Satya N. Atluri
Igor Vušanović *Editors*

Computational and Experimental Simulations in Engineering

Proceedings of ICCES 2020. Volume 2

Springer

Mechanisms and Machine Science

Volume 98

This book series establishes a well-defined forum for monographs, edited Books, and proceedings on mechanical engineering with particular emphasis on MMS (Mechanism and Machine Science). The final goal is the publication of research that shows the development of mechanical engineering and particularly MMS in all technical aspects, even in very recent assessments. Published works share an approach by which technical details and formulation are discussed, and discuss modern formalisms with the aim to circulate research and technical achievements for use in professional, research, academic, and teaching activities.

This technical approach is an essential characteristic of the series. By discussing technical details and formulations in terms of modern formalisms, the possibility is created not only to show technical developments but also to explain achievements for technical teaching and research activity today and for the future.

The book series is intended to collect technical views on developments of the broad field of MMS in a unique frame that can be seen in its totality as an Encyclopaedia of MMS but with the additional purpose of archiving and teaching MMS achievements. Therefore, the book series will be of use not only for researchers and teachers in Mechanical Engineering but also for professionals and students for their formation and future work.

The series is promoted under the auspices of International Federation for the Promotion of Mechanism and Machine Science (IFToMM).

Prospective authors and editors can contact Mr. Pierpaolo Riva (publishing editor, Springer) at: pierpaolo.riva@springer.com

Indexed by SCOPUS and Google Scholar.

Satya N. Atluri · Igor Vušanović
Editors

Computational and Experimental Simulations in Engineering

Proceedings of ICCES 2020. Volume 2

Springer

Editors
Satya N. Atluri
Mechanical Engineering
Texas Tech University
Lubbock, TX, USA

Igor Vušanović
Mechanical Engineering
University of Montenegro
Podgorica, Montenegro

Preface

This book gathers the latest advances, innovations, and applications in the field of computational engineering, as presented by leading international researchers and engineers at the 26th International Conference on Computational & Experimental Engineering and Sciences (ICCES). ICCES covers all aspects of applied sciences and engineering: theoretical, analytical, computational, and experimental studies and solutions of problems in the physical, chemical, biological, mechanical, electrical, and mathematical sciences. As such, the book discusses highly diverse topics, including data-driven computational modeling; biomedical engineering and biomechanics; sound and vibration; computational and experimental materials and design; engineering and experimental sciences; modern computational methods; modern developments in mechanics of materials and structures; multi-scale and multi-physics fluid engineering; structural integrity and longevity; materials design and simulation. The contributions, which were selected by means of a rigorous international peer review process, highlight numerous exciting ideas that will spur novel research directions and foster multidisciplinary collaborations.

Contents

Effects of Multi-factor on Casing Stress Under Transient Force-Thermal Coupling

Hui Zhang, Jun Li, and Xin Zhang

Abstract In shale gas reservoir, casing eccentricity often occurs during casing cementing. In addition, the fracturing times are more and the downhole temperature field changes greatly during later reservoir reconstruction. Therefore, this paper establishes casing eccentric thermal-mechanical coupling model, determines the change law of temperature field at the heel end of casing and that of instantaneous stress of casing with time, and analyzes the effects of injection temperature, internal pressure, displacement and elastic modulus of cement sheath on the instantaneous stress of casing. Results indicate that the casing temperature decreases rapidly to nearly constant temperature during fracturing. The maximum stress of inner wall of casing increases with the increase of eccentricity angle. In the thermal-mechanical coupling simulation, the maximum stress of the casing first stabilizes briefly with time, then decreases sharply, and finally stabilizes at a constant temperature. The heat transfer coefficient increases with the increase of displacement, and then the maximum stress of casing also increases. The results can be used as reference for the analysis of influencing factors of casing damage during the fracturing of horizontal shale gas Wells.

Keywords Shale gas · Casing damage · Finite element · Casing eccentricity · Thermal-mechanical coupling

1 Introduction

At present, Large displacement casing fracturing methods are used with deep shale gas in Weirong area in China, and serious casing deformation problems appear in the fracturing process. By November 2018, 3,854 stages of 240 Wells had been

H. Zhang · J. Li (✉) · X. Zhang
MOE Key Laboratory of Petroleum Engineering, China University of Petroleum, Beijing, China

J. Li
China University of Petroleum-Beijing at Karamay, Karamay, China

fractured in the Weiyuan-Changning area, and 99 Wells had been damaged, with the casing deformation rate as high as 40%. In addition, the percentage of deformation points of casing increased significantly from toe end to heel end, and the percentage of deformation points of heel end was as high as 46.8%. After one or more stages of fracturing, the casing is damaged and deforms, which makes the fracturing tool unable to enter, and the bridge plug cannot be drilled and ground, thus affecting the later construction and production [1–3].

During cementing, casing eccentricity is caused by the weight of casing in inclined section or horizontal section, instability and collapse of the borehole walls, improper position of centralizer and other reasons [4, 5]. The influence of temperature field on casing stress needs to be taken into account because of the number of fracturing times, large pump pressure and large downhole temperature differences. Therefore, this paper establishes casing eccentric thermal-mechanical coupling model, determines the change law of temperature field at the heel end of casing and that of instantaneous stress of casing with time, and analyzes the effects of injection temperature, internal pressure, displacement and elastic modulus of cement sheath on the instantaneous stress of casing. The results can be used as reference for the analysis of influencing factors of casing damage during the fracturing of horizontal shale gas Wells.

2 Calculation Model of Temperature Field During Fracturing

During the fracturing of horizontal shale gas Wells, large displacement fracturing fluids are pumped from the wellhead into the wellbore and into the formation. During this process, there is heat transfer between the fracturing fluid and the casing, and the temperature of the fracturing fluid and the wellbore vary over time. In view of this phenomenon, many experts and scholars at home and abroad have modeled to simulate the variation law of wellbore temperature. Ramey Jr [6] used steady-state treatment for wellbore fluids and established an analytical wellbore heat transfer model. Eickmeier et al. [7] established a numerical method model of wellbore heat transfer, which can simulate the instantaneous temperature change of wellbore. Wang Hongxun [8] improved it and put forward the numerical calculation method of unsteady wellbore heat transfer. However, these models are mainly applied to tubing fracturing, not to large-displacement casing fracturing of shale gas Wells, and also ignore the heat generated by friction between large-displacement fracturing fluid and casing wall. Therefore, these two factors are taken into account in this paper to establish a wellbore temperature variation model during the volumetric fracturing of shale gas Wells.

In order to establish the calculation models of temperature field, the following assumptions are made [9]:

- The residual fluid temperature in the wellbore has reached equilibrium with the formation temperature.
- Ignored longitudinal heat transfer between casing and formation.
- The thermodynamic parameters used in the equation do not vary with temperature.
- With the casing center as the axis, the heat transfer mediums are symmetric, isotropic and homogeneous.
- Assume that the temperature (T_b) at depth $H = b$ below the surface is constant and unaffected by the surface temperature. At $H = a$, the original temperature of the formation changes linearly with the depth:

$$T_a = T_b + \alpha(a - b), a \geq b \tag{1}$$

α is geothermal gradients, °C/m.

Figure 1 is a schematic diagram of the wellbore and its surrounding rock during fracturing. According to the principle of energy conservation, the energy conservation equation of fluid unit $\pi r_0^2 \Delta H_j$ in the wellbore is:

$$Q\rho_0 C_0 T_{0,j-1}^{n+1} - Q\rho_0 C_0 T_{0,j}^{n+1} + 2\pi r_0 \Delta H_j U\left(T_{1,j}^{n+1} - T_{0,j-\frac{1}{2}}^{n+1}\right) + W_j$$

$$= \pi r_0^2 \Delta H_j \rho_0 C_0 \frac{\left(T_{0,j-\frac{1}{2}}^{n+1} - T_{0,j-\frac{1}{2}}^{n}\right)}{\Delta t} \tag{2}$$
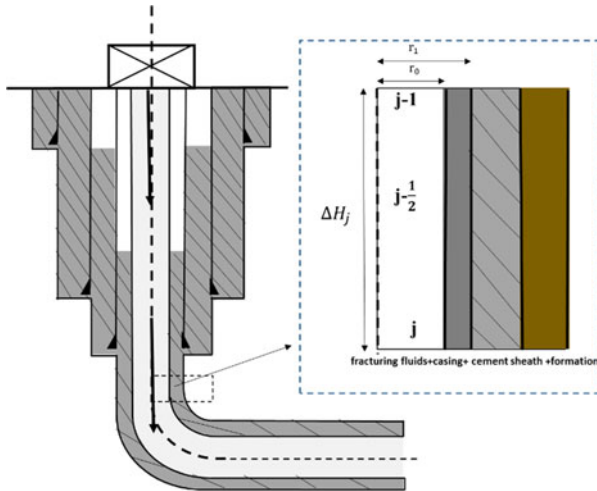
Fig. 1 Schematic diagram of heat conduction among casing, cement sheath and surrounding rock

$$W_j = \mu_{fj} \frac{\Delta H_j}{r_0} \frac{\rho_0}{2} \frac{Q^3}{\pi^2 r_0^2} \tag{3}$$

$$\frac{T_{0,j}^{n+1} - T_{0,j-1}^{n+1}}{2} = T_{0,j-\frac{1}{2}}^{n+1} \tag{4}$$

Q is the amount of fracturing fluid pumped, m^3/s; ρ_0 is the fracturing fluid density, kg/m^3; C_0 is specific heat of fracturing fluid, $J/(kg \cdot {}^{\circ}C)$; U is Convection heat transfer coefficient between fracturing fluid and casing surface, $W/(m^2 \cdot {}^{\circ}C)$; W_j is the heat generated by contact friction between fracturing fluids and wellbore walls, J.

the energy conservation equation of unit $\pi(r_1{}^2 - r_0{}^2)\Delta H_j$ on the casing is:

$$-2\pi r_0 \Delta H_j U \left(T_{1,j}^{n+1} - T_{0,j-\frac{1}{2}}^{n+1} \right) + 2\pi r_1 \Delta H_j K_1 \frac{\left(T_{2,j}^{n+1} - T_{1,j}^{n+1} \right)}{\frac{r_2-r_0}{2}}$$

$$= \pi\left(r_1{}^2 - r_0{}^2\right)\Delta H_j \rho_1 C_1 \frac{\left(T_{1,j}^{n+1} - T_{1,j}^{n} \right)}{\Delta t} \tag{5}$$

ρ_1 is the casing density, kg/m^3; C_1 is specific heat of casing, $J/(kg \cdot {}^{\circ}C)$; K is heat conductivity coefficient, $W/(m \cdot {}^{\circ}C)$.

3 Eccentric Casing—Cement Sheath—Formation Finite Element Model

3.1 Build FEM Numerical Model

In the cementing process, casing eccentricity is common in downhole, and formation-cement sheath-casing is regarded as a combined system. The following assumptions are:

- Assume that casing, cement sheath and formation are all elastomers.
- Assume that the components are closely connected and have no relative sliding.
- In the direction of perpendicular to the central axis of casing, the thermodynamic properties of casing, cement sheath and formation remain unchanged.

The center of the hole is located at point O, the X axis is the maximum horizontal geostress σ_H direction, the Y axis is the vertical geostress σ_v direction and the Z axis is the minimum horizontal geostress σ_h direction, that is, the borehole axis direction. The center point of the casing is set as O', the hole radius is R, the inner diameter of the casing is r_0, the outer diameter of the casing is r_1, the eccentric

distance is e and the eccentric angle is φ. In Fig. 2, select the cross section of the assembly at the heel end of the horizontal segment as the research object, according to the saint-venant's theorem, and establish a fully coupled finite element model of casing, cement sheath and formation with a size of 3 m × 3 m.

3.2 Parameter Settings

Take Weiye 23-1HF Well as an example. The depth of the well is 5555 m, the vertical depth is 3867 m, the length of the horizontal section is 1500.54 m, the length of the fracturing section is 1485.6 m, the surface temperature is 17.9 °C, the initial temperature of the fracturing fluid is 20 °C, the reservoir temperature of the Longmaxi section is 135 °C, internal pressure is 70 MPa, displacement is 10–20 m³/min, maximum displacement is 20 m³/min, and the fracturing time is 4 h.

The borehole diameter is 215.9 mm, the casing diameter is 139.7 mm, the casing steel grade is BG125V, the casing thickness is 12.7 mm, the elastic modulus is 210 GPa, and the Poisson's ratio is 0.3. The elastic modulus of the cement sheath is 10 GPa, and the Poisson's ratio is 0.17; the elastic modulus of the formation is 19.7 GPa, the Poisson's ratio is 0.246, and the formation pressure coefficient is 1.9 MPa/100 m. The specific stress parameters and bottom hole temperature calculation parameters are shown in Table 1 and Table 2.

3.3 Results Analysis

During the fracturing of Weiye 23-1HF Well, the fracturing pump pressure was 73–97 MPa, the displacement was 10 m³/min, the injection temperature was 20 °C, the geothermal gradient was 3.16 °C/100 m, and the fracturing was continued for 4 h.
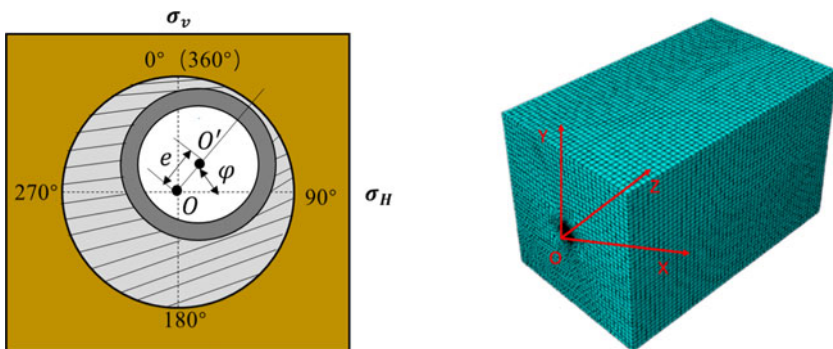
Fig. 2 Simplified casing—cement sheath—formation combination model and finite element model

Table 1 Stress parameters of Weiye 23-1HF well

Formation pressure coefficient	Surface temperature/°C	Reservoir temperature/°C	Rock type	Geostress /MPa			Stress type
				σ_H	σ_v	σ_h	
1.9	17.9	135	Brittle shale	48	35	29	Slip stress

Table 2 Bottom hole temperature calculation parameters

Name	Outer diameter/mm	Elastic modulus /GPa	Poisson's ratio	Density /kg·m^{-3}	Thermal conductivity/ W·(m·°C)$^{-1}$	Specific heat/J· (kg·°C)$^{-1}$	Thermal expansion coefficient/ 10^{-6}·°C^{-1}
Casing	139.7	210.0	0.300	7850	45.55	461	13.0
Cement sheath	215.9	10.0	0.170	2250	0.98	837	11.0
Formation	\	19.7	0.246	2300	1.59	1256	10.5
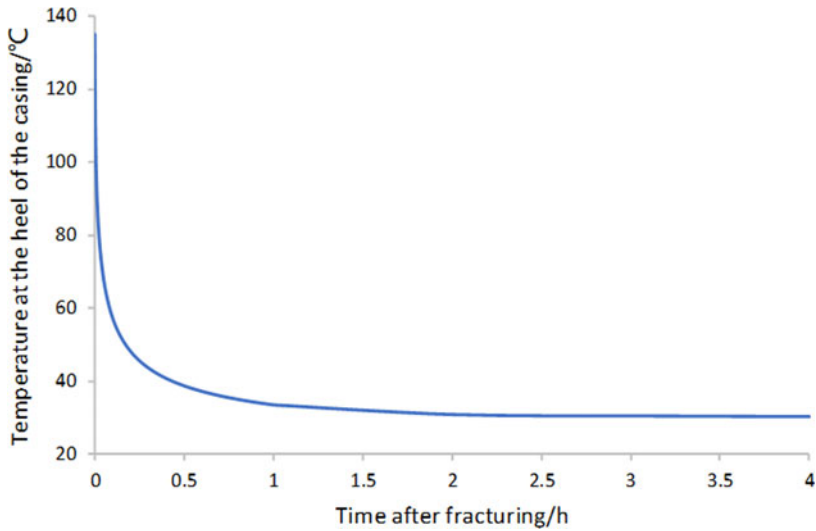
Fig. 3 Temperature curve of the inner wall of the casing at the heel end of the horizontal section during fracturing

According to the wellbore temperature field model, obtain the temperature change curve of the casing inner wall at the heel end of the horizontal section during fracturing by numerical calculation methods, as shown in Fig. 3.

When the eccentricity of the casing is 20 mm, study the extreme state of stress on the inner wall of the casing under eccentric angles of 0°, 30°, 60°, and 90°. The study shows that the extreme stress of the eccentric casing occurs on the inner wall of the circumferential angle corresponding to the eccentric angle. And the larger the eccentric angle, the greater the extreme value of the stress on the inner wall of the casing. When the eccentric angle is 90°, the maximum stress on the inner wall of the casing reaches 771 MPa. Figure 4 shows the stress calculation results with eccentric angles of 0° and 90° respectively.

Considering the thermal-mechanical coupling, when the eccentric angle is 90° and the eccentric distance is 20 mm, the change law of the maximum Mises stress on the inner wall of the casing with time is Fig. 5. It can be seen from the figure that the maximum Mises stress of the casing during the first 7.66 s is maintained at 771 MPa, then the stress drops sharply, and the stress is maintained at about 492 MPa after 2372 s. This is because immediately after the fracturing fluid is pumped in, the temperature difference between the fracturing fluid and the inner wall of the casing is large and temperature gradient of the casing change greatly, which leads to high temperature stress. With a period of fracturing, due to the ever-decreasing temperature difference between the fracturing fluid and the casing and the gradient of the casing temperature, the thermal stress is gradually reduced, which makes the maximum stress gradually decrease and eventually stabilizes.
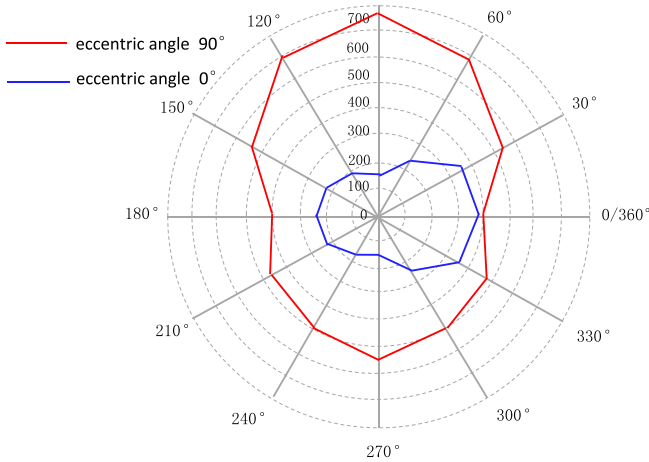
Fig. 4 Maximum stress of casing wall at eccentric angles of 0° and 90° respectively
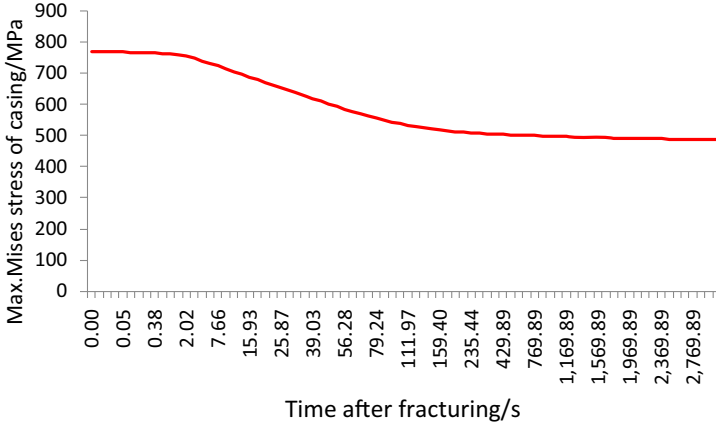
Fig. 5 Considering the thermal-mechanical coupling, change of maximum Mises stress of casing with fracturing time under eccentric angle of 90° and eccentric distance of 20 mm

4 Sensitivity Analysis of Influencing Factors

When the eccentric angle of the casing is 90° and the eccentric distance is 20 mm, under thermal-mechanical coupling, this paper discusses the influence of the elastic modulus of the cement sheath, the injection temperature, the displacement and the anisotropy of the geostress on the maximum stress on the inner wall of the casing by abaqus finite element software.

4.1 Effect of Injection Temperature

From Fig. 6 and Fig. 7, when the temperature of the injected fracturing fluid is 0 °C, the maximum casing stress is increased by 3.9% compared with the actual. Therefore, it can be seen that the lower the temperature of the fracturing fluid is and the faster the bottom hole temperature decreases, which results in an increase in the maximum casing stress.

