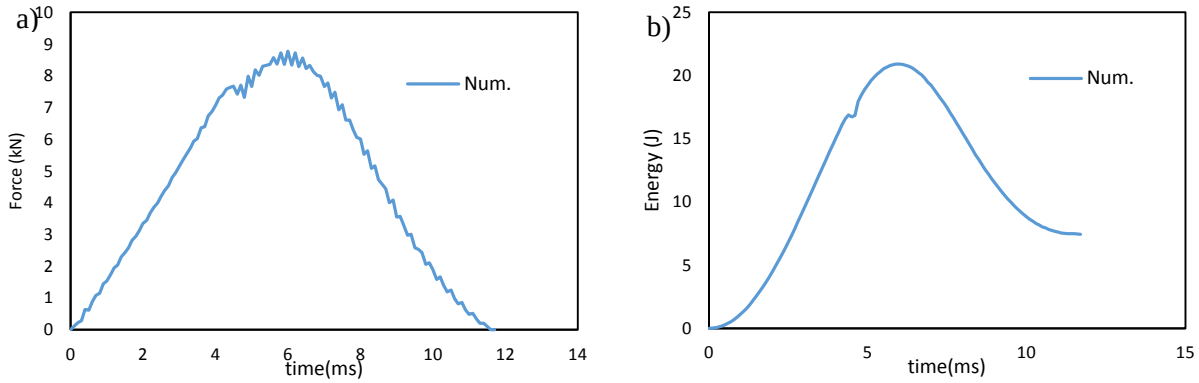


Fig. 11. Impact response of NCF laminated for an impact energy of 21J

Fig. 13 shows the predicted impact damage pattern in the NCF composite laminates. As can be seen in Fig. 13, the damage area is roughly circular with a diameter of 30 mm, which is relatively large damage area. The numerical results reveal that the matrix damage and delaminations are the main failure mechanisms in the NCF laminates for 21J impact energy. The high tensile stresses due to the large bending deformation are the main reason behind the matrix damage propagation.

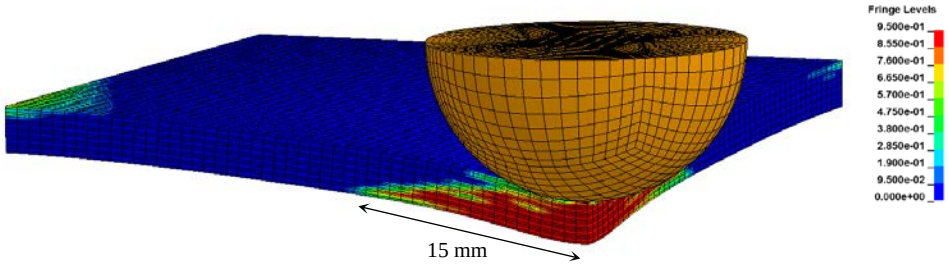


Fig. 12. Damage zone in NCF laminated

Impact response of the sandwich

Figs. 14a and b present the contact force-time and impact energy-time of the sandwich panel. As can be seen from Fig. 14a, there is a reasonable correlation between numerical predictions and experimental data. The maximum force is well predicted and its value is almost close to that achieved for the ATH/Epoxy specimen. The contact time, which is related to the material's resistance, is slightly shorter than that of ATH/Epoxy specimen. It seems that sandwich panel is a little stiffer than the ATH/Epoxy. From the experimental and numerical results, it was clear that the core material played an important role in the impact response of sandwich panel.

There is a smaller difference between the predicted and measured absorbed energy as shown in Fig. 14b. This difference is probably due to the plastic deformation in the core material. Beyond this, these results demonstrate the capacity of core material and sandwich panel to absorb energy. The sandwich panel absorbs more than half of impact energy. The absorbed energy is dissipated through face-sheets damage and core damage. The damage in the face-sheets was considerably reduced due to present of the ATH/Epoxy core (compare to NCF laminates only). Indeed, the nature of stress distribution is different from that of NCF laminates. The flexural deformation in

the face-sheets decreased due to core stiffness, and hence, the amount of the bending cracks significantly decreased. In contrast, the shear cracks, which result from the high transverse shear stresses, are more pronounced in this case.

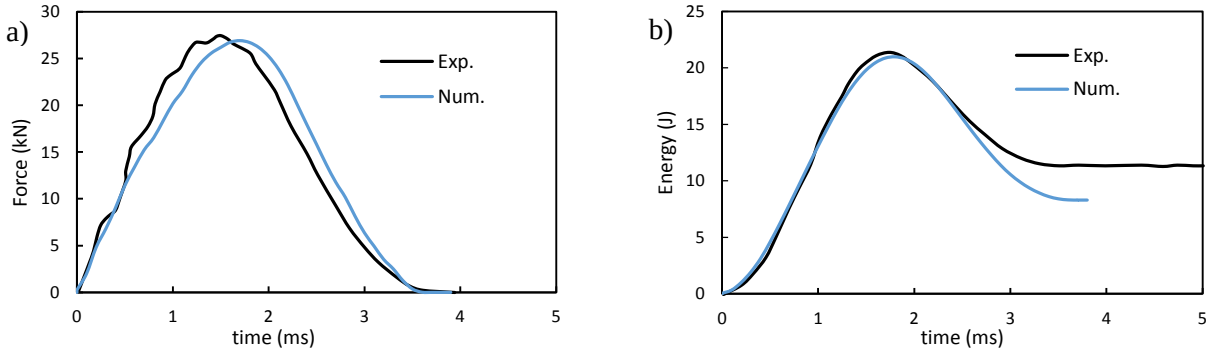
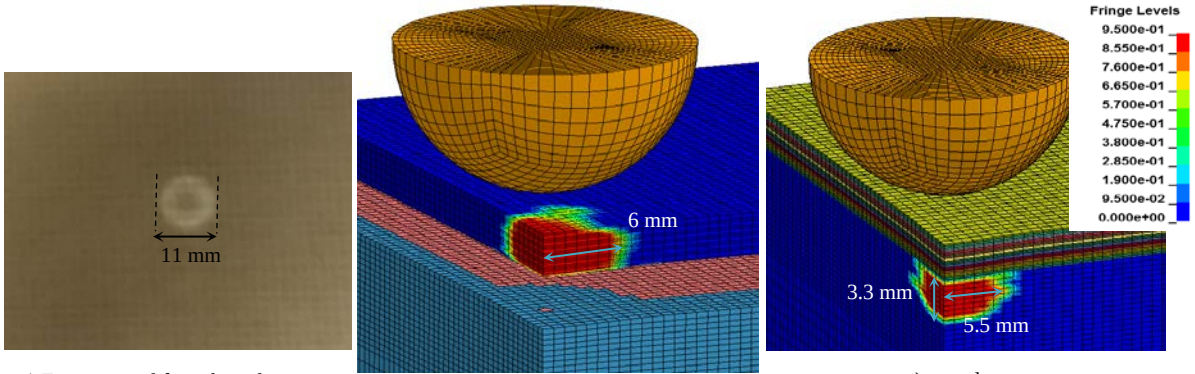


Fig. 13. Impact response of sandwich panel for an impact energy of 21J

The impact-damage areas in both the upper face-sheet and core are shown in Fig. 15. The damage pattern in the upper face-sheet is well predicted in terms of shape and size. Because of some experimental limitations, it was difficult to assess the impact-damage inside the core. However, since the damage model of the core was previously compared and validated with experimental data, the predicted damage in the core must be reasonably considered as reliable.

As expected, the size of damage area in the core is smaller than that of ATH/Epoxy specimen (without face-sheets) as shown in Fig. 15.

Moreover, the numerical results show a debonding failure at face-sheet/core interface close to the impact zone (Fig. 15a) where the shear stresses are the highest. It can therefore deduce that the sliding mode is the main cause of the interface debonding between the upper face-sheet and the core. The debonding zone is meanwhile relatively small. This could be due to the high elastic modulus of the ATH/Epoxy material. Indeed, the core's elastic modulus has a considerable effect on the interface debonding resistance [18].



a) Experimental face sheet damage

b) Face sheet damage

c) core damage

Fig. 14. Damage zone in sandwich panel

In order to highlight the role of the plastic core on the energy dissipation process under impact loads, the energy dissipation in each component of the sandwich panel is tracked. Fig. 16 displays the energy dissipated in the core and the face-sheets along with the total energy dissipated in the sandwich panel for 21J impact energy. According to these energy curves, it was found that the energy dissipated in the ATH/Epoxy core is almost two times more than that dissipated in the face-sheets. Furthermore, more than 25% of the initial kinetic energy is absorbed in core crush (which was about 65% of the overall absorbed energy). However, less than 12% of the initial kinetic energy is absorbed in the upper face-sheet damage. These numerical findings are consistent with the previous results that reveal that the ATH/Epoxy has a good ability to locally deform and hence can absorbed a considerable amount of the energy dissipated in the whole structure.

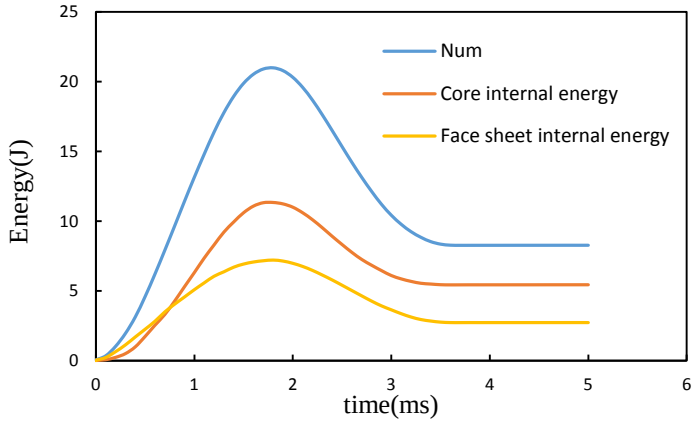


Fig. 15. Damage dissipation mechanism for an impact energy of 21J

Conclusions

A 3D progressive damage model was implemented into FEM software LS-DYNA/Explicit to predict the face-sheets and core damage in ATH/Epoxy core sandwich panels subjected to low-velocity impact loads. A continuum damage model was used to describe the behaviour and failure of the NCF glass/epoxy composite face-sheets, accounts for matrix damage, delamination, and fiber failure. Besides this, a damage model was developed to simulate the ATH/Epoxy behaviour which includes damage initiation and propagation and residual compressive strength. Experimental tests were conducted to validate the numerical model. In general, a reasonable correlation between the experimental data and the numerical simulations was achieved. The damage model used to simulate damage propagation in the face-sheets, has reflected accurately the experimental damage in the face-sheet. The numerical model of ATH/Epoxy predicted the damage and the absorbed energy in ATH/Epoxy specimens precisely.

Form experimental and numerical results, it can be drawn that ATH/Epoxy core sandwich panels are effective structure at withstanding low-velocity impact with relatively high impact energy. The ability of the ATH/Epoxy material to locally deform and absorb a large amount of impact energy makes them a suitable choice for the sandwich core when impact damage resistance is the main design issue.

The work presented in this paper is the first step in the development of a novel generation of hydraulic turbines components made from composite sandwich structures capable to better withstands impact loads.

References

[1] Chai, G. B., and Zhu, S., 2011, "A review of low-velocity impact on sandwich structures," Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials Design and Applications, 225(4), pp. 207-230.
[2] Mines, R., Worrall, C., and Gibson, A., 1998, "Low velocity perforation behaviour of polymer composite sandwich panels," International Journal of Impact Engineering, 21(10), pp. 855-879.
[3] Crupi, V., Kara, E., Epasto, G., Guglielmino, E., and Aykul, H., 2015, "Prediction model for the impact response of glass fibre reinforced aluminium foam sandwiches," International Journal of Impact Engineering, 77, pp. 97-107.
[4] Evans, A. G., He, M., Deshpande, V. S., Hutchinson, J. W., Jacobsen, A. J., and Carter, W. B., 2010, "Concepts for enhanced energy absorption using hollow micro-lattices," International Journal of Impact Engineering, 37(9), pp. 947-959.
[5] Torre, L., and Kenny, J., 2000, "Impact testing and simulation of composite sandwich structures for civil transportation," Composite structures, 50(3), pp. 257-267.
[6] Zhou, J., Hassan, M. Z., Guan, Z., and Cantwell, W. J., 2012, "The low velocity impact response of foam-based sandwich panels," Composites science and Technology, 72(14), pp. 1781-1790.
[7] Foo, C., Chai, G., and Seah, L., 2008, "A model to predict low-velocity impact response and damage in sandwich composites," Composites science and Technology, 68(6), pp. 1348-1356.
[8] Aktay, L., Johnson, A. F., and Holzapfel, M., 2005, "Prediction of impact damage on sandwich composite panels," Computational Materials Science, 32(3), pp. 252-260.
[9] Feng, D., and Aymerich, F., 2013, "Damage prediction in composite sandwich panels subjected to low-velocity impact," Composites Part A: Applied Science and Manufacturing, 52, pp. 12-22.
[10] ASTM, 2010, "Standard Test Method for Compressive Properties of Rigid Plastics," ASTM International.
[11] Donadon, M., Iannucci, L., Falzon, B. G., Hodgkinson, J., and de Almeida, S. F., 2008, "A progressive failure model for composite laminates subjected to low velocity impact damage," Computers & Structures, 86(11), pp. 1232-1252.
[12] Catalanotti, G., Camanho, P., and Marques, A., 2013, "Three-dimensional failure criteria for fiber-reinforced laminates," Composite structures, 95, pp. 63-79.
[13] Puck, A., and Schürmann, H., 1998, "Failure analysis of FRP laminates by means of physically based phenomenological models," Composites science and Technology, 58(7), pp. 1045-1067.
[14] Benzeggagh, M., and Kenane, M., 1996, "Measurement of mixed-mode delamination fracture toughness of unidirectional glass/epoxy composites with mixed-mode bending apparatus," Composites science and Technology, 56(4), pp. 439-449.

[15] Besant, T., Davies, G., and Hitchings, D., 2001, "Finite element modelling of low velocity impact of composite sandwich panels," Composites Part A: Applied Science and Manufacturing, 32(9), pp. 1189-1196.
[16] Kärger, L., Baaran, J., and Teßmer, J., 2007, "Rapid simulation of impacts on composite sandwich panels inducing barely visible damage," Composite structures, 79(4), pp. 527-534.
[17] ASTM, 2006, "ASTM D3763–2006. Standard test method for high speed puncture properties of plastics using load and displacement sensor," USA Standards Association International, USA.
[18] Goswami, S., and Becker, W., 2001, "The effect of facesheet/core delamination in sandwich structures under transverse loading," Composite structures, 54(4), pp. 515-521.

Seismic behavior of a caisson type breakwater on non-homogeneous soil deposits composed of liquefiable layer under earthquake loading

***X.H. Bao[1,2], †Dong Su[1], Y.B. Fu[1], and F. Zhang[3]**

[1]Department of Civil Engineering, Shenzhen University, China.
[2] Key Laboratory of Geotechnical and Underground Engineering, Ministry of Education, Tongji University, China
[3] Department of Civil Engineering, Nagoya Institute of Technology, Japan

*Presenting author: bxh@szu.edu.cn
†Corresponding author: sudong@szu.edu.cn

Abstract

Damage of breakwaters during earthquakes is mainly attributed to the liquefaction of foundation soil. Most of the studies have investigated the dynamic response of breakwaters considering uniform sand foundation and a single earthquake event. However, the foundation of a breakwater usually consists of many sub-layers of soil from liquefiable sand to relatively impermeable clay. Moreover, during earthquakes a main shock may trigger numerous aftershocks within a short time which may have the potential to cause additional damage to soil and structures. In this study, the performance of an existing caisson type breakwater on the natural ground composed of discontinuous liquefiable sand layer and impermeable clay layer is investigated using an effective based soil-water coupling finite element method. In the calculation, a real recorded seismic wave in the 2011 Great East Japan earthquake which composed of a main shock and two aftershocks is adopted as the input earthquake wave. The results reveal that time histories of excess pore water pressure is the governing factors to estimate the behavior of breakwater during and after an earthquake, and the repeated earthquake shakings have a significant effect on the accumulated displacement of breakwater and ground. Eventually the settlement is the most important aspect for the tsunami resistance capacity of breakwater structures.

Keywords: Caisson type breakwater, Repeated Earthquake shakings, Excess pore water pressure, Settlement, FEM.

Introduction

Earthquake induced liquefaction has become a major problem to offshore structures such as breakwaters, river dykes, levees, earth dams etc., supported on a cohesionless foundation soil. Previous studies have shown that the wide spread damage to offshore structures occurred mainly due to the liquefaction of foundation soil, resulting in settlement, tilting, slumping and lateral spreading (Seed 1968, Adalier et al. 1998, Huang & Yu 2013) [1]-[3]. Despite the extensive research and development of remedial measures to prevent the large deformation of soil structures, offshore structures have suffered severe damage during 2011 Great East Japan Earthquake (Oka et al. 2012, Mori et al. 2013, Mori et al. 2015) [4]-[6]. The minor to major damage was attributed due to the liquefaction of foundation soil. This event elucidates the further need to understand the deformation behavior of offshore structures resting on non-homogeneous liquefiable foundations. However, attention given to the seismic response of offshore structures under strong seismic loading is limited. Among these offshore structures, breakwaters may damage or lose their normal ability to resist tsunami loading during strong earthquake loading before the arriving of tsunami. To date, most of the investigations on breakwaters concentrated on the tsunami wave and the mechanical behavior of rubble mound (Fujima 2006, Imase et al. 2012, Susumu 2012, Takahashi et al. 2014) [7]-[10]. Experimental and numerical investigations on seismic behaviors of a composite breakwater under

earthquake loading are still limited, which can be found in the works (Memos et al. 2003, Yuksel et al. 2004, Jafarian et al 2010, Ye 2012) [11]-[14].

On the other hand, most of the experimental studies and numerical analyses have been conducted previously to examine the behavior of offshore structures resting on uniform cohesionless soil during earthquakes (Aydingun & Adalier 2003, Adalier & Sharp 2004, Ye & Wang 2015) [15]-[17]. However, it is noted that natural soil deposits normally consist of many sub-layers with different soil particles and properties, ranging from sand to cohesive clay and coarse sand layers, referred to as non-homogeneous soil deposits. Huang et al. (2015) [18] point out that liquefaction in the saturated layer was the contributing factor to large settlement and sliding of the structures. Thus, the dynamic behavior of the breakwater on a liquefiable non-homogeneous foundation, consisting of discontinuous low permeability layers of silt or clay at different depths should be well understood.

During the earthquake that repeated ground-motion sequences occurring after short intervals of time, resulting from a main shock and aftershocks earthquakes (Zhang et al. 2013) [19], it was found that the low amplitude aftershock can accumulate large lateral deformation and continue for several minutes on the liquefied soil (Maharjan & Takahashi 2014) [20]. However, in most of the previous experimental and numerical studies seismic performance of soil structures is investigated by applying only a single earthquake, ignoring the influence of repeated shake phenomena. Among the limited studies considering the repeated earthquake shakings, Ye et al. (2007) [21] conducted shaking table tests and numerical analyses on saturated sandy soil to investigate the mechanical behavior of liquefiable foundations considering repeated shaking and consolidation processes. Xia et al. (2010) [22] presented numerical analysis of an earth embankment on liquefiable foundation soils under repeated shake and consolidation condition. During 2011 Great East Japan Earthquake, some structures continued to shake after the onset of soil liquefaction for more than two minutes. Moreover, during the reconnaissance survey after the earthquake, Sasaki et al. (2012) [23] found that the more severe deformation and subsidence of levees was due to the occurrence of aftershock, 30 min after the main shock. However, no previous study has examined the effects of repeated earthquake shakings on breakwaters lying on non-homogeneous soil deposits. Therefore, to understand the deformation mechanism of breakwaters resting on non-homogeneous soil deposits under main shock and sequential aftershocks is of great importance.

In this study, the co-seismic and post-seismic behavior of an existing caisson type breakwater resting on the natural ground composed of discontinuous clay and sand layer under the recorded seismic wave in the 2011 Great East Japan Earthquake which composed of a main shock and two aftershocks is investigated using an effective based soil-water coupling numerical model DBLEAVES (Ye 2011) [24]. In the analysis, an advanced elasto-plastic soil constitutive model named as Cyclic Mobility model (Zhang et al. 2007, Zhang et al. 2011) [25] [26] is used to describe the complicated nonlinear dynamic behavior of the foundation soils. The results show that the used numerical method is capable of capturing the progressive ground liquefaction and long-term consolidation process of the breakwater and foundation system during and after earthquake loading. The influence of earthquake can significantly reduce the capacity of breakwater to resist tsunami loading. In engineering practice, the settlement maybe a serious problem for the breakwater when its foundation ground composes of discontinuous impermeability clay and liquefiable sand soils.

Constitutive model

Using a proper constitutive model to accurately describe soil behaviors including the development of excess pore water pressure during earthquakes becomes a key factor when

assessing the dynamic behavior of ground and foundation. In the studies using numerical methods, most of the previous investigations on seismic dynamics of offshore structures used simple constitutive models such as elastic or Mohr–Coulomb model to model the seabed soil (Ye & Wang 2015) [17]. These simple models are not capable of simulating the complicated nonlinear cyclic behaviors of soils and the failure process of offshore structures. Intensive nonlinear interaction between foundation and the structure cannot be effectively captured. Iai et al. (1998) [27] conducted effective stress analyses of port structures in Kobe port during the Hyogoken–Nambu earthquake in 1995. The numerical analyses calculated that the composite breakwater constructed on loose seabed soil settled about 2m during the event, which is consistent with the field observation. The work highlighted the importance of using effective stress analyses with well-calibrated cyclic soil model to realistically capture the nonlinear structure–foundation interaction. Therefore, it is very important to estimate the co-seismic and post-seismic behavior of breakwaters using an effective numerical method with proper constitutive model, for tsunami associated with the earthquake would cause serious damage to the structures especially when the foundation composed of liquefiable layer and might experience large deformation by earthquakes.

For this reason, by adopting the concepts of subloading (Hashiguchi & Ueno 1977) [28] and superloading (Asaoka et al. 2002) [29], Zhang et al. (2007) [25] proposed a rotational kinematic hardening elasto-plastic model named as Cyclic Mobility model (CM model) which can describe the mechanical behavior of soils under different drainage and loading conditions. Zhang et al. (20110 [26] and Ye B. et al. (2012) [30] extended the CM model to describe the mechanical behavior of soils under general three-dimensional stress conditions to consider the intermediate principal stress (Ye G.L. et al. 2012, Ye G.L. et al. 2013) [31] [32]. According to the work of shaking-table tests and numerical simulation under a repeated liquefaction-consolidation process by Ye B. et al. (2007) [33], it was confirmed that the static and dynamic behavior of sand could be well described by the CM model, considering the effect of the stress-induced anisotropy, the density and the structure of the soil formed in the natural sedimentary process, different loading conditions and drained conditions in a unified way.

In this study, the clay and sand are modeled with the above mentioned CM model. Eight parameters are employed in the model, among which five parameters, M, N, λ, κ and ν, are the same as those in the Cam-clay model. The other three parameters, a: the parameter controlling the collapse rate of the structure, m: the parameter controlling the loosing rate of the overconsolidation ratio or the change in density of the soil, and b_r: the parameter controlling the developing rate of the stress-induced anisotropy, have clear physical meanings and can be easily determined by undrained triaxial cyclic loading tests and drained triaxial compression tests. The values of eight parameters involved in the model are fixed in all loading process once they are determined from the laboratory tests. A detailed description of the CM model can be found in the references (Zhang et al. 2007, Zhang et al. 2011, Zhang et al. 2010) [25] [26] [34].

FEM model and parameters

Analysis range and soil profiles

The analysis range is shown in Fig.1, in which, the breakwater consisting of a caisson and rubble mound beneath, is constructed on a natural ground mainly composed of clay soil noted as Ac and sand soil noted as As. The original clay soil beneath rubble mound was replaced by sand noted as Rs. The caisson is made of concrete, and can be practically treated as an impermeable; while the rubble mound, which made of stones, is permeable. The total length of the analysis range is 240 m, and the distances from the centerline of breakwater to lateral

sides of the ground foundation are both 120 m, which is considered to be large enough. The whole depth of the ground is 31 m, which composed of clay noted as Ac, sand noted as As and bottom sand noted as Ds. The depth of each soil layer and the size of breakwater are listed in Fig.1. Obviously, the liquefiable sand lay lied beneath thick clay layer which may prohibit the dissipation of pore water pressure. To improve the ground bearing capacity for structures, the original clay soil was replaced by sand beneath breakwater during project construction.

Some typical points on the breakwater and in the ground are chosen to illustrate the co-seismic and post-seismic behaviors of breakwater and foundation system. As shown in Fig. 1, the points on the breakwater are P-1 at the top of breakwater and P-2 at the bottom of caisson on rubble mound; the points beneath breakwater at the centerline are C-1 (GL-5 m) and C-2 (GL-15 m); the points in the near-filed of the ground (20 m away from the caisson) are N-0 (GL-0 m), N-1 (GL-5 m), N-2 (GL-15 m); the points in the far-filed of the ground (100 m away from the caisson) are F-0 (GL-0 m), F-1 (GL-5 m), F-2 (GL-15 m). Here, the locations with depth of 5 m and 15 m below ground surface in free field and beneath the breakwater are representative for the seismic behavior in upper clay layer and middle sand layer.

Ground parameters

As is known that the identification of parameters from laboratory and in situ tests is convincible, since no cyclic tests data of soils are available, some of these parameters were determined by element simulation with reference to the standard penetration tests. The average N-value and permeability for soils are listed in Table 1, while the eight ground parameters of each soil layer used in calculation are listed in Table 2. The initial values of the state variables employed in the constitutive model are listed in Table 3. On the other hand, the caisson which made of concrete is modeled as impermeable elastic solid element. The rubble mound, which made of stones, is modeled as permeable elastic solid element. The Physical properties of breakwater are listed in Table 4.

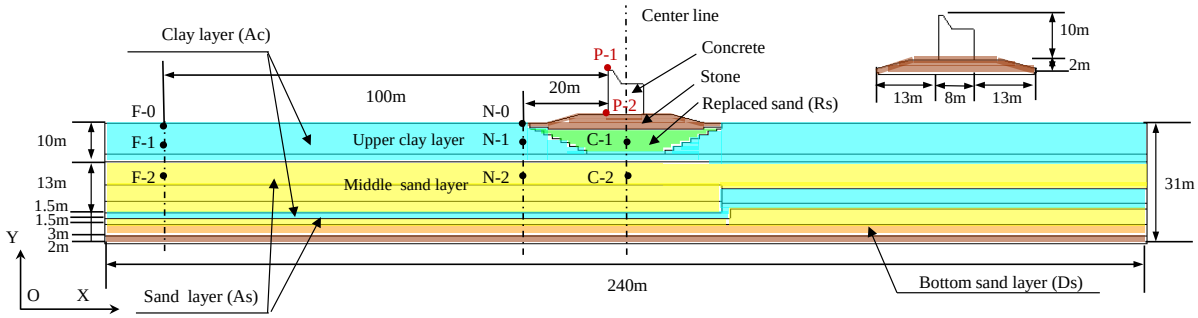


Figure 1. Soil profiles and section view of the caisson type breakwater

Analysis program and boundary condition

The numerical analysis was conducted using an effective stress based 2D/3D soil-water coupling program named as DBLEAVES (Ye 2011) [24], whose applicability and accuracy was firmly verified by the investigation on group-pile foundations in real scale (Jin et al. 2010) [35] and model tests (Bao et al. 2012, Bao et al. 2014) [36] [37]. Not only the instant reaction of ground and structure system when subjected to a strong earthquake but also the consequential long-term settlement of an alternately layered ground can be well examined using a sophisticated constitutive model and effective stress based soil-water coupling finite element method (Bao et al. 2016) [38].

For the boundary conditions, the base nodes of the ground foundation were assumed to be fixed in both x and y direction. The side boundary nodes at the same elevation were all “tied” together to experience the same accelerations. The earthquake loading is applied as a time-

varying input acceleration to the foundation base. A constant water level is assumed and the drained boundary is set at the surface of the ground. As a large ocean wave is unlikely to occur simultaneously with earthquake, the wave loading is not considered in this study.

Table 1. The average N-value and permeability of ground soils

Layer	Clay Ac	Sand As	Replaced Rs	Ds
N-value	3	13	20	above 50
Permeability $\boldsymbol{k}$ (m/sec)	1×10^{-9}	1×10^{-4}	1×10^{-4}	4×10^{-5}

Table 2. Material parameters of ground soils

Layer	Ac	As	Rs	Ds
Compression index $\boldsymbol{\lambda}$	0.13	0.05	0. 05	0.046
Swelling index $\boldsymbol{\kappa}$	0.026	0.062	0.065	0.0061
Stress ratio of critical state **M**	1.21	1.41	1.42	1.42
Void ratio $\boldsymbol{N}$ (p'=98 kPa on *N.C.L.*)	1.08	0.93	0.92	0.88
Poisson's ratio $\boldsymbol{\nu}$	0.38	0.35	0.35	0.35
Degradation parameter of overconsolidation state $\boldsymbol{m}$	2.20	0.10	0.10	0.10
Degradation parameter of structure $\boldsymbol{a}$	0.10	2.20	2.20	2.20
Evolution parameter of anisotropy $\boldsymbol{b_r}$	0.10	1.50	1.50	1.50

Table 3. Initial values of the state variables of ground soils

Layer	Ac	As	Rs	Ds
Void ratio $\boldsymbol{e_0}$	0.97	0.98	0.91	0.81
Degree of structure $\boldsymbol{R_0^*}$	0.80	0.60	0.60	0.70
Overconsolidation $\boldsymbol{OCR}$ ($\boldsymbol{1/R_0}$)	2.00	3.00	4.00	20.0
Anisotropy $\boldsymbol{\zeta_0}$	0.0	0.0	0.0	0.0

Table 4. Physical properties of breakwater

Item	Elastic modulus (kPa)	Poisson's ratio $\boldsymbol{\nu}$	Density $\boldsymbol{\rho}$(t/m^3)	Permeability $\boldsymbol{k}$ (m/sec)
Caisson	1.0×10^{8}	0.25	2.5	1.0×10^{-11}
Rubble mound	1.0×10^{6}	0.30	2.0	1.0×10^{-2}

Earthquake loading and simulation stages

Input Earthquake wave

In the calculation, the seismic wave induced by the 2011 Great East Japan Earthquake (ML =9.0) is used as the earthquake loading to applied to the breakwater and foundation system. One of the main features of this earthquake is that the aftershock activity was extremely vigorous. The input earthquake motion recorded 2,300 m below ground surface at Urayasu in E-W direction is considered as being representative in Chiba Prefecture (source: www.k-net.bosai.go.jp) as shown in Fig. 2. This observation station is near to the coastal line of pacific ocean, therefore, the chosen input earthquake wave in the analysis is similar as close as possible with the real seismic wave propagating to the breakwater foundation.

It is noted that the earthquake composed of a major shock and two aftershocks lasts for 42.25 munities. The first shock (major shock) lasted for 5 min with a maximum acceleration of 85 gal and the second shock (first aftershock) also lasted for 5 min with a maximum acceleration

of 25 gal while the third shock (second aftershock) lasted for 2.25 min with a maximum acceleration of 3 gal as shown in Fig.3. The interval between the first shock and the second shock was approximate 24 min, and the interval between the second shock and the third shock was approximate 6 min. It should be mentioned herein that such a long duration of motions has been the major cause of the severe liquefaction and ground deformation.

Newmark-method is used and the integration time interval is 0.01s. Rayleigh type of initial-rigidity-proportional attenuation is used and the damping values of the soils, the structure and the piles are assumed to be 2% and 10% for the first and second modes respectively in the dynamic analysis of the breakwater and foundation system.

Calculation steps

The analysis was performed in three steps:
Step 1: The static analysis considering the ground foundation-breakwater as a whole system is carried out to get the initial effective stress of the ground before the dynamic analysis. The distribution of initial mean effective stress caused by the gravity of ground and breakwater is shown in Fig. 4.

Step 2: Effective stress based soil water fully coupled dynamic analysis to investigate the seismic behavior of ground and breakwater during earthquake loading. In this step, static consolidation process followed by each earthquake shock is considered. Excess pore water pressure would develop in liquefiable sand layer, and the ground deformation would begin to accumulate.

Step 3: The long-term static analysis after earthquake loading, considering a complete consolidation in 3.5 years to examine the post-seismic behavior of breakwater and ground soil. The detailed loading process is listed in Table 5.

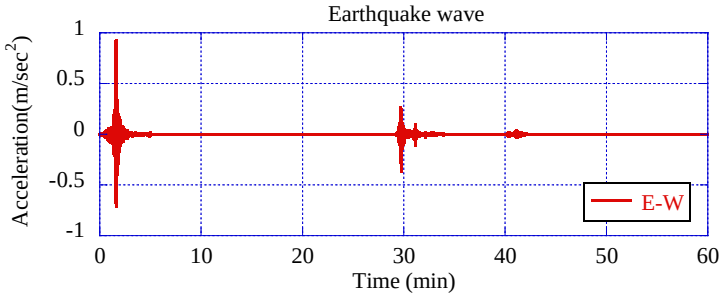


Figure 2. Recorded earthquake loading in E-W direction during the 2011 Great East Japan Earthquake

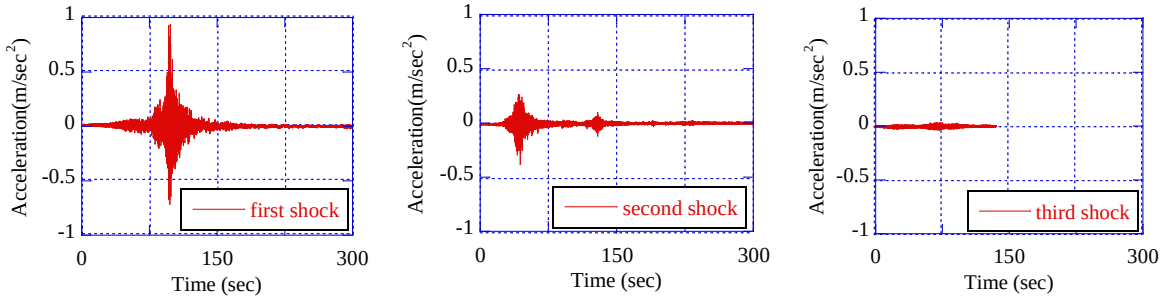


Figure 3. Three shocks of the earthquake loading during the 2011 Great East Japan Earthquake

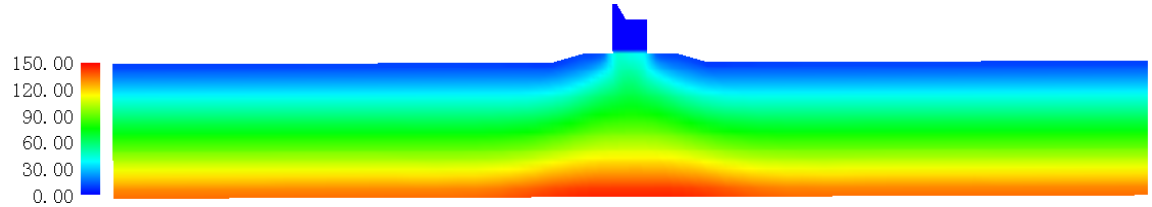


Figure. 4 Distribution of initial mean effective stress in the breakwater and foundation system due to gravity (unit: kPa)

Results and discussions

Seismic responses of breakwater and foundation soil

The seismic responses of breakwater and foundation soil under the earthquake loading are investigated. Fig.5 shows the horizontal acceleration responses of P-1 at the top of breakwater and P-2 at the bottom of caisson under the earthquake loading. The acceleration seismic responses at the two points are very similar and the amplification from the bottom of caisson to the top of breakwater is not obvious. However, the acceleration seismic responses are damped out by soil in the middle sand layer comparing with the input earthquake wave as shown in Fig. 6. The peak value of horizontal acceleration decreases obviously for the soil in liquefiable sand layer (GL-15 m), while the seismic wave was transmitted well in the upper clay layer (GL-5 m & GL-0 m). The amplitude of acceleration decreased as the building up of excess pore water pressure (Su et al. 2013) [39], and the soil's shear strength is reduced, which hampers effective propagation of shear waves to the soil surface. As the EPWPR value was larger at the middle sand layer (Fig. 8), the accelerations were highly attenuated relative to the base input (Fig. 6). Moreover, the attenuation of acceleration due to the loss of soil stiffness and strength was more significant in the near filed than that in the far field at the up clay layer. It was confirmed by Fig. 7 of the relationship between shear strain and shear stress, that larger shear strain in near field than that in far field was considered to be influenced by the replaced sand soil with high permeability below the breakwater structure.

Table 5. Loading process in liquefaction-consolidation analysis (a major shock followed by two aftershocks)

Step	Analysis type	Loading type	Calculation time (min)
1	Dynamic analysis	Major shock	5.00 (300 sec)
2	Static analysis	Consolidation	24.00 (1440 sec)
3	Dynamic analysis	First aftershock	5.00 (300 sec)
4	Static analysis	Consolidation	6.00 (360 sec)
5	Dynamic analysis	Second aftershock	2.25 (135 sec)
6	Static analysis	Consolidation	3.5 years

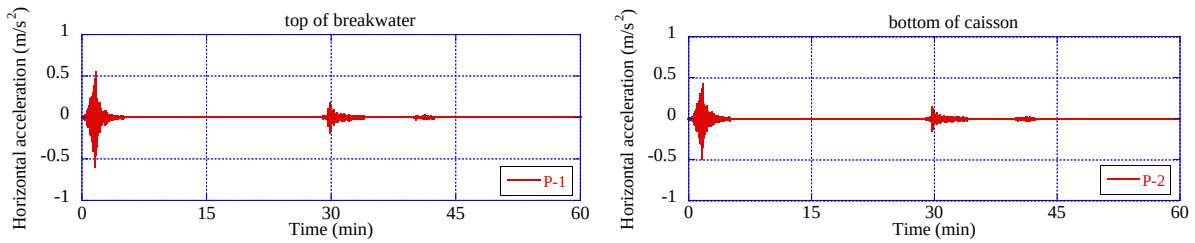
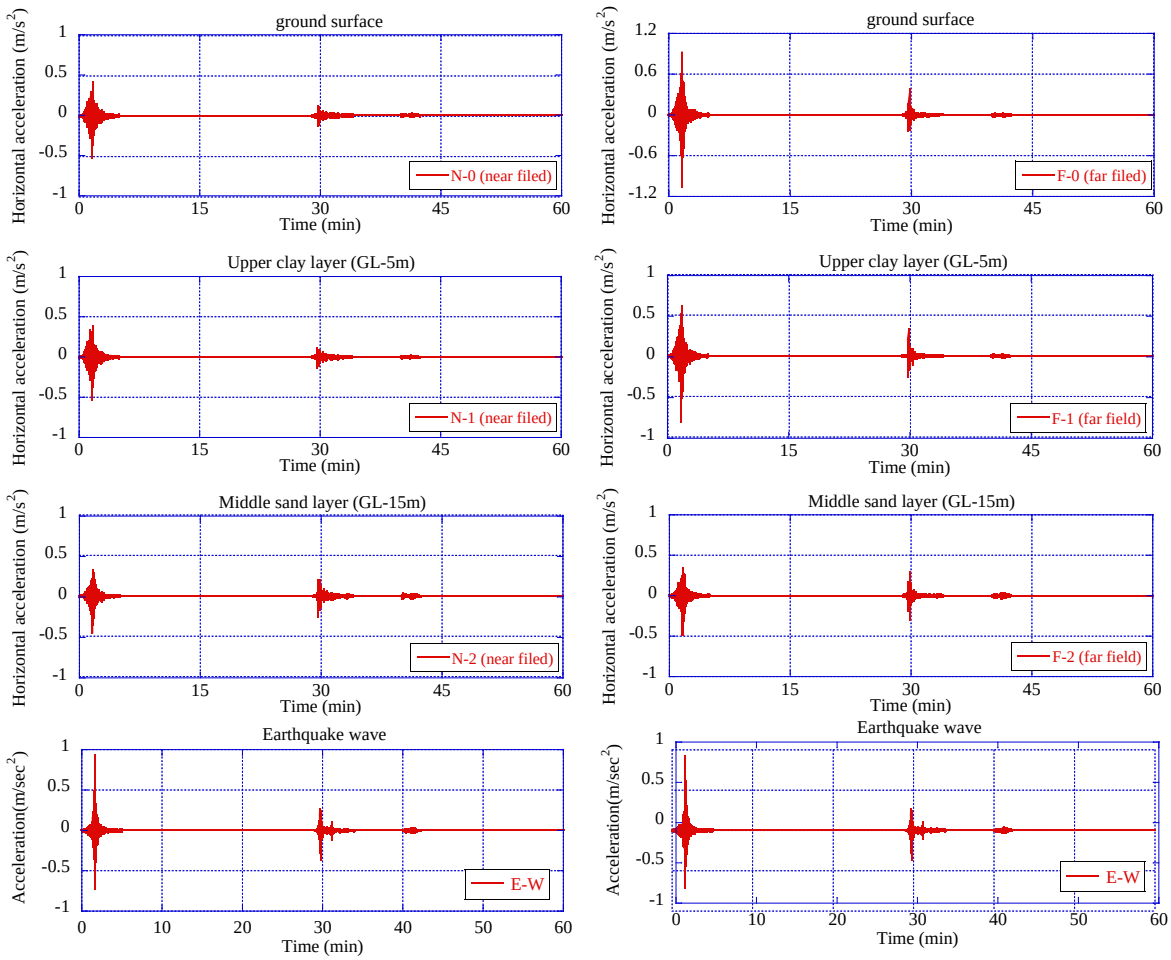
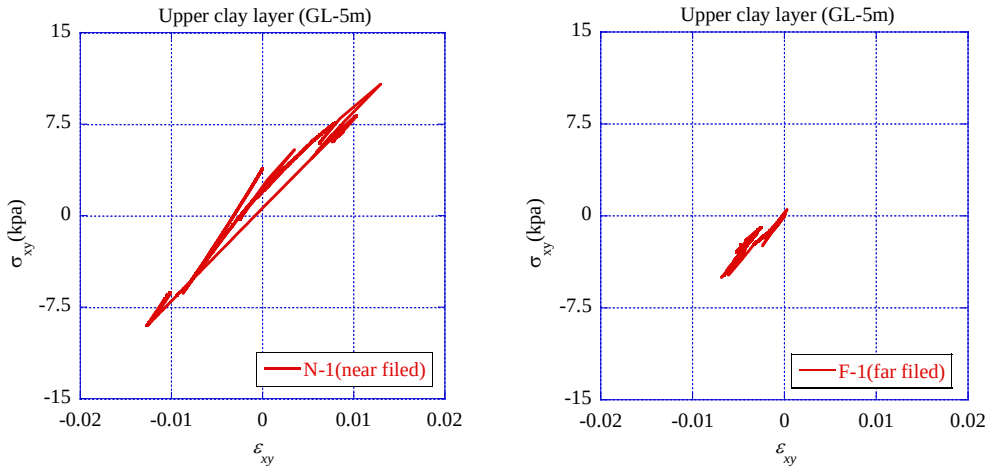


Figure 5. Horizontal acceleration responses of breakwater under earthquake loading



(a) Near field (b) Far field

Figure 6. Horizontal acceleration responses at different depth of foundation soil under earthquake loading



(a) Near field (b) Far field

Figure 7. Comparison of shear stress-strain relationship in upper clay layer in near field and far field

Liquefaction analysis

Fig. 8 shows the time history of excess pore water pressure ratio (EPWPR), which is defined as the ratio of excess pore water pressure (EPWP) to the initial vertical effective stress, at the

selected location (see Fig. 2 for locations of these points). Comparing the results in upper clay with that in middle sand layer, it was clear that liquefaction occurred seriously in middle sand layer. The EPWPR values were significantly smaller at upper clay layer and replaced sand region (GL-5 m) throughout the shaking, revealing the clay soil and replaced sand had not yet liquefied. A small aftershock (the second shock) caused rapid increase in EPWPR, re-liquefying the middle sand layer (GL-15 m) at both near and far field and beneath the breakwater. EPWPR continued to increase and remained significantly larger until the end of earthquake. The dissipation of EPWP was in a slower rapid in near and far field than that in the region beneath breakwater, which could cause a slower rate of the settlement accumulation in near and far field than that beneath the breakwater. Obviously, this was attributed to the high permeability of the replaced sand soil beneath the breakwater structure.

As shown in Fig. 9, excess pore water generated rapidly with the highest value in middle sand layer below breakwater, in near field and far field. It is clear that liquefaction occurs at the end of the first shock, however, the liquefaction area become large at the end of the second shock, and at the end of the third shock, large area of liquefaction still remains in the middle sand layer. The thicker the upper clay layer is, the longer the duration of liquefaction is. This is because the dissipation of large excess pore pressures generated in the deeper depth leads to a longer duration of flow to the shallower depth. In addition, the replaced sand soil beneath breakwater is not fully liquefied during the whole earthquake loading because of its high permeability and the overlying breakwater structure which constricted soil liquefaction. Moreover, the replaced sand soil beneath breakwater might have reduced the degree of liquefaction of the soil lying below around the centerline and allowed the lateral stretching of the soil below the replaced sand towards the free field.

Fig. 10 shows the dissipation process of EPWP. The pore water was accumulated in middle sand layer beneath the clay layer as the clay layer acted as the barrier for vertical dissipation of EPWP. It was found that EPWP remains for a longer period of time in middle sand layer below upper clay layer compared with the region below the replaced sand soil. In the region around centerline below breakwater, EPWP become much lesser and the dissipation was quite faster after earthquake shakings (after t = 4 hours shown in Fig. 10). This might be due to the reason that the presence of replaced sand region underneath breakwater distributes the out flow of pore water. Overall, the dissipation of pore water was concentrated through the discontinuity region below the breakwater and finally towards the ground surface, contracting the foundation soil below breakwater and inducing additional settlement after shaking. In another word, EPWP remained for a longer period of time at discontinuous regions in non-homogeneous soil deposits, manifesting a larger settlement at that corresponding region causing non-uniform settlements. A significant amount of non-uniform settlement took place during and after earthquake shaking as shown in Figs.11&12. The value of EPWP build-up beneath breakwater was larger than that in other locations, which caused larger amount of settlement at breakwater than at ground surface in near and far field. The total amount of settlements at ground surface in near field and far field are 0.697 m and 0.688 m respectively, which was smaller than the settlement of the breakwater with a value of 0.815 m after complete consolidation of the ground as shown in Table 6. Obviously, the aftershock (the second shock) caused additional amount of settlement to breakwater structure (Fig. 11). The settlements occurred during earthquake shakings are almost the same at ground surface in both near field and far filed except for the small amount of heave at ground surface in the near filed (Fig. 12). However, the settlement developed faster in near filed than that in far field under post-earthquake consolidation process because of the quick out flow of pore water from near filed to the replaced sand region. As the pore water pressure dissipated mainly through the discontinuity, the complete dissipation took a long period of time, about 3.5 years (Fig.

10). An additional breakwater settlement of 0.277 m was measured due to post-seismic and dissipation of EPWP. The heaving at ground surface in near field occurring during the main shock shaking also settled down to a final settlement of 0.697m.

The total amount of settlements of ground in near field and far field were smaller than the settlement of the breakwater after complete consolidation of the ground. This might be due to the lager volume strain of replaced sand soil underneath breakwater during the dissipation of pore water. As the settlement induced due to dissipation of pore water after earthquake shaking were significantly larger in near field (0.549 m) and far field (0.547 m) than that at breakwater (0.277 m), dissipation of EPWP became the major factor after the earthquake shaking stopped, which caused larger amount of additional ground settlement than that during earthquake shaking.

Table 6. Amount of settlement at different positions (Unit: m)

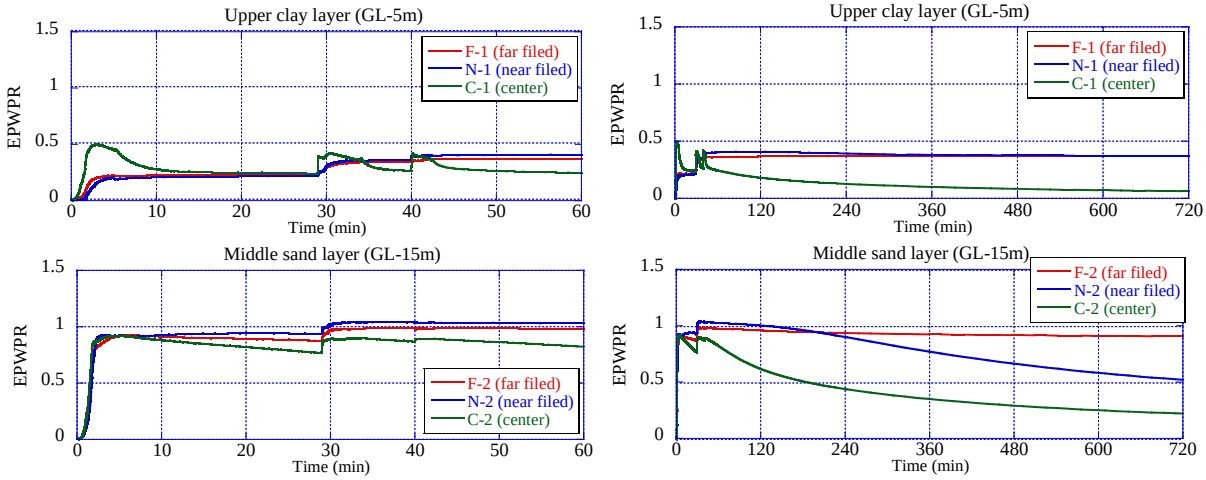
Time	P-1	N-0	F-0
At the end of earthquake	0.538	0.148	0.141
3.5 years after earthquake	0.815	0.697	0.688

Deformation of the breakwater and foundation system

From Figs. 11&12 of the time histories of vertical displacements at top of breakwater and ground surface in both near field and far field, as mentioned above, a total settlement of 0.815 m for breakwater structure was observed, of which 0.538 m (66%) was measured during the main shock shaking (Table. 6). The main mechanisms that contribute to the settlement of foundation on a liquefied soil layer are volumetric compaction and shear deformation of the soil mass underneath the foundation. Shear deformation is accompanied by the lateral spreading of the non-liquefied soil below the structure which is initiated when the soil in the free field adjacent to the underlying soil on either side of the breakwater liquefies and loses its shear strength and allows the newly unconstrained soil below the breakwater collapse vertically and spread outwards. This type of settlement causes considerable vertical strain with no volume change. Concurrently, volumetric compaction of the sand mass under the upper clay layer occurs which results in both vertical and volumetric strains. This settlement results in the disruption of soil structure and rearrangement of soil grains and is the main mechanism responsible for the settlement in the free field (Maharjan & Takahashi 2014) [20]. It is difficult to separate the volumetric compaction effect from the shear deformation effect beneath the breakwater as they both happen at the same time. The final mechanism involved in the foundation soil settlement is the long-term dissipation of the excess pore pressure (consolidation).

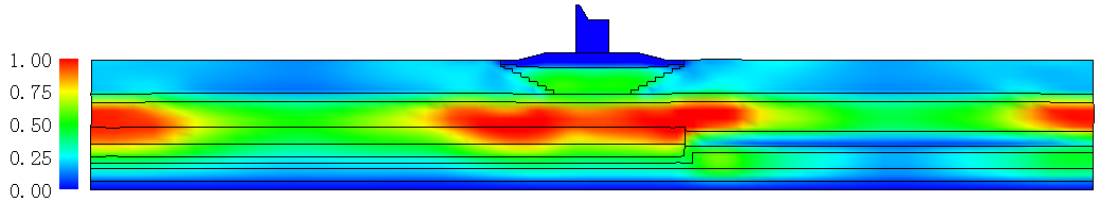
From Table 6 of the calculated values for the settlement at the end of earthquake and final settlements of the breakwater and ground surface, as mentioned above, most part of the breakwater settlement (66%) accumulated during earthquake shaking. After ending of the earthquake shaking, the settlement of breakwater increases with a lower rate and ceases to increase when dissipation of excess pore pressure is completed. However, for the settlement of ground surface in both near field and far field, it is notable that the significant part of the settlement takes place in the process of pore pressure dissipation after the end of earthquake shaking (78.8% and 79.5% in the near field and far field, respectively). The calculated settlements of breakwater and ground surface differ to some extent from each other. This difference can be attributed to the upper clay layer that hindered the dissipation of pore water for near and far filed soil, and conversely, the replaced sand soil beneath the breakwater structure that accelerated the dissipation of pore water around centerline below the breakwater.

Obviously, the overall deformation of the ground around breakwater was large as shown in Fig.13 of displacement vector of breakwater and foundation system. The soil near breakwater translated sideways and lateral deformation was observed at the two sides of breakwater during earthquake shaking, especially in the middle sand layer that was found to laterally spread on both sides towards the free field. This caused serious settlement of breakwater. Shear deformation of underlying liquefied sand and volumetric change due to pore water dissipation are also factors for breakwater and ground settlements. As the presence of the upper clay layer acted as a hindrance and it took about 3.5 years for the water complete its dissipation through the discontinuous region according to the calculation results.

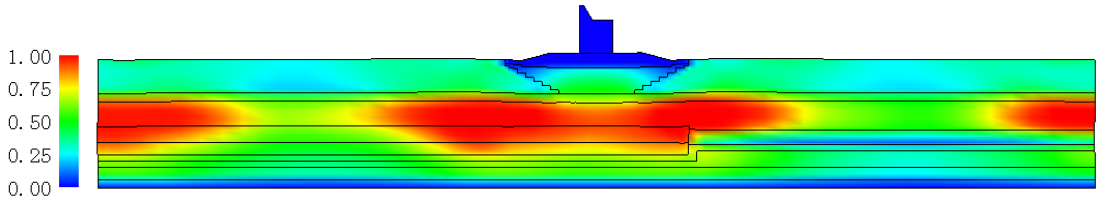


(a) Within 60 minutes (b) Within720 minutes

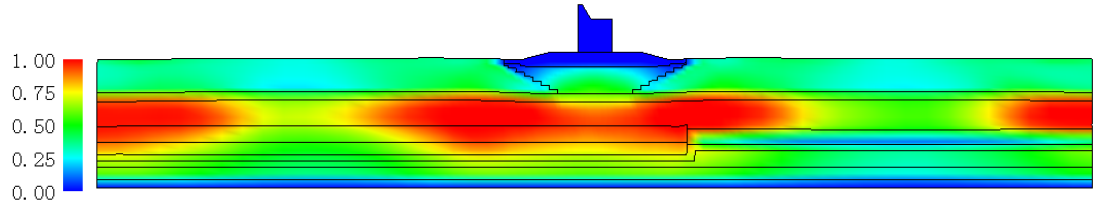
Figure 8. Time history of EPWPR at different depth of foundation soil



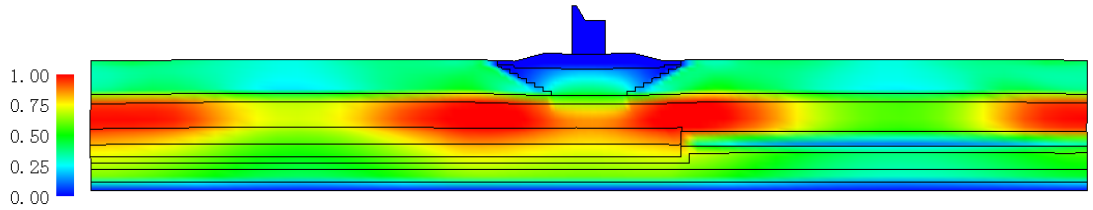
(a) At the end of the first shock (t = 5 minutes)



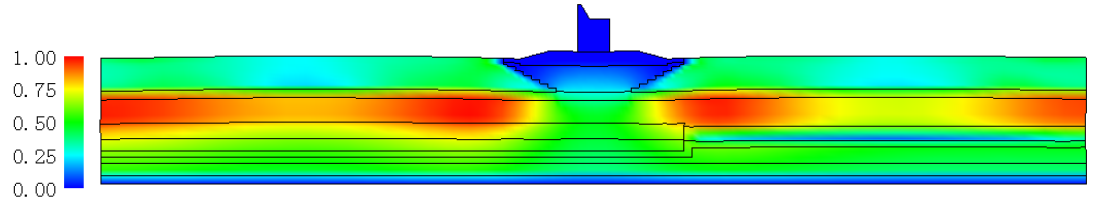
(b) At the end of the second shock (t = 34 minutes)



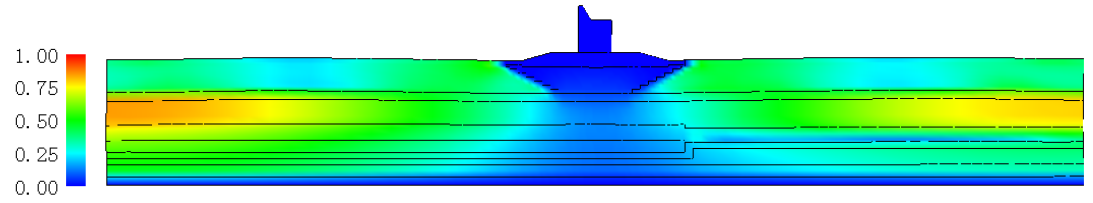
(c) At the end of the third shock (t = 42.25 minutes)



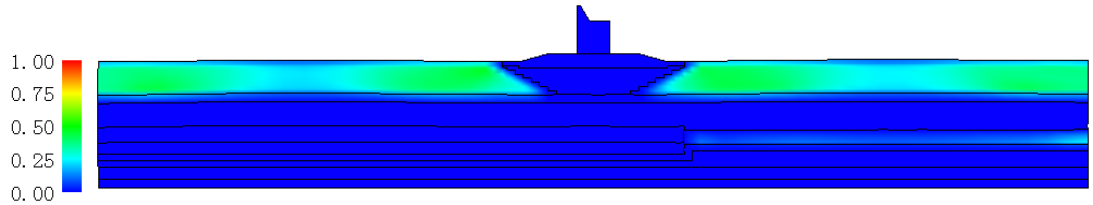
(d) After earthquake shaking (t = 1 hour)



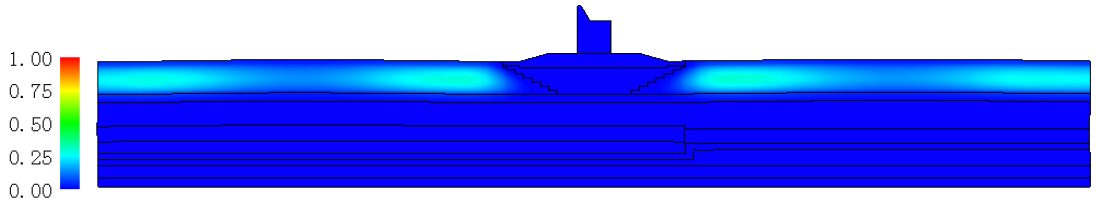
(e) After earthquake shaking (t = 4 hours)



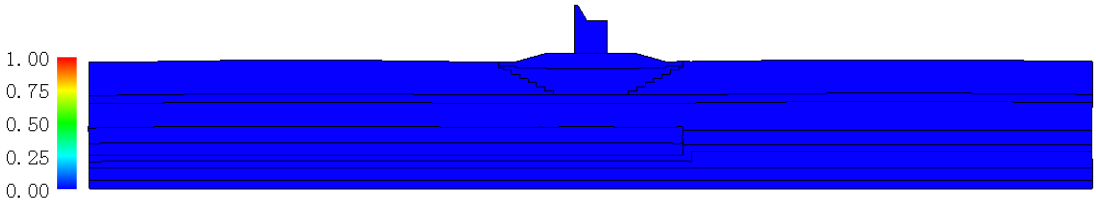
(f) After earthquake shaking (t = 1.3 days)



(g) After earthquake shaking (t = 13 days)

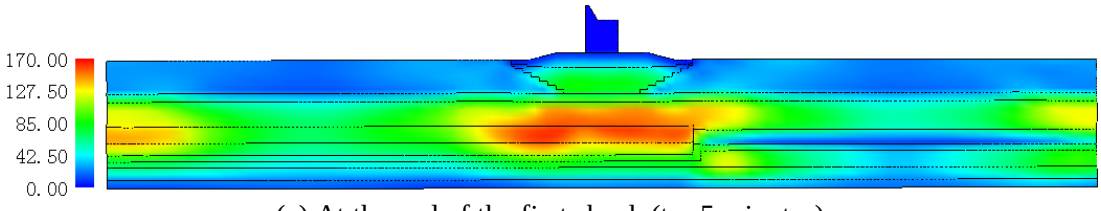


(h) After earthquake shaking (t = 130 days)

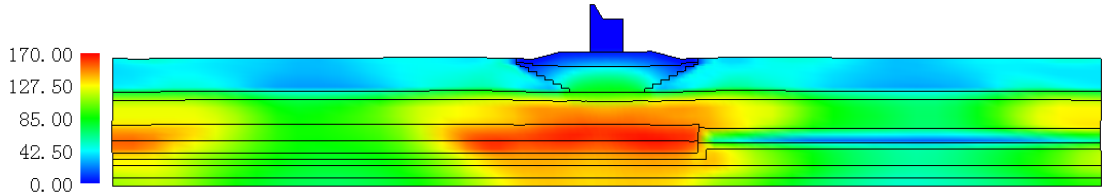


(i) After earthquake shaking (t = 3.5 years, complete dissipation of EPWP)

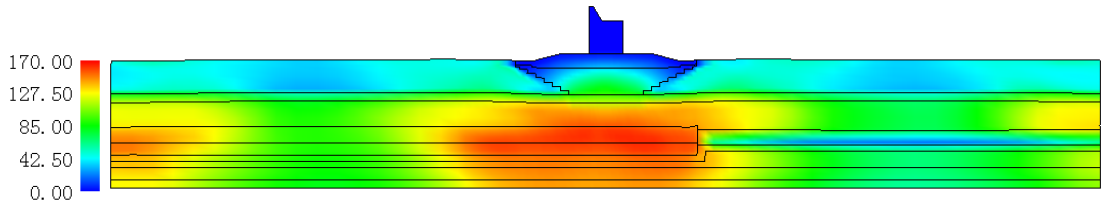
Figure 9. Distribution of Excess pore water pressure ratio at different time



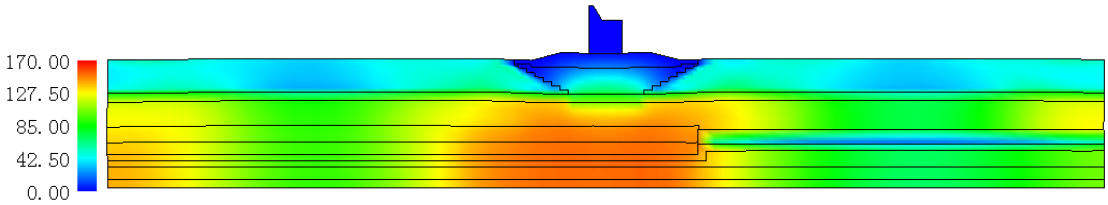
(a) At the end of the first shock (t = 5 minutes)



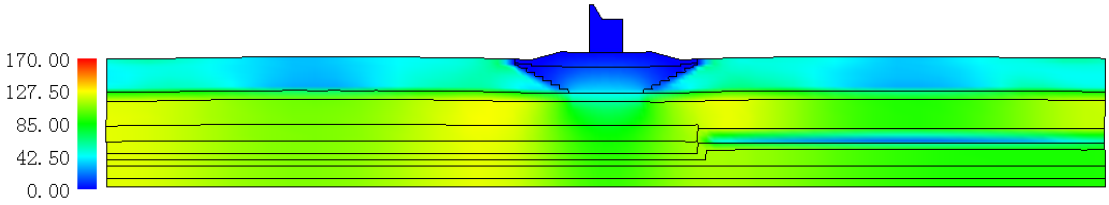
(b) At the end of the second shock (t = 34 minutes)



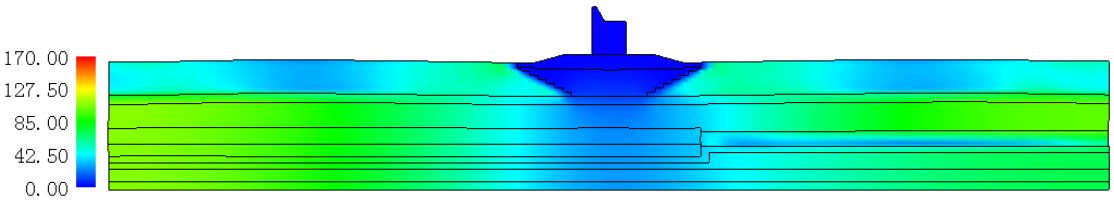
(c) At the end of the third shock (t = 42.25 minutes)



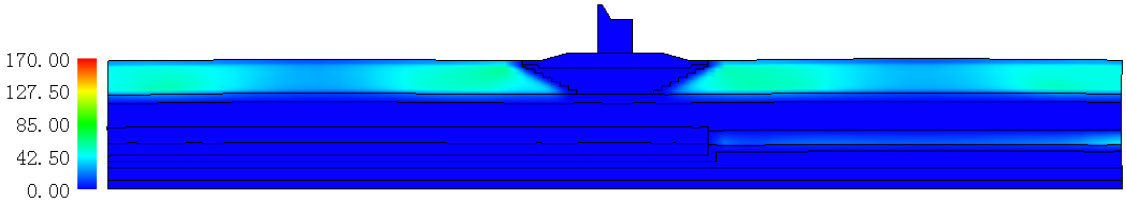
(d) After earthquake shaking (t = 1 hour)



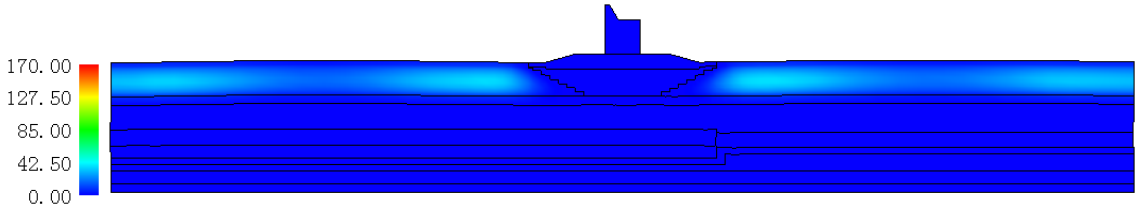
(e) After earthquake shaking (t = 4 hours)



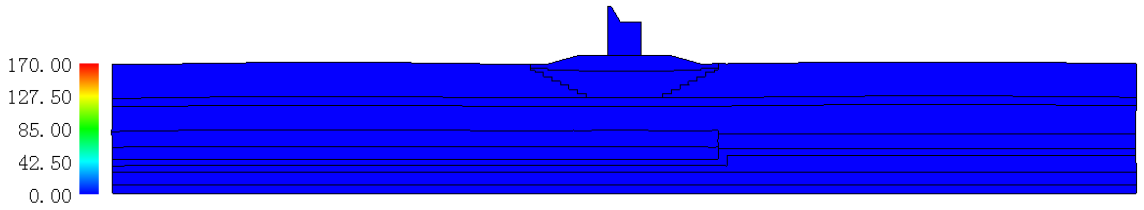
(f) After earthquake shaking (t = 1.3 days)



(g) After earthquake shaking (t = 13 days)

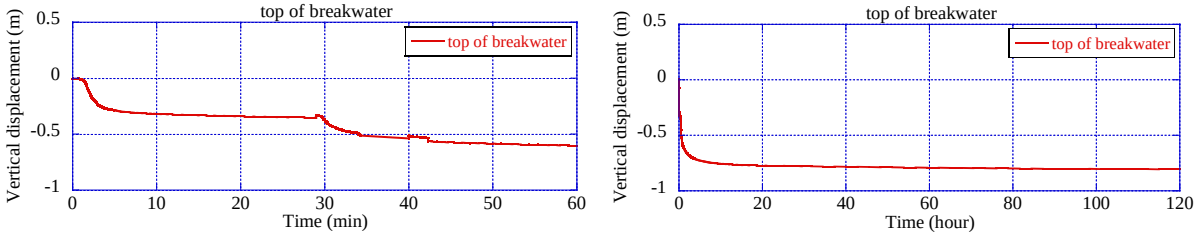


(h) After earthquake shaking (t = 130 days)



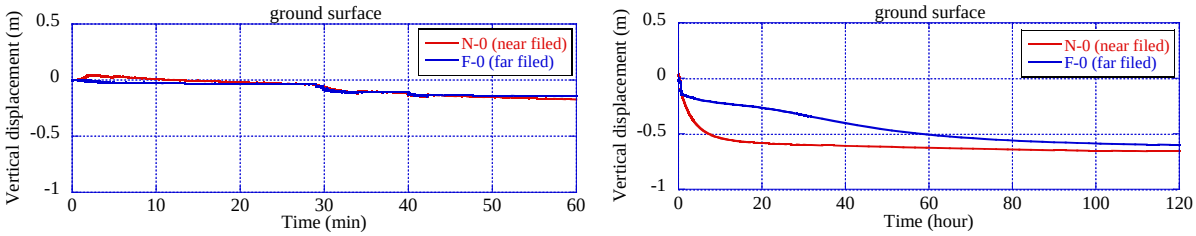
(i) After earthquake shaking (t = 3.5 years, complete dissipation of EPWP)

Figure 10. Dissipation process of excess pore water pressure (unit: kPa)



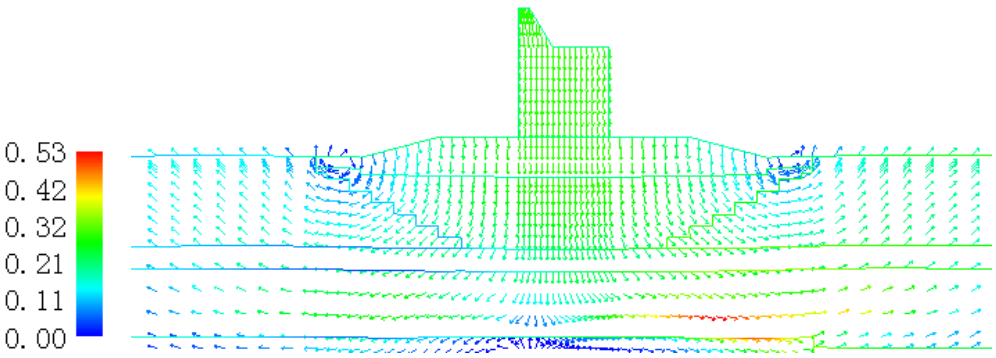
(a) Within 60 minutes (b) Within120 hours

Figure 11. Time history of vertical displacement at the top of breakwater

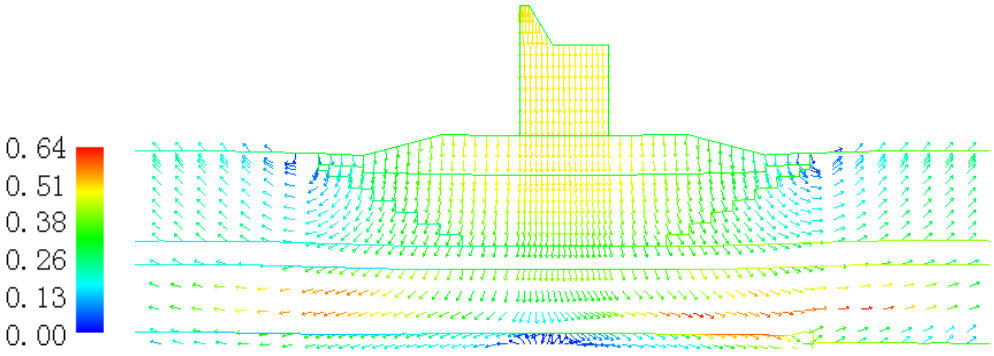


(a) Within 60 minutes (b) Within120 hours

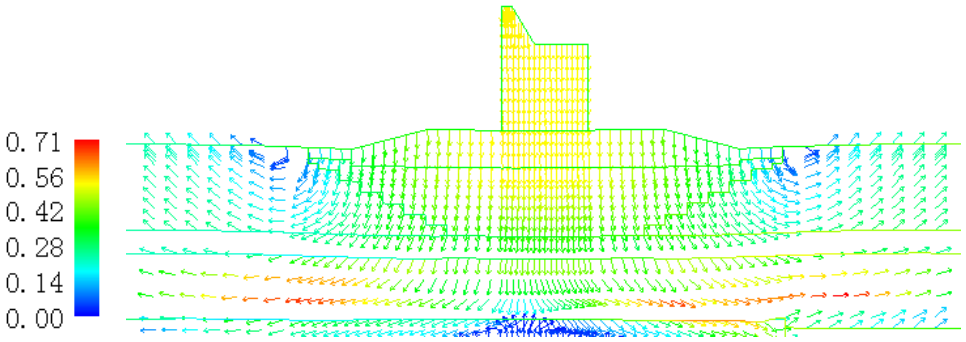
Figure 12. Time history of vertical displacement at ground surface in near field and far field



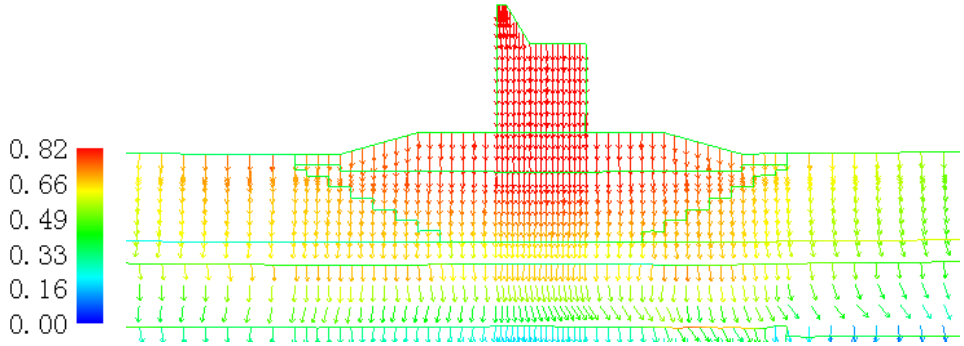
(a) At the end of the first shock (unit: m)



(b) At the end of the second shock (unit: m)



(c) At the end of the third shock (unit: m)



(d) After complete dissipation of EPWP (unit: m)

Figure 13. Displacement vector of breakwater and foundation system during and after earthquake loading (A part of mesh)

Conclusions

In this study, the co-seismic and post-seismic performance of a caisson type breakwater resting on the natural ground with discontinuous low permeability an liquefiable layers subjected to the 2011 Great East Japan Earthquake is investigated using soil-water coupled finite element method. Based on the calculated results, the following conclusions can be drawn:

1. The repeated earthquake shaking has a significant effect on the accumulated deformation of embankments. The second aftershock caused an increase in EPWP generation and an additional settlement. Moreover, the effects of aftershocks were more pronounced in the non-homogeneous liquefiable foundations, leading to the post-liquefaction delayed settlement and this conclusion was also confirmed by Maharjan and Takahashi (2014) [20].
2. The replaced sand region with a high permeability has faster dissipation of pore water while the dissipation continued for a longer time period in near and far field of ground, accumulating delayed displacements. Overall, the dissipation of pore water was concentrated through the discontinuity region below the breakwater and finally towards the ground surface, contracting the foundation soil and inducing additional settlement after shaking and causing larger amount of settlement on breakwater than that on ground surface in near and far filed.
3. The accumulation of pore water beneath the low permeability upper clay layer induced large shear strain in middle sand layer, resulting large amount of lateral spreading. Lateral spread, shear deformation of underlying liquefied sand and volumetric change due to pore water dissipation are the main factors for breakwater and ground

settlements. The presence of the upper clay layer acted as a hindrance and it took about 3.5 years for the water complete its dissipation through the discontinuous region according to the calculation results.

4. The thick clay layer may cause long term consolidation process while the thick sand layer may bring a large area of liquefaction and severe ground deformation. Although the replaced sand soil beneath breakwater structure can improve ground bearing capacity, it may cause the risk of large amount of settlement to breakwater, which can reduce capacity of the breakwater to resist tsunami after earthquake loading.

Acknowledgements

This work was supported by National Nature Science Foundation of China (Grant No. 51308346), Guangdong Natural Science Foundation (Grant No. 2014A030313551), Natural Science Foundation of SZU (Grant No.201420), Foundation of Key Laboratory of Geotechnical and Underground Engineering (KLE-TJGE-B1504) and Scientific and Technical Innovation Foundation of Shenzheng (Grant No. JCYJ20150324140036839).

References

[1] Seed, H.B. (1968) Landslides during earthquakes due to liquefaction, *J Mech Found Div* 1968; **94**(5), 1055–123.
[2] Adalier, K., Elgamal, A.W., Martin, G.R. (1998) Foundation liquefaction countermeasures for earthembankments, *J Geotech Geoenviron Eng* **124** (6), 500–17.
[3] Huang, Y., Yu, M. (2013) Review of soil liquefaction characteristics during major earthquakes of the twenty-first century, *Natural Hazards* **65**, 2375-2384.
[4] Oka, F., Tsai, P., Kimoto, S., Kato, R. (2012) Damage patterns of river embankments due to the 2011off the Pacific Coast of Tohoku earthquake and a numerical modeling of the deformation of river embankments with a clayey subsoil layer, *Soils Found* **52**(5), 890–909.
[5] Mori, N., Cox, D.T., Yasuda, T., Mase, H. (2013) Overview of the 2011 Tohoku Earthquake Tsunami damage and relation with coastal protection along the Sanriku coast, *Earthquake Spectra* **29**(S1), S127–S143.
[6] Mori, N., Yoneyama, N., Pringle, W. (2015) Effects of the Offshore Barrier Against the 2011 Off the Pacific Coast of Tohoku Earthquake Tsunami and Lessons Learned, Post-Tsunami Hazard, *Advances in Natural and Technological Hazards Research* **44**, 121-132.
[7] Fujima, K. (2006) Effect of a submerged bay-mouth breakwater on tsunami behavior analyzed by 2D/3D hybrid model simulation, *Natural Hazards* **39** (2), 179-193.
[8] Imase, T., Maeda, K., Miyake, M., Sawada, Y., Sumida. H., Tsurugasaki, K. (2012) Destabilization of a caisson-type breakwater by scouring and seepage failure of the seabed due to a tsunami, *Proceedings of International Conference on Scour and Erosion*, 128–135.
[9] Susumu, I. (2012) Combined Failure Mechanism of a Breakwater Subject to Tsunami during 2011 East Japan Earthquake, *Geological and Earthquake Engineering* **37**, 177-186.
[10] Takahashi, H., Sassa, S., Morikawa, Y., Takano, D., Maruyama, K. (2014) Stability of caisson-type breakwater foundation under tsunami-induced seepage, *Soils and Foundations* **54** (4), 789-805.
[11] Memos, C., Kiara, A., Pavlidis, E. (2003) Coupled seismic response analysis of rubble-mound breakwater, *Water and Maritime Engineering* ICE, 23–31
[12] Yuksel, Y., Cetin, K.O., Ozguven, O., Isik, N.S., Cevik, E., Sumer, B.M. (2004) Seismic response of a rubble mound breakwater in Turkey, *Proc Inst Civil Eng: Marit Eng* **157**(4), 151–161.
[13] Jafarian. Y., Alielahi. H., Abdollahi, A.S., Vakili, B. (2010) Seismic numerical simulation of breakwater on a liquefiable layer: Iran LNG port, *Electron J Geotech Eng* 15D, 1–11.
[14] Ye, J.H. (2012) Seismic response of poro-elastic seabed and composite breakwater under strong earthquake loading, *Bulletin of Earthquake Engineering* **10**(5), 1609-1633.
[15] Aydingun, O., Adalier. K. (2003) Numerical analysis of seismically induced liquefaction in earth embankment foundations. PartI. Bench mark model, *CanGeotech J* **40**(4), 753–65.
[16] Adalier, K., Sharp, M.K. (2004) Embankment dam on liquefiable foundation-dynamic behavior and densification remediation, *J Geotech Geoenviron Eng* **130** (11), 1214–24.
[17] Ye, J.H., Wang, G. (2015) Seismic dynamics of offshore breakwater on liquefiable seabed foundation, *Soil Dynamics and Earthquake Engineering* **76**, 86-99.
[18] Huang, Y., Bao, Y.J., Zhang, M., Liu, C., Lu, P. (2015) Analysis of the mechanism of seabed liquefaction induced by waves and related seabed protection, *Natural Hazards* **79**, 1399-1408.
[19] Zhang, S., Wang, G., Sa, W. (2013) Damage evaluation of concrete gravity dams under mainshock–aftershock seismic sequences, *Soil DynEarthqEng* **50**, 16–27.

[20] Maharjan, M., Takahashi, A. (2014) Liquefaction-induced deformation of earth embankments on non-homogeneous soil deposits under sequential ground motions, *Soil Dynamics and Earthquake Engineering* **66**,161–9.
[21] Ye, B., Ye, G.L., Zhang, F., Yashima, A. (2007) Experiment and numerical simulation of repeated liquefaction-consolidation of sand, *Soils Found* **47**(3), 547–58.
[22] Xia, Z.F., Ye, G.L., Wang, J.H., Ye, B., Zhang, F. (2010) Fully coupled numerical analysis of repeated shake-consolidation process of earth embankment on liquefiable foundation, *Soil Dyn Earthq Eng* **30**(11), 1309–18.
[23] Sasaki, Y., Towhata, I., Miyamoto, K., Shirato, M., Narita, A., Sasaki, T., Sako, S. (2012) Reconnaissance report on damage in and around river levees caused by the 2011off the Pacific coast of Tohoku earthquake, *Soils Found* **52** (5), 1016–32.
[24] Ye, G.L. (2011) DBLEAVES: User's manual Version 1.6. *Shanghai Jiaotong University* (In Japanese and Chinese) China.
[25] Zhang, F., Ye, B., Noda, T., Nakano, M., Nakai, K. (2007) Explanation of cyclic mobility of soils: approach by stress-induced anisotropy, *Soils and Foundations* **47**(4), 635-648.
[26] Zhang, F., Ye, B., Ye, G.L. (2011) Unified description of sand behavior, *Frontiers of Architecture and Civil Engineering in China* **5**(2), 121–150.
[27] Iai, S., Ichii, K., Liu, H.L. (1998) Effective stress analysis of port structures, *Soils Found* (Special Issue of Geotechnical Aspects of the January 17, 1995 Hyogoken Nambu Earthquake) (2), 97–114.
[28] Hashiguchi, K., Ueno, M. (1977) Elastoplastic constitutive laws of granular material, Constitutive Equations of Soils, *Proc. 9th Int. Conf. Soil Mech. Found. Engrg., Spec*. Ses. 9, Murayama, S. and Schofield, A. N. (eds.) Tokyo, JSSMFE: 73-82.
[29] Asaoka, A., Noda, T., Yamada, E., Kaneda, K., Nakano, M. (2002) An elasto-plastic description of two distinct volume change mechanisms of soils, *Soils and Foundations* **42**(5), 47-57.
[30] Ye, B., Ye, G.L., Zhang, F. (2012) Numerical modeling of changes in anisotropy during liquefaction using a generalized constitutive model, *Computers and Geotechnics* **42**, 62-72.
[31] Ye, G.L., Sheng, J.R., Ye, B., Wang, J.H. (2012) Automated true triaxial apparatus and its application to over-consolidated clay, *Geotechnical Testing Journal* **35**(4), 517-528.
[32] Ye, G.L., Ye, B., Zhang, F. (2013) Strength and dilatancy of overconsolidated clays in drained true triaxial tests, *Journal of Geotechnical and Geo-environmental Engineering*, **140**(4), 06013006.
[33] Ye, B., Ye, G.L., Zhang, F., Yashima, A. (2007) Experiment and numerical simulation of repeated liquefaction-consolidation of sand, *Soils and Foundations*, **47** (3), 547-558.
[34] Zhang, F., Jin, Y., Ye, B. (2010) A try to give a unified description of Toyoura sand, *Soils and Foundations* **50** (3), 679-693.
[35] Jin, Y., Bao, X.H., Kondo, Y., Zhang, F. (2010) Soil-water coupling analysis of real-scale field test for 9-pile foundation subjected to cyclic horizontal loading. *Geotechnical Special Publication*, *Deep Foundation and Geotechnical in situ Test ASCE* **205**,111-118.
[36] Bao, X.H., Morikawa, Y., Kondo, Y., Nakamura, K., Zhang, F. (2012) Shaking table test on reinforcement effect of partial ground improvement for group-pile foundation and its numerical simulation, *Soils and Foundations* **52**(6), 1043-1061.
[37] Bao, XH., Ye, B., Ye, G.L., Sagou, Y. and Zhang, F. (2014) Seismic performance of SSPQ retaining wall-centrifuge model tests and numerical evaluation, *Soil Dynamics and Earthquake Engineering* **61-62**, 63-82.
[38] Bao, X.H., Ye, G.L., Ye, B. and Zhang, F. (2016) Co-seismic and post-seismic behavior of an existed shallow foundation and super structure system on a natural sand/silt layered ground, *Engineering Computations* **33**(1), 288-304.
[39] Su, D., Ming, H.Y., Li, X.S. (2013) Effect of shaking strength on the seismic response of liquefiable level ground, *Engineering Geology* **166**, 262-271.

Numerical Simulation for Combined Blast & Fragment Effects on RC Slabs

†Shengrui Lan and Kenneth B. Morrill

Karagozian & Case, Inc., 700 N. Brand Blvd., Suite 700, Glendale, CA 91203, USA

†Corresponding author: lan@kcse.com

Abstract

The objective of this study is to evaluate the residual loading capacity of the damaged RC slabs by the combined blast and fragment loading (CBFL) effects. High-fidelity physics-based (HFPB) finite element analysis technique is used for the numerical simulations in this study, which takes into account material nonlinearity, strain rate effects, large deformation behavior, "real time" blast and fragment loading, and actual supporting boundary conditions. Numerical model and simulation techniques have been validated through five tests including the quasi-static tests, the blast loading only test and the CBFL tests, in comparison of the deformation and pristine/residual loading capacity of the RC slabs, which was carried out without knowing the test results. Using a fast running tool, fragments and blast loading are generated on full scale RC slabs. A parametric study has been done to investigate dynamic response of the full-scale RC slabs under CBFL effects.

Keywords: Blast loading, fragment loading, RC slab, dynamic response, residual capacity.

Introduction

When a cased munition or an improvised explosive device (IED) detonates nearby a structure, the structure is subjected to a combination of blast and fragment loading (CBFL). Dynamic response of reinforce concrete (RC) slabs under the CBFL may be different from that under the air blast loading. Some studies have been done in this area for the formation of fragments loading and their effects on the structural members, e.g., the works reported in reference [1-5]. A series of small scale experiments using bare charge with pre-formed ball bearings were conducted by Swedish Research Institute (FOI) to investigate the effects of the combined blast and fragment loadings [4]. The objective of the study was to develop a fast running tool to account for the fragmentation effect on the doubly reinforced concrete slab. As part of the study, numerical simulations were conducted for five of the experiments. This paper will focus on the numerical simulation to predict the residual capacity of the RC slabs after CBFL effect.

High-fidelity physics-based (HFPB) finite element analysis (FEA) technique is used for the simulations in this study, which can take into account many physical behaviors of materials and structures, such as: (a) material nonlinearity and geometry nonlinearity (large deformation); (b) dynamic strain rate effects for material strength increase; (c) structural 3D behavior with complex stress states – not only flexural but also axial, shear and torsional behaviors/responses; (c) multi-components with multi-materials and structural details for connections – not only global but also localized response; (d) "Real time" blast loading – blast loads are generally applied with different arrival times and pressure time histories at different locations (i.e., non-uniform loading); (e) more realistic boundary conditions of the structure, instead of the artificial boundary conditions in SDOF model.

LS-DYNA (www.lstc.com) is used for the simulations in this study. LS-DYNA is a general-purpose finite element program capable of simulating complex dynamic structural problems, which has been widely used in blast and impact effects analysis communities.

K&C concrete material model (i.e., MAT_072 in LS-DYNA) has been incorporated in LS-DYNA, which enables that more reliable analysis results can be obtained for reinforced concrete (RC) structures under blast and impact effects. This is because this concrete model has implemented many key features of concrete materials: (a) three-invariant strength surfaces to reflect the pressure-dependent and difference in triaxial extension and compression; (b)

effects of confinement - compressive strength is significantly enhanced by confinement; (c) non-linearity - Elastic, plastic with hardening and softening (or damage), for which a damage metric is used in K&C model to gauge the evolution; (d) strain rate effects – significant material strength enhancement by high strain rate, which is important for blast loading effects where the concrete strength could be more than doubled; (e) Fracture energy – important tensile behavior of concrete; (f) Shear-dilatancy – Concrete's expansion upon cracking provides increased strength /ductility where confinement is adequate.

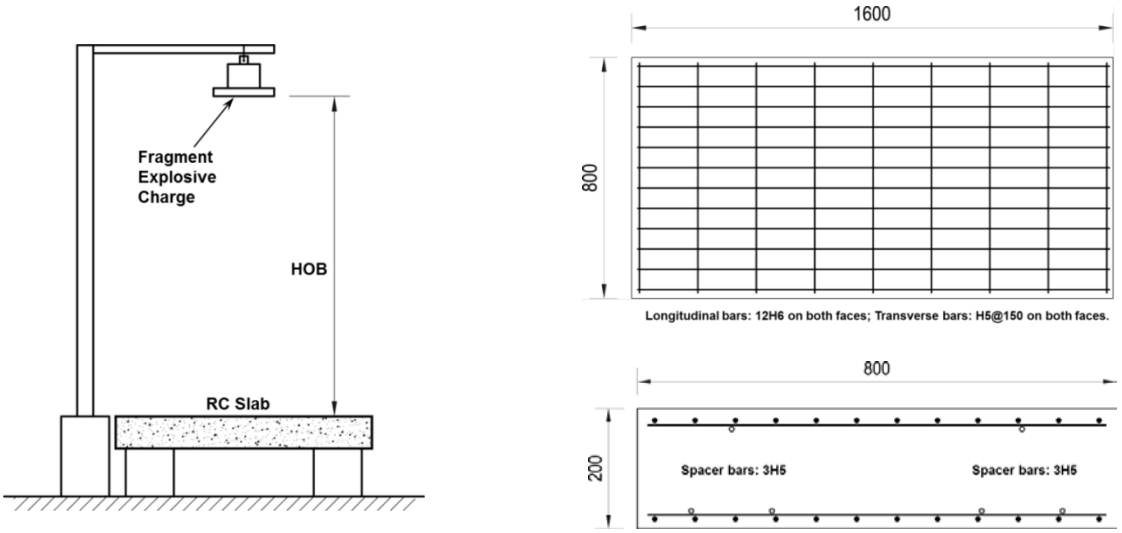


Figure 1. Test set-up and RC slab specimen.

RC Slab Tests Under Fragment Loading

A series of test has been conducted for RC slabs using the test set-up shown in Figure 1 [4]. Tables 1 and 2 summarize the information about the specimens and test data from five tests, which include two quasi-static loading (QSL) tests of the pristine slabs, two fragment loading tests followed by QSL tests of the damaged slabs for the residual capacity, and one test under CBFL.

Three approaches were taken using three types of set-ups: (a) the QSL test (three point loading flexural test) for loading capacity of the pristine slab specimens as a baseline control, (b) expose the specimens to the fragments loading then conduct QSL test of the damaged slabs for their residual loading capacity, and (c) Expose the specimens to the CBFL effects and record dynamic displacement histories; QSL test was not conducted for the damaged slab.

Table 1. Test specimens of RC slabs for quasi-static and dynamic tests.

Test No.	Slab No.	Spacer bar	Concrete Strength (MPa)		Pristine Capacity (kN)	Residual Capacity (kN)	Loading
			Cube	Cylinder			
24	20	No	39.9	31.9	178	-	Quasi-static
40	19	One side	43.2	34.5	183	-	Quasi-static.
19	18	Both sides	37.3	29.8	-	172	Fragment
41	16	Both sides	37.9	30.3	-	185	Fragment
52	22	No	33.2	26.6	-	-	Blast & Fragment

Note: a) The dimension of all slabs is 1600 x 800 x 200 mm. b) Longitudinal bars in all slabs are 12Φ6 as shown in Figure 1.

Table 2. Dynamic test results under fragment explosive charges.

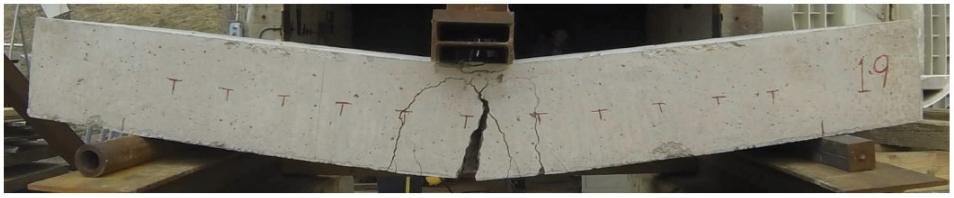
Test No.	Charge Weight (kg)	Fragment size (mm)	No. of Balls	Height of Burst (m)	Average Velocity from Test (m/s)	Fragment Density (kg/m^2)
19	8.847	Φ8	345	2.1	1,880	0.25
41	8.969	Φ8	346	1.9	1,880	0.30
52	8.877	Φ8	345	2.7	1,815	0.17

Blind Prediction of Test Results

Quasi-static Loading for Pristine Slabs

Two QSL tests (i.e., Test 24 and Test 40 in Table 1) were conducted for the pristine slabs as the baseline control data. The loading capacity was evaluated through three points loading flexural test on the simply supported slabs as shown in Figure 2. In those tests, the slabs were loaded quasi-statically till all the bottom longitudinal rebars fractured, which captured the post-peak low strength as well.

(a) Test 24 as described in Table 1.



(b) Test 40 as described in Table 1.

Figure 2. Two quasi-static tests for the pristine slabs.

Finite element models have been developed for the five specimen slabs (Table 1) to simulate the tests. As an example, simulation results for test 24 presented in Figure 3 show that concrete damage is concentrated at the middle span of the slab and rebars are fractured, which agrees the test failure mode shown in Figure 2a. The predicted loading capacities of 165 kN (Figure 4a) in Test 24 and 185 kN in Test 40 are quite close to the test results of 178 kN in Test 24 and 183 kN in Test 40, respectively. Those simulation results indicate the numerical model developed for the slab QSL test is valid. In addition, the predicted loading capacities of the five pristine slabs (Table 1) are presented in Figure 4b, which indicate the spacer bars have some influence.

Fragment Loading Effects

Two tests (i.e., Test 19 and Test 41) were carried out by the fragment loading for dynamic response and then QSL on the damaged slabs for their residual capacities. The tested specimen of Test 41 is shown in Figure 5. These two tests are simulated with following procedures: Step 1: Gravity loading is applied from 0 to 100 ms (t1); Step 2: Blast/fragment loading is applied from 100 to 200 ms (t2), where the fragments' velocities and positions are outputted from the fragment explosive charge simulation described in the previous section according to the height of burst in Table 2 and mapped on the test specimens; and Step 3:

Posttest QSL with a displacement loading on the loading bar of 50 mm/s from 200 ms until the rebars fracture.

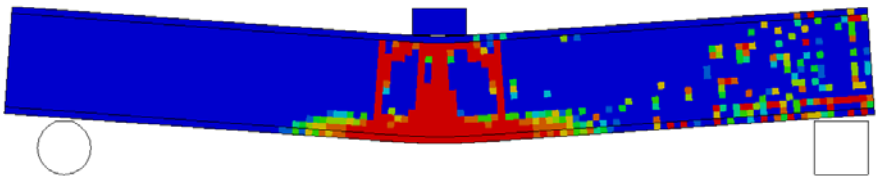
The simulation results for Test 41 in Figure 6 indicate that fragment damage is mainly on the top surface of the slab and the failure of the damaged slab in QSL simulation is due to rebar fracture. The residual capacity and the pristine capacity of the slab are compared in Figure 6d, which indicates that the residual capacity is about 91% of the pristine capacity. This is probably because the loading capacity is governed by the rebar fracture and the concrete damage has relatively less influence to the loading capacity.

An interesting observation from the simulation on the residual capacity is the damaged slab behaved more ductile than the pristine slabs, i.e., the force-displacement curve has a clear "softening" stage, instead of dropping immediately to the lowest value in the pristine slab as shown in Figure 6d. This is probably because all rebars in the pristine slabs are at the same stress status and break at the same time, whereas the rebars may be in slightly different stress status and break at the different time due to the non-uniform damage of concrete.

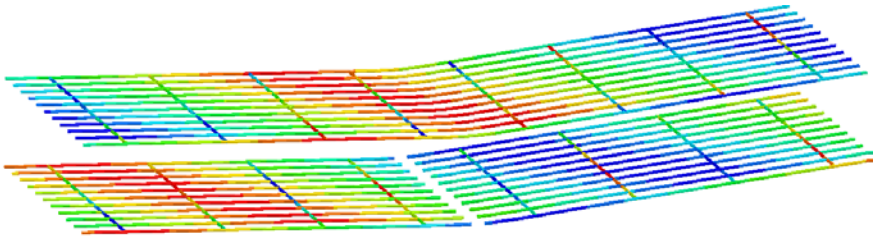
From the simulation results for Test 19, the same observations and conclusions can be drawn as those in Test 41. The residual capacity of the damage slab in Test 19 is about 88%. A comparison of the predicted results and test results in Table 3 indicates that the numerical model can reasonably predict the loading capacity of the damaged slabs. In addition, the fragment damage in these two cases doesn't significantly reduce the loading capacity of the slabs.

Slab Response by Combined Blast and Fragment Loading

Displacement-time histories from the simulation results of Test 52 are presented in Figure 7, which indicates the response of the slab specimen under the CBFL is basically a significant rebound and followed by some oscillations. The rebound displacement was not captured during the CBFL test, which is probably the displacement gages were not set for the rebound displacement. Nevertheless, it may be considered that the predicted overall response is still reasonable when compared to the entire global response curves from test (Figure 7a) and simulation (Figure 7b).



(a) Concrete damage at failure.

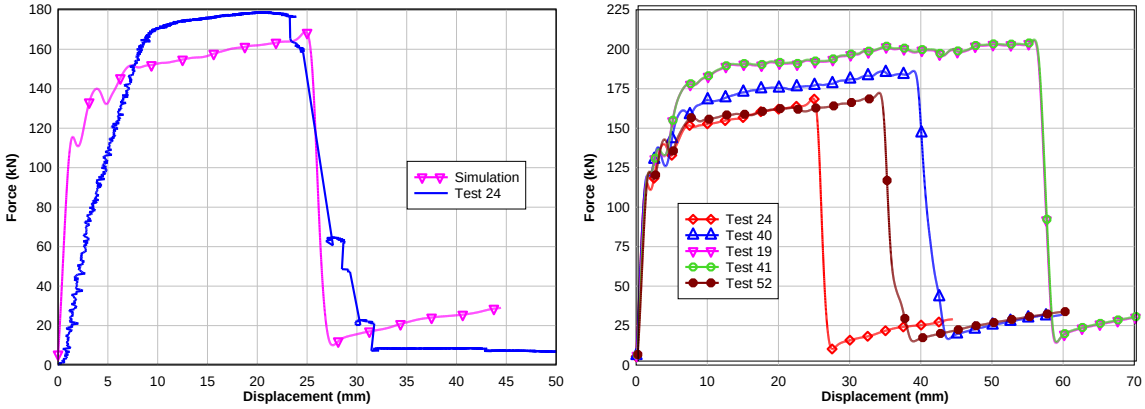


(b) Rebar fracture.

Figure 3. Simulation of quasi-static loading test for pristine slab - Test 24.

Table 3. Comparison of loading capacity from pristine and fragment damaged slabs.

Test	Pristine slab (predicted)	Damaged slab (predicted)	Damaged slab (Test)	Percentage of Residual Capacity
Test 19	205	180	172	88%
Test 41	200	182	185	91%



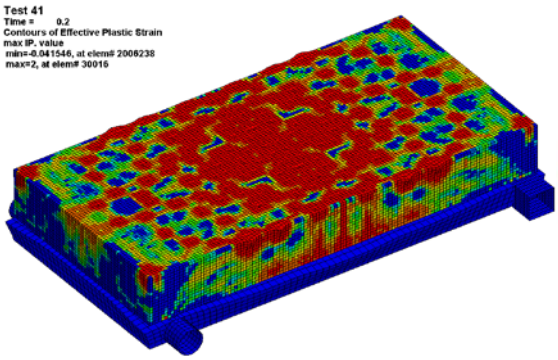
(a) Test 24 (b) All five specimens.

Figure 4. Loading capacity of pristine slabs in five tests.

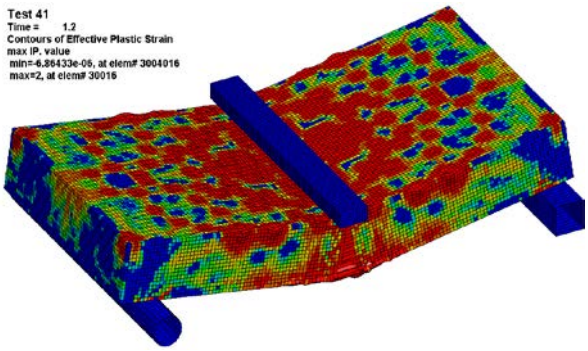
(a) Damaged by the fragment loading.

(b) Failure in quasi-static test for the damaged slab.

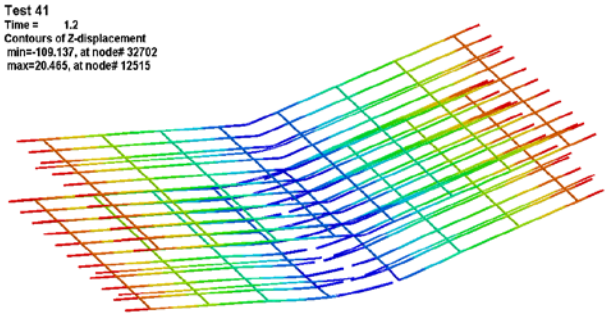
Figure 5. Test 41 – slab under fragment loading and posttest quasi-static test.



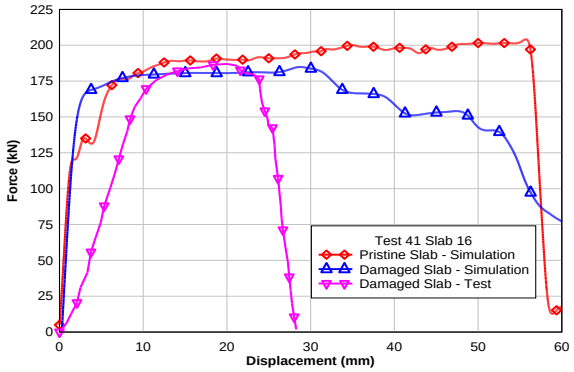
(a) Damage by fragment loading



(b) Failure model after QSL

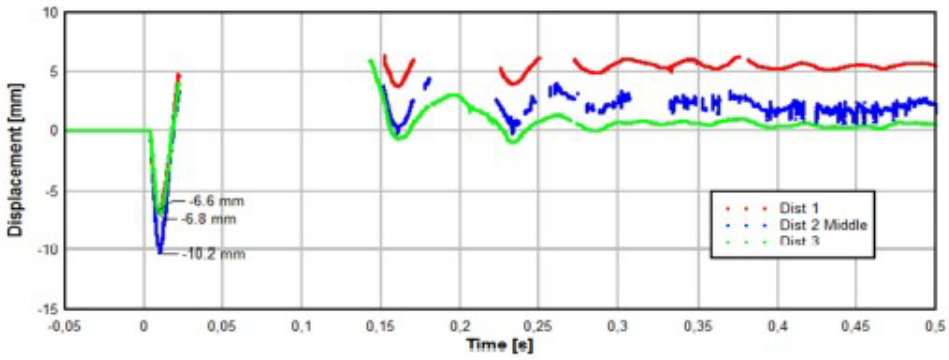


(c) Rebar fracture in QSL

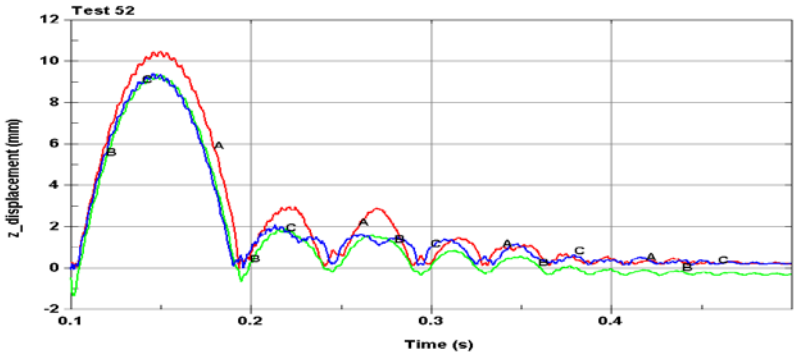


(d) Compare loading capacity

Figure 6. Simulation results for Test 41.



(a) Test measurement.



(b) Simulation

Figure 7. Test and simulation results for Test 52.

Parametric Study

Case Description

Following the successful validation of the numerical model described in the foregoing sections, this section summarizes the simulation results from the ten cases for a parametric study. The ten cases are defined in Table 4, including Cases D1 to D5 for Bomb B and Cases E1 to E5 for Bomb C at different standoff distances and different orientation angles α (Figure 8). The slabs to be analyzed are 3.0 m long by 1.6 m wide, while their thickness and reinforcement bars (rebars) are different in the two series as shown in Table 4. Concrete of the slabs is Grade C32/40 (f'_c= 32 MPa), and the rebar is Grade 500C (fy = 500 MPa).

The fragment loading from these two bombs are calculated by KC-Frag [5], which has been developed by K&C for characterizing fragment loading from a pipe bomb. The blast loading for the two cases is calculated based on the reduced charge weight, which is determined by the Fano Equation [9]:

$$\frac{W_1}{W} = 0.2 + \frac{0.8}{1 + \frac{M}{C}}$$

Where, W1/W is the ratio of the reduced charge weight to the actual charge weight;
M/C is the case to charge weight ratio.

The loading characteristics from these ten cases are summarized in Tables 5 and 6. The two key parameters are the total momentums due to air blast and fragments, which is usually dominant the damage of the RC slabs. The total momentum due to fragments is much greater than that due to air blast in all cases, which indicates that the fragment loading may produce greater damage on the slab.

The fragment loading for each fragment generated by KC-Frag is a triangle pressure pulse with a high pressure peak and a short duration based on the momentum from a fragment, which will be applied to a single element. About two thousands of loading curves calculated from the effective charge weight are generated and mapped on the slabs to mimic the "real time" blast and fragment loading as each load curve has its own arrival time, peak pressure and duration.

Table 4. Parameters of the ten cases.

Case No.	Bomb (Effective Charge)	Orientation Angle, *a*	Standoff (m)	Scaled Distance ($m/kg^{1/3}$)	Slab Dimension & Reinforcements
D1	Bomb B (46.05 kg)	0^o	5	1.395	3.0 x 1.6 x 0.6 m (clear span = 2.8m) Longitudinal rebars: 11H16 (150 mm c/c) Transverse rebars: 15H10 (200 mm c/c)
D2		10^o	5	1.395	
D3		17^o	5	1.395	
D4		0^o	7.5	2.092	
D5		0^o	10.0	2.790	
E1	Bomb C (2.29 kg)	0^o	2.8	2.124	3.0 x 1.6 x 0.25 m (clear span = 2.8m) Longitudinal rebars: 16H13 (100 mm c/c) Transverse rebars: 15H10 (200 mm c/c)
E2		10^o	2.8	2.124	
E3		17^o	2.8	2.124	
E4		0^o	5.0	3.793	
E5		0^o	7.5	5.690	

Note: Concrete is grade C32/40 (fpc f_c' = 32 MPa); Reinforcement bar is Grade 500C (f_y=500 MPa).

Table 5. Summary of blast and fragment loading for Cases D1 to D5 with Bomb B.

Description	D1	D2	D3	D4	D5
Standoff (m)	5	5	5	7.5	10
Orientation (degree)	0	10	17	0	0
Charge Explosive Weight (kg)			90		
Casing Weight (kg)			141		
Reduction Factor			0.512		
Reduced Charge Weight (kg)			46.05		
Peak AirBlast Pressure (MPa)	1.82	1.82	1.82	0.51	0.24
Peak AirBlast Impulse (MPa-msec)	1.65	1.65	1.65	1.01	0.73
Total Momentum due to AirBlast (N-sec)	7633	7633	7633	4704	3477
Total number of fragment	1184	764	298	798	602
Smallest fragment weight impacting slab (g)	0.16	0.17	0.17	0.16	016
Largest fragment weight impacting slab (g)	139	102	85	139	139
Total fragment weight impacting slab (g)	7827	5594	2289	5432	4044
Lowest fragment normal impact velocity (m/s)	868	875	850	868	868
Highest fragment normal impact velocity (m/s)	3499	3445	3078	3499	3499
Average fragment normal impact velocity (m/s)	2171	2137	2079	2184	2179
Fragment velocity per Gurney Equation (m/s)			2187		
Average fragment impact momentum (N-sec)	14.2	15.7	15.9	14.5	14.8
Total Momentum due to Fragment (N-sec)	16861	12029	4738	11571	8875

Finite Element Model

The finite element models for Cases D1 to D5 (Model D) and Cases E1 to E5 (Model E) are shown in Figures 9 and 10, respectively. In these models, 25 mm cube solid elements are employed for concrete material and 25 mm long beam elements are employed for reinforcement bars.

Fragment Loading

As an example, the blast and fragment loading in Case D1 and D2 is shown in Figures 10, which provide information about the fragment distribution on the slab and the momentum of each fragment. This figures also clearly exhibit how the orientation angle influences the fragment distribution on the slab, i.e., when the orientation angle increases, the affected area reduced from the entire top face (Figure 10c) to about two third (Figure 10d). The key parameters of the blast and fragment loading in Cases D1 to D5 summarized in Table 5 indicate that the total number of fragments and the total momentum due to the fragments are significantly reduced from Case D1 to Case D3, while the average normal impact velocities of the fragments are almost identical. Table 5 also indicates that the fragments numbers are reduced when the standoff distance is increased in Case D4 and D5 in comparison with Case D1.

The key parameters of the fragment loadings in Cases E1 to E5 are summarized in Table 6 and exhibit the same characteristics as in Case D1 to D5 mentioned in the above paragraph.

As each fragment loading is represented by a triangle pressure pulse, a lot of loading curves are generated according to the total number of fragments in Tables 5 and 6 (i.e., from 73 to 1184) and applied on the slab, where applicable.

Table 6. Summary of blast and fragment loading for Cases E1 to E5 with Bomb C.

Description	E1	E2	E3	E4	E5
Standoff (m)	2.8	2.8	2.8	5.0	7.5
Orientation (degree)	0	10	17	0	0
Charge Explosive Weight (kg)			7		
Casing Weight (kg)			37		
Reduction Factor			0.327		
Reduced Charge Weight (kg)			2.29		
Peak AirBlast Pressure (MPa)	0.49	0.49	0.49	0.12	0.057
Peak AirBlast Impulse (MPa-msec)	0.355	0.355	0.355	0.186	0.117
Total Momentum due to AirBlast (N-sec)	1568	1568	1568	886	555
Total number of fragment	431	320	73	272	194
Smallest fragment weight impacting slab (g)	0.26	0.31	0.31	0.26	0.26
Largest fragment weight impacting slab (g)	224	224	194	194	127
Total fragment weight impacting slab (g)	5471	4290	1053	3707	2589
Lowest fragment normal impact velocity (m/s)	148	145	269	148	148
Highest fragment normal impact velocity (m/s)	2587	2547	2070	2586	2586
Average fragment normal impact velocity (m/s)	1356	1355	1318	1378	1393
Fragment velocity per Gurney Equation (m/s)			1355		
Average fragment impact momentum (N-sec)	17.0	18.2	19.7	18.6	18.2
Total Momentum due to Fragment (N-sec)	7343	5820	1436	5064	3534

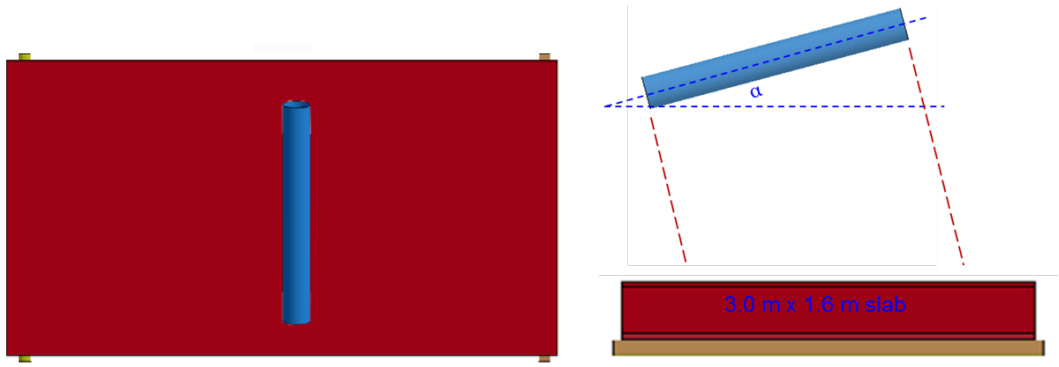


Figure 8. Model for parametric study (not to scale).

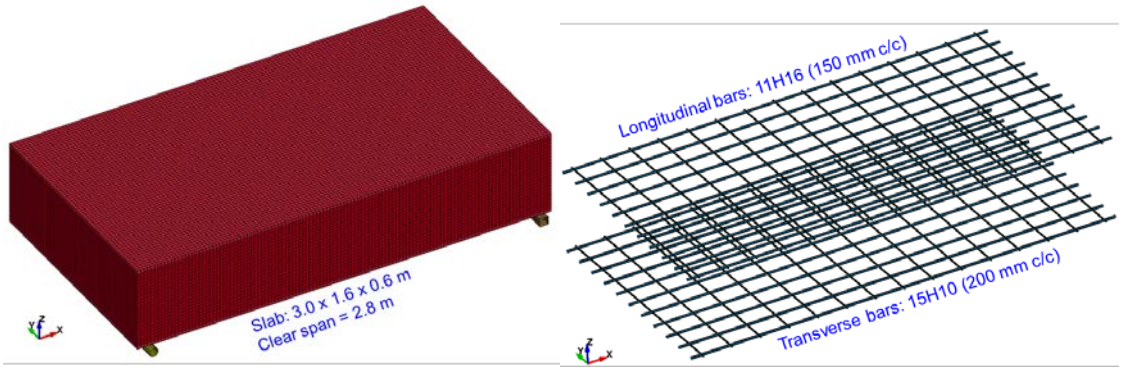


Figure 9. Finite element model for Cases D1 to D5.

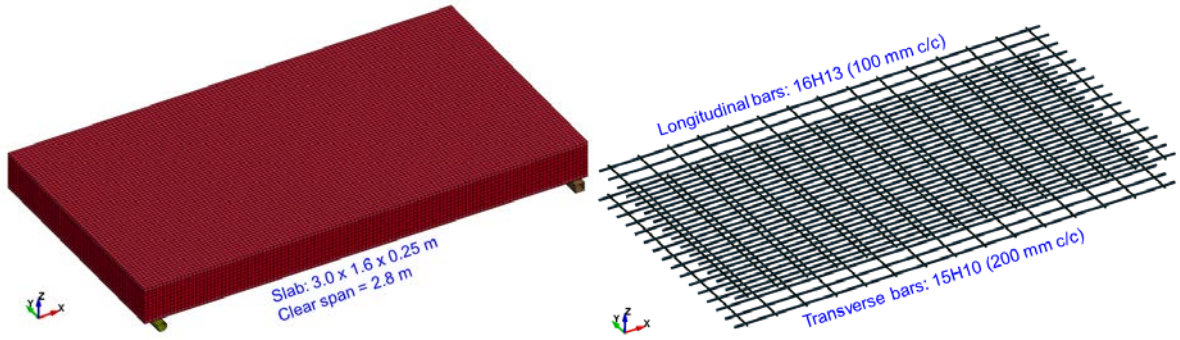


Figure 10. Finite element model for Cases E1 to E5.

Analysis Results

Analysis Results for Cases D1 to D5

The analysis results from Case D1 are presented in Figures 11. The analyses results exhibit that the fragment loading dominates the slab damage, e.g., the entire top face of the slab is damaged by the fragments in Case D1, which results in only 39% of residual capacity of the damaged slab (Table 7). In Case D2, the fragments hit only about two third of the top face and only this area is badly damaged (Figure 12), which results in 42% of residual capacity. In Case D3, nearly one third of the top face is badly damaged by fragments and the damaged slab remains 95% residual capacity. From the QSL simulations for the damaged slabs (Figures 11 and 12), all slabs lose their loading capacities due to the concrete shear failure without rebar fracture in Cases D1 to D3.

When the standoff distance is increased in Cases D4 and D5 compared to Case D1, both blast and fragment momentums decrease significantly (Table 5). Consequently, the concrete damage is less severe and the slab residual capacity in these two cases are increased significantly, i.e., 75% in Case D4 and 83% in Case D5 (Table 7).

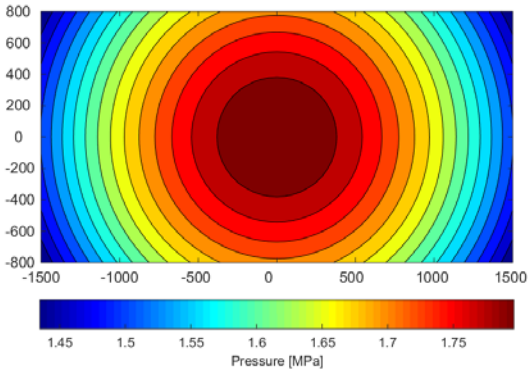
The loading capacities of the pristine and damaged slabs shown in Figure 13 indicate that the slab residual capacity in Case D1 and D2 is less than a half of the pristine capacity and the residual capacity in Case D3 has no significant reduction. In Cases D4 and D5, substantial residual capacities still exist. Furthermore, the loading capacity of the damaged slab by blast loading only has almost no reduction compared to the pristine slab.

Table 7 summarizes the loading capacities, the peak dynamic displacements and corresponding support rotation angle [4] in Cases D1 to D5. A relationship between the support rotation and residual capacity is plotted in Figure 13, which indicates that the residual capacity of the damaged slabs can be significantly reduced (less than 50%) when the support rotation is greater than 0.7 degree.

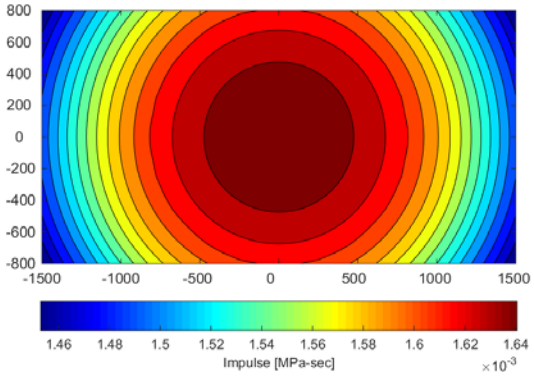
Table 7. Loading capacities of pristine and damaged slabs for Case D1 to D5.

	Loading Capacity (kN)						
	Pristine	Blast	D1	D2	D3	D4	D5
Value	1800	1800	710	763	1710	1346	1497
percentage	100%	100%	39%	42%	95%	75%	83%
D_{max} (mm)	-	1.4	24	19	6	7.6	5.2
θ_{max}		0.05^0	0.98^0	0.78^0	0.25^0	0.31^0	0.21^0

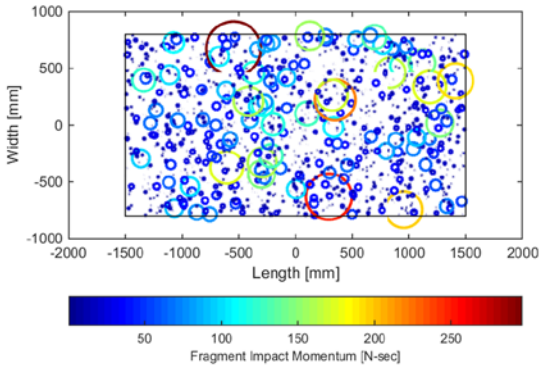
Note: D_{max} = the peak dynamic displacement at the slab center.
θ_{max} = the maximum support rotation angle (degree).



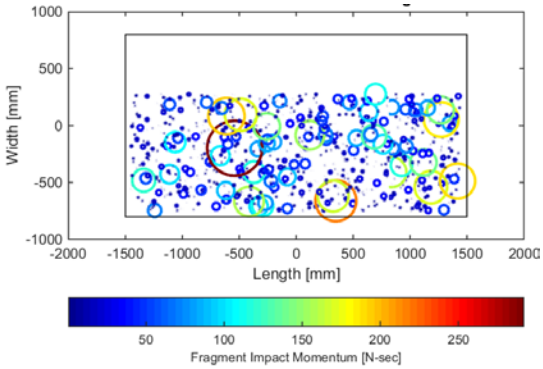
(a) Peak pressure.



(b) Maximum Impulse.

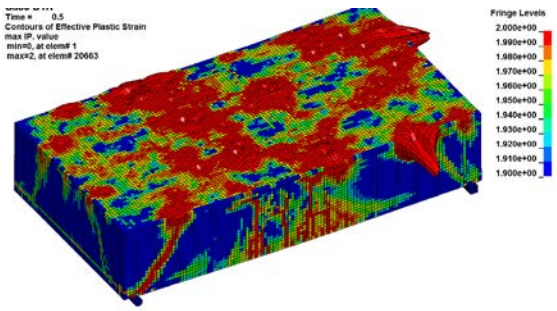


(c) Fragment Loading in Case D1.

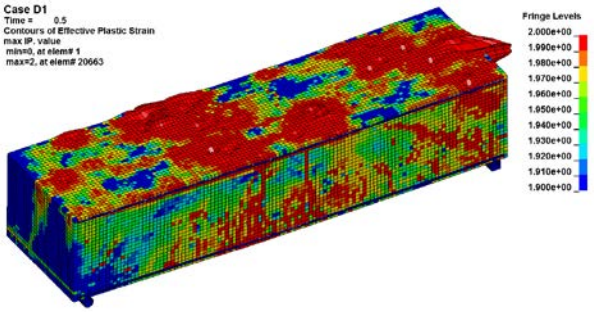


(d) Fragment Loading in Case D2.

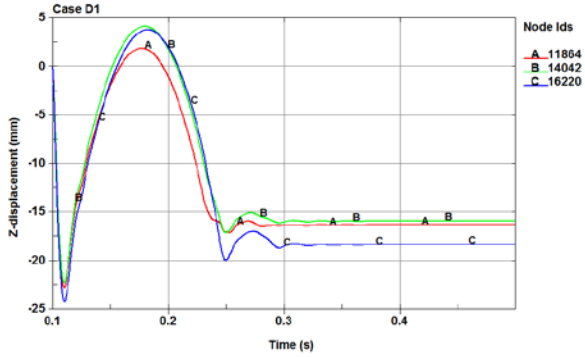
Figure 11. Blast and fragment loading distribution in Cases D1 and D2.



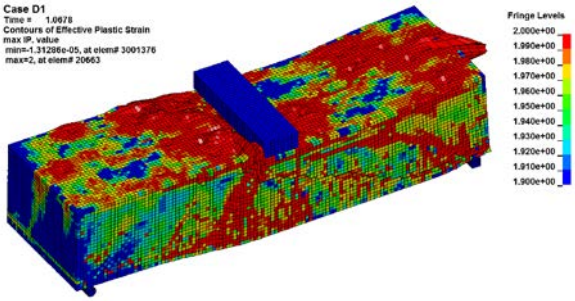
(a) Damage by CBFL



(b) Damage by CBFL (view through the middle section).

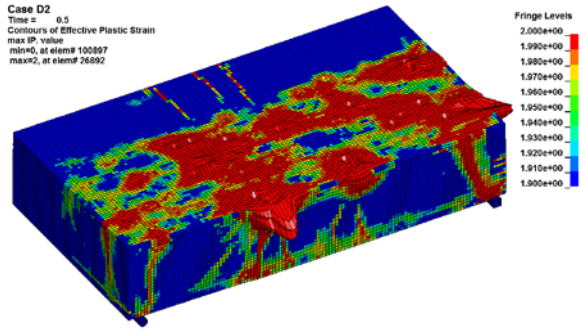


(c) Displacement history (B – middle span; A and C – at quarter spans on left and right, respectively).

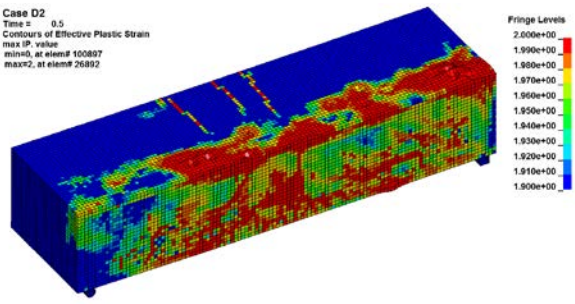


(d) Failure of the damaged slab in posttest QSL (view through the middle section).

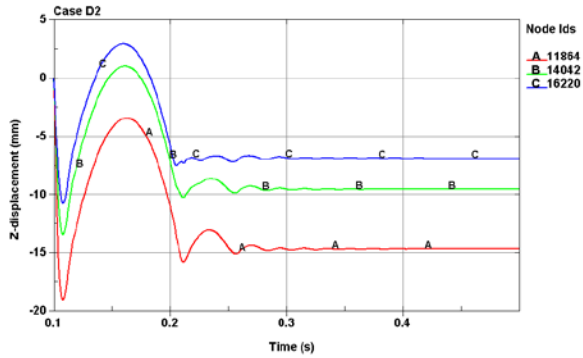
Figure 12. Case D1: slab response under CBFL (Continued).



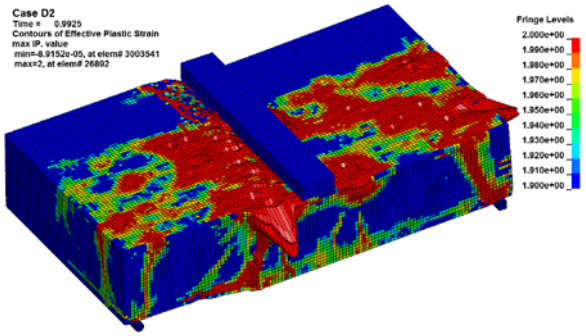
(a) Damage by CBFL



(b) Damage by CBFL (view through the middle section).

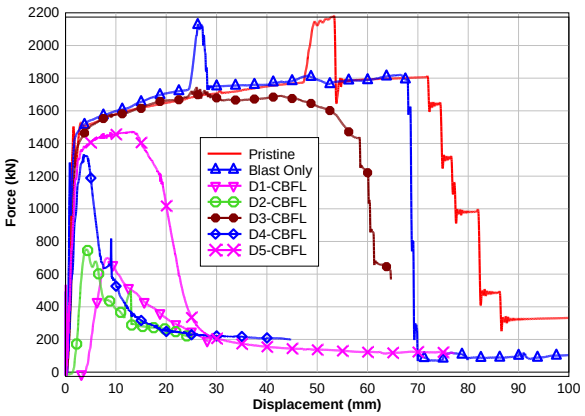


(c) Displacement history (B – middle span; A and C – at quarter spans on left and right, respectively).

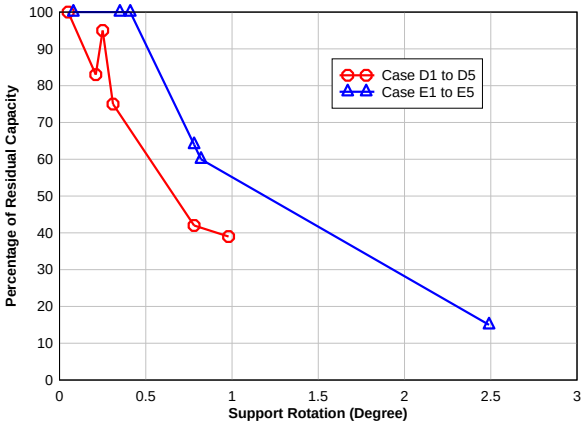


(d) Failure of the damaged slab in posttest QSL (view through the middle section).

Figure 13. Case D2: slab response under CBFL.



(a) Loading Capacity



(b) Support rotation vs. residual capacity.

Figure 14. Loading capacities of pristine and damage slabs in Cases D1 to D5.

Analysis Results for Cases E1 to E5

As an example, the blast and fragment loadings on the slab is shown in Figure 14 and the analysis results from Cases E1 are presented in Figures 15. In Case E1, the slab damage is in the middle along the longitudinal span as the fragment loading distribution from Bomb B (Figure 14b). However, the right side of the slab (Figures 15a) undergoes severer damage due to larger fragment momentum in this area as shown in Figure 4-14b. Damage patterns from other cases are not shown here.

The analysis results summarized in Table 8 indicate that the residual capacities of the damaged slabs in Cases E1, E2 and E4 are 15%, 60% and 64%, respectively, when compared

to the pristine slab. The damaged slabs in Cases E3 and E5 and by blast loading only have almost the same loading capacity with the pristine slab. The relationship between the support rotation and the residual capacity in Figure 16 indicates that the slabs lose about 50% loading capacity when the support rotation is greater than 1.2 degree.

When the standoff distances are increased in Case E4 and E5, the global damage to the slab is less significant compared with that in Case E1, although a fragment near the slab edge may cause severe local damage (Figure 16).

The loading capacities of the pristine and damaged slabs are evaluated and their load displacement curves are presented in Figure 17. Similar with Cases D1 to D3, from the failure model of the damaged slab in posttest QSL, all slabs lose the loading capacity due to concrete shear failure and no reinforcement bars fracture.

Table 8. Loading capacities of pristine and damaged slabs for Case E1 to E5.

	Loading Capacity (kN)						
	Pristine	Blast	E1	E2	E3	E4	E5
Value	460	460	70	273	458	295	470
Percentage	100%	100%	15%	60%	100%	64%	100%
D_{max} (mm)	-	2	61	20	8.5	19	10
θ_{max}		0.08^0	2.49^0	0.82^0	0.35^0	0.78^0	0.41^0

Note: D_{max} = the peak dynamic displacement at the slab center.
θ_{max} = the maximum support rotation angle (degree).

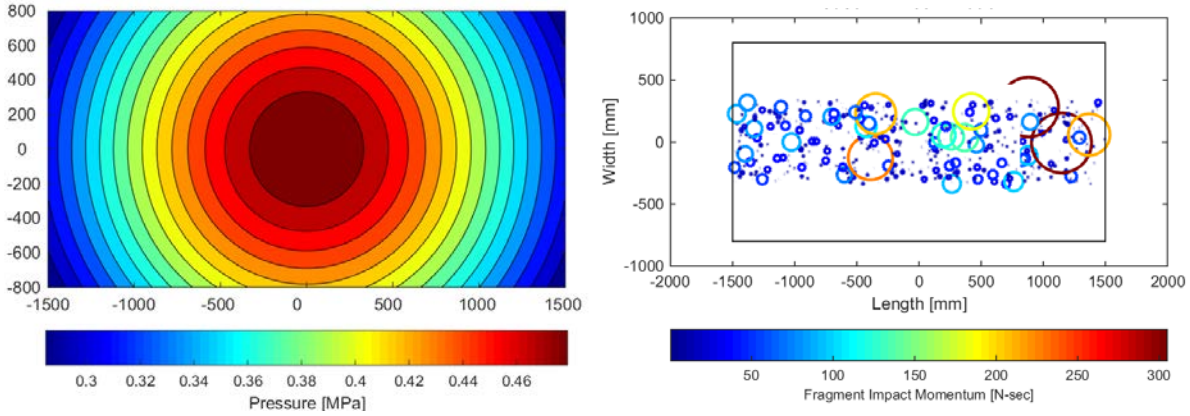


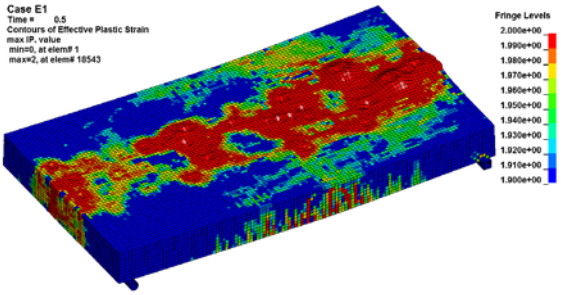
Figure 15. Case E1: Blast and fragment loading on slab.

Conclusions

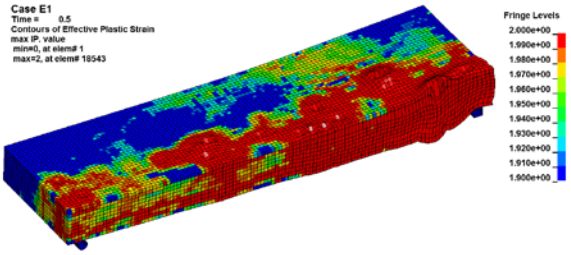
In this study, the HFPB finite element techniques and procedures for evaluating the residual capacities of RC slabs after the combined blast and fragment loading effects have been validated. A parameter study has been conducted to evaluate the residual capacity of the RC slabs subjected to various blast and fragment loadings. The calculated loading shows the total momentum from fragments can be greater than that from air blast loading generated by pipe bombs. The simulation results show that the fragment loading can dominate the damage of the RC slabs and their residual capacities.

Acknowledgement

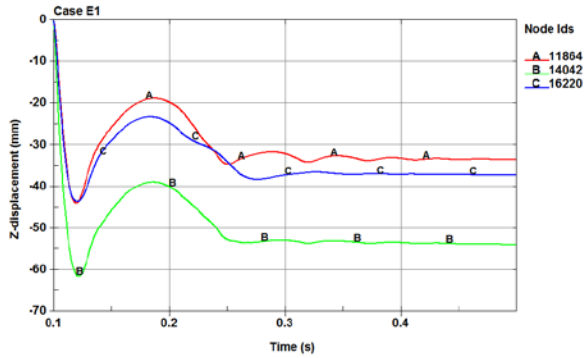
The authors acknowledge the financial and technical support by Defence Science & Technology Agency (DSTA) and Nanyang Technological University (NTU).



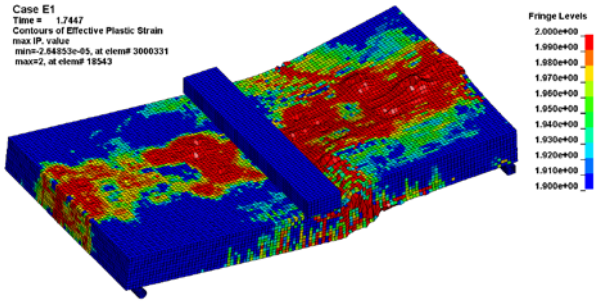
(a) Damage by CBFL



(b) Damage by CBFL (view through the middle section).



(c) Displacement history (B – middle span; A and C – at quarter spans on left and right, respectively).



(d) Failure of the damaged slab in posttest QSL (view through the middle section).

Figure 16. Case E1: slab response under CBFL.

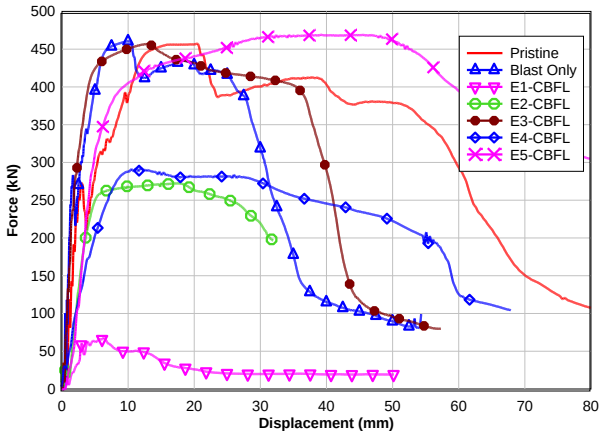


Figure 17. Loading capacities of pristine and damage slabs in Cases E1 to E5.

References

[1] Karpenko, A. and Ceh, M. (2007, April), Experimental simulation of fragmentation effects of an improvised explosive device. 23rd International Symposium on Ballistics, 1413-1420.

[2] Nordstrom, M. and Forsen, R. Damage to reinforced concrete slabs due to fragment loading with different fragment velocities, fragment areal densities and sizes of fragments. National Defense Research, FOA Report S-172-90, Sundyberg, Sweden.

[3] Forsen, R., & Nordstrom, M. (1992, September). Damage to reinforced concrete slabs due to the combination of blast and fragment loading. FOA Report B 20101-2.6, Tumba, Sweden.

[4] Forsén, R., Response to RC Slabs Subjected to Combined Blast and Fragment Loading, Fifth Internal conference on Design and Analysis of Protective Structures, May 2015, Singapore

[5] Crawford, J. E., Fu, S., Magallanes, J. M., and Zhang, Y. "Determining the Effects of Cased Explosives on the Response of RC Columns" for the MABS21, Jerusalem, Israel, October 3-8, 2010.

[6] Joseph M. Magallanes, Youcai Wu, Shengrui Lan, and John E. Crawford, “Parameters Influencing Finite Element Results for Concrete Structures”, ACI SP-306-1, 2016.
[7] Malvar, L. J., Crawford, J. E., Wesevich, J. W., and Simons, D. “A Plasticity Concrete Material Model for DYNA3D”, Int. J. Impact Engng, Vol. 19, Nos. 9-10. pp. 847-873, 1997.
[8] Lan, S., Lim, H. S., Ow, M. C., and Morrill, K. B. "Reinforced Concrete Slab Under Combined Blast and Fragment Loading," for the 16th International symposium on the Interaction of the Effects on Munitions with Structures, Florida, November 9-13, 2015
[9] Fisher E.M., “The Effects of the Steel Case on the Air Blast from High Explosives”, U.S. Nabal Ordnance Laboratory, White Oak, Maryland, February 1953.
[10] U.S. Department of Defense UFC 3-340-02 - "Structures to Resist the Effects of Accidental Explosions", Dec. 2008. (Previously Known as TM5-1300).

EXPLICIT MODELLING OF FIBRE PULLOUT IN CEMENTITIOUS COMPOSITES

***†Hui Zhang[1], Rena C. Yu[2], Shilang Xu[1]**

[1]College of Civil Engineering and Architecture, Zhejiang University, 310058, China.
[2]School of Civil Engineering, University of Castilla-La Mancha, Ciudad Real, Spain

*Presenting author: huizhangzju@zju.edu.cn
†Corresponding author: huizhangzju@zju.edu.cn

Abstract

It is well-known that fibres improve the performance of cementitious composites by acting as bridging ligaments in cracks. Such bridging behaviour is often studied through the fibre pullout tests. The relation between the pullout force versus slip end displacement is characteristic of the fibre-matrix interface. However, such a relation varies significantly with the fibre inclination angle. In the current work, we establish a numerical model to explicitly represent the fibre, matrix and the interface for arbitrary fibre orientations. Cohesive elements endorsed with mixed-mode fracture capacities are implemented to represent the bond-slip behaviour at the interface. Contact elements with Coulomb's friction are placed at the interface to simulate frictional contact. Matrix spalling is modelled through material erosion. The bond-slip behaviour is first calibrated through pull-out curves for fibres aligned with loading direction, then validated against experimental results carried out by Leung and Shapiro in 1999 for steel fibres oriented at 30° and 60°. The proposed methodology provides the necessary pull-out curves for a fibre oriented at a given angle for multi-scale models to study fracture in fibre-reinforced cementitious materials.

Keywords: Fibre-reinforced concrete, Pullout response, Cohesive model, Matrix spalling,

Introduction

Cementitious materials, known as a quasibrittle, have almost no ductility, additionally, have very low tensile strength. The addition of fibers in a cement-based matrix enables a considerable amount of energy dissipated during structural cracking.

The effectiveness of force transmission of certain fibres is often assessed by a pullout test, in which the force required to pull a fibre out of the hardened concrete is measured. This force is derived from interfacial bond, defined as the shear stress at the interface between the fibre and the surrounding matrix [1]–[4].

Classical approaches assume that the perfect bond on the interface between fibre and matrix will be maintained unless a failure criterion is achieved [5]. Stress criterion [6][7] or energy criterion [5][8][9] have been adopted, as well as cohesive approaches where bond stress is determined by relative slip between fibre and matrix [2][5][10][11]. Naaman et al. [2] indicated that for aligned fibres whose load direction is along the fibre direction, there are two types of shear bond at the interface: the elastic shear bond and the frictional one. If the elastic one exceeds the bond strength of the interface, bond becomes frictional in nature.

Based on the interfacial properties, Chanvillard [12] took into consideration the different phenomena existed in a non-straight fibre with a new micro-mechanical model. Ellis [13]

carried out a simulation with emphasis of fibre morphology. Despite of the good agreement, only aligned fibres were considered in these two models.

In fibre reinforced cementitious material, most fibres lie at an angle to the load direction. For inclined fibres, besides bond strength and friction along the interface, additional phenomena such as fibre bending, matrix spalling and local friction effects need to be considered [14]–[17]. Furthermore, these micro-mechanisms are sensitive to fibre inclination angle and fibre material properties [3][9][14][15][16][18][19]. Fibre bending contributes increasingly more for larger inclination angles and the fibre curvature has an impact on the pressure distribution against the surrounding matrix. Because of the fibre curvature and residual stress at the interface, matrix is likely to crack and spall [14][16][19], which in turn influences the effective embedment length and deformation within the fibre.

A great deal of efforts have been put to study the above phenomena in the pullout process of inclined fibres. For example, Mortons and Groves [20] calculated the force needed to produce a plastic hinge in the fibre based on an elementary beam theory. The model reproduced well the experimental observations for lower inclination angles, but failed to do so for steep inclinations due to the fact that matrix spalling was not accounted for. Regarding the fibre as a beam bent on an elastic foundation with variable stiffness, Leung and Li [9] studied the coupled fibre bending-matrix spalling mechanism in random brittle fibre-reinforced brittle matrix composites. Afterwards, the micro-mechanical model was extended to ductile fibres [21]. However, the whole pullout curve was not simulated.

More comprehensive models such as Cailleux et al. [19] and Fantilli et al. [11], require a number of parameters which can only be obtained through pullout experiments in aligned as well as inclined fibres, and the numerical iterative procedures involved are tedious.

In spite of continuous efforts during the last decades, models that cover all the aforementioned phenomena, however, have been seldom developed to explicitly consider the whole pullout process for fibres at a random inclination angle. In this study, we endeavour to do so. Cohesive models able to represent mixed-mode fracture and Coulomb's friction at the interface between fibre and matrix are employed. In addition, fibre bending and matrix spalling are naturally taken into account owing to the explicit representation of the fibre, the matrix and the interface in between.

The paper is structured as follows. Section 2 describes the interfacial bond characteristics and matrix spalling. Afterwards, model calibration and validation are given in Section 3. The numerical results and relevant conclusions are respectively presented in Section 4 and Section 5.

Bond characterisation and matrix spalling

In this section, two major factors that determine the pullout response of a fibre with arbitrary orientation are explained in detail: bond characterisation and quantification of matrix spalling.

Interface bond characterisation

As a constitutive property of the interface, the shear stress versus slip relationship is very important for predicting both the mechanical and fracture properties of fibre reinforced composites. Naaman et al. [1] ascribed the presence and combination of four bond components: physical and chemical adhesion, the mechanical contribution of deformed or

hooked fibres, the entanglement of fibres and friction. In this work, the concept of internal friction is illustrated and quantified through fitting with the test data of Leung and Shapiro [16].

Regardless of the fibre orientation, after debonding during pulling out, there exists certain frictional resistance which is mainly determined by the surface roughness. This component is denominated as internal fraction resis tance, τ_0. As for inclined fibres, pullout load is decomposed of a parallel force and a perpendicular force. The former pulls the fibre out while the latter bends the fibre and changes the direction of fibre during the pullout process. A constitutive law involving three constituents to govern the interface evolution is proposed as follows

$$\tau(\theta, s) = \tau_f(s) + \tau_b(s) + \mu p(\theta) \tag{1}$$

where μ is the friction coefficient of Coulomb, $p(\theta)$ is the pressure against the matrix when the fibre is inclined at an angle θ with respect to the external load direction. The first term, $\tau_f(s)$, is contributed by the internal friction, acting as a resistance between the fibre and the matrix. The second term, $\tau_b(s)$, represents the interfacial bonding caused by internal physical and chemical cohesion. The third term describes the shear stress due to dry friction, which works only when the fibre is oriented at a non-zero angle.

In the case of constant friction, see Fig. 1, $\tau_f(s)$ is unchanging with the slip displacement s. In the current work, both $\tau_f(s)$ and $\tau_b(s)$ are decaying functions of s and they are assumed as linear-decreasing:

$$\tau_f(s) = \tau_0 \left(1 - \frac{s}{s_0}\right) \tag{2}$$

$$\tau_b(s) = (\tau_{\max} - \tau_0)\left(1 - \frac{s}{s_c}\right) \tag{3}$$

where τ_0 represents the internal frictional resistance, τ_{max} is the bond strength, defined as the maximum shear stress resisted at the interface, covering both the internal bond and internal friction.

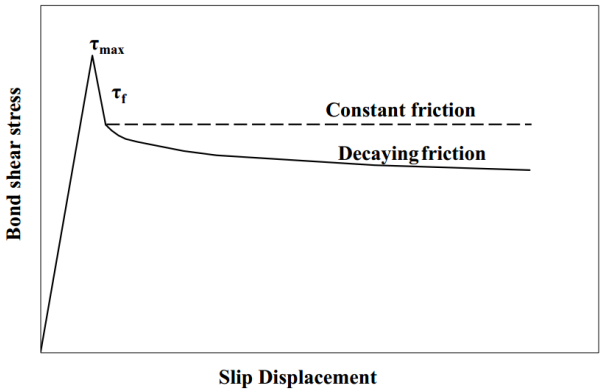


Figure 1. Interface bonding.

Eqs. (1-3) presents a gradual failure at the interface which is of vital importance, since in most actual fibre applications, the slip displacement is less than 1 mm, see Yu et al. [22]. When

working in a corrosive environment, the maximum allowed crack opening in steel-fibre reinforced composites is 0.3 mm making it paramount to trait the detailed failure process at small slips.

Matrix spalling

The local curvature and stretching of the fibre segment at the free end will inevitably lead to spalling of matrix. Because the failure process starts near the interface within a narrow band, convergence problems led by excessive mesh distortion impedes the further modelling of the entire pullout process. As a result, simplifications are often assumed so that the spalling part is removed once the matrix tensile strength is reached [17][19][23][24]. The length of the eroded matrix along the fibre direction is the so-called spalling length, denoted as L_{sp}. The spalling length of the matrix can be estimated according to Laranjeira et al. [23] as follows:

$$aL_{sp}^2 + bL_{sp} + c = 0 \tag{4}$$

where

$$a = \frac{\sqrt{2}}{\sin\theta} + \frac{\cos\theta}{\sin^2\theta}, b = \frac{d_f}{\sin\theta}, c = -\frac{P_{\max}\sin\theta}{f_t} \tag{5}$$

in which, d_f is the fibre diameter, θ is the inclination angle, P_{max} is the peak pullout load of the aligned fibre. The results from Eq. (4) closely match the ones measured by scanning electron microscopy (SEM) by Leung and Shapiro [16].

In this work, Eq. (4) serves as a first approximation to obtain the size of the matrix wedge to be spalled off. Then trial runs are conducted to determine the moment to deactivate the matrix wedge. Then the elements within the matrix wedge stop to contribute to the overall stiffness and state variables. Furthermore, the first principal stress within the matrix is checked and the spalled length is adjusted if necessary.

Model calibration and validation

To explicitly model the physical phenomena in the pullout process, matrix is represented as solid elements with elastic constitutive law while fibre is by solid elements with bilinear kinematic plastic constitutive law. Cohesive elements representing mixed-mode fracture [25]-[29] are employed as the interface in between. Furthermore, contact pairs coincident with the cohesive elements are implemented to model friction after de-bonding.

Experimental setup of Leung and Shapiro,1999

To assess the effect of fibre yield strength on the maximum crack bridging force and total energy absorption, Leung and Shapiro [16] performed pullout tests for steel fibres of different yield strengths. All the fibres are of 0.5 mm in diameter and 22 mm in length. The pullout specimens are blocks of 25.4 mm× 12.7 mm × 9.5 mm in dimension, with effective embedment length of 10 mm. The material parameters of the matrix, fibre and the interface are given in Table 1, whereas the yield and tensile strengths of the four fibre types are listed in Table 2. Additionally listed in Table 2 is the critical fibre length, the maximum embedded length for a fibre to be pulled out from a matrix without rupture [14]. It is related with the maximum shear stress τ_{max} as follows

$$L_c = \frac{\pi d_f^2 / 4 f_y}{\pi d_f \tau_{\max}} = \frac{d_f f_y}{4\tau_{\max}} \tag{6}$$

Note that this estimation is for aligned fibres only, in the case of inclined ones, this length is smaller due to fibre bending.