4.2 Anisotropic Effects of Geostress

The fracturing of long shale gas horizontal wells is characterized by high pressure, large displacement and long time [10]. When the fracturing fluid is pumped into the formation, the non-uniformity of the formation stress field may be exacerbated. In order to study the effect of inhomogeneity of geostress on the maximum casing stress under temperature changes, the ratio $K = \sigma_v/\sigma_H$ of the overlying rock pressure to the horizontal maximum principal stress is defined as the reference value
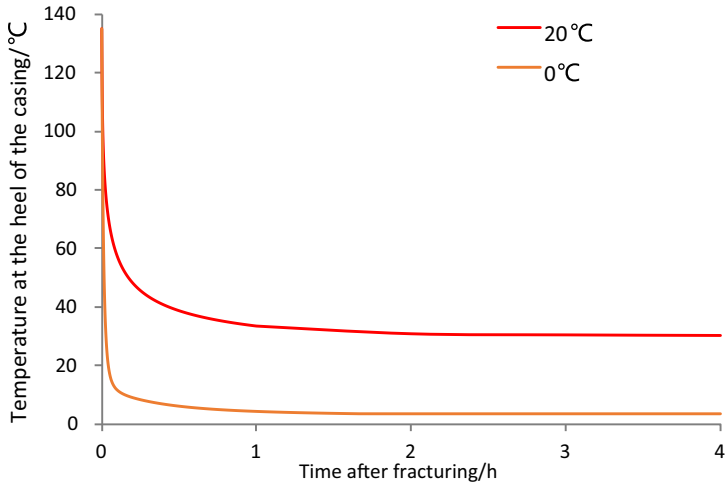
Fig. 6 Temperature curve of heel end when fracturing fluid is injected at different temperatures
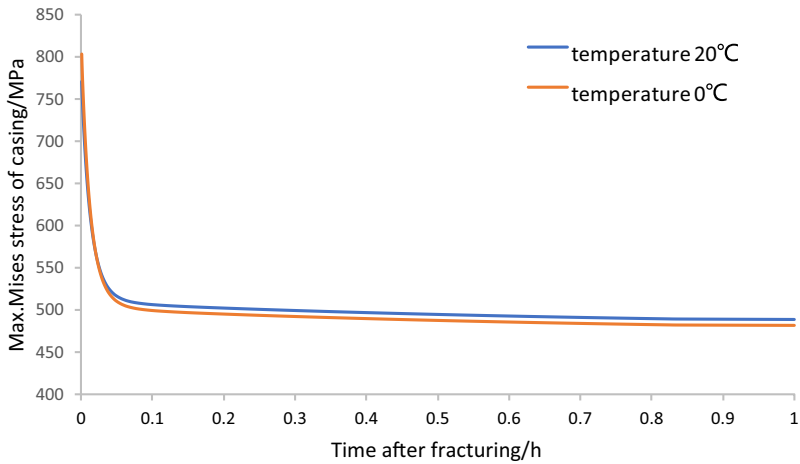
Fig. 7 Curve of maximum Mises stress at the heel end of horizontal well over time when fracturing fluids are injected at different temperatures

of the formation non-uniformity. And the larger the K value, the stronger the non-uniformity of the geostress. The actual minimum horizontal geostress of this well is 29 MPa, the maximum horizontal geostress is 48 MPa, the overburden pressure is 35 MPa, and K = 0.8. Now change σ_v so that K is 1, 1.2, 1.3, 1.5.
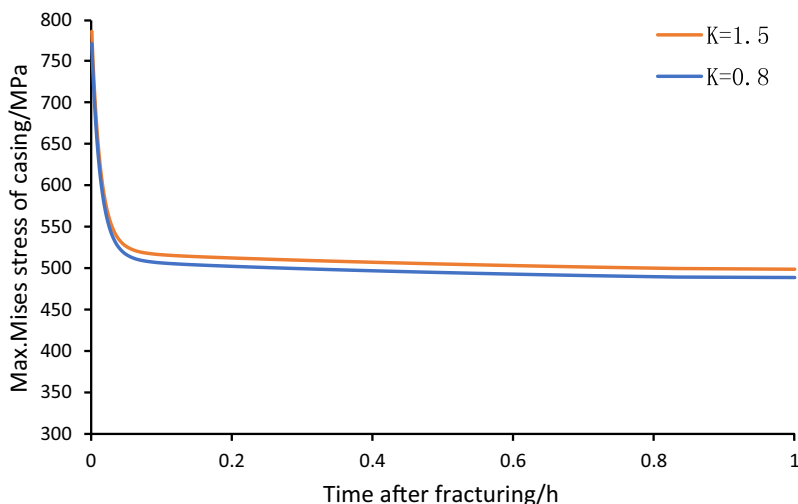
Fig. 8 Curve of maximum Mises stress at the heel end of horizontal well over time when K is different

It can be obtained from Fig. 8 that when K = 1.5, the maximum Mises stress of the casing is increased by 1.9% compared with the actual. And the non-uniformity of the geostress field increases, and the stress on the inner wall of the casing increases.

4.3 Effect of Cement Sheath Elastic Modulus

Figure 9 shows the influence of the elastic modulus of the cement sheath on the stress of the assembly. As shown in Fig. 9, the elastic modulus of the cement sheath increases and the maximum stress on the casing decreases. And when the elastic modulus of the cement sheath is increased to 20 GPa, the maximum Mises stress of the casing is reduced by 47.9%. Therefore, in actual construction, the elastic modulus of the cement sheath can be appropriately increased according to the properties of the stratum, which can effectively reduce the probability of casing damage during high-pressure liquid injection.

4.4 Effect of Internal Pressure

From Fig. 10, when the internal pressure is 80 MPa, the maximum casing stress is increased by 6.7% compared with the actual. It can be seen that under the condition
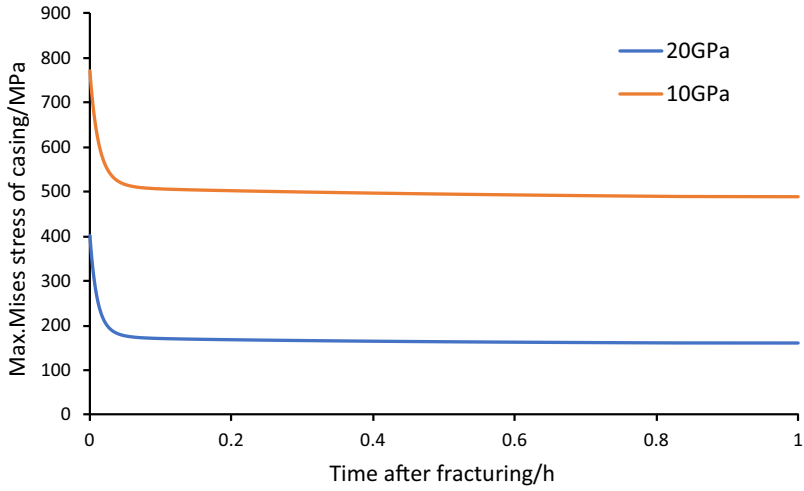
Fig. 9 Curve of maximum Mises stress at the heel end of horizontal well over time when the elastic modulus of the cement sheath is different
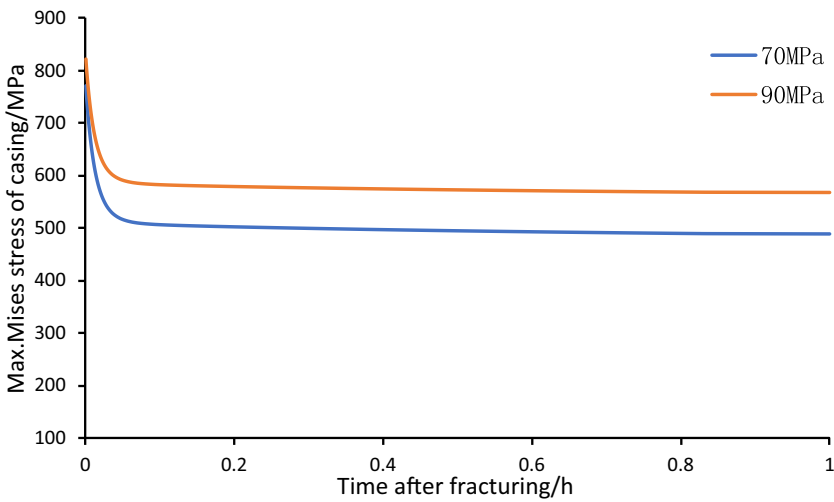
Fig. 10 Curve of maximum Mises stress at the heel end of horizontal well over time when the internal pressure is different

of thermal-mechanical coupling, the maximum Mises stress on the inner wall of the casing increases as the internal pressure increases.
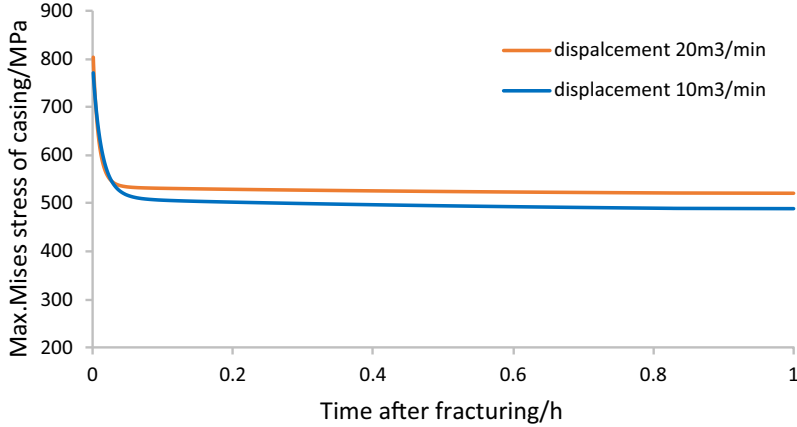
Fig. 11 Curve of maximum Mises stress at the heel end of horizontal well over time when the displacement is different

4.5 Effect of Displacement

The displacement affects the heat transfer coefficient between the fracturing fluid and the casing, which affects the change in casing temperature during the fracturing, and ultimately affects the maximum stress on the inner wall of the casing. From Fig. 11, as the fracturing fluid displacement increases, the maximum stress on the inner wall of the casing also increases. It can be seen that during the fracturing, the larger the displacement, the greater the maximum stress of the casing, and the greater the risk of casing damage.

5 Conclusion

1. Based on the actual construction data at the site, a transient temperature-pressure coupling model of an eccentric casing-cement sheath-rock combination considering the changes in the transient temperature field is established, and the change of casing maximum stress with temperature under different construction parameters is studied.
2. Obtained the law of temperature change at the heel of the horizontal section and the maximum stress of the inner wall of the casing over time during fracturing: It first stabilizes briefly, then drops rapidly, and finally stabilizes at a constant value.
3. The calculation results show: As the displacement of fracturing fluid increases, the heat transfer coefficient will increase, and casing stress will increase accordingly; the casing stress will increase as the temperature of the fracturing

fluid decreases; the elastic modulus of the cement sheath increase, and the maximum stress on the casing will decrease; the non-uniformity of the geostress field increases, and the stress on the inner wall of the casing will increases; reasonable construction pressure helps reduce the casing stress. Therefore, selecting a reasonable construction pump displacement, fracturing fluid injection temperature, and construction pressure can effectively reduce the temperature difference at the bottom of the well during fracturing and ensure the safety of the casing during fracturing.

References

1. Xue, C.: Technical advance and development proposals of shale gas fracturing. Pet. Drill. Tech. **39**(3), 24–29 (2011)
2. Dai, Q.: A tentative discussion on the cause of casing damage during the shale gas well cementing. Drill. Prod. Technol. **38**(3), 22–25 (2015)
3. Tian, Z., Shi, L., Qiao, L.: Research of and countermeasure or wellbore integrity of shale gas horizontal well. Nat. Gas. Ind. **35**(9), 70–76 (2015)
4. Wang, G., Chen, Z., Xiong, J.: Effect of casing eccentricity on casing strength under non-uniform load. J. Pet. Nat. Gas Technol. **34**(10), 105–107 (2012)
5. Li, J., Feng, S., Li, M.: The main problems and research progress affecting cementing displacement efficiency. Fault-Block Oil Gas Field **23**(3), 393–396 (2016)
6. Ramey, H.J., Jr.: Wellbore Heat Transmission. J. Pet. Technol. **14**(04), 427–435 (1962)
7. Eickmeier, J.R., Ersoy, D., Ramey, H.J.: Wellbore temperatures and heat losses during production or injection operations. J. Can. Pet. Technol. **9**(02), 115–121 (1970)
8. Wang, H., Li, P.: Numerical calculation method of wellbore temperature during hydraulic fracturing. Acta Petrolei Sinica **8**(02), 91–99 (1987)
9. Li, J., Chen, M., Liu, G., Zhang, H.: Elastoplastic analysis of casing, cement ring and surrounding. Acta Petrolei Sinica Rock **26**(6), 99–103 (2005)
10. Xi, Y., Liu, G., Li, J.: Study on casing transient stress under thermo-mechanical coupling effect. China Pet. Mach. **45**(6), 8–12 (2017)

Study of Supersonic Microjet
from a Square Convergent Nozzle
by the Mach-Zehnder Interferometer

**Tatsuya Sakanashi, Shinichiro Nakao, Yoshiaki Miyazato,
and Yojiro Ishino**

Abstract The three-dimensional density field of a supersonic microjet issued from a square convergent nozzle with a cross-sectional area of 1 mm $\times$ 1 mm at the exit is measured by the Mach-Zehnder interferometer with the finite fringe setting. The experiment is performed at a nozzle pressure ratio of 4.0 to produce underexpanded free jets. Interferogram analyses for reconstructing the density field of the microjet are performed using the convolution back projection (CBP) algorithm in which the analysis of the phase shift of the deformed fringe relative to the background fringe is carried out by the Fourier-transform method. A flow topology of the microjet is clarified with various visual representations.

Keywords Supersonic microjet · Square nozzle · Mach-Zehnder interferometer

1 Introduction

There has been considerable research on the subject of the dynamics of a supersonic microjet for the application of microscale devices, including a small satellite thruster in space engineering [1], a critical nozzle for obtaining mass-flow rate at a low Reynolds number [2], and a micro-propulsion nozzle [3]. A detailed comprehension of the flow characteristics through such devices requires information regarding the quantitative velocity, density, and temperature measurements in the flow field. Although, the structure of supersonic microjets was systematically studied by Scroggs & Settles [4], a detailed understanding of the flow structure of rectangular supersonic microjets is lacking in the current literature because of a challenging issue.

T. Sakanashi (✉) · S. Nakao · Y. Miyazato
The University of Kitakyushu, 1-1, Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan

Y. Ishino
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

For a rectangular supersonic jet operating at slightly off-design conditions, an analytical model predicting the jet flow properties has been proposed by Tam [5]. However, no exact solutions exist for a rectangular supersonic free jet with strong shock waves such as Mach shocks and intercepting shocks, or with shock wave-vortex interactions. On the other hand, as a computational fluid dynamic (CFD) analytical tool, the Reynolds averaged Navier-Stokes (RANS) simulations have been used to provide reasonably accurate results within a relatively short period of computational time [6]. The RANS models rely on additional physical approximations. Therefore, it is critical that the capabilities of these models are assessed before the numerical results are accepted. To validate a computational model, reliable experimental data sets are thus very important.

In the present study, a square convergent nozzle with an exit area of 1 mm 1 mm is used for obtaining the density field of a supersonic microjet that shows complex shock structures. To this aim, the Mach-Zehnder interferometer is utilized to investigate the three-dimensional density fields of a square supersonic microjet, which could be useful for the CFD community as the validation data against their numerical simulations of shock-containing square microjets.

2 Experimental Apparatus

The experiments were conducted in a blowdown compressed-air facility of the High-Speed Gasdynamics Laboratory at the University of Kitakyushu. A schematic diagram of the experimental apparatus with the Mach-Zehnder interferometer system is shown in Fig. 1. Ambient air is pressured by the compressor up to 1 MPa, and then stored in the high-pressure reservoir consisting of two storage tanks with a total capacity of 2 m^3 after being filtered and dried. The high-pressure dry air from the reservoir is stagnated in a plenum chamber as shown in Fig. 1, and then discharged into the atmosphere through a test nozzle. In the present experiment, the plenum pressure was controlled and maintained constant at a value of p_{os} = 407.6 kPa $\pm$ 0.5 kPa during the testing by a solenoid valve. As schematically shown in Fig. 2, a square convergent nozzle with 6 mm and 1 mm in height at the inlet and exit was used as a test nozzle. The nozzle wall contour from the inlet to exit was designed based on a sinusoidal curve so as to realize smooth uniform flows at the inlet and exit. The nozzle has a constant width of 5 mm over the entire length from the inlet to the exit. The experiment was carried out at a nozzle pressure ratio (NPR) of 4.0 within an accuracy of $\pm$1.0% to produce an underexpanded free jet with a Mach shock. The total temperature in the plenum chamber was equal to the room temperature (T_b = 300.3 K) within an accuracy of about 0.1 centigrade during the experiment. To obtain density fields in a shock-containing microjet, the quantitative flow visualization was made using the Mach-Zehnder interferometer system with a field of view of 50 mm diameter. The Reynolds number at the nozzle exit is Re_d = 6.06 $\times$ 10^4, which is calculated based on the assumption of an isentropic flow from the nozzle inlet to exit. The Mach-Zehnder interferometer is an optical
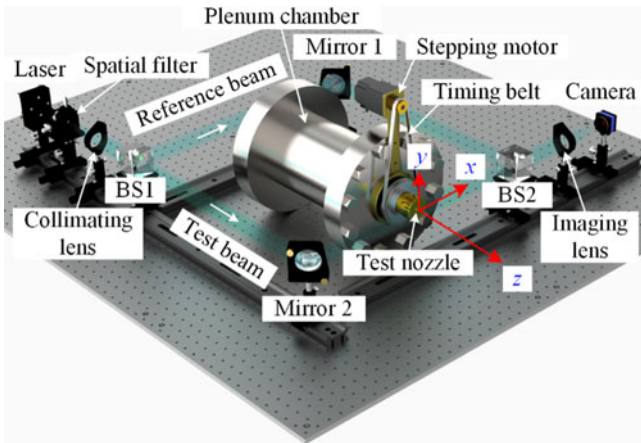
Fig. 1 Schematic drawing of experimental set up

instrument of high precision and versatility with the associated optical equipment that uses a blue semiconductor laser with a wavelength of 405 nm as a light source. The laser beam is expanded by a spatial filter to make a coherent beam before collimated into a parallel beam, and then is split into reference and test beams by a beamsplitter (BS1). The beamsplitter reflects roughly half of the intensity of the wavefront in one direction and transmits the rest in another direction. The reference beam passes through the still air and travels to another beamsplitter (BS2) after being reflected at Mirror 1. The test beam passes through a refracted index field produced by a free jet that is issued from a test nozzle and also travels to BS2 after being reflected at Mirror 2. The two beams are combined before being focused by the imaging lens and then produce interferograms on the recording medium in a digital camera.

3   Reconstruction of Jet Density Fields

3.1   Fourier Transform Method

When the test beam passes through a free jet with a variable refractive-index field, the background fringes are changed into the deformed fringe patterns because of the phase shift caused by the variations of the light speed as the beam passes through the test field. Typical profiles showing background and deformed fringe patterns are illustrated in Fig. 3. The parallel and equally spaced fringes shown as blue solid lines in Fig. 3(a) are also referred to as the wedge fringes. The interval b in Fig. 3(a) denotes the distance between two successive crests of the background fringes, and it is a function of the intersection angle between reference and test beams and the wavelength of the laser light used in the experiment. The red dashed line in Fig. 3(b)

Fig. 2 Schematic of test
nozzle

Part A to B: $y_w = 1.25\sin\left[\dfrac{\pi}{7.5}(z-3.75)\right]+1.75$

(a) Fringe patterns on recording medium   (b) Fringe intensity profiles at $z = z_0$
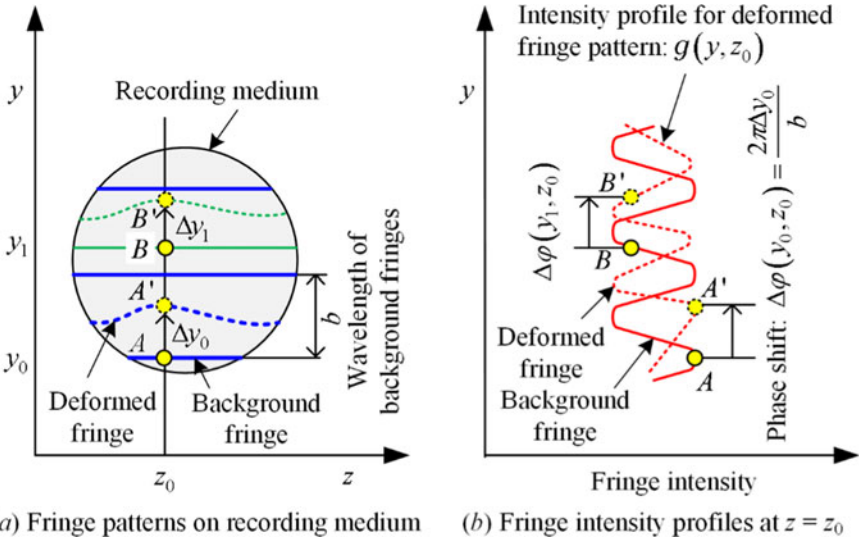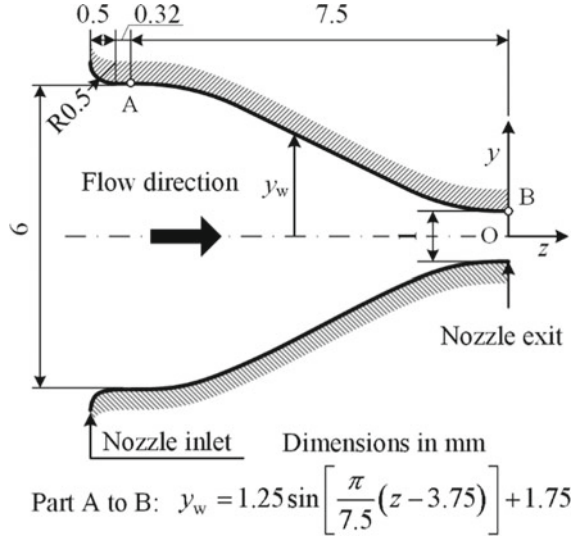
Fig. 3 Variation of background and deformed fringes by finite fringe method

shows the intensity profile $g(y, z_0)$ corresponding to the deformed fringe pattern at a
particular axial position z_0, and it can be given by Takeda et al. [7], Yagi et al.[8], and
Sugawara et al. [9]

$$g(y, z_0) = g_0(y, z_0) + g_1(y, z_0)\ \cos[k_0 y - \Delta\varphi(y, z_0)] \tag{1}$$

Here, the phase shift $\Delta\varphi(y, z_0)$ contains the desired information on the density field in the free jet, and $g_0(y, z_0)$ and $g_1(y, z_0)$ represent unwanted irradiance variations arising from the nonuniform light reflection or transmission when the test beam passes through a free jet, and $k_0 = 2/b$. The coordinates y and z form the vertical plane, which is perpendicular to the test beam propagation direction, and the x-axis is taken as the direction, in which the test beam propagates after being reflected at Mirror 2 as shown in Fig. 1.

Equation (1) can be rewritten in the following expression:

$$g(y, z_0) = g_0(y, z_0) + c(y, z_0) \ \exp(ik_0y) + c*(y, z_0) \ \exp(-ik_0y) \qquad (2)$$

With

$$c(y, z_0) = g_1(y, z_0) \ \exp\left[-i\Delta\varphi(y, z_0)\right] \qquad (3)$$

where i is the imaginary unit and the asterisk * denotes the complex conjugate. The Fourier transform of Eq. (2) with respect to y is given by

$$G(k, z_0) = G_0(k, z_0) + C(k - k_0, z_0) + C*(k + k_0, z_0) \qquad (4)$$

where the capital letters denote the Fourier transforms of the respective primitive functions, and k is the spatial wavenumber in the y direction. Since the spatial variations of $g_0(y, z_0)$, $g_1(y, z_0)$, and $\Delta\varphi(y, z_0)$ are slow compared with the spatial frequency k_0 when the interval between fringes is sufficiently small, the Fourier spectra in Eq. (4) are separated by the wavenumber k_0 and have the three independent peaks as schematically shown in Fig. 4(a). We make use of either of these two spectra on the carrier, say $C(k - k_0, z_0)$, and translate it by k_0 on the wavenumber axis toward the origin to obtain $C(k, z_0)$ as shown in Fig. 4(b). The unwanted background variation $G_0(k, z_0)$ has been filtered out in this stage by the pertinent bandpass filter. Applying the inverse Fourier transform of $C(k, z_0)$ with respect to the k to obtain $c(y, z_0)$ defined by Eq. (3), and taking the logarithm of Eq. (3) leads to

$$\ln c(y, z_0) = \ln g_1(y, z_0) - i\Delta\varphi(y, z_0) \qquad (5)$$

Consequently, the phase shift $\Delta\varphi(y, z_0)$ in the imaginary part of (5) can be completely separated from the unwanted amplitude variation $g_1(y, z_0)$ in the real part. It is necessary to make the interval b between parallel fringes as narrow as possible to avoid islands in the fringe pattern. Wide fringes move farther away from the original location than narrower fringes, and as a result, they cover regions, which are considerably different from the optical retardations. This results in the creation of islands [10].

(a) Separated Fourier spectra

(b) Parallel translation of single spectrum selected
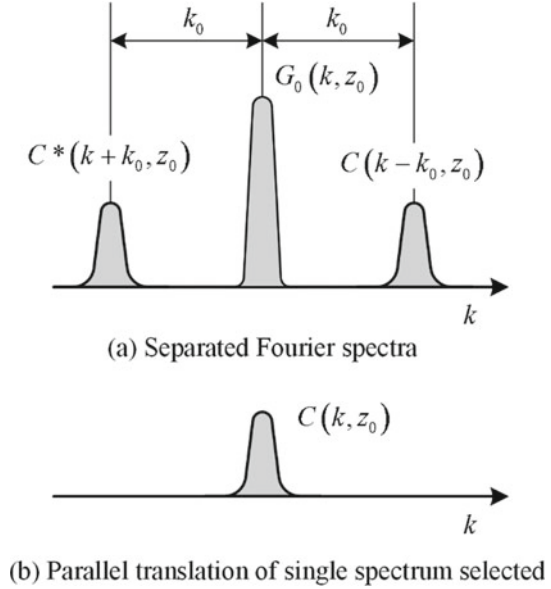
3.2 Convolution Back Projection Method

Figure 5 shows a light ray travelling in the jet cross-section (*x-y* plane) at a fixed streamwise location $z = z_0$ of an asymmetric jet issued from a test nozzle and the phase shift from the original location of an interferogram on a recording medium where the *x*, *y*, *z* rectangular Cartesian coordinate system is used and the *z*-axis is parallel to the streamwise direction, which is also perpendicular to the *x,y* plane. The refractive index or density fields will now be investigated for a cross-section of *z* = constant. Accordingly, we omit the independent variable *z* in the subsequent analysis to reconstruct two-dimensional density fields.

As shown in Fig. 5, let us consider the rotated coordinates (*s*, *t*) inclined at an angle θ away from the fixed-original coordinates (*x*, *y*) where $n(x, y)$ and n_a in Fig. 5 indicate the refractive index in the jet cross-section and that of the surrounding air, respectively. Then, for a light ray traveling in the *s* direction with an offset of *t* from the axis *s*, the optical path difference $\Lambda_\theta(t)$ of the test beam which passed through the field with and without the jet, can be expressed by the Radon transform

$$\Lambda_\theta(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta n(x, y)\delta(x\sin\theta - y\cos\theta + t)dxdy \tag{6}$$

where δ denotes the Dirac delta function and $\Delta n(x, y) \equiv n(x, y) - n_a$.

The optical path difference $\Lambda_\theta(t)$ is taken for a range of various angles from $\theta = 0°$ to $180°$. The task of tomographic reconstruction in the present investigation
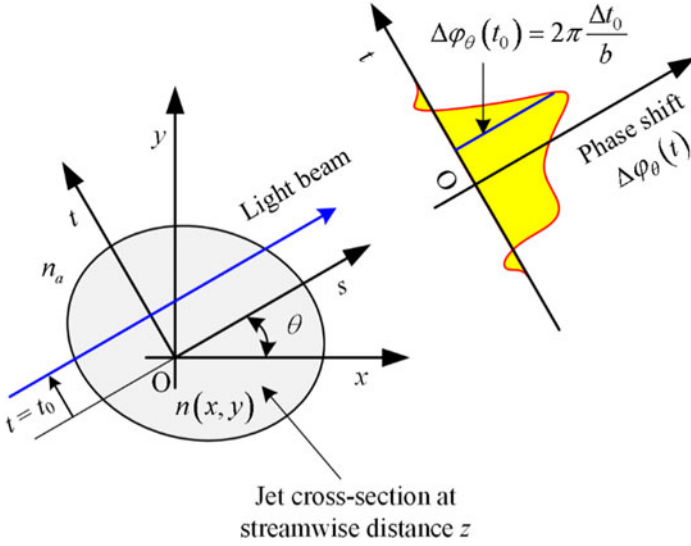
Fig. 5 Light beam passing through asymmetric refractive index field

is to find $\Delta n(x, y)$ based upon a given knowledge of $\Lambda_\theta(t)$ and then density fields can be obtained through a well-known linear relation between refractive index and density, as described later.

The transfer from the (x, y) to (s, t) coordinate systems for the integral of the right-hand side of Eq. (6) leads to

$$\Lambda_\theta(t) = \int_{-\infty}^{\infty} \Delta n(s \cos \theta - t \sin \theta, s \sin \theta + t \cos \theta) ds \tag{7}$$

After the one-dimensional Fourier transform of Eq. (7) with respect to the t variable and using the transform from the (s, t) to (x, y) coordinate systems, it reduces to a relation,

$$
\begin{aligned}
FT[\Lambda_\theta(t)] &\equiv \int_{-\infty}^{\infty} \Lambda_\theta(t) \exp(-2\pi i t \zeta) dt \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta n(x, y) \exp[-2\pi i(ux + vy)] dx dy
\end{aligned} \tag{8}
$$

with $u = \zeta\sin\theta$ and $v = \zeta\cos\theta$.

Equation (8) expresses the two-dimensional Fourier transform of $\Delta n(x, y)$. Hence, the inverse Fourier transform of Eq. (8) with $u = \zeta \sin\theta$ and $\upsilon = \zeta \cos\theta$ and after some rearrangement, it reduces to

$$\Delta n(x, y) = \int_0^\pi \left\{ \int_{-\infty}^\infty |\zeta| FT[\Lambda_\theta(t)] \exp(2\pi i t\zeta) d\zeta \right\}_{t=-\sin\theta x + \cos\theta y} d\theta \qquad (9)$$

The shift magnitude Δt of the deformed fringe relative to the background fringe at a fixed location ($t = const.$) on the recording medium can be given by

$$\Delta t = b\frac{\Lambda_\theta(t)}{\lambda_0} \qquad (10)$$

and the phase shift $\Delta\varphi_\theta(t)$ of the deformed fringe on the recording medium can be
expressed by

$$\Delta\varphi_\theta(t) = 2\pi \frac{\Delta t}{b} \qquad (11)$$

Accordingly, combination of Eqs. (9)–(11) with the Gradstone-Dale relation yields the jet two-dimensional density field at a cross-section perpendicular with respect to a fixed streamwise location z as follows:

$$\rho(x, y) = \rho_a + \frac{\lambda_0}{2\pi K} \int_0^\pi \left\{ \int_{-\infty}^\infty |\zeta| FT[\Delta\varphi_\theta(t)] \exp(2\pi i t\zeta) d\zeta \right\}_{t=-\sin\theta x + \cos\theta y} d\theta \quad (12)$$

This equation can expressed in alternative form with the convolution theorem as follows:

$$\rho(x, y) = \rho_a + \frac{\lambda_0}{2\pi K} \int_0^\pi [k(t) * \Delta\varphi_\theta(t)]_{t=-\sin\theta x + \cos\theta y} d\theta \qquad (13)$$

where

$$k(t) \equiv FT^{-1}[|\zeta|] = \int_{-\infty}^\infty |\zeta| \exp(2\pi i t\zeta) d\zeta \qquad (14)$$

and the symbol * denotes convolution between $k(t)$ and $\Delta\varphi_\theta(t)$.

Therefore, with the use of Eq. (13) as well as the Fourier transform method for the phase shift analysis in Sect. 3.1, the two-dimensional density fields at various streamwise locations z can be stacked together to from the three dimensional density field in a shock-containing jet. In the present study, considering the nozzle symmetry, only one-quarter between $\theta = 0$ and $45°$ was visualized in an equal interval of $5°$ to reconstruct the jet three-dimensional density field.

4 Results and Discussion

The density contour plots at the cross-section including the jet centerline are depicted in Figs. 6(a) and 6(b) where the normalized contour levels with an interval of 0.1 are shown at the top, and the spatial resolution in the experimental density map is around 4 µm. The density contour plot illustrates the various flow features of the shock cell structures quantitatively, such as the shape and size of the expansion and compression regions, the shock cell intervals, the jet boundaries, the outer shear layers near the jet boundary, and slip lines from the triple points of the Mach shock with a small Mach stem. For the jet just downstream of the nozzle exit, the jet boundary at the symmetry plane expands outwards (Fig. 6(a)), but that at the diagonal plane converges gradually toward the jet centerline (Fig. 6(b)).
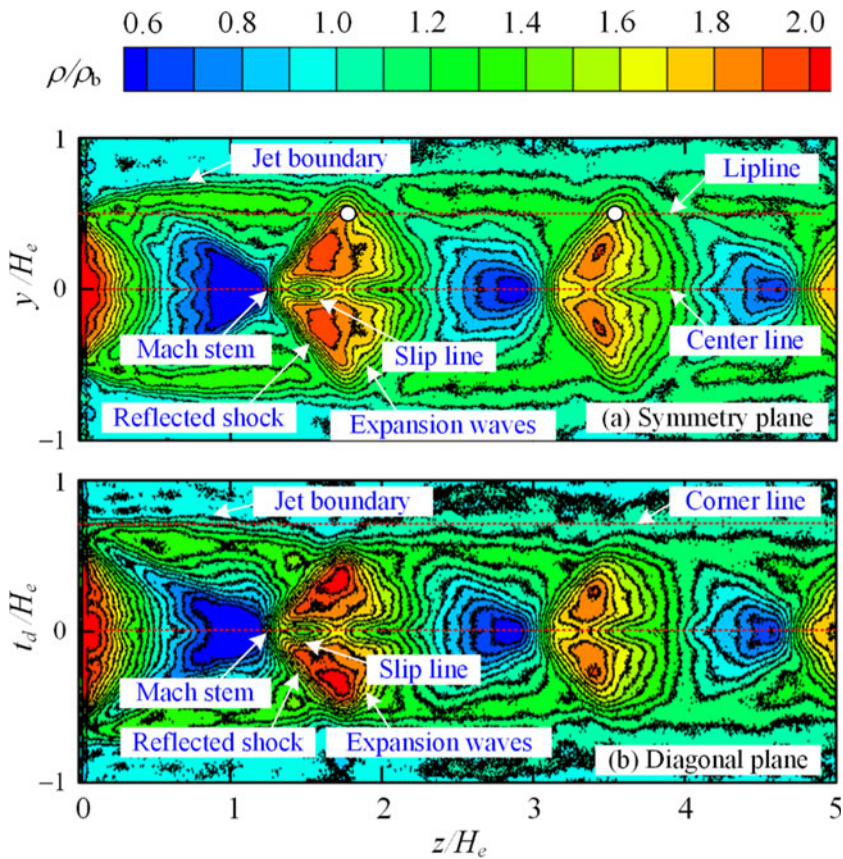
Fig. 6 Density contour plots at cross-sectional area including jet centerline

The normalized density profiles along the streamwise direction are shown in Figs. 7(a)–(c) where the theoretical density at the nozzle exit plane, which is estimated based on the assumption of the one-dimensional isentropic flow from the nozzle inlet to the exit, is shown as a leftward arrow on the vertical axis in Fig. 7(a) as a reference. The centerline density profile in Fig. 7(a) shows a representative distribution appearing in the shock-cell structure in an underexpanded free jet, i.e., the density decreases rapidly below the ambient level by expansion waves originating from the nozzle lip. This shows a sharp increase just downstream of a local minimum at z/H_e = around 1.1 and then the flow expansion and compression are repeated downstream. The local minima in the density profile gradually increase with increasing streamwise distance. Two distinct spikes (z/H_e = around 1.3 and 1.7) can be recognized just downstream of the density jump caused by the Mach stem in the first shock-cell. A similar waveform can be observed in the density
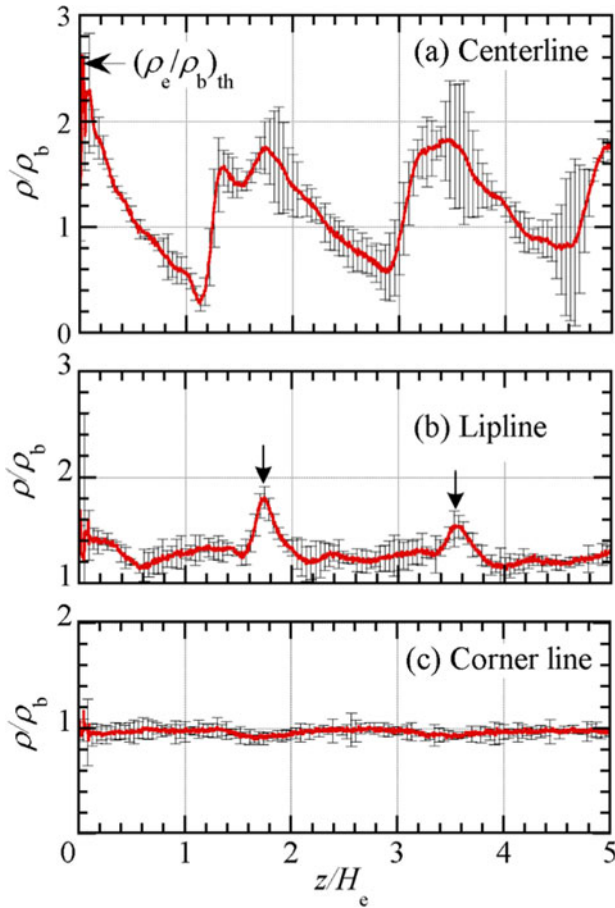
Fig. 7 Streamwise density profiles of square supersonic microjet

profile along the jet centerline obtained by the rainbow schlieren deflectometry of Takano et al. [11]. As shown with the downward arrows in Fig. 7(b), the density profile at lipline ($x/H_e = 0$, $y/H_e = 1.0$) exhibits two distinct bumps due to the reflection of the shock waves at the jet free boundary. The corresponding locations are illustrated with the open symbols in Fig. 6(a). The density profile at the corner line (Fig. 7(c)) shows almost constant variation in the streamwise direction because it is located outside the jet free boundary (see Fig. 6(b)).

5   Concluding Remarks

The Mach-Zehnder interferometer system with the finite fringe method was applied for the first time to "quantitatively" measure the density profiles over the whole density field of a square shock-containing microjet. In this experiment, a nozzle pressure ratio was constant at 4.0 to produce an underexpanded free jet with a Mach shock in the first shock cell. The microjet structures were demonstrated with the density contour plot at the cross-section including the jet centerline. It was shown that the experimental density contour plot illustrates the various flow features of the square microjet, including the shape and size of the Mach shock as well as the expansion and compression regions, the shock cell intervals, the jet boundaries, the slip streams produced from the triple points of the Mach shock and the outer shear layers near the jet boundaries. The spatial feature of the complex shock structure of the square supersonic microjet can be clearly exhibited at a high spatial resolution of 4 µm.