Table 1. Material parameters for the matrix and the fibre given in [16].

	ρ [kg/m^3]	E [GPa]	υ -	f_c [MPa]
Matrix	2100	30	0.20	36.5 ± 2.5
Steel fibre	7800	200	0.33	-

Table 2. Yield and tensile strength of the four types of fibres tested in [16], the corresponding L_c is listed for a diameter of 0.5 mm.

Fibre type	1	2	3	4
f_y [MPa]	275	469	635	954
f_t [MPa]	-	783	847	1023
L_c [mm]	12.7	21.7	29.4	44.2

Identification of the fibre-matrix interface properties

With the assumption that fibre-matrix interface property is uniform, the peak pullout load and maximum fictional load are respectively calculated as

$$P_{\max} = \pi d_f L_e \tau_{\max}, P_f = \pi d_f L_e \tau_0 \tag{7}$$

The values for P_{max} and P_f are determined from pullout response of aligned fibres, as shown in Fig. 2 and Fig. 3. The critical slip displacement for internal frictional resistance, s_0, is directly assumed as the final slip length, 9.0 mm approximately. The straight dotted line in Fig. 2a starts at the point (s_0, 0), follows the mean slope of the experimental curves and intercepts the load axis at (0, P_f). The values for P_{max} and P_f are averaged for the four types of fibres listed in Table 2 to obtain those of τ_{max} and τ_0 as well as their standard deviations in Table 3. The critical slip for interfacial bond, s_c, is determined through trial and error so that the first decaying branch of the numerical pullout responses, as demonstrated in Fig. 2b, should fall within the experimental range. As regards the friction coefficient given in Table 3, it is estimated according to the experimental results of Chanvillard [12], which was also adopted by Laranjeira et al. [17].

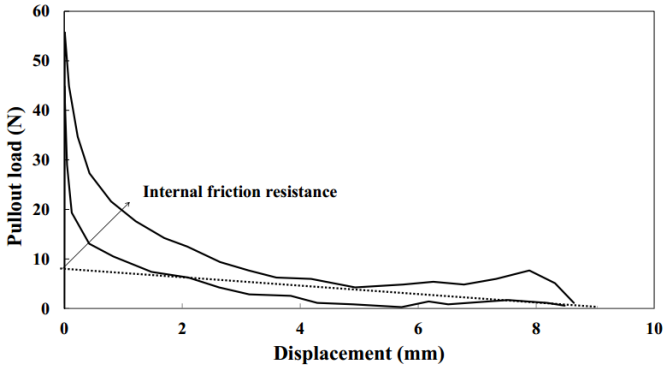


Figure 2. Experimental range for pullout curves for aligned fibre type 2 (yield strength 469 MPa) [16], where the dotted line represents the contribution from internal friction.

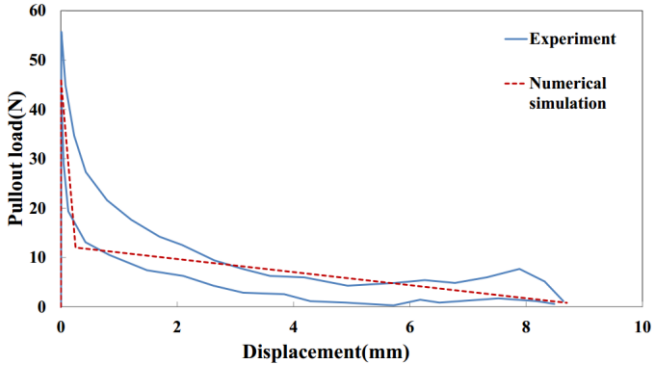


Figure 3. The corresponding numerical pullout curve plotted against the experimental range for aligned fibre type 2 (yield strength 469 MPa) [16].

Table 3. Extracted parameters of the fibre-matrix interface from the experimental data of Leung and Shapiro [16].

$\tau_{\max}$	τ_0	s_0	s_c	μ
[MPa]	[MPa]	[mm]	[mm]	-
2.7 ± 0.1	-	783	847	1023

Numerical model

Fig. 4 illustrates the in-plane dimensions and boundary conditions to simulate the pullout tests performed by Leung and Shapiro [16]. Note that within a two-dimensional plain stress framework, the fibre thickness, T_f, is calculated through Eq. (8) so that the contact area at the interface is the same as that of the tested one.

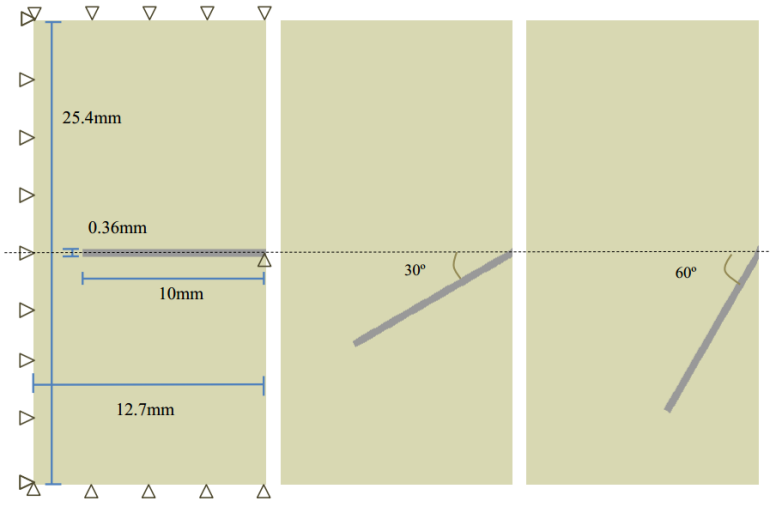


Figure 4. In-plane dimensions and boundary conditions for the pullout tests performed by Leung and Shapiro [16], with fibre inclination angle of 0°, 30° and 60°.

Similiarly, the fibre height, H_f, is determined via Eq. (9) so that the second moment of inertia is the same as the original fibre. For the case of fibre diameter of 0.5 mm, T_f and H_f are computed as 0.785 and 0.36 mm respectively.

$$T_f = \frac{\pi d_f L_f}{2L_f} \tag{8}$$

$$H_f = \left(\frac{\pi d_f^4 / 64}{T_f / 12} \right)^{1/3} \tag{9}$$

Boundary conditions are described in Fig. 4. Vertical displacements are prevented on the top and bottom sides, whereas horizontal movements are impeded on the left. The right end of the fibre is fixed in the vertical direction so that only horizontal movement is permitted. The pulling process is carried out with intervals of 0.001 mm in the horizontal direction until 0.3 mm, followed with increments of 0.01 mm until the end.

Mesh description

A typical mesh and detailed element distribution around the fibre for the inclination angle of 30° is demonstrated in Fig. 5. Note that the right end of fibre leans on a matrix wedge which will spall later on. For this particular case, the matrix and the fibre consist of 2198 and 154 solid elements respectively, whereas 289 contact pairs are placed at the interface. The mesh sensitivity analysis performed to achieve a balance between the computational efficiency and accuracy is going to be presented in Section 4.

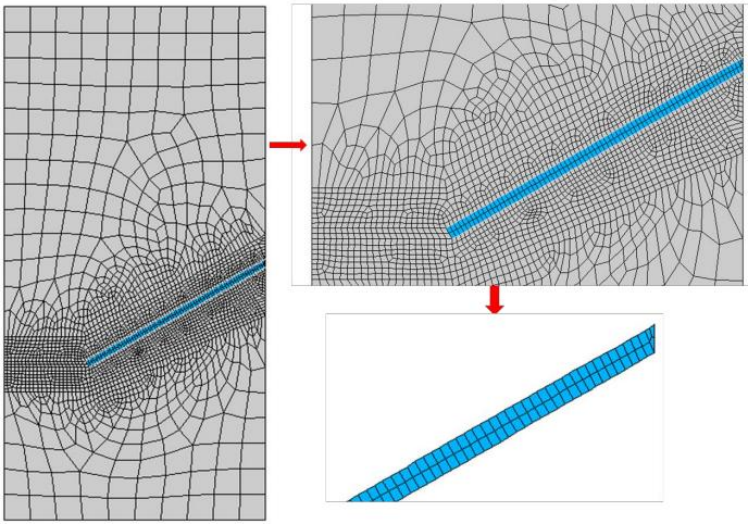


Figure 5. Typical mesh (left), zoomed in around the fibre (top right) and discretisation of the fibre (bottom right).

Numerical results and discussion

In this section, we first conduct the mesh sensitivity analysis along the fibre transverse and longitudinal directions to determine the particular mesh to employ for further studies. Second, the entire pullout load vs slip displacement curves are extracted to compare with those obtained experimentally by Leung and Shapiro [16]. Third, the von Mises stress and the first principle stress evolutions are explored both for the fibre and the matrix. Finally, the pullout work is obtained.

Mesh sensitivity analysis

The mesh-sensitivity analysis is carried out for the inclination angle of 30° and the yield strength of 635 MPa type 3 in Table 2). Two kinds of mesh sensitivities are studied: the refinement in the transverse direction and along axial direction of the fibre. The former is to check the capacity of the mesh in the fibre to bear bending moments, whereas the latter is to assess if the discretisation is fine enough to resolve the slip length.

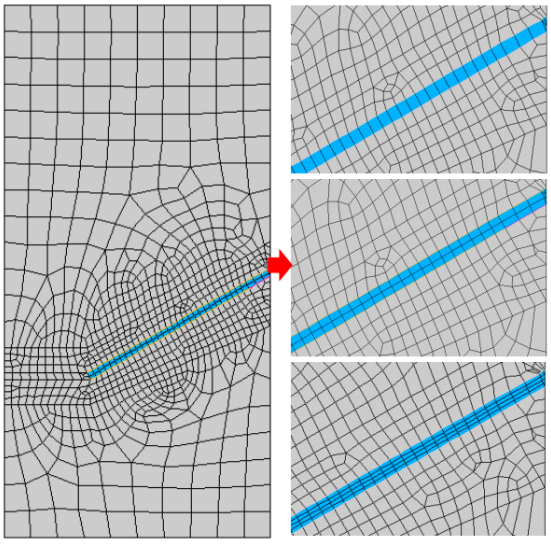


Figure 6. Different number of divisions in fibre transverse direction (one, two and four).

In the transverse direction, the fibre is split into one, two or four divisions, see Fig. 6, the corresponding load-displacement curves are plotted in Fig. 7. Meshes of two divisions across the transverse direction are employed for further studies. Along the longitudinal direction of the fibre, four different element sizes are considered: 0.303 mm, 0.222 mm, 0.135 mm and 0.068 mm, which lead to 33, 45, 74 and 148 divisions along the 10-mm length, see Fig. 8, the corresponding pullout load versus slip end displacement curves are depicted in Fig. 9. In order to keep a balance between the computational efficiency and accuracy of sought results, the mesh size of 0.135 mm is selected for further studies.

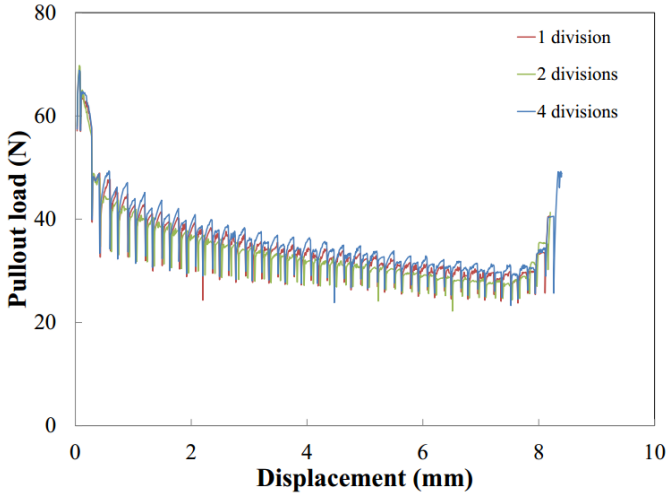


Figure 7. The pullout load vs displacement responses corresponding to different number of divisions in fibre transverse direction (one, two and four) .

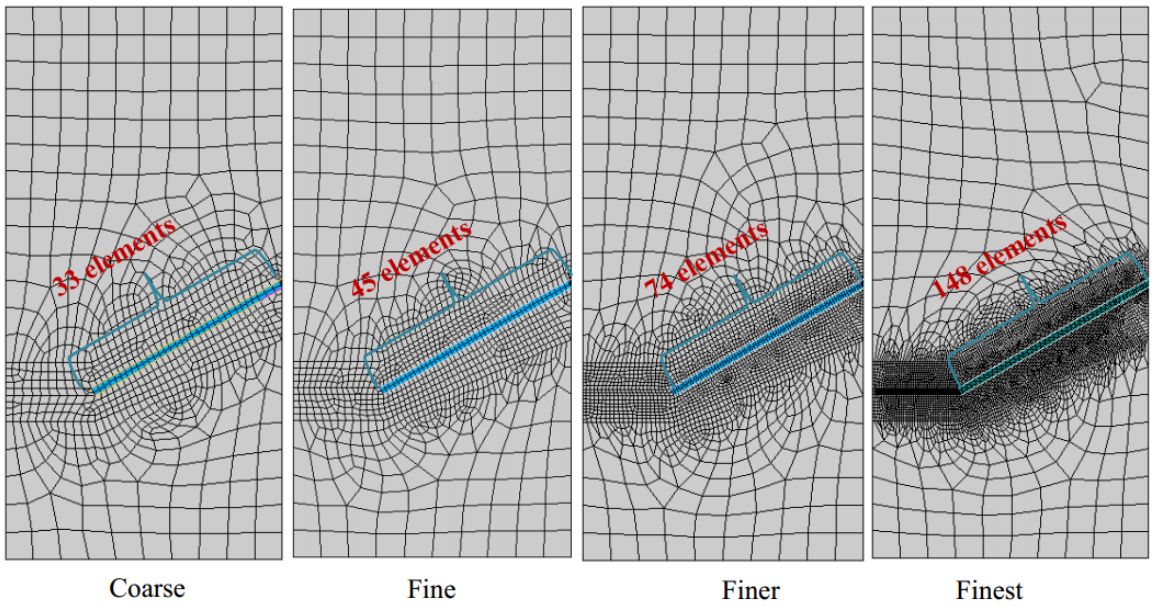


Figure 8. Different element sizes (number of divisions) along the fibre.

In addition, From Fig.9, it is observed that the maximum pullout load was achieved at slip displacement of 0.07 mm, this verifies the statement of Morton and Groves [20], who claimed that this value should be of the order of, but less than half a fibre diameter.

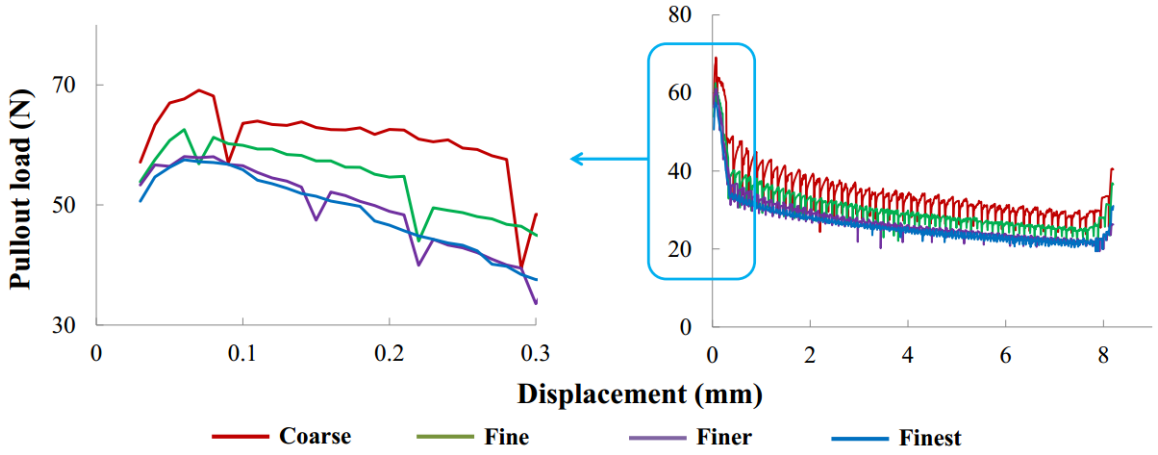


Figure 9. The pullout load-displacement responses corresponding to different element sizes (number of divisions) along the fibre.

Validation against experimental pullout load vs displacement response

In order to verify the previously developed methodology, we compare the entire pullout curves with their experimental counterparts given by Leung and Shapiro [16].
This comparison is displayed in Fig. 10 for fibres inclined at 30° and 60° with four different yield strengths given in Table 2. Note that both the peak loads and the general tendency are well captured, the numerical curves fall within the experimental range, in particular the rising tail at the end of each pullout process is also reproduced.

Stress evolution within the fibre

Taking fibre type 2 (yield strength 469 MPa) with inclination angle of 30° as an example, the von Mises stress evolution for several characteristic points within the fibre are examined. These points are the pullout end A, the embedded end D, two intermediate ones B (location of

matrix spalling) and C, as depicted in Fig. 11. Note that at point A, the first peak stress was obtained when the pullout load reached its maximum due to interface debonding. Then after a slight decrease, this stress increased again until yielding at the slip end displacement of 3 mm. Similar peaks are observed for B and C at slip displacement of 0.3 mm and a second peak upon yielding at 2 mm for point B and 5 mm for point C respectively. The second peak is attributed to the stress concentration due to the cusp formed by matrix spalling. This is the snubbing effect introduced by Li et al. [14].

In Fig. 12 and Fig. 13, the first principal stress distributions in the fibre at different loading stages are plotted for type-2 fibre inclined at 30° and type-3 fibre inclined at 60° respectively. Note that during the pullout process, there are stress gradients both in the transversal direction and along the longitudinal one, which indicates bending contribution. Stress concentration is also observed at the fibre exit point. In addition, the maximum tensile stress is always inferior to the fibre tensile strength. This means that the fibre was pulled out but not broken.

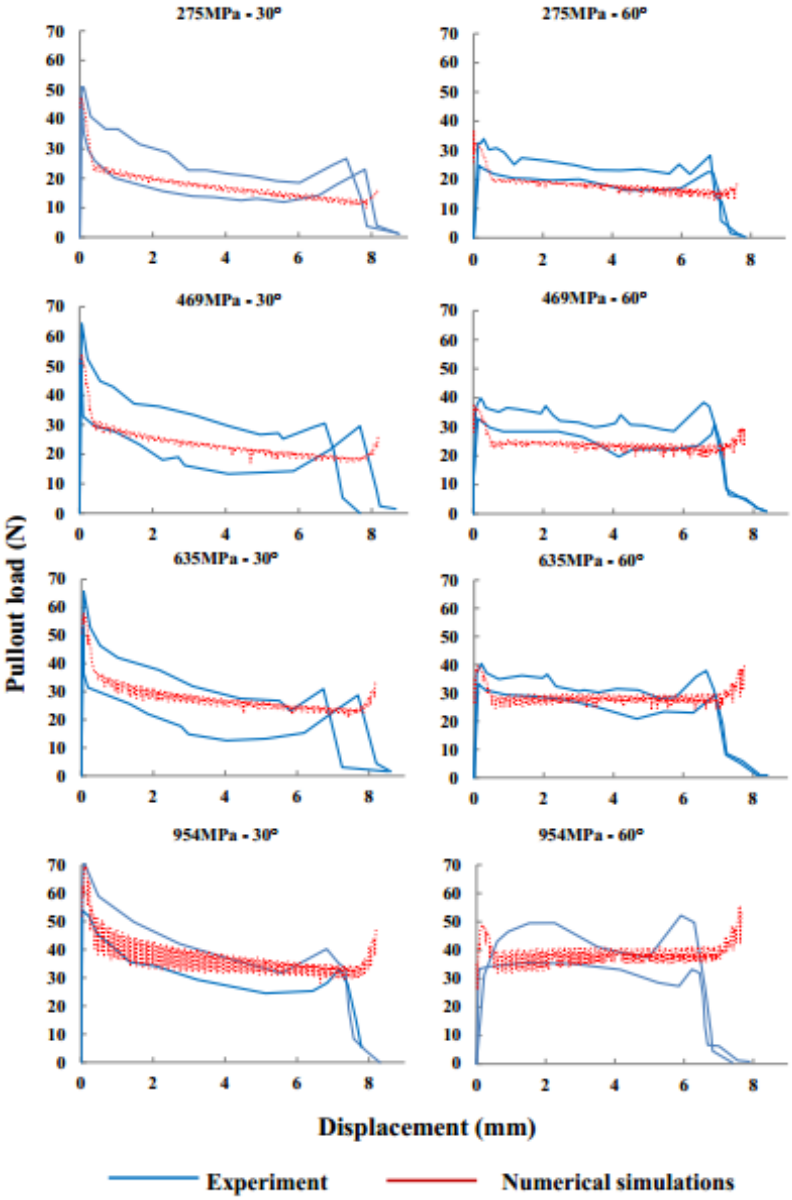


Figure 10. Numerical-experimental comparison: complete pullout curves for the four yield strengths given in Table 2, the fibre is inclined either at 30° (left column) or 60° (right column).

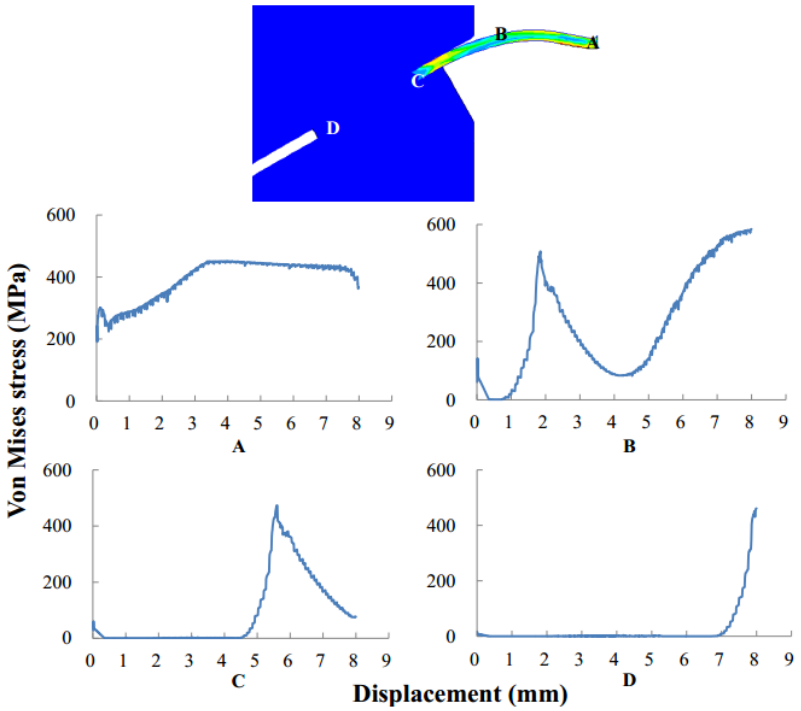


Figure 11. Four positions (A, B, C and D) within the fibre during pullout and the corresponding von Mises stress evolution for type 2 (fibre yield strength 469 MPa).

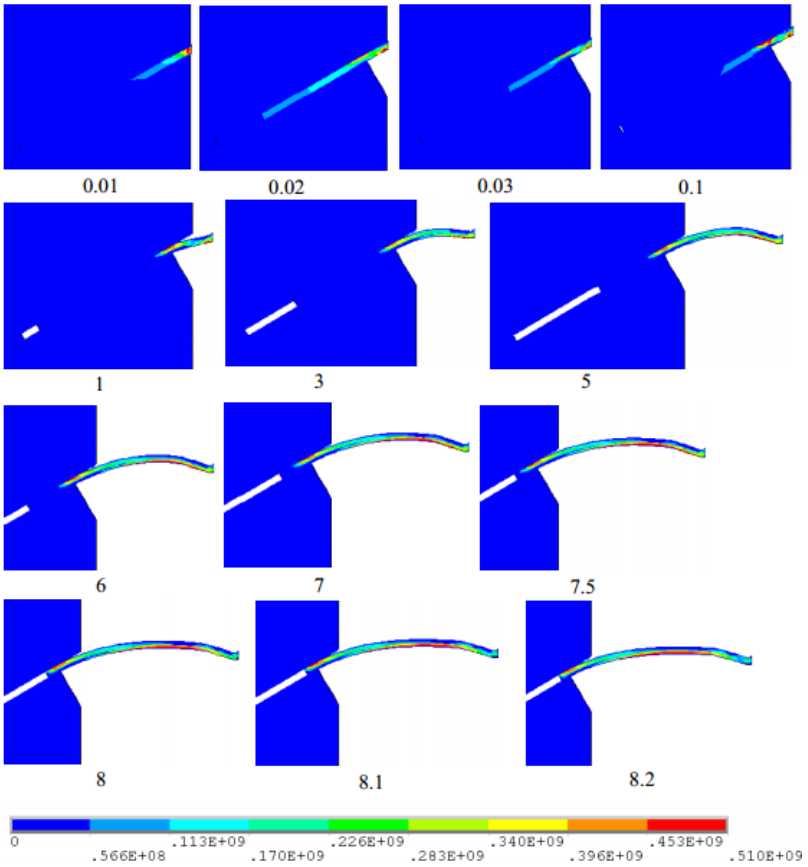


Figure 12. The first principle stress evolution (in Pa) for pullout displacement from 0.01 mm to 8.2 mm for fibre inclination of 30° and yield strength of 469 MPa (type 2 in Table 2).

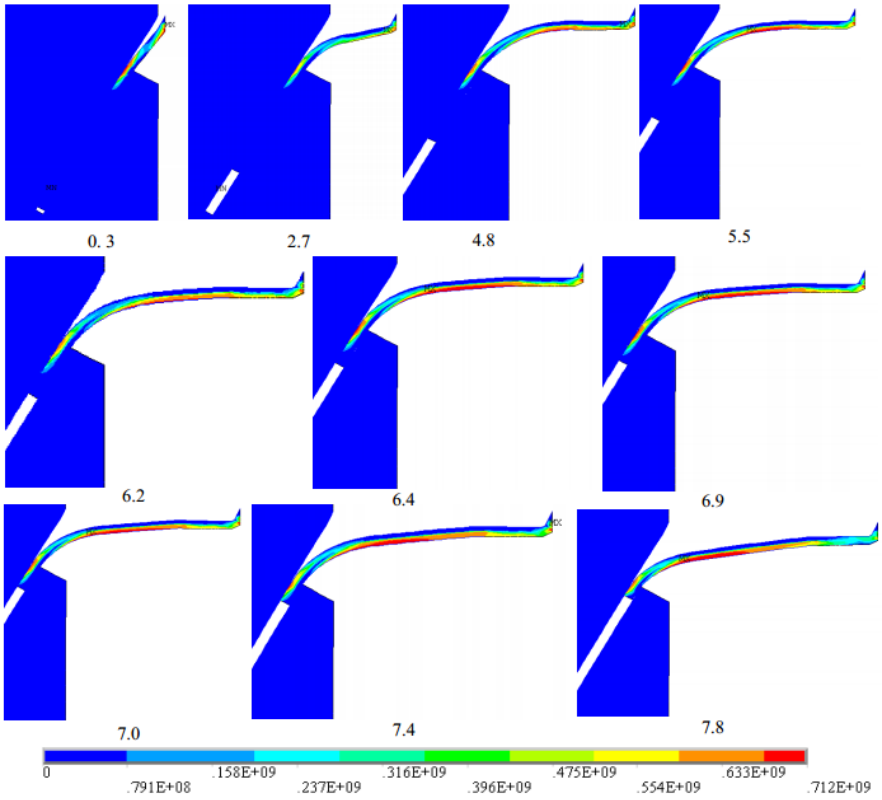


Figure 13. The first principle stress evolution (in Pa) for pullout displacement from 0.3 mm to 7.8 mm for fibre inclination of 60° and yield strength of 635 MPa (type 3 in Table 2).

Stress evolution in the matrix

For the matrix, we are more concerned on the tensile stress distribution to ensure that no fracture should take place where matrix spalling is not expected. Three representative points, E, F and G, see Fig. 14, are selected to display the first principal stress evolution in the matrix. The point E is where the matrix is expected to spall. The point G is the location where the fibre is anchored, whereas F is the point in the matrix close to the fibre centre.

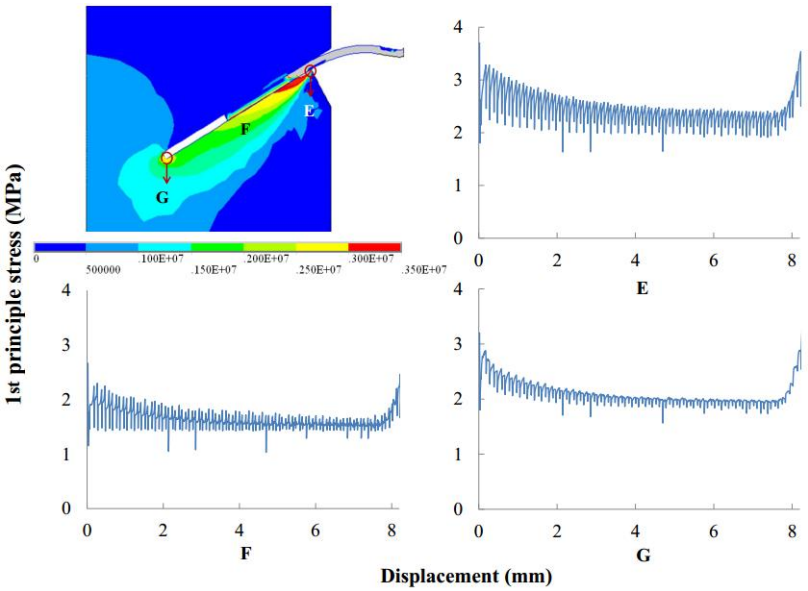


Figure 14. Three positions (E, F and G) in the matrix and the corresponding first principle stress evolution during pullout (the top left one is in Pa while the other three in MPa).

Since the tensile strength was not measured, we estimate it to be 1/12 of the compressive strength, which is 3.0 MPa. Note that at point E, there is significant stress fluctuation during the pullout process. This indicates, on the one hand, the stressing-relaxing cycle endured by the matrix. On the other hand, it can be attributed to the fact that the spalled matrix is assigned with a zero stiffness at the moment of spalling, whereas the real failure process is gradual. The stress evolution curves at points F and G, assimilate those of global pullout curves in Fig. 10, each with a different amplitude.

Furthermore, it is noted from Fig. 10 that the maximum tensile stress due to axial pull out of the fibre and bending load occurs at the close region at the fibre exit point. This confirms the assumption adopted by Zhang and Li [30] in their study on the effect of inclination angle on fibre rupture load in fibre reinforced cementitious composites.

Variation of the pullout work with respect to fibre yield strength

The pullout work is calculated as the area under the load vs slip displacement curve. In Fig. 15, both experimenta and numerical values for fibres inclined at 30° and 60° are depicted. Note that as a general trend, the pullout work increases with the increase of fibre yield strength, and such a tendency is correctly captured by our numerical model.

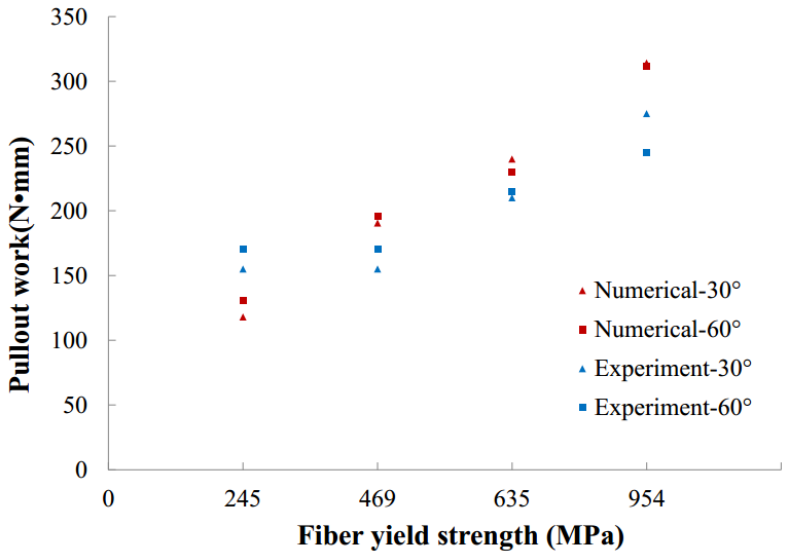


Figure 15. Pullout work vs fibre yield strength.

The maximum pullout load

To explore the effect of fibre inclination, a spectrum of angles up to 85° are simulated by keeping the fibre, matrix and interface properties fixed. It is known that the length of spalled matrix and the time when the matrix spalls both matter in the pullout responses. According to Laranjeira et al. [23], spalling is considered to take place just after the beginning of fibre debonding but prior to its full accomplishment. After some trial runs, this slip displacement is estimated, which is around 0.01 mm. Simulations of different fibre yield strengths are carried out and the obtained pullout curves are plotted in Fig. 16, the corresponding fitting parameters are given in Table 4.

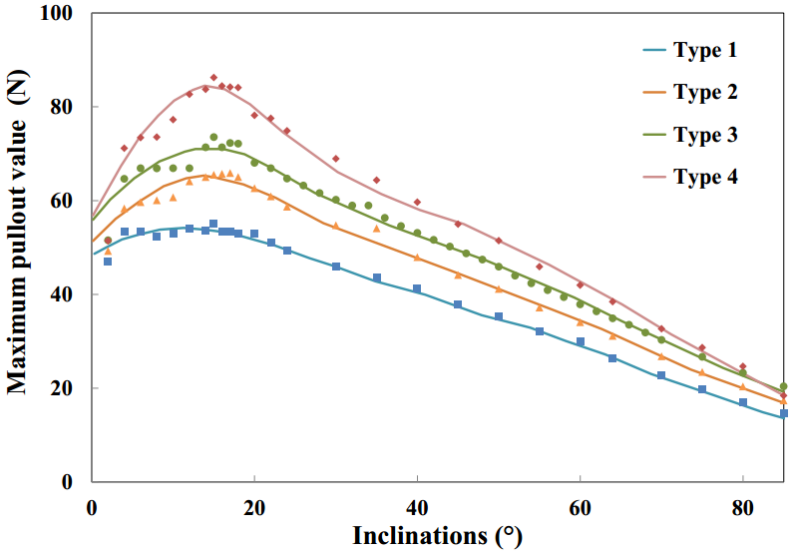


Figure 16. Maximum pullout load vs. inclination angle for the four yield strengths given in Table 2 and fitted curves using Eq. 10 with parameters given in Table 4.

$$P_{\max}(\theta) = F_1 \exp\left[-\left(\frac{\theta-\alpha_1}{\beta_1}\right)^2\right] + F_2 \exp\left[-\left(\frac{\theta-\alpha_2}{\beta_2}\right)^2\right] \tag{10}$$

Table 4. Fitted parameters for the maximum pullout load vs fibre inclination angle.

Fibre type	F_1 [N]	F_2 [N]	α_1 [°]	α_2 [°]	β_1 [°]	β_2 [°]
1	15.5	42.0	7.9	26.2	19.7	56.4
2	18.1	49.4	12.7	26.7	16.2	55.7
3	20.1	54.2	12.6	27.4	16.0	55.9
4	33.7	59.0	12.7	33.0	14.6	48.2

From Fig. 16, when the fibre is inclined at angles around 12° or 15°, the maximum pullout load is the largest. After that, the maximum pullout load goes down almost linearly. This differs from the result of Morton and Groves [20], who claimed that θmax is about 45° for polyester resin matrix of rather high tensile strength and steel fibres of high yield strength. This indicates that optimum inclination angle for maximum pullout resistance varies with both fibre and matrix strength as well as the interface properties.

CONCLUSIONS

We have proposed a numerical model to explicitly reproduce the pullout behaviour of a single fibre embedded within a cement-based matrix. This model takes into consideration of the gradual deterioration of interface bond, internal and dry friction as well as matrix spalling. In particular, a constitutive law which isolates the contributions of internal bond, internal friction and dry friction is formed and validated. Cohesive elements endorsed with mixed-mode fracture capacities are implemented to represent the bond-slip behaviour at the interface. Contact elements with Coulomb's friction are placed at the interface to simulate frictional contact. Matrix spalling is modelled through material erosion. The bond-slip behaviour is first calibrated through pull-out curves for fibres aligned with loading direction, then validated against experimental results carried out by Leung and Shapiro for steel fibres oriented at 30° and 60°. The influence of fibre yield strength on the stress distribution within the fibre and the matrix, the effect of the inclination angle on the pullout response are all explored in detail. The proposed methodology provides the necessary pull-out curves for a fibre oriented at a given angle for multi-scale models to study fracture in fibre-reinforced cementitious materials.

References

[1] Naaman, A. E., Namur, G., Najm, H., and Alwan, J. (1989). Bond Mechanisms in fibre reinforced cement-based composites. *Report UMCE 89-9*, Dept. of Civil Engineering, Univ. of Michigan, Ann Arbor, Michigan:1–233.

[2] Naaman, A. E., Namur, G., Alwan, J. M., and Najm, H. (1991). Fiber pullout and bond slip. I: Analytical study. *Journal of Structural Engineering*, **117**:2769–90.

[3] Banthia, N. and Trottier, J. (1994). Concrete reinforced with deformed steel fibers, part I: bond-slip mechanisms. *ACI Materials Journal*, **91**:435–446.

[4] Groth, P. (2000). *Fibre reinforced concrete fracture mechanics methods applied on self-compacting concrete and energetically modified binders*. PhD thesis, Lulea University of Technology, Sweden.

[5] Shah, S. P. and Ouyang, C. (1991). Mechanical-behavior of fiber-reinforced cement-based composites. *Journal of the American Ceramic Society*, **74**:2727–2953.

[6] Lawrence, P.(1972). Sometheoreticalconsiderationsoffiberpull-outfromanelasticmatrix. *Journal of Materials Science*, 7:1–6.

[7] Leung, C. K. Y. and Li, V. C. (1991). New strength-based model for the debonding of discontinuous fibers in an elastic matrix. *Journal of Materials Science*, **26**:5996–6010.

[8] Stang, H., Li, Z., and Shah, S. P. (1990). Pullout problem - stress versus fracture mechanical approach. *Journal of Engineering Mechanics-ASCE*, **116**:2136–2150.

[9] Leung, C. and Li, V. (1992). Effect of fiber inclination on crack bridging stress in brittle fiber reinforced brittle matrix composites. *Journal of the Mechanics and Physics of Solids*, **40**:1333–1362.

[10] Wang, Y., Li, V. C., and Backer, S. (1988). Modelling of fibre pull-out from a cement matrix. *International Journal of Cement Composites and Lightweight Concrete*, **10**:143–149.
[11] Fantilli, A. P. and Vallini, P. (2007). A cohesive interface model for the pullout of inclined steel fibers in cementitious matrixes. *Journal of Advanced Concrete Technology*, **5**:247–258.
[12] Chanvillard, G. (1999). Modeling the pullout of wire-drawn steel fibers. *Cement and Concrete Research*, **29**:1027–1037.
[13] Ellis, B., McDowell, D, L., and Zhou, M. (2014). Simulation of single fiber pullout response with account of fiber morphology. *Cement and Concrete Composites*, **48**:42–52.
[14] Li, V., Wang, Y., and Backer, S. (1990). Effect of inclining angle, bundling and surface treatment on synthetic fibre pull-out from a cement matrix. *Composites*, **21**:132–140.
[15] Ouyang, C., Pacios, A., and Shah, S. (1994). Pullout of inclined fibers from cementitious matrix. *Journal of Engineering Mechanics*, **120**:2641–2659.
[16] Leung, C. and Shapiro, N. (1999). Optimal steel fiber strength for reinforcement of cementitious materials. *Journal of Materials in Civil Engineering*, **11**:116–123.
[17] Laranjeira, F. (2011). *Design-oriented constitutive model for steel fiber reinforced concrete*. PhD thesis, Universitat Politècnica de Catalunya.
[18] Naaman, A. and Shah, S. (1976). Pull-out mechanism in steel fiber-reinforced concrete. *Journal of the Structural Division*, **102**(8):1537–1548.
[19] Cailleux, E., Cutard, T., and Bernhart, G. (2005). Pullout of steel fibres from a refractory castable: experiment and modelling. *Mechanics of materials*, **37**:427–445.
[20] Morton, J. and Groves, G. (1974). The cracking of composites consisting of discontinuous ductile fibres in a brittle matrix - effect of fibre orientation. *Journal of Materials Science*, **9**:1436–1445.
[21] Leung, C. and Chi, J. (1995). Crack-bridging force in random ductile fiber brittle matrix composites. *Journal of Engineering Mechanics*, **121**:1315–1324.
[22] Yu, R., Cifuentes, H., Rivero, I., Ruiz, G., and Zhang, X. (2016). Dynamic fracture behavior in fibre-reinforced cementitious composites. *Journal of the Mechanics and Physics of Solids*, pageDOI:10.1016/j.jmps.2015.12.025.
[23] Laranjeira, F., Aguado, A., and Molins, C. (2010). Predicting the pullout response of inclined straight steel fibers. *Materials and Structures*, **43**:875–895.
[24] Brandt, A. (1985). On the optimal direction of short metal fibres in brittle matrix composites. *Journal of Materials Science*, **20**:3831–3841.
[25] Camacho, G. and Ortiz, M. (1996). Computational modelling of impact damage in brittle materials. *International Journal of Solids and Structures*, **33** (20-22):2899–2938.
[26] Ruiz, G., Pandolfi, A., and Ortiz, M. (2001). Three-dimensional cohesive modeling of dynamic mixed-mode fracture. *International Journal for Numerical Methods in Engineering*, **52**(1-2):97–120.
[27] Alfano, G. and Crisfield, M. (2001). Finite element interface models for the delamination analysis of laminated composites: mechanical and computational issues. *International Journal for Numerical Methods in Engineering*, **50**:1701–1736.
[28] Yu, R. C. and Ruiz, G. (2006). Explicit finite element modeling of static crack propagation in reinforced concrete. *International Journal of Fracture*, **141**:357–372.
[29] Yu, R., Zhang, X., and Ruiz, G. (2008). Cohesive modeling of dynamic fracture in reinforced concrete. *Computers and Concrete*, **5**(4):389–400.
[30] Zhang, J. and Li, V. (2002). Effect of inclination angle on fiber rupture load in fiber reinforced cementitious composites. *Composites Science and Technology*, **62**:775–781.

Minimum volume of the longitudinal fin with rectangular and triangular profile by a modified Newton-Raphson method

***Nguyen Quan[1], †Nguyen Hoai Son[2], and Nguyen Quoc Tuan[2]**

[1]Department of Engineering Technology, Pham Van Dong University, Viet Nam.
[2]GACES, University of Technical Education in Ho Chi Minh City, Viet Nam
†Corresponding author: sonnh@hcmute.edu.vn

Abstract

The minimum volume of nonlinear longitudinal fin with rectangular and triangular profile by using the modified Newton-Raphson method is presented in this paper. The dimension of the fin profile is regarded as optimization variables. Furthermore, a mechanism called "volume updating" is added into the modified Newton-Raphson algorithm to obtain the minimum volume of the fin. Two examples are illustrated to demonstrate the proposed method. The obtained results showed that the proposed method use efficiently and accurately in finding the minimum volume of the nonlinear longitudinal fin problem with the rectangular and triangle profile.

Keywords: Shape Optimization; Modified Newton-Raphson; Rectangular fin; Triangular fin.

Introduction

Fin or extended surface is used widely in various industrial applications when we want to improve the convective heat transfer from a hot surface where cooling is required [1]. However, the use of fins increases the volume or mass of systems and rise the costs of production. Consequently, the optimization of fins for light weight and high efficiency and compact heat exchanger system is of great interested and have been done in the past several decades.

Fin optimization problems can be divided into two approaches. The first approach of optimization problem is to select a simple profile (i.e. rectangular or triangular) and then determine the dimensions of fin so that either maximize the heat transfer rate for a given volume or minimize the volume of fin for a specified heat dissipation. In the second approach, the shape of fin is determined so that the volume of the material used is minimum for a given heat loss. For this second approach, the criterion of fin optimization problems was first proposed by Schmidt [2]. For purely conduction and convection fins, the author suggested that the minimum volume of the optimization fins is a parabolic shape. Unfortunately, the parabolic profiles of optimization fin in the second approach are curved surface with zero tip thickness which is too complex and expensive to manufacture. Thus, the first approach of problems is more relevant and important than the second type of problems.

In fact, the rectangular and triangular profiles are widely used in the heat exchanger system due to the ease of fabrication. As a result, more studies have been performed to determine the optimal size of these types. Under the assumption constant thermal parameters, negligible effects of heat transfer from the tip, and approximation of one-dimensional heat transfer equation, the optimization of rectangular and triangular profiles is treated in the book by Kraus [1]. Aziz [3] published an article which present a literature survey on optimum dimension of this object. Aziz [4, 5] optimized the rectangular and triangular fins with convective boundary conditions and presented the optimum design of a rectangular fin with a step change in cross-sectional area under the constant thermal parameters. Under the variable

thermal parameters, Yu [6] studied the optimization of rectangular by applying Taylor transformation method. Recently, by using the differential transformation method, Poozesh [7] presented the efficiency of convective-radiative fin with temperature-dependent thermal conductivity. In Poozesh's paper, the effects of convection-conduction parameter, thermal conductivity parameter and the radiation-conduction parameter on efficiency of fin are considered and discussed. For a bi-dimensional analysis, Kang [8, 9] estimated the optimum dimension of annular fins with rectangular profile under thermally asymmetric convective and radiating condition as well as optimized an annular trapezoidal fin using a new approach to a two-dimension analytical method. However, these above researches were performed based on the analytical method under assumption of constant thermal parameters. The drawback of analytical methods is that they can not solve the general non-linear fin design problem. However, none of previous published papers however propose the effective methods to minimize the volume of rectangular and triangle fins for general non-linear fin design problem until now.

In this paper, an effective method is presented to find the minimum volume of longitudinal fin of rectangular and triangular profiles for general high non-linear fin design problem based on modified Newton Raphson method (MNR). A mechanism called as "volume updating" is added in MRN algorithm to obtain the minimum volume of optimum fin. The potential and feasibility of applying MNR as an optimization method on the fin problems will be demonstrated in this work.

Problem Statement

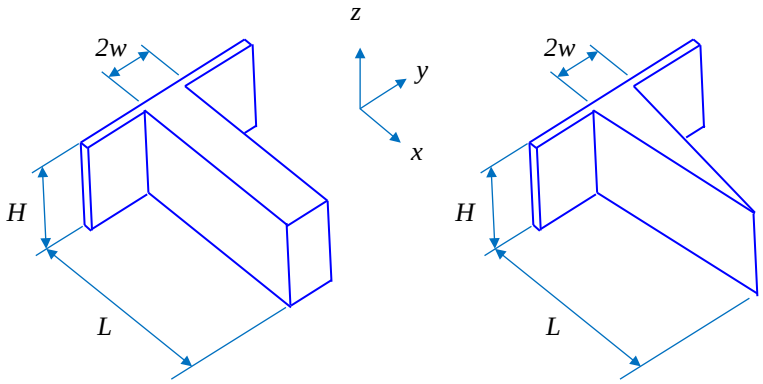


Figure 1. The longitudinal fin with rectangle and triangle shape

Consider a longitudinal symmetric fin model with rectangular and triangular profiles as Figure 1, in the steady state condition, the general heat transfer equation without internal heat source for the two-dimensional model given by the semi-cross-section of the natural convection and radiation cooled fin takes on the following forms:

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)=0 \text{ in domain of fin} \tag{1}$$

$$-kA_b\frac{\partial T}{\partial x}=q_{flow} \text{ at fin base} \tag{2}$$

$$-k\left(\frac{\partial T}{\partial x}+\frac{\partial T}{\partial y}\right).\mathbf{n}=h(T-T_{\infty})+\varepsilon\sigma(\mathrm{T}^4\text{-}\mathrm{T}_{sur}^4) \text{ at convective surface} \quad (3)$$

$$k\frac{\partial T}{\partial y}=0 \text{ at symmetric of fin} \quad (4)$$

where T is the unknown temperature field over the cross-section domain of fin , k is the heat thermal conductivity, A_b is the fin cross section area at the base, q_{flow} is the inward total heat loss at the base, h is the convective heat transfer coefficient, ε is emissivity coefficient, σ is Estefan-Boltzmann constant, T_{∞} and T_{sur} is the ambient and surrounding temperature respectively, and n is the exterior normal vector of the convective surface. In general, the coefficients k, h, ε are constant or functions of temperature.

When the shape of fin and all boundary condition is known and given, the temperature field of fin and the base temperature could be estimated by solving the non-linear fin design problem (Eqs.(1-4)). This direct problem is solved by the finite element method (FEM) [10].

The Optimization Problem

Modified Newton Raphson

In this paper, the purpose of optimization process is to minimize the volume of the longitudinal fins of rectangular and triangular profiles for a given heat loss and the specified base temperature. Therefore, the dimensions of fin profiles are regarded as optimization variables. For this two-dimensional geometric problem, the dimension of fin is determined by the length and width of fin. We will thus have 2 optimization variables for both of rectangular and triangle profiles as shown in Figure 2. Besides, to remain the continuity during the optimization process, the position of control points must satisfy the following conditions:

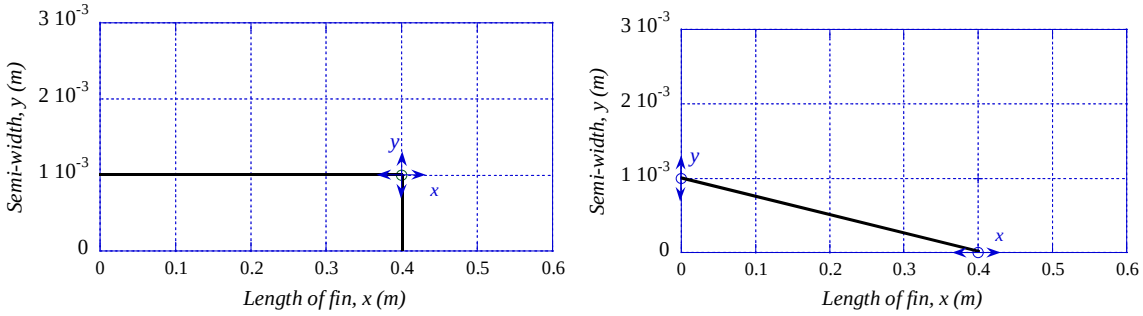


Figure 2. The profile of fins and their optimization variables: a) rectangular fin; b) triangular fin

$$\begin{cases} x>0 \\ y>0 \end{cases} \quad (5)$$

MNR method [11] is used to find out the minimum volume by finding the optimal position of control points. The proposed method directly formulates the problem from two comparisons between the calculated and the expected temperature at the base, and between the calculated

and expected volume of the fin. Therefore, the expected base temperature T_x^i and the expected fin volume V_x are necessary to be given first; the calculated temperature T_c^i and the calculated fin volume V_c are evaluated from direct problem. Then, the estimation of optimal fin shape can be recast as the solution of a set of nonlinear equations as following:

$$\begin{cases} T_c^i - T_x^i = 0 \\ V_c - V_x = 0 \end{cases} \quad i = 1, 2...M \tag{6}$$

where, M is the number of the temperature equation which is obtained from the base. As a result, there are M+1 equations in Eq. (6).

The characteristic of fin is that the fin width compared to the fin length is very small. Therefore, the variation of the base temperature along with the width of fin could be neglected. Consequently, the expected temperature at the base would be assigned to one expected value. Furthermore, since the value of fin volume is very small compared to the base temperature, the volume value is converted into the temperature value so that the influence of fin volume and base temperature in Eq. (6) is the same. Subsequently, Eq. (6) can be re-written as following:

$$\begin{cases} T_c^i - T_x = 0 \\ \hat{V}_c - T_x = 0 \end{cases} \tag{7}$$

where, T_x is the expected temperature at the base and $\hat{V}_c$ is the converted volume given by:

$$\hat{V}_c = \frac{V_c}{V_x}.T_x \tag{8}$$

The detail procedure to solve Eq. (7) can be shown as following:

$$\mathbf{T} = [\{T_c^1 - \mathrm{T}_x\}, \{T_c^2 - \mathrm{T}_x\}, \cdots, \{T_c^M - \mathrm{T}_x\}, \{\hat{V}_c - \mathrm{T}_x\}]^T = \{\hat{T}_c\}^T \tag{9}$$

where, $\hat{T}_c$ is the component of vector $\mathbf{T}$.

The optimization variables are set as folowing:

$$\boldsymbol{\chi} = \{x, \mathrm{y}\}^T = \{\chi_1, \chi_2\}^T = \{\hat{\chi}_v\} \tag{10}$$

where, x, y are the size of the fin (as Fig. 1), $\hat{\chi}_v$ is the component of vector $\boldsymbol{\chi}$

The derivative of $\hat{\Phi}_c$ with respect to $\hat{\chi}_v$ is can be expressed as following:

$$\mathbf{S} = \frac{\partial \hat{T}_c}{\partial \hat{\chi}_v} \tag{11}$$

where, $\mathbf{S}$ is the sensitivity matrix.

With the above derivatives from Eq. (6) to Eq. (11), we have the following equation:

$$\boldsymbol{\Delta}_k = -\left[\mathbf{S}^T(\boldsymbol{\chi}_k)\mathbf{S}(\boldsymbol{\chi}_k)\right]^{-1}\mathbf{S}^T(\boldsymbol{\chi}_k)\mathbf{T}(\boldsymbol{\chi}_k) \tag{12}$$

$$\boldsymbol{\chi}_{k+1} = \boldsymbol{\chi}_k + \lambda\boldsymbol{\Delta}_k \tag{13}$$

where, λ is the factor to adjust the step size of $\boldsymbol{\Delta}_k$ so that the constraints of Eq. (5) are satisfied.

From Eq. (9), it is claimed that the solution can be achieved when the base temperature and the appropriate volume of fin is given. However, the minimum volume of the fin is unknown prior and is the optimization goal. To solve this problem, an approach called “volume

updating" is added into the modified Newton Raphson algorithm. This approach is based on "curve fitting" mechanism of the modified Newton Raphson method. In this mechanism, the obtained solution is the best approximation which is defined as that which minimizes the sum of squared differences between the computed and expected value. As a result, in Eq. (9), the larger value of *N* is, the closer solution to the expected temperature compared to the expected volume is. Consequently, "volume updating" approach is performed as following:

Step 1: Set a large value for M and guess a small initial value of fin volume.
Step 2: Use Eqs. (9-13) to find the best solution.
Step 3: Update the new volume obtained from the best solution of step 2 and return to step 1.
Step 4: Terminate the process if the stopping criterion is satisfied.

The stopping criteria

The modified Newton Raphson method from Eq. (11) to Eq. (15) is used to determine the optimal location of the control points which are presented as the unknown variables, $\boldsymbol{\chi}$. The step size $\boldsymbol{\Delta}_k$ goes from $\boldsymbol{\chi}_k$ to $\boldsymbol{\chi}_{k+1}$ and it is determined from Eq. (16). Once $\boldsymbol{\Delta}_k$ is calculated, the iterative to determine $\boldsymbol{\chi}_{k+1}$ is executed until the stopping criterion is satisfied. There are two stopping criteria used in the proposed method. One is for updating the volume and another is for modified Newton Raphson method. Base on the discrepancy principle [1], the volume would be updated when both of two criteria are satisfied as following:

$$\begin{cases} \mathbf{T}_c - T_x \geq 0 \\ \left\| \mathbf{J}\left(\chi_{k+1}\right) - \mathbf{J}\left(\chi_{\kappa}\right) \right\| \leq \delta \left\| \mathbf{J}\left(\chi_{k+1}\right) \right\| \end{cases} \tag{14}$$

where,

$$\left\| \mathbf{J}\left(\boldsymbol{\chi}_{k+1}\right) \right\| = \sum_{i=1}^{M} \left[\mathbf{T}_c^i - T_x \right]^2 + \left[\hat{V}_c - T_x \right]^2 \tag{15}$$

and the stopping criteria is given by

$$\left\| \mathbf{T}_c - T_x \right\| \leq e \left\| T_x \right\| \tag{16}$$

or

$$\left\| \mathbf{J}\left(\chi_{k+1}\right) - \mathbf{J}\left(\chi_{\kappa}\right) \right\| \leq \delta \left\| \mathbf{J}\left(\chi_{k+1}\right) \right\| \tag{17}$$

where, e and δ are small positive value known as the convergence tolerances.

Computational Algorithm

The procedure for the proposed method can be summarized as following:

Given overall convergence tolerance e and δ, the initial control point $\boldsymbol{\chi}_0$, the initial volume of fin V_x^0, and the adjusting factor λ (say $\lambda = 1$ in the present work). The value $\boldsymbol{\chi}_k$ is known at the iteration as following:

Step 1: Solve the direct problem Eqs. (1-4), and compute $\mathbf{T}_c$.

Step 2: Integrate $\mathbf{T}_c$ with $\mathbf{T}_x$ through Eq. (9) to construct $\mathbf{T}$.

Step 3: Calculate the sensitivity matrix $\mathbf{S}$ through Eq. (11).

Step 4: Knowing $\mathbf{S}$ and $\mathbf{T}$, calculate the step size $\boldsymbol{\Delta}_k$ from Eq. (12).

Step 5: Calculate $\boldsymbol{\chi}_{k+1}$ through Eq. (13).

Step 6: If condition of Eq. (5) is not satisfied, replace $\lambda = 0.1\lambda$ and return to step 5. Otherwise, accept the new control points $\boldsymbol{\chi}_{k+1}$ and set $\lambda = 1$ again.

Step 8: Update the fin volume if the updating criterion Eq. (14) is satisfied, and replace k by $k+1$ and return to step 2.

Step 9: Terminate the process if the stopping criterion Eq. (16) or Eq. (17) is satisfied. Otherwise, replace k by $k+1$ and return to step 2.

Results and Discussions

In this section, two cases with the triangle and rectangle profile of the longitudinal fin are deal with to demonstrate the proposed method. The two-dimensional model will be considered in two cases. Additionally, the optimal results by the proposed method are discussed and compared with the theory results by Kruas [1]. The longitudinal fin with the height of fin of $H = 0.2[\text{m}]$ and the thermal conductivity of $k = 58.3[\text{W/mK}]$ is considered in two cases. It is assumed that our purpose is to find the minimum volume of the fin so that the fin can dissipate a given heat flow of $Q = 20[W]$ with the base temperature of $T_b = 400[K]$ in the surrounding ambient with the temperature $T_a = 300[K]$. The convective heat transfer coefficient is considered to be constants and obtained from Eq. (18) by Dobaru [13] as following:

$$h = \frac{8k\,\text{Pr}^{1/2}}{3H\left[336\left(\text{Pr}+\frac{9}{5}\right)\right]}\left(\frac{g\beta\left[T(x)-T_\infty\right]H^3}{\nu^2}\right)^{1/4} \tag{18}$$

where, all the fluid properties are computed at a mean temperature, $T_m = (T_b + T_a)/2$. With the given thermal parameters above, the mean convective heat transfer coefficient is $\overline{h} = 5.2564\ [\text{W/m}^2\text{K}]$.

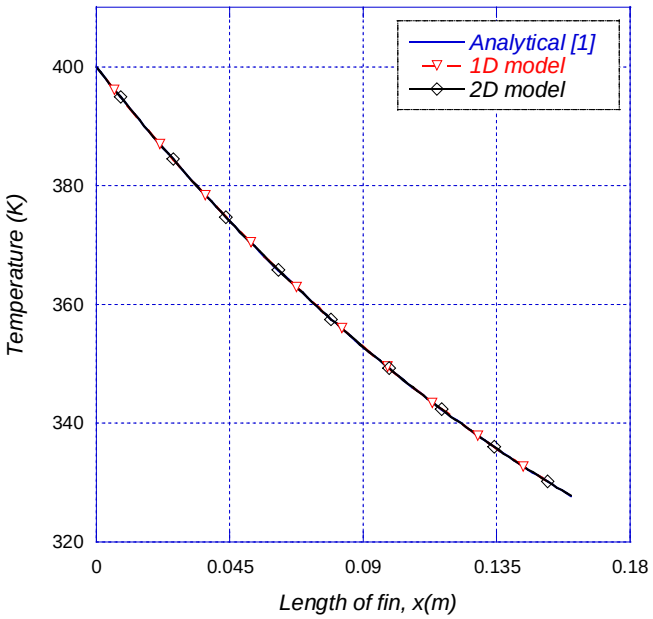


Figure 3. The temperature distribution along the length of the optimal fin for triangle profile.

In Case 1, a fin design problem with triangle profile of the longitudinal fin is considered. For optimization procedure by MNR, the value of updating and stopping criteria used were

$\delta = 10^{-4}$ and $\varepsilon = 10^{-4}$ respectively. The initial volume is $V_{xpcd}^{0} = 4e-5(m^3)$. The initial dimensions of the fin with triangle profile are $x_0 = 0.3[m]$ and $y_0 = 0.001[m]$. In this case, the optimal results obtained by the proposed method and the theory optimal results are shown in Table 1. The relative difference of the geometrical parameters of the optimal fin between the theoretical value and MNR's results are also shown in Table 2. Furthermore, the temperature distribution along the length of the optimal triangle fin is presented in Figure 3.

Table 1. The geometrical parameters of the optimal fin by MRN method and theoretical results for the triangle profile

Dimension of optimum fin	Theoretical Results	The proposed Method
The length, *x[m]*	$x = 1.6022e-1$	$x = 1.6020e-1$
The semi-width, *y[m]*	$y = 1.3498e-3$	$y = 1.3503e-3$
The min volume, *V[m^3]*	$V = 4.3255e-5$	$V = 4.3265e-5$

Table 2. The relative difference of fin geometrical parameters between MRN method and theoretical results for the triangle profile

Dimension of optimum fin	Error (%)
The length, %*x*	0.01%
The semi-width, %*y*	0.04%
The min volume, %*V*	0.023%

As shown, with specified thermal properties and given boundary conditions above, the minimum volume of the optimal triangle fin is about $V = 4.325e-5$. The optimal length of triangle fin is about $x = 0.16[m]$ and the optimal semi-width triangle fin is about $y = 1.35e-3[m]$. Table 2 show that the relative error of the geometrical parameters of the optimum fin between MRN method and the theoretical is small. The relative error for the minimum volume is 0.23% and that for the length and semi-width are 0.01% and 0.04% respectively. This mean that the results obtained by the proposed method satisfied the given condition and are in high agreement with the theoretical values.

Table 3. The geometrical parameters of the optimal fin by MRN method and theoretical results for the rectangle profile

Dimension of optimum fin	Theoretical result	The proposed Method	
		2D model (no tip convection)	2D model (tip convection)
The length, *x[m]*	$x = 1.5178e-1$	$x = 1.5179e-1$	$x = 1.5022e-1$
The semi-width, *y[m]*	$y = 1.0313e-3$	$y = 1.0313e-3$	$y = 1.0350e-3$
The min volume, *V[m^3]*	$V = 6.2613e-5$	$V = 6.2614e-5$	$V = 6.2188e-5$

In Case 2, the fin design problem with the rectangle profile is investigated. In this case, the initial dimensions of the rectangle fin are $x_0 = 0.2[m]$ and $y_0 = 0.001[m]$. Two cases of no tip convection and tip convection are considered in this case. Table 3 showed the optimal results

achieved by the proposed method and the theoretical formulation. Table 4 illustrated the relative deviation of the geometrical parameters of the optimal rectangle fin between MNR's method and theoretical formulation. In addition, the temperature distribution along the length of the optimal rectangle fin is drawn in Figure 4.

Table 4. The relative difference of fin geometrical parameters between MRN method and theoretical results for the rectangle profile

Dimension of optimum fin	The proposed Method	
	2D model (without convective tip)	2D model (convective tip)
The length, %x	0.02%	1%
The semi-width, %y	0%	0.3%
The min volume, %V	0.0016%	0.68%

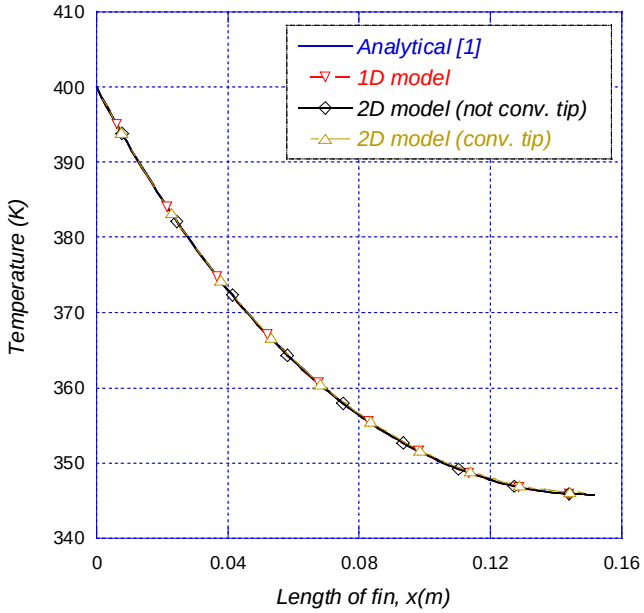


Figure 4. The temperature distribution along the length of the optimal fin for the rectangle profile.

The obtained results showed that there is good approximation between the optimal result by MNR's method and theoretical method for the case of insulated tip. Particularly, the minimum volume the case of insulated tip is about $V = 6.26e-5[m^3]$ for both of the theory method and the proposed method. The length and semi-width of the optimal rectangle fin are respectively about $x = 1.52e-1[m]$ and about $y = 1.03e-3[m]$ with the very small relative deviation between the methods (as Table 4). For the case of the convective tip, the minimum volume of optimal fin is about $V = 6.22e-5[m^3]$. As shown in Table 4, the value of the optimal rectangle fin volume with the convective tip is 0.68% less than that with the insulated tip. This is due to the face that the consideration of convective tip leads the increase of heat dissipation comparing with the assumption of the insulated tip. Thus, the volume of the optimal rectangle fin with the convective tip is less than that with the insulated tip.

With the obtained results from two cases, it can be said that the proposed method is potential and feasible in finding the minimum volume of the optimum fin with rectangle and triangle profile. Furthermore, the proposed method does not depend on the type of the direct problem

(linear or non-linear direct problem). In the other words, the proposed method can be utilized in finding the minimum volume for any fin design problem with rectangle and triangle profile.

Conclusions

In this work, the minimum volume of the longitudinal fin with the rectangular and triangular profile for the given heat flow and the expected temperature at the base by using the modified Newton Raphson method was presented. A mechanism called as "volume updating" was added in the proposed algorithm to obtain the minimum volume of the optimum fin. Two cases with the rectangle and triangle profile were performed to validate the proposed method. The obtained results by MNR's method have been compared with the results of Kraus [1]. The results showed that the values of the volume of the optimal fin are in good agreement with that of Kraus [1] in all two cases. In the other words, it can be declared that the proposed method is an efficient and accurate method to find the minimum volume of the optimal fin with triangle and rectangle profile for the given heat flow and the expected temperature at the base. Furthermore, the proposed method do not depend upon the type of the direct problem. Thus, this method can be applied for any linear or non-linear fin design problem.

References

[1]. Kraus, A. D., Aziz, A., & Welty, J. (2002). *Extended surface heat transfer*. John Wiley & Sons.

[2]. Schmidt, E. (1926), Die Warmeubertragung durch Rippen, *Zeitschrift des Verein Deutscher Ingenieure*, **70**, 885 – 947.

[3]. Aziz, A. (1992). Optimum dimensions of extended surfaces operating in a convective environment. *Applied Mechanics Reviews*, ***45*****(5)**, 155-173.

[4]. Aziz, A. (1985). Optimization of rectangular and triangular fins with convective boundary condition. *International communications in heat and mass transfer*, ***12*****(4)**, 479-482.

[5]. Aziz, A. (1994). Optimum design of a rectangular fin with a step change in cross-sectional area. *International communications in heat and mass transfer*, ***21*****(3)**, 389-401.

[6]. Yu, L. T. (1998). Application of Taylor transformation to optimize rectangular fins with variable thermal parameters. *Applied Mathematical Modelling*, ***22*****(1)**, 11-21.

[7]. Poozesh, S., Nabi, S., Saber, M., Dinarvand, S., & Fani, B. (2013). The efficiency of convective-radiative fin with temperature-dependent thermal conductivity by the differential transformation method. *Research Journal of Applied Sciences, Engineering and Technology*, ***6*****(8)**, 1354-1359.

[8]. Kang, H. S., & Look Jr, D. C. (2007). Optimization of a thermally asymmetric convective and radiating annular fin. *Heat transfer engineering*, ***28*****(4)**, 310-320.

[9]. Kang, H. S., & Look Jr, D. C. (2009). Optimization of a trapezoidal profile annular fin. *Heat transfer engineering*, ***30*****(5)**, 359-367.

[10]. Baskharone, E. A. (2013). *The Finite Element Method with Heat Transfer and Fluid Mechanics Applications*. Cambridge University Press.

[11]. Nguyen, Q., & Yang, C. Y. (2016). Design of a longitudinal cooling fin with minimum volume by a modified Newton–Raphson method. *Applied Thermal Engineering*, ***98***, 169-178.

[12]. Beck, J. V., Blackwell, B., & Clair Jr, C. R. S. (1985). *Inverse heat conduction: Ill-posed problems*. James Beck.

[13]. Bobaru, F., & Rachakonda, S. (2004). Boundary layer in shape optimization of convective fins using a meshfree approach. *International journal for numerical methods in engineering*, ***60*****(7)**, 1215-1236..

The transient of Visco-elastic MHD fluid through Stokes oscillating porous plate: an exact solution

Bhaskar Kalita: drbhaskarkalita@yahoo.com

Department of Mathematics: T. H. B. College, Sonitpur, Assam, India

Abstract

The magnetohydrodynamic transient free convection flow of a visco-elastic fluid (Rivlin - Ericksen) caused by the sinusoidal oscillation of a plane flat porous plate has been studied in this paper. The constitutive equations of continuity and mass conservation of visco-elastic fluid are solved by Laplace transform technique. The Velocity profiles of transient and steady-state due to porous plate in presence Magnetic Hartmann Number and porosity of the medium are obtained in exponential forms and Complementary Error Functions. The results got for velocity profiles are shown through graphs and discussed in the concluding section.

Key Words: MHD, Transient flow, Porous Plate, Rivlin – Ericksen fluid, Sinusoidal Oscillation.

Introduction

Stokes first studied the unsteady free convection flow of a viscous incompressible fluid past an impulsively started infinite horizontal plate. The plate oscillates in its own plane. The plate has two natures- one is of impulsively starting in its own plane suddenly set into motion which creates a start - up flow and other one is of oscillating – that oscillates in its own plane. H. Schlichting[*] called the farmer problem as "Stokes first problem" and later one as "Stokes second problem". Stokes presented exact solutions to both the problems. These problems being of fundamental in nature are referred in all the text books of viscous flow. Stokes result for the oscillating plate is the steady- state solution which applies after the effect of any initial velocity profile has died out. But this solution is not a complete solution, since it does not satisfy the initial condition. The complete solution for the problem requires the transient solution as well as steady state solution. And this is given by Panton (7). He presented the solution to transient problem in exact from in terms of standard mathematical functions and velocity distributions for the plate either oscillating as $Sin(T)$ or$-Cos(T)$. Later on, $Deka\ et.al.$ (2) studied this problem considering semi – infinite incompressible viscous fluid in the presence of a uniform magnetic field applied transversely to the plate. I determined to extend this paper by considering the fluid as Visco – elastic electrically conducting and the flat plate as Porous. In section 2, the mathematical formulation and a solution to transient component is presented in terms of standard mathematical functions. In section 3, characteristics of the solutions are cited, while in section 4, the problem is concluded with outcome of investigation.

Mathematical Formulation

The constitutive equation of second order Visco – elastic (extended by Rivlin - Ericksen) fluid in tensor notation is as follows -

$$\tau_{ij} = -p\delta_{ij} + \mu_1 A_{(1)ij} + \mu_2 A_{(2)ij} + \mu_3 A_{(1)i\alpha} A_{(1)\alpha j} \qquad (1)$$

Where τ_{ij} is the stress tensor, p is the hydrostatic pressure, δ_{ij} is the Kronecar delta, $A_{(1)}$ and $A_{(2)}$ are Rivlin – Ericksen tensors of order 1 and 2, and $\mu's$ are coefficients of viscosity. Here $A_{(1)}$ and $A_{(2)}$ are given by symmetric tensors and they are defined by

$$A_{(1)ij} = v_{i,j} + v_{j,i} \qquad (2)$$

$$A_{(2)ij} = a_{i,j} + a_{j,i} + 2v_{m,i}v_{m,j} \qquad (3)$$

$$a_{i,j} = \frac{\partial v_i}{\partial t} + v_j v_{i,j} \quad (i,j,m = 1,2,3)$$

$v_i =$ Component of velocity, $a_i =$ component of acceleration.

Here the plate is porous and semi-infinite horizontal. The $X^{/} -$ axis is taken along the flat plate while the $Y^{/}$ - axis is taken normal to the plate. Let $u^{/} and\ v^{/}$ be the fluid velocities along $X^{/} and\ Y^{/}$ axis, respectively. Then since the plate is semi infinite in extent, the fluid is taken to occupy the upper half plane. $u^{/}$ is a function of $y^{/}$ and $t^{/}$ and $v^{/}$ is independent of $y^{/}$. The fluid is electrically conducting and the plate is non- conducting. Let a uniform magnetic field H_0 be applied in a direction perpendicular to $X^{/}$- axis. The fluid is assumed to be of low conductivity; so induced magnetic field is negligible. The Lorentz's force is $-\sigma H_0^2 u^{/}$. At time $t(\leq 0)$, the plate and fluid are at rest. At $t\ (> 0)$, the plates start oscillating in its own plane. For boundary condition it is assumed that there is no slip at the wall.

Under these assumptions, we can write the continuity and momentum equations which governs the flow field as –

$$\frac{\partial v^{/}}{\partial y^{/}} = 0 \qquad (4)$$

$$\frac{\partial u^{/}}{\partial t^{/}} + v^{/}\frac{\partial u^{/}}{\partial y^{/}} = \vartheta^{/}\frac{\partial^2 u^{/}}{\partial y^{/2}} - \frac{\sigma H_0^2 u^{/}}{\rho} + \frac{K_0}{\rho}\frac{\partial^3 u^{/}}{\partial t^{/}\partial y^{/2}} \qquad (5)$$

Where,

$\rho =$ density of the fluid, $H_0 =$ uniform magnetic field applied transversely to the plate, $\sigma =$ electrical conductivity of the fluid, $\vartheta =$co-efficient of Kinematic viscosity of the fluid, $K_0 =$ coefficient of elasticity.

The initial and boundary conditions are

$$u'(y',0)=0; \quad u'(\infty,t')<\infty; \quad u'(0,t')=Usin(\omega' t') \tag{6}$$

The non-dimensional quantities are defined as follows:

$$y=\frac{y'U}{\vartheta}, u=\frac{u'}{U}, t=\frac{t'U^2}{\vartheta}, R_c=\frac{K_0U^2}{\alpha_0}, \ M=\frac{\sigma H_{0\vartheta}^2}{\rho U^2}, \ \omega'=\frac{U^2}{\vartheta}, \ V=\frac{v'}{U}, \ P=\frac{\alpha_0 C}{K} \tag{7}$$

Hence, the equations of continuity, motion and boundary conditions reduces to-

$$U\frac{\partial V}{\partial y}=0 \tag{9}$$

$$\frac{\partial u}{\partial t}+V\frac{\partial u}{\partial y}=\frac{\partial^2 u}{\partial y^2}-Mu+R_c\frac{\partial}{\partial t}\left(\frac{\partial^2 u}{\partial y^2}\right), \qquad R_c=\frac{K_{0U^2}}{\alpha\vartheta} \tag{10}$$

and

$$u(y,0)=0, \ u(0.t)=\sin(t), \ u(\infty,t)<\infty \tag{11}$$

Solution of Equation

Solving equation (9), we obtain

$$V=constant$$

For constant suction, we consider $V=-V_0$ (12)

where the negative sign indicates that the suction is towards the plate.

Hence the equation (10) reduces to

$$\frac{\partial u}{\partial t}-V_0\frac{\partial u}{\partial y}=\frac{\partial^2 u}{\partial y^2}+R_c\frac{\partial}{\partial t}\left(\frac{\partial^2 u}{\partial y^2}\right)-Mu \tag{13}$$

Equation (13) is a 3rd order differential equation due to the presence of the elastic parameter R_c. If the elastic parameter R_c would zero, we would have it as second order differential equation, and thereby the fluid reduces to Newtonian case (viscous fluid). To have the complete solution of (13), we require another boundary condition. But we have only two boundary conditions as specified above. However, we overcome this difficulty by considering the physical condition of the fluid. As, R_c, the elastic parameter is a small quantity based on vanishing memory, it is always $\ll 1$. So we can expand u in powers of R_c as

$$u=u_0+R_cu_1 \tag{14}$$

Substituting (14) in equation (13), and equating the coefficients of equal powers of R_c , and neglecting those of R_c^2, we have the following equations

$$\frac{\partial u_0}{\partial t}-V_0\frac{\partial u_0}{\partial y}=\frac{\partial^2 u_0}{\partial y^2}-Mu_0 \tag{15}$$

$$\frac{\partial u_1}{\partial t} - V_0 \frac{\partial u_1}{\partial y} = \frac{\partial^2 u_1}{\partial y^2} + \frac{\partial}{\partial t}\left(\frac{\partial^2 u_0}{\partial y^2}\right) - M u_1 \qquad (16)$$

The boundary conditions (11) now modified as

$$u_0(y,0) = 0,\ \ u_1(y,0) = 0,\ \ for\ all\ y\ ,\ \text{t=0} \qquad (17a)$$

$$u_0(\infty,t) < \infty,\quad u_1(\infty,t) < \infty\ \ ,\ t > 0 \qquad (17b)$$

$$u_0(0,t) = \sin(t),\ \ u_1(0,t) = \sin(t)\ ,\ t > 0 \qquad (17c)$$

The velocity may be decomposed into steady – state and transient components satisfying equations (15) and (16) as:

$$u_0 = u_0^s + u_0^t \quad \text{and} \quad u_1 = u_1^s + u_1^t \qquad (18)$$

The steady-state components can be derived as:

$$u_0^s = \exp(-ay/\sqrt{2})sin\left(t - by/\sqrt{2}\right) \ \text{ and } \ u_1^s = \exp(-ay/\sqrt{2})sin\left(t - by/\sqrt{}\right) \qquad (19)$$

Where $a = \sqrt{M + \sqrt{1+M^2}}$, $b = {}^1\!/_a$

The solutions (19) satisfy the boundary conditions (17b) and (17c), but not the initial condition (17a). If the transient solution satisfies the following boundary conditions

$$u_0^t(y,0)[= -e^{-ay/\sqrt{2}}sin\left(-by/\sqrt{2}\right)] = Ime^{-cy/\sqrt{2}} = u_1^t(\text{y, 0}) \qquad (19a)$$

$$u_0^t(\infty,T) = \infty = u_1^t(\infty,T) \qquad (19b)$$

$$u_0^t(0,T) = 0 = u_1^t(0,T) \qquad (19c)$$

where $c = a - ib$, then the composition of the both transient and steady – state solutions will completely satisfy eq. (13) or (15) and (16). For the transient solution, we apply Laplace Transform Technique on the transient part of (15) and (16), and on boundary conditions [19(a,b,c)]. The final results are found as –

$$u_0^t(Y,T) = Im\left[\frac{1}{2}e^{\frac{T}{2}(b^2-a^2)-\left(\frac{Y^2}{4T}+\frac{V_0^2 T}{4}\right)}\{w(z_1) - e^{-YV_0}w(z_2)\}\right] \qquad (20a)$$

$$u_1^t(Y,T) = \frac{1}{2}Im\left[\begin{array}{c}(M-1-i)\left\{e^{\frac{T}{2}(b^2-a^2)-\left(\frac{Y^2}{4T}+\frac{TV_0^2}{4}\right)}\left(w(z_3) + e^{i\frac{Yb}{\sqrt{2}}-\frac{Ya}{\sqrt{2}}}w(z_4)\right)\right\} \\ -2e^{-\frac{a}{\sqrt{2}}(Y+TV_0)}e^{-i\frac{b}{\sqrt{2}}(Y+\frac{\sqrt{2}T}{b}+V_0T)}\end{array}\right] \qquad (20b)$$

Where,

$$z_2 = \sqrt{\frac{T}{2}}b + i(\frac{Y}{2\sqrt{T}} - \frac{\sqrt{T}}{2}V_0 + \sqrt{\frac{T}{2}}a) = z_3$$

$$z_4 = -\sqrt{\frac{T}{2}}b + i\left(\frac{Y}{2\sqrt{T}} - \sqrt{\frac{T}{2}}a + \frac{\sqrt{T}}{2}V_0\right) = z_1$$

The complete steady state and transient solutions are respectively

$$u^t(Y,T) = u_0^t(Y,T) + R_c u_1^t(Y,T) \quad (21)$$

$$u^s = u_0^s(Y,T) + R_c u_0^s(Y,T) \quad (22)$$

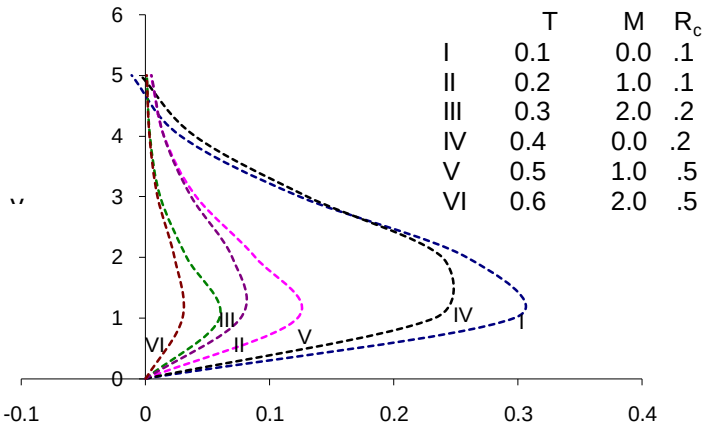


Figure1: Transient vel. distribution; Plate velocity sin (T)

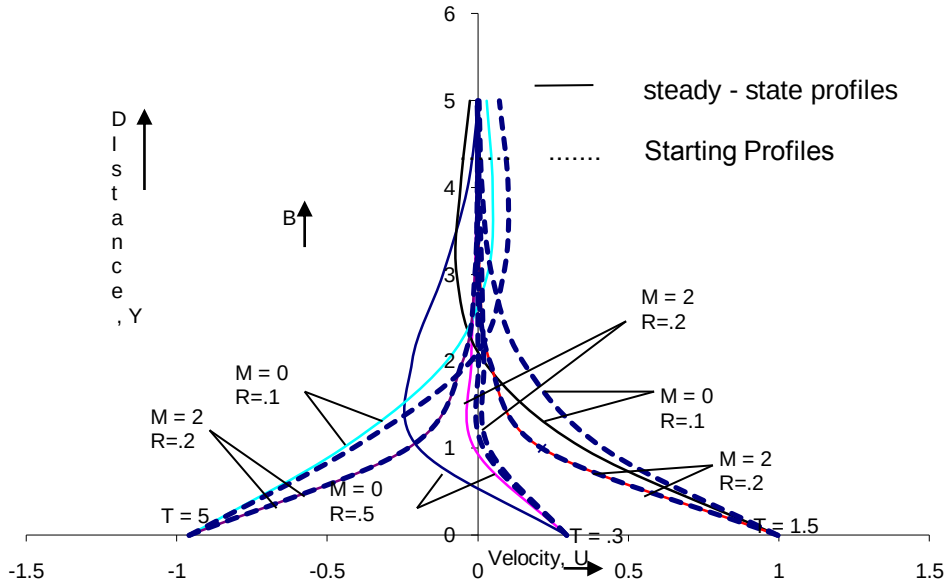


Figure. Starting Phase vel. Distribution; Plate velocity sin (T)

Conclusion

If we go through the figures presented by Panton [5] and Deka *et. Al.* [1], in their respective papers with our obtained figures, a clear difference can be seen. We see that the fluid trying penetrating towards the plate at and near the plate. However at far distance from the plate this nature cannot be seen. We feel that this effect is due to the magnetic parameter (M), the porosity (V_0) of the plate on the flow field and due to elastic property of the fluid. However, the effect of Rivlin-Ericksen fluid is very clear.

References:

1. Abramowitz B.M. and. Stegum I.A, (1964): Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, *Washington, U.S. Govt. printing*
2. Deka R.K., Das U.N. and Soundalgekar V.M.(1995): Transient free convection flow of an elastic-viscous fluid past an infinite vertical plate, *Engineering Transaction*, Vol. 43,3, 487-491
3. Puri P(1984): Impulsive motion of a flat plate in a Rivlin-Ericken fluid, *Rheological ACTA*, Vol.23:451-453
4. Kalita, B(2012): The transient for MHD Stokes's oscillating porous plate: an exact solution, *Far East Journal of Applied Mathematics*, Vol.65, No.127-36,
5. Datta, N. et. Al. (1995): Magnetohydrodynamic unsteady free convection flow and heat transfer of a visco-elastic fluid past an impulsively started porous flat plate with heat source/sinks; *Indian Journal of Theoretical Physics*, Pp. 45-59
6. Hetnarski, B. R. (1975): An algoritham for generating some Inverse Laplace Transform of exponential form, *Journal of Applied Math. Physics (ZAMP)*, Vol.26,
7. Panton R. (1968): The Transient for Stokes's oscillating plate: an exact solution in terms of tabulated functions, *Journal of Fluid Engineering*, Vol. 31 PP. 819-825.

Numerical instability of staggered electromagnetic and structural coupled analysis using time integration method with numerical damping

†*T. Niho[1], T. Horie[1], J. Uefuji[1] and D. Ishihara[1]

[1]Department of Mechanical Information Science and Technology, Kyushu Institute of Technology, Japan

*Presenting author: niho@mse.kyutech.ac.jp
†Corresponding author: niho@mse.kyutech.ac.jp

Abstract

In electromagnetic and structural coupled problems such as magnetic damping vibration, the staggered method is used for coupled analyses because of its low computational cost. However, numerical instability may occur as a result of the time lag in coupled effect evaluation even if the time integration method for each phenomenon is unconditionally stable.

In this study, the stability of staggered coupled analyses is evaluated based on the spectral radius, and the stable regions of time increments with the intensity of the coupling effect are obtained. The numerical stability of the coupled analysis methods is compared for various coupling effect intensities based on the stable region.

The coupled analysis method with the conventional serial staggered algorithm and generalized–α method is most stable. The stability of the conventional parallel staggered algorithm is much improved if the generalized–α method is used.

Keywords: Numerical instability, Electromagnetic and structural coupled analysis, Coupled algorithm, Time integration method, Numerical damping.

Introduction

The use of coupled finite element analyses such as fluid–structure interaction analysis and electromagnetic–structural coupled analysis is increasing in the design of mechanical components. Coupled finite element analysis methods are classified as simultaneous (or monolithic) and staggered (or partitioned) methods. In simultaneous methods, the coupled finite element equations are obtained by combining each finite element equation for multi-physics phenomena and then solved. However, high computational cost is incurred because the matrix size becomes large. In staggered methods, multiple finite element equations are solved separately. Because the computational cost of the staggered method is low, this method is used in many coupled analyses. However, numerical instability may occur owing to time lag in coupled effect evaluation even if the time integration method for each phenomenon is unconditionally stable.

Many studies of staggered methods have been performed for fluid–structure interaction problems. In addition to the conventional serial staggered (CSS) algorithm, which is widely used for staggered analysis, several coupled algorithms have been proposed such as the conventional parallel staggered (CPS) algorithm, improved serial staggered algorithm and improved parallel staggered algorithm; and then the numerical stability, result accuracy and computing time of these methods have been discussed[1].

Magnetic damping vibration is one type of electromagnetic and structural coupled problem. Studies have focused on magnetic damping vibration analysis, which is required for the design of conductive structures located in a strong magnetic field, such as those in future fusion reactors or magnetically levitated vehicles. Several coupled analysis methods have been compared

for magnetic damping vibration with the bending mode[2] and with the bending and torsional mode[3] from the viewpoint of the modeling, formulation, type of element, and time integration method. In the past few years, the geometrical nonlinearity of magnetic damping vibration has been discussed[4], and a coupled analysis method using a Lagrangian approach has been proposed[5]. However, numerical instability occurs in magnetic damping vibration analysis even if unconditionally stable time integration methods are used.

In this study, a stability evaluation method is proposed for the coupled finite element analysis of magnetic damping vibration. In this method, the stability is evaluated by the spectral radius obtained from the coupled eigenmode and the time integration scheme. Next, the numerical stability is examined by the stable region for various coupled analysis methods that are combined with a coupled algorithm and a time integration method with numerical damping.

Coupled Finite Element Analysis Method for Magnetic Damping Vibration Problem

Magnetic Damping Vibration

Magnetic damping vibration occurs in a conductive structure located in a magnetic field. A conductive structure is vibrated by the Lorentz force which is induced by an eddy current and a magnetic field. While the structure is vibrating, the electromotive force reduces the eddy current and vibration.

Finite Element Equations

The T method is used for eddy current analysis of the magnetic damping vibration problem of a thin shell structure[6]. The matrix equation of the eddy current analysis is expressed using the nodal point normal component $\boldsymbol{T}$ of the current vector potential and nodal point deformation vector $\mathbf{u}$:

$$\mathbf{U}\dot{\boldsymbol{T}} + \mathbf{R}\boldsymbol{T} = \mathbf{C}_\mathbf{e}\dot{\mathbf{u}} + \dot{\boldsymbol{B}}^{\boldsymbol{ex}}. \tag{1}$$

Here, $\mathbf{U}$, $\mathbf{R}$, $\mathbf{C}_\mathbf{e}$, and $\dot{\boldsymbol{B}}^{\boldsymbol{ex}}$ are the inductance matrix, the resistance matrix, the coupling sub-matrix of electromotive force, and the time-varying external magnetic field, respectively.

The matrix equation of the structural analysis is expressed by

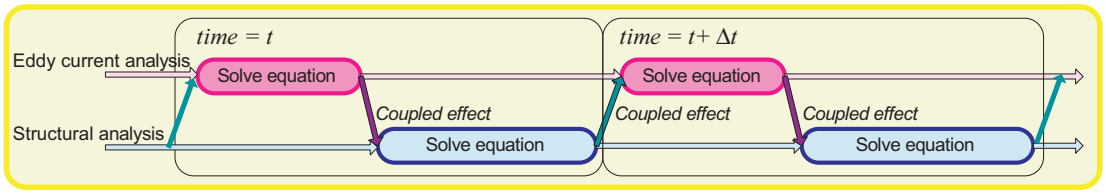
$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{C}_\mathbf{s}\boldsymbol{T} + \boldsymbol{F}^{\boldsymbol{ex}}, \tag{2}$$

where $\mathbf{M}$, $\mathbf{K}$, $\mathbf{C}_\mathbf{s}$, and $\boldsymbol{F}^{\boldsymbol{ex}}$ are the mass matrix, the stiffness matrix, the coupling sub-matrix of the Lorentz force, and the external force, respectively.

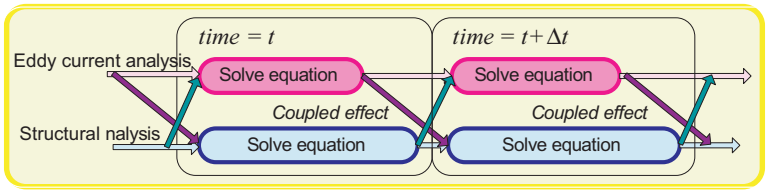
Coupled Algorithms

The coupled analysis methods for magnetic damping vibration are classified as simultaneous and staggered methods. In the simultaneous method, the coupled finite element equation obtained by combining Eqs. (1) and (2) has been solved[6] and shown to be unconditionally stable[7]. In the staggered method, Eqs. (1) and (2) are solved separately and alternately. However, it is conditionally stable even if unconditionally stable time integration methods are used for each equation because the solution diverges by numerical instability under specific conditions, for example, according to the intensity of the magnetic field and the time increment. In addition to the CSS algorithm, the CPS algorithm have been proposed for fluid–structure interaction analysis[1]. According to the previous studies, the CPS algorithm has weak stability. In this study, the numerical stability of these coupled algorithm are discussed for staggered methods of magnetic damping vibration analysis.

Fig. 1 shows the data flow between the eddy current analysis and the structural analysis using the CSS and CPS algorithms for magnetic damping vibration analysis. In the CSS algorithm,



(a) Conventional serial staggered (CSS) algorithm



(b) Conventional parallel staggered (CPS) algorithm

Figure 1: Procedures of staggered coupled algorithms for eddy current and structural coupled analyses.

Eq. (1) for the eddy current analysis is solved using the results from the previous time step of the structural analysis to evaluate the coupling term in Eq. (1). Then, Eq. (2) for the structural analysis is solved using the results of eddy current analysis to evaluate the coupling term in Eq. (2). In the CPS algorithm, Eq. (1) for the eddy current analysis and Eq. (2) for the structural analysis are solved simultaneously and separately in each time step. The terms for the coupled effect in Eqs. (1) and (2) are evaluated using the results from the previous time step.

Coupld Analysis Methods

For eddy current analysis, the backward difference method is applied. Eq. (1) becomes

$$(\mathbf{U} + \Delta t\mathbf{R})\,\boldsymbol{T}_{t+\Delta t} = \Delta t\mathbf{C}_e\dot{\mathbf{u}}_{t+\Delta t} + \mathbf{U}\boldsymbol{T}_t + \Delta t\dot{\boldsymbol{B}}_t^{ex}. \tag{3}$$

The backward difference method is unconditionally stable for uncoupled eddy current analysis.

For structural analysis, two types of time integration methods are applied. By using the parameter ρ_∞ to control the numerical dissipation, Eq. (2) becomes

$$\begin{aligned}
&\left\{(1-\alpha_m)\,\frac{1}{\beta\Delta t^2}\mathbf{M} + (1-\alpha_f)\,\mathbf{K}\right\}\mathbf{u}_{t+\Delta t}\\
=\;& (1-\alpha_f)\,\boldsymbol{F}_{t+\Delta t}^{ex} + \alpha_f\boldsymbol{F}_t^{ex}\\
+\;& \mathbf{C}_s\left\{(1-\alpha_f)\,\boldsymbol{T}_{t+\Delta t} + \alpha_f\boldsymbol{T}_t\right\}\\
-\;& \mathbf{M}\left[(1-\alpha_m)\left\{\left(1-\frac{1}{2\beta}\right)\ddot{\mathbf{u}}_t - \frac{1}{\beta\Delta t}\dot{\mathbf{u}}_t - \frac{1}{\beta\Delta t^2}\mathbf{u}_t\right\} + \alpha_m\ddot{\mathbf{u}}_t\right]\\
-\;& \alpha_f\mathbf{K}\mathbf{u}_t,
\end{aligned} \tag{4}$$

where

$$\alpha_m = \frac{2\rho_\infty - 1}{\rho_\infty + 1},\ \alpha_f = \frac{\rho_\infty}{\rho_\infty + 1},\ \delta = \frac{1}{2} - \alpha_m + \alpha_f,\ \beta = \frac{(1-\alpha_m+\alpha_f)^2}{4}$$

for Newmark's β method ($\delta = 1/2,\ \beta = 1/4$) with $\alpha_m = \alpha_f = 0$, and the asymptotic annihilation case $\rho_\infty = 0$ of the generalized–α method ($\delta = 3/2,\ \beta = 1$)[8]. These time integration methods are unconditionally stable for uncoupled structural analysis.

For the electromotive force or the coupled effect in eddy current analysis, the coupling term $\mathbf{C_e}\dot{\boldsymbol{u}}$ is evaluated under the assumption that $\dot{\boldsymbol{u}}$ is equal to $\dot{\boldsymbol{u}}_t$ for both coupled algorithms. The Lorentz force or the coupled effect in structural analysis is evaluated in a different way for each coupled algorithm. In the CSS algorithm, the coupling term for structural analysis $\mathbf{C}_s\boldsymbol{T}$ can be evaluated using $\boldsymbol{T}_{t+\Delta t}$ obtained from the eddy current analysis in the same time step. In the CPS algorithm, $\boldsymbol{T}$ is assumed to be $\boldsymbol{T}_t$ to evaluate the coupling term.

Stability Analysis Method of Magnetic Damping Vibration Analysis

The stability of a time integration method for uncoupled analysis can be generally evaluated using the spectral radius[9]. The stability of magnetic damping vibration analysis, which is one type of coupled analysis, is also evaluated using the spectral radius[10].

The stability analysis method for the combination of the vibration mode m and the eddy current mode n is described below. By ignoring the term of the external transient magnetic field in Eq. (3) for eddy current analysis and using the mode amplitude factor $\bar{u}^{(m)}$ for the vibration mode m, the mode amplitude factor $\bar{T}^{(n)}$ for the eddy current mode n is expressed as

$$\left(\bar{U}^{(n)} + \Delta t \bar{R}^{(n)}\right) \bar{T}^{(n)}_{t+\Delta t} = \Delta t \bar{C}^{(m)(n)}_e \dot{\bar{u}}^{(m)}_{t+\Delta t} + \bar{U}^{(n)} \bar{T}^{(n)}_t, \tag{5}$$

where $\bar{U}^{(n)}$ and $\bar{R}^{(n)}$ are respectively the modal inductance and modal resistance of eddy current mode n, and $\bar{C}^{(m)(n)}_e$ is the modal electromotive force of the coupling effect between vibration mode m and eddy current mode n. On the other hand, by ignoring the term of the external force in Eq. (4) for structural analysis, $\bar{u}^{(m)}$ is expressed as

$$\begin{aligned}
& \left\{ (1-\alpha_m) \frac{1}{\beta \Delta t^2} \bar{M}^{(m)} + (1-\alpha_f) \bar{K}^{(m)} \right\} \bar{u}^{(m)}_{t+\Delta t} \\
= \; & \bar{C}^{(m)(n)}_s \left\{ (1-\alpha_f) \bar{T}^{(n)}_{t+\Delta t} + \alpha_f \bar{T}^{(n)}_t \right\} \\
- \; & \bar{M}^{(m)} \left[(1-\alpha_m) \left\{ \left(1 - \frac{1}{2\beta}\right) \ddot{\bar{u}}^{(m)}_t - \frac{1}{\beta \Delta t} \dot{\bar{u}}^{(m)}_t - \frac{1}{\beta \Delta t^2} \bar{u}^{(m)}_t \right\} + \alpha_m \ddot{\bar{u}}^{(m)}_t \right] \\
- \; & \alpha_f \bar{K}^{(m)} \bar{u}^{(m)}_t,
\end{aligned} \tag{6}$$

where $\bar{M}^{(m)}$ and $\bar{K}^{(m)}$ are respectively the modal mass and the modal stiffness for vibration mode m, and $\bar{C}^{(m)(n)}_s$ is the modal Lorentz force for the coupled effect between vibration mode m and eddy current mode n.

By combining Eqs. (5) and (6) and moving terms according to the time, the recurrence equation of the magnetic damping vibration analysis becomes

$$\left\{ \ddot{\bar{u}}^{(m)}_{t+\Delta t} \; \dot{\bar{u}}^{(m)}_{t+\Delta t} \; \bar{u}^{(m)}_{t+\Delta t} \; \bar{T}^{(n)}_{t+\Delta t} \right\}^T = \mathbf{A} \left\{ \ddot{\bar{u}}^{(m)}_t \; \dot{\bar{u}}^{(m)}_t \; \bar{u}^{(m)}_t \; \bar{T}^{(n)}_t \right\}^T. \tag{7}$$

The stability of the magnetic damping vibration analysis can be evaluated using the modulus of the complex eigenvalue $|\lambda^{(m)(n)}|$, which is the spectral radius of the amplitude matrix $\mathbf{A}$. If any $|\lambda^{(m)(n)}|$ is greater than 1.0, the coupled analysis method is considered unstable.

The stability analysis method is applied to the coupled analysis method with CSS algorithm and

Newmark's β method. The eigenvalues λ of $\mathbf{A}$ are obtained from the characteristic equation

$$\lambda^4 + \left\{-2 + \frac{4\omega^2\Delta t^2}{4+\omega^2\Delta t^2} - \frac{1}{1+\phi\Delta t} - \frac{2\Delta t^2}{4+\omega^2\Delta t^2}\frac{1}{1+\phi\Delta t}\frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}\right\}\lambda^3$$
$$+\left\{1+\left(2-\frac{4\omega^2\Delta t^2}{4+\omega^2\Delta t^2}\right)\frac{1}{1+\phi\Delta t}\right\}\lambda^2$$
$$+\left\{-\frac{1}{1+\phi\Delta t}+\frac{2\Delta t^2}{4+\omega^2\Delta t^2}\frac{1}{1+\phi\Delta t}\frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}\right\}\lambda = 0, \tag{8}$$

where

$$\omega = \sqrt{\frac{\bar{K}}{\bar{M}}}, \quad \phi = \frac{\bar{R}}{\bar{U}},$$

and the superscripts $^{(m)}$ and $^{(n)}$ of the modal coefficients are omitted. The characteristic equations for other coupled analysis methods are obtained in the same way as described above.

The values of ω, ϕ, and $\frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}$ in Eq. (8) depend on the material properties, geometric configuration, and intensity of the coupled effect, so they are obtained using theoretical and finite element solutions. The value of ω is obtained from Young's modulus, mass density, length, width, and thickness of the plate by using the theoretical solution for a thin flexible plate. For the values of ϕ and $\frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}$, the characteristic equation of the magnetic damping vibration[11] is used. By combining modal Eqs. (1) and (2), the characteristic equation becomes

$$\alpha_c^3 + \phi\,\alpha_c^2 + \left(\omega^2 - \frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}\right)\alpha_c + \omega^2\phi = 0, \tag{9}$$

where α_c is coupled eigenvalue that depends on the geometry. By using the result of the eigenvalue of the coupled finite element monolithic matrix equation[11] combined with Eqs. (1) and (2), Eq. (9) becomes a complex linear equation with unknown variables ϕ and $\frac{\bar{C}_e\bar{C}_s}{\bar{M}\bar{U}}$, which are determined through this equation. Therefore, the eigenvalue of the characteristic equation Eq. (8) can be solved numerically for each Δt using the Newton method, and the stability can be evaluated using the spectral radius $|\lambda|$.

Results of stability analysis of coupled analysis methods

Magnetic Damping Vibration of Elastic Plate

The coupled analyses are performed for a magnetic damping vibration problem, as shown in Fig. 2[2]. A copper rectangular plate clamped at one end is placed in a longitudinal steady magnetic field B_x and transient magnetic field

$$B_z = 5.5\times10^{-2}\exp\frac{-t}{6.6\times10^{-3}}\ [\mathrm{T}], \tag{10}$$

that is applied perpendicularly to the plate surface. The Lorentz force produced by both the eddy current induced by B_z and B_x causes bending vibration. While the plate is vibrating, the electromotive force induced by the vibration velocity and B_x induces a coupling effect to reduce the eddy current and vibration.

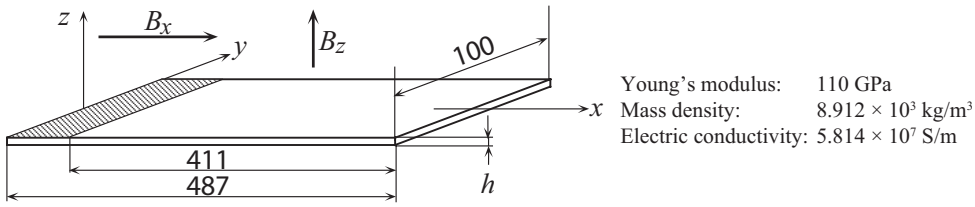


Figure 2: Schematic diagram of a clamped plate placed in electromagnetic field.

Verification of the stability analysis method

The spectral radii $|\lambda|$ obtained by the stability analysis for the coupled analysis methods using the CSS algorithm are shown in Fig. 3 for $B_x = 0.5$ T. The time increment Δt is normalized by the natural period $\tau_0 = 9.37 \times 10^{-2}$ s for the first vibration mode. The critical time increment $\Delta t_c^{(s)}$ is defined from the limit of Δt when all $|\lambda|$ values become less than or equal to 1.0. If any value of $|\lambda|$ is greater than 1.0, the coupled solution is unstable. For coupled analysis method with generalized–α method, $|\lambda|$ is always less than 1.0.

The validity of the stability analysis method should be confirmed using the coupled finite element analysis for various values of Δt, in which the staggered method for both the vibration mode response analysis and the eddy current mode response analysis is used. For coupled analysis methods with Newmark's β method, coupled finite element analyses are performed under the time increment conditions of both $\Delta t < \Delta t_c^{(s)}$ and $\Delta t > \Delta t_c^{(s)}$. For coupled analysis method with the generalized–α method, coupled finite element analysis is performed using large Δt, such as the natural period τ_0.

Fig. 4 shows the deflections at the free end of the plate. According to Fig. 4(a), the results obtained using the method with Newmark's β method is stable when $\Delta t < \Delta t_c^{(s)}$, but it is unstable when $\Delta t > \Delta t_c^{(s)}$. For the method with the generalized–α method, instability is not observed in Fig. 4(b) even if Δt is set to be as large as the natural period.

The $|\lambda|$ values obtained by the stability analysis are shown in Fig. 5 for the coupled analysis methods with the CPS algorithm. For the method with the generalized–α method, $|\lambda|$ is always less than 1.0. Fig. 6 shows the deflections of the plate obtained using the CPS algorithm. According to Fig. 6(a), the results obtained using the method with Newmark's β method is stable when $\Delta t < \Delta t_c^{(s)}$, but it is unstable when $\Delta t > \Delta t_c^{(s)}$. For the method with the generalized–α method, instability is not observed in Fig. 6(b) even if Δt is set to be as large as the natural period. Therefore, the validity of the stability evaluation method using the spectral radius is confirmed for the coupled analysis methods for the magnetic damping vibration.

Comparison of Numerical Stability

The numerical stability of the coupled analysis methods is compared for various intensities of the coupling effect. Fig. 7 shows the normalized critical time increment $\Delta t_c^{(s)}/\tau_0$ for various steady magnetic fields B_x, which is proportional to the intensity of the coupling effect. The lower left region of each curve is stable, whereas the upper right region is unstable because of the high intensity of the coupling effect. Although Newmark's β method, the generalized–α method and the backward difference method are unconditionally stable for uncoupled analysis, all coupled analysis methods using these time integration methods are conditionally stable on account of the staggered coupled analysis method. Because $\Delta t_c^{(s)}/\tau_0$ becomes smaller with increasing intensity of the coupling effect, the stability deteriorates with the coupling effect. The stable regions of the coupled analysis methods with the generalized–α method are larger than those with Newmark's β method. This is because the numerical damping of the generalized–α method may suppress the instability induced by these coupled analysis methods.

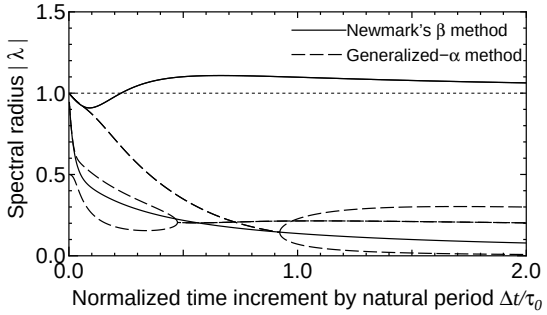


Figure 3: Results of spectral radii $|\lambda|$ of coupled analysis methods with CSS algorithm.

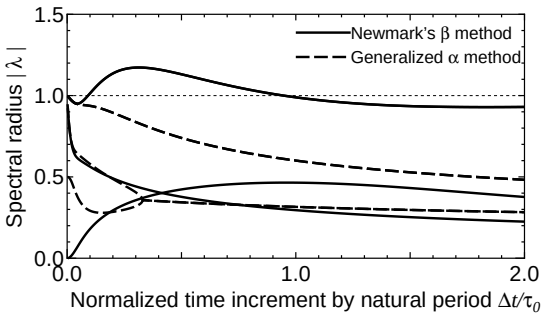
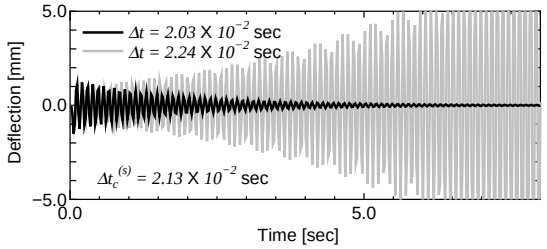
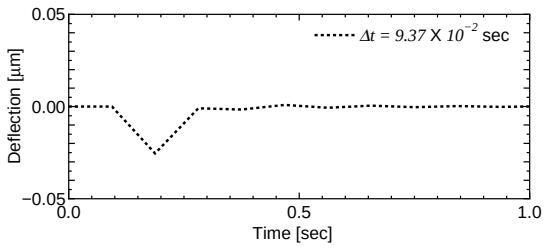


Figure 5: Results of spectral radii $|\lambda|$ of coupled analysis methods with CPS algorithm.

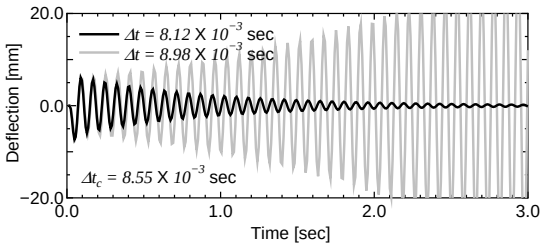


(a) Newmark's β method

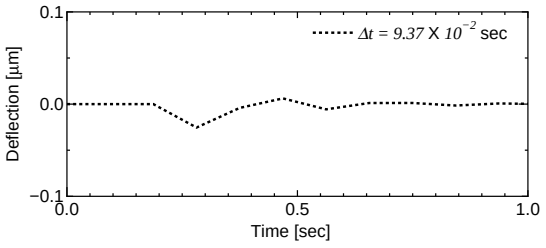


(b) Generalized–α method

Figure 4: Deflection of the plate obtained by coupled analysis methods with CSS algorithm.



(a) Newmark's β method



(b) Generalized–α method

Figure 6: Deflection of the plate obtained by coupled analysis methods with CPS algorithm.

When results for the CPS algorithm are compared with those for the CSS algorithm, the stable regions of the CPS algorithm are smaller than those of the CSS algorithm, which is the same tendency as in the fluid–structure interaction analysis with the CPS algorithm[1]. This may be because the time lag of coupled effect evaluation for the CSS algorithm treats only the electromotive force, whereas that for the CPS algorithm treats both the electromotive force and the Lorentz force. Although the stability of the CPS algorithm was worse than that of the CSS algorithm in general, it was much improved when using the generalized–α method, and this offers the advantage of a shorter computing time.

Conclusions

A stability evaluation method using the spectral radius was proposed and applied to the coupled finite element analysis of magnetic damping vibration. The stability was evaluated for coupled analysis methods that were combined with a coupled algorithm and time integration method. The validity of the stability evaluation method and the results of stability analysis were confirmed through comparisons with the results of coupled finite element analyses.

The coupled analysis method with the CSS algorithm and generalized–α method is the most

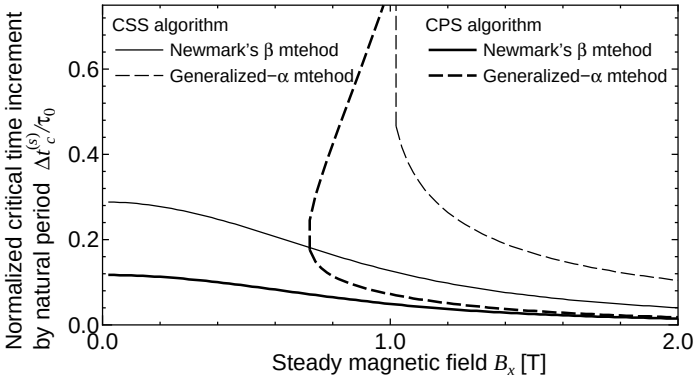


Figure 7: Stability limit of coupled analysis methods as a function of intensity of the coupled effect. Lower left region of the curve is stable region, and upper right region of the curve is unstable region.

suitable for the coupled finite element analysis of the magnetic damping problem. The generalized–α method is superior to Newmark's β method from the viewpoint of the stability of the coupled analysis. The CPS algorithm is considered inferior to the CSS algorithm in terms of numerical stability, but the stability is much improved if the generalized–α method is used. Then, the advantage of parallel computing can be better utilized when the intensity of the coupling effect is low.

References

[1] Farhat, C. and Lesoinne, M. (2000) Two efficient staggered algorithms for the serial and parallel solution of three–dimensional nonlinear transient aeroelastic problems, *Computer Methods in Applied Mechanics and Engineering* **182**, 499–515.

[2] Turner, L. R. and Hua, T. Q. (1990) Results for the cantilever beam moving in crossed magnetic fields, *COMPEL* **9**, 205–216.

[3] Takagi, T. (1995) Summary of the results for magnetic damping in torsional mode (TEAM problem 16), *COMPEL* **14**, 77–89.

[4] Zhang, J., Yan, Z., Ding, Q., Wu, H. and Pan, M. (2013) Analysis of magnetoelastic interaction of cantilever conductive thin plate with nonlinear dynamic response, *European Journal of Mechanics. A/Solids* **37**, 132–138.

[5] Li, W., Dong, H., Yuan, Z. and Chen, Z. (2014) Numerical analysis of electromagneto-mechanical coupling using Lagrangian approach and adaptive time stepping method, *International Journal of Applied Mechanics* **6**, 1450051.

[6] Horie, T. and Niho, T. (1994) Electromagnetic and mechanical interaction analysis of a thin shell structure vibrating in an electromagnetic field, *International Journal of Applied Electromagnetics in Materials* **4**, 363–368.

[7] Niho, T., Horie, T. and Tanaka, Y. (2000) Numerical instability of magnetic damping problem of elastic plate *IEEE Transactions on Magnetics* **36**, 1373–1376.

[8] Chung, J. and Hulbert, G. M (1993) A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized–α method, *Journal of Applied Mechanics* **60**, 371–375.

[9] Bathe, K. J. and Wilson, E. L. (1973) Stability and accuracy of direct time integration methods, *Earthquake Engineering and Structural Dynamics* **1**, 283–291.

[10] Tanaka, Y., Horie, T., Niho, T., Shintaku, E. and Fujimoto, Y. (2004) Stability of augmented staggered method for electromagnetic and structural coupled problem, *IEEE Transactions on Magnetics* **40**, 549–552.

[11] Horie, T., Niho, T. and Kawano, T. (1995) *Coupling intensity parameter and its dependence for magnetic damping problem*, *Electromagnetic Phenomena Applied to Technology, Enokizono, M. and Todaka, T. eds.*, JSAEM, Tokyo, Japan, 61–68.

Modeling, Computation and simulation of non-linear soft-tissue interaction with flow dynamics with application to biological systems*

Manal Badgaish and Padmanabhan Seshaiyer

Department of Mathematical Sciences, George Mason University, Fairfax VA 22030, USA.

Abstract

In this work, a computational model for the interaction of blood flow with the wall of an intracranial saccular aneurysm that is surrounded by cerebral spinal fluid is considered. The coupled fluid-structure interaction model presented includes growth and remodeling effects within the soft-tissue by incorporating elastin and collagen dynamics which are two of the main layers in the arterial wall. The resulting nonlinear system of coupled differential equations are solved numerically using implicit finite difference methods coupled with the Newton's method. The linearized version of the nonlinear system was also considered and solved both analytically using Laplace transformation and numerically using implicit finite difference methods. The nonlinear effects on rupture was studied and compared for benchmark studies and the computational results indicate that the model proposed is robust and reliable.

Keywords: Mechanics, Computation, Aneurysm, Rupture, bio-mechanics.

Introduction

Over the last three decades there has been a lot of efforts to study intracranial saccular aneurysms which are focal dilatation of the arterial wall that are found in the Circle of Willis. The specific mechanisms responsible for their genesis, enlargement, and rupture has been a prominent area of research during these years. There have been competing hypothesis in the literature on the pathogenesis and lesion development involving limit point instabilities, [12, 1, 7], equilibrium wall stress and wall strength comparisons [2] and instability of the wall in response to pulsatile blood flow [8, 18, 13, 17, 19, 11].

Intracranial Saccular aneurysm which is a soft tissue interacts with a variety of flows including blood as well as the Cerebral spinal fluid. Based on the influence of various bio-mechanical factors, the growing aneurysm can be potentially ruptured and that leads to either a neurological disorder or death. About 80% to 90% of ruptured aneurysms leads to death [21].

*Contact Emails: mbadgai2@masonlive.gmu.edu, pseshaiy@gmu.edu (Corresponding Author)

In the last two decades, several researchers have tried to investigate different aspects of biomechanics of aneurysms [14, 15, 16, 20]. Different groups of researchers had identified the elastodynamics of the arterial wall interaction with the blood flow to be the main reason for the rupture of an aneurysm [8, 18, 13]. A coupled fluid-structure model to understand the elastodynamics better was studied more extensively in the past few years [17, 19, 4, 11]. These models introduced mathematical models of increasing complexity for intracranial saccular aneurysms that described the coupled interaction between blood, arterial wall, and Cerebral Spain Fluid (CSF). In [19], the CSF was modeled using simplified Navier-Stokes equations, whereas the arterial wall structure was modeled using a spring mass system. A Fourier series was used to model the interaction between blood pressure and inner wall. While the model developed yielded good insight into understanding rupture, there was a great need to incorporate the growth and remodeling effects of the soft-tissue that will help to introduce important attributes and constituent of the arteries wall which will be the focus of this work. There are three main constituents of the artery wall, namely, the elastin, the collagen, and the smooth muscle [9, 5]. The elastin is a stable protein and is considered the most load bearing element that functions as resistance to the formation of an anuerysm, whereas the collagen is the protein that is responsible for preventing rupture after formation of an aneurysm. The growth of the aneurysm is associated with deficiency of elastin and weakening of the artery wall [6]. Hence, elastin and collagen should be incorporated into the modeling of arterial wall in order to obtain an accurate biological model of the aneurysm that can lead to better interpretation and prediction for this disease. This is one of the main contributions of this work.

In Section 2, we will describe the mathematical model that we will consider to solve a coupled fluid structure problem. Section 3 describes the implicit finite difference implementation of the coupled system. In section 4 we include some results from our computational experiments indicating the influence of collagen and elastin. Finally in section 5, we conclude and present some future work.

Mathematical Models and Background

The current work will build on models developed in [19] which helped to develop a very simple mathematical model of a thin-walled, spherical intracranial aneurysm surrounded by cerebral spinal fluid which is referred to as CSF (See Figure 1). This model involved solving coupled partial differential equations for fluids (modeling blood and cerebral spinal fluid) interacting with elastic structures modeling aneurysms using novel approaches. These models in [19, 4] were validated using analytical techniques and computational tools.

Next we describe briefly the models that were proposed which will be considered in this

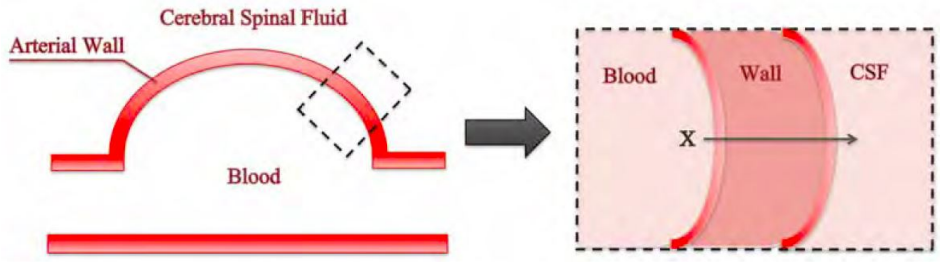


Figure 1: Model of an aneurysm in an arterial wall with blood inside and CSF outside

work and how they will be enhanced in this work using effects of growth and remodeling.

Model of the Cerebral Spinal Fluid

The model of Cerebral Spinal Fluid (CSF) considered in this paper is the simplified one dimensional Navier-stokes equation. Assuming the CSF is inviscid and slightly compressible with negligible non-linear effects, one can derive the following wave equation [19]:

$$v_t = c^2 u_{xx} \tag{1}$$

$$u_t = v \tag{2}$$

Here $u(x,t)$ is assumed to be the displacement of the CSF with $v(x,t)$ as the velocity. Since we are looking to find the movement of outer wall due to the interaction with CSF, we consider $x = 0$ to denote the outer wall (See Figure 1) and therefore we are interested in finding the solution to equations (1) and (2) at $x = 0$ that will describe the movement of the wall at any time $t \geq 0$. In order to solve the system, we will assume that the displacement and velocity of the CSF is zero initially. This is given by the initial conditions:

$$u(x,0) = v(x,0) = 0. \tag{3}$$

The boundary conditions will be described later after the discussion of the modeling of the blood pressure and the arterial wall which are discussed next.

Model of the Blood Pressure

The blood pressure is modeled using Fourier series since we consider the behavior to be pulsatile [3, 10, 17]. This relation can be described as:

$$P_B(t) = P_m + \sum_{n=1}^{N}(A_n \cos(nwt) + B_n \sin(nwt)) \tag{4}$$

where P_m is the mean blood pressure, A_n, B_n are Fourier coefficients, and w is the fundamental circular frequency [10].

Model of the Arterial Wall

We consider the arterial wall to be modeled using a simple spring-mass system that incorporates the elastin and collagen effects in the outer wall of the arteries. The force of this system maybe denoted by F_S which is given by $F_O - F_I$ where F_O and F_I are the forces of outer and inner wall respectively. This maybe expressed as:

$$F_S = K_E A_E(t)\sigma_E(\epsilon_E) + K_C A_C(t)\sigma_C(\epsilon_C) - aP_B(t) \tag{5}$$

where K_E, K_C are the scaling coefficients, $A_E(t), A_C(t)$ are the cross-sectional areas, and $\sigma_E(\epsilon_E)$, $\sigma_C(\epsilon_C)$ are the stresses for elastin and collagen respectively. These stresses are related to the respective strains through nonlinear constitutive laws given by:

$$\epsilon_E = (((L + u(0,t))/L)^2 - 1)/2 \qquad \epsilon_C = (\epsilon_E + (1 - r^2)/2)/r^2$$

where L denotes the length of the unstrained tissue, u its extension, and r is the stretched factor of unstrained tissue of collagen fiber.

Governing Equations of Motion

In order to solve the system (1)-(2), we need two boundary conditions. The first boundary condition is at point $x = 0$, and it can be derived from the model of blood pressure and the arterial wall that we have discussed. Note that the force balance equation at $x = 0$ maybe written as:

$$F_T = F_F - F_S. \tag{6}$$

where $F_T = mv_t(0,t)$ which is the inertial term corresponding to the product of mass of the wall m and acceleration, $F_F = \rho c^2 u_x(0,t)a$ is the fluid force, with a is the cross-sectional area and ρ, the density of the CSF.

Substituting equation (5) into (6) we obtain the following boundary condition at $x = 0$:

$$\begin{aligned} mv_t(0,t) = aP_m &- K_E A_E(t)\sigma_E(\epsilon_E) - K_C A_C(t)\sigma_C(\epsilon_C) + \rho c^2 a u_x(0,t) \\ &+ \sum_{n=1}^{N}(aA_n \cos(nwt) + aB_n \sin(nwt)) \end{aligned} \tag{7}$$

The second boundary condition can be obtained using the plane wave approximation that states that the waves from the wall will die down some fixed distance away from the wall. If this can be applied at point $x = L$, then the second boundary condition becomes [19]:

$$v(L,t) = -cu_x(L,t) \tag{8}$$

Combining (1), (2), (3), (7) and (8), we obtain the following system of coupled fluid-structure interaction problem:

$$\begin{aligned} v_t &= c^2 u_{xx} \\ u_t &= v \\ u(x,0) = v(x,0) &= 0 \\ mv_t(0,t) &= aP_B(t) - K_E A_E(t)\sigma_E(\epsilon_E) \\ &\qquad -K_C A_C(t)\sigma_C(\epsilon_C) + \rho c^2 a u_x(0,t) \\ v(L,t) &= -cu_x(L,t) \end{aligned} \tag{9}$$

For simplicity, we will assume that the cross-sectional areas are constant and a materially linear constitutive relationship between stress and strain is considered. In particular, we consider $A_E(t) = \gamma_E, A_C(t) = \gamma_C$, and $\sigma_E(\epsilon_E) = \epsilon_E$, $\sigma_C(\epsilon_C) = \epsilon_C$. Note that we still consider the soft-tissue to be geometrically non-linear which is the relation between the strains and the respective displacements. Given that system (9) is a coupled nonlinear system, it requires a numerical solution which will be discussed next.

An Implicit Finite Difference Solution Method

In order to solve system (9), we use an implicit finite difference method wherein we will replace the derivatives of the terms in the system by their corresponding finite difference approximations in a discretized domain. We employ the following second order finite difference approximation:

$$
\begin{aligned}
u'(y_i) &= \frac{u(y_i+\Delta y)-u(y_i-\Delta y)}{2\Delta y}+O(\Delta y^2), && \Delta y \le y_i \le Y-\Delta y \\
u''(y_i) &= \frac{u(y_i+\Delta y)-2u(y_i)+u(y_i-\Delta y)}{\Delta y^2}+O(\Delta y^2) && \Delta y \le y_i \le Y-\Delta y \\
u'(0) &= \frac{-3u(0)+4u(\Delta y)-u(2\Delta y)}{2\Delta y}+O(\Delta y^2) && (y_i=0) \\
u'(Y) &= \frac{u(Y-2\Delta y)-4u(Y-\Delta y)+3u(Y)}{2\Delta y}+O(\Delta y^2) && (y_i=Y)
\end{aligned}
$$

where $\Delta x = \dfrac{L}{M},\ \Delta t = \dfrac{tF}{N}, 0 \le x \le L$, and $0 \le t \le tF$

Then the system (9) can be rewritten implicitly as:

$$\frac{v_i^{j+1}-v_i^{j-1}}{2\Delta t} = \frac{c^2(u_{i+1}^{j+1}-2u_i^{j+1}+u_{i-1}^{j+1})}{\Delta x^2}+O(\Delta x^2,\Delta t),\ 1 \le i \le M-1 \quad (10)$$

$$\frac{u_i^{j+1}-u_i^{j-1}}{2\Delta t} = v_i^{j+1}+O(\Delta t),\ 0 \le i \le M \quad (11)$$

$$\begin{aligned}
\frac{m(v_0^{j+1}-v_0^{j-1})}{2\Delta t} &= aP_B(t(j+1))+\frac{\rho c^2 a(-3u_0^{j+1}+4u_1^{j+1}-u_2^{j+1})}{2\Delta x}-\frac{K_E\gamma_E}{L}u_0^{j+1} \\
&\quad -\frac{K_E\gamma_E}{L}(u_0^{j+1})^2-\frac{K_C\gamma_C}{Lr^2}u_0^{j+1} \quad (12) \\
&\quad -\frac{K_C\gamma_C}{2L^2r^2}(u_0^{j+1})^2-\frac{K_C\gamma_C(1-r^2)}{2r^2}+O(\Delta x^2,\Delta t)
\end{aligned}$$

$$v_M^{j+1} = \frac{-c(u_{M-2}^{j+1}-4u_{M-1}^{j+1}+3u_M^{j+1})}{2\Delta x}+O(\Delta x^2) \quad (13)$$

Rewriting this nonlinear system as $F(\mathbf{u})=\mathbf{0}$ after dropping the higher order terms we get:

$$\left(\frac{2c^2}{\Delta x^2}\right)u_i^{j+1}-\left(\frac{c^2}{\Delta x^2}\right)(u_{i-1}^{j+1}+u_{i+1}^{j+1})+\left(\frac{1}{2\Delta t}\right)v_i^{j+1}-\left(\frac{1}{2\Delta t}\right)v_i^{j-1}=0 \quad (14)$$

$$u_i^{j+1}-2\Delta t v_i^{j+1}-u_i^{j-1}=0 \quad (15)$$

$$\left(\frac{K_E\gamma_E}{2L^2}+\frac{K_C\gamma_C}{2L^2r^2}\right)(u_0^{j+1})^2+\left(\frac{3\rho c^2 a}{2\Delta x}+\frac{K_E\gamma_E}{L}+\frac{K_C\gamma_C}{Lr^2}\right)u_0^{j+1}-\left(\frac{4\rho c^2 a}{2\Delta x}\right)u_1^{j+1}+\left(\frac{\rho c^2 a}{2\Delta x}u_2^{j+1}\right)$$
$$+\left(\frac{m}{2\Delta t}v_0^{j+1}\right)-\left(\frac{m}{2\Delta t}\right)v_0^{j-1}-aP_B(t(j+1))+\frac{K_C\gamma_C(1-r^2)}{2r^2}=0 \quad (16)$$

$$cu_{M-2}^{j+1}-4cu_{M-1}^{j+1}+3cu_M^{j+1}+2\Delta x v_M^{j+1}=0 \quad (17)$$

The system can be solved at each time step $J+1$ for $J \geq 1$ using the Newton's method for solving nonlinear system:

$$\mathbf{u}^{n+1}=\mathbf{u}^n-J(\mathbf{u})^{-1}F(\mathbf{u}) \quad (18)$$

where $J(\mathbf{u})$ is the Jacobian matrix of the system, n is the Newton iteration number, and $F(\mathbf{u})$ is the system above. Here,

$$J(\mathbf{u})=\begin{bmatrix} B(\mathbf{u}) & C \\ D & E \end{bmatrix}$$

$$B(\mathbf{u})=\begin{bmatrix}
\frac{3\rho c^2 a}{2\Delta x}+\frac{K_E\gamma_E}{L}+\frac{K_C\gamma_C}{Lr^2}+\left(\frac{K_E\gamma_E}{L^2}+\frac{K_C\gamma_C}{L^2r^2}\right)u_0^{j+1} & \frac{-4\rho c^2 a}{2\Delta x} & \frac{\rho c^2 a}{2\Delta x} & 0 & \dots & 0 \\
-\frac{c^2}{\Delta x^2} & 2\frac{c^2}{\Delta x^2} & -\frac{c^2}{\Delta x^2} & 0 & \dots & 0 \\
0 & -\frac{c^2}{\Delta x^2} & 2\frac{c^2}{\Delta x^2} & -\frac{c^2}{\Delta x^2} & \dots & 0 \\
\vdots & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & \ddots & \ddots & \ddots & \\
0 & \dots & 0 & -\frac{c^2}{\Delta x^2} & 2\frac{c^2}{\Delta x^2} & -\frac{c^2}{\Delta x^2} \\
0 & \dots & 0 & c & -4c & 3c
\end{bmatrix}$$

$$C=\begin{bmatrix}
\frac{m}{2\Delta t} & 0 & 0 & & \dots & 0 \\
0 & \frac{1}{2\Delta t} & 0 & 0 & \dots & 0 \\
0 & 0 & \frac{1}{2\Delta t} & 0 & \dots & 0 \\
\vdots & & \ddots & \ddots & & \vdots \\
0 & \dots & 0 & 0 & \frac{1}{2\Delta t} & 0 \\
0 & \dots & 0 & 0 & 0 & 2\Delta x
\end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} -2\Delta t & 0 & \dots & 0 \\ 0 & -2\Delta t & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & -2\Delta t \end{bmatrix}$$

To solve using the Newton's method, we require a guess which we will use from the solution at first two time steps.

For $1 \leq i \leq M-1$,

$$\frac{v_i^1 - v_i^0}{\Delta t} = \frac{c^2(u_{i+1}^1 - 2u_i^1 + u_{i-1}^1)}{\Delta x^2} + O(\Delta x^2, \Delta t) \tag{19}$$

for $0 \leq i \leq M$,

$$\frac{u_i^1 - u_i^0}{\Delta t} = v_i^1 + O(\Delta t) \tag{20}$$

$$\frac{m(v_0^1 - v_0^0)}{\Delta t} = aP_{Blood}(t) + \frac{\rho c^2 a(-3u_0^1 + 4u_1^1 - u_2^1)}{2\Delta x} - \frac{K_E\gamma_E}{L}u_0^1 - \frac{K_E\gamma_E}{2L^2}(u_0^1)^2 - \frac{K_C\gamma_C}{Lr^2}u_0^1 - \frac{K_C\gamma_C}{2L^2r^2}(u_0^1)^2 - \frac{K_C\gamma_C(1-r^2)}{2r^2} + O(\Delta x^2, \Delta t) \tag{21}$$

$$v_M^1 = \frac{-c(u_{M-2}^1 - 4u_{M-1}^1 + 3u_M^1)}{2\Delta x} + O(\Delta x^2) \tag{22}$$

Then substituting the initial condition and drooping higher order terms, we get:

$$\left(\frac{1}{\Delta t}\right)v_i^1 - \left(\frac{c^2}{\Delta x^2}\right)u_{i+1}^1 + \left(\frac{2c^2}{\Delta x^2}\right)u_i^1 - \left(\frac{c^2}{\Delta x^2}\right)u_{i-1}^1 = 0 \tag{23}$$

for $0 \leq i \leq M$,

$$u_i^1 - \Delta t v_i^1 = 0 \tag{24}$$

$$\left(\frac{m}{\Delta t}\right)v_0^1 + \left(\frac{3\rho c^2 a}{2\Delta x} + \frac{K_E\gamma_E}{L} + \frac{K_C\gamma_C}{Lr^2}\right)u_0^1 + \left(\frac{K_E\gamma_E}{2L^2} + \frac{K_C\gamma_C}{2L^2r^2}\right)(u_0^1)^2 - \frac{4\rho c^2 a}{2\Delta x}u_1^1 + \frac{\rho c^2 a}{2\Delta x}u_2^1 + \frac{K_C\gamma_C(1-r^2)}{2r^2} - aP_{BLOOD}(t) = 0 \quad (25)$$

$$cu_{M-2}^1 - 4cu_{M-1}^1 + 3cu_M^1 + (2\Delta x)v_M^1 = 0 \quad (26)$$

Computational Experiments

In this section, we perform some computational studies to validate the numerical solution to the geometrically nonlinear model that introduces the effects of the elastin and collagen. Since this nonlinear system can only be solved numerically using nonlinear solvers, the following steps are applied in order to validate this solution. First, the nonlinear model is linearized using Taylor series expansion, and this linearized version of the model was solved both analytically using Laplace transform and numerically using implicit finite difference approximation. The behavior of numerical solution against the analytical solution was validated. After the validation, the influence of various parameters on the displacement of the wall $u(0,t)$ was investigated. Secondly, the numerical solution for the linear model is used as initial guess for the nonlinear model to solve system numerically using Newton's method with implicit finite difference approximation. Finally the influence of some parameters on the displacement of wall is also considered.

In this experiment, the following realistic values are utilized. For the CSF, $p = 1000kg/m^3$, $c = 1500m/s$ are used. For the Wall, $a = 0.01m^2$, $k_E = 800\ N/m$, $k_C = 3.52N/m$, $A_E = 20\ m^2$, $A_C = 10\ m^2$, $r = 2\ m$, and $L = 1.5m$ are used. Finally, $Pm = 8759.279403mmHg$, $w = 1rad/s$ are used for the blood pressure model, and for the harmonics, $A_1 = -7.13, A_2 = -3.08, A_3 = -0.130, A_4 = -0.205, A_5 = 0.0662, B_1 = 4.64, B_2 = -1.18, B_3 = -0.564, B_4 = -0.346, B_5 = -0.120$, all in mmHg.

First, in Figure 2, we compare the linear solution without growth and remodeling obtained in [19] in comparison to both the analytical solution obtained by linearization of coupled non-linear system with growth and remodeling (9) as well as the numerical solution to (9) obtained via the implicit finite difference method. The figure shows that the inclusion of growth and remodeling does have an effect even though the solution seems to have the same shape. Their inclusion yields a decreased displacement of the outer wall which seems to suggest that including elastin and collagen can help prevent rupture.

The Influence of length of unstrained tissues

Next, we wanted to investigate the effect of the length of the column where the CSF lives on the displacement of the outer wall. As Figure 3 illustrates, we noted that as the length

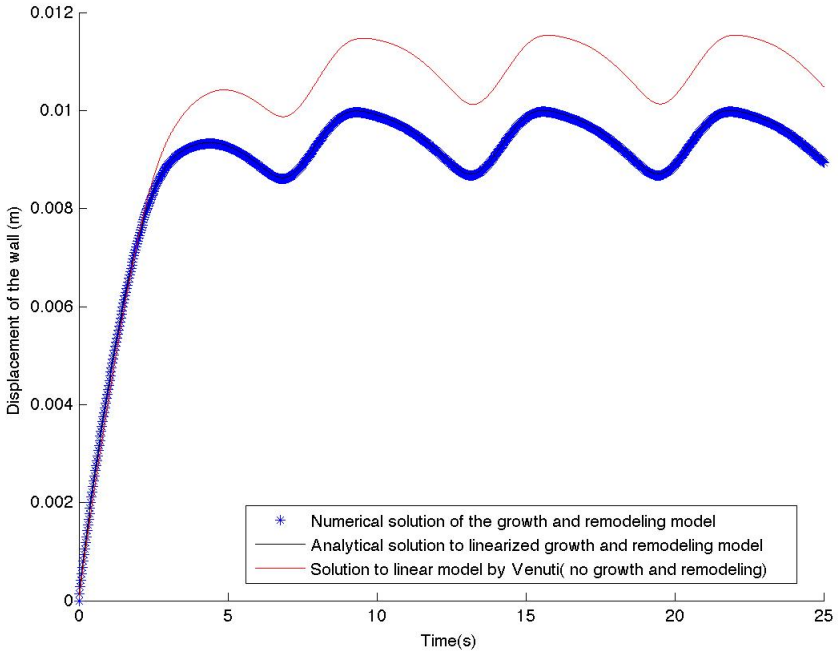


Figure 2: Nonlinear Growth and Remodeling solution VS Linear Solutions

is reduced, the movement of the wall declines dramatically. Figure 3 illustrates the motion of the wall for decreasing length from $L = 1.5\ m$ to $L = 0.1\ m$. The results seem to agree with what is expected intuitively.

Influence of Elastin and Collagen parameters

The elastic and collagen parameters (K_E,K_C) seem to play an important role in the modeling of the arterial wall since they are responsible for the elasticity and strength of wall tissue. Figure 4 shows the solution for different values of K_E starting from $300N/M$ till $800N/M$ while figure 5 represent the solution for different values of K_C starting from $1.52N/M$ till $6.52N/M$. Figure 4 suggests that the displacement increases and takes longer to stabilize into a periodic motion as K_E decreases. However, Figure 5 shows that the displacement increases in a steady periodic motion as K_C increases. Both these computational observations seem to correspond to what has been observed in the literature.

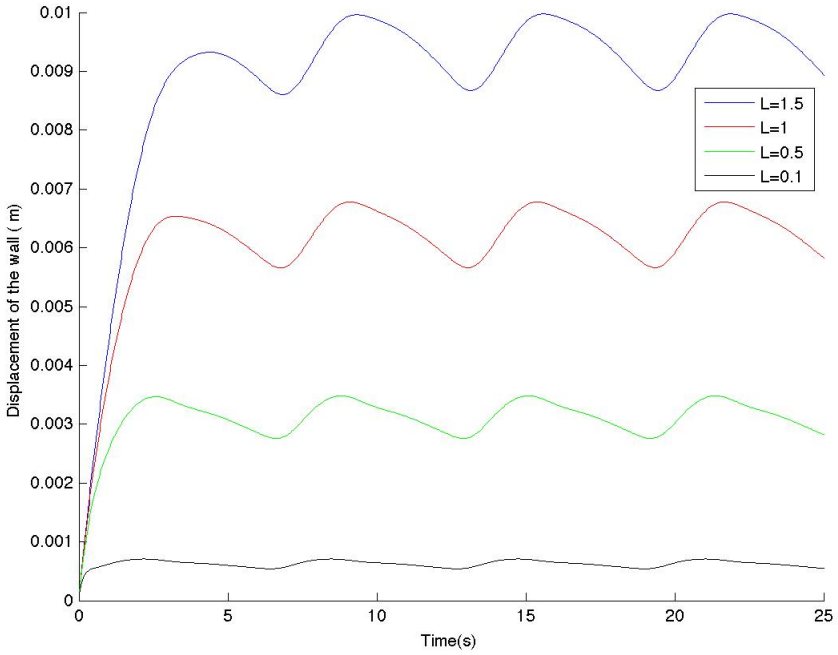


Figure 3: Influence of Length of Unstrained Tissues

Conclusions and Future work

The model developed in this work studies the influence of growth and remodeling on the rupture of an aneurysm In this model, three important components of aneurysm modeling that were considered include the blood pressure,the CSF, and arterial wall. The specific contribution of this paper was to expand on an earlier work to incorporate more relevant features of the arterial wall to stimulate the complex biological structure of the human arteries. The collagen and elastin are the most important fibers located in the wall layers that are incorporated herein in the model of the wall. This new incorporation results in a new nonlinear system that is solved numerically using implicit finite difference approximation and Newton's method for solving system of nonlinear equations. The results obtained in this work is encouraging to understand and provides a better insight into the rupture of an aneurysm. The model for the fluid considered herein is a linear model and we hope to expand our work to incorporate non-linearities in the fluid as well as develop similar models in higher dimensions which are aspects that will be considered in forthcoming papers.

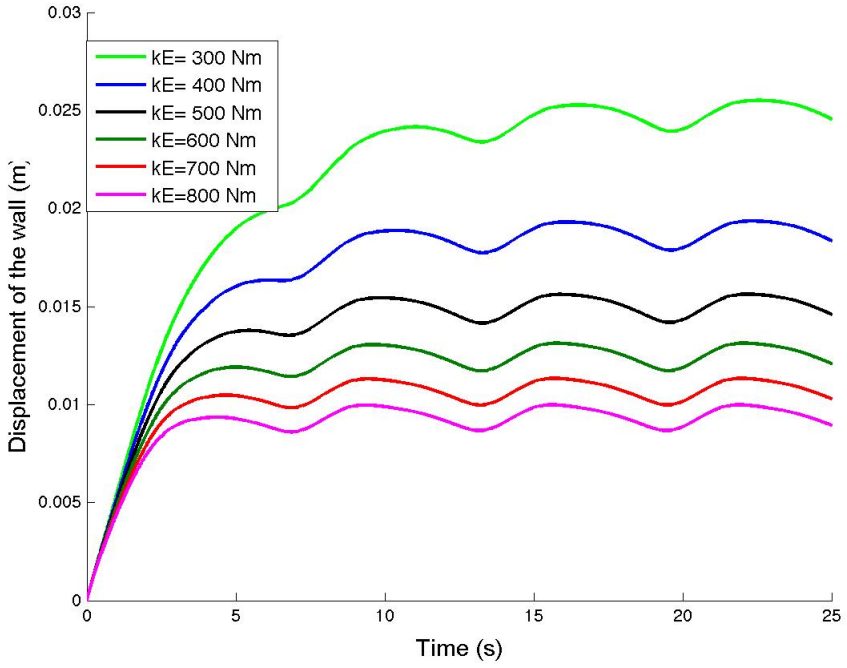


Figure 4: Influence of parameter k_E on the displacement of the outer-wall

References

[1] Akkas, N. (1990) Aneurysms as a biomechanical instability problem. *Biomechanical Transport Processes*. Plenum Press, New York, 303—311.

[2] Canham, P.B. and Ferguson, G. G. (1985) A mathematical model for the mechanics of saccular aneurysms. *Neurosurgery* **17** , 291—295, .

[3] Ferguson, G.G. (1972) Direct Measurment of Mean and Pulsatile Blood Pressure at Operation in Humman Intracranial Saccular Aneurysms. *Journal of Neurosurgery* **36**, 560—563.

[4] Foster, A., Anderson, D. , and Seshaiyer, P. (2011) Numerical Modeling and Analysis of Fluid Structure Interaction in Application to Cerebral Arteries. *GMU Review* **20**, 62—73.

[5] Fratzl, P. (2008) Collagen: Structure and Mechanics. *Springer*.

[6] He, CM. , and Roach, M. (1993) The Composition and Mechanical Properties of Abdominal Aortic Aneurysms. *J Vasc Surg* **20** (1) , 6—13.

[7] Humphrey, J.D. (1995) Arterial wall mechanics: review and directions. *Critical Reviews in Biomedical Engineering* **23**, 1—162.

[8] Jain, J.J. (1963) Mechanism of Rupture in Interacrinal Anueurysms. Surgery **54**, 347—350.

[9] Lanzer, P., and Topal, E. (2002) Pan Vascular Medicine: Integrated Clinical Management. *Springer*.

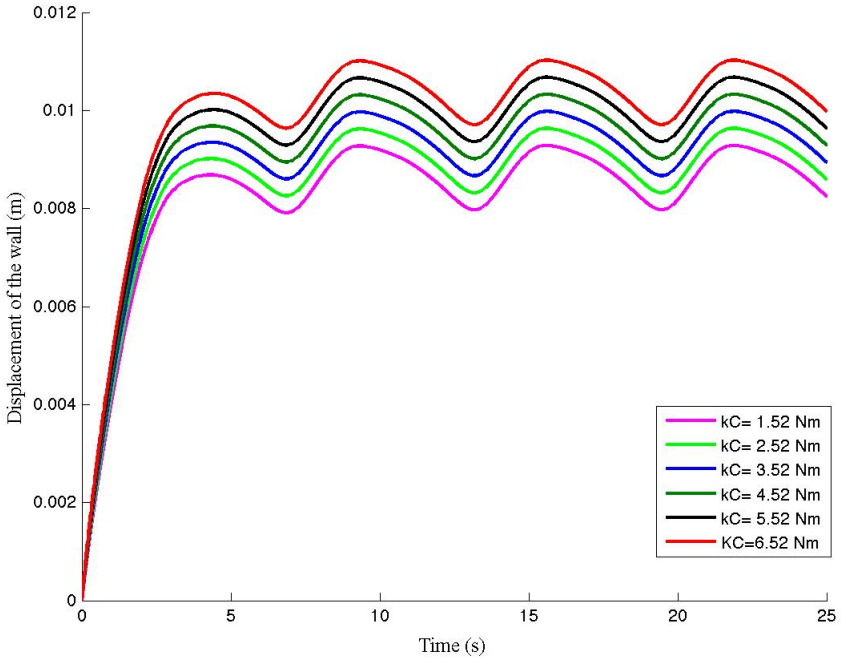


Figure 5: Influence of parameter k_C on the displacement of the outer-wall

[10] Milnor, W.R. (1982) Hemodynamics. *Williams and Wilkens,Baltimore*.

[11] Samuelson, A. , and Seshaiyer, P. (2001) Stability of Membrane Elastodynamics with applications to Cylindrical Aneurysms. *Journal of Applied Mathematics*,doi:10.1155/2011/906475.

[12] Sekhar, L.N. , and Heros, R. C. (1981) Origin, growth and rupture of saccular aneurysms: a review. *Neurosurgery* **8**, 248—260.

[13] Sekhar, L.N., Sclabassi, R.J. , Sun,M., Blue,H.B., and Wasserman, J.F. (1988) Intraaneurysmal Pressure measurements in Experimental Saccular Aneurysms in Dogs. *Stroke***19**, 352—356.

[14] Seshaiyer, P. , and Humphrey, Jay D. (2001) On the protective role of contact constraints in saccular aneurysms. *Journal of Biomechanics* **34**, 607—612.

[15] Seshaiyer, P. , Hsu,F. P. K., Shah, A. D., Kyriacou, S. K., and Humphrey, J. D. (2001) Multiaxial mechanical behavior of human saccular aneurysms. *Computer Methods in Biomechanics and Biomedical Engineering* **4**, 281—289.

[16] Seshaiyer, P. , and Humphrey, J. D. (2003) A sub-domain inverse finite element characterization of hyperelastic membranes including soft tissues. *J. of Biomech. Engg.* **125** (3),363—371.

[17] Shah, A.D., and Humphrey, J.D. (1999) Finite Strain Elastodynamics of Intracranial Saccular Aneurysms. *Journal Of Biomechanics* **32**, 593—599.

[18] Simkins, T.E., and Stehbens, W.E. (1973) Vibrational Behavior of Arterial Aneurysms. *Let-*

ters in Applied and Engineering Sciences **1**, 85—100.

[19] Venuti, S. , and Seshaiyer, P. (2010) Modeling, Analysis and Computation of Fluid-structure interaction models for biological systems. *SIAM Undergrad. Research Online* **3**, 1—17.

[20] Watton, P., Hill, N., and Heil, M. (2004) A Mathematical Model for the Growth of the Abdominal Aortic Aneurysm. *Biomech. Model. Mechanobiology* **3** (2), 98—133.

[21] Wilmink, W.B.M., Quick,C.R.G., Hubbard, C.S., and Day, N.E. (1999) The Influence of Screening on the Incidence of Ruptured Abdominal aortic Aneurysms. *J. Vasc surg* **30**(2), 203—208.

The numerical manifold method for two-dimensional transient heat conduction problems

*†H.H. Zhang, S.Y. Han, G.D. Hu, and Y.X. Tan

School of Civil Engineering and Architecture, Nanchang Hangkong University, Nanchang, Jiangxi, China

*Presenting author: hhzhang@nchu.edu.cn
†Corresponding author: hhzhang@nchu.edu.cn

Abstract

Due to the use of two cover systems, i.e., the mathematical cover system and the physical cover system, the numerical manifold method (NMM) is able to solve both continuous and discontinuous problems within the same framework. In the present paper, the NMM is developed to analyze unsteady heat conduction problems in two-dimensional settings. The NMM discrete equations are derived using the weighted residual method in Galerkin form. The spatial integration is performed through triangulation and Gauss quadrature while time integration is realized by the backward Euler scheme. The proposed approach is verified through a typical numerical example.

Keywords: Numerical manifold method (NMM), Two-dimensional heat conduction, Transient, Temperature

Introduction

In the past two decades, considerable efforts have been put on to the development of the numerical manifold method (NMM) proposed by Shi [Shi (1991)]. The outstanding performance of the NMM originates from the use of finite cover concept. Benefiting from the use of dual cover systems, that is, the mathematical cover system and the physical cover system, the NMM is able to solve both continuous and discontinuous problems in a unified framework. The major highlights of the NMM can be summarized in the following aspects: (1) the mathematical cover system can be independent of both external and internal boundaries; (2) the local property of physical field can be manifested in essence or through the proper choice of cover functions; (3) Higher-order approximation can be achieved at a fixed mathematical cover system by the use of higher-order cover functions.

Since the advent, the NMM has been applied and developed to solve various problems in many fields. Tsay et al applied the NMM to predict crack growth trajectory combined with the local remeshing technique [Tsay et al. (1999)]. Chiou et al adopted the NMM to investigate mixed mode crack propagation together with the virtual crack extension method [Chiou et al. (2002)]. Li et al developed the enriched meshless manifold method to solve two-dimensional (2D) crack problems [Li et al. (2005)]. Terada et al applied the NMM (called finite cover method therein) to analysis progressive failure processes involving cohesive zone fracture in heterogeneous solids and structures [Terada et al. (2007)]. Kurumatani and Terada extended the NMM to crack simulations for quasi-brittle heterogeneous solids by using only a regular structured mathematical mesh [Kurumatani and Terada (2009)]. Ma and his co-authors tackled 2D complex crack problems using singular physical covers in the NMM [Ma et al. (2009)], and then they further studied multiple crack propagation problems [Zhang et al. (2010)]. Zhao et al applied the NMM to consider the microstructure influence of materials in plane micropolar elasticity [Zhao et al. (2010)]. An et al introduced weak-discontinuous physical covers to describe material discontinuities within the framework of NMM [An et al. (2011)]. Zhang and Zhang computed the SIFs on polygonal mathematical elements by the NMM [Zhang and Zhang (2012)]. Wu and Wong studied the effects

of the friction and cohesion on the crack growth from a closed crack under compression with the NMM [Wu and Wong (2012)]. An et al solved 2D bimaterial interface crack problems by the NMM [An et al. (2013)]. Fan et al simulated the stress wave propagation through fracture rock with the NMM [Fan et al. (2013)]. Zhang and Ma investigated the fracture of functionally graded materials by the NMM [Zhang and Ma (2014)]. Zhang et al focused on 2D crack problems under thermo-mechanical loading [Zhang et al. (2014)]. Hu et al developed a discontinuous approach for the simulation of fluid flow in heterogeneous media by the NMM [Hu et al. (2015)]. Zheng et al proposed a mixed solution to the unconfined seepage problems with the NMM [Zheng et al. (2015)].Wang et al proposed a second-order NMM to study free surface flow containing inner drains [Wang et al. (2016)].

In the present paper, the NMM is further developed to study 2D unsteady heat conduction problems. To this end, the remaining of the paper is addressed as follows. Firstly, the governing equations and associated boundary and/or initial conditions for concerned problems are provided. Secondly, the NMM formulations for transient heat conduction analysis are derived; then, to verify the proposed method, a typical numerical example is tested. Finally, the corresponding conclusions are drawn.

Governing equations

Ignoring the heat source, the governing equations for transient heat conduction problems is [Prasad et al. (1996)]

$$\rho c \frac{\partial T(\mathbf{x},t)}{\partial t} + \nabla \mathbf{q}(\mathbf{x},t) = 0 \tag{1}$$

where ρ is the mass density and c is the specific heat at constant pressure. ∂ denotes partial derivative. $T(\mathbf{x},t)$ is the temperature with $\mathbf{x} \in \Omega$ (Ω denotes the physical domain) and t the time. The heat flux $\mathbf{q}$ is determined by the Fourier's law as $\mathbf{q} = -k\nabla T$ with k the thermal conductivity for isotropic material and ∇ the gradient operator.

The associated boundary conditions are

$$T(\mathbf{x},t) = \overline{T}(\mathbf{x},t) \qquad (\mathbf{x} \in \Gamma_T) \tag{2}$$

$$\mathbf{q}(\mathbf{x},t) \cdot \mathbf{n} = \overline{q}(\mathbf{x},t) \quad (\mathbf{x} \in \Gamma_q) \tag{3}$$

where Γ_T is the temperature boundary and Γ_q is the flux boundary. $\overline{T}$ and $\overline{q}$ are, respectively, the prescribed temperature and flux on corresponding boundary. $\mathbf{n}$ is the outward unit normal to the domain.

The initial condition for Eq. (1) is

$$T(\mathbf{x},0) = T_0 \quad (\mathbf{x} \in \Omega) \tag{4}$$

The NMM for unsteady heat conduction

A brief introduction of the NMM

In the NMM, to solve a given problem, the mathematical cover (MC) system is firstly built. Broadly speaking, the MC composed of mathematical elements can be of any shape and the MC system may be independent of all domain boundaries (including internal ones) but must be large enough to cover the whole domain. On each MC, a partition of unity (PU) [Melenk and Babuska (1996)] weight function is defined. Next, the physical cover (PC) system is formed by the intersection of MCs and physical domain. On each PC, the cover function is constructed to represent the local physical property. Then, the manifold elements (MEs) are generated through the shared region of

PCs. Accordingly, the NMM approximation on each ME is obtained by pasting the cover functions using the associated weight functions. More details about the above process can be found in the previous work [Zhang et al. (2010)].

For the present problem, the temperature in any ME *e* is approximately expressed as

$$T^h(\mathbf{x},t)=\sum_{i=1}^{n_t} w_i(\mathbf{x})T_i(\mathbf{x},t) \tag{5}$$

where n_t is the amount of PCs shared by *e*. $w_i(\mathbf{x})$ is the PU weight function defined on the MC containing the *i*th PC. $T_i(\mathbf{x},t)$ is the cover functions defined on the *i*th PC. For 2D continuous problems, $T_i(\mathbf{x},t)$ is frequently chosen as

$$T_i(\mathbf{x},t)=\mathbf{P}(\mathbf{x})\mathbf{a}_i(\mathbf{x},t) \tag{6}$$

where $\mathbf{a}_i$ is the thermal degrees of freedom (DOFs) defined on the *i*th PC. $\mathbf{P}(\mathbf{x})$ is the polynomial basis being

$$\mathbf{P}(\mathbf{x})=\begin{bmatrix}1 & x & y & \cdots\end{bmatrix} \tag{7}$$

NMM Discrete equations

The NMM discrete equations can be derived using the weighted residual method in Galerkin form [Lin (2003)]. Let $T\in H^1(\Omega)$ be the temperature trial function and $\delta T\in H^1(\Omega)$ be the corresponding test function with H^1 the first Hilbert space and δ the first order variation. A weak form of the discrete problem on a ME *e* is to find T^h in the finite dimensional subspace $V^h\in H^1(\Omega)$, $\forall \delta T^h\in V^h$ so that

$$\int_{\Omega^e}\left(\rho c\frac{\partial T}{\partial t}\delta \mathrm{T}^h+\mathbf{q}(T^h)k\cdot\mathbf{q}(\delta \mathrm{T}^h)\right)d\Omega+\lambda_T\int_{\Gamma_T^e}(T^h-\overline{T})\delta T^h d\Gamma-\int_{\Gamma_q^e}\overline{q}\delta T^h d\Gamma=0 \tag{8}$$

where λ_T is the penalty numbers adopted to enforce the essential boundary conditions due to the inconsistence of MC system with the physical boundary. Ω^e and Γ_m^e (*m* denotes *T* and *q*) are, respectively, the domain and/or boundary occupied or shared by the ME *e*.

Through Eq. (5), the test functions δT^h is expressed as

$$\delta T^h(\mathbf{x},t)=\sum_{i=1}^{n_t} w_i(\mathbf{x})\delta T_i(\mathbf{x},t) \tag{9}$$

On substituting Eqs. (5) and (9) into Eq. (8) and considering the arbitrariness of variation of DOFs, the NMM discrete equations for transient thermal conduction problems are derived as

$$\mathbf{K}_T\mathbf{T}+\mathbf{C}_T\dot{\mathbf{T}}=\mathbf{F}_T \tag{10}$$

where $\mathbf{T}$ and $\dot{\mathbf{T}}$ are, respectively, the vector of thermal DOFs and their time derivatives. $\mathbf{K}_T$, $\mathbf{C}_T$ and $\mathbf{F}_T$ are, respectively, the thermal conductivity matrix, the heat capacity matrix and the equivalent thermal load vector as

$$\mathbf{K}_T=\int_{\Omega^e}\mathbf{B}_T^{\mathrm{T}}k\mathbf{B}_T d\Omega+\lambda_T\int_{\Gamma_T^e}\mathbf{N}_T^{\mathrm{T}}\mathbf{N}_T d\Gamma \tag{11}$$

$$\mathbf{C}_T=\int_{\Omega^e}\mathbf{N}_T^{\mathrm{T}}\rho c\mathbf{N}_T d\Omega \tag{12}$$

$$\mathbf{F}_T=\lambda_T\int_{\Gamma_T^e}\mathbf{N}_T^{\mathrm{T}}\overline{T}d\Gamma-\int_{\Gamma_q^e}\mathbf{N}_T^{\mathrm{T}}\overline{q}d\Gamma \tag{13}$$

where the superscript T denotes the matrix transpose. The entries of $\mathbf{N}_T$ and $\mathbf{B}_T$ are

$$\mathbf{N}_T=\begin{bmatrix}\mathbf{N}_T^1 & \mathbf{N}_T^2 & \ldots & \mathbf{N}_T^i & \ldots & \mathbf{N}_T^{n_t}\end{bmatrix} \tag{14}$$

$$\mathbf{B}_T = \left[\mathbf{B}_T^1 \quad \mathbf{B}_T^2 \quad ... \quad \mathbf{B}_T^i \quad ... \quad \mathbf{B}_T^{n_t} \right] \tag{15}$$

with

$$\mathbf{N}_T^i = \left[w_i \mathbf{P} \right] \tag{16}$$

$$\mathbf{B}_T^i = \begin{bmatrix} (w_i \mathbf{P})_{,x} \\ (w_i \mathbf{P})_{,y} \end{bmatrix} \tag{17}$$

Numerical integration

Although the shape of MCs is user-defined, the highly developed elements in the finite element method are widely chosen. In view that the MC system can be independent of the physical domain, in this work, square elements are adopted. Further, for simplicity, the polynomial basis in Eq. (7) is set to be constant. In addition, since the shape of MEs may be diversified due to the inconsistence of MCs and physical boundary, to conveniently and accurately calculate the corresponding spatial integration in Eqs. (11) and (12), each non-triangular ME is firstly partitioned into several sub-triangles, and then the 3-point Gaussian quadrature rules are applied on each sub-triangle, the corresponding result on which finally adds up to the integration of the ME. As for the time integration, the widely used Euler backward difference method [Cebeci (2002)] is used.

Numerical examples

In this section, to verify the accuracy of the proposed method, unsteady heat conduction in an isotropic square plate is considered.

As shown Fig. 1a, the side length of the plate is L. The associated boundary and initial conditions are prescribed as

$$T(0, y, t) = T(x, 0, t) = T(L, y, t) = T(x, L, t) = 0.0 \tag{18}$$

$$T(x, y, 0) = 10 \sin(x) \sin(y) \tag{19}$$

When modeling, corresponding parameters are set as: $L = \pi$, $\rho = 1.0, c = 1.0$ and $k = 1.0$. Accordingly, the theoretical temperature solution to this problem is [Li et al. (2011)]

$$T(x, y, t) = 10 \sin(x) \sin(y) \exp(-2t) \tag{20}$$

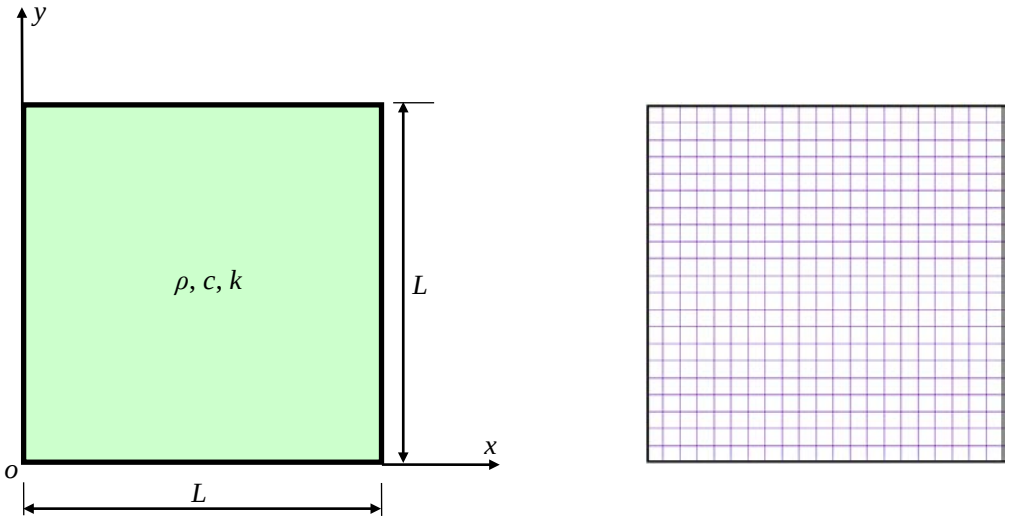


Figure 1. Transient heat conduction in a square plate: (a) physical domain and (b) discretization when *h*=0.15

In the simulation, mathematical cover system of element size (defined as the edge length of the square mathematical elements) h =0.15 is used to cover the whole plate and the associated discretized domain is illustrated in Fig. 1b, which contains 484 PCs and 441MEs. As for the time step, three values, i.e., $\Delta t = 0.1, 0.05$ and 0.02, are examined. The penalty number λ_T in Eq. (8) is taken as 1.0×10^6. The computed temperatures of two sample points A: $(\pi/4, \pi/4)$ and B: $(\pi/2, \pi/2)$ at different instants by the present method are, respectively, plotted in Fig. 2 and 3. For comparison, the exact results from Eq. (20) are also provided therein. Obviously, the temperatures at all time steps match well with the exact solution; what's more, with the decrease of time step, our results are getting closer to the analytical ones, which conforms to the convergence rule of the backward difference method.

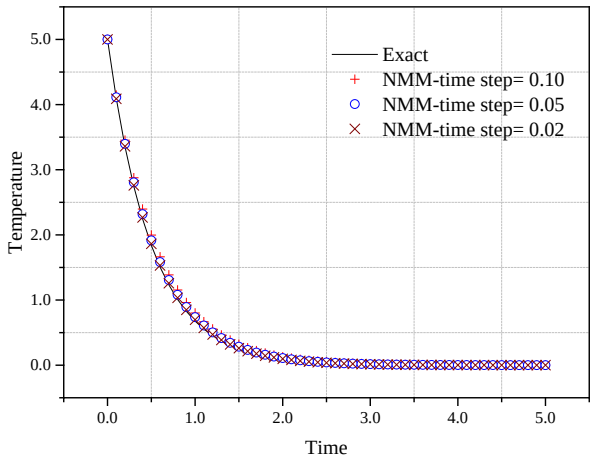


Figure 2. Computed temperatures of point A at different instants

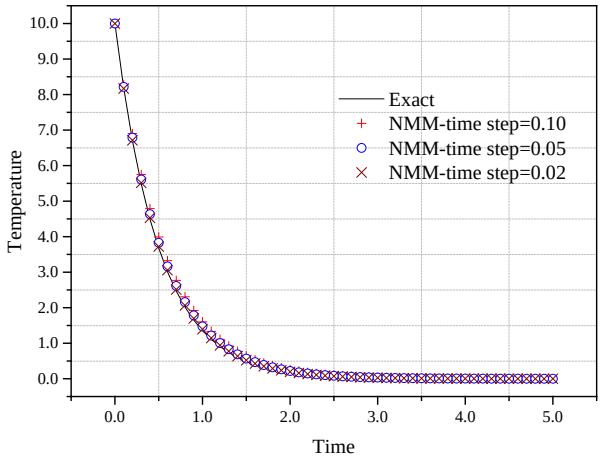


Figure 3. Computed temperatures of point B at different instants

Conclusions

In this work, the numerical manifold method has been developed to study 2D unsteady heat conduction problems. The NMM discrete equations are derived and the numerical integration schemes in the spatial and time domain are presented. A typical example is conducted to validate the proposed method. Mathematical covers formed by square elements are adopted for numerical

modeling due to the inconsistence of the mathematical cover system and physical boundaries. It's found that the accuracy of the present method is satisfactory compared with the reference solutions.

Acknowledgements

The present work was supported by the National Natural Science Foundation of China (Grant No. 11462014), the Provincial Natural Science Foundation of Jiangxi, China (Grant No. 20151BAB202003) and the Science and Technology Program of Educational Committee of Jiangxi Province of China (Grant No. GJJ14526 and GJJ150752).

References

An, X. M., Ma, G. W., Cai, Y. C. and Zhu, H. H. (2011) A new way to treat material discontinuities in the numerical manifold method, *Computer Methods in Applied Mechanics and Engineering* **200**, 3296–3308.

An, X. M., Zhao, Z. Y., Zhang, H. H. and He, L. (2013) Modeling bimaterial interface cracks using the numerical manifold method, *Engineering Analysis with Boundary Elements* **37**, 464–474.

Cebeci, T. (2002) *Convective heat transfer*, 2nd Revised edn, Horizons Publishing and Springer.

Chiou, Y. J., Lee, Y. M. and Tsay, R. J. (2002) Mixed mode fracture propagation by manifold method, *International Journal of Fracture* **114**, 327–347.

Fan, L. F., Yi, X. W. and Ma, G. W. (2013) Numerical manifold method (NMM) simulation of stress wave propagation through fractured rock, *International Journal of Applied Mechanics* **5**, 1350022.

Hu, M. S., Wang, Y. and Rutqvist, J. (2015) Development of a discontinuous approach for modeling fluid flow in heterogeneous media using the numerical manifold method, *International Journal for Numerical and Analytical Methods in Geomechanics* **39**, 1932–1952.

Kurumatani, M. and Terada, K. (2009) Finite cover method with multi-cover layers for the analysis of evolving discontinuities in heterogeneous media, *International Journal for Numerical Methods in Engineering* **79**, 1–24.

Lin, J. S. (2003) A mesh-based partition of unity method for discontinuity modeling, *Computer Methods in Applied Mechanics and Engineering* **192**, 1515–1532.

Li, Q. H., Chen, S. S. and Kou, G. X. (2011) Transient heat conduction analysis using the MLPG method and modified precise time step integration method, *Journal of Computational Physics* **230**, 2736–2750.

Li, S. C., Li, S. C. and Cheng, Y. M. (2005) Enriched meshless manifold method for two-dimensional, *Theoretical and Applied Fracture Mechanics* **44**, 234–248.

Ma, G. W., An, X. M., Zhang, H. H. and Li, L. X. (2009) Modeling complex crack problems using the numerical manifold method, *International Journal of Fracture* **156**, 21–35.

Melenk, J. M. and Babuska, I. (1996) The partition of unity finite element method: Basic theory and applications, *Computer Methods in Applied Mechanics and Engineering* **139**, 289–314.

Prasad, N. N. V., Aliabadi, M. H. and Rooke, D. P. (1996) The dual boundary element method for transient thermoelastic crack problems, *International Journal of Solids and Structures* **33**, 2695–2718.

Shi, G. H. (1991) Manifold method of material analysis, *Proceedings of the Transcations of the Ninth Army Confernece on Applied Mathematics and Computing*, 57–76.

Terada, K., Ishii, T., Kyoya, T. and Kishino, Y. (2007) Finite cover method for progressive failure with cohesive zone fracture in heterogeneous solids and structures, *Computational Mechanics* **39**, 191–210.

Tsay, R. J., Chiou, Y. J. and Chuang, W. L. (1999) Crack growth prediction by manifold method, *Journal of Engineering Mechanics-ASCE* **125**, 884–890.

Wang, Y., Hu, M. S. and Zhou, Q. L. Rutqvist, J. (2016) A new second-order numerical manifold method model with an efficient scheme for analyzing free surface flow with inner drains, *Applied Mathematical Modelling* **40**, 1427–1445.

Wu, Z. J. and Wong, L. N. Y. (2012) Frictional crack initiation and propagation analysis using the numerical manifold method, *Computers and Geotechnics* **39**, 38–53.

Zhang, H. H., Li, L. X. An, X. M. and Ma, G. W. (2010) Numerical analysis of 2-D crack propagation problems using the numerical manifold method, *Engineering Analysis with Boundary Elements* **34**, 41–50.

Zhang, H. H. and Zhang, S. Q. (2012) Extract of stress intensity factors on honeycomb elements by the numerical manifold method, *Finite Elements in Analysis and Design* **59**, 55–65.

Zhang, H. H. and Ma, G. W. (2014) Fracture modeling of isotropic functionally graded materials by the numerical manifold method, *Engineering Analysis with Boundary Elements* **38**, 61–71.

Zhang, H. H., Ma, G. W. and Ren, F. (2014) Implementation of the numerical manifold method for thermo-mechanical fracture of planar solids, *Engineering Analysis with Boundary Elements* **44**, 45–54.

Zhao, G. F., Zhao, J., Zhang, H. H. and Ma, G. W. (2010) A numerical manifold method for plane micropolar elasticity, *International Journal of Computational Methods* **7**,151–166.

Zheng, H., Liu, F. and Li, C. G. (2015) Primal mixed solution to unconfined seepage flow in porous media with numerical manifold method, *Applied Mathematical Modelling* **39**, 794–808.

Numerical analysis of optimum packing structure of particles on a spherical surface

***Takuya Uehara[1]**

[1]Department of Mechanical Systems Engineering, Yamagata University, Japan.

*Presenting author: uehara@yz.yamagata-u.ac.jp

Abstract

Numerical analyses on the particle-packing structures on a spherical surface are performed. The particles are assumed to have uniform radius and connected by inter-particle interaction. The positions of the particles are justified to minimize the interaction energy, which is assumed to be represented by two-body potential function. The number of particles N and radius ratio x of the particles to the central large sphere are varied as the model parameters. As a result, it revealed that filling all the space by the hexagonal packing on a sphere surface is impossible and some defects are needed. Also the optimized arrangement of the defects presents a regular pattern. In conclusion, the availability of the present method for the optimization of the particle packing structure is fairly validated.

Keywords: Particle packing, Optimization problem, Close packed structure, Molecular dynamics method, Computer simulation.

Introduction

Advanced materials using nano- or micro-particles have been developed and applied for various engineering fields such as electronics, biological and medical engineering, and so on. The packing structure of particles plays an important role on the functionality of the material, and the arrangement of the particles is one of the most definitive factors in the material design. In general, to set the particles artificially on designed sites is difficult, and hence self-organization process is often utilized. In such processes, however, the structure obtained is limited, and more suitable structures may exist to generate much higher performance. Therefore, we have been investigating the optimum structures of particle packing for specified purposes. Considering when arranging particles with uniform radius on a planar face, for instance, it is well known that the densest packing is achieved when the centers of particles are disposed on the regular triangular positions making regular hexagonal arrangement, which is often referred to as honeycomb arrangement. Concerning the three-dimensional structures, the closest packed structure is well known as the face-centered cubic (fcc) or hexagonal close packed (hcp) structures, which is achieved by accumulating this hexagonal plane to the perpendicular direction.

This kind of simple problem is, however, very complicated if the applied condition is varied; for instance, when the shape of particles is not sphere, when the size of particles are not identical, and when the particles are arranged on a curved surface. Since the analytical solutions for these problems are difficult to obtain, we have been approaching them using numerical analysis [1][2]. Particularly, in this study, the packing structures of small spherical particles with uniform radius on a spherical surface with relatively large radius are investigated, and the optimum structures are discussed on the basis of relation between the number of particles and the radius of the surface.

Simulation model and conditions

In this study, completely spherical particles with uniform radius r are considered. These particles are adhered on a convex surface of a sphere of radius R. The particles are assumed to

have interaction to the other particles with strong repulsion and weak attraction, and also elastic attraction on the spherical surface. The optimum arrangement of the particles is numerically searched by moving the particles. This procedure is similar to that of molecular dynamics method, and hence the algorithm is employed. The interaction between two particles are represented by Lennard-Jones type potential function, $\phi = 4\varepsilon((\sigma/r)^{12}-(\sigma/r)^{6})$, and linear spring connection with the sphere surface is assumed. The interacting forces between particles and the elastic force from the central sphere are calculated, and every particle is moved depending on the force vector. The interacting force is relaxed as repeating the motion, and finally stable arrangement is expected to be obtained.

The initial positions of particles are randomly provided, while the center of the large sphere on which particles are adhered is set as the origin of the coordinate system. Standardized dimension is employed so that the particle diameter r = 1.0, and the radius R of the large sphere are varied. L-J parameters are taken as σ = 0.893 and ε = 0.002, where the value in σ is taken as the equilibrium inter-particle distance to be 1.0. The elastic force on the surface is assumed to be $F = k\ (d - d_0)$ where d is the distance from the center of the large sphere, and $d_0 = r + R$ (see Fig. 1).

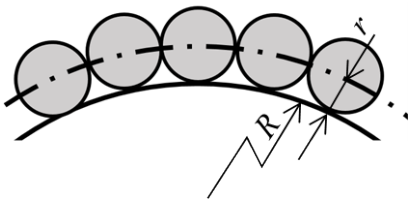


Figure 1. Particles on a large sphere surface.

Total number of particles, N, to be put on the surface is varied. Theoretical maximum number is defined for a planar problem as follows. The area of the unit hexagon in Fig. 2 is calculated as $S_{hex} = (6/\sqrt{3})\ r^2 \approx 3.46\ r^2$, and hence the maximum number of particle can be calculated as S/S_{hex}, while some influence of the considering domain area should be taken into account especially when S is relatively small. Anyway, in the spherical case, the area of the surface is represented as $S = 4\pi(R+r)^2$, and hence the optimum number of particles is derived as

$$N_{opt} = 4\pi(R+r)^2 / (\sqrt{3}\ r^2) = 3.63\ ((R+r)/r)^2 . \quad (1)$$

In the present model, r = 1.0 and hence it is represented as

$$N_{opt} = 3.63\ (R+1)^2. \quad (2)$$

When the number of particle N is given, on the other hand, the optimum radius R_{opt} is provided as

$$R_{opt} = (0.275\ N)^{1/2} - 1. \quad (3)$$

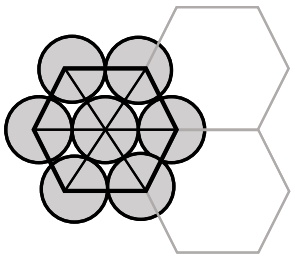


Figure 2. Hexagonal close packed structure.

These values indicate the ideal structure under hexagonal packing, but it is impossible to achieve on the spherical surface due to the effect of curvature and periodicity. Additionally, soft-core particle is assumed instead of hard sphere, and hence these values are used only as a referential one.

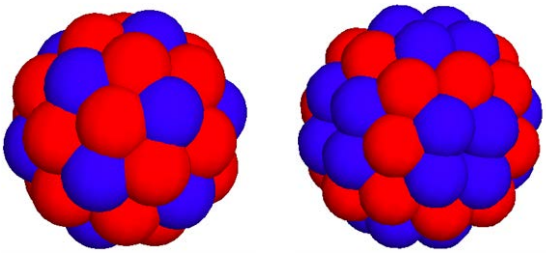
From a different viewpoint, geometric feature of regular polyhedron is helpful for predicting the particle packing structure on the spherical surface. Regular polyhedron consists of regular polygonal planes, and the surface of a sphere can be approximated to be consisting of these planes. Then the particle packing manner is similar to those in the planar problem, i.e. honeycomb arrangement on each polygon. In this context, it is not necessarily regular polyhedron, but semi-regular polyhedron consisting of two or more types of polyhedra is regarded. This procedure is not complete because the irregularity on the edges and vertexes are unavoidable, but the similarity will be worth noting.

Total number of particles N and the radius R of the sphere on which the particles are adhered are employed as the simulation parameters. The particle radius r is kept constant to be $r = 1.0$ for all cases. The simulations are carried out for a given radius R with various numbers of particles N. The initial positions of particles are set randomly, and hence several trials are demonstrated. The optimum configuration of particles are selected from all data for the given R and presented in the following sections.

Results and discussion − Case 1: relatively large particles

Firstly, the results for the cases when the radius R is relatively small, i.e. the particle radius is relatively large, are shown in this section. Figure 3 (a) shows the particle arrangement for R = 2.0 obtained by N = 32. Color indicates the number of particles in the nearest-neighbor distance, n_d, which is 6 for ideal hexagonal arrangement. In Fig. 3, blue and red represents n_d = 5 and 6, respectively. The particles are arranged regularly; every red particle is connected by three red particles and three blue ones, and every blue particle is surrounded by five red ones. The predicted optimum number for this radius is calculated as N_{opt} = 32.7 from Eq. (2), and actually the result shows the optimum structure for this radius. If taking these particles on the center of certain polyhedra, the red and blue particles are corresponding to regular hexagons and pentagons, respectively; i.e. the sphere corresponds to be approximated by a truncated icosahedron, and the number of particles N = 32 is identical to the number of plane of the truncated icosahedron. This structure is similar to that observed as fullerene C60 or well known as soccer ball pattern.

Figure 3 (b) shows the result for R = 2.5, for which N_{opt} = 44.5, and a regular pattern is observed when N = 44. The blue particles (n_d = 5) are assembled together making square arrangement, and red particles (n_d = 6) surround the squared four blue particles. Also the number of square assembly is 6, and they are disposed in the orthogonal orientation, like the Cartesian x, y and z axes. This structure seems on the basis of a cube; four particles in each 6 square face, one on each 8 vertex, and one on each 12 edge of the cube.



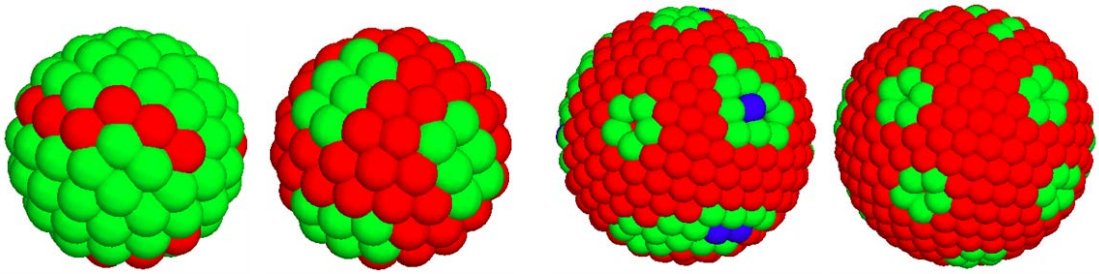
(a) R = 2.0, N = 32 (b) R = 2.5, N = 44

Figure 3. Simulation results obtained for relatively large particles.

In this way, the optimum packing structure is observed on regular polyhedron or some other regular pattern, when the relative particle radius is large.

Results and discussion − Case 2: relatively small particles

Next, in this section, the results for the case when the particle radius is relatively small and many particles are adhered on the sphere surface. Both Figs. 4 (a) shows the result for R = 4.0, for which N_{opt} = 90.8, and Figs. (i) and (ii) show the result for N = 90 and 100, respectively. In these figures, blue, green, and red particles represent n_d = 4, 5, and 6, respectively. Generally, as the number of particles becomes larger, it becomes more difficult to explore the completely optimum structure. Nevertheless, candidate structures can be found. In Fig. 4 (a)(i) for N = 90, most particles have 5 neighbors (n_d = 5), and regularity in the arrangement cannot be observed. The value of 5 in n_d indicates less density than the ideal packing, and more particles should be adhered on the surface. Then the result for N = 100 shows better results with more number of ideal density. The regularity of the particles are not perfect but it shows similar tendency to that in Fig. 3(b); six particles of n_d = 5 depicted in green color assembled together making rectangular shape, and each assembly is surrounded by red particles.

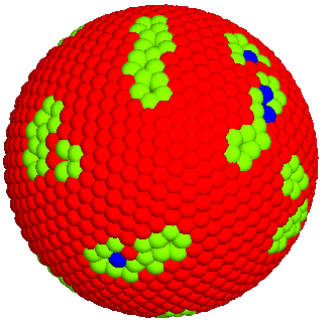


(i) N = 90 (ii) N = 100 (i) N = 294 (ii) N = 325

(a) R = 4.0 (b) R = 8.0

Figure 4. Simulation results obtained for relatively small particles.

For the case of R = 8.0 (N_{opt} = 294.0), the result for N = 294 is shown in Fig. 4(b)(i). A few particles have only four neighbors, and some vacant spaces are also observed, while the ratio of the particles of n_d = 6 is higher than that shown in Fig. 4(a)(i). This feature implies that the formation of local defects tends to be unavoidable. In this case, the better regularity was found for N = 325, as shown in Fig. 4(b)(ii). Many particles have 6 neighbors, and some particles with 5 neighbors are scattered. Characteristic feature in this case is that the particles surrounding the vacancy form pentagonal assembly of particles with n_d = 5, and the vacancy sites are regularly dispersed.



R = 16.0, N = 1120

Figure 5. Simulation results for very small particles.

As the radius of the particles becomes much smaller and the number of particles gets larger, it becomes more difficult to obtain the complete regularity and resultant optimum structure. For example, Figure 5 represents the result for R = 16.0 (N_{opt} = 1049.1) by N = 1120. Several vacancy and pentagonal arrangement can be observed, but the regularity in their arrangement was not clarified so far, and further analysis will be reported in the near future.

Conclusions

In this paper, numerical scheme for analyzing the optimum structure when small particles are adhered on a spherical surface was presented. A simple model with inter-particle interaction and adhesion with a simple elastic connection was assumed, and the effectivity was shown. In the case that the particle radius is relatively small, regular pattern was obtained. The geometrical similarity to the polyhedral structure was also found. When the particle radius is relatively small and number of particles is large, then it was difficult to find absolutely optimum structure, but it revealed that several specific local structures are formed. As a conclusion, the method presented in this paper is effective for exploring the optimum particle packing structures, and further analysis is to be continued.

References

[1] Uehara, T. (2015) Molecular dynamic analysis of particle aggregation structure on a spheric surface, *Proc. 20th Symposium on Molecular Dynamics Simulation, JSMS*, #P15 (in Japanese).

[2] Uehara, T. (2016), Numerical simulation of a domain-tessellation pattern on a spherical surface using a phase field model, *Open Journal of Modelling and Simulation* **4**, 24-33.

Semilocal convergence of a parameter based iterative method for operator with bounded second derivative

†P.Maroju[1], R.Behl[2] and S.S.Motsa[3]

[1,2,3]Department of Mathematics, Statistics and Computer science, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

†Corresponding author: maroju.prashanth@gmail.com

Abstract

A parameter based set of third order iterative method and the semilocal convergence analysis of this methods using majorizing sequence approach for solving nonlinear equations in Banach spaces is investigated by Ezquerro and Hernandez [4]. This method is a weighted mean between the Chebyshev and the Halley methods, the weight being α and $1-\alpha$, where $\alpha \in \mathbb{R}$. A convergence theorem and corresponding error bounds provided. We have recurrence relation approach to discuss the semilocal convergence of iterative methods. This is motivated us to discuss the semilocal convergence. In this paper, mainly we focus on to discuss the semilocal convergence of parameter based iterative method developed by [4] using recurrence relations approach under the assumption that F'' is bounded and a punctual condition. Also, we established the R-order of convergence and provided some a priori error bounds. Finally, we discuss some numerical examples that where the Smale-like theorem fails but our bounded condition satisfy. We calculate the existence and uniqueness region for the Numerical examples. Also, we calculate the error bounds for parameter $\alpha = 0, 1, 2$. We observed that the existence region obtained by our approach is superior than Ezquerro and Hernandez [4] for each value of parameter $\alpha = 0, 1, 2$.

Keywords: The Halley's method, The Convex acceleration of Newton's method, A Continuation method, Banach space, Lipschitz condition, Fréchet derivative.

Introduction

Let $F : \Omega \subseteq X \to Y$ be a nonlinear twice Fréchet differentiable operator in an open convex domain Ω and X, Y Banach spaces. In many years passed, one of the main problem in numerical analysis is to solve the nonlinear equation

$$F(x) = 0. \tag{1}$$

Many scientific and engineering problems, Kinetic theory of gases, elasticity, applied mathematics can be brought in the form of a nonlinear equation (1) and solved by using iterative methods. Newton in 1669 and Raphson in 1690 was proposed a procedure for solving nonlinear equation (1). Now, this method is called Newton's method or Newton-Raphson method and it is a central technique for solving nonlinear equations. The Newton's method is quadratically convergent. Basic results concerning that the semilocal convergence of Newton's method, the error estimates and the existence and uniqueness of solution are given by Kantorovich theorem. Kantrovich [9] established two different approaches to provide the proof of his theorem. Those are majorizing sequences and recurrence relations approaches.

Methods using higher order derivatives may be advantageous for special types of problems, if it is not particularly expensive to evaluate the involved derivatives in these methods. The well-known third-order methods of this type are Chebyshev, the Halley and the Super-Halley methods. These methods are of third order and can be successfully applied to solve (1). We have

three different ways to study the convergence analysis of iterative methods. In the first technique, the convergence analysis have been studied under the assumption that first/second order Fréchet derivative satisfies Lipschitz/Hölder/ω-continuity conditions. This type of convergence analysis discussed by [2][3][7] using recurrence relations approach. This technique developed by these authors is an extension of technique followed by kantorovich and other authors [9][12] to study the Newton's method. In second technique, Smale [13] obtained the convergence of Newton's method for analytic maps from data at one point instead of Lipschitz continuity condition. Another technique is to discuss the convergence of (1) assume that F'' is bounded and a punctual condition, instead of Lipschitz continuity condition. Gutierrez and Hernandez [8] discussed the convergence analysis of third order iterative method under the assumption that F'' is bounded and a punctual condition.

Continuation, embedding or homotopy methods have long served as useful theoretical tools in modern mathematics. According to the basic idea of continuation methods [10][1], a homotopy $\alpha G(x) + (1-\alpha)H(x)$, where $\alpha \in [0,1]$, can be defined between two operators $G(x)$ and $H(x)$. Prashnath and Gupta [11] studied the semilocal convergence of continuation method between the Chebyshev and the Super-Halley methods by using recurrence relations approach. J.A.Ezquerro et.al [4][5][6] discussed the convergence analysis of continuation method between different third order iterative methods namely the Chebyshev, the Halley and the Super-Halley methods using majorizing sequence approach. Based on this idea, uniparametric family of iteration between the Chebyshev and the Halley's method derived by Ezquerro and Hernandez [4] is

$$\left.\begin{array}{l} x_{\alpha,n+1} = x_{\alpha,n} - [I + \frac{1}{2}L_F(x_{\alpha,n})G_\alpha(x_{\alpha,n})]F'(x_{\alpha,n})^{-1}F(x_{\alpha,n}) \\ G_\alpha(x_{\alpha,n}) = I + \frac{\alpha}{2}L_F(x_{\alpha,n})J(x_{\alpha,n}) \\ J(x_{\alpha,n}) = (I - \frac{1}{2}L_F(x_{\alpha,n}))^{-1} \\ L_F(x_{\alpha,n}) = F'(x_{\alpha,n})^{-1}F''(x_{\alpha,n})F'(x_{\alpha,n})^{-1}F(x_{\alpha,n}). \end{array}\right\} \tag{2}$$

This method (2) is parameter based method of order three which contain both methods for specific choice of the parameter. For $\alpha = 0$ the family mentioned above reduces to the Chebyshev method and for $\alpha = 1$ we get the Halley method. Ezquerro and Hernandez [4] discussed the convergence of this method using majorizing sequence approach under the assumptions that the second order Fréchet derivative satisfies Lipschitz continuity condition. Until now, we know that convergence of these methods is established assuming that the second order derivative F'' satisfies a Lipschitz continuity condition.

The main goal of this paper is to discuss the semilocal convergence of (2) using recurrence relation approach. We assume that F'' is bounded and a punctual condition instead of Lipschitz continuity condition. An existence-uniqueness theorem is given. We have also derived a closed form of error bounds in terms of parameter $\alpha \in \mathbb{R}$. We given some numerical applications to demonstrate our approach.

We end this section briefly by describing the organization of this paper. Section 1, is the introduction. In Section 2, the recurrence relations are derived. The a convergence theorem with the existence and uniqueness ball and error estimates for the solution is established in Section 3. In Section 4, two numerical examples are worked out to demonstrate the efficacy of our approach and the results obtained are compared with the results obtained in [4]. Finally, conclusions from the section 5.

Recurrence relations for the method

Let us suppose that $\Gamma_{\alpha,0} = F'(x_{\alpha,0})^{-1} \in \mathcal{L}(X, y)$ exists at some $x_{\alpha,0} \in \Omega$, where $\mathcal{L}(X, Y)$ is the set of bounded linear operators from Y into X. Moreover, we assume that following assumptions:

$$\left.\begin{array}{l}(i)\ \|\Gamma_{\alpha,0}\| = \|F'(x_{\alpha,0})^{-1}\| \leq \beta, \\ (ii)\ \|F'(x_{\alpha,0})^{-1}F(x_{\alpha,0})\| \leq \eta, \\ (iii)\ \|F''(x)\| \leq M,\ \forall\ x \in \Omega,\end{array}\right\} \tag{3}$$

Let us denote $a = M\beta\eta$. Then for $\alpha \in \mathbb{R}$ define the following real sequences for $n = 0, 1, 2, \ldots$

$$a_0 = 1, b_0 = 1, c_0 = a, d_0 = \frac{(\alpha-1)a^2+4}{2(2-a)}$$

$$a_{n+1} = \frac{a_n}{1-aa_nd_n},\ \ b_{n+1} = \frac{aa_{n+1}d_n^2}{2}\Big[1+\frac{4+c_n(2\alpha-4)-(\alpha-1)c_n^2}{(2+c_n+(\alpha-1)c_n^2)^2}\Big]$$

$$c_{n+1} = aa_{n+1}b_{n+1},\ \ d_{n+1} = \Big(\frac{2+c_{n+1}+(\alpha-1)c_{n+1}^2}{2-c_{n+1}}\Big)b_{n+1}.$$

Let $\{x_{\alpha,n}\}$ a sequence of family. Based on these sequences, we now prove the following inequalities

(I) $\|\Gamma_{\alpha,n}\| = \|F'(x_{\alpha,n})^{-1}\| \leq a_n\beta$.

(II) $\|\Gamma_{\alpha,n}F(x_{\alpha,n})\| \leq b_n\eta$.

(III) $\|L_F(x_{\alpha,n})\| \leq c_n$.

(IV) $\|x_{\alpha,n+1} - x_{\alpha,n}\| \leq d_n\eta$.

The conditions (I), (II) and (III) for $n = 0$ hold from the assumptions (i), (ii) and

$$\|L_F(x_{\alpha,0})\| \ = \ \|F'(x_{\alpha,0})^{-1}F(x_{\alpha,0})F'(x_{\alpha,0})^{-1}F''(x_{\alpha,0}\| \leq M\beta\eta = a = c_0 < 1.$$

Using Banach Lemma, this gives

$$\|(I-\frac{1}{2}L_F(x_{\alpha,0}))^{-1}\| \leq \frac{1}{1-\frac{1}{2}\|L_F(x_{\alpha,0})\|} = \frac{1}{1-\frac{c_0}{2}} = \frac{1}{1-\frac{a}{2}} = \frac{2}{2-a}.$$

From

$$G_\alpha(x_{\alpha,0}) = I + \frac{\alpha}{2}L_F(x_{\alpha,0})J(x_{\alpha,0})$$

we get

$$\|G_\alpha(x_{\alpha,0})\| \ \leq \ 1+\frac{\alpha}{2}\|L_F(x_{\alpha,0})\|\|J(x_{\alpha,0})\| \leq \frac{2+(\alpha-1)a}{(2-a)}.$$

Using (2) and condition (II) we get

$$\|x_{\alpha,1} - x_{\alpha,0}\| \leq \Big[\frac{4+(\alpha-1)a^2}{2(2-a)}\Big]\eta \leq d_0\eta.$$

Hence, the condition (IV) also hold true for $n = 0$. Let us assume that the conditions (I)-(IV) hold true for $n = k$. To prove that they also hold true for $n = k + 1$, we use $x_{\alpha,k} \in \Omega,\ c_k < 1$ and $aa_k d_k < 1$ to get $\|I - \Gamma_{\alpha,k} F'(x_{\alpha,k})\| \leq aa_k d_k < 1$. Now, by using Banach's theorem, we find that $\Gamma_{\alpha,k+1} = F'(x_{\alpha,k+1})^{-1}$ exists and

$$\begin{aligned} \|\Gamma_{\alpha,k+1}\| &\leq \frac{\|\Gamma_{\alpha,k}\|}{1 - \|I - \Gamma_{\alpha,k} F'(x_{\alpha,k}\|} \\ &\leq \frac{a_k \beta}{1 - aa_k d_k} = a_{k+1}\beta. \end{aligned} \tag{4}$$

Now from (2),

$$\begin{aligned} F(x_{\alpha,k+1}) &= \int_0^1 [F'(x_{\alpha,k} + t(x_{\alpha,k+1} - x_{\alpha,k})) - F'(x_{\alpha,k})](x_{\alpha,k+1} - x_{\alpha,k})dt \\ &\quad -\frac{1}{2} F''(x_{\alpha,k}) F'(x_{\alpha,k})^{-1} F(x_{\alpha,k}) G_\alpha(x_{\alpha,k}) F'(x_{\alpha,k})^{-1} F(x_{\alpha,k}) \end{aligned}$$

From this,

$$\|F(x_{\alpha,k+1})\| \leq \frac{M\eta^2 d_k^2}{2} + \frac{M\eta^2 b_k^2 (2 + (\alpha - 1)c_k)}{2(2 - c_k)} \tag{5}$$

and

$$\begin{aligned} \|\Gamma_{\alpha,k+1} F(x_{\alpha,k+1})\| &\leq \|\Gamma_{\alpha,k+1}\| \|F(x_{\alpha,k+1})\| \\ &\leq a_{k+1}\beta M\eta^2 \Big[\frac{d_k^2}{2} + \frac{b_k^2(2 + (\alpha - 1)c_k)}{2(2 - c_k)}\Big] \\ &= \frac{aa_{k+1}d_k^2}{2}\Big[1 + \frac{b_k^2(2 + (\alpha - 1)c_k)}{d_k^2(2 - c_k)}\Big]\eta \\ &= \frac{aa_{k+1}d_k^2}{2}\Big[1 + \frac{4 + c_k(2\alpha - 4) - (\alpha - 1)c_k^2}{(2 + c_k + (\alpha - 1)c_k^2)^2}\Big]\eta \end{aligned}$$

This gives

$$\|\Gamma_{\alpha,k+1} F(x_{\alpha,k+1})\| \leq b_{k+1}\eta. \tag{6}$$

Also from,

$$\begin{aligned} \|L_F(x_{\alpha,k+1})\| &\leq \|F'(x_{\alpha,k+1})^{-1}\| \|F'(x_{\alpha,k+1})^{-1} F(x_{\alpha,k+1})\| \|F''(x_{\alpha,k+1})\| \\ &\leq a_{k+1}\beta b_{k+1}\eta M = M\beta\eta a_{k+1}b_{k+1} = aa_{k+1}b_{k+1} \end{aligned}$$

we get

$$\|L_F(x_{\alpha,k+1})\| \leq c_{k+1} \tag{7}$$

Again using,

$$\begin{aligned}\|x_{\alpha,k+2}-x_{\alpha,k+1}\| &\leq [1+\frac{1}{2}\|L_F(x_{\alpha,k+1})\|\|G_\alpha(x_{\alpha,k+1})]\|\Gamma_{\alpha,k+1}F(x_{\alpha,k+1})\| \\ &= \Big[\frac{2+c_{k+1}+(\alpha-1)c_{k+1}^2}{(2-c_{k+1})}\Big]b_{k+1}\eta\end{aligned}$$

we get

$$\|x_{\alpha,k+2}-x_{\alpha,k+1}\| \leq d_{k+1}\eta. \tag{8}$$

From (4),(6), (7) and (8) conclude that the conditions (I)-(IV) hold true for $n=k+1$.

Convergence Analysis

In this section, discuss the properties of real sequences and establish a convergence theorem and the existence and uniqueness region along with an estimation of the error bounds for the method (2). First at all we give a technical lemma including the results concerning one and two variable functions that we are going to need. We omit the proof to the reader could get it patiently but without any difficulty.

Lemma 1 *The following recurrence relation holds for the sequence* $\{c_n\}$.

$$c_{n+1}=\frac{c_n^2}{2}\Big[\frac{c_n^4(\alpha^2-2\alpha+1)+c_n^3(2\alpha-2)+c_n^2(3\alpha-2)+2c_n\alpha+8}{(2-3c_n-c_n^2-(\alpha-1)c_n^3)^2}\Big]$$

Lemma 2 *Let* $a_0=0.291481$ *be the smallest positive root of polynomial* $-2x^6+5x^5+8x^4-22x^3-10x^2+32x-8=0$ *and define the functions*

$$h(x)=\frac{-2-11a+10a^2+6a^3-4a^4+\sqrt{4+76a-111a^2+52a^3-8a^4}}{2(a^3-a^4)},$$

$$H(x,y)=\frac{y^4(x^2-2x+1)+x^3(2x-2)+x^2(3x-2)+2xy+8}{(2-3y-y^2-(x-1)y^3)^2},$$

$$g_\alpha(x)=\frac{(2-x)}{2-3x-x^2-(\alpha-1)x^3},$$

$$f_\alpha(x)=\frac{2+x+(\alpha-1)x^2}{(2-x)}.$$

then

(i) $h(x)$ *is a decreasing function.*

(ii) $H(x,y)$ *is increasing as a functions of* y *in* $(0,a_0]$ *and* $0\leq x\leq h(y)$.

(iii) $f_\alpha(x)$ *and* $g_\alpha(x)$ *are increasing for all* $\alpha\geq 0$.

Proof: This proof is simple then omitted for the readers.

Lemma 3 Let $0 < a \leq a_0$ and $0 \leq \alpha \leq h(a)$, then the sequence $\{c_n\}$ is decreasing.

Proof. This Lemma can be proved by induction. From Lemma 2, $c_{n+1} \leq c_n$ if

$$\frac{c_n}{2}\Big[\frac{c_n^4(\alpha^2-2\alpha+1)+c_n^3(2\alpha-2)+c_n^2(3\alpha-2)+2c_n\alpha+8}{(2-3c_n-c_n^2-(\alpha-1)c_n^3)^2}\Big] \leq 1, n \geq 0$$

for $n = 0$, we get

$$a^5(\alpha^2-2\alpha+1)+a^4(2\alpha-2)+a^3(3\alpha-2)+2a^2\alpha+8a \leq 2(2-3a-a^2-(\alpha-1)a^3)^2$$

This gives,

$$(-2a^6+a^5)\alpha^2+(4a^6-6a^5-10a^4+11a^3+2a^2)\alpha-2a^6+5a^4+8a^4-22a^3-10a^2+32a-8 \geq 0$$

This hold true for, $0 \leq \alpha \leq h(a)$. Hence $c_1 \leq c_0$. Let us assume that $c_k \leq c_{k-1} \ldots \leq c_1 \leq c_0$. Since, $h(x)$ is a decreasing function, so that $\alpha \leq h(a) = h(c_0) \leq h(c_k)$. Hence, $c_{k+1} \leq c_k$.

Lemma 4 *Under the hypothesis of Lemma ,$aa_nd_n < 1$ for $n \geq 0$ and $\{a_n\}$ is an increasing sequence.*

Proof. We have,

$$aa_nd_n = \frac{c_n(2+c_n+(\alpha-1)c_n^2)}{(2-c_n)}.$$

Then, $aa_nd_n < 1$ if $\alpha < q(c_n)$, where $q(x) = (x^3-x^2-3x+2)/x^3$. As $q(x)$ is decreasing and $c_n \leq c_0, q(c_n) \geq q(c_0)$. Besides, $\alpha < h(a)$ for $a \in (0, a_0]$. Indeed $q(a) - h(a) > 0$. Hence, $aa_nd_n < 1$ for $n \geq 0$. Finally, $a_0 = 1$, $a_1 = \frac{a_0}{1-aa_0d_0} > a_0 = 1$ and inductively, $a_{n+1} = a_n/(1-aa_nd_n) \geq a_n \geq a_{n-1} \geq \ldots \geq a_1 \geq a_0$.

Lemma 5 *Under the assumptions, $0 < a \leq a_0$ and $0 \leq \alpha \leq h(a)$. Then $c_{n+1} \leq \gamma^{2^n}\frac{c_0}{\gamma}$, where $\gamma = c_1/c_0$. Also the sequence $\{c_n\}$ converges to 0 and $\sum_{n=0}^{\infty} c_n < \infty$.*

Proof. First we prove the first part of Lemma. Let $c_1 = \gamma c_0$, with $\gamma < 1$. We prove that $c_n \leq \gamma c_{n-1}$ implies $c_{n+1} \leq \gamma^2 c_n$. From Lemma 1 we get

$$c_{n+1} = \frac{c_n^2}{2}H(\alpha, c_n) \leq \frac{\gamma^2 c_{n-1}^2}{2}H(\alpha, c_n).$$

As $H(\alpha, y)$ is increasing in the second variable and $c_n < c_{n-1}$, we get

$$c_{n+1} = \frac{c_n^2}{2}H(\alpha, c_n) \leq \gamma^2 c_n.$$

Then we have $c_{n+1} \leq \gamma^{2^n} c_n$ and using this inequality, $c_n \leq \gamma^{2^n} c_0/\gamma$. As $\gamma < 1$, the first part proved. The second part of the proof is simple and omitted for readers. Hence the Lemma is proved.

Lemma 6 *The sequence $\{a_n\}$ is bounded above, that is, there exists a constant $M > 0$ such that $a_n \leq M \quad \forall n \in \mathbb{N}$*

Proof. From $a_{n+1} = \frac{a_n}{1-aa_nd_n}$ and $g_\alpha(c_n) = \frac{(2-c_n)}{2-3c_n-c_n^2-(\alpha-1)c_n^3}$ which gives,

$$a_{n+1} = a_n\Big[1 + c_n g_\alpha(c_n)\Big] = \Pi_{k=0}^n\Big[1 + c_k g_\alpha(c_k)\Big]$$

Taking log on both sides, we get

$$\log a_{n+1} = \Sigma_{k=0}^n \log(1 + c_k g_\alpha(c_k)) \leq \Sigma_{k=0}^n c_k g_\alpha(c_k) < \infty.$$

Hence, $\{a_n\}$ is a bounded sequence.

Lemma 7 *The sequence $\{d_n\}$ is a cauchy sequence and satisfies the condition $d_n \leq \gamma^{2^n-1}d_0$ for $0 < a \leq a_0$.*

Proof. From

$$d_n = f_\alpha(c_n)\tfrac{c_n}{aa_n}, \quad \text{where, } f_\alpha(c_n) = \tfrac{2+c_n+(\alpha-1)c_n^2}{(2-c_n)}.$$

Since $a_n > 1$, so we get, $d_n \leq c_n f_\alpha(c_n)/a \leq \gamma^{2^n-1}d_0$ for $\gamma < 1$. Thus,the sequence $\{d_n\}$ converges to 0. Hence it is a cauchy sequence.

Theorem 1 *Let X and Y be two Banach spaces and let $F : \Omega \subseteq X \to Y$ be a nonlinear twice Fréchet differentiable on a non-empty open convex subset Ω. Assume that $\Gamma_{\alpha,0} = F'(x_{\alpha,0})^{-1}$ exist at some $x_{\alpha,0} \in \Omega$ and the assumptions (i)-(iii) are satisfied. Let us denote $a_0 = M\beta\eta$. Suppose that $0 < a \leq a_0 = 0.291481$ and $0 \leq \alpha \leq h(a)$, where $h(x)$ is the function defined in Lemma 1. Then, if $\overline{\mathcal{B}}(x_{\alpha,0}, r\eta) = \{x \in X : \|x - x_{\alpha,0}\| \subseteq \Omega$, where, $r = \Sigma_{n=0}^\infty d_n$, the sequence $\{x_{\alpha,n}\}$ defined in (2) and starting at $x_{\alpha,0}$ converge to a solution x^* of the equation (1). In this case the solution x^* and the iterates $x_{\alpha,n}$ lies in $\overline{\mathcal{B}}(x_{\alpha,0}, r\eta)$, and the solution x^* is unique in the open ball $B(x_{\alpha,0}, 2/M\beta - r\eta)$. Further, the error estimate of the method in terms of real sequence $\{d_n\}$ is given by*

$$\|x^* - x_{\alpha,n+1}\| \leq \sum_{k=n+1}^{\infty} d_k\eta.$$

Proof. For $0 < a < a_0$, $0 \leq \alpha < h(a)$ and using above Lemmas, the sequence $\{x_{\alpha,n}\}$ converge to the solution. For $\alpha = h(a)$, we have $c_n = c_0 = a$, for $n \geq 0$, From

$$a_{n+1} = \frac{a_n}{1 - aa_nd_n}$$

and

$$d_n = \Big[\frac{2 + c_n + (\alpha-1)c_n^2}{2 - c_n}\Big]\frac{c_n}{aa_n}$$

we get

$$a_{n+1} = a_n\Big[1 + \frac{(2-c_0)}{2 - 3c_0 - c_0^2 - (\alpha-1)c_0^3}\Big]$$

Taking, $w = \left[1 + \frac{(2-c_0)}{2-3c_0-c_0^2-(\alpha-1)c_0^3}\right]$. This can be written as $a_{n+1} = wa_n = w^{n+1}a_0$. Since $a_0 = 1$, this gives $a_{n+1} = w^{n+1}$ and

$$\begin{aligned} d_n &= \left[\frac{2+c_n+(\alpha-1)c_n^2}{2-c_n}\right]\frac{c_n}{aa_n} \\ &= \left[\frac{2+c_0+(\alpha-1)c_0^2}{2-c_0}\right]\frac{c_0}{aa_0} \\ &= \frac{1}{w^n}\left[\frac{2+c_0+(\alpha-1)c_0^2}{2-c_0}\right]\frac{c_0}{aa_0} \end{aligned}$$

Hence, $\lim_{n\to\infty} d_n = 0$. Thus, $\{d_n\}$ is a cauchy sequence. From condition (IV), we get $\{x_{\alpha,n}\}$ is also a cauchy sequence and hence there exists a x^* such that $\lim_{n\to\infty} x_{\alpha,n} = x^*$. Now from the equation (5) , we get

$$\|F(x_{\alpha,n+1})\| \leq \frac{M\eta^2}{2}\left[d_n^2 + \frac{b_n^2(2+(\alpha-1)c_n)}{2-c_n}\right], \tag{9}$$

the limit of the sequence $\{b_n\}$ and $\{d_n\}$ is 0 and the continuity of F, we prove that $F(x^*) = 0$. Thus, x^* is a solution of equation (1). Also

$$\begin{aligned} \|x_{\alpha,n+1} - x_0\| &\leq \|x_{\alpha,n+1} - x_{\alpha,n}\| + \|x_{\alpha,n} - x_{\alpha,n-1}\| + ... + \|x_{\alpha,1} - x_{\alpha,0}\| \\ &\leq \sum_{k=0}^{n} d_k\eta \\ &\leq r\eta \end{aligned}$$

This gives $x_{\alpha,n} \in \overline{\mathcal{B}}(x_{\alpha,0}, r\eta)$. Now taking limit as $n \to \infty$, we get $\|x^* - x_{\alpha,0}\| \leq r\eta$ and hence $x^* \in \overline{\mathcal{B}}(x_{\alpha,0}, r\eta)$. Also for every $m \geq n+1$, we get

$$\begin{aligned} \|x_{\alpha,m} - x_{\alpha,n+1}\| &\leq \|x_{\alpha,m} - x_{\alpha,m-1}\| + \|x_{\alpha,m-1} - x_{\alpha,m-2}\| + ... + \|x_{\alpha,n+2} - x_{\alpha,n+1}\| \\ &\leq \sum_{k=n+1}^{\infty} d_k\eta < r\eta \end{aligned}$$

by taking $m \to \infty$, we get $\|x^* - x_{\alpha,n+1}\| \leq \sum_{k=n+1}^{\infty} d_k\eta < r\eta$.

To prove the uniqueness of the solution, if y^* be the another solution of (1) then we have

$$0 = F(y^*) - F(x^*) = \int_0^1 F'(x^* + t(y^* - x^*))dt(y^* - x^*)$$

Clearly, $y^* = x^*$, if $\int_0^1 F'(x^* + t(y^* - x^*))dt$ is invertible. This follows from

$$\begin{aligned} \|\Gamma_{\alpha,0}\|\| \int_0^1 [F'(x^* + t(y^* - x^*)) - F'(x_{\alpha,0}]dt\| &\leq M\beta \int_0^1 \|x^* + t(y^* - x^*) - x_{\alpha,0}\|dt \\ &\leq M\beta \int_0^1 (1-t)\|x^* - x_{\alpha,0}\| + t\|y^* - x_{\alpha,0}\|dt \\ &\leq \frac{M\beta}{2}(r\eta + \frac{2}{k_1\beta} - r\eta) = 1 \end{aligned}$$

and by Banach's theorem. Thus, $y^* = x^*$.

Numerical Examples

Example 1 *Consider the function* $F(x) = 0$*, where,*

$$F(x) = 9x^{7/3} + 4x^2 - 36x + 9, \tag{10}$$

defined in $X = [-1, 1]$ *and initial approximation* $x_0 = 0$

Solution: From this, we observed that $F^{(k)}(x)$ does not defined at x_0 for $k \geq 3$. So Smale-like condition do not work. Hence, Using the assumptions (i)-(iii) for the initial value $x_0 = 0$, we get $\beta = 1/36$, $\eta = 1/4$, and $M = 36$. Hence, $a = M\beta\eta = 0.25 < a_0$ and we can take the real sequences defined in (2) for $0 \leq \alpha \leq h(a) = 2.12382$. We calculate the real sequences for $\alpha = 0$, $\alpha = 1$ and $\alpha = 2$ displayed in following Tables.

Table-1 : Real sequences for $\alpha = 0$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.00000	1.00000	0.25000	1.12500	1.12500
1	1.39130	0.360978	0.125558	0.406302	1.53130
2	1.62029	0.0598264	0.024234	0.0612762	1.59258
3	1.66153	0.0015232	0.00063271	0.00152416	1.59410
4	1.66258	9.64964e-007	4.01083e-007	9.64964e-007	1.59410
5	1.66258	3.87031e-013	1.60868e-013	3.87031e-013	1.59410
6	1.66258	6.22607e-026	2.58784e-026	6.22607e-026	1.59410
7	1.66258	1.6112e-051	6.6969e-052	1.6112e-051	1.59410
8	1.66258	1.07901e-102	4.48484e-103	1.07901e-102	1.59410
9	1.66258	4.83917e-205	2.01138e-205	4.83917e-205	1.59410
10	1.66258	0.	0.	0.	1.59410

Table-2 : Real sequences for $\alpha = 1$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.	1.	0.25000	1.14286	1.14286
1	1.40000	0.386596	0.135309	0.442702	1.58556
2	1.6567	0.0737824	0.0305588	0.0760721	1.66163
3	1.71059	0.00241948	0.00103469	0.00242198	1.66406
4	1.71237	2.50924e-006	1.07419e-006	2.50924e-006	1.66406
5	1.71237	2.6954e-012	1.15388e-012	2.6954e-012	1.66406
6	1.71237	3.11017e-024	1.33144e-024	3.11017e-024	1.66406
7	1.71237	4.141e-048	1.77273e-048	4.141e-048	1.66406
8	1.71059	7.34087e-096	3.14257e-096	7.34087e-096	1.66406
9	1.71237	2.30692e-191	9.87574e-192	2.30692e-191	1.66406
10	1.71237	0.	0.	0.	1.66406

Table-3 : Real sequences for $\alpha = 2$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.	1.	0.25	1.16071	1.16071
1	1.40881	0.411943	0.145087	0.48106	1.64177
2	1.69619	0.0906748	0.0384504	0.0942979	1.73607
3	1.76684	0.00385091	0.00170098	0.00385747	1.73993
4	1.76986	6.57829e-006	2.91066e-006	6.5783e-006	1.73993
5	1.76986	1.91473e-011	8.472e-012	1.91473e-011	1.73993
6	1.76986	1.62216e-022	7.17749e-023	1.62216e-022	1.73993
7	1.76986	1.1643e-044	5.15163e-045	1.1643e-044	1.73993
8	1.76986	5.99805e-089	2.65393e-089	5.99805e-089	1.73993
9	1.76986	1.59184e-177	7.04334e-178	1.59184e-177	1.73993
10	1.76986	0.	0.	0.	1.73993

From Table-1 for $\alpha = 0$ we get $r = \sum d_n = 1.59410$. So the existence and uniqueness solution of (10) are $\overline{\mathcal{B}}(x_{0,0}, 0.398525) \subseteq \Omega$, $\mathcal{B}(x_{0,0}, 1.60148) \bigcap \Omega$. From Table-2 for $\alpha = 1$ we get $r = \sum d_n = 1.66406$. So the existence and uniqueness solution of (10)respectively are $\overline{\mathcal{B}}(x_{1,0}, 0.416015) \subseteq \Omega$, $\mathcal{B}(x_{1,0}, 1.58399) \bigcap \Omega$. From Table-3 for $\alpha = 2$ we get $r = \sum d_n = 1.73993$. So the solution of (10) exists in $\overline{\mathcal{B}}(x_{2,0}, 0.434983) \subseteq \Omega$ and unique in $\mathcal{B}(x_{2,0}, 1.56502) \bigcap \Omega$. However, solving (10) by using majorizing sequence [4], for $\alpha \in (-15, 2)$ we find that the solution exists in the ball $\overline{\mathcal{B}}(x_{\alpha,0}, 0.292893) \subseteq \Omega$ and unique in $\mathcal{B}(x_{\alpha,0}, 1.70711) \bigcap \Omega$. From this result, we can easily conclude that our existence region of solution is greater than the existence region obtained by majorizing sequences. Also, we calculated error bounds by our approach and with majorizing sequence approach [4] given in Table-4.

Table-4:Error bounds for $\alpha = 0$ and $\alpha = 1$

n	$\alpha = 0$	$\alpha = 1$	$\alpha = 0$ by [4]	$\alpha = 1$ by [4]
0	0.11727600	0.13030000	0.292893	0.292893
1	0.01570000	0.01962400	0.0144311	0.00717893
2	0.00038100	0.00060600	2.91523e-6	1.82202e-007
3	2.41241e-007	6.27312e-007	2.47752e-017	3.02432e-021
4	9.67577e-014	6.7385e-013	1.52073e-050	1.3831e-062
5	1.55652e-026	7.77542e-025	3.5169e-150	1.32291e-186
6	4.02801e-052	1.03525e-048	4.3498e-449	1.157605e-558

Example 2 *Let $X = C[0, 1]$ be the space of all continuous functions on the interval $[0, 1]$ and consider the H-equation called integral equation of Chandrasekhar*

$$F(x)(s) = 1 - x(s) + \frac{1}{4}x(s)\int_0^1 \frac{s}{s+t}x(t)dt \tag{11}$$

If we choose $x_0 = x_0(s) = s$ and the norm $\|x\| = \max_{s\in[0,1]} |x(s)|$. Then we get, $M = 0.3465, \beta = 1.5304$ and $\eta = 0.2652$. Hence, we get $a = M\beta\eta = 0.1406312 < a_0$. Also, we can take the real sequence (2) for $0 \leq \alpha \leq h(a) = 46.1089$. The real sequences for $\alpha = 0$, $\alpha = 1$ and $\alpha = 2$ is given in following Table-5, Table-6 and Table-7. For $\alpha = 0$, from the Table-5 the solution of (11) exists in the ball $\overline{\mathcal{B}}(x_{0,0}, 0.330869)$ and is unique in the ball $\mathcal{B}(x_{0,0}, 3.4407)$. For $\alpha = 1$, from the Table-6 the solution of (11) exists in the ball $\overline{\mathcal{B}}(x_{1,0}, 0.33407)$ and is unique in the ball $\mathcal{B}(x_{1,0}, 3.4375)$. For $\alpha = 2$, from the Table-7 the

solution of (11) exists in the ball $\overline{\mathcal{B}}(x_{2,0}, 0.337268)$ and is unique in the ball $\mathcal{B}(x_{2,0}, 3.4343)$. However, solving (11) by using majorizing sequence [4], for $\alpha \in (-15, 2$ we find that the solution exists in the ball $\overline{\mathcal{B}}(x_{\alpha,0}, 0.287047) \subseteq \Omega$ and is unique in $\mathcal{B}(x_0, 3.48452)$. From this result, we can easily conclude that our existence region of solution is greater than the existence region obtained by majorizing sequences.

Table-5 : Real sequences for $\alpha = 0$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.00000	1.00000	0.140631	1.07032	1.07032
1	1.17719	0.167709	0.0277641	0.172365	1.24268
2	1.21177	0.00492798	0.000839789	0.00493212	1.24762
3	1.21279	4.14542e-006	7.07026e-007	4.14543e-006	1.24762
4	1.21279	2.93093e-012	4.99887e-013	2.93093e-012	1.24762
5	1.21279	1.46513e-024	2.49887e-025	1.46513e-024	1.24762
6	1.21279	3.66117e-049	6.24435e-050	3.66117e-049	1.24762
7	1.21279	2.28616e-098	3.89919e-099	2.28616e-098	1.24762
8	1.21279	8.91419e-197	1.52037e-197	8.91419e-197	1.24762
9	1.21279	0.	0.	0.	1.24762

Table-6 : Real sequences for $\alpha = 1$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.000000	1.000000	0.140631	1.07563	1.07563
1	1.17823	0.173644	0.028772	0.178713	1.25434
2	1.21418	0.00533859	0.000911574	0.00534346	1.25969
3	1.21529	4.87651e-006	8.33434e-007	4.87652e-006	1.25969
4	1.21529	4.06426e-012	6.94615e-013	4.06426e-012	1.25969
5	1.21529	2.8231e-024	4.8249e-025	2.8231e-024	1.25969
6	1.21529	1.36212e-048	2.32796e-049	1.36212e-048	1.25969
7	1.21529	3.17095e-097	5.41941e-098	3.17095e-097	1.25969
8	1.21529	1.71847e-194	2.937e-195	1.71847e-194	1.25969
9	1.21529	0.	0.	0.	1.25969

Table-7 : Real sequences for $\alpha = 2$

n	a_n	b_n	c_n	d_n	$\sum d_n$
0	1.	1.	0.140631	1.08095	1.08095
1	1.17927	0.179515	0.029771	0.185021	1.26597
2	1.2166	0.00576852	0.0009869	0.005774	1.27175
3	1.2178	5.70729e-006	9.77435e-007	5.7073e-006	1.27175
4	1.2178	5.57852e-012	9.55382e-013	5.57852e-012	1.27175
5	1.2178	5.32961e-024	9.12754e-025	5.32961e-024	1.27175
6	1.2178	4.86462e-048	8.3312e-049	4.86462e-048	1.27175
7	1.2178	4.05281e-096	6.94088e-097	4.05281e-096	1.27175
8	1.2178	2.81301e-192	4.81759e-193	2.81301e-192	1.27175
9	1.2178	0.	0.	0.	1.27175

Conclusions

In this paper, we discussed the semilocal convergence of parameter based iterative method under the assumption that second order Fréchet derivative satisfies bounded condition instead of Lipschitz continuity condition. The analysis discussed using recurrence relation approach. Based

on this approach, the existence and uniqueness region with priori error bounds established. Finally, Numerical examples are worked out to demonstrate our approach. we observed that our approach have more superior error bounds than the other approach [4].

References

[1] Allgower , E.L.,Georg, K. (1990) Numerical Continuation Methods, an Introduction, *Springer-Verlag, New York*.

[2] Candela, V, Valencia, A. M. (1990) Recurrence Relation for Rational Cubic Methods I: The Halley Method, *Computing* **44**, 169-184.

[3] Candela, V, Valencia, A. M. (1990) Recurrence Relation for Rational Cubic Methods II: The Chebyshev Method, *Computing* **45**, 355-367.

[4] Ezquerro, J.A, Hernández, M. A. (1999) On a class of iteration containing the chebyshev and the Halley methods, *Publ.Math.Debrecen* **54**, 403-415.

[5] Ezquerro, J.A, Gutiérrez, J.M., Hernández, M. A. (1997) A Construction Procedure of Iterative Methods with Cubical Convergence, *Journal of Applied Mathematics and Computation* **85**, 181-199.

[6] Ezquerro, J.A, Hernández, M. A. (2003) A New class of third order methods in Banach spaces, *Journal of Applied Mathematics and Computation* **31**, 181-199.

[7] Gutiérrez, J.M., Hernández, M. A. (1998) Recurrence Relations for the Super-Halley Method, *Journal of Computer Math.Applications* **36**, 1-8.

[8] Hernández, M. A.,Gutiérrez, J.M. (1997) Third-order iterative methods for operators with bounded second derivative, *Journal of Computational and Applied Mathematics* **82**, 171-183.

[9] Kantorovich , L.V.,Akilov, G.P. (1982) Functional Analysis. *Pergamon Press, Oxford.*

[10] Ortega , J.M.,Rheinboldt.W.C (1970) Iterative solution of nonlinear equations in several variables, *Academic Press*.

[11] Prashnath, M, Gupta, D. K. (2013) A Continuation method and its convergence for solving nonlinear equations in banach spaces, *International Journal of Computational Methods* **10**.

[12] Rall , L.B (1979) Computational Solution of Nonlinear Operator Equations, *Krieger, New York*.

[13] Smale,S. (1986) Newton's method estimates from data at one point, *Proc. Conf. in Honor of Gail Young*.

Efficient family of sixth-order iterative methods for nonlinear models which require only one inverse Jacobian matrix

†R. Behl[1], P. Maroju[2] and S.S. Motsa[3]

[1,2,3]Department of Mathematics, Statistics and Computer science, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

†Corresponding author: ramanbehl87@yahoo.in

Abstract

In this study, we design a new efficient families of sixth-order iterative methods for solving scalar as well as system of nonlinear equations. The main beauty of the proposed family is that we have to calculate only one inverse of the Jacobian matrix in the case of nonlinear system which reduce the computational cost. The convergence properties are fully investigated along with two main theorems describing their order of convergence. In addition, we also presented a numerical work which confirm the order of convergence of the proposed family is well deduced for scalar as well as system of nonlinear equations. Further, we have also shown the the implementation of the proposed techniques on real world problems like, Van der Pol equation, Hammerstein integral equation, etc.

Keywords: Nonlinear equations and systems, iterative methods, Newton's method, order of convergence.

Introduction

Construction of higher-order multi-point iterative methods which provide the accurate and efficient approximate solution to the form of

$$F(x) = 0, \tag{1}$$

(where $F : I \subset \mathbb{R}^n \to \mathbb{R}^n$ is a univariate function when $n = 1$ or multivariate function when $n > 1$ on an open domain I.) is one of the most basic and important problem of the numerical analysis.

The reason behind the importance of this topic is the applicability of these iterative methods in the real world and applied science problems. In the literature, we can find several examples where we can see the applicability of these iterative methods to the real world problems and nonlinear models can be transformed in to the system of nonlinear equations. For example, More presented the set nonlinear model like variational inequalities, the Bratu problem, a shallow arch, etc. in his paper [17]. However, most of them are pharased in the terms of system of nonlinear equations of the form (1). Recently, Rangan et al. [23] discussed the applicability of the nonlinear system on the problem of investigating coarse-grained dynamical properties of neuronal networks in kinetic theory. In addition, Nejat and Ollivier-Gooch [18] presented the problem to study the effect of discretization order on preconditioning and convergence of high-order Newton-Krylor unstructured flow solver in computational fluid dynamics. On the other hand, Grosan and Abraham [11], also shown the applicability of the system of nonlinear equations in neurophysiology, kinematics syntheses problem, chemical equilibrium problem, combustion problem and economics modeling problem. Very recently, Awawdeh [3] and Tsoulos and Stavrakoudis [29], solved the reactor and steering problems by phrasing them in the

system of nonlinear equations. Moreover, Lin et al. [16] also discussed the applicability of the system of nonlinear equations in transport theory.

There are two main ways to develop new iterative methods for system of nonlinear equations. Firstly, researchers proposed new iterative methods in order to approximate the zeros of univariate function. Then, they tried to extend the same scheme to the multidimensional case preserving the same order of convergence. For example, Cordero et al. [5], proposed the extension of the classical fourth-order Jarratt's method [13] for scalar equations to system of nonlinear equations. In addition, Abad et al. [1], Cordero et al. [6], Ren et al. [24] and Wang et al. [30], proposed some higher-order extension for systems of nonlinear equations of the previously published work for the scalar equations. Moreover, Sharma and Arora [25] and Hueso et al. [12], also proposed the extension of higher-order Jarratt like method for scalar equation to nonlinear system. We can say that it is one of simple way to develop new scheme for system of nonlinear equations. But, it is not always possible to retain the same order of convergence and the same form of body structure. One of the main reason behind this is that in the case of scalar functional evaluation of the involved function and its derivative consume the same computational cost. However, this is not true in the multidimensional case.

Secondly, researchers tried some other approaches and procedures to develop new and higher-order methods for system of nonlinear equations. In 2010, Sharma et al. [26] proposed fourth and six-order iterative methods based on weighted-Newton iteration. On the other hand, Artidiello et al. [2] proposed fourth-order methods based on the weight function approach. Moreover, Noor et al. [19] also presented several higher-order iterative methods for system of nonlinear with the aid of decomposition technique. We can also use the different approaches like quadrature formulae, Adomian polynomial, divided difference approach, etc. for constructing iterative schemes to solve nonlinear systems. For the details of the other approaches one refers some standard text books [20, 22, 28].

In the earlier proposed schemes by some scholars like Ren et al. [24], Alicia et al. [5], Sharama and Arora [25], Noor et al. [19], Artidiello et al. [2] and Hueso et al. [12], required the evaluation of more than one inverse Jacobian matrix. It is not an easy task to find the inverse of the complicated Jacobian matrix because it requires a lot of computational work. Therefore, we need the higher-order families of iterative methods which require only one evaluation of the Jacobian matrix. Because, it will be very beneficial from the computational point of view.

The principal aim of this study is to propose a new efficient family of sixth-order iterative methods which required only one inverse of the Jacobian matrix for the system of nonlinear equations. Therefore, we propose firstly a new family of sixth-order iterative methods for a scalar equation. Then, we extend this family for the multidimensional case preserving the same order of convergence. The convergence behavior of the proposed methods is tested on a concrete variety of nonlinear equations with same initial guess as other scholars mentioned in their own papers (for the more details please see the section 4). Further, we observed that our proposed methods perform better than the existing ones. Further, we have also shown the applicability of our proposed schemes in the multidimensional case on some real world problems like, Van der pol equation, Hammerstein integral equations and etc.

Development of the scheme for scalar equations

In this section, we propose a new sixth-order family of iterative methods, which is defined as follows:

$$\begin{aligned} y_n &= x_n - \frac{2}{3}\frac{f(x_n)}{f'(x_n)}, \\ z_n &= x_n - \left[\theta_1 + \theta_2\frac{f'(y_n)}{f'(x_n)} + \theta_3\left(\frac{f'(y_n)}{f'(x_n)}\right)^2\right]\frac{f(x_n)}{f'(x_n)}, \\ x_{n+1} &= z_n - \left[\theta_4 + \theta_5\frac{f'(y_n)}{f'(x_n)} + \theta_6\left(\frac{f'(y_n)}{f'(x_n)}\right)^2\right]\frac{f(z_n)}{f'(x_n)}, \end{aligned} \tag{2}$$

where $\theta_i \in \mathbb{R},\ i = 1,\ 2,\ \ldots, 6$ are free disposable parameters. The following result demonstrates that the order of convergence reaches sixth-order with some conditions on the disposable parameters.

Theorem 1 *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be a sufficiently differentiable function in an interval D containing a simple root α of the equation $f(x) = 0$. Further, we also assume that an initial guess x_0 is sufficiently close to α. Then, the family of iterative methods* (2) *reaches a sixth-order convergence when*

$$\theta_1 = \theta_3 + \frac{7}{4},\ \theta_2 = -2\theta_3 - \frac{3}{4},\ \theta_3 = \frac{9}{8},\ \theta_4 = 1 - \theta_5 - \theta_6,\ \theta_5 = -2\theta_6 - \frac{3}{2}, \tag{3}$$

where $\theta_6 \in \mathbb{R}$, is a free disposable parameter.

Proof. Let us assume that $e_n = x_n - \alpha$ be the error in the n^{th} iteration. Further, let us also expand the functions $f(x_n)$ and it's first order derivative $f'(x_n)$ around the point $x = \alpha$ by using Taylor's series expansion with the assumption $f'(\alpha) \neq 0$, which are defined as follows:

$$f(x_n) = f'(\alpha)\left(e_n + c_2e_n^2 + c_3e_n^3 + c_4e_n^4 + c_5e_n^5 + c_6e_n^6 + O(e_n^7)\right), \tag{4}$$

where $c_k = \frac{f^{(k)}(\alpha)}{k!f'(\alpha)}$ for $k = 2,\ 3,\ \ldots$ and

$$f'(x_n) = f'(\alpha)\left(1 + 2c_2e_n + 3c_3e_n^2 + 4c_4e_n^3 + 5c_5e_n^4 + 6c_6e_n^5 + O(e_n^7)\right), \tag{5}$$

respectively.
With the aid of the expressions (4) and (5), we get

$$\begin{aligned} \frac{f(x_n)}{f'(x_n)} &= e_n - c_2e_n^2 + 2(c_2^2 - c_3)e_n^3 - (4c_2^3 - 7c_3c_2 + 3c_4)e_n^4 + (8c_2^4 - 20c_3c_2^2 + 10c_4c_2 + 6c_3^2 \\ &\quad - 4c_5)e_n^5 + \left(52c_3c_2^3 - 16c_2^5 - 28c_4c_2^2 + \left(13c_5 - 33c_3^2\right)c_2 + 17c_3c_4 - 5c_6\right)e_n^6 + O(e_n^7). \end{aligned} \tag{6}$$

By inserting the above expression (6) in the first sub step of scheme (2), we further obtain

$$\begin{aligned} y_n - \alpha &= \frac{1}{3}e_n + \frac{1}{3}c_2e_n^2 - \frac{4}{3}(c_2^2 - c_3)e_n^3 + \frac{2}{3}(4c_2^3 - 7c_3c_2 + 3c_4)e_n^4 - \frac{4}{3}\Big(4c_2^4 - 10c_3c_2^2 + 5c_4c_2 + 3c_3^2 \\ &\quad - 2c_5\Big)e_n^5 + \frac{2}{3}\left(16c_2^5 - 52c_3c_2^3 + 28c_4c_2^2 + \left(33c_3^2 - 13c_5\right)c_2 - 17c_3c_4 + 5c_6\right)e_n^6 + O(e_n^7). \end{aligned} \tag{7}$$

Now, we expand the Taylor series expansion of the function $f'(y_n) = f'\left(x_n - \frac{2}{3}\frac{f(x_n)}{f'(x_n)}\right)$ about the point $x = \alpha$ by using (6), which is given as follows:

$$f'(y_n) = f'(\alpha)\left[1 + \frac{2c_2e_n}{3} + \frac{1}{3}\left(4c_2^2 + c_3\right)e_n^2 + \sum_{i=1}^{4} P_i e_n^{i+2} + O(e_n^7)\right], \tag{8}$$

where $P_i = P_i(c_2,\ c_3,\ \ldots,\ c_6)$.
Now, by using the above expressions namely, (4), (5), (6) and (8) in the second sub step, we get

$$z_n - \alpha = (1 - \theta_1 - \theta_2 - \theta_3)e_n + \frac{1}{3}c_2(3\theta_1 + 7\theta_2 + 11\theta_3)e_n^2 + \sum_{l=1}^{4} Q_j e_n^{j+2} + O(e_n^7), \tag{9}$$

where $Q_j = Q_j(\theta_1,\ \theta_2,\ \theta_3,\ c_2,\ c_3,\ \ldots,\ c_6)$.
It is clear from the above equation that for obtaining at least cubic convergence the coefficient of e_n and e_n^2 should be zero simultaneously. Therefore, we have

$$\theta_1 = \theta_3 + \frac{7}{4},\ \ \theta_2 = -2\theta_3 - \frac{3}{4}. \tag{10}$$

Using the above values of θ_1 and θ_2 in $Q_1 = 0$, we obtain the following independent relation

$$8\theta_3 - 9 = 0, \tag{11}$$

which further yields

$$\theta_3 = \frac{9}{8}. \tag{12}$$

By inserting the values of $\theta_1,\ \theta_2$ and θ_3, in the expression (9), we get

$$\begin{aligned} z_n - \alpha = &\left(5c_2^3 - c_3c_2 + \frac{c_4}{9}\right)e_n^4 + \left(-36c_2^4 + 32c_3c_2^2 - \frac{20c_4c_2}{9} - 2c_3^2 + \frac{8c_5}{27}\right)e_n^5 \\ &+ \frac{2}{27}\left(2295c_2^5 - 3537c_3c_2^3 + 633c_4c_2^2 + 9\left(99c_3^2 - 5c_5\right)c_2 - 99c_3c_4 + 7c_6\right)e_n^6 + O(e_n^7). \end{aligned} \tag{13}$$

In this way, we obtain a new optimal fourth-order iterative method. In order to obtain sixth-order convergent family of iterative methods, we expand the Taylor's series expansion of the function $f(z_n)$ about a point $x = \alpha$ with the aid of expression (13), we obtain

$$\begin{aligned} f(z_n) = f'(\alpha)\Bigg[&\left(5c_2^3 - c_3c_2 + \frac{c_4}{9}\right)e_n^4 + \left(-36c_2^4 + 32c_3c_2^2 - \frac{20c_4c_2}{9} - 2c_3^2 + \frac{8c_5}{27}\right)e_n^5 \\ &+ \frac{2}{27}\left(2295c_2^5 - 3537c_3c_2^3 + 633c_4c_2^2 + 9\left(99c_3^2 - 5c_5\right)c_2 - 99c_3c_4 + 7c_6\right)e_n^6 + O(e_n^7)\Bigg]. \end{aligned} \tag{14}$$

By using the equations (4), (5), (8), (13) and (14), in the last sub step of (2), we obtain

$$e_{n+1} = -\frac{1}{9}(45c_2^3 - 9c_3c_2 + c_4)(\theta_4 + \theta_5 + \theta_6 - 1)e_n^4 + \sum_{l=1}^{2} R_l e_n^{l+4} + O(e_n^7), \tag{15}$$

where $R_l = R_l(\theta_4,\ \theta_5,\ \theta_6,\ c_2,\ c_3,\ \ldots,\ c_6)$.
In order to obtain at least fifth-order of convergence, we have to substitute the following value of the disposable parameter θ_4

$$\theta_4 = -\theta_5 - \theta_6 + 1. \tag{16}$$

Now, we will use the above value of θ_4 in $R_1 = 0$, we have

$$2\theta_5 + 4\theta_6 + 3 = 0, \tag{17}$$

which further yields

$$\theta_5 = -2\theta_6 - \frac{3}{2}. \tag{18}$$

By using the values of θ_4 and θ_5 in the expression (15), we get

$$e_{n+1} = -\frac{1}{81}(45c_2^3 - 9c_3c_2 + c_4)\Big(2c_2^2(8\theta_6 - 27) + 9c_3\Big)e_n^6 + O(e_n^7), \quad \theta_6 \in \mathbb{R}. \tag{19}$$

Hence, it is straightforward to say from the above error equation that the proposed scheme (2) reaches the sixth-order convergence. This completes the proof. □

Development of the scheme for multi-dimensional case

The previous scheme (2) for scalar equation can be written for the multi-dimensional case as follows:

$$\begin{aligned}
y^{(n)} &= x^{(n)} - \frac{2}{3}F'(x^{(n)})^{-1}F(x^{(n)}),\\
z^{(n)} &= y^{(n)} - \left[\theta_1 I + \theta_2 F'(x^{(n)})^{-1}F'(y^{(n)}) + \theta_3\left(F'(x^{(n)})^{-1}F'(y^{(n)})\right)^{-2}\right]F'(x^{(n)})^{-1}F(x^{(n)}),\\
x^{(n+1)} &= z^{(n)} - \left[\theta_4 I + \theta_5 F'(x^{(n)})^{-1}F'(y^{(n)}) + \theta_6\left(F'(x^{(n)})^{-1}F'(y^{(n)})\right)^{-2}\right]F'(x^{(n)})^{-1}F(z^{(n)}),
\end{aligned} \tag{20}$$

where I is the identity matrix of order n and $\theta_i,\ i = 1,\ 2,\ \ldots, 6$ are free disposable parameters. With the values of the parameters obtained in Theorem 1 we design a parametric family of sixth-order iterative methods for solving nonlinear systems as shows the following theorem. In the proof of this result we use the tools and procedure introduced in [5].

Theorem 2 *Let $F : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a sufficiently differentiable function in an open neighborhood D of its zero α. Suppose that $F'(x)$ is continuous and nonsingular in α and the initial guess $x^{(0)}$ is close enough to α. Then, the iterative schemes defined by* (20) *have order of convergence six when*

$$\theta_1 = \theta_3 + \frac{7}{4},\ \theta_2 = -2\theta_3 - \frac{3}{4},\ \theta_3 = \frac{9}{8},\ \theta_4 = 1 - \theta_5 - \theta_6,\ \theta_5 = -2\theta_6 - \frac{3}{2},$$

where θ_6 is a free disposable parameter.

Proof. Let us assume that $e^{(n)} = x^{(n)} - \alpha$ be the error in the nth-iteration. Further, by developing $F(x^{(n)})$ in a neighborhood of α, we have

$$F(x^{(n)}) = F'(\alpha)\left[e^{(n)} + C_2(e^{(n)})^2 + C_3(e^{(n)})^3\right] + O((e^{(n)})^4), \tag{21}$$

$C_k = \frac{1}{k!}F'(\alpha)^{-1}F^{(k)}(\alpha)$, $k \geq 2$.
Similarly, we obtain

$$F'(x^{(n)}) = F'(\alpha)\left[I + 2C_2e^{(n)} + 3C_3(e^{(n)})^2 + 4C_4(e^{(n)})^3\right] + O((e^{(n)})^4). \tag{22}$$

By using the above expression (22), we further obtain

$$F'(x^{(n)})^{-1} = \left[I - 2C_2e^{(n)} + (4C_2^2 - 3C_3)(e^{(n)})^2\right] F'(\alpha)^{-1} + O((e^{(n)})^3), \tag{23}$$

With the help of equation (21) and (23), we have

$$F'(x^{(n)})^{-1}F(x^{(n)}) = e^{(n)} - C_2(e^{(n)})^2 + 2\left(C_2^2 - C_3\right)(e^{(n)})^3 + O((e^{(n)})^4), \tag{24}$$

By using the above expression (24) in the first step of (20), we get

$$y^{(n)} - \alpha = \frac{1}{3}e^{(n)} + \frac{2}{3}C_2(e^{(n)})^2 - \frac{2}{3}(2C_2^2 - 2C_3)(e^{(n)})^3 + O((e^{(n)})^4). \tag{25}$$

With aid of the expression (25), we further obtain

$$F'(y^{(n)}) = F'(\alpha)\left[I + \frac{4}{3}C_2e^{(n)} + \frac{1}{3}(4C_2^2 + C_3)(e^{(n)})^2\right] + O((e^{(n)})^3) \tag{26}$$

By using the equations (23) and (26), we further yield

$$\begin{aligned} F'(x^{(n)})^{-1}F'(y^{(n)}) = I - \frac{4C_2}{3}e^{(n)} + \left(4C_2^2 - \frac{8C_3}{3}\right)(e^{(n)})^2 \\ - \frac{8}{27}(36C_2^3 - 45C_3C_2 + 13C_4)(e^{(n)})^3 + O((e^{(n)})^4). \end{aligned} \tag{27}$$

By using equations (24), (27) and the values of disposable parameters $\theta_1,\ \theta_2$ and θ_3, in the second sub step of the scheme (20), we obtain

$$z^{(n)} - \alpha = A_1(e^{(n)})^4 + A_2(e^{(n)})^5 + O((e^{(n)})^6), \tag{28}$$

where A_1 and A_2 depend on constants C_j.

Now, we want to prove that the proposed scheme will reach sixth-order convergence when we will use the previous values of the disposable parameters (which are mentioned in the previous theorem). For this, we develop $F(z^{(n)})$ in a neighborhood of α

$$F(z^{(n)}) = F'(\alpha)\left[A_1(e^{(n)})^4 + A_2(e^{(n)})^5\right] + O((e^{(n)})^6). \tag{29}$$

With the aid of expressions (23), (24), (27), (29) and the values of disposable parameters θ_4 and θ_5 (which are display in the previous theorem), we have

$$\begin{aligned} \left[\left(\theta_6 + \frac{5}{2}\right)I + \left(-2\theta_6 - \frac{3}{2}\right)F'(x^{(n)})^{-1}F'(y^{(n)}) + \theta_6\left(F'(x^{(n)})^{-1}F'(y^{(n)})\right)^{-2}\right] F'(x^{(n)})^{-1}F(z^{(n)}) \\ = A_1(e^{(n)})^4 + A_2(e^{(n)})^5 + \frac{A_1}{9}\left(2C_2^2\left(8\theta_6 - 27\right) + 9C_3\right)(e^{(n)})^6 + O((e^{(n)})^7) \end{aligned} \tag{30}$$

Finally, by using (28) and (29) in the last sub step of the proposed scheme (20), we obtain

$$\begin{aligned} x^{(n+1)} - \alpha &= z^{(n)} - \alpha - \left[A_1(e^{(n)})^4 + A_2(e^{(n)})^5 + \frac{A_1}{9}\left(2C_2^2\left(8\theta_6 - 27\right) + 9C_3\right)(e^{(n)})^6 + O((e^{(n)})^7)\right] \\ &= \frac{A_1}{9}\left(2C_2^2\left(8\theta_6 - 27\right) + 9C_3\right)(e^{(n)})^6 + O((e^{(n)})^7). \end{aligned} \tag{31}$$

Therefore, (20) is a new family of sixth-order iterative methods. □

Numerical experiments

This section is devoted to verify the convergence behavior and computational efficiency of the proposed family of iterative methods which we have proposed in the earlier sections.

Most of the times, some researchers who want to claim that their methods are superior than other existing methods available in the literature. They consider some well-known or standard or self-made examples and manipulate the initial approximations to claim that their methods are superior than other methods. To halt this practice, we consider six numerical examples; first one is chosen from Guem et al. [7]; second one is chosen from Grau and Díaz-Barrero [8]; third one is chosen from Parhi and Gupta [21], fourth one is chosen from Soleymani [27] and fifth one is consider from Ren et al. [24], with same initial guesses which are mentioned in their papers. Further, we also want to see what will happen if we consider different examples and with different initial guesses, which are not mentioned in their papers. Therefore, we consider one more nonlinear equation from Behl et al. [14]. The details of chosen examples or test functions are available in Table 1. Moreover, the considered test functions with their corresponding zeros and initial guesses are also displayed in the same table.

Now, we employ the new sixth-order scheme (2) $\left(\text{for } \theta_6 = 0,\ \frac{27}{8} \text{ and } \theta_6 = \frac{55}{16}\right)$ denoted by (PM_1), (PM_2) and (PM_3), respectively to see the convergence behavior and effectiveness. We shall compare our methods with a higher-order family of double-Newton methods with a bivariate weighting function that is very recently presented by Guem et al. [7], out of them we choose one of their best method (3.8), called by (GKN). In addition, we consider a sixth-order variants of Ostrowski's method proposed by Grau and Díaz-Barrero [8], out of them we choose expression (4–6), described as (GB). Further, we also compare them with a sixth-order multi-point iterative method (2.7) proposed by Parhi and Gupta [21], called by (PG). Moreover, we will compare them with a sixth-order Jarratt method presented by Soleymani [27], out of which we consider method (10), denoted by (SM). Finally, we also compared our methods with some new sixth-order variants of Jarratt's method designed by Ren et al. [24], out of them we choose method (54) (for $\alpha = \frac{5}{10},\ \beta = \frac{12}{10},\ \gamma = \frac{2}{10},\ \delta = \frac{2}{10}$), described as (RWB).

For better comparisons of our proposed methods, we have displayed the errors between the two consecutive iterations $|x_{n+1} - x_n|$, the estimation of the computational order of convergence $\rho = \frac{\log|(x_n - x_n)/(x_{n-1} - x_{n-2})|}{\log|(x_{n-1} - x_{n-2})/(x_{n-2} - x_{n-3})|}$ or $\frac{\log|(x_n - \alpha)/(x_{n-1} - \alpha)|}{\log|(x_{n-1} - \alpha)/(x_{n-2} - \alpha)|}$ and residual error of the corresponding function $(|f(x_n)|)$, corresponding to each test function in Tables 2 and 3.

Further, we also consider a variety of applied examples to further check the validity of theoretical results for nonlinear system. Therefore, we employ the new sixth-order scheme (20) for $\theta_6 = 0,\ \frac{27}{8}$ and $\theta_6 = \frac{55}{16}$ denoted by $(\widehat{PM_1})$, $(\widehat{PM_2})$ and $(\widehat{PM_3})$, respectively, to verify the performance of these methods on the examples 1–3. We shall compare them with a fourth-order Jarratt's method [5] for system of nonlinear equations, denoted by (JM). In addition, we shall compare them with a method (61) that is recently presented by Ren et al. [24], denoted by $(\overline{RWB})$. Further, we also compared our methods with Ostrowski type methods for solving systems of nonlinear equations designed by Grau et al. [9], out of them we consider methods namely, method (5) and method (7), denoted by (GM_1) and (GM_2), respectively. Moreover, we also compared our methods with sixth-order family of iterative method designed by Cordero et al. [5], out of them we choose method (6), denoted by (CM). Finally, we compare our methods with an efficient Jarratt-like methods presented by Sharma [25], we consider method (13) called by (SA).

In the following Tables 4, 5, 7–10, we have displayed the error between two consecutive error in

the iterations $\|x^{(n+1)}-x^{(n)}\|$, the computational order of convergence $\rho = \frac{\log\left[\|x^{(n+1)}-x^{(n)}\|/\|x^{(n)}-x^{(n-1)}\|\right]}{\log\left[\|x^{(n)}-x^{(n-1)}\|/\|x^{(n-1)}-x^{(n-2)}\|\right]}$ and residual error of the corresponding function $(\|F(x^{(n)})\|)$.

During the current numerical experiments with programming language Mathematica (Version 9), all computations have been done with multiple precision arithmetic with 1000 digits of mantissa, which minimize round-off errors. Let us remark that, in all tables, $a\ e(\pm b)$ denotes $a \times 10^{(\pm b)}$.

Table 1: Test problems

$f(x)$	$Zeros(\alpha)$	x_0
$f_1(x) = 2\cos(x^2) - \log(1 + 4x^2 - \pi) - \sqrt{2}$; [7]	$\sqrt{\frac{\pi}{4}}$	1
$f_2(x) = \left[1 + (1-\gamma)^4\right] x - (1-\gamma x)^4$ $[\gamma = 5]$; [8]	0.00361710817890406354076835 1 . . .	0.05
$f_3(x) = x^2 - e^x - 3x + 2$; [21]	0.25753028543986076045536 73 . . .	2
$f_4(x) = \tan x$; [27]	0	1.2
$f_5(x) = e^{-x} + \cos x$; [24]	1.74613953040801241765070 3 . . .	2
$f_6(x) = x^3 + \sin x + 2x$; [14]	0	1

Table 2: Comparison of $|x_{n+1} - x_n|$ for the functions $f_i(x),\ i = 1,\ 2, \ldots, 6$ among listed methods

f_i	x_0	$\lvert x_{n+1} - x_n\rvert$ $\backslash\rho$	GKN	GB	PG	SM	RWB	PM_1	PM_2	PM_3
f_1	1	$\lvert x_2 - x_1\rvert$	$1.3e(-4)$	$1.5e(-5)$	$1.2e(-4)$	$1.1e(-5)$	$9.7e(-6)$	$2.4e(-6)$	$2.0e(-5)$	$2.0e(-5)$
		$\lvert x_3 - x_2\rvert$	$1.1e(-28)$	$3.2e(-28)$	$7.7e(-22)$	$1.9e(-28)$	$5.1e(-29)$	$3.2e(-32)$	$1.1e(-26)$	$1.3e(-26)$
		$\lvert x_4 - x_3\rvert$	$3.1e(-167)$	$3.7e(-164)$	$6.4e(-125)$	$3.6e(-165)$	$9.9e(-169)$	$2.1e(-187)$	$3.5e(-154)$	$9.4e(-154)$
		ρ	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
f_2	0.05	$\lvert x_2 - x_1\rvert$	$8.0e(-9)$	$6.8e(-9)$	$1.6e(-8)$	$5.0e(-9)$	$4.5e(-9)$	$3.0e(-106)$	$7.9e(-10)$	$7.9e(-10)$
		$\lvert x_3 - x_2\rvert$	$3.5e(-49)$	$9.3e(-50)$	$4.5e(-47)$	$1.8e(-50)$	$8.4e(-51)$	$4.1e(-60)$	$1.6e(-56)$	$1.8e(-56)$
		$\lvert x_4 - x_3\rvert$	$2.6e(-291)$	$6.3e(-295)$	$2.2e(-278)$	$4.5e(-299)$	$3.3e(-301)$	$2.3e(-359)$	$1.3e(-336)$	$2.2e(-336)$
		ρ	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
f_3	2	$\lvert x_2 - x_1\rvert$	$3.5e(-2)$	$9.4e(-2)$	$7.6e(-2)$	$9.8e(-15)$	$3.4e(-1)$	$5.2e(-3)$	$1.4e(-2)$	$1.5e(-2)$
		$\lvert x_3 - x_2\rvert$	$9.3e(-13)$	$1.4e(-10)$	$9.2e(-11)$	$1.8e(-4)$	$5.4e(-7)$	$3.1e(-19)$	$13e(-15)$	$1.5e(-15)$
		$\lvert x_4 - x_3\rvert$	$3.2e(-76)$	$1.1e(-63)$	$2.9e(-64)$	$1.4e(-26)$	$9.0e(-42)$	$1.3e(-116)$	$7.5e(-94)$	$2.1e(-93)$
		ρ	6.0005	6.0054	6.0004	5.9147	6.0008	6.0039	5.9979	5.9979
f_4	1.2	$\lvert x_2 - x_1\rvert$	$2.7e(-1)$	$6.4e(-1)$	$4.4e(-1)$	$3.4e(-1)$	$3.7e(-1)$	$4.2e(-1)$	$3.2e(-1)$	$3.2e(-1)$
		$\lvert x_3 - x_2\rvert$	$6.0e(-6)$	$3.5e(-3)$	$3.4e(-4)$	$4.8e(-5)$	$7.6e(-5)$	$7.3e(-7)$	$9.7e(-6)$	$9.6e(-6)$
		$\lvert x_4 - x_3\rvert$	$1.6e(-38)$	$5.0e(-19)$	$4.3e(-26)$	$3.6e(-32)$	$9.1e(-31)$	$6.9e(-45)$	$5.1e(-37)$	$4.6e(-37)$
		ρ	7.0150	7.0180	7.0244	7.0436	7.0359	6.6011	6.9218	6.9244
f_5	2.0	$\lvert x_2 - x_1\rvert$	$2.9e(-6)$	$1.9e(-6)$	$4.9e(-7)$	$6.1e(-7)$	$1.0e(-6)$	$1.e(-5)$	$8.2e(-7)$	$6.5e(-7)$
		$\lvert x_3 - x_2\rvert$	$9.2e(-37)$	$1.9e(-37)$	$2.2e(-41)$	$9.2e(-41)$	$3.0e(-39)$	$1.2e(-32)$	$1.3e(-39)$	$2.8e(-40)$
		$\lvert x_4 - x_3\rvert$	$9.4e(-220)$	$1.4e(-223)$	$2.2e(-247)$	$1.1e(-243)$	$1.9e(-234)$	$2.1e(-194)$	$1.6e(-236)$	$2.1e(-240)$
		ρ	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
f_6	1	$\lvert x_2 - x_1\rvert$	$3.7e(-3)$	$1.8e(-3)$	$1.8e(-2)$	$4.3e(-2)$	$1.8e(-2)$	$1.0e(-2)$	$2.2e(-4)$	$4.1e(-4)$
		$\lvert x_3 - x_2\rvert$	$2.9e(-19)$	$3.1e(-21)$	$3.1e(-14)$	$1.1e(-11)$	$2.4e(-14)$	$4.7e(-16)$	$9.2e(-28)$	$7.9e(-26)$
		$\lvert x_4 - x_3\rvert$	$5.3e(-132)$	$1.3e(-145)$	$1.3e(-96)$	$1.1e(-78)$	$2.2e(-97)$	$2.0e(-109)$	$2.3e(-191)$	$8.0e(-178)$
		ρ	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000

Table 3: Comparison of resiual error $|f(x_n)|$ in among listed methods

f_i	x_0	$\|x_{n+1}-x_n\|$ $\backslash\rho$	GKN	GB	PG	SM	RWB	PM_1	PM_2	PM_3
f_1	1	$\|f(x_1)\|$	$1.3e(-5)$	$1.4e(-4)$	$1.1e(-3)$	$1.1e(-4)$	$9.3e(-5)$	$2.3e(-5)$	$1.9e(-4)$	$2.0e(-4)$
		$\|f(x_2)\|$	$1.0e(-27)$	$3.1e(-27)$	$7.4e(-21)$	$1.8e(-27)$	$4.8e(-28)$	$3.1e(-31)$	$1.1e(-25)$	$1.3e(-25)$
		$\|f(x_3)\|$	$3.0e(-166)$	$3.6e(-163)$	$6.1e(-124)$	$3.4e(-164)$	$9.5e(-168)$	$2.0e(-186)$	$3.3e(-153)$	$9.0e(-153)$
f_2	0.05	$\|f(x_1)\|$	$2.2e(-6)$	$1.9e(-6)$	$4.5e(-6)$	$1.4e(-6)$	$1.3e(-6)$	$8.4e(-8)$	$2.2e(-7)$	$2.2e(-7)$
		$\|f(x_2)\|$	$9.7e(-47)$	$2.6e(-47)$	$1.3e(-44)$	$5.1e(-48)$	$2.3e(-48)$	$1.1e(-57)$	$4.5e(-54)$	$4.9e(-54)$
		$\|f(x_3)\|$	$7.1e(-289)$	$1.7e(-292)$	$6.1e(-276)$	$1.2e(-296)$	$9.1e(-299)$	$6.4e(-357)$	$3.7e(-334)$	$6.2e(-334)$
f_3	2	$\|f(x_1)\|$	$1.3e(-1)$	$3.5e(-1)$	$2.9e(-1)$	4.2	1.3	$2.0e(-2)$	$5.4e(-2)$	$5.5e(-2)$
		$\|f(x_2)\|$	$3.5e(-12)$	$5.2e(-10)$	$3.5e(-10)$	$6.8e(-4)$	$2.1e(-6)$	$1.2e(-8)$	$4.9e(-15)$	$5.8e(-15)$
		$\|f(x_3)\|$	$1.2e(-75)$	$4.3e(-63)$	$1.1e(-63)$	$5.1e(-26)$	$3.4e(-41)$	$4.8e(-116)$	$2.8e(-93)$	$7.8e(-93)$
f_4	1.2	$\|f(x_1)\|$	$2.7e(-1)$	$7.5e(-1)$	$4.8e(-1)$	$3.6e(-1)$	$3.9e(-1)$	$4.5e(-1)$	$3.3e(-1)$	$3.3e(-1)$
		$\|f(x_2)\|$	$6.0e(-6)$	$3.5e(-3)$	$3.4e(-4)$	$4.8e(-5)$	$7.6e(-5)$	$7.3e(-7)$	$9.7e(-6)$	$9.6e(-6)$
		$\|f(x_3)\|$	$1.6e(-38)$	$5.0e(-19)$	$4.3e(-26)$	$3.6e(-32)$	$9.1e(-31)$	$6.9e(-45)$	$5.1e(-37)$	$4.6e(-37)$
f_5	2.0	$\|f(x_1)\|$	$3.4e(-6)$	$2.3e(-6)$	$5.6e(-7)$	$7.0e(-7)$	$1.2e(-6)$	$1.2e(-5)$	$9.6e(-7)$	$7.5e(-7)$
		$\|f(x_2)\|$	$1.1e(-36)$	$2.1e(-37)$	$2.6e(-41)$	$1.1e(-40)$	$3.5e(-39)$	$1.3e(-32)$	$1.5e(-39)$	$3.3e(-40)$
		$\|f(x_3)\|$	$1.1e(-219)$	$1.6e(-223)$	$2.5e(-247)$	$1.3e(-243)$	$2.2e(-234)$	$2.5e(-194)$	$1.9e(-236)$	$2.4e(-240)$
f_6	1	$\|f(x_1)\|$	$1.1e(-2)$	$5.5e(-3)$	$5.4e(-2)$	$1.3e(-1)$	$5.3e(-2)$	$3.0e(-2)$	$6.5e(-4)$	$1.2e(-4)$
		$\|f(x_2)\|$	$8.7e(-9)$	$9.4e(-21)$	$9.3e(-14)$	$3.4e(-11)$	$7.3e(-14)$	$1.4e(-35)$	$2.8e(-27)$	$2.8e(-27)$
		$\|f(x_3)\|$	$1.6e(-131)$	$3.8e(-145)$	$4.0e(-96)$	$3.2e(-78)$	$6.6e(-7)$	$6.1e(-109)$	$7.0e(-191)$	$2.4e(-177)$

Example 1 *Let us consider the Van der Pol equation [4, 19], which is defined as follows:*

$$y'' - \mu(y^2-1)y' + y = 0,\ \mu > 0, \tag{32}$$

which governs the flow of current in a vacuum tube, with the boundary conditions $y(0) = 0,\ y(2) = 1$. *Further, we consider the partition of the given interval* $[0,\ 2]$, *which is given by*

$$x_0 = 0 < x_1 < x_2 < x_3 < \cdots < x_n,\ where\ x_i = x_0 + ih,\ h = \frac{2}{n}.$$

Moreover, we assume that

$$y_0 = y(x_0) = 0,\ y_1 = y(x_1),\ \ldots,\ y_{n-1} = y(x_{n-1}),\ y_n = y(x_n) = 1.$$

If, we discretized the above problem (32) *by using the numerical formula for the first derivative and second derivative, which are given by*

$$y'_k = \frac{y_{k+1} - y_{k-1}}{2h},\ y''_k = \frac{y_{k-1} - 2y_k + y_{k+1}}{2h},\ k = 1,\ 2,\ \ldots,\ n-1,$$

then, we obtain a $(n-1)\times(n-1)$ *system of nonlinear equations*

$$2h^2 x_k - h\mu\left(x_k^2 - 1\right)(x_{k+1} - x_{k-1}) + 2\left(x_{k-1} + x_{k+1} - 2x_k\right) = 0.$$

Let us consider $\mu = \frac{1}{2}$ *and initial approximation* $y_k^{(0)} = \left(\frac{1}{2},\ \frac{1}{2},\ \ldots, \frac{1}{2}\right)$. *In this problem, we consider the value of* $n = 7$ *so that we can obtain a* 6×6 *system of nonlinear equations. The*

Table 4: (Comparison of $\|x^{(n+1)} - x^{(n)}\|$ among listed methods in the Van der Pol equation)

$\|x^{(n+1)} - x^{(n)}\|$ $\backslash\rho$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|x^{(2)} - x^{(1)}\|$	$1.4e(-3)$	$1.7e(-5)$	$7.0e(-4)$	$7..7e(-4)$	$1.7e(-5)$	$8.5e(-5)$	$6.4e(-5)$	$1.7e(-5)$	$1.5e(-5)$
$\|x^{(3)} - x^{(2)}\|$	$3.0e(-15)$	$2.2e(-34)$	$4.4e(-17)$	$6.5e(-17)$	$2.2e(-34)$	$1.1e(-29)$	$2.0e(-30)$	$2.2e(-34)$	$1.4e(-34)$
$\|x^{(4)} - x^{(3)}\|$	$1.2e(-61)$	$7.9e(-208)$	$9.3e(-70)$	$4.8e(-69)$	$7.9e(-208)$	$2.8e(-179)$	$6.3e(-183)$	$2.3e(-206)$	$9.3e(-208)$
ρ	3.9826	6.0017	3.9883	3.9874	6.0017	6.0153	5.9768	5.9551	5.9989

Table 5: (Comparison of residual error $\|F(x^{(n)})\|$ among listed methods in the Van der Pol equation)

$\|F(x^{(n)})\|$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|F(x^{1})\|$	$1.4e(-4)$	$2.0e(-5)$	$5.6e(-4)$	$6.0e(-4)$	$2.0e(-5)$	$8.9e(-5)$	$9.2e(-5)$	$2.3e(-5)$	$2.2e(-5)$
$\|F(x^{(2)})\|$	$5.1e(-15)$	$3.4e(-34)$	$7.0e(-17)$	$1.1e(-16)$	$3.4e(-34)$	$2.8e(-29)$	$6.0e(-30)$	$1.1e(-33)$	$7.2e(-34)$
$\|F(x^{(3)})\|$	$1.5e(-61)$	$1.1e(-207)$	$1.5e(-69)$	$7.8e(-69)$	$1.1e(-207)$	$5.5e(-179)$	$1.2e(-182)$	$3.5e(-206)$	$2.4e(-207)$

solutions of this problem is

$$\alpha = (0.3822666\ldots,\ 0.6911725\ldots,\ 0.9234664\ldots,\ 1.076325\ldots,\ 1.143815\ldots,\ 1.118869\ldots)^t.$$

Example 2 *In this example, we consider one of the famous applied science problem which is known as Hammerstein integral equation (see [20, pp. 19-20] to check the effectiveness and applicability of our proposed methods as compared to the other existing methods, is given as follows:*

$$x(s) = 1 + \frac{1}{5}\int_0^1 F(s,t)x(t)^3 dt$$

where $x \in C[0,1]; s,t \in [0,1]$ and the kernel F is

$$F(s,t) = \begin{cases} (1-s)t, t \leq s, \\ s(1-t), s \leq t. \end{cases}$$

To transform the above equation into a finite-dimensional problem by using Gauss Legendre quadrature formula given as $\int_0^1 f(t)dt \simeq \sum_{j=1}^8 w_j f(t_j)$, where the abscissas t_j and the weights w_j are determined for $t = 8$ by Gauss Legendre quadrature formula. Denoting the approximations of $x(t_i)$ by $x_i (i = 1, 2, ..., 8)$, one gets the system of nonlinear equations $5x_i - 5 - \sum_{j=1}^8 a_{ij}x_j^3 = 0$, where $i = 1, 2, ..., 8$

$$a_{ij} = \begin{cases} w_j t_j(1-t_i), j \leq i, \\ w_j t_i(1-t_j), i < j. \end{cases}$$

Where the abscissas t_j and the weights w_j are known and given in following table for $t = 8$. The convergence of the methods towards the root

$$X = (1.00209\ldots, 1.00990\ldots, 1.01972\ldots, 1.02643\ldots, 1.02643\ldots, 1.01972\ldots, 1.00990\ldots, 1.00209\ldots)^t,$$

Table 6: (Abscissas and weights of Gauss Legendre quadrature formula for $t = 8$)

j	t_j	w_j
1	0.019855071751231884158219 57...	0.050614268145188129576265 67...
2	0.101666761293186630204223 03...	0.111190517226687235272178 00...
3	0.237233795041835507091130 47...	0.156853322938943643668981 10...
4	0.408282678752175097530261 93...	0.181341891689180991482575 22...
5	0.591717321247824902469738 07...	0.181341891689180991482575 22...
6	0.762766204958164492908869 52...	0.156853322938943643668981 10...
7	0.898333238706813369795776 96...	0.111190517226687235272178 00...
8	0.980144928248768115841780 43...	0.050614268145188129576265 67...

Table 7: (Comparison of $\|x^{(n+1)} - x^{(n)}\|$ among listed methods in the Hammerstein integral equation)

$\|x^{(n+1)} - x^{(n)}\|$ $\backslash\rho$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|x^{(2)} - x^{(1)}\|$	$1.2e(-4)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$	$5.7e(-6)$
$\|x^{(3)} - x^{(2)}\|$	$4.3e(-20)$	$6.5e(-38)$	$6.5e(-38)$	$8.7e(-38)$	$6.5e(-38)$	$1.7e(-37)$	$1..7e(-37)$	$7.8e(-38)$	$7.8e(-38)$
$\|x^{(4)} - x^{(3)}\|$	$8.1e(-82)$	$1.6e(-229)$	$1.6e(-229)$	$1.2e(-228)$	$1.6e(-229)$	$1.4e(-226)$	$1.4e(-226)$	$5.5e(-229)$	$5.5e(-229)$
ρ	3.9983	5.9987	5.9987	5.9987	5.9987	5.9989	5.9989	6.0067	5.9989

Table 8: (Comparison of residual error $\|F(x^{(n)})\|$ among listed methods in the Hammerstein integral equation)

$\|F(x^{(n)})\|$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|F(x^{(1)})\|$	$5.4e(-4)$	$2.7e(-5)$	$2.7e(-5)$	$2.7e(-6)$	$2.7e(-5)$	$2.7e(-5)$	$2.7e(-5)$	$2.7e(-5)$	$2.7e(-5)$
$\|F(x^{(2)})\|$	$2.0e(-19)$	$3.1e(-37)$	$3.1e(-37)$	$4.1e(-37)$	$3.1e(-37)$	$8.0e(-37)$	$8.1e(-37)$	$3.6e(-37)$	$3.6e(-37)$
$\|F(x^{(3)})\|$	$3.8e(-81)$	$7.6e(-229)$	$7.6e(-229)$	$5.7e(-228)$	$7.6e(-229)$	$6.5e(-226)$	$6.5e(-226)$	$2.6e(-228)$	$2.6e(-228)$

is tested in the following Tables 4 and 5 on the basis of the initial guess $\left(-\frac{1}{2},\ -\frac{1}{2},\ -\frac{1}{2},\ -\frac{1}{2}\right)$.

Example 3 *Let us consider the following nonlinear system of nonlinear equation [10]*

$$f_i(x) = x_i - \cos\left(2x_i - \sum_{j=1}^{4} x_j\right), \tag{33}$$

where $i = 1, 2, 3, 4$. *We choose the initial guess* $x^{(0)} = (1,\ 1,\ 1,\ 1)^t$ *for this problem for obtaining the required solution* $\alpha = (0.5149333\ldots,\ 0.5149333\ldots,\ 0.5149333\ldots,\ 0.5149333\ldots)^t$.

Concluding remarks

The main beauty of the proposed family of iterative methods for the system of nonlinear equations is that we have to calculate only one inverse of the Jacobian matrix (i.e. $F'(x^{(n)})$) in the

Table 9: (Comparison of $\|x^{(n+1)} - x^{(n)}\|$ among listed methods in example (3))

$\|x^{(n+1)} - x^{(n)}\|$ $\backslash\rho$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|x^{(2)} - x^{(1)}\|$	$3.7e(-3)$	$3.6e(-4)$	$3.5e(-4)$	$3.5e(-4)$	$3.6e(-4)$	$3.9e(-4)$	$3.9e(-4)$	$3.6e(-4)$	$3.6e(-4)$
$\|x^{(3)} - x^{(2)}\|$	$4.6e(-12)$	$9.3e(-24)$	$8.3e(-24)$	$1.4e(-23)$	$9.3e(-24)$	$5.5e(-23)$	$5.6e(-23)$	$1.2e(-23)$	$1.2e(-23)$
$\|x^{(4)} - x^{(3)}\|$	$1.2e(-47)$	$2.8e(-141)$	$1.6e(-141)$	$5.4e(-140)$	$2.8e(-141)$	$4.9e(-136)$	$4.9e(-136)$	$2.0e(-140)$	$1.4e(-140)$
ρ	4.0004	6.0000	5.9987	6.0000	6.0000	6.0000	6.0000	6.0000	5.9989

Table 10: (Comparison of residual error $\|F(x^{(n)})\|$ among listed methods in example (3))

$\|F(x^{(n)})\|$	JM	$\overline{RWB}$	$GM1$	$GM2$	CM	SA	$\widehat{PM_1}$	$\widehat{PM_2}$	$\widehat{PM_3}$
$\|F(x^{(1)})\|$	$1.0e(-2)$	$9.7e(-4)$	$9.4e(-4)$	$9.6e(-4)$	$9.7e(-4)$	$1.0e(-3)$	$1.1e(-3)$	$9.8e(-4)$	$9.8e(-4)$
$\|F(x^{(2)})\|$	$1.3e(-11)$	$2.5e(-23)$	$2.3e(-23)$	$3.8e(-23)$	$2.5e(-23)$	$1.5e(-22)$	$1.5e(-22)$	$3.4e(-24)$	$3.2e(-23)$
$\|F(x^{(3)})\|$	$3.2e(-47)$	$7.7e(-141)$	$4.2e(-141)$	$1.5e(-139)$	$7.7e(-141)$	$1.3e(-135)$	$1.3e(-135)$	$5.3e(-140)$	$3.9e(-140)$

case of nonlinear system which reduce the computational cost. The convergence properties are fully investigated along with two main theorems describing their order of convergence. We also tested the order of convergence of our proposed families on a concrete variety of numerical experiments and it is found that the order of convergence of the proposed family is well deduced for scalar as well as system of nonlinear equations. Further, our proposed methods perform better than the existing methods on the mentioned numerical examples even though if we choose the same problems with same initial guesses.

Further, the computational accuracy of the iterative methods dependent on several factors like; body structures of the iterative methods, initial guesses, test functions and the sought zeros. We have shown in the numerical experiments that our proposed iterative methods perform better than the existing ones of the same order. But, these results are not always expected because there is no iterative methods till date which shows best accuracy for every test functions. Further, it is also important to note that the behavior of iterative methods for convergence to the required root is depend on asymptotic error constant c_j, test function $f(x)$ and the required root α.

References

[1] Abad, M.F., Cordero, A. and Torregrosa, J.R. (2014) A family of seventh-order schemes for solving nonlinear systems. *Bull. Math. Soc. Sci. Math. Roum.* **57 (105)(2)**, 133–145.

[2] Artidiello, S., Cordero, A., Torregrosa, J.R. and Vassileva, M.P. (2015) Multidimensional generalization of iterative methods for solving nonlinear problems by means of weight-function procedure. *Appl. Math. Comput.* **268**, 1064–1071.

[3] Awawdeh, F. (2010) On new iterative method for solving systems of nonlinear equations. *Numer. Algor.* **54**, 395–409.

[4] Burden, R.L. and Faires, J.D. (2001) Numerical Analysis. *PWS Publishing Company, Boston.*

[5] Cordero, A., Hueso, J.L. and E. Martínez, Torregrosa, J.R. (2010) A modified NewtonJarratt's composition. *Numer. Algor.* **55**, 87–99.

[6] Cordero, A., Maimó, J.G., Torregrosa, J.R. and Vassileva, M.P. (2014) Solving nonlinear problems by Ostrowski-Chun type parametric families. *J. Math. Chem.* **52**, 430–449.

[7] Geum, Y.H., Kim, Y.I. and Neta, B. (2015) On developing a higher-order family of double-Newton

methods with a bivariate weighting function. *Appl. Math. Comput.* **254**, 277–290.

[8] Grau, M. and Díaz-Barrero, J.L. (2006) An improvement to Ostrowski root-finding method. *Appl. Math. Comput.* **173**, 450–456.

[9] Grau-Sánchez, M., Grau, Á. and Noguera, M.(2011) Ostrowski type methods for solving systems of nonlinear equations. *Appl. Math. Comput.* **218**, 2377–2385.

[10] Grau-Sánchez, M., Grau, Á. and Noguera, M. (2011) Frozen divided difference scheme for solving systems of nonlinear equations. *J. Comput. Appl. Math.* **235**, 1739–1743.

[11] Grosan, C. and Abraham A. (2008) A new approach for solving nonlinear equations systems. *IEEE Trans. Syst. Man Cybernet Part A: Syst. Humans* **38**, 698–714.

[12] Hueso, J.L., Martínez, E. and Teruel, C. (2015) Convergence, efficiency and dynamics of new fourth and sixth order families of iterative methods for nonlinear systems *J. Comput. Appl. Math.* **275**, 412–420.

[13] Jarratt, P. (1966) Some fourth order multipoint iterative methods for solving equations. *Math. Comput.* **20**, 434–437.

[14] Kim, Y.I., Behl, R. and Motsa, S.S. (2016) Higher-order efficient class of Chebyshev-Halley type methods. *Appl. Math. Comput.* **273**, 1148–1159.

[15] Kou, J. and Li, Y. (2007) An improvement of the Jarratt method. *Appl. Math. Comput.* **189**, 1816–1821.

[16] Lin, Y., Bao, L. and Jia, X. (2010) Convergence analysis of a variant of the Newton method for solving nonlinear equations. *Comput. Math. Appl.* **59**, 2121–2127.

[17] Moré, J.J. (1990) A collection of nonlinear model problems. *in:E.L. Allgower, K. Georg(Eds.), Computational Solution of Nonlinear Systems of Equations, Lectures in Applied Mathematics, Amer. Math. Soc. Providence, RI* **26**, 723–762.

[18] Nejat, A. and Ollivier-Gooch, C. (2008) Effect of discretization order on preconditioning and convergence of a high-order unstructured Newton-GMRES solver for the Euler equations. *J. Comput. Phys.* **227(4)**, 2366–2386.

[19] Noor, M.A., Waseem, M. and Noor, K.I. (2015) New iterative technique for solving a system of nonlinear equations. *Appl. Math. Comput.* **271**, 446–466.

[20] Ortega, J.M. and Rheinboldt, W.C. (1970) Iterative solution of nonlinear equations in several variables. *Academic Press, New-York.*

[21] Parhi, S.K. and Gupta, D.K. (2008) A sixth order method for nonlinear equations. *Computational Mechanics* **203**, 50–55.

[22] Petković, M.S., Neta, B., Petković, L.D. and J. Džunić, (2012) Multipoint methods for solving nonlinear equations. *Academic Press.*

[23] Rangan, A.V., Cai, D. and Tao, L. (2007) Numerical methods for solving moment equations in kinetic theory of neuronal network dynamics. *J. Comput. Phys.* **221**, 781–798.

[24] Ren, H., Wu,Q. and Bi, W. (2009) New variants of Jarratt's method with sixth-order convergence. *Numer. Algor.* **52**, 585–603.

[25] Sharma, J.R. and Arora, H. (2014) Efficient Jarratt-like methods for solving systems of nonlinear equations. *Calcolo* **51**, 193–210.

[26] Sharma, J.R., Guna, R.K. and Sharma, R. (2013) An efficient fourth order weighted-Newton method for systems of nonlinear equations. *Numer. Algor.* **2**, 307–323.

[27] Soleymani, F. (2011) Revisit of Jarratt method for solving nonlinear equations. *Numer. Algor.* **57**, 377–388.

[28] Traub, J.F. (1964) Iterative methods for the solution of equations. *Prentice-Hall, Englewood Cliffs.*

[29] Tsoulos, I.G. and Stavrakoudis, A. (2010) On locating all roots of systems of nonlinear equations inside bounded domain using global optimization methods. *Nonlinear Anal. Real World Appl.* **11**, 2465–2471.

[30] Wang, X. and Zhang, T. (2013) A family of Steffensen type methods with seventh-order convergence. *Numer. Algor.* **62**, 429–444.

Dynamic Analysis of Heat Exchanger Piles in Offshore Environment

***†Arundhuti Banerjee[1], Tanusree Chakraborty[2], and Vasant Matsagar[3]**

[*†1]Research Scholar, Department of Civil Engineering, Indian Institute of Technology (IIT) Delhi, Hauz Khas, New Delhi - 110016, E-mail: arundhuti.banerjee@iitd.civil.ac.in

[2]Assistant Professor, Department of Civil Engineering, Indian Institute of Technology (IIT) Delhi, Hauz Khas, New Delhi - 110016, E-mail: tanusree@civil.iitd.ac.in.

[3]Associate Professor, Department of Civil Engineering, Indian Institute of Technology (IIT) Delhi, Hauz Khas, New Delhi - 110016, E-mail: matsagar@civil.iitd.ac.in.

**Corresponding author*

*Presenting and †corresponding author: arundhuti.banerjee@iitd.civil.ac.in

Abstract

In order to secure the future generations from energy crisis, it is widely accepted that implementation of renewable energy resources is the urgent need of the day. Renewable energy resources, e.g. wind, ocean wave and geothermal energy provide substantial benefits towards our climate, health, and economy. Heat exchanger piles are deep foundations that combine the structural function as a foundation with a heat exchanger for extracting heat from the earth's crust. In the proposed study, a novel concept of combined offshore wind turbine-heat exchanger pile foundation technology will be investigated. Coupled temperature-displacement analysis of geothermal energy pile foundations will be carried out using finite element (FE) software ANSYS to understand the interaction between geothermal pile and the surrounding soil subjected to random wave loading and thermal loading-unloading cycles. In this paper, thermal characteristics are evaluated in the ground thermal energy system with steel foundation pile. The average effective thermal conductivity of the surrounding offshore soil is estimated by conducting an extensive literature survey. Pile will be modeled using the linear elastic model where as the soil is modeled using Drucker Prager constitutive model. The thermal loading-unloading cycle will be applied on the pile using temperature cycles. The random wave loading on the pile would be modeled using the Pierson-Moskowitz spectrum. The results of the analyses will be studied for stress, strain and displacement response of the heat exchanger pile foundation for offshore wind turbine and the surrounding soil. Temperature changes in steel and surrounding soil during thermal pile operation will lead to additional steel stresses and displacements within the pile-soil system. Hence proper care has to be taken that the temperatures remain within acceptable limits, while the pile geotechnical analysis should demonstrate that any adverse thermal stresses are within design safety factors and that any additional displacements do not affect the serviceability of the offshore structure.

Keywords: ANSYS, Coupled temperature-displacement analysis, Drucker Prager Cap Model, Offshore Wind Turbine.

1. Introduction

Renewable energy is derived from natural sources that are replenished constantly. In its various forms, the renewable energy may be derived directly from the sun, wind or from heat generated deep within the earth. Moreover, electricity and heat may be generated from different types of renewable energy, e.g. solar, wind, ocean wave, hydropower, biomass, geothermal resources and biofuels. As per the world energy consumption report [1], the consumption of wind energy is only 0.51% and the consumption of geothermal energy is only

0.12%. The energy consumption from ocean wave is even lesser, only 0.001%. Thus there is lot of scope in increasing the consumption of these energy resources.

In offshore environment, the wind energy may be captured through offshore wind turbines whereas the geothermal energy may be extracted from the heat stored in the earth's crust and sea water. Thus, the total world energy consumption combining wind, ocean wave and geothermal energy comes down to a value which is even lesser than 1% of the total energy consumption. Hence, it is necessary to explore the possibilities of deriving electricity and heat from the renewable energy. In the proposed work, the possibility of using both the offshore wind turbine and its foundation in deriving energy from wind, ocean wave and geothermal resources will be investigated. A combined offshore wind turbine-geothermal pile technology will be studied for its response under random wind and wave loading on the turbine along with thermal loading-unloading of the geothermal energy pile foundation.

2. Combined Offshore-Geothermal-Wind Turbine System

A mono-pile foundation consists of a large-diameter steel pile, which is in principle simply a prolongation of the tower shaft into the ground. The mono-pile must be able to transfer both lateral and axial loads from the structure into the seabed. The steel piles are of simple tubular construction which is inexpensive to produce and provide a low cost fabrication option. In the present work, mono-pile foundation in form of energy pile extracts the geothermal heat from the sea crust and the heat is utilized in generating electricity. Energy piles in general contain high-density polyethylene pipes for carrying the fluid used for heat transfer as shown in Figure.2. The pipes circulate the geothermal fluid from the surface to the targeted depth and brings the heated fluid back to surface where a heat exchanger converts the temperature difference between the fluids and finally goes into a thermoelectric generator which works on the principle of Seebeck effect [2] and converts this temperature difference to power as shown in Figure.3. In the present study, a single tube is considered instead of a U-tube loop used in conventional energy piles for simplicity as shown in Fig.2 and the thermal loading is applied on the tube. Finally Lot of research has been reported in the literature [3-8] related to the response of the geothermal energy piles under heating and cooling operations. It is observed that the use of geothermal energy piles affect the load-displacement response of the piles, the axial stresses generated in the pile and the relative displacement at the pile-soil interface. Also, these piles may undergo significant uplift when subjected to heating. However, none of these studies have focused on the dynamic stress-strain response of the geothermal piles used in offshore applications.

3. Modeling of the Structure

3.1 Steel-Monopile-Heat Exchanger

A steel hollow monopile of diameter 7.5 m with 9 cm thickness is embedded 30 m below sea bed as shown in Fig. 1 has been considered as the energy pile. The steel hollow casing overlies the concrete grout with its mechanical properties shown in Table 1. The soil bed is considered to be heated upto 200^{o}C which is the minimum temperature required for extracting heat out of a geothermal reservoir and then cooled as the fluid is removed from the system. More details about the heating- cooling cycle has been discussed in section 4 of the paper. The time history of heating - cooling cycle has been presented in Figure 3. Assuming a water depth of 30 m and a maximum design wave height of 14.5 m, the design horizontal load for a monopile with a diameter of 7.5 m amounts to about 8 MN, the resultant horizontal force acting about 30 m above sea level, i e. nearly at still water level. Additionally a vertical load of 10 MN representing the own weight of the turbine, the blades and the tower was assumed. Such loads have to be considered analyzing the behaviour of monopiles. The current model

which involves characteristics of both offshore monopile and an energy pile has arrived after the successful validation of two separate models by Laloiu [3] and Achmus [9]. Appropriate offshore soil thermal parameters have been chosen after devoting significant amount of time in going through a large number of parameters of sea sediments.

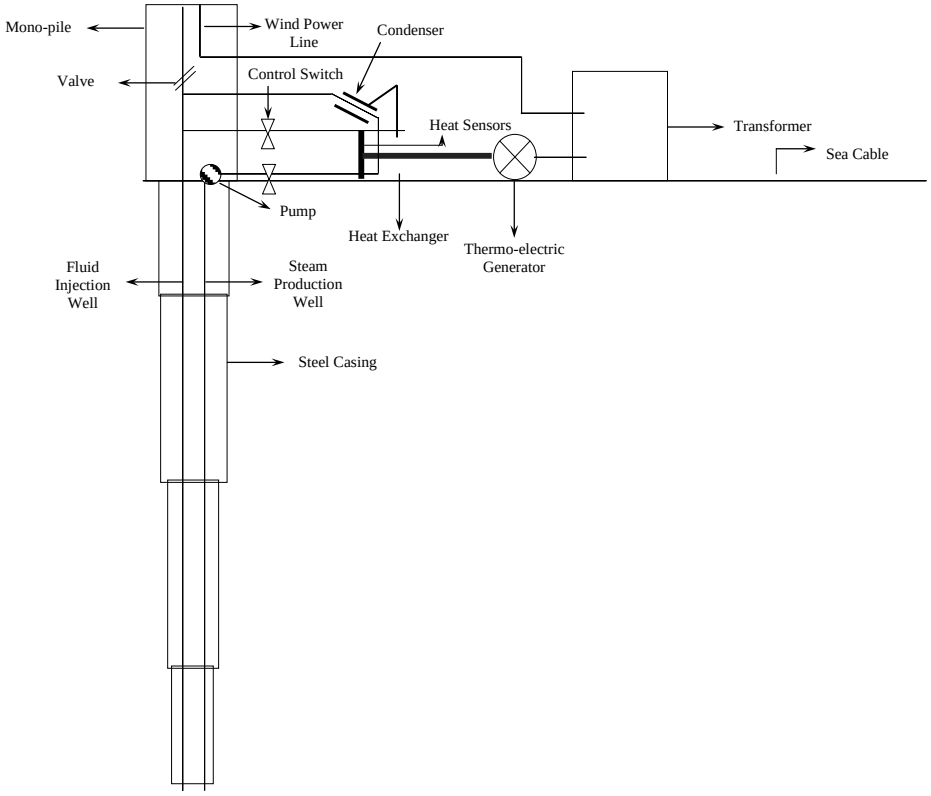


Figure 1. Schematic representation of offshore wind turbine geothermal system.

3.2 Finite Element Modelling

To simulate the behavior of the heat exchanger pile and the surrounding saturated soil, a numerical model, coupling the mechanical behavior to the thermal and hydraulic phenomena is needed. Here, a thermo-hydro-mechanical (THM) model for saturated porous media is used. Fully-coupled equations govern the evolution of pore water pressure, solid displacement and heat flow under mechanical, hydraulic and thermal loading. This can be implemented in ANSYS using the coupled poro-pressure element CPT215 which is based on Biot's theory of consolidation. The pile is modeled as a thermo-elastic solid using an eight noded SOLID185 brick elements. The thermo-mechanical data of steel has been presented in Table 4. Soil is assumed to have the thermo-poro-elastic properties of sand. The software used for model simulations in thermo-hydro-mechanical analysis is able to calculate simultaneously heat transfer from the grout to steel pile shaft and surrounding soil and the mechanical behavior of domains. An initial value of 80^{o}C was selected based on the ambient ground temperature at the targeted depth of geothermal energy heat extraction. Infinite boundary is simulated using spring and dashpots derived from Lysner[10] to simulate radiation damping of the soil. Only half of the pile and soil domain was considered as axi-symmetric condition. Soil and the pile domain is meshed using an element size of 0.75 m. The soil layer below the pile is modeled using a finer element size of 0.5 m. Element sizes coarser than 0.75 m gives an element shape distortion error. The structural and thermal properties of the steel pile has been presented in Table 1.

Initial stresses are generated in the soil medium using the INISTATE command in ANSYS. The computations were done using the finite element program system ANSYS .In order to carry out many calculations for loading conditions, a large computer system with parallel processor technology was used to minimize the time effort. The aim of the investigation was to analyse the behaviour of a large monopile under thermal loading-unloading cycle for energy storage.

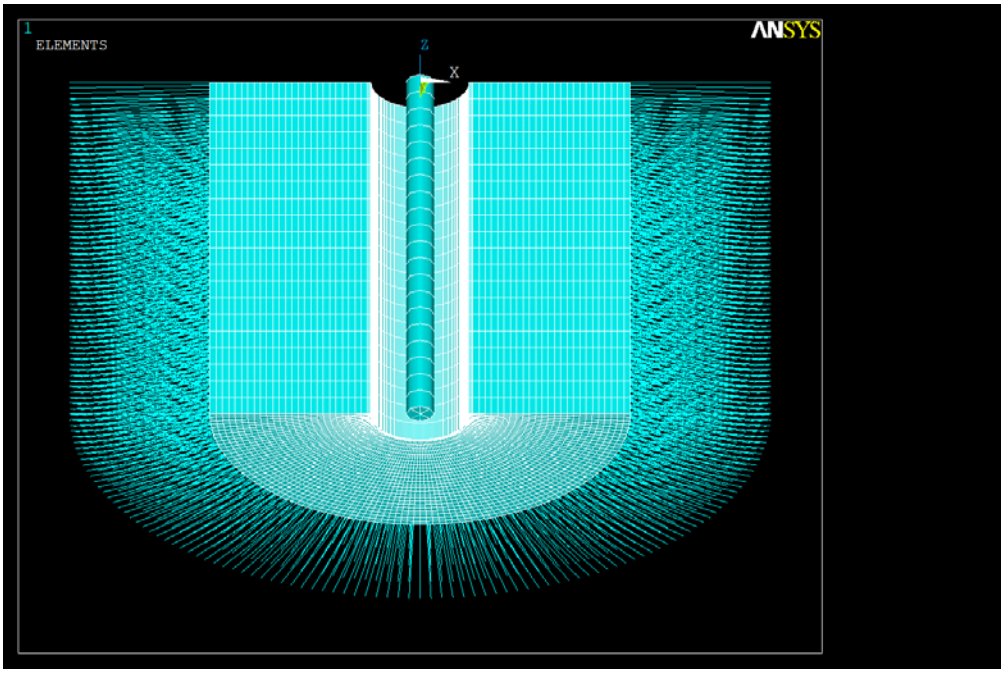


Figure 2. Finite element model in ANSYS.

Table.1 Summary of properties.

Steel density ρ	7850 kg/m^3
Thermal expansion coefficient steel	1.6×10^{-4}
Specific Heat Steel	419 J/kgK
Thermal conductivity	43 W/mK
Foundation radius r_0	3.75 m
Foundation depth z_0	40 m
Moment of inertia I_0	4.1368 m^4
Thermal expansion coefficient of sand	3.33×10^{-5}
Specific Heat of Sand	1090 J/kgK
Thermal conductivity of Sand	6 /mK

4. Thermal Loading

The load-settlement behavior of foundation piles directly impacts on the serviceability and safety of the structure above it. To determine the amount of pile displacement associated with cyclic thermal loading of energy piles, finite element simulation of an energy pile has been performed. The validity of the numerical analysis has been ensured by comparing the numerical simulation results with the field pile load test data and the results of numerical simulations performed by Laloui et al.[3]. Axisymmetric finite element analysis of piles in marine sand have been performed in two steps - (i) a static step to apply the gravity loading and to bring the model in geostatic equilibrium and (ii) a coupled temperature-displacement step to apply the thermal loading (iii) pore pressure loading. The thermal load is applied on the steel pile is generated by convection load due to fluid circulating through pipes. The heating-cooling load applied to the steel pile has been presented in Fig.3a.

5. Results and Discussion

Results are presented in the form of plots predicting changes in displacement, radial strain and axial stress in the steel pile under heating-cooling period.
Analysis has been carried out in three parts, where firstly an initial state load has been applied. Secondly a thermal loading has been applied for a duration of 16 days of heating and cooling cycle as shown in Fig.3a. Thirdly the pore pressure load has been applied on the steel pile structure. It is seen from Fig.3b that under thermal loading, the pore pressure keeps fluctuating with the highest value at the beginning of the heating period and then it keeps decreasing. The value of the pore pressure increases again in the beginning of the cooling period. After the initial loading has been applied, thermal load is applied on the structure and the thermal strain is shown in Fig.3c. The minor fluctuations in the curve shows the expansion and contractions throughout the heating and cooling period in the steel pile. Fig.3d shows the axial stress in the steel pile at three different depths of 0 m (pile head), 9 m and 36 m. It is seen from the figure that the highest axial stress occurs at the base of the pile. Fig.3e and 3f shows the translational and axial displacement in the steel pile throughout the heating and cooling period. The axial displacement at a depth of 9 m is almost half of that at the base of the steel pile. The axial displacement at the base of the pile follows a negative pattern as compared to that at the depth of 9m and that at the pile head. The radial displacement shown in Fig.3e shows that the displacement is almost same without too much of change for all the different depths.

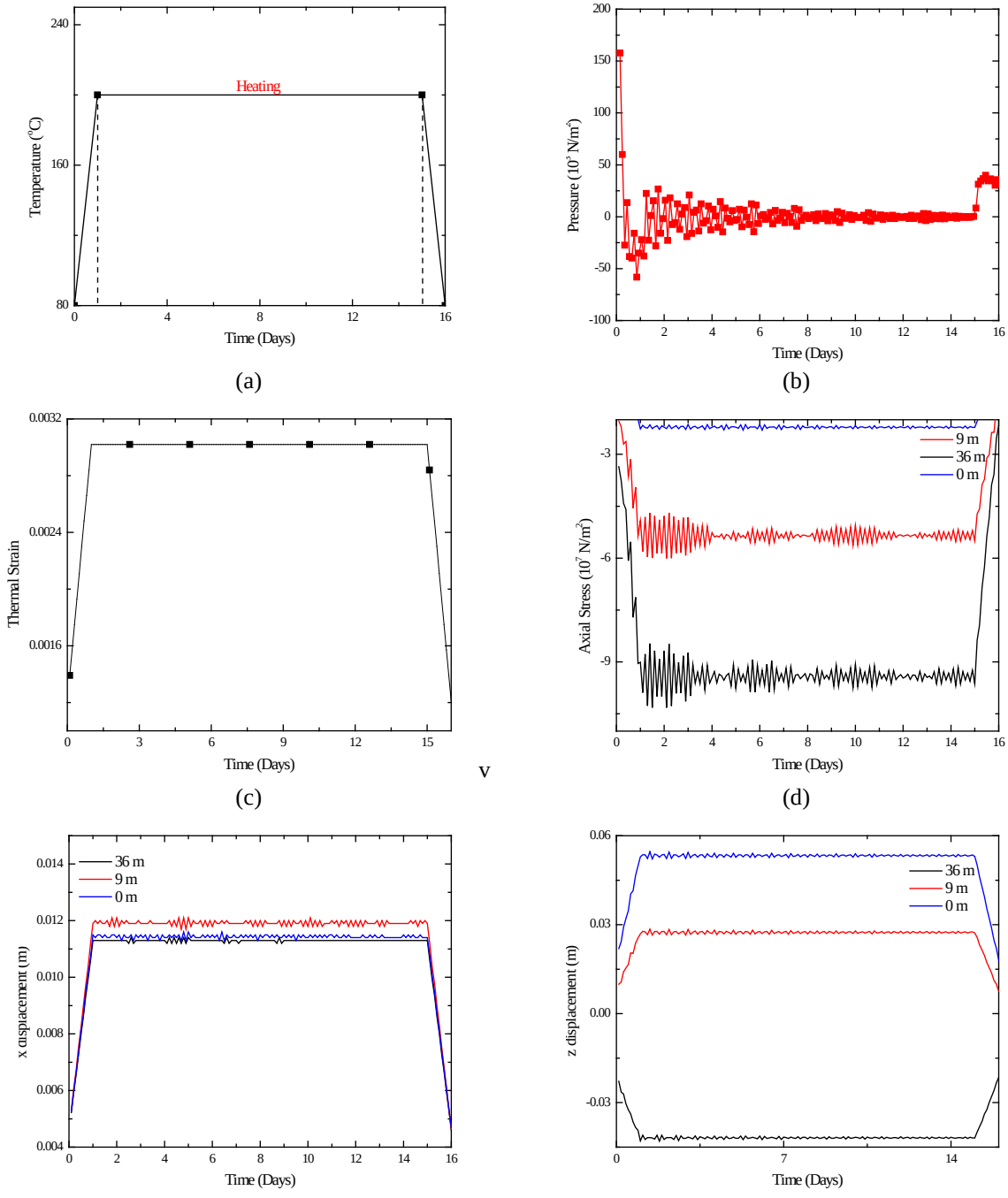


Figure 3. Results of (a) thermal load, (b) pore pressure under heating-cooling cycle , (c) thermal strain, (d) axial stress, (e) radial displacement, and (f) axial displacement.

References

[1] Renewables 2012 Global Status Report, http://www.nrel.gov/docs/fy08osti/41958.pdf, seen on 26th September 2014.

[2] Thomson, William (1851).On a mechanical theory of thermoelectric currents, Proceedings of .Royal .Society of .Edinburgh, 91-98.

[3] Laloui, L., Nuth, M.,Vulliet, L. (2006) “Experimental and numerical investigations of the behaviour of a heat exchanger pile.”International Journal for Numerical and Analytical Methods in Geomechanics, Vol.30(8), pp. 763-781.

[4] Raymond, J., Therrien, R., Gosselin, L., Lefebvre, R. (2011) Numerical analysis of thermal response tests with a groundwater flow and heat transfer model, Renewable Energy, 36, 315-332.