References

1. Lempert, W.R., Boehm, M., Jiang, N., Gimelshein, S., Levin, D.: Comparison of molecular tagging velocimetry data and direct simulation Monte Carlo simulations in supersonic micro jet flows. Exp. Fluids **34–3**, 403–411 (2003)
2. Nakao, S., Takamoto, M.: Choking phenomena of sonic nozzles at low Reynolds numbers. Flow Meas. Instrum. **11–4**, 285–291 (2000)
3. Louisos, W.F., Hitt, D.L.: Heat transfer & viscous effects in 2D & 3D supersonic micro-nozzle flows. AIAA Paper, No. 2007-3987 (2007)
4. Scroggs, S.D., Settles, G.S.: An experimental study of supersonic microjets. Exp. Fluids **21**, 401–409 (1996)
5. Tam, C.K.W.: The shock-cell structures and screech tone frequencies of rectangular and non-axisymmetric supersonic jets. J. Sound Vib. **121–1**, 135–147 (1988)
6. Franquet, E., Perrier, V., Gibout, S., Bruel, P.: Free underexpanded jets in a quiescent medium: a review. Prog. Aerosp. Sci. **77**, 25–53 (2015)

7. Takeda, M., Ina, H., Kobayashi, S.: Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. J. Opt. Soc. Am. **72–1**, 156–160 (1982)
8. Yagi, S., Inoue, S., Nakao, S., Ono, D., Miyazato, Y.: Optical measurements of shock waves in critical nozzles at low Reynolds numbers. J. Flow Control Meas. Visual. **5–2**, 36–50 (2017)
9. Sugawara, S., Nakao, S., Miyazato, Y., Ishino, Y., Miki, K.: Three-dimensional reconstruction of a microjet with a Mach disk by Mach-Zehnder interferometers. J. Fluid Mech. **893**, A25 (2020)
10. Winckler, J.: The Mach interferometer applied to studying an axially symmetric supersonic air jet. Rev. Sci. Instrum. **19–5**, 307–322 (1948)
11. Takano, H., Kamikihara, D., Ono, D., Nakao, S., Yamamoto, H., Miyazato, Y.: Three-dimensional rainbow schlieren measurements in underexpanded sonic jets from axisymmetric convergent nozzles. J. Thermal Sci. **25–1**, 78–83 (2016)

Joint Probability Density Immune Algorithm for Multi-satellite Earth Observation Scheduling

Lili Ren, Xin Ning, Shi Chao Ma, and Jian Ping Yuan

Abstract As a multi-objective combinatorial optimization problem, multi-agile satellite scheduling of earth observation, which with long observation time and multiple observation windows under complex constraints, has been a hotspot research problem in recent years. As a scarce resource, improving the in-orbit utilization of satellites and optimizing the imaging observation plan have been the goals of many researchers. For small scale task scheduling, the optimal solution is obtained by precise algorithm. However, for the large-scale task scheduling, it is difficult to find the exact solution, and the precise algorithm no longer has the advantage. Intelligent algorithms have been widely studied to obtain approximate optimal solutions. Genetic algorithm and immune algorithm are the most widely used intelligent algorithm. But the immune algorithm follows the evolutionary mode of the genetic algorithm, when confronted with large-scale task scheduling, the evolution operations such as crossover and mutation tend to make the solution fall into local optimization easily, and the randomness of these two operations make the scheduling time increases. Therefore, the traditional evolutionary approaches not only resulted in a waste of resources but also extended the evolutionary generations. In this paper we developed an improved immune algorithm, which the evolutionary idea of probability distribution instead. The specific evolution process are as follows. First, discretized the observation time of the all tasks, then coded all observation periods of tasks as antibody genes. Secondly, assigned an average probability value to the different observation periods of tasks, so the initial joint probability density matrix was constructed. Finally, updates the joint probability

L. Ren (✉) · X. Ning · S. C. Ma · J. P. Yuan
National Key Laboratory of Aerospace Flight Dynamics,
Northwestern Polytechnical University, Xi'an, China
e-mail: liliren@mail.nwpu.edu.cn

X. Ning
e-mail: ningxin@nwpu.edu.cn

S. C. Ma
e-mail: shichaoma@nwpu.edu.cn

J. P. Yuan
e-mail: jyuan@nwpu.edu.cn

density matrix until convergence according to the fitness value of antibodies, then the optimal scheduling solution was obtained. The simulation results show that compared with the genetic algorithm, both scheduling time and scheduling efficiency have a significant improvement.

Keywords Probability distribution · Immune algorithm · Multi-satellite

1 Introduction

As the key research object of the new generation of earth observation satellite, Agile earth observation satellite (AEOS) is of great economic and military value and has become an important way to obtain information in many fields, such as agriculture and forestry planning, disaster monitoring, resource exploration and military reconnaissance. For the excess mission observation requirements and the relatively scarce satellite resources, making a reasonable scheduling plan is helpful to improve the utilization rate of satellites and meet the needs of users to the maximum extent.

With the development of Agile satellites in recent years, many scholars have studied the scheduling problem of AEOS. M. Lemaitre defined the agile satellite scheduling problem for the first time and simplify the scheduling problem, then solved the problem by local search algorithm, constraint programming algorithm, dynamic programming algorithm and greedy algorithm [1]. According to different observation requirements and task scale, the study can be divided into single AEOS scheduling and multiple AEOS scheduling. Such as E. J. Kuipers, X. Wang and X. Liu respectively solved the single satellite scheduling problem by using simulated annealing algorithm, fast approximate scheduling algorithm and adaptive large neighborhood search algorithm [2–4]. And the agile satellite constellation scheduling problem has been studied based on priority heuristic algorithm and ant colony optimization meta-heuristic algorithm by P. Wang and E.V. Ntagiou [5, 6]. At the level of scheduling algorithm, the research has undergone many evolutions. R. Xu solved the satellite scheduling problem by using the priority-based sequential construction algorithm [7]. N. Bianchessi proposed a deterministic construction algorithm with forward-looking and retrospective features [8]. In addition, many hybrid algorithms have been studied, such as enumeration search algorithm hybrid the tabu search algorithm [9], and forward construction algorithm hybrid three pruning strategies [10]. Other works related to the hybrid algorithms also solved the AEOS scheduling problems [11–13]. Furthermore, there were some studies on intelligent algorithms, such as different improvements of genetic algorithms to solve satellite scheduling problems [11, 12, 14–19].

However, for the above intelligent algorithms, especially for different improvements of genetic algorithms, the search process has the disadvantages of slow convergence speed and easy to fall into local optimization in large-scale task scheduling. In view of the above shortcomings, this paper proposes an improved

immune algorithm, the joint probability density immune algorithm (JPDI), to solve the large-scale task scheduling problem of multi-agile satellites. The algorithm has the following advantages. Firstly, the optimal evolution generations have greatly reduced, and the scheduling efficiency also improved. Secondly, under the same iterations, the scheduling results were obviously better than the traditional evolutionary ways. Thirdly, scheduling process do not easily fall into local optimization.

The following contents are divided into the following parts. The second part is the description of the problem and the construction of the model, the introduction of the algorithm and the solution method in the third part, the fourth part is the simulation results, and the conclusion and declaration in the fifth part.

2 Problem Description

In this study, we considered the process of observing multiple imaging tasks on the earth's surface by multiple agile satellites to develop an optimal imaging plan. The simulated observation process is shown in Fig. 1. Since agile satellite has three rotational degrees of freedom, the observation efficiency has greatly improved. At the same time, the solution space of the scheduling scheme increases exponentially, which increases the difficulty of the solution process.

Based on this understanding, we detail the scheduling process and the variables or constraints involved. In general, agile satellites have multiple visible time windows (VTW) for each task, when there are M AEOS observing N tasks, if task i has been arranged to be observed by the kth VTW under satellite j, then $x_{ijk} = 1$, otherwise $x_{ijk} = 0$. The start time of the kth VTW of task i under satellite j is expressed as ws_{ijk} and the end time is we_{ijk}. In the actual observation, the length of VTW is much longer than the actual observation time window (OTW). If the observation time length of task i is expressed as dur_i, the start time of OTW is st_i,

Fig. 1 Simulation of the earth observation process of three agile satellites orbiting the earth
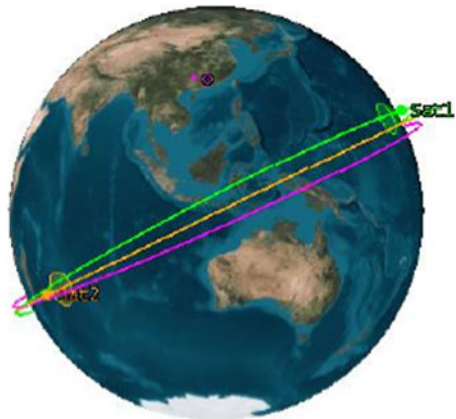
then the end time is $et_i = st_i + dur_i$, and must satisfy that $ws_{ijk} \leq st_i \leq et_i \leq we_{ijk}$. Obviously, the image quality observed in different OTW is different, and the image quality obtained in the middle of each VTW is set to be the best. q_i is used to represent the quality of the observation image of task i to measure the distance between OTW and the midpoint of the corresponding VTW. Each task has an observation priority level p_i. If the user has a demand for stereo observation of a task i, it is denoted as $ste_i = 1$. We specify that two observations with an interval of three times OTW of the task are the observation of completing the stereo task. If tasks i and j are executed continuously, then denoted as $y_{ij} = 1$. Moreover, the observation start time of task j is greater than the sum of the observation end time of task I and the attitude angle conversion time d_{ij} between the two tasks. In addition, if there are clouds in the observation process, the observation will be affected. Note the time when the n_i th clouds begin to appear is cs_i and the time when they stop is ce_i. Our ultimate goal is to maximize the total return and also called the fitness in the combinatorial probability density function:

$$\max \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{L_i} x_{ijk} p_i q_i \tag{1}$$

Assume

In the scheduling process of multi-AEOS, in addition to the oversubscribed observation tasks and the different observation needs from different users, the maneuvering constraints of satellites make the scheduling process very complicated. Since the purpose of our study concentrates more on the task planning level, the simplified model of satellite is adopted under the following assumptions in this paper:

1. The satellite has enough energy for mission observation.
2. The storage capacity of the satellite is sufficient for the satellite to operate until it reaches the ground receiving station to transmit the image down.
3. Set the stereo observation task as the highest observation priority.

The key of this study lies in how to obtain an approximate optimal scheduling plan that maximizes the total scheduling profit and minimizes the scheduling time. This paper will solve this problem in the following part.

3   Scheduling Algorithm

Based on the above problem description, we will give the solution process in this part. First, pre-processing the task, then constructing the combined probability density immune algorithm for the multi-AEOS scheduling problem, Immune algorithm is an intelligent computing method inspired by biological immune

system. Self-organizing and self-learning ability and powerful information processing ability of immune make it effectively applied in many fields such as intelligent control pattern recognition, optimization design and machine learning [20]. Considering that the immune algorithm inherits the disadvantages of the genetic algorithm, we will replace the crossover and mutation evolution strategy in the traditional immune algorithm with the combined probability density estimation strategy, and the performance of the combined probability density immune algorithm should be further improved. And finally presenting the general flow of the algorithm.

3.1 Scheduling Preprocessing

Please Scheduling preprocessing refers to select an observation satellite for each task to be observed and determine a VTW under that satellite. Then the first step is to select an AEOS for each task based on the total observed collision time heuristic. The second step is to choose an VTW for each observation task based on the equal opportunity heuristic operator which means ignore the difference in the number of Windows. The conflict calculation function is as follows:

$$\sum_{i=1}^{t_2} \sum_{j=1}^{t_1} c_{ij} \tag{2}$$

t_1 is the number of VTW under task i, t_2 is the number of tasks assigned to satellite. c_{ij} refers to the length of OTW overlap between the two tasks.

And the tasks with closed observation windows are removed from the database, which can reduce the total computation time.

3.2 JPDI Algorithm Description

After the scheduling preprocessing above, the process of the JPDI algorithm scheduling is summarized as shown in Table 1.

3.3 JPDI Algorithm

In order to encode the task into the antibodies required by the algorithm, we need to divide all the tasks into m segments corresponding to the OTW of the task according to the VTW selected by the above preprocessing process. Take task i as an example to introduce the specific segmentation method, first subtract $dur_i/2$

Table 1 The flow of combined probability density immune algorithm

Combined Probability Density Immune Algorithm
Input: Learning rate α, Group size T, number of iterations C, visible observation window VTW.
Output: optimal solution s^*
Begin: i =1. // Evolution generations count
Task execution time matrix S
Task end execution matrix E
Image quality matrix Q
Initial population G^0
Initial population quality matrix Q'
Initial probability matrix D
Excellent group database matrix $E = G^0$ // Scheduling initialization
Repeat: $G^{i+1} \leftarrow$ n_pop(m, G^i, E, D) // Create a new population
$\quad\quad Q'^{(i+1)} \leftarrow$ im_group quality(G^{i+1}, $Q'^{(i)}$) // Calculate new population's quality
$\quad\quad F_v \leftarrow$ im_affinity(G^{i+1}, $Q'^{(i+1)}$, S, E) // Fitness affinity calculation
$\quad\quad G_v \leftarrow$ im_concen($G^{(i+1)}$) // Calculate antibody concentration
$\quad\quad R_v = \dfrac{F_v}{G_v}$ // Calculate the expected reproduction rate
$\quad\quad [E, F] \leftarrow$ n_exc(G^{i+1}, $Q'^{(i+1)}$, R_v, E, D)
$\quad\quad\quad\quad$ // Calculate excellent individual database and frequency matrix
$\quad\quad D^{i+1} \leftarrow$ im_update prob(D^i, F, α) // Update probability matrix
while($i > C$)
output S^*, optimal solution
end

before and after the corresponding VTW, then divide the rest of the period into $m - 1$ segments, all the segmentation points and endpoints in this segment have a total of m, these m points are the midpoints of m OTWs for task i. The start time, end time and corresponding observation quality of m OTWs respectively corresponding to n tasks were respectively stored into the matrix S, E and Q. The quality calculation is measured by the distance between the midpoint of each OTW and the corresponding VTW midpoint, and If they overlap, then the mass equals one.

Thus, the coding method is introduced. n tasks are randomly assigned the integer value between one and m according to the task order, which represents the OTW selected for the corresponding task. And repeat several times to form an initial population P^0. Then the quality of the corresponding antibody gene in the population was calculated and put into the matrix Q. Finally, we build an initial probability density matrix D, let it be a discrete probability density matrix with uniform distribution, which is of n rows and m columns with all same elements $1/m$.

Now we describe the algorithm update iteration strategy. In the population update iteration process, a matrix E is required to store better antibodies, which is assigned as $E = P^0$. And we use the expected reproductive rate to evaluate the quality of antibodies. It is defined by

$$R_v = \frac{f_v}{g_v} \tag{3}$$

f_v is the affinity between antibody and antigen, and it is defined by

$$f_v = \frac{1}{1 + 1/\textit{fitness}} \tag{4}$$

g_v is the antibody concentration. Antibody concentration is determined by affinity between antibodies. Before we talk about antibody concentration let's talk about the entropy value of the genes. For a population composed of T antibodies, each antibody includes t genes, and the entropy of gene j is defined by

$$H_j(T) = -\sum_{i=1}^{m} p_{ij} \cdot \log p_{ij} \tag{5}$$

Where p_{ij} is the probability that the jth gene belongs to i in T antibodies. If all the genes of antibodies are the same, the entropy equals zero. And the information entropy of these M antibodies is defined by

$$H(M) = \frac{1}{n} \sum_{j=1}^{m} H_j(M) \tag{6}$$

Then the affinity between antibody A_e and A_f is defined by $A_{ef} = \frac{1}{1+H(2)}$, The higher the value of A_{ef}, the closer the two antibodies are. Thus the concentration of antibody A_e in population is defined by

$$g_v = \frac{1}{T} \sum_{f=1}^{M} C_{ef} \tag{7}$$

C_{ef} is a counting variable defined by $C_{ef} = \begin{cases} 1, \alpha * \max_{1<f<T} A_{ef} < A_{ef} < \max_{1<f<T} A_{ef} \\ 0, \textit{others} \end{cases}$.

α depends on the distribution of antibodies. After introduced the calculation steps of expected reproduction rate, select the antibodies with higher expected reproduction rate to replace the suboptimal antibodies in E, thus completed the update of the excellent antibody library. Furthermore, the new generation probability density matrix D is obtained by updating $\beta * D + (1 - \beta) * F$, where β is the learning rate, and

the frequency matrix F is also obtained by the distribution of OTW in the antibodies with higher expected reproduction rate in the current generation.

More importantly, the generation of the new population consists of two parts: the first part is the random selection of some antibodies in E; the second part is the generation of new antibodies through the new generation D, and the new population is obtained by combining the two parts.

We will carry out simulation verification of this algorithm in the following part.

4 Simulation

The current section will verify the effectiveness of the joint probability density immune algorithm (JPDI) proposed in this paper. The simulation test data is provided by the high resolution AEOS called SPOT-6 and Pleiades-1A. The details of the two satellites are given in Table 2. STK is used to simulate the satellite and calculate the VTWs of the all the tasks under the two satellites.

4.1 Test Data

We simulated 150 common static tasks on the surface of the earth. Tasks are randomly distributed over land areas of the earth. We set the priority of the task as an integer, no generality, it's evenly distributed on a scale of 1 to 10. In addition, the imaging time of each target generated randomly between 30 seconds and 70 seconds.

The maneuvering angle velocity of satellite SPOT-6 is 2.5°/S, and the maneuvering angle range is ±30°; the maneuvering angular velocity of satellite Pleiades-1A is 2.4°/S, and the maneuvering angle range is ±60°. The orbital parameters of the satellite constellation are shown in Table 3.

Table 2 The parameters of the two satellites

Satellite designation	SSC number	Launch time	TLE format track parameters
SPOT-6	38,755	2012	1 38755U 12047A 13,037.07500000.00000403 00,000-0 86,692-400,003 2 38,755 098.1869 106.6959 0,001,101 018.0310 123.7197 14.58593013021858
Pleiades-1A	38,012	2011	1 38012U 11076F 13,031.16666667 0.00000268 00,000-0 57,796-4 0 00,008 2 38,012 098.1744 108.2165 0,001,176 018.8207 163.0471

Table 3 The maneuver parameters of the two satellites

Satellite	GM(km)	OR(m)	MA(°)	MAV(°/s)
SPOT-6	60	2	±30	30/12
Pleiades-1A	20	0.7	±60	60/25

Fig. 2 The convergence comparison of the algorithm
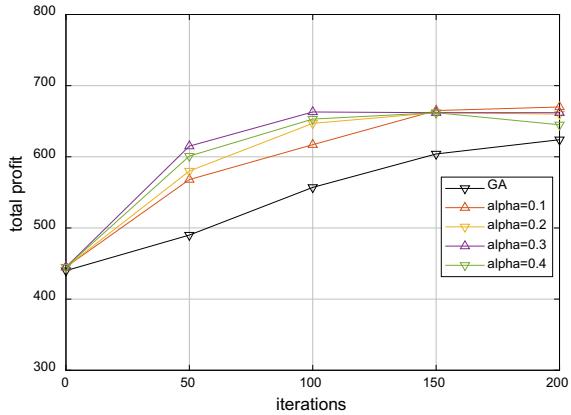
4.2 Simulation of GA and JPDI

In this section, we designed the satellite scheduling simulation process in task scale 150. The computational results of GA and JPDI in four different learning rates are in Fig. 2. Our reason for studying GA is that it's evolutionary way similar to the immune algorithm and thus can be used as a baseline to test the JPDI against.

As can be seen from Fig. 2, the improved immune algorithm is superior to the traditional genetic algorithm not only in convergence speed but also in scheduling total benefits. Moreover, the convergence rate is different at different learning rates, and the evolution of probability distribution immune algorithm by the 150th iteration stops at all learning rates. In the four simulation data experiments we did, when alpha equals 0.3, the probability distribution immune algorithm has the fastest convergence rate of our objective function.

The scheduling results show that in addition to the fact that the improved immune algorithm is superior to the genetic algorithm in the scheduling objective function value, there is a significant difference in the scheduling time. The time pairs of the two algorithms under different iteration times are shown in Fig. 3.

As the iteration increases, time consuming with the improved immune algorithm shows a linear growth trend with a low slope, while the genetic algorithm shows an exponential growth trend. When traditional genetic algorithms solve continuous problems, time consumption increased dramatically in global search process, which leads to the low efficiency of satellite scheduling.

The scheduling results generated by different VTW segments of the task are different. Figure 4 shows the scheduling benefit results at three different segments.