[5] Abdelaziz, S., Olgun, C., and Martin, J. (2011) Design and Operational Considerations of Geothermal Energy Piles, Proceedings of the American Society of Civil Engineers, Geo-Frontiers, Dallas, USA, 48, 450-459.
[6] Diersch, H, Bauer, D, Heidemann, W., Rühaak, W., Schätzl, P. (2011) Finite element modeling of borehole heat exchanger systems - part 1: fundamentals, Computational Geosciences,37, 1122-1135.
[7] Ping, C., Xin,C., Yi-Man, Zhaohong, F. (2011) Heat transfer analysis of pile geothermal heat exchangers with spiral coils, Elsevier Applied, 88, 4113-4119.
[8] Suryatriyastuti, M.E., Mroueh,H., Burlon S. (2014) A load transfer approach for studying the cyclic behavior of thermo-active pile, Computers and Geotechnics, 55, 378-391.
[9] Achmus, M., Abdel-Rahman, K. & Kuo, Y.-S. (2007). Numerical Modelling of large Diameter Steel Piles under Monotonic and Cyclic Horizontal Loading, 10th International Symposium on Numerical Models in Geomechanics, Greece, 453-459.
[10] Lysmer J, Kuhlemeyer R L.(1969) Finite dynamic model for infinite media. Journal of Engineering Mechanics Division ASCE 95: 850- 877.

Model Free Deep Learning With Deferred Rewards For Maintenance Of Complex Systems.

***Alan DeRossett[1], Pedro V Marcal[2]** , Inc.,
[1]Boxx Health Inc.., 3538 S. Thousand Oaks Blvd., Thousand Oaks CA. 91362
[2]MPACT Corp.,5297 Oak bend Lane, Suite 105, Oak Park, CA. 91377
*Presenting and Corresponding author : alan@mb1.com

Abstract

This paper reviews progress in Deep Learning and the successful Application of Deep Q Networks to Competitive Games. The basis of the method is Watkins' Deferred Rewards Learning[1]. However its implementation in its current form is due to D. Hassabis et al [2]. In its current form the technology is heavily dependent on image processing. This suggests that the method may be adapted for the establish ment of Optimal Maintenance Policies for Complex systems such as Airliners and/or racing cars with the addition of monitoring of Sound.

Keywords : deep learning, Deferred Rewards Learning, game theory, maintenance of complex systems.

Introduction

The authors started off by investigating the recent explosive growth of the GPU in numerical processing. We soon discovered that the Deep Learning Community provided the largest growth in using the GPU. In fact it may be said that progress in Deep Learning owes its recent progress to massive computing which was able to achieve the scale necessary to solve problems in imaging. Two programs that enabled this was Theano [3] from the University of Montreal and the recent open source system from Google, Tensor Flow [4]. Theano may be looked upon as a system for code optimization and deployment on GPUs. The program was developed and used for neural network problems. The Tensor Flow program achieves the same means by allowing itd users to make use of a data flow model. Once users cast their algorithms in data flow format. The program will allow the user to bring all the available computing power to bear on the data flow object to be completed. The goal of massive parallelization has been elusive. For a long time the people in this audience have tried to apply it to FEM for example. The existence of multiple core machines have sped up the solution of our problems. However the Artificial Neural Network Community has shown us that concurrency is much simpler when problems are increased by two orders of magnitude. It is interesting to speculate that the FEM community may be able to take advantage of the two programs for FEA. In order to achieve our objective of explaining the technology behind the success of DQN [2] and its spectacular achievement of beating the world Go Champion, we will start introduce the three technologies behind it. We should also note that one of us (ADR) has spent considerable time and effort downloading and installing Theano and Tensor Flow on local computers as well as on the Cloud.

Theoretical Considerations.

Deep Learning

Nielsen [5] has pre-released the first Chapter of a book in preparation by Bengio et al [6] that explains the theory of deep learning applies it to decyphering handwritten numbers The exposition is particularly instructive because of the demonstration of the theory in the form of a Python computer program.

Following [5] we start by defining a sigmoid network.

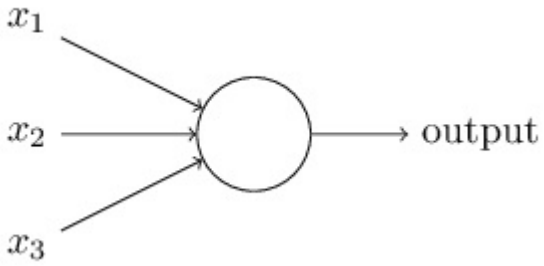


Fig. 1 Sigmoid neuron

The neuron has 3 inputs x and a bias b (scalar) with weights w.
The input to the neuron is w.x+ b. The output is modified by the sigmoid or logistic function written as,

$$\frac{1}{1+\exp(-\sum_j w_j x_j - b)}.$$

Fig.2 Sigmoid function

The sigmoid function σ (z) has the desirable property of small input changes resulting in small and smooth changes in output.
In deep networks, we introduce additional layers between the input and output.

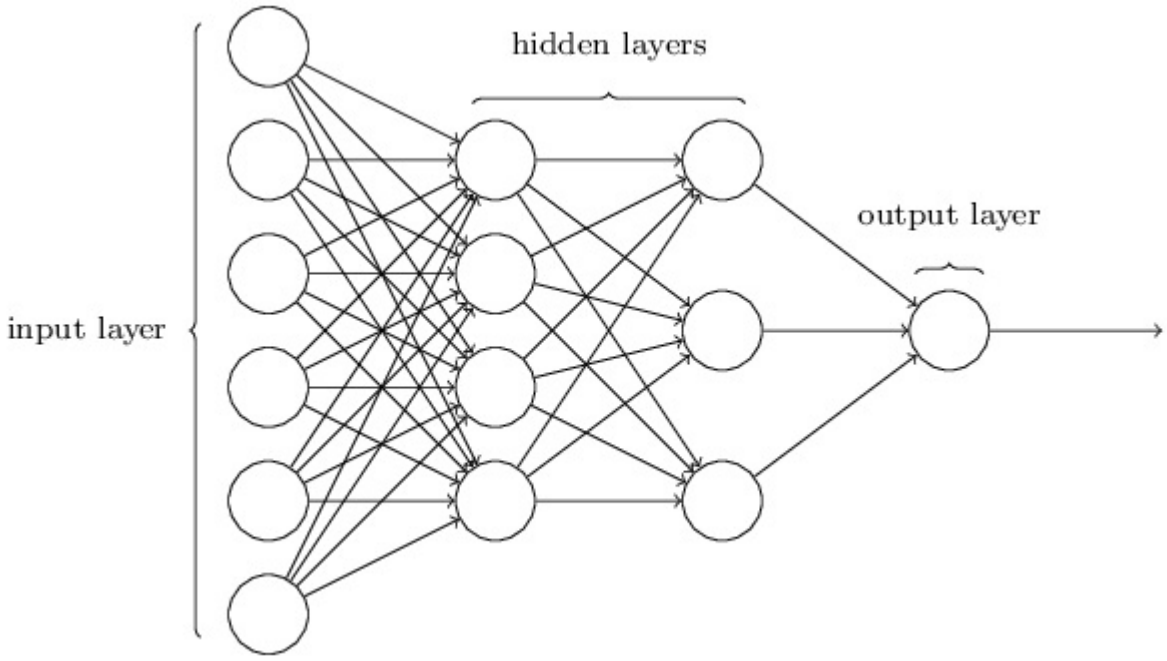


Fig. 3 Deep network with hidden layers (2).
We introduce an additional index to indicate the layer, then the input to a neuron in a hidden layer is $z_{ij} = \Sigma\ x_{ij}.w_{ij} + b_{ij}$ and its output is $\sigma_{ij}(z)$ for j= 0 to d layers where d is the output layer.

Hence an input x_{i0} results in a nonlinearly mapped output $\sigma_{id}(z)$.
The network is defined by a number of training inputs x with n outputs y(x).
We define a cost function $C(w,b)= 1/2\ n\ .\ \Sigma\ (y_{id}-\sigma_{id}(z))^2$
then the objective is to vary the weights w and biases b to reduce the value of C for all the members of the training set. So we see the training problem as the minimization of the quadratic function C. The numerical procedure for solving such problems is well known. However the most used method is that of back propagation of errors. As mentioned earlier, we are at a stage where the combined power of software and hardware can be applied to large problems with many layers (typically up to 10). I think of this process as building a large interpolation function so that any other input x will result in an output that conforms to the output function defined by the input output set x, y respectively. This concludes our brief discussion of deep learning.

Learning With Delayed Rewards
In his thesis[1], Watkins investigated animal behavior in both the wild and under lab controlled condition. There is a rich set of experiences and theories. One may view an animal's reaction as an intelligent response to a set of stimuli as an immediate as well as a long term reaction with a view to survival, (the ultimate response). Watkins framed the problem as a number of variables x_i with state a_{it} at time t. At any stage the problem was assumed to be a Markov process. The problem is framed in turns of a cycle or epoch in terms of which all the states are changed in sequence according to some set policy. Initially this could even be a random one. The time step within the epoch is assumed to take n steps. At a step t we assume that some action $Q(a_{it})$ results in some reward r_{it} at every time step but depreciated by a factor γ. Because $\gamma < 1$ the return tends to 0 with n the number of steps being large. Hence we have the n-step truncated return given by a change in a_i at time t.

$$r_t^{[n]} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots + \gamma^{n-1} r_{t+n-1}$$

The rewards for actions $Q(a_{it})$ over n steps is given by the shifted n-step rewards

$$R_n = \Sigma\ r_{it}^{[n]t}$$

At this stage the framing of the problem has introduced a further number of unknowns. We note that the state a changes with every action Q. The reward function r is not known and the policy for selecting Q is also not known.The problem is however given specificity by selecting the changes $Q(a_{it})$ by the principle of dynamic programming(DP).[7]. Watkins showed that with DP the action Q can be achieved iteratively and Watkins and Dayan[8] gave further proof governing the iterative procedure. Watkins also assumed that the reward function could also be defined by repeated observation of the actual game over time. The thesis did not go into the application of the theory. One must assume that a large amount of numerical calculations must have been performed for Watkins to be able to speak so authoritatively on the problem. The reader is referred to [9] to see a tutorial on learning with deferred rewards.

Theory of Games with deep learning and DQN.
It was up to Hassabis and his colleagues at Deep Mind[2] to bring substance to Watkins' methods. Hassabis reformulated the problem in neural network terms. This was demonstrated by applying it to a whole set of Atari Games. In many ways the Atari games were the perfect form as a project. The games already had a scoring system that provided the Reward Function R and the games were built for a user to specify an action Q at any stage. The pixels on the screen were used as input. They were turned into values by applying a convolution mapping to the screen. The input state a_{i0} were defined by the screen image. The Q function was defined as a square matrix giving all the possible combination of the

changes in the state a sequence with time. The preprocessing for input to the neural net is shown in Fig. 6. The Q functions were given by the coding on the right while the conversion to convolutions were obtained on the left side. Both results were fed into the neural network.

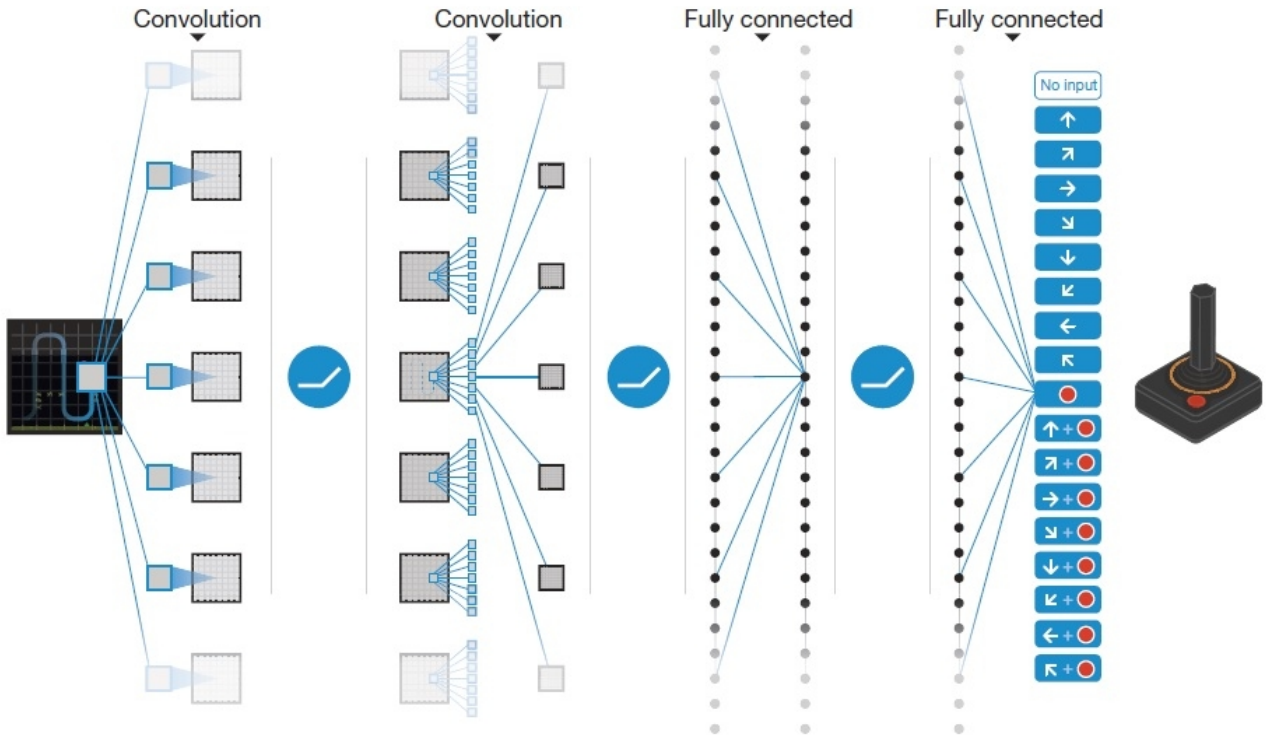


Fig. 6 Schematic for preprocessing raw input for input to the neural net.

The project proceeded in two phases. In the first training phase the program was set to collect a massive amount of data resulting from changes in the Q functions for different starting state values a. The program stores all the experiences in a database D_e where e is the total sum of recordings for training a particular game. The training of DQN networks is known to be unstable. In the second phase the database D was used to train the neural net in what is known as a experience replay developed in [2] and using a biologically inspired mechanism that randomizes over the data, thereby removing correlations in the observation sequence and smoothing over changes in the data distribution. The second improvement used an iterative update that adjusts Q towards targets that are only periodically updated.

Results

Here we show the results for the Space Invader game using the two useful metrics of average score and averagepredicted action-value Q.

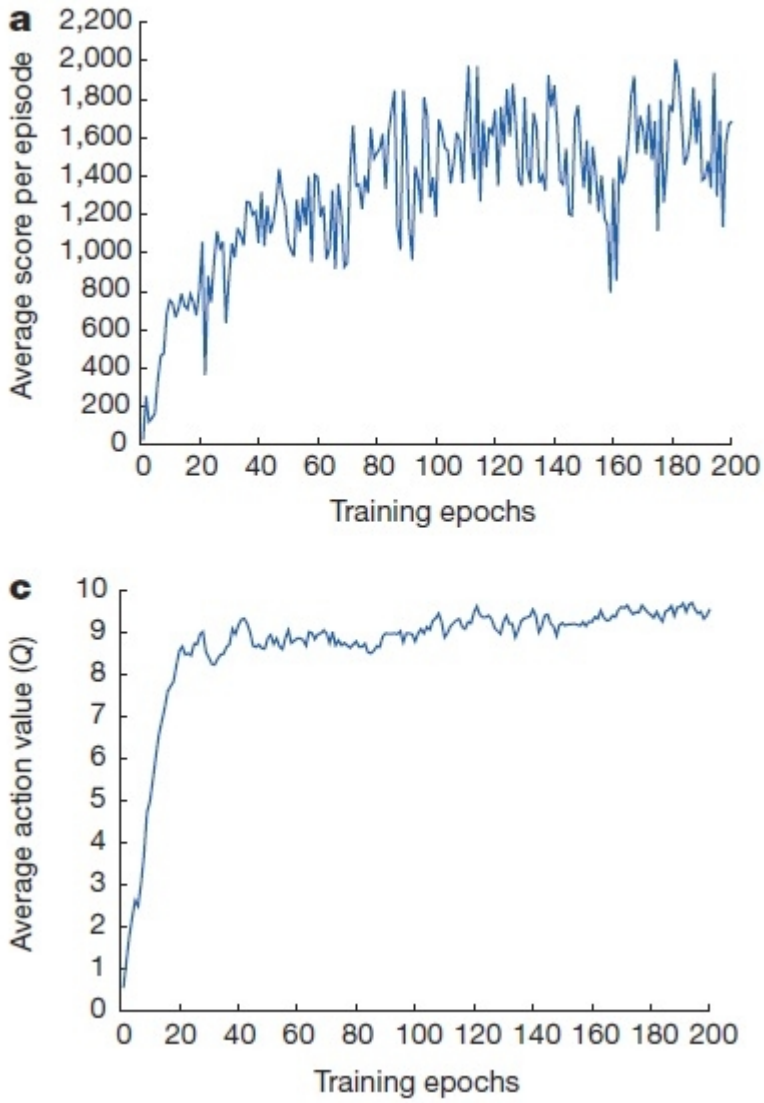


Fig. 7 Scores for space invader.

We conclude by noting that the results for all the Atari Games beat the results obtained by specialized computer playing games (tailored to one game).Hence the project even at this first step, proved the power of the method for implementing the DQN as a deep learning network.

Possible Applications in Complex Engineering Systems.
Such a general approach to applying DQN to Model Free problems has many potential applications. So we should pick the more important problems. One such problem that requires little alteration in the Lua Open Source program provided by the authors could be in the study of the maintenance problem in Airlines. Currently the fleet is overhauled and serviced on a regular basis. At such times parts are examined and sometimes replaced. The maintenance actions have an impact on the performance of an aircraft. Perhaps this improvement can be detected by video cameras that record the visual performanc of the plane during taxiing and parking. Such records could replace those captured for the Atari Games. In another similar vein, we could probably also record the roar of the racing car engines and add this to figure out the best maintenance policy.

References

[1] CJCH Watkins, 'Learning From Delayed Rewards', Ph. D. Thesis, King's College, Cambridge, 1989
[2] Google Deep Mind[1]
Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]
'Human-level Control Through Deep Reinforcement Learning', Nature, Vol 518, February, 2015
[3]. I. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley and Y. Bengio. "Theano: A CPU and GPU Math Expression Compiler".*Proceedings of the Python for Scientific Computing Conference (SciPy) 2010. June 30 – July 3, Austin, TX* .
[4] Google Brain Team,'Tensor Flow', Open source, 2016.
[5] Michael Nielsen,'Deep Learning', Chapter 1. of [6], January, 2016
[6] Y. Bengio, I. Goodfellow, A. Courville, 'Deep Learning', MIT Press, (in Press), 2016.
[7] R.E. Bellman and S.E. Dreyfus, 'Appied Dynamic Programming', Rand Corp., 1962.
[8] CJCH. Watkins and P. Dayan, 'Q-Learning', Machine Learning, 8, 272-279, 1992.
[9] 'A Painless Q Learning Tutorial',http://mnemstudio.org/path-finding-q-learning-tutorial.htm, 2016

Small defining sets in $n \times n$ Sudoku squares

†**Mohammad Mahdian**[1] **and** ***Ebadollah S. Mahmoodian**[2]

[1]Google Research, Mountain View, CA, USA
[2]Department of Mathematical Sciences, Sharif University of Technology, Tehran, I. R. Iran

*Presenting and Corresponding author: emahmood@sharif.edu

Abstract

Over the last decade, Sudoku, a combinatorial number-placement puzzle, has become a favorite pastimes of many all around the world. Recently it is shown that this concept has many mathematical and computational relations and applications. In this puzzle, the task is to complete a partially filled 9 by 9 square with numbers 1 through 9, subject to the constraint that each number must appear once in each row, each column, and each of the nine 3 by 3 blocks. Sudoku squares can be considered a subclass of the well-studied class of Latin squares. Actually a Sudoku square of order $n = k^2$ is a Latin square of order n such that every element in $[n] = \{1, \ldots, n\}$ appears exactly once in each block. A partial Sudoku square P is a defining set for a Sudoku square S if S is the unique Sudoku square that is an extension of P. A central problem is to determine the size of the smallest defining set for Sudoku squares of order n. For $n = 9$ (regular Sudoku) extensive computer search showed that this number is 17 (McGuire et al, 2014), but the asymptotics of this value is unknown. For Latin squares, this number is conjectured to be $\lfloor n^2/4 \rfloor$ (Mahmoodian 1995, Van Rees and Bates 1999). A construction based on back-circulant Latin squares shows that this number is at most $\lfloor n^2/4 \rfloor$, but the best proven lower bound is just slightly superlinear. Also, the $\lfloor n^2/4 \rfloor$ conjecture is proved if "defining set" is replaced by a more strict notion called "forcing set".

For Sudoku squares, we show that the same construction (with a permutation on the rows of the matrix) works, giving an upper bound of $\lfloor n^2/4 \rfloor$. We also show that the size of the smallest forcing set for Sudoku squares of order n is at least $\Theta(n^2)$. Our conjecture is that the size of the smallest defining set for Sudoku squares of order n is also $\Theta(n^2)$. Finally, we discuss open problems related to Sudoku squares, their defining sets, and the computational complexity of Sudoku completion.

Keywords: Computation, Sudoku, Latin squares, defining set, forcing set, extension.

Introduction

A Latin square of order n is an $n \times n$ matrix with entries from $[n] = \{1, \ldots, n\}$ such that every element in $[n]$ appears exactly once in each row and in each column.

A partial Latin square of order n is an $n \times n$ matrix with entries from $[n] \cup \{*\}$ such that every element in $[n]$ appears at most once in each row and in each column. A partial Latin square P_1 is an extension of a partial Latin square P_2 if for every $(i, j) \in [n]^2$, if $P_2(i, j) \neq *$, then $P_1(i, j) = P_2(i, j)$.

A partial Latin square P is a defining set for a Latin square L if L is the unique Latin square that is an extension of P. A critical set is a minimal defining set. A forcing set (also called a strong critical set) is a partial Latin square P such that there is a sequence $P = P_0, P_1, \ldots, P_\ell$ such that P_ℓ is a Latin square and for every r,

- P_r is a partial Latin square and an extension of P_{r-1},

- the difference between P_r and P_{r-1} is in precisely one entry, i.e., there is $(x, y) \in [n]^2$ such that $P_r(i, j) = P_{r-1}(i, j)$ for every $(i, j) \neq (x, y)$ and $P_{r-1}(x, y) = *$ and $P_r(x, y) \neq *$, and
- for every $z \in [n]$ and $z \neq P_r(x, y)$, the matrix obtained from P_r by setting $P_r(x, y)$ to z is not a partial Latin square.

In a Latin square of order $n = k^2$, the (i, j)'th block (for $i, j \in [k]$) is the set of entries with coordinates in $((i-1)k+x, (j-1)k+y)$ for $x, y \in [k]$. We say that (i, j) are the coordinates of this block. These blocks partitions the set of entries in the matrix into n blocks, each containing n entries. A Sudoku square of order $n = k^2$ is a Latin square of order n such that every element in $[n]$ appears exactly once in each block. We say that the (i, j)'th block belongs to the i'th row block and the j'th column block.

Notions of partial Sudoku square, extensions of a partial Sudoku square, defining sets, critical sets, and forcing sets for Sudoku squares can be defined similarly.

A central problem is to determine the size of the smallest defining set for Sudoku squares of order n. For $n = 9$ (regular Sudoku) extensive computer search showed that this number is 17 (McGuire et al [6]), but the asymptotics of this value is unknown. For Latin squares, this number is conjectured to be $\lfloor n^2/4 \rfloor$ (Mahmoodian [5], Bate and Van Rees [1]). A construction based on back-circulant Latin squares shows that this number is at most $\lfloor n^2/4 \rfloor$, but the best proven lower bound is just slightly superlinear. Also, the $\lfloor n^2/4 \rfloor$ conjecture is proved if "defining set" is replaced by "forcing set". For Sudoku square, we show that the same construction (with a permutation on the rows of the matrix) works, giving an upper bound of $\lfloor n^2/4 \rfloor$. We also show that the size of the smallest forcing set for Sudoku squares of order n is at least $\Theta(n^2)$. Our conjecture is that the size of the smallest defining set for Sudoku squares of order n is also $\Theta(n^2)$. We conclude with the discussion of many Sudoku-related problems that remain open.

Lower bound on the size of forcing sets

In this section, we prove the main result of this paper, which is the following lower bound on the size of the smallest forcing set in Sudoku squares. This result, combined with the observation that essentially the same construction as the one for back-circulant Latin squares gives us a forcing set of size $\lfloor n^2/4 \rfloor$ for an equivalent Sudoku square, shows that the smallest forcing set of Sudoku squares of order n is precisely $\Theta(n^2)$.

Theorem 1 *For every n, the size of the smallest forcing set for Sudoku squares of order $n = k^2$ is at least $\Omega(n^2)$.*

Proof. Let F be a partial Sudoku square that is a forcing set, and consider the forcing order on the entries not specified by F. Let S denote this ordering, i.e., S_1 is an entry that is forced by F, S_2 is an entry that is forced by $F \cup \{S_1\}$, and so on.

We start by defining a subsequence S' of S as follows: $S'_1 = S_1$, and for every $i > 1$, S'_i is the first element in S after S'_{i-1} that is not in the same row, the same column, or the same block as any of $S'_1, S'_2, \ldots, S'_{i-1}$. In other words, S' is obtained from S by removing elements that are in the same row, same column, or same block. Therefore, the sequence S' has at most n elements, and contains at most one element from each row, each column, and each block of the Sudoku square.

We now transform S' into an ordering of a $k \times k$ square. More formally, we define a permutation π of the set $[k]^2$ as follows: for every i where S'_i is defined, π_i is the coordinates of the block

containing S'_i. Since S' contains at most one element from each block, the π_i's defined based on S'_i's are distinct. There can be blocks with no element present in S'; we add the coordinates of such blocks in an arbitrary order to the end of π. This completes the definition of the permutation π of $[k]^2$. The proof of the theorem is based on two lemmas. The first lemma bounds the size of the forcing set in terms of a quantity associated with the permutation π, and the second lemma bounds this quantity for every such permutation.

To state the first lemma, we need a few notations. For every permutation π of $[k]^2$ and every $u, v \in [k]^2$ ($u \neq v$), we say $u \prec_\pi v$ if u comes before v in π. Let $B^r_\pi(v)$ denote the number of $u \in [k]^2$ such that $u \prec_\pi v$ and u and v are in the same row (i.e., $u = (i, j)$ and $v = (i, j')$ for $i, j, j' \in [k]$). Similarly, let $B^c_\pi(v)$ denote the number of $u \prec_\pi v$ that are in the same column as v. Finally, let $B_\pi(v) = B^r_\pi(v) + B^c_\pi(v)$. We are now ready to state the first lemma.

Lemma 1 *Let π be the permutation defined based on a forcing set F using the above procedure. Then,*

$$|F| \geq \sum_{i=1}^{n} \max(0, n + 1 - 2i - (2k - 2)B_\pi(\pi_i)).$$

Let $L(\pi)$ denote the quantity on the right-hand side of the inequality in Lemma 1. The second lemma bounds this quantity for every permutation π.

Lemma 2 *There is a constant c such that for every permutation π of $[k]^2$, we have $L(\pi) \geq cn^2$.*

We start by proving the first lemma.

[Proof of Lemma 1] Let $i \in [n]$ be an index for which S'_i exists. Therefore, π_i is the coordinates of the block containing S'_i. We argue that to uniquely force S'_i, we need at least $n + 1 - 2i - (2k - 2)B_\pi(\pi_i)$ *new* elements in F (i.e., elements other than the ones needed to force S'_j for $j < i$).

Let A_i denote the set of entries of the Sudoku square that are in the same row, same column, or the same block as S'_i. The following lemma bounds the cardinality of the intersection of these sets.

Lemma 3 *For every i, j, $j \neq i$, if S'_i and S'_j are not in the same row block or the same column block, then $|A_i \cap A_j| = 2$. If they are on the same row block or same column block, then $|A_i \cap A_j| = 2k$.*

Proof. Proof is easy. Omitted for now.

Since F is a forcing set, by the time S'_i is forced, there must be at least $n-1$ entries in A_i whose values are uniquely specified. We argue that out of these $n-1$, at most $2(i-1)+(2k-2)B_\pi(\pi_i)$ are either forced in previous steps or already counted in F, and therefore there must be at least $n + 1 - 2i - (2k - 2)B_\pi(\pi_i)$ of them that are in F and are not previously counted in F. Note that any entry that is either forced in previous steps or already counted in F must be in A_j for a $j < i$. This is because any entry that is forced before S'_i must either be present in the sequence $S'_1, \ldots, S'_{i-1}$, or be in the same row, same column, or same block as one of the elements of this sequence. Either way, this element belongs to $\bigcup_{j<i} A_j$. Also, in each step $j < i$, we

count elements of F that are used to force S'_j, and these elements belong to A_j. Therefore, the number of elements in A_i that either forced before S'_i or are already counted in F is at most $|A_i \cap (\bigcup_{j<i} A_j)|$. To bound this cardinality, we use Lemma 3. By this lemma and the definition of $B_\pi(\pi_i)$, the value of $|A_i \cap A_j|$ is equal to $2k$ for precisely $B_\pi(\pi_i)$ values of j and is equal to 2 for the remaining $i - 1 - B_\pi(\pi_i)$. Therefore,

$$|A_i \cap (\bigcup_{j<i} A_j)| \leq \sum_{j<i} |A_i \cap A_j| = 2kB_\pi(\pi_i) + 2(i - 1 - B_\pi(\pi_i)).$$

Therefore, there must be at least $\max(0, n + 1 - 2i - (2k - 2)B_\pi(\pi_i))$ elements in F that are used to force S'_i and are not counted in previous steps.

Next, we consider i's for which S'_i does not exist. Recall that when the length of S' is less than n, we append a list of block coordinates that contain no element of S' at the end of π in an arbitrary order. Therefore π_i is the coordinate of a block none of whose elements appears in S'. This means that all of the n elements of the block at coordinates π_i must either be in F, or in the same row, column, or block as an element of S', since otherwise they would have been included in S'. We can now repeat the same argument with A_i replaced by the set of entries in the block at coordinates π_i.

Putting these cases together, we get that in total F must contain at least $\sum_{i=1}^{n} \max(0, n + 1 - 2i - (2k - 2)B_\pi(\pi_i))$ elements.

Next, we prove Lemma 2, which gives a bound on the quantity $L(\pi)$ for every permutation π of $[k]^2$.

[Proof of Lemma 2] Let $\alpha \in [0, 1]$ be a parameter that will be fixed later. For convenience we assume that $(1 - \alpha)k/2$ (and therefore $(1 - \alpha)n/2$) is an integer. We use the following lower bound on $L(\pi)$:

$$L(\pi) \geq \sum_{i=1}^{(1-\alpha)n/2} \max(0, n + 1 - 2i - (2k - 2)B_\pi(\pi_i)).$$

Since for every $i \leq (1 - \alpha)n/2$, we have $n + 1 - 2i > \alpha n$, the above inequality implies:

$$L(\pi) \geq \sum_{i=1}^{(1-\alpha)n/2} \max(0, \alpha n - (2k - 2)B_\pi(\pi_i)).$$

For every i, we define

$$L(\pi, i) = \begin{cases} 0 & \text{if } \max\{B^r_\pi(\pi_i), B^c_\pi(\pi_i)\} > \frac{\alpha n}{4k-4} \\ \alpha n - (2k - 2)B_\pi(\pi_i) & \text{otherwise.} \end{cases}$$

It is easy to see that $\max(0, \alpha n - (2k - 2)B_\pi(\pi_i)) \geq L(\pi, i)$ for every i. Therefore,

$$L(\pi) \geq \sum_{i=1}^{(1-\alpha)n/2} L(\pi, i).$$

Let $L'(\pi)$ denote the right-hand side of the above inequality. We will show how the permuta-

tion π can be transformed into a structurally simpler permutation π' such $L'(\pi) \geq L'(\pi')$. Let $t = \lfloor \frac{\alpha n}{4k-4} \rfloor$. Consider the smallest index i such that $\max\{B^r_\pi(\pi_i), B^c_\pi(\pi_i)\} = t$, and assume, without loss of generality, that $B^r_\pi(\pi_i) = t$. This means that there are t indices $i_1, i_2, \ldots, i_t = i$ such that π_{i_ℓ}'s, for all $\ell = 1, \ldots, t$, are on the same row in $[k]^2$. It is not hard to see that moving all these π_{i_ℓ}'s to the beginning of the permutation does not change the value of $L'(\pi)$. Furthermore, all other elements of the same row can be added after these elements without increasing $L'(\pi)$. Therefore, by moving all entries that are on the same row as π_i to the beginning of the permutation, we obtain another permutation whose L' value is not more than the L' value of the original permutation. We can continue this process, by finding the first index i' such that $\max\{B^r_\pi(\pi_{i'}), B^c_\pi(\pi_{i'})\} = t$ and $\pi_{i'}$ is not on the same row as π_i. Using the same argument, depending on whether $B^r_\pi(\pi_{i'}) = t$ or $B^c_\pi(\pi_{i'}) = t$, elements of the row or column of $\pi_{i'}$ (except possibly the ones that were on the same row as π_i) can be moved right after the elements of the row of π_i. Continuing with this process, we can build a permutation π' such that $L'(\pi) \geq L'(\pi)$, and π' has the following structure: it starts with the list of all elements of a row/column of $[k]^2$, then all elements of another row/column of $[k]^2$ except the ones that have appeared before, and so on.

What remains is to prove that for a permutation π' that has the above structure, $L'(\pi') = \Omega(n^2)$. Using the structure of π', we can decompose it into segments, where each segment lists all elements of a row/column of $[k]^2$ except the ones that are listed that are listed in previous segments. We call a segment a row/column segment, depending on whether it is a list of elements in a row or a column of $[k]^2$. The value of a segment is the sum of $L(\pi', i)$ for all i that belong to that segment. Let ℓ^r_j (ℓ^c_j, respectively) denote the number of row (column, respectively) segments *before* the j'th segment. Therefore, if the j'th segment is a column segment, its value can be written as:

$$\begin{aligned} V_j &= (\alpha n - (2k-2)\ell^r_j) + (\alpha n - (2k-2)(\ell^r_j + 1)) + \cdots + (\alpha n - (2k-2)t) \\ &= (t - \ell^r_j + 1)(\alpha n - (k-1)(t + \ell^r_j)), \end{aligned} \tag{1}$$

if $\ell^r_j \leq t$. We also have $V_j = 0$ if $\ell^r_j > t$. If the j'th segment is a row segment, we get a similar expression for V_j, with ℓ^r_j replaced by ℓ^c_j.

Since each segment contains at most k elements, there are at least $\frac{(1-\alpha)n}{2k} = (1-\alpha)k/2$ segments that are entirely contained in the first $(1-\alpha)n/2$ elements of π. Therefore, $L'(\pi')$ is at least the sum of the values of the first $(1-\alpha)k/2$ segments, i.e., $L'(\pi') \geq \sum_{j=1}^{(1-\alpha)k/2} V_j$. We let $L''(\pi') := \sum_{j=1}^{(1-\alpha)k/2} V_j$.

The final step is to change π' to another permutation π'' (with a similar segmented structure) such that $L''(\pi') \geq L''(\pi'')$. We do this as follows: assume, for some j, $\ell^r_j > \ell^c_j$ and the j'th segment is a row segment. Find the smallest index $j' \in [j, (1-\alpha)k/2]$ such that the j''th segment is a column segment, if such an index exists. We can write down the difference in the total L'' value if we replace the order of the segments j' and $j'-1$ (i.e., first list all elements in the column corresponding to segment j' and then list all elements in the row corresponding to segment $j'-1$). It is easy to see that the inequality $\ell^r_j > \ell^c_j$ implies that this swap cannot increase the L'' value of the permutation. If such an index j' does not exist, we can change the last segment to a column segment. Again, it is not hard to see that the assumption $\ell^r_j > \ell^c_j$ implies that this change does not increase the L'' value of the permutation. Similar statements hold if we switch the role of row segments and the column segments. By repeatedly using this

procedure, we get a permuation π'' that consists of alternating row and column segments, and satisfies $L''(\pi') \geq L''(\pi'')$.

All that remains is to write down the value of $L''(\pi'')$. This permutation satisfies $\ell_j^r = \ell_j^c = \lfloor j/2 \rfloor$ for j odd and $\ell_j^r = \lfloor j/2 \rfloor = \ell_j^c + 1$ for j even. Using Equation (1), the value of $L''(\pi'')$ can be written as follows:

$$\begin{aligned} L''(\pi'') &= \sum_{s=0}^{p}(t-s+1)(\alpha n-(k-1)(t+s))+ \\ &\quad \sum_{s=0}^{p}(t-s+1)(\alpha n-(k-1)(t+s)) \\ &\geq 2\sum_{s=0}^{p}(t-s+1)(\alpha n-(k-1)(t+s)), \end{aligned}$$

where $p = \min\{t, \lfloor (1-\alpha)k/4 \rfloor\}$. Recall that $t = \lfloor \frac{\alpha n}{4k-4} \rfloor$. Therefore,

$$\begin{aligned} L''(\pi'') &\geq 2(k-1)\sum_{s=0}^{p}(t-s)\left(\frac{\alpha n}{k-1}-t-s\right) \\ &\geq 2(k-1)\sum_{s=0}^{p}(t-s)^2. \end{aligned}$$

If we pick α in such a way that $t \leq \lfloor (1-\alpha)k/4 \rfloor$, we have $p = t$ and therefore,

$$L''(\pi'') \geq \frac{2(k-1)t^3}{3} \geq \frac{2\alpha^3 k^4}{3 \cdot 4^3}.$$

Now, it suffices to pick any $\alpha < 1/2$. It is easy to see that this satisfies the inequality $t \leq \lfloor (1-\alpha)k/4 \rfloor$, and gives us $L''(\pi'') \geq \frac{1}{3.4^4} n^2$.

The theorem follows by putting Lemmas 1 and 2 together.

Conclusions (Open Problems and Future Directions)

Sudoku is a fascinating source of new interesting open questions in combinatorics. The obvious open question is whether the result in this paper can be strengthened to defining sets. Our conjecture is that this is true, i.e., the size of the smallest defining set of Sudoku squares of order n is $\Theta(n^2)$. If true, this is probably a difficult problem, since the similar question for Latin squares has been open for years.

A simpler problem is to strengthen the result to a notion like "semi-strong critical set" ([1]), as defined similarly to Latin squares. Also, finding any super-linear lower bound is an interesting open question. Note that in the case of Latin squares, the best lower bounds we know are just barely superlinear.

As mentioned earlier in the paper, for Latin squares, there is a construction for a defining set of size $\lfloor n^2/4 \rfloor$. This defining set has a unique extension to a back-circulant Latin square. In fact, it is proved that for even n, this is the smallest defining set of a back-circulant Latin square. It is not hard to show that by permuting rows and columns of a back-circulant Latin square, one can obtain a Sudoku square. This gives a construction for a defining set of size $\lfloor n^2/4 \rfloor$ for Sudoku.

Two questions remain open: Are there Sudoku squares with smaller defining sets, and are there smaller defining sets for this particular Sudoku squares. The answer to both of these questions are conjectured to be negative in the case of Latin squares (and proved to be so in the case of the second question for n even). For Sudoku, however, these conjectures might not be true, since the block constraint could reduce the size of the smallest defining set.

There are also many computational open questions arising from the Sudoku puzzle. The first question is whether the problem of Sudoku completion (given a partial Sudoku square, is there a completion to a Sudoku square) is NP-hard. Our conjecture, of course, is that it is. A more difficult problem is the complexity of completing a defining set (i.e., a set that is guaranteed to have a unique completion) to a full Sudoku square. As a less mathematical problem, it would be interesting if one can define a measure of difficulty for Sudoku puzzles that roughly correspond to how hard the puzzle is for humans. An online search reveals many 9×9 Sudoku puzzles that are claimed to be the hardest Sudoku puzzle. It would be interesting to have a quantitative measure of such puzzles.

Finally, there are many open combinatorial conjectures for Latin squares for which the corresponding Sudoku problem might be more approachable. Two example are two long-standing conjectures of Brualdi-Stein and Ryser.

Conjecture 1 ([3, 8]) *Every Latin square of even order n contains a partial transversal of length $n - 1$.*

Conjecture 2 ([7]) *Every Latin square of odd order contains a transversal.*

Another interesting question is whether Galvin's theorem about list colorability of the Latin squares ([4]) to Sudoku squares.

References

[1] J. A. Bate and G. H. J. van Rees. The size of the smallest strong critical set in a Latin square. *Ars Combin.*, 53:73–83, 1999.

[2] Richard A. Brualdi and Herbert J. Ryser. *Combinatorial matrix theory*, volume 39 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1991.

[3] J. Dénes and A. D. Keedwell. *Latin squares and their applications*. Academic Press, New York, 1974.

[4] Fred Galvin. The list chromatic index of a bipartite multigraph. *J. Combin. Theory Ser. B*, 63(1):153–158, 1995.

[5] E. S. Mahmoodian. Some problems in graph colorings. In *Proceedings of the 26th Annual Iranian Mathematics Conference, Vol. 2 (Kerman, 1995)*, pages 215–218. Shahid Bahonar Univ. Kerman, Kerman, 1995.

[6] Gary McGuire, Bastian Tugemann, and Gilles Civario. There is no 16-clue Sudoku: solving the Sudoku minimum number of clues problem via hitting set enumeration. *Exp. Math.*, 23(2):190–217, 2014.

[7] H. J. Ryser. Neuere Probleme in der Kombinatorik (prepared by D.W. Miller). *Vortrage uber Kombinatorik*, pages 69–91, 1967. (cited in [2]).

[8] S. K. Stein. Transversals of Latin squares and their generalizations. *Pacific J. Math.*, 59(2):567–575, 1975.

Performance Evaluation of Various Smoothed Finite Element Methods with Tetrahedral Elements in Large Deformation Dynamic Analysis

Ryoya Iida[1,a)], Yuki Onishi[1] and Kenji Amaya[1]

[1] Department of Systems and Control Engineering, Tokyo Institute of Technology, Japan

[a)]Corresponding and Presenting author: iida.r.ad@m.titech.ac.jp

;

Abstract

It is known that Selective ES/NS-FEM-T4 and F-barES-FEM-T4 show far better results than standard FEM with first-order tetrahedral elements in static analysis. These formulations resolve the pressure oscillation and locking problems in finite element (FE) analysis for nearly incompressible materials without increasing DOF. In this paper, we apply these formulations to modal and dynamic analysis and evaluate the accuracy and stability. Some demonstration analyses confirm these methods can show the as good accuracy in dynamic and modal analysis as in static one. They reveal that the time evolution of total energy of F-barES-FEM-T4 diverge exponentially due to the asymmetric components of the stiffness matrices of this formulation.

Keywords: Smoothed finite element method, F-bar method, Large deformation, Pressure oscillation, Locking-free, Modal analysis, Dynamic analysis.

Introduction

Tetrahedral elements are commonly used in practical FE analyses because arbitrary shapes cannot be meshed into hexahedral elements automatically. Additionally, because intermediate nodes easily pop out in large deformation problems, second or higher-order elements are not preferable[8, 9, 3, 10]. However, T4 elements suffer from pressure oscillation and locking in the FE analysis for nearly incompressible materials. Therefore, high-accuracy FE analysis with T4 elements have been in demand.

Recently, Smoothed Finite Element Method (S-FEM) have been proposed. This technique is known for the high-accuracy FE formulations with T4 elements. Edge-based S-FEM (ES-FEM-T4)[4, 1] has good accuracy in isovolumetric deformation without shear locking. However, it suffers from pressure oscillation and volumetric locking, as well as Standard FEM-T4, in the analysis for nearly incompressible materials[4, 6]. Node-based S-FEM (NS-FEM-T4)[4] has good accuracy in volumetric deformation without locking nor strong pressure oscillation. However, it has spurious low energy modes; therefore, it may cause instability in large strain problems[4].

Considering these features, Selective ES/NS-FEM-T4[4, 5] and F-barES-FEM-T4[7], which combine some classical S-FEMs, have been proposed. In Selective ES/NS-FEM-T4, the hydrostatic part of Cauchy stress tensor is evaluated by using NS-FEM-T4 and the deviatoric part of that is calculated by using ES-FEM-T4. On the other hand, in F-barES-FEM-T4, the isovolumetric part of deformation gradient is derived from that of ES-FEM-T4 and the volumetric part is derived through the multiple smoothing among nodes and elements. It is known that these two methods have good accuracy in static analysis.

In this paper, we evaluate the accuracy and stability of NS-FEM-T4, Selective ES/NS-FEM-T4 and F-barES-FEM-T4 in modal and dynamic analysis. A modal analysis of a multi-material cylinder and a dynamic bending analysis of a cantilever are performed with these formulations.

Methods

In this section, we describe the way to calculate the nodal internal force of F-barES-FEM-T4. That of the other methods is referred to in papers of Liu[4] and Onishi[4, 5].

Concept of F-barES-FEM-T4

It is known that NS-FEM-T4 can suppress pressure oscillation to a certain degree in FE analysis for nearly incompressible materials. This implies that the node-based smoothing of the relative volume change J have the effect of low-pass filter for the pressure distribution. Then, in F-barES-FEM-T4, it is expected that a repetitive node-based smoothing suppresses the pressure oscillation more strongly.

Figure 1 illustrates the outline of F-barES-FEM-T4 in 2D problem (i.e., F-barES-FEM-T3) for simplicity. In F-barES-FEM-T4, the deformation gradient $\boldsymbol{F}$ at each edge is divided into the isovolumetric part $\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$ and the volumetric part $\overline{J}$. The isovolumetric part $\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$ at each edge is calculated as weighted mean of only the adjacent elements in the same manner as ES-FEM-T4. The volumetric part $\overline{J}$ at each edge is calculated as weighted mean of the relative volumetric changes of some surrounding elements which are defined by cyclic smoothings. The smoothed deformation gradient $\overline{\boldsymbol{F}}$ is calculated with $\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$ and $\overline{J}$ in the manner of F-bar method[2]. Thus, a good accuracy of ES-FEM-T4 in isovolumetric part and stronger suppression of pressure oscillation than NS-FEM-T4 are expected.

Cyclic Smoothings

In this section, we describe the way to calculate the deformation gradient of volumetric part $\overline{J}$ with cyclic smoothings.

1. Calculate the relative volume change at each element ${}^{\mathrm{Elem}}J$ in the same manner as Standard FEM-T4.
2. Calculate the smoothed relative volume change at each node ${}^{\mathrm{Elem}}\widetilde{J}$ in the same manner as NS-FEM-T4:

$$ {}^{\mathrm{Node}}_{n}\widetilde{J} = \frac{\sum_{k\in {}^{\mathrm{Node}}_{n}\mathbb{E}} {}^{\mathrm{Elem}}_{k}J\,{}^{\mathrm{Elem}}_{k}V^{\mathrm{ini}}/4}{\sum_{k\in {}^{\mathrm{Node}}_{n}\mathbb{E}} {}^{\mathrm{Elem}}_{k}V^{\mathrm{ini}}/4}, \tag{1} $$

   where ${}^{\mathrm{Elem}}_{k}V^{\mathrm{ini}}$ is the initial volume of element k, ${}^{\mathrm{Node}}_{n}\mathbb{E}$ is the set of elements attached to node n.
3. Calculate the smoothed relative volume change at each element ${}^{\mathrm{Elem}}\widetilde{J}$ in the same way as NS-FEM-T4:

$$ {}^{\mathrm{Elem}}_{e}\widetilde{J} = \frac{1}{4}\sum_{k\in {}^{\mathrm{Elem}}_{e}\mathbb{N}} {}^{\mathrm{Node}}_{k}\widetilde{J}, \tag{2} $$

   where ${}^{\mathrm{Elem}}_{e}\mathbb{N}$ is the set of nodes attached to element e.
4. Repeat step 2. and 3. as necessary to calculate multi-smoothed relative volumetric change ${}^{\mathrm{Elem}}\overline{J}$. In the second or later evaluation of Eq. (1), ${}^{\mathrm{Elem}}\widetilde{J}$ is substituted for ${}^{\mathrm{Elem}}J$.

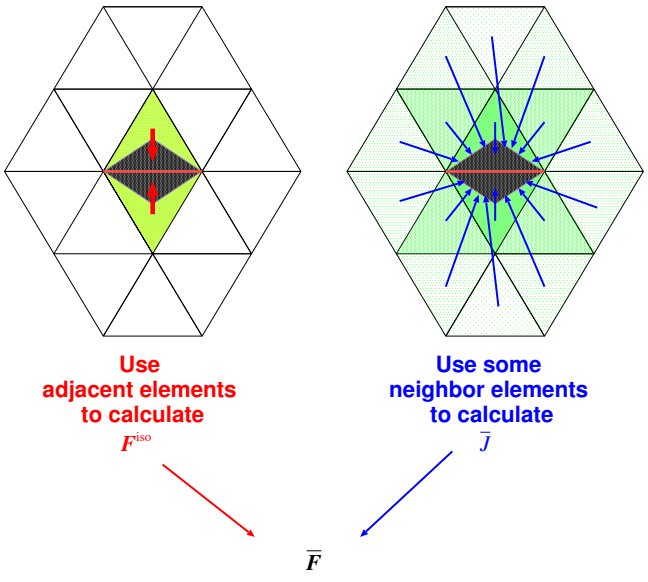


Figure 1. Outline of F-barES-FEM in 2D for simplicity.

5. Calculate the relative volume change at each edge ${}^{\mathrm{Edge}}\overline{J}$ in the similar way to ES-FEM-T4:

$$ {}^{\mathrm{Edge}}_{h}\overline{J} = \frac{\sum_{k\in {}^{\mathrm{Edge}}_{h}\mathbb{E}} {}^{\mathrm{Elem}}_{k}\widetilde{J}\ {}^{\mathrm{Elem}}_{k}V^{\mathrm{ini}}/6}{\sum_{k\in {}^{\mathrm{Edge}}_{h}\mathbb{E}} {}^{\mathrm{Elem}}_{k}V^{\mathrm{ini}}/6}, \tag{3} $$

where ${}^{\mathrm{Edge}}_{h}\mathbb{E}$ is the set of elements attached to edge h.

These process means the relative volumetric change ${}^{\mathrm{Edge}}\overline{J}$ is calculated as the weighted mean of that of the surrounding elements ${}^{\mathrm{Elem}}J$.

Smoothed Deformation Gradient and Nodal Internal Force

We describe the way to calculate the smoothed deformation gradient ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$ and the nodal internal force. The isovolumetric part ${}^{\mathrm{Edge}}\widetilde{\boldsymbol{F}}^{\mathrm{iso}}$ is the same with that of ES-FEM-T4 ${}^{\mathrm{Edge}}\boldsymbol{F}^{\mathrm{iso}}$:

$$ {}^{\mathrm{Edge}}\widetilde{\boldsymbol{F}}^{\mathrm{iso}} = {}^{\mathrm{Edge}}\boldsymbol{F}^{\mathrm{iso}} = \left(\frac{1}{{}^{\mathrm{Edge}}J}\right)^{1/3} {}^{\mathrm{Edge}}\boldsymbol{F}, \tag{4} $$

where the relative volume change at an edge ${}^{\mathrm{Edge}}J$ is det(${}^{\mathrm{Edge}}\boldsymbol{F}$). The volumetric part ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{vol}}$ is calculated by using ${}^{\mathrm{Edge}}\overline{J}$ denoted in previous section:

$$ {}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{vol}} = {}^{\mathrm{Edge}}\overline{J}^{1/3}\boldsymbol{I}, \tag{5} $$

where $\boldsymbol{I}$ is the second-order identity tensor. Smoothed deformation gradient ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$ is calculated by using these two part in the same manner of F-bar method as follows:

$$ {}^{\mathrm{Edge}}\overline{\boldsymbol{F}} = {}^{\mathrm{Edge}}\widetilde{\boldsymbol{F}}^{\mathrm{vol}} \cdot {}^{\mathrm{Edge}}\overline{\boldsymbol{F}}^{\mathrm{iso}} = \left(\frac{{}^{\mathrm{Edge}}\overline{J}}{{}^{\mathrm{Edge}}J}\right)^{1/3} {}^{\mathrm{Edge}}\boldsymbol{F}. \tag{6} $$

The smoothed Cauchy stress tensor $\overline{\boldsymbol{T}}$ is derived from a material constitutive model and ${}^{\mathrm{Edge}}\overline{\boldsymbol{F}}$.

The nodal internal force vector at an edge ${}^{\mathrm{Edge}}f^{\mathrm{int}}$ is calculated as follows:

$$ {}^{\mathrm{Edge}}_{h}f^{\mathrm{int}}_{P:p} = \frac{\partial {}^{\mathrm{Edge}}_{h}D_{ij}}{\partial \dot{u}_{P:p}} {}^{\mathrm{Edge}}_{h}\overline{T}_{ij}\ {}^{\mathrm{Edge}}_{h}V, \tag{7} $$

where $\square_{P:p}$ means the p-th component of node P, $\dot{u}$ is the nodal velocity and ${}^{\mathrm{Edge}}_{h}\boldsymbol{D}$ is the stretching tensor derived in the same way as ES-FEM-T4.

Results

Modal Analysis of Multi-Material Cylinder

Figure 2 illustrates the outline of the modal analysis of a multi-material cylinder. The analysis domain is a quarter of the cylinder of ϕ 2 × 6 m; its bottom is fixed completely. The upper part of the cylinder is made of steel and the bottom one is made of rubber. The material constitutive model for the analysis domain is linear elastic material. The density, Young's modulus and Poisson's ratio of the steel are 7800 kg/m^3, 200 GPa and 0.3 respectively, and those of the rubber are 920 kg/m^3, 5.0 MPa and 0.499, respectively. The analysis with F-barES-FEM-T4, Selective ES/NS-FEM-T4, NS-FEM-T4 and ABAQUS C3D4 (4-node tetrahedral elements) are performed with unstructured tetrahedral elements of 0.2 m global mesh seed size. The analysis with ABAQUS C3D8 (8-node hexahedral elements with selective reduced integration method) of 0.2 m global mesh seed size is performed to obtain reference solutions. The number of cyclic smoothings for the steel part is 0 and the one for the rubber part is 1 or 2, in the analysis with F-barES-FEM-T4. The results of F-barES-FEM-T4 are labeled with 'c', which is the number of cyclic smoothings for the rubber part.

Figure 3 shows the comparison of the natural frequencies between various S-FEMs and two ABAQUS elements. The natural frequencies of ABAQUS C3D4 are far higher than reference solution, the result of ABAQUS C3D8, because of

the volumetric locking, and those of NS-FEM-T4 are far lower because of spurious low energy modes. F-barES-FEM-T4 and Selective ES/NS-FEM-T4 show good accuracy of natural frequencies without locking and spurious mode.

Figure 4 and 5 show the mode shapes of the 1st and 11th modes. NS-FEM-T4 shows strange mode shape due to the spurious low energy modes in 11th mode. F-barES-FEM-T4 and Selective ES/NS-FEM-T4 show the similar shapes to ABAQUS C3D8 without locking and spurious modes.

It is known that the stiffness matrix of F-barES-FEM-T4 is asymmetric. The asymmetric property cause complex eigenvalue in modal analysis. Figure 6 shows the distributions of natural frequency of F-barES-FEM-T4(1) and (2).

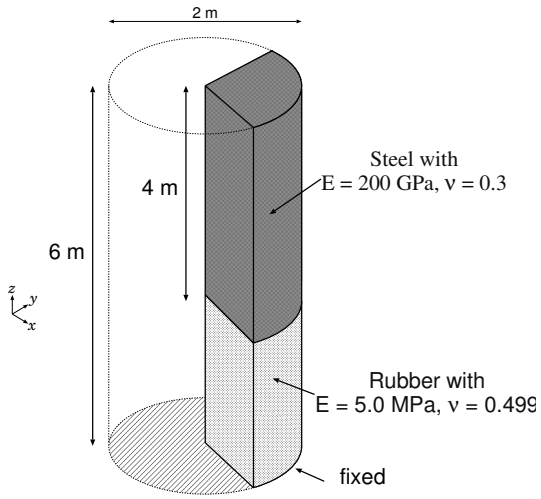


Figure 2. Outline of the modal analysis for a multi-material cylinder.

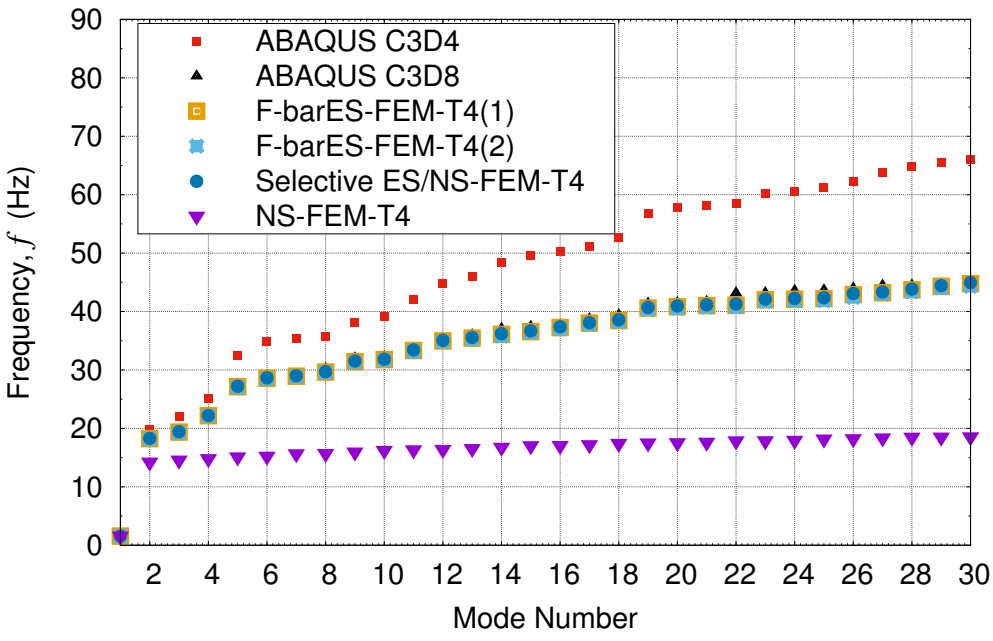


Figure 3. Comparison of the natural frequencies vs. mode numbers. The frequencies of ABAQUS C3D4 are higher than those of the others because of the volumetric locking.

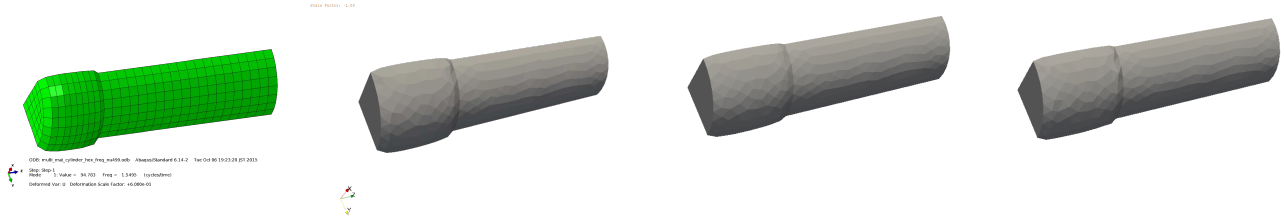


Figure 4. Mode shapes of the 1st mode.

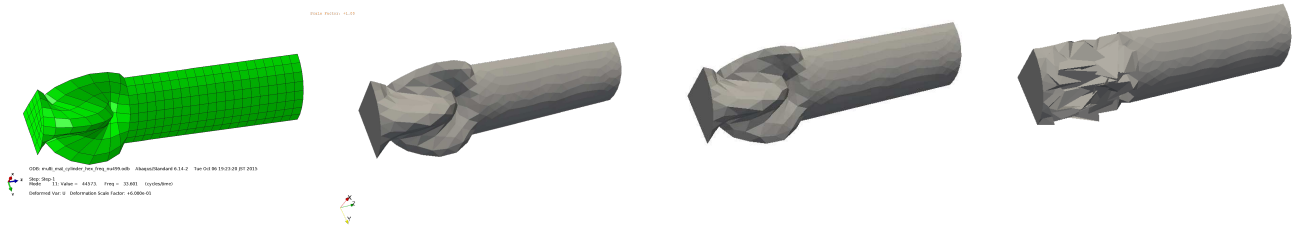


Figure 5. Mode shapes of the 11th mode.

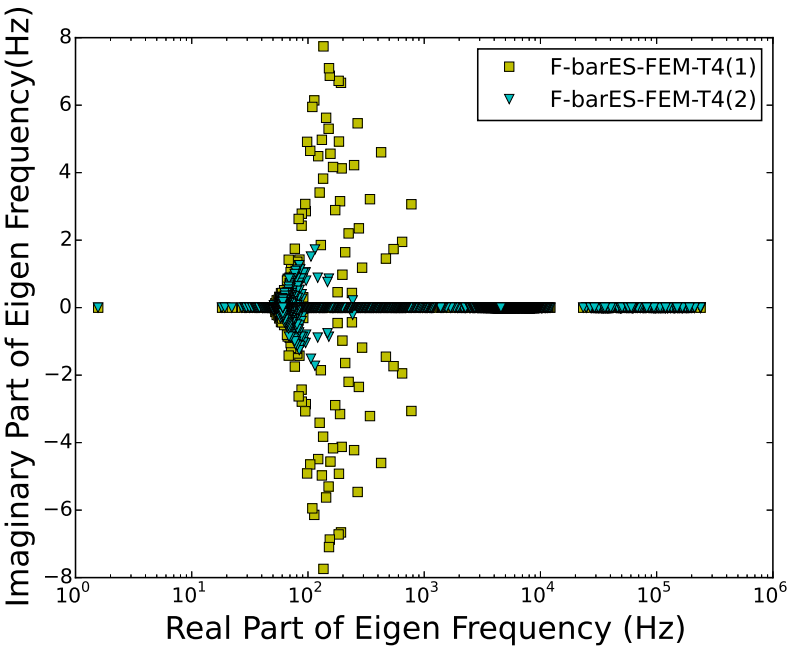


Figure 6. Distribution of the eigenvalue in the multi material cylinder model with F-barES-FEM-T4(1) and F-barES-FEM-T4(2).

Dynamic Bending Analysis of Cantilever

Figure 7 illustrates the outline of the dynamic bending analysis of a cantilever. The analysis domain is a cuboid of 10×1×1 m; its left side is perfectly constrained; a uniform initial velocity of 2.0 m/s in $-z$ direction is applied. The material constitutive model for the analysis domain is Neo-Hookean hyperelastic model. The density, initial Young's modulus and initial Poisson's ratio are 10000 kg/m^3, 6.0 MPa and 0.499, respectively. The analysis with F-barES-FEM-T4, Selective ES/NS-FEM-T4, NS-FEM-T4 and ABAQUS/Explicit C3D4 are performed with unstructured tetrahedral elements of 0.2 m global mesh seed size. The analysis with ABAQUS/Explicit C3D8 of 0.2 m global mesh seed size is also performed to obtain a reference solution. The number of cyclic smoothings is 1 to 3, in the analysis with F-barES-FEM-T4. The results of F-barES-FEM-T4 are labeled with 'c', which is the number of cyclic smoothings. In this analysis, the time integration scheme is Velocity Verlet, and the time increment is 1.0×10^{-4} s.

The comparison of the vertical displacements (u_z) at one of the corner node (◯ in Figure 7) is shown in Figure 8. The results of ABAQUS/Explicit C3D4 and NS-FEM-T4 differ from the reference solution due to the locking and spurious low energy modes, respectively. Those of F-barES-FEM-T4s and Selective ES/NS-FEM-T4 agree with the reference; therefore, these formulations have good accuracy without locking in dynamic analysis.

Figure 9 shows the pressure distributions at 1.5 s. In these figures, the value above the range is colored dark red, the one below the range is colored in dark blue and the contour range is [-0.223, 0.2813] (MPa). The results of NS-FEM-T4, Selective ES/NS-FEM-T4 and F-barES-FEM-T4(1) are different from ABAQUS/Explicit C3D8 at a certain level. Those of F-barES-FEM-T4(2) and (3) agree with the reference; therefore, F-barES-FEM-T4 with sufficient number of cyclic smoothings have good accuracy without pressure oscillation in dynamic analysis.

Figure 10 illustrates the comparison of time-histories of total energy among several formulations. The results of F-barES-FEM-T4s diverge exponentially, and the divergence speeds decrease as the number of cyclic smoothings increasing. This is caused by the imaginary part of the natural frequencies of F-barES-FEM-T4s. As shown in Figure 6, F-barES-FEM-T4 causes complex natural frequencies due to the asymmetric component of the stiffness matrix unlike other methods. When the k-th natural frequency is a complex number, $\omega_k = a + \mathrm{i}b$ $(a, b \in \mathbb{R})$, time evolution of the k-th mode shape $u_k(t)$ is expressed as

$$u_k(t) = \mathrm{Re}[\{u_{k0}\} \exp(-\mathrm{i}\omega_k t)] \tag{8}$$

$$= \mathrm{Re}[\{u_{k0}\} \exp(-\mathrm{i}at)] \exp(bt), \tag{9}$$

where the constant vector $\{u_{k0}\}$ is defined by initial conditions. If $b > 0$, $\exp(bt)$ part diverges with time evolution and that is why the results of F-barES-FEM-T4s diverge exponentially. The long time analysis with F-barES-FEM-T4 requires more ingenuity and it is our future work.

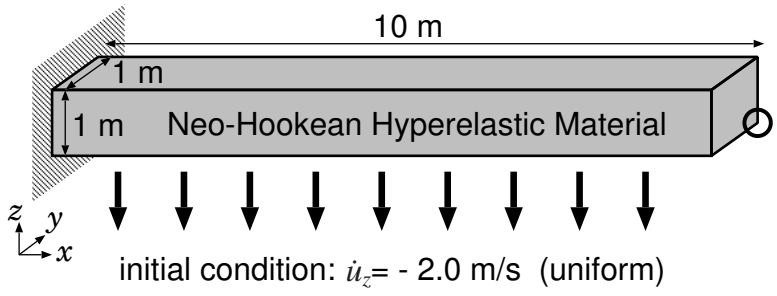


Figure 7. Outline of the dynamic bending analysis of a cantilever. The initial uniform velocity is -2.0 m/s in z direction.

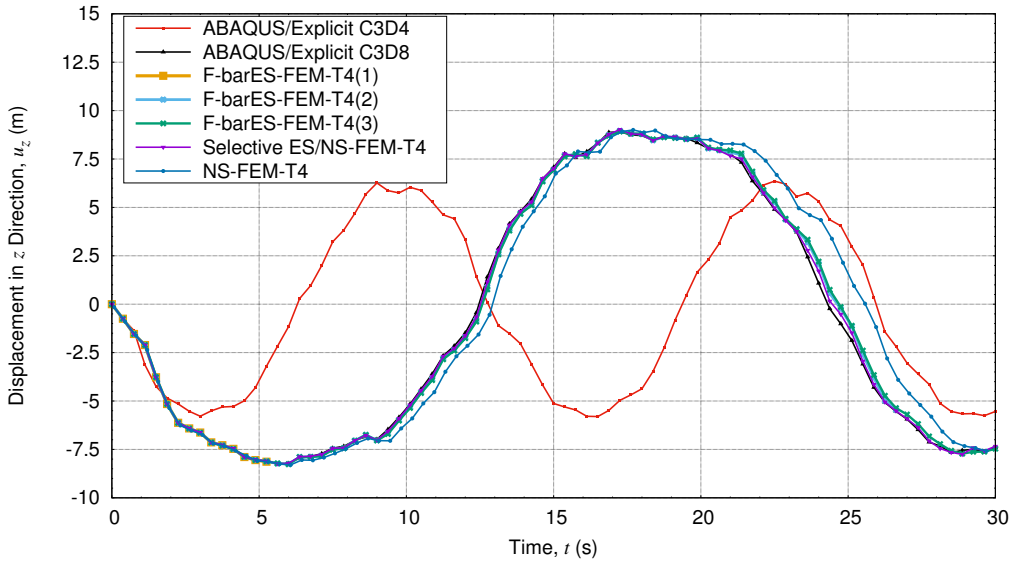


Figure 8. Comparison of the vertical displacement at the corner vs. time in the cantilever bending analysis.

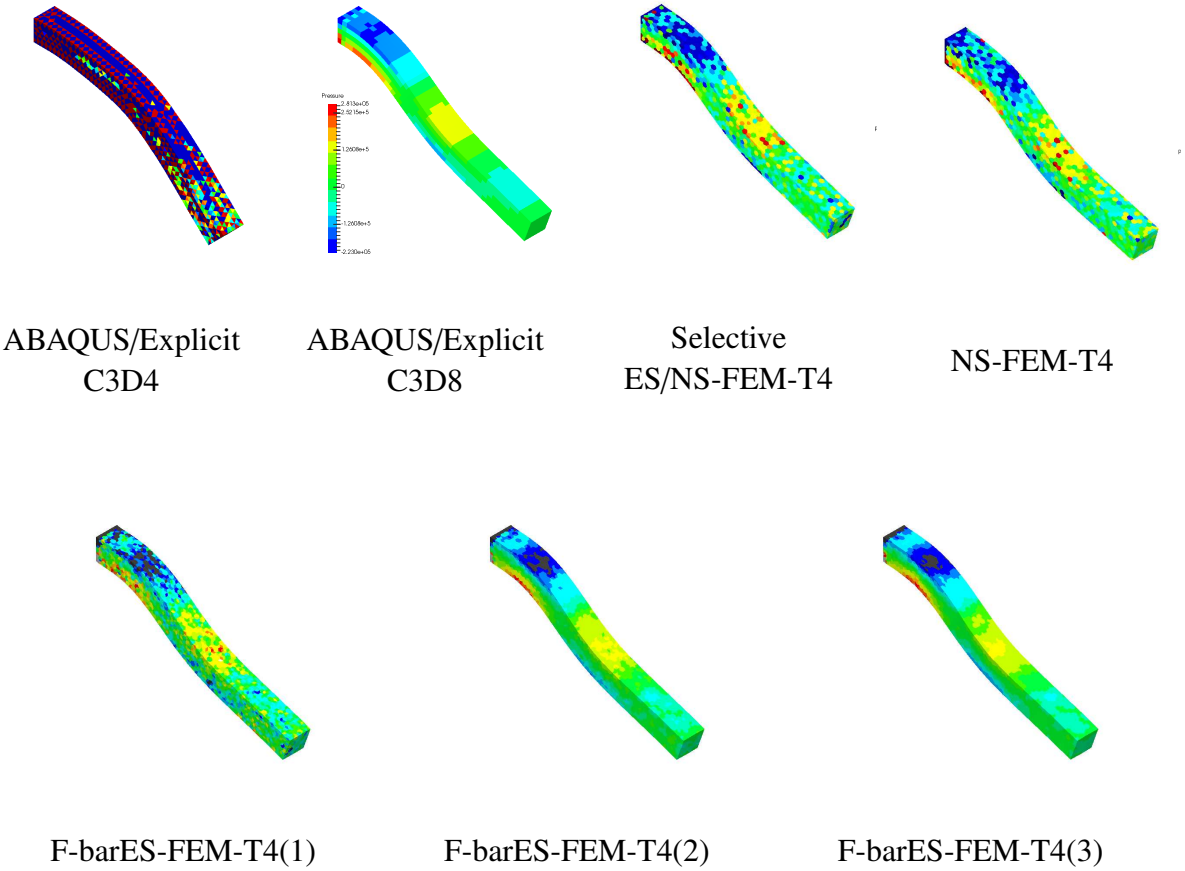


Figure 9. Deformed shapes and pressure distributions of the dynamic cantilever bending analysis at 1.5 s.

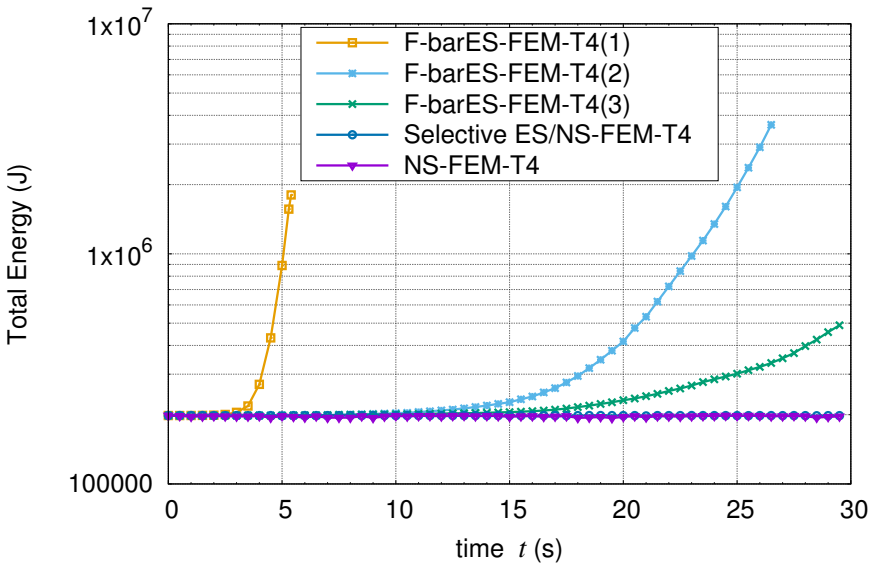


Figure 10. Comparison of the total energy vs. time for different FEM formulations in the dynamic cantilever bending analysis.

Conclusions

We adapted NS-FEM-T4, Selective ES/NS-FEM-T4 and F-barES-FEM-T4 to the modal and dynamic analysis. The features of these formulations are summarized as follows.

- NS-FEM-T4
  - **Modal analysis**: low accuracy due to the spurious low energy modes.
  - **Dynamic analysis**: slightly soft solution in displacement; a little pressure oscillation; no divergence of energy.
- Selective ES/NS-FEM-T4
  - **Modal analysis**: good accuracy without locking nor spurious modes.
  - **Dynamic analysis**: good accuracy in displacement; a little pressure oscillation; no divergence of energy.
- F-barES-FEM-T4
  - **Modal analysis**: good accuracy without locking nor spurious modes.
  - **Dynamic analysis**: good accuracy in displacement; no pressure oscillation with sufficient number of cyclic smoothings; divergence of energy due to the asymmetric property of the stuffiness matrix.

The long time analysis with F-barES-FEM-T4 is our future work.

References

[1] Cazes, F. and Meschke, G. (2013). An edge-based smoothed finite element method for 3D analysis of solid mechanics mechanics problems. *International Journal for Numerical Methods in Engineering*, 94(8):715–739.

[2] de Souza Neto, E., Peric, D., Dutko, M., and Owen, D. (1996). Design of simple low order finite elements for large strain analysis of nearly incompressible solids. *International Journal of Solids and Structures*, 33(20-22):3277–3296.

[3] Lee, M. C., Joun, M. S., and Lee, J. K. (2007). Adaptive tetrahedral element generation and refinement to improve the quality of bulk metal forming simulation. *Finite Elements in Analysis and Design*, 43(10):788–802.

[4] Liu, G. R. and Nguyen-Thoi, T. (2010). *Smoothed Finite Element Methods*. CRC Press, Boca Raton, FL, USA.

[5] Onishi, Y. and Amaya, K. (2014). A locking-free selective smoothed finite element method using tetrahedral and triangular elements with adaptive mesh rezoning for large deformation problems. *International Journal for Numerical Methods in Engineering*, 99(5):354–371.

[6] Onishi, Y. and Amaya, K. (2015). Performance evaluation of the selective smoothed finite element methods using tetrahedral elements with deviatoric/hydrostatic split in large deformation analysis. *Theoretical and Applied Mechanics Japan*, 63:55–65.

[7] Onishi, Y., Iida, R., and Amaya, K. (under review). F-bar aided edge-based smoothed finite element method using tetrahedral elements for finite deformation analysis of nearly incompressible solids. *International Journal for Numerical Methods in Engineering*.

[8] Son, I.-H. and Im, Y.-T. (2006). Localized remeshing techniques for three-dimensional metal forming simulations with linear tetrahedral elements. *International Journal for Numerical Methods in Engineering*, 67(5):672–696.

[9] Wan, J., Kocak, S., and Shephard, M. (2005). Automated adaptive 3D forming simulation processes. *Engineering with Computers*, 21(1):47–75.

[10] Wicke, M., Ritchie, D., Klingner, B. M., Burke, S., Shewchuk, J. R., and O'Brien, J. F. (2010). Dynamic local remeshing for elastoplastic simulation. *ACM Transactions on Graphics*, 29(4):49:1–11. Proceedings of ACM SIGGRAPH 2010, Los Angles, CA.

LES of oscillating boundary layers under surface cooling

Mario J. Juha, Andrés E. Tejada-Martínez*†, and Jie Zhang

Department of Civil and Environmental Engineering, University of South Florida, USA.

*Presenting author: aetejada@usf.edu
†Corresponding author

Abstract

Large-eddy simulation (LES) of open channel flow driven by an oscillating pressure gradient with zero surface shear stress was performed. The flow is representative of an oscillating tidal boundary layer. Under neutrally stratified conditions, during certain phases of the oscillating pressure gradient corresponding, for example, to peak tide, the flow develops secondary structures, characterized by coherent, full-depth, streamwise-elongated counter-rotating cells. These structures are similar to the classical Couette cells found in Couette flow driven by parallel no-slip plates moving in opposite direction. A constant cooling flux at the surface with an adiabatic bottom wall leads to more intense and coherent streamwise-elongated cells characterized by greater crosswind width, which we term *convective supercells*. The signature of the convective supercells is observed even during times when the oscillating mean flow is decelerating, unlike in cases without surface cooling. Investigation of these coherent structures (with and without surface cooling) is deemed important due to their strong influence on vertical mixing and their potential role in determining the wake behind tidal turbines.

Keywords: Large-eddy simulation, oscillating boundary layer flow, tidal flow, surface cooling, convective supercell, vertical mixing of momentum

Introduction

Large-eddy simulation (LES) of neutrally stratified open channel flows driven with either a constant [1] or oscillating [2] pressure gradient have revealed the presence of secondary, coherent, streamwise-elongated roll cells occupying the full-depth of the water column. These cells, sketched in Figure 1, are similar nature to the well-known Couette cells occurring in channel flow driven by no-slip plates moving in opposite direction [4]. Furthermore, in [1], application of a surface cooling flux to the initially neutrally stratified open channel flow driven by constant pressure gradient led to an unstably stratified flow characterized by wider, intensified streamwise-elongated roll cells. The latter were termed *convective supercells* due to their greater intensity and cross-stream size resulting in greater vertical mixing (e.g. of momentum) throughout the water column.

The goal of the present work is to re-visit open channel flow with an oscillating pressure gradient and to apply, for the first time, a surface cooling flux in order to understand its effect on the structure of the cells throughout the pressure gradient cycle. The pressure gradient is chosen so as to drive an oscillating boundary layer flow with period characteristic of tidal boundary layers. Results will be analyzed via visualizations of the coherent cells revealed by the averaging of instantaneous velocity fluctuations over the streamwise direction (x_1). Additional analysis is presented in terms of the instantaneous streamwise velocity averaged over x_1 and x_2 (the cross-stream direction) at various instances during the tidal cycle in order to understand the vertical mixing of momentum induced by the cells. More in-depth analysis of the flows, for example in terms of their turbulent structure as revealed through Lumley invariant maps [5] at different phases of the tide, will be reserved for a more in-depth journal article.

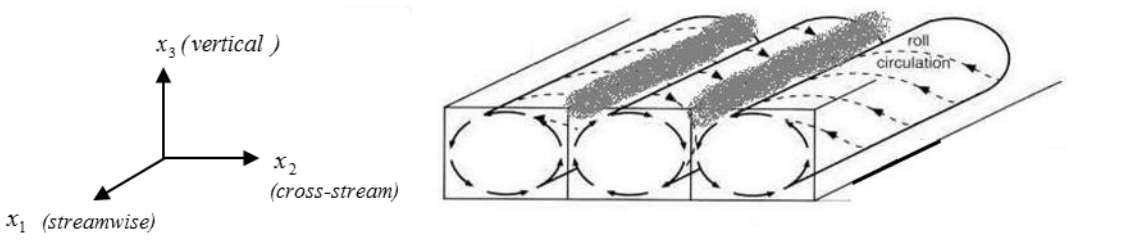


Figure 1. Sketch of streamwise-elongated counter-rotating cells occurring secondary to the mean flow in open channel flow driven by a pressure gradient. In the field such roll cells, characterizing boundary layer flow, could potentially lead to accumulation of lines of floating material along the surface convergence of the cells, as depicted above, similar to the action of Langmuir circulations [3].

In addition to enhancing vertical mixing of momentum and scalars, the previously discussed coherent structures (with and without surface cooling) could potentially have a strong impact on the wake behind a tidal turbine. From a computational engineering analysis perspective, the modeling of the tidal flow and the turbulent wake generated by the turbine are equally important. The most sophisticated simulations of turbines without an accurate model of the tidal flow become as limited as much simpler turbine models. In a similar fashion to wind energy, wake meandering is caused by the large eddies of the ambient flow, such as the coherent cells studied here. Therefore, an accurate model of these large eddies is the most accurate way to capture wake meandering [6] and should be pursued in the future.

Governing LES equations

The spatially filtered or LES equations consisting of conservation of momentum, continuity and temperature (scalar) transport are given by Eqns. (1)-(3), respectively, as

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j} = -\frac{\partial \bar{p}}{\partial x_i} + \frac{1}{Re_\tau}\frac{\partial^2 \bar{u}_i}{\partial x_j^2} + \frac{\partial \tau_{ij}^{SGS}}{\partial x_j} + Ra_\tau \bar{\theta}\delta_{i3} + \frac{1}{Ro}\left(\frac{U_O^{max}}{u_\tau^{max}}\right)^2 \cos\left(\frac{1}{Ro}\frac{U_O^{max}}{u_\tau^{max}}t\right)\delta_{i1} \quad (1)$$

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0 \quad (2)$$

$$\frac{\partial \bar{\theta}}{\partial t} + \frac{\partial \bar{\theta}\bar{u}_j}{\partial x_j} = \frac{1}{PrRe_\tau}\frac{\partial^2 \bar{\theta}}{\partial x_j^2} + \frac{\partial \lambda_j^{SGS}}{\partial x_j} \quad (3)$$

In these equations, an over-bar denotes a filtered quantity with $\bar{u}_i$, $\bar{p}$ and $\bar{\theta}$ being the filtered velocity, pressure and temperature, respectively. Time and the spatial coordinates in the streamwise, cross-stream and vertical (depth) directions are given by t, x_1, x_2 and x_3, respectively. The third term on the right hand side of Eqn. (1) is comprised of the LES subgrid-scale (SGS) stress, τ_{ij}^{SGS}, modeled here through a dynamic Smagorinsky model (not shown). The fourth term on the right side of Eqn. (1) represents buoyancy acting on the

vertical (x_3) momentum equation and the fifth term is an oscillating pressure gradient (or tidal body force) driving the mean flow in the positive or negative x_1 direction. Note that the body force has been defined to drive a maximum dimensional free stream velocity U_O^{max} following the formulation in [2]. The second term on the right hand side of Eqn. (3) contains the SGS scalar flux λ_j^{SGS} modeled in terms of an eddy diffusivity taken as the dynamic Smagorinsky eddy viscosity (not shown) divided by turbulent Prandtl number, the latter set to 1.

Equations (1)-(3) have been made dimensionless with characteristic velocity and length scales given by the maximum wall friction velocity u_τ^{max} and water column half-depth δ, respectively. Temperature has been non-dimensionalized via the magnitude of the vertical temperature gradient at the surface (i.e. at the top of the water column) defined as Q/k where Q is the surface cooling flux and k is the thermal conductivity. Specifically, the characteristic temperature is taken as $Q\delta/k$.

Non-dimensionalization of the governing equations gives rise to the Reynolds, Rayleigh, Rossby and molecular Prandtl numbers defined as $Re_\tau = u_\tau^{max}\,\delta/\nu$, $Ra_\tau = g\beta\delta^2 Q/(u_\tau^{max\,2} k)$, $Ro = U_O^{max}/(\delta\,\omega)$ and $Pr = \nu/\kappa$, respectively. In these definitions, ν is kinematic viscosity, β is the coefficient of thermal expansion, ω is tidal frequency and κ is thermal diffusivity in water. Note that the Rayleigh number is indicative of the strength of surface buoyancy forcing relative to wall shear forcing. The Rossby number is proportional to the oscillatory boundary layer thickness relative to the water column half-depth.

Computational setup

Figure 2 shows a sketch of the computational domain for the LES. The domain consists of an open channel with periodic boundary conditions in the streamwise (x_1) and cross-stream (x_2) (i.e. the horizontal) directions. The top surface of the channel is open with imposed shear-free and cooling boundary conditions. The bottom of the channel consists of an adiabatic wall. The channel is $4\pi\delta$ long in the streamwise direction and $8\pi\delta/3$ wide in the cross-stream direction. The latter length was chosen so as to resolve one pair of convective supercells, as sketched in Figure (1), in simulations with surface cooling.

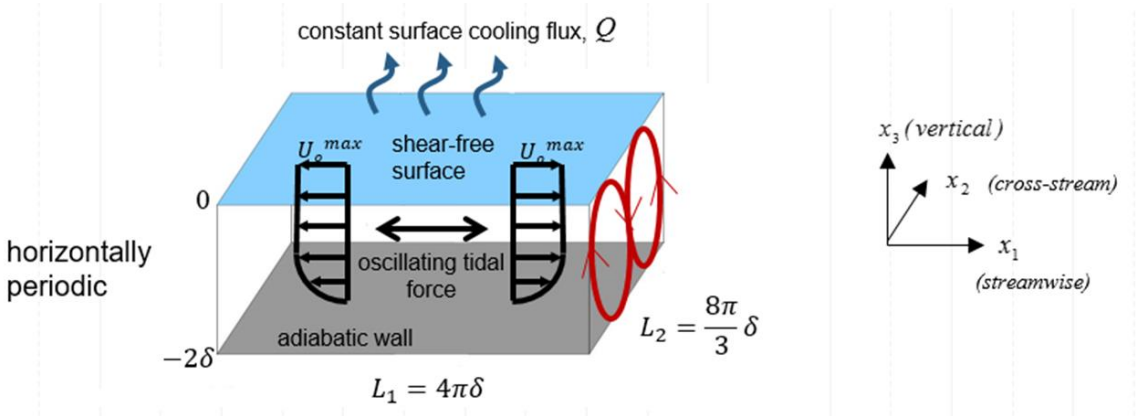


Figure 2. Sketch of computational domain displaying boundary conditions, the driving oscillating pressure gradient (or tidal force), the resulting mean velocity vectors and the secondary general circulation associated with convective supercells (red circles).

The pressure gradient (tidal force) frequency has been set such that the Rossby number Ro is equal to 878, following the tidal boundary layer simulations in [1]. This Rossby number value was obtained for a water column 10 meters deep (corresponding to δ = 5 m) with $U_O^{max} = 0.5$ m s^{-1} and frequency ω corresponding to a tidal period of approximately 12.5 hr. Furthermore, in the oscillating pressure gradient on the right hand side of Eqn. (1), the ratio U_O^{max}/u_τ^{max} has been set equal to the corresponding value obtained in a preliminary open channel simulation with constant pressure gradient prescribed such that Reynolds number Re_τ is equal to 395. Simulations with either Raleigh number Ra_τ set to 0 or 250 were performed. Ra_τ = 0 corresponds to zero surface heat flux Q and Ra_τ = 250 corresponds to Q of about 200 Watts m^{-2}. Molecular Prandtl number Pr was set to 1.

The computational domain was discretized with 32 uniformly spaced points in the streamwise, 64 uniformly spaced points in the cross-stream direction and 65 uniformly spaced points in the vertical direction. The vertical distribution of grid points is such that the viscous and buffer wall sublayers are not resolved, and thus a wall model was used. The numerical discretization consisted of the finite volume method along with time integration implemented in the popular open source code openFOAM [7].

Results

First we take a look at results for the flow with Ra_τ = 0. In Figure 3, on the panel on the right we can see the instantaneous velocity profiles averaged over the streamwise and crosswind directions at various instances during the tidal cycle. Velocity profiles in the LES are shown in solid and log-law fits of the LES solution through the first 6 grid points from the wall are shown with dots. Throughout the instances during the tidal cycle being shown, it is seen that the LES velocity is well-approximated by a log law.

On the panels on the left, in Figure 3, we can see instantaneous velocity fluctuations averaged over the streamwise direction. On the panels on the left the vertical axis covers the vertical extent of the water column and the horizontal axis covers the crosswind extent of the domain.

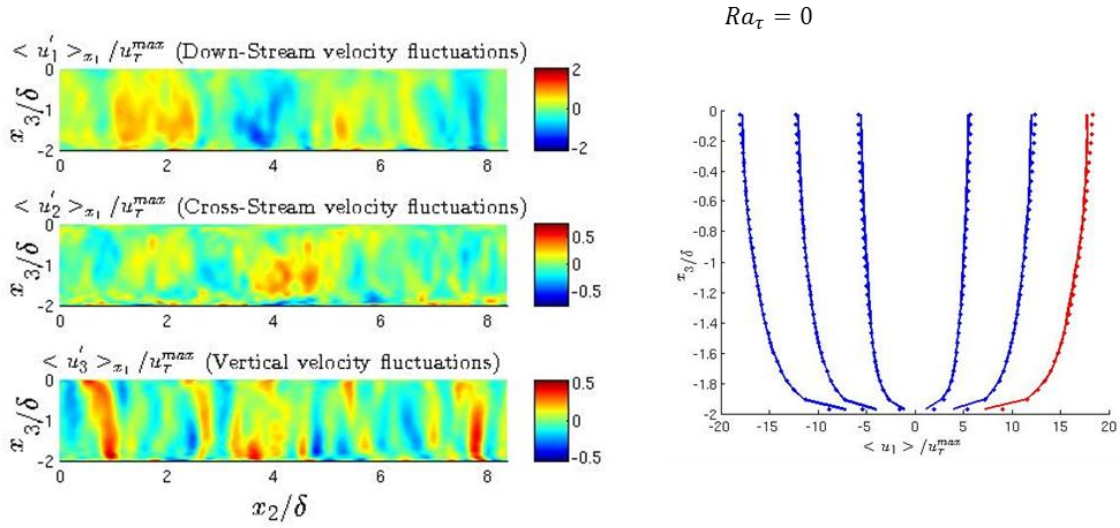


Figure 3. Panel on right: Depth profiles of instantaneous streamwise velocity averaged over horizontal directions (panel on right) in flow with Ra_τ = 0. Panels on left: Instantaneous, streamwise-averaged velocity fluctuations at the time corresponding to the velocity profile colored in red on the panel on the right.

The instantaneous snapshot of the velocity fluctuations seen on the left corresponds to the peak tide mean velocity profile shown in red on the panel on the right. At peak tide, the flow with Ra_τ = 0 is characterized by Couette-like cellular structures spanning the entire depth of the water column, as described earlier. For example, full-depth regions of negative and positive vertical velocity fluctuations can be seen on the lower panel on the left. The full-depth regions of negative vertical velocity fluctuations correspond to the downwelling limbs of the cells being resolved. These dowelling limbs of the cells generally coincide with a full-depth region of positive downwind velocity fluctuations as seen on the top panel on the left.

As the tidal current decelerates, the structures previously described become weaker, which can be seen in Figures 4 and 5. Furthermore, note that during the acceleration stage the cells are not as coherent (i.e. visible) as those during the deceleration stage. These less coherent cells are not shown here for brevity.

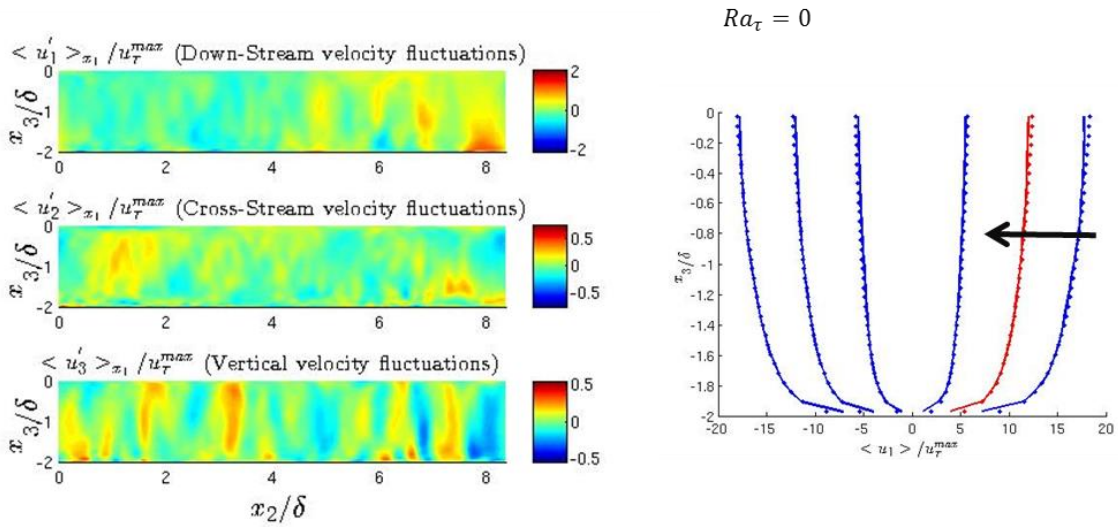


Figure 4. Same as caption of Figure 3 but for flow during deceleration stage.

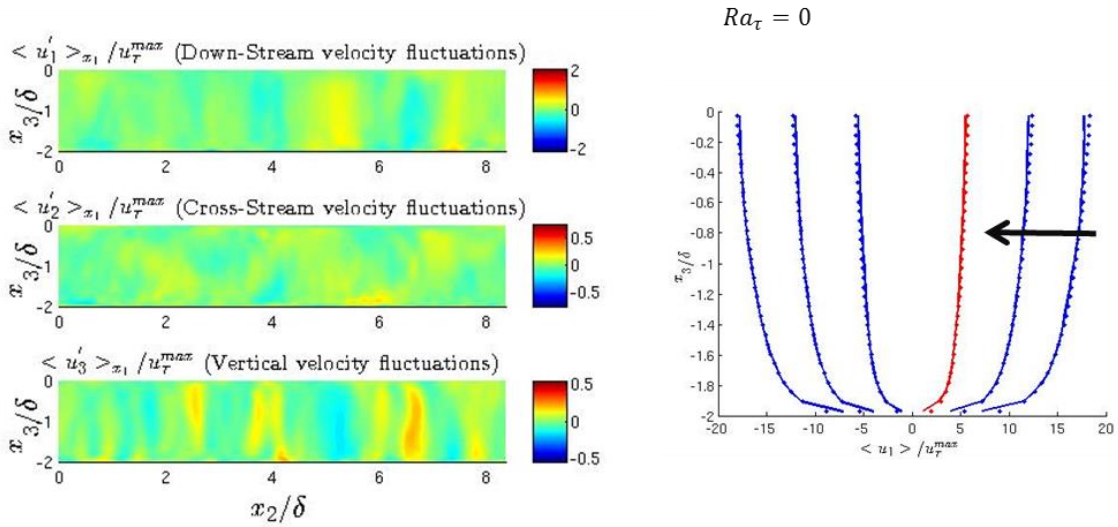


Figure 5. Same as caption of Figure 3 but for flow during deceleration stage.

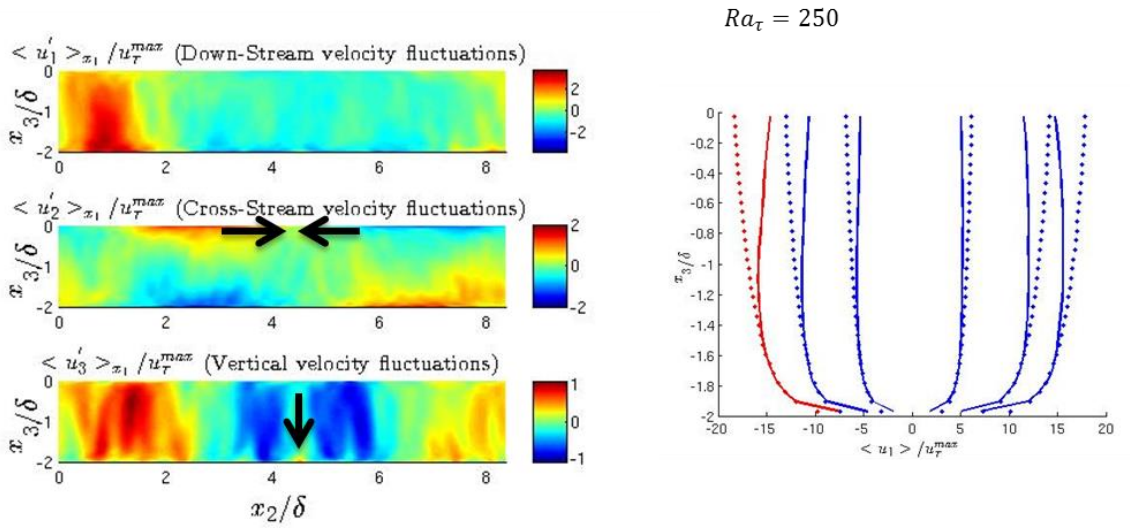


Figure 6. Panel on right: Depth profiles of instantaneous streamwise velocity averaged over horizontal directions (panel on right) in flow with Ra_τ = 250. Panels on left: Instantaneous, streamwise-averaged velocity fluctuations at the time corresponding to the velocity profile colored in red on the panel on the right.

Next we explore the flow with Ra_τ = 250. Throughout the cycle, the mean velocity is well homogenized and thus deviates from the classical log-law as seen on the panel on the right in Figure 6. Recall that the log-law is given by the dots and the LES velocity is given by the solid lines. In Figure 6, on the panels on the left we now see the presence of one cell, more coherent and wider (over the crosswind direction) than the cells described earlier with Ra_τ = 0. We term this cell as a *convective supercell*, due to its greater intensity.

The surface convergence of the convective supercell resolved is indicated by the black arrows appearing on the middle left panel in Figure 6. These arrows follow the orientation of the partially averaged crosswind velocity fluctuation. The surface convergence of the supercell leads into the full-depth downwelling limb of the cell characterized by negative vertical velocity fluctuation shown in the lower panel on the left indicated by the downward pointing black arrow. The increase in strength of the cellular structure resolved with surface cooling is

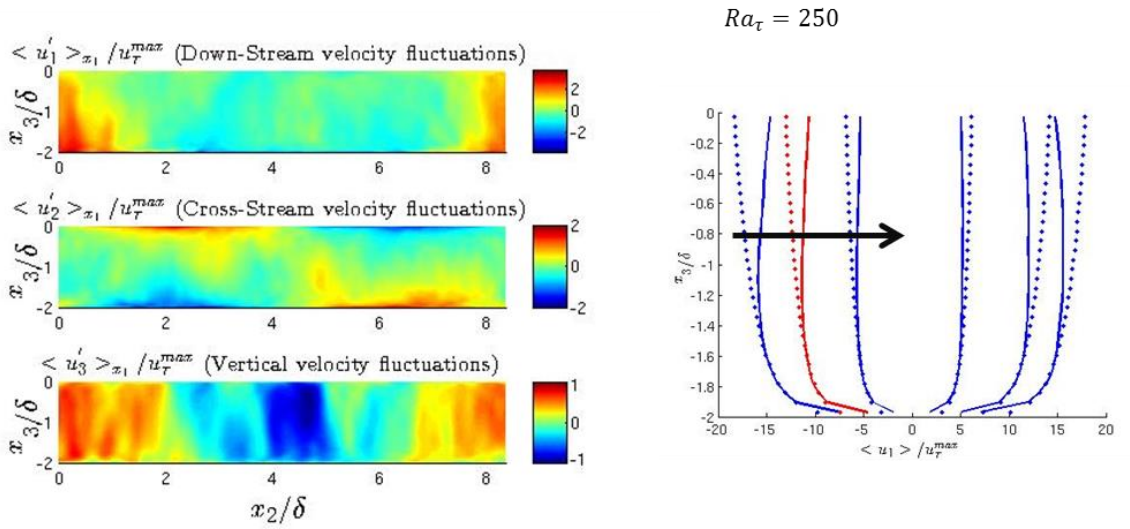


Figure 7. Same as caption of Figure 6 but for flow during deceleration stage.

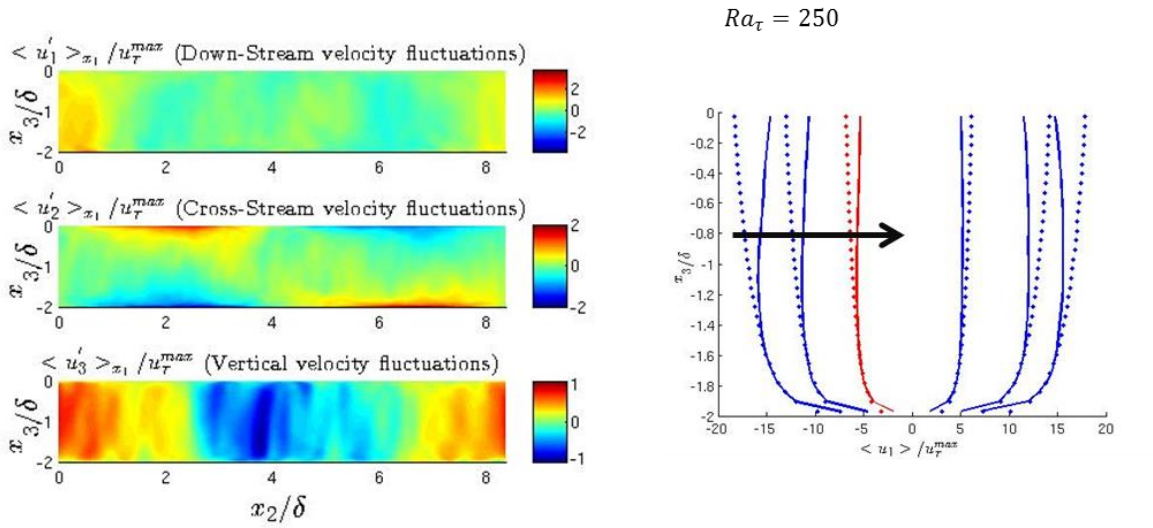


Figure 8. Same as caption of Figure 6 but for flow during deceleration stage.

responsible for the greater homogenization of the velocity profiles shown on the right panel of Figure 6 (i.e. greater vertical mixing of momentum), with respect to the velocity profiles for the flow without surface cooling in Figure 5.

In Figures 7 and 8, it can be seen that despite losing some strength as the flow decelerates, the convective supercell resolved remains visible and significantly serves to increase vertical mixing of momentum throughout the entire tidal cycle, relative to the flow without surface cooling.

As the flow transitions from deceleration to acceleration, remnants of the convective supercell can be observed in Figure 9, regaining strength as the flow accelerates back towards peak tide (Figure 10).

Conclusions

Unstratified open channel flow driven by an oscillating pressure gradient tidal (body) force and zero surface heat flux was shown to be characterized by weakly coherent streamwise-elongated counter-rotating cells occupying or engulfing the bulk of the water column. Surface

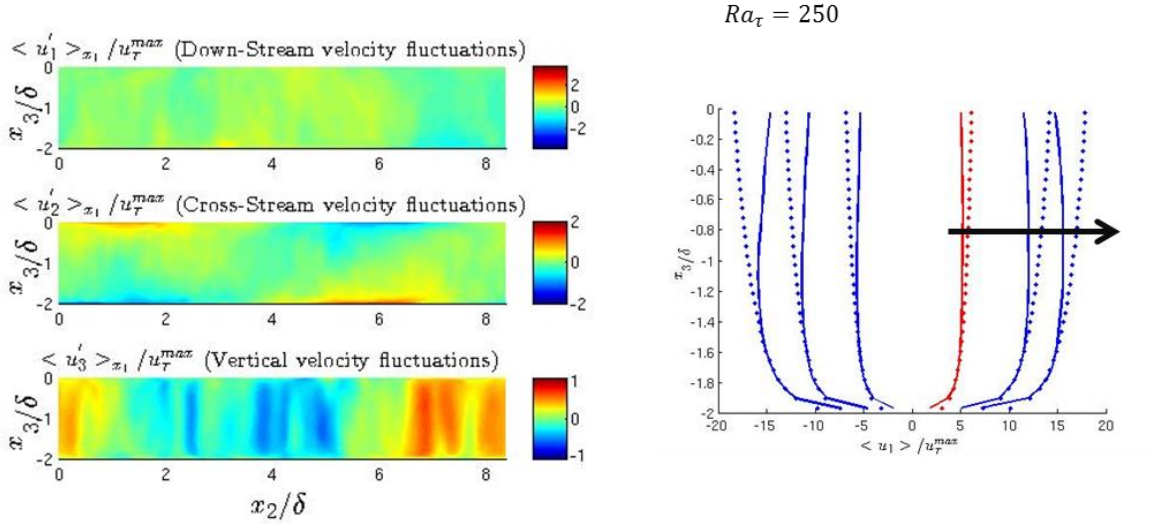


Figure 9. Same as caption of Figure 6 but for flow during acceleration stage.

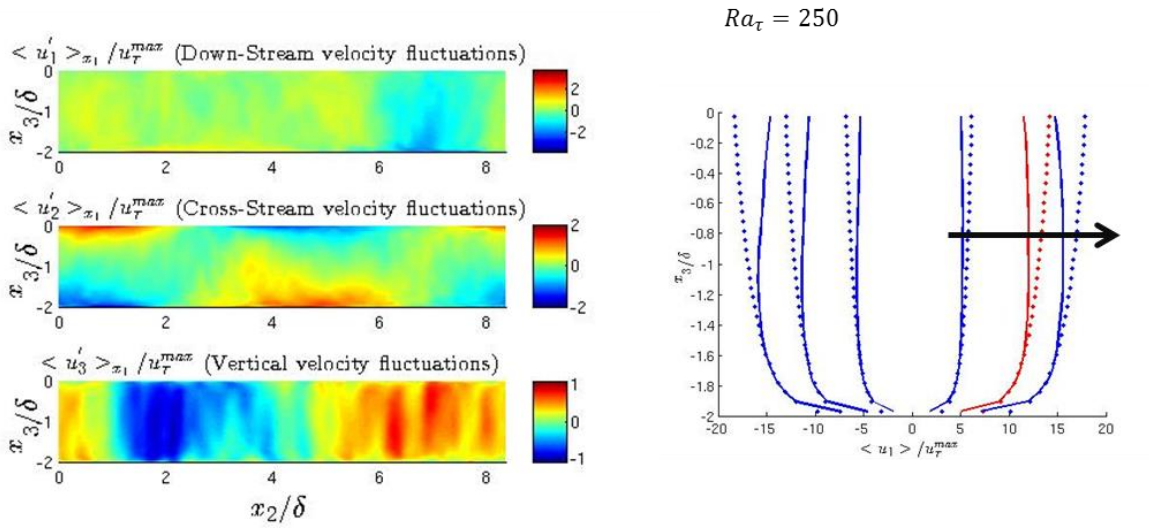


Figure 10. Same as caption of Figure 6 but for flow during acceleration state.

cooling with a heat flux of about 200 Watts m^{-2} led to convective supercells characterized by greater crosswind width and greater intensity and coherency. These cells are able to homogenize the depth profile of the mean velocity, causing deviation from the classical log-law velocity profile, throughout the entire tidal cycle. Although it was not shown here, the surface cooling and associated convective supercells are able to alter the turbulence structure throughout the water column. For example in flow without surface cooling, the middle of the water column is characterized by shear-dominated turbulence throughout the tidal cycle. In the flow with surface cooling of about 200 Watts m^{-2}, the stronger cells are able to induce higher vertical velocity fluctuations in the middle of the water column leading towards an isotropic turbulence structure as the flow transitions from deceleration to acceleration. In a future journal article we will explore the changing turbulence structure throughout the tidal cycle with and without surface cooling with the aid of Lumley invariant maps [5] in addition to the effects of Reynolds, Rossby and Raleigh number on the structure of the convective supercells and the turbulence.

References

[1] Walker, R, Tejada-Martínez, A.E., Martinat, G. and Grosh, C.E. (2015) Large-eddy simulation of open chanel flow with surface cooling, *International Journal of Heat and Fluid Flow* **50**, 209 – 224.

[2] Li. M, Sanford, L., and Chao, S.Y (2005) Effects of time dependence in unstratified tidal boundary layers: results from large eddy simulations, *Estuarine Coastal and Shelf Science* **62**, 193 – 204.

[3] Thorpe, S.A. (2004) Langmuir Circulation, *Annu. Rev. Fluid Mech* **36**, 55 – 79.

[4] Papavassiliou, D.V., Hanratty, T.J., (1997) Interpretation of large-scale structures observed in a turbulent plane Couette flow. *Int. J. Heat Fluid Flow* **18**, 55–69.

[5] Pope, S.B. (2000) Turbulent Flows. *Cambridge University Press*.

[6] Fernandez, A. Personal communication.

[7] openFOAM User Manual.

Adaptive Central-upwind Weighted Compact Non-linear Scheme with Increasing Order of Accuracy

†Kamyar Mansour[1], Kaveh Fardipour[1]

[1]Department of Aerospace Engineering, Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran.

†Corresponding author: mansour@aut.ac.ir

Abstract

In this work, effect of using an adaptive central-upwind (ACU) interpolation on weighted compact non-linear scheme (WCNS) is investigated. Based on the smoothness of solution, this type of interpolation adapts between central and upwind stencils by a weighting relation and combination of smoothness indicators of the optimal high-order stencil and its sub-stencils. The coefficients of sixth to tenth order ACU-WCNS are calculated. To evaluate basic numerical characteristics of this new schemes truncation error analysis and wavenumber analysis is performed and by applying ACU-WCNS on several benchmark problems, its shock-capturing abilities, its behavior in presence of severe discontinuity and its numerical resolution in shock-entropy interaction are investigated.

Keywords: High-order numerical method; Weighted compact nonlinear scheme; Shock-capturing; Compressible flow.

Introduction

Over past three decades there were many efforts for development of high-order numerical methods that simultaneously have the capability to capture flow discontinuity and resolve small-scale features of flow. Weighted Essentially Non-oscillatory (WENO) [1] scheme and Weighted Compact Nonlinear Scheme (WCNS) [2] scheme are two families of such numerical methods.

The WENO scheme is based on Essentially Non-oscillatory (ENO) scheme [3], but instead of using only one of sub-stencils, it uses a weighted combination of all sub-stencils. This scheme was developed in finite volume framework by Liu et al. [4]. Jiang and Shu [1] extended the WENO scheme to finite difference framework and proposed a new formulation for nonlinear weights to increase order of accuracy and later Balsara and Shu [5] and Gerolymos et al. [6] studied the high order behavior of the WENO scheme.

Despite having high order of accuracy and good shock capturing capabilities, the WENO scheme also has some shortcomings. One of the problems with the original WENO scheme of Jiang and Shu [1] is loss of accuracy near critical points. Analysis of Henrick et al. [7] showed this loss of accuracy is because of nonlinear weights and they purposed a mapping method for computation of nonlinear weights to prevent loss of accuracy. Borges et al. [8] also purposed a new method for computation of nonlinear weights of fifth order WENO to avoid loss of accuracy and later expanded it for higher order of accuracy [9], their method has lower computational cost in comparison to Henrick et al. [7] mapping method.

There are several ways to reduce numerical dissipation of the WENO scheme. One them is hybrid methods which only use the WENO scheme in vicinity of discontinuities and use another scheme with lower or no numerical dissipation in smooth regions [10]-[12]. Another way is to optimize dissipation and dispersion error [13]-[15], this usually achieved by finding optimal coefficients or linear weights by minimizing integral error following optimizing procedures of Tam and Webb [16] and Zhuang and Chen [17]. A more recent way for reduction of numerical dissipation of the WENO scheme is adaptive central-upwind WENO (ACU-WENO) scheme [18]-[20]. Based on smoothness of solution, this new family of WENO scheme can adapt between central and upwind stencils and achieves higher order of

accuracy, numerical resolution and lower dissipation by using a central stencil in smooth regions of solution.

The WCNS was originally developed by Deng and Zhang [21] and later extended to higher order of accuracy by Nonomura et al. [22] and Zhang et al. [23]. This scheme is a combination of compact scheme [24] and WENO interpolation [25]. This scheme includes a node-to-midpoint weighted interpolation and a midpoint-to-node differencing and has three advantages over finite difference WENO scheme [26]:

1- Slightly higher numerical resolution;
2- Compatibility with different flux treatments;
3- Better performance on general curvilinear grids [27].

Nonomura and Fujii [28] studied effects of different types of midpoint-to-node differencing methods on WCNS and they showed it does not significantly change numerical resolution and shock capturing capabilities of WCNS. Nonomura and Fujii [26] proposed a new formulation for midpoint-to-node differencing, which significantly increases robustness of WNCS. Recently Sumi and Kurotaki [29] used a sixth order adaptive central-upwind interpolation with robust formulation of a tridiagonal midpoint-to-node differencing to improve numerical resolution and robustness of original WCNS [21]. Some studies [22][30]-[32] showed increasing the order of accuracy of numerical method, will increase computational efficiency. Therefore in this paper we intend to study ACU-WCNS with order of accuracy higher than sixth.

Construction of the Numerical Scheme

For numerical solution of a conservation law as

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=0, \tag{1}$$

where t is time, x is a spatial dimension, u is function of x and t and f is flux function, Eq. (1) can be written in a semidiscretized form as

$$\left(\frac{\partial u}{\partial t}\right)_i=-f_i', \tag{2}$$

where f_i' is an approximation of spatial derivative of f on grid point x_i. Following Lele [24], for computation of f_i' in Eq. (2) we can use a linear formulation as

$$f_i'+\sum_{j=1}^{M}a_j\left(f_{i+j}'+f_{i-j}'\right)=\frac{1}{h}\sum_{k=1}^{N}b_l\left(f_{i+k-\frac{1}{2}}-f_{i-k+\frac{1}{2}}\right), \tag{3}$$

where M and N are positive integers. We can derive the coefficients a_j and b_l by matching the Taylor series coefficients [24]. The robust formulation of Nonomura and Fujii [26], which uses a midpoint-and-node-to-node differencing, can be written in a general form as

$$f_i'+\sum_{j=1}^{M}a_j\left(f_{i+j}'+f_{i-j}'\right)=\frac{1}{h}\sum_{k=1}^{N}c_k\left(f_{i+\frac{k}{2}}-f_{i-\frac{k}{2}}\right), \tag{4}$$

Following Nonomura and Fujii [28] we only use explicit form of Eq. (3) and Eq. (4) (i.e. $a_j=0$). For explicit form of these equations, the coefficients are listed in [28] and we list them again in Table 1 and Table 2, respectively for Eq. (3) and Eq. (4).

Table 1. Coefficients for Eq. (3) [28]

Coefficients	b_1	b_2	b_3	b_4	b_5
Fourth-order explicit	$\frac{9}{8}$	$-\frac{1}{24}$	0	0	0
Sixth-order explicit	$\frac{75}{64}$	$-\frac{25}{384}$	$\frac{3}{640}$	0	0
Eighth-order explicit	$\frac{1225}{1024}$	$-\frac{245}{3072}$	$\frac{49}{5120}$	$-\frac{5}{7168}$	0
Tenth-order explicit	$\frac{19845}{16384}$	$-\frac{735}{8192}$	$\frac{567}{40960}$	$-\frac{405}{229376}$	$\frac{35}{294912}$

Table 2. Coefficients for Eq. (4) [28]

Coefficients	c_1	c_2	c_3	c_4	c_5
Fourth-order explicit	$\frac{4}{3}$	$-\frac{1}{6}$	0	0	0
Sixth-order explicit	$\frac{3}{2}$	$-\frac{3}{10}$	$\frac{1}{30}$	0	0
Eighth-order explicit	$\frac{8}{5}$	$-\frac{2}{5}$	$\frac{8}{105}$	$-\frac{1}{140}$	0
Tenth-order explicit	$\frac{5}{3}$	$-\frac{10}{21}$	$\frac{2}{42}$	$-\frac{5}{252}$	$\frac{1}{630}$

To interpolate midpoint values from node values (to save space we only write formulation for left-biased interpolation, which is shown by superscript L and the right-biased interpolation could be derived by mirroring the left-biased interpolation around $x_{i+\frac{1}{2}}$), in a stencil $S^{(2r-1)} = \left(x_{i-r+1},...,x_{i+r-1}\right)$ with $\left(2r-1\right)$ points and r substencils as $S_k^{(2r-1)} = \left(x_{i+k-r+1},...,x_{i+k}\right)$, we can use a linear formulation as

$$\hat{f}^{L}_{i+\frac{1}{2}} = \sum_{k=-r+1}^{r+1} d_k \, f_{i+k}, \tag{5}$$

where d_k is constant coefficient. If we consider r substencils as $S_k^{(2r-1)} = \left(x_{i+k-r+1},...,x_{i+k}\right)$ in $S^{(2r-1)}$, we can use a linear formulation as

$$\hat{f}^{L}_{i+\frac{1}{2}} = \sum_{k=0}^{r-1} d_k^r \, \hat{f}^{r}_{k,i+\frac{1}{2}}, \tag{6}$$

where d_k^r is linear weight and $\hat{f}^r_{k,i+\frac{1}{2}}$ is the interpolated value for each substencil. We can write $\hat{f}^r_{k,i+\frac{1}{2}}$ as

$$\hat{f}^r_{k,i+\frac{1}{2}} = \sum_{l=0}^{r-1} a^r_{k,l} f_{i-k+j}, \tag{7}$$

where $a^r_{k,l}$ is constant coefficient. The linear relation in Eq. (6) cannot capture discontinuities accurately. To solve this problem we can combine the substencil values of Eq. (7) by a nonlinear formulation as

$$\hat{f}^L_{i+\frac{1}{2}} = \sum_{k=0}^{r-1} \omega^r_k \hat{f}^r_{k,i+\frac{1}{2}}, \tag{8}$$

where ω^r_k is nonlinear weight and is given by

$$\omega^r_k = \frac{\alpha^r_k}{\sum_{l=0}^{r-1} \alpha^r_l}, \tag{9}$$

$$\alpha^r_k = \frac{d^r_k}{\left(\varepsilon + IS^r_k\right)^p}, k = 0,...,r-1, \tag{10}$$

where ε is small positive value to avoid division by zero, p is a positive integer and IS^r_k is smoothness indicator and is given by

$$IS^r_k = \sum_{l=1}^{r-1} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \Delta x^{2l-1} \left(\frac{\partial^l f^{(r)}(x)}{\partial x^l} \right)^2 dx. \tag{11}$$

Hu et al. [19] proposed an alternative procedure for the WENO scheme which smoothly adapts between central and upwind stencils. According to this concept, to interpolate midpoint values from node values, instead of using biased stencil $S^{(2r-1)}$, we use a central stencil $S^{(2r)} = \left(x_{i-r+1},...,x_{i+r}\right)$ with $\left(2r\right)$ points and $r+1$ substencils. To include the new substencil, we should rewrite Eq. (5) to Eq. (8) as

$$\hat{f}^L_{i+\frac{1}{2}} = \sum_{k=-r+1}^{r} d_k f_{i+k}, \tag{12}$$

$$\hat{f}^L_{i+\frac{1}{2}} = \sum_{k=0}^{r} d^r_k \hat{f}^r_{k,i+\frac{1}{2}}, \tag{13}$$

$$\hat{f}^L_{i+\frac{1}{2}} = \sum_{k=0}^{r} \omega^r_k \hat{f}^r_{k,i+\frac{1}{2}}. \tag{14}$$

the nonlinear weight is given by

$$\omega_k^r = \frac{\alpha_k^r}{\sum_{l=0}^{r} \alpha_l^r}, \tag{15}$$

$$\alpha_k^r = d_k^r \left(C + \frac{\tau_{2r}}{\varepsilon + IS_k^r} \right), k = 0, ..., r, \tag{16}$$

where C is a constant and $C \gg 1$. τ_{2r} is a new reference smoothness indicator. To avoid oscillations near discontinuities, instead of using IS_r^r in Eq. (16), we use IS_{2r} which is smoothness indicator of the complete stencil. To increase numerical resolution Hu and Adams [20] proposed a new formulation for computation of α_k^r

$$\alpha_k^r = d_k^r \left(C_q + \frac{\tau_{2r}}{IS_k^r + \varepsilon \Delta x^2} \frac{IS_{ave}^r + \chi \Delta x^2}{IS_k^r + \chi \Delta x^2} \right)^q, k = 0, ..., r, \tag{17}$$

where C_q is a constant and $C_q \gg C$. $\chi = \frac{1}{\varepsilon}$ and IS_{ave}^r is an average of smoothness indicator of different substencils and there is a relation between τ_{2r}, IS_{2r} and IS_{ave}^r as

$$\tau_{2r} = IS_{2r} - IS_{ave}^r. \tag{18}$$

Some values and formulas for d_k^r, $\hat{f}^r_{k,i+\frac{1}{2}}$, IS_k^r, IS_{2r} and IS_{ave}^r are given in appendix A.

Truncation Error Analysis

Following Hu et al. [19], in this we perform a truncation error to find sufficient condition for ACU-WCNS to achieve the designed order of accuracy. We can write below relations between $f_{i+\frac{1}{2}}$ and interpolated values $\hat{f}_{i+\frac{1}{2}}$ from Eq. (13) and $\hat{f}^r_{k,i+\frac{1}{2}}$ from Eq. (7)

$$\hat{f}_{i+\frac{1}{2}} = f_{i+\frac{1}{2}} + O(\Delta x^{2r}), \tag{19}$$

$$\hat{f}^r_{k,i+\frac{1}{2}} = f_{i+\frac{1}{2}} + A_k^r \Delta x^r + O(\Delta x^{r+1}). \tag{20}$$

We can rewrite Eq. (14) as

$$\hat{f}_{i+\frac{1}{2}} = \sum_{k=0}^{r} d_k^r \hat{f}^r_{k,i+\frac{1}{2}} + \sum_{k=0}^{r} \left(\omega_k^r - d_k^r \right) \hat{f}^r_{k,i+\frac{1}{2}}, \tag{21}$$

Using the first linear term on the right-hand-side of Eq. (21) in Eq. (3) or Eq. (4) leads to derivative of $O(\Delta x^{2r})$. Therefore the sufficient condition for Eq. (3) or Eq. (4) to be of $O(\Delta x^{2r})$ is that the term on the right-hand-side of Eq. (21) is at least $O(\Delta x^{2r+1})$. Using Eq. (20) we can expand this term as

$$\sum_{k=0}^{r} \left(\omega_k^r - d_k^r \right) \hat{f}^r_{k,i+\frac{1}{2}} = f_{i+\frac{1}{2}} \sum_{k=0}^{r} \left(\omega_k^r - d_k^r \right) + A_k^r \Delta x^r \sum_{k=0}^{r} \left(\omega_k^r - d_k^r \right) + O(\Delta x^{r+1}) \sum_{k=0}^{r} \left(\omega_k^r - d_k^r \right). \tag{22}$$

The first term on the right-hand-side of Eq. (22) is zero because of normalization of the weights, therefore the sufficient condition for having a $O(\Delta x^{2r})$ derivative is

$$\omega_k^r - d_k^r = O(\Delta x^{r+1}). \tag{23}$$

Numerical Examples

In this section, we provide some numerical examples to show shock-capturing capabilities and numerical resolution of the proposed ACU-WCNS scheme. These problems are described by compressible Euler equations

$$\frac{\partial}{\partial t}\begin{bmatrix}\rho\\ \rho u\\ E\end{bmatrix}+\frac{\partial}{\partial x}\begin{bmatrix}\rho u\\ \rho u^2+p\\ (E+p)u\end{bmatrix}=0, \tag{24}$$

where ρ is density, p is pressure, u is x component of velocity vector, E is total energy and related to pressure as $e=\frac{p}{\gamma-1}+\frac{1}{2}\rho u^2$ and $\gamma=1.4$ is the ratio of specific heats. To reduce numerical oscillations we use local characteristic decomposition by Roe averaged variables and we use the Lax-Friedrichs method for flux vector splitting. For time integration we use a third order TVD Runge-Kutta method [33]. We used Eq. (4) for midpoint-and-node-to-node differencing and Eq. (17) for calculation of α_k^r, $C_q=1000$ and $q=2(r-1)$.

In a series of benchmark problems, results of ACU-WCNS with six, eight and ten point stencils are compared with ninth order WENO-Z [9]. The first problem is Sod shock tube [34] and initial condition is defined as

$$(\rho,u,p)=\begin{cases}(1,0,1) & if \quad 0<x<0.5,\\ (0.125,0,0.1) & if \quad 0.5<x<1.\end{cases}$$

The second problem is Lax shock tube [35] and initial condition is

$$(\rho,u,p)=\begin{cases}(0.445,0.698,0.3528) & if \quad 0<x<0.5,\\ (0.5,0,0.571) & if \quad 0.5<x<1.\end{cases}$$

The third problem is 123 shock tube [36] with initial condition as

$$(\rho,u,p)=\begin{cases}(1,-2,0.4) & if \quad 0<x<0.5,\\ (1,2,0.4) & if \quad 0.5<x<1.\end{cases}$$

We choose the fourth problem from Nonomura and Fujii [26]. This shock tube has a very high pressure ratio and includes a severe shock. Initial condition is

$$(\rho,u,p)=\begin{cases}(1,0,10000) & if \quad 0<x<0.5,\\ (0.125,0,0.1) & if \quad 0.5<x<1.\end{cases}$$

The fifth problem is from Toro [37] and this problem also includes a severe shock. Initial condition is

$$(\rho,u,p)=\begin{cases}(1,-19.59745,1000) & if \quad 0<x<0.5,\\ (1,-19.59745,0.01) & if \quad 0.5<x<1.\end{cases}$$

The last problem is the Shu-Osher problem [38] and initial condition is

$$(\rho, u, p) = \begin{cases} (3.857, 2.629, 10.333) & if \quad 0 < x < 1, \\ (1 + 0.2\sin(5x), 0, 1) & if \quad 1 < x < 10. \end{cases}$$

Fig. (1) and Fig. (2) respectively show density distribution for the Sod and the Lax problems. All ACU-WCNS show good capturing abilities and there are no visible numerical oscillations. It should be noted some adaptive central-upwind [18] or optimized [15] WENO schemes have numerical oscillations in these problem and therefore we could conclude Hu et al [19] adaption mechanism also works well in ACU-WCNS.

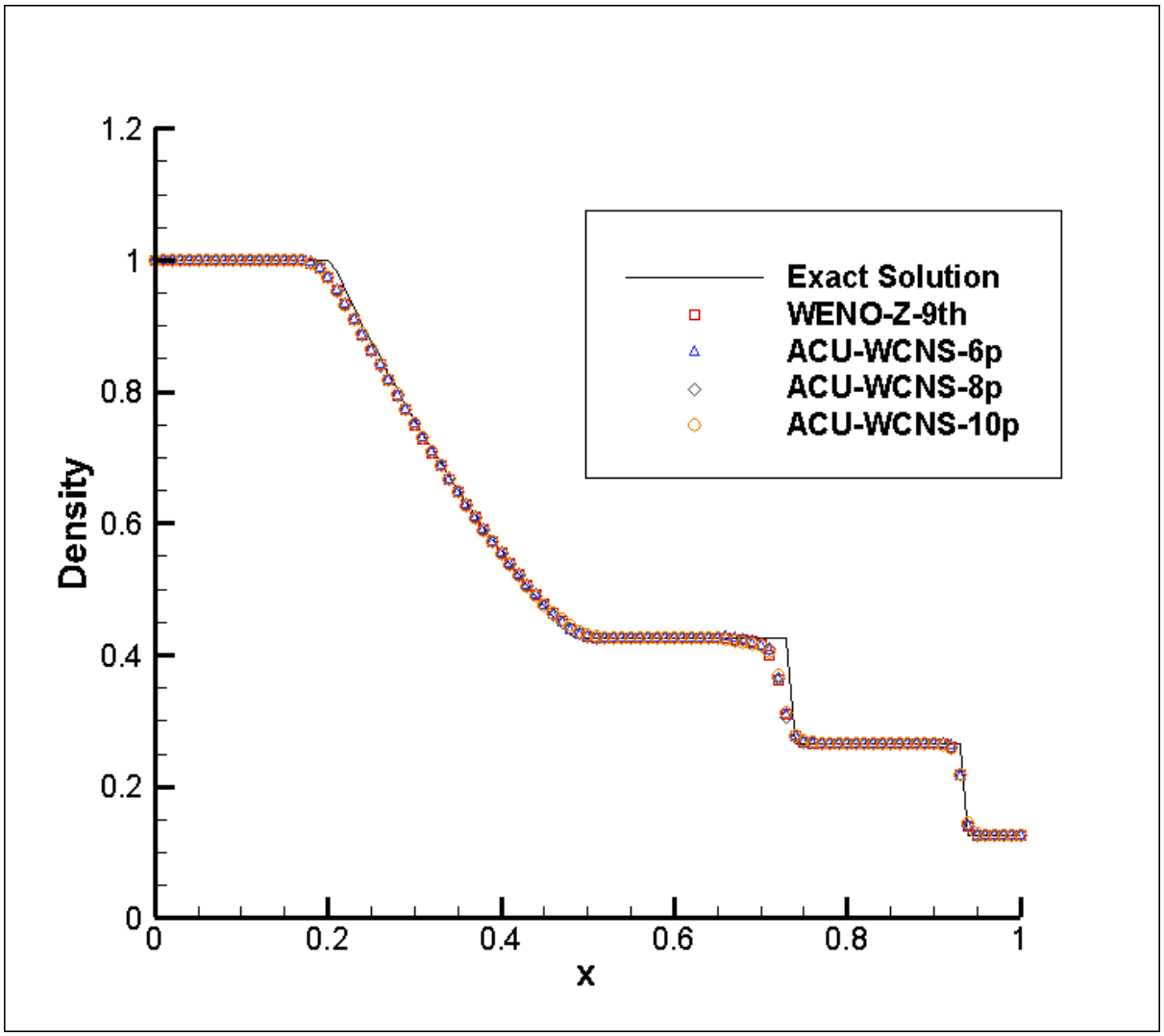


Figure 1. Density distribution for the Sod problem with 100 grid points at t=0.25 s

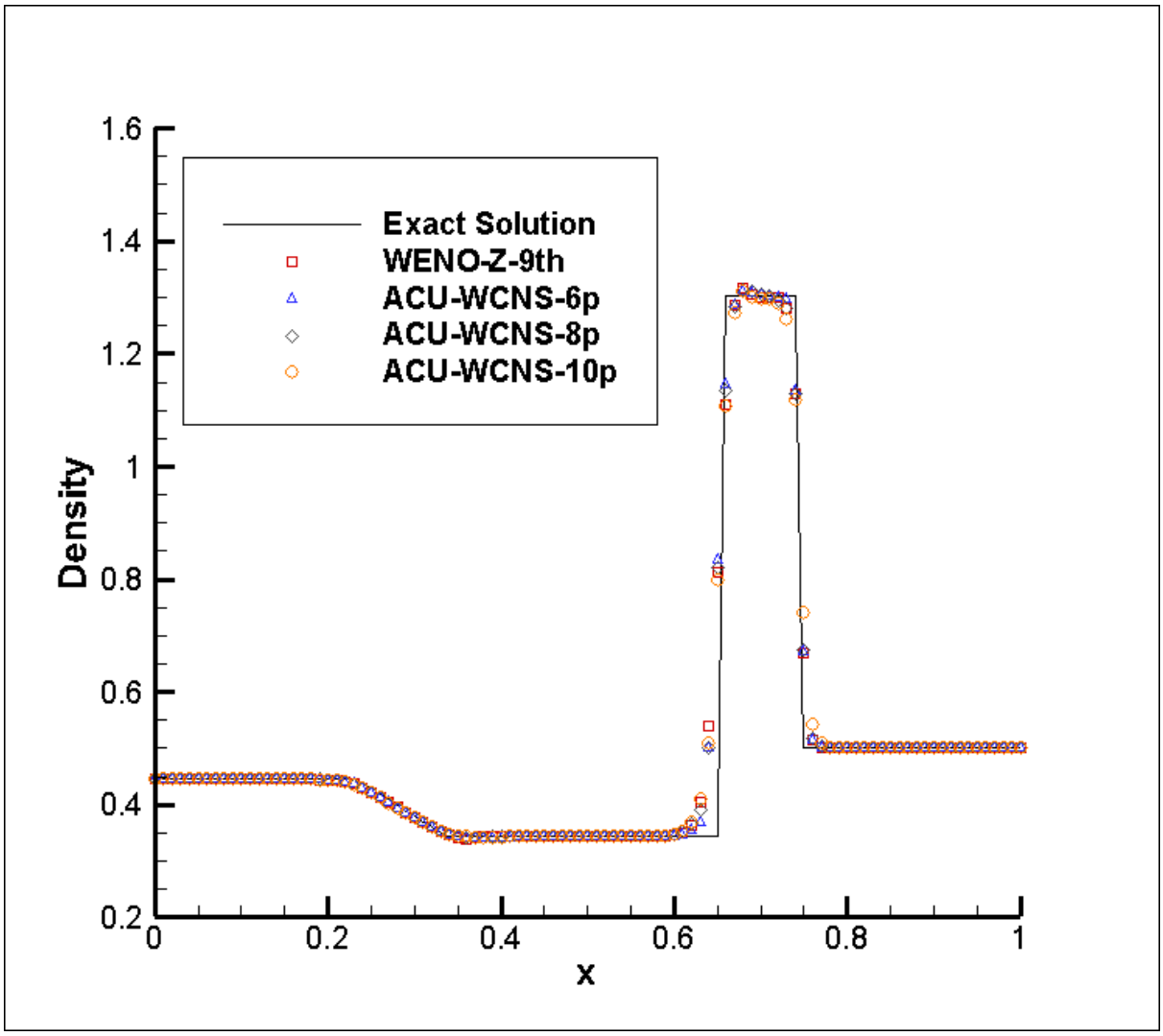


Figure 2. Density distribution for the Lax problem with 100 grid points at t=0.1 s

The third to fifth problems are cases with severe conditions and Fig. (3) to Fig. (5) show their density distribution. The 123 problem contains a near-vacuum condition and is suitable for assessment of numerical methods in low pressure and density situations. All ACU-WCNS show good results for this problem. The fourth and fifth problems contain strong shocks. All methods show good results except ACU-WCNS with ten points stencil, therefore this method is not as robust as other methods.

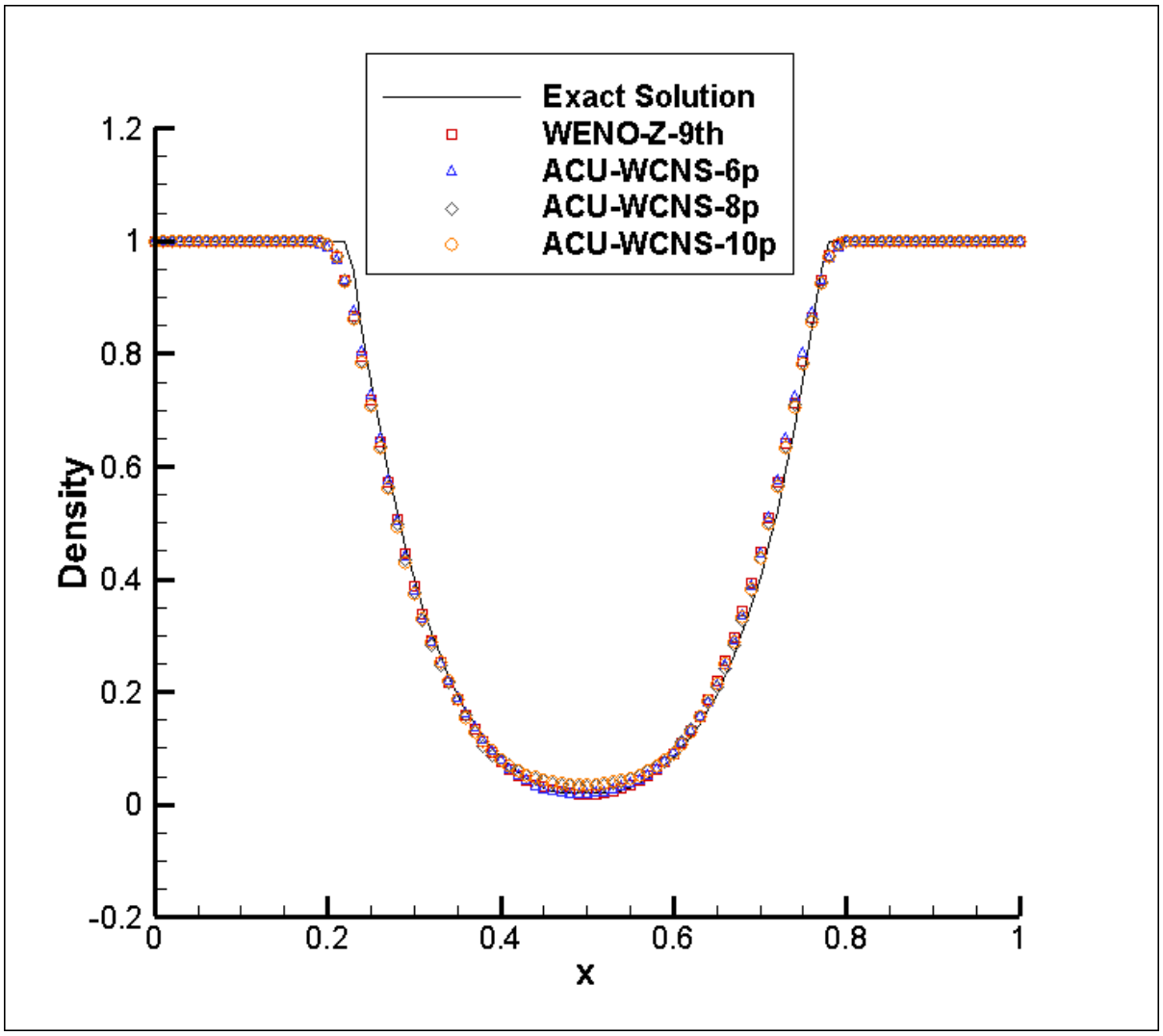


Figure 3. Density distribution for the 123 problem with 100 grid points at t=0.1 s

S

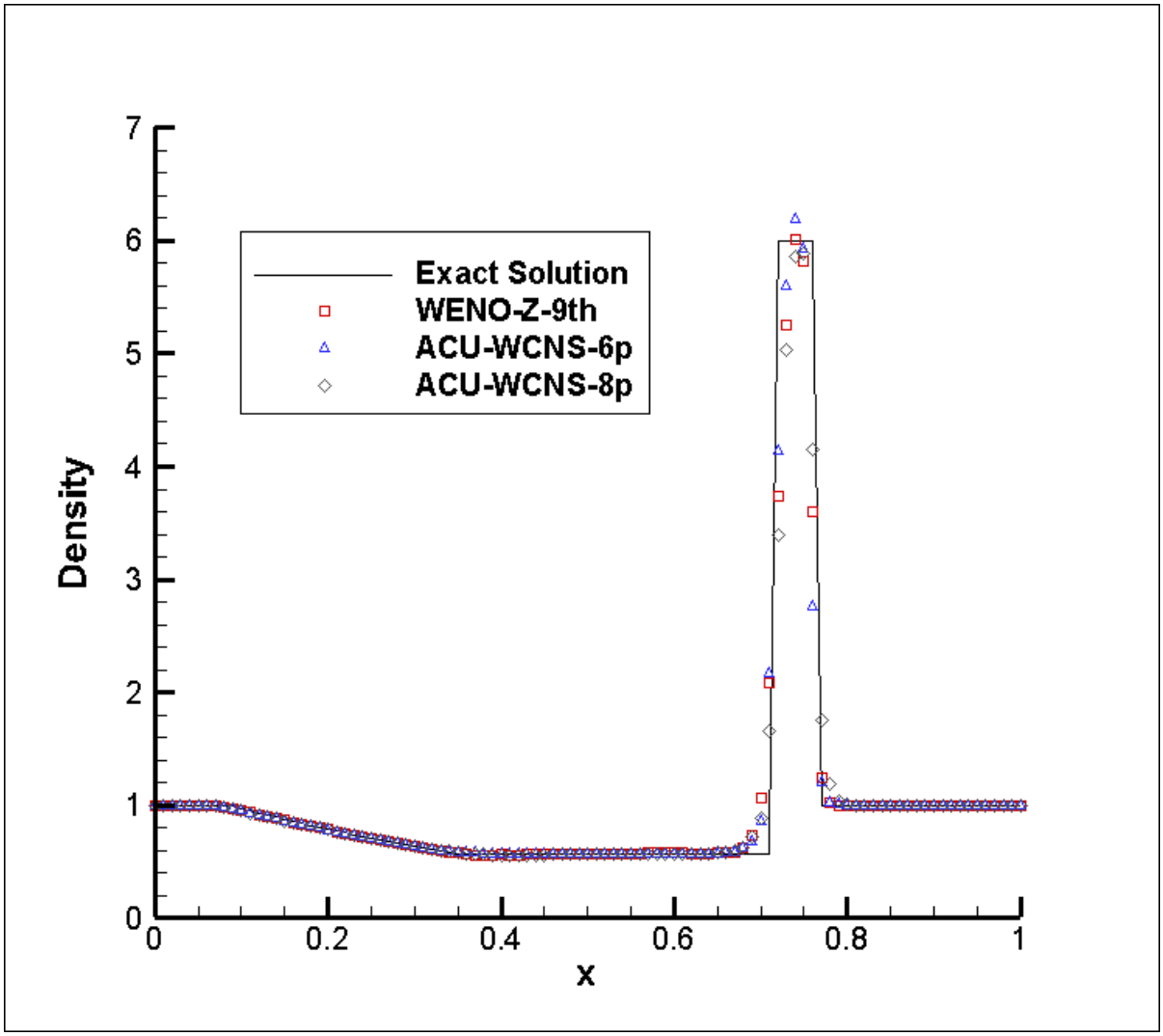


Figure 4. Density distribution for the forth problem with 100 grid points at t=0.0035 s

S

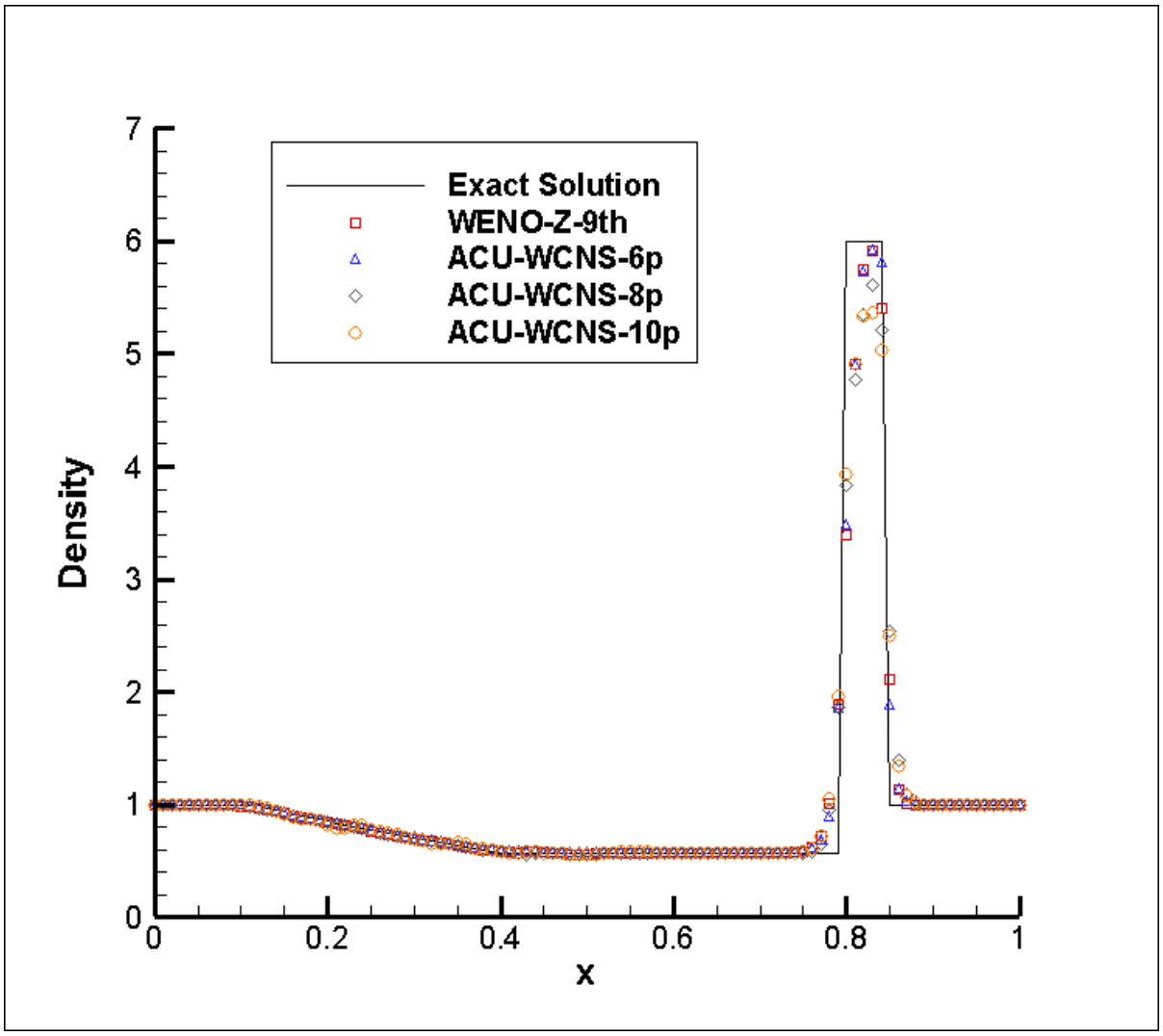


Figure 5. Density distribution for the fifth problem with 100 grid points at t=0.012 s

Fig. (6) shows the density distribution for the Shu-Osher problem. This problem includes an interaction between an entropy wave and a shock wave and resolution of density oscillations after the shock is a good criteria for investigate the resolution of a numerical method. The reference solution this problem is calculated by a fifth order WENO-JS [1] scheme. All ACU-WCNS show good numerical resolution and their resolution is superior to that of ninth order WENO-Z.

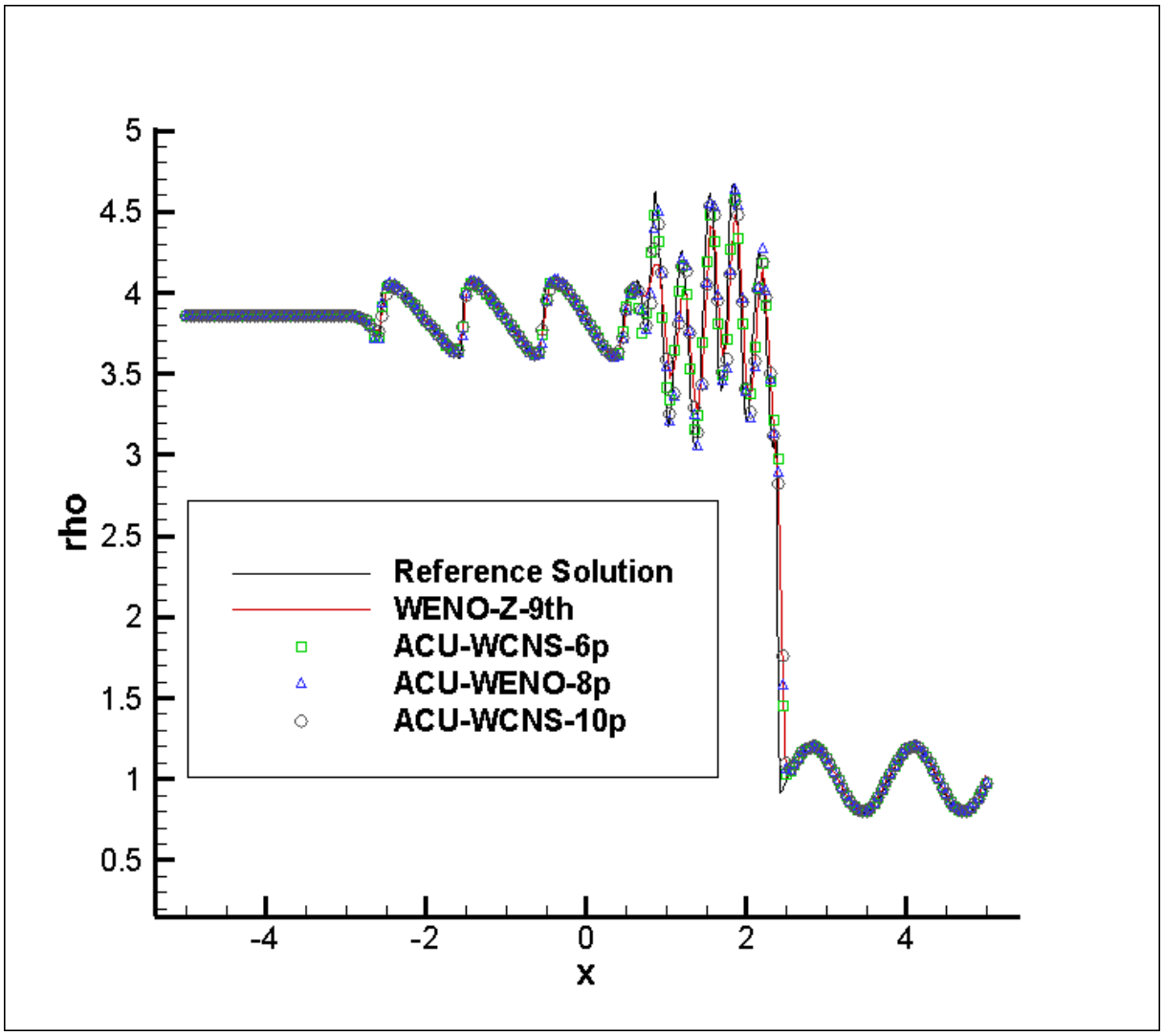


Figure 6. Density distribution for the Shu-Osher problem with 200 grid points t=1.8 s

Conclusions

In this paper we developed an adaptive interpolation procedure for WCNS scheme which adapts between upwind and central stencil based on smoothness of solution. The shock-capturing capabilities of the new scheme and its robustness was tested by solving several benchmark problems. The results of benchmark problems shows the new scheme has good shock capturing capabilities and high numerical resolution.