The experimental results show that as iteration increases, the more segments the better the scheduling benefits. When VTW is divided into 15 and 11 segments, the scheduling benefit difference is less than that 11 and 7 segments. Therefore, it can be assumed that when the number of segments reaches a certain level, the income growth gradually slows to stop. Therefore, the selection for suitable segments should be considered with the balance of scheduling time limit and profit value.
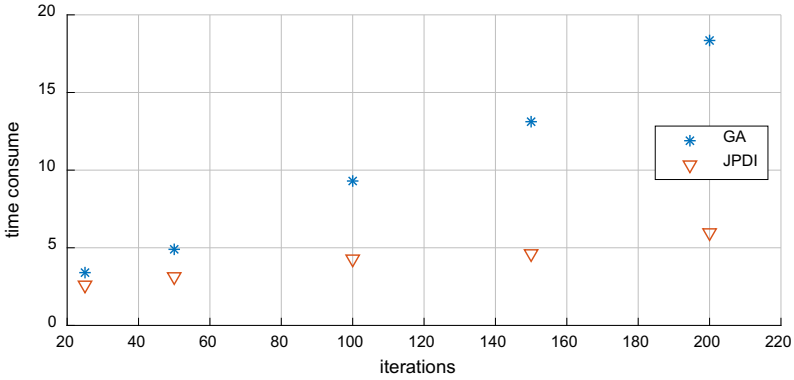
Fig. 3 The time consuming of the two algorithms
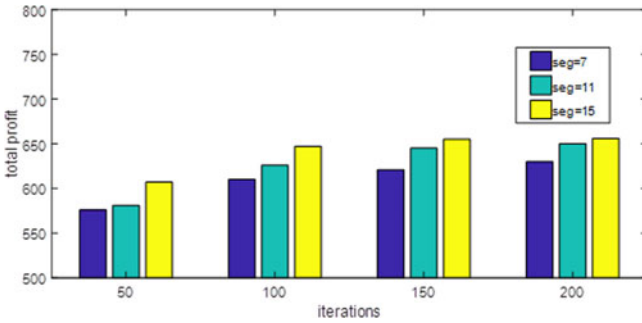
Fig. 4 The comparison under different VTW segments

5 Conclusions

This study focuses on the observing schedule problem of multiple agile satellites, and a detailed multi-satellite task planning constraint satisfaction model is established, which includes stereoscopic image requirements, cloud cover constraints and other factors. Our main goal was to maximize the scheduling efficient, reduce the waste of satellite resources and meet the different needs of users to the greatest extent. To achieve this, based on the evolution mode of classical intelligent algorithms, such as genetic algorithm and immune algorithm, making the solution easily falls into the local optimal solution and convergences slowly. We studied a new evolutionary approach, then the combined probability density immune algorithm was adopted. The numerical experiments which conducted on 150 tasks under two AEOS shows that when achieved the same scheduling efficiency, the algorithm proposed in this paper requires less evolutionary generations and as the algorithm iterates, the time required for scheduling convergence is greatly reduced.

In the future, the combined probability density immune algorithm can be further extended to other applications. The most pertinent direction for future work would be coordinated scheduling of multi-agent satellites with more practical constraints considered. Agent satellites will be more applicable for a better scheduling system. We will explore these potential studies in our future work.

References

1. Lemaitre, M., Verfaillie, G., Jouhaud, F., Lachiver, J.M., Bataille, N.: Selecting and scheduling observations of agile satellites. Aerosp. Sci. Technol. **6**(5), 367–381 (2002)
2. Kuipers, E.J.: An algorithm for selecting and timeabling requests for an earth observation satellite. Bulletin de la Société Françaisede Recherche Opérationnelle, etd' AideàlaDécision 7–10 (2003)
3. Wang, X., Chen, Z., Han, C.: Scheduling for single agile satellite, redundant targets problem using complex networks theory. Chaos Solitons Fract. Res. **83**, 125–132 (2016)
4. Liu, X., Laporte, G., Chen, Y., He, R.: An adaptive large neighborhood search metaheuristic for agile satellite scheduling with time-dependent transition time. Comput. Oper. Res. **86**, 41–53 (2017)
5. Wang, P., Tan, Y.: A heuristic method for selecting and scheduling observations of satellites with limited agility. In: 7th World Congress on Intelligent Control and Automation, pp. 5292–5297. IEEE (2008)
6. Ntagiou, E.V., Palmer, P., Iacopino, C., Policella, N., Donati, A.: Traffic estimation in road networks via compressive sensing. Coverage planning for agile EO constellations using ant colony optimisation. In: 14th International Conference on Space Operations (2016)
7. Xu, R., Chen, H., Liang, X., Wang, H.: Priority based constructive algorithms for scheduling agile Earth observation satellites with total priority maximization. Expert Syst. Appl. **51**, 195–206 (2016)

8. Bianchessi, N., Righini, G.: Planning and scheduling algorithms for the COSMO-SkyMed constellation. Aerosp. Sci. Technol. **12**(7), 535–544 (2008)

9. Habet, D., Vasquez, M., Vimont, Y.: Bounding the optimum for the problem of scheduling the photographs of an agile earth observing satellite. Comput. Optim. Appl. **47**(2), 307–333 (2010)

10. Chu, X., Chen, Y., Tan, Y.: A branch and bound algorithm for agile earth observation satellite scheduling. Adv. Space Res. **2017**(9), 1–15 (2017)

11. Geng, X., Li, J., Yang, W., Gong, H.: Agile satellite scheduling based on hybrid coding genetic algorithm. In: 12th World Congress on Intelligent Control and Automation, pp. 2727–2731. IEEE (2016)

12. Li, Y., Xu, M., Wang, R.: Scheduling observations of agile satellites with combined genetic algorithm. In: International Conference on Natural Computation, pp. 29–33. IEEE (2007)

13. Li, G., Chen, C., Yao, F.: Hybrid differential evolution optimisation for earth observation satellite scheduling with time-dependent earliness-tardiness penalties. Math. Probl. Eng. **(pt.8)**, 1–10 (2017)

14. Globus, A., Crawford, J., Lohn, J., Pryor, A.: A comparison of techniques for scheduling earth observing satellites. In: Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, pp. 836–843. AAAI (2004)

15. Wolfe, W.J., Sorensen, S.E.: Three scheduling algorithms applied to the Earth observing systems domain. Manag. Sci. **46**(1), 148–168 (2000)

16. Baek, S.W., Seung, W., Han, S.M., Cho, K.R.: Development of a scheduling algorithm and GUI for autonomous satellite missions. Acta Astronaut. **68**(7–8), 1396–1402 (2011)

17. Tangpattanakul, P., Jozefowiez, N., Lopez, P.: A multi-objective local search heuristic for scheduling Earth observations taken by an agile satellite. Eur. J. Oper. Res. **245**(2), 542–554 (2015)

18. Tangpattanakul, P., Jozefowiez, N., Lopez, P.: Multi-objective optimization for selecting and scheduling observations by agile earth observing satellites. In: International Conference on Parallel Problem Solving from Nature, pp. 112–121. Springer, Heidelberg (2012)

19. Zhang, D., Guo, L., Zhang, D.: A memetic algorithm for single-orbit image scheduling of agile satellite. In: International Conference on Computer Sciences and Automation Engineering. IEEE (2016)

20. Overill, R., Richard, E.: Artificial Immune Systems and their Applications. J. Logic Comput. **11**(6), 961–962 (2001)

The Radial Point Interpolation Mixed Collocation (RPIMC) Method for the Solution of the Reaction-Diffusion Equation in Cardiac Electrophysiology

Konstantinos A. Mountris and Esther Pueyo

Abstract The Radial Point Interpolation Mixed Collocation (RPIMC) method is developed for the solution of the reaction-diffusion equation in cardiac electrophysiology simulations. RPIMC is an efficient and purely meshfree technique which is expected to be a valuable alternative to the Finite Element Method (FEM) for cardiac electrophysiology applications where models with large number of degrees of freedom and high geometric complexity are common. We propose applying the operator splitting technique to solve the decoupled reaction-diffusion equation. In this way, the reaction (cardiac cell dynamics) and diffusion (action potential propagation) terms are solved independently. We evaluate the RPIMC in a simulation of the cardiac action potential (AP) propagation in a two-dimensional square tissue composed of human ventricular epicardium cells. The state-of-art O'Hara Rudy cell dynamics model is used to solve the reaction term while the diffusion term is solved using the standard forward Euler method. The simulation of the AP propagation using the RPIMC method is compared against a FEM simulation using isoparametric bilinear elements. Comparable results between RPIMC and FEM are obtained for both normal AP propagation and spiral wave generation conditions (expected in arrhythmic events). The convergence of the RPIMC solution to the FEM solution is evaluated for varying nodal spacing and varying dilatation coefficient during support domain nodes identification.

Keywords Radial point interpolation · Mixed collocation · Meshfree · Cardiac electrophysiology

K. A. Mountris (✉) · E. Pueyo
Aragón Institute of Engineering Research, University of Zaragoza, IIS Aragón, 50018 Zaragoza, Spain
e-mail: kmountris@unizar.es

K. A. Mountris · E. Pueyo
CIBER in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain

1   Introduction

The action potential propagation in the human heart (cardiac electrophysiologic function) is a complex multiscale phenomenon [1] that can be described mathematically by the reaction-diffusion Eq. (1a) and the boundary condition Eq. (1b).

$$\partial V/\partial t = -I_{ion}/C + \nabla \cdot (\boldsymbol{D}\nabla V) \qquad in\ \Omega, \tag{1a}$$

$$\boldsymbol{n} \cdot (\boldsymbol{D}\nabla V) = 0 \qquad\qquad in\ \partial\Omega, \tag{1b}$$

where Ω and $\partial\Omega$ are the domain of interest and its boundary, $\boldsymbol{n}$ is the outward unit vector to the boundary, V is the action potential (AP), I_{ion} is the sum of the cardiac cell ionic currents, C is the cell capacitance, and D is the diffusivity tensor of the cardiac tissue given by:

$$\boldsymbol{D} = d_0[(1 - \rho)\boldsymbol{f} \otimes \boldsymbol{f} + \rho\boldsymbol{I}], \tag{2}$$

where d_0 expresses the conductivity coefficient, $\rho \leq 1$ is the transversal to longitudinal conductivity ratio, $\boldsymbol{f}$ is the fiber direction vector, $\boldsymbol{I}$ is the identity matrix, and $\otimes$ denotes the tensor product operation. The diffusive term $\nabla \cdot (\boldsymbol{D}\nabla V)$ describes the propagation of the cellular AP in the tissue, while the reactive term $-I_{ion}/C$ describes the cellular AP dynamics. Due to the high complexity of the cardiac cell structure, realistic AP models employ a large number of "stiff" ordinary differential equations (ODE) to model accurately the cardiac gate variables and the produced ionic currents [2]. The reaction term "stiffness" requires a time integration step with sufficiently small size to ensure the stability and accuracy of the numerical solution of Eq. (1). To allow for larger time step size without reducing the numerical stability and accuracy, the solution can be decoupled using the operator splitting technique [3]. The two terms of the reaction-diffusion equation are decoupled and solved sequentially. A larger time integration step can be used then for the integration of the diffusion term while the reaction term can be integrated adaptively using a smaller step. State-of-art numerical solvers in cardiac electrophysiology employ the operator splitting approach and use the Finite Element Method (FEM) to derive the numerical solution. However, due to the complexity of the human heart geometry, meshfree methods that alleviate the mesh requirement are of great interest. In this work, we propose the Radial Point Interpolation Mixed Collocation [4] for the simulation of AP propagation in cardiac electrophysiology.

2   Methodology

The Radial Point Interpolation Mixed Collocation (RPIMC) method is a purely meshfree method based on the Meshless Local Petrov Galerkin (MLPG) method [5, 6], where the Radial Point Interpolation (RPI) is used to construct trial functions

and the Dirac delta function is used to construct test functions. Using RPIMC, the weak form of Eq. (1) is evaluated directly on the discretization nodes of the domain of interest and is given in Eq. (3). The use of the Dirac delta function to construct test functions results to reducing the spatial integration of the weak form to nodal summation over the support domain nodes.

$$\sum_{i=1}^{n} \phi^i(\boldsymbol{x_I}) \partial V^i / \partial t = -I_{ion}/C + \sum_{i=1}^{n} \nabla \cdot \boldsymbol{D} \nabla^T \phi^i(\boldsymbol{x_I}) V^i, \tag{3}$$

where n is the number of support domain nodes of the Ith discretization node and $\phi^i(x_I)$ is the RPI basis function given by:

$$\phi^T = \{ \boldsymbol{r_I}^T \boldsymbol{p_I}^T \} \boldsymbol{G_I}^{-1}, \tag{4}$$

where $\boldsymbol{r_I}$ is a radial basis function (RBF) and $\boldsymbol{p_I}$ is the polynomial basis, both evaluated at the Ith discretization node. $\boldsymbol{G_I}$ is composed by the RBF and polynomial basis moment matrices, $\boldsymbol{R_I}$ and $\boldsymbol{P_I}$ respectively:

$$\boldsymbol{G_I} = \begin{bmatrix} \boldsymbol{R_I P_I} \\ \boldsymbol{P_I}^T 0 \end{bmatrix}. \tag{5}$$

In this work, we use the linear polynomial basis $\boldsymbol{p_I} = \{1, x_I, y_I\}$ and the polyharmonic RBF ($\boldsymbol{r_I} = r_{Ii}^5, i = 1 \ldots n$) proposed in [7], with r_{Ii} being the radial distance between the Ith discretization node and its ith support domain node. Explicit time integration is performed using the forward Euler method and applying the operator splitting technique to decouple the solution of Eq. (3). Applying operator splitting, the solution of Eq. (3) at a time step k is obtained by:

(i) Solving $\partial V_I^{k'} / \partial t = -I_{ion} \left(V_I^{k-1} \right) / C$, and then
(ii) Solving $\sum_{i=1}^{n} \phi^i(\boldsymbol{x_I}) \partial V_I^k / \partial t = \sum_{i=1}^{n} \nabla \cdot \boldsymbol{D} \nabla^T \phi^i(\boldsymbol{x_I}) V_I^{k'}$.

In the following numerical examples, we consider a tissue composition of human cardiac ventricular epicardium. The O'Hara Rudy model [2] is used to simulate the cell dynamics in step *(i)* of the decoupled RPIMC solution.

3   Numerical Examples

In the first example, we consider a 5 cm × 5 cm square tissue with human cardiac ventricular epicardium composition. The direction of the cardiac fibers is considered perpendicular to the X-axis ($\boldsymbol{f} = [10]^T$). We use conductivity coefficient $d_0 = 0.0013$ mS/cm and transversal to longitudinal conductivity ratio $\rho = 1/4$.
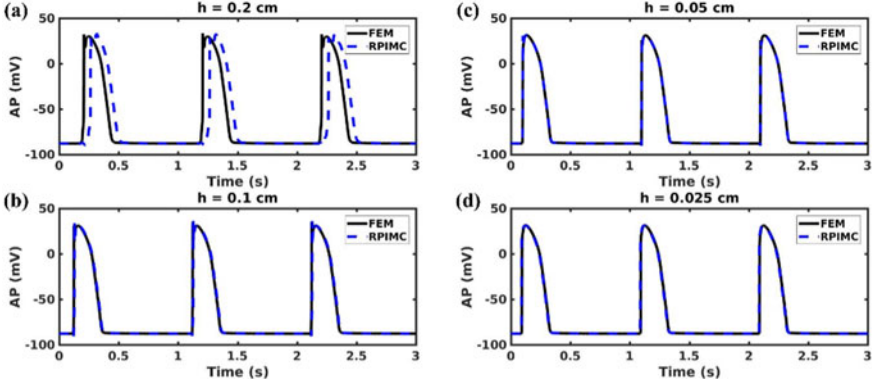
Fig. 1 Comparison of action potential (AP) for time interval $t = [0, 3]$ s for nodal spacing (a) $h = 0.2$ **cm**, (b) $h = 0.1$ **cm**, (c) $h = 0.05$ **cm**, (d) $h = 0.025$ **cm**

A periodic stimulus with period $t_T = 1$ s, duration $t_d = 2$ ms, and amplitude (A) of twice diastolic threshold is applied on the left side of the tissue ($x = 0$ cm). The AP propagation is simulated for time ($t_s = 3$ s). We validate the solution of the RPIMC method comparing it with a FEM solution using bilinear isoparametric elements. Regular nodal discretizations and quadrilateral meshes with nodal spacing $h = \{0.2, 0.1, 0.05, 0.025\}$ cm are considered. The support domain size $s_d = \alpha * h$, with $\alpha = 2.8$, is used for the support domain construction in RPIMC. A comparison of the generated AP by RPIMC and FEM in the time interval $t = [0, 3]$ s for all the nodal discretizations is given in Fig. 1.

To further evaluate the quality of the simulated AP we measure the AP duration (APD) metric for 90%, 50%, and 20% repolarization. The APD_{90} metric denotes the time between the maximum value of the potential's time derivative $(dV/dt)_{max}$ and the time of 90% repolarization from peak amplitude. The APD_{50} and APD_{20} metrics are defined similarly. The highest value of the percentage error of the RPIMC APD compared to the FEM APD is found 0.45%, 2.25%, and 2.27% for the APD_{90}, APD_{50}, APD_{20} and nodal spacing $h = 0.2$ cm. The percentage errors are reduced monotonically for reducing nodal spacing and are equal to zero for nodal spacing $h = 0.025$ cm. We further investigate the effect of the dilatation coefficient α by computing the Normalized Root Mean Square (NRMS) error between the RPIMC solution at $t = 2.2$ s for a nodal discretization with $h = 0.04$ cm and varying dilatation coefficient $a = \{1.2, 1.6, 2.0, 2.4, 2.8\}$ with a FEM solution. The NRMS error is computed using the formula:

$$NRMS = \frac{\left(\sum_{x_i \in \Omega} (u^{RPIMC}(x_i) - u^{FEM}(x_i))^2\right)^{1/2}}{max|u^{FEM}(x_i)| - min|u^{FEM}(x_i)|}, \tag{6}$$

Where x_i is the vector denoting the spatial coordinates of the ith node in the discretization of the domain Ω, $u^{RPIMC}(x_i)$ is the RPIMC solution at x_i, and u^{FEM} is the FEM solution at x_i. The NRMS error convergence plot with respect to a is given in Fig. 2.

In the next example, a S1-S2 cross stimulation protocol [8] is simulated to investigate the ability of RPIMC to generate and maintain spiral wave effects that are usually observed in cardiac arrhythmic events. The same tissue geometry and parameters as in the previous example are used with nodal spacing $h = 0.025$ cm. An initial stimulus (S1) is applied at the left edge of the tissue ($x = 0$ cm) at $t = 50$ ms. A second stimulus (S2) is applied at a square region located at the left bottom corner of the tissue with width 1.25 cm and height 2.50 cm at $t = 290$ ms. A spiral wave is generated due to the interaction of the S2 wave front with the S1 wave tail. The spiral wave is considered maintainable if at least 2 spirals are generated during the simulation time $t_s = 1$ s. The spiral wave simulation using RPIMC and FEM at different time intervals is plotted in Fig. 3. Maintainable spiral

Fig. 2 Normalized Root Mean Square (NRMS) error convergence for varying dilatation coefficient $a = \{1.2, 1.6, 2.0, 2.4, 2.8\}$ and nodal spacing $h = 0.04$ cm
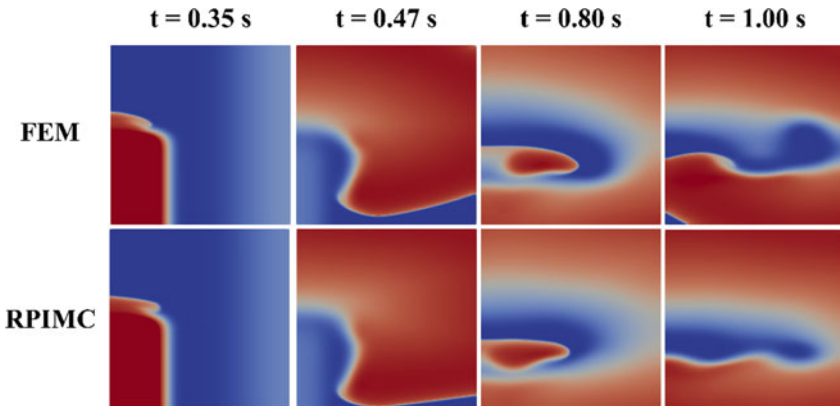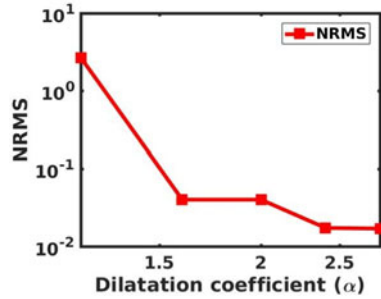
Fig. 3 Spiral wave propagation at different instants in the time interval $t = [0, 1]$ s for the S1-S2 cross stimulation protocol

waves are generated both in RPIMC and FEM with high similarity. The degradation of the similarity between the RPIMC and FEM spiral waves with time may be associated with the slightly slower conduction velocity of the AP in the FEM simulation compared to RPIMC.

References

1. Pueyo, E., Corrias, A., Virág, L., Jost, N., Szél, T., Varró, A., Szentandrássy, N., Nánási, P.P., Burrage, K., Rodríguez, B.: A multiscale investigation of repolarization variability and its role in cardiac arrhythmogenesis. Biophys. J . **101**(12), 2892–2902 (2011)
2. O'Hara, T., Virág, L., Varró, A., Rudy, Y.: Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. PLoS Comput. Biol. **7**(5), 1002061 (2011)
3. Qu, Z., Garfinkel, A.: An advanced algorithm for solving partial differential equation in cardiac conduction. IEEE Trans. Biomed. Eng. **46**(9), 1166–1168 (1999)
4. Mountris, K.A., Pueyo, E.: The Radial Point Interpolation Mixed Collocation (RPIMC) Method for the Solution of Transient Diffusion Problems. arXiv preprint arXiv:2001.01027 (2020)
5. Atluri, S.N., Zhu, T.: A new meshless local Petrov-Galerkin (MLPG) approach in computational mechanics. Comput. Mech. **22**(2), 117–127 (1998)
6. Zhang, T., He, Y., Dong, L., Li, S., Alotaibi, A., Atluri, S.N.: Meshless local petrov-galerkin mixed collocation method for solving cauchy inverse problems of steady-state heat transfer. CMES: Comput. Modeling Eng. Sci. **97**(6), 509–553 (2014)
7. Liu, G.R., Gu, Y.T.: An Introduction to Meshfree Methods and Their Programming. Springer, Heidelberg (2005)
8. Gomez, J.F., Cardona, K., Martinez, L., Saiz, J., Trenor, B.: Electrophysiological and structural remodeling in heart failure modulate arrhythmogenesis. 2D simulation study. PLoS ONE **9**(7), e103273 (2014)

Constraint Sensitivity Data Driven Automatic Optimal Design of Steel Frame Shear-Wall Structure

Jie Yao, Xin Zhao, and Junchen Guo

Abstract The traditional design method of the structure is commonly conceived by designers based on demands, thoughts, and experiences. However, this process is often time-consuming and hard to get the optimal solution. Automatic optimal design is the future development trend of structural design. The main purpose of this study is to propose an automatic optimization design method for building structures. The optimization method is based on constraint sensitivity data. Design constraint sensitivities can indicate the optimal changing directions of design variables. Based on the initial design, the sensitivity data can provide the optimal direction of structural material adjustment. Employing sensitivity data can realize the compliance design by the minimum material increment, and realize the optimization design by the maximum material decrement, thus to meet the performance requirements of super high-rise building structures with high efficiency and low cost. The design automation process is also described based on a modeling-analysis-design methodology. A frame shear wall structure is to be employed to show the viability and process of constraint sensitivity data-driven structural design automation.

Keywords Optimal Design · Sensitivity analysis · Frame shear wall structure · Constraint sensitivity data · Design Automation

J. Yao · X. Zhao (✉) · J. Guo
Department of Structural Engineering, Tongji University, No. 1239 Siping Road, Shanghai 200092, China
e-mail: 22zx@tjadri.com

X. Zhao
Tongji Architectural Design Group, No. 1230 Siping Road, Shanghai 200092, China

45

1   Introduction

Structure optimal design is attracting research interest in the building industry, especially in the design of tall buildings. Optimal design method includes objectives, constraints, and design variables. Objective function is minimized (or maximized) during optimization. Design objectives can be weight, structural cost, or seismic energy. Constraint functions are the criteria that the system has to satisfy for each feasible design. Constraints include stress, displacement, drift, etc. Design variables generally are member sizes in structure optimal design.

Constraint sensitivity data is usually used as the driving direction of structure optimal design. According to the sensitivity of the constraint conditions to the size of each component, the size of the component can be modified reasonably to make the final constraint in a reasonable range of redundancy. Based on the principle of virtual work, Jiang Xiang [1] deduced the relationship between the top acceleration and the component size of super tall building under lateral wind vibration. And, based on the sensitive data, the actual super tall building is optimized. Sensitivity analysis based optimization method to optimize structures subject to story drift constraint under earthquake action was studied [2].

Structure optimal design method is generally based on optimization algorithm. XieYimin [3] used the bi-directional evolutionary structural optimization algorithm to carry out architectural design, and through a series of cases, elaborated the use of bi-directional evolutionary optimization algorithm to generate beautiful and efficient forms, and revealed the important role of modern structural topology optimization algorithm in architectural design. Dong Yaoming proposed an optimal design for the structure lateral system of super tall buildings under multiple constraints [4]. Cao Benfeng developed a multilevel constrained optimal structural design method for reinforced concrete tall buildings [5].

In this paper, a multi-level optimization model design method is proposed. The method classifies the design purpose, design variables and constraints into component, component assemble and global levels. With the constraint sensitive data as the driving direction, and combined with the bi-directional evolutionary structural optimization idea, the automatic optimal design of the structure is finally realized. The design automation process is also described based on a modeling-analysis-design methodology (MAD).

2   Digital Modeling

The general structure design can be divided into four stages: planning, design, construction and maintenance. This paper introduces the concept of digital modeling in design stage. Digital modeling can be divided into four stages: geometric modeling, physical modeling, structural analysis and structural design, which can be summarized as modeling-analysis-design methodology (MAD). Figure 1

describes the specific process of MAD digital modeling method. The design elements of each modeling stage are shown in Table 1.

From Fig. 1 and Table 1, Mad method makes modeling standardized, which is conducive to the automation of structural design and optimization. This is the purpose of this paper.

3 Theoretical Basis

This paper introduces a structural automatic optimization method based on sensitive data-driven and bi-directional evolutionary optimization. The sensitive data can provide a reasonable direction for optimization, while the idea of bi-directional evolutionary optimization provides an algorithm to realize automatic optimization.

3.1 Sensitivity Method

As a tool of structure optimization, sensitivity analysis can provide reasonable optimization direction for structures. In this paper, the sensitivity analysis method of virtual work is used to obtain the first derivative of constraint conditions with respect to design variables, i.e. sensitivity data.

The sensitivity coefficient is defined as follows:

$$s_i^k = \Delta g_i / \Delta v^k \tag{1}$$

When Δv^k tends to 0, the formula (1) can be expressed as a formula (2):
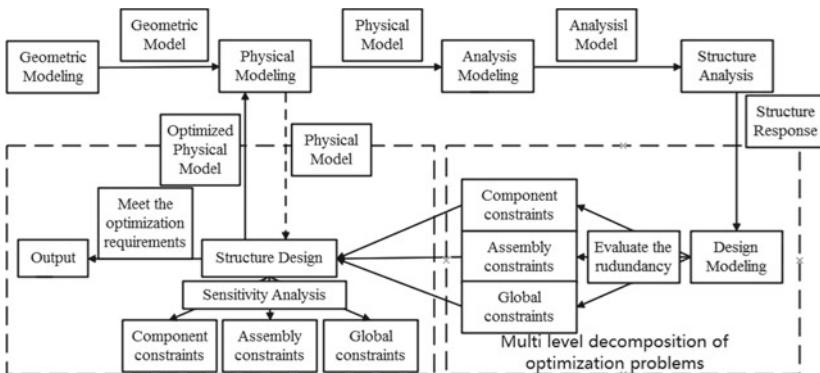
Fig. 1 MAD method process

Table 1 Modeling elements of MAD method

Digital modeling	Content
Structural geometric modeling	Point, Line, Surface, Body…
Structural physical modeling	Physical properties, Section…
Structural analysis modeling	Finite Element Mesh, Load, Boundary Finite Element Analysis Load Effect Combination
Structural design modeling	Design Criteria, Structure Response Sensitivity Analysis, Compliant design Optimal design, Design Automation

$$s_i^k = dg_i/dv^k \tag{2}$$

where s_i^k refers to the sensitivity coefficient of i-th constraint on the k-th design variable; g_i refers to the value of i-th constraint; v^k refers to the value of k-th design variable.

The relationship between constraints and design variables can be expressed by the principle of virtual work. According to the principle of virtual work, the virtual work done by the load under the real working condition under the displacement generated by the virtual working condition is equal to the virtual work of the internal force, which is expressed as follows.

$$e_{\delta i} = \int_0^{L_{bi}} \left(\frac{F_X f_X}{EA} + \frac{F_Y f_Y}{GA_Y} + \frac{F_Z f_Z}{GA_Z} + \frac{M_X m_X}{GI_X} + \frac{M_Y m_Y}{EI_Y} + \frac{M_Z m_Z}{EI_Z} \right)_i dx \quad \text{(Beam, column)} \tag{3}$$

where L_{bi} refers to the length of the i-th beam and column elements, F_X, F_Y, F_Z, M_X, M_Y, M_Z refer to the internal forces of the element under external loads; f_X, f_Y, f_Z, m_X, m_Y, m_Z refer to the internal force of the element under the virtual load; A refers to the cross section area of beam and column element; A_Y and A_Z refer to the shear area of beam and column elements; I_X, I_Y and I_Z refers to the torsion and bending moment of beam and column under inertia force.

It is assumed that the model is a statically determinate structure, that is to say, the internal force distribution of the structure will not be affected by changing the element size. Then the sensitivity coefficient of constraint conditions with respect to component size can be expressed as follows:

$$S_i = \frac{d\delta}{dvo_i} = \frac{de_{\delta i}}{dvo_i} \tag{4}$$

The analysis process of shell element is the same as above. We can get a simplified expression of the sensitivity coefficient.

$$S_i = -\frac{C_1}{(co^k)^2} - \frac{2C_2}{(co^k)^3} \quad \text{(Bar element)} \tag{5}$$

$$S_i = -\frac{C_3}{(co^k)^2} - \frac{3C_4}{(co^k)^4} \quad \text{(Shell element)} \tag{4}$$

Where C_1, C_2, C_3, C_4 are constants related to internal force, material properties, height, length and unit volume cost of materials.

3.2 Automatic Optimal Design

Based on the ascending bi-directional evolutionary optimization method, Design Modeling can be divided into three levels and check part according to the different level of constraints, which are component level, assembly component level and global structure level. The optimization process of each structure level is shown in Fig. 2.

The general process of realizing structure automatic optimal design by computer is shown in Fig. 2.

After calculating the sensitivity coefficient of the component, the optimization strategy is shown in Fig. 3.

Status a+ refers to that the structure is under-constrained and a- refers to the structure is over-constrained. A indicates those components that are sensitive to the controlling constraint and B indicate those components that are insensitive to the controlling constraint. The arrow indicates the adjustment direction of component size, up indicates increasing component size, and downward indicates decreasing component size.

4 Case Study

This paper takes a single story plane steel frame shear wall as an example. The design parameters and initial member section dimensions are shown in Table 2 and Table 3 (Fig. 4).

Taking the minimum cost as the optimization objective, the constraint conditions are the stress ratio and mid span deflection of the beam, the stress ratio and slenderness ratio of the column, the axial compression ratio of the shear wall, and story

Fig. 2 The automatic optimal
design of component process
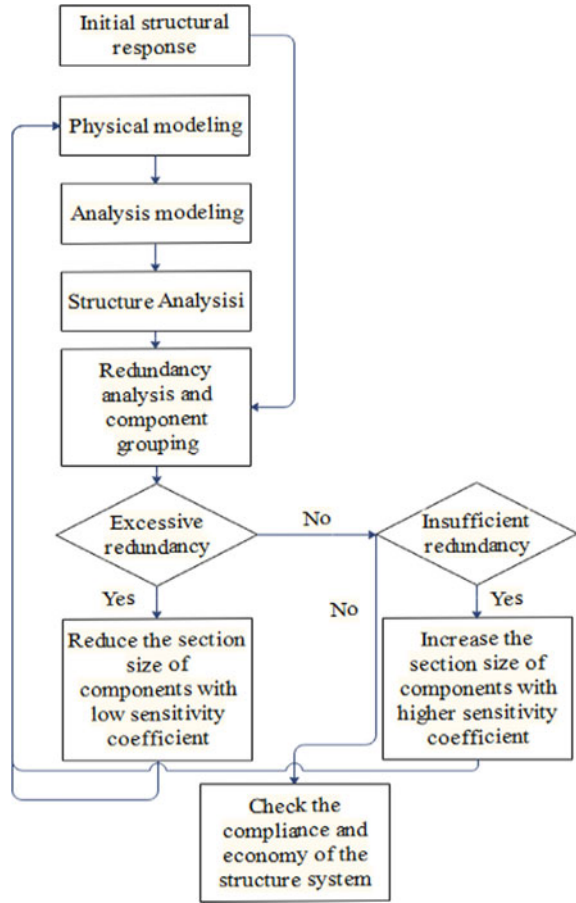
Fig. 3 Optimization
strategies

a^+ status a^- status
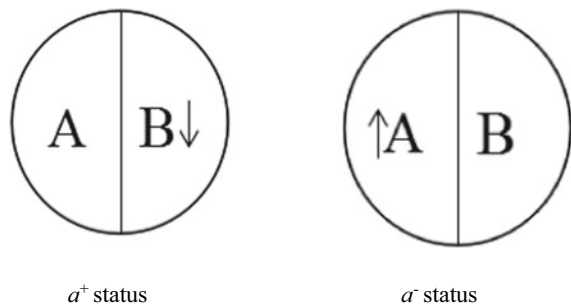
drift..The initial calculation results and the redundancy of each constraint are shown
in Table 4.
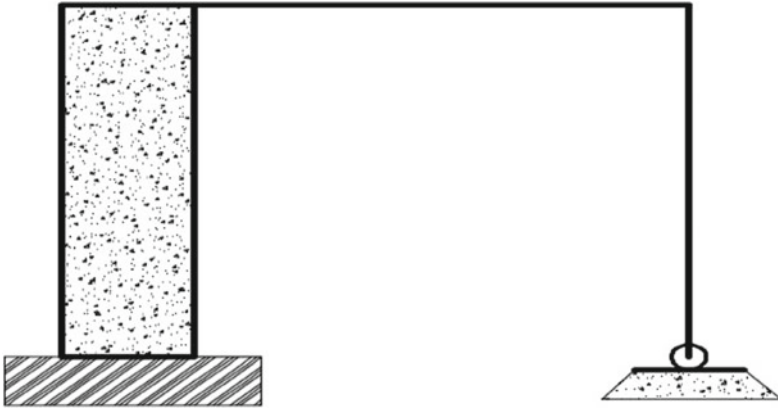
Fig. 4 Plane steel frame shear wall model

Table 2 Calculation parameters

Additional dead load (N/m)	Live load (N/m)	Basic wind pressure (kN/m^2)	Shape coefficient	Damping ratio	Characteristic site period
27,000	12,000	7.5	1.3	0.04	0.4

Table 3 Initial structural member size

	Height	Width	Web thickness	Flange thickness
Size of beam/mm	250	250	10	10
Size of column/mm	250	250	10	10
Wall thickness/mm	250			

It can be seen that the initial redundancy of each constraint is relatively large.For the structure, the amount of material is not economical enough. The constraint conditions at the member level only need to modify the corresponding member size. Sensitivity analysis is needed for the constraint conditions of the global structure level. The sensitivity index (The index of maximum sensitivity coefficient is designated as 100) of story drift with respect to each member is shown in Fig. 5.

It can be seen that in this model, the sensitivity coefficient of the beam is the largest, which indicates that changing the size of the beam has the greatest impact on the story drift. After obtaining the sensitive data, according to the automatic optimization method in Fig. 2, the optimized component size is shown in Table 5.

The redundancy of each constraint after optimization is shown in Table 6.

Table 4 Initial calculation results

Constraint level	Constraints at the component level					Global constrint	
Component	Beam		Column		Wall	–	
Constraint	Stress ratio	Deflection/ m^{-1}	Stress ratio	Slenderness ratio	Axial compression ratio	Story drift	
Initial calculation constraint value	0.55	0.0012	0.39	35	0.159	9.66×10^{-4}	2.09×10^{-4}
Constraint limits	0.9	0.0025	0.9	100	0.4	0.004	0.004
Redundancy	39%	52%	57%	–	60%	–	–
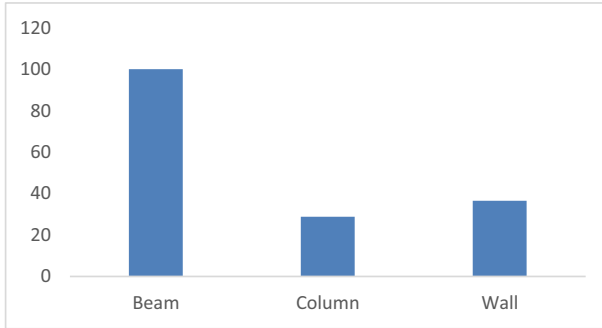
Fig. 5 Sensitivity index of story drift to component

Table 5 Optimized results

	Height	Width	Web thickness	Flange thickness
Size of beam/mm	220	220	8.8	8.8
Size of column/mm	130	5.2	5.2	5.2
Wall thickness/mm	200			

Table 6 Calculation results after optimization

Constraint level	Constraints at the component level					Global constraint	
Component	Beam		Column		Wall	--	
Constraint	Stress ratio	Deflection/m^{-1}	Stress ratio	Slenderness ratio	Axial compression ratio	Story drift	
Redundancy	2%	48%	23%	5%	50%	-	-

It can be seen that the optimized material consumption is more economical, and the constraint redundancy is in a more reasonable range. It is verified that the optimization result is at least a local optimal solution.

5   Conclusions

This paper introduces the MAD modeling method, which decomposes the model into several levels, which is beneficial to the computer programming to realize the goal of automatic optimization of structure. The relationship between constraints and design variables (component size) is established by using the principle of virtual work. Applying sensitive data to the direction of structural optimization can improve the search speed of size modification and save the optimization time. The case results show that the results obtained by this method are reasonable and economical, and the method is expected to be applied to larger structures.

References

1. Zhao, X., Jiang, X.: Sensitivity analysis of wind-induced acceleration on design variables for super tall buildings. In: Proceedings of IABSE Conference on Structural Engineering: Providing Solutions to Global Challenges, pp. 593–599 (2015)
2. Qin, L., Zhao, X.: Sensitivity analysis based optimal seismic design of tall buildings under story drift constraint. In: Proceedings of IABSE Conference on Bridges and Structures Sustainability: Seeking Intelligent Solutions, pp. 873–880 (2016)
3. Xie, Y., Zuo, Z., Lv, J.: Architectural Design through Bi-directional Evolutionary Structural Optimization, Time + Architecture,pp. 20–25 (2014)
4. Dong, Y.: Optimal design for structure lateral system of super tall buildings under multiple constraints, Tongji University (2015)
5. Cao, B.: Optimal design for reinforced concrete structure of tall building under multiple constraints, Tongji University (2015)

The Minimum Control Area Characterizing Hydrodynamic Macro Properties Dispersion Systems

Haoran Zhang, Xiaohan Yan, Huixia Feng, and Yongtu Liang

Abstract Studies on various dispersion systems have been developed with their widespread in engineering production, consisting of numerical simulation and experimental means. Experimental methods focus on hydrodynamic macro properties while simulation methods can be employed for interparticle interaction that closely relates with the non-Newtonian behaviors of dispersions. In this paper, a theoretical study method of the minimum control area (MCA) for dispersion systems was first proposed from the meso-scale perspective. It is a smallest infinitesimal extracted from a uniformly distributed dispersion system, which can characterize hydrodynamic macro properties of the system. The MCA method considers the interparticle interaction and is characterized by micro infinity and macro infinitesimal. It is representative through ensuring the key factor, involving the number and size distribution of dispersed particles in this area, conforms to the real dispersion system. We anticipated this method can be combined with computer technique to improve the study accuracy for the hydrodynamic macro properties of dispersion systems in the future.

Nomenclature

i	A class of dispersed particles with the same size i
$S_{a\ min}$	The minimum control area (MCA)
p_i	The probability whether the dispersed particles of a size i will be divided

H. Zhang (✉)
Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
e-mail: zhang_ronan@csis.u-tokyo.ac.jp

X. Yan
CNOOC Research Institute Co. Ltd., Beijing 100028, China

H. Feng · Y. Liang
Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Beijing 102249, China

R_i The radius of dispersed particles of the same size i
V The dispersion sample volume
h The height from the cutting position to the center of a dispersed particle
$\overline{S}_i$ The average cut profile of dispersed particles of a size i
S_i The cut profile of dispersed particles of a size i
k_i The cut numbers of dispersed particles of a size i
n_i The total dispersed particles of a size i in the sample dispersion
P_i The binomial distribution of the possibility that dispersed particles of a size i can be cut
k_i The number of dispersed particles of a size i displayed in the MCA
ϕ The dispersed particle volume fraction of the sample dispersion
C_i The number of dispersed particles of a size i per volume of dispersed particles in the dispersion, which can be known from experiments
M The size number of all the dispersed particles
γ The derived dispersed particle volume fraction in $S_{a\,min}$
e The variance between γ and ϕ
b The parameter valued as $1/e$

1   Introduction

Dispersion system is a kind of multi-medium fluid where dispersed medium distributes in the other immiscible continuous medium in the form of particles, which consists of solid-liquid, liquid-liquid, and gas-liquid systems. It is ubiquitous but also important in a variety of industries ranging from food, cosmetic, pharmaceutical, oil and mine, which produce an appreciable fraction of dispersed substances to be disposed. Thus, the prediction and analysis on hydrodynamic macro properties of dispersion systems is of considerable significance for stable and effective production.