Appendix A

To avoid exceeding the limit for number of pages in a paper, we omitted the values and formulas for d_k^r, $\hat{f}^r_{k,i+\frac{1}{2}}$, IS_k^r, IS_{2r} and IS_{ave}^r are given for $r=3$ and $r=4$ and only give these values and formulas for $r=5$.

$$d_0^5=\frac{1}{512}, d_1^5=\frac{45}{512}, d_2^5=\frac{105}{256}, d_3^5=\frac{105}{256}, d_4^5=\frac{45}{512}, d_5^5=\frac{1}{512}. \tag{A11}$$

$$
\begin{cases}
\hat{f}^{5}_{0,i+\frac{1}{2}} = \dfrac{35}{128} f_{i-4} - \dfrac{45}{32} f_{i-3} + \dfrac{189}{64} f_{i-2} - \dfrac{105}{32} f_{i-1} + \dfrac{315}{128} f_{i} \\
\hat{f}^{5}_{1,i+\frac{1}{2}} = -\dfrac{5}{128} f_{i-3} + \dfrac{7}{32} f_{i-2} - \dfrac{35}{64} f_{i-1} + \dfrac{35}{32} f_{i} + \dfrac{35}{128} f_{i+1} \\
\hat{f}^{5}_{2,i+\frac{1}{2}} = \dfrac{3}{128} f_{i-2} - \dfrac{5}{32} f_{i-1} + \dfrac{45}{64} f_{i} + \dfrac{15}{32} f_{i+1} - \dfrac{5}{128} f_{i+2} \\
\hat{f}^{5}_{3,i+\frac{1}{2}} = -\dfrac{5}{128} f_{i-1} + \dfrac{15}{32} f_{i} + \dfrac{45}{64} f_{i+1} - \dfrac{5}{32} f_{i+2} + \dfrac{3}{128} f_{i+3} \\
\hat{f}^{5}_{4,i+\frac{1}{2}} = \dfrac{35}{128} f_{i} + \dfrac{35}{32} f_{i+1} - \dfrac{35}{64} f_{i+2} + \dfrac{7}{32} f_{i+3} - \dfrac{5}{128} f_{i+4} \\
\hat{f}^{5}_{5,i+\frac{1}{2}} = \dfrac{35}{128} f_{i+5} - \dfrac{45}{32} f_{i+4} + \dfrac{189}{64} f_{i+3} - \dfrac{105}{32} f_{i+2} + \dfrac{315}{128} f_{i+1}
\end{cases}
. \qquad \text{(A12)}
$$

$$
\begin{cases}
IS_0^5 = -\dfrac{2569471 f_{i-4} f_{i-3}}{60480} + \dfrac{1501039 f_{-4} f_{i-2}}{20160} - \dfrac{3568693 f_{i-4} f_{i-1}}{60480} + \dfrac{1076779 f_{i-4} f_i}{60480} + \dfrac{5951369 f_{i-3}^2}{60480} \\
- \dfrac{1751863 f_{i-3} f_{i-2}}{5040} + \dfrac{8405471 f_{i-3} f_{i-1}}{30240} - \dfrac{5121853 f_{i-3} f_i}{60480} + \dfrac{2085371 f_{i-2}^2}{6720} - \dfrac{2536843 f_{i-2} f_{i-1}}{5040} \\
+ \dfrac{3141559 f_{i-2} f_i}{20160} + \dfrac{12627689 f_{i-1}^2}{60480} - \dfrac{8055511 f_{i-1} f_i}{60480} + \dfrac{668977 f_i^2}{30240} + \dfrac{139567 f_{i-4}^2}{30240} \\
IS_1^5 = \dfrac{221869 f_{i-3} f_{i+1}}{60480} - \dfrac{1079563 f_{i-2} f_{i+1}}{60480} + \dfrac{671329 f_{i-1} f_{i+1}}{20160} - \dfrac{1714561 f_i f_{i+1}}{60480} + \dfrac{139567 f_{i+1}^2}{30240} \\
+ \dfrac{20591 f_{i-3}^2}{15120} - \dfrac{725461 f_{i-3} f_{i-2}}{60480} + \dfrac{395389 f_{i-3} f_{i-1}}{20160} - \dfrac{847303 f_{i-3} f_i}{60480} + \dfrac{1650569 f_{i-2}^2}{60480} \\
- \dfrac{57821 f_{i-2} f_{i-1}}{630} + \dfrac{2027351 f_{i-2} f_i}{30240} + \dfrac{539351 f_{i-1}^2}{6720} - \dfrac{306569 f_{i-1} f_i}{2520} + \dfrac{2932409 f_i^2}{60480} \\
IS_2^5 = \dfrac{20591 f_{i+2}^2}{15120} + \dfrac{98179 f_{i-2} f_{i+2}}{60480} - \dfrac{461113 f_{i-1} f_{i+2}}{60480} + \dfrac{266659 f_i f_{i+2}}{20160} - \dfrac{601771 f_{i+1} f_{i+2}}{60480} \\
- \dfrac{461113 f_{i-2} f_{i+1}}{60480} + \dfrac{1050431 f_{i-1} f_{i+1}}{30240} - \dfrac{291313 f_i f_{i+1}}{5040} + \dfrac{1228889 f_{i+1}^2}{60480} + \dfrac{20591 f_{i-2}^2}{15120} \\
- \dfrac{601771 f_{i-2} f_{i-1}}{60480} + \dfrac{266659 f_{i-2} f_i}{20160} + \dfrac{1228889 f_{i-1}^2}{60480} - \dfrac{291313 f_{i-1} f_i}{5040} + \dfrac{299531 f_i^2}{6720} \\
IS_3^5 = \dfrac{221869 f_{i-1} f_{i+3}}{60480} - \dfrac{847303 f_i f_{i+3}}{60480} + \dfrac{395389 f_{i+1} f_{i+3}}{20160} - \dfrac{725461 f_{i+2} f_{i+3}}{60480} + \dfrac{20591 f_{i+3}^2}{15120} \\
+ \dfrac{1650569 f_{i+2}^2}{60480} - \dfrac{1079563 f_{i-1} f_{i+2}}{60480} + \dfrac{2027351 f_i f_{i+2}}{30240} - \dfrac{57821 f_{i+1} f_{i+2}}{630} + \dfrac{671329 f_{i-1} f_{i+1}}{20160} \\
- \dfrac{306569 f_i f_{i+1}}{2520} + \dfrac{539351 f_{i+1}^2}{6720} + \dfrac{139567 f_{i-1}^2}{30240} - \dfrac{1714561 f_{i-1} f_i}{60480} + \dfrac{2932409 f_i^2}{60480} \\
IS_4^5 = -\dfrac{5121853 f_i f_{i+3}}{60480} + \dfrac{8405471 f_{i+1} f_{i+3}}{30240} - \dfrac{1751863 f_{i+2} f_{i+3}}{5040} + \dfrac{5951369 f_{i+3}^2}{60480} + \dfrac{1076779 f_i f_{i+4}}{60480} \\
- \dfrac{3568693 f_{i+1} f_{i+4}}{60480} + \dfrac{1501039 f_{i+2} f_{i+4}}{20160} - \dfrac{2569471 f_{i+3} f_{i+4}}{60480} + \dfrac{139567 f_{i+4}^2}{30240} + \dfrac{2085371 f_{i+2}^2}{6720} \\
+ \dfrac{3141559 f_i f_{i+2}}{20160} - \dfrac{2536843 f_{i+1} f_{i+2}}{5040} - \dfrac{8055511 f_i f_{i+1}}{60480} + \dfrac{12627689 f_{i+1}^2}{60480} + \dfrac{668977 f_i^2}{30240} \\
IS_5^5 = \dfrac{14813989 f_{i+1} f_{i+3}}{20160} - \dfrac{2982247 f_{i+2} f_{i+3}}{1260} + \dfrac{9836471 f_{i+3}^2}{6720} - \dfrac{24804943 f_{i+1} f_{i+4}}{60480} + \dfrac{40203671 f_{i+2} f_{i+4}}{30240} \\
- \dfrac{4166159 f_{i+3} f_{i+4}}{2520} + \dfrac{28344089 f_{i+4}^2}{60480} + \dfrac{5324029 f_{i+1} f_{i+5}}{60480} - \dfrac{17334403 f_{i+2} f_{i+5}}{60480} + \dfrac{7212409 f_{i+3} f_{i+5}}{20160} \\
- \dfrac{12302761 f_{i+4} f_{i+5}}{60480} + \dfrac{668977 f_{i+5}^2}{30240} + \dfrac{58316969 f_{i+2}^2}{60480} - \dfrac{36559021 f_{i+1} f_{i+2}}{60480} + \dfrac{724873 f_{i+1}^2}{7560}
\end{cases}
\tag{A13}
$$

$$IS_{10} = \frac{625652246900527564859}{72844785274060800} f_{i-1} f_{i+3} - \frac{9505458619090627988189}{728447852740608000} f_i f_{i+3}$$
$$+ \frac{9507517967943533533133}{728447852740608000} f_{i+1} f_{i+3} - \frac{3130783545165186871223}{364223926370304000} f_{i+2} f_{i+3}$$
$$+ \frac{1309941690301995642707}{728447852740608000} f_{i+3}^2 + \frac{887205217073247700925 3}{2913791410962432000} f_i f_{i+4}$$
$$- \frac{8969401895441394272981}{2913791410962432000} f_{i+1} f_{i+4} + \frac{2715884744243791857 41}{132445064134656000} f_{i+2} f_{i+4}$$
$$- \frac{25300800689519630695 9}{29137914109624320 0} f_{i+3} f_{i+4} + \frac{1236790723120202233859}{11655165643849728000} f_{i+4}^2$$
$$+ \frac{932643697717251472957}{2913791410962432000} f_{i+1} f_{i+5} - \frac{3875228907897276887}{17986366734336000} f_{i+2} f_{i+5}$$
$$+ \frac{10340996628529541551}{112068900421632000} f_{i+3} f_{i+5} - \frac{4927899975011358913}{215836400812032000} f_{i+4} f_{i+5}$$
$$+ \frac{39555919576041596 59}{31786815392317440 00} f_{i+5}^2 - \frac{1034236639532760969 67}{26489012826931200 0} f_{i-4} f_{i+1}$$
$$+ \frac{5299705740000020791}{22413780084326400} f_{i-4} f_{i+2} - \frac{14827167383164575419}{161877300609024000} f_{i-4} f_{i+3}$$
$$+ \frac{1201492985433636796 27}{58275828219248640 00} f_{i-4} f_{i+4} - \frac{3595521736053749053 7}{17482748465774592000} f_{i-4} f_{i+5}$$
$$- \frac{2479311724775857034 51}{11206890042163200 0} f_{i-3} f_{i+2} + \frac{9693122937839276243 9}{11206890042163200 0} f_{i-3} f_{i+3}$$
$$- \frac{1143858248597364187859}{5827582821924864000} f_{i-3} f_{i+4} + \frac{1149408483211501830 83}{58275828219248640 00} f_{i-3} f_{i+5}$$
$$- \frac{2668291129300573832 3}{7433141354496000} f_{i-2} f_{i+3} + \frac{3683521346038177823}{4482756016865280} f_{i-2} f_{i+4}$$
$$- \frac{1212215737967781346 43}{14568957054812160 00} f_{i-2} f_{i+5} - \frac{2223776828902720186747}{112068900421632000} f_{i-1} f_{i+4}$$
$$+ \frac{7567425717440384561}{37356300140544000} f_{i-1} f_{i+5} - \frac{3955591957604159659}{12613815631872000} f_i f_{i+5}$$
$$+ \frac{9838048439101619007 1}{9460361723904000} f_{i+2}^2 + \frac{3311611232916529587463}{364223926370304000} f_{i-2} f_{i+2}$$
$$- \frac{1120112685403347800557}{52031989481472000} f_{i-1} f_{i+2} + \frac{3365674159325425003583}{104063978962944000} f_i f_{i+2}$$
$$- \frac{2661470759469739877 9}{8325118317035 52} f_{i+1} f_{i+2} + \frac{2106642936305224401793}{582758282192486400} f_{i-3} f_{i+1}$$
$$- \frac{1070206666549807347257 3}{72844785274060800 0} f_{i-2} f_{i+1} + \frac{3577106583060361769087}{104063978962944000} f_{i-1} f_{i+1} \qquad \text{(A14)}$$
$$- \frac{1061511834159706036637 3}{20812795792588800 0} f_i f_{i+1} + \frac{1479832118857263560579}{59465130835968000} f_{i+1}^2$$
$$+ \frac{7120487564252807947}{317868153923174400 0} f_{i-4}^2 - \frac{4443403029532537814 3}{11655165643849728 00} f_{i-4} f_{i-3}$$

$$+\frac{5335640432737222309\,f_{i-4}f_{i-2}}{37356300140544000}-\frac{450698762943845856103\,f_{i-4}f_{i-1}}{1456895705481216000}$$
$$+\frac{1244753877809442799517\,f_{i-4}f_{i}}{2913791410962432000}+\frac{1933473108524561292707\,f_{i-3}^{\,2}}{11655165643849728000}$$
$$-\frac{1846150261009678724267\,f_{i-3}f_{i-2}}{1456895705481216000}+\frac{3695406612826630487 81\,f_{i-3}f_{i-1}}{132445064134656000}$$
$$-\frac{1138600296553219536590 9\,f_{i-3}f_{i}}{2913791410962432000}+\frac{1795138265821207839347\,f_{i-2}^{\,2}}{728447852740608000}$$
$$-\frac{8036823462504383880 43\,f_{i-2}f_{i-1}}{72844785274060800}+\frac{1631626449639313364747\,f_{i-2}f_{i}}{104063978962944000}$$
$$+\frac{1187058512647118810 47\,f_{i-1}^{\,2}}{9460361723904000}-\frac{3766828957892473535791\,f_{i-1}f_{i}}{104063978962944000}$$
$$+\frac{1102594454913534130066 1\,f_{i}^{\,2}}{416255915851776000}.$$

$$IS_{ave}^{5}=\frac{1}{16}\left(3IS_{1}^{5}+10IS_{2}^{5}+3IS_{3}^{5}\right). \tag{A15}$$

References

[1] Jiang, G. S. and Shu, C. W. (1996) Efficient Implementation of Weighted ENO Schemes, *Journal of Computational Physics* **126(1)**, 202-228.

[2] Deng, X. and Zhang, H. (2000) Developing High-Order Weighted Compact Nonlinear Schemes, *Journal of Computational Physics* **165(1)**, 22-44.

[3] Harten, A., Engquist, B., Osher, S. and Chakravarthy, S. R. (1987) Uniformly high order accurate essentially non-oscillatory schemes, III, *Journal of Computational Physics* **71(2)**, 231-303.

[4] Liu, X. D., Osher, S. and Chan, T. (1994) Weighted essentially non-oscillatory schemes, *Journal of Computational Physics* **115(1)**, 200-212.

[5] Balsara, D. S. and Shu, C. W. (2000) Monotonicity Preserving Weighted Essentially Non-oscillatory Schemes with Increasingly High Order of Accuracy, *Journal of Computational Physics* **160(2)**, 405-452.

[6] Gerolymos, G. A., Senechal, D. and Vallet, I. (2009) Very-high-order WENO schemes, *Journal of Computational Physics* **228(23)**, 8481–8524.

[7] Henrick, A. K., Aslam, T. D. and Powers, J. M. (2005) Mapped weighted essentially non-oscillatory schemes-Achieving optimal order near critical points, *Journal of Computational Physics* **207(2)**, 542-567.

[8] Borges, R., Carmona, M., Costa, B. and Don, W. S. (2008) An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws, *Journal of Computational Physics* **227(6)**, 3191-3211.

[9] Castro, M., Costa, B. and Don, W. S. (2011) High order weighted essentially non-oscillatory WENO-Z schemes for hyperbolic conservation laws, *Journal of Computational Physics* **230(5)**, 1766-1792.

[10] Pirozzoli, S. (2002) Conservative hybrid compact-WENO schemes for shock-turbulence interaction, *Journal of Computational Physics* **178(1)**, 81-117.

[11] Ren, X. Y., Liu, M, and Zhang, H. (2003) A characteristic-wise hybrid compact-WENO scheme for solving hyperbolic conservation laws, *Journal of Computational Physics* **192(2)**, 365-386.

[12] Kim, D. and Kwon, J. H. (2005) A high-order accurate hybrid scheme using a central flux scheme and a WENO scheme for compressible flow field analysis, *Journal* of *Computational Physics* **210(2)**, 554-583.

[13] Weirs, V. G. and Candler, G. V. (1997) Optimization of Weighted ENO Schemes for DNS of Compressible Turbulence, *AIAA Paper 97–1940*.

[14] Wang, Z. J. and Chen, R. F. (2001) Optimized weighted essentially nonoscillatory schemes for linear waves with discontinuity, *Journal of Computational Physics* **174(1)**, 381-404.

[15] Martín, M.P., Taylor, E.M., Wu, M. and Weirs, V.G. (2006) A bandwidth-optimized WENO scheme for the effective direct numerical simulation of compressible turbulence, *Journal of Computational Physics*, **220(1)**, 270-289.

[16] Tam, C. K. and Webb, J. C. (1993) Dispersion-relation-preserving finite difference schemes for computational acoustics, *Journal of computational physics*, **107(2)**, 262-281.

[17] Zhuang, M. and Chen, R. F. (1998) Optimized upwind dispersion-relation-preserving finite difference scheme for computational aeroacoustics. *AIAA journal*, **36(11)**, 2146-2148.
[18] Yamaleev, N. K. and Carpenter, M. H. (2009) A systematic methodology for constructing high-order energy stable WENO schemes. *Journal of Computational Physics*, ***228*(11)**, 4248-4272.
[19] Hu, X. Y., Wang, Q. and Adams, N. A. (2010) An adaptive central-upwind weighted essentially non-oscillatory scheme, *Journal of Computational Physics* **229(23)**, 8952-8965.
[20] Hu, X. Y. and Adams, N. A. (2011) Scale separation for implicit large eddy simulation. *Journal of Computational Physics*, **230(19)**, 7240-7249.
[21] Deng, X. and Zhang, H. (2000) Developing high-order weighted compact nonlinear schemes. *Journal of Computational Physics*, **165(1)**, 22-44.
[22] Nonomura, T., Iizuka, N. and Fujii, K. (2007) Increasing order of accuracy of weighted compact nonlinear scheme, *AIAA paper 2007-893*.
[23] Zhang, S., Jiang, S. and Shu, C. W. (2008) Development of nonlinear weighted compact schemes with increasingly higher order accuracy *Journal of Computational Physics*, **227(15)**, 7294-7321.
[24] Lele, S.K. (1992) Compact finite difference schemes with spectral-like resolution, *Journal of computational physics*, **103(1)**, 16-42.
[25] Shu, C.W. (2009) High order weighted essentially nonoscillatory schemes for convection dominated problems, *SIAM review*, **51(1)**, 82-126.
[26] Nonomura, T. and Fujii, K. (2013) Robust explicit formulation of weighted compact nonlinear scheme, *Computer & Fluids* **85**, 8-18.
[27] Nonomura, T., Iizuka, N. and Fujii, K. (2010) Freestream and vortex preservation properties of high-order WENO and WCNS on curvilinear grids, *Computers & Fluids*, ***39*(2)**, 197-214.
[28] Nonomura, T. and Fujii, K. (2009) Effects of difference scheme type in high-order weighted compact nonlinear schemes, *Journal of Computational Physics*, ***228*(10)**, 3533-3539.
[29] Sumi, T. and Kurotaki, T. (2015) A new central compact finite difference formula for improving robustness in weighted compact nonlinear schemes, *Computers & Fluids*, **123**, 162-182.
[30] Shi, J., Zhang, Y.T. and Shu, C.W. (2003) Resolution of high order WENO schemes for complicated flow structures, *Journal of Computational Physics*, **186(2)**, 690-696.
[31] Zhang, Y.T., Shi, J., Shu, C.W. and Zhou, Y. (2003) Numerical viscosity and resolution of high-order weighted essentially nonoscillatory schemes for compressible flows with high Reynolds numbers, *Physical Review E*, **68(4)**, 046709.
[32] Berland, J., Bogey, C. and Bailly, C. (2008) A study of differentiation errors in large-eddy simulations based on the EDQNM theory, *Journal of Computational Physics*, **227(18)**, 8314-8340.
[33] Shu, C.W. and Osher, S. (1988) Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, **77(2)**, 439-471.
[34] Sod, G.A. (1978) A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws, *Journal of computational physics*, **27(1)**, 1-31.
[35] Lax, P.D. (1954) Weak solutions of nonlinear hyperbolic equations and their numerical computation, *Communications on pure and applied mathematics*, **7(1)**, 159-193.
[36] Einfeldt, B., Munz, C.D., Roe, P.L. and Sjögreen, B. (1991) On Godunov-type methods near low densities, *Journal of computational physics*, **92(2)**, 273-295.
[37] Toro, E. F. (2009) *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*, 3rd edn, Springer, Berlin, Germany.
[38] Shu, C.W. and Osher, S. (1989) Efficient implementation of essentially non-oscillatory shock-capturing schemes, II, *Journal of Computational Physics*, **83(1)**, 32-78.

Interval-based analysis and word-length optimization of non-linear systems with control-flow structures

***J.A. López, E. Sedano, C. Carreras, and C. López**

Dpto. Ingeniería Electrónica, Universidad Politécnica de Madrid. 28040 Madrid, Spain.

*Corresponding author: juanant@die.upm.es

Abstract

The techniques based on extensions of interval computations allow fast and accurate analysis of the behavior of complex systems. Some of the most recent works in this area have presented procedures to evaluate systems with smooth non-linearities. We take this approach a step further by introducing a methodology that combines Multi-Element Generalized Polynomial Chaos (ME-gPC) and Statistical Modified Affine Arithmetic (MAA). This methodology allows modeling systems with highly non-linear operators and/or control-flow structures. It has been implemented in our modular and automated analysis framework, HOPLITE, so that it can be used to estimate the dynamic range, quantization noise and sensitivity of systems containing the aforementioned control-flow blocks. With this approach we have obtained in case studies with non-linear operators a deviation of only 0.04% with respect to the simulation-based reference values, which proves the accuracy of our approach.

Keywords: Interval Computation, Polynomial Chaos, Affine Arithmetic, Digital Signal Processing, Fixed-Point, Quantization, FPGA Implementation.

Introduction

In an industry where time-to-market is critical, the design and implementation of efficient and reliable Digital Signal Processing (DSP) systems can make the difference between success and failure. In addition, fixed-point computations are preferred when such systems are implemented on FPGAs and ASICs due to the lower implementation cost and power consumption, and higher performance with respect to its floating-point alternative. However, finding a fast and general way for transforming floating-point system descriptions to efficient fixed-point implementations remains an open issue. The analysis and selection of optimized word-lengths is an important and time-consuming step in the design of DSP and VLSI systems. Studies indicate that fixed-point refinement can take up to 25% to 50% of the overall development time [1]. Thus, automating and accelerating this process is strongly desirable.

During the past decades there has been a lot of work on the analytical characterization of the different structures of the DSP subsystems using mathematical expressions [1-20]. These studies provide guidelines to optimize these blocks, but they fail to provide results for the newly-developed (typically complex) structures, as well as for the complete (large) systems. To try to overcome this issue, a number of proposals has recently appeared. They are aimed at developing fast and accurate computation models aimed at providing the optimized word-lengths for the specific system that will be implemented.

Figure 1 outlines the main parts of this Word-Length Optimization (WLO) process [16]. Three major areas are easily identified: (i) Determining the dynamic range of the signals of the system, in order to allocate the integer word-length of each variable; (ii) assigning the

number of bits of the fractional word-lengths; and (iii) obtaining the statistical deviation (quantization noise) and determining the validity of the results. None of these three areas is trivial, and each of them is a large field of research on its own.

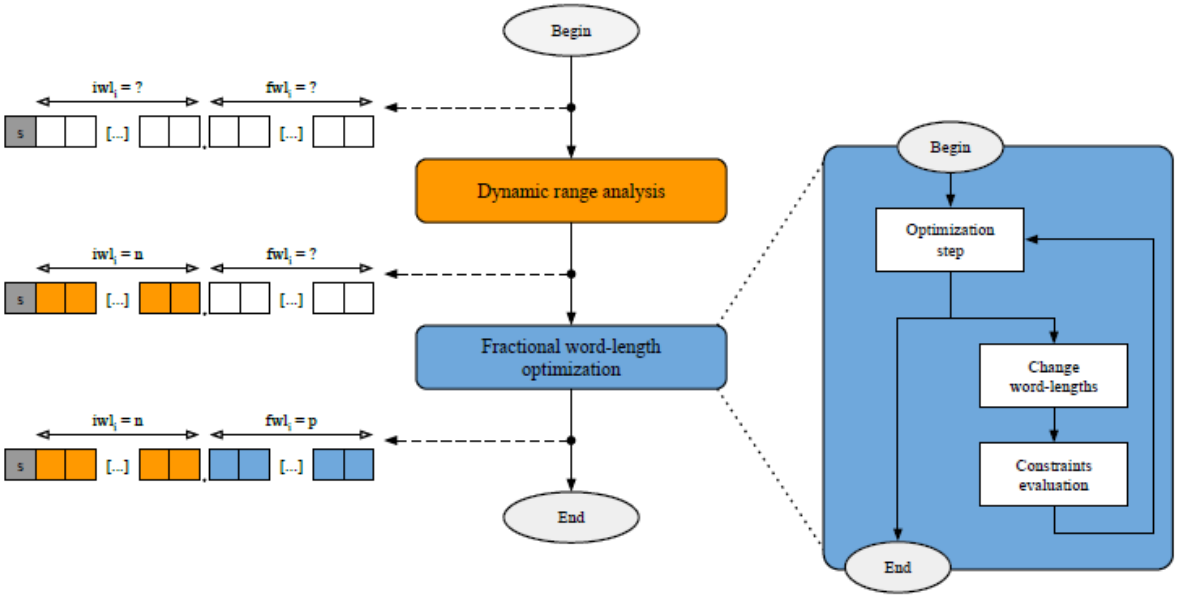


Figure 1. Fixed-point word-length optimization flow

In practice, the WLO process is commonly split in two parts: First, a computational accuracy constraint is determined according to the application performance, and then a WLO technique is applied using this constraint. Such modern WLO techniques are classified in two groups: simulation-based approaches, and analytical (or hybrid) ones.

Simulation-based techniques [2, 3] for modeling the quantization are the most reliable and general approaches, but also the slowest. In order to obtain accurate models, large input data sets are usually required. This makes simulation-based methods impractical for WLO, since estimations must be repeated many times with different combinations of word-lengths as the optimization progresses.

Modern analytical or hybrid techniques are several orders of magnitude faster than the simulation-based ones, but they are limited a given type of systems [4-7]. They perform separate analysis of the word-lengths required for the integer part (to represent the dynamic range of the signals) and the fractional one (to comply with the specified round-off constraint). The integer word-lengths are determined using range propagation or interval arithmetic. The fractional word-lengths are determined using a number of techniques, such as the Perturbation Theory [4], System Transformations [7], Arithmetic Transformations [17], and Handelman Representations [18]. Different Extensions of Interval Computations based on Affine Arithmetic (AA) [5, 6] have also provided very fast and accurate results, but they must be applied according to the characteristics of the system to be evaluated (linear, quasi-linear, polynomial, or strongly non-linear) [16].

The structure of the full version of the paper will be as follows: The models based on extensions of AA used to evaluate the different types of systems will be explained in separate subsections of Section 2. It will also be shown that the non-linear computations need the

application of Polynomial Chaos techniques to provide accurate results. Section 3 will explain some of the main applications that can be performed using our AA-based analysis, such as the sensitivity-driven optimization. The tool used for the propagation and computation of the results will be briefly desbribed in Section 4. Two of its main features will be highlighted: its modular implementation and the gradual computation of the results, since they are of particular importance for High Performance Computing (HPC) and the analysis of big data applications. Finally, Section 5 will provide the conclusions and summarizes this work.

Theoretical background on Extensions of Interval Computations

The evaluation of the quantization techniques using Extensions of Interval Computations has been rapidly progressing during the past years, and different new methodologies have been suggested to improve the quality and accuracy of the solutions, as well as to broaden the scope of the systems that can be addressed using them.

The first of such extensions is Affine Arithmetic (AA). AA has been originally suggested for the evaluation and characterization of the linear systems, and has shown to provide among the fastest computation times [10, 11]. However, AA is not able to capture of the correlations of the nonlinear operations. To overcome this fact, Modified Affine Arithmetic (MAA) has been proposed instead [5, 11, 19]. MAA contains higher-order terms that keep track of the results of the non-linear operations. However, these higher-order terms are not orthonormal, so the propagation of the affine terms provides misleading results.

A key feature for the accurate propagation of the higher-order terms is the incorporation of the Polynomial Chaos Expansions (PCE) techniques. The intervals of AA are included in the computation as parameters of the orthonormal polynomials of PCE, thus allowing easy propagation of the coefficients through the nonlinear system [16, 20]. This approach has been applied to dynamic range estimation [20], and to the analysis of the quantization noise for small, sequential systems [16]. However, PCE still fails to efficiently handle systems with discontinuities, and is not capable of modeling control-flow operations. Multi-Element generalized Polynomial Chaos (ME-gPC) is able to produce accurate models for discontinuous systems [16], as will be explained below. In this Section the mathematical background for AA, MAA, PCE and ME-gPC is given.

Affine Arithmetic (AA)

An affine form is defined as a polynomial expansion of order one where the independent variables are uniformly distributed in the interval [−1, 1]. Affine arithmetic is capable of capturing the correlation between intervals after affine operations (i.e. linear). A first-order affine form is expressed as [6]:

$$\hat{a} = a_0 + \sum_{i=1}^{n_a} a_i \varepsilon_i \tag{1}$$

The mean value is given by a_0, the terms ε_i are the independent sources of uncertainty and the coefficients a_i are the amplitudes of these uncertainties. The uncertainty sources can represent the variations of the signal or the RON. The basic operations between two affine forms $\hat{a}$ y $\hat{b}$ are summarized in Table 1 [5, 6]. The instructions supported by this methodology are either linear [5, 6] or smooth non-linear [9], meaning that their behavior can be approximated by linear models. The terms n_{max} refer to the maximum number of noise terms present in the affine forms.

Table 1. Coefficient propagation rules of Affine Arithmetic

Operation	Coefficient propagation rule
Addition	$\hat{a}+\hat{b}=(a_0+b_0)+\sum_{i=1}^{max(n_a,n_b)}(a_i+b_i)\cdot\epsilon_i$
Subtraction	$\hat{a}-\hat{b}=(a_0-b_0)+\sum_{i=1}^{max(n_a,n_b)}(a_i-b_i)\cdot\epsilon_i$
Constant multiplication	$c\cdot\hat{a}=c\cdot a_0+\sum_{i=1}^{n_a}c\cdot a_i\cdot\epsilon_i$
Multiplication	$\hat{a}\cdot\hat{b}=(a_0\cdot b_0)+\sum_{i=1}^{max(n_a,n_b)}(a_0\cdot b_i+a_i\cdot b_0)\cdot\epsilon_i+(\sum_{i=1}^{n_a}\lvert a_i\rvert\sum_{i=1}^{n_b}\lvert b_i\rvert)\epsilon_{n_{max}+1}$
Rounding	$Q_f^R(a)=(a_0-2^{-f-1})+\sum_{i=1}^{n_a}a_i\cdot\epsilon_i+2^{-f-1}\cdot\epsilon_{n_{max}+1}$
Truncation	$Q_f^T(a)=a_0+\sum_{i=1}^{n_a}a_i\cdot\epsilon_i+2^{-f-1}\cdot\epsilon_{n_{max}+1}$

Linear operations (addition, subtraction and constant multiplication) are executed in a precise manner. However, after performing the nonlinear operations the temporal correlations of the input signals are lost [5]. The result of executing non-linear operations over uniform distributions is typically non-uniform so it is theoretically impossible to represent it as a linear combination of uniform distributions. In order to alleviate this shortage, MAA [21] introduces higher order polynomials to capture the correlations among the signals.

Modified Affine Arithmetic (MAA)

MAA was initially used for polynomial evaluation and algebraic curve plotting in 2D [21]. Given two affine forms:

$$\hat{a}=a_0+a_1\varepsilon_a,\ \hat{b}=b_0+b_1\varepsilon_b \tag{2}$$

ε_a and ε_b are the noise terms bounded in the interval [−1, 1], a_0 and b_0 are the means of both variables and a_1 and b_1 represent the variations of the signals over the mean values. The simplest nonlinear operation is a multiplication of both affine forms:

$$f(\hat{a},\hat{b})=\hat{a}\cdot\hat{b}=a_0b_0+a_0b_1\varepsilon_b+a_1b_0\varepsilon_a+a_1b_1\varepsilon_a\varepsilon_b \tag{3}$$

Generalizing for any order, the centered form of the output polynomial is given by:

$$f(\hat{a},\hat{b})=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}a_ib_j\varepsilon_a^i\varepsilon_b^j \tag{4}$$

It can be seen that this solution is an extension of AA in which all the high-order terms are taken into account [5]. In [21], this technique is just applied in the case of multiplications and other non-linear operations are obviated. Nevertheless, since the monomials of MAA are not orthonormal, the incorporation of the PCE techniques that take into account higher order terms when considering different types of operations is also required. Without them, the propagation of the gains throughout the system under analysis would not be accurately performed.

Polynomial Chaos Expansions (PCE)

Given a set of independent random variables of dimension *N*, $\Phi = \{\phi_1, \phi_1, \ldots, \phi_N\}$, and another random variable *Y*, square integrable, such that $Y = f(\Phi)$, then *Y* can be expressed as a weighted sum of polynomials as

$$Y = \sum_{i}^{\infty} \alpha_i \psi_i(\Phi) \tag{5}$$

where each α_i is a constant coefficient and each ψ_i is the *i*-th polynomial from an orthogonal basis [20]. The terms α_i are the spectral coefficients of the expansion, and the terms $\psi_i(\Phi)$ are the orthonormal polynomial basis, which satisfy the condition

$$< \psi_i, \psi_j > = \begin{cases} \psi_i^2 & if\ i = j \\ 0 & if\ i \neq j \end{cases} \tag{6}$$

In practice, the number of terms of the PCE is truncated to a finite number. It depends on the dimension of the expansion *n* (number of independent variables in vector Φ) and the maximum order of the polynomials used, *p*. The selection of the basis depends on the probability density functions (gaussian, uniform, gamma, beta, etc.) of the RVs present in the system. In particular, for the analysis of a given system with gaussian random variables, Hermite basis polynomials provide the most accurate results [22].

The coefficients of the expansion α_i in Eq. (5) are computed by applying a Galerkin projection operation [16, 19], and solved by applying Monte-Carlo techniques with a small number of samples.

Once the random input signals have been defined, and expressed as a function of the $\psi_i(\Phi)$ basis polynomials, the next step is to propagate the coefficients through the data flow graph. This procedure is exploits the orthogonality properties of the polynomials. The basic operations are performed as follows.

Consider two input RVs $\hat{x}$ and $\hat{y}$ expanded in a PCE,

$$\hat{x} = \sum_{i=1}^{m} x_i \cdot \psi_i\ ,\ \ \hat{y} = \sum_{i=1}^{m} y_i \cdot \psi_i \tag{7}$$

The computation of the linear operations is straightforward, i.e.:

$$\hat{z} = a\hat{x} \pm b\hat{y} = \sum_{i=1}^{m} (ax_i \pm by_i) \cdot \psi_i \quad \Rightarrow \quad z_i = ax_i \pm by_i \tag{8}$$

The propagation through the non-linear operations such as the multiplication is not so direct. Considering that $\hat{z} = \hat{x} \cdot \hat{y}$ and substituting each variable by its correspondent PCE:

$$\hat{z} = \sum_{k=1}^{m} z_k \cdot \psi_k = \sum_{i=1}^{m} x_i \cdot \psi_i \sum_{j=1}^{m} y_j \cdot \psi_j \tag{9}$$

The coefficients z_k are calculated by performing a Galerkin projection [19]:

$$z_k = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{<\psi_i \psi_j \psi_k>}{\psi_k^2} x_i y_j = \sum_{i=1}^{m}\sum_{j=1}^{m} C(i,j,k)\, x_i y_j, \tag{10}$$

which constitutes a linear system of *m* equations. It can be expressed in matrix form as:

$$Z = A \cdot X, \text{ with } A = C \cdot Y \tag{11}$$

where *A* is an *m*×*m* matrix and X, Y and Z are the column vectors that correspond to the $\hat{x}$, $\hat{y}$ and $\hat{z}$ coefficients, respectively. Tensor *C*(*i*, *j*, *k*) is the same for a given dimension and order, so it only has to be calculated once (for instance in a pre-processing stage), and afterwards reused when needed, thus notably reducing the required computation time [16]. In addition, a number of techniques for accelerating the computation of the *C* matrix can be applied, speeding the overall process even further. The interested reader may find detailed examples of the propagation of affine forms using combined PCE + MAA in [16, 19].

Multi-Element generalized Polynomial Chaos (ME-gPC)

In many cases, PCE requires an excessively large basis to accurately represent the set of values. This happens particularly in the presence of discontinuities, or when many non-linear operations appear following each other. To overcome this, ME-gPC is formulated [WK05]. This technique partitions the input domain in smaller sub-domains, decomposing the complex functions into a set of simpler ones. This enables the efficient use of lower PCE orders to model the sub-domains, while still providing very accurate results [16].

Being $B = [-1, 1]^n$ the domain in which $\Xi = [\xi_1, \xi_2, \ldots, \xi_n]$ is defined, the ME-gPC method proposes its decomposition in a regular set of non-overlapping elements. Each element will be now contained in the domain $B_k = [a_1^k, b_1^k) \times [a_2^k, b_2^k) \times \ldots \times [a_n^k, b_n^k]$, where a_i and b_i are respectively the upper and lower bounds of the *i*-th local random variable.

From this decomposition of the global domain, a local random vector for each element is now defined as $\zeta^k = [\zeta_1^k, \zeta_2^k, \ldots, \zeta_n^k]$. Next, in order to take advantage of the properties of the Legendre Chaos, each ζ^k is re-scaled into a new random vector $\xi^k = [\xi_1^k, \xi_2^k, \ldots, \xi_n^k]$. This vector is equivalent to ζ^k but in the domain $[-1, 1]^n$, instead of B_k.

Once a dimension has been partitioned, the new PCE expansions for each sub-domain are generated. Each of these expansions has the form

$$\tilde{u}(\tilde{\xi}) = \sum_{i=1}^{m} \tilde{u}_i \Phi_i(\tilde{\xi}) \tag{12}$$

where $\tilde{\xi}$ is defined $[-1, 1]^d$. To calculate the coefficients $\tilde{u}_i$ of each new expansion, a linear system of equations is solved. This system is generated by choosing *m*+1 uniform grid points $\tilde{\xi}_i$ in $[-1, 1]^d$.

With the expansions $\tilde{u}(\tilde{\xi})$ obtained with this method, PCE can be locally applied to the different elements. Once the expansions have been computed, the statistical global moments can be reconstructed applying Bayes' theorem and the law of total probability.

Figure 2 shows an example of the domain decomposition using ME-gPC for the conditional inequality $x^2 \geq y$.

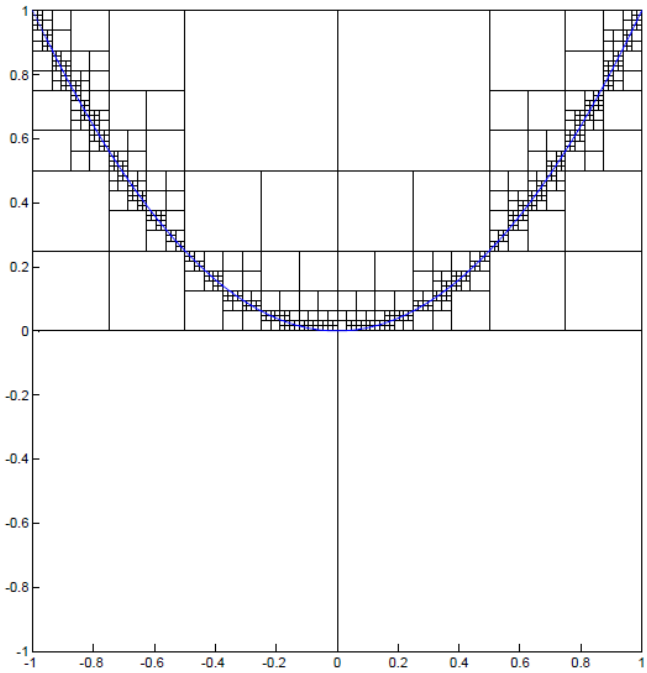


Figure 2. Example of domain decomposition using ME-gPC.

So far MEgPC has only been used to estimate the dynamic range in systems without control-flow structures, and it has been only applied to numerical procedures. In the following Sections we will combine MEgPC and MAA to estimate the sensitivity and the quantization noise in fixed-point digital systems with control-flow structures, extending the initial analysis carried out for linear systems in [6] to non-linear operations and control structures in the Data Flow Graph.

This largely broadens the applicability of the probabilistic interval analysis in word-length optimization, as it allows for an entire new class of systems to be targeted for modelling and optimization [16].

DSP Applications of the Extensions of Interval Computations

Some of the main applications of the Extensions of Interval Computations will be explained here, in different subsections, such as Dynamic Range Estimation, Quantization Noise, and Sensitivity Analysis in the different types of structures, including systems with discontinuities and control-flow structures.

The HOPLITE framework

In this Section a modular automated word-length optimization tool, HOPLITE, is introduced. One of its main objectives is to provide designers flexibility to perform modelling and search policies that best suit their objectives [16]. In the different subsections a general overview of the HOPLITE work flow will be provided, some of the implementation decisions, modules and interfaces of the framework, and a detailed execution example.

Table 2 provides a preliminary analysis of the languajes evaluated for its implementation, and Figure 3 shows a general overview of the functions included in the HOPLITE framework.

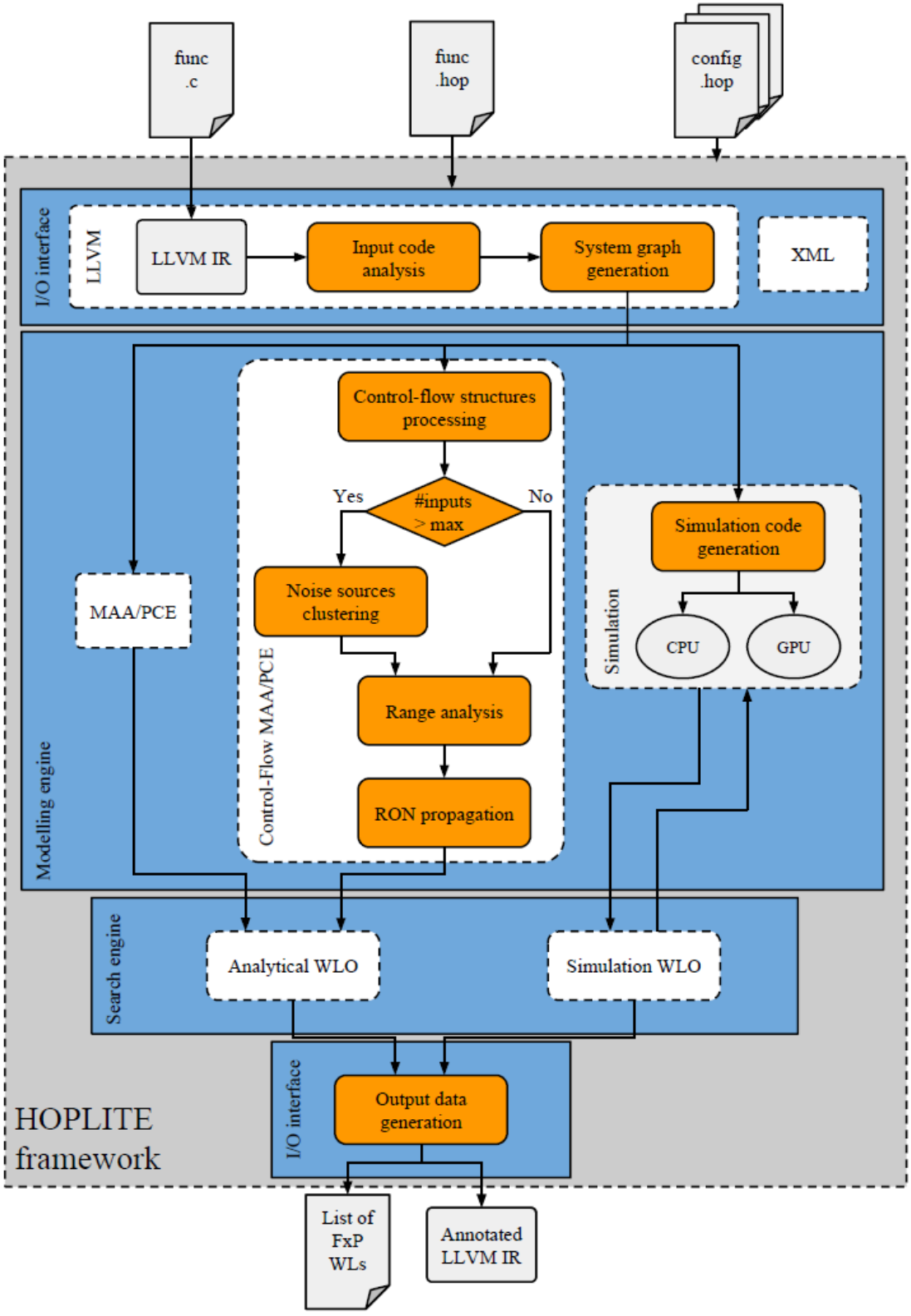


Figure 3. The HOPLITE framework work flow

Table 2. Language selection: requisites and availability

Requisite	MATLAB	Octave	C/C++	Python
Simple	No	No	Some	Yes
Algebra & symbolic systems	Yes	Some	Yes	Yes
Memory management	Some	Some	No	Yes
Support for external tools	Yes	Yes	Yes	Yes
Free	No	Yes	Yes	Yes

References

[1] M. Clark, M. Mulligan, D. Jackson, D. Linebarger. Accelerating fixed-point design for mb-ofdm uwb systems. *Comms Design*, EE times (online), 2005.

[2] R. Cmar, L. Rijnders, P. Schaumont, S. Vernalde, and I. Bolsens, "A Methodology and Design Environment for DSP ASIC Fixed Point Refinement," in *Proc. conf. Design, automation and test in Europe, DATE '99*, p. 56, 1999.

[3] K. I. Kum and W. Sung, "Combined Word-Length Optimization and High-Level Synthesis of Digital Signal Processing Systems," *IEEE Trans. Circuits Syst.*, vol. 20, pp. 921–930, Aug. 2001.

[4] G. A. Constantinides, P. Y. K. Cheung, and W. Luk, "Wordlength Optimization for Linear Digital Signal Processing," *IEEE Trans. Comput.- Aided Design Integr. Circuits Syst.*, vol. 22, n. 10, pp. 1432–1442, 2003.

[5] J. A. Lopez, C. Carreras, and O. Nieto-Taladriz, "Improved interval-based characterization of fixed-point lti systems with feedback loops," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 11, pp. 1923–1933, 2007.

[6] J. Lopez, G. Caffarena, C. Carreras, and O. Nieto-Taladriz, "Fast and accurate computation of the roundoff noise of linear time-invariant systems," *IET Circuits, Devices and Systems*, vol. 2, no. 4, pp. 393–408, 2008.

[7] D. Menard, R. Rocher, and O. Sentieys, "Analytical Fixed-Point Accuracy Evaluation in Linear Time-Invariant Systems," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 55, November 2008.

[8] R. Rocher, D. Menard, N. Herve, and O. Sentieys, "Fixed-Point Configurable Hardware Components," *EURASIP Journal on Embedded Systems*, vol. 2006, pp. Article ID 23197, 13 pages, 2006. doi:10.1155/ES/2006/23197.

[9] G. Caffarena, J. Lopez, G. Leyva, Carreras, and O. Nieto-Taladriz, "Architectural Synthesis of Fixed-Point DSP Datapaths Using FPGAs," *Int. J. of Reconfigurable Computing*, vol. 2009, pp. 1–14, 2009.

[10] C. Fang, R. Rutenbar, and T. Chen, "Fast, accurate static analysis for fixed-point finite-precision effects in dsp designs," in *Int. Conf. on Computer-Aided Design, 2003 (ICCAD '03)*. pp. 275–282, 2003.

[11] J. Lopez, Evaluacion de los Efectos de Cuantificacion en las Estructuras de Filtros Digitales Utilizando Tecnicas de Cuantificacion Basadas en Extensiones de Intervalos. PhD thesis, Univ. Politecnica de Madrid, Madrid, 2004.

[12] C. Shi and R. Brodersen, "Floating-point to fixed-point conversion with decision errors due to quantization," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, 2004.

[13] S. Roy and P. Banerjee, "An algorithm for trading off quantization error with hardware resources for matlab-based fpga design," *IEEE Trans.* Computers, vol. 54, no. 7, pp. 886–896, 2005.

[14] D.-U. Lee, A. Gaffar, R. Cheung, W. Mencer, O. Luk, and G. Constantinides, "Accuracy-Guaranteed Bit-Width Optimization," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 10, pp. 1990–2000, 2006

[15] L. Zhang, Y. Zhang, and W. Zhou, "Floating-point to fixed-point transformation using extreme value theory," in *Eighth IEEE/ACIS Int. Conf. Computer and Information Science, 2009 (ICIS 2009).* pp. 271–276, 2009.

[16] E. Sedano, *Automated word-length optimization framework for multi-source statistical interval-based analysis of non-linear systems with control-flow structures*. PhD thesis, Univ. Politecnica de Madrid, Madrid, 2016.

[17] Y. Pang, K. Radecka, Z. Zilic. Optimization of imprecise circuits represented by taylor series and real-valued polynomials. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 29(8):1177-1190, 2010.

[18] D. Boland, G.A. Constantinides. A scalable precision analysis framework. *IEEE Trans. Multimedia*, 15(2), 242-256, 2013.

[19] L. Esteban. *High precision FPGA based phase meters for infrared interferometers fusion diagnostics.* PhD thesis, Universidad Politecnica de Madrid, 2011.
[20] B. Wu, J. Zhu, and F.N. Najm. Dynamic-range estimation. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* 25(9), 1618-1636, 2006.
[21] H. Shou, H. Lin, R. Martin, G. Wang. Modified Affine Arithmetic Is More Accurate than Centered Interval Arithmetic or Affine Arithmetic. *Mathematics of Surfaces,* Lecture Notes in Computer Science, Springer. vol. 2768 pages 355-365. 2003.
[22] D. Xi,u G.E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of computational physics,* 187(1), 137-167, 2003.
[23] X. Wan, G.E. Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics,* 209(2), 617-642, November 2005.

Axial Green's function Methods on Free Grids

Junhong Jo[1], Hong-Kyu Kim[2] and Do Wan Kim[3*†]

[1]Inha University, Department of Mathematics, Incheon, Republic of Korea
[2]Korea Electrotechnology Research Institute, Changwon, Republic of Korea
[3]Inha University, Department of Mathematics, Incheon, Republic of Korea

*Presenting author: dokim@inha.ac.kr
†Corresponding author: dokim@inha.ac.kr

Abstract

We are going to talk about axial Green's function methods (AGMs) on free grids called axial lines. These are novel approaches in numerical computations. AGMs that we have developed for elliptic boundary value problems [3] and the steady Stokes flows [2] in complicated geometry use axial lines for discretization. These axial lines are parallel to axes and there is no restriction on their distribution. The salient feature of the methods is that not only one-dimensional Green's function for the axially split differential operators is sufficient to solve the multi-dimensional problems but also the free grids are available. In this talk, short introduction to AGMs is presented and then we show that the localization [1] of axial lines enables us to enforce Neumann boundary condition, and refinement of axial lines on separated regions are readily available as well.

Keywords: Axial Green's function, Free grids, Axial lines, One-dimensional Green's function, Multi-dimensional problem, Boundary condition, Refinement.

Introduction

By the axial Green's function, we mean that it is one-dimensional Green's function of an ordinary differential operator defined on lines parallel to axis, belonging to the multi-dimensional domain. In general, the finite difference method uses this kind of lines, called the grids, but the admissible grids in this method are so restrictive that the method cannot work unless the domain is simple or the grids are gradually changing in space. The axial Green's function methods(AGM) we have developed work fine in arbitrary domains without deterioration of accuracy, and furthermore they do even in randomly spacing axial lines.

The use of Green's function take place in the boundary element(BEM) method, which can reduce the dimension of the problem by discretizing the boundary of the domain. This is possible only when finding the fundamental solution or Green's function of the multi-dimensional differential operator, called partial differential operator. The BEM has been successful in Laplace operator, Lame operator in linear elasticity, Stokes operator in fluid mechanics, Helmholtz operator, and so on. However, if the material coefficients are functions of space variable, then the BEM suffers from finding the multi-dimensional Green's function in the domain or even a fundamental solution in entire space.

The advantages of AGMs are obvious in two points: (1) Arbitrarily distributed axial lines are available, which is inconvenient in FDMs, and (2) It is much easier than BEMs to find one-dimensional Green's functions. Based on these facts, we are able to implant these advantages to the refinements of axial lines in some regions of interest. The refined regions can be independently handled by using the representation formula for the solution in terms of axial Green's functions.

Axial Green's function method

For the sake of simplicity, we consider the Poisson problem in 2-dimensional domain Ω as an example:

$$-\Delta u = f, \quad \text{in } \Omega, \tag{1}$$
$$u = u^{\partial\Omega}, \quad \text{on } \partial\Omega. \tag{2}$$

Our interest is laid on the point that this multi-dimensional problem can be reformulated by one-dimensional problems. First of all, decomposing the multi-dimensional operator $-\Delta$ as two parts by introducing a new variable $\phi(x, y)$ as follows:

$$-u_{xx} = \phi, \quad \text{in } \Omega, \tag{3}$$
$$-u_{yy} = f - \phi, \quad \text{in } \Omega, . \tag{4}$$

From the first equation in (3), we find one-dimensional Green's functions to represent the

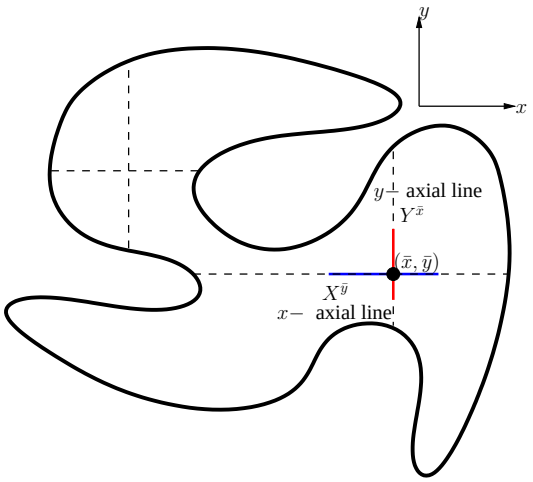


Figure 1: Axial lines for AGMs

solution $u(x, y)$ on $x-$axial line $X^{\bar{y}}$ and $y-$axial line $Y^{\bar{x}}$ associated with a given cross point $(\bar{x}, \bar{y}) \in \Omega$ as shown in Fig. 1:

$$u(\xi, \bar{x}) = \int_{X^{\bar{y}}} G(x, \xi; X^{\bar{y}})\phi(x, \bar{y})\, dx + u(x_-, \bar{y})B_-^X(\xi) + u(x_+, \bar{y})B_+^X(\xi), \ (\xi, \bar{y}) \in X^{\bar{y}}, \tag{5}$$
$$u(\bar{x}, \eta) = \int_{Y^{\bar{x}}} G(y, \eta; Y^{\bar{x}})(f - \phi)(\bar{x}, y)\, dy + u(\bar{x}, y_-)B_-^Y(\eta) + u(\bar{x}, y_+)B_+^Y(\eta), \ (\bar{x}, \eta) \in Y^{\bar{x}}. \tag{6}$$

In this case, these representations can be unified in the following form:

$$u(\tau) = \int_{t_-}^{t_+} G(t, \tau)g(t)\, dt + u(t_-)B_-(\tau) + u(t_+)B_+(\tau), \quad \tau \in (t_-, t_+), \tag{7}$$

where $G(t, \tau)$ is the corresponding one-dimensional Green's function and $B^{\pm}(\tau)$ is the function related to the boundary values $u(t_{\pm})$. Instead of directly attacking the multi-dimensional problem in (1) with boundary condition (2), we pay attention to the equations of integral form

in (3) and (4). That is, after discretizing these integral equations for the unknown ϕ and u, we can solve the resultant system of equations well using GMRES.

In an analogous way, AGM can be applied to more general problems, for instance, general elliptic problem with function coefficient, the Stokes flow, and the convection-diffusion problem with variable coefficients, etc. For more effective computations, we need grid refinements in

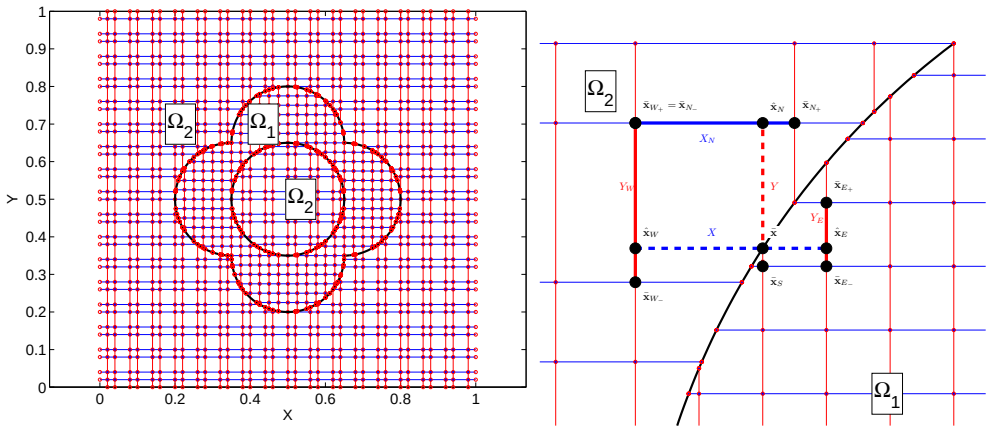


Figure 2: (left) Split domains Ω_1 and Ω_2 and (right) Interfacial configutration between split domains for refinement

different subdomains of interest. Let us consider the following problem:

$$-\nabla \cdot (\epsilon \nabla u) + \mathbf{U} \cdot \nabla \mathbf{u} = f, \quad \text{in } \Omega, \tag{8}$$

$$u = u^{\partial\Omega}, \quad \text{on } \partial\Omega. \tag{9}$$

In this case, of course, we can find the axial Green's function associated with the convection operator in (8) and thus AGM can be applied on it. If we split domain Ω into Ω_1 and Ω_2 as in the left panel in Fig. 2, then the axial lines can be distributed as the right panel in Fig. 2 near the interface between two domains. Since we have the best approximations (5) and (6) of the solution, these equations enable us to merge the AGM solutions on both domains, Ω_1 and Ω_2. It is in fact an obvious advantage that there is no need for the conformity to axial lines across the interface. Assume $\mathbf{U} = (2, 3)$ and the exact solution $u(x, y)$ satisfying (8) and (9) in

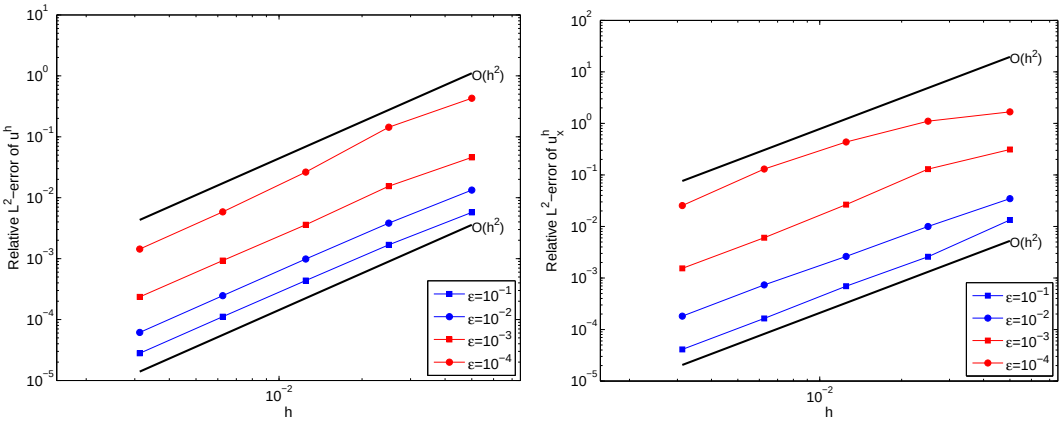


Figure 3: (left) $O(h^2)$-convergence for u^h in L^2-sense and (right) $O(h^2)$-convergence for the derivative of u^h in L^2-sense

$\Omega = [0,1] \times [0,1]$ as follows:

$$u(x,y) = 16x(10x)y(1-y)\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(2/\sqrt{\epsilon}(0.25^2 - (x-0.5)^2 - (y-0.5)^2)\right)$$

which has interior steep layer. On the axial lines in Fig. 2, we obtain the second order convergence for the numerical solution u^h and its $x-$derivative u^h_x which are illustrated in Fig. 3.

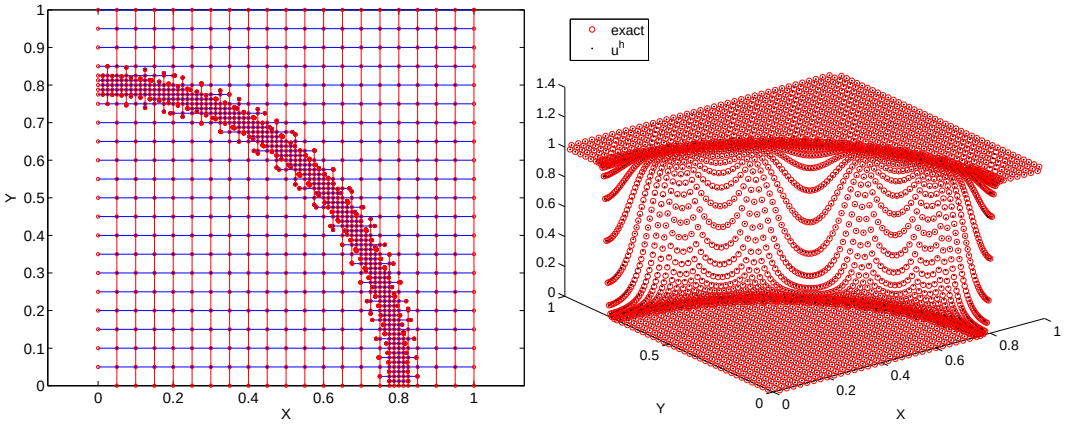


Figure 4: (left) Refined axial lines and (right) the corresponding numerical solution

In Fig. (4), two types of axial lines are drawn in the left panel, one is refinement near the steep interior layer of the solution and the other is coarse in a slowly varying region. Both the exact solution u and the computed solution u^h are depicted in the right panel of Fig. 4.

Conclusions

We present the axial Green' function method called the AGM which has two marked features. Firstly, arbitrarily distributed axial lines are available for the numerical computation without any degradation of accuracy. Second, the axial Green's function can be found easily compared to the multi-dimensional one. Using these features, we can devise an adaptive refinement of axial lines for the purpose of the effective computations. We expect that it can be applied to 3-dimensional problems as well.

References

[1] Lee, W. and Kim, D.W. (2014) Localized Axial Green's Function Method for the Convection-Diffusion Equations in Arbitrary Domains. *Journal of Computational Physics* **275**, 390-414

[2] Jun, S. and Kim, D.W. (2011) Axial Green's Function Method for Steady Stokes Flow in Geometrically Complex Domains. *Journal of Computational Physics* **230**, 2095-2124

[3] Kim, D. W., Park, S.-K., and Jun, S. (2008) Axial Green's function method for multi-dimensional elliptic boundary value problems. *International Journal for Numerical Methods in Engineering* **76**, 697-726

Perspective into Model-based Genetic Programming

†P. He[1, 2] and A. C. Hu[1]

[1]School of Computer Science and Educational Software, Guangzhou University, P. R. China
[2]School of Computer and Communication Engineering, Changsha University of Science and Technology, P. R. China

*Presenting author:bk_he@126.com
†Corresponding author: bk_he@126.com

Abstract

Representation is an open issue in GP (Genetic Programming) research area, having close relationships with its performance improvements. This paper introduces a novel GP framework called model-based grammatical evolution (MGE) as well as the principle it obeys. In MGE, individuals take the form of sequences of productions, therefore providing means for structural analysis and semantic reuses. To certify the effectiveness of MGE, comparisons with some other GP variants like classical grammatical evolution (CGE), integer representation GE are also conducted.

Keywords: Genetic Programming, Grammatical Evolution, Model, Finite Sate Automaton.

Introduction

Genetic Programming (GP) [1] as one of the most important automatically programming approaches constructs programs by means of evolution principle. It generates populations of chromosomes in terms of genetic algorithm (GA) [2], chooses at last the fittest individual from the final population for the desired solution. So, GP could be recognized as a GA variant, but it is much simpler than GA in delineating complex structures, therefore having been applied in a wide range of fields like mathematical modeling, circuit design, pattern recognition, and financial prediction, etc. [1][3]-[8].

Up to now, GP grows up into a big family comprising of a large number of variants such as classical GP, gene expression programming (GEP), multi-expression programming (MEP), grammatical evolution (GE), and so on [4]-[8]. However, while using them extensively, we should take notice of the following deficiencies.

- Many GPs like tree based GP and grammar based GP are difficult to use for the sake of their complex representations.
- Most of existing GPs are devised from the principle of software testing, providing few means dealing with semantics.
- Some GP variants like GEP are easy to use, but their expressiveness is very limited. For instance, GEP, as far as the expressiveness is concerned, can essentially be described by GE.

In view of these, we will provide a novel GP framework, which was called model based GE, for coping with the abovementioned problems. It borrows some ideas of model checking. We will introduce the principle abided by and a sample model-based GE in the following parts. Finally, to demonstrate the effectiveness of the present approach, comparisons with some other GP variants like IGE (Integer representation GE) [7], PIGE and CGE (Classical GE) [4][5][8] are conducted.

Modeling Principle

Model approach has long been regarded as a powerful solution to system representation, system analysis, and software development. GP as program generation tool can naturally benefit from using of model approach. By model approach, we mean [9]:

1. Delineating both the problem and property of concern, say M and φ, in the context of some description model;
2. Establishing that $M \models \varphi$ holds. When M is a transition system, our goal is to prove $M, S \models \varphi$ holds for some special state S in M.

Consequently, GP can be modeled as follows based on the above model checking strategy.

1. Constructing a finite state transition system M for the concerned GP;
2. Delineating what we are interested;
3. Designing an algorithm suitable to check the satisfaction of $M, S \models \varphi$.

Model-based GP

So far, we have obtained two model-based GP variants called HGP (Hoare Logic-based Genetic Programming) [10][11] and MGE (Model-based Grammatical Evolution) [12]-[14] in terms of the principle of part II. The unified method is summarized as the following steps. If having further interest, one can refer to [12] [13] for the details.

1. Constructing a transition diagram G= <V, E> with some vertex $v_0 \in V$ as the start symbol by steps 2 through 5. Here V and E are sets of vertices and edges, respectively.
2. Regarding the states of V either as sets of logic formulas or as sets of sentential forms;
3. Regarding e in E either as programs or as productions of some context-free grammar. In this case, both states and edges could be used to define Hoare triples or grammatical deviations.
4. Defining relations among states to be connected.
5. The formal framework obtained from steps 2 through 4 is suitable for either verifying and generating the desired programs or deriving programs grammatically;
6. Constructing genetic operations over the formal framework of step 5, we obtain either HGP or MGE [10]-[13].

For instance, the transition matrix given in table 1 is the model of languages of the grammar in Figure 1. This model covers all the leftmost derivations of the concerned grammar. According to the matrix, we can solve certain regression problems (see the following part) as shown in Figure 2.

(1) <expr>::= <expr><op><expr>	(11)	(3) <pre_op>::= sin	(31)
| (<expr><op><expr>)	(12)	|cos	(32)
| <pre_op> (<expr>)	(13)	| exp	(33)
| <var>	(14)	| log	(34)
(2) <op>::= +	(21)	(4) <var>::= y	(41)
|-	(22)	|1.0	(42)
| *	(23)		
| /	(24)		

Figure 1 grammar of expression

Table 1 Transition Matrix

S	1	2	3	4	5	6	7	8	9	10
ε	1	1,2	3,4	4	5,6	6	1,7	8,9	9	1,10
11	2									
12	2									
13	8									
14	3									
21						7				
22						7				
23						7				
24						7				
31									10	
32									10	
33									10	
34									10	
41				5						
42				5						

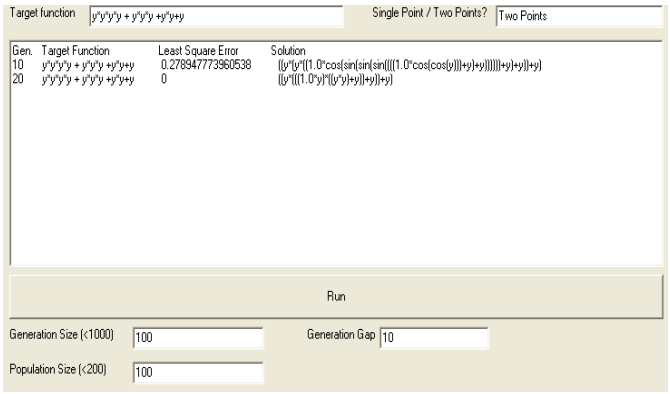


Figure 2 Screenshot of the `method` with population size=100, generation size=100.

Experiments

In this part, we will demonstrate the performance improvement of the present approach through comparisons of it with CGE[5], IGE [7] and PIGE in solving regression problems. The grammar used here is given in figure 1, and the objective is to find Eq. 1 based on 20 sample input values {-1, -0.9, -0.8, -0.76, -0.72, -0.68, -0.64, -0.4, -0.2, 0, 0.2, 0.4, 0.63, 0.72, 0.81, 0.90, 0.93, 0.96, 0.99, 1} in the range [-1…1].

$$f(y) = y^4 + y^3 + y^2 + y \qquad \text{(Eq. 1)}$$

The method is as follows: constructing the grammar model as shown in table 1; analyzing the structure of the model, and constructing building-block based GE; running the obtained GE over sample input values will result in figures 3 to 4 [12]. It follows from these figures that the present approach has advantages over the other GE variants in efficiency.

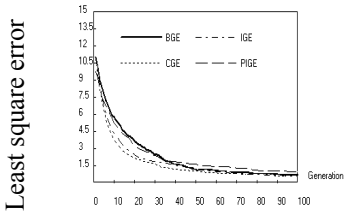


Figure 3 Average fitness of 100 runs of the four GEs in Eq. 1

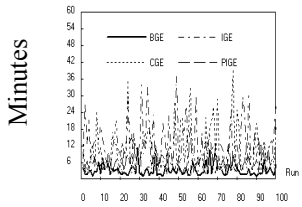


Figure 4 Time used of 100 individual runs of the four GEs in Eq. 1

Conclusions

This paper introduces the principle MGE abides by, and application method in solving real-world problems. Experiment demonstrates that MGE has advantage over classical GE, integer representation GE, and PIGE in performance improvement. Our future work will focus deeply on its semantic computing, and unifications with other GP variants.

Acknowledgements

The research work was supported in part by National Natural Science Foundation of China (Grant No. 61170199), the Scientific Research Fund of Education Department of Hunan Province, China (Grant No.11A004), and the Natural Science Foundation of Guangdong Province, China (Grant No. 2015A030313501)

References

[1] Koza J.R (1992). Genetic Programming. MIT Press, Cambridge M. A.
[2] Holland J.(1975) Adaptation in Natural and Artificial Systems, The University of Michigan Press: Michigan.
[3] Langdon W.B, Harman M. (2015). Optimizing Existing Software with Genetic Programming, IEEE Transactions on Evolutionary Computation, 19(1): 118-135
[4] Oltean M, Grosan C, Diosan L, Mihaila C (2009). Genetic Programming with Linear Representation: A Survey, International Journal on Artificial Intelligence Tools, 19(2): 197-239
[5] O'Neill M., Ryan C. (2001) Grammatical Evolution. IEEE Transactions on Evolutionary Computation., 5(4): 349-358.
[6] Ferreira C. (2001) Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex Systems, 13(2): 87-129
[7] Hugosson J, Hemberg E, Brabazon A, O'Neill M (2010). Genotype Representation in Grammatical Evolution. Applied Soft Computing. 10: 36-43
[8] Alfonseca M., Gil F.J.S. (2013). Evolving an ecology of Mathematical expressions with Grammatical Evolution, BioSystems, 111: 111-119
[9] Michael Huth, Mark Ryan.(2004) Logic in Computer Science: Modelling and Reasoning about System. Cambridge University Press: England.
[10] He P. Kang L.S., Fu M. (2008). Formality Based Genetic Programming, IEEE CEC.
[11] He P, Kang L.S, Johnson C. G. and Ying S (2011), Hoare logic-based genetic programming, Science China Information Sciences, 52, 623-637
[12] He P., Johnson C.G. Wang H.F (2011), Modeling Grammatical Evolution by Automaton, Science China Information Sciences, 52, 2544-2553
[13] He P. Deng Z.L. Wang H.F, Liu Z.S (2015), Model Approach to grammatical evolution: theory and case study, Soft Comput, 2015. DOI 10.1007/s00500-015-1710-9
[14] He P. Deng Z. L., Gao C.Z., Wang X. N., Li J. (2016). Model Approach to grammatical evolution: Deep-Structured Analyzing of Model and Representation, Soft Comput, 2016. DOI 10.1007/s00500-016-2130-1

Dynamic crack analysis of fiber reinforced piezoelectric composites by a Galerkin BEM

†*M. Wünsche[1], J. Sladek[1], V. Sladek[1] and Ch. Zhang[2]

[1]Institute of Construction and Architecture, Slovak Academy of Sciences, 84503 Bratislava, Slovakia
[2]School of Science and Technology, Chair of Structural Mechanics, University of Siegen, D-57068 Siegen, Germany

*Presenting author: wuensche@bauwesen.uni-siegen.de
†Corresponding author: wuensche@bauwesen.uni-siegen.de

Abstract

In this paper, transient dynamic analysis of micro-cracks of arbitrary shape in two-dimensional, linear piezoelectric fiber reinforced composite materials is presented. Interface cracks between fiber and matrix as well as cracks inside the matrix and fibers are analyzed. A symmetric Galerkin time-domain boundary element method in conjunction with a multi-domain technique is developed for this purpose. The time discretization is performed by a collocation method and time-domain fundamental solutions for piezoelectric materials are applied. An explicit time-stepping scheme is obtained to compute the discrete boundary data including the generalized crack-opening-displacements (CODs). Iterative solution algorithms are implemented to treat the non-linear semi-permeable electrical crack-face boundary conditions and for a crack-face contact analysis at time-steps when a physically unacceptable crack-face intersection occurs. Numerical examples are presented to reveal the effects of the micro-cracks, the material combinations and the dynamic loading on the intensity factors and the scattered wave fields.

Keywords: piezoelectric fiber composites, interface cracks, impact loading, complex intensity factors, time-domain BEM.

Introduction

Piezoelectric materials are widely applied in smart structures like transducers, actuators and sensors by utilizing the property of converting electrical energy into mechanical energy and vice versa. In recent years piezoelectric fiber reinforced materials have received increasing attention. A special class of such composites combines piezoelectric ceramics or polymers as active fibers with passive non-piezoelectric materials as matrix. Fiber reinforced materials can be optimized to satisfy the high performance requirements by taking advantages of the most beneficial properties of each constituent. Piezoelectric ceramics are very brittle with low fracture toughness and micro as well as macro cracks may be induced during the manufacturing and under the in-service condition. Beside cracks inside the homogeneous matrix and fibers, interface cracks play an important role for the design and safety of real structures. Since the electrical permittivity of the crack medium has a significant influence on the intensity factors the crack-face boundary conditions have to be described properly. Although the analysis of cracks in homogenous piezoelectric solids under static and dynamic loadings has been presented by many authors the corresponding analysis of interface cracks in piezoelectric fiber reinforced materials is rather limited due to the problem complexity. This paper presents such an analysis by using a hypersingular symmetric Galerkin boundary element method (SGBEM) for crack problems in two-dimensional (2D), fiber reinforced and linear piezoelectric solids.

Problem statement and numerical solution algorithm

We consider a piecewise homogeneous linear piezoelectric fiber-matrix structure with cracks of arbitrary shape. In the absence of body forces, free electric charges and using quasi-electrostatic

assumption, the cracked solid satisfies the generalized constitutive equations

$$\sigma_{iJ}(\mathbf{x},t) = C^{\lambda}_{iJKl} u_{K,l}(\mathbf{x},t) \tag{1}$$

and the generalized equations of motion

$$\sigma_{iJ,i}(\mathbf{x},t) = \rho^{\lambda} \delta^{*}_{JK} \ddot{u}_K(\mathbf{x},t), \quad \delta^{*}_{JK} = \begin{cases} \delta_{jk}, & J = j;\ K = k, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

the initial conditions

$$u_I(\mathbf{x}, t=0) = \dot{u}_I(\mathbf{x}, t=0) = 0, \tag{3}$$

the boundary conditions

$$u_I(\mathbf{x},t) = \bar{u}_I(\mathbf{x},t), \quad \mathbf{x} \in \Gamma_u, \tag{4}$$

$$t_I(\mathbf{x},t) = \bar{t}_I(\mathbf{x},t), \quad \mathbf{x} \in \Gamma_t, \tag{5}$$

and the continuity as well as the equilibrium conditions on the interface between the fiber and the matrix except the crack-faces

$$u^{I}_{I}(\mathbf{x},t) = u^{II}_{I}(\mathbf{x},t), \quad \mathbf{x} \in \Gamma_{if}, \tag{6}$$

$$t^{I}_{I}(\mathbf{x},t) = -t^{II}_{I}(\mathbf{x},t), \quad \mathbf{x} \in \Gamma_{if}, \tag{7}$$

with the lower case letter subscripts $j \in \{1,2\}$ and the capital letter subscripts $J \in \{1,2,4\}$, respectively. The generalized displacements u_I, the generalized tractions t_I, the generalized stresses σ_{iJ} and the generalized elasticity tensor C^{λ}_{iJKl} for a homogenous domain Ω^{λ} ($\lambda = 1,2,...,N$) are defined by

$$u_I = \begin{cases} u_i, & I = i \quad \text{(mechanical displacements)} \\ \varphi, & I = 4 \quad \text{(electrical potential)} \end{cases}, \tag{8}$$

$$\sigma_{iJ} = \begin{cases} \sigma_{ij}, & J = j \quad \text{(mechanical stresses)} \\ D_i, & J = 4 \quad \text{(electrical displacements)} \end{cases}, \tag{9}$$

$$C_{iJKl} = \begin{cases} c_{ijkl}, & J = j;\ K = k \quad \text{(elasticity tensor)} \\ e_{lij}, & J = j;\ K = 4 \quad \text{(piezoelectric tensor)} \\ e_{ikl}, & J = 4;\ K = k \quad \text{(piezoelectric tensor)} \\ -\kappa_{il}, & J = K = 4 \quad \text{(electrical permittivity tensor)} \end{cases}, \tag{10}$$

$$t_I(\mathbf{x},t) = \sigma_{jI}(\mathbf{x},t) e_j(\mathbf{x}). \tag{11}$$

In the Eqs. (1)-(11), e_j, u_i, σ_{ij}, φ and D_i are the outward unit normal vector, the mechanical displacements, the stresses, the electrical potential and the electrical displacements. Further, ρ, C_{ijkl}, e_{ijk} and κ_{ij} represent the mass density, the elasticity tensor, the piezoelectric tensor and the dielectric permittivity tensor. Γ_t and Γ_u define the external boundaries where the tractions t_I and the displacements u_I are prescribed, while Γ_{if} is the interface between the homogenous domains Ω^{λ} ($\lambda = 1,2,...,N$).

On the crack-faces three different electrical boundary conditions are considered. As an extension of the mostly applied traction-free crack-face boundary condition in linear elastic fracture mechanics it has been suggested in [4] to consider the crack as impermeable for the electrical field

$$D_2(\mathbf{x} \in \Gamma_{c^+}, t) = D_2(\mathbf{x} \in \Gamma_{c^-}, t) = 0. \tag{12}$$

$\Gamma_{c\pm}$ denotes the upper and the lower crack-faces. Another in [5] introduced model treats the crack as fully electrical permeable

$$D_2(\mathbf{x} \in \Gamma_{c^+}, t) = D_2(\mathbf{x} \in \Gamma_{c^-}, t), \quad \varphi(\mathbf{x} \in \Gamma_{c^+}, t) - \varphi(\mathbf{x} \in \Gamma_{c^-}, t) = 0. \tag{13}$$

This implies identical potentials on both crack-faces or in other words the crack exists only for the mechanical and not for the electrical field. In both models, the limited dielectric properties of the interior of the crack are not taken into account. Due to this fact a more realistic semi-permeable crack-face boundary condition has been introduced as [2]

$$D_2(\mathbf{x} \in \Gamma_{c^+}, t) = D_2(\mathbf{x} \in \Gamma_{c^-}, t) = -\kappa_c \frac{\varphi(\mathbf{x} \in \Gamma_{c^+}, t) - \varphi(\mathbf{x} \in \Gamma_{c^-}, t)}{u_2(\mathbf{x} \in \Gamma_{c^+}, t) - u_2(\mathbf{x} \in \Gamma_{c^-}, t)}, \tag{14}$$

where $\kappa_c = \kappa_r \kappa_0$ is the product of the relative permittivity of the considered crack medium κ_r and the permittivity of the vacuum $\kappa_0 = 8.854 \cdot 10^{-12} C/(Vm)$. D_2 and u_2 are the normal components of the electrical displacements and the mechanical displacements on the crack-faces. This crack-face boundary condition has been further improved by including electrostatic tractions [3], [1]. The generalized crack-opening-displacements (CODs) are defined by

$$\Delta u_I(\mathbf{x}, t) = u_I(\mathbf{x} \in \Gamma_{c^+}, t) - u_I(\mathbf{x} \in \Gamma_{c^-}, t). \tag{15}$$

Throughout the paper, a comma after a quantity represents spatial derivatives while a dot over the quantity denotes time differentiation. Lower case Latin indices take the values 1 and 2 (elastic), while capital Latin indices take the values 1, 2 (elastic) and 4 (electric). Unless otherwise stated, the conventional summation rule over repeated indices is implied.

Time-domain boundary integral equations and fundamental solutions

A spatial Galerkin-method is implemented to solve the initial-boundary value problem with the boundary element method. This demands that the time-domain boundary integral equations (BIEs) are treated in a weighted residual sense. The generalized time-domain displacement and traction BIEs can be written as [8]

$$\begin{aligned}
&\int_\Gamma \psi(\mathbf{x}) u_J(\mathbf{x}, t) \mathrm{d}\Gamma_x = \\
&\int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_b} \Big[u^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * t_I(\mathbf{y}, t) - t^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * u_I(\mathbf{y}, t) \Big] \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x \\
&+ \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_{c^+}} t^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * \Delta u_I(\mathbf{y}, t) \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x,
\end{aligned} \tag{16}$$

$$\begin{aligned}
&\int_\Gamma \psi(\mathbf{x}) t_J(\mathbf{x}, t) \mathrm{d}\Gamma_x = \\
&\int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_b} \Big[v^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * t_I(\mathbf{y}, t) - w^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * u_I(\mathbf{y}, t) \Big] \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x \\
&+ \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_{c^+}} w^G_{IJ}(\mathbf{x}, \mathbf{y}, t) * \Delta u_I(\mathbf{y}, t) \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x,
\end{aligned} \tag{17}$$

where $\psi(x)$ is the weight or test function, $\Gamma_b = \Gamma_u + \Gamma_t + \Gamma_{if}$, an asterisk denotes the Riemann convolution

$$g(\mathbf{x},t) * h(\mathbf{x},t) = \int_0^t g(\mathbf{x},t-\tau)h(\mathbf{x},\tau)\mathrm{d}\tau \tag{18}$$

and the dynamic displacement, traction and higher-order traction fundamental solutions are defined by

$$t^G_{IJ}(\mathbf{x},\mathbf{y},t) = C_{qIKr}e_q(\mathbf{y})u^G_{KJ,r}(\mathbf{x},\mathbf{y},t), \tag{19}$$

$$v^G_{IJ}(\mathbf{x},\mathbf{y},t) = -C_{pIKs}e_p(\mathbf{x})u^G_{KJ,s}(\mathbf{x},\mathbf{y},t), \tag{20}$$

$$w^G_{IJ}(\mathbf{x},\mathbf{y},t) = C_{pIKs}e_p(\mathbf{x})C_{qJLr}e_q(\mathbf{y})u^G_{KL,sr}(\mathbf{x},\mathbf{y},t). \tag{21}$$

The fundamental solutions possess the following spatial symmetry properties

$$u^G_{IJ}(\mathbf{x},\mathbf{y},t) = u^G_{JI}(\mathbf{y},\mathbf{x},t), \tag{22}$$

$$t^G_{IJ}(\mathbf{x},\mathbf{y},t) = -v^G_{IJ}(\mathbf{x},\mathbf{y},t) = v^G_{JI}(\mathbf{y},\mathbf{x},t), \tag{23}$$

$$w^G_{IJ}(\mathbf{x},\mathbf{y},t) = w^G_{JI}(\mathbf{y},\mathbf{x},t). \tag{24}$$

These symmetry properties (22)-(24) can be used to derive a spatial symmetric Galerkin-method. This is achieved if the displacement Galerkin-BIEs (16) are applied on the external boundary Γ_u where the generalized displacements are known and the interface Γ_{if} for the generalized tractions, while the traction Galerkin-BIEs (17) are used on the external boundary Γ_t where the generalized tractions are prescribed and the interface Γ_{if} for the generalized displacements.

The time-domain fundamental solutions for homogeneous linear piezoelectric solids [7] are implemented in this work. They are expressed in the 2D case by a line integral over a unit circle as

$$u^G_{IJ}(\mathbf{x},\mathbf{y},t) = \frac{\mathrm{H}(t)}{4\pi^2}\int_{|\mathbf{n}|=1}\sum_{m=1}^{3}\frac{P^m_{IJ}}{\rho c_m}\frac{1}{c_m t + \mathbf{n}\cdot(\mathbf{y}-\mathbf{x})}\mathrm{d}\mathbf{n}, \tag{25}$$

where $H(t)$, $\mathbf{n}$, c_m and P^m_{IJ} denote the Heaviside step function, the wave propagation vector, the phase velocities of the elastic waves and the projector. By integration by parts and applying the properties of the time convolution the time-domain generalized displacement fundamental solutions can be divided into a singular static and a regular dynamic part as

$$u^G_{IJ}(\mathbf{x},\mathbf{y},t) * f(t) = u^S_{IJ}(\mathbf{x},\mathbf{y})f(t) + u^D_{IJ}(\mathbf{x},\mathbf{y},t) * \dot{f}(t). \tag{26}$$

In the same way, the traction and the higher-order traction fundamental solutions can also be divided into their singular static and regular dynamic parts [8].

Numerical solution algorithm

To solve the time-domain BIEs (16) and (17) a numerical solution procedure is presented in the following. The Galerkin-method is used for the spatial discretization while a collocation method is utilized for the temporal discretization [9]. The piezoelectric solid is divided into several sub-domains with homogeneous material properties and to each sub-domain the time-domain BIEs (16) and (17) are applied. For the spatial discretization, the crack-faces, the external boundary of each homogeneous sub-domain and the interfaces of the cracked solid are discretized by linear elements. Linear shape functions are also used for the temporal discretization in the present analysis. At the crack-tips inside a homogeneous sub-domain, special crack-tip elements are applied to describe the local behaviour of the generalized CODs near the crack-tips

properly. This ensures an accurate and a direct calculation of the intensity factors from the numerically computed CODs. On the other hand, the asymptotic crack-tip field in the case of an interfacial crack between two dissimilar piezoelectric materials shows different oscillating and non-oscillating singularities in the generalized stress field [6], which makes an implementation of special crack-tip elements quite cumbersome. For this reason, only standard elements are applied at the crack-tips for interface cracks. The strongly singular and hypersingular boundary integrals can be computed analytically. By using linear temporal shape-functions, time integrations can also be performed analytically. Only the line integrals over the unit circle arising in the regular parts of the dynamic fundamental solutions have to be computed numerically by the standard Gaussian quadrature.

After temporal and spatial discretizations and considering the initial conditions the following systems of linear algebraic equations can be obtained for each sub-domain Ω^ζ $(\zeta = 1, 2, ..., N)$

$$\begin{aligned}\mathbf{C}_\zeta \mathbf{u}_\zeta^K = \quad & \mathbf{U}_\zeta^S \mathbf{t}_\zeta^K - \mathbf{T}_\zeta^S \mathbf{u}_\zeta^K + \mathbf{T}_\zeta^S \Delta \mathbf{u}_\zeta^K \\ & + \sum_{k=1}^{K} \left[\mathbf{U}_\zeta^{D;K-k+1} \mathbf{t}_\zeta^k - \mathbf{T}_\zeta^{D;K-k+1} \mathbf{u}_\zeta^k + \mathbf{T}_\zeta^{D;K-k+1} \Delta \mathbf{u}_\zeta^k \right], \end{aligned} \tag{27}$$

$$\begin{aligned}\mathbf{D}_\zeta \mathbf{t}_\zeta^K = \quad & \mathbf{V}_\zeta^S \mathbf{t}_\zeta^K - \mathbf{W}_\zeta^S \mathbf{u}_\zeta^K + \mathbf{W}_\zeta^S \Delta \mathbf{u}_\zeta^K \\ & + \sum_{k=1}^{K} \left[\mathbf{V}_\zeta^{D;K-k+1} \mathbf{t}_\zeta^k - \mathbf{W}_\zeta^{D;K-k+1} \mathbf{u}_\zeta^k + \mathbf{W}_\zeta^{D;K-k+1} \Delta \mathbf{u}_\zeta^k \right]. \end{aligned} \tag{28}$$

By invoking the continuity conditions (6) and (7) on the interface Γ_{if} as well as (12), (13) or (14) on the crack-faces Γ_{c+} and Γ_{c-} and by considering the boundary conditions (4) and (5), the following explicit time-stepping scheme can be obtained

$$\mathbf{x}^K = (\Xi^1)^{-1} \left[\Upsilon^1 \mathbf{y}^K + \sum_{k=1}^{K-1} \left(\Lambda^{K-k+1} \mathbf{t}^k - \Theta^{K-k+1} \mathbf{u}^k \right) \right], \tag{29}$$

where Ξ^1 and Υ^1 are the system matrices, $\mathbf{y}^K$ is the vector of the prescribed boundary data while $\mathbf{x}^K$ represents the vector of the unknown boundary data, which can be computed time-step by time-step.

The dynamic intensity factors for a crack-tip inside a homogeneous domain or on the interface are defined in [6] and [8]. They are obtained directly from the numerically computed generalized CODs.

Numerical examples

In the following, numerical examples are presented and discussed. To measure the intensity of the electrical loading the parameter

$$\chi = \frac{e_{22}}{\kappa_{22}} \frac{D_0}{\sigma_0} \tag{30}$$

is introduced, with σ_0 and D_0 being the mechanical and electrical loading amplitudes. For convenience, the mode-I, the mode-II and the mode-IV dynamic intensity factors for crack-tips inside a homogeneous sub-domain are normalized by

$$K_I^*(t) = \frac{K_I(t)}{K_0}, \quad K_{II}^*(t) = \frac{K_{II}(t)}{K_0}, \quad K_{IV}^*(t) = \frac{e_{22}}{\varepsilon_{22}} \frac{K_{IV}(t)}{K_0}. \tag{31}$$

In the same way, the real part K_1 and the imaginary part K_2 of the complex dynamic stress intensity factor and the electrical displacement intensity factor K_4 for interface cracks are normalized by

$$K_1^*(t) = \frac{K_1(t)}{K_0}, \quad K_2^*(t) = \frac{K_2(t)}{K_0}, \quad K_4^*(t) = \frac{e_{22}^I}{\varepsilon_{22}^I}\frac{K_4(t)}{K_0}, \tag{32}$$

with $K_0 = \sigma_0\sqrt{\pi a}$ and a is the half length of an internal crack.

A fiber reinforced plate with a crack near the fiber

In the first example as shown in Fig. 1, we consider a fiber reinforced plate with a crack of length $2a$ near the fiber. The geometry of the cracked plate is determined by $h = 16.0mm$, $w = 20.0mm$, $r = 5.0mm$ and $a = r$.

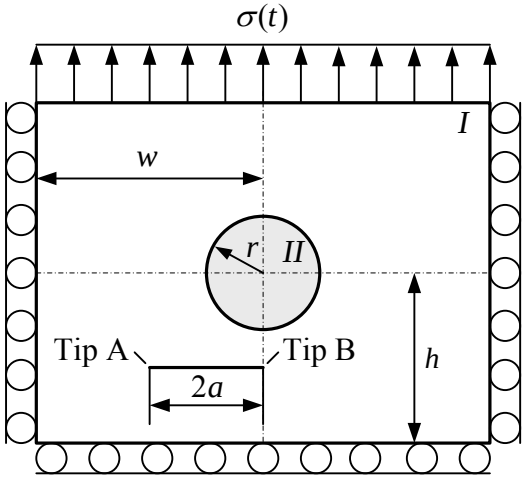


Figure 1: A fiber reinforced plate with a crack near the fiber

A tensile impact loading of the form $\sigma(t) = \sigma_0 \mathrm{H}(t)$ is applied on the upper boundary, where $\mathrm{H}(t)$ denotes the Heaviside step function. The normal components of the mechanical displacements are fixed on the left, right and lower boundary. As material for the matrix Epoxy is chosen, which has the following material parameters

$$\begin{aligned} &C_{11} = 8.0\text{GPa}, \quad C_{12} = 4.4\text{GPa}, \quad C_{22} = 8.0\text{GPa}, \quad C_{66} = 1.8\text{GPa}, \\ &\kappa_{11} = 0.0372\text{C/(GVm)}, \quad \kappa_{22} = 0.0372\text{C/(GVm)} \end{aligned} \tag{33}$$

and the mass density $\rho = 1260kg/m^3$. For the fiber three different configurations are investigated. In the first case we consider a circular hole. In contrast, a piezoelectric Zirconate Titanate (PZT-5H) with the material constants

$$\begin{aligned} &C_{11} = 126.0\text{GPa}, \quad C_{12} = 84.1\text{GPa}, \quad C_{22} = 117.0\text{GPa}, \quad C_{66} = 23.0\text{GPa}, \\ &e_{21} = -6.5\text{C/m}^2, \quad e_{22} = 23.3\text{C/m}^2, \quad e_{16} = 17.0\text{C/m}^2, \\ &\kappa_{11} = 15.04\text{C/(GVm)}, \quad \kappa_{22} = 13.0\text{C/(GVm)} \end{aligned} \tag{34}$$

and the mass density $\rho = 7500kg/m^3$ is applied in the second case for the fiber. To point out the influence of the hole and the piezoelectric fiber on the dynamic intensity factors Epoxy is chosen for the fiber in the third computation. This corresponds to a crack in a homogeneous

plate. The spatial discretization of the external boundary is performed by an element-length of $1.0mm$. The circular interface and the upper crack-face are approximated by 20 elements. A normalized time-step of $c_L\Delta t/h = 0.06$ is chosen, where c_L is the longitudinal wave velocity. The numerical results of the time-domain BEM are shown in Fig. 2.

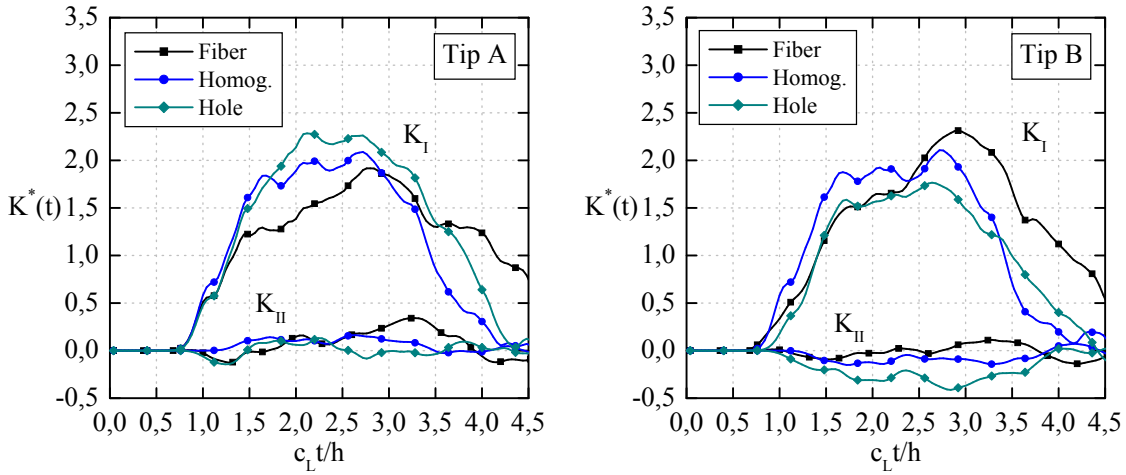


Figure 2: Normalized dynamic stress intensity factors of the three investigated configurations

The normalized dynamic mode-I and mode-II stress intensity factors of both crack-tips show a similar behavior. The curves of the homogeneous case are between the corresponding results of the fiber and hole configuration. The peak values of the left and the right crack-tip are nearly identical. In contrast, the mode-I stress intensity factors of the plate with the fiber and the hole show significant differences between both crack-tips. The right crack-tip is shielded by the hole which results in the lowest maximum peak value of all normalized dynamic mode-I stress intensity factor curves. On the other side the highest dynamic mode-II stress intensity factor is obtained. As clearly seen in Eqs. (33) and (34) the piezoelectric Zirconate Titanate has significant higher elastic constants than Epoxy. As a consequence the fiber increases the stiffness of the whole rectangular plate. Nevertheless the highest normalized dynamic mode-I stress intensity factor is obtained at the right crack-tip for the fiber configuration.

A square plate with a crack across the interface between the fiber and the matrix

In the next example a square plate with a crack across the interface between the fiber and the matrix is investigated. As depicted in Fig. 3 the cracked plate is subjected to an impact tensile loading $\sigma(t) = \sigma_0 \mathrm{H}(t)$ normal to the crack-faces on the upper and the lower boundary. On the left and the right boundary the mechanical stresses are zero. The geometrical data are $h = 20.0mm$, $r = h/2$ and $2a = 4.8mm$.

As in the first example the material properties given in Eqs. (33) and (34) are considered for the matrix and the piezoelectric fiber. For spatial discretization the external boundary and the interface are discretized by a uniform mesh with an element-length about $1.0mm$. The upper crack-face is divided into 16 elements. A normalized time-step $c_L\Delta t/h = 0.06$ is used. The crack-faces are treated as electrically impermeable described by Eq. (12). The normalized dynamic intensity factors obtained by the time-domain BEM are given in Fig. 4.

As clearly observed the normalized dynamic stress intensity factors for the left and the right crack-tip show a quite different behavior. The dynamic stress intensity factor for the right crack-tip is considerably larger than that for the left crack-tip. This is very interesting since the

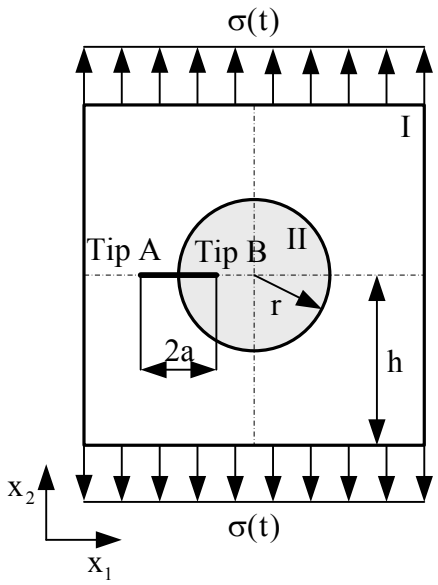


Figure 3: A crack across the interface between the fiber and the matrix in a square plate

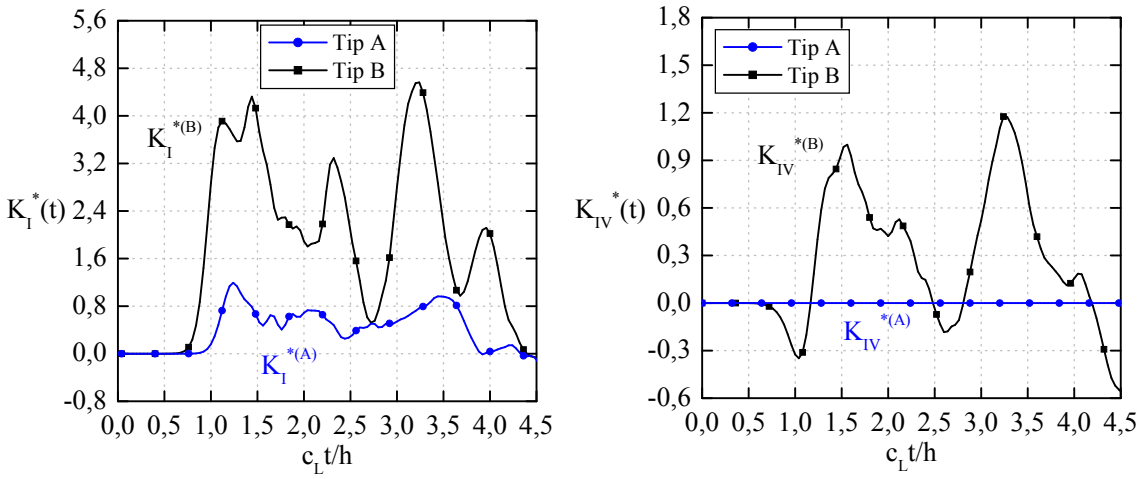


Figure 4: Normalized dynamic intensity factors of the left (A) and the right (B) crack-tip

elastic constants of the piezoelectric fiber (PZT-5H) are much higher than those of the matrix (Epoxy). The left crack-tip is inside the passive non-piezoelectric matrix and as a consequence the electrical displacement intensity factor is zero. Although the cracked plate is subjected to a pure mechanical impact loading a significant electrical displacement intensity factor is obtained at the right crack-tip. This is mainly induced by the coupling between the mechanical and the electrical field as well as the transient dynamic effects.

Interface crack in a square plate between the fiber and the matrix

In the last numerical example, we consider an interface crack in a square plate between the central fiber and the matrix as shown in Fig. 5. The geometry is prescribed by $h = 20.0mm$, $r = h/2$ and $2a = 14.0mm$. On the upper and the lower boundary an impact tensile loading $\sigma(t) = \sigma_0 \mathrm{H}(t)$ is applied.

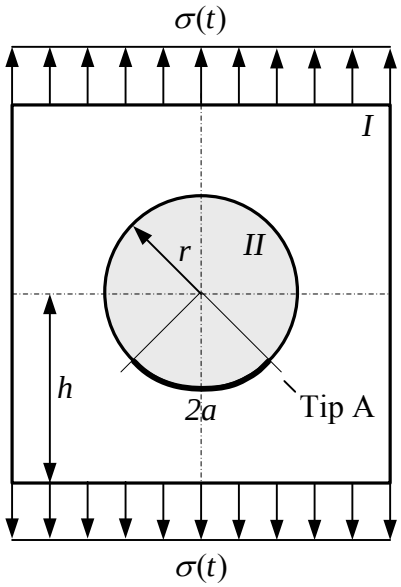


Figure 5: An interface crack in a square plate between the central fiber and the matrix

PZT-5H with the material properties given in Eq. (34) is used for the fiber (domain II). Barium Titanate ($\mathrm{BaTiO_3}$) with the material constants

$$\begin{aligned} &C_{11} = 150.0\text{GPa}, \quad C_{12} = 66.0\text{GPa}, \quad C_{22} = 146.0\text{GPa}, \quad C_{66} = 44.0\text{GPa}, \\ &e_{21} = -4.35\text{C/m}^2, \quad e_{22} = 17.5\text{C/m}^2, \quad e_{16} = 11.4\text{C/m}^2, \\ &\kappa_{11} = 9.87\text{C/(GVm)}, \quad \kappa_{22} = 11.2\text{C/(GVm)} \end{aligned} \tag{35}$$

and the mass density $\rho = 5800kg/m^3$ is chosen for the matrix (domain I). The external boundary and the interface are divided into elements with a length about $1.0mm$. The interface crack is divided into 20 elements. A normalized time-step $c_L\Delta t/h = 0.06$ is used.

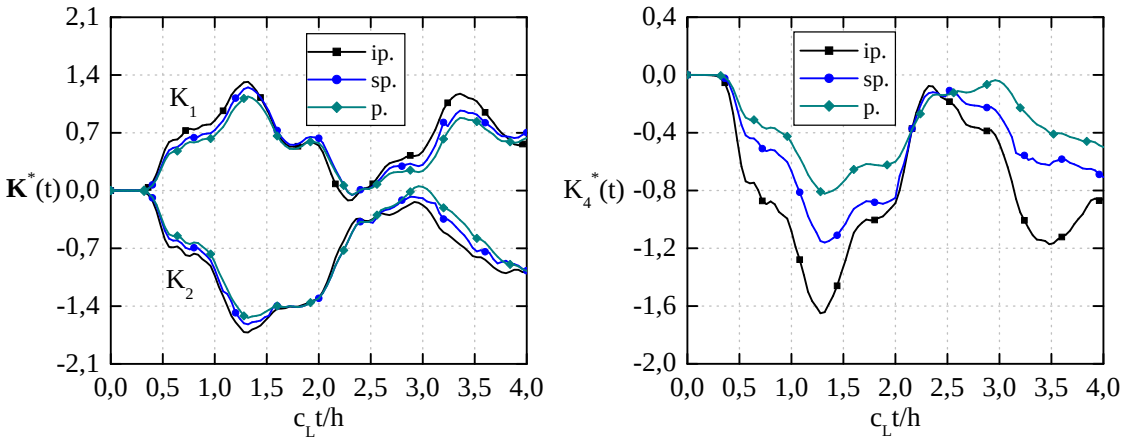


Figure 6: Normalized dynamic intensity factors of the interface crack

The normalized dynamic intensity factors for the impermeable (ip.), permeable (p.) and semi-permeable (sp.) crack-face boundary conditions (12)-(14) are shown in Fig. 6. The relative permittivity $\kappa_r = 40$ is used in the computations for the semi-permeable crack-face conditions. The elastic waves induced by the mechanical impact need some time to reach and excite the

crack. The global behavior of the dynamic intensity factors is very similar. It can be clearly seen, that the electrical permittivity of the medium inside the crack has a significant influence. Here again a high electrical displacement intensity factor is obtained even for a pure mechanical impact loading.

Conclusions

The transient dynamic analysis of piezoelectric fiber composites with cracks of arbitrary shape is presented in this paper. The developed symmetric Galerkin time-domain BEM is an attractive tool to compute the dynamic intensity factors. The formulation is general without limitations on the crack geometry, loading configuration and poling directions. The investigated numerical examples indicate a significant influence of the piezoelectric fiber and the transient dynamic loading on the normalized intensity factors.

Acknowledgement

This work is supported by the Slovak Academy of Sciences Project (SASPRO) 0106/01/01. The financial support is gratefully acknowledged.

References

[1] Gellmann R., Ricoeur A. (2012) Some new aspects of boundary conditions at cracks in piezoelectrics. *Archive of Applied Mechanics* **82**, 841-852.

[2] Hao T.H., Shen Z.Y. (1994) A new electric boundary condition of electric fracture mechanics and its applications. *Engineering Fracture Mechanics* **47**, 793-802.

[3] Landis C.M. (2004) Energetically consistent boundary conditions for electromechanical fracture. *International Journal of Solids and Structures* **41**, 6291-6315.

[4] Pak Y.E. (1990) Crack extension force in a piezoelectric material. *Journal of Applied Mechanics* **57**, 647-653.

[5] Parton V.Z. (1976) Fracture mechanics of piezoelectric materials. *Acta Astronautica* **3**, 671-683.

[6] Suo Z., Kuo C-M., Barnett D.M., Willis J.R. (1992) Fracture mechanics for piezoelectric ceramics. *Journal of the Mechanics and Physics of Solids* **40**, 739-765.

[7] Wang C.-Y., Zhang Ch. (2005) 3-D and 2-D dynamic Green's functions and time-domain BIEs for piezoelectric solids. *Engineering Analysis with Boundary Elements* **29**, 454-465.

[8] Wünsche M., García-Sánchez F., Sáez A., Zhang Ch. (2010) A 2D time-domain collocation-Galerkin BEM for dynamic crack analysis in piezoelectric solids. *Engineering Analysis with Boundary Elements* **34**, 377-387.

[9] Wünsche M., Zhang Ch., García-Sánchez F., Sáez A., Sladek J., Sladek V. (2011) Dynamic crack analysis in piezoelectric solids with non-linear electrical and mechanical boundary conditions by a time-domain BEM. *Computer Methods in Applied Mechanics and Engineering* **50**, 2848-2858.

Recursive Formulas, Fast Algorithm and Its Implementation of Partial Derivatives of the Beta Function

†H.Z. Qin[1], Youmin Lu[2] and Nina Shang[1]

[1]Institute of Applied Mathematics, Shandong University of Technology Zibo, Shandong, P. R. China
[2]Department of Mathematics and Computer Science Bloomsburg University Bloomsburg, PA 17815, USA
†Corresponding author: $qinhz\_000@163.com$

Abstract

In this paper, the values of Beta function $B(x,y)$ at $(-n,y),(x,-m),(-n,-m)$ for $n,m=0,1,2,\cdots,x,y\neq 0,1,2,\cdots$ are redefined and some recurrence formulas on the partial derivatives $B_{p,q}(x,y)=\frac{\partial^{q+p}}{\partial x^p\partial y^q}B(x,y)$ of the Beta function are established in Mathematica,where p,q are the positive integers, and x,y are complex numbers, When $x=n,n+\frac{1}{2},y=m,m+\frac{1}{2}$ and $n,m=0,\pm1,\pm2,\cdots$, $B_{p,q}(x,y)$ can be expressed as Riemann zeta function. We provide a fast algorithm, give its implementation in Mathematica, obtain closed forms of many generalized integrals and achieve high-precision calculation of these integrals.

Keywords: Riemann zeta function; Beta Function; Partial derivatives of the Beta Function; high-precision.

Introduction

In Mathematical software such as Mathematica, Maple and Matlab there are special functions, and the Beta function is one of them. By partial derivatives of the Beta function some generalized integral can be calculated. For example

$$\int_0^1 t^{x-1}(1-t)^{y-1}\ln^p t\ln^q(1-t)dt=B_{p,q}(x,y), \tag{1.1}$$

where $B_{p,q}(x,y)=\frac{\partial^{p+q}}{\partial x^p\partial y^q}B(x,y)$. However, we note that although the following integral exists

$$\int_0^1 t^{-2}(1-t)^{-2}\ln^p t\ln^q(1-t)dt \tag{1.2}$$

for integer $p,q\geq 2$, but $B_{p,q}(-1,-1)=\infty(D[D[Beta[xx,yy],\{xx,p\}]/.xx\to -1\{yy,q\}]/.yy\to -1)$ in Mathematica. By Mathematica symbolic integral, the closed form of the integral (1.2) can also be obtained for smaller p,q, but very time consuming, and the closed form of the integral (1.2) are difficult to obtain for larger integer p,q. By closed form, we mean that the integral can be expressed analytically in terms of a finite number of Riemann zeta functions and some constant π and the Euler-Mascheroni constant γ, etc.

The Beta function was the first known scattering amplitude in string theory, first conjectured by Gabriele Veneziano. It also occurs in the theory of the preferential attachment process, a type of stochastic urn process[1,2], the supersymmetric gauge theories[3] and other physical[4-5].

For $B_{p,q}(x,y)$ we have established a recurrence formula by the neutrix calculus[6-8]. In this article, in Mathematica, we give the function DBeta of the calculating $B_{p,q}(x,y)$ for positive integers p and q and complex numbers x and y. Through a number of examples show that our program is very effective is better in the calculation of the closed form and the numerical integration.

In the following sections, we introduce additional definitions of the Beta function, some recurrence formulas and an algorithm for calculating the values of partial derivatives of the Beta function.

Software Summary

Manuscript title: Remark on Beta Function and it's Partial Derivatives in mathematca.

Authors: Huizeng Qin, Youmin Lu Nina Shang

Title of program: BetaAll (for computing the Beta Function $B(x, y)$ in all complex values of x and y), DBeta (for computing partial derivatives $\frac{\partial^{p+q}}{\partial x^p \partial y^q} B(x, y)$ of the Beta Function in all complex values of x and y).

Licensing provisions: None

Computer: ACPI Multiprocessor PC.

Operating system: Microsoft windows XP, but does not depend on the particular operating system.

Programming language used: Mathematica 9

Memory required to execute with typical data: 2 Megabytes.

CPC Library Classification: 6.5 Software including Parallel Algorithms

Solution method: For the partial derivatives of the Beta Function, the recurrence formulas (2.5),(2.6), (2.7)-(2.9) and (2.12) in this paper are employed. BetaAll is composed of the following five key subprograms: DBeta, PolyGammaAmend, DPochhammer, DBeta1 and DBeta2. BetaAll is based on the formulas (2.2)-(2.4) and Beta in Mathematica. PolyGammaAmend is based on the formulas (2.13)-(2.20). DPochhammer is based on the formula (2.11). DBeta1 is based on the formulas (2.5) and (2.6) for x and $x + y \neq 0, -1, -2, \cdots$. DBeta2 is based on the formulas (2.7)-(2.9) and (2.12) for $x = -1, -2, \cdots$ or $y = -1, -2, \cdots$ or $x+y = -1, -2, \cdots$.

Nature of the problem: The Beta function $B(x, y)$ is a very important special function. Many mathematical softwares have defined inherent function (for example $Beta[x, y]$ in Mathematica) for computing the Beta function $B(x, y)$. However, wnen $x = -1, -2, \cdots$ or $y = -1, -2, \cdots$, $Beta[x, y]$ is not defined in Mathematica, and similar problem exists in other mathematics software. In addition, it is possible to use symbolic deferentiation and integration in Mathematica to obtain the partial derivatives of the Beta function, but it is very inefficient in speed and can rarely get the closed forms(it cannot get the closed form although it exists). Therefore, we give an algorithm that calculates the values of Beta function and its partial derivetives in the entire complex plane. In this way, one can obtain the closed forms of all integrals that can be expressed in terms of partial derivatives of Beta function.

Typical running time: The running time of BetaAll depends strongly on p, q, x, y and the number of bits required by computation precision. BetaAll is 30-10000 times faster than Integrate in Mathematica. As the number of bits for precision increases, the advantage of BetaAll becomes more significant.

The purpose of the program design: This process is designed to calculate the values of the partial derivatives of the Beta function. Thus, it can be used to achieve fast and high-precision calculation of generalized integrals that can be represented in terms of partial derivatives of the Beta function, regardless of computing power. The speed of this process is far superior to Integrate and NIntegrate in Mathematica.

Additional Definition and a Recurrence Formula of Partial derivatives of the Beta Function

The values of x and y must be real and non-negative for the Beta function $B(x,y)$ in Matlab. Although they may be complex in Mathematica and Maple, the definitions there

$$B(-n,y)=\infty, B(x,-m)=\infty, B(-n,-m)=\infty, n,m=0,1,2,\cdots, \tag{2.1}$$

where x and y is not an integer, lead to the following unreasonable results:

$$B(-1,\frac{1}{2})=\infty,\ B(-\frac{3}{2},\frac{1}{2})=0, B(-1,\frac{5}{2})=\infty,\ B(-\frac{3}{2},\frac{5}{2})=\pi.$$

To remedy this problem, it is necessary to modify (2.1). For this reason, we do give some additional definitions and results[9-12].

For $B(x,y)$ the following definitions are given for $x>0, y>0$ and $n,m=1,2,\cdots$:

$$\begin{aligned} B(n,-m)=B(-m,n)&=\sum_{l=0,l\neq m}^{n-1} C_{n-1}^{l}\frac{(-1)^l}{l-m}, m=0,1,2,\cdots, n=1,2,\cdots \\ &=\begin{cases} \frac{(-1)^m(m-1)!(n-m)!}{n!}, n=1,2,\cdots,m, m=1,2,\cdots \\ \frac{(-1)^n(m-1)!(H_n-H_{m-n-1})}{n!(m-n-1)!}, n=m+1,m+2,\cdots,m=1,2,\cdots,n \end{cases} \end{aligned} \tag{2.2}$$

$$\begin{gathered} B(-n,y)=(-1)^n C_{y-1}^{n}\left((y-n-1)B_{0,1}(y-n-1,1)+H_n\right), \\ y\neq 0,-1,-2,\cdots, \\ B(x,-m)=B(-m,x), x\neq 0,-1,-2,\cdots, \end{gathered} \tag{2.3}$$

where $H_n=\sum\limits_{l=1}^{n}\frac{1}{l}$, and

$$B(-n,-m)=-\sum_{i=0}^{m-1}\left(\begin{array}{c} n+i \\ i \end{array}\right)\frac{1}{m-i}-\sum_{j=0}^{n-1}\left(\begin{array}{c} m+j \\ j \end{array}\right)\frac{1}{n-j}. \tag{2.4}$$

We obtain the following three groups of recurrence formulas of $B_{p,q}(x,y)$.

I. For integers $q,p\geq 1$ and complex numbers x,y satisfying $x,y,x+y\neq 0,-1,-2,\cdots$,

$$\begin{aligned} B_{0,q}(x,y)&=\sum_{j=0}^{q-1}C_{q-1}^{j}\left(\psi^{(q-1-j)}(y)-\psi^{(q-1-j)}(x+y)\right)B_{0,j}(x,y), \\ B_{p,q}(x,y)&=\sum_{j=0}^{q-1}C_{q-1}^{j}\left(\psi^{(q-1-j)}(y)-\psi^{(q-1-j)}(x+y)\right)B_{p,j}(x,y) \\ &\quad-\sum_{k=0}^{p-1}C_{p}^{k}\sum_{j=0}^{q-1}C_{q-1}^{j}\psi^{(p+q-1-k-j)}(x+y)B_{k,j}(x,y). \end{aligned} \tag{2.5}$$

or

$$\begin{aligned} B_{p,0}(x,y)&=\sum_{k=0}^{p-1}C_{p-1}^{k}(\psi^{(p-1-k)}(x)-\psi^{(p-1-k)}(x+y))B_{k,0}(x,y), \\ B_{p,q}(x,y)&=\sum_{k=0}^{p-1}C_{p-1}^{k}(\psi^{(p-1-k)}(x)-\psi^{(p-1-k)}(x+y))B_{k,q}(x,y) \\ &\quad-\sum_{j=0}^{q-1}C_{q}^{j}\sum_{k=0}^{p-1}C_{p-1}^{k}\psi^{(p+q-1-k-j)}(x+y)B_{k,j}(x,y) \end{aligned} \tag{2.6}$$

where $\psi(x)$ is the digamma function defined by

$$\psi(x)=\frac{d}{dx}\ln\Gamma(x)=-\gamma-\frac{1}{x}+\sum_{l=1}^{\infty}\left(\frac{1}{l}-\frac{1}{l+x}\right).$$

II. For integers $q,p,n,m\geq 0$ and complex numbers x,y satisfying $x,y\neq 0,-1,-2,\cdots$,

$$\begin{array}{c}B_{p,q}(-n,y)=\frac{1}{(p+1)a_{n+1,1}(-n)}\sum\limits_{u=0}^{p+1}C_{p+1}^{u}\sum\limits_{v=0}^{q}C_{q}^{v}a_{n+1,p+q+1-u-v}(y-n)B_{u,v}(1,y)\\ -\frac{1}{(p+1)a_{n+1,1}(-n)}\sum\limits_{u=0}^{p-1}C_{p+1}^{u}a_{n+1,p+1-u}(-n)B_{u,q}(-n,y).\end{array}\tag{2.7}$$

$$\begin{array}{c}B_{p,q}(x,-m)=\frac{1}{(q+1)a_{m+1,1}(-m)}\sum\limits_{u=0}^{p}C_{p}^{u}\sum\limits_{v=0}^{q+1}C_{q+1}^{v}a_{m+1,p+q+1-u-v}(x-m)B_{u,v}(x,1)\\ -\frac{1}{(q+1)a_{m+1,1}(-m)}\sum\limits_{v=0}^{q-1}C_{q+1}^{v}a_{m+1,q+1-v}(-m)B_{p,v}(-m,y)\end{array}.\tag{2.8}$$

and

$$\begin{array}{c}B_{p,q}(-n,-m)=\frac{(-1)^{n+m}}{(q+1)(p+1)n!m!}\sum\limits_{u=0}^{p+1}C_{p+1}^{u}\sum\limits_{v=0}^{q+1}C_{q+1}^{v}a_{n+m+2,p+q+2-u-v}(-n-m)B_{u,v}(1,1)\\ -\frac{(-1)^{n}}{(p+1)n!}\sum\limits_{u=0}^{p-1}C_{p+1}^{u}a_{n+1,p+1-u}(-n)B_{u,q}(-n,-m)\\ -\frac{(-1)^{m}}{(q+1)m!}\sum\limits_{v=0}^{q-1}C_{q+1}^{v}a_{m+1,q+1-v}(-m)B_{u,v}(-n,-m)\\ -\frac{(-1)^{n+m}}{(q+1)(p+1)n!m!}\sum\limits_{u=0}^{p-1}C_{p+1}^{u}\sum\limits_{v=0}^{q-1}C_{q+1}^{v}a_{n+1,p+1-u}(-n)a_{m+1,q+1-v}(-m)B_{u,v}(-n,-m).\end{array}\tag{2.9}$$

where

$$a_{n,i}(x)=\frac{d^{i}}{dx^{i}}\left(x\right)_{n}=i!\sum_{k=i}^{n}C_{k}^{i}(-1)^{n-k}s(n,k)x^{k-i},i=1,2,\cdots,\tag{2.10}$$

$$\left(x\right)_{n}=x(x+1)\cdots(x+n-1)=\sum_{k=1}^{n}(-1)^{n-k}s(n,k)x^{k},\tag{2.11}$$

and $s(n,k)$ is the Stirling number of the first kind.

III. For integers $q,p,n,m\geq 0$ and complex numbers x,y satisfying $x+y=0,-1,-2,\cdots,Rex\neq 0,-1,-2,\cdots$, we have the following recurrence relations

$$\begin{array}{c}B_{p,q}(x,y)=\frac{1}{(x)_{n}(y)_{m}}\sum\limits_{u=0}^{p}C_{p}^{u}\sum\limits_{v=0}^{q}C_{q}^{v}a_{n+m,p+q-u-v}(x+y)B_{u,v}(x+n,y+m)\\ -\frac{1}{(y)_{m}}\sum\limits_{v=0}^{q-1}C_{q}^{v}a_{m,q-v}(y)B_{p,v}(x,y)-\frac{1}{(x)_{n}}\sum\limits_{u=0}^{p-1}C_{p}^{u}a_{n,p-u}(x)B_{u,q}(x,y)\\ -\frac{1}{(x)_{n}(y)_{m}}\sum\limits_{u=0}^{p-1}C_{p}^{u}\sum\limits_{v=0}^{q-1}C_{q}^{v}a_{n,p-u}(x)a_{m,q-v}(y)B_{u,v}(x,y).\end{array}\tag{2.12}$$

Now we are ready to consider the closed form of $B_{p,q}(x,y)$. It is well-known that the digamma

function $\psi(x)$ has the following identities:

$$\psi(n+x)=\psi(x)+\sum_{l=0}^{n-1}\frac{1}{(l+x)},\psi(x-n)=\psi(x)+\sum_{l=1}^{n}\frac{1}{(l-x)}, \tag{2.13}$$

$$\psi^{(k)}(x)=k!(-1)^{k+1}\zeta(k+1,x),k>0, \tag{2.14}$$

and

$$\begin{array}{c}\psi^{(k)}(n+x)=k!(-1)^{k+1}\zeta(k+1,x)+(-1)^{k}k!\sum\limits_{l=0}^{n-1}\frac{1}{(l+x)^{k+1}},k>0\\ \psi^{(k)}(x-n)=k!(-1)^{k+1}\zeta(k+1,x)+k!\sum\limits_{l=1}^{n}\frac{1}{(l-x)^{k+1}},k>0,\end{array} \tag{2.15}$$

where $\zeta(s,x)$ is the Hurwitz zeta function defined by

$$\zeta(s,x)=\sum_{l=0}^{\infty}\frac{1}{(l+x)^{s}},\zeta(s,0)=\zeta(s,1).$$

The Hurwitz zeta function $\zeta(s,x)$ also has the following identity

$$\begin{array}{c}\zeta(s,n+x)=\zeta(s,x)-\sum\limits_{l=0}^{n-1}\frac{1}{(l+x)^{s}},\zeta(s,-n+x)=\zeta(s,x)+\sum\limits_{l=1}^{n}\frac{1}{(x-l))^{s}},\\ \zeta(s,\frac{1}{2})=(2^{s}-1)\zeta(s).\end{array} \tag{2.16}$$

particularly[13],

$$\zeta(k,0)=\left\{\begin{array}{c}\gamma,k=1\\ \zeta(k),k>1\end{array}\right.,\zeta(k,\frac{1}{2})=\left\{\begin{array}{c}\gamma+2\ln 2,k=1\\ \left(2^{k}-1\right)\zeta(k),k>1\end{array}\right., \tag{2.17}$$

and

$$\begin{array}{c}\left.\begin{array}{c}\zeta(2n+1,\frac{1}{3})\\ \zeta(2n+1,\frac{2}{3})\end{array}\right\}=\frac{3^{2n+1}-1}{2}\zeta(2n+1)\\ \pm\frac{\sqrt{3}}{2\pi}\left(\left(2n+2+3^{2n+2}\right)\zeta(2n+2)-2\sum\limits_{l=0}^{n-1}3^{2n-2l}\zeta(2n-2l)\zeta(2l+2)\right)\end{array} \tag{2.18}$$

$$\begin{array}{c}\left.\begin{array}{c}\zeta(2n+1,\frac{1}{4})\\ \zeta(2n+1,\frac{3}{4})\end{array}\right\}=2^{2n}(2^{2n+1}-1)\zeta(2n+1)\\ \pm\frac{1}{2\pi}\left(2n+2+4^{2n+2}\right)\zeta(2n+2)-2\sum\limits_{l=0}^{n-1}4^{2n-2l}\zeta(2n-2l)\zeta(2l+2)\end{array} \tag{2.19}$$

$$\begin{array}{c}\left.\begin{array}{c}\zeta(2n+1,\frac{1}{6})\\ \zeta(2n+1,\frac{5}{6})\end{array}\right\}=\frac{6^{2n+1}-3^{2n+1}-2^{2n+1}+1}{2}\zeta(2n+1)\\ \pm\frac{1}{2\sqrt{3}\pi}\left(6^{2n+2}-3^{2n+2}\right)\zeta(2n+2)-2\sum\limits_{l=0}^{n-1}\left(6^{2n-2l}-3^{2n-2l}\right)\zeta(2n-2l)\zeta(2l+2)\end{array} \tag{2.20}$$

where $\zeta(1)=\gamma$.

Remark If $x=\pm n,\frac{1}{2}\pm n,y=\pm m,\frac{1}{2}\pm m,n,m=0,1,2,\cdots$, $B_{p,q}(x,y)$ certainly has closed form. If $x,y=\frac{1}{3}\pm n,\frac{1}{4}\pm n,\frac{1}{6}\pm n,n=0,1,2,\cdots$, $B_{p,q}(x,y)$ may have closed form. Otherwise, $B_{p,q}(x,y),p,q>1$ does not seem to have closed form.

Algorithms for Calculating $B_{p,q}(x,y)$ and comparison with the symbolic(numerical) integration in Mathematica

Algorithm

The source code $BetaAll[x,y]$ and $DBeta[x,y,p,q,all]$ that calculates the values of $B(x,y)$ and $B_{p,q}(x,y)$ is placed in the file beta.nb(Mathematica file format). $DBeta[x,y,p,q,all]$ calls five key subprograms BetaAll[x,y], PolyGammaAmend[k,x], DPochhammer[k,x], DBeta1[x,y,p,q,all] and DBeta2[x,y,p,q,all]. The following is our specific algorithm.

1) To obtain closed form, PolyGammaAmend performs the calculation of (2.13)-(2.18) when x is a real number or $x = a \pm n, a = 0, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{6}, \frac{5}{6}$, $n = 0, 1, 2, 3, \cdots$. Otherwise, PolyGamma from Mtathematica replaces PolyGammaAmend.

2) BetaAll does the calculation of (2.2)-(2.4) when $x$or $y = 0, -1, -2$. Otherwise, Beta from Mtathematica replaces BetaAll.

3) DPochhammer[k, x] does calculation of (2.10).

4) DBeta1[x, y, p, q, all] is to calculate the values of $B_{p,q}(x,y)$ by using (2.5) and (2.6) when $x, y, x+y \neq 0, -1, -2, \cdots$. The parameter all indicates whether all the values of $B_{i,j}(x,y), i = 0, 1, 2, \cdots, p, j = 0, 1, 2, \cdots q$ are displayed or only the value of $B_{p,q}(x,y)$ is displayed depending on all is positive or zero.

5) DBeta[x, y, p, q, all] calls BetaDl[x, y, p, q, all] directly when $x, y, x + y \neq 0, -1, -2, \cdots$. Otherwise, DBeta[$x, y, p, q, all$] calls DBeta2[$x, y, p, q, all$] that calculates the values of $B_{p,q}(x,y)$ by using (2.7)-(2.9) and (2.12) and calling two subprograms BetaAll[x, y, p, q, all] and DPochhammer[k, x].

The above algorithm is run in the mathematics symbolic computation system. If the numerical calculation, we will be in front of the source code to add a "N", for example, change BetaAll to NBetaAll and the above algorithm is run in the specified precision Prec. Therefore, in beta.nb there are a public constants: Prec, which is for the calculation precision.

Comparison with the symbolic(numerical) integration in Mathematica

In order to show how much more efficient of DBeta and NDBeta is than the corresponding programs of Mathematica, we apply it, the Mathematica symbolic integration (Integrate) and the Mathematica numerical integration (NIntegrate) to a couple more integrals with different parameters and display the running results in Tables 1 and 2.

Table 1. Comparison of Several Algorithms

x,y	p,q		Time	p,q		Time	p,q		Time
		I	113.1787		I	137.9672		I	159.0118
$2,2$	$4,4$	B	0.046800	$6,4$	B	0.078001	$6,6$	B	0.156001
		BD	0.873606		DB	2.199614		DB	6.162039
		I	57.08076		I	87.87536		I	123.6307
$2,-\frac{5}{2}$	$3,3$	B	0.483603	$4,5$	B	1.341609	$5,5$	B	5.132432
		BD	6.162039		BD	80.32491		BD	187.9500
		I	106.6266		I	126.7664		I	127.8272
$-1,\frac{5}{2}$	$3,4$	B	1.887612	$4,5$	B	5.226033	$5,5$	B	17.61251
		BD	$*0.031200$		BD	$*0.062400$		BD	$*0.046800$
		I	59.29598		I	69.42044		I	161.2114
$-1,-1$	$2,2$	B	0.062400	$3,2$	B	0.312002	$4,4$	B	3.915625
		BD	$*0.046800$		BD	$*0.031200$		BD	$*0.078001$

In Table 1, letters I, B and BD represent Integrate[$t^{x-1}(1-t)^{y-1}Log[t]^pLog[1-t]^q, [t,0,1\}$],DBeta-

$[x, y, p, q, 0]$ and $D[D[Beta[xx, yy], \{xx, p\}]/.xx \to x, \{yy, q\}]/.yy \to y]$, respectively. The data shows that B is much more efficient in time than I and BD, and the rate ranges from 7 to 2400. An asterisk in front the time consumed indicates that the algorithm is valid.

Table 2 Comparison of NDBeta$[x, y, p, q, 0]$ and NIntegrate[$t^{x-1}(1-t)^{y-1}$Log[t]^{p}Log[1-t]$^q, \{t, 0, 1\}, WorkingPrecision- > Prec]$

x, y	p, q		T_{32}, rr	T_{64}, rr	T_{128}, rr	T_{256}, rr
$2, 2$	$6, 6$	NI	$0.046800, 10^{-32}$	$0.156001, 10^{-65}$	$0.499203, 10^{-129}$	$1.856412, 10^{-257}$
		NB	$0.015600, 10^{-47}$	$0.031200, 10^{-100}$	$0.015600, 10^{-207}$	$0.015600, 10^{-420}$
$-\frac{5}{2}, -\frac{7}{3}$	$6, 6$	NI	$0.062400, 10^{-32}$	$0.280802, 10^{-65}$	$0.795605, 10^{-129}$	$2.761218, 10^{-257}$
		NB	$0.031200, 10^{-34}$	$0.031200, 10^{-87}$	$0.031200, 10^{-194}$	$0.031200, 10^{-407}$
$-4, -\frac{5}{2}$	$6, 6$	NI	$0.062400, 10^{-32}$	$0.202801, 10^{-65}$	$0.717605, 10^{-117}$	$2.511616, 10^{-257}$
		NB	$0.062400, 10^{-39}$	$0.062400, 10^{-92}$	$0.062400, 10^{-199}$	$0.062400, 10^{-412}$
$-4, -5$	$6, 6$	NI	$0.046800, 10^{-32}$	$0.187201, 10^{-64}$	$0.577204, 10^{-129}$	$2.402415, 10^{-257}$
		NB	$0.093601, 10^{-38}$	$0.093601, 10^{-91}$	$0.093601, 10^{-198}$	$0.093601, 10^{-411}$

In this table, NI and NB represent

$$NIntegrate[t^{x-1}(1-t)^{y-1}Log[t]^p Log[1-t]^q, \{t, 0, 1\}, WorkingPrecision \to Prec]$$

and NDBeta$[x, y, p, q, 0]$, respectively. In order to reduce the accumulated calculation accuracy take [5Prec/3]. The subindex of T indicates the computing accuracy requirement and rr is the relative error. From Table 2, we see that the running time of NDBeta$[x, y, p, q, 0]$ is not substantially affected by the specified accuracy and much smaller than NIntegrate[$t^{x-1}(1-t)^{y-1}Log[t]^p Log[1-t]^q$, $[t, 0, 1\}$,WorkingPrecision-¿Prec] and its efficiency is more significant especially in high-precision. It is noteworthy that the relative error of the BetaD$[x, y, p, q, 0]$ is always less than the specified one.

Partial derivatives of the Beta Function used in some generalized integral calculation

Many generalized integrals can be expressed in terms of Beta function and its partial derivatives. Thus, a faster and high accuracy algorithm for calculating the values of the Beta function and its partial derivatives can speed up and increase the accuracy of the calculation of generalized integrals. There are many identities in this respect. For Example [14],

$$\begin{gathered}
\int_0^1 t^{x-1}(1-t)^{y-1}dt = B(x,y) \qquad [Rex, Rey > 0] \\
\int_0^\infty \frac{t^{x-1}}{(1+t)^{x+y}}dt = B(x,y) \qquad [Rex > 0, Rey > 0] \\
\int_{-1}^1 \frac{(1+t)^{2x-1}(1-t)^{2y-1}}{(1+t^2)^{x+y}}dt = 2^{x+y-2}B(x,y) \quad [Rex > 0, Rey > 0] \\
\int_0^1 \frac{t^{x-1}+t^{y-1}}{(1+t)^{x+y}}dt = \int_1^\infty \frac{t^{x-1}+t^{y-1}}{(1+t)^{x+y}}dt = B(x,y) \quad [Rex, Rey > 0] \\
\int_0^1 \frac{(1+t)^{x-1}(1-t)^{y-1}+(1+t)^{y-1}(1-t)^{x-1}}{2^{x+y-1}}dt = B(x,y) \quad [Rex > 0, Rey > 0] \\
\int_0^{\frac{\pi}{2}} \sin^{2x-1}t \cos^{2y-1}t dt = \frac{1}{2}B(x,y) \qquad [Rex, Rey > 0] \\
\int_{-\infty}^\infty \frac{e^{2iyt}}{(2\cosh t)^{2x}}dt = \frac{B(x+iy,x-iy)}{2} \quad [Rex > 0, y \text{ is a real}] \\
\int_{-\infty}^\infty \frac{e^{-2yt}}{(2\cosh t)^{2x}}dt = \frac{B(x-y,x+y)}{2} \quad [Rex > 0, Rex > |Rey|] \\
\int_0^\infty \frac{\cosh 2yt}{(2\cosh zt)^{2x}}dt = \frac{B(x+\frac{y}{z},x-\frac{y}{z})}{4z} \quad [Rex > |Rey|] \\
\int_0^1 (1-t^z)^{x-1}t^{y-1}dt = \frac{1}{z}B(x,\frac{y}{z}) \quad \left[Re\frac{y}{z}, Rex > 0\right] \\
\int_0^\infty e^{-xt}\sinh^y ztdt = \frac{1}{2^{y+1}z}B(\frac{x}{2z}-\frac{y}{2}, y+1) \quad [Rez > -1, Rezy > 0] \\
\int_0^\infty \frac{e^{-xt}}{\cosh^{2y+1}xt}dt = \frac{2^{2y-2}}{x}B(y,y) - \frac{1}{2xy} \quad [x, y > 0] \\
\int_0^\infty e^{-xt}(\cosh zt - 1)^y dt = \frac{B(\frac{x}{z}-y,2y+1)}{2^y z} \quad \left[Rey > -\frac{1}{2}, Rex > Rezy, Rez > 0\right]
\end{gathered} \tag{4.1}$$

and

$$\begin{array}{c}\int_0^1 \frac{\left(\frac{t}{t+z}\right)^x\left(\frac{1-t}{t+z}\right)^y}{t(1-t)}dt=\left\{\begin{array}{c}\frac{B(x,y)}{z^y(1+z)^x},Rex,Rey>0,Re(x+y)<1,-1<z<0\\ \left(\frac{1}{z}\right)^y\left(\frac{1}{1+z}\right)^x B(x,y),\ Rex,Rey>0,z\notin[-1,0]\end{array}\right.\\ \int_0^{\frac{\pi}{2}} \frac{\left(\frac{\cos^2 t}{\cos^2 t+z}\right)^x\left(\frac{\sin^2 t}{\cos^2 t+z}\right)^y}{\sin t\cos t}dt=\left\{\begin{array}{c}\frac{B(x,y)}{2z^y(1+z)^x},Rex,Rey>0,Re(x+y)<1,-1<z<0\\ \frac{1}{2}\left(\frac{1}{z}\right)^y\left(\frac{1}{1+z}\right)^x B(x,y),\ Rex,Rey>0,z\notin[-1,0]\end{array}\right.\end{array} \tag{4.2}$$

By the means of (4.1) and (4.2), we can express many generalized integrals in terms of partial derivatives of the Beta function. We give several examples here.

1) When p and q are non-negative integers, $p+Rey>0$ and $q+Rex>0$, we have

$$\int_0^1 \left(\begin{array}{c}(-1)^q t^{x-1}(1+t)^{-x-y}\ln^p\frac{t}{1+t}\ln^q(1+t)\\ +(-1)^p t^{y-1}(1+t)^{-x-y}\ln^q\frac{t}{1+t}\ln^p(1+t)\end{array}\right)dt=B_{p,q}(x,y), \tag{4.3}$$

and

$$\int_0^1 \frac{t^{x-1}}{(1+t)^{2x}}\ln^p\frac{t}{1+t}\ln^p(1+t)dt=\frac{(-1)^p}{2}B_{p,p}(x,x). \tag{4.4}$$

2) When p and q are non-negative integers,we have

$$\begin{array}{c}\int_0^1 \frac{1}{t(1-t)}\left(\frac{t}{t+z}\right)^x\left(\frac{1-t}{t+z}\right)^y\ln^p\frac{t}{t+z}\ln^q\frac{1-t}{t+z}dt\\ =\left(\frac{1}{z}\right)^y\left(\frac{1}{1+z}\right)^x\sum_{j=0}^{p}C_p^j\ln^{p-j}(\frac{1}{1+z})\sum_{k=0}^{q}C_q^k\ln^{q-k}\frac{1}{z}B_{j,k}(x,y).\end{array} \tag{4.5}$$

for $Rex,Rey>0,z\notin[-1,0]$.

3) When p and q are non-negative integers, and $Rex>|y|$, y is real, we have

$$\begin{array}{c}\int_0^\infty \frac{t^{2q}\cosh 2yt\ln^p\cosh t}{(2\cosh zt)^{2x}}\ln^p(2\cosh zt)\,dt\\ =\frac{1}{2^{p+2q+2}z^{2q+1}}\sum_{j=0}^{p}C_p^j\sum_{k=0}^{2q}(-1)^k C_{2q}^k B_{j+2q-k,p-j+k}(x+\frac{y}{z},x-\frac{y}{z})\end{array} \tag{4.6}$$

and

$$\begin{array}{c}\int_0^\infty \frac{t^{2q+1}\sinh 2yt\ln^p(2\cosh zt)}{(2\cosh zt)^{2x}}dt\\ =\frac{1}{2^{p+2q+3}z^{2q+2}}\sum_{j=0}^{p}C_p^j\sum_{k=0}^{2q+1}(-1)^k C_{2q+1}^k B_{j+2q+1-k,p-j+k}(x+\frac{y}{z},x-\frac{y}{z}).\end{array} \tag{4.7}$$

4) When p and q are non-negative integers, $\alpha>0$ and $Rex>|Imy|$, we have

$$\begin{array}{c}\int_{-\infty}^\infty \frac{t^q e^{-2yt}\ln^p(2\cosh t)}{(2\cosh t)^{2x}}dt=\frac{(-1)^{p+q}}{2^{p+q+1}}\sum_{j=0}^{p}C_p^j\sum_{k=0}^{q}(-1)^k C_q^k B_{k+j,p+q-k-j}(x-y,x+y)\\ p,q\text{ are integer},\ p,q\geq 0,\ Rex>|Rey|.\end{array} \tag{4.8}$$

5) When p and q are non-negative integers, $Rez>0,p+Re\frac{y}{z}$ and $q+Rex>0$, we have

$$\int_0^1 (1-t^z)^{x-1}t^{y-1}\ln^p(1-t^z)\ln^q t\,dt=\frac{1}{z^{q+1}}B_{p,q}(x,\frac{y}{z}). \tag{4.9}$$

6) Letting $t = \sin^2 u$ or $\cos^2 u$ in (4.5), we have

$$\begin{aligned}&\int_0^{\frac{\pi}{2}} \frac{\left(\frac{\cos^2 t}{\cos^2 t+z}\right)^x \left(\frac{\sin^2 t}{\cos^2 t+z}\right)^y}{\sin t \cos t} \ln^p \frac{\cos^2 t}{\cos^2 t+z} \ln^q \frac{\sin^2 t}{\cos^2 t+z} dt \\ &= \frac{1}{2}\left(\frac{1}{z}\right)^y \left(\frac{1}{1+z}\right)^x \sum_{j=0}^{p} C_p^j \ln^{p-j}(\tfrac{1}{1+z}) \sum_{k=0}^{q} C_q^k \ln^{q-k} \tfrac{1}{z} B_{j,k}(x,y),\end{aligned} \tag{4.10}$$

for integer, $p, q \geq 0, Rex, Rey > 0, z \notin [-1, 0]$.

7) When p and q are non-negative integer, $q + Rex > 0$ and $Rey > 0$, we have

$$\int_0^{\infty} t^{x-1}(1+t)^{-x-y} \ln^p \frac{t}{1+t} \ln^q(1+t) dt = (-1)^q B_{p,q}(x,y). \tag{4.11}$$

For $x = \pm n, \frac{1}{2} \pm n, y = \pm m, \frac{1}{2} \pm m, n, m = 0, 1, 2, \cdots$, $B_{p,q}(x,y)$ and $B_{p,q}(x+y, x-y)$ always exists closed form, so the generalized integral (4.3)-(4.11), which also exist closed form. However, the use of symbolic integration (Integrate) in Mathematica, closed forms of these integrals are difficult to obtain. For example, in Mathematica we have the following results for the generalized integral (4.3).

$$\begin{aligned}&x = 2; y = 1/2; p = 1; q = 1; \\ &Timing[s1 = Integrate[\tfrac{t\hat{}(x-1) Log[\frac{t}{1+t}]\hat{}p * Log[\frac{1}{1+t}]\hat{}q}{(1+t)\hat{}(x+y)} + \tfrac{p*t\hat{}(y-1) Log[\frac{t}{1+t}]\hat{}q * Log[\frac{1}{1+t}]\hat{}p}{(1+t)\hat{}(x+y)}, \{t, 0, 1\}]] \\ &Timing[s2 = Simplify[DBeta[x, y, p, q, 0]]] \\ &N[s1 - s2, Prec] \\ &\{96.736220, \tfrac{1}{27}(320 - 116\sqrt{2}] - 30\pi^2 + 72 ArcSin[\sqrt{2}]^2 - 72 ArcSinh[1] - 27 Log[2]^2 - \\ &4i\sqrt{2} HypergeometricPFQ[\{-\tfrac{3}{2}, -\tfrac{3}{2}, -\tfrac{3}{2}, \tfrac{1}{2}\}, \{-\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}\}, 2] - \\ &12i\pi(-13 + Log[8]) - 64 Log[8] - 72 I ArcSin[\sqrt{2}](1 + Log[4] + 2Log[4 - 2\sqrt{2}]) - \\ &96 Log[-1 + \sqrt{2}] + 108 Log[2] Log[1 + \sqrt{2}] + 144 Log[1 + \sqrt{2}]^2 - \\ &144 Log[1 + \sqrt{2}] Log[2 + \sqrt{2}] - 216 PolyLog[2, -\tfrac{1}{\sqrt{2}}] + \\ &288 Log[1 + \sqrt{2}] + 216 PolyLog[2, 1 - \sqrt{2}] + 72 PolyLog[2, -3 + 2\sqrt{2}])\} \\ &\{0., -\tfrac{2}{27}(9\pi^2 + 16(-10 + Log[64]))\} \\ &0. * 10^{-111} + 0. * 10^{-112} i\end{aligned}$$

When $p, q > 1$, the use of the symbolic integration even above complex can not be obtained.

However, the right-hand sides($(4.*)_R$) of the equations (4.3)-(4.11) give a high accuracy and fast algorithm to calculate the integrals of the left-hand side($(4.*)_L$). In order to verify the correctness of the formulas (4.3)-(4.11) and further show the high accuracy and time efficiency of our algorithm, the following numerical results are given in Mathematica.

Table 3 Comparison of numerical integration for (4.3)-(4.5)

	p, q, x, y	T_{32}, rr	T_{64}, rr	T_{128}, rr	Integral value
$(4.3)_L$	$4, 4, -2, i-3$	$0.1716, 10^{-32}$	$0.5616, 10^{-64}$	$1.5600, 10^{-128}$	$-1.554939\cdots$
$(4.3)_R$		$0.0156, 10^{-46}$	$0.0156, 10^{-99}$	$0.0156, 10^{-206}$	$+1.779627\cdots i$
$(4.4)_L$	$4, 4, -\frac{5}{2}+i, -3+i$	$0.1404, 10^{-32}$	$0.3900, 10^{-64}$	$1.2636, 10^{-128}$	$-2.114631\cdots$
$(4.4)_R$		$0.0156, 10^{-50}$	$0.0156, 10^{-102}$	$0.0156, 10^{-210}$	$-2.863698\cdots i$
$(4.5)_L$, z=$\frac{1}{2}$	$3, 2, \frac{1}{3}, \frac{1}{5}$	$0.1248, 10^{-14}$	$0.2028, 10^{-23}$	$0.4836, 10^{-27}$	$-279.808345\cdots$
$(4.5)_R$, z=$\frac{1}{2}$		$0., 10^{-51}$	$0.0156, 10^{-104}$	$0.0156, 10^{-211}$	
$(4.5)_L$, z$=-2$	$3, 3, \frac{1}{3}, \frac{1}{4}$	$0.1872, 10^{-13}$	$0.4056, 10^{-18}$	$0.7800, 10^{-27}$	$-54904.915\cdots$
$(4.5)_R$, z$=-2$		$0., 10^{-51}$	$0.0156, 10^{-104}$	$0.0156, 10^{-211}$	$+28376.28\cdots i$

In particular, for the left integral($(4.5)_L$) of the formula (4.5), the error of numerical integration always exists regardless of the calculation precision. When the numerical computation in Mathematica is used for calculating values of integrals, the error does not much improve no matter how the accuracy requirement is increased.

It is noteworthy that we found symbol integration and numerical integration of inconsistent results in Mathematica.

$$z = -1/2; x = 1/4; y = 1/5; Prec = 32;$$
$$Timing[s1 = NIntegrate[\left(\frac{t}{z+t}\right)\hat{}x\left(\frac{1-t}{z+t}\right)\hat{}y\frac{1}{t(1-t)}, \{t, 0, 1\}, WorkingPrecision->Prec]]$$
$$Timing[s2 = N[Integrate[\left(\frac{t}{z+t}\right)\hat{}x\left(\frac{1-t}{z+t}\right)\hat{}y\frac{1}{t(1-t)}, \{t, 0, 1\}], Prec]]$$
$$Timing[s0 = \frac{1}{z\hat{}y}\left(\frac{1}{1+z}\right)\hat{}y * NBetaAll[x, y]]$$
$$\{s1 - s0, s2 - s0, s3 - s0\}$$
$$\{0.124801, 9.3442494430451904923601953855976 + 6.7887335956884325541842812213085I\}$$
$$\{1.419609, 9.3462960217122886259805243287725 - 6.7904815391012934935031911000193I\}$$
$$\{0., 9.3462960217122886259805243287725073193521592331001509-$$
$$6.7904815391012934935031911000192799983188576389449694I\}$$
$$\{-0.0020465786670981336203289431749 + 13.579215134789726047687472321327 8i,$$
$$0. \times 10^{-32} + 0. \times 10^{-32}\}$$

Conclusions

By giving additional definition of the Beta function, the domain of the Beta function has been extended to the entire complex plane. In the entire complex plane, we have established recursive formulas on the partial derivatives of the Beta function. Applying these recursive formulas, we give the conditions of the closed form of the generalized integral, which are expressed in terms of partial derivatives of the Beta function. And for the numerical calculation, calculation speed and accuracy have some improvements.

References

[1] http://en.wikipedia.org/wiki/Gabriele_Veneziano#cite_note-5.

[2] Riddhi D., Beta Function and its Applications, http://sces.phys.utk.edu/˜moreo/mm08/Riddi.pdf.

[3] V.A. Novikov, M.A. Shifman,A.I. Vanshtein and V.i. Zakhrov, The Beta Function in Supersymmetric Gauge Theories. Instantons Versus Traditional Approach,Physics Letters, Volume 166B,number 3 16 January (1986) ,329–333.

[4] Akio Morita, Haruyo Koiso, Yukiyoshi Ohnishi, and Katsunobu Oide, Measurement and correction of on-and off-momentum beta functions at KEKB, Phys. Rev. ST Accel. Beams 10, 072801 Published 27 July 2007

[5] G.Benfatto G.Gallavottie, A.Procacci B.Scoppola, Beta Function and Schwinger Functions for a Many Fermions System in One Dimension. Anomaly of the Fermi Surface, Commun. Math.Phys. 160,93–171 (1994)

[6] Y. Jack Ng, H. van Dam, Neutrix calculus and finite quantum field theory, J. Phys. A 38 (2005) 317–323.

[7] Y. Jack Ng, H. van Dam, An application of neutrix calculus to quantum field theory, Int.J.Mod.Phys. A21 (2006) 297–312

[8] Youn-Sha Chan, Albert C. Fannjiang, Glaucio H. Paulino, and Bao-Feng Feng, Finite Part Integrals and Hypersingular Kernels,DCDIS A Supplement, Advances in Dynamical Systems, Vol. 14(S2)264–269

[9] N. Shang and H. Qin, The Closed Form on a Kind of Log-cosine and Log-sine Integral, Journal of Mathematics in Practice and Theory,Volume 42, No.23, December 2012 234–246.

[10] N. Shang, A. Li, Z. Sun, and H. Qin,A Note on the Beta Function And Some Properties of Its Partial Derivatives, IAENG International Journal of Applied Mathematics, 44:4, IJAM_44_4_06.

[11] Z. Sun, H. Qin,A. Li, Extension of the partial derivatives of the incomplete beta function for complex values, Applied Mathematics and Computation 275(2016)63–71.

[12] A Li, Z Sun, H Qin,The Algorithm and Application of the Beta Function and Its Partial Derivatives, Engineering Letters, 23:3, EL_23_3_04

[13] Huizeng Qin, Nina shang, Aijuan Li, Some identities on the Hurwitz zeta function and the extended Euler sums. Integral Transforms and Special Functions. iFirst, 2012,1–21.

[14] I.S.Gradshteyn I. M. Ryzhik,Table of Integrals Series and Products, Seventh Ediyion,Academic Press is an imprint of Elsevier, 908–909

Kernel-based Collocation Method for Deformable Image Registration Model

*S. M. Wong[1], T. S. Li[1] and †K. S. NG[1]
[1]The school of Science and Technology, the Open University of Hong Kong
*Presenting author: anwong@ouhk.edu.hk
†Corresponding author: dng@ouhk.edu.hk

Abstract

The deformable image registration (DIR) is a medical imaging device which is used to assess the growth of tumor and adjust the target region in radiotherapy treatment accordingly. During the treatment, the patient could suffer a significant weight loss, and the shrinkage of tumor could also bring a mass transformation over the target region. These anatomical changes could affect the medical outcome.

The satisfactory treatment results could be achieved if the exact location and the spatial extent of target region are accurately measured. Thus the surrounding health organs could suffer less impact.

The DIR has been studied over 30 years in the discipline of biomedical science. Over these years, the progress of deformable image models has achieved a diversity development. However, the human body is quite complicated and its complexity obstructs the development of a precision scheme. In order to accomplish the clinically satisfactory level, a high level of accuracy is indispensable in delivering right medical dose to radiotherapy treatment.

This paper adopts the meshless kernal-based collocation method together with the Demo's iterative algorithm to solve the concerned DIR model. The formulation to the classical model of DIR algorithms are outlined. The proposed algorithm is applied to a real-life data from a patient who suffered liver cancer. The aim of study is to trace the growth of the tumor of liver cancer.

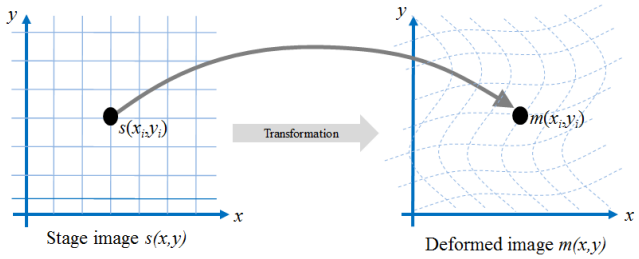
Keywords: Deformable image, Meshless, Kernel-based collocation.

Introduction

Cancer is a fatal disease in and is one of leading killers in the world. Many cancer patients were treated with external beam radiotherapy treatment during diseases management. Identifying the target region of tumor accurately can avoid damages to healthy organs. This can inhibit the tumor growth and minimize the side effects to the patient. This can be done by Image Guided Radiation Therapy (IGRT). In order to optimize IGRT, a segmentation and registration method has to be used to delineate the clinically critical objects in the computed tomography images obtained from the radiation treatment.

Deformable image registration is a process in medical image analysis. Even though there has been a good progress in the development of deformable registration image methods, this topic remains a challenging problem in the field of radiotherapy.

Usually image registration is formulated as an optimization problem. Given that the static image $s(x, y)$ and deformed image $m(x, y)$ were taken at time t_0 and t_1 respectively, the target object had undergone some changes from time t_0 to t_1. We want to find the spatial displacement between the static image $s(x_i, y_i)$ and the deformed image $m(x_i, y_i)$. An example of 2D deformable transformation mapping is illustrated in *Figure 1.*



Figire 1 An example of 2D deformable transformation mapping

In this paper, the meshless scheme based on the kernel collocation will be used to approximate the displacements between the two images. Radial basis function (RBF) is a well known meshless algorithm and had been proved to be effective in solving various kinds of partial differential equations. RBF was first used in neural network and later used for multivariate interpolation. Osorio *et al* [1] combined the feature matching and RBF interpolation in registration for radiotherapy; the numerical approximation was obtained iteratively by using the so called thin plate spline robust point matching technique.

Reference case study

A real-life deformable image registration from a patient with liver cancer is used as a reference case study. One of the original static registered images is depicted in *Figure (2) at time* t_1 and the deformed images are obtained from two different treatment periods *at time* t_2 and t_3.

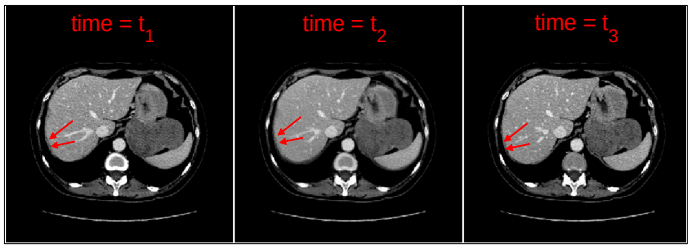


Figure 2 The growth of tumor of liver cancer as indicated by an arrow from different time $t_1 < t_2 < t_3$.(The local ethics committee approval was obtained for a waiver of informed consented for retrospective analysis. The CT images were collected from collaborative hospital in year 2010.)

The computational region is set up according to the original registered image. The geometrical structure of the original static image contains 512×512 pixels. In the suggested algorithm, the computational region is simplified by removing the insignificant backgrounds. These include the air with intensity equals to 0, and the bone with intensity equals to 1. After removing the insignificant pixels, the study region contains $82,776$ valid pixels.

The objective of the present study is to use a global kernel-based approximation method to predict the progression of tumor changes in a future time t_3 by using the known information given by the deformable image registration obtained from t_1 and t_2. Similarly, the subsequent changes of tumor can be predicted by using the deformed relationship obtained in t_2 and t_3. Our model aims to give a prediction to the changes of tumor so that an appropriate medical treatment can be applied according to the stage of the cancer. For example, the cancer is at early stage, say at stage A when diagnosed, a complete medical treatment may be possible by means of liver transplation, resection of liver using radiotherapy treatment.

Deformable Image Registration Model

This section introduces the basic deformable image registration problem which was first proposed by Broit [2] in 1981. The image is modelled as a 2D homogeneous infinite elastic medium under the influence of some forces. It is assumed that the external forces are applied at every control point, where the strengths, directions and influence areas of the forces as well as the positions of the control points are considered.

Several forms of deformable image registration models are established in different numerical methods. One of the classical DIR models is called Demon's algorithm developed by [3] in 1996. The displacement function $\mathbf{D} = (u_x, u_y)^T$ between the static and deformed images is derived from Navier nonlinear equation as given by

$$\mathbf{D} = \frac{(\mathbf{m} - \mathbf{s})\nabla\mathbf{s}}{|\nabla\mathbf{s}|^2 + (\mathbf{m} - \mathbf{s})^2}, \tag{1}$$

where $\mathbf{m}$ is moving image and $\mathbf{s}$ is the static image, $(\mathbf{m} - \mathbf{s})$ is the differential forces between the moving and deformed images. ∇ is the gradient operator of the static image defined by

$$\nabla\mathbf{s}(x, y) = \left(\frac{\partial\mathbf{s}(x,y)}{\partial x}, \frac{\partial\mathbf{s}(x,y)}{\partial y}\right).$$

The equation in (1) can be rearranged to a homogeneous PDE

$$\mathbf{D}\left(|\nabla\mathbf{s}|^2 + (\mathbf{m} - \mathbf{s})^2\right) - (\mathbf{m} - \mathbf{s})\nabla\mathbf{s} = \mathbf{0}$$

subject to the given initial conditions $\mathbf{s}^0 = \mathbf{0}$ and $\mathbf{m}^0 = \mathbf{0}$.

Cachier *et al* [4] in 1999 revised model to improves the registration convergence speed and stability. The revised equation is

$$\mathbf{D} = \frac{(\mathbf{m} - \mathbf{s})\nabla\mathbf{s}}{|\nabla\mathbf{s}|^2 + \xi^2\,|(\mathbf{m} - \mathbf{s})|^2} + \frac{(\mathbf{m} - \mathbf{s})\nabla\mathbf{m}}{|\nabla\mathbf{s}|^2 + \xi^2\,|(\mathbf{m} - \mathbf{s})|^2}. \tag{2}$$

The revised mode includes the image edge forces of the deformed image. The normalization factor ξ is added to adjust the force strengths. This attempts to normalize the relations between the moving and static images so as to improve the registration.

The classical demon's algorithm is an iterative finite difference scheme. The incremental displacement matrix $\mathbf{D}$ in equation (1) is simulated by equation (3)

$$\mathbf{D}^j(x, y) = \mathbf{D}^{j-1}(x, y) - \frac{(\mathbf{m}^{j-1}(x, y) - \mathbf{s}^0(x, y))\,\nabla\mathbf{s}^0}{|\nabla\mathbf{s}^0|^2 + [\mathbf{m}^{j-1}(x, y) - \mathbf{s}^0(x, y)]^2} \tag{3}$$

for the j^{th} iteration, $j = 1, 2, \ldots n$. The initial values of $\mathbf{D}^0$ and the initial moving image $\mathbf{m}^0$ and statics image $\mathbf{s}^0$ are given as

$$\begin{aligned}\mathbf{D}^0(x, y) &= \mathbf{0},\\ \mathbf{m}^0(x, y) &= \tilde{\mathbf{m}}^0(x, y),\\ \mathbf{s}^0(x, y) &= \tilde{\mathbf{s}}^0(x, y).\end{aligned}$$

The deformed image resulted at the j^{th} iteration can be determined by substituting $\mathbf{D}^{j-1}$

and $\mathbf{m}^{*j-1}$ into the following equation

$$\mathbf{m}^{*j}(x,y) = \mathbf{m}^{*j-1}(x,y) + \mathbf{D}^{j-1}(x,y)\nabla\mathbf{s}^0.$$

The spatial gradients $\nabla\mathbf{s}^0 = (\nabla^0\mathbf{s}_x, \nabla^0\mathbf{s}_y)$ is the coefficient matrices computed only once by using the original static image $\mathbf{s}^0$. The result of $\mathbf{D}^j(x,y)$ will be continuously updated until the estimated displacements reaches the following preset stopping criteria:

$$\left\|\mathbf{D}^j(x,y) - \mathbf{D}^{j-1}(x,y)\right\| < \varepsilon,$$

where ε is the upper bound of iterative error. The present study will use the results from the classical Demon's iterative algorithm as a reference guide for the radial kernel-based collocation algorithm.

Kernel-based RBF Collocation Method

This paper focuses on kernel approximations in the form of radial basis function. It will be used to solve the differential equation involved in the deformable image model. The method of kernel radial approximation method have been refined and diversified for facilitating the needs of various types of differential equations. The radial basis functions were originally devised for scattered geographical data interpolation by Hardy [5], who introduced a class of functions called multiquadric function in the early 1970's.

In this study, the basic idea of the radial kernel-basis interpolation is used to approximate an unknown displacement function $\{\mathbf{D}(\mathbf{x}) : \mathbf{x} \in \Omega\}$ by a RBF interpolant, say $\{\mathbf{s}(\mathbf{x}) : \mathbf{x} \in \Omega\}$ at a given set of N distinct nodal points $X = \{\mathbf{x}_i \in \Omega : i = 1, 2, \cdots, N\}$.

Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ be a set of positive definite radial basis functions defined by

$$\Phi = \{\phi\left(\|\mathbf{x} - \tilde{\mathbf{x}}_i\|\right)\} \quad \mathbf{x}, \tilde{\mathbf{x}}_i \in \Omega,$$

on a fixed space on Ω. Here ϕ refers to a specific choice of RBF functions that is solely dependent on the Euclidean distance $\|\mathbf{x} - \tilde{\mathbf{x}}_i\|$ between $\mathbf{x}$ and a fixed centre $\tilde{\mathbf{x}}_i \in \mathbb{R}^d$. A suitable choice of the function for $\{\mathbf{D}(\mathbf{x}_i), i = 1, 2, \cdots, N\}$ can ensure the interpolation smoothly passing through the given nodal points in X.

The chosen RBF interpolant for $\mathbf{D}$ can be expressed as a finite linear combination by the following equation

$$\mathbf{D}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \phi(\|\mathbf{x} - \tilde{\mathbf{x}}_i\|), \quad \mathbf{x} \text{ and } \tilde{\mathbf{x}}_i \in \Omega. \tag{4}$$

The unknown coefficients $\{\alpha_i : i = 1, 2, \cdots, N\}$ can be determined by collocating

$$\mathbf{s}(\mathbf{x}_i) = \mathbf{D}(\mathbf{x}_i), \text{ for } i = 1, 2, \ldots, N, \tag{5}$$

at a set of N distinct nodal points $\{\mathbf{x}_i \in \Omega, 1, 2, \cdots, N\}$. This yields a system of linear equations which can be expressed in the following matrix form

$$\mathbf{A}_\phi \boldsymbol{\alpha} = \mathbf{D}, \tag{6}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ are the unknown coefficients and

$$\mathbf{D} = [D(\mathbf{x}_1), D(\mathbf{x}_2), \ldots, D(\mathbf{x}_N)]^T.$$

Both $\boldsymbol{\alpha}$ and $\mathbf{D}$ are $N \times 1$ column matrices, and $\mathbf{A}_\phi = [\phi(\|\mathbf{x}_j - \tilde{\mathbf{x}}_i\|)]_{1 \le i,j \le N}$ is an $N \times N$ coefficient matrix.

Generally, the interpolation points in interior and boundaries are distinct and the chosen radial basis function $\phi \in \mathbb{R}^d$ is positive definite, the matrix $\mathbf{A}_\phi$ is always non-singular, so the linear system in (6) has a unique solution as proved by [6]. The unknown coefficients $\boldsymbol{\alpha}$ can then be obtained uniquely by solving the system of linear equations as

$$\boldsymbol{\alpha} = \mathbf{A}_\phi^{-1}\mathbf{D}.$$

The approximated displacement matrix $\mathbf{D}$ can be evaluated once the unknown coefficients $\{\alpha_i \mid i = 1, \cdots, N\}$ are found.

A classical theory on the existence, uniqueness and convergence of the RBF interpolation was proved by Micchelli [7] in 1986. Later, Powell [6] and Madych *et al* [8] extended the study and developed the important non-singularity properties of the RBF interpolation. Their analysis proved that the RBF interpolation methods hold a truly mesh-free algorithm and a super-convergent property. The accuracy of the RBF interpolant has an order of convergence $\mathcal{O}\left(h^{d+1}\right)$, where h is the density of the collocation points and d is the spatial dimension.

The most popular types of radial basis functions are listed below:

$$\phi(r_j) = \left\{ \begin{array}{lll} (r_j^2)\log r_j, & \text{Thin plate splines in } \mathbb{R}^2 & (a) \\ e^{-\sigma r_j^2}, & \text{Gaussian, } \sigma > 0 & (b) \\ (r_j^2 + \delta^2)^{\frac{1}{2}}, & \text{Multiquadric, } \delta \in \mathbb{R} & (c) \\ (r_j^2 + \delta^2)^{-\frac{1}{2}}, & \text{Reciprocal multiquadric, } \delta \in \mathbb{R} & (d) \end{array} \right\} \tag{7}$$

where $\{r_j = \|\mathbf{x} - \mathbf{x}_j\|,\ j = 1, 2, \cdots, N\}$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_j \in \mathbb{R}^d$, and $\delta^2 \in \mathbb{R}$ is called the shape parameter of the multiquadric functions in (c) & (d). This shape parameter uses to control the fitting of a smooth surface to the data.

To avoid having singularity, the radial kernel approximation is formulated by adding a finite polynomials $\{q_k(\mathbf{x}), k = 1, 2, \cdots, M\}$ into the interpolation system in (4). The RBFs interpolant $\mathbf{D}(\mathbf{x})$ in (4) is extended to the following equation:

$$\mathbf{D}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \phi(\|\mathbf{x} - \tilde{\mathbf{x}}_i\|) + \sum_{k=1}^{M} b_k q_k(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2, \quad 0 \le M < N, \tag{8}$$

where $\phi(\|\mathbf{x} - \tilde{\mathbf{x}}_i\|)$ is a class of chosen radial kernel functions, $\{\alpha_i\}$ and $\{b_k\}$ are the coefficients to be determined. Given a set of N distinct nodes $X = \{\mathbf{x}_i \in \Omega, i = 1, 2, \cdots, N\} \subseteq \mathbb{R}^d$, the approximation function in (8) will produce a unique solution if the system satisfies the following condition

$$\mathbf{s}(\mathbf{x}_i) = \mathbf{D}(\mathbf{x}_i), \quad i = 1, 2, \cdots, N \tag{9}$$

and the constraints

$$\sum_{i=i}^{N} \alpha_i q_k(\mathbf{x}) = 0, \quad k = 1, 2, \cdots, M \text{ and } i = 1, 2, \cdots, N.$$

The resulting system can be written concisely in matrix form,

$$\begin{bmatrix} \mathbf{A}_\phi & \mathbf{Q} \\ \mathbf{Q}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix} \tag{10}$$

where $\mathbf{A}_\phi = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$ is a square matrix , $\mathbf{a}$ and $\mathbf{c}$ are column vectors. $\mathbf{Q} = [q_k(\mathbf{x}_i)]$ is an $N \times M$ matrix and the unknown coefficients $[\mathbf{b}]$ is an $M \times 1$ matrix:

$$\mathbf{Q} = \begin{bmatrix} q_1(\mathbf{x}_1) & q_2(\mathbf{x}_1) & \cdots & q_M(\mathbf{x}_1) \\ q_1(\mathbf{x}_2) & q_2(\mathbf{x}_2) & \cdots & q_M(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ q_1(\mathbf{x}_N) & q_2(\mathbf{x}_N) & \cdots & q_M(\mathbf{x}_N) \end{bmatrix}, \quad [\mathbf{b}] = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}.$$

The interpolation problem in equation (9) is solvable if the matrix of this system is

$$[\tilde{\Phi}] = \begin{bmatrix} \mathbf{A}_\phi & \mathbf{Q} \\ \mathbf{Q}^T & 0 \end{bmatrix}$$

is non-singular. In the application of deformable image registration models, the concerned displacement matrix $\mathbf{D}$ can then be determined by the above basis function subject to the given initial conditions for $\mathbf{D}$, $\mathbf{m}$ and $\mathbf{s}$**:**

$$\begin{aligned} \mathbf{D}^0(x, y) &= \mathbf{0}, \\ \mathbf{m}^0(x, y) &= \tilde{\mathbf{m}}(x, y), \\ \mathbf{s}^0(x, y) &= \tilde{\mathbf{s}}^0(x, y). \end{aligned}$$

We choose one of the radial kernel functions listed in (7) for the interpolant $\mathbf{D}(\mathbf{x})$ in (8). From our numerical experiences presented in the paper [9] , using multiquadric function $\phi(r_j) = (r_j^2 + \delta^2)^{\frac{1}{2}}$ usually results in a higher degree of accuracy and stability in similar studies. According to the classic literature written by Micchelli's [7] in 1986, it have been shown that the multiquadric interpolation is always solvable for distinct dataset, owing to the fact that multiquadric possess an exponentially convergent property.

We use the Demon algorithm (3) in the image registration process but use the radial kernel interpolation instead of using the finite difference method. The Demon algorithm radial kernel-based interpolation for the displacement matrix $\mathbf{D}$ is formulated as

$$\mathbf{D}^j(x, y) = \sum_{i=1}^{N} \alpha_i^j \phi(\|\mathbf{r}\|) + \sum_{k=1}^{M} b_k q_k^j(\mathbf{x})$$

is the displacement at j^{th} iteration, and

$$\mathbf{D}^j(x, y) = \sum_{i=1}^{N} \alpha_i^{j-1} \phi(||\mathbf{r}||) + \sum_{k=1}^{M} b_k q_k^{j-1}(\mathbf{x}) - \frac{([\mathbf{m}^{j-1}(x, y) - \mathbf{s}^0(x, y)]\, \nabla \mathbf{s}(x, y)}{\|\nabla \mathbf{s}^0(x, y)\|^2 + [\mathbf{m}^{j-1}(x, y) - \mathbf{s}^0(x, y)]^2}, \tag{11}$$

where $\|\mathbf{r}\| = \|\mathbf{x} - \widetilde{\mathbf{x}}_i\|$. In order to determine the deformable image $\mathbf{m}^j$ at the j^{th} iteration, the forward iterative scheme is applied according to the following equation

$$\mathbf{m}^{*j}(x, y) = \mathbf{m}^{*j-1}(x, y) + \mathbf{D}^{j-1}(x, y) \nabla \mathbf{s}^0.$$

The required stopping criteria is

$$\left\|\mathbf{D}^{j}(x,y) - \mathbf{D}^{j-1}(x,y)\right\| < \varepsilon,$$

where ε is the preset upper bound of iterative error.

Conclusions

In this report, we developed a radial basis function Demon's scheme to approximate the deformation image registration in medical fields. The proposed model combined the RBF scheme and the classical demon's algorithm. The method was applied to simulate a 2D deformable image registration in tracing the growth of tumor of liver cancer. The numerical investigations and the performance using this kernel-based collocation. Further investigations on the computational efficiency and accuracy as well as support sizes of the kernel-based RBFs will be explored and discussed in the future report.

Acknowledgements

The development of the research material was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/IDS16/14).

References

[1] E. M. V. Osorio, M. S. Hoogeman, L. Bondar, P. C. Levendag, and B. J. M. Heijmen, "A novel flexible framework with automatic feature correspondence optimization for non-rigid registration in radiotherapy". Medical Physics, 36(7): 2848-2859, 2009.

[2] Broit, C. Optimal Registration of Deformed Images. PhD thesis, University of Pennsylvania, 1981.

[3] Thirion J. P. Non-rigid matching using demons. Proc. Conf. Computer Vision and Pattern Recognition, pp. 245-251, June 1996.

[4] X. Pennec, P. Cachier, and N. Ayache, Understanding the Demon's algorithm: 3D non-rigid registration by gradient descent, Proc. of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '99), pp. 597–605.

[5] R. L. Hardy, Multiquadric equations of topography and other irregular surfaces, J. Geophys, Res, 176 (1971) 1905-1915.

[6] M. J. D. Powell, The theory of radial basis functions approximations in 1990, Chapter 3, Wavelets, subdivision algorithms and radial basis functions, ed. Will Light, Vol. II, Oxford University Press, (1990) pp. 105-210.

[7] C. A. Micchelle, "Interpolation of scattering data: distance matrices and conditionally positive definite functions", Constr. Approx. 2 (1986) pp. 11-22.

[8] W. R. Madych and S. A. Nelson, "Multivariate interpolation and conditionally positive definite functions", Approx. Theory Appl., Vol. 4 (1988) pp. 77-89.

[9] S. M. Wong, K. S. NG and T. S. Li, "Numerical Solution for Deformable Image Registration Using Radial Basis Functions", accepted by Dynamics of Continuous, Discrete and Impulsive Systems in Apirl 2016.

Optimization of stiffened composite plate using adjusted different evolution algorithm

†Thuan Lam-Phat[1,*], Son Nguyen-Hoai[1], Vinh Ho-Huu[2], Trung Nguyen-Thoi[2]

[1] GACES, HCMC University of Technology and Education, Vietnam
[2] Division of Computational Mathematics and Engineering (CME), Institute of Computational Science (INCOS), Ton Duc Thang University, Hochiminh City, Viet Nam

*Presenting author: lamphatthuan@gmail.com
†Corresponding author(*): *lamphatthuan@gmail.com*

Abstract

Stiffened composite plate has been widely used in many braches of engineering area and the demand of optimizing the cost of manufacturing is also very high. One of many approaches to minimize the cost is to optimize the weight of the structure. In this paper, an improved version of the Differential Evolution (DE) algorithm is adopted to solve for suitable values of the fiber angle and the thickness of the stiffened composite plate to achieve the structure with minimum weight. For computing the constrained conditions of stress and strain in the optimization process, the finite element analysis using the CS-DSG3 element is used. To verify the accuracy and the effectiveness of the algorithm, the numerical solutions obtained from the proposed method are compared with those of other available approaches.

Keywords: *Stiffened composite plate, Differential Evolution (DE), Cell-based smoothed discrete shear gap method (CS-DSG3), Optimization analysis.*

1. Introduction

Nowadays, stiffened composite plates have been widely used in many branches of structural engineering such as aircraft, ships, bridges, buildings, etc. For its advantages in both bending stiffness and the amount of material in comparison with common bending plate structures, stiffened composite plate usually has higher economic efficiency in practical applications. However, choosing the best design that satisfies the working requirement is difficult. In addition, the complex mechanical behavior of composite materials also increases the difficulty of the problems related to their design [1]. In this case, the design optimization tools combined with numerical methods must be utilized. Design optimization is one of the most interesting research directions that brings a lots of profits in both life and industry. And so, methods for design optimization are also quickly developed. The optimization methods can be classified into two main groups: gradient-based and popular-based approach. Methods based on gradient information is fast but usually stuck in local solution and depended too much on a good initial point to obtain global optimal solution. T. Nguyen-Thoi et al [2] used SQP to find the optimal fiber orientations for stiffened composite plate, but the results still depend on the initial point to get the exact solution. To deal with such disadvantages, population-based global optimization methods are utilized alternatively. Marin et al. [3] used the genetic algorithm, including the application of elitism, which preserved the use of the Pareto front to optimize the design of a composite material-stiffened panel. Falzon and Faggiani [4] applied the genetic algorithm to improve the post-buckling strength of stiffened composite panels. And among many proposed global optimization algorithms, Differential Evolution (DE) firstly introduced by Storn and Price in 1997 [5] was one of the most potential algorithms. The DE has demonstrated excellently performance in solving many different engineering problems. Wang et al. [6] applied the DE for designing optimal truss structures with continuous and discrete variables. Wu and Tseng [7] applied a multi-population differential evolution with a penalty-based, self-adaptive strategy to solve the COP of the truss structures. Le-Anh et al. [8] using an adjusted Differential Evolution algorithm and a smoothed triangular plate element for static and frequency optimization of folded laminated composite plates. Ho-Huu et al. [9] proposed a new version of the DE to optimize the shape and size of truss with discrete variables. However, using the method in finding the global optimum

solution still gets highly computational cost. Hence, many approaches have been proposed to increase the effectiveness of the algorithm. Most recently, Ho-Huu et al. [10] also introduced two new improvement steps to increase the convergence of DE algorithm based on roulette wheel selection (ReDE). The new modified DE algorithm is applied for solving shape-and-size optimization problem of truss structure with frequency constraints and show its high effectiveness.

In this paper, this new improved version of the Differential Evolution (ReDE) algorithm is adopted to solve for suitable values of the fiber angle and the thickness of the stiffened composite plate to achieve the structure with minimum weight. For computing the constrained conditions of stress and train in the optimization process, the finite element analysis using the cell-based smoothed discrete shear gap technique with triangular elements (CS-DSG3) proposed by T. Nguyen-Thoi et al [11,12] is used. The numerical solutions obtained from the method are compared with references to show the effectiveness and the accuracy of the algorithm.

2. Theory Fundamental

An optimization problem can be expressed as follows:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \begin{cases} h_i(\mathbf{x}) = 0 & i = 1,...,l \\ g_j(\mathbf{x}) \le 0 & j = 1,...,m \end{cases} \tag{1}$$

where $\mathbf{x}$ is the vector of design variables; $h_i(\mathbf{x}) = 0$ and $g_j(\mathbf{x}) \le 0$ are inequality and equality constraints; *l, m* are the number of inequality and equality constraints, respectively; $f(\mathbf{x})$ is the objective function which can be the function of weight, cost, etc.

Design optimization of a structure is to find optimal values of design variables in design space such that the objective function is minimum [2]. Dealing with such problems, many optimization methods are used including gradient-based and population-based approach to find the solution. In this paper, the Differential Evolution is utilized to solve the problem of finding optimal fiber orientations and thickness of the stiffened composite plate.

2.1 Brief on the differential evolution algorithm [10,9]

The original differential evolution algorithm firstly proposed by Storn and Price [5] has been widely used to solve many kinds of optimization problems. The scheme of this algorithm consists of four main phases as follows:

Phase 1: Initialization
Create an initial population by randomly sampling from the search space
Phase 2: Mutation
Generate a new mutant vector v_i from each current individual x_i based on mutation operations.
Phase 3: Crossover
Create a trial vector u_i by replacing some elements of the mutant vector v_i via crossover operation.
Phase 4: Selection
Compare the trial vector u_i with the target vector x_i. One with lower objective function value will survive in the next generation

To improve the effectiveness of the algorithm, the *Mutation phase* and the *Selection phase* are modified to increase the convergence rate as follow:

In the *mutation phase*, parent vectors are chosen randomly from the current population. This may make the DE be slow at exploitation of the solution. Therefore, the individuals participating in mutation should be chosen following a priority based on their fitness. By doing this, good information of parents in offspring will be stored for later use, and hence will help to increase the convergence speed. To store good information in offspring populations, the individuals is chosen based on Roulette wheel selection via acceptant stochastic proposed by Lipowski and Lipowska [13] instead of the random selection.

In the *selection phase*, the elitist operator introduced by Padhye et al. [14] is used for the selection progress instead of basic selection as in the conventional DE. In the elitist process, the children population C consisting of trial vectors is combined with parent population P of target vectors to create a combined population Q. Then, best individuals are chosen from the combined population Q to construct the population for the next generation. By doing so, the best individuals of the whole population are always saved for the next generation.
The modified algorithm Roulette-wheel-Elitist Differential Evolution is then expressed as below:

1:	Generate the initial population
2:	Evaluate the fitness for each individual in the population
3:	while <the stop criterion is not met> do
4:	Calculate the selection probability for each individual
5:	for i =1 to *NP* do {*NP*: Size of population}
6:	Do mutation phase based on Roulette wheel selection
7:	j_{rand} = randi(1,D) {D: number of design variables}
8:	for j =1 to D do
9:	if rand[0,1] < CR or j == j_{rand} then {CR: crossover control parameter}
10:	$u_{i,j} = x_{r1,j} + Fx(x_{r2,j} - x_{r3,j})$ {*F*:randomly chosen within [0,1] interval}
11:	else
12:	$u_{i,j} = x_{i,j}$
13:	end if
14:	end for
15:	Evaluate the trial vector u_i
16:	end for
17:	Do selection phase based on Elitist selection operator
18:	end while

2.2 Brief on the behavior equation of stiffened composite plate [2]

Stiffened composite plate can be seen as the combination between composite plate elements and the stiffening Timoshenko composite beam elements, as illustrated in Figure 1. The stiffening composite beam is set parallel with the axes in the surface of plate and the centroid of beam has a distance *e* from the middle plane of the plate. The plate-beam system is discretized by a set of node. The degree of freedom (DOF) of each node of the plate is $\mathbf{d} = [u, v, w, \beta_x, \beta_y]^T$, in which u, v, w are the displacements at the middle of the plate and β_x, β_y are the rotations around the *y*-axis and *x*-axis. The DOF of each node of the beam is $\mathbf{d}_{st} = [u_r, u_s, u_z, \beta_r, \beta_s]^T$. The centroid displacements of beam are expressed as

$$u = u_r(r) + z\beta_r(r) \quad ; \quad v = z\beta_s(r) \quad ; \quad w = u_z(r) \tag{2}$$

where u_r, u_s, u_z are respectively centroid displacements of beam and β_r, β_s are the rotations of beam around *r*-axis and *s*-axis.

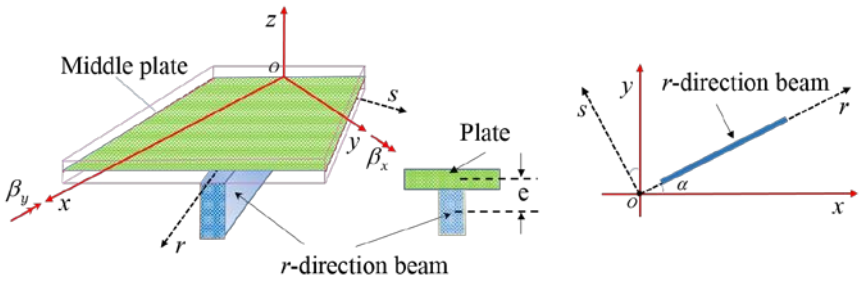


Figure 1. A plate composite stiffened by an *r*-direction stiffener

* Energy equation of stiffened composite plates

The strain energy of composite plate is given by

$$U_P = \frac{1}{2}\iint_A \left(\boldsymbol{\varepsilon}_0^T \mathbf{D}^m \boldsymbol{\varepsilon}_0 + \boldsymbol{\varepsilon}_0^T \mathbf{D}^{mb} \boldsymbol{\kappa}_b + \boldsymbol{\kappa}_b^T \mathbf{D}^{mb} \boldsymbol{\varepsilon}_0 + \boldsymbol{\kappa}_b^T \mathbf{D}^b \boldsymbol{\kappa}_b + \boldsymbol{\gamma}^T \mathbf{D}^s \boldsymbol{\gamma}\right) \mathrm{d}A \tag{3}$$

where $\boldsymbol{\varepsilon}_0, \boldsymbol{\kappa}_b, \boldsymbol{\gamma}$ are respectively membrane, bending and shear strains of composite plate and are expressed as follows

$$\boldsymbol{\varepsilon}_0 = [u_{,x}, v_{,y}, u_{,y} + v_{,x}]^T; \boldsymbol{\kappa}_b = [\beta_{x,x}, \beta_{y,y}, \beta_{x,y} + \beta_{y,x}]^T; \boldsymbol{\gamma} = [w_{,x} + \beta_x, w_{,y} + \beta_y]^T. \tag{4}$$

$\mathbf{D}^m, \mathbf{D}^{mb}, \mathbf{D}^b, \mathbf{D}^s$ are material matrices of plate

The strain energy of composite stiffener is given by

$$U_{st} = \tfrac{1}{2}\int_l \left((\boldsymbol{\varepsilon}_{st}^b)^T \mathbf{D}_{st}^b \boldsymbol{\varepsilon}_{st}^b + (\boldsymbol{\varepsilon}_{st}^s)^T \mathbf{D}_{st}^s \boldsymbol{\varepsilon}_{st}^s\right) \mathrm{d}x \tag{5}$$

where $\boldsymbol{\varepsilon}_{st}^b, \boldsymbol{\varepsilon}_{st}^s$ are respectively bending, shear strain of beam and are expressed as follows

$$\boldsymbol{\varepsilon}_{st}^b = [u_{r,r} + z_0 \beta_{r,r}, \beta_{r,r}, \beta_{s,r}]^T; \boldsymbol{\varepsilon}_{st}^s = [u_{z,r} + \beta_r]^T \tag{6}$$

$\mathbf{D}_{st}^b, \mathbf{D}_{st}^s$ are material matrices of composite beam

Using the superposition principle, total energy strain of stiffened composite plate is obtained by

$$U = U_P + \sum_{i=1}^{N_{si}} U_{st} \tag{7}$$

where N_{st} is the number of stiffeners.

For static analysis, the global equations for the stiffened composite plate $[\mathbf{K}]\{\Delta\} = \{\mathbf{F}\}$ can found in [16] for detail.

3. Numerical Results

3.1 Unconstrained problem for fiber angle optimization

Consider an optimization analysis of a composite plate stiffened by a composite beam according to *x*-direction as in Figure 2 under simply-supported condition. The parameters of the problem are given by a = 254 mm, h = 12.7 mm, c_x = 6.35 mm and d_x = 25.4 mm. The analysis is carried out with two cases of square (b = 254 mm) and rectangular (b = 508 mm) plate.

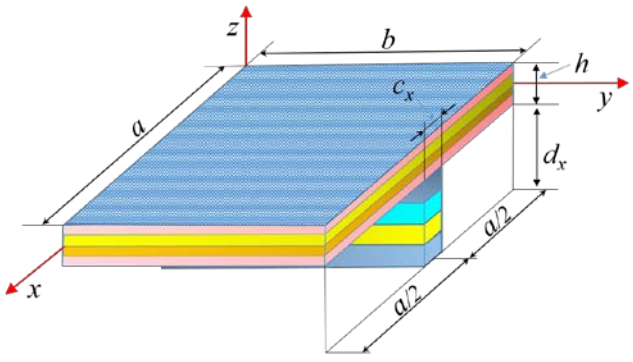


Figure 2. Model of a stiffened composite plate

Both plate and beam have four symmetric layers. The fiber orientation for layers of the plate is a set [θ_1 θ_2 θ_2 θ_1], and for the layers of the beam is [θ_3 θ_4 θ_4 θ_3]. The plate and beam are made by the same materials with $E_1 = 144.8\text{GPa}$, $E_2 = E_3 = 9.65\text{GPa}$, $G_{12} = G_{13} = 4.14\text{GPa}$, $G_{23} = 3.45\text{GPa}$, $\upsilon_{12} = \upsilon_{13} = \upsilon_{23} = 0.3$. The plate is subject to a uniform load f = 0.6895 (N/mm^2).

The optimization problem is now expressed as:

$$\begin{cases} \min\limits_{\boldsymbol{\theta}} & \mathbf{U} = \dfrac{1}{2}\mathbf{d}^T\mathbf{K}\mathbf{d} \\ \text{subject to} & 0 \le \theta_i \le 180, \quad i = 1,...,4 \end{cases}$$

where U is strain energy and θ_I is fiber orientation of *i*th layer.

Firstly, static analysis for the case of square plate is carried out to verify the reliability of the finite element solution using CS-DSG3 [15]. The results compared with those by Li Li [17] and M. Kolli [16] are presented in Table 1 and show good agreement.

Table 1. Comparison of central deflection (mm) of the simply-supported square stiffened composite plates subjected to a uniform load f = 0.6895 N/mm^2

Orientation angle for both beam and plate	$[0^0/90^0/90^0/0^0]$			$[45^0/-45^0/-45^0/45^0]$	
Method	CS-DSG3	[16]	[17]	CS-DSG3	[11]
Central deflection	1.0917	1.0396	1.0892	2.5049	2.4912

In Table 2, a comparison of different types of DE algorithm for the case of rectangular plate is presented. The first two versions are the original different evolution (DE) and the adjusted one (ReDE). Both of them are used with continuous variables. We can see that, the difference of computational cost between the two versions is rather big. The cost from DE is nearly double in comparison with that of ReDE algorithm. The third version is ReDE algorithm with integer variables. And the result obtained from this type just equals 43% of that of ReDE with continuous variables. However, the values of the solution are still nearly the same. Therefore, in this paper, the ReDE with integer variables is utilized for the optimization process for saving the cost.

Table 2. Comparison of different types of DE

Type of stiffened plate	Method	Optimal angle [Degree]				Strain energy (N.m)	Computational cost (seconds)
		θ_1	θ_2	θ_3	θ_4		
Rectangular (a = 254 mm, b = 508 mm)	DE	159.2	37	0	179.9	30300	11223
	ReDE	159.2	37	0	179.9	30300	6787
	Int_ReDE	160	37	0	180	30366	2851

Next, the optimization analysis for two cases of square plate and rectangular plate is carried out. The results of fiber orientations obtained from ReDE are presented in Table 3. In this analysis, the value of the design variables is chosen to be integer for saving the time of computing. The results from the Table 3 show that the solutions by the DE agree very well with those by the GA. However, the computational cost for the case of square plate with the mesh size of 20x20 is less than 188 seconds. And in the case of rectangular plate with the mesh size of 20x40, the cost from GA is nearly double in comparison with the one from DE. This shows a big difference and proves the effectiveness of the proposed method.

It is also seen that the optimal fiber orientations of the square plate problem are quite different from those of the rectangular plate case under the same conditions. This implies that the geometric parameters of the structures also have influence to the results of the optimization problems.

Table 3. The optimal results of two problems

Type of stiffened plate	Method	Optimal angle [Degree]				Strain energy (N.m)	Computational cost (seconds)
		θ_1	θ_2	θ_3	θ_4		
Square ($a = b$ = 254 mm)	ReDE	135	48	0	180	6183.2	2065
	GA	135	48	0	180	6183.1	2253
Rectangular (a = 254 mm, b = 508 mm)	ReDE	160	37	0	180	30366	2851
	GA	159	37	0	180	30300	4995

3.2 Constrained problem with thickness optimization

Consider the same composite plate stiffened by a composite beam according to *x*-direction as in Figure 2 under simply-supported condition. But in this case, the fiber orientations for layers of the plate and the beam are given. The problem here is to find the optimal thickness of the plate (t_p) and the beam (t_b) to minimize the weight of the stiffened composite plate under the constraints of displacement and stress. The analysis is also carried out with two cases of square and rectangular plate. For both cases, the optimal fiber angles found in the above unconstrained problems are used, respectively. In particular, the fiber angles of [*135 48 0 180*] is used for the square plate case and the fiber angles of [*160 37 0 180*] is used for the rectangular plate.

For composite materials, many failure criteria proposed to predict lamina failure. In this paper, the Tsai-Wu index defined below is used to predict the most likely failure point in a layer.

$$S_{tw} = \frac{\sigma_{11}^2}{X_t X_c} + \frac{\sigma_{22}^2}{Y_t Y_c} + \frac{\tau_{12}^2}{S^2} - \frac{\sigma_{11}\sigma_{22}}{\sqrt{X_t X_c Y_t Y_c}} + \left(\frac{1}{X_t} - \frac{1}{X_c}\right)\sigma_{11} + \left(\frac{1}{Y_t} - \frac{1}{Y_c}\right)\sigma_{22} \le 1 \tag{8}$$

The point with the highest Tsai-Wu index is the point that will most likely fail. And this is considered as the stress constraint in this problem.
The optimization problem is then expressed as

$$\begin{cases} \min\limits_{t_p, t_b} & \text{Weight}(t_p, t_b) \\ \text{subject to} & \text{Displacement is less than 1 mm} \\ & S_{tw} \le 1 \end{cases}$$

Table 4. The optimal results of two problems

Type of stiffened plate	Method	Optimal thickness		Weight (kg)	Computational cost (seconds)
		t_p	t_b		
Square ($a = b$ = 254 mm)	ReDE	13	83	1.5269	1065
	GA	13	83		3659
Rectangular (a = 254 mm, b = 508 mm)	ReDE	18	20	4.6593	2606
	GA	18	20		7482

The results from the Table 4 show that the solutions by the ReDE agree very well with those by the GA. The objective function is almost the same but the computational costs from GA

are about 3 times bigger. This shows that the effectiveness of ReDE in comparison with GA is much better.

It is also seen that the optimal thicknesses of the square plate are quite different from those of the rectangular plate under the same conditions. In the case of square plate, when the thickness of the plate decreases about 27% (from 18 to 13), the thickness of the stiffened beam increases 4 times (from 20 to 83). This implies that the thickness of the stiffened beam has not too much influence to the response of the whole structure as of the thickness of the plate. Therefore, in the problem of weight optimization, we can adjust the thickness of the plate and focus only on optimizing the thickness of the plate for saving the cost.

4. Conclusion

In this paper, the unconstrained and constrained optimization analysis with integer variables for the stiffened composite plate using new modified version of DE is presented. In both problems, the results obtained are agreed well with those of GA. However, the computational cost of ReDE algorithm is much cheaper than the one from GA. The results illustrated the efficiency and the accuracy of the adjusted Differential Evolution in solving the optimization problem of the stiffened composite plate.

References

[1] Shun-Fa, H., Ya-Chu H., & Yuder, C,. (2014) A genetic algorithm for the optimization of fiber angles in composite laminates. *Journal of Mechanical Science and Technology*, Vol. 28 (8), p. 3163-3169

[2] Nguyen-Thoi, T., Ho-Huu, V., Dang-Trung, H., Bui-Xuan, T., Lam-Phat, T. (2013) Optimization analysis of stiffened composite plate by sequential quadratic programming. *Journal of Science and Technology*, Vol. 51(4B), p. 156-165.

[3] Marin, L., Trias, D., Badallo, P., Rus, G., & Mayugo, J. A. (2012) Optimization of composite stiffened panels under mechanical and hygrothermal loads using neural networks and genetic algorithms. *Composite Structures*, Vol. 94, p. 3321-3326.

[4] Falzon, B. G., & Faggiani, A. (2012) The use of a genetic algorithm to improve the postbuckling strength of stiffened composite panels susceptible to secondary instabilities, *Composite Structures*, Vol. 94, p. 883-895.

[5] Storn, R., Price, K. (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, *J. Glob. Optim*, p. 341–359. doi:10.1023/A:1008202821328.

[6] Z.W.Z. Wang, H.T.H. Tang, P.L.P. Li. (2009) Optimum Design of Truss Structures Based on Differential Evolution Strategy, 2009 Int. Conf. Inf. Eng. Comput. Sci. 0–4. doi:10.1109/ICIECS.2009.5365996.

[7] C.-Y. Wu, K.-Y. Tseng (2010) Truss structure optimization using adaptive multi-population differential evolution, Struct. Multidiscip. Optim. 42 (2010) 575–590. doi:10.1007/s00158-010-0507-9.

[8] Le-Anh, L., Nguyen-Thoi, T., Ho-Huu, V., Dang-Trung, H., & Bui-Xuan, T. (2015) Static and frequency optimization of folded laminated composite plates using an adjusted Differential Evolution algorithm and a smoothed triangular plate element. *Compos. Struct*, Vol. 127, p. 382–394. doi:10.1016/j.compstruct.2015.02.069.

[9] Ho-Huu, V., Nguyen-Thoi, T., Nguyen-Thoi, M. H., Le-Anh, L. (2015) An improved constrained differential evolution using discrete variables (D-ICDE) for layout optimization of truss structures, *Expert Syst. Appl*, doi:10.1016/j.eswa.2015.04.072.

[10] Ho-Huu, V., Nguyen-Thoi, T., Khac-Truong, T., Le-Anh, L., Nguyen-Thoi, M. H. (2015) A fast efficient differential evolution based on roulette wheel selection for shape and sizing optimization of truss with frequency constraints.

[11] Liu GR, Nguyen Thoi Trung. (2010) *Smoothed Finite Element Methods*. NewYork: CRC Press, Taylor and Francis Group.

[12] Bui-Xuan, T., Nguyen-Thoi, T., Pham-Duc, T., Phung-Van, P., & Ngo-Thanh, P. (2012) An analysis of eccentrically stiffened plates by CS-FEM-DSG3 using triangular elements. *The international conference on advances in computational mechanics*, 629-643.

[13] Lipowski, A., & Lipowska, A. (2012) Roulette-wheel selection via stochastic acceptance, *Physica A*, Vol. 391, p. 2193–2196. doi:10.1016/j.physa.2011.12.004.

[14] Padhye, N., Bhardawaj, P., & Deb, K. (2013) Improving differential evolution through a unified approach, *J. Glob. Optim*, Vol. 55, p.771–799. doi:10.1007/s10898-012-9897-0.

[15] Nguyen-Thoi, T., Phung-Van, P., Nguyen-Xuan, H., Thai-Hoang, C. (2012) A cell-based smoothed discrete shear gap method (CS-DSG3) using triangular elements for static and free vibration analyses of Reissner-Mindlin plates. *International Journal for Numerical Methods in Engineering*, p. 705-741.

[16] Kolli, M., & Chandrashekhara. K. (1996) Finite element analysis of stiffened laminated plates under transverse loading. *Composites Science and Technology 56*, p. 1355-1361.
[17] Li Li, Ren Xiaohui. (2010) Stiffened plate bending analysis in terms of refined triangular laminated plate element. *Composite Structures*, p. 2936-2945.
[18] Javidrad, F., & Nouri, R. (2011) A simulated annealing method for design of laminates with required stiffness properties. *Composite Structures*, Vol. 93, p. 1127-1135.

Seismic resistance for high-rise buildings using water tanks considering the liquid - tank wall interaction

†*Bui Pham Duc Tuong[1], Phan Duc Huynh[1], Son Nguyen-Hoang[2,3]

[1]Department of Civil and Applied Mechanic, University of Technical and Education, HCMC, Vietnam.

[2]CIRTech, Ho Chi Minh City University of Technology (HUTECH), Vietnam.

[3]Department of Mechanical & Automotive Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul, South Korea.

†Presenting author: tuongbpd@hcmute.edu.vn

*Corresponding author: tuongbpd@hcmute.edu.vn

Abstract

In recent years, considerable attentions have been paid to research for the development of structural control devices, particularly focusing on the mitigation of wind and seismic's effects. The use of water tanks at roof as resistant solutions, which are known as Tuned Liquid Dampers (TLD), for high-rise buildings is considered in this paper. In the literature, TLD has shown significant advantages and can be one of excellent methods to control high-rise building's vibration. Liquid storage tank is designed to achieve its natural frequency same as that of the building. As a result, the resonant phenomenon will occur and contribute to the building's balance.

Besides using TLD to analyze the seismic resistance for high-rise buildings, this paper is also considered the interaction between the liquid and tank wall for with/without using water tank at roof as seismic resistance devices. Results showed that the maximum displacements at the top of buildings can be decreased from 50% to 80% and internal stresses are also reduced meaningfully.

Keywords: Tuned Liquid Dampers, sloshing, dynamic control, finite element method, liquid-structure interaction.

Introduction

In general, TLD is a tank with a part of liquid inside relied on liquid sloshing to dissipate vibration energy (Figure 1, 2). In fact, the liquid is employed to provide all of the necessary characteristics of a secondary system. Meanwhile, its gravity provides the required restoring mechanism. Therefore, the secondary system has characteristic periods that can be tuned for optimal performance, in the same way as a tuned mass damper (TMD). TLD is a passive mechanical damper and has been used in marine for centuries, and in the 20th it was applied in aerospace. The advantages of TLD are low cost, easy install & maintenance and the most advantage is that it can apply for almost kind of structure including existing building or tower.

A liquid storage tank on a fixed offshore platform was first used as a TLD to suppress the wind-induced vibration of the platform structure by Vandiver et al. (1979) [17], and was shown to be effective. Yozo Fujino (1989) [7] was one of the first researching TLD's wind resistance with full scale testing. Sun LM (1992) [8] analyzed TLD's capacity under wind and earthquake by theory and compared with experiment but in his research, the amplitude of liquid sloshing is not large. Bui Thanh Tam (1997) [2], showed that TLD can reduced 60% of the vibration of structure by using finite element method (FEM). Dorothy Reed (1998) [15] published his research about TLD under very large liquid sloshing's amplitude and Jin Kyu Yu (1999) [5][6] modeled TLD as an equivalent TMD with non-linear stiffness and damping.

In mechanical, civil and aerospace engineering, fluid-structure interactions with a moving free surface can have significant influence on the dynamic behavior of the structure and needs to be properly taken into account. Large amounts of work deal with linearization on the free surface, or other simplifications because of the complexity of the problem. However, for fluid-structure interactions including large scale sloshing motion of the fluid and large displacement motion of the structure, advanced theory and numerical methods are required [5] [6]. In recent years, advances have been made in this respect. In some circumstances, especially when the deformations of the container are small compared to its displacements, it is reasonable to simplify the structure as a rigid container supported by a system consisting of elastic springs and dashpots. Typical situations include TLD to passively suppress vibrations of high-rise buildings or towers of cable-suspension bridges, and liquid-loaded vehicle systems. In these fluid structure interaction problems involving large-amplitude sloshing, the nonlinear characters caused by the free surface motion and the dynamic boundary conditions need to be considered. The nonlinearity can also be inherited from the dynamics of the structure.

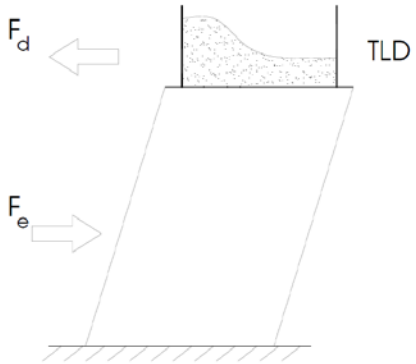


Figure 1. Mechanism of building with TLD

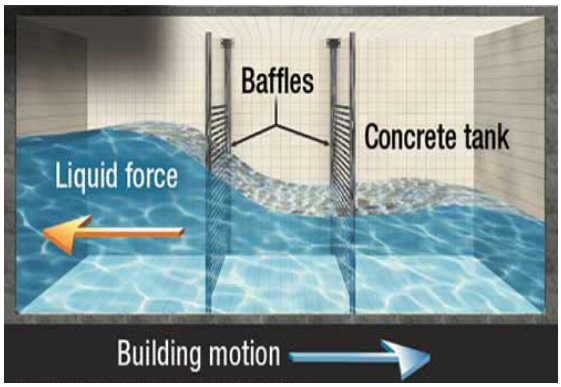


Figure 2. Inside a liquid tank as TLD

All of previous TLD's design assumed that the tank's wall or TLD is rigid to ignore tank's flexibility in that the complicate at the liquid-tank interface. In fact, there are many tank's failures (Figure 3, 4) because of this assumption so that it is attracted researchers and engineers in the last few years (Praveen K. Malhotra *et al*. (2000)[5]; M. Gradinscak (2009)[10] to study. Andersson (2001)[16] first investigated the possibility of using container flexibility for control of liquid sloshing. Recently, M. Gradinscak (2009)[10] presented that flexible container partially filled with water, as the sloshing absorber, and it can be advantages over a rigid container for effective control. However, in this study the building is modeled as a single degree of freedom, and the mass ratio of TLD over structure is 10% (this ratio is too much to practically apply especially in high rise building).

Figure 3. Sloshing damage to upper shell of tank (courtesy of UC Berkeley)

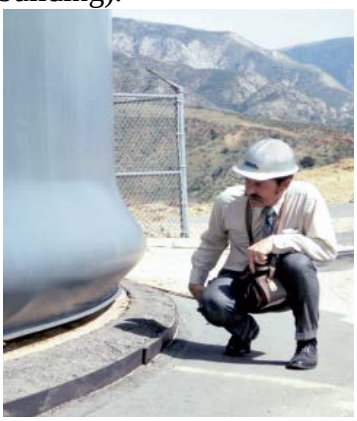


Figure 4. Elephant-foot buckling of tank wall (courtesy of UC Berkeley)

The work in this paper is to analyze building under harmonic load and earthquake with and without TLD. Addition, the effects of flexible tank's wall is considered throught several main parameters in container such as natural frequency of liquid sloshing, shear forces, moments in tanks wall or in building for instance column's moments, top displacements. In this work, **Ansys** V.11 is used to model the hold structure and investigate the thick of tank wall to describe the relation of the rigid and flexible tank.

The liquid-tank's wall interaction in TLD

The TLD is designed to have the same natural frequencies with structure and achieved the resonant phenomenon. One side helps to promote maximum ability of the damper but the other side it changes the TLD's dynamic properties through the liquid-tank's wall interaction. The main problem in studying of the liquid-structure interaction is solving the boundary condition at tank's wall. The equations described this condition is re-written by Biswal (2003)[18] as:

$$\begin{bmatrix} [M_s] & [0] \\ \rho_f [S] & [M_f] \end{bmatrix} \begin{Bmatrix} \ddot{d} \\ \ddot{p} \end{Bmatrix} + \begin{bmatrix} [K_s] & -[S]^T \\ [0] & [K_f] \end{bmatrix} \begin{Bmatrix} d \\ p \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{1}$$

with:

- M_s, M_f are mass matrices of structure and liquid
- K_s, K_f are stiffness matrices and liquid
- $\ddot{d}, d$ are acceleration and deformation of the structural boundary
- p is the liquid hydro-dynamic pressure.

Equation (1) leads to a non-standard, unsymetric eigenvalue. It is more difficult to find the eigenvalue for large size matrices. Many studies were presented to deal with this problem, and **Ansys** V.11 can be used to model liquid, container and main structure. The natural frequencies, amplitude of liquid sloshing are selected to emphasize the importance of the interaction. The tank's wall, columns and beams were modeled by using "*Beam3*" element, liquid by "*Fluid 79*" element. The liquid–container interaction was achieved by coupling the displacements of the liquid and container walls in the normal direction to the container walls.

The effect of tank's wall to natural frequency

Four types of containers with different thickness of tank's wall, *t*, from thin to thick are analyzed to find the natural frequencies and then compared with Housner's formulation (1967)[1] for rigid container. The containers are $T0.59\times0.03$ (container's width is 0.59m, height of liquid is 0.03m), $T1.00\times0.10$, $T3.00\times0.20$, $T6.00\times0.50$. The natural frequency of tank [5]:

$$f = \frac{1}{2\pi}\sqrt{\frac{\pi g}{2a}\tanh\left(\frac{\pi h_f}{2a}\right)} \tag{2}$$

The relation between flexible tank and rigid tank can be set up though ψ by Duc Tuong et al (2010)[1] as the flexibility parameter which is depended on thickness of tank wall, height and modulus of liquid:

$$\psi = E \times \frac{t_{\text{tank}}^3}{h_{\text{liquid}}^3} \tag{3}$$

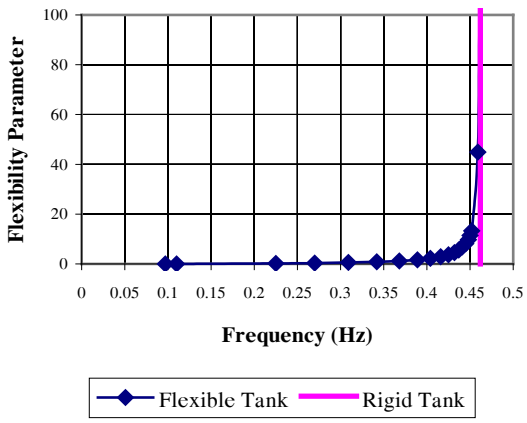


Figure 5. Frequency of $T0.59\times0.03$

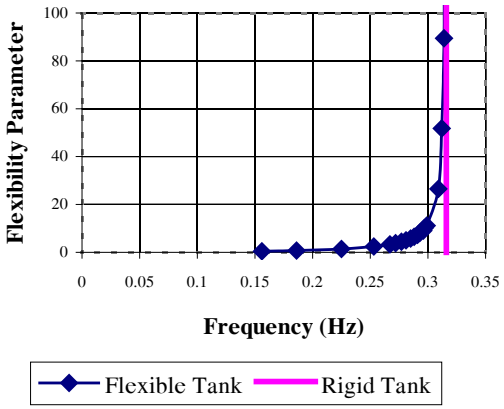


Figure 6. Frequency of $T1.00\times0.10$

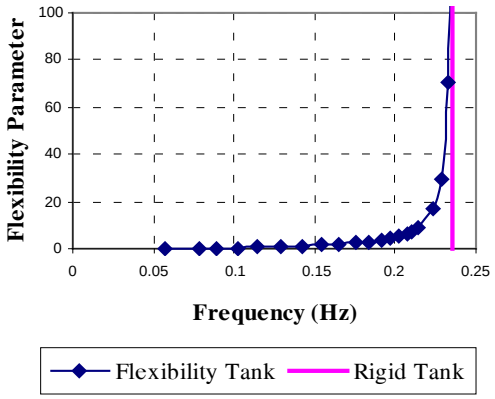


Figure 7. Frequency of $T0.59\times0.03$

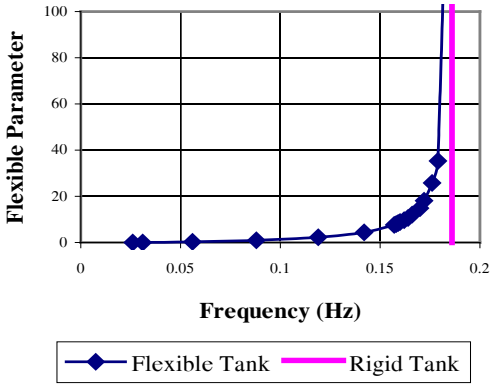


Figure 8. Frequency of $T1.00\times0.10$

Figure 5, 6, 7, 8 showed that the container is rigid if ψ is more than *100* otherwise it is flexible. And easily see that the thicker tank's wall, the higher frequency of TLD.

The effect of tank's wall to sloshing amplitude

To see clearly the effect of the liquid–tank's wall interaction, a numerical example is considered. A rectangular concrete container has the sections $6.0m\times1.0m\times0.5m\times t$, in length, height, liquid height and wall's thickness, and is applied harmonic load $x=A_o\sin\omega t$ with $A_o=5(mm)$. The properties of concrete are: $\rho=2400kg/m^3$, $E=2.65e10(kN/m^2)$, $\upsilon=0.2$. The mass of the structure was 28.941*kN*. From (2), $f_{\tan k}^{rigid}=0.18582(Hz)$ and in **Ansys** V.11 $f_{\tan k}^{rigid}=0.18250(Hz)$. Based on the flexibility parameter ψ [1], this container is rigid when $\psi\geq100\Leftrightarrow t\geq1.78mm$.

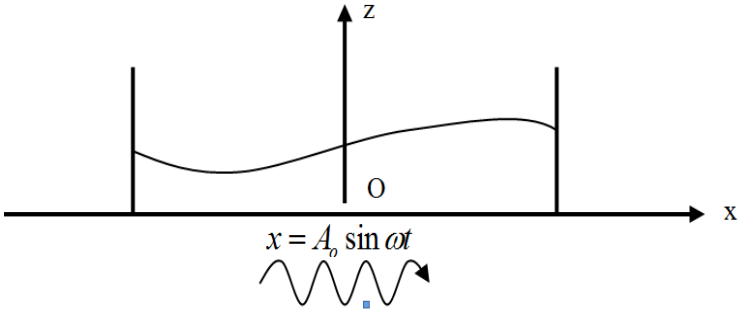


Figure 9. Container $T6.0m\times1.0m\times0.5m\times t$

Under harmonic loading $x = A_o \sin \omega t$ with $A_o = 5(mm)$ and $f = 0.25Hz$. The sloshing can be expressed as Figure 10 and 11, it is clear to see that when *t* is rigid or near rigid ($\psi \geq 100$), sloshing amplitude is merely the same (Figure 10) and the amplitude in flexible container is much more than in rigid one (Figure 11).

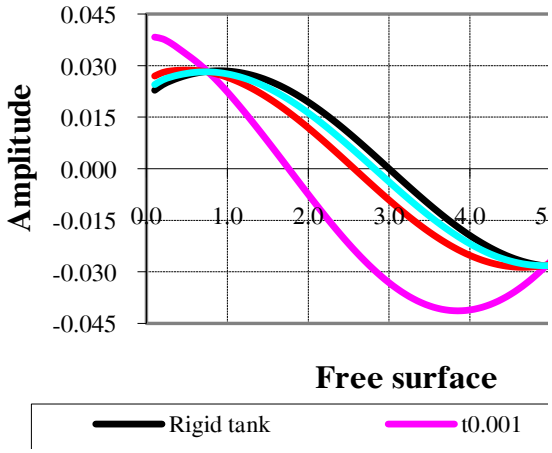


Figure 10. Sloshing amplitude in near rigid container

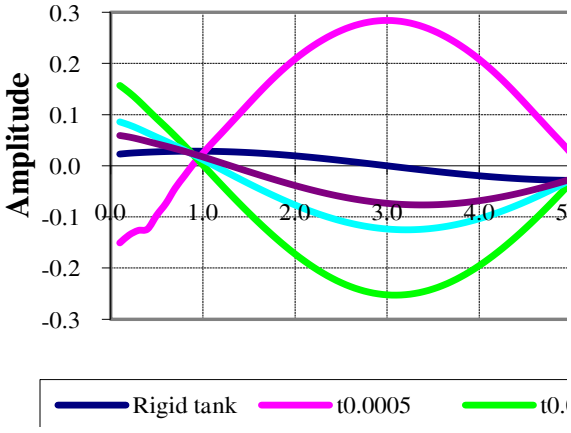


Figure 11. Sloshing amplitude in flexible container

Come to conclusion, the liquid-tank's wall interaction is important in container design because of two reasons: (1) the interaction leads to change the dynamic properties of tank and the natural frequencies of container can adjust easily by increasing or reducing the tank wall's thickness; (2) under the same load, sloshing amplitude in flexible tank is higher in rigid tank. This implied flexible tank is carried more load than in the rigid one so that when design TLD the flexibility should be checked by parameter ψ to protect the stability of container. Because of the rigid tank wall assumption, there are a lot of failure containers especially with the dynamic loads.

Numerical Example in Designing TLD Considering Liquid-Tank's Wall Interaction

Two examples are analyzed to investigate the seismic resistance of TLD for high-rise building. The first one presents the main point in designing damper and the second shows the TLD's capacity and emphasize the importance of liquid-container interaction.

Example 1

Design TLD for a steel building $70m$ in height under harmonic and seismic load, El-Centro earthquake data is used to analyze the seismic resistance of the building in **Ansys** V.11 and Newmark's method is used to predict sloshing and top building's deformation. Building has 14 storeys with each storey $5m$ in height and one span $3m$ in length. All of beams and columns section are the same and the tank is $T0.6m \times 0.8m$ with $E_{steel} = 2.1 \times 10^{11} N/m^2$, $\rho_{steel} = 7800kg/m^3$, $\upsilon = 0.3$. Mass of structure is $P_{building} = 6685000N = 6685kN$.

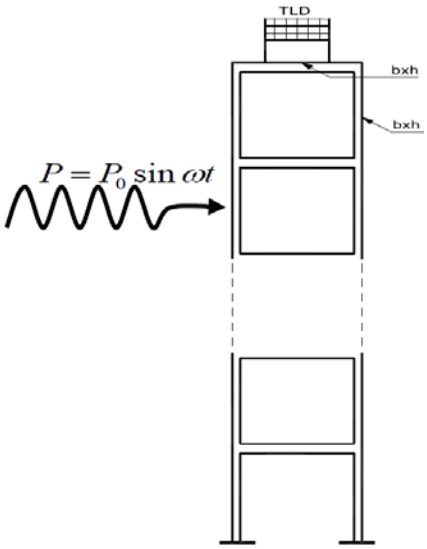


Figure 12. 14-storeys building with TLD

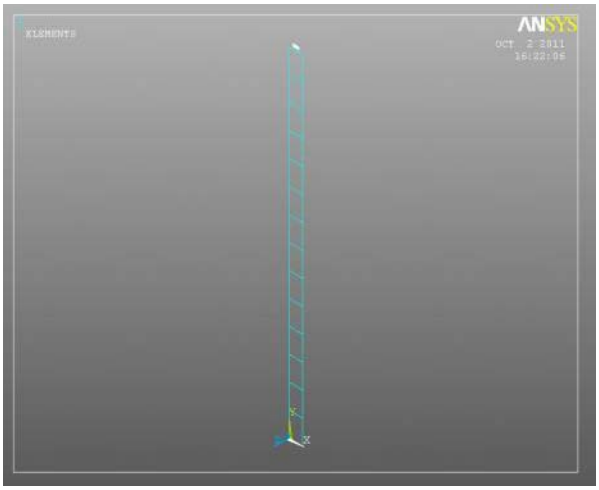


Figure 13. 14-storeys building in Ansys

The TLD is designed based on Sun LM's guides (1992) [8], that is: $f_{TLD} \approx f_{structure}$ and $P_{TLD} \approx \frac{1}{100} P_{structure} = 6685N$. And two conditions can be described as:

$$\begin{cases} P_{TLD} = \gamma \times g \times b_t \times h_f = 9810 \times b_t \times h_f = 6650 \\ f_{TLD} = \dfrac{1}{2\pi} \sqrt{\dfrac{\pi g}{b_t} \tanh\left(\dfrac{\pi h_f}{b_t}\right)} \approx f_s = 0.70873 \end{cases} \tag{4}$$

Where b_t and h_f are tank's width and liquid height. The liquid in TLD is water with $E_{water} = 2.2 \times 10^9 N/m^2$, $\rho_{water} = 1000 kg/m^3$, $\upsilon = 0.5$. Withdraw from(4), we have $b_t \approx 1.2m, h_f = 0.5m$. Then the natural frequency of TLD followed (2) is $f_{TLD} = 0.749Hz$, the building is under harmonic load $P = P_0 \sin \omega t = 1000 \sin \omega t (N)$ with frequencies of load from $0 \to 1.2Hz$ and El-Centro earthquake. Figure (14) and (15) shows the response of structure with and without TLD.

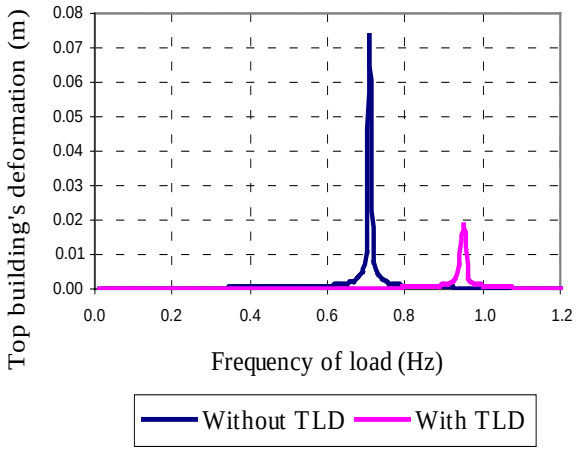


Figure 14. Structure under harmonic load

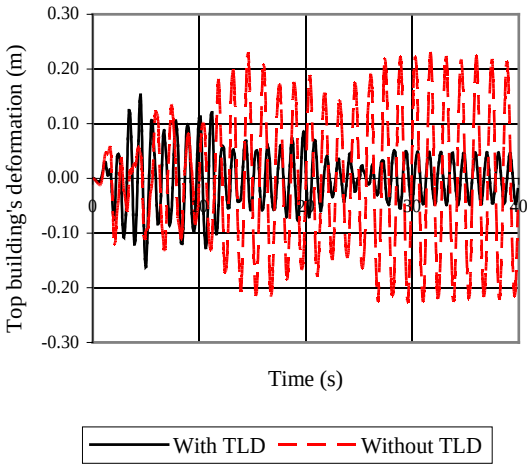


Figure 15. Structure under El-Centro

In Figure (14) the top building's deformation reduces 4 times when using TLD and the resonant occurs at frequency $f = 0.94Hz$. Figure (15) shows that the building's vibration

reduced 80% under seismic load if TLD is used. Beside, the moments in the left column of the structure with TLD are less than without TLD 25%.

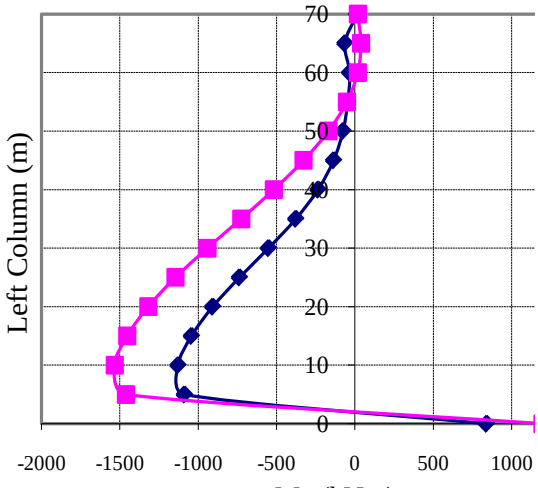


Figure 16. Moments in left column of the building

Example 2

An eight storeys steel building has one span 3.0m in length and each storey is 3.0m in height with $E_{steel} = 2.1\times10^{11} N/m^2$, $\rho_{steel} = 7800kg/m^3$, $\upsilon = 0.3$. Mass of structure is $P_{building} = 881762.8N \approx 881.763kN$. Figure 17 showed the natural frequency of structure $f_1^{building} = 0.29(Hz)$. The transient analysis is carried out to find the response vibration of the building with the frequencies of load from $0 \rightarrow 1(Hz)$ and Figure 18 illustrated the maximum response vibration of the structure without TLD is 1.8m at $f = 0.29(Hz)$.

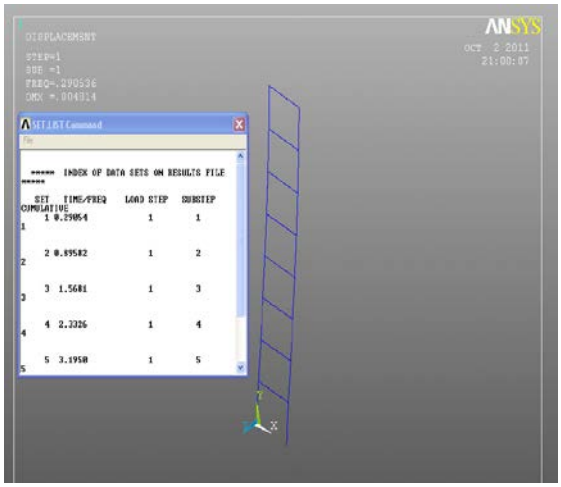


Figure 17. 8-storeys building's frequency in Ansys

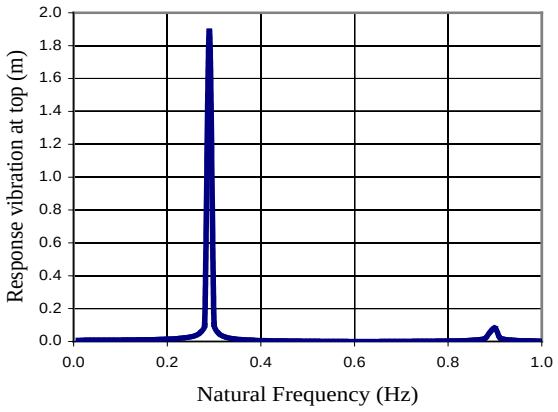


Figure 18. Response vibration of building without TLD

The TLD is designed by the same progress with Example 1 in **Ansys** V.11 to suppress the vibration and its section is $L_{TLD} \times h_{liquid} = 2.0m \times 0.2m$, $f_{TLD}^{rigid} = 0.277(Hz) \approx 0.95 f_{building}$ and $P_{TLD} = P_{tank} + P_{water} = 10079N \approx 1.13\% P_{building}$. The thickness t of container is changed from thin to thick to investigate the effective of liquid–tank's wall interactions. The thickness t of container is separated in two types which are rigid and flexible. Figure 19 described the response vibration of building with rigid TLD that means $\psi \geq 100$ is reduced 50% and the

resonance was occurred at $f = 0.29(Hz)$, the same with the natural frequency of the building. But in the Figure 20, the resonance was occurred uncontrollably and at the undefined value. Thus, the flexibility of TLD must be checked when designed.

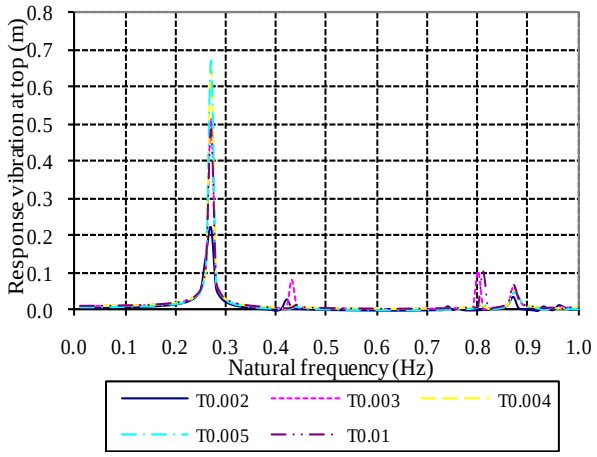


Figure 19. Top displacement with rigid TLD

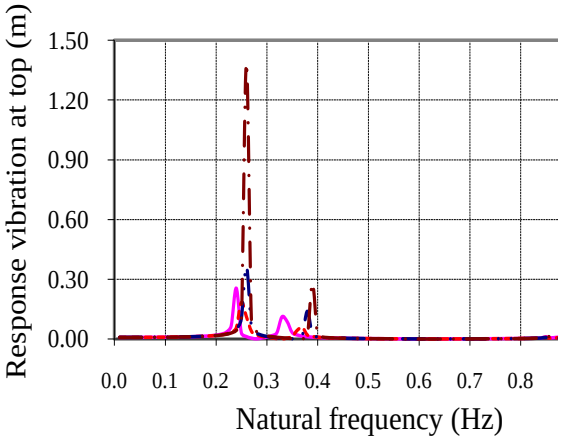


Figure 20. Top displacement with flexible TLD

To continue this part, the seismic resistance of TLD is analyzed. El-Centro and Newmark is used as data and tool to track liquid sloshing and displacements of building. When occurred earthquake, TLD is activated and the liquid sloshing is oscillated as shown in Figure 21, and contributes to reduce 67% the vibration of building as shown in Figure 22. However, with the different container's thickness, there were various top deformations as illustrated in Figure 23.

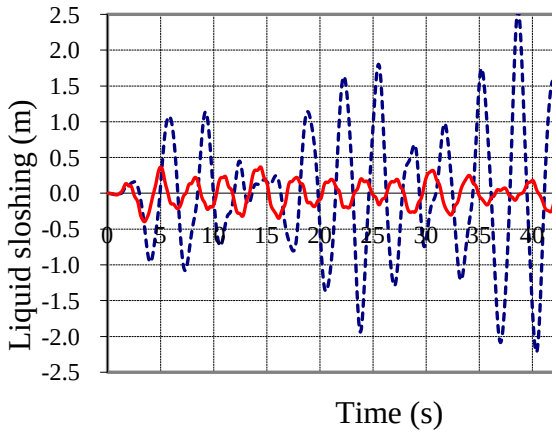


Figure 21. Sloshing at top building

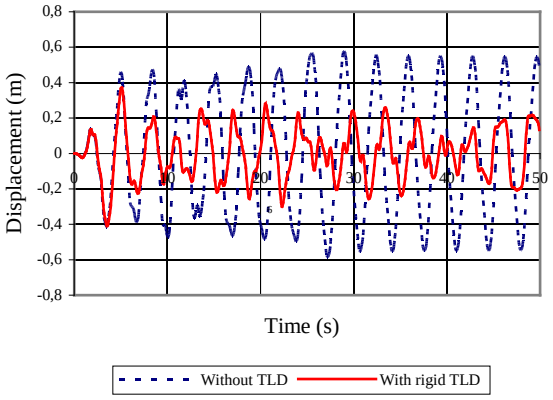


Figure 22. Top building with & without TLD

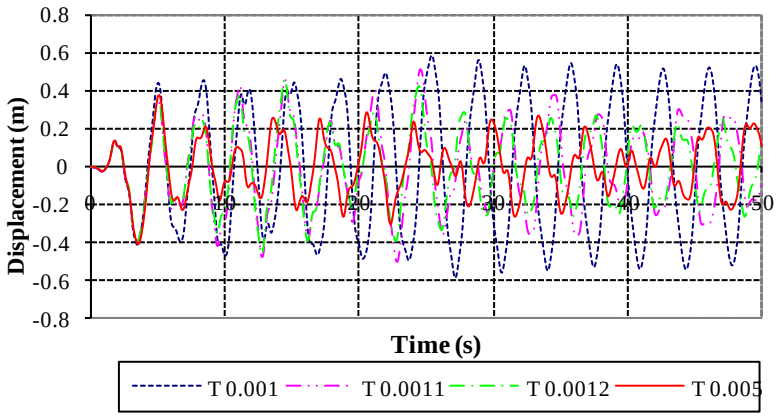


Figure 23. Top building with flexible TLD

To see more clearly the seismic resistance capacity of the damper, the moments in column of the building are illustrated in the Figure 24 and 25 in cases with and without TLD. That figure showed TLD can reduced from 50 to 75% moment in column (good agreement with Sun and Fujino's experiment in 1989 [7]).

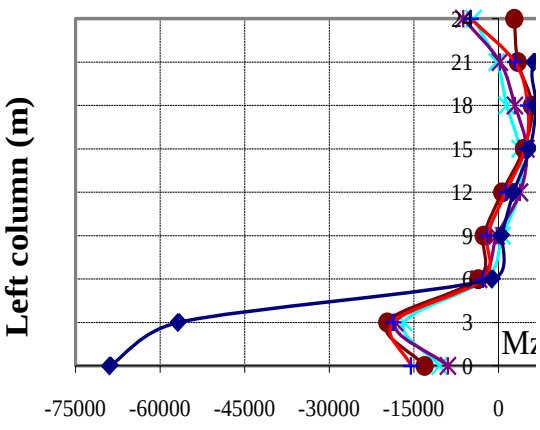


Figure 24. Moment in left column of building

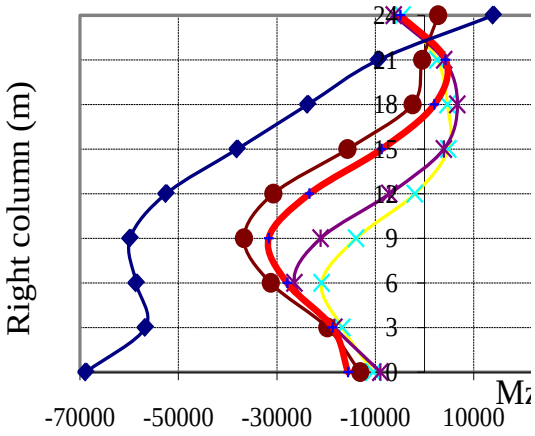


Figure 25. Moment in left column of building

Conclusions

The seismic resistance capacity of TLD is enough good if the ratio of TLD over the suppressed modal mass of the structure is 1-3%. With 1-3%, the TLD's weight does not significant affect the dynamic characteristics of the structure. Then, the vibration at the top of the building is reduced from 67% to 75% based on TLD. This leads to the moment in columns also reduced meaningfully up to 80% (good agreement with Fujino's results (1988) [13]).
When TLD is activated, there is no different of the building's internal forces between rigid and flexible tank's wall as showed in Figure 24 and 25. So the recommendation is that TLD should be designed to have a rigid wall to avoid the tank's deformation because of the interaction.

Using TLD to seismic resistance will re-distribute the internal forces so that the maximum moment may be not appear at the column base at Figure 16 and 25.
The interaction at liquid-tank's wall is very important so that it must be considered carefully. The interaction can change the dynamic properties of container wall then the water tank could not be TLD. Otherwise, one can use the flexibility to control the natural frequency of the tank [10].

The TLD can be designed easily as in example 1 by adjust the size of tank and height of liquid. It also can be applied for almost structure.

References

[1] Bùi Phạm Đức Tường. (2010) Highrise building's vibration reduced considered tank wall and fluid interaction, *Master thesis, HCM university of technology*.
[2] Bùi Thanh Tâm. (1997) A displacement based formulation of nearly incompressible fluid element for analysis of large amplitude liquid sloshing for tuned liquid damper applications, PhD thesis, Asian Institute of Technology-AIT.
[3] Han Jun và Li Yingmin. (2006) Numerical modelling on the damping control of TLD structure, *4th International Conference on Earthquake Engineering, Taipei, Taiwan*, 183-186.
[4] Hsiung và Weingarten. (1973) Dynamic analysis of hydro-elastic systems using the finite element method. Report USCCE013, Civil Engineering, Southern California University, USA.
[5] Jin Kyu Yu. (1999) A non linear numerical model of the tuned liquid dampers, *Earthquake Engineering and Structural Dynamics*, **28**, 671-686.

[6] Jin Kyu Yu. (1997) Non-linear characteristic of tuned liquid damper (TLD), *PhD thesis*, University of Washington.
[7] Li Min Sun và Yozo Fujino. (1989) Nonlinear waves and dynamic pressures in rectangular tuned liquid damper (TLD), *JSCE*, **410**, I-12.
[8] Li Min Sun. (1992) Semi-analytical modelling of tuned liquid damper (TLD) with emphasis on damping of liquid sloshing, *PhD thesis*, University of Tokyo.
[9] Lương Văn Hải. (2008) Modelling, simulation and behaviour of sloshing liquid-tank-ship coupled system, *PhD thesis*, National University of Singapore.
[10] Marija Gradinscak. (2009) Liquid sloshing in containers with flexibility, *PhD thesis*, Victoria University, Australia.
[11] Modi và Welt. (1987) Vibration control using Nutation Damper, *International Conference on Flow Induced Vibration, England*.
[12] T.T. Soong và Spencer. (2002) Supplemental energy dissipation: state of the art and state of the practice, *Engineering Structures*, **24**, 243-259.
[13] Yozo Fujino và Pacheno. (1988) Parametric studies on tuned liquid damper (TLD) using circular container by free oscillation experiments, JSCE, **398**, I-10.
[14] Zienkiewicz. (1971) The finite element method in engineering science. *McGraw Hill, New York*
[15] Dorothy Reed. (1998) Investigation of tuned liquid dampers under large amplitude excitation, *Journal of Engineering Mechanics*, 405-413
[16] Anderson, J. G., Semercigil, S. E. and Turan, Ö. F., (2001), A Standing-Wave Type Sloshing Absorber to Control Transient Oscillations, *Journal of Sound and Vibration*, **232(5)**, 839-856.s
[17] Vandiver, (1979), Effect of storage tank on the dynamic response of offshore platform, Journal of Petroleum Technology, 1231-1240
[18] Biswal, (2003), Dynamic response of liquid filled composite containers with baffles considering liquid – structure interaction, PhD thesis, Indian Institute of Technology.

Molecular communication in Nano Networks

Sidra Zafar[1],Mohsin Nazir[1], Aneeqa Sabah[2]

[1]Department of Computer Science, Lahore College for Women University ,Pakistan.
[1]Department of Physics, Lahore College for Women University ,Pakistan.
Email: sidzafar.88@gmail.com
mohsinsage@gmail.com

Abstract

This article examines the current research in nano communication networks specifically Molecular Communication (MC).Molecular Communication is an emerging communication paradigm where molecules are used to exchange information. Unlike traditional communication paradigms, molecules are transmitted as messages between biological nano machines. Key research challenges in molecular communication include design of system components and mathematical modeling of each system component as well as entire systems. Recent research in molecular communication and its propagation medium has been reviewed in this article.

Keywords: Nano machine; Nano communication networks; Nano communication networks applications; Molecular communication; Molecular Propagation Systems.

Introduction

Nanotechnology is miniaturization and fabrication of devices in a scale ranging from 1 to 100 nanometers. The prefix nano means one billionth i.e., ($1x10^{-9}$). Nanotechnology has been defined in a number of ways in literature. However according to [1] nanotechnology is "a branch of technology dealing with the manufacture of objects with dimensions of less than 100 nano-meters and the manipulation of individual molecules and atoms". Nano machines are fully functional devices and capable of performing trivial tasks like sensing, actuation, computing and data storage. Single nano machines are only capable of performing trivial tasks therefore to perform more complex tasks they must be interconnected to form a network [2], [24]. Nano-machines can be manufactured using three approaches top-down, bottom-up and bio-hybrid.

Molecular communication (MC) is the paradigm in nano communication networks that utilizes molecules for communication among nano machine [2][3][20]. Molecular communication is biologically inspired i.e., it adapts the communication mechanisms already existent in nature for communication among living organisms. Human body is composed of a large scale heterogeneous network where molecular communication takes place for intra body communication [25].There are a number of intra body applications where small scale communication is necessary e-g targeted drug delivery, BMI (Brain Imaging Interface), tissue engineering and cell repair etc [2][20][25]. Various communication and networking aspects of MC are currently being explored by research community. We have investigated the current research in propagation medium of Molecular communication, after giving brief introduction of nano communication network.

Section 2 discusses the current research in Molecular Communication. Section 3 defines the architecture of nano machines between which Molecular communication takes place. Section 4 highlights the applications of nano networks. Section 5 discusses different communication mediums used for communication between nano machines. In Section 6 detailed architecture of Molecular communication has been discussed and in Section 7 is the conclusion.

Background Study

Molecular communication is as old as the existence of nature, as communication has been taking place between living organisms since then. However with the advancement in computer networking the significance of Molecular communication has been brought to light all over. The research on molecular communication from the networking perspective is almost two decades old but still is immature. There are still a number of open issues regarding design and mathematical modeling of system components that needs to be addressed.

Molecular communication is able to take place in three ranges according to [2]. (1) Short range communication using molecular motors is the mechanism where inter cell and intra cell communication takes place using molecular motors, which are carriers of information encoded molecule. (2)Short range communication using calcium ions is another mechanism of molecular communication where communication might take place either between physically adjacent cells or distant cells using calcium ions (Ca^{2+}).(3)Long range communication using pheromones is the communication mechanism that takes place between sender and receiver nano machines that might be millimeters to kilometers apart. Application domains long range communication is military field and environmental applications [2].

Architecture of Nano-machines

Nano-machine consists of five components in its complete form. In order to develop efficient and novel nano-machines and to understand the communication mechanism between nano-machine, study of biological cell architecture and their interactions has been proved helpful. Following architectural components are included in most complete nano-machines and their biological cell counterparts are identified by [2], [6] and compared in the table below:

Table 1. Mapping between synthesized nano-machines and nano-machines found in biological cells

Synthesized nano-machines	Biological nano-machines
Control Unit. It contains the embedded software, which aims to perform the intended task of nano machine.	**Control Unit**. Similar to software conditional expressions biological control unit encodes protein structures, data units and regulatory sequences.
Communication Unit. Communication mechanism of nano machine is realized through transceivers. Transceivers allow the embedded system to exchange information by transmitting and receiving messages at nano level.	**Communication Unit**. The inter-cellular communication is realized through the gap junctions, hormonal and pheromonal receptors placed on the membrane of cell.
Reproduction unit. It contains the instructions to fabricate the components of nano-machines and then to replicate them.	**Reproduction**. This process takes place when nano machines are replicated by saving the code of nano machine in molecular sequences .
Power Unit. Power unit supplies stored energy to all the other components of nano-machines, to maintain the electrical current in embedded software.	**Power Unit**. Mitochondrion, chloroplast and Adenosien Tri phosphate are some of the substances of cells that correspond to the external chemical reactions to produce energy. This chemical energy is stored in the cell reservoirs and supplied to regulate the other components of cell.

Sensor and Actuators. This unit provides interface between environment and nano machine.	**Sensors and Actuators**. Sensing and actuation is the ability of biological cell to distinguish external molecules or stimuli e-g chloroplast of plants and flagellum of bacteria.

The most complete nano machine consists of all the components described above. However according to application domain nano machines might be changed in shape such as nano robots in medical applications.

Nano Network Applications

Nanonetworks applications are unlimited and are used extensively in almost every field. However they are classified in following broad groups in [2].

Biomedical applications

The size of nano devices makes them feasible for a number of bio medical and health monitoring applications including diagnostics, treatment and prevention of diseases. Another advancement in the field of healthcare is the nano machine deployed inside the human body which can remotely be controlled from the nanoscale and over the internet by an external user (healthcare provider) [7].

Industrial applications

Nano devices are showing potential in a number of industrial and consumer good applications. Interconnected nano-machines are used by video gaming industry for increased thrill and realistic gaming experience. It provides the functionality of transporting molecules from one location to another, mixing different types of molecules and separating specific kind of molecules from a mixture [10][11] .

Military Applications

Nanotechnology also has several applications in the military field. Nano devices such as imperceptible nano cameras, ultrasonic nano phones, and biological nano-sensors are devices that show potential in battlefield monitoring and actuation [2][7].

Environmental Applications

The bio inspired nature of nano technology makes it feasible to detect and sense contaminated materials found in nature. The problem of handling and dumping garbage is increasing around the world; this problem can be dealt by biodegradation process that uses nano-networks [2].
Nanonetworks can also be used to monitor air, thus controlling air pollution and nano filters can be developed to improve air quality and remove harmful materials from air [12].

Communication between Nano Machines

Nano-machines are only able to perform trivial tasks on their own; therefore communication among nano-machines is very important to realize more complex tasks .Nano-machines can be interconnected to execute collaborative tasks in a distributed manner resulting in nano-networks that expand the capabilities and applications of single nano-machines [2].
Nano-machine communication technologies are divided into four groups namely:

- Electromagnetic communication
- Acoustic Communication
- Nano Mechanical Communication
- Molecular Communication.

Electromagnetic communication
This type of communication based on the transmission and reception of electromagnetic waves between novel nano materials such as carbon nanotubes and graphene based nanoribbons [2][13]. The traditional transceiver of classical wireless communication is not feasible for nano-scale communication, however novel graphene based nano-materials have shown potential to overcome this limitation [13].

Acoustic Communication
Acoustic communication is realized by the transmission of ultrasonic waves through nano machine integrated transducers .These transducers should be capable to sense the variety of pressure and then react accordingly. Currently the size of transducers is the major barrier to implement this communication mechanism at nano-scale [2].

Nano Mechanical Communication
In nano mechanical communication, the information is sent through nano machines that are linked physically. One of the major drawbacks for this communication technique in nano communication context is physical connection between devices. Therefore it is not feasible for the applications where nano-machines have to be placed at distant locations [4].

Molecular Communication
Molecular Communication (MC) is a molecule based communication paradigm that enables transmission of bio-chemical information (e.g. status of living organisms), which is not feasible using traditional communication [14]. Molecules encoded with information to be transmitted, are called information molecules. The information molecules activate bio-chemical reaction at receiver and may recreate phenomena and/or chemical status, which sender then transmits [9][14] Molecular communication (MC) is considered the most promising nano networking mechanism due to its nano-sized transceivers that can easily integrate into nano machine [2][15].

Molecular Communication Architecture
Molecular communication architecture consists of information molecules that contain information to be transmitted, sender bio-nano machines that send information molecules, and receiver bio-nano machines that receive information molecules. Other types of molecules might be included in the system such as transport molecules which move information molecules, guide molecules which guides the movement of transport molecules, interface molecules for selective transport of information molecules [17].MC communication architecture is presented in the figure below. Different phases of molecular communication are described below [2][17]:

- **Encoding** in this phase sender nano machine encodes the information into the information molecules in various forms.
- **Sending:** In this phase sender bio nano machine releases information molecules in the environment.
- **Propagation:** It is the phase in which molecules travel from sender nano machine towards receiver nano machine. This transport can be either passive or active. Passive transport is the through diffusion of molecules in the environment without chemical energy, where as in active transport information molecules bind to molecular motors.
- **Receiving:** Transmitted molecules are received from the aqueous medium in this phase usually with the help of chemical receptors [38].
- **Decoding:** In this phase the captured molecules are decoded by receiver nano-machines into the form of chemical energy.

Conclusion

Molecular Communication is a novel communication paradigm which uses molecules for information transmission. Unlike traditional communication MC is capable to transmit information over short distances [22]. As MC is inspired from the communication among living cells and other biological materials it provides a number of biomedical and environmental applications. Nano Communication inside human body can poses a number of health applications e.g., targeted drug delivery, tissue engineering, BMI (Brain Machine Interface) and enhanced immune system [22].Interdisciplinary research is needed to develop theoretical and mathematical models for end-to-end communication between bio-nano machines. However authors in [18][24] have done wonderful work to explain layered and TCP like molecular communication.

References

[1] Nanotechnology. (n.d.).*Collins Engish Dictionary-Complete & Unabridged 10th Edition*.Retreived January 04, 2016 from Dictionary.com website http://dictionary.refference.com/browse/nanotechnology

[2] Akyildiz, I., Brunetti, F., & Blázquez, C. (2008). Nanonetworks: A new communication paradigm. *Computer Networks*, *52*(12), 2260-2279. http://dx.doi.org/10.1016/j.comnet.2008.04.001

[3] Minoli, D. (2006). *Nanotechnology applications to telecommunications and networking*. Hoboken, N.J.: Wiley-Interscience.

[4] Akyildiz, I., Jornet, J., & Pierobon, M. (2011). Nanonetworks:A New Frontier in Communications. *Communications Of The ACM*, *54*(11), 84. http://dx.doi.org/10.1145/2018396.2018417.

[5] Lim, H. (2004). Nanotechnology in diagnostics and drug delivery. *Medicinal Chemistry Research*, *13*(6-7), 401-413. http://dx.doi.org/10.1007/s00044-004-0044-4

[6] Liu, G. R., Zhang, G. Y., Dai, K. Y., Wang, Y. Y., Zhong, Z. H., Li, G. Y. and Han, X. (2005) A linearly conforming point interpolation method (LC-PIM) for 2D mechanics problems, *International Journal for Computational Methods* **2**, 645–665.

[7] Jornet, J. & Akyildiz, I. (2012). The Internet of Multimedia Nano-Things. *Nano Communication Networks*, *3*(4), 242-251. http://dx.doi.org/10.1016/j.nancom.2012.10.001

[8] Moritani, Y., Hiyama, S., & Suda, T. (2006, May). Molecular communication among nanomachines using vesicles. In *Proceedings of NSTI nanotechnology conference*

[9] Moritani, Y., Hiyama, S., & Suda, T. (2006, March). Molecular communication for health care applications. In *Pervasive computing and communications workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on* (pp. 5-pp). IEEE

[10] Yager, P., Edwards, T., Fu, E., Helton, K., Nelson, K., Tam, M., & Weigl, B. (2006). Microfluidic diagnostic technologies for global public health. *Nature*, *442*(7101), 412-418. http://dx.doi.org/10.1038/nature05064

[11] Dittrich, P. & Manz, A. (2006). Lab-on-a-chip: microfluidics in drug discovery. *Nature Reviews Drug Discovery*, *5*(3), 210-218. http://dx.doi.org/10.1038/nrd1985

[12] Han, J., Fu, J., & Schoch, R. (2008). Molecular sieving using nanofilters: Past, present and future. *Lab Chip*, *8*(1), 23-33. http://dx.doi.org/10.1039/b714128a.

[13] Akyildiz, I. F., Jornet, J. M., & Pierobon, M. (2010, April). Propagation models for nanocommunication networks. In *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on* (pp. 1-5). IEEE.

[14] Hiyama, S. & Moritani, Y. (2010). Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems. *Nano Communication Networks*, *1*(1), 20-30. http://dx.doi.org/10.1016/j.nancom.2010.04.003

[15] Akyildiz, I. & Jornet, J. (2010). The Internet of nano-things. *IEEE Wireless Commun.*, *17*(6), 58-63. http://dx.doi.org/10.1109/mwc.2010.5675779

[16] Loscri, V., Marchal, C., Mitton, N., Fortino, G., & Vasilakos, A. (2014). Security and Privacy in Molecular Communication and Networking: Opportunities and Challenges. *IEEE Transactions On Nanobioscience*, *13*(3), 198-207. http://dx.doi.org/10.1109/tnb.2014.2349111

[17] Nakano, T., Moore, M., Fang Wei, Vasilakos, A., & Jianwei Shuai,. (2012). Molecular Communication and Networking: Opportunities and Challenges. *IEEE Transactions On Nanobioscience*, *11*(2), 135-148. http://dx.doi.org/10.1109/tnb.2012.2191570

[18] Davis, S. (2000). Giant Vesicles. *Journal Of Controlled Release*, *68*(1), 135-136. http://dx.doi.org/10.1016/s0168-3659(00)00254-6

[19] Jesorka, A. & Orwar, O. (2008). Liposomes: Technologies and Analytical Applications. *Annual Review Of Analytical Chemistry*, *1*(1), 801-832. http://dx.doi.org/10.1146/annurev.anchem.1.031207.112747

[20] Walsh, F., Balasubramaniam, S., Botvich, D., Suda, T., Nakano, T., Bush, S. F., & Foghlú, M. Ó. (2008). Hybrid DNA and enzyme based computing for address encoding, link switching and error correction in molecular communication. In *Nano-Net* (pp. 28-38). Springer Berlin Heidelberg.
[21] Nakano, T., Suda, T., Okaie, Y., Moore, M., & Vasilakos, A. (2014). Molecular Communication Among Biological Nanomachines: A Layered Architecture and Research Issues. *IEEE Transactions On Nanobioscience*, *13*(3), 169-197. http://dx.doi.org/10.1109/tnb.2014.2316674
[22] Freitas, R. A. (1999). *Nanomedicine, volume I: basic capabilities* (pp. 17-8). Georgetown, TX: Landes Bioscience.'
[23] Ladd, A., Gang, H., Zhu, J., & Weitz, D. (1995). Time-Dependent Collective Diffusion of Colloidal Particles. *Phys. Rev. Lett.*, *74*(2), 318-321. http://dx.doi.org/10.1103/physrevlett.74.318
[24] ShahMohammadian, H., Messier, G., & Magierowski, S. (2012). Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels. *Nano Communication Networks*, *3*(3), 183-195. http://dx.doi.org/10.1016/j.nancom.2012.09.006.
[25] Felicetti, L., Femminella, M., Reali, G., Nakano, T., & Vasilakos, A. (2014). TCP-Like Molecular Communications. *IEEE J. Select. Areas Commun.*, *32*(12), 2354-2367. http://dx.doi.org/10.1109/jsac.2014.2367653
[26] Malak, D. & Akan, O. (2012). Molecular communication nanonetworks inside human body. *Nano Communication Networks*, *3*(1), 19-35. http://dx.doi.org/10.1016/j.nancom.2011.10.002
[27] Felicetti, L., Femminella, M., Reali, G., & Liò, P. (2016). Applications of molecular communications to medicine: A survey. *Nano Communication Networks*, *7*, 27-45. http://dx.doi.org/10.1016/j.nancom.2015.08.004.
[28] Dressler, F. & Kargl, F. (2012). Towards security in nano-communication: Challenges and opportunities. *Nano Communication Networks*, *3*(3), 151-160. http://dx.doi.org/10.1016/j.nancom.2012.08.00
[29] Dressler, F. & Fischer, S. (2015). Connecting in-body nano communication with body area networks: Challenges and opportunities of the Internet of Nano Things. *Nano Communication Networks*, *6*(2), 29-38. http://dx.doi.org/10.1016/j.nancom.2015.01.006
[30] Dressler, F., & Kargl, F. (2012, June). Security in nano communication: Challenges and open research issues. In *Communications (ICC), 2012 IEEE International Conference on* (pp. 6183-6187). IEEE.
[31] L u, P., & Wu, Z. (2016). Continuous Molecular Communication in one dimensional situation. *arXiv preprint arXiv:1603.03495*.
[32] Neri, I., Travasso, F., Vocca, H., & Gammaitoni, L. (2011). Nonlinear noise harvesters for nanosensors. *Nano Communication Networks*, *2*(4), 230-234. http://dx.doi.org/10.1016/j.nancom.2011.09.001.
[33] Chou, C. (2013). Noise properties of linear molecular communication networks. *Nano Communication Networks*, *4*(3), 87-97. http://dx.doi.org/10.1016/j.nancom.2013.06.001.
[34] Jabbari, A. & Balasingham, I. (2013). Noise characterization in a stochastic neural communication network. *Nano Communication Networks*, *4*(2), 65-72. http://dx.doi.org/10.1016/j.nancom.2013.04.002.
[35] Gul, E., Atakan, B., & Akan, O. (2010). NanoNS: A nanoscale network simulator framework for molecular communications. *Nano Communication Networks*, *1*(2), 138-156. http://dx.doi.org/10.1016/j.nancom.2010.08.003.
[36] Llatser, I., Demiray, D., Cabellos-Aparicio, A., Altilar, D., & Alarcón, E. (2014). N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks. *Simulation Modelling Practice And Theory*, *42*, 210-222. http://dx.doi.org/10.1016/j.simpat.2013.11.004.
[37] Chahibi, Y., Akyildiz, I., Balasubramaniam, S., & Koucheryavy, Y. (2015). Molecular Communication Modeling of Antibody-Mediated Drug Delivery Systems. *IEEE Transactions On Biomedical Engineering*, *62*(7), 1683-1695. http://dx.doi.org/10.1109/tbme.2015.2400631.
[38] Chou, C. (2015). Impact of Receiver Reaction Mechanisms on the Performance of Molecular Communication Networks. *IEEE Transactions On Nanotechnology*, *14*(2), 304-317. http://dx.doi.org/10.1109/tnano.2015.2393866.

The Effects of Quality and Shortages on the Economic Production Quantity Model in a Two-Layer Supply Chain

***†Abdul-Nasser El-Kassar**

[1]Information Technology & Operations Management, Lebanese American University, Lebanon

*Presenting author: abdulnasser.kassar@lau.edu.lb
†Corresponding author: abdulnasser.kassar@lau.edu.lb

Abstract

The purpose of this paper is to develop an economic production quantity model (EPQ) in a coordinated supplier-produced supply chain. This collaborative supply chain accounts for the quality of both finished product and raw material used in the production process. It is assumed that the raw material acquired from the supplier contains a percentage of good quality items. These items are detected through a screening process at the beginning of the production period. The quality of the finished items produced is checked continuously throughout the production period. The imperfect quality items are either reworked or rejected. The nature of this production/inventory problem necessitates the consideration of shortages. The mathematical model is formulated and the supply chain is optimized by determining the order quantity that maximizes the collaborative profit function. Numerical examples are provided to illustrate the model and the collaborative and non-collaborative models are compared.

Keywords: Inventory Control, Economic Production Quantity, Supply Chain, Quality, Screening, Rework, Reject.

Introduction

The two classical inventory control techniques known as economic order quantity (EOQ) and economic production quantity (EPQ) models have been widely used among researchers and industries (Bedworth and Bailey [1]) and (Simpson, [2]). The EOQ model aims to optimize the order size by balancing or trading off the ordering and holding costs. The EPQ model seeks to determine the optimal lot size by minimizing the total setup and carrying costs. Despite of their widespread usage and implementation, these models are based on idealistic assumptions that are rarely met in real life situations. In the past few decades, considerable research has been published whereby the underlying assumptions are relaxed so that the EOQ/EPQ models are examined under situations that closely resemble the actual inventories encountered in real life. The modified models account for factors that influence the inventory costs. These factors include deterioration, shortages, probabilistic demand, order quantity and demand dependent costs, inflation, time discounting and credit facilities.

One of the unrealistic assumptions of the classical EOQ/EPQ models is that all items received from the supplier and all items produced by the manufacturer are of a perfect quality. These assumptions initiated a new line of research in the field of inventory management that ensures quality. Porteus [3] studied the relationship between the lot size, process quality, and setup cost. Rosenblatt and Lee [4] examined a production system with defective finished items. Salameh and Jaber [5] introduced an EOQ model where each lot delivered by the supplier contains imperfect items, not necessarily defectives, which can be salvaged at a discounted price. This modeling approach has triggered numerous research papers extending this model. Hayek and Salameh [6] studied an EPQ model where the imperfect quality items are reworked. Chiu [7] considered a production process with random defective rate where the defective items are reworked and unsatisfied demand is backlogged. Ozdemir [8] proposed an

EOQ model with defective items where shortages are backordered. Khan et al. [9] presented an extensive survey of such articles.

A typical supply chain consists of suppliers, producers, distributors and retailers. The inventory problems of partners in a supply chain have been treated separately. Recently, numerous studies have been published in the field of inventory management dealing with the interaction between partners in a supply chain and aiming for the improvement of their joint performance. Khan et al. [9] reviewed articles related to EOQ/EPQ models in supply chains with imperfect quality items.

In a different direction, El-Kassar et al. [10] introduced an EPQ model with imperfect quality items of raw material used in the production process. El-Kassar et al. [11] examined the effect of time value of money on this model. The purpose of this paper is to examine the effects of the interaction between the supplier of raw material and the producer of the finished product on the joint performance of both partners in this supply chain. This is done by developing an EPQ model in a coordinated supplier-produced supply chain. This model accounts for the quality of both finished product and raw material in a collaborative supply chain. The raw material acquired from the supplier contains a percentage of good quality items. At the beginning of the production period, the good quality items are detected through a 100% screening process. Throughout the production period, the quality of the finished product is checked continuously. The imperfect quality items are either reworked or rejected. This model allows for shortages. The unsatisfied demand is assumed to be fully backordered.

The rest of this paper is organized as follows. In section two, the needed assumptions and the used notations are presented and the mathematical model is formulated so that the supply chain is optimized by determining the order quantity that maximizes the collaborative profit function. In section three, a numerical example is provided to illustrate the model and the collaborative and non-collaborative models are compared. In section four, a conclusion and some managerial implications are provided. Also, future research suggestions are stated.

The Mathematical Model

Assumptions:

1. Finished product items produced are checked for quality through a 100% error free screening process conducted throughout the production process.
2. The production rate of perfect quality items is greater than the demand rate.
3. Planned shortages are permitted and fully backordered.
4. Rework process starts at the end of the production process with no setup time.
5. Reworked items are processed at the same production rate.
6. The percentage of good quality items of raw material, the percentage of perfect quality finished items, and the percentage of scrap items are known constants.

Notation:

P	Production rate
D	Demand rate
x	Raw material screening rate
Q	Order size of raw material
S	Maximum shortage per cycle
T	Cycle length
t_1	Time to fulfill the backorder of size S
t_2	Time to build up the maximum inventory of perfect quality finished items
t_3	Rework time
t_4	Time to deplete on-hand inventory after rework
t_5	Time to build up the maximum shortage level of size S
t_s	Raw material screening period

t_p	Production time ($t_p = t_1 + t_2$)
A_p	Producer's fixed ordering cost of raw material
A_s	Supplier's fixed ordering cost of raw material
K	Production fixed setup cost
c_{ms}	Supplier's cost of one unit of raw material
c_{mp}	Producer's cost of one unit of raw material
c_p	Cost of producing one unit of finished product
c_r	Cost of reworking one unit of finished product
c_d	Cost of disposing of one unit of scrap item of the finished product
c_s	Cost per unit shortage per unit time
d_m	Cost of screening one unit of raw material
d_f	Cost of screening one unit of finished product
r	Selling price per unit of finished product
r_d	Discounted selling price per unit of raw material.
h_{ms}	Supplier's holding cost of raw material per unit per unit time
h_{mp}	Producer's holding cost of raw material per unit per unit time
h_p	Holding cost due to production per unit per time
h_r	Holding cost due to rework per unit per time
γ	Percentage of good quality items of raw material
π	Percentage finished product that are of perfect quality
ρ	Percentage of imperfect quality items that are reworked
$1-\rho$	Percentage of imperfect quality items that are scrapped (defective)
λ	Proportion of reworked items used to meet the demand
G_s	Supplier's profit function per unit time
G_p	Producer's profit function per unit time
G_c	Chain's profit function per unit time
N	Number of production cycles per one supplier's inventory cycle

In this coordinated two layer supply chain, the producer orders from the supplier Q units of raw material at the beginning of each production cycle. The raw material acquired contains a percentage γ of imperfect quality items. The γQ units of good quality items are detected through a 100% error free screening process and used in the production of the finished product. At the end of the screening period, the remaining $(1-\gamma)Q$ units of raw material are returned to the supplier who sells the items at a discounted unit price r_d. Since raw material is screened at rate of x units per unit time, the screening period is $t_s = Q/x$. Also, the γQ units of good quality items of raw material are processed into finished product at a rate of P, where $x>P$, so that the production period is $t_p = \gamma Q/P$.

Throughout the production period, the finished product is screened to detect perfect quality items. Since the percentage of perfect quality finished items is π, the number of perfect quality items produced is $\gamma\pi Q$ and the remaining $(1-\pi)\gamma Q$ finished items are of imperfect quality. A percentage ρ of the imperfect quality finished items are can be reworked into perfect quality finished items and the remaining $1-\rho$ are scraped. The number of reworked and scraped items are $\rho(1-\pi)\gamma Q$ and $(1-\rho)(1-\pi)\gamma Q$, respectively.

The $\gamma\pi Q$ perfect quality items are produced at a rate of $P\pi$ units per unit time. Since shortages are allowed, the perfect quality finished items produced at the beginning of the production period will be used to meet the demand, at a rate D, and to fulfill backorders at a rate $P\pi - D > 0$. Assuming that the inventory cycle begins with S units short, the time required to fulfill the backorders is $t_1 = S/(P\pi - D)$. Once all backorders are fulfilled, inventory of perfect quality finished items is accumulated at a rate of $P\pi - D$ until a level of $z_1 = t_2 (P\pi - D)$, where $t_2 = t_p - t_1$, is reached at the end of the production period.

When regular production stops, the scraped items are disposed of at a unit cost of c_d. The remaining imperfect quality finished are reworked into perfect quality items at the same

production rate P. During the rework period, from $t = t_p$ until $t = t_p + t_3 = t_p + \rho(1-\pi)\gamma Q/P$, the perfect quality finished items inventory increases at a rate of $P–D$ until a maximum level of z_2 is reached where $z_2 = z_1 + (P–D)t_3$. This accumulated inventory will be used to meet the demand at a rate D so that the time required to deplete this inventory is $t_4 = z_2/D$. During the remainder of the inventory period, the demanded items are backordered. The time required to build up the maximum shortage level of size S is $t_5 = S/D$. The inventory behavior of perfect quality items is depicted in Fig. 1.

In order to calculate the inventory holding cost, the behavior of both imperfect quality items and reworked items inventories must be determined. Since imperfect quality items are reworked at the end of the production period, such items are accumulated throughout this period at a rate of $(1-\pi)P$ until a maximum level of $z_3 = t_p\,(1-\pi)P = (\gamma Q/P)(1-\pi)P = (1-\pi)\gamma Q$ is reached. After the disposal of scraped items, the imperfect quality items inventory drops to a level of $z_4 = \rho(1-\pi)\gamma Q$. During the rework period, between time $t = t_p$ and time $t = t_p + t_3$, these items are reworked into perfect quality items at a rate P. The inventory behavior of the imperfect quality items is illustrated in Fig. 2.

In the following we construct the producer's profit function by determining the relevant cost and revenue. At the beginning of each production/inventory cycle the producer places an order of size Q of raw material at a unit purchasing cost of c_{mp} and an ordering cost of A_p. These items are screened to detect the good quality at a unit screening cost of d_m. The γQ good quality items are used to produce γQ units of the finished product at a unit production cost of c_p and a setup cost of K. The remaining $(1-\gamma)Q$ items are returned to the supplier at the end of the screening period. Therefore, the purchasing cost of raw material is $c_{mp}\gamma Q$, the screening cost is $d_m Q$, and the production cost is $c_p\gamma Q$. Throughout the production period, the finished items are screened to detect the perfect and imperfect quality items at a unit screening cost d_f so that the screening cost of items produced is $d_f\,\gamma Q$. The perfect quality items $\pi\gamma Q$ are sold at a unit selling price of r. The remaining $(1-\pi)\gamma Q$ imperfect quality items are classified as scrap items or as items that be reworked into perfect quality. The $(1-\rho)(1-\pi)\gamma Q$ the scrap items are disposed of at a unit cost of c_d. The remaining $\rho(1-\pi)\gamma Q$ items of imperfect quality are reworked at a unit cost of c_r and sold at the same unit selling price of r. The shortage cost per cycle is obtained by multiplying the average number of units short per cycle by the cycle length by the cost of having of unit short per unit time. Similarly, the holding costs of the various types of items on hand are calculated by multiplying the average number of units on hand per cycle by the cycle length by the holding cost per unit per unit time. In summary, the revenues and cost components per cycle are:

Sales of good quality items	$= r\pi\gamma Q + r\rho(1-\pi)\gamma Q$
Ordering/Setup Cost	$= A_p + K$
Purchasing cost of raw material	$= c_{mp}\,\gamma Q$
Screening cost of raw material	$= d_m\,Q$
Finished items production cost	$= c_p\,\gamma Q$
Screening cost of finished product	$= d_f\,\gamma Q$
Disposal cost of scrap items	$= c_d\,(1-\rho)(1-\pi)\gamma Q$
Imperfect quality items rework cost	$= c_r\,\rho(1-\pi)\gamma Q$

In addition, the shortage cost is given by

$$\text{Shortage cost per cycle} = \frac{1}{2}S(t_1 + t_5)c_s = \frac{1}{2}S\left(\frac{S}{P\pi - D} + \frac{S}{D}\right)c_s = \frac{S^2}{2D\left(1 - \dfrac{D}{P\pi}\right)}c_s, \qquad (1)$$

and the various holding costs are

$$\begin{aligned}
&\text{Holding cost of raw material} = Q^2\left(\frac{\gamma^2}{2P}+\frac{1-\gamma}{x}\right)\times h_{mp}\\
&\text{Perfect quality items holding cost} = \frac{1}{2}\left(z_1t_2+(z_1+z_2)t_3+z_2t_4\right)\times(h_{mp}+h_p)\\
&\text{Imperfect quality items holding cost} = \frac{1}{2}\left(z_3t_p+z_4t_3\right)\times(h_{mp}+h_p)\\
&\text{Holding cost due to rework} = \frac{1}{2}z_4(t_3+t_4)\times h_r
\end{aligned} \tag{2}$$

From the above revenue and cost components as well as Eqs. (1) and (2), we have that the total profit per cycle function is

$$\begin{aligned}
TP(Q,S) = {} & r\pi\gamma Q + r\rho(1-\pi)\gamma Q - A_p - K - c_{mp}\gamma Q - d_m Q - c_p\gamma Q - c_r\rho(1-\pi)\gamma Q\\
& - d_f\gamma Q - c_d(1-\rho)(1-\pi)\gamma Q - \frac{S^2}{2D\left(1-\frac{D}{P\pi}\right)}c_s - Q^2\left(\frac{\gamma^2}{2P}+\frac{1-\gamma}{x}\right)\times h_{mp}\\
& - \frac{1}{2}z_4(t_3+t_4)\times h_r - \frac{1}{2}\left(z_1t_2+(z_1+z_2)t_3+z_2t_4+z_3t_p+z_4t_3\right)
\end{aligned} \tag{3}$$

Dividing by the inventory cycle length $T = Q(\pi+\rho-\pi\rho)/D$, we obtain the producer total profit per unit time function $G_p(Q,S)$.

Next, all revenue and cost components for the supplier are determined by assuming that the supplier inventory cycle is a multiple of the producer production cycle T. Let N be the number of production cycles in one supplier's inventory cycle. At the beginning of the cycle, the supplier orders NQ units of raw material at an ordering cost A_s and a unit cost of c_{ms}. These items will be delivered to the producer in batches each of size Q, where the first batch is delivered at the start of the supplier cycle so that the supplier maximum inventory level is $(N-1)Q$. The supplier inventory behavior is shown in Fig. 3. The producer keeps the γNQ good quality items and the producer sells the $(1-\gamma)NQ$ returned items at a discounted price r_d, where $r_d < c_{ms}$.

The supplier cost and revenue components per cycle are:

Sales of good quality items	$= c_{mp}\, N\gamma Q$
Sales of returned items	$= r_d\, N(1-\gamma)Q$
Ordering	$= A_s$
Purchasing cost of raw material	$= c_{ms}\, NQ$
Holding cost	$= QT\, hms\, N(N-1)/2$

The supplier total profit per cycle function is

$$TP(Q,S) = TP(Q,S) = c_{mp}N\gamma Q + r_d N(1-\gamma)Q - A_s - c_{ms}NQ - QTh_{ms}N(N-1)/2 \tag{4}$$

Dividing by the supplier inventory cycle length NT, we obtain the supplier total profit per unit time function $G_s(Q,S)$. The supply chain total profit per unit time function is obtained by adding Eqs. (3) and (4) so that

$$Gc(Q) = Gp(Q) + Gs(Q). \tag{5}$$

In a non-collaborative supply chain, the producer is the decision maker. In this case, the optimal solution Q^* is determined by maximizing the function $G_p(Q)$. The supplier then determines the integer N that maximizes $G_s(Q^*)$. In the case of a coordinated supply chain, the optimal solution is determined by maximizing the $G_c(Q)$ for each value of N and selecting the value corresponding to the largest maximum total profit for the supply chain.

Numerical Example

Consider a production process where the demand rate for an item is 100 units per day and the production rate is 400 units per day. The raw material used in production is ordered from a supplier where 80% of the items received are of good quality. Screening for good quality items of the raw material is conducted at a rate of 1000 items per day and at a cost of $0.25 per unit. The ordering cost for the raw material is $5,000 and the production setup cost is $5,000. The holding cost of raw material is $0.02 per unit per day while the holding cost due to production is $0.05 per unit per day. Hence, the holding cost of one unit of the finished product is $0.07 per day. 75% of the items produced are of perfect quality. 80% of the imperfect quality items produced can be reworked and the remaining 20% are scrap items. The screening cost for detecting imperfect quality finished items is $0.5 per unit. If an item is reworked, an additional holding cost of $0.01 per unit per day is incurred. The purchasing cost of one item of raw material is $10, the unit production cost is $20, and the rework cost per unit is $5. The selling price is $50 per unit. The scrap items are disposed of at the end of production period at a cost of $2 per unit. Planned shortages are permitted, where the cost of having one perfect quality finished short is $0.3 per day. The supplier cost of one item of raw material is $5. A returned item of raw material can be sold at a discount price of $2. The supplier holding cost of raw material is $0.01 per unit per day.

The parameters of the problem are D = 100, P = 400, x = 1000, γ = 0.8, π = 0.75, ρ = 0.80, Ap = 5000, K = 5000, h_{mp} = 0.02, h_p = 0.05, h_r = 0.01, C_{mp} = 10, C_p = 20, C_r = 5, C_d = 2, d_m = 0.25, d_f = 0.50, and r = 50, As = 3000, h_{ms} = 0.01, C_{ms} = 10, and r_d = 50.

In a non-collaborative supply chain, the optimal solution obtained by maximizing the function $G_p(Q,S)$ via a numerical search. The search resulted in the following:

Optimal Order Size = Q^* = 8000

Optimal Planned Shortage = S^* = 800

Producer Total Daily Profit = $1318.77

Using the optimal order size, the supplier determines the best value of N=3 with a supplier total profit of $64.43 per day so that the supply chain's total profit is equal to $1383.20.

On the other hand, if the supply chain is coordinated, the best value of N is found to be 1, and the optimal solution is:

Optimal Order Size = Q^* = 10,000

Optimal Planned Shortage = S^* = 1000

Supply Chain Total Daily Profit =$1693.86

Conclusion

The effects of the interaction between the supplier of raw material and the producer of the finished product on the joint performance of both partners in this supply chain were examined. An EPQ model in a coordinated supplier-produced supply chain was developed. The model accounted for both the quality of finished product and raw material in a collaborative supply chain. A mathematical model was formulated so that the supply chain is optimized by determining the order quantity that maximizes the collaborative profit function. A numerical example was provided to compare the collaborative and non-collaborative models. This study showed how collaboration between supply chain members can increase their overall profits.

It is recommended that future research consider probabilistic percentages of good quality items in both raw material and finished products. In another direction, this model can be extended to incorporate other factors such as time value of money, and credit facilities.

References

[1] Bedworth, D. D., and Bailey, J. E. (1987) Integrated Production Control Systems, 2nd Ed. (New York, NY: John Wiley & Sons).

[2] Simpson, N.C. (2001) Questioning the rules of dynamic lot sizing rules, Computers & Operations Research, 28(9), 899–914.

[3] Porteus, E.L. (1986) Optimal lot sizing, process quality improvement and setup cost reduction, Operations Research 34(1), 137-144.

[4] Rosenblatt, M.J., and Lee, H.L. (1986) Economic production cycles with imperfect production processes, IEE Transactions 18, 48–55.

[5] Salameh, M. K., and Jaber, M. Y. (2000) Economic production quantity model for items with imperfect quality, International Journal of Production Economics 64, 59–64.

[6] Hayek, P.A. and Salameh, M.K. (2001) Production lot sizing with the reworking of imperfect quality items produced, Production Planning and Control, 12(6), 584–90.

[7] Chiu, Y.P. (2003) Determining the optimal lot size for the finite production model with random defective rate, the rework process, and backlogging, Engineering Optimization 35, 427-437.

[8] Ozdemir, A.E. (2007) An economic order quantity model with defective items and Shortages, International. Journal of Production Economics 106 (2), 544-549.

[9] Khan, M., Jaber, M.Y., Guiffrida, A.L., and Zolfaghari, S. (2011) A review of the extensions of a modified EOQ model for imperfect quality items, International Journal of Production Economics 132 (1), 1-12.

[10] El-Kassar, A.N., Salameh M., and Bitar, M. (2012a) EPQ model with imperfect quality raw material, Math Balkanica, 26, 123-132.

[11] El-Kassar, A.N., Salameh M., and Bitar, M. (2012b) Effects of time value of money on the EPQ Model with the imperfect quality items of Raw material, Proceedings of Academy of Information and Management Sciences, New Orleans, 16 (1), 11-18.

Figure 1: Perfect Quality Inventory Level

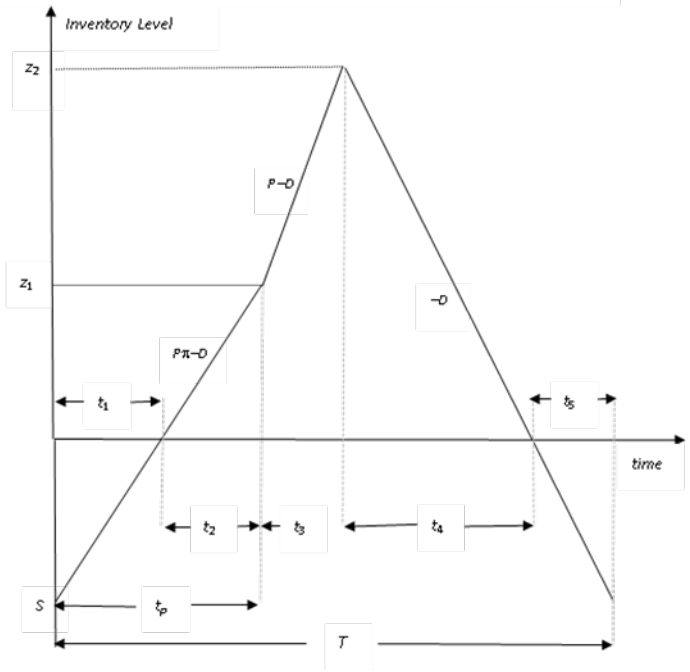


Figure 2: Finished Product Inventory Level

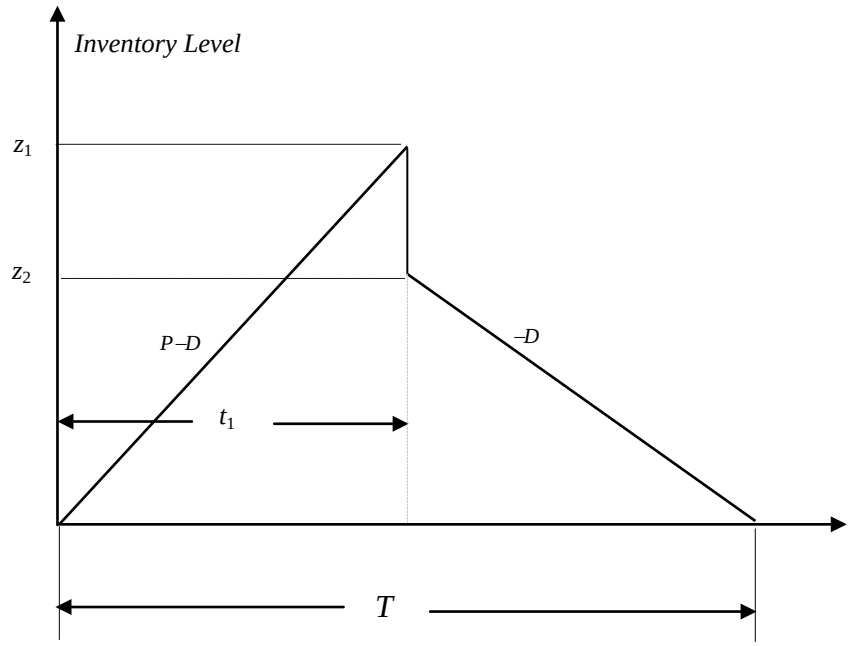
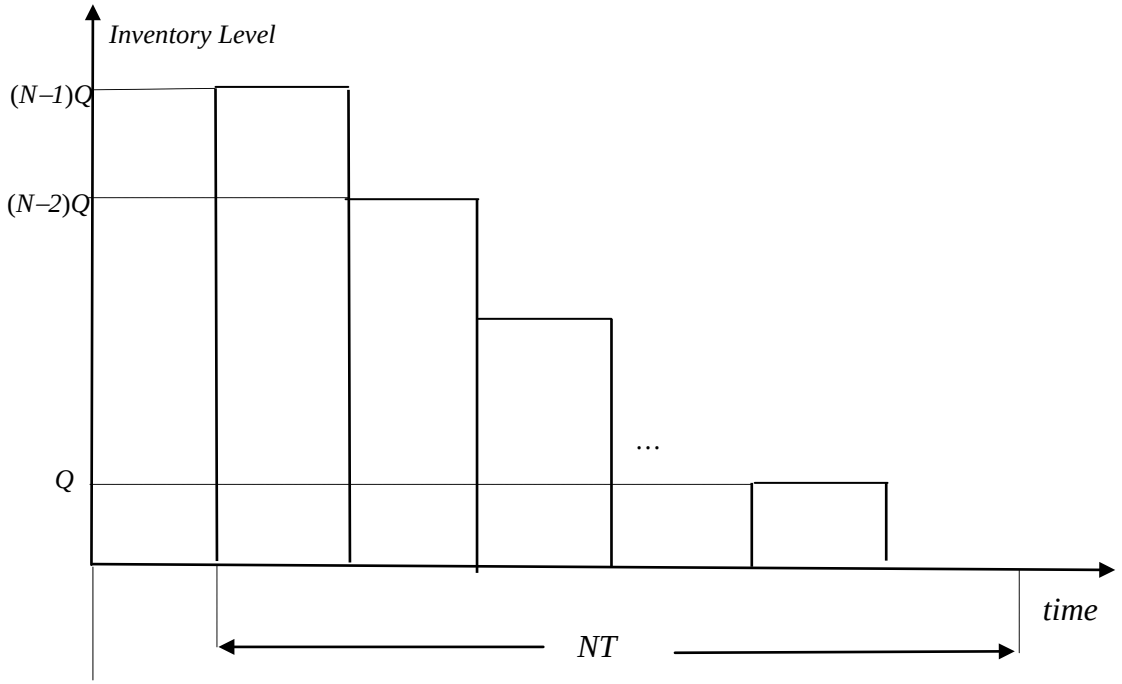


Figure 3: Supplier Inventory Level



Using the Basic Math and the Drawing Software for Calculating the Length of Tube for a Cane of Personalized Dimensions

[1]Z. Damián-Noriega*, F. Beltrán-Carbajal*, E. Montes-Estrada*, G.D. Alvarez-Miranda*, R. Pérez-Moreno*.

***Departamento de Energía, Universidad Autónoma Metropolitana Unidad Azcapotzalco
[1] Presenting and corresponding author: zdn@correo.azc.uam.mx

Abstract

A walking stick is an external help to maintain balance and maintain mobility during the walk of a person who is recovering from an injury to one of his legs. In this paper the analytical method is described for determining the length of pipe needed for the manufacture of a cane custom design, using basic math such as calculating perimeters and trigonometry, as a training exercise for students of mechanical engineering who take the course theoretical and practical manufacturing process because they start the course with deficiencies in basic math. It also briefly describes the use of Autocad software for this purpose. To design the stick, geometry and material commercial canes were considered, and anthropometric measurements of the palm of hand and height to the stick were taken. For a stick 800 mm height and 90 mm grip, a length of 937.14 mm tube is required.

Keywords: Trigonometry, Drawing software, Length of Tube, Walking Stick, Personalized Dimensions.

Introduction

In the courses of Manufacturing Processes I and Workshop Manufacturing Processes I for students of mechanical engineering, is studied and applied the process of bending sheet and pipe, and for enabling the material in this process, it is necessary to use basic math such as calculating perimeters and trigonometry, or the use of engineering software such as Autocad.

Because students who start these courses have gaps in knowledge of basic mathematics, this paper is intended that students see the practical use of basic math, or use Autocad if students already handle it, to determine the tube length required for manufacturing a cane custom dimensions.

Function and correct use of a cane

A cane (Fig. 1) is an external aid for balance and maintain mobility while driving when a person is recovering from an injury to one of his legs. The cane is lighter walking aid, which part of the body weight is transferred, wrist support him; a cane can not and should not hold most of the body weight [1].

The cane should be held with the hand that is on the same side of the functional leg: if the left leg is injured, then the stick must be held with the right hand, and vice versa (Fig 2). When step with the injured leg is given, you must move the stick forward at the same time, supporting part of the body weight on the leg but on his cane.

Method

For custom design cane is considered in principle the design and material of rods available commercially.

Material Selection

The material of commercial canes is aluminum tube of outer diameter 22.2 mm. This material is commercially available as extruded tube and calibrated tube, the tube length is 6.10 m and

the wall thickness of 1.24 mm [2]. The extruded tube was suitable for the bending process, as the calibrated tube fractured to start bending.

Figure 1. Forearm position recommended for use cane

Figure 2. Support on the pole on the right side, left foot injury.

Design parameters

Bending angles. The geometric contour of a commercial cane has five segments (Fig. 3); the internal angle between segments *a* and *c* is 60 ° and between the segments c and *f* is 150 ° (Fig. 3.a). For the initial design of cane, internal angles of 45° and 13° between the segments *a* and c, and between the segments *c* and *f* respectively, were considered (Figure 3.b). But with these angles about 2% over tube length is required (Fig. 3.c).

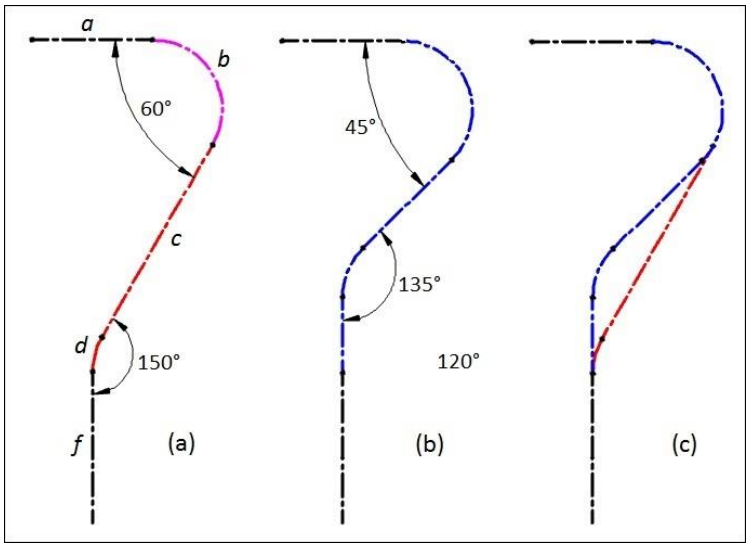


Figure 3. Outline of a stick: a) commercial design and custom design, b) design initially considered, c) Overlapping designs.

Therefore, to our final design the bending angles of commercial sticks were considered (Fig. 3.a). The segment *a* is the segment in which the stick is held, and its length depends on the size of the hand palm of the person (Fig. 4).

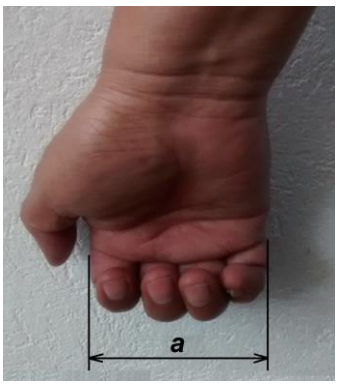


Figure 4. Width of the hand palm.

Bending radii. The internal radius *R* of bending of the two segments *b* and *d* was considered 46.2 mm (Fig. 5), which is the radius having the disk to bend tube 22.2 mm diameter, disk of the manual bender [3] with which practices account for students Manufacturing Process Workshop I.

Cane height H. To determine the appropriate height *H* of the cane, the person standing still with your arms at his sides, to hold the stick (Fig. 1) forearm must be flexed between 15° and 20 ° to the vertical [1] so that the support on the stick to be effective.

Dimension A. For the calculation of the dimension A (Fig. 5), the length of the straight segment *a*, the radius *R* of arc *b* and the diameter ϕ of the tube, are considered. Therefore, using vector algebra [ ], we have (Fig. 6):

$$\boldsymbol{A = a + R + \phi} \qquad (1)$$

Four dimensions are considered for segment ***a***, according to hand measurements: 69.6, 79.6, 89.6 and 99.6 mm. ***R*** = 46.2 mm and ϕ = 22.2 mm.

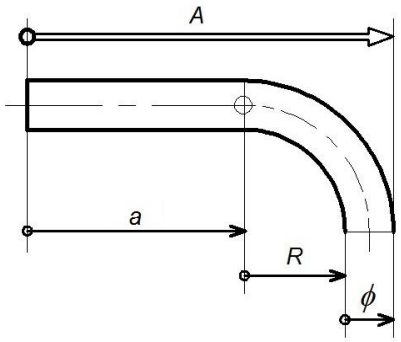


Figure 5. Dimension string for the vector calculation of *A*.

Dimensions *H* and *A* can be checked after the manufacture of the cane.

Dimension B. If one considers that the supporting force of the person (*0.5P*) is applied to the center of *a*, the line of action of the force must coincide with the longitudinal axis of the straight segment ***f*** to avoid unnecessary bending moments.

To ensure the above, the dimension *B* must be checked after the manufacture of the cane.

Therefore, using vector algebra (Fig. 6), we have:

$$\boldsymbol{B = 0.5\,(a - \phi)} \qquad (2)$$

Substituting values, we have (Fig. 7):

B = 23.7, 28.7, 33.7 and 38.7 mm

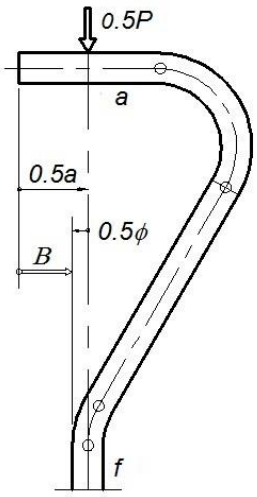


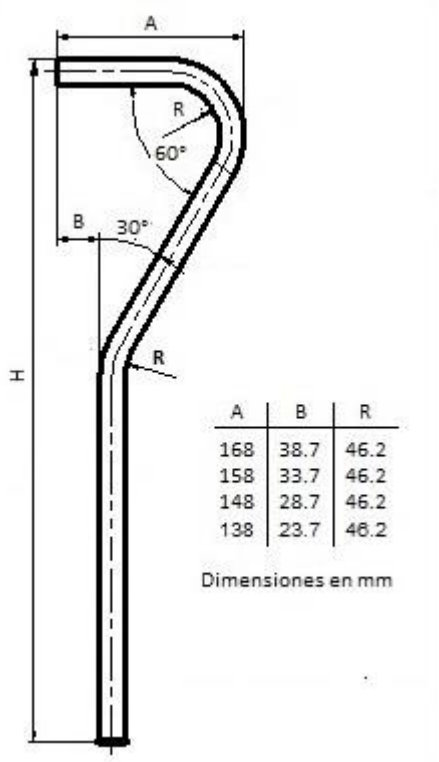
Figure 6. Dimension string for the vector calculation of *B*.

Calculation of the tube length

For the calculation of the tube length, the analytical method or drawing software (Autocad) can be used.

Analytical method

This method involves applying basic math: trigonometry, calculus of perimeters, and vector algebra.



A	B	R
168	38.7	46.2
158	33.7	46.2
148	28.7	46.2
138	23.7	46.2

Dimensiones en mm

Fig. 7. Custom dimensions for the cane.

Fig. 8 shows the identification of the cane segments, considering the neutral axis of the tube (where no have any effort), to calculate the tube length *L* necessary for the manufacture of cane, given by Eq. (3):

$$\boldsymbol{L} = a + b + c + d + f \qquad (3)$$

$$a = \text{data}$$

Calculation of the segment b:

According to the calculation perimeters:

$$\boldsymbol{b} = 2\,\pi.c3 / 3 \qquad (4)$$

$$c3 = (R + 0.5\phi) \qquad (5)$$

Calculation of the segment c:

First the length of its horizontal projection ***Ch*** is determined using vector algebra [4], the dimension string to calculate Ch is (Fig.8):

$$\boldsymbol{Ch} = (\boldsymbol{a - c2}) + (\boldsymbol{c3 - c4 - c1}) \qquad (6)$$

$$\boldsymbol{c2} = B + 0.5\ \phi \qquad (7)$$

Using vector algebra, Fig. 9 shows that:

$$\boldsymbol{c4 = c3 - c5} \qquad (8)$$

By trigonometry , Figs. 8 and 9 show that:

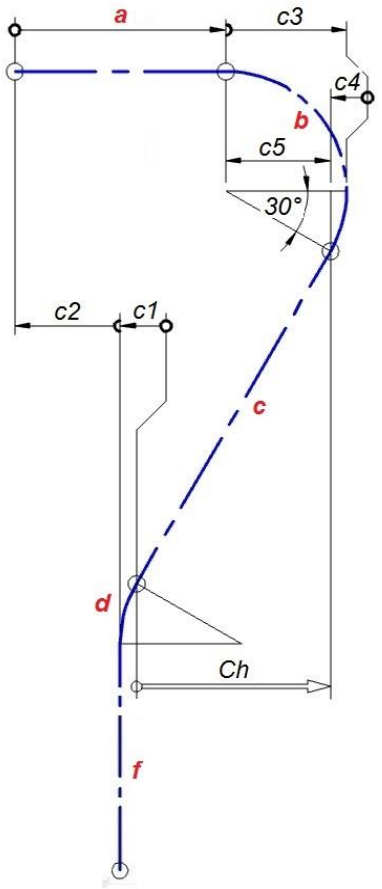


Fig. 8. Minimum dimension string *Ch* to calculate the cane segment *c*.

*

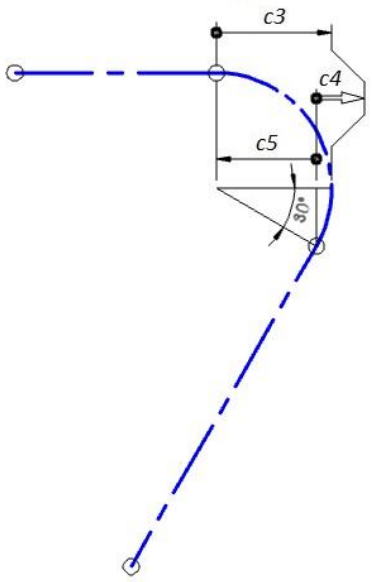


Fig. 9. Dimension string *c4*.

$$\boldsymbol{c5} = c3.\cos 30° \quad (9)$$

Fig. 8 shows that:

$$\boldsymbol{c1} = \boldsymbol{c4} \quad (10)$$

The value of the parameters *c1*, *c3*, *c4* and *c5* is constant for any size stick.

Finally, also by trigonometry we have:

$$\boldsymbol{c} = Ch / \cos 60° \quad (11)$$

Calculation of the segment d:

According to the calculation perimeters:

$$\boldsymbol{d} = 2\,\pi.c3 / 12 \quad (12)$$

Calculation of the segment f:

Using vector algebra, Fig. 10 shows that:

$$\boldsymbol{f} = \boldsymbol{h} - \boldsymbol{c7} \quad (13)$$

where ***h*** is the height of cane considering the position of the neutral axis thereof:

$$\boldsymbol{h} = \boldsymbol{H} - \mathbf{0.5}\,\phi \quad (14)$$

Fig. 11 shows that:

$$\boldsymbol{c7} = \boldsymbol{c6} + \boldsymbol{Cv} + \boldsymbol{c6} + \boldsymbol{c3} \quad (15)$$

By trigonometry (Fig. 11):

$$c6 = c3.\ \mathrm{Sen}30° \quad (16)$$

Observing Figs. 8 and 11:

$$Cv = Ch.\tan 60° \quad .(17)$$

Substituting Eq. 14 and c7 values in Eq. 13, the segment f have values dependent height H of the stick.

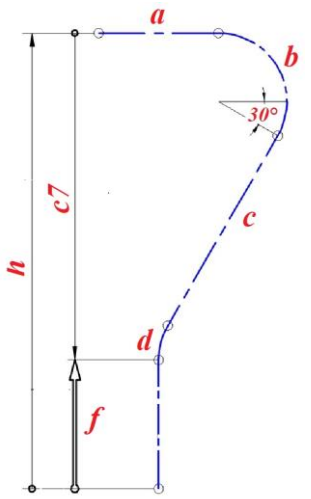


Fig. 10. Dimension string *f*.

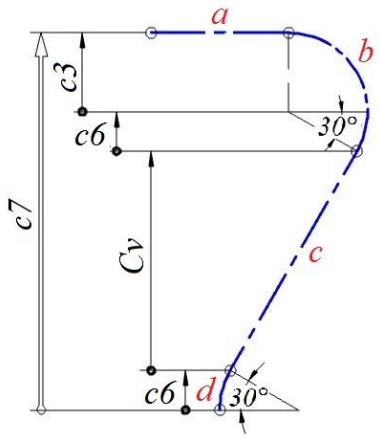


Fig. 11. Dimension string *c7*.