There is much research on experimental methods for hydrodynamic macro properties of dispersion systems [1–3, 41]. In the view of large amounts of literatures [4–7], the volume fraction and size distribution of dispersed-phase have been claimed to the most important factor governing the non-Newtonian hydrodynamic macro properties. Recently, many scholars devoted themselves to experimenting on dispersions with different physical structures of interphases and analyzing the results from the micro perspective, in order to construct a relationship between the macro experimental phenomenon with the basic micro mechanism [8, 9]. However, it is of great difficulty to extent to micro-scale study merely relying on experimental means, and thereby some numerical simulation methods are necessary [10–14].

The multi-particle interaction in dispersions is a hot issue since it is a complex process associating with diverse certain forces such as buoyancy force, drag force, gravity [15–17] and uncertain forces such as Brownian force [18], and special physical features of dispersed and continuous phases [19]. To cope with it, the

direct numerical simulation (DNS) [20] can be utilized to model steady or turbulent flow fields, and the Lagrangian tracking means are required to model the dispersed particle movement. This DNS method facilitates the study and analysis on basic physical behaviors of dynamics systems. Carrier fluid and particle characteristics are obtained simultaneously. Later, the exchanges between the traditional numerical methods and the discrete element method (DEM) have been put forward [21]. The coupled the Lattice-Boltzmann numerical simulation and DEM, and the coupled DNS and DEM were developed and became popular in terms of a relatively higher simulation efficiency [22]. And there is an improved approach combining the DEM with the computational fluid dynamics (CFD) [23, 24]. DEM is used to model discrete particles while CFD is used to model the continuous phases based on the N-S equation, which can play an important role in the explicitly particle-particle, particle-wall, as well as particle-liquid collisions.

Most of the simulation methods coupled with the particle-discrete ideas are employed for single or several particles or dilute dispersions and can be well implemented, whereas larger amounts of dispersed particles need to be involved when considering a real condition. These methods mainly focus on the theoretical mechanism research but relatively ignore the connection between theory and reality. Further improvements need to be made by considering the movement of more dispersed particles. The micro-scale interaction should be displayed through macro visible phenomena. Although there are other computational simulation methods such as CFD, not only considering the micro interaction among dispersed particles but also focusing on macro fluid flow in dispersions [25–27], the simplification and artificial experience decrease the simulation accuracy [28] and the micro mechanism cannot be comprehensively taken into consideration.

From this point, we put forward a study method of the minimum control area (MCA) on the meso scale. It is proved to be the smallest study unit that can characterize the hydrodynamic macro property of a whole dispersion system, taking the number and size distribution of dispersed particles into account. This MCA allows micro study methods to be applied to the more accurate exploration of macro properties, since the study object scale and computational load can be considerably shrunk.

The present paper is organized as follows. The details of the methodology used are presented in Sect. 2. The derivation of the MCA analytical expression is performed in Sect. 3. The discussion for the application development and value of the MCA method is presented in Sect. 4.

2   Methodology Description

With respect to a homogenous dispersion system as a sample, it can be infinitely large to be studied but unrealistic, so there is a way to try to obtain a proper volume on behalf of the whole system. If half of the dispersion sample is taken, its property can be ensured to represent the whole system. And if half of the half dispersion is determined, it may be representative as well. While if we continue to divide the dispersion by half for each time, there must be an infinitely small divided volume failing to represent the whole dispersion, since the dispersed particles containing in the infinite small volume is so locally distributed leading to a distortion from the real system. In this sense, there must be a lower limit volume representative of the hydrodynamic macro properties of the whole dispersion.

To further simplify the volume concept for universal application, the three-dimensional volume can be converted into a two-dimensional area by means of cutting the volume in any direction, and the obtained infinitely small profile of the volume is defined as the MCA. Figure 1 displays this conversion procedure. And as seen in the Fig. 1(b) and (c), it is possible that the sphere particles can be cut at any position, thus, they can be regarded as circular profiles of different diameters in the MCA.

The hydrodynamic macro properties of dispersions are mostly subjected to the hydrodynamic interaction among dispersed particles [15–17], which is mainly associated with two factors: the particle number and size distribution. The particle number determines the frequency of particles' colliding and flocculating, while the particle size distribution determines the strength and effect of colliding and flocculating [29–31]. Based on this, the MCA is representative since it ensures the two factors of this area almost accord with those of the whole dispersion. According to Li et al. [32–34], the MCA can be regarded on the meso scale that can be perceived as a study perspective between micro and macro scales. If a study area is larger than the MCA, it belongs to a macro problem, leading to computational cost and the micro mechanism difficult to be involved; while a study area is smaller than the MCA, it belongs to a micro problem, leading to the theoretical mechanism study focusing on single or several particles and failing to characterize the whole systems.

(a) Dispersion system   (b) Infinitely small volume   (c) MCA
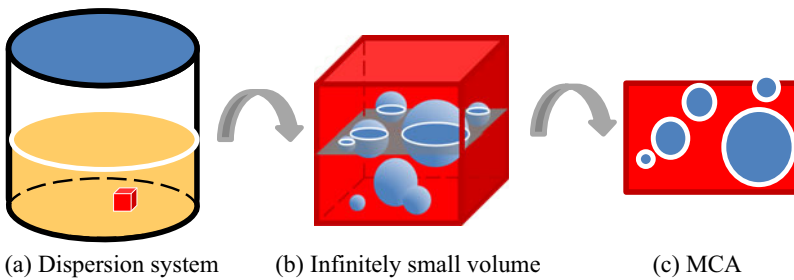
Fig. 1 Conversion procedure for MCA

The meso scale has both the macro and micro characters; in other words, the micro mechanism such as multiple interactions among dispersed particles can be extended into a macro scale and illustrated by macro-study means.

3 Methodology Derivation

The MCA analytical equation was derived based on probability theories from the particle-scale perspective. And the derivation principle accords with the particle volume content and distribution of the extracted study object agreeable with the original dispersion system. Firstly, some assumptions are given:

- The dispersion should be regarded as a homogenous system.
- The dispersed particles should be treated as rigid spheres with no deformation.
- With specified sample dispersions and environment conditions, the hydrodynamic interaction among particles is only subjected to the number and size distribution of dispersed particles in the MCA.

3.1 MCA

The conversion from three-dimension to two-dimension can be regarded as the volume being cross-cut to get its profile. Whereas it is inevitable that some of the dispersed particles in the volume will be divided into two halves possibly of different sizes with each one. And the probability that the dispersed particles of a size i will be divided is given by the following:

$$p_i = \frac{2S_{amin}R_i}{V} \tag{1}$$

where: $2S_{a\,min}R_i$ represents a cube that will be cut in a dispersion sample. $2R_i$ is the height and $S_{a\,min}$ is the base of the cube. Suppose that some dispersed particles of size i are cut at a distance h to the particle center, then the particle spheres are transformed into profiles and yield:

$$S_i = \pi(R_i^2 - h^2) \tag{2}$$

And the probability that this size of dispersed particles will be cut to be profiles is expressed by:

$$p(S_i) = \frac{1}{2R_i} \tag{3}$$

The profiles of one size of particles are different with one another since the cut positions (i.e. h) are different. But we can obtain the average profiles of one size of dispersed particles regardless of cutting position by combining Eqs. (2) and (3), shown as below:

$$\overline{S_i} = \int_{-R_i}^{R_i} S_i \frac{1}{p(S_i)} dh = \frac{2\pi R_i^3 - \frac{2}{3}\pi R_i^3}{2R_i} = \frac{2}{3}\pi R_i^2 \tag{4}$$

If the dispersed particles of size i fall in the range of the MCA, $S_{a\ min}$, the probability will be p_i, shown as Eq. (1); otherwise it will be $1 - p_i$, which can be described by a binomial distribution. Suppose that the cut number of one size of dispersed particles is k_i and the total dispersed particle number of the whole dispersion is n_i. Then the binomial distribution can be established as:

$$P_i(k_i) = C_{n_i}^{k_i} p_i^{k_i} (1 - p_i)^{n_i - k_i} \tag{5}$$

which can also be denoted by the expression:

$$k_i \sim B(n_i, p_i) \tag{6}$$

When the sample dispersion is characterized by higher concentration, the total dispersed particle number n_i will tend to be infinitely large, contributing to the binomial distribution converging towards a normal distribution, namely:

$$k_i \sim N(n_i p_i, n_i p_i(1 - p_i)) \tag{7}$$

The number of total dispersed particles n_i in the sample dispersion can be expressed as:

$$n_i = V\emptyset C_i \tag{8}$$

And there is another equation defining C_i as:

$$\frac{4}{3}\pi \sum_i^M C_i R_i^3 = 1 \tag{9}$$

where: $\frac{4}{3}\pi C_i R_i^3$ can be interpreted as the ratio of the total particle volume of a size i to a volume of dispersed particles of the whole sample, V_s, and summing up it can carry out the ratio of total particle volume of all sizes to V_s, which equals to 1.

And next is to deal with Eq. (7). First of all, multiplying Eq. (7) by the mean profile $\overline{S_i}$ obtained from Eq. (4) yields:

$$\frac{2}{3}\pi R_i^2 k_i \sim N(\frac{2}{3}\pi R_i^2 n_i p_i, \frac{4}{9}\pi^2 R_i^4 n_i p_i(1-p_i)) \tag{10}$$

And sum up Eq. (10) to include all sizes of dispersed particles:

$$\sum_i^M \frac{2}{3}\pi R_i^2 k_i \sim N(\sum_i^M \frac{4}{3}\pi R_i^2 n_i p_i, \sum_i^M \frac{4}{9}\pi^2 R_i^4 n_i p_i(1-p_i)) \tag{11}$$

Then combining Eq. (11) with Eqs. (1) and (8) yields:

$$\sum_i^M \frac{2}{3}\pi R_i^2 k_i \sim N(\sum_i^M \frac{4}{3}\pi R_i^3 \emptyset C_i S_{amin}, \sum_i^M \frac{8}{9}\pi^2 R_i^5 \emptyset C_i S_{amin}(1-p_i)) \tag{12}$$

Since the MCA is particle-scale and infinitely small, the probability p_i can be perceived as infinitely small as well. On the contrary, the probability of $(1-p_i)$ will approach to 1, then Eq. (12) can be simplified:

$$\sum_i^M \frac{2}{3}\pi R_i^2 k_i \sim N(\sum_i^M \frac{4}{3}\pi R_i^3 \emptyset C_i S_{amin}, \sum_i^M \frac{8}{9}\pi^2 R_i^5 \emptyset C_i S_{amin}) \tag{13}$$

Equation (13) divided by S_{amin} yields:

$$\gamma = \frac{\sum_i^M \frac{2}{3}\pi R_i^2 k_i}{S_{amin}} \sim N(\sum_i^M \frac{4}{3}\pi R_i^3 \emptyset C_i, \frac{\sum_i^M \frac{8}{9}\pi^2 R_i^5 n_i \emptyset C_i}{S_{amin}}) \tag{14}$$

According to Eq. (9), the mean value $\sum_i^M \frac{4}{3}\pi R_i^3 \phi C_i$ equals to ϕ constantly. And incorporating Eq. (9) with Eq. (14) yields:

$$\gamma \sim N(\emptyset, \frac{\sum_i^M \frac{8}{9}\pi^2 R_i^5 n_i \emptyset C_i}{S_{amin}}) \tag{15}$$

It is known that the γ in Eq. (15) represents the volume fraction of dispersed particles of S_{amin} that we have derived, which illustrates γ keeps the same physical significance with ϕ. From the equation, the mean value of the derived fraction γ equals to the volume fraction ϕ of dispersed particles of the whole dispersion that can be measured by experiments. Since it is expected that the volume fraction of S_{amin} should conform to the dispersion, a specified minimum error $\sqrt{e}$ is given here:

$$\frac{\sum_i^M \frac{8}{9}\pi^2 R_i^5 n_i \emptyset C_i}{S_{amin}} = e \tag{16}$$

And S_{amin} can be given by:

$$S_{amin} \frac{\sum_i^M \frac{8}{9}\pi^2 R_i^5 n_i \emptyset C_i}{e})$$ (17)

It is the variance to indicate how much the difference is between γ and ϕ. And S_{amin} is obtained from Eq. (17), where all the parameters including R_i, n_i, ϕ, C_i, and M can be known from experimenting on the sample dispersion.

3.2   Particle Number in MCA

As shown in Fig. 1(c), there are dispersed particles of different sizes in the MCA, which are especially important in the MCA similarity with the original dispersion. Under the premise of a known MCA equation, we can continue to carry out the analytical equation of the particle numbers of different sizes.

Set $b = \frac{1}{e}$, and incorporating it and Eq. (4) into Eq. (17) yields:

$$S_{amin} = b\frac{8}{9}\pi^2\emptyset \sum_i^M R_i^5 n_i C_i = b\frac{2}{3}\pi\emptyset \sum_i^M R_i^2 = b\emptyset \sum_i^M \bar{S}_i$$ (18)

According to Eq. (7), summing up k_i yields the normal distribution for total particle number that are transformed to profiles, as below:

$$\sum_i^M k_i \sim \mathrm{N}(\sum_i^M n_i p_i, \sum_i^M n_i p_i(1-p_i))$$ (19)

Sum up the multiplication of Eqs. (1) and (8):

$$\sum_i^M n_i p_i = \sum_i^M 2S_{amin}\emptyset R_i C_i$$ (20)

$$\sum_i^M n_i p_i(1-p_i) = \sum_i^M 2S_{amin}\emptyset R_i C_i(1 - \frac{2S_{amin}R_i}{V_T})$$ (21)

Then coupling Eq. (19) with Eqs. (20) and (21) yields:

$$\sum_i^M k_i \sim N\left[\sum_i^M 2S_{amin}\emptyset R_i C_i, \sum_i^M 2S_{amin}\emptyset R_i C_i(1 - \frac{2S_{amin}R_i}{V_T})\right]$$ (22)

As seen from Eq. (22), it indicates that $\sum\limits_{i}^{M} k_i$, the total particle number that is cut, should fluctuate around the mean value $\sum\limits_{i}^{M} 2S_{a\,min}\phi R_i C_i$ with the variance of $\sum\limits_{i}^{M} 2S_{a\,min}\phi R_i C_i \left(1 - \frac{2S_{a\,min}R_i}{V_T}\right)$. While the value of $1 - \frac{2S_{a\,min}R_i}{V_T}$ approaches to 1, thus k_i can be calculated by the following equation:

$$\sum_{i}^{M} k_i \approx \sum_{i}^{M} 2S_{amin}\emptyset R_i C_i \tag{23}$$

The parameter, k_i is perceived as a key factor deciding the MCA similarity and exerting great influence on the hydrodynamic macro properties of dispersion, as a result of both the total particle number and size distribution involved.

4   Conclusion Remarks

To our knowledge, considerable investigations on hydrodynamic macro properties of dispersion systems have been done, which drives a lot of research directions such as the rheology and viscosity prediction [35–37], the relevant numerical simulation methods [12, 38], the interparticle micro mechanism [15, 39], etc. In the future, the MCA method proposed in this paper aims to be applied to these study areas by coupling with computational simulation methods such as computational fluid dynamics (CFD), molecular dynamic (MD) simulation, dissipative particle dynamic (DPD) simulation, etc. For example, when modelling a dispersion system, the basic data for numerical simulation, especially the MCA size is required. On the one hand, the simulation should conform to the real condition; on the other, the calculation time is better to be shortened as much as possible. Thus, the MCA is able to work out a minimum dispersion area consisting of the smallest number and a real size distribution of dispersed particles. It cannot only improve the simulation efficiency but also increase the implementation value of the simulation, in virtue of the macro properties of the modelled dispersion satisfactory with real conditions.

Furthermore, there is a key factor k_i illustrating the number and the size distribution of dispersed particles in the MCA, which have tremendous impact on hydrodynamic macro properties of dispersions [40, 41]. For example, it is essential that the dispersion rheology is mainly dominated by the interparticle interaction that is closely related to k_i for the interaction possibility and effect are determined by the particle number and size distribution respectively. With more concentrated dispersions (i.e. larger value of k_i), there is more intense interaction highlighting the rheological behaviors. Hence, studies on dispersion rheology have paid much attention to the influence of dispersed particle concentration and size distribution.

From this point, the key factor k_i, worked out by the MCA method, will be another novel study idea provided for theoretical analysis, which is able to be combined with experiments and contribute to more reliable hydrodynamic property prediction for non-Newtonian dispersion systems.

Author Contributions Haoran Zhang wrote the initial paper, Xiaohan Yan and Huixia Feng checked grammar and drew. Yongtu Liang provided the overall idea of the article and revised the paper. All authors contributed to the final manuscript.

Additional Information Competing interests: The authors declare no competing financial interests.

References

1. Caricchi, L., Burlini, L., Ulmer, P., Gerya, T.V., Vassalli, M., Papale, P.: Non-Newtonian rheology of crystal-bearing magmas and implications for magma ascent dynamics. Earth Planet. Sci. Lett. **264**, 402–419 (2007)
2. Costa, A., Caricchi, L., Bagdassarov, N.: A model for the rheology of particle-bearing suspensions and partially molten rocks. Geochem. Geophys. Geosyst. **10**(3) (2009)
3. Lejeune, A.-M., Richet, P.: Rheology of crystal bearing silicate melts: an experimental study at high viscosities. J. Geophys. Res. **100**(B3), 4215–4229 (1995)
4. Mooney, M.J.: The viscosity of a concentrated suspension of spherical particles. J. Colloid Sci. **6**(2), 162–170 (1951)
5. Pal, R., Rhodes, E.: Viscosity/concentration relationships for emulsions. J. Rheol. **33**(7), 1021–1045 (1978-present) (1989)
6. Pal, R.: Viscous behavior of concentrated emulsions of two immiscible Newtonian fluids with interfacial tension. J. Colloid Interface Sci. **263**(1), 296–305 (2003)
7. Wang, W., Wang, P.Y., Li, K., et al.: Prediction of apparent viscosity of non-Newtonian water-in-crude oil emulsions. Pet. Explor. Dev. **40**(1), 121–124 (2013)
8. Huang, Q.Y., Wang, L.: Effect of droplet distribution on rheological properties of water-in-oil emulsion in waxy crude oils. Acta Petrolei Sinica **34**(4), 765–774 (2013)
9. Moradi, M., Alvarado, V., Huzurbazar, S.: Effect of salinity on water-in-crude oil emulsion: evaluation through drop-size distribution proxy. Energy Fuels **25**(25), 260–268 (2011)
10. Derksen, J.J.: Numerical simulation of solids suspension in a stirred tank. AIChE J. **49**(11), 2700–2714 (2003)
11. Frising, T., Noïk, C., Dalmazzone, C.: The liquid/liquid sedimentation process: from droplet coalescence to technologically enhanced water/oil emulsion gravity separators: a review. J. Dispersion Sci. Technol. **27**(7), 1035–1057 (2006)
12. Chou, Y.J., Fringer, O.B.: Modeling dilute sediment suspension using large-eddy simulation with a dynamic mixed model. Phys. Fluids **20**(11), 703–710 (2008)
13. Delnoij, E., Kuipers, J.A.M., Swaaij, W.P.M.V.: Dynamic simulation of gas-liquid two-phase flow: effect of column aspect ratio on the flow structure. Chem. Eng. Sci. **52**(21–22), 3759–3772 (1997)