Autocad method

Initially drawing stick to scale 1 is prepared (Fig. 12), with the measures corresponding to the palm (*a*) and height (*H*) of the stick, and the *LIST* command is used pointing each segments to display the value thereof, and having displayed the value of the length of each of the segments, the values are added and the tube length L necessary for the manufacture of the stick is obtained. They may also bind all segments to form a single identity and then applies *LIST*.

Each student makes calculation for his cane, which will serve for one of their relatives.

Results

Analytical method

As an example, for a = 89.6 mm and H = 830 mm, the the tube lenght L is 967.14 mm. See Table 1, that shows the value of each segment and the general dimensions of cane.

Table 2 has been prepared to quickly verify that student calculations are correct.

Table 1. Calculation example: tube length *L* for *a* = 79.6 mm and *H* = 830 mm.

Segment	Value (mm)	Calculated with Equation
A	158	1
B	33.7	2
B	120	5, 4
C	173.49	9, 8, 10, 7, 6, and 11
D	30	12
F	554	17, 16, 15, 14 and 13
L	967.14	3

Table 2. Tube length *L*, dimensions for the cane, and values of each segment.

Dimension or segment, and tube length (mm)	Dimension *a* (mm)			
	69.6	79.6	89.6	99.6 mm
A	138	148	158	168
B	23.7	28.7	33.7	38.7
B	120.00			
C	153.49	163.49	173.49	183.49
D	30.00			
F	H – 258.63	H – 267.29	*H* – 275.95	*H* – 284.61
L	*L* = H + 114.5	H + 125.8	*L* = *H* + 137.14	*L* = *H* + 148.48

Autocad method

For the above analytical example, Fig. 12 shows the baston drawing scale and the Autocad text window, where the application of the LIST command can be observed, to display the value of segment *c*. As expected, the above value is the same as that obtained analytically.

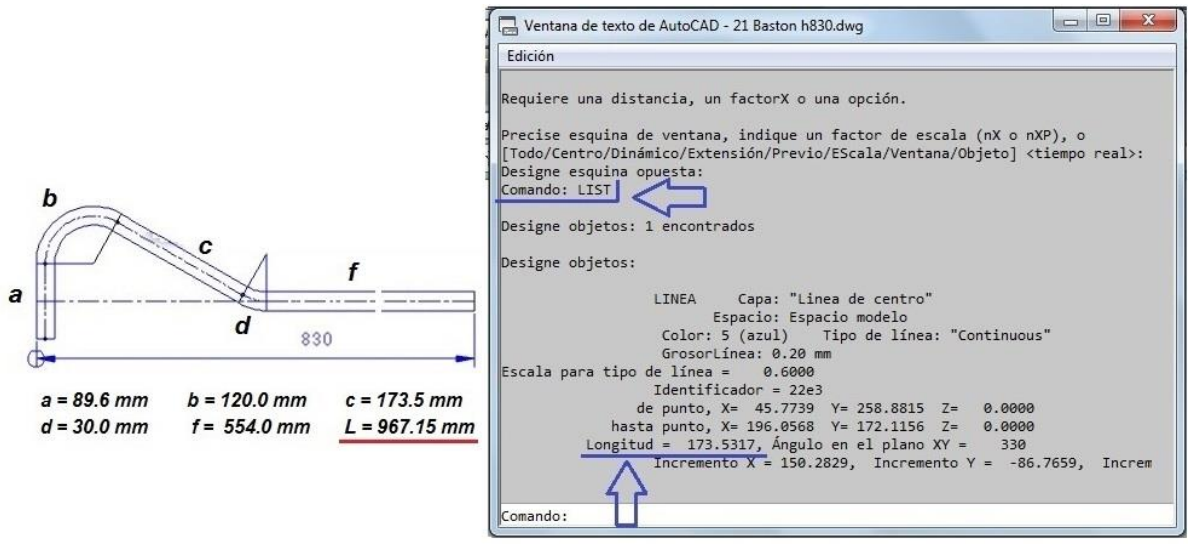


Fig. 12. Autocad text window showing the value of segment *c* of the cane.

Conclusions

This paper presented two methods for calculating the length of pipe required for the manufacture of a cane custom dimensions. For the analytical method, equations have been deduced for calculating dimensions *A* and *B* of the cane, which can be verified after manufacture, and the height *H* depends on the stature of the person. Dimension *A* depends on the width of the palm of the person who will use the cane.

The student must measure both the width of the palm as the appropriate cane height according to the stature of the person who will use the cane, and using this data, the student must establish the corresponding equations to calculate the value of each segment, applying the calculation of perimeters, trigonometry and vector algebra.

Therefore, the analytical method is useful to the student to remember basic math, and with these, calculate the length of tube required to manufacture his own cane.

References

1. Cómo usar correctamente el bastón. http://es.wikihow.com/usar-un-bast%C3%B3n-correctamente. (consultation October 30, 2015):
2. Catálogo de Metales Diaz S.A. de C.V. (2015). Ave. Dr. Gustavo Baz No. 224, Col. Bella Vista, Tlalnepantla de Baz, Estado de México, México. C.P. 54080. Tel. 0155-5360-3545. www.metalesdiaz.com.mx
3. Villanueva-Pruneda Sergio *et al*. Manual de fichas técnicas. Máquinas-herramienta. Taller Mecánico. 1a Edición 2013. Comité editorial de la División de Ciencias Básicas e Ingeniería. UAM-A. ISBN: 978-607-28-0207-0
4. Mendez-Hinojosa A. Matemáticas I enfoque por competencias DGB. Editorial Santillana. ISBN: 978-607-012-5638X

An Average Nodal Pressure Face-based Smoothed Finite Element Method (FS-FEM) for 3D nearly-incompressible solids

†Chen Jiang[1,2] , G.R. Liu[2]

[1]College of Mechanical Engineering and Vehicle Engineering, Hunan University, Changsha, China.
[2]Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, USA.

*Presenting author: jc2007@hnu.edu.cn
†Corresponding author: jc2007@hnu.edu.cn

Abstract

Average Nodal Pressure (ANP) is a simple and useful technique to alleviate the volumetric locking for all element types of standard FEM, including linear 3-node Triangles (T3) and 4-node Tetrahedrons (T4). However, standard FEM using T3 and T4 elements has shown interior accuracy and convergence than FS-FEM using same elements in previous literatures. In this paper, we combine FS-FEM and ANP to propose FS-FEM/ANP using linear T4 element for nearly-incompressible solids. The proposed FS-FEM/ANP-T4 is used to calculate a benchmark, 3D Lame problem. This 3D Lame benchmark proves that FS-FEM/ANP-T4 is free of volumetric locking, more accurate and converging faster than FEM/ANP-T4. Meanwhile, FS-FEM/ANP-T4 still possesses the remarkable endurance of mesh distortion. Also, a rubber beam applied with pressure is calculated to verify the good stability of FS-FEM/ANP-T4 on large deformation. In addition, proposed FS-FEM/ANP-T4 is used to simulate an application, a rubber hanger loaded with exhaust gravity. Comparisons in these examples with analytical results and other methods results show FS-FEM/ANP-T4 is a better alternative of FEM/ANP-T4.

Keywords: Average Nodal Pressure, FS-FEM, Nearly-incompressible, Tetrahedron.

Introduction

Linear 3-nodes triangles (T3) and 4-nodes tetrahedrons (T4) are simplest elements for 2D and 3D problems. Because the piecewise linear shape function is used, the stress and strain are uniformly distributed within element. Consequently, gauss integration with one gauss point is enough. Therefore, T3 and T4 element have fastest speed. More importantly, T3 and T4 elements can be automatic generated and h-adaptive mesh refined for any geometry. On the contrary, quadrilaterals and hexahedrons can only mesh certain topology types of geometry automatically.

However, the over-stiff linear shape function of the standard FEM using T3 and T4 elements cause poor accuracy and convergence and volumetric locking issue. Therefore, linear T3 and T4 elements are not recommended by most FEM software packages. To safely use triangles and tetrahedrons for complex geometry, second-order 6-node Triangle (T6) and 10-node Tetrahedron (T10) are often suggested. But the much more Degrees Of Freedom (DOFs) of T6 and T10 than T3 and T4 cause much more memory usage and computation cost. Another approach to improve T3 and T4 element is to use the Smoothed Finite Element Method (S-FEM), based on G-space theory and weakened weak form (W2) [1]–[3].

S-FEM adopts the gradient smoothing to gain the improvement for T3 and T4 element. The gradient smoothing is a generalization of the strain smoothing technique for Element-Free

Galerkin (EFG) method [4]. Based on different gradient smoothing techniques applied to T3 and T4 element, we will have different types of S-FEMs. For 3D compressible problem with T4 element, S-FEM is classified as cell-based S-FEM (CS-FEM) [5,6], face-based S-FEM (FS-FEM) [7], node-based S-FEM (NS-FEM) [8], alpha S-FEM (αS-FEM) [9] and 3D-edge-based S-FEM (3D-ES-FEM) [10,11]. Among these variations of S-FEMs, FS-FEM and 3D-ES-FEM have been demonstrated with better accuracy and convergence than FEM. Meanwhile, all these variations of S-FEMs are spatial stable and temporal stable, except for NS-FEM which is only temporal instable. However, due to the "sufficient softness", only NS-FEM is volumetric locking free. Hence, a selective S-FEM [12–14] is developed by combining advantages of FS-FEM or 3D-ES-FEM and NS-FEM to deal with volumetric locking of incompressible solids. The selective S-FEM is still temporal stable. Recently, some temporal stabilization techniques are also proposed for NS-FEM [15–18]. Also, a bubble enriched S-FEMs are also proposed to further alleviate pressure instability when solid has very high bulk modulus [19–21].

On the other hand, in FEM, many researchers endeavored to rectify the volumetric locking of linear T3 and T4 elements. In this paper, all these approaches are classified into six types, (1) Mixed-enhanced elements. Different approximations of displacement field and pressure field are used to yield more displacement Degrees Of Freedom (DOFs) than pressure DOFs, like MINI element enriched with "bubble function" [23] and element using Hu-Washizu three fields variational theorem [24]; (2) Pressure stabilizations. Additional stabilization term is applied to interpolated pressure field to satisfy the Babuška-Brezzi conditions, like Finite Increment Calculus (FIC) [25], Galerkin Least Square (GLS) method and direct pressure stabilization [26] and so on; (3) Composite pressure fields. Reduce the incompressible constraint by enforcing a constant pressure or strain on a patch of T3 or T4 elements, like F-bar method [27] and so on; (4) Average nodal pressure/strain. Compute the pressure or strain at nodes by averaging pressure and strain of surrounding T3 and T4 elements [22,28–30]; (5) Fractional time stepping. Calculate an intermediate displacement field using governing equation without pressure term, then use the intermediate displacement to calculate pressure at current time step and correct the intermediate displacement field to obtain displacement at current time step, like Characteristic-based Split (CBS) method [31] and fractional time stepping [32]; (6) Selective S-FEM. Like selective integration for 4-node Quadrilaterals (Q4) and 8-node Hexahedrons (H8), Selective S-FEM [12–14,33] use NS-FEM to calculate volumetric part for T3 and T4 elements.

Definitely, the most straightforward methods are definitely the Average Nodal Pressure/Strain (ANP/ANS). Meanwhile, ANP/ANS [30] can also cure the bending locking. Similar to the selective integration, the ANP/ANS can directly be used in explicit dynamic time stepping.

In this paper, the ANP is applied to alleviate volumetric locking for FS-FEM with linear T4 element. We named this method as FS-FEM/ANP-T4. Likewise, we named standard ANP as FEM/ANP-T4. Because FS-FEM/ANP inherits some merits of FS-FEM, a superior performance of FS-FEM/ANP-T4 than FEM/ANP-T4 can be expected. In addition, an Adaptive Dynamic Relaxation (ADRM) is also introduced to speed up the analysis of quasi-static process using explicit time stepping.

The rest sections of this paper are outlined as: section 2 presents the theoretical basis of FS-FEM/ANP-T4; Section 3 mainly presents the computer implementations of explicit FS-FEM/ANP-T4 and FS-FEM/ANP-T4 with ADRM; Section 4 provides examples for verification and performance test; Section 5 draws conclusions.

Theoretical Basis

In this paper, proposed FS-FEM/ANP-T4 incorporates the gradient smoothing and the average nodal pressure. The gradient smoothing brings outperforming accuracy and robustness to S-FEM. But S-FEMs are still volumetric locking except for node-based gradient smoothing. On the other hand, average nodal pressure method [29] is able to cure t volumetric locking of S-FEMs for nearly-incompressible solids.

Gradient Smoothing

Although this paper use T4 element for 3D problem, we still illustrate the gradient smoothing in two dimensional systems. The extension of 2D gradient smoothing to three dimensions is straightforward and trivial. Give a 2D domain Ω , the smoothing gradients of displacement $u_i(\mathbf{x})$ in sub-domain Ω_L of Ω are expressed as

$$\frac{\partial u_i(\mathbf{x}_L)}{\partial x_j} \approx \int_{\Omega_L} \frac{\partial u_i(\mathbf{x}_L)}{\partial x_j} \tilde{w}(\mathbf{x}-\mathbf{x}_L) d\Omega. \tag{1}$$

Use the Gauss-Green's theorem to above equation,

$$\frac{\partial u_i(\mathbf{x}_L)}{\partial x_j} \approx \int_{\partial\Omega_L} u_i(\mathbf{x}_L)\tilde{w}(\mathbf{x}-\mathbf{x}_L)\mathbf{n}d\Gamma - \int_{\Omega_L} u(\mathbf{x}_L)\frac{\partial \tilde{w}(\mathbf{x}-\mathbf{x}_L)}{\partial x_j} d\Omega. \tag{2}$$

where $\tilde{w}$ is the smoothing function whose requirements will be described later, $\partial\Omega_L$ is the outer boundary of sub-domain Ω_L which is also call smoothing domain here, and **n** is the unit outward normal of $\partial\Omega_L$, as illustrated in Figure 1.

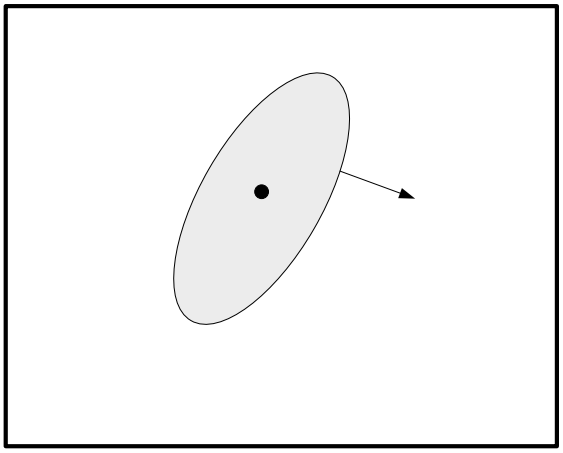


Figure 1 Generic smoothing domain.

The smoothing function in Eq.(1) can be any derivable function. Here, we adopt the suggested piecewise constant function in references [5],

$$\tilde{w} = \begin{cases} 1/A_L, x \in \Omega_L \\ 0. x \notin \Omega_L \end{cases} \tag{3}$$

where A_i is the area of smoothing domain.

With the piecewise constant smoothing function in Eq.(3), the second domain integral will be zero as,

$$\frac{\partial u_i(\mathbf{x}_L)}{\partial x_j} \approx \int_{\partial\Omega_L} u_i(\mathbf{x}_L)\tilde{w}(\mathbf{x}-\mathbf{x}_L)\mathbf{n}d\Gamma \tag{4}$$

As we can see, the calculation of spatial derivatives of displacements is boundary integral now and only need the displacement value. If we further discretize the displacement by FEM, the displacements can be approximated by,

$$u_i(\mathbf{x}) = \sum_{I \in G_L} \Phi(\mathbf{x}_I) u_i(\mathbf{x}_I), i = 1,2,3. \tag{5}$$

where $\Phi(\mathbf{x}_I)$ is the FEM shape function of node I, $u_i(\mathbf{x}_I)$ is the value of displacement at node I. G_L means supporting nodes of the smoothing domain Ω_L.

Hence, the discretized gradient of displacement is derived as,

$$\frac{\partial u_i(\mathbf{x}_I)}{\partial x_j} \approx \sum_{I \in G_L} \left(\frac{1}{A_L} \int_{\partial \Omega_L} \Phi(\mathbf{x}) n_j d\Gamma \right) u_i(\mathbf{x}_I). \tag{6}$$

where n_j is the j-th component of outward unit normal.

Compare Eq.(6) with standard calculation of gradient of displacement, the smoothed derivatives of shape functions $\bar{\Phi}_{I,j}$ are defined as,

$$\frac{\partial \bar{\Phi}_I}{\partial x_j} = \bar{\Phi}_{I,j} = \frac{1}{A_L} \int_{\partial \Omega_L} \Phi(\mathbf{x}) n_j d\Gamma. \tag{7}$$

where only shape function itself is used here, so corresponding mapping of standard FEM is no longer needed which will bring much better robustness of element distortion [6].

We have mentioned several S-FEMs for T4 element in introduction section, such as Cell-based S-FEM (CS-FEM-T4), Node-based S-FEM, Face-based S-FEM (FS-FEM-T4), Edge-based S-FEM (ES-FEM-T4) and alpha S-FEM (αS-FEM). In our previous experience, the FS-FEM-T4 is more accurate and efficient than FEM-T4. The definition of smoothing domain of FS-FEM-T4 is drawn in Figure 2(a). Also, the node-based smoothing domain of NS-FEM-T4 is also presented in Figure 2(b).

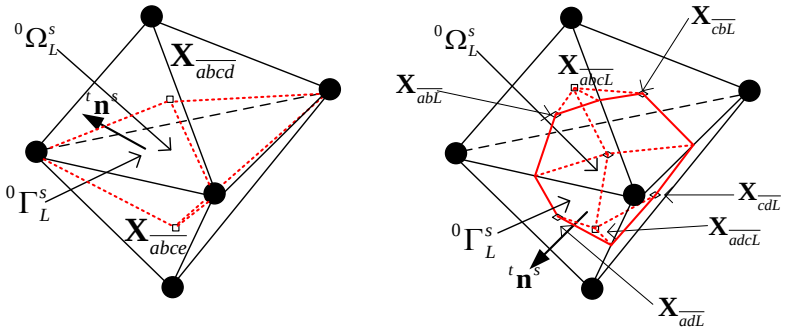


Figure 2 Smoothing domains for FS-FEM-T4 (a) and NS-FEM-T4 (b).

Average Nodal Pressure

We have already mentioned many techniques to alleviate volumetric locking to make non-locking FEM-T4 in introduction section. Among them, the average nodal pressure (ANP) is the simplest [7]–[10].

In standard ANP formulation [7], [8], the pressure is assumed as a constant within the volume associated with one node. For T4 element case, the nodal volume of node I is computed by,

$$V_I = \sum_{e=1}^{N_e} \frac{V_I^e}{4}. \tag{8}$$

where V_I is the nodal volume, N_e is the number of associated elements with node *I*, V_I^e is the volume of element which associate with node *I*.

For geometric nonlinear problems, the nodal volumetric ratio can be calculated by,

$$J_I = \frac{V_I^0}{V_I^n}. \tag{9}$$

in which, V_I^0 is the nodal volume at the initial configuration, V_I^n is the nodal volume at the current configuration.

Then, if problem is homogeneous without other materials, the ANP is given as below,

$$p_I = \kappa(J_I - 1). \tag{10}$$

where κ is the bulk modulus.

Finally, we can use this ANP to get the pressure value at the Gauss points of T4 elements. The whole process will be demonstrated more clear in later sections. In fact, the further investigation about the selective S-FEM shows the NS-FEM has some similarities with ANP to achieve volumetric locking free. The reason is that node-based gradient smoothing also gives a constant strain in node-based smoothing domain which also overlaps the same nodal volume of ANP.

Hyperelastic constitutive models

In this section, we briefly review the finite deformed hyperelasticity. Consider a solid with domain ${}^0\Omega$ at initial configuration, see Figure 3. Then after a large deformation, this solid moves and deforms to current configuration ${}^t\Omega$. The deformation is represented by the motion $\mathbf{x} = \chi(\mathbf{X}, t)$, where $\mathbf{x}$ is the current coordinates and $\mathbf{X}$ denotes initial or reference coordinates.

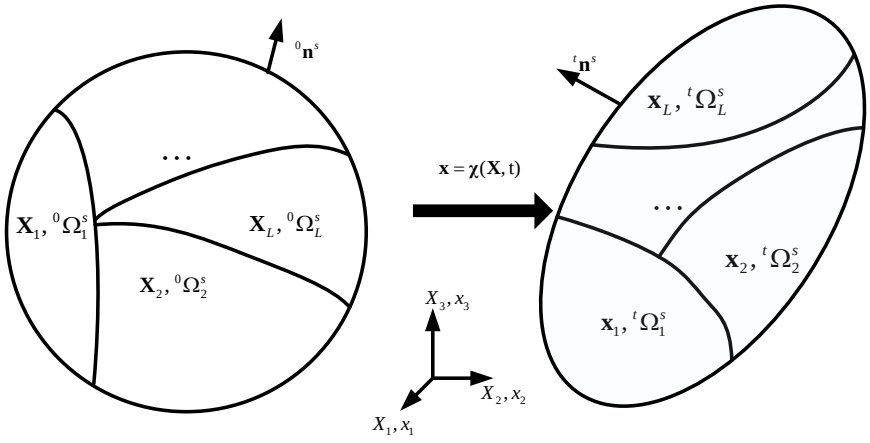


Figure 3 Configurations and deformations of a solid.

In finite deformation, the deformation gradient is important which is defined as,

$$F_{ij} = \frac{\partial x_i}{\partial X_j} + \delta_{ij} = \frac{\partial u_i}{\partial X_j} + \delta_{ij}. \tag{11}$$

After substituting smoothed strain of S-FEM, the smoothed deformation gradient is given as,

$$\bar{F}_{ij} = \sum_{I \in G_L} \bar{\Phi}_{I,j}(\mathbf{X}_L) u_{Ii} + \delta_{ij}. \tag{12}$$

With Eq.(12), we can get the smoothed Green strain as follows,

$$\bar{E}_{ij} = \frac{1}{2}\left(\bar{F}_{ki}\bar{F}_{kj} - \delta_{ij}\right). \tag{13}$$

Meanwhile, the smoothed right Cauchy-Green tensor $\bar{\mathbf{C}}$ is calculated as below,

$$\bar{C}_{ij} = \bar{F}_{ki}\bar{F}_{kj} \tag{14}$$

We can also get the three invariants of $\bar{\mathbf{C}}$ which are often treated as basic variables of hyperelastic material models,

$$\bar{I}_1 = \bar{C}_{ii},\ \bar{I}_2 = \frac{1}{2}\left[\left(\bar{C}_{ii}\right)^2 - \left(\bar{C}_{ij}\bar{C}_{ij}\right)\right],\ \bar{I}_3 = J^2 = \det(\bar{\mathbf{C}}). \tag{15}$$

where the third invariant $\bar{I}_3$ of $\bar{\mathbf{C}}$ also relates to the volumetric ratio.

The strain energy density of hyperelastic material is often decoupled into deviatoric and volumetric parts. Here the general isotropic strain energy density functions is expressed as,

$$\Psi(\bar{I}_1, \bar{I}_2, \bar{J}) = \Psi^{dev}(\bar{J}_1, \bar{J}_2) + \Psi^{vol}(\bar{J}). \tag{16}$$

where *dev* and *vol* denote the deviatoric and volumetric part of strain energy, respectively. And $\bar{J}_1 = \bar{I}_1 \bar{I}_3^{-1/3}$ is the first invariant of modified $\hat{\mathbf{C}} = \bar{I}_3^{-2/3}\bar{\mathbf{C}}$, $\bar{J}_2 = \bar{I}_2 \bar{I}_3^{-2/3}$ is the modified second invariant of $\bar{\mathbf{C}}$.

Although many isotropic hyperelastic strain energy density functions are proposed, the most widely used form of Ψ^{vol} is,

$$\Psi^{vol}\left(\bar{J}\right) = \frac{1}{2}\kappa\left(\bar{J} - 1\right)^2. \tag{17}$$

where κ is the bulk modulus and this part will be cared by ANP technique.

For a given hyperelastic strain energy function, the second Piola-Krichhoff (PK2) stress tensor which is also the stress measure in Total Lagrangian formulation can be calculated by FS-FEM/ANP-T4 and ES-FEM/ANP-T4,

$$\bar{\mathbf{S}} = 2\frac{\partial \Psi^{dev}}{\partial \bar{\mathbf{C}}^{FS}} + 2\frac{\partial \Psi^{vol}}{\partial \mathbf{C}^{ANP}} = \bar{J}^{-2/3}\mathrm{Dev}\tilde{\mathbf{S}} + Jp^{ANP}\mathbf{C}^{-1}. \tag{18}$$

where $\bar{\mathbf{S}}$ is the smoothed PK2 stress tensor, the *FS* is short for FS-FEM-T4, operator $\mathrm{Dev}(\bullet) = (\bullet) - (1/3)\left[(\bullet) : \bar{\mathbf{C}}\right]\bar{\mathbf{C}}^{-1}$.

In above equation, a new fictitious PK2 stress tensor is also introduced. It can be expressed as,

$$\tilde{\mathbf{S}} = 2\frac{\partial \Psi^{dev}(\bar{J}_1, \bar{J}_2)}{\partial \hat{\mathbf{C}}}. \tag{19}$$

Readers can find more details about calculating the PK2 stress tensor of hyperelastic material models in reference [11].

Total Lagrangian formulations of explicit S-FEM/ANP-T4

For FEM discretization of finite deformation with Lagrangian mesh, we select the Total Lagrangian (T.L.) formulation. For temporal discretization, the explicit time integration is selected which only needs internal nodal forces.

Still consider the domain ${}^0\Omega$ in Figure 3 at reference configuration with boundary ${}^0\Gamma$. The density is ρ_0, and a body force is applied. On the velocity boundaries ${}^t\Gamma_v$, $v_i(\mathbf{x},t)=\hat{v}_i(\mathbf{x},t)$. On the traction boundaries, $n_j\sigma_{ij}=h_i$ is applied. And the initial conditions are $\mathbf{v}(\mathbf{X},0)=\mathbf{v}_0(\mathbf{X})$ and $\mathbf{u}(\mathbf{X},0)=\mathbf{u}_0(\mathbf{X})$.

The energy in T.L. formulation for explicit dynamic is expressed as follow (without damping),

$$\prod=\overline{\prod}^{int}(\mathbf{u})-\prod^{ext}(\mathbf{u})+\prod^{kin}(\mathbf{u}),\ \overline{\prod}^{int}=\int_\Omega \Psi d\Omega. \tag{20}$$

where $\overline{\prod}^{int}$ is the internal energy, $\prod^{ext}$ is the external energy and $\prod^{kin}$ is the kinetic energy.

In this paper, because the ANP/S-FEM is used, the strain energy is split into deviatoric and volumetric parts like below,

$$\overline{\prod}^{int}=\overline{\prod}^{int,dev}+\overline{\prod}^{int,vol} \tag{21}$$

We directly give the semi-discrete equations of Eq.(20) after taking variation with smoothed Galerkin weak form [3],

$$\mathbf{M}\ddot{\mathbf{u}}=\mathbf{f}^{ext}-\overline{\mathbf{f}}^{int,dev}-\mathbf{f}^{int,vol}. \tag{22}$$

where,

$$\mathbf{M}_{IJ}=\int_{{}^0\Omega}\rho_0[\mathbf{\Phi}_I(\mathbf{X})]^T[\mathbf{\Phi}_J(\mathbf{X})]d\Omega. \tag{23}$$

$$\overline{\mathbf{f}}_I^{int,dev}=\int_{{}^0\Omega}\left[\overline{\mathbf{B}}_I^{FS}\right]^T\left\{\overline{\mathbf{P}}^{dev}\right\}d\Omega=\sum_L\left[\overline{\mathbf{B}}_I^{FS}\left(\mathbf{X}_L\right)\right]^T\left\{\overline{\mathbf{P}}^{dev}\left(\mathbf{X}_L\right)\right\}{}^0A_L^s. \tag{24}$$

$$\mathbf{f}_I^{int,vol}=\int_{{}^0\Omega}\left[\mathbf{B}_I^{ANP}\right]^T\left\{\mathbf{P}^{ANP}\right\}d\Omega=\sum_L\left[\mathbf{B}_I^{ANP}\left(\mathbf{X}_L\right)\right]^T\left\{\mathbf{P}^{ANP}\left(\mathbf{X}_L\right)\right\}{}^0A_L^s. \tag{25}$$

$$\mathbf{f}_I^{ext}=\int_{{}^0\Omega}\left[\mathbf{\Phi}_I\left(\mathbf{X}\right)\right]^{\mathrm{T}}\left\{\mathbf{b}\right\}_I d\Omega+\int_{{}^t\Gamma}\left[\mathbf{\Phi}_I\left(\mathbf{X}\right)\right]^{\mathrm{T}}\left\{\mathbf{h}\right\}_I d\Gamma. \tag{26}$$

In above equations, $\mathbf{M}$ can be lumped mass matrix or consistent mass matrix. $\overline{\mathbf{f}}_I^{int,dev}$ is the smoothed deviatoric internal force vector which is calculated by FS-FEM-T4. $\overline{\mathbf{f}}_I^{int,vol}$ is the volumetric internal force vector calculated by ANP method. $\mathbf{P}$ is the first Piola-Krichhoff (PK1) stress tensor. $\overline{\mathbf{B}}_I^{FS}$ is the strain-displacement relation matrix of *I*-th node using FS-FEM-T4. $\mathbf{B}_I^{ANP}$ is the strain-displacement relation matrix of *I*-th node using ANP which is identical to corresponding matrix of FEM-T4. More detailed equations can be found in reference [12], [13].

Then, we can use the explicit central difference scheme to implement the time integration. First, calculate the acceleration at step *n* using Eq.(22),

$$\mathbf{M}\ddot{\mathbf{u}}^{n} = \mathbf{f}^{ext}(\mathbf{u}^{n}, t^{n}) - \bar{\mathbf{f}}^{int,dev}(\mathbf{u}^{n}, t^{n}) - \mathbf{f}^{int,vol}(\mathbf{u}^{n}, t^{n}). \tag{27}$$

Then, update velocity,

$$\mathbf{v}^{n+1/2} = \mathbf{v}^{n-1/2} + \Delta t^{n}\ddot{\mathbf{u}}^{n}. \tag{28}$$

where Δt is the time step which is constant here.

Finally, update displacement,

$$u^{n+1} = u^{n} + v^{n+1/2}\Delta t. \tag{29}$$

From the procedures of central difference scheme, it is no need to solve linear equations systems. And when lumped mass matrix is employed, the calculation of acceleration $\ddot{\mathbf{u}}$ is purely element-by-element division of two arrays which is fast and also much lesser memory usage. However, we should satisfy the conditional temporal stability of explicit central difference scheme. In the whole analysis, time step must always smaller than the critical time step which is expressed below,

$$\Delta t < \Delta t_{crit} \le \min\left(l_e / c_e\right). \tag{30}$$

where, l_e is the characteristic length of element, c_e is sound speed of this element. The calculations of these two quantities can be found in nonlinear FEM book [14].

Adaptive Dynamic Relaxation of ANP/S-FEM

Explicit time stepping can simulate the quasi-static deformation by using a quite number of time steps. To accelerate the calculation, an adaptive dynamic relaxation (ADR) method in reference [15] are adopted by introducing the mass-scaling and mass-proportional artificial damping into governing equations. Meanwhile, the loads are divided into several load steps to apply. Furthermore, in every load step, the pseudo time stepping is used to achieve quasi-static state. The equilibrium equation at *m* pseudo time step in *n* load step is given as,

$$\mathbf{M}^{fict}\ddot{\mathbf{u}}^{n,m} = \mathbf{f}^{ext}(\mathbf{u}^{n}, t^{n}) - \bar{\mathbf{f}}^{int}(\mathbf{u}^{n,m}, t^{m}) - f^{damping}(\mathbf{u}^{n,m}, t^{m}). \tag{31}$$

where $\mathbf{M}^{fict}$ is the fictitious mass matrix by scaling from original mass matrix, $f^{damping}(\mathbf{u}^{n,m}, t^{m}) = c_d\mathbf{M}^{fict}\mathbf{v}(t^{m-1/2})$ is the damping force with mass-proportional damping coefficient c_d, m is counter for the pseudo time step in ADR.

To check if system has reached the quasi-static state, the following criterion for displacement residual r^u is applied,

$$r^{u} = \left\|u(t^{n+1}) - u(t^{n})\right\| / \left\|u(t^{n})\right\| < e_{adm}. \tag{32}$$

where e_{adm} is a very small positive value which is set as 10^{-6} for all cases in this study, $\|\bullet\|$ is the L2-norm.

Theoretically, $\mathbf{M}^{fict}$ and c_d can be any values in calculation. However, there exist optimal values to achieve fastest convergence to quasi-static state. Many literatures provide massive methods to evaluate the desire $\mathbf{M}^{fict}$ and c_d. In this paper, we select one of simplest ADR algorithm from reference [15]. This ADR only needs to scale the mass matrix to make the

critical pseudo time step always larger than 1 for every element， see Eq.(30). However, other ADRs scale the mass matrix based on the element tangent stiffness matrices [16], [17] which are not necessary for explicit dynamic FEM.

When evaluating the optimal damping coefficient c_d , this ADR is using a estimation of stiffness matrix. The calculation of optimal damping coefficient at *m*-th pseudo time step is given as below,

$$c_d = 2\sqrt{\frac{[\mathbf{u}^{m+1}]^T \mathbf{K}^{esti} \mathbf{u}^{m+1}}{[\mathbf{u}^{m+1}]^T \mathbf{M}^{fict} \mathbf{u}^{m+1}}}. \tag{33}$$

where, $\mathbf{K}^{esti}$ denotes the estimation of stiffness which is calculated as below,

$$K_{ii}^{esti} = \frac{F_{int}^{m} - F_{int}^{m-1}}{\Delta t \cdot v_i^{m-1/2}}. \tag{34}$$

Implementation

Flowchart of Explicit FS-FEM/ANP

I. Initialization:
   A. Set initial conditions v^0 and u^0.
   B. $u^0 = 0, n = 0, t = 0$
   C. Compute lumped mass matrix $\mathbf{M}$
   D. Calculate smoothed gradient of shape functions $\bar{\Phi}_{I,i}^{FS}(\mathbf{X})$.
   E. Assemble smoothed $\tilde{\mathbf{B}}^L$ using $\bar{\Phi}_{I,i}^{FS}(\mathbf{X})$.
   F. Call subroutine ***Calculate_Nodal_Force_ANP*** to calculate nodal force vector $\mathbf{f}\left(\mathbf{u}^0, 0\right)$
   G. Calculate acceleration $\mathbf{a}^0 = \mathbf{M}^{-1}\mathbf{f}\left(\mathbf{u}^0, 0\right)$.

II. Temporal loop, *n* = 1: *n_max*
   A. $t^{n+1} = t^n + \Delta t^{n+1/2}, t^{n+1/2} = 1/2\left(t^n + t^{n+1}\right).$
   B. $\mathbf{v}^{n+1/2} = \mathbf{v}^n + \Delta t^{n+1/2}\mathbf{a}^n.$
   C. Impose velocity boundary conditions.
   D. $\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta t^{n+1/2}\mathbf{v}^{n+1/2}.$
   E. Call subroutine ***Calculate_Nodal_Force_ANP*** to calculate nodal force vector $\mathbf{f}\left(\mathbf{u}^{n+1}, t^{n+1}\right)$.
   F. Calculate acceleration $\mathbf{a}^{n+1} = \mathbf{M}^{-1}\mathbf{f}\left(\mathbf{u}^{n+1}, t^{n+1}\right)$.
   G. $\mathbf{v}^{n+1} = \mathbf{v}^{n+1/2} + \left(t^{n+1} - t^{n+1/2}\right)\mathbf{a}^{n+1}.$
   H. Update the time step counter $n+1 \to n$, $\mathbf{v}^{n+1} \to \mathbf{v}^n$, $\mathbf{u}^{n+1} \to \mathbf{u}^n$, $\mathbf{a}^{n+1} \to \mathbf{a}^n$.

Flowchart of FS-FEM/ANP with ADRM

I. Initialization:
   A. Set initial conditions v^0 and u^0.

B. $u^0 = 0, n = 0, t = 0$

C. Compute lumped mass matrix $\mathbf{M}$.

D. Calculate smoothed gradient of shape functions $\bar{\Phi}_{I,i}^{FS}(\mathbf{X})$.

E. Assemble smoothed $\tilde{\mathbf{B}}^L$ using $\bar{\Phi}_{I,i}^{FS}(\mathbf{X})$.

F. Calculate original nodal volume V_a^0 for each node.

G. Call subroutine ***Calculate_Nodal_Force_ANP*** to calculate nodal force vector $\mathbf{f}\left(\mathbf{u}^0, 0\right)$

H. Change the density to make critical time step of every element as $\Delta t_{critial} = 1.05$.

I. Calculate acceleration $\mathbf{a}^0 = \mathbf{M}^{-1}\mathbf{f}\left(\mathbf{u}^0, 0\right)$.

II. Load step loop, *nLS* = 1 : *nLS_max*

A. Calculate external nodal force $\mathbf{f}_{nLS}^{ext}$ at current load step.

B. Check critical time step, if $\min(\Delta\mathbf{t}_{critial}) < 1.001$, change density to retain $\Delta t_{critial} = 1.05$; else, continue.

C. Pseudo temporal loop, *pn* = 1 : *pn_max*

1. $t^{pn+1} = t^{pn} + \Delta t^{pn+1/2}, t^{pn+1/2} = 1/2\left(t^{pn} + t^{pn+1}\right)$.
2. $\mathbf{v}^{pn+1/2} = \mathbf{v}^{pn} + \Delta t^{pn+1/2}\mathbf{v}^{pn+1/2}$.
3. Impose velocity boundary conditions.
4. $\mathbf{u}^{pn+1} = \mathbf{u}^{pn} + \Delta t^{pn+1/2}\mathbf{v}^{pn+1/2}$. .
5. Call subroutine ***Calculate_Nodal_Force_ANP*** to calculate internal nodal force vector $\mathbf{f}^{int}\left(\mathbf{u}^{pn+1}, t^{pn+1}\right)$.
6. Calculate optimal damping coefficient c_d using Eq.(33) and Eq.(34).
7. Calculate the damping nodal force $\mathbf{f}^{damping}\left(\mathbf{v}^{pn+1/2}, t^{pn+1/2}\right) = c_d\mathbf{M}\mathbf{v}^{pn+1/2}$.
8. Calculate acceleration $\mathbf{a}^{pn+1} = \mathbf{M}^{-1}\left(\mathbf{f}^{ext} - \mathbf{f}^{damping} - \mathbf{f}^{int}\right)$.
9. $\mathbf{v}^{pn+1} = \mathbf{v}^{pn+1/2} + \left(t^{pn+1} - t^{pn+1/2}\right)\mathbf{a}^{pn+1}$.
10. Update pseudo-time step counter $pn+1 \to pn$, $\mathbf{v}^{pn+1} \to \mathbf{v}^{pn}$, $\mathbf{u}^{pn+1} \to \mathbf{u}^{pn}$, $,\mathbf{a}^{pn+1} \to \mathbf{a}^{pn}$.
11. Check displacement residual, if $r_d < e_{adm}$, back to step C; else, continue pseudo temporal loop.

Flowchart of Subroutine Calculate_Nodal_Force_ANP in S-FEM (SD-by-SD)

Deviatoric part:

I. **For each SD**: calculate smoothed deformation gradient $\tilde{\mathbf{F}}_{SD}^{n+1}$.

II. **For each SD**: Calculate smoothed right Cauchy-Green strain tensor $\tilde{\mathbf{C}}^{n+1}$.

III. **For each SD**: Calculate the invariants $\tilde{I}_i (i = 1, 2, 3)$ of smoothed right Cauchy-Green strain tensor $\tilde{\mathbf{C}}^{n+1}$.

IV. **For each SD:** Calculate smoothed PK2 stress $\tilde{\mathbf{S}}^{n+1}$ using selected hyperelastic strain energy density function.

V. **For each SD**: Calculate $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}^L + \tilde{\mathbf{B}}^{NL}\left(u^{n+1}, t^{n+1}\right)$.

VI. **For each node**: Calculate smoothed deviatoric internal force vector $\tilde{\mathbf{f}}_{dev}^{int}\left(u^{n+1},t^{n+1}\right)$.

Volumetric part:

I. **For each element**: Calculate volume V_e .

II. **For each node**: Calculate nodal volume $V_a = V_a + V_e / 4.$.

III. **For each node**: Calculate nodal pressure $p_a = \kappa(J_a - 1) = \kappa\left(V_a / V_a^0 - 1\right)$.

IV. **For each element**: Calculate element's pressure $p_e = \frac{1}{n}\sum_{a=1}^{4} p_a$.

V. **For each element**: Calculate volumetric PK2 stress $\tilde{\mathbf{S}}^{n+1}$, and $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}^{L} + \tilde{\mathbf{B}}^{NL}\left(u^{n+1},t^{n+1}\right)$.

VI. **For each node**: Calculate smoothed volumetric internal force vector $\tilde{\mathbf{f}}_{vol}^{int}\left(u^{n+1},t^{n+1}\right)$.

VII. **For each node**: Calculate external force vector $\mathbf{f}^{ext}(t^{n+1})$.

VIII. **For each node**: Calculate $\mathbf{f}\left(u^{n+1},t^{n+1}\right) = \mathbf{f}^{ext}\left(t^{n+1}\right) - \tilde{\mathbf{f}}_{dev}^{int}\left(u^{n+1},t^{n+1}\right) - \tilde{\mathbf{f}}_{vol}^{int}\left(u^{n+1},t^{n+1}\right)$.

Numerical Examples

3D Lame problem

The 3D Lame problem, a 1/8 sphere inflated with internal pressure, is widely used to validate and benchmark numerical methods for 3D solid mechanics. The accuracy and convergence of proposed FS-FEM/ANP with ADRM are tested by comparing with analytical solution. The inner radius $a = 1m$ and outer radius $b = 2m$. The internal pressure applied is $P = 1pa$. This small internal pressure applied here is to coincide with analytical solution from small deformation theory. The mesh of this 3D Lame problem with 2553 nodes is presented in Figure 4. The surfaces on the symmetry planes are all imposed with symmetrical boundary conditions. The material model in this example is the nearly-incompressible Neo-Hookean hyperelastic model with following strain energy density function,

$$\Psi(\bar{I}_1,\bar{I}_2,\bar{J}) = \Psi^{dev}(\bar{J}_1,\bar{J}_2) + \Psi^{vol}(\bar{J}) = C_{10}(\bar{J}_1 - 3) + \frac{1}{2}\kappa(\bar{J}-1)^2. \tag{35}$$

where $C_{10} = 500\,pa$, the value of κ is calculated by user-defined Poisson's ratio ν as below,

$$\kappa = \frac{4(1+\nu)}{3(1+2\nu)}C_{10}.$$

The analytical solution of 3D Lame problem with Neo-Hookean material is available in spherical coordinate system as below,

$$\begin{cases} u_r = \dfrac{Pa^3 r}{4C_{10}(1+\nu)(b^3-a^3)}\left[(1-2\nu)+(1+\nu)\dfrac{b^3}{2r^3}\right], \\ \sigma_r = \dfrac{Pa^3(b^3-r^3)}{r^3(a^3-b^3)}, \\ \sigma_\theta = \dfrac{Pa^3(b^3+2r^3)}{2r^3(b^3-a^3)}. \end{cases} \tag{36}$$

Validation

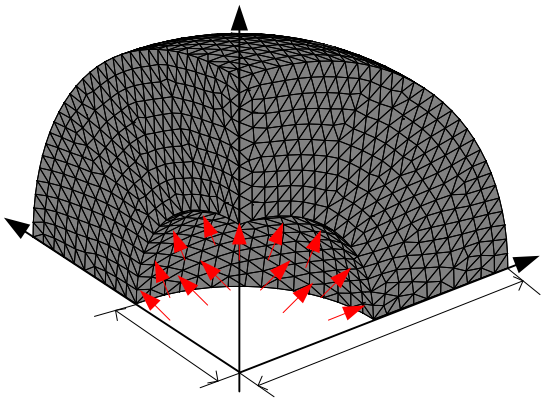


Figure 4 3D Lame configuration and mesh with 2553 nodes.

As a validation, proposed FS-FEM/ANP-T4 with the ADRM is used to solve the 3D Lame problem with Poisson's ration 0.49. Two load steps are used for methods using ADRM. The steady state of each load step is reached when displacement residual is smaller than 1e-6. The radial displacement, radial and tangential stresses on $x \in (0.0,1.0)$ of FS-FEM/ANP-T4 are compared with analytical solution. Besides, displacement and stress solutions of FS/NS-FEM-T4 with ADRM, FS/NS-FEM-T4 with static solver, FEM/ANP-T4 with DRM and FEM-T4 with static solver are also compared in Figure 5 and Figure 6.

FEM-T4 has the worst displacement and stresses accuracies. Besides, the radial and tangential stresses on $x \in (0.0,1.0)$, S11 and S33 are components of pressure. Therefore, the oscillations of S11 and S33 are just the pressure check-board issue. On the other hand, methods with ANP and NS-FEM to deal with volumetric deformation show much better performances than FEM-T4 on accuracy and stability of pressure.

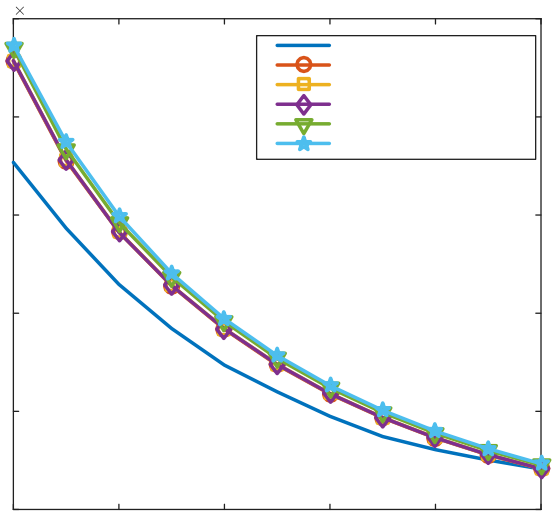


Figure 5 The radial displacements on line x =(0.0, 1.0) of different methods.

To further quantify the oscillation level of these methods, the absolute S33 errors on each node of line $x \in (0.0,1.0)$ are plotted in Figure 7. Proposed FS-FEM/ANP has smoother changes of S33 than the rest with averaged absolute errors as 0.02773. The averaged absolute errors are 0.03701 for FEM/ANP and 0.0458 for FS/NS-FEM. Hence, our implementation of FS-FEM/ANP with ADRM is correct and ANP can obtain smoother pressure distribution than NS-FEM.

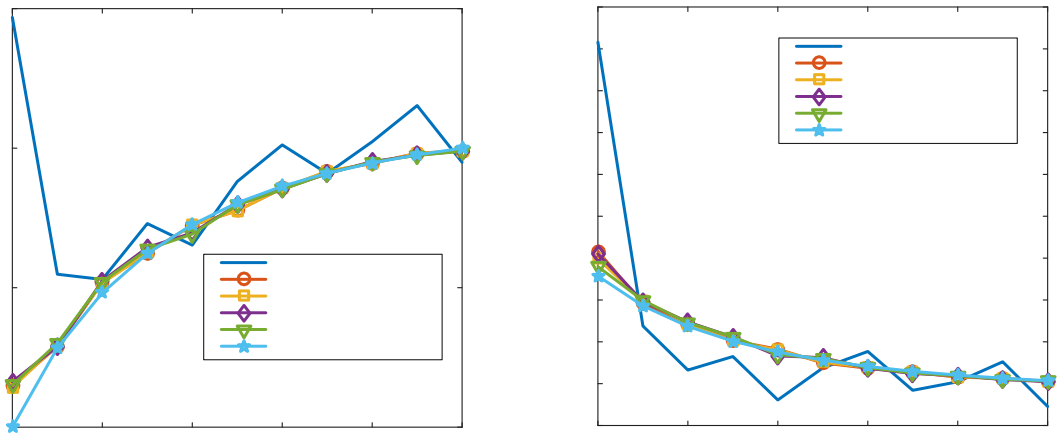


Figure 6 The radial stress σ_{xx} (a) and tangential stress σ_{zz} (b) on line x = (0.0, 1.0) of different methods.

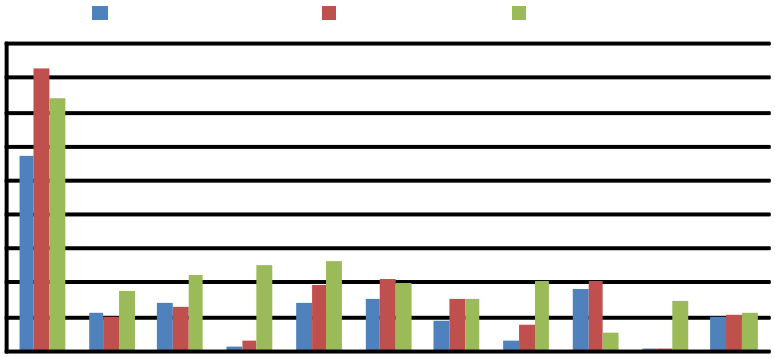


Figure 7 Absolute errors of σ_{zz} on line x =(0.0, 1.0) of different methods, and average absolute errors of different methods are 0.02773 (FS-FEM/ANP), 0.03701 (FEM/ANP), and 0.0458 (FS/NS-FEM).

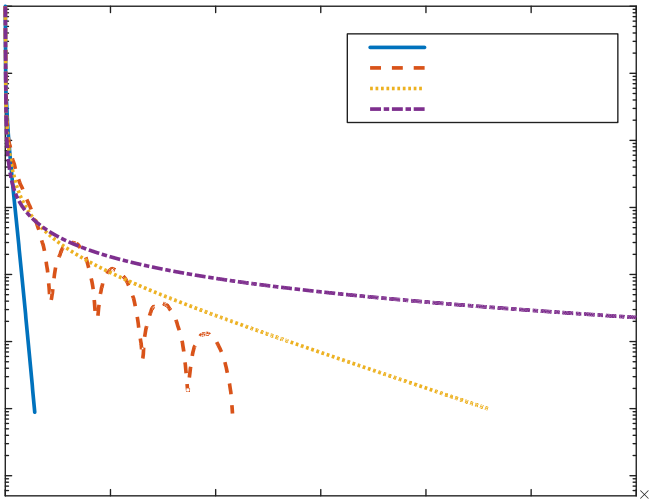


Figure 8 The displacement residual histories of Adaptive Dynamic Relaxation (ADRM) and Conventional Explicit Dynamic Relaxation (CEDRM) using FS-FEM/ANP-T4.

We also tested presented ADRM using this 3D Lame problem. As mentioned before, three load steps are used here to gradually apply the external pressure loading. The Conventional Explicit Dynamic Relaxation (CEDRM) with different damping coefficients is also adopted as comparisons. The number of pseudo time steps of the first load step to reach the steady state is used as indicator of the performance of ADRM and CEDRM. The residual histories of different DRMs are plotted in Figure 8. CEDRMs with damping coefficient 10.0 and 100.0 are with over damping effects; the latter can't satisfy the criterion even after 100,000 pseudo time steps. CEDRM with damping coefficient 1.0 reaches steady state much faster despite of the under damping effect. As supposed, ADRM can straightly reach steady state without need to tune damping coefficient.

As Average Nodal Pressure (ANP) technique has been incorporated into FS-FEM, first time for S-FEM family, its endurance of volumetric locking is also tested, see 错误!未找到引用源。 and Figure 9. Here, the L2-norm of relative radial displacement is used to indicate the accuracy,

$$e_d = \sqrt{\sum_{i=1}^{N_n} (\mathbf{u}_i^{exact} - \mathbf{u}_i^{numerical})^2} \Big/ \sqrt{\sum_{i=1}^{N_n} (\mathbf{u}_i^{exact})^2}. \tag{37}$$

where $\mathbf{u}_i^{exact}$ is analytical displacement, $\mathbf{u}_i^{numerical}$ is the displacement obtained by given numerical methods.

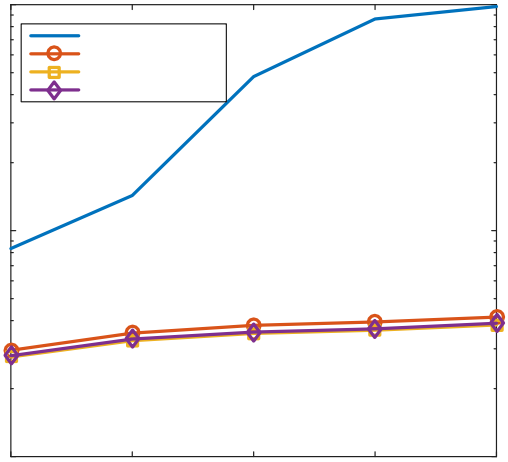


Figure 9 Volumetric locking test for FS-FEM-T4, FEM/ANP-T4, FS-FEM/ANP-T4 and FS/NS-FEM-T4.

Although the errors all tested methods are increasing when Poisson's ratio increasing, they are still under control and less than 5% for all chosen Poisson's ratios except for FS-FEM-T4. With ANP or NS-FEM, FS-FEM suffers a high volumetric locking with increasing Poisson's ration. Another observation is that both FS-FEM/ANP and FS/NS-FEM has higher accuracies than FEM-ANP. It may be caused by the higher accuracy of FS-FEM for the deviatoric deformation. In fact, there lacks of such volumetric locking endurance test for ANP in previous literatures [7], [18].

Table 1. Volumetric locking test: the radial displacement L2-norm ed of different methods versus several Poisson's ratios.

Poisson's ratio	FS-FEM/ANP	FEM/ANP	FS/NS-FEM	FS-FEM
0.4	0.0276	0.0296	0.0280	0.08337
0.49	0.0326	0.0352	0.0331	0.14307
0.499	0.0351	0.0381	0.0356	0.48062
0.4999	0.0363	0.0394	0.0368	0.86474
0.49999	0.0386	0.0414	0.0389	0.98159

The convergences of displacement and strain energy of FS-FEM/ANP are also studied and compared with convergences of FEM/ANP and FS/NS-FEM. Here, the relations between number of nodes and radial displacement L2-norm error of tested methods are plotted in Figure 10(a). In Figure 10(a), all three methods can converge to analytical solution. Among them, FS-FEM/ANP and FS/NS-FEM get the almost identical convergence curves. This means that ANP has almost same performance to NS-FEM when selectively used for volumetric deformation. In addition, FS-FEM/ANP can always get smaller displacement error than FEM/ANP on all meshes. This comparison proves the higher displacement accuracy of FS-FEM than FEM again. In Figure 10(b), strain energy convergence curves of three methods show same features of previous displacement convergence curves.

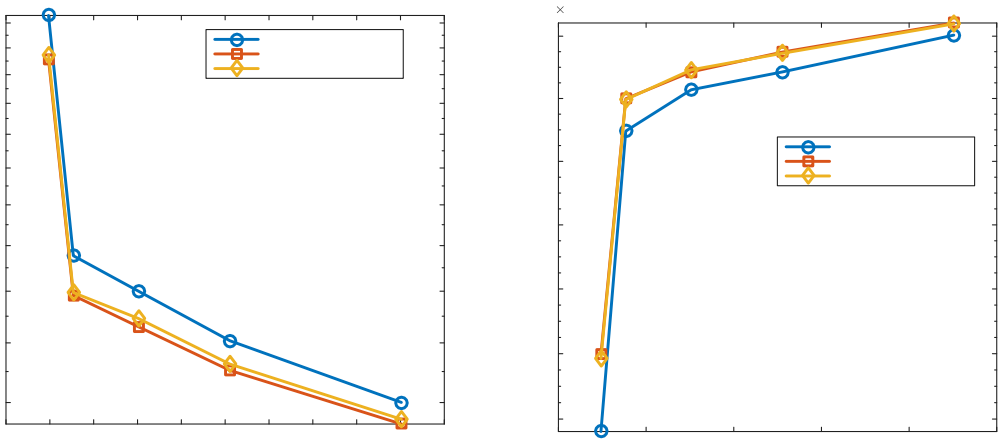


Figure 10 The radial displacement and strain energy convergences of FS-FEM/ANP, FEM/ANP and FS/NS-FEM.

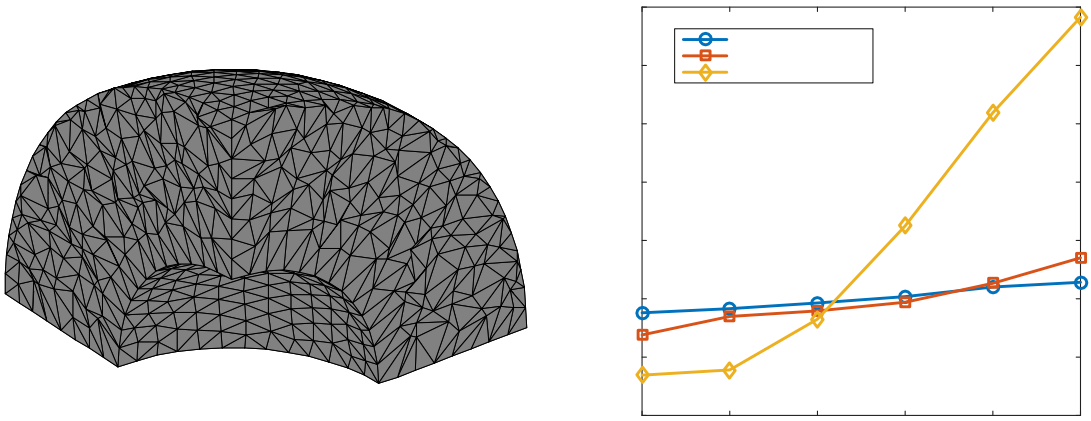


Figure 11 (a) The distorted mesh with distortion coefficient 0.5, (b) Radial displacement errors versus the distortion coefficient.

Another extraordinary capability of S-FEM family is the remarkable mesh distortion robustness. Previous studies have shown tiny accuracy deterioration even when some elements are collapsed [12], [19]. For the first time of S-FEM family embracing ANP, the evaluation of FS-FEM/ANP mesh distortion robustness is necessary. Like previous works, the artificial distortion of mesh is conducted by updating node coordinates of the non-distorted mesh with following equation,

$$\begin{cases} x' = x + h \cdot r_c \cdot \alpha \\ y' = y + h \cdot r_c \cdot \alpha \\ z' = z + h \cdot r_c \cdot \alpha \end{cases} \tag{38}$$

where α is the distortion coefficient from 0 to 1, h is the characteristic length of initial element. r_c is a random number between -1 to 1.

As a further development upon FEM/ANP, we also evaluate the mesh distortion robustness of FEM-ANP with T4 element which is also never evaluated before. After cure the volumetric locking of FEM-T4 with ANP, we can expect the similar mesh distortion robustness of FEM/ANP-T4 to FS-FEM/ANP-T4 for nearly-incompressible solids. This expectation is based on the fact that FEM-T4 is just a special case of Cell-Based S-FEM (CS-FEM) for T4 element [5]. In Figure 11 (a), one mesh of 3D Lame problem with distortion coefficient 0.5 is presented. We can see several elements are severe distorted. Then, to be a more comprehensive comparison, the mesh distortion robustness of FEM-T10 with Selective Reduced Integration (FEM/SRI-T10) is also evaluated. The relation between distortion coefficient and radial displacement L2-norm error of all evaluated methods is plotted in Figure 11(b). When mesh quality is good, the second-order FEM/SRI-T10 has smallest displacement error, then FS-FEM/ANP–T4 and FEM/ANP-T4. However, the increasing distortion coefficients aggravate the error of FEM/SRI-T10 much faster than FS-FEM/ANP-T4 and FEM/ANP-T4. Therefore, we conclude that ANP has no influence on extraordinary mesh distortion robustness of S-FEM. By the way, due to random element distortion, the mesh with distortion coefficient 0.5 for FS-FEM/ANP-T4 may locally more severe than the mesh with distortion coefficient 0.5 for FEM/ANP-T4. Therefore, the error of FS-FEM/ANP-T4 may be slightly larger than FEM/ANP-T4. In summary, researchers should pay meticulous attention to mesh quality when using FEM/SRI-T10.

Conclusions

In this paper, the FS-FEM/ANP-T4 has been proposed to solve 3D explicit dynamic and quasi-static problems of nearly-incompressible solids. In FS-FEM/ANP-T4, the FS-FEM is used for deviatoric deformation. And the ANP responds to the volumetric deformation. Several features of FS-FEM/ANP-T4 have been confirmed by selected numerical examples.

The ANP can provide the “under integration” effects to FS-FEM which is ideal for volumetric part deformation of nearly-incompressible solids.

Although FS-FEM/ANP-T4 still encounters pressure oscillation issue, it shows more mild pressure oscillation than FEM/ANP-T4 and FS/NS-FEM-T4.

FS-FEM/ANP-T4 has higher accuracy and convergence than FEM/ANP-T4. The “overly-stiff” behavior of linear T4 element is relieved by FS-FEM. FS-FEM can improve the performance for the deviatoric part deformation of nearly-incompressible solids.

FS-FEM/ANP-T4 is still very robust for mesh distortion as FS-FEM-T4. Because the ANP is based on FEM-T4 which is also special case of CS-FEM-T4.

Since the ANP is not too “soft”, FS-FEM/ANP-T4 also works well for large deformation of nearly-incompressible solids.

FS-FEM/ANP-T4 use much less computational time than FEM/RI-T10 with same mesh in explicit dynamic simulation.

Acknowledgement

The first author is partially supported by the scholarship No.201406130010 of China Scholarship Council and the National Natural Science Foundation of China under grant No. 11372106 and No. 11202075. This theoretical basic research by the senior author is partially sponsored by NSF under the Award No. DMS-1214188 and the National Natural Science Foundation of China (Grant No. 11472184). The authors would like to thank the valuable supports leading to this paper.

References

[1] G. R. LIU, "ON G SPACE THEORY," *Int. J. Comput. Methods*, vol. 06, no. 02, pp. 257–289, 2009.
[2] G. R. Liu, "A G space theory and a weakened weak (W 2) form for a unified formulation of compatible and incompatible methods: Part I theory," *Int. J. Numer. Methods Eng.*, vol. 81, no. 9, pp. 1093–1126, 2010.
[3] G. R. Liu and G. Y. Zhang, *The Smoothed Point Interpolation Methods – G Space Theory and Weakened Weak Forms*. WorldScientific, 2013.
[4] J. Chen, C. Wu, S. Yoon, and Y. You, "A stabilized conforming nodal integration for Galerkin mesh-free methods," *Int. J. Numer. Methods Eng.*, vol. 50, no. February 2000, pp. 435–466, 2001.
[5] G. R. Liu and N. T. Trung, *Smoothed Finite Element Methods*. Boca Raton: CRC Press, 2010.
[6] G. R. Liu, K. Y. Dai, and T. T. Nguyen, "A Smoothed Finite Element Method for Mechanics Problems," *Comput. Mech.*, vol. 39, no. 6, pp. 859–877, May 2006.
[7] J. Bonet and A. J. Burton, "A simple average nodal pressure tetrahedral element for incompressible and nearly incompressible dynamic explicit applications," *Commun. Numer. Methods Eng.*, vol. 14, no. 5, pp. 437–449, May 1998.
[8] J. Bonet, H. Marriott, and O. Hassan, "An averaged nodal deformation gradient linear tetrahedral element for large strain explicit dynamic applications," *Commun. Numer. Methods Eng.*, vol. 17, no. 8, pp. 551–561, Jul. 2001.
[9] G. R. Joldes, A. Wittek, and K. Miller, "Non-locking tetrahedral finite element for surgical simulation," *Commun. Numer. Methods Eng.*, vol. 25, no. September 2008, pp. 827–836, 2009.
[10] C. R. Dohrmann, M. W. Heinstein, J. Jung, S. W. Key, and W. R. Witkowski, "Node-based uniform strain elements for three-node triangular and four-node tetrahedral meshes," *Int. J. Numer. Methods Eng.*, vol. 47, no. June 1999, pp. 1549–1568, 2000.
[11] G. Holzapfel, "Nonlinear Solid Mechanics: A Continuum Approach for Engineering," 2000.
[12] C. Jiang, Z.-Q. Zhang, G. R. Liu, X. Han, and W. Zeng, "An edge-based/node-based selective smoothed finite element method using tetrahedrons for cardiovascular tissues," *Eng. Anal. Bound. Elem.*, vol. 59, pp. 62–77, 2015.
[13] C. Jiang, Z.-Q. Zhang, X. Han, and G. R. Liu, "Selective smoothed finite element methods for extremely large deformation of anisotropic incompressible bio-tissues," *Int. J. Numer. Methods Eng.*, pp. 1–39.
[14] T. Belytschko, W. . Liu, and B. Moran, *Nonlinear Finite Elements for Continua and Structures*. Hoboken,NJ: John Wiley & Sons, 2013.
[15] R. G. Sauvé and D. R. Metzger, "Advances in Dynamic Relaxation Techniques for Nonlinear Finite Element Analysis," *J. Press. Vessel Technol.*, vol. 117, no. 2, p. 170, 1995.
[16] D. R. Oakley and N. F. Knight, "Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures Part II. Single-processor implementation," *Comput. Methods Appl. Mech. Eng.*, vol. 126, no. 1–2, pp. 91–109, Sep. 1995.
[17] R. Oakley and F. Knight, "Adaptive Dynamic Relaxation algorithm for non-linear hyperelastic structures Part I . Formulation," *Comput. Methods Appl. Mech. Eng.*, vol. 25, no. 95, pp. 67–89, 1993.
[18] J. Bonet, H. Marriott, and O. Hassan, "An averaged nodal deformation gradient linear tetrahedral element for large strain explicit dynamic applications," *Commun. Numer. Methods Eng.*, vol. 17, no. 8, pp. 551–561, Jul. 2001.
[19] C. Jiang, G.-R. Liu, X. Han, Z.-Q. Zhang, and W. Zeng, "A Smoothed finite element method for analysis of anisotropic large deformation of passive rabbit ventricles in diastole," *Int. j. numer. method. biomed. eng.*, vol. 31, 2015.

A cell-based smoothed finite element method for free vibration analysis of a rotating plate

***C.F. Du[1], †D.G. Zhang[1], G.R. Liu[2]**

[1] School of Sciences, Nanjing University of Science and Technology, Nanjing 210094, China
[2] Department of Aerospace Engineering and Engineering Mechanics
University of Cincinnati, Cincinnati, OH 45221, USA

*Presenting author: duchaofan1987@163.com
†Corresponding author:zhangdg419@mail.njust.edu.cn

Abstract

A cell-based smoothed finite element method (CS-FEM) is formulated for non-linear free vibration analysis of a plate attached to a rigid rotating hub. The first–order shear deformation theory which is known as Mindlin plate theory is used to model the plate. In the process of formulating the system stiffness matrix, the discrete shear gap (DSG) method is used to construct the strains to overcome the shear locking issue. The effectiveness of the CS-FEM is first demonstrated in some static cases and then extended for free vibration analysis of a rotating plate considering the non-linear effects arising from the coupling of vibration of the flexible structure with the undergoing large rotational motions. The nonlinear coupling dynamic equations of the system are derived via employing Lagrange's equations of the second kind. The effect of different parameters including thickness ratio, aspect ratio, hub radius ratio and rotation speed on dimensionless natural frequencies are investigated. The dimensionless natural frequencies of CS-FEM are compared with those other existing method including the finite element method (FEM) and the assumed modes method (AMM). It is found that the CS-FEM based on Mindlin plate theory provides more accurate and "softer" solution compared with those of other methods even if using coarse meshes. In addition, the frequency loci veering phenomena associated with the mode shape interaction are examined in detail.

Keywords: cell-based smoothed finite element method, rotating Mindlin plate, discrete shear gap method, shear locking, natural frequencies, frequency veering.

1 Introduction

A lot of engineering structures consist of a flexible appendage attached to a rigid body, which are called rigid-flexible coupled structures, such as space robotic manipulators, satellite antenna, helicopter rotors, solar energy panels and aircraft engine blades and so on. Such structures can often be simplified to a rotating hub-beam or rotating-plate for dynamic analysis. Compared to the modal characteristics of non-rotating structures, those of rotating structures behave significantly, due to the coupling of the non-linear effects arising from the coupling of vibration of the flexible structure with the undergoing large rotational motions. Therefore, it is essential to conduct accurate analysis of natural frequencies and mode shapes

of these rotating structures in the design stages, considering the nonlinear effects.

The rotating structures are often idealized as rotating beams in early stage researching. The earliest works on the natural frequency of rotating beams was performed by Southwell and Gough[1] in 1921. The famous Southwell equation was presented in their work. Later, a lot of research achievement has been obtained about rotating beams [2-7]. However, there are many structures with low aspect ratios that behave like plates rather than beams. It's obvious that beam models can't obtain accurate modal characteristics and rotating plate models are more appropriate for those plate-like structures. Dokainish and Rawtani [8] used a finite element technique to determine the natural frequencies and the mode shapes of a cantilever plate mounted on the periphery of a rotating disc. The effect of the aspect ratio, the speed of rotation, the disc radius and the setting angle for the natural frequencies were discussed. Ramamurti and Kielb [9] used FEM to analyze the natural frequencies of twisted rotating plates. Yoo [10,11] used AMM to investigate the modal characteristics of a rotating cantilever plate and dimensionless parameters were identified through dimensional analysis. Hashemi [12] developed a finite element formulation for vibration analysis of rotating thick plates. The effect of different dimensionless parameters on dimensionless natural frequencies were investigated and discussed. In these references, there are two things in common: the discrete methods are FEM or AMM; the modeling theory is most based on classic plate theory (Kirchhoff plate theory), which does not work for thicker plates. Thus, more effective discrete methods based on higher order theory are necessary.

Recently, a new discrete method named smoothed FEM (S-FEM) has been proposed [13]. This method combines with the conventional FEM and the strain smoothing technique used in meshfree methods. It possesses the features of both FEM and meshfree methods. According to the different smoothing domain creation, there are a series of S-FEM models: the cell-based S-FEM (CS-FEM) [14-16], the node-based S-FEM (NS-FEM) [17-19], the edge-based S-FEM (ES-FEM) [20-22] and the face-based S-FEM (FS-FEM) [23,24], each of which has especial properties.

This paper extends the CS-FEM for non-linear free vibration analysis of a rotating plate based on Mindlin plate theory. In the present CS-FEM, we use triangular elements that can be automatically generated. The discrete shear gap (DSG) method is used to construct the strains to overcome the shear locking issue. The effectiveness of the CS-FEM is first demonstrated in some static cases and then extended for free vibration analysis of a rotating plate considering the non-linear effects arising from the coupling of vibration of the flexible structure with the undergoing large rotational motions. The effect of different parameters including thickness ratio, aspect ratio, hub radius ratio and rotation speed on dimensionless natural frequencies are investigated. The dimensionless natural frequencies of CS-FEM are compared with those other existing method including the finite element method (FEM) and the assumed modes method (AMM). It is found that the CS-FEM based on Mindlin plate theory provides "softer" solution compared with those of other methods. In addition, the frequency loci veering phenomena associated with the mode shape interaction are examined in detail.

2 Formulation of FEM for the Mindlin plate

Consider a plate under bending deformation as shown in Fig.1. The middle (neutral) surface of plate is chosen as the reference plane and the problem domain. The plate is discretized with a set of three nodes triangular element as shown in Fig.2.

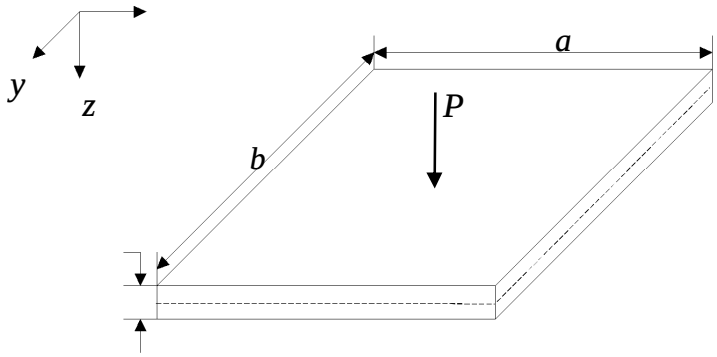


Fig. 1 Mindlin plate with uniform thickness

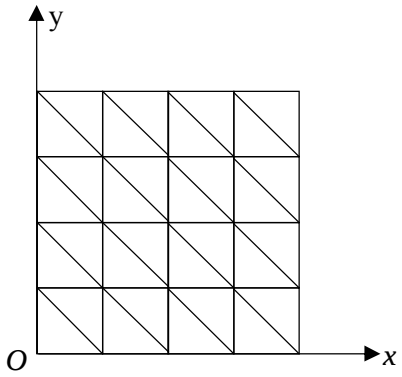


Fig. 2 Discretization of the plate using triangular elements

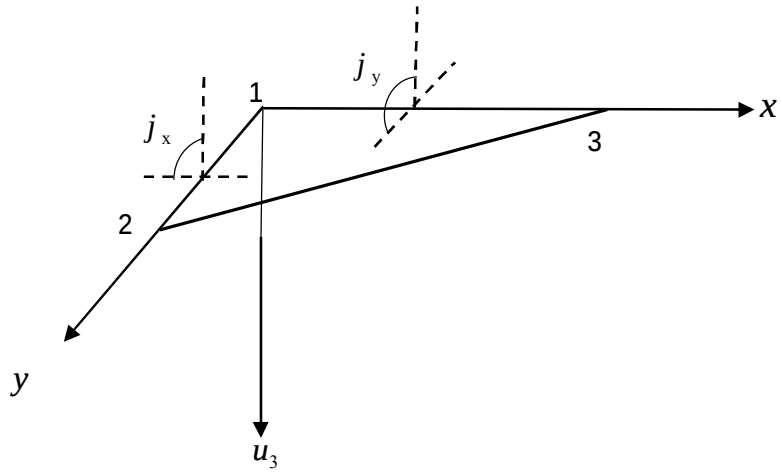


Fig. 3 A typical three nodes triangular element

In one triangular element as shown in Fig.3, the displacement of an arbitrary point can be expressed as

$$\boldsymbol{u}_e = \begin{bmatrix} u_3 & j_x & j_y \end{bmatrix}^{\mathrm{T}} = \sum_{i=1}^{3} \begin{bmatrix} f_{3i} & 0 & 0 \\ 0 & f_{3i} & 0 \\ 0 & 0 & f_{3i} \end{bmatrix} \begin{bmatrix} u_{3i} \\ j_{xi} \\ j_{yi} \end{bmatrix} = f\boldsymbol{d}_e \tag{1}$$

where u_3 is transverse deflection, and j_x, j_y are the rotations of the middle plane around y-axis and x-axis, respectively. f_{3i} (i=1,2,3) are the shape functions corresponding to three nodes of the triangular element and their expressions are

$$f_{3i} = a_i + b_i x + c_i y \tag{2}$$

$$\begin{cases} a_i = \dfrac{1}{2A_e}(x_j y_k - x_k y_j) \\ b_i = \dfrac{1}{2A_e}(y_j - y_k) \\ c_i = \dfrac{1}{2A_e}(x_k - x_j) \end{cases} \tag{3}$$

where A_e is the area of the triangular element. x_j and y_j (j=1,2,3) are the coordinate values at the jth node. The subscript i, j and k vary from 1 to 3 and are determined by cyclic permutation in the order of i, j and k. For example, if i=1, then j=2, k=3; or if i=2, then j=3, k=1.

The nodal displacement vector associated to node i can be expressed as $\boldsymbol{d}_i = \begin{bmatrix} u_{3i} & j_{xi} & j_{yi} \end{bmatrix}^{\mathrm{T}}$. Then the bending and shear strains in the matrix forms are

$$z = \sum_{i=1}^{3} \boldsymbol{B}_i \boldsymbol{d}_i = \boldsymbol{B}\boldsymbol{d}_e \tag{4}$$

$$g = \sum_{i=1}^{3} \boldsymbol{S}_i \boldsymbol{d}_i = \boldsymbol{S}\boldsymbol{d}_e \tag{5}$$

Where

$$\boldsymbol{B}_i = \begin{bmatrix} 0 & f_{3i,x} & 0 \\ 0 & 0 & f_{3i,y} \\ 0 & f_{3i,y} & f_{3i,x} \end{bmatrix} \tag{6}$$

$$\boldsymbol{S}_i = \begin{bmatrix} f_{3i,x} & f_{3i} & 0 \\ f_{3i,y} & 0 & f_{3i} \end{bmatrix} \tag{7}$$

$$\boldsymbol{d}_e = \begin{bmatrix} \boldsymbol{d}_1 & \boldsymbol{d}_2 & \boldsymbol{d}_3 \end{bmatrix}^{\mathrm{T}} \tag{8}$$

Substituting Eq.(3) into Eq.(6) and Eq.(7), the bending strain matrix can be written as

$$\boldsymbol{B} = \frac{1}{2A_e}\begin{bmatrix} 0 & b-c & 0 & 0 & c & 0 & 0 & -b & 0 \\ 0 & 0 & d-a & 0 & 0 & -d & 0 & 0 & a \\ 0 & d-a & b-c & 0 & -d & c & 0 & a & -b \end{bmatrix} = \frac{1}{2A_e}\begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 \end{bmatrix} \tag{9}$$

Where

$$\begin{aligned} a &= x_2 - x_1 \quad b = y_2 - y_1 \\ c &= y_3 - y_1 \quad d = x_3 - x_1 \end{aligned} \tag{10}$$

As reported in many literatures, the shear locking issue often occurs when using thick plate theory to analyze thin plates. To avoid this problem, many numerical techniques have been well developed [25-30]. Recently, the discrete shear gap (DSG) method was proposed by Bletzinger *et al*. [31]. This method can be applied to both triangular and rectangular elements of different polynomial order. According to DSG method, the shear strain matrix can be written as

$$\boldsymbol{S} = \frac{1}{2A_e}\begin{bmatrix} b-c & A_e & 0 & c & \frac{ac}{2} & \frac{bc}{2} & -b & -\frac{bd}{2} & -\frac{bc}{2} \\ d-a & 0 & A_e & -d & -\frac{ad}{2} & -\frac{bd}{2} & a & \frac{ad}{2} & \frac{ac}{2} \end{bmatrix} = \frac{1}{2A_e}\begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{S}_2 & \boldsymbol{S}_3 \end{bmatrix} \tag{11}$$

Then the discretized system equation of Mindlin plate with FEM for static analysis can be expressed as

$$K = \int_\Omega \boldsymbol{B}^{\mathrm{T}} \boldsymbol{D}_{\mathrm{b}} \boldsymbol{B} \mathrm{d}\Omega + \int_\Omega \boldsymbol{S}^{\mathrm{T}} \boldsymbol{D}_{\mathrm{s}} \boldsymbol{S} \mathrm{d}\Omega \tag{12}$$

Where the matrices $\boldsymbol{D}_{\mathrm{b}}$ and $\boldsymbol{D}_{\mathrm{s}}$ are related to the bending deformation and shear deformation, respectively. They are given by

$$\boldsymbol{D}_{\mathrm{b}} = \frac{Eh^3}{12(1-m^2)}\begin{bmatrix} 1 & m & 0 \\ m & 1 & 0 \\ 0 & 0 & \frac{(1-m)}{2} \end{bmatrix} \tag{13}$$

$$\boldsymbol{D}_{\mathrm{s}} = \frac{kEh}{2(1+m)}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{14}$$

Where E is the Young's modulus, h is the thickness of the Mindlin plate, m is the Poisson's ratio and k is the shear correction factor which is given by $k = \frac{5}{6}$. In order to improve the accuracy of the solutions and stabilize shear force oscillations, Bischoff [32] suggested that a

stabilization term should be added to the element. Such a modification can be simply achieved by replacing $\boldsymbol{D}_s$ in Eq.(14) with the following equation

$$\boldsymbol{D}_s = \frac{k E h^3}{2(1+m)(h^2 + a\, he^2)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{15}$$

Where h_e is the longest length of the edges of the element and a is a ppositive constant which is called stabilized parameter [33]. In this paper, a is fixed at 0.1.

For the free vibration analysis, the discretized system equation of Mindlin plate can be expressed as

$$(K - w^2 \boldsymbol{M})\boldsymbol{d} = \boldsymbol{0} \tag{16}$$

Where w is the natural frequency and $\boldsymbol{d}$ is the global displacement vector. $\boldsymbol{M}$ is the global mass matrix and defined by

$$\boldsymbol{M} = \int_\Omega f^{\mathrm{T}} \boldsymbol{m} f \, \mathrm{d}\Omega = \sum_{i=1}^{N_e} \int_{\Omega_i^e} f^{\mathrm{T}} \boldsymbol{m} f \, \mathrm{d}\Omega_i^e \tag{17}$$

In which $\boldsymbol{m}$ is a constant matrix about the mass density ρ and thickness of the plate, which is given by

$$\boldsymbol{m} = r \begin{bmatrix} h & 0 & 0 \\ 0 & \dfrac{h^3}{12} & 0 \\ 0 & 0 & \dfrac{h^3}{12} \end{bmatrix} \tag{18}$$

3 Formulation of CS-FEM for the Mindlin plate

In CS-FEM, each triangular element domain is further devided into three triangular smoothing domains by simply connecting three field nodes of the element to the central point of the element, as shown in Fig.4. These smoothing domains are not overlapping and there are no gaps between them.

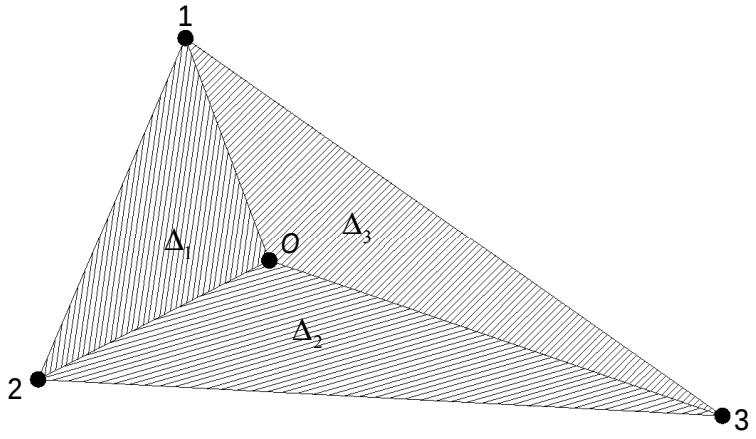


Fig. 4 Three triangular smoothing domains

Define the triangular element domain as Ω_e and three triangular smoothing domains as Δ_1, Δ_2 and Δ_3. Then we have $\Omega_e = \mathbf{U}_{i=1}^{3} \Delta_i$ and $\Delta_i \ \mathbf{I} \ \Delta_j = \emptyset \ (i \neq j)$. The coordinates of the central point $\boldsymbol{x}_o = [x_o \quad y_o]^{\mathrm{T}}$ are calculated by

$$\begin{cases} x_o = \dfrac{1}{3}(x_1 + x_2 + x_3) \\ y_o = \dfrac{1}{3}(y_1 + y_2 + y_3) \end{cases} \tag{19}$$

Where $\boldsymbol{x}_i = [x_i \quad y_i]^{\mathrm{T}}$ with i=1,2,3 are the three field nodes of the element. The displacement vector $\boldsymbol{d}_o$ at the central point O is assumed to be the simple average of three displacement vectors at three field nodes of the element

$$\boldsymbol{d}_o = \frac{1}{3}(\boldsymbol{d}_1 + \boldsymbol{d}_2 + \boldsymbol{d}_3) \tag{20}$$

On the first subtriangle Δ_1(1-2-O), the displacement of an arbitrary point in the element can be expressed as

$$\boldsymbol{u}_e^{\Delta_1} = \begin{bmatrix} f_{31} & 0 & 0 \\ 0 & f_{31} & 0 \\ 0 & 0 & f_{31} \end{bmatrix} \boldsymbol{d}_1 + \begin{bmatrix} f_{32} & 0 & 0 \\ 0 & f_{32} & 0 \\ 0 & 0 & f_{32} \end{bmatrix} \boldsymbol{d}_2 + \begin{bmatrix} f_{33} & 0 & 0 \\ 0 & f_{33} & 0 \\ 0 & 0 & f_{33} \end{bmatrix} \boldsymbol{d}_o \tag{21}$$

Substituting Eq.(20) into Eq.(21), then $\boldsymbol{u}_e^{\Delta_1}$ can be rewritten as

$$\boldsymbol{u}_e^{\Delta_1} = \left(\begin{bmatrix} f_{31} & 0 & 0 \\ 0 & f_{31} & 0 \\ 0 & 0 & f_{31} \end{bmatrix} + \frac{1}{3} \begin{bmatrix} f_{33} & 0 & 0 \\ 0 & f_{33} & 0 \\ 0 & 0 & f_{33} \end{bmatrix} \right) \boldsymbol{d}_1 + \left(\begin{bmatrix} f_{32} & 0 & 0 \\ 0 & f_{32} & 0 \\ 0 & 0 & f_{32} \end{bmatrix} + \frac{1}{3} \begin{bmatrix} f_{33} & 0 & 0 \\ 0 & f_{33} & 0 \\ 0 & 0 & f_{33} \end{bmatrix} \right) \boldsymbol{d}_2 + \frac{1}{3} \begin{bmatrix} f_{33} & 0 & 0 \\ 0 & f_{33} & 0 \\ 0 & 0 & f_{33} \end{bmatrix} \boldsymbol{d}_3 \tag{22}$$

Then the strain matrices in the subtriangle Δ_1 can be obtained

$$z^{\Delta_1} = \left[\boldsymbol{B}_1^{\Delta_1} + \frac{1}{3}\boldsymbol{B}_3^{\Delta_1} \quad \boldsymbol{B}_2^{\Delta_1} + \frac{1}{3}\boldsymbol{B}_3^{\Delta_1} \quad \frac{1}{3}\boldsymbol{B}_3^{\Delta_1} \right] \boldsymbol{d}_e = \boldsymbol{B}^{\Delta_1}\boldsymbol{d}_e \tag{23}$$

$$g^{\Delta_1} = \left[\boldsymbol{S}_1^{\Delta_1} + \frac{1}{3}\boldsymbol{S}_3^{\Delta_1} \quad \boldsymbol{S}_2^{\Delta_1} + \frac{1}{3}\boldsymbol{S}_3^{\Delta_1} \quad \frac{1}{3}\boldsymbol{S}_3^{\Delta_1} \right] \boldsymbol{d}_e = \boldsymbol{S}^{\Delta_1}\boldsymbol{d}_e \tag{24}$$

Where $\boldsymbol{B}^{\Delta_1}$ and $\boldsymbol{S}^{\Delta_1}$ are calculated similarly as Eq.(9) and Eq.(11). The only difference is that the corresponding functions are computed in the domain of subtriangle Δ_1, which means the three field nodes are $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_o$, respectively. Similarly, the strain matrices in the subtriangles Δ_2 (2-3-*O*) and Δ_3 (3-1-*O*) can be obtained by cyclic permutation like described in section 2. Their expressions are as follows

$$z^{\Delta_2} = \left[\frac{1}{3}\boldsymbol{B}_1^{\Delta_2} \quad \boldsymbol{B}_2^{\Delta_2} + \frac{1}{3}\boldsymbol{B}_1^{\Delta_2} \quad \boldsymbol{B}_3^{\Delta_2} + \frac{1}{3}\boldsymbol{B}_1^{\Delta_2} \right] \boldsymbol{d}_e = \boldsymbol{B}^{\Delta_2}\boldsymbol{d}_e \tag{25}$$

$$g^{\Delta_2} = \left[\frac{1}{3}\boldsymbol{S}_1^{\Delta_2} \quad \boldsymbol{S}_2^{\Delta_2} + \frac{1}{3}\boldsymbol{S}_1^{\Delta_2} \quad \boldsymbol{S}_3^{\Delta_2} + \frac{1}{3}\boldsymbol{S}_1^{\Delta_2} \right] \boldsymbol{d}_e = \boldsymbol{S}^{\Delta_2}\boldsymbol{d}_e \tag{26}$$

$$z^{\Delta_3} = \left[\boldsymbol{B}_1^{\Delta_3} + \frac{1}{3}\boldsymbol{B}_2^{\Delta_3} \quad \frac{1}{3}\boldsymbol{B}_2^{\Delta_3} \quad \boldsymbol{B}_3^{\Delta_3} + \frac{1}{3}\boldsymbol{B}_2^{\Delta_3} \right] \boldsymbol{d}_e = \boldsymbol{B}^{\Delta_3}\boldsymbol{d}_e \tag{27}$$

$$g^{\Delta_3} = \left[\boldsymbol{S}_1^{\Delta_3} + \frac{1}{3}\boldsymbol{S}_2^{\Delta_3} \quad \frac{1}{3}\boldsymbol{S}_2^{\Delta_3} \quad \boldsymbol{S}_3^{\Delta_3} + \frac{1}{3}\boldsymbol{S}_2^{\Delta_3} \right] \boldsymbol{d}_e = \boldsymbol{S}^{\Delta_3}\boldsymbol{d}_e \tag{28}$$

By applying the strain smoothing technique, the smoothed bending strain and smoothed shear strain in each triangular element can be obtained as

$$\begin{aligned} z^{\%} &= \frac{1}{A_e}\int_{\Omega_e} z\,(x,y)\mathrm{d}\Omega_e \\ &= \frac{1}{A_e}\left(\int_{\Delta_1} z^{\Delta_1}\mathrm{d}\Delta_1 + \int_{\Delta_2} z^{\Delta_2}\mathrm{d}\Delta_2 + \int_{\Delta_3} z^{\Delta_3}\mathrm{d}\Delta_3 \right) \end{aligned} \tag{29}$$

$$\begin{aligned} \tilde{g} &= \frac{1}{A_e}\int_{\Omega_e} g(x,y)\mathrm{d}\Omega_e \\ &= \frac{1}{A_e}\left(\int_{\Delta_1} g^{\Delta_1}\mathrm{d}\Delta_1 + \int_{\Delta_2} g^{\Delta_2}\mathrm{d}\Delta_2 + \int_{\Delta_3} g^{\Delta_3}\mathrm{d}\Delta_3\right) \end{aligned} \tag{30}$$

Because the strain in the subtriangles are constant, Eq.(29) and Eq.(30) can be rewritten as

$$\tilde{z} = \frac{A_{\Delta_1} z^{\Delta_1} + A_{\Delta_2} z^{\Delta_2} + A_{\Delta_3} z^{\Delta_3}}{A_e} = \tilde{\boldsymbol{B}}\boldsymbol{d}_e \tag{31}$$

$$\tilde{g} = \frac{A_{\Delta_1} g^{\Delta_1} + A_{\Delta_2} g^{\Delta_2} + A_{\Delta_3} g^{\Delta_3}}{A_e} = \tilde{\boldsymbol{S}}\boldsymbol{d}_e \tag{32}$$

Where the smoothed strain matrices are as follows

$$\tilde{\boldsymbol{B}} = \frac{A_{\Delta_1}\boldsymbol{B}^{\Delta_1} + A_{\Delta_2}\boldsymbol{B}^{\Delta_2} + A_{\Delta_3}\boldsymbol{B}^{\Delta_3}}{A_e} \tag{33}$$

$$\tilde{\boldsymbol{S}} = \frac{A_{\Delta_1}\boldsymbol{S}^{\Delta_1} + A_{\Delta_2}\boldsymbol{S}^{\Delta_2} + A_{\Delta_3}\boldsymbol{S}^{\Delta_3}}{A_e} \tag{34}$$

In which A_e is the area of triangular element. A_{Δ_1} , A_{Δ_2} and A_{Δ_3} are the areas of three subtriangles, respectively. Substituting Eqs.(33) and (34) into Eq.(12), the smoothed element stiffness matrix can be given by

$$\tilde{\boldsymbol{K}}_e = \int_{\Omega_e} \tilde{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{D}_b\tilde{\boldsymbol{B}}\mathrm{d}\Omega_e + \int_{\Omega_e} \tilde{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{D}_s\tilde{\boldsymbol{S}}\mathrm{d}\Omega_e \tag{35}$$

Where the matrices $\boldsymbol{D}_b$ and $\boldsymbol{D}_s$ are the same as Eqs.(13) and (15). Then the global stiffness matrix of the CS-FEM can be assembled by

$$\hat{\boldsymbol{K}} = \sum_{e=1}^{N_e} \tilde{\boldsymbol{K}}_e \tag{36}$$

It should be mentioned that the central point in each element is only used to form the smoothing domain. Finally, the displacement vector of this point will be replaced by those of three field nodes of the element as shown in Eq.(22). Hence, there are no extra DOFs, which means the DOFs of CS-FEM are the same as FEM if using the same mesh.

4 Nonlinear dynamic equations of rotating plates based on Mindlin plate theory

4.1 Dynamic equations of a Mindlin plate undergoing overall motion

In Mindlin plate theory, the plate doesn't demand the cross section be perpendicular to the neutral plane after deformation. The transverse shear strain which is neglected in classical thin plate theory is taken into account. In this section, the nonlinear dynamic equations of a rotating rectangular Mindlin plate undergoing overall motion in three-dimensional space will be presented in detail.

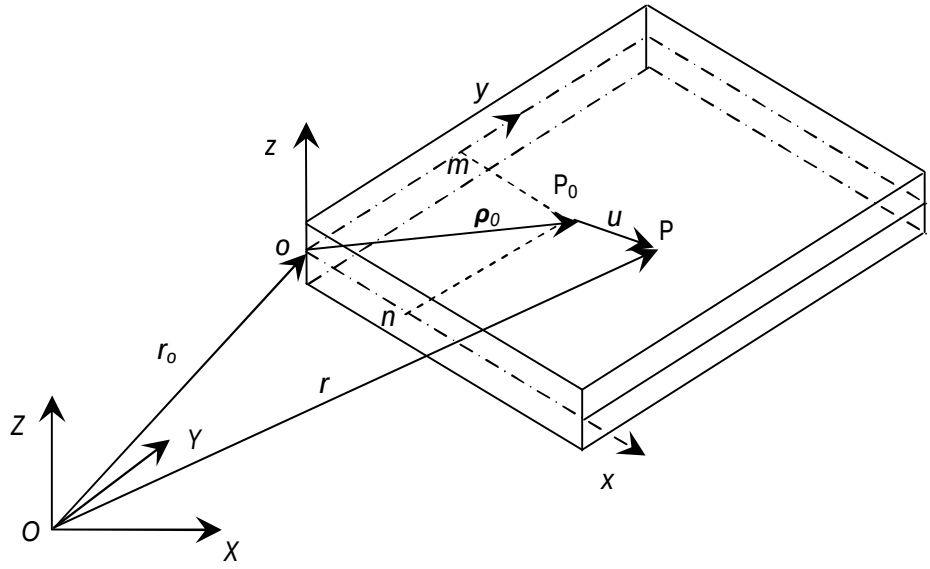


Fig. 5 The configuration of a rectangular Mindlin plate

Consider a flexible plate undergoing overall motion as shown in Fig.5. The inertial coordinate system and the local coordinate system which is fixed to the neutral surface of the plate are denoted by *O-XYZ* and *o-xyz*, respectively. The physical parameters of the plate are as follows: length *a,* width *b*, thickness *h*, Young's modulus *E*, mass density ρ and Poisson's ratio μ. P_0 is an arbitrary point on the undeformed neutral surface of the plate in the local coordinate system. After deformation, it moves to *P* and the displacement vector is denoted by $\boldsymbol{u} = (u_1, u_2, u_3)^{\mathrm{T}}$ where u_1, u_2 and u_3 are the displacement components along the *x*, *y* and *z* axis in the local coordinate system, respectively. The displacement components u_1 and u_2 can be expressed as

$$\begin{cases} u_1 = w_1 - \dfrac{1}{2}\int_0^x (\dfrac{\partial u_3}{\partial x})^2 \mathrm{d}x + z j_x \\ u_2 = w_2 - \dfrac{1}{2}\int_0^y (\dfrac{\partial u_3}{\partial y})^2 \mathrm{d}y + z j_y \end{cases} \tag{37}$$

Where w_1 and w_2 are neutral surface stretch along the *x* and *y* axis, respectively. $-\dfrac{1}{2}\int_0^x (\dfrac{\partial u_3}{\partial x})^2 \mathrm{d}x$ and $-\dfrac{1}{2}\int_0^y (\dfrac{\partial u_3}{\partial y})^2 \mathrm{d}y$ are the coupling terms of the deformation which are caused the transverse deformation. In the traditional approximate model, these two coupling terms are ignored because of the small deformation assumption. j_x and j_y are the rotations of the middle plane around *y*-axis and *x*-axis, respectively. The velocity vector of an arbitrary point P in the inertial coordinate system can be expressed as

$$\boldsymbol{V}_P = \boldsymbol{V}_o + \boldsymbol{\omega}_A \times (\boldsymbol{\rho}_0 + \boldsymbol{u}) + \boldsymbol{V}^{PA} \tag{38}$$

Where $\boldsymbol{V}_o$ and $\boldsymbol{\omega}_A$ are the velocity and angular velocity of the local coordinate system relative

to the inertial coordinate system, respectively. $\boldsymbol{\rho}_0$ and $\boldsymbol{V}^{PA}$ are the position vector of point P_0 and the velocity vector of point P in the local coordinate system, respectively. These vectors are as follows

$$\boldsymbol{V}_o = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + v_3\boldsymbol{e}_3 \qquad \boldsymbol{\omega}_A = w_1\boldsymbol{e}_1 + w_2\boldsymbol{e}_2 + w_3\boldsymbol{e}_3 \tag{39}$$

$$\boldsymbol{\rho}_0 = x\boldsymbol{e}_1 + y\boldsymbol{e}_2 \quad \boldsymbol{u} = u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + u_3\boldsymbol{e}_3 \quad \boldsymbol{V}_{PA} = \dot{u}_1\boldsymbol{e}_1 + \dot{u}_2\boldsymbol{e}_2 + \dot{u}_3\boldsymbol{e}_3 \tag{40}$$

Where $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ are the unit vectors along x, y, and z axis, respectively. Substituting Eqs.(39) and (40) into Eq.(38), the velocity vector of an arbitrary point P in the inertial coordinate system can be obtained as

$$\begin{aligned}\boldsymbol{V}_P = {} & [v_1 + \dot{u}_1 + w_2u_3 - w_3(y+u_2)]\boldsymbol{e}_1 + [v_2 + \dot{u}_2 + w_3(x+u_1) - w_1u_3]\boldsymbol{e}_2 \\ & + [v_3 + \dot{u}_3 + w_1(y+u_2) - w_2(x+u_1)]\boldsymbol{e}_3\end{aligned} \tag{41}$$

Then the kinetic energy of the system is

$$T = \frac{1}{2}\int_V r\,\boldsymbol{V}_P^{\mathrm{T}}\boldsymbol{V}_P \mathrm{d}V = \frac{1}{2}\iint_A r\,h\boldsymbol{V}_P^{\mathrm{T}}\boldsymbol{V}_P \mathrm{d}A \tag{42}$$

According to Mindlin plate theory, the strain vectors can be obtained as

$$e = \begin{cases} e_{xx} = \dfrac{\partial w_1}{\partial x} + z\dfrac{\partial j_x}{\partial x} \\ e_{yy} = \dfrac{\partial w_2}{\partial y} + z\dfrac{\partial j_y}{\partial y} \\ g_{xy} \approx \dfrac{\partial w_1}{\partial y} + \dfrac{\partial w_2}{\partial x} + z\left(\dfrac{\partial j_x}{\partial y} + \dfrac{\partial j_y}{\partial x}\right) \\ g_{xz} = \dfrac{\partial u_3}{\partial x} + j_x \\ g_{yz} = \dfrac{\partial u_3}{\partial y} + j_y \end{cases} \tag{43}$$

Compared to the Kirchhoff plate theory, the strain components g_{xz} and g_{yz} are not equal zero. Then the stress vectors for isotropic materials are obtained as

$$s = \frac{E}{1-m^2}\begin{bmatrix} 1 & m & 0 & 0 & 0 \\ m & 1 & 0 & 0 & 0 \\ 0 & 0 & (1-m)/2 & 0 & 0 \\ 0 & 0 & 0 & k(1-m)/2 & 0 \\ 0 & 0 & 0 & 0 & k(1-m)/2 \end{bmatrix}\begin{bmatrix} e_{xx} \\ e_{yy} \\ g_{xy} \\ g_{xz} \\ g_{yz} \end{bmatrix} \tag{44}$$

Where k is the shear correction factor which is given by $k = \dfrac{5}{6}$. Now the total potential energy of the system is as follows

$$\begin{aligned} U &= \frac{1}{2}\iiint_V e^{\mathrm{T}} s \, \mathrm{d}V \\ &= \frac{1}{2}\iiint_V \frac{E}{1-m^2}(e_{xx}^2 + 2m e_{xx}e_{yy} + e_{yy}^2 + \frac{1-m}{2}g_{xy}^2 + \frac{k(1-m)}{2}g_{xz}^2 + \frac{k(1-m)}{2}g_{yz}^2)\mathrm{d}V \\ &= U_1 + U_2 \end{aligned} \tag{45}$$

Where U_1 represents the bending strain energy and U_2 represents the in-plane strain energy of the plate. They can be denoted as

$$U_1 = \frac{1}{2}\iint_A \{\frac{Eh}{1-m^2}[(\frac{\partial w_1}{\partial x})^2 + (\frac{\partial w_2}{\partial y})^2 + 2m(\frac{\partial w_1}{\partial x})(\frac{\partial w_2}{\partial y})] + \frac{Eh}{2(1+m)}(\frac{\partial w_1}{\partial y} + \frac{\partial w_2}{\partial x})^2\}\,\mathrm{d}A \tag{46}$$

$$\begin{aligned} U_2 = {} & \frac{Eh^3}{24(1-m^2)}\iint_A [(\frac{\partial^2 j_x}{\partial x^2})^2 + (\frac{\partial^2 j_y}{\partial y^2})^2 + 2m(\frac{\partial j_x}{\partial x})(\frac{\partial j_y}{\partial y}) + \frac{1-m}{2}(\frac{\partial j_x}{\partial y} + \frac{\partial j_y}{\partial x})^2]\,\mathrm{d}A \\ & + \frac{k\,Eh}{4(1+m)}\iint_A [(\frac{\partial u_3}{\partial x} + j_x)^2 + (\frac{\partial u_3}{\partial y} + j_y)^2]\mathrm{d}A \end{aligned} \tag{47}$$

Discretize the plate using triangular elements as shown in Fig.2. The unknown deformation variables can be expressed as

$$\begin{cases} w_1(x,y,t) = \sum_{i=1}^{3} f_{1i}(x,y)q_{1i}(t) = f_1(x,y)\boldsymbol{q}_1(t) \\ w_2(x,y,t) = \sum_{i=1}^{3} f_{2i}(x,y)q_{2i}(t) = f_2(x,y)\boldsymbol{q}_2(t) \\ u_3(x,y,t) = \sum_{i=1}^{3} f_{3i}(x,y)\boldsymbol{q}_{3i}(t) = F_3(x,y)\boldsymbol{q}_3(t) \\ j_x(x,y,t) = \sum_{i=1}^{3} f_{4i}(x,y)\boldsymbol{q}_{3i}(t) = f_4(x,y)\boldsymbol{q}_3(t) \\ j_y(x,y,t) = \sum_{i=1}^{3} f_{5i}(x,y)\boldsymbol{q}_{3i}(t) = f_5(x,y)\boldsymbol{q}_3(t) \end{cases} \tag{48}$$

Where $f_{1i}(x,y)$, $f_{2i}(x,y)$, f_{3i}, f_{4i} and f_{5i} are shape functions in one element corresponding to the node i. $\boldsymbol{q}_i(t)\,(i=1,2,3)$ are generalized coordinates. Let $F_4 = zf_4$ and $F_5 = zf_5$, then substituting them into Eq.(37), we can have the displacement and velocity components

$$\begin{cases} u_1 = f_1\boldsymbol{q}_1 - \frac{1}{2}\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{H}_1(x,y)\boldsymbol{q}_3 + F_4\boldsymbol{q}_3 \\ u_2 = f_2\boldsymbol{q}_2 - \frac{1}{2}\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{H}_2(x,y)\boldsymbol{q}_3 + F_5\boldsymbol{q}_3 \end{cases} \tag{49}$$

$$\begin{cases} \dot{u}_1 = f_1\dot{\boldsymbol{q}}_1 - \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{H}_1(x,y)\dot{\boldsymbol{q}}_3 + F_4\dot{\boldsymbol{q}}_3 \\ \dot{u}_2 = f_2\dot{\boldsymbol{q}}_2 - \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{H}_2(x,y)\dot{\boldsymbol{q}}_3 + F_5\dot{\boldsymbol{q}}_3 \end{cases} \tag{50}$$

Where $\boldsymbol{H}_1(x,y)$ and $\boldsymbol{H}_2(x,y)$ are coupling shape functions which are defined by

$$\begin{cases} \boldsymbol{H}_1 = \boldsymbol{R}_{j3}^{\mathrm{T}} \int_{x_{jl}}^{x} f_{3,x}^{\mathrm{T}} f_{3,x} \mathrm{d}x \boldsymbol{R}_{j3} + \sum_{i \in mP_0} \boldsymbol{R}_{i3}^{\mathrm{T}} \int_{x_{ik}}^{x_{il}} f_{3,x}^{\mathrm{T}} f_{3,x} \mathrm{d}x \boldsymbol{R}_{i3} \\ \boldsymbol{H}_2 = \boldsymbol{R}_{j3}^{\mathrm{T}} \int_{y_{jl}}^{y} f_{3,y}^{\mathrm{T}} f_{3,y} \mathrm{d}y \boldsymbol{R}_{j3} + \sum_{i \in nP_0} \boldsymbol{R}_{i3}^{\mathrm{T}} \int_{y_{il}}^{y_{il}} f_{3,y}^{\mathrm{T}} f_{3,y} \mathrm{d}y \boldsymbol{R}_{i3} \end{cases} \tag{51}$$

If these coupling terms are neglected, the results will be divergent when the angular velocity is high. There will be a so-called dynamic stiffening problem [34].In Eq.(51), $\boldsymbol{R}_{j3}$ is the orientation matrix decided by nodal numbering of the element. The comma means the first shape derivative of shape function versus x or y. mP_0 and nP_0 denote the collection of elements through these two segments.

Let $\boldsymbol{q} = [\boldsymbol{q}_1^{\mathrm{T}}, \boldsymbol{q}_2^{\mathrm{T}}, \boldsymbol{q}_3^{\mathrm{T}}]^{\mathrm{T}}$ be the generalized coordinate vector. Substituting Eqs.(42) and (45) into Lagrange's equations of the second kind

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\boldsymbol{q}}}\right) - \frac{\partial T}{\partial \boldsymbol{q}} + \frac{\partial U}{\partial \boldsymbol{q}} = \boldsymbol{0} \tag{52}$$

Then the strong-coupled and nonlinear dynamic equations of the plate undergoing overall motion can be given by

$$\begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{0} & \boldsymbol{M}_{13} \\ \boldsymbol{0} & \boldsymbol{M}_{22} & \boldsymbol{M}_{23} \\ \boldsymbol{M}_{31} & \boldsymbol{M}_{32} & \boldsymbol{M}_{33} \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{q}}_1 \\ \ddot{\boldsymbol{q}}_2 \\ \ddot{\boldsymbol{q}}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{G}_{12} & \boldsymbol{G}_{13} \\ \boldsymbol{G}_{21} & \boldsymbol{0} & \boldsymbol{G}_{23} \\ \boldsymbol{G}_{31} & \boldsymbol{G}_{32} & \boldsymbol{G}_{33} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{q}}_1 \\ \dot{\boldsymbol{q}}_2 \\ \dot{\boldsymbol{q}}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} & \boldsymbol{K}_{13} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} & \boldsymbol{K}_{23} \\ \boldsymbol{K}_{31} & \boldsymbol{K}_{32} & \boldsymbol{K}_{33} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \boldsymbol{q}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \\ \boldsymbol{Q}_3 \end{bmatrix} \tag{53}$$

Where

$$\boldsymbol{M}_{11} = \boldsymbol{W}_{11}, \quad \boldsymbol{M}_{22} = \boldsymbol{W}_{22}, \quad \boldsymbol{M}_{33} = \boldsymbol{W}_{33} + \boldsymbol{W}_{44} + \boldsymbol{W}_{55} \tag{54}$$

$$\boldsymbol{M}_{31} = \boldsymbol{M}_{13}^{\mathrm{T}} = \boldsymbol{W}_{41} \qquad \boldsymbol{M}_{32} = \boldsymbol{M}_{23}^{\mathrm{T}} = \boldsymbol{W}_{52} \tag{55}$$

$$\boldsymbol{G}_{12} = -\boldsymbol{G}_{21}^{\mathrm{T}} = -2w_3 \boldsymbol{W}_{12} \quad \boldsymbol{G}_{23} = -\boldsymbol{G}_{32}^{\mathrm{T}} = -2w_1 \boldsymbol{W}_{23} \tag{56}$$

$$\boldsymbol{G}_{13} = -\boldsymbol{G}_{31}^{\mathrm{T}} = 2(w_2 \boldsymbol{W}_{13} - w_3 \boldsymbol{W}_{15}) \tag{57}$$

$$\boldsymbol{G}_{33} = 2w_2(\boldsymbol{W}_{43} - \boldsymbol{W}_{34}) + 2w_3(\boldsymbol{W}_{54} - \boldsymbol{W}_{45}) + 2w_1(\boldsymbol{W}_{35} - \boldsymbol{W}_{53}) \tag{58}$$

$$\boldsymbol{K}_{11} = \boldsymbol{K}_{f11} - (w_2^2 + w_3^2)\boldsymbol{W}_{11} \qquad \boldsymbol{K}_{12} = \boldsymbol{K}_{f12} + (w_1 w_2 - \dot{w}_3)\boldsymbol{W}_{12} \tag{59}$$

$$\boldsymbol{K}_{13} = (w_1 w_3 + \dot{w}_2)\boldsymbol{W}_{13} - (w_2^2 + w_3^2)\boldsymbol{W}_{14} + (w_1 w_2 - \dot{w}_3)\boldsymbol{W}_{15} \tag{60}$$

$$\boldsymbol{K}_{21} = \boldsymbol{K}_{f21} + (w_1 w_2 + \dot{w}_3)\boldsymbol{W}_{21} \qquad \boldsymbol{K}_{22} = \boldsymbol{K}_{f22} - (w_1^2 + w_3^2)\boldsymbol{W}_{22} \tag{61}$$

$$\boldsymbol{K}_{23} = (w_2 w_3 - \dot{w}_1)\boldsymbol{W}_{23} + (\dot{w}_3 + w_1 w_2)\boldsymbol{W}_{24} - (w_1^2 + w_3^2)\boldsymbol{W}_{25} \tag{62}$$

$$\boldsymbol{K}_{31} = (w_1 w_3 - \dot{w}_2)\boldsymbol{W}_{31} - (w_2^2 + w_3^2)\boldsymbol{W}_{41} + (w_1 w_2 + \dot{w}_3)\boldsymbol{W}_{51} \tag{63}$$

$$\boldsymbol{K}_{32} = (w_2 w_3 + \dot{w}_1)\boldsymbol{W}_{32} + (w_1 w_2 - \dot{w}_3)\boldsymbol{W}_{42} - (w_1^2 + w_3^2)\boldsymbol{W}_{52} \tag{64}$$

$$\begin{aligned}\boldsymbol{K}_{33} = \boldsymbol{K}_{f33} &- (w_1^2 + w_2^2)\boldsymbol{W}_{33} - (w_2^2 + w_3^2)\boldsymbol{W}_{44} - (w_1^2 + w_3^2)\boldsymbol{W}_{55} - \dot{w}_2\boldsymbol{W}_{34} + \dot{w}_1\boldsymbol{W}_{35} \\ &+ (2w_1 w_3 + \dot{w}_2)\boldsymbol{W}_{43} + (2w_1 w_2 - \dot{w}_3)\boldsymbol{W}_{45} + (2w_2 w_3 - \dot{w}_1)\boldsymbol{W}_{53} + \dot{w}_3\boldsymbol{W}_{54} \\ &+ (w_2^2 + w_3^2)\boldsymbol{D}_{11} + (w_1^2 + w_3^2)\boldsymbol{D}_{22} - (w_1 w_2 + \dot{w}_3)\boldsymbol{D}_{12} - (w_1 w_2 - \dot{w}_3)\boldsymbol{D}_{21} - a_{01}\boldsymbol{C}_1 - a_{02}\boldsymbol{C}_2\end{aligned} \tag{65}$$

$$\boldsymbol{Q}_1 = (w_2^2 + w_3^2)\boldsymbol{S}_{11}^{\mathrm{T}} - (w_1 w_2 - \dot{w}_3)\boldsymbol{S}_{21}^{\mathrm{T}} - a_{01}\boldsymbol{Y}_1^{\mathrm{T}} \tag{66}$$

$$\boldsymbol{Q}_2 = (w_1^2 + w_3^2)\boldsymbol{S}_{22}^{\mathrm{T}} - (w_1 w_2 + \dot{w}_3)\boldsymbol{S}_{12}^{\mathrm{T}} - a_{02}\boldsymbol{Y}_2^{\mathrm{T}} \tag{67}$$

$$\begin{aligned}\boldsymbol{Q}_3 = &-(w_1 w_3 - \dot{w}_2)\boldsymbol{S}_{13}^{\mathrm{T}} + (w_2^2 + w_3^2)\boldsymbol{S}_{14}^{\mathrm{T}} - (w_1 w_2 + \dot{w}_3)\boldsymbol{S}_{15}^{\mathrm{T}} - (w_2 w_3 + \dot{w}_1)\boldsymbol{S}_{23}^{\mathrm{T}} \\ &-(w_1 w_2 - \dot{w}_3)\boldsymbol{S}_{24}^{\mathrm{T}} + (w_1^2 + w_3^2)\boldsymbol{S}_{25}^{\mathrm{T}} - a_{03}\boldsymbol{Y}_3^{\mathrm{T}} - a_{01}\boldsymbol{Y}_4^{\mathrm{T}} - a_{02}\boldsymbol{Y}_5^{\mathrm{T}}\end{aligned} \tag{68}$$

In which a_{01}, a_{02} and a_{03} are the acceleration of point O in the local coordinate system which are denoted as

$$a_{01} = \dot{v}_1 + (w_2 v_3 - w_3 v_2) \qquad a_{02} = \dot{v}_2 + (w_3 v_1 - w_1 v_3) \quad a_{03} = \dot{v}_3 + (w_1 v_2 - w_2 v_1) \tag{69}$$

The constant matrices in Eq.(53) are defined by

$$\boldsymbol{W}_{ij} = \iiint_V r F_i^{\mathrm{T}} F_j \mathrm{d}V \quad (i = 1, \ \mathbf{L}\ , 5; j = 1, \ \mathbf{L}\ , 5) \tag{70}$$

$$\boldsymbol{C}_i = \iint_A r h \cdot \boldsymbol{H}_i \, \mathrm{d}A \quad (i = 1, 2) \tag{71}$$

$$\boldsymbol{D}_{1i} = \iint_A r h \cdot x \cdot \boldsymbol{H}_i \, \mathrm{d}A \quad (i = 1, 2) \tag{72}$$

$$\boldsymbol{D}_{2i} = \iint_A r h \cdot y \cdot \boldsymbol{H}_i \, \mathrm{d}A \quad (i = 1, 2) \tag{73}$$

$$\boldsymbol{S}_{1i} = \iiint_V r x F_i \, \mathrm{d}V \quad (i = 1, \ \mathbf{L}\ , 5) \tag{74}$$

$$\boldsymbol{S}_{2i} = \iiint_V r y F_i \, \mathrm{d}V \quad (i = 1, \ \mathbf{L}\ , 5) \tag{75}$$

$$\boldsymbol{Y}_i = \iiint_V r F_i \, \mathrm{d}V \quad (i = 1, \ \mathbf{L}\ , 5) \tag{76}$$

$$\boldsymbol{K}_{f11} = \iint_A \frac{Eh}{1 - m^2}\left(F_{1,x}^{\mathrm{T}} F_{1,x} + \frac{1 - m}{2} F_{1,y}^{\mathrm{T}} F_{1,y}\right)\mathrm{d}A \tag{77}$$

$$\boldsymbol{K}_{f12} = \boldsymbol{K}_{f21}^{\mathrm{T}} = \iint_A \frac{Eh}{1 - m^2}\left(m F_{1,x}^{\mathrm{T}} F_{2,y} + \frac{1 - m}{2} F_{1,y}^{\mathrm{T}} F_{2,x}\right)\mathrm{d}A \tag{78}$$

$$\boldsymbol{K}_{f22} = \iint_A \frac{Eh}{1 - m^2}\left(F_{2,y}^{\mathrm{T}} F_{2,y} + \frac{1 - m}{2} F_{2,x}^{\mathrm{T}} F_{2,x}\right)\mathrm{d}A \tag{79}$$

$$\begin{aligned}\boldsymbol{K}_{f33} = \iint_A \frac{Eh^3}{12(1-m^2)}[&F_{4,x}^{\mathrm{T}}F_{4,x} + F_{5,y}^{\mathrm{T}}F_{5,y} + m(F_{4,x}^{\mathrm{T}}F_{5,y} + F_{5,y}^{\mathrm{T}}F_{4,x}) \\ &+ \frac{1-m}{2}(F_{4,y}^{\mathrm{T}}F_{4,y} + F_{5,x}^{\mathrm{T}}F_{5,x} + F_{4,y}^{\mathrm{T}}F_{5,x} + F_{5,x}^{\mathrm{T}}F_{4,y})]\mathrm{d}A \\ &+ \frac{kEh}{2(1+m)}\iint_A [F_{3,x}^{\mathrm{T}}F_{3,x} + F_{3,y}^{\mathrm{T}}F_{3,y} + F_4^{\mathrm{T}}F_4 + F_5^{\mathrm{T}}F_5 \\ &+ F_{3,x}^{\mathrm{T}}F_4 + F_4^{\mathrm{T}}F_{3,x} + F_{3,y}^{\mathrm{T}}F_5 + F_5^{\mathrm{T}}F_{3,y}]\mathrm{d}A\end{aligned} \tag{80}$$

Nonlinear dynamic equation (53) can be used not only for the analysis of thin plates, but also for the analysis of thick plates. These underlined terms in Eq.(65) are additional dynamic stiffness terms which are caused by considering the coupling shape functions. If these coupling terms are neglected, the results will be divergent when the angular velocity is high which is mentioned in reference [34].

4.2 Formulation for vibration analysis of a rotating cantilever Mindlin plate

Consider a flexible plate attached to a rigid hub with radius R, and rotating around the y axis with a constant rotation speed W in the local coordinate system xyz which is fixed to the neutral surface of the plate as shown in Fig.5. The physical parameters of the plate are as follows: length a, width b, thickness h, Young's modulus E, mass density ρ and Poisson's ratio μ.

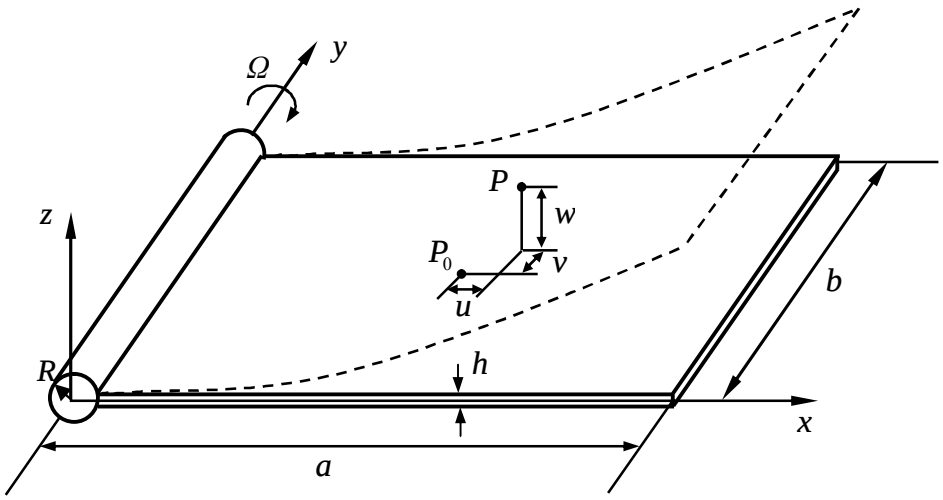


Fig. 6 The configuration of a rotating cantilever Mindlin plate

In the local coordinate system xyz, the velocity and angular velocity of point O in the direction of x, y, and z axis are

$$v_1 = v_2 = 0, v_3 = -RW, w_1 = w_3 = 0, w_2 = W \tag{81}$$

Ignoring the in-plane motions of the plate and the right-hand side terms in Eq.(53), the dynamic equation for the free vibration analysis of the rotating plate can be obtained as

$$\boldsymbol{M}_{33}\ddot{\boldsymbol{q}}_3 + [\underline{W^2(\boldsymbol{RC}_1 + \boldsymbol{D}_{11})} - W^2(\boldsymbol{W}_{33} + \boldsymbol{W}_{44}) + \boldsymbol{K}_{f33}]\boldsymbol{q}_3 = \boldsymbol{0} \tag{82}$$

Note that the underlined term is the dynamic stiffness term due to rotation, the second term is

the dynamic softness term and the last term is the static stiffness term. Rewrite Eq.(82) in a non-dimensional form. The following non-dimensional variables are defined:

$$d=\frac{a}{b},\quad h=\frac{h}{a},\quad s=\frac{R}{a},\quad v=wT,\quad g=WT \tag{83}$$

Where $T=\sqrt{r ha^4/D}$. Then Eq.(82) can be rewritten in the non-dimensional form:

$$\bar{\boldsymbol{M}}_{33}\ddot{\boldsymbol{q}}_3+[\underline{W^2(R\bar{\boldsymbol{C}}_1+\bar{\boldsymbol{D}}_{11})}-W^2(\bar{\boldsymbol{W}}_{33}+\bar{\boldsymbol{W}}_{44})+\bar{\boldsymbol{K}}_{f33}]\boldsymbol{q}_3=\boldsymbol{0} \tag{84}$$

Their expressions can be found in Eqs.(70),(71),(72) and (80). The difference is that the integration of constant matrices is from 0 to 1 in Eq.(84).

5 Numerical results

5.1 Elimination of shear locking

To examine the efficiency of CS-FEM for static deflection analyses, consider a rectangular plate with uniform load f=1N/m^2 as shown in Fig.1. The geometric and material property parameters are as follows: $a=10.0$m , $b=10.0$m , $E=1.0\times10^9$ N/m^2 and $m=0.3$. Define a deflection coefficient $x=w_{\max}D/fb^4$, where $w_{\max}$ is the maximum deflection at the center of the plate and the elastic rigidity of the plate is $D=\dfrac{Eh^3}{12(1-m^2)}$.

Table 1 shows the deflection coefficient of the clamped plate against the different mesh densities $N\times N$ for the thin plate (aspect ratio h/a=0.001) and thick plate(aspect ratio h/a=0.1). It is seen that the CS-FEM and FEM with DSG method both provide a locking-free solution. The plate is meshed by more elements, the results will be more accurate. In addition, the results of CS-FEM are more accurate and softer than those of FEM with the same DOFs for both thin and thick plates. The deflection coefficient of the simply supported plate against the different mesh densities $N\times N$ for the thin plate (aspect ratio h/a=0.001) and thick plate(aspect ratio h/a=0.1) is presented in Table2. The same comments as in the clamped plate can be obtained again.

Table 1 The deflection coefficient of the clamped plate

h/a	Method	Mesh densities $N\times N$					Analytical solutions[35]
		4×4	8×8	10×10	16×16	20×20	
0.001	FEM	0.000906	0.001121	0.001167	0.001225	0.001239	0.001266
	CS-FEM	0.001123	0.001227	0.001241	0.001256	0.001259	
0.1	FEM	0.001203	0.001425	0.001456	0.001487	0.001493	0.001499
	CS-FEM	0.001357	0.001467	0.001480	0.001495	0.001498	

Table 2 The deflection coefficient of the simply supported plate

h/a	Method	Mesh densities $N\times N$					Analytical solutions[35]
		4×4	8×8	10×10	16×16	20×20	
0.001	FEM	0.002949	0.003728	0.003844	0.003975	0.004006	0.004062
	CS-FEM	0.003517	0.003928	0.003977	0.004030	0.004042	
0.1	FEM	0.003349	0.004058	0.004142	0.004225	0.004243	0.004273
	CS-FEM	0.003748	0.004143	0.004190	0.004240	0.004252	

5.2 Free vibration analysis of the plate with different boundary conditions

Consider a rectangular plate as shown in Fig.1. The geometric and material property parameters are as follows: $a=10.0$m , $b=10.0$m , $E=2.0\times10^{11}$ N/m^2 , r=8000 kg/m^3 and $m=0.3$. Define a dimensionless frequency coefficient $v=(w^2 r a^4 h/D)^{1/4}$, where w is the natural frequency and D is the elastic rigidity of the plate which is the same as mentioned in last section. The combined boundary condition is defined by different symbols. The symbols F, S and C represent the free, simply supported and clamped boundary conditions, respectively. For example, SFCF means a combined boundary condition for a plate whose four edges are simply supported, free, clamped and free. In order to get more accurate results [36], the well-known lumped mass matrix is used in this paper.

Tables 3 and 4 show the six lowest dimensionless frequencies of thin plate (aspect ratio h/a=0.005) and thick plate (aspect ratio h/a=0.1)with SSSS boundary condition. It is seen that the results of the CS-FEM agree well with the results of reference [37]. For the thin plate, there is no shear locking phenomenon because of using DSG method. The results of CS-FEM are more accurate than those of FEM with the same DOFs. In particular, the CS-FEM can provide accurately values of frequencies even if using coarse meshes. Tables 5 and 6 show the six lowest dimensionless frequencies of thin plate (aspect ratio h/a=0.005) and thick plate (aspect ratio h/a=0.1) with CCCC boundary condition and the obtained comments in the SSSS plate are confirmed for the CCCC plate again. The other five different boundary conditions: SSSF, SFSF, CCCF, CFCF and CFSF for thin plate (aspect ratio h/a=0.005) are listed in Table 7. The plate is discretized with 2×16×16 triangular elements. It is again observed that the results of CS-FEM agree well with the results of reference. The six lowest shape modes of thin plate (aspect ratio h/a=0.005) with SSSS boundary condition are plotted in Fig.7. It is seen that they express exactly the real physical modes and there is no spurious energy modes are found.

Table 3 The dimensionless frequency coefficient v of a SSSS plate (h/a=0.005)

Mesh	Method	Mode number					
		1	2	3	4	5	6
4×4	FEM	4.9382	7.8786	8.9396	10.5890	12.2902	12.9570

8×8	CS-FEM	4.4965	7.1241	7.2503	9.0931	10.0933	10.1619
	FEM	4.5708	7.2889	7.5694	9.6691	10.8368	11.0471
16×16	CS-FEM	4.4543	7.0536	7.0791	8.9750	10.0418	10.0477
	FEM	4.4745	7.0941	7.1603	9.1230	10.1731	10.1872
20×20	CS-FEM	4.4453	7.0310	7.0367	8.9051	9.9590	9.9592
	FEM	4.4629	7.0691	7.1108	9.0396	10.0871	10.0928
	CS-FEM	4.4443	7.0284	7.0320	8.8972	9.9492	9.9493
	Exact [37]	4.4430	7.0250	7.0250	8.8860	9.9350	9.9350

Table 4 The dimensionless frequency coefficient v of a SSSS plate (h/a=0.1)

Mesh	Method	Mode number					
		1	2	3	4	5	6
4×4	FEM	4.7556	7.4493	8.1649	9.6534	10.9393	11.3686
	CS-FEM	4.4032	6.7790	6.8435	8.3901	9.0714	9.0889
8×8	FEM	4.4526	6.9470	7.0927	8.8583	9.8241	9.9006
	CS-FEM	4.3743	6.7560	6.7712	8.3830	9.2329	9.2341
16×16	FEM	4.3861	6.7950	6.8252	8.4862	9.3746	9.3788
	CS-FEM	4.3683	6.7470	6.7511	8.3623	9.2256	9.2257
20×20	FEM	4.3788	6.7766	6.7954	8.4386	9.3195	9.3211
	CS-FEM	4.3676	6.7460	6.7486	8.3595	9.2242	9.2243
	Exact [37]	4.37	6.74	6.74	8.35	9.22	9.22

Table 5 The dimensionless frequency coefficient v of a CCCC plate (h/a=0.005)

Mesh	Method	Mode number					
		1	2	3	4	5	6
4×4	FEM	6.8025	9.5749	10.4873	11.7647	13.1176	13.5040
	CS-FEM	6.1712	8.6783	8.9731	10.3804	11.0673	11.2107
8×8	FEM	6.2735	9.0386	9.3955	11.5044	12.7057	12.9603
	CS-FEM	6.0475	8.6471	8.7198	10.5863	11.7010	11.7459
16×16	FEM	6.0709	8.6976	8.7894	10.7738	11.8234	11.8616
	CS-FEM	6.0101	8.5862	8.6030	10.4502	11.5285	11.5556
20×20	FEM	6.0446	8.6504	8.7090	10.6462	11.6973	11.7265

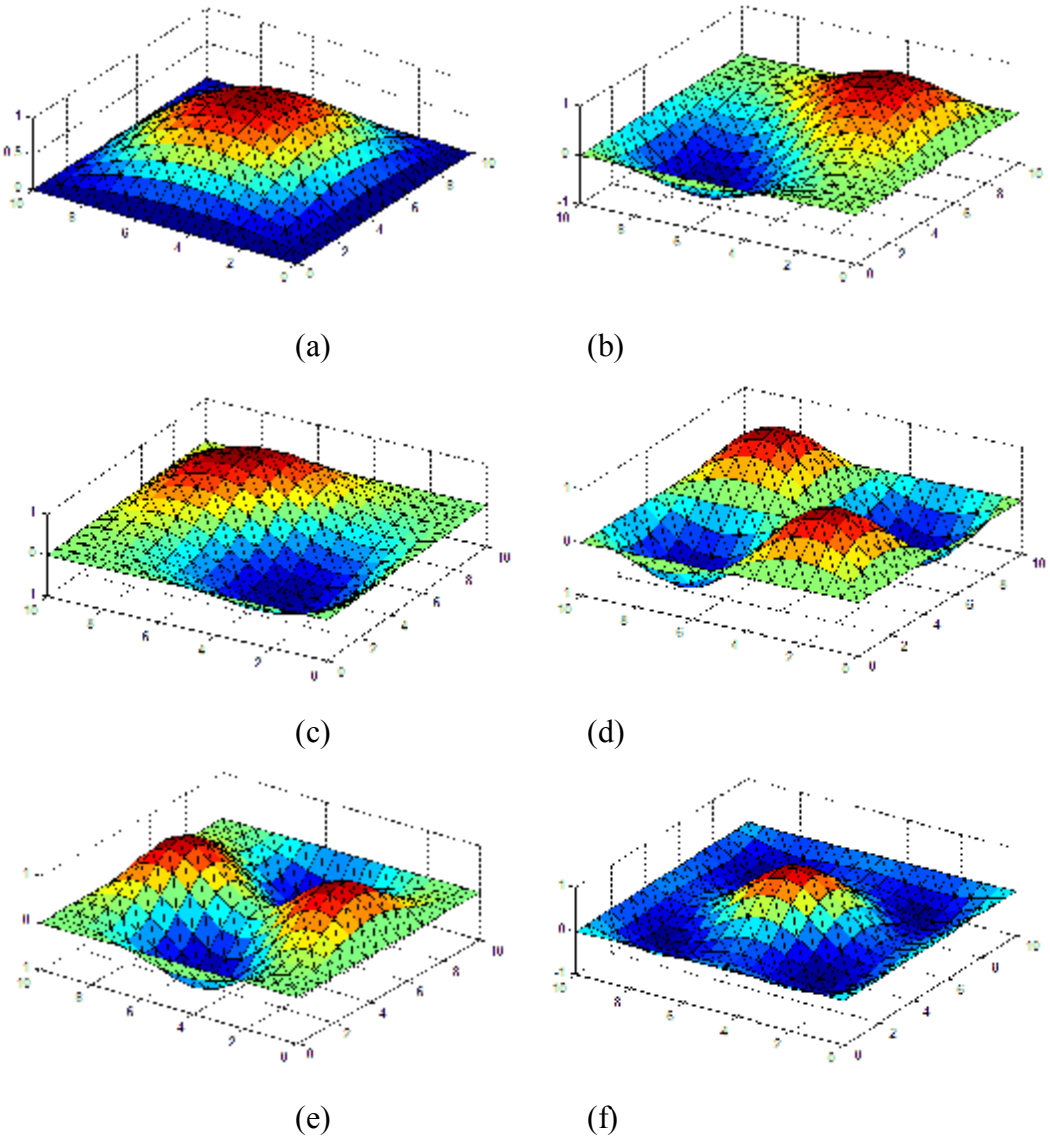
	CS-FEM	6.0057	8.5784	8.5891	10.4317	11.5059	11.5328
	Exact [37]	5.999	8.568	8.568	10.407	11.472	11.498

Table 6 The dimensionless frequency coefficient v of a CCCC plate (h/a=0.1)

Mesh	Method	Mode number					
		1	2	3	4	5	6
4×4	FEM	6.2704	8.5690	9.2591	10.3480	11.3969	11.7235
	CS-FEM	5.8163	7.8647	8.0481	9.2126	9.6233	9.7156
8×8	FEM	5.8578	8.1344	8.3480	9.8978	10.7828	10.9282
	CS-FEM	5.7350	7.9027	7.9501	9.3817	10.1442	10.2003
16×16	FEM	5.7402	7.9437	7.9937	9.4901	10.3007	10.3528
	CS-FEM	5.7117	7.8846	7.8965	9.3442	10.1319	10.1803
20×20	FEM	5.7267	7.9197	7.9513	9.4319	10.2369	10.2858
	CS-FEM	5.7087	7.8819	7.8895	9.3378	10.1284	10.1766
	Exact [37]	5.71	7.88	7.88	9.33	10.13	10.18

Table 7 The dimensionless frequency coefficient v of a thin plate (aspect ratio h/a=0.005) with various boundary conditions

Boundary type	Mode number			
	1	2	3	4
SSSF	3.4168	5.2565	6.4200	7.6802
Exact[37]	3.4176	5.2684	6.4185	7.6854
SFSF	3.1033	4.0072	6.0235	6.2430
Exact[37]	3.1034	4.0168	6.0602	6.2406
CCCF	4.8945	6.3176	7.9672	8.7349
Exact[37]	4.9010	6.3276	7.9682	8.7613
CFCF	4.7131	5.1372	6.5755	7.8384
Exact[37]	4.7193	5.1506	6.6079	8.0291
CFSF	3.8994	4.5307	6.2707	7.0404
Exact[37]	3.9096	4.5468	6.3152	7.0356



(a) (b)

(c) (d)

(e) (f)

Fig. 7 The six lowest shape modes of thin plate (aspect ratio h/a=0.005) with SSSS boundary condition. (a)-(h): 1-8 shape modes

5.3 Free vibration analysis of a rotating cantilever Mindlin plate

In order to examine the efficiency of the CS-FEM, the results are compared with those of FEM and AMM which are based on Kirchhoff plate theory. The plate is discretized with 2×16×16 triangular elements in CS-FEM and FEM. Five cantilever beam mode functions and seven free-free beam mode functions are combined to generate 35 plate mode functions in AMM according to reference [11]. Table 8 shows the lowest five dimensionless natural frequencies with $d=1, h=0.01$ and $s=0$. It is seen that the dimensionless natural frequencies increase with the increasing angular velocity. Under the same angular velocity, the results of AMM is bigger than those of FEM, which means AMM provides stiffer results if using the same modeling theory. The results of CS-FEM are always smaller than the other two methods. That means the Mindlin plate theory make the structure become softer because of considering the shear deformation. In other words, the Kirchhoff plate theory is always overvalued on the

natural frequencies of the structure. Table 9 shows the lowest five dimensionless natural frequencies with d=1 , h=0.01 and s =1 . The same comments obtained above can be confirmed again. Compared with the results in Table 8, it is observed that the dimensionless natural frequencies increase with the increasing hub radius.

Table 8 Five lowest dimensionless natural frequencies of a rotating plate

(d=1,h=0.01, s =0)

Non-dimensional angular velocity	Mode	CS-FEM	FEM	AMM
g =1	1	3.4963	3.4983	3.5156
	2	8.4799	8.5215	8.5328
	3	21.3255	21.4703	21.520
	4	27.0454	27.1473	27.353
	5	30.8780	31.0911	31.206
g =2	1	3.5751	3.5760	3.5963
	2	8.4901	8.5357	8.5507
	3	21.6413	21.8101	21.865
	4	27.0245	27.1756	27.384
	5	31.1047	31.3537	31.477

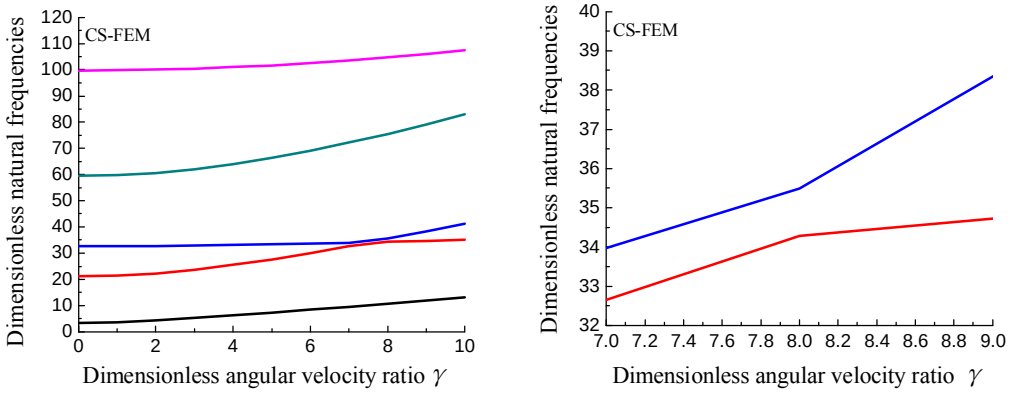
Table 9 Five lowest dimensionless natural frequencies of a rotating plate

(d=1,h=0.01, s =1)

Non-dimensional angular velocity	Mode	CS-FEM	FEM	AMM
g =1	1	3.7127	3.7151	3.7324
	2	8.5944	8.6112	8.6240
	3	21.4835	21.6533	21.706
	4	27.0333	27.183	27.394
	5	30.9789	31.2315	31.350
g =2	1	4.3615	4.3670	4.3805
	2	8.8492	8.8878	8.9087
	3	22.3442	22.5107	22.580
	4	27.2023	27.3434	27.557
	5	31.6681	31.9054	32.043

Fig.8 shows the five lowest dimensionless natural frequencies of a rotating cantilever plate(d = 5， h = 0.01， s =1) using CS-FEM versus angular velocity. It is seen that the

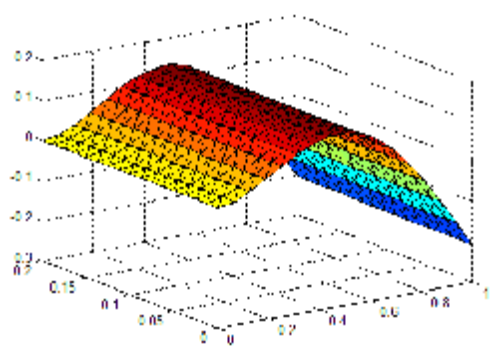
second and third frequency loci approach each other as the angular velocity increases and then veer away from each other. This interesting phenomenon is referred to as eigenvalue loci veering and was first discussed by Leissa [38]. Yoo [11] said these two frequency loci crossed each other between symmetric and skew-symmetric modes. However, it can be found from Fig.8(b) that they are only very closely each other and the eigenvalue crossing doesn't occur. Compared with reference [11] using AMM, the results of CS-FEM are milder.



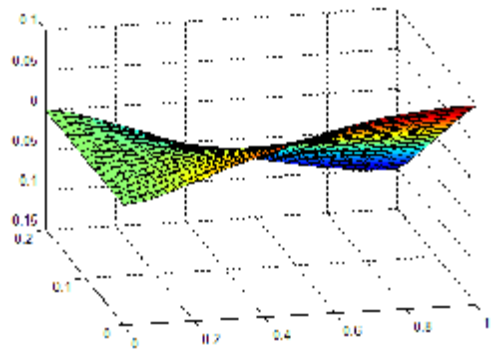
(a) (b)

Fig. 8 The five lowest dimensionless natural frequencies of a rotating cantilever plate versus angular velocity ($d=5$, $h=0.01$, $s=1$)

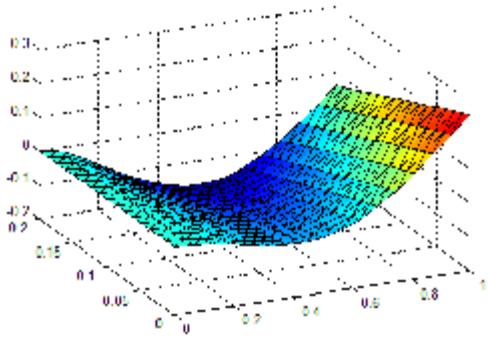
Fig.9 shows the second and third vibration modes in the veering region when the non-dimensional angular velocity is 7, 7.58,7.6 and 8 with $d=5$, $h=0.01$ and $s=1$, respectively. When the non-dimensional angular velocity is 7, it is clearly observed from figs.(a) and (b) that the second mode is bending mode and the third mode is torsion mode. When the non-dimensional angular velocity is 8, it is clearly observed from figs. (g) and (h) that the second mode is torsion mode and the third mode is bending mode. This phenomenon means that the second and third modes switch their shapes when the non-dimensional angular velocity changes from 7 to 8. This switching phenomenon does not occur suddenly but has a process. It is clearly seen from figs.(c),(d),(e) and (f) that there are both bending and torsion modes in the second and third modes. When the non-dimensional angular velocity is 7.58, the bending effect is greater than the torsion effect for the second mode and the torsion effect is greater than the bending effect for the third mode. That means the torsion effect is increasing for the second mode and the bending effect is increasing for the third mode when the non-dimensional angular velocity increases in the veering region. When the non-dimensional angular velocity is 7.6, we see the opposite situation, which the torsion effect is greater than the bending effect for the second mode and the bending effect is greater than the torsion effect for the third mode. Finally, when the non-dimensional angular velocity increases to 8, the switching process is complete.



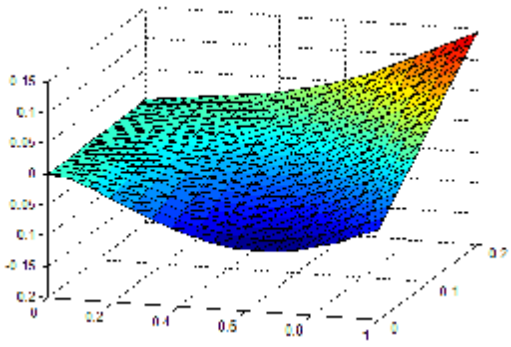
(a) g =7 Second mode



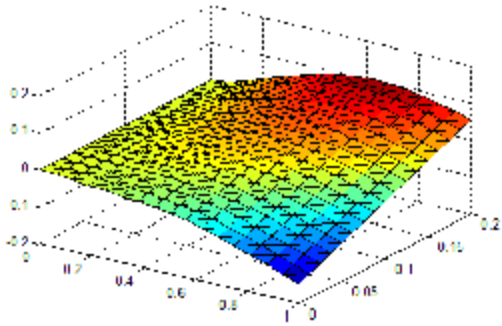
(b) g =7 Third mode



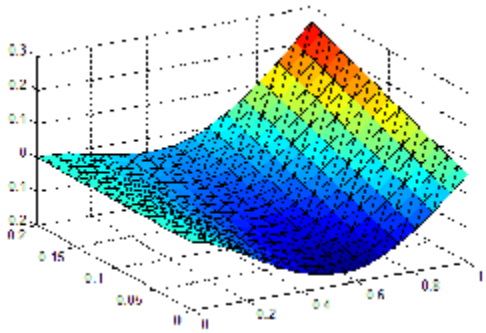
(c) g =7.58 Second mode



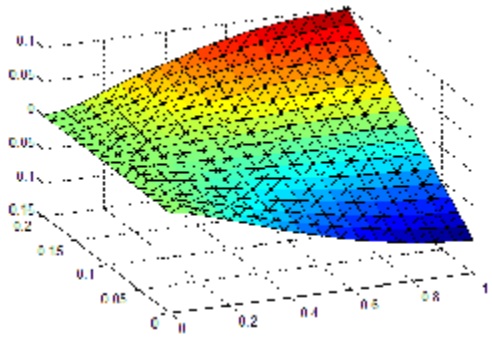
(d) g =7.58 Third mode



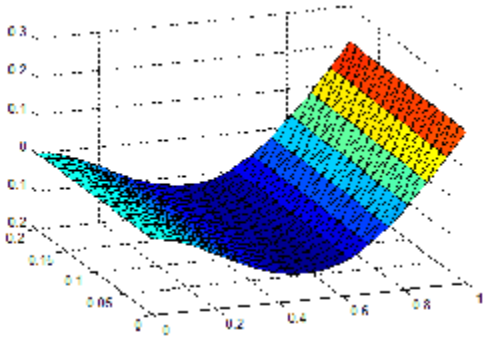
(e) g =7.6 Second mode



(f) g =7.6 Third mode



(g) g =8 Second mode



(h) g =8 Third mode

Fig. 9 The second and third vibration modes of a rotating cantilever plate

$(d = 5,\ h = 0.01,\ s = 1)$

Fig.10 shows the eight lowest dimensionless natural frequencies of a rotating cantilever

plate($d=1$， $h=0.01$， $g=10$) versus hub radius ratio s . It is seen that the dimensionless natural frequencies increase as the hub radius ratio increases. The first two frequencies are very closely. The frequency loci veering occurs from the fourth to seventh natural frequencies and there are two veering phenomena in the sixth natural frequency. Fig.11 shows the eight lowest dimensionless natural frequencies of a rotating cantilever plate($s=0$， $h=0.01$， $g=10$) versus aspect ratio d . It is observed that there are many abrupt frequency loci veering phenomena. Compared with fig.10, the aspect ratio has a greater effect on the frequency loci veering phenomena than the hub radius ratio. Fig.12 shows the eight lowest dimensionless natural frequencies of a rotating cantilever plate ($d=1$， $s=0, g=10$) versus thickness ratio h . The results of FEM are based on Kirchhoff plate theory and those of CS-FEM are based on Mindlin plate theory. It is seen that the results of FEM are constant and they don't change with the thickness ratio. However, the results of CS-FEM decrease as the thickness ratio increases. In the low order frequencies, the results of these two modeling theories are very closely, which means different modeling theories have a small effect on low order frequencies but have a great effect on high order frequencies. In addition, the results of CS-FEM are always smaller than those of FEM and this is confirmed again that the Kirchhoff theory overestimates the structural dynamic characteristics.

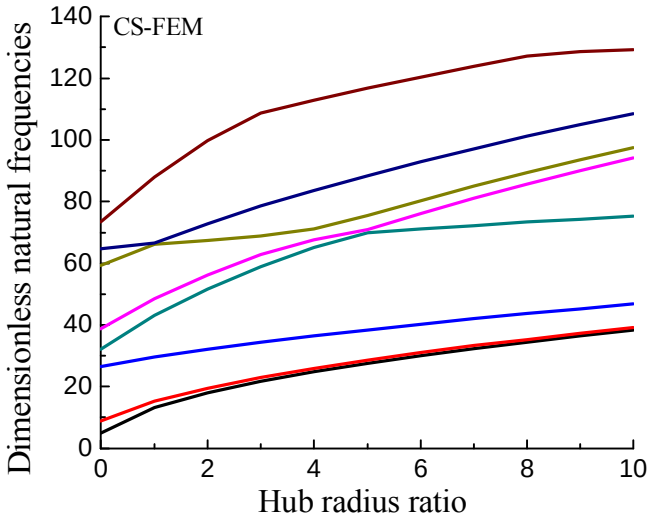


Fig. 10 The eight lowest dimensionless natural frequencies of a rotating cantilever plate versus hub radius ratio s $(d=1, h=0.01, g=10)$

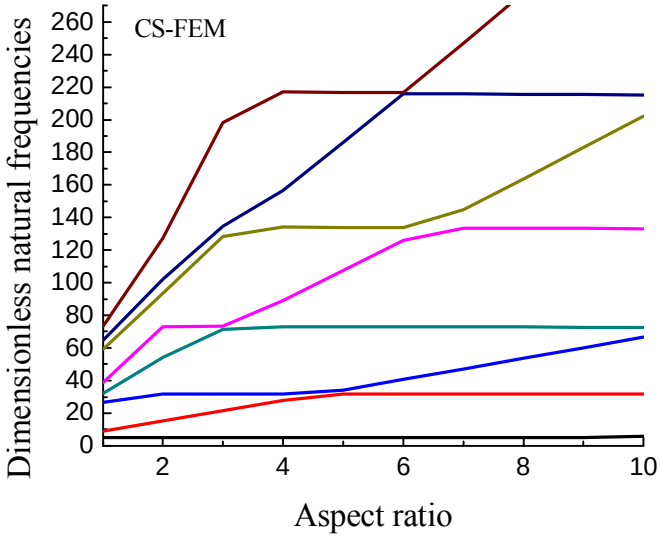


Fig. 11 The eight lowest dimensionless natural frequencies of a rotating cantilever plate versus aspect ratio d $(s=0,\ h=0.01,\ g=10)$

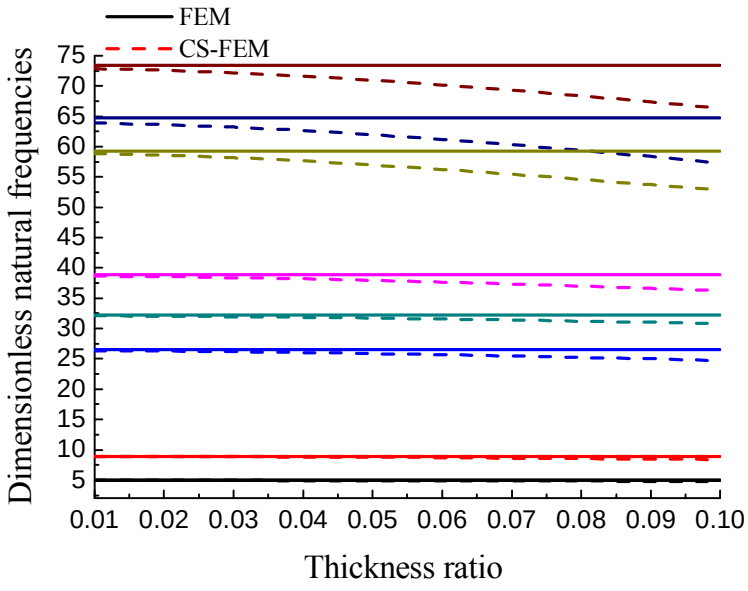


Fig. 12 The eight lowest dimensionless natural frequencies of a rotating cantilever plate versus thickness ratio h $(d=1,\ s=0, g=10)$

6 Conclusion

In this paper, a cell-based smoothed finite element method (CS-FEM) is formulated for non-linear free vibration analysis of rotating Mindlin plates. In order to overcome the shear locking problem, the discrete shear gap (DSG) method is used. The static cases and free vibration analysis of plates with various boundary conditions demonstrate the effectiveness of the CS-FEM. It is found that the CS-FEM based on Mindlin plate theory can provide more

accurate and "softer" solution compared with those of the conventional FEM even if using coarse meshes. For the analysis of free vibration of a rotating cantilever plate, the CS-FEM results are compared with the FEM and AMM. It is found that the natural frequencies of neighboring modes may "kissing" each other, when the angular velocity, aspect ratio and hub radius ratio changes, but they do not go cross. At the frequency kissing point, the vibration modes switch. It is also found that because of the use of the Mindlin plate theory, the natural frequencies decrease as the thickness ratio increases, which is not observed when the Kirchhoff plate theory is used. Moreover, the effect of thickness ratio is more significant in high order frequencies.

Acknowledgment

The authors are grateful for the support from the National Natural Science Foundation of China (Grant Nos. 11272155, 11132007, 11502113), the 333 Project of Jiangsu Province, China (Grant No. BRA2011172), the Fundamental Research Funds for the Central Universities of China (Grant No. 30920130112009), the innovation of postgraduate education project of Jiangsu Province, China (Grant No. KYLX15_0404), and the China Scholarship Council for one year study at the University of Cincinnati.

References

[1] R. Southwell and F. Gough, The free transverse vibration of airscrew blades, *British A.R.C. Reports and Menoranda* No. 766 (1921).

[2] H.H. Yoo, R.R. Ryan, R.A. Scott, Dynamics of flexible beams undergoing overall motions, *Journal of Sound and Vibration* 181 (1995) 261-278.

[3] H.H. Yoo, S.H. Shin, Vibration analysis of rotating cantilever beams, *Journal of Sound and Vibration* 212 (1998) 807-828.

[4] L. Li, D.G. Zhang, W.D. Zhu, Free vibration analysis of a rotating hub-functionally graded material beam system with the dynamic stiffening effect, *Journal of Sound and Vibration* 333 (2014) 1526-1541.

[5] J. Chung, H.H. Yoo, Dynamic analysis of a rotating cantilever beam by using the finite element method, *Journal of Sound and Vibration* 249 (2002) 147-164.

[6] H. Du , M.K. Lira, K.M. Liew, A nonlinear finite element model for dynamics of flexible manipulators, *Mech. Mach. Theory* 31 (1996) 1109-1119.

[7] G.G. Sanborn, A.A. Shabana, A rational finite element method based on the absolute nodal coordinate formulation, *Nonlinear Dynamics* 58 (2009) 565-572.

[8] M.A. Dokainish, S. Rawtani, Vibration analysis of rotating cantilever plates, *International Journal for Numerical Methods in Engineering* 3 (1971) 233-248.

[9] V. Ramamurti, R. Kielb, Natural frequencies of twisted rotating plates, *Journal of Sound and Vibration* 97 (1984) 429-449.

[10] H.H. Yoo, J. Chung, Dynamics of rectangular plate undergoing prescribed overall motions, *Journal of Sound and Vibration* 280 (2005) 531-553.

[11] H.H. Yoo, C. Pierre, Modal characteristic of a rotating rectangular cantilever plate, *Journal of Sound and Vibration* 259 (2003) 81-96.

[12] S.H. Hashemi, S. Farhadi, S. Carra, Free vibration analysis of rotating thick plates, *Journal of Sound and Vibration* 323 (2009) 366-384.

[13] G.R. Liu, T. Nguyen-Thoi, Smoothed finite element methods, CRC Press: Taylor and Francis Group, New York, 2010.

[14] G.R. Liu, K.Y. Dai, T. Nguyen-Thoi, A smoothed finite element for mechanics problems, *Computational Mechanics* 39 (2007) 859-877.

[15] K.Y. Dai, G.R. Liu, Free and forced vibration analysis using the smoothed finite element method (SFEM), *Journal of Sound and Vibration* 301 (2007) 803-820.

[16] G.R. Liu, T. Nguyen-Thoi, K.Y. Dai, *et al*, Theoretical aspects of the smoothed finite element method (SFEM), *International Journal for Numerical Methods in Engineering* 71 (2007) 902-930.

[17] G.R. Liu, T. Nguyen-Thoi, H. Nguyen-xuan, *et al*, A node-based smoothed finite element method (NS-FEM) for upper bound solutions to solid mechanics problems, *Computers and Structures* 87 (2009) 14-26.

[18] T. Nguyen-Thoi, G.R. Liu, H. Nguyen-xuan, *et al*, Adaptive analysis using the node-based smoothed finite element method (NS-FEM), *Communications in Numerical Methods in Engineering* 27 (2009) 198-218.

[19] T. Nguyen-Thoi, G.R. Liu, H. Nguyen-xuan, Additional propertices of the node-based smoothed finite element method (NS-FEM) for solid mechanics problems, *International Journal of Computational Methods* 6 (2009) 633-666.

[20] G.R. Liu, T. Nguyen-Thoi, K.Y. Lam, An edge-based smoothed finite element method (ES-FEM) for static, free and forced vibration analyses of solids, *Journal of Sound and Vibration* 320 (2009) 1100-1130.

[21] T. Nguyen-Thoi, G.R. Liu, H.C. Vu-Do, *et al*, An edge-based smoothed finite element method (ES-FEM) for visco-elastoplastic analyses of 2D solids using triangular mesh, *Computational Mechanics* 45 (2009) 23-44.

[22] T.N. Tran, G.R. Liu, T. Nguyen-Thoi, *et al*, An edge-based smoothed finite element method for primal-dual shakedown analysis of structures, *International Journal for Numerical Methods in Engineering* 82 (2010) 917-938.

[23] T. Nguyen-Thoi, G.R. Liu, K.Y. Lam, *et al*, A face-based smoothed finite element method (FS-FEM) for 3D linear and nonlinear solid mechanics problems using 4-node tetrahedral elements, *International Journal for Numerical Methods in Engineering* 78 (2009) 324-353.

[24] T. Nguyen-Thoi, G.R. Liu, H.C. Vu-Do, *et al*, A face-based smoothed finite element method (FS-FEM) for visco-elastoplastic analyses of 3D solids using triangular mesh, *Communications in Numerical Methods in Engineering* 198 (2009) 3479-3498.

[25] R.H. Macneal, Derivation of element stiffness matrices by assumed strain distributions, *Nuclear Engineering and Design* 7 (1982) 3-12.

[26] K.C. Park, G.M. Stanley, A curved C0 shell element based on assumed natural-coordinate strains, Journal of Applied Mechanics 53 (1986) 278-290.

[27] F. Brezzi, K.J. Bathe, M. Fortin, Mixed-interpolated elements for Reissner-Mindlin plates, *International Journal for Numerical Methods in Engineering* 28 (1989) 1787-1801.

[28] K.J. Bathe, E.N. Dvorkin, A four-node plate bending element based on Mindlin-Reissner plate theory and a mixed interpolation, *International Journal for Numerical Methods in Engineering* 21 (1985) 367-383.

[29] E. Onate, O.C. Zienkiewicz, B. Suarez, R.L. Taylor, A general methodology fir deriving shear constrained Reissner-Mindlin plate elements. *International Journal for Numerical Methods in Engineering* 33 (1992) 345-367.

[30] Zengjie C, Wanji C, Refined triangular discrete Mindlin flat shell elements, *Computational Mechanics* 33 (2003) 52-60.

[31] K.U. Bletzinger, M. Bischoff, E. Ramm, A unified approach for shear-locking free triangular and rectangular shell finite element, *Computational Mechanics* 75 (2000) 321-334.

[32] M. Bischoff, K.U. Bletzinger, Stabilized DSG plate and shell elements, *Intrends in computational structural mechanics*. CIMNE: Barcelona, Spain, 2001.

[33] M. Lyly, R. Stenberg, T. Vihinen, A stable bilinear element for the Reissner-Mindlin plate model, *Computer*

Methods Applied Mechanics Engineering 110 (1993) 343-357.

[34] A.K. Banerjee, T.R. Kane, Dynamics of a plate in large overall motion, *Journal of Applied Mechanics* 56 (1989) 887-892.

[35] R.L. Taylor, F. Auricchio, Linked interpolation for Reissner-Mindlin plate element. Part Ⅱ-a simple triangle, *International Journal for Numerical Methods in Engineering* 36 (1993) 3057-3066.

[36] T. Nguyen-Thoi, P. Phung-Van, H. Nguyen-Xuan *et al*, A cell-based smoothed discrete shear gap method using triangular elements for static and free vibration analyses of Reissner-Mindlin plates, *International Journal for Numerical Methods in Engineering* 91 (2012) 705-741.

[37] F. Abbassian, D.J. Dawswell, N.C. Knowles, Free vibration benchmarks, Atkins Engineering Science: Glasgow, 1987.

[38] A. Leissa, On a curve veering aberration, *Journal of Applied Mathematics and Physics* 25 (1974) 99-111.

Design of a Speed Adaptive Controller for DC Shunt Connected Motors using Neural Networks

†R. Tapia-Olvera[1], F. Beltran-Carbajal[2], Z. Damián-Noriega[2] and *G.D. Alvarez-Miranda[2]

[1]Departamento de Ingeniería, Universidad Politécnica de Tulancingo, México.
[2]Departamento de Energía, Universidad Autónoma Metropolitana, Unidad Azcapotzalco, México City, México.

*Presenting author: gdam@azc.uam.mx
†Corresponding author: ruben.tapia@upt.edu.mx

Abstract

Improving the applicability of electrical machines depends on knowing their performance on different operation conditions. In this paper a technique based on B-spline neural networks for obtaining a high performance of direct current shunt motors is proposed. This algorithm sets the control signal on line without the need to know a system model and, therefore, their performance is not dependent on the equilibrium point of design and prior knowledge of the parameters. Motor operation is subjected to highly demanding conditions for variant speed reference, also takes advantage of the feature of including a load torque from zero to full with minimal impact on the rotor speed. Time domain simulations and laboratory measurements in a test direct current shunt motor demonstrate the applicability of the proposal.

Keywords: Automatic Learning, DC shunt motors, Model-Free Control, Neural Networks.

Introduction

Currently, electrical machines are presented in a wide variety of applications from use in homes to remote research applications on land, in air, in water and finally in space, each with its own characteristics and specific protections [1]. However, the demanding operation conditions are increasingly, consequently it is necessary to develop new proposals for operation, control and protection [2][3].

Although direct current, DC, machines have been studied, there are still many possibilities for its use as motors and generators [4]. It is an open research topic mainly for implementation purpose with robust performance and low digital control demand. Therefore, it must be demonstrated satisfactory performance in a wide range of operating conditions and adaptability characteristics in face to the changing demands of the load torque and environmental conditions. DC motors are used in many areas such as mobile robotics, industrial robotic arms, elevators, cranes, drills, in addition, its simple model (some configurations) facilitates their use as test systems for evaluating new drivers [4].

The control area is widely in electric motors [4]-[9]. There are several controllers based on conventional PID linear techniques or a combination, robust based on sliding mode, adaptive algorithms, asymptotic differentiation, neural networks, and fuzzy logic. Most of them require full or partial information of the mathematical model and motor parameters, limiting its application because control laws are dependent on having available these values [4]-[6]. Consequently, if are not known, the performance of the control law is degraded. Additionally, some of them have a high computational cost which restricts its operation in real time applications [5].

In particular the analysis and design of DC motors controllers is emphasized for so called direct current permanent magnet, or separate excitation, reaching simple linear models, and justifying its performance against certain operating conditions [4]-[6][9]. Its operation is guaranteed around equilibrium point and also is highly dependent on prior knowledge of the motor parameters. In this configuration the inrush current is limited only by the armature resistance, this resistance is relatively high for small motors and there are no problems. But for motors of several kilowatts, the armature resistance is small, then an excessively high armature current is presented in starter condition at rated voltage [10]. Therefore, a starter resistor is connected in

series with the armature winding, causing losses and the need to include and disconnect starting resistors.
The DC motor configuration discussed in this paper has the distinction of having a nonlinear model, which makes its control a non-simple task [10]. Also, conventional controllers in some cases may not be sufficient due that the wide range of operating conditions. With this configuration is intended that the change in the mechanical load torque from zero to full, have a minimal impact on the rotor speed, also it changes over time. This analysis allows visualizing that the electrical DC motors have wide applicability, but is required to expand the existing studies with the inclusion of adaptive control techniques to cover highly demanding conditions, which the motor could be subjected.
The use of artificial neural networks, ANN, offers an attractive alternative for tracking speed of DC motors. The ANN are able to model and control on line nonlinear and non-stationary systems. Technique nature makes it robust, adaptive and optimal controller that can be used in independent or hybrid configurations with existing techniques. These features allow being an important option for practice engineers who are in face with the physical systems changing and high demands of connected loads. ANN are particularly attractive for controlling electric motors. At the same time, they consider the complexity of the physical system and provide a realistic control with less computational time for an effective and robust control in a wide operating range. B-spline neural networks, BSNN, are a particular class of neural networks that have exhibited important results in various physical systems [11]-[13]. This paper presents the design and performance with the qualities required for a real time application. The results exhibit its ability to adapt and how to face the change in load torque and motor conditions.
The proposed controller shows that only a previous off line training for some operating conditions is required and based on weights updating together with the base functions shape, it adapts to changes in the original design without losing its high performance. The result is an adaptive controller that enhances the motor operation even in different operation condition where the design was done.

DC Shunt Motor Model

There are various configurations of DC motors, where one in particular provides important operating characteristics where the rotor speed does not change appreciably as the load torque varies from zero to its nominal value [10]. Fig. 1 shows a connection diagram of the motor under study.

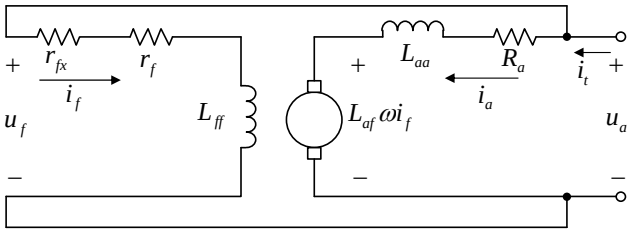


Figure 1. Equivalent circuit of a direct current shunt motor

We can see that the voltage source supplies both the field winding and the armature winding, therefore the total current is the sum of the two circulating currents. Considering a linear voltage-current relation for resistive and inductive elements, the DC motor model is obtained by,

$$L_f \frac{d}{dt} i_f = -R_f i_f + u_f \tag{1}$$

$$L_a \frac{d}{dt} i_a = -R_a i_a - L_{af} i_f \omega + u_a \tag{2}$$

The relation among electrical system and the mechanical system is determined by,

$$J\frac{d}{dt}\omega = -b\omega + L_{af}i_f i_a - \tau_L \tag{3}$$

where the electric torque is $\tau_e = L_{af}i_f i_a$. The mathematical model exhibits that it is a coupled nonlinear system. Therefore, conventional linear controllers guarantee the operation around an equilibrium point. The equilibrium points of the system can be determined by,

$$i_f = \frac{u_{in}}{R_f} \tag{4}$$

$$i_a = \frac{\left(bR_f^2 + R_f L_{af}\tau_L\right)u_{in}}{L_{af}^2 u_{in}^2 + bR_a R_f^2} \tag{5}$$

$$\omega = \frac{-R_a R_f^2 \tau_L + L_{af}R_f u_{in}^2}{L_{af}^2 u_{in}^2 + bR_a R_f^2} \tag{6}$$

where $u_{in} = u_a = u_f$. It can be seen that the equilibrium points depend on the voltage applied on terminals, the load torque and motor parameters. Therefore, the motor has multiple equilibrium points; it depends precisely on the operating condition.
The conventional PI controller used in this study is designed to have stable poles by $k_p = 10$ and $k_i = 100$.

Adaptive B-Spline Controller

Considering the nonlinear nature of DC motors described in section two and a linear controller, a problem arises with the regulation of the interest variables. If it is possible the control law must get a driver that is robust even tracking speed trajectories over time. In that sense, in this work an adaptive controller based on B-spline neural networks is proposed. Its design consists of two stages: first in defining the structure and characteristics of the inputs and the training rule; the second part is an on line learning where the ANN can determine changes in the reference signal, load torque and motor parameters.
In the off line training corresponding to the first part of the design, data of the interest variables, armature voltage and instantaneous rotor speed are used. With these data the neural network structure is validated, if the closed loop control meets expected performance proceeds to its on line operation, where continuous learning of new operational and/or parametric variations of the motor is done.
Among the objectives of the proposed controller we are looking to have a robust but simple design features and implementation on an experimental level. Time domain simulations and laboratory results demonstrate these aspects. In this work the diagram, Fig. 2, defines the proposed neural controller and the output is defined by [14],

$$y = \sum_{i=1}^{p} a_i w_i = \boldsymbol{a}^T \boldsymbol{w} \tag{7}$$

where w_i and a_i are the *i*-th weighting factor and the *i*-th basis function output, respectively; *p* is the number of weights of the neural network structure. The base function output changes with a nonlinear relationship of the input values, defined by the base function shape. For the proposed controller two monovariable functions of third order are used. The weight vector is updated by an instantaneous learning rule, defined by [14]

$$\Delta w(t) = \frac{\eta e_\omega(t)}{\|\boldsymbol{a}(t)\|_2^2}\boldsymbol{a}(t) \tag{8}$$

where η is the learning rate and e_ω is the error between the desired and actual rotor speed. The update of the weights depends on the base functions output and the learning rule; therefore, the neural network performance is not conditional on the reference type (constant or variable) or to the actual operating condition. The operational test conditions for defining the neural network structure are shown in Table 1.

Table 1. Operation Conditions for off line Training

Load Torque τ_L (Nm)	**Rotor Speed** ω (rad/sec)
0.5	30
0.3	100
1.25	50
1.1	70
0.4	120
0.75	45

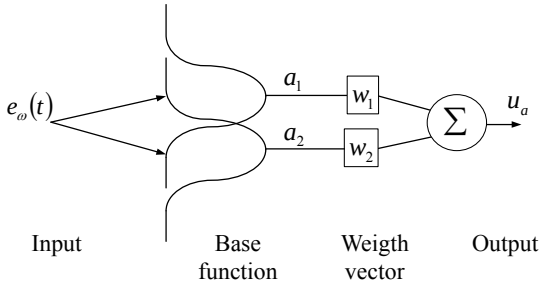


Figure 2. Proposed adaptive controller structure with the main elements

There are some applications of adaptive controllers based on B-spline neural networks where how to define the base functions, neural network structure and a training rule is explained [11]-[13]. Clearly simple structure facilitates the form of implementation and adaptation to different systems, in addition, the number of neurons, structure and shape of base functions have similarity in all these cases, therefore, and same structure is able to extend systems of different characteristics.

It is important to note that the implementation of these controllers prior knowledge of the operation and control system analysis is required. Finally, this particular BSNN structure makes them a very attractive structure that can be exploited in hybrid with other control strategies that can be linear, robust or adaptive configurations.

Test DC Shunt Motor

In order to evaluate the speed regulation, digital simulations and laboratory tests are accomplished using the DC motor arrangement described in Fig. 1, under different disturbances. The parameters used in the simulation studies are presented in Table 2; these values are approximate of the physical system. It is shown that the proposed with an initial off line training controller for tracking reference is sufficient to face the change on motor kind and the operation change.

To verify the proposed controller robustness the analysis is divided in two parts: a) simulation with BSNN controller; b) laboratory tests with BSNN controller. For simplicity the load in simulation study is represented proportional to velocity as,

$$\tau_L = B\omega \tag{9}$$

where the constant B is calculated by trial and error procedure, considering the current and voltage measured in laboratory tests in steady state with different load values.

Table 2. Motor Parameters of a DC Shunt Motor

Parameter	Value	Unit
R_a	7.5	Ω
R_f	469.75	Ω
L_a	55.3	mH
L_f	2.4123	H
L_{af}	2.2881	H
J	0.0013	Kg-m^2
b	1e-4	N.m.s

Simulation Results

Two scenarios are presented for adaptive controller analysis based on DC shunt motor, section 2. These results exemplify the adaptive controller evolution in possible real world scenarios, but strictly they are not necessary for laboratory experiments. DC motor is subjected to two tests scenarios: case 1 with tracking reference speed from 0 to 52 rad/sec considering a constant load torque 0.5 Nm is applied, after that a load torque change in $t = 5$ to 1.2 Nm, and finally is reduced to 0.1 Nm; case 2 a similar speed reference but with different magnitude is considered, also in $t = 5$ sec a variable load between 0.2 and 1 Nm is added.
Fig. 3 displays the DC motor speed controlled by the B-spline neural network. The transient and steady state condition exhibits a good performance for tracking a reference shape. Similar performance is maintained in the presence of load torque change, the modification considers a load constant for each adjustment. The reference signal is in blue color.

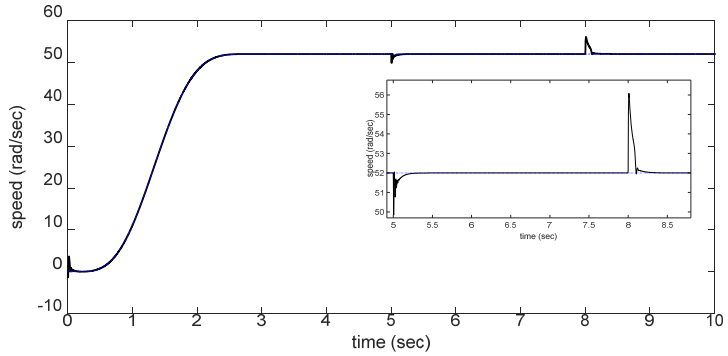


Figure 3. Rotor speed performance for reference tracking, case 1.

Fig. 4 presents the electrical torque required for operating conditions demanded in case 1. The armature current evolution is in accordance with DC motor performance, it is clear that the main overshoot is presented when the rotor speed change from 0 to 52 rad/sec with a constant load torque, Fig. 5. The oscillations are eliminated faster with the proposed adaptive controller. The settling time is near to 0.1 sec for BSNN.
The adaptive neural controller performance is guarantee by two main features: the off line training and the second the continuous learning in each sample time, and is reflected in the weight factor. This evolution is exhibited in Fig. 6 by one on the two weights of the B-spline neural network structure, the main change is presented when some of the motor system

configuration change, and obviously in steady state condition its value is maintained constant due that the error magnitude is near to zero. The learning rate is related to the velocity response of the weight factors, in this case it has a initial value equal to 5100 obtained in the previous training. This performance is expected while the continuous learning rule (8) is operating.

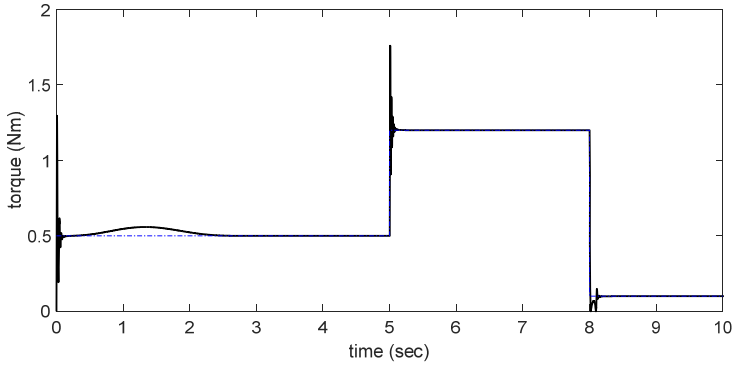


Figure 4. Electrical torque evolution, case 1.

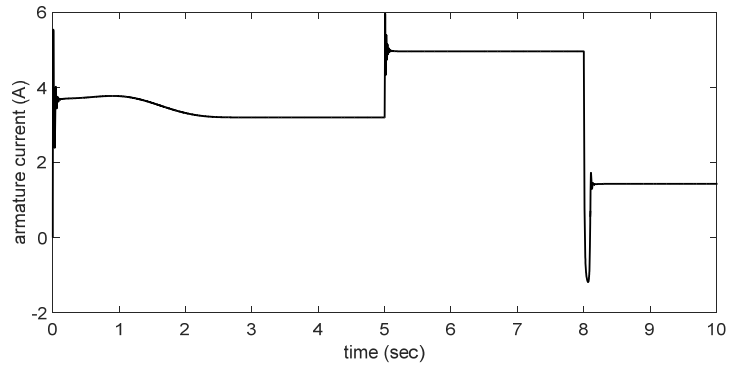


Figure 5. Armature current behavior, case 1.

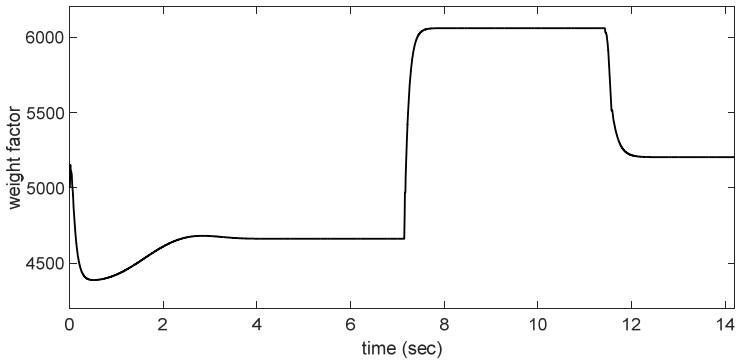


Figure 6. Weight factor evolution, case 1.

In Fig. 7 the rotor speed is presented, when the motor is exposed to operation conditions described in case 2. The BSNN has the ability of being updated to a new operating condition, improving its performance. The propose technique has a good performance because a speed tracking operation is demanded, additionally when the load torque is modified as a time function, the adaptive controller also achieves the reference speed requirements. This kind of performance is one of the main advantages of the adaptive controller when a design structure was well conditioned. The error between reference and actual speed is less than 1.5 rad/sec in transient period.

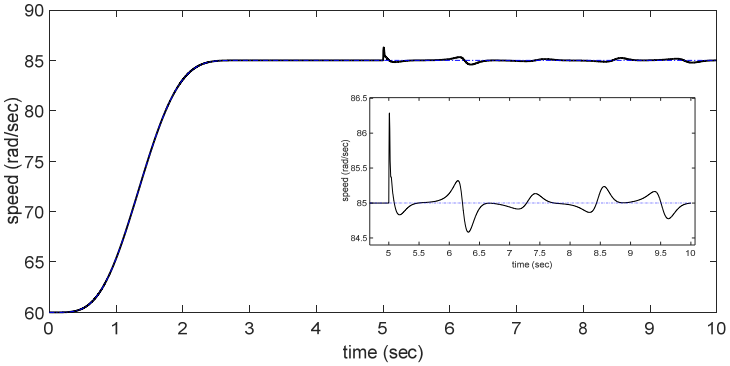


Figure 7. Rotor speed performance for reference tracking, case 2.

The input current for armature winding applied to the motor has an evolution as shows in Fig. 8. It is evident that the proposed controller confirms its faster response, therefore it could maintain similar behavior for both cases and different system requirements. In this study the input voltage is the only control variable, initially the motor is operating at 60 rad/sec with constant load torque.

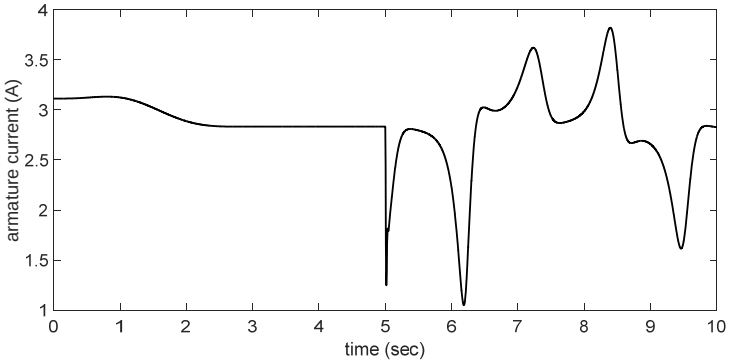


Figure 8. Armature current performance, case 2.

As a DC shunt motor, the field current is variable but the magnitude presented is fewer that armature current, Fig. 9. This feature allows complex load torque specifications for variable rotor speed. The behavior is a consequence of external load demands.

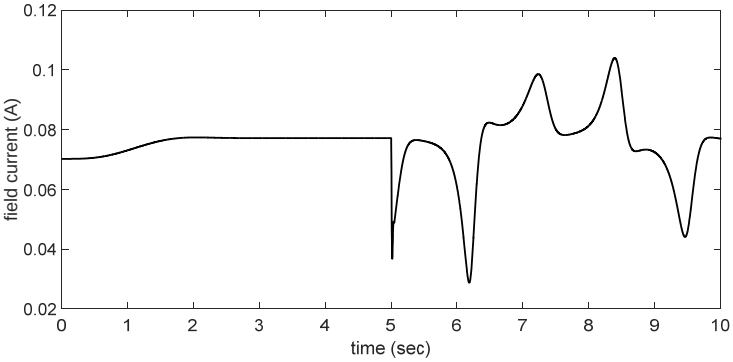


Figure 9. Field current evolution, case 2.

Measurement variables in laboratory test motor

The performance and applicability of the proposition are proved by hardware implementation on a laboratory DC motor. This strategy allows controlling appropriately the motor speed where

the load and set point is modified. The neural control is able to adapt by itself to different operating conditions, in other strategies turn out to be diminished in some situations, especially under different operating conditions for which its parameters have been tuned. Thus, the feedback signals to the BSNN are pertinent for a suitable control of the DC motor (shunt connected) velocity exhibiting a well performance for different operating points without modifications in control law. The applicability is demonstrated by laboratory results, some interest variables are presented in Fig. 10-13.

The DC motor to three scenarios was exposed. First case a, all variables are zero, after the speed reference is changed to 52 rad/sec, considering a constant load torque. An AC synchronous generator connected to the motor rotor is used as load. In this case the generator has an excited constant system, no electrical elements are wire to AC terminals.

The rotor speed and total current are showed in Fig. 10 and 11. The DC motor performance is in accordance with the simulations results. The proposed BSNN controller is able to regulate the speed with a desired behavior without knowledge of the system model and parameters. It is enough a collected data from the physical system in this case input voltage a rotor speed or and approximate mathematical model to know some possible characteristics about interest variables in transient and steady state condition.

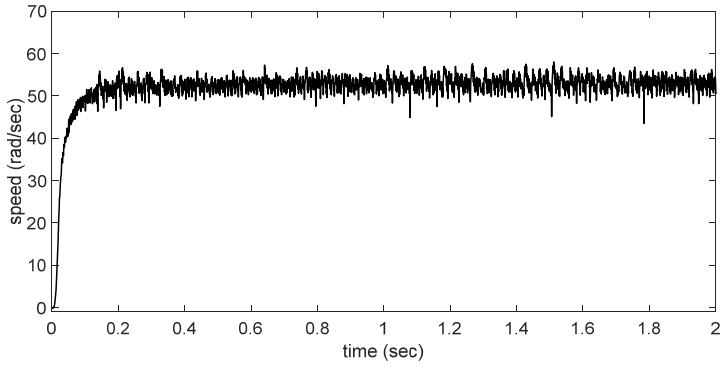


Figure 10. Rotor speed performance for reference tracking, case a.

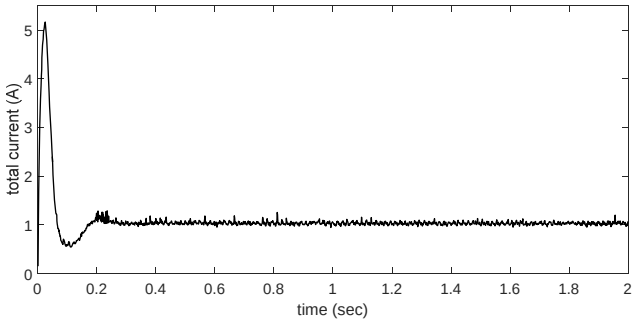


Figure 11. Total current performance, case a.

The second case b exhibits the reference tracking performance when the value is changed from 80 to 60 rad/sec. The controller has a good evolution when the speed reference is diminished, Fig. 12. The laboratory test where develop with operation conditions similar to simulation test. Discrepancies with simulations are due to a rough estimation of parameters. The comparison is only for demonstrating the behavior of the proposed adaptive controller.

The final case c includes a load torque with different values, first the rotor speed achieves 80 rad/sec; after that at t = 3 sec an AC synchronous generator is included. At t = 4.5, 6 and 8 seconds a three phase resistive load is inserted in the generator terminals, the resistive load is increased in each time. In the last part at t = 11.7 sec the resistive load is disconnected from the

generator terminals. The total current exemplify the system performance, all variables attain a behavior with similar features with all presented study cases.

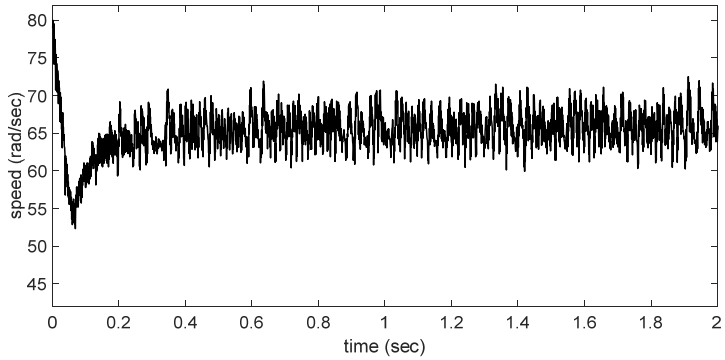


Figure 12. Rotor speed performance for reference tracking, case b.

The adaptability of the proposed controller has been presented by prior off line design. A mathematical model with approximate motor parameters was used, and the design was performed by data collected by the model and laboratory test system. The performance of the proposed adaptive controller is evaluated by simulation model analysis and results obtained experimentally. The observed performance validates the initial design methodology.

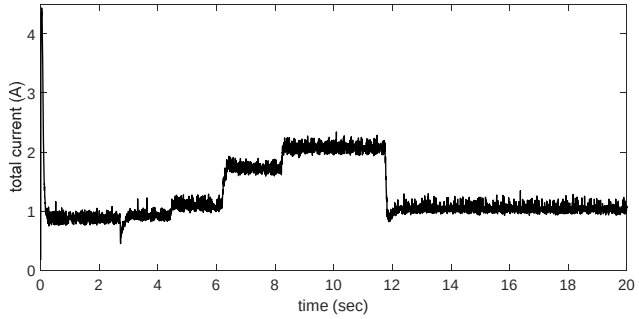


Figure 13. Total current behavior, case c.

Conclusions

It has been shown that with a prior off line neural network design we achieve a robust and adaptive controller, the instantaneous learning rule permits that the controller adapts by itself in each demanded operation condition. The same behavior is exhibited with real world operation conditions for the motor. For this proposition is not need the use of mathematical model furthermore the dependency of motor parameters is omitted. The main feature is an adaptive nature and an easy way to implement in physical motor. The hardware implementation validates the proposed design; several operation conditions are taking into account.

Acknowledgment

This work has been supported in part by CONACYT project under grant 266333.

References

[1] Jian Sun, Yi Chai, Chunxiao Su, Zhiqin Zhu, Xianke Luo, "BLDC motor speed control system fault diagnosis based on LRGF neural network and adaptive lifting scheme," Applied Soft Computing, vol. 14, pp. 609-622, 2014.

[2] Hemza Mekki, Omar Benzineb, Djamel Boukhetala, Mohamed Tadjine, Mohamed Benbouzid, "Sliding mode based fault detection reconstruction and fault tolerant control scheme for motor systems," ISA Transactions, vol. 57, pp. 340-351, 2015.
[3] Chun-Lai Li, Wen Li, Fu-Dong Li, "Chaos induced in Brushless DC Motor via current time-delayed feedback," Optik, vol. 125, pp. 6589-6593, 2014.
[4] Vincent Léchappé, Oscar Salas, Jesús de León, Franck Plestan, Emmanuel Moulay, Alain Glumineau, "Predictive control of disturbed systems with input delay: experimental validation on a DC motor," IFAC, vol. 48, pp. 292-297, 2015.
[5] Rui Bai, "Neural network control-based adaptive design for a class of DC motor Systems with the full state constraints," Neurocomputing, vol. 168, pp. 65-69, 2015.
[6] Ahmad M. Zaki, Mohammad El-Bardini, F.A.S. Soliman, Mohammed Mabrouk Sharaf, "Embedded two level direct adaptive fuzzy controller for DC motor speed control," Ain Shams Engineering Journal, in press, 2015.
[7] Rohit G. Kanojiya, P. M. Meshram, "Optimal tuning of PI controller for speed control of DC motor drive using particle swarm optimization," IEEE International Conference on Advances in Power Conversion and Energy Technologies (APCET), 2012.
[8] S. Jayaprakash, Comparison of solar based closed loop DC-DC converter using PID and Fuzzy Logic Control for Shunt motor drive," IEEE, 2014.
[9] Jia-Jun Wang, "A common sharing method for current and flux-linkage controlof switched reluctance motor," Electric Power Systems Research, vol. 131, pp. 19-30, 2016.
[10] Paul C. Krause, Oleg Wasynczuk, Scott D. Sudhoff, "Analysis of Electric Machinery and Drive Systems," 2nd Edition, Wiley-IEEE Press, 2002.
[11] R. Tapia-Olvera, O. Aguilar-Mejía, H. Minor-Popocatl, C. Santiago-Tepantlan, "Power System Stabilizer and Secondary Voltage Regulator Tuning for Multimachine Power Systems," Electric Power Components and Systems, vol. 40, No. 16, pp. 1751-1767, 2012.
[12] Omar Aguilar-Mejía, Rubén Tapia-Olvera, Antonio Valderrabano-González, Iván Rivas Cambero, "Adaptive neural network control of chaos in permanent magnet synchronous motor," Intelligent Automation & Soft Computing, pp. 1-6, 2015.
[13] O. Aguilar, R. Tapia, A. Valderrabano, H. Minor, "Design and Performance Comparison of PI and Adaptive Current Controllers for a WECS," IEEE Latin America Transactions, vol. 13, No. 5, pp. 1361-1368, 2015
[14] Brown, and C. Harris, Neurofuzzy Adaptive Modelling and Control, Prentice Hall International, 1994.

Active Vibration Control of a Vehicle Suspension System Based on Signal Differentiation

†F. Beltran-Carbajal[1], A. Favela-Contreras[2], I. Lopez-Garcia[1], R. Tapia-Olvera[3], Z. Damian-Noriega[1] and G. Alvarez-Miranda[1]

[1]Departamento de Energía, Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Mexico City, Mexico
[2]Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, C.P. 64849, Monterrey N.L., Mexico
[3]Department of Engineering, Universidad Politécnica de Tulancingo, Hidalgo, Mexico
*Presenting author: gdam@correo.azc.uam.mx
†Corresponding author: fbeltran@azc.uam.mx

Abstract

Real-time estimation and differentiation of signals are common tasks in diverse applications of active vibration control. In this paper, an asymptotic approach for signal differentiation is applied in an active vehicle suspension system. The synthesis of the differentiation approach evades the use of a mathematical model of the suspension system. Estimation of unknown exogenous disturbances due to irregular road surfaces are also estimated. Estimates of time derivatives of the output variable and disturbances are then used for the implementation of an active vibration control scheme. Some numerical results are provided to show the effectiveness of the real-time estimation of the unavailable signals as well as a reasonable vibration attenuation level on a linear quarter-vehicle active suspension system.

Keywords: Active Vibration Control, Vehicle Suspension System, Differential Flatness, Signal Differentiation, Disturbance Rejection.

Introduction

Real-time estimation of parameters and signals is an active research subject in vibration control. Several approaches about parameter and signal estimation for mass-spring-damper systems, vibration absorbers and rotor-bering systems have been proposed in [1, 2, 3, 4, 5, 6]. Time derivatives of some system variables (e.g., velocity and acceleration) could be also required for implementation of active vibration control schemes. In fact, error signal differentiation is demanded in classical Proportional-Integral-Derivative (PID) control which is applied in many industrial engineering systems. Moreover, availability of signal derivatives can be used to reconstruct disturbance forces affecting a vibration mechanical system. State vector estimation is commonly based on asymptotic observers designed for specific dynamical systems. In practice, differentiation of signals is also performed by real-time numerical computations from samplings of the available output signals. Nevertheless, numerical differentiation could deteriorate the efficiency and robustness of system identification or control when measurements are corrupted by noise.

Recently, an asymptotic differentiation approach of signals for angular acceleration estimation for DC motors has been proposed in [7]. This paper describes the application of this signal differentiation approach to approximately estimate time derivatives and disturbances in vibrating mechanical systems. Signal differentiation is applied to control an active vehicle suspension system as well. The synthesis of the differentiation approach evades the use of a mathematical model of the suspension system. Hence, the differentiation approach can be employed in vibration mechanical systems where time derivative of some signal is required. It is shown

that unknown exogenous disturbances due to irregular road surfaces can be algebraically reconstructed from estimates of time derivatives. Estimates of time derivatives of the output variable and disturbances are then used for the implementation of an active vibration control scheme. Some numerical results are provided to show the effectiveness of the real-time estimation of the unavailable signals. A reasonable level of forced vibration attenuation on an active linear quarter-vehicle suspension system is also verified.

1 Mathematical Model of a Vehicle Suspension System

Firstly, consider the mathematical model (1) of the active quarter-vehicle suspension system schematically shown in Fig.1:

$$\begin{aligned} m_s\ddot{z}_s + c_s(\dot{z}_s - \dot{z}_u) + k_s(z_s - z_u) &= u \\ m_u\ddot{z}_u + k_t(z_u - z_r) - c_s(\dot{z}_s - \dot{z}_u) - k_s(z_s - z_u) &= -u \end{aligned} \tag{1}$$

where the sprung mass m_s represents the mass of the car-body part, the unsprung mass m_u denotes the mass of the assembly of the axle and wheel, c_s is the damper coefficient of suspension, k_s and k_t are the spring coefficients of the suspension and tire, respectively. The generalized coordinates are the displacements of both masses z_s and z_u, z_r is the terrain disturbance and u is the control force input provided by some (electromagnetic or hydraulic) actuator.

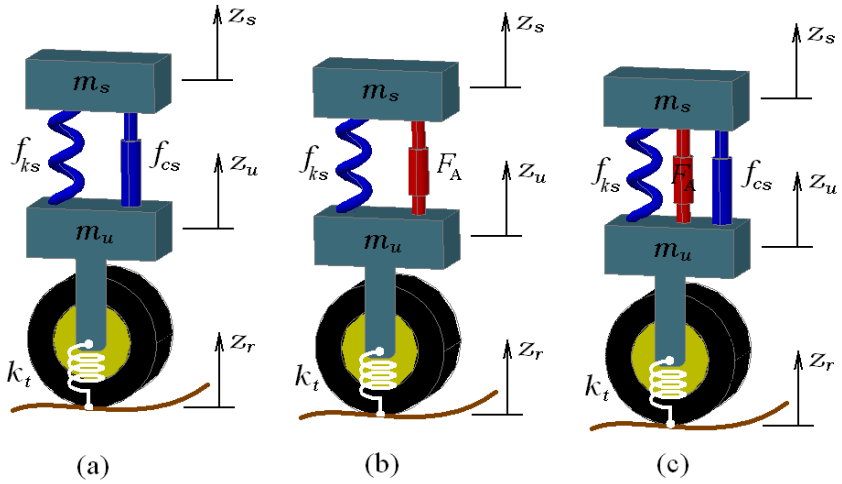


Figure 1: Quarter-vehicle suspension system: (a) passive suspension system, (b) active suspension system with an electromagnetic actuator, (c) active suspension system with a hydraulic actuator.

Defining the state variables as $x_1 = z_s$, $x_2 = \dot{z}_s$, $x_3 = z_u$ and $x_4 = \dot{z}_u$, mathematical model (1) adopts the state-space description

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{k_s}{m_s}x_1 - \frac{c_s}{m_s}x_2 + \frac{k_s}{m_s}x_3 + \frac{c_s}{m_s}x_4 + \frac{1}{m_s}u \\ \dot{x}_3 &= x_4 \\ \dot{x}_4 &= \frac{k_s}{m_u}x_1 + \frac{c_s}{m_u}x_2 - \frac{k_s + k_t}{m_u}x_3 - \frac{c_s}{m_u}x_4 - \frac{1}{m_u}u + \frac{k_t}{m_u}z_r \end{aligned} \tag{2}$$

The active suspension system (2) is a differentially flat system, where a flat output y is given by [8, 9]:

$$y = m_s x_1 + m_u x_3 \tag{3}$$

Therefore, state and control variables can be expressed in terms of the flat output y and a finite number of its time derivatives. Indeed, from y and its time derivatives up to fourth order:

$$\begin{aligned} \dot{y} &= m_s x_2 + m_u x_4 \\ \ddot{y} &= k_t \left(z_r - x_3\right) \\ y^{(3)} &= k_t \left(\dot{z}_r - x_4\right) \\ y^{(4)} &= \frac{1}{m_u} u + \frac{k_t}{m_u} x_3 - \frac{1}{m_u} \left(\mathcal{F}_{sc} + \mathcal{F}_{sk}\right) - \frac{k_t}{m_u} z_r + k_t \ddot{z}_r \end{aligned} \tag{4}$$

with

$$\begin{aligned} \mathcal{F}_{sk} &= k_s \left(x_1 - x_3\right) \\ \mathcal{F}_{sc} &= c_s \left(x_2 - x_4\right) \end{aligned} \tag{5}$$

the differential parameterization results as follows

$$\begin{aligned} x_1 &= \frac{1}{m_s} y + \frac{m_u}{k_t m_s} \ddot{y} - \frac{m_u}{m_s} z_r \\ x_2 &= \frac{1}{m_s} \dot{y} + \frac{m_u}{k_t m_s} y^{(3)} - \frac{m_u}{m_s} \dot{z}_r \\ x_3 &= -\frac{1}{k_t} \ddot{y} + z_r \\ x_4 &= -\frac{1}{k_t} y^{(3)} + \dot{z}_r \\ u &= \frac{1}{b} \left(a_0 y + a_1 \dot{y} + a_2 \ddot{y} + a_3 y^{(3)} + y^{(4)} - \xi\right) \end{aligned} \tag{6}$$

with

$$\begin{aligned} a_0 &= \frac{k_s k_t}{m_s m_u}, \quad a_1 = \frac{c_s k_t}{m_s m_u} \\ a_2 &= \frac{k_s}{m_s} + \frac{k_s + k_t}{m_u}, \quad a_3 = \frac{c_s}{m_s} + \frac{c_s}{m_u} \\ b &= \frac{k_t}{m_u} \end{aligned} \tag{7}$$

and

$$\xi\left(t\right) = \left(\frac{k_t}{m_s} + \frac{k_t}{m_u}\right) k_s z_r + \left(\frac{k_t}{m_s} + \frac{k_t}{m_u}\right) c_s \dot{z}_r + k_t \ddot{z}_r \tag{8}$$

Thus from (6) the flat output is governed by the perturbed input-output differential equation

$$y^{(4)} + a_3 y^{(3)} + a_2 \ddot{y} + a_1 \dot{y} + a_0 y = b u + \xi \tag{9}$$

Hence, the following active vibration control scheme based on differential flatness can be di-

rectly synthesised:

$$u = \frac{1}{b}(\upsilon + a_3 y^{(3)} + a_2 \ddot{y} + a_1 \dot{y} + a_0 y - \xi) \tag{10}$$

with

$$\upsilon = -\alpha_3 y^{(3)} - \alpha_2 \ddot{y} - \alpha_1 \dot{y} - \alpha_0 y$$

Nevertheless, implementation of the control law (10) needs measurements or estimates of some time derivatives of the flat output variable y. In addition, information of the profile of irregular road surfaces z_r could be also demanded.

On the other hand, note that from (4) the flat output y and its derivatives up to third order can be computed from state variables and disturbance z_r. Otherwise, time derivatives of the flat output can be also estimated directly. Moreover, the disturbance z_r can be calculated by

$$z_r = \frac{1}{k_t}\ddot{y} + x_3 = \frac{1}{k_t}\left(m_s \ddot{x}_1 + m_u \ddot{x}_3\right) + x_3 \tag{11}$$

Thus, in the next section it is described a signal differentiation approach to get approximate derivatives for some stable dynamical system [7].

2 A Signal Differentiation Approach

The synthesis of the signal differentiation scheme with respect to time is based on the local approximation of some bounded signal $\mathscr{Y}$ by a family of Taylor polynomials of forth degree as

$$\mathscr{Y}(t) \approx \sum_{i=0}^{4} q_i t^i \tag{12}$$

where coefficients q_i are assumed to be unknown.

Therefore, the signal $\mathscr{Y}$ can be locally reconstructed by the dynamical system

$$\begin{aligned}
\dot{\mathscr{y}}_f &= \mathscr{y}_1 \\
\dot{\mathscr{y}}_1 &= \mathscr{y}_2 \\
\dot{\mathscr{y}}_2 &= \mathscr{y}_3 \\
\dot{\mathscr{y}}_3 &= \mathscr{y}_4 \\
\dot{\mathscr{y}}_4 &= \mathscr{y}_5 \\
\dot{\mathscr{y}}_5 &= \mathscr{F}
\end{aligned} \tag{13}$$

where $\mathscr{y}_1 = \mathscr{Y}$, $\mathscr{y}_2 = \dot{\mathscr{Y}}$, $\cdots$, $\mathscr{y}_5 = \mathscr{Y}^{(4)}$, $\mathscr{y}_f = \int_0^t \mathscr{Y}\, dt$, and $\mathscr{F}$ is considered as an unknown bounded perturbation signal including the influence of high frequency noise and small residual terms of the truncated Taylor polynomial expansion (12) (see [7]). Moreover, we have assumed that the time derivatives up to fifth order of $\mathscr{Y}$ are uniformly absolutely bounded.

Hence, from (13) we propose the following state observer for asymptotic estimation of some

time derivatives of the signal $\mathscr{Y}$:

$$\begin{aligned}
\widehat{\dot{\mathscr{Y}}}_f &= \widehat{\mathscr{Y}_1} + \beta_5 \left(\mathscr{Y}_f - \widehat{\mathscr{Y}_f} \right) \\
\widehat{\dot{\mathscr{Y}}}_1 &= \widehat{\mathscr{Y}_2} + \beta_4 \left(\mathscr{Y}_f - \widehat{y}_f \right) \\
\widehat{\dot{\mathscr{Y}}}_2 &= \widehat{\mathscr{Y}_3} + \beta_3 \left(\mathscr{Y}_f - \widehat{\mathscr{Y}_f} \right) \\
\widehat{\dot{\mathscr{Y}}}_3 &= \widehat{\mathscr{Y}_4} + \beta_2 \left(\mathscr{Y}_f - \widehat{\mathscr{Y}_f} \right) \\
\widehat{\dot{\mathscr{Y}}}_4 &= \widehat{\mathscr{Y}_5} + \beta_1 \left(\mathscr{Y}_f - \widehat{\mathscr{Y}_f} \right) \\
\widehat{\dot{\mathscr{Y}}}_5 &= \beta_0 \left(\mathscr{Y}_f - \widehat{\mathscr{Y}_f} \right)
\end{aligned} \tag{14}$$

which only uses information of the filtered output signal $\mathscr{Y}_f$. Here, we use the notation $\widehat{(\cdot)}$ for the estimated signals.

Then, the estimation error dynamics is governed by

$$\begin{aligned}
\dot{e}_f &= e_1 - \beta_5 e_f \\
\dot{e}_1 &= e_2 - \beta_4 e_f \\
\dot{e}_2 &= e_3 - \beta_3 e_f \\
\dot{e}_3 &= e_4 - \beta_2 e_f \\
\dot{e}_4 &= e_5 - \beta_1 e_f \\
\dot{e}_5 &= -\beta_0 e_f
\end{aligned} \tag{15}$$

which is completely independent of any coefficients q_i of the Taylor polynomial expansion of the output signal $\mathscr{Y}$. Here, $e_i = \mathscr{Y}_i - \widehat{\mathscr{Y}_i}$, $i = 1, 2, \ldots, 5$, $e_f = \mathscr{Y} - \widehat{\mathscr{Y}_f}$. Notice that, estimator gains should be properly selected in order to have a stable characteristic polynomial for the observer-based closed-loop system dynamics. Additionally, the estimation dynamics should be sufficiently fast to get estimates opportunely to be used by the active vibration control scheme. Note that, it is widely known that delays in measurements or estimations could become unstable a dynamical systems. Better estimates can be obtained by employing a Taylor polynomial model of higher order.

3 Simulation results

Effectiveness of the differentiation approach for approximate estimation of time derivatives and road disturbance signals required for implementation of the active vibration control scheme (10) for an active linear quarter-vehicle suspension system was verified by some preliminary computer simulations. The vehicle suspension system is characterized by the set of parameters described in Table 1 [10].

In Fig. 2 is shown the unknown exogenous disturbance excitation due to irregular road surfaces which is described by [11]:

$$z_r(t) = \begin{cases} f_1(t) + f(t) & \text{for } t \in [3.5, 5) \\ f_2(t) + f(t) & \text{for } t \in [5, 6.5) \\ f_3(t) + f(t) & \text{for } t \in [8.5, 10) \\ f_3(t) + f(t) & \text{for } t \in [10, 11.5) \\ f(t) & \text{else} \end{cases} \tag{16}$$

Table 1: Parameters of the vehicle suspension system.

Parameter	Value
Sprung mass m_s	216.75 kg
Unsprung mass m_u	28.85 kg
Spring stiffness k_s	21700 N/m
Damping constant c_s	1200 Ns/m
Tire stiffness k_t	184000 N/m

with

$$\begin{aligned} f_1(t) &= -0.0592\,(t-3.5)^3 + 0.1332\,(t-3.5)^2 \\ f_2(t) &= 0.0592\,(t-6.5)^3 + 0.1332\,(t-6.5)^2 \\ f_3(t) &= 0.0592\,(t-8.5)^3 - 0.1332\,(t-8.5)^2 \\ f_3(t) &= -0.0592\,(t-11.5)^3 - 0.1332\,(t-11.5)^2 \\ f(t) &= 0.002\sin(2\pi t) + 0.002\sin(7.5\pi t) \end{aligned}$$

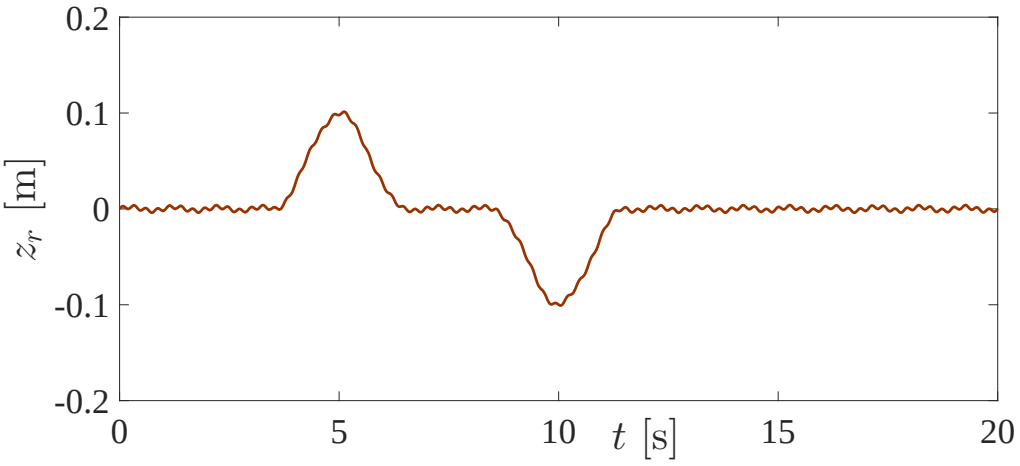


Figure 2: Irregular profile of the road surface.

Fig. 4 describes a reasonable attenuation level of vibrations induced by irregular road surfaces (16) using the active vibration control scheme based on high-gain signal differentiation. To get a fast signal estimation the characteristic polynomial of the estimation error dynamics was set as

$$P_O(s) = (s^2 + 2\zeta_o\omega_o s + \omega_o^2)^3 \tag{17}$$

with $\omega_o = 2000$ rad/s and $\zeta_o = 5$.

Acceptable approximate estimation of the disturbance signal z_r is depicted in Fig. 4. On the other hand, the active control force applied to the vehicle suspension system is illustrated in Fig. 5. The control gains were chosen to have the closed loop characteristic polynomial

$$P_c(s) = (s^2 + 2\zeta_c\omega_c s + \omega_c^2)^2 \tag{18}$$

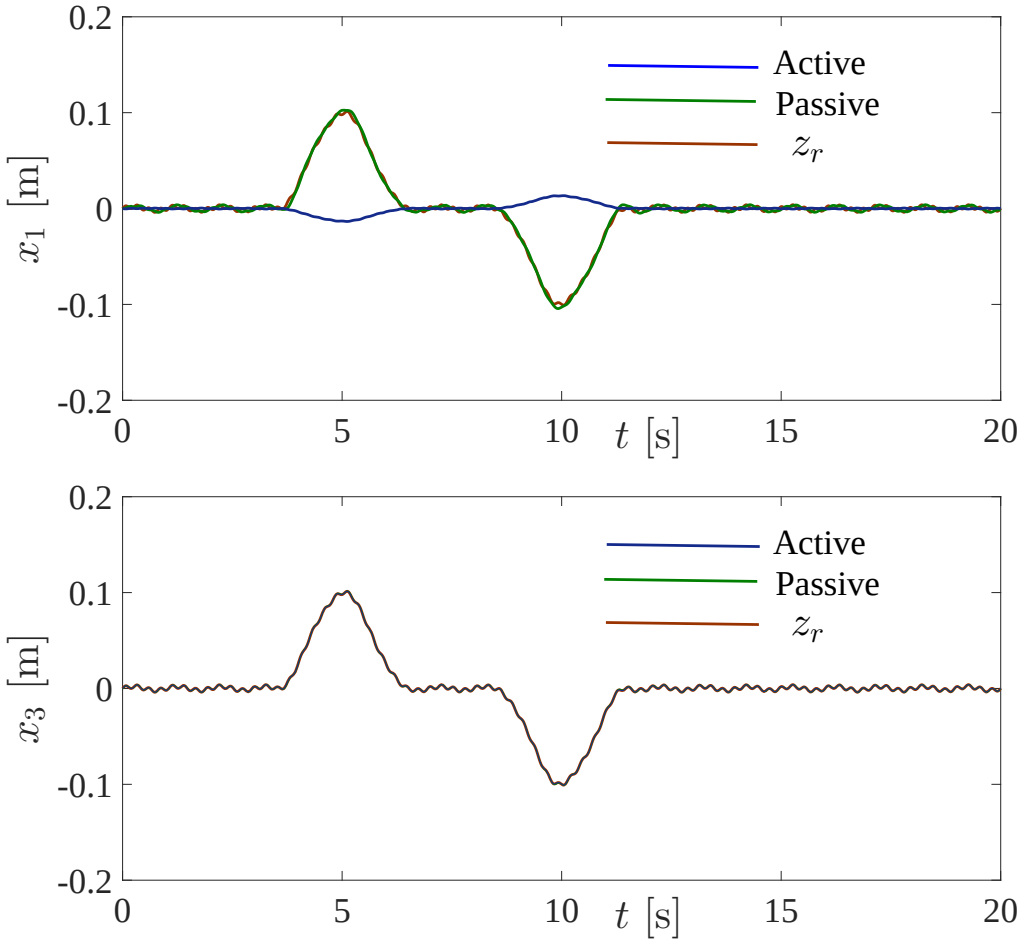


Figure 3: Position responses of sprung and unsprung masses.

with $\omega_c = 10$ rad/s and $\zeta_c = 0.7071$.

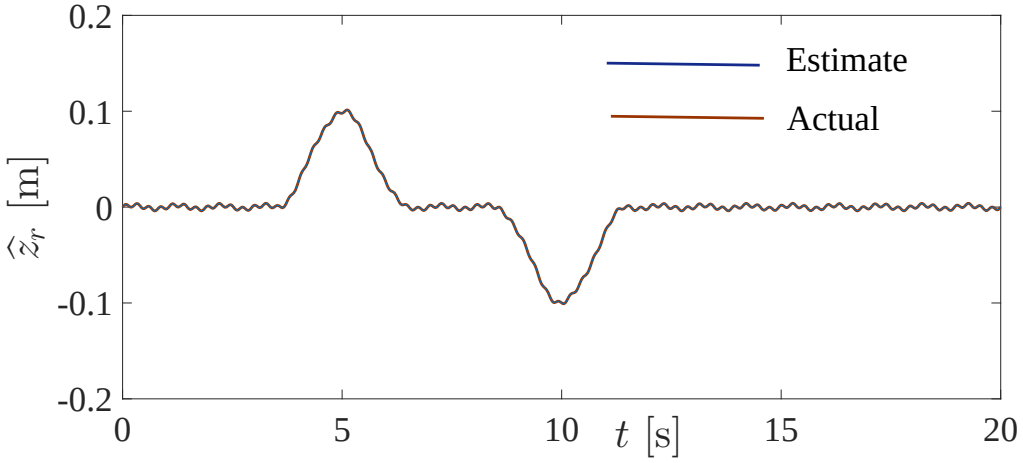


Figure 4: High-gain fast estimation of the irregular profile of the road surface.

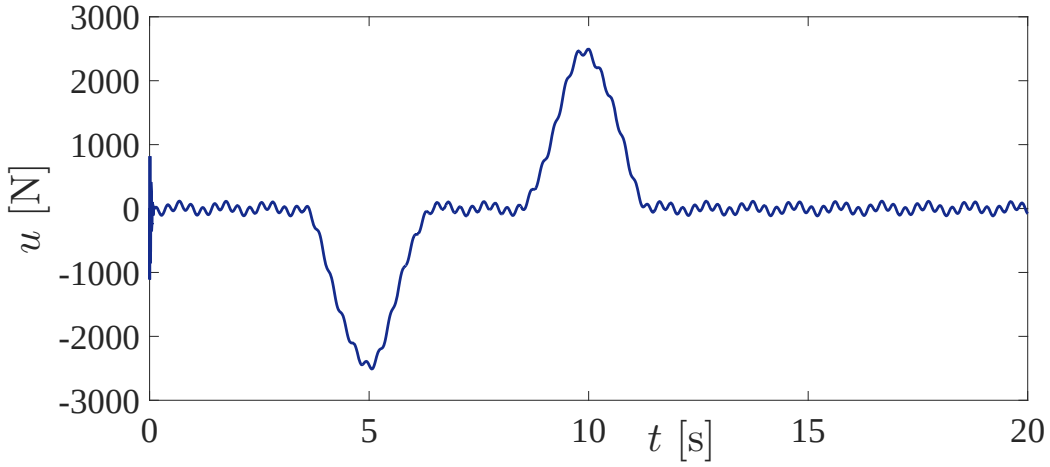


Figure 5: Active control force applied to the vehicle suspension system.

4 Conclusions

The application of a signal differentiation approach with respect to time has been described for an active linear quarter-vehicle suspension system. Certain signal derivatives and unknown exogenous disturbances due to irregular road surfaces were estimated. Approximate estimates were then used for the implementation of an active vibration control scheme based on differential flatness. Numerical results illustrate an acceptable estimation of the disturbance signal due to irregular road surfaces. It was also shown that the active vibration control scheme archives a reasonable vibration attenuation level on a linear quarter-vehicle active suspension system when the estimation error dynamics is sufficiently fast with respect to the closed loop vehicle suspension system and disturbances. Thus, the effectiveness of the on-line signal estimation algorithm without employing some specific mathematical model of the controlled dynamical system requires fast velocities for signal processing and high estimation gains. Actually, high speed and precise sensors, DSP boards, and software with high computational performance operating at high sampling rates are now available. Hence, the described differentiation approach represents a good choice to approximately estimate disturbances and time derivatives for scenarios where evasion of the use of some mathematical model for the system or a minimal number of sensors are desired.

References

[1] Beltran-Carbajal, F. and Silva-Navarro, G. On the Algebraic Parameter Identification of Vibrating Mechanical Systems, *International Journal of Mechanical Sciences*, **92**, 178-186, (2015).

[2] Arias-Montiel, M., Beltran-Carbajal, F. and Silva-Navarro, G. On-line algebraic identification of eccentricity parameters in active rotor-bearing systems, *International Journal of Mechanical Sciences*, **85**, 152-159, (2014).

[3] Beltran-Carbajal, F. and Silva-Navarro, G. Active Vibration Control in Duffing Mechanical Systems Using Dynamic Vibration Absorbers, *Journal of Sound and Vibration*, **333**, 319-330, (2014).

[4] Beltran-Carbajal, F. and Silva-Navarro, G. and Trujillo-Franco, L.G., Evaluation of on-line algebraic modal parameter identification methods, In R. Allemang (ed.), *Topics in Modal Analysis II*, **15** (8), Springer, NY, pp. 145-152, (2014).

[5] Beltran-Carbajal, F. and Silva-Navarro, G., Adaptive-like vibration control in mechanical systems with unknown parameters and signals, *Asian Journal of Control*, **15** (6), 1613-1626 (2013).

[6] Beltran-Carbajal, F., Silva-Navarro, G. and Arias-Montiel, M. Active Unbalance Control of Rotor Systems Using On-line Algebraic Identification Methods, *Asian Journal of Control*, **15** (6), 1613-1626, (2013).

[7] Beltran-Carbajal, F., Valderrabano-Gonzalez, A., Rosas-Caro, J.C. and Favela-Contreras, A. An asymptotic differentiation approach of signals in velocity tracking control of DC motors, *Electric Power Systems Research*, **122**, 218-223, (2015).

[8] Chavez-Conde, E., Beltran-Carbajal, F., Valderrabano-Gonzalez, A. and Favela-Contreras, A. Active vibration control of vehicle suspension systems using sliding modes, differential flatness and generalized proportional-integral control, *Rev. Fac. Ing. Univ. Antioquia*, **61**, 104-113, (2011).

[9] Beltran-Carbajal, F., Chavez-Conde, E., Favela-Contreras, A. and Chavez-Bracamontes, R. Active nonlinear vehicle suspension control based on real-time estimation of perturbation signals, *Proceedings of the 2011 IEEE International Conference on Industrial Technology (ICIT)*, Auburn, AL, United States, 14–16 March, (2011).

[10] Tahboub, K. A. Active nonlinear vehicle-suspension variable-gain control, *Proceedings of the 13th Mediterranean Conference on Control and Automation*, Limassol, Cyprus, 27–29 June, (2005).

[11] Chen, P. and Huang, A. Adaptive sliding control of non-autonomous active suspension systems with time-varying loadings, *Journal of Sound and Vibration*, **282**, 1119–1135, (2005).

Closed Loop Algebraic Parametric Identification of a DC Shunt Motor

†F. Beltran-Carbajal[1], R. Tapia-Olvera[2], A. Favela-Contreras[3], I. Lopez-Garcia[1], Z. Damian-Noriega[1] and G. Alvarez-Miranda[1]

[1]Departamento de Energía, Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Mexico City, Mexico
[2]Department of Engineering, Universidad Politécnica de Tulancingo, Hidalgo, Mexico
[3]Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, C.P. 64849, Monterrey N.L., Mexico
*Presenting author: gdam@correo.azc.uam.mx
†Corresponding author: fbeltran@azc.uam.mx

Abstract

Real-time system identification for electric machinery is an active research topic. Diagnostic or adaptive control tasks could demand the knowledge of the energy conversion system parameters and, possibly, load mechanical torque as well. In this paper, an on-line identification scheme is proposed for estimation of all parameters and load torque for an efficiently controlled DC shunt motor. The parameters of the nonlinear electromechanical system and load torque are estimated algebraically and quickly. A PI control law is also described for regulation and tracking tasks on this nonlinear energy conversion system. Some numerical simulation results are provided to show the effective closed-loop estimation of all system parameters and mechanical torque.

Keywords: DC shunt motor, System identification, Algebraic identification, PI control.

Introduction

In recent decades, several applications of electric motors can be found at industry and homes. In fact, motor-driven equipment is approximately 60% of manufacturing final electricity worldwide [1]. Among them are direct current (DC) motors and, particularly, shunt connection allows advantages over those well-known permanent magnet motors. This configuration is commonly applied for operation conditions of variable load torque with a reduced effect on the rotor speed. Further, it does not handle high currents as series DC motors, therefore, it is a useful configuration that allows starting and nominal torques with relatively low currents in transient and steady state operation. Several control schemes for DC electrical machines have been reported in the literature (see, e.g., [2, 3, 4]). However, most of them are focused on permanent magnet or separately excited, limiting the load torque operation mainly in starting and tracking variable speed condition. In addition, a priori knowledge of the motor parameters and, possibly, load torque are required to get an efficient control performance under variable velocity operation scenarios. Thus, parameter identification techniques have been commonly employed [5, 6, 7]. Nevertheless, this requirement is a difficult aspect to guarantee because a DC shunt motor is a nonlinear dynamical system with parameters changing in time. The parameter identification area for electrical machines is very extensive, where important aspects of implementation are searched. In general, closed loop parametric identification should be performed on-line and fast to be used simultaneously with some control technique applied to the motor [8].

There are numerous research works that propose different parameter identification techniques [9, 10, 11]. Recent contributions based on neural networks, fuzzy logic, Kalman filter, complementary, and using optimization procedures such as genetic algorithms, ant colony, particle swarm have been proposed for motor parameters estimation [10, 11]. They are suitable for

on-line or off-line application depending mostly on high computational requirements. One of the drawbacks of some of the proposed strategies is that correct parameters estimation is not guaranteed because a nonlinear and coupled nature of the interest variables are not included. Additionally, identification schemes have a weak performance in some operation conditions due to in many cases the complex behavior presented in electrical motors are not considered in the design stage. In general, there are some considerations that must be taken into account in parameters estimation as continuous variations of the load torque, the impact of the electronic controllers in transient response and noise included in measured variables, and tracking speed. Such schemes must meet high precision in face to continuous motor changes with low computational cost for implementation in real time platform. On the other hand, recent algebraic parametric identification methods have shown an excellent and fast performance [12]. In fact, algebraic identification have been successfully applied to estimate parameters and signals in flexible mechanical systems [13, 14, 15]. Numerical and experimental results have confirmed that algebraic identification represents an very good choice for the synthesis of on-line parameter estimators.

In this paper, an on-line identification scheme is proposed for estimation of all parameters and load torque for an efficiently controlled DC shunt motor. The parameters of the nonlinear electromechanical system and load torque are estimated algebraically and quickly. A PI control law is also described for regulation and tracking tasks on this nonlinear energy conversion system. Some numerical simulation results are provided to show the effective closed-loop estimation of all system parameters and mechanical torque.

1 Mathematical Model of a Controlled DC Shunt Motor

Consider the nonlinear mathematical model of a DC motor with field and armature circuits connected in parallel

$$\begin{aligned} L_f \frac{d}{dt} i_f &= -R_f i_f + u \\ L_a \frac{d}{dt} i_a &= -R_a i_a - L_{af} i_f \omega + u \\ J \frac{d}{dt} \omega &= -b\omega + L_{af} i_f i_a - \tau_L \\ y &= \omega \end{aligned} \tag{1}$$

where the positive parameters of the field circuit are the inductance L_f and resistance R_f. L_a and R_a are the inductance and resistance of the armature circuit, respectively, and L_{af} is the mutual inductance. J and b are the inertia moment and viscous damping of the mechanical subsystem. Here, u is the voltage control input, $y = \omega$ is the controlled output angular velocity and τ_L is the constant load torque. The field and armature current signals are respectively i_f and i_a.

From basic control fundamentals, one can very that the dynamics of the output angular velocity $y = \omega$ around some desired equilibrium operation state $(\omega^e, i_a^e, i_f^e, u^e)$ is governed by

$$\ddot{y} + a_1 \dot{y} + a_0 y = \gamma u + \phi \tag{2}$$

with

$$\begin{aligned} a_1 &= \frac{b}{J} + \frac{R_a}{L_a} \\ a_0 &= \frac{L_{af}^2}{JL_a}\left(i_f^e\right)^2 + \frac{R_a b}{JL_a} \\ \gamma &= \frac{L_{af}}{JL_a} i_f^e + \frac{L_{af}}{JL_f} i_a^e \end{aligned} \tag{3}$$

$$\begin{aligned} \phi &= \left[\frac{L_{af}^2}{JL_a}\left(i_f^e\right)^2 + \frac{R_a b}{JL_a}\right]\omega^e - \left(\frac{L_{af}}{JL_a} i_f^e + \frac{L_{af}}{JL_f} i_a^e\right) u_a^e \\ &\quad + \left[\frac{L_{af}}{J}\left(\frac{R_a}{L_a} - \frac{R_f}{L_f}\right) i_a^e - \frac{L_{af}^2}{JL_a} i_f^e \omega^e\right] i_{f\delta} \end{aligned} \tag{4}$$

Notice that, constants a_0, a_1 and γ depend on the system parameters and desired operation state. Thus, high operation efficiency levels could require information about some approximated values of the parameters of the motor subjected to completely unknown load torque disturbances.

In the design of some classical control law, ϕ could be considered as an completely unknown disturbance signal depending on the equilibrium operation state specified for the electromechanical system. Moreover, for a constant operation velocity $y = \overline{\omega}$, $i_f \longrightarrow i_f^e$ and, as a consequence, $\phi \longrightarrow \phi^e =$ constant. Therefore, we propose the following PI angular velocity tracking controller:

$$u = -k_p e - k_i \int_0^t e \, dt \tag{5}$$

where the proportional and integral control gains, k_p and k_i, should be chosen such as the characteristic polynomial associated to the closed loop tracking error dynamics, $e = \omega - \omega^\star$,

$$P(s) = s^3 + a_1 s^2 + (a_0 + \gamma k_p)\, s + \gamma k_i \tag{6}$$

is a Hurwitz polynomial. Hence, closed loop system stability can be verified. Notice that, the control gains should be also selected properly in accordance with the equilibrium operation point for the motor. Certainly, the on-line knowledge of the system parameters and load torque allows to tune easily the control gains during the operation of the machine.

The main objective of this paper is to propose an alternative choice for on-line estimation of all parameters and load torque for a DC Shunt Motor. Thus, in the next section estimators are synthesized to get estimates of the system parameters algebraically and on-line.

2 On-line Algebraic Parameter Identification

Firstly, consider the dynamics of the electrical subsystem. Multiplication of the first two equations of model (1) by $\Delta = t - t_i$, and integrating the resulting expressions twice with respect to

time yields to

$$\begin{aligned} L_f\left[\Delta i_f - \int_{t_i}^{t} i_f\, dt\right] + R_f \int_{t_i}^{t} \Delta i_f\, dt \\ = \int_{t_i}^{t} \Delta u_f\, dt \\ L_a\left[\Delta i_a - \int_{t_i}^{t} i_a\, dt\right] + R_a \int_{t_i}^{t} \Delta i_a\, dt + L_{af} \int_{t_i}^{t} \Delta i_f \omega\, dt \\ = \int_{t_i}^{t} \Delta u_a\, dt \end{aligned} \tag{7}$$

where $t_i > 0$ is the start time to perform the parameter identification process.

By integrating up to twice Eqs. (7), we get the following equation systems

$$A_i \theta_i = B_i, \quad i = f, a \tag{8}$$

where $\theta_f = \begin{bmatrix} L_f & R_f \end{bmatrix}^T$ and $\theta_a = \begin{bmatrix} L_a & R_a & L_{af} \end{bmatrix}^T$ are the parameter vectors associated with the electrical subsystem to be identified on-line. Matrices A_i and B_i are given by

$$\begin{aligned} A_f &= \begin{bmatrix} a_{11,f} & a_{12,f} \\ a_{21,f} & a_{22,f} \end{bmatrix} \\ A_a &= \begin{bmatrix} a_{11,a} & a_{12,a} & a_{13,a} \\ a_{21,a} & a_{22,a} & a_{23,a} \\ a_{31,a} & a_{32,a} & a_{33,a} \end{bmatrix} \\ B_f &= \begin{bmatrix} b_{1,f} \\ b_{2,f} \end{bmatrix} \\ B_a &= \begin{bmatrix} b_{1,a} \\ b_{2,a} \\ b_{3,a} \end{bmatrix} \end{aligned} \tag{9}$$

with

$$\begin{aligned}
a_{11,f} &= \Delta i_f - \int_{t_i}^{t} i_f \, dt \\
a_{12,f} &= \int_{t_i}^{t} \Delta i_f \, dt \\
a_{21,f} &= \int_{t_i}^{t} a_{11,f} \, dt \\
a_{22,f} &= \int_{t_i}^{t} a_{12,f} \, dt \\
b_{1,f} &= \int_{t_i}^{t} \Delta u \, dt \\
b_{2,f} &= \int_{t_i}^{t} b_{1,f} \, dt \\
a_{11,a} &= \Delta i_a - \int_{t_i}^{t} i_a \, dt \\
a_{12,a} &= \int_{t_i}^{t} \Delta i_a \, dt \\
a_{13,a} &= \int_{t_i}^{t} \Delta i_f \omega \, dt \\
b_{1,a} &= \int_{t_i}^{t} \Delta u_a \, dt \\
a_{kh,a} &= \int_{t_i}^{t} a_{k-1h,a}(\tau_1) d\tau_1 \\
b_{k,a} &= \int_{t_i}^{t} b_{k-1,a}(\tau_1) d\tau_1
\end{aligned} \tag{10}$$

with $k = 2, 3$ and $h = 1, 2, 3$.

Therefore, from (8) the electrical subsystem parameters can be computed algebraically as

$$\theta_i = A_i^{-1} B_i \tag{11}$$

Nevertheless, parameter identifiers (11) could present problems of singularities when $detA_i = 0$. Hence, we propose the following algebraic identifiers to get estimates of the parameters of the electrical subsystem without singularities $\forall t_i > 0$:

$$\begin{aligned}
\widehat{L}_f &= \frac{\int_{t_i}^{t} |\Delta_{1,f}| \, dt}{\int_{t_i}^{t} |\Delta_f| \, dt} \\
\widehat{R}_f &= \frac{\int_{t_i}^{t} |\Delta_{2,f}| \, dt}{\int_{t_i}^{t} |\Delta_f| \, dt} \\
\widehat{L}_a &= \frac{\int_{t_i}^{t} |\Delta_{1,a}| \, dt}{\int_{t_i}^{t} |\Delta_a| \, dt} \\
\widehat{R}_a &= \frac{\int_{t_i}^{t} |\Delta_{2,a}| \, dt}{\int_{t_i}^{t} |\Delta_a| \, dt} \\
\widehat{L}_{af} &= \frac{\int_{t_i}^{t} |\Delta_{3,a}| \, dt}{\int_{t_i}^{t} |\Delta_a| \, dt}
\end{aligned} \tag{12}$$

Now, consider the dynamics of the mechanical subsystem described by third equation of model (1). By applying the same procedure explained before, one can get the following estimators for the mechanical parameters and load torque:

$$\begin{aligned} \widehat{J} &= \frac{\widehat{L}_{af}}{\sigma_1} \\ \widehat{b} &= \frac{\sigma_2}{\sigma_1}\widehat{L}_{af} \\ \widehat{\tau}_L &= \frac{\sigma_3}{\sigma_1}\widehat{L}_{af} \end{aligned} \tag{13}$$

where σ_j, $j = 1, 2, 3$, are given by

$$\begin{aligned} \sigma_1 &= \frac{\int_{t_i}^{t} |\Delta_{1,m}| \ dt}{\int_{t_i}^{t} |\Delta_m| \ dt} \\ \sigma_2 &= \frac{\int_{t_i}^{t} |\Delta_{2,m}| \ dt}{\int_{t_i}^{t} |\Delta_m| \ dt} \\ \sigma_3 &= \frac{\int_{t_i}^{t} |\Delta_{3,m}| \ dt}{\int_{t_i}^{t} |\Delta_m| \ dt} \end{aligned} \tag{14}$$

with

$$\theta_m = \left[\begin{array}{ccc} \sigma_1 & \sigma_2 & \sigma_3 \end{array} \right]^T = A_m^{-1} B_m \tag{15}$$

$$\begin{aligned} A_m &= \left[\begin{array}{ccc} a_{11,m} & a_{12,m} & a_{13,m} \\ a_{21,m} & a_{22,m} & a_{23,m} \\ a_{31,m} & a_{32,m} & a_{33,m} \end{array} \right] \\ B_m &= \left[\begin{array}{c} b_{1,m} \\ b_{2,m} \\ b_{3,m} \end{array} \right] \end{aligned} \tag{16}$$

and

$$\begin{aligned} a_{11,m} &= \int_{t_i}^{t} \Delta i_f i_a \ dt \\ a_{12,m} &= -\int_{t_i}^{t} \Delta\omega \ dt \\ a_{13,m} &= -\int_{t_i}^{t} \Delta \ dt \\ b_{1,m} &= \Delta\omega - \int_{t_i}^{t} \omega \ dt \\ a_{kh,m} &= \int_{t_i}^{t} a_{k-1h,m}(\tau_1) d\tau_1 \\ b_{k,m} &= \int_{t_i}^{t} b_{k-1,m}(\tau_1) d\tau_1 \end{aligned} \tag{17}$$

with $k = 2, 3$ and $h = 1, 2, 3$.

3 Simulation results

Effectiveness of the proposed parameter estimation scheme was verified by computer simulations. The parameter values of the DC motor are described in Table 1.

Table 1: Parameters of the DC motor.

$R_a = 7.5\ \Omega$	$L_{af} = 2.2881$ H
$L_a = 0.0553$ H	$J = 0.0013$ Kg m^2
$R_f = 469.75\ \Omega$	$b = 0.001$ Nms
$L_f = 2.4123$ H	$\tau_L = 0.5$ Nm

The reference velocity trajectory $\omega^\star$ planned for the electromechanical system is shown in Fig. 1 and described by

$$\omega^\star(t) = \begin{cases} 0 & \text{for } 0 \leq t < T_i \\ \varpi\left(t, T_i, T_f\right)\bar{\omega} & \text{for } T_i \leq t \leq T_f \\ \bar{\omega} & \text{for } t > T_f \end{cases} \tag{18}$$

where $\bar{\omega} = 10$ rad/s, $T_i = 0$ s, $T_f = 5$ s, $\varpi\left(t, T_i, T_f\right)$ is a Bézier interpolation polynomial, with $\varpi\left(T_i, T_i, T_f\right) = 0$ and $\varpi\left(T_f, T_i, T_f\right) = 1$, given by

$$\begin{aligned} \varpi(t) \;=\; & \left(\frac{t-T_i}{T_f-T_i}\right)^5 \left[d_1 - d_2\left(\frac{t-T_i}{T_f-T_i}\right)\right. \\ & \left. + d_3\left(\frac{t-T_i}{T_f-T_i}\right)^2 - \ldots - d_6\left(\frac{t-T_i}{T_f-T_i}\right)^5\right] \end{aligned}$$

with $d_1 = 252$, $d_2 = 1050$, $d_3 = 1800$, $d_4 = 1575$, $d_5 = 700$, $d_6 = 126$. This profile was established to efficiently take the motor from a rest state to a low operation velocity of 10 rad/s in 5 seconds.

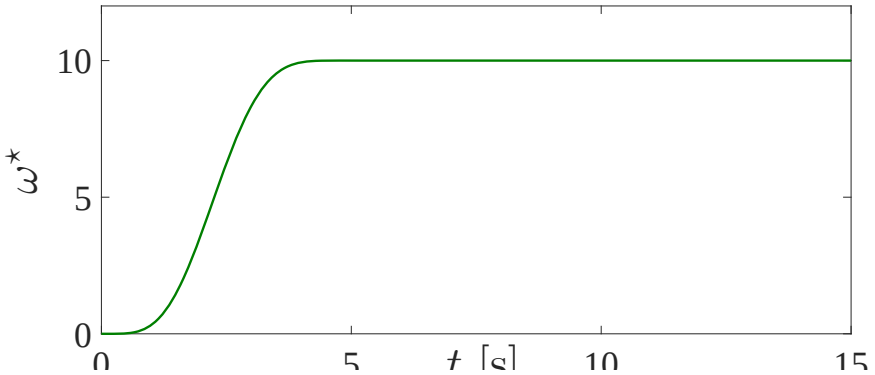


Figure 1: Reference angular velocity planned for the DC motor.

Fig. 2 depicts the satisfactory closed loop tracking of the desired velocity reference trajectory (18). A small velocity tracking error is clearly observed. The responses of the control voltage, current signals and electric powers are shown in Figs. 3 and 4. It can be observed that the properly controlled motion planning (18) avoids high peaks of the electric signals of voltage,

currents and powers in presence of load torque from the start. Moreover, it is known widely that large fluctuations of voltage could cause control saturations and system instability. Thus, planning motion tracking control represents an excellent choice to reduce these undesirable issues. PI controller gains were conveniently set as: $k_p = 100$ and $k_i = 10$. Nevertheless, the controller gains can be easily adjusted on-line for diverse operation states for the electric motor, including uncertain changes in the system parameters and load mechanical torque.

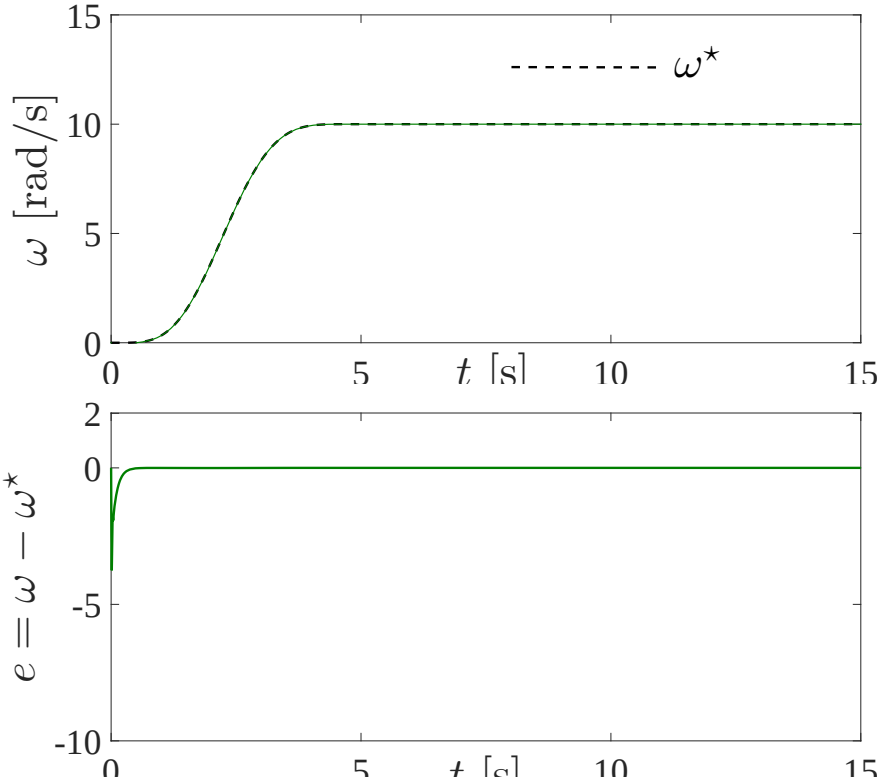


Figure 2: Closed loop tracking of the reference velocity trajectory.

On the other hand, the efficient performance of the parameter and torque identification scheme is presented in Figs. 5-7. An effective and fast estimation of the system parameters and load torque is confirmed. Estimates of all parameters and mechanical torque are quickly obtained before 1 second. Thus, those estimates can be used to tune the controller gains to improve the dynamic performance of the closed loop system and guarantee asymptotic stability around possibly varying-time desired operation states for the electric machine. Moreover, estimators could be updated continually for possible changes of the parameters and load torque during the motor operation.

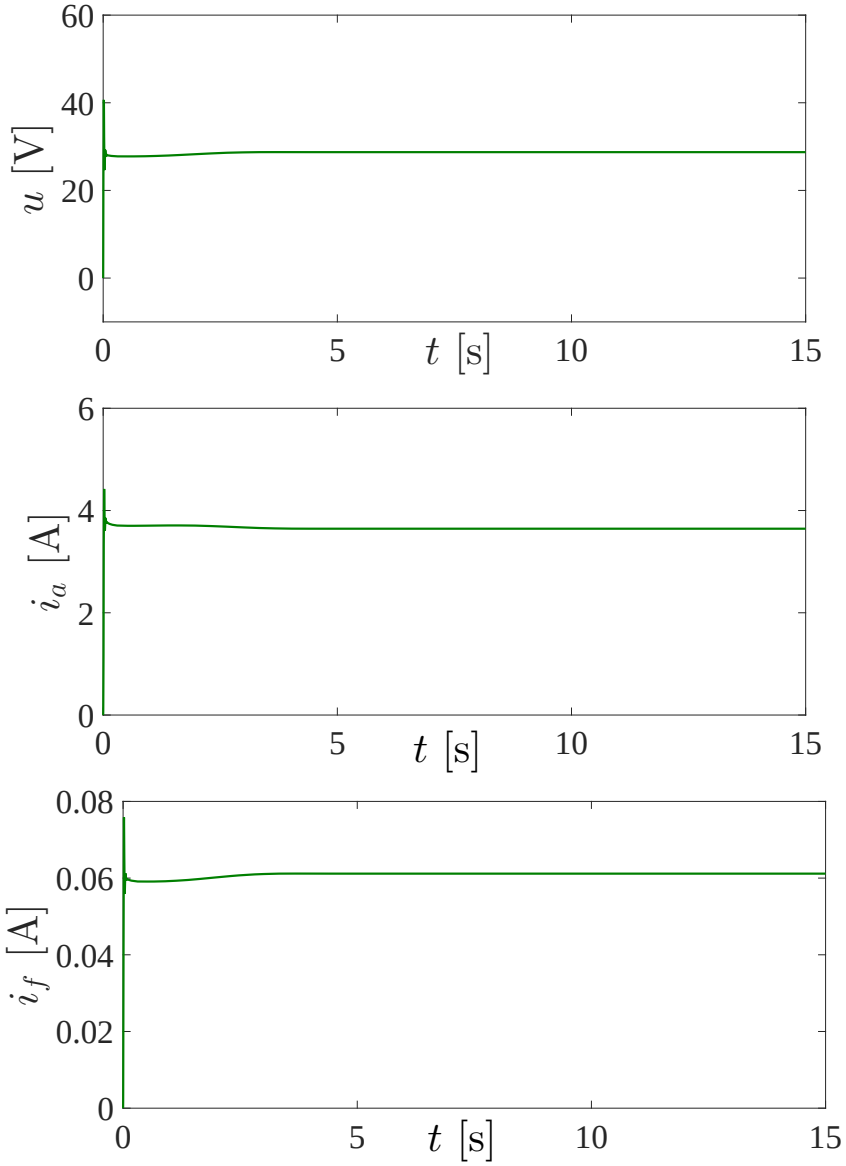


Figure 3: Closed loop responses of the control voltage and electric current signals.

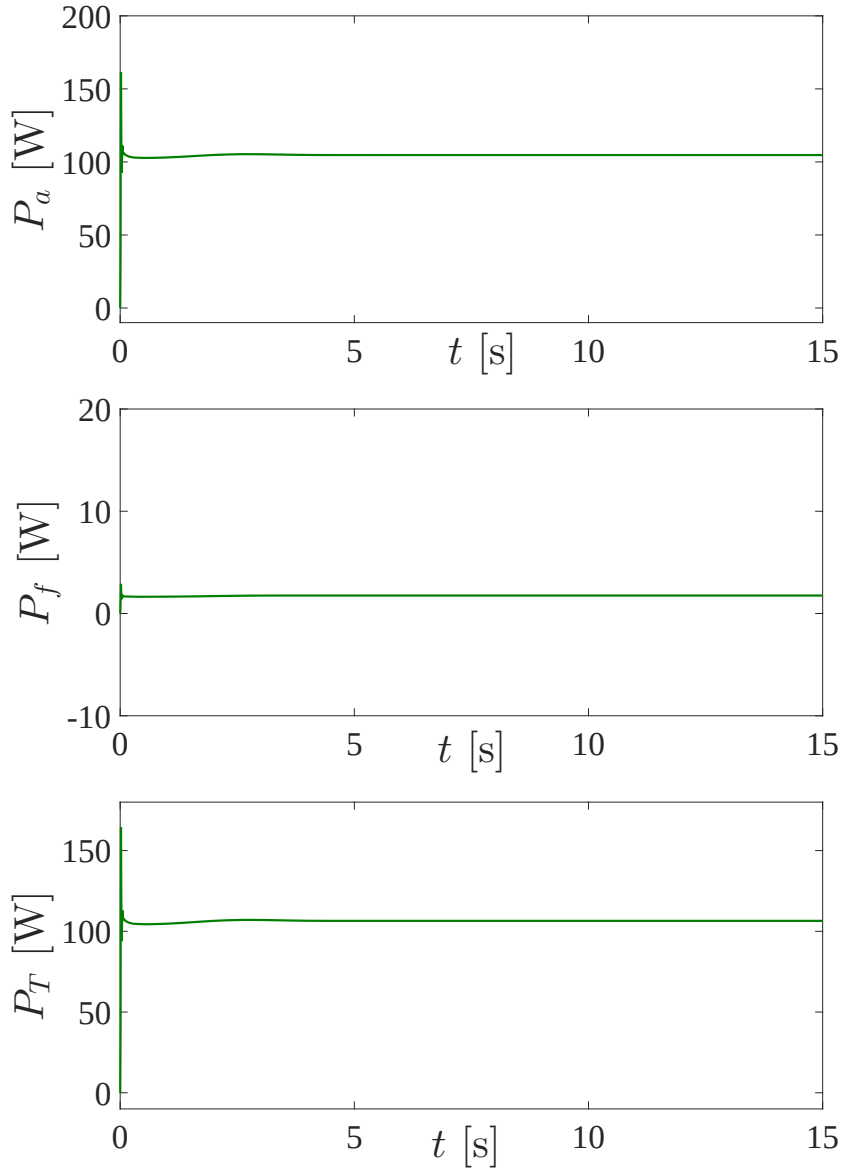


Figure 4: Closed loop responses of the electric powers of the armature circuit P_a, field P_f and total $P_T = P_a + P_f$.

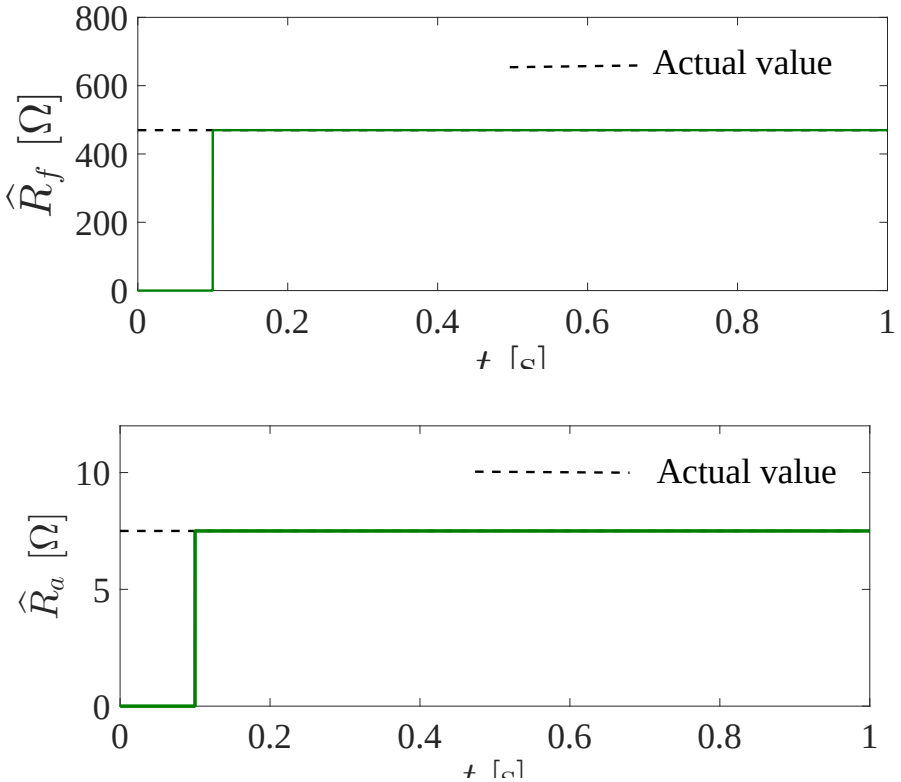


Figure 5: Algebraic estimation of the resistance parameters.

4 Conclusions

In this paper we have proposed an on-line estimation scheme for parameters and load torque for DC shunt motors. Connection in parallel of the field and armature windings of the motor results in a nonlinear system dynamics. The proposed estimation approach is performed algebraically and on-line. In addition, a PI tracking control law was also described to take the motor from a rest state toward a desired operation velocity. Controlled motion planning was established by a Bézier interpolation polynomial. It was shown that the suitable trajectory tracking avoid large fluctuations of voltage and, as a consequence, in the electric current signals as well. Analytical and numerical results show the effectiveness of the parameter and torque estimation for tracking tasks of reference angular velocity trajectories. Therefore, tracking control can be combined with on-line and algebraic estimation of system parameters and load mechanical torque to get satisfactory efficiency levels for DC shunt motors.

References

[1] Aimee McKane, Ali Hasanbeigi, Motor systems energy efficiency supply curves: A methodology for assessing the energy efficiency potential of industrial motor systems, *Energy Policy*, vol. 39, pp. 6595-6607, 2011.

[2] Aleksei Tepljakov, Emmanuel A. Gonzalez, Eduard Petlenkov, Juri Belikov, C. A. Monje, Ivo Petras, Incorporation of fractional-order dynamics into an existing PI/PID DC motor control loop, *ISA Transactions*, vol. 60, pp. 262-273, 2016.

[3] I. G. A. P. Raka Agung, S. Huda, I. W. Arta Wijaya, Speed control for DC motor with pulse width modulation (PWM) method using infrared remote control based on ATmega16 Microcontroller, IEEE International Conference on Smart Green Technology in Electrical and Information Systems, (ICSGTEIS), pp. 108-112, 2014.

[4] F. Beltran-Carbajal, A. Favela-Contreras, A. Valderrabano-Gonzalez, J. C. Rosas-Caro, Output feed-

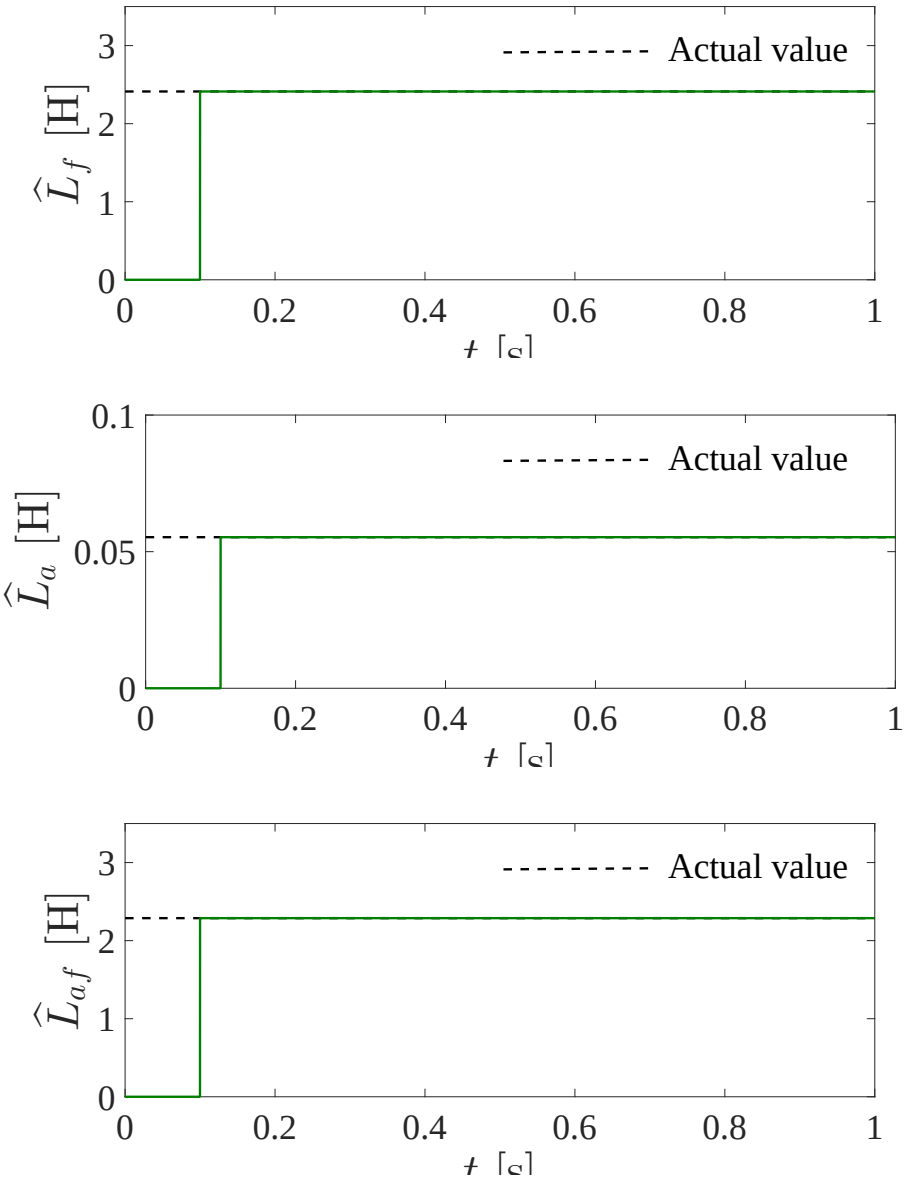


Figure 6: Algebraic estimation of the inductance parameters.

back control for robust tracking of position trajectories for DC electric motors, *Electric Power Systems Research*, vol. 107, pp. 183-189, 2014.

[5] R. Isermann, M. Munchhof, *Identification of Dynamic Systems*, Springer-Verlag, Berlin (2011).

[6] L. Ljung, *Systems Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, NJ (1987).

[7] T. Soderstrom, P. Stoica, *System Identification*, Prentice-Hall, New York, NY (1989).

[8] P. Dhinakaran, D. Manamalli, Novel strategies in the Model-based Optimization and Control of Permanent Magnet DC motors, *Computers & Electrical Engineering*, vol. 44, pp. 34-41, 2015.

[9] T. Boileau, N. Leboeuf, B. Nahid-Mobarakeh, F. Meibody-Tabar, Online identification of PMSM parameters: parameter identifiability and estimator comparative study, *IEEE Trans Indust Appl*, vol. 47, No. 4, 2011.

[10] A. Rahimi, F. Bavafa, S. Aghababaei, M. Hassan Khooban, S. Vahid Naghavi, The online parameter identification of chaotic behaviour in permanent magnet synchronous motor by Self-Adaptive Learning Bat-inspired algorithm, *Electrical Power and Energy Systems*, vol. 78, pp. 285-291, 2016.

[11] M. Jirdehi, A Rezaei, Parameters estimation of squirrel-cage induction motors using ANN and

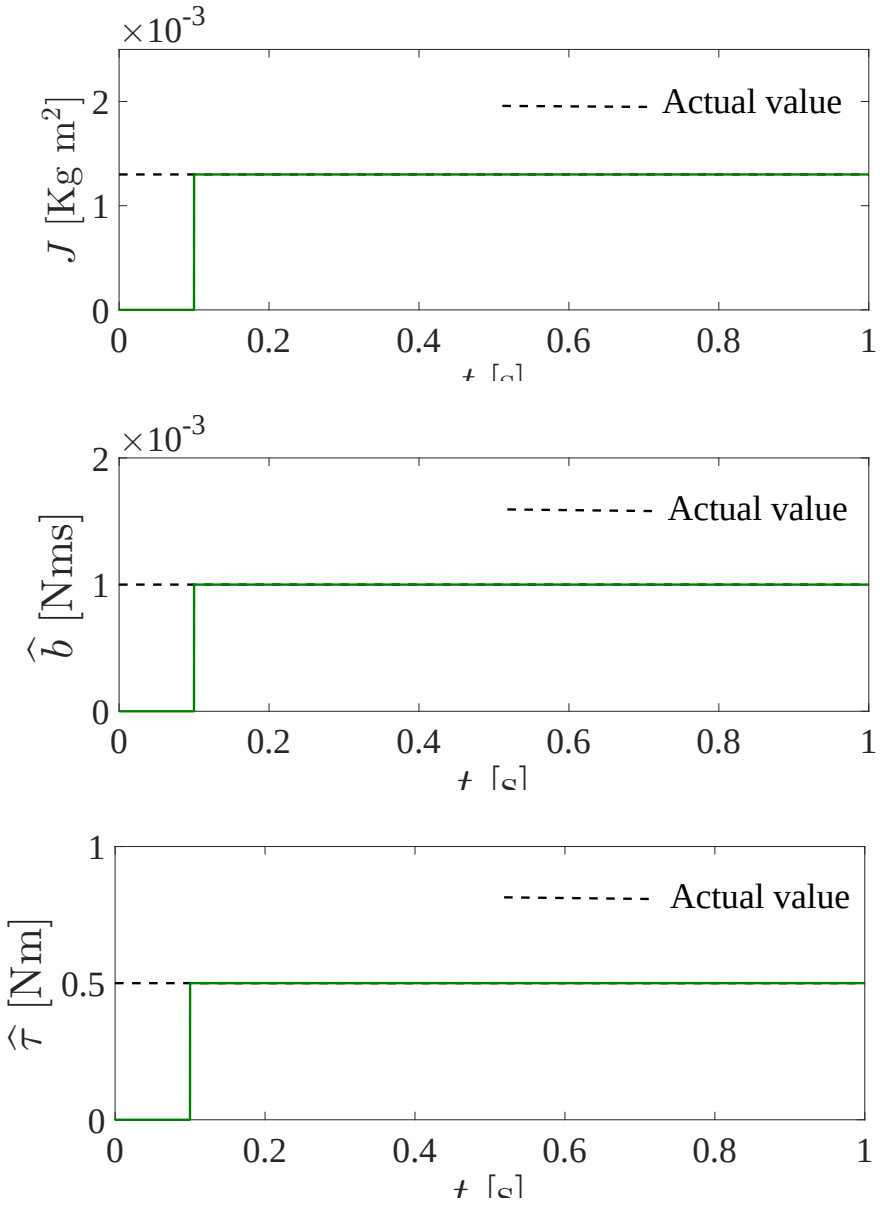


Figure 7: Algebraic estimation of the mechanical subsystem parameters and load torque.

ANFIS, *Alexandria Engineering Journal*, vol. 55, pp. 357-368, 2016.

[12] M. Fliess and H. Sira-Ramirez, An algebraic framework for linear identification, *ESAIM: Control, Optimization and Calculus of Variations*, 9 (2003) 151-168.

[13] F. Beltran-Carbajal, G. Silva-Navarro, Adaptive-like vibration control in mechanical systems with unknown parameters and signals, *Asian Journal of Control* 15 (6) (2013), 1613-1526.

[14] F. Beltran-Carbajal, G. Silva-Navarro, M. Arias-Montiel, Active unbalance control of rotor systems using on-line algebraic identification methods, *Asian Journal of Control*, 15 (6) (2013), 1627-1637.

[15] F. Beltran-Carbajal, G. Silva-Navarro, L. G. Trujillo-Franco, Evaluation of on-line algebraic modal parameter identification methods, *Proc. of the 32nd IMAC, A conference and Exposition on Structural Dynamics*, Orlando, Florida, USA, February 3-6, 2014, 145-152.

Novel 6-DoF dexterous parallel manipulator with CRS kinematic chains

†M.A. Hosseini[1]

[1]Department of Mechanical Engineering, University of Mazandaran, Babolsar, Iran.

*Presenting author: ma.hosseini@umz.ac.ir
†Corresponding author: ma.hosseini@umz.ac.ir

Abstract

In this research work, a novel parallel manipulator with 6 degrees of freedom (DoF) and high positioning and orienting rate is introduced. Kinematics and Jacobian analysis are investigated. Workspace of mechanism considering different rotation capabilities are computed and illustrated in Cartesian coordinates. Defining *global maximum* and *minimum singular values* of homogenized jacobian matrix through the workspace has been utilized in order to synthesis positioning and orienting rates capability of mechanism. Thus, improving high rates of displacement is achieved by elimination of moving elements and changing kinematic chains compared with general stewart-gough mechanism, which makes it suitable in pick and place or motion stabilizer devices and high speed machining applications with lower payload**.**

Keywords: Kinematics, Workspace, 6-CRS, Parallel robot.

Introduction

Potential superior properties of parallel manipulators such as low inertia, high stiffness, high precision and high load carrying capacity [1]-[2] of parallel manipulators lead to extensive attention over the last three decades of them. Performance indices such as manipulability, condition number, conditioning and dexterity are useful for comparison studies of different robot structures. Manipulability at first was introduced by Yoshikawa [3] as the square root of the determinant of the product of the manipulator Jacobian by its transpose.

The Jacobian matrix maps a unit ball in the joint space into a rotated or reflected ellipsoid in the Cartesian space. The geometric interpretation of the mapping is proportional to the volume of the ellipsoid or the manipulability [3]. Moreover, the volume is equal to the products of the singular values of the Jacobian [3]. Salisbury and Craig [4] introduced the ratio between the maximum and minimum singular values as the condition number. The inverse of the Euclidean condition number is defined as conditioning index which varies from 0 to 1. if the entries of the Jacobian have different units for the manipulators with both positioning and orientation tasks, which is the case here, one faces a problem of ordering singular values of different units from largest to smallest. Ranjbaran and Angeles [5] introduced carachteristic length to resolve this issue. Gosselin [6] introduced a method for formulating dimensionally homogeneous Jacobian matrix for a planar mechanism with one rotational and two translational degreeof-freedom (dof). Kim and Ryu [7] furthered this work by using the velocities of three points on the endeffector platform to develop a dimensionally homogeneous Jacobian matrix. Pond and Corretero [8] furthered this method again by using three independent coordinates of three points on an end-effector platform. Moreover, Angeles [9] introduced engineering characteristic length for a rigid body transformation matrix to make it homogeneous. Finally, Hosseini et. al. [10]-[11], introduced a weighting factor method to make it homogeneous.

Here a novel mechanism with high positioning and orienting rate is introduced. Its kinematic is studied and its Jacobian matrices are derived from these equations. Because of complexity of DoF, Jacobian matrix is homogenized by using weighted factor method [10].

Moreover, kinematic indices for a trajectory have been investigated and compared with the similar size of stewart-gough mechanism, as a case study. Although decreasing the moving elements leads to better dynamic performances, this investigation could demonstrate kinematic indices improvement due to structural transformation at all.

I. 6-CRS Parallel Manipulator

As depicted in Fig. 1, 6-CRS parallel manipulator consists of two platforms connecting to each other by six identical active C-R-S (Cylindrical-Revolute-Spherical) legs. The active legs consist of a fixed length link connected to the mobile platform by a passive spherical joint. On the other extremity of the leg there is an actuated prismatic joint followed by a passive revolute joint.

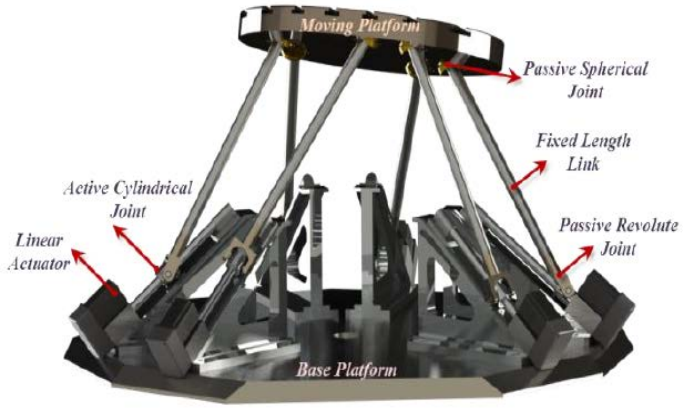


Figure 1. CAD model of 6-CRS parallel manipulator

II. Kinematic Analysis

Geometrical model of the mechanism is illustrated in Fig. 2. Two moving and global frames ({*P* (*uvw*)} and {*O* (*xyz*)}) are attached to the moving and base platforms, respectively.

The kinematic close loop equation can be written as follow for each leg:

$$\mathbf{x} + a\mathbf{R}\mathbf{n}_{ai} = b\mathbf{n}_{bi} + q_i\mathbf{n}_{qi} + \mathrm{l}\mathbf{n}_{li} \quad . \tag{1}$$

where **x** is the vectors from *O* to *P*, i.e. the end effector position vector. Moreover, **R** is rotation matrix carrying frame {*P*} into an orientation coincident with that of frame {*O*}; $\mathbf{n}_{ai}$ is the i^{th} spherical joint position unit vector in the moving frame. Similarly, $\mathbf{n}_{bi}$, $\mathbf{n}_{qi}$ and $\mathbf{n}_{li}$ are the unit vectors from *O* to B_i, B_i to Q_i and Q_i to A_i, respectively; while *a* and *b* are the radius of the moving and base platform that joints are posed on. Furthermore, the moving part of the limbs length is *l*.

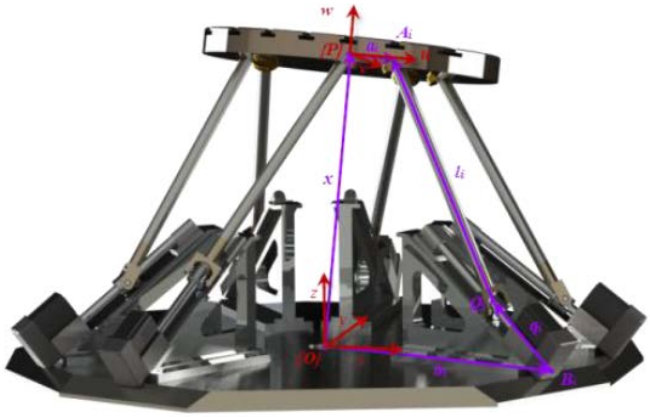


Figure 2. Geometrical Model of 6-CRS

A. Inverse Kinematic

In the Inverse kinematic problem the pose of the end-effector (EE) is given and the joint variables that produce this pose are to be found. Considering the i^{th} leg as depicted in Fig. 3; it is obvious that Q_i is on the surface of a sphere with the centre A_i and radius of l. Then the intersection of this sphere with the slant base concludes the inverse kinematic problem roots.

The position vector of A_i can be defined by the following equation.

$$\mathbf{a}_i = \mathbf{x} + a\mathbf{R}\mathbf{n}_{ai} . \tag{2}$$

Considering spherical and universal joints position vector as $\mathbf{a}_i=[x_{ai}\ y_{ai}\ z_{ai}]^T$ and $\mathbf{b}_i=[x_{bi}\ y_{bi}\ 0]^T$ the parametric equation of GB_i can be written as follow, in which the intersection of all slant bases is illustrated by G.

$$x = -x_{bi}t_i + x_{bi}; y = -y_{bi}t_i + y_{bi}; z = ht_i \tag{3}$$

where h is the height of G point.

Substituting the above equations in the parametric equation of sphere as the following:

$$(x - x_{ai})^2 + (y - y_{ai})^2 + (z - z_{ai})^2 - l^2 = 0 \tag{4}$$

Leads to the following equation

$$m_i t_i^{\,2} - 2n_i t_i + p_i = 0 \tag{5}$$

In which coefficients are given as:

$$m_i = (x_{bi}^{\,2} + y_{bi}^{\,2} + h^2) \tag{6}$$

$$n_i = (x_{bi}^{\,2} + y_{bi}^{\,2} - x_{bi}x_{ai} - y_{bi}y_{ai} + hz_{ai}) \tag{7}$$

$$p_i = (x_{bi}^2 + x_{bi}^2 - 2x_{bi}x_{ai} + y_{bi}^2 + y_{ai}^2 - 2y_{bi}y_{ai} + z_{ai}^2 - l^2) \tag{8}$$

Solving Eq. (5) for t_i and substituting the values of Eq. (3) led to the inverse kinematic problem solution. This approach could help to avoiding impossible roots such as R_{i2} in Fig 3. Thus, only the roots are acceptable in which associated t_i lie in desired interval satisfied by the linear actuator stroke.

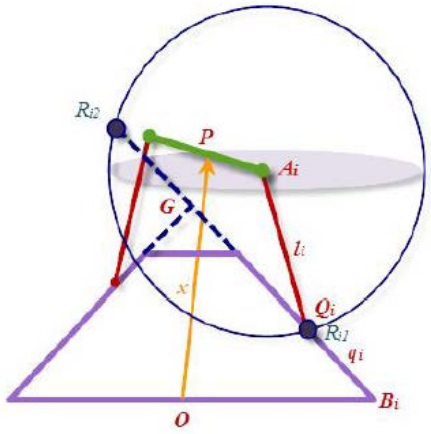


Figure 3. Schematic configuration of 6-CRS kinematic

The following cases may occur:

Case 1) The slant guide way does not intersect the associated sphere. Thus there is no solution for IKP (Inverse Kinematic Problem), i.e., the assumed position would be out of reach by the EE(End Efector).

Case 2) The slant guide way intersects with the associated sphere at one point. Therefore, IKP leads to only one solution for the corresponding leg.

Case 3) The slant guide way intersects with the associated sphere at two points. Therefore, IKP leads to two solutions for the corresponding leg, as depicted in Fig 3, by R_{i1} and R_{i2}.

Therefore, the IKP might leads to 2^6 solutions (with considering dual roots) or no solution at all.

III. Jacobian Matrix and Velocity Analysis

The first time derivative of Eq. (1) leads to:

$$\dot{\mathbf{x}} + \boldsymbol{\omega}_p \times a\mathbf{R}\mathbf{n}_{ai} = \dot{q}_i \mathbf{n}_{qi} + \boldsymbol{\omega}_l \times l\,\mathbf{n}_{li} \tag{9}$$

In which $\boldsymbol{\omega}_l$ and $\boldsymbol{\omega}_p$ are the angular velocities of the fixed length link and the moving platform, respectively. Inner product of the both sides of Eq.9 by $\mathbf{n}_{li}$, upon simplifications leads to:

$$\dot{\mathbf{x}}\mathbf{n}_{li}^T + \boldsymbol{\omega}_p \times a\mathbf{R}\mathbf{n}_{ai}\mathbf{n}_{li}^T = \dot{q}\mathbf{n}_{qi}\mathbf{n}_{li}^T \tag{10}$$

Equation (10) can be rewritten as bellow

$$\mathbf{n}_{li}^{T}\dot{\mathbf{x}} + a(\mathbf{n}_{li} \times \mathbf{R}\mathbf{n}_{ai})\boldsymbol{\omega}_{p} = \dot{q}\mathbf{n}_{li}^{T}\mathbf{n}_{qi} \tag{11}$$

Writing the foregoing equation for the three legs yields to:

$$\mathbf{A}\dot{\mathbf{x}} = \mathbf{B}\dot{\mathbf{q}} \tag{12}$$

In which $\dot{\mathbf{x}}$ and $\dot{\mathbf{q}}$ are EE twist array and joint space velocity vector, respectively. Moreover, **A** and **B** are two Jacobian matrices which are given as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{n}_{li} & \mathbf{n}_{li} \times a\mathbf{R}\mathbf{n}_{ai} \end{bmatrix}_{6\times 6} \tag{13}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{n}_{l1}^{T}\mathbf{n}_{q1} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \mathbf{n}_{l6}^{T}\mathbf{n}_{q6} \end{bmatrix} \tag{14}$$

The Jacobian matrix can be determined by Eq. 15.

$$\mathbf{J} = \mathbf{B}^{-1}\mathbf{A} \tag{15}$$

IV. Singularity Analysis

Generally, singularity occurs whenever the manipulator loses some DoF or gains some uncontrollable DoF. In parallel manipulators singularities occur whenever **A**, **B** or both become singular. Thus, for the manipulator at hand a distinction can be made among three types of singularities, which have different kinematic interpretations.

For the 6-CRS parallel manipulator, singularity occurs in four cases, namely;

Case 1) First type of singularity or Inverse Singularity; in this case **B** is invertible and **A** is singular, i.e. when

$$\det(\mathbf{B}) = 0 \,\&\, \det(\mathbf{A}) \neq 0 \tag{16}$$

The physical condition happens when one of the fixed length link is perpendicular to the direction of the associated linear guide way.

Case 2) Second type of singularity or Direct Singularity; arises when **B** is singular and **A** is invertible, i.e. when

$$\det(\mathbf{B}) \neq 0 \,\&\, \det(\mathbf{A}) = 0 \tag{17}$$

This case occurs when the z coordinates of the fixed-length links vector is equal to zero. In this condition all three legs lie in the plane of the moving platform which is parallel to the base one, as well. Hence, by increasing or decreasing the actuator length, there are two options for A_i to locate, as depicted in Fig. 4, by 1 and 2.

Case 3) Third type of singularity; this type of singularity arises even if both **B** and **A** are simultaneously singular. Under a singularity of this type the manipulator can undergo finite motions even if the actuators are locked. As well, a finite motion of actuators produces no motion for EE in some directions.

Case 4) Constraint singularity; this case will occur when the moving platform rotates 90 degrees around x or y axis. In this case the platform will lose one rotational dof. Zalatanov et. al. [12] illustrated some constraint singularities, as well.

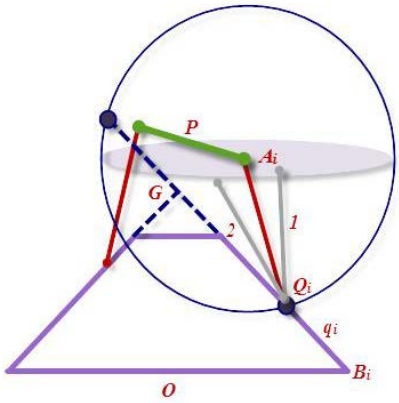


Figure 4. Schematic for direct singularity

V. Workspace and Optimization

Applying the inverse kinematic equations and a search algorithm in different height leads to the bound of reachable workspace [13]. This operation will be continuing as the geometric constraints are satisfied, subject to Table 1.

Table 1. Geometrical constraint for mechanism

Actuator (mm)	*l (mm)*	*λ (deg)*	*b (mm)*	*a (mm)*
0-600	*100-300*	*10-80*	*300-500*	*100-300*

As a case study, the Cartesian workspace of the structure according to Table 2, with the foregoing constraints is depicted in Fig. 5 in which the workspaces are depicted considering different rotation capabilities around three axes. Moreover, sub workspaces include bounded local conditioning indices into a minimum allowable of 0.0003 are depicted in Fig. 6 which singularity avoidance is performed.

Table 2. The case study design parameters

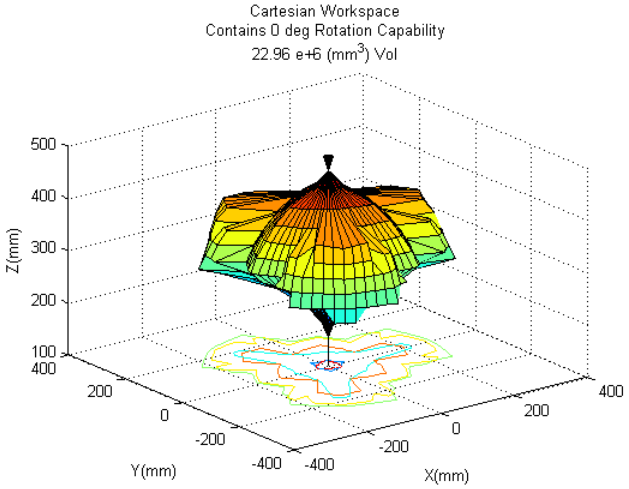
l (mm)	*λ (deg)*	*b (mm)*	*a (mm)*
300	30	300	100

Considering 100 (mm) weight factor for homogenized jacobian matrix, for the workspace with 20 degree rotation capability, the performance indices such as global conditioning index (GCI), average minimum and maximum singular values are depicted in Table 3.

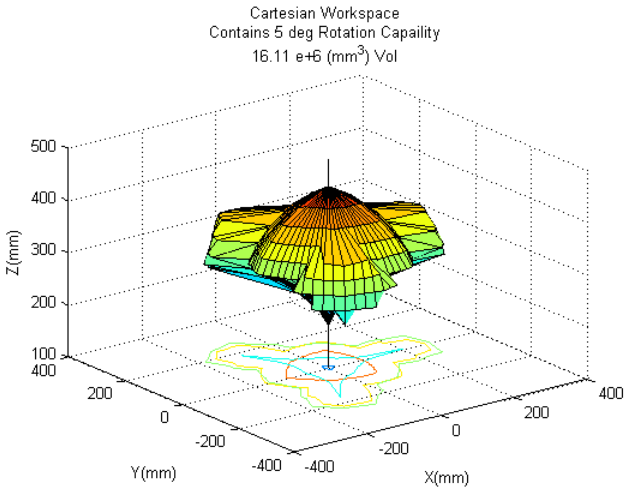
Table 3. The case study performance indices

$V(mm^3)$	GCI	$\bar{\sigma}_{max}$	$\bar{\sigma}_{min}$
6.51e+6	0.9895	1.5016e+4	2.5899

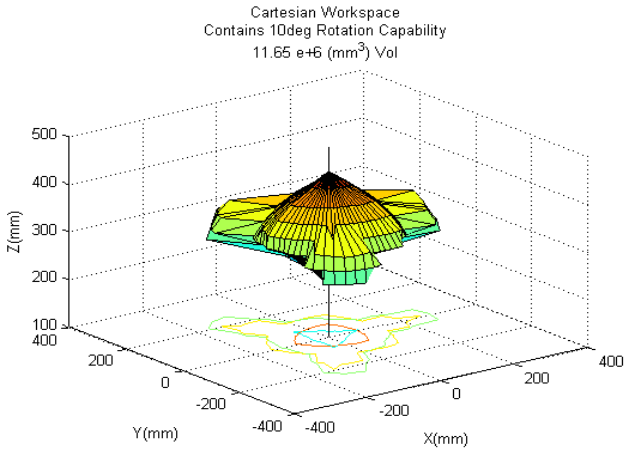
Global conditioning index (GCI) [6], are defined as following equations.



a. *0 deg Rotation Capability Cartesian Workspace*

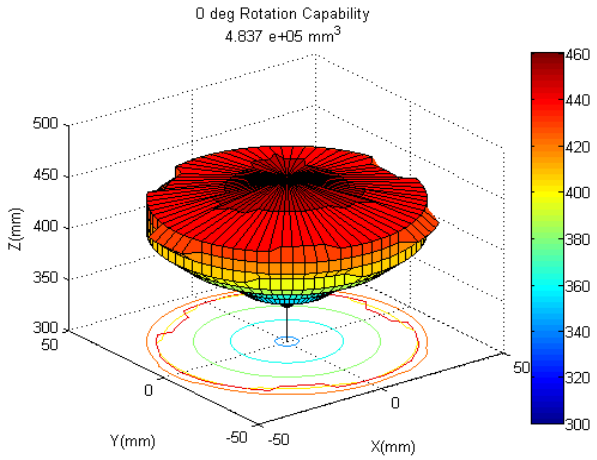


b. *5 deg Rotation Capability Cartesian Workspace*

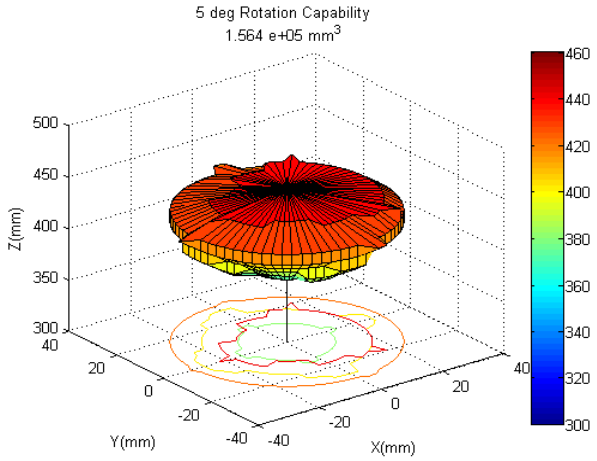


b. *10 deg Rotation Capability Cartesian Workspace*

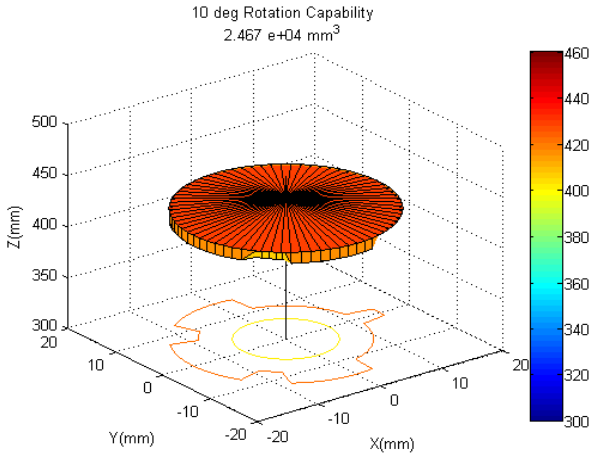
Figure 5. The case study workspaces with different rotation capabilities



d. *Subworkspace with 0 deg Rotatioon Capability with minimum 0.0003 LCI*



e. *Subworkspace with 5 deg Rotatioon Capability with minimum 0.0003 LCI*



f. *Subworkspace with 10 deg Rotatioon Capability with minimum 0.0003 LCI*

Figure 6. Sub workspaces with different rotation capabilities

$$GCI = \frac{\int \kappa dv}{\int dv} \tag{18}$$

In which local conditioning index (κ) for the workspace element is determined by the respective of minimum and maximum singular values of homogenized jacobian matrix using weighted factor method.

$$\kappa = \frac{\sigma_{\min}}{\sigma_{\max}} \tag{19}$$

Respectively, the average maximum singular value and average minimum singular value indices as the performances indices for positioning and orienting rates are defined as follow.

$$\bar{\sigma}_{\max} = \frac{\int \sigma_{\max} dv}{\int dv} \tag{20}$$

And

$$\bar{\sigma}_{\min} = \frac{\int \sigma_{\min} dv}{\int dv} \tag{21}$$

Lower value of $\bar{\sigma}_{\max}$ led to higher end-effector positioning and orienting resolution and higher value of $\bar{\sigma}_{\min}$ led to higher positioning and orienting rates [6].

Conclusions

In this research work a novel parallel manipulator with 6-CRS kinematic chains is introduced. The mechanism has 6 degrees of freedom. Inverse kinematic equations with a geometrical approach have been solved and used to workspace evaluation. Proposed parametric solution method leads to avoidance of actuators to locate into other inverse kinematic solutions sets. Jacobian matrix is derived by taking the first time derivation respect to time. Jacobian entries inhomogeneity has resolved by weighted factor approach equal with moving platform radius. Considering minimum desired rotation angles workspaces estimated in Cartesian workspaces. Bounding minimum local conditioning indices to the minimum allowable value led to sub workspaces with different rotation capabilities. Finally for the case study structure, some global indices are calculated in order to have performance indices for comparison between other same-dof parallel manipulator.

References

[1] Merlet, J. –P. (2006) Parallel Robots, Springer.

[2] Masory, O. and Wang, J. (1995) Workspace evaluation of Stewart platforms, *Advanced Robotics Journal* **9**, 443- 461.

[3] Yoshikawa, T. (1985) Manipulability of robotic mechanisms, Int. J. Robot. Res., **4**, 3–9.

[4] Salisbury, J. K., and Craig, J. J. (1982) Articulated hands: Force control and kinematic issues, Int. J. Robot. Res., **4**, 4–17.

[5] Ranjbaran, F., Angeles, J., Gonzalez-Palacios, M. A. and Patel, R. V. (1995) The mechanical design of a seven-axes manipulator with kinematic isotropy, *Journal of Intelligent and Robotic Systems.*, **14**, 21-41.

[6] Gosselin, C.M. (1992) The optimum design of robotic manipulators using dexterity indices, *Journal of Robotics and Autonomous Systems*, **9**, 213–226.

[7] Kim, S. G. and Ryu, J. (2003) New dimensionally homogeneous jacobian matrix formulation by three end-effector points for optimal design of parallel manipulators, IEEE Transactions on Robotics and Automation, **19**, 731–737.

[8] Pond, G. and Carretero, J.A. (2007) Quantitative dexterous workspace comparison of parallel manipulators, Mechanism and Machine Theory, **42**, 1388-1400.

[9] Angeles, J. (2006) Is there a characteristic length of a rigid-body displacement?, Mechanism and Machine Theory, **41**, 884–896.

[10] Hosseini, M.A. and Daniali, H.M. (2011) Weighted local conditioning index of a positioning and orienting parallel manipulator, Sientica Iranica B, **8**, 115-120.

[11] Hosseini, M.A., Daniali, H.R. M. and Taghirad, H.D. (2011) Dexterous Workspace optimization of Triceps Parallel Manipulator, Advanced Robotics, **25**, 1697-1712.

[12] Zlatanov, D., Bonev, I.A. and Gosselin, C.M. (2002) Constraint Singularities of Parallel Mechanisms, IEEE International Conference on Robotics and Automation (ICRA 2002), Washington, D.C., USA, May, 11–15.

[13] Hosseini, M.A. and Daniali, H.M. (2011) Machine Tool Design Optimization of High Resolution Parallel Hexapod in Cartesian Workspace, *Majlesi Journal of Mechanical Engineering*, **4**, 75-84.

Topology Optimization of the Interior Structure of Blades with Optimized

***†G.R. Liu, Dustin McClanahan, and Dr. Mark Turner**

Department of Aerospace Engineering and Engineering Mechanics
University of Cincinnati
Cincinnati, OH45221, USA.

*Presenting author: nguyettnmail@gmail.com
†Corresponding author: liugr@uc.edu

Abstract

For any given geometry of blade-type structure with desired outer-surface shape that may be determined by a CFD software for desired performance for thermal and fluid flows, a three-dimensional solid of the blade is converted into a CAD file. An optimization process is then designed to produce optimal interior structure of the blade that follows the proposed step-by-step procedure, considering both the pressure on the outer surface and centrifugal forces produced by the rotational movements of the blade. The optimized blade will be hollow with minimum materials needed to take the pressure loading on the outer skin of the blade and the centrifugal force. 3D printers were used to produce the optimized blades.

Keywords: Optimization procedure, FEM, engine blade, topology optimization, hollow blade, centrifugal force

Optimization procedure

The proposed optimization process to produce optimal interior structure of the blade follows the following step-by-step procedure.

Step 1: Read in the CAD file of the solid blade into a finite element method (FEM) software package with standard meshing and topology design capability (such as ABAQUS® that is commercially available). A typical blade generated in this step is shown in Figure 1.

Step 2: Designate a non-design space for the blade, which is a very thin skin of the surface of the solid blade. The blade tip may not have a skin, if so desired.

Step 3: Designate the interior part (solid blade excluded the thin skin) as the design space for the topology optimization.

Step 4: Create FEM elements for both the non-design and design spaces of the blade solid (see, Figure 2)

Step 5: Assign material properties to all FEM elements for this blade.

Step 6: Specify the boundary conditions on the blade base.

Step 7: Apply loads on the blade surface, including the pressure from the CFD solution when determining the outer surface of the blade.

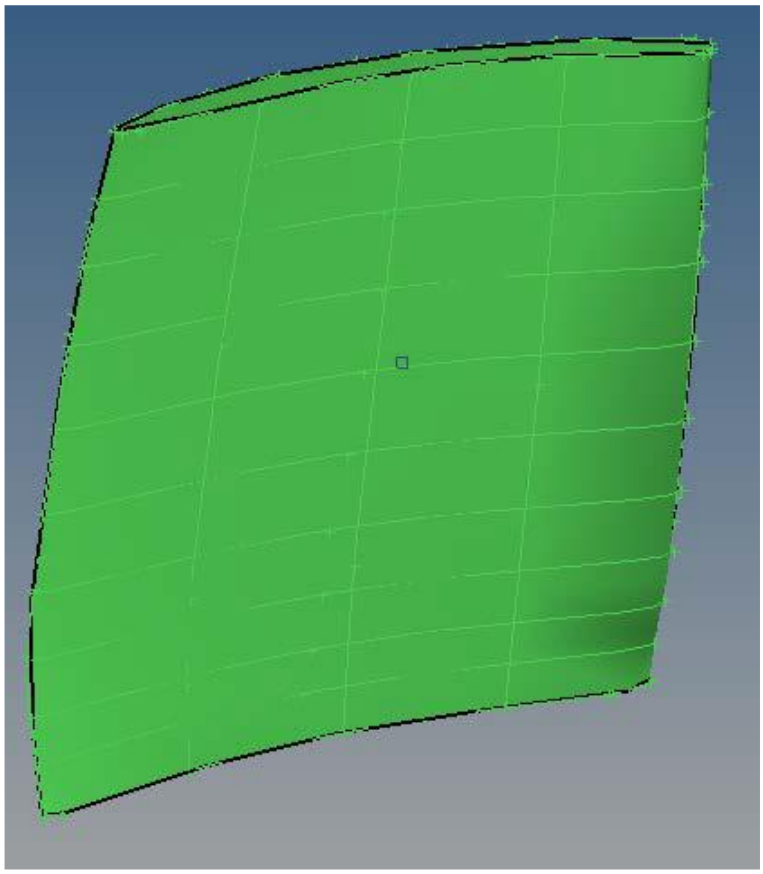


Figure 1 Geometry of a typical solid blade with outer surface determined by a CFD solver for desired performance for thermal and fluid flows under cruise conditions.

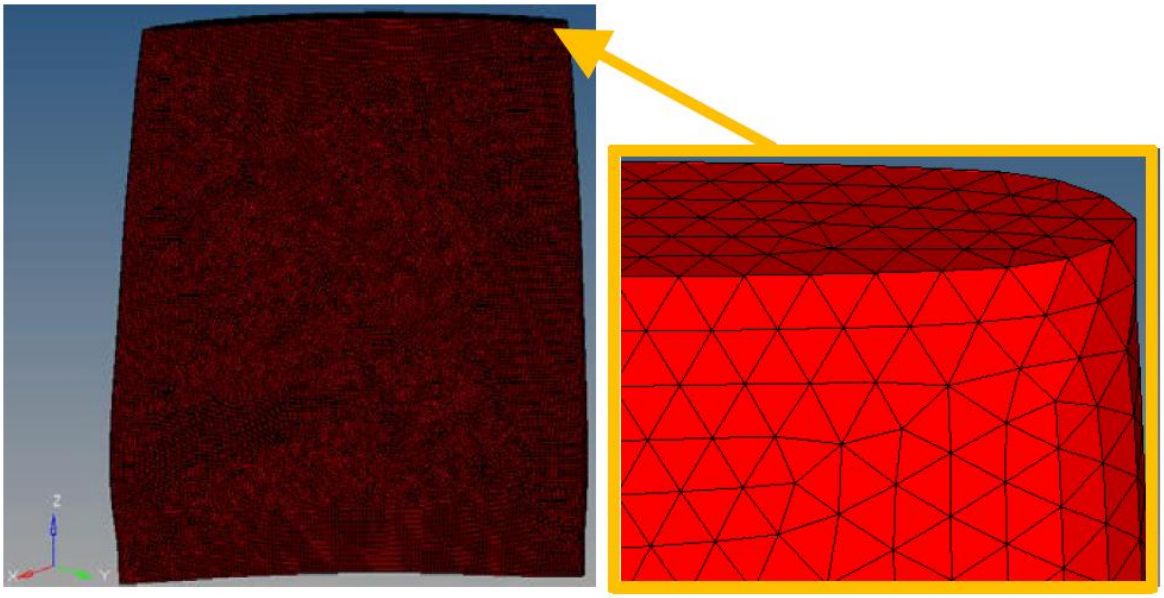


Figure 2 A typical finite element mesh for the solid blade. Left: fine mesh with dense tetrahedral elements; right: a zoomed in view

Step 8: Apply centrifugal loading for any given rotation of speed that the blade experiences at a steady state operational cruise condition.

Step 9: Set topology constraints, including limiting the stress below the material yield stress with a proper safety factor, frequency constraints, and life cycle constraints.

Step 10: Set topology optimization goal, such as aiming to create the stiffest possible structure.

Step 11: Run optimization using the standard FEM package to generate the optimized topological structure that is partially hollow, and satisfy all the design condition imposed. Figure 3 shows such a topologically optimized hollow blade

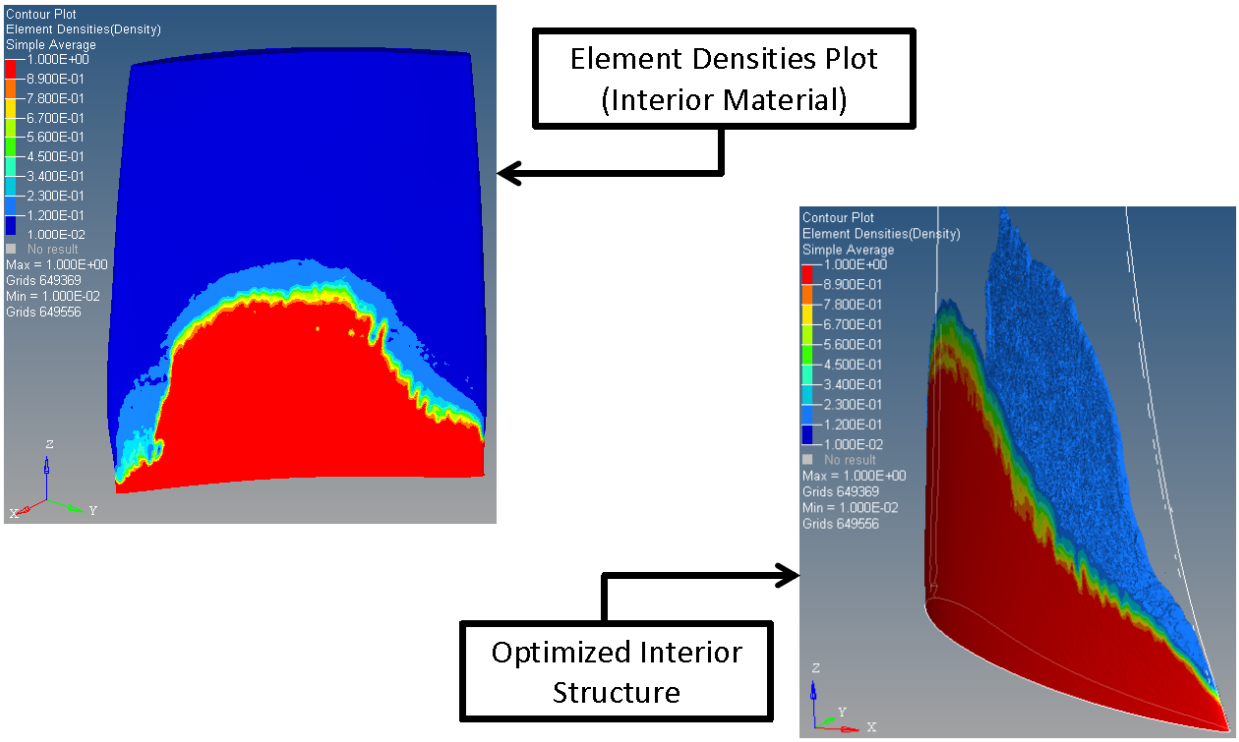


Figure 3 Topologically optimized hollow blade. The non-design space of the thin skin remains to achieve the desired performance of the blade considering thermal and fluid flow conditions. The design space of the interior solids is partially removed with materials only on the front and back surfaces of the blade near the base.

Step 12: Produce the topologically optimized blade using 3D printers.

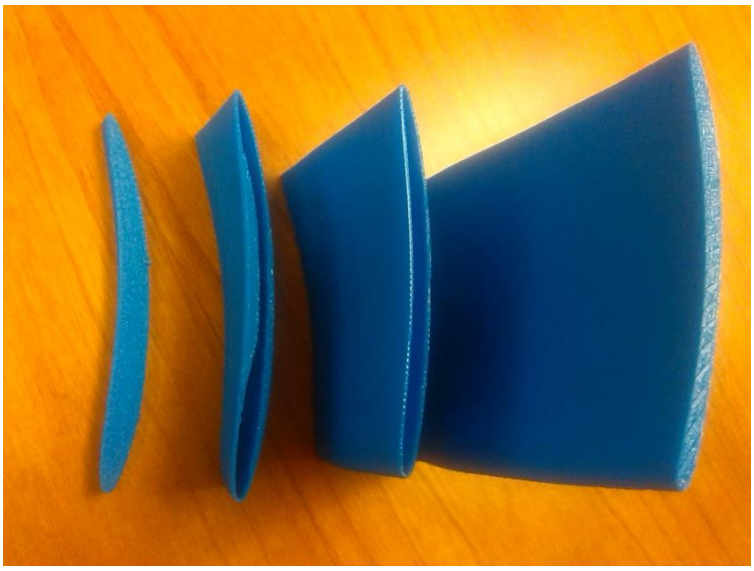


Figure 3 Topologically optimized hollow blade printed using 3D printer.

Conclusions

A topology optimization procedure has been developed using a commercially available FEM software package such ABAQUS® or any other FEM codes. For one particular example of an engine blade, the outcome of the weight reduction was 60-65%, and the max stress reduced by as much as 70%. The blade design was printed using 3D printers, and proven practical for topology optimal design for rotating structures for optimal performance with minimum weight.

References

[1] Liu GR & Quek SS, ***Finite element method***–*a practical course*, Elsevier, UK, 2nd Edition, 2013.

[2] Liu GR, ***Meshfree Methods****: Moving Beyond the Finite Element Method*. CRC Press: Boca Raton, USA, 2nd Edition, 2009. (CRC bestseller).

[3] Jiang, C, Han, X, Liu, GR, et al. (2008). A nonlinear interval number programming method for uncertain **optimization** problems. *Eur J Oper Res, 188*(1), 1-13.

[4] Tong, W. H., Jiang, J. S., & Liu, G. R. (2000). Solution existence of the **optimization** problem of truss structures with frequency constraints. *Int J Solids Struct, 37*(30), 4043.

[5] Tong, WH, & Liu, GR (2001), An **optimization** procedure for truss structures with discrete design variables and dynamic constraints. *Comput Struct, 79*(2), 155-162.

[6] Zheng, H, Cai, C, Pau, GSH., & Liu, GR (2005). Minimizing vibration response of cylindrical shells through layout **optimization** of passive constrained layer damping treatments. *Journal of Sound and Vibration, 279*(3-5), 739-756.

[7] Jiang, C, Han, X, & Liu, GR (2007). **Optimization** of structures with uncertain constraints based on convex model and satisfaction degree of interval. *Comput Method Appl Mech Eng, 196*(49-52), 4791-4800.

[8] J Yao, GR Liu, et. al. (2012), Immersed smoothed finite element method for fluid–structure interaction (**FSI**) simulation of aortic valves. *Comput Mech*, 50(6), 789-804.

[9] ABAQUS® user manuals.

Authors Index