14. Becker, S., Sokolichin, A., Eigenberger, G.: Gas-liquid flow in bubble columns and loop reactors: Part II. Comparison of detailed experiments and flow simulations. Chem. Eng. Sci. **49**(24), 5747–5762 (1994)
15. Dueck, J., Minkov, L.L.: Non-stokesian sedimentation as applied to the analysis of the interaction of particles in a suspension. J. Eng. Phys. Thermophys. **85**(1), 19–28 (2012)
16. Vanroyen, C., Omari, A., Toutain, J., et al.: Interactions between hard spheres sedimenting at low Reynolds number. Eur. J. Mech. B. Fluids **24**(5), 586–595 (2005)
17. Gromer, A., Gunning, A.P.: Atomic force spectroscopy of interactions between oil droplets in emulsions. J. Microsc. Anal. **1**, 9–12 (2011)
18. Deming, N.: Research on the Particles Sedimentation and Brownian Motion. Zhe Jiang University, Zhe Jiang (2011)
19. Liu, D., Bu, C., Chen, X.: Development and test of CFD–DEM model for complex geometry: a coupling algorithm for Fluent and DEM. Comput. Chem. Eng. **58**, 260–268 (2013)
20. Pan, T.W., Glowinski, R., Hou, S.: Direct numerical simulation of pattern formation in a rotating suspension of non-Brownian settling particles in a fully filled cylinder. Comput. Struct. **85**(11–14), 955–969 (2007)
21. Suzuki, K., et al.: Simulation of upward seepage flow in a single column of spheres using discrete-element method with fluid–particle interaction. J. Geotech. Geoenviron. Eng. **133**(1), 104–110 (2007)
22. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., Yu, A.B.: Discrete particle simulation of particulate systems: a review of major applications and findings. Chem. Eng. Sci. **63**, 5728–5770 (2008)
23. Qiu, L., Wu, C.Y.: Gravitational sedimentation and separation of particles in a Liquid: a 3D DEM/CFD study. Special Publication-Royal Society of Chemistry (2012)
24. Zhao, J., Shan, T.: Coupled CFD–DEM simulation of fluid–particle interaction in geomechanics. Powder Technol. **239**(17), 248–258 (2013)
25. Chou, Y.J., Gu, S.H., Shao, Y.C.: An Euler-Lagrange model for simulating fine particle suspension in liquid flows. J. Comput. Phys. **299**, 955–973 (2015)
26. Balakin, B.V., Hoffmann, A.C., Kosinski, P., et al.: Eulerian-Eulerian CFD model for the sedimentation of spherical particles in suspension with high particle concentrations. Eng. Appl. Comput. Fluid Mech. **4**(1), 116–126 (2014)
27. Tamburini, A., Cipollina, A., Micale, G., et al.: CFD simulations of dense solid–liquid suspensions in baffled stirred tanks: prediction of suspension curves. Chem. Eng. J. **178**(1), 324–341 (2011)
28. Ali, B.A., Pushpavanam, S.: Analysis of unsteady gas–liquid flows in a rectangular tank: comparison of Euler-Eulerian and Euler-Lagrangian simulations. Int. J. Multiph. Flow **37**(3), 268–277 (2011)
29. Kollár, L.E., Farzaneh, M., Karev, A.R.: The role of droplet collision, evaporation and gravitational settling in the modeling of two-phase flows under icing conditions. In: Proceedings of the 11th International Workshop on Atmospheric Icing of Structures, Montreal, QC, Canada, Paper IW38 (2005)
30. Jurado, E., Bravo, V., Camacho, F., et al.: Estimation of the distribution of droplet size, interfacial area and volume in emulsions. Colloids Surf. A: Physicochem. Eng. Aspects **295**, 91–98 (2007)
31. Kang, W., Guo, L., Fan, H., et al.: Flocculation, coalescence and migration of dispersed phase droplets and oil–water separation in heavy oil emulsion. J. Petrol. Sci. Eng. **81**, 177–181 (2012)
32. Li, J., Ge, W., Kwauk, M.: Meso-scale phenomena from compromise – a common challenge, not only for chemical engineering. Eprint Arxiv (2010)
33. Li, J., Ge, W., Wang, J., et al.: From multiscale modeling to meso-science. Springer, Heidelberg (2013)
34. Bona, Lu., Wei, W., Jinghai, Li.: Eulerian simulation of gas–solid flows with particles of Geldart groups A, B and D using EMMS-based meso-scale model. Chem. Eng. Sci. **66**(20), 4624–4635 (2011)

35. Liu, D.M.: Particle packing and rheological property of highly-concentrated ceramic suspensions: φm determination and viscosity prediction. J. Mater. Sci. **35**(21), 5503–5507 (2000)
36. Mueller, S., Llewellin, E.W., Mader, H.M.: The rheology of suspensions of solid particles. Proc. Math. Phys. Eng. Sci. **466**(2116), 1201–1228 (2010)
37. Zhang, J., Zhao, H., Li, W., et al.: Multiple effects of the second fluid on suspension viscosity. Sci. Rep. **5**, 16058 (2015)
38. Aoyi, O., Onyango, M.S.: CFD simulation of solids suspension in stirred tanks: review. Hemijska Industrija **64**(5), 365–374 (2010)
39. Naso, A., Prosperetti, A.: The interaction between a solid particle and a turbulent flow. New J. Phys. **12**(3), 033040 (2010)
40. Graham, A.L., Steele, R.D., Bird, R.B.: Particle clusters in concentrated suspensions. 3. Prediction of suspension viscosity. Ind. Eng. Chem. Fundam. **23**(4), 420–425 (1984)
41. Durlofsky, L., Brady, J.F., Bossis, G.: Dynamic simulation of hydrodynamically interacting particles. J. Fluid Mech. **180**, 21–49 (1987)

The Optimization for Natural Gas Supply Chain Under Multi-source Pattern

Zhengbing Li, Yongtu Liang, Zhigang Duan, Qi Liao, Haoran Zhang, and Yi Wang

Abstract As clean and green energy, natural gas has an immense potential for utilization and is becoming dominant in global energy consumption structure. To satisfy the large energy demand in China, natural gas industry is developing towards large scale, diversification and complication. In recent years, the optimization of natural gas supply chain has been studied extensively. However, most of the studies focused on supply security and the optimal distribution of single supplier to demand sides, while the optimization of natural gas supply chain under multi-party competition is rarely considered. Aiming at the competition between the multiple suppliers for the same market, the paper develops a universal mixed integer linear programming (MILP) model for the optimal design and operation of the natural gas supply chain system. The model takes the maximum profit of each supplier as the objective function, and considers the constraints of balance between supply and demand, pipeline construction, station construction, underground reservoir construction and so on. Through using particle swarm optimization (PSO) algorithm, the model is successfully solved to obtain the gas supply scheme and construction scheme of each supplier. Finally, the applicability of the model is illustrated by applying it to a real world natural gas supply chain in China. The results show that the solution is reasonable and can converge to a better value. The research has guiding significance for optimizing natural gas supply chain and ensuring the economic and stability of natural gas supply.

Z. Li · Y. Liang (✉) · Q. Liao · Y. Wang
National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Fuxue Road No. 18, Changping District, Beijing 102249, China

Z. Duan
SINO-Pipeline International Company Limited, Anding Road No. 5, Chaoyang District, Beijing 100029, China

H. Zhang
Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

67

Nomenclature

Sets and Indices

$i, i' \in I = \{1, 2, \ldots \ldots, i_{\max}\}$	Set of supply node.
$J = \{i_{\max} + 1, i_{\max} + 2, \ldots \ldots, i_{\max} + j_{\max}\}$	Set of demand node.
$j, j' \in I \cup J$	Set of all nodes.
$k \in K = \{1, 2, 3, 4\}$	Set of seasons.

Continuous Parameters

M	A big number.
$co_i^{production}$	Unit price of gas production of supplier i(CNY/m^3).
$co^{station}$	Station construction cost (CNY).
co^{pipe}	Unit price of pipeline construction (CNY/km).
$co^{maintenace}$	Unite price of pipeline maintenance (CNY/km).
co^{energy}	Unit price of energy consumption for pipeline transportation (CNY/(km·m^3/h)).
$co^{storage}$	Storage construction cost (CNY).
$d_{k,j}$	Demand at node j in season k(m^3).
$dy^{station}$	Depreciation rate of station.
dy^{pipe}	Depreciation rate of pipeline.
$dy^{storage}$	Depreciation rate of storage.
$l_{j,j'}$	Distance between nodes j and j'(km).
n^{season}	The number of days in a season.
$p_{k,i}^{purchase}$	Unit purchase price of natural gas of supplier i in the season k(CNY/m^3).
$p_{k,i}^{sale}$	Sale price of natural gas of supplier i in the season k(CNY/m^3).
$p_{i'}^{pipe}$	Unit price of transportation through the pipeline of supplier i'(CNY/m^3).
$q_{j,j'}^{min}$	Lower boundary of flowrate between node j and node j'(m^3/h).
$q_{j,j'}^{max}$	Upper boundary of flowrate between node j and node j'(m^3/h).
sv_i^{max}	Maximum production of gas field of supplier i(m^3).
vos_j^{min}	Lower boundary of the storage ability at node j(m^3).
vos_j^{max}	Upper boundary of the storage ability at node j(m^3).
$vpr_{k,i}^{max}$	Maximum purchase of natural gas by supplier i during the season k (m^3).

Binary Parameters

$s_{i,j,j'}^{pipe}$ Binary parameters of existing pipelines, $s_{i,j,j'}^{pipe} = 1$ if supplier i has constructed the pipeline between nodes j and j', or else $s_{i,j,j'}^{pipe} = 0$.

$s_j^{station}$ Binary parameters of existing stations, $s_j^{station} = 1$ if there is a station at node j, or else $s_j^{station} = 0$.

$s_j^{storage}$ Binary parameters of existing underground gas storages, $s_j^{storage} = 1$ if there is a storage at node j, or else $s_j^{storage} = 0$.

$sc_j^{storage}$ Binary parameters of underground gas storage construction, $sc_j^{storage} = 1$ if node j is suitable for constructing a underground gas storage, or else $sc_j^{storage} = 0$.

$fl_{j,j'}$ Binary parameters of gas flow, $fl_{j,j'} = 1$ if gas flows from node j to node j', or else $fl_{j,j'} = 0$.

Positive Continuous Variables

$Q_{k,j,j'}$ Flowrate between node j and node j' in the season $k(\text{m}^3/\text{h})$.

$VTF_{k,i,j}$ Daily peak shaving volume of the storage j constructed by supplier i in the season $k(\text{m}^3/\text{d})$.

$V_{k,i}^{purchase}$ Purchased volume of natural gas by supplier i during the season $k(\text{m}^3)$.

$VS_{k,i,i',j}$ The volume supplied by the supplier i to node j through the pipeline of the supplier i' in season $k(\text{m}^3)$.

Binary Variables

$B_{i,j,j'}$ Pipeline construction binary variables, $B_{i,j,j'} = 1$ if supplier i will construct a pipeline between nodes j and j', or else $B_{i,j,j'} = 0$.

$C_{i,j}^{station}$ Station construction binary variables, $C_{i,j}^{station} = 1$ if supplier i will construct a station at node j, or else $C_{i,j}^{station} = 0$.

$C_{i,j}^{storage}$ Binary variables of underground gas storage construction, $C_{i,j}^{storage} = 1$ if supplier i will construct a storage at node j, or else $C_{i,j}^{storage} = 0$.

1 Introduction

1.1 Background

Natural gas, as clean and green energy, has been widely used in various countries around the world [1]. In China, the use of natural gas is gradually increasing [2]. According to the statistical data of national energy administration, the natural gas

consumption of China has increased from $245 \times 10^8 m^3$ to $3025 \times 10^8 m^3$, which indicates that China's natural gas market has entered a stage of rapid development. Nowadays, the multi-source pattern consisiting of domestic natural gas, imported LNG, imported pipeline natural gas has been formed in China [3]. Due to the existence of multi-source pattern, the suppliers will compete in the downstream market. To maximize their own interests, each supplier will constantly adjust the strategies and optimize resource allocation. Thus, it is of great significance to consider market competition in the optimization of natural gas supply chain.

As the main transportation mode of natural gas, pipelines are safe and economical, which makes most of the suppliers adopt this mode to transport natural gas [4]. A well-designed supply chain system should be able to resist outside interference and ensure the secure downstream supply [5–7], which will increase the profits of energy companies. Therefore, each supplier should not only ensure the stable supply in the downstream market while constructing the supply chain system, but also consider how to obtain the optimal pipeline layout scheme under the market competition, so as to minimize the construction cost and maximize their own profits.

1.2 Related Work

Currently, most of the research on the optimization of the natural gas supply chain focused on the supply security [8–10], price mechanism [11, 12] and so on. Among them, the studies on the optimal distribution of single supplier to demand side are the majority [13, 14], while the optimization of natural gas supply chain under market competition is rarely considered. Some scholars have worked out the integrated optimization of natural gas supply chain which mainly includes the pipeline layout scheme and detailed supply scheme [15, 16]. Since the operational optimization of natural gas networks is a complex nonlinear problem, Zhang et al. [17] introduced a multi-period optimization framework for the large scale natural gas network. On this basis, a non-convex MINLP model is established to improve the economic performance of the network. Azadeh et al. [18] considered the impact of greenhouse gas produced during natural gas production and transportation on the environment, and then presented a dual-objective fuzzy linear programming model with minimum economic cost and minimum greenhouse gas emission cost to optimize a natural gas supply chain. Finally, the solution was evaluated. Aiming at a small-scale LNG supply system in coastal area, Jokinen et al. [19] developed a MILP model that minimized the cost of fuel purchase to obtain the optimal distribution scheme. By taking the minimum daily cost including transportation cost, liquefaction cost, purchase cost and construction cost as the objective function, Zhang et al. [20] put

forward a three-stage stochastic programming method to optimize infrastructure development and inventory routing of the LNG supply system.

There are also studies on the game between different levels of natural gas supply chain [21–23]. According to the link-based flexible network allocation rule, Nagayama et al. [24] developed a network game model to analyze the relative power structure among gas-trading countries. Castillo et al. [25] proposed a framework of LNG supply chain with multiple levels, and established a multi-objective model based on the game theory. Considering the competition among third-party marketers, Steven et al. [26] put forward a linear complementarity model for the North American natural gas market to obtain the demand distribution scheme. Daniel et al. [27] proposed a dynamic upstream gas supply model which could stimulate the investment and operation decision of the upstream gas industry. By imposing an agent-based framework, Guo et al. [28] developed a global-scale model Gas-GAME to analyze the impact of US liquefied natural gas exportation strategies on global gas market. The above research mainly concentrated on policy analysis, but did not optimize the design of supply chain system.

In summary, the previous works on the optimization of natural gas supply chain were relatively mature, with a wide range of research directions and rich achievements. However, at present, few studies account for the optimization of natural gas supply chain under multi-source pattern. Aiming at this issue, a universal MILP model is proposed for the design and operation of the natural gas supply chain under multi-source. Through solving the model, the construction scheme of infrastructure and the detailed seasonal supply and storage scheme can be obtained.

2 Problem Description

2.1 Background

Similar to other fuel supply chains, the structure of natural gas supply chain also can be divided into three parts [29, 30]. The upstream part of natural gas supply chain is the gas production or import, the middle part mainly includes the transportation and storage of natural gas, and the downstream part is the sale of natural gas products [14]. The problem studied in this paper is based on the supply chain system in Fig. 1. It can be seen that there are multiple natural gas suppliers in upstream part. Each supplier relies on its own gas fields to produce natural gas or import natural gas from abroad to supply the downstream market, thus forming the pattern of multi-party competition. The market is the downstream part of natural gas supply chain, and its demand may fluctuate due to the influence of seasonal factors. To satisfy the seasonal demand of the downstream market and reduce the risk caused by demand fluctuation, it is indispensable to construct the underground gas storage near the market [31].

Fig. 1 A natural gas supply chain system under multi-source
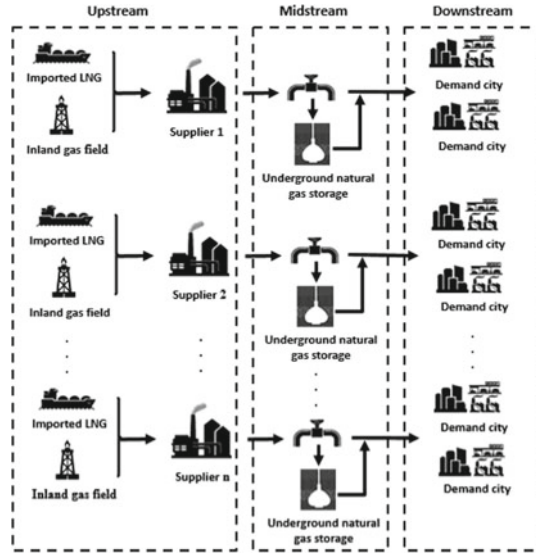
Transportation is the main part of the midstream of natural gas supply chain, which has large impact on economy [32]. Natural gas can be transported by many different modes, among which pipeline is the most used one [33]. Pipeline has a large transport capacity and requires less space [34]. In particular, it is not affected by any extreme weather conditions and can operate continuously for a long time [35, 36].

2.2 Model Requirement

Given:

- The seasonal demand of each city.
- The maximum production and external production of each gas field.
- Geographical locations of gas fields, cities, existing pipelines, stations, and underground gas storages.
- Various costs, including station construction cost, pipeline construction cost and maintenance cost, gas storage construction cost and so on.
- The annual depreciation rate of each equipment.
- The maximum and minimum capacity of underground gas storages.

Determine:

- Construction scheme of infrastructure of each supplier.
- Seasonal supply scheme and purchase scheme of natural gas of each supplier.
- Seasonal storage and peak shaving volume of each gas storage.

Objective:

The objective function is to maximize the profit of each supplier under market competition. To develop the model effectively, this paper proposes the following assumptions:

- The gas fields and cities are set as one node.
- The energy consumption cost is considered to be linearly related to the distance and the flow.
- The difference of gas quality between different gas sources is not considered.

3 Mathematical Model

3.1 Objective Function

In this paper, the infrastructure investment of each supplier is dynamically depreciated in the way of equivalent payment by installments. The objective function $(\max F_i = f_{i,1} - (f_{i,2} + f_{i,3} + f_{i,4} + f_{i,5} + f_{i,6} + f_{i,7} + f_{i,8}))$ of the model is to maximize the total profit of each supplier. The total profit includes the annual net profit on sales of natural gas $(f_{i,1} = \sum_{k \in K} \sum_{j \in J} \sum_{i' \in I} VS_{k,i,i',j}(p_{k,i}^{sale} - co_i^{production})\ i \in I)$, the annual depreciation cost of station construction $(f_{i,2} = co^{station}dy^{station} \sum_{j \in J} C_{i,j}^{station}\ i \in I)$, the annual depreciation cost of pipeline construction $(f_{i,3} = co^{pipe}dy^{pipe} \sum_{j \in I \cup J} \sum_{j' \in I \cup J} B_{i,j,j'} l_{j,j'}\ i \in I)$, the annual maintenance cost of pipeline $(f_{i,4} = co^{maintenace} \sum_{j \in I \cup J} \sum_{j' \in I \cup J} (B_{i,j,j'} + s_{i,j,j'}^{pipe})l_{j,j'}\ i \in I)$, the annual cost of energy consumption $(f_{i,5} = co^{energy} \sum_{k \in K} \sum_{j \in I \cup J} \sum_{j \in I \cup J} Q_{k,j,j'} l_{j,j'}\ i \in I)$, the annual depreciation cost of storage construction $(f_{i,6} = co^{storage}dy^{storage} \sum_{j \in J} C_{i,j}^{storage}\ i \in I)$, the purchase cost for natural gas $(f_{i,7} = \sum_{k \in K} p_{k,i}^{purchase} V_{k,i}^{purchase}\ i \in I)$ and the transportation cost $(f_{i,8} = \sum_{k \in K} \sum_{i' \in I} \sum_{j \in J} p_{i'}^{pipe} VS_{k,i,i',j}\ i \in I, i \neq i')$.

3.2 Pipeline Construction Constraints

If there is a pipeline between nodes j and j', it is not indispensable to construct a new one, as stated in constraints (1–2). Equation (3) indicates that no pipeline will be constructed among the gas sources. To avoid the repeated construction, if the supplier i has already constructed the pipeline between node j and node j', other

suppliers cannot construct more(i.e., (4)). The topology of pipeline network considered in this paper is branch pipeline network, so the relationship between the number of pipeline segments and nodes can be expressed by Eq. (5).

$$B_{i,j,j'} \leq 1 - \sum_{i \in I} s_{i,j,j'}^{pipe} \quad i \in I j, j' \in I \cup J \tag{1}$$

$$\sum_{j' \in I \cup J} \left(B_{i,j,j'} + \sum_{i \in I} s_{i,j,j'}^{pipe} \right) \geq 1 \quad i \in I j, j' \in I \cup J \tag{2}$$

$$B_{i,j,j'} = 0 \quad i,j,j' \in I \tag{3}$$

$$\sum_{i \in I} \left(B_{i,j,j'} + B_{i,j',j} \right) \leq 1 \quad i \in I j, j' \in I \cup J, j \neq j' \tag{4}$$

$$\sum_{i \in I} \sum_{j \in I \cup J} \sum_{j' \in I \cup J} \left(B_{i,j,j'} + s_{i,j,j'}^{pipe} \right) = |I \cup J| - 1 \tag{5}$$

3.3 Volume Constraints

The variable $VS_{k,i,i',j}$ indicates the volume supplied by the supplier i to node j through its own pipeline or other suppliers' pipeline in the season k. Specially, if $(C_{i,j}^{station} + s_j^{station})$ equals 0, $VS_{k,i,i',j}$ will equal 0 (i.e., (6)). The supplier i can supply natural gas to node j through its own pipeline when the supplier i has constructed the pipeline to node j ($\sum_{j' \in I \cup J} B_{i,j,j'} = 1$). Or else, if the supplier i' has constructed the pipeline to node j, the supplier i can adopt the pipeline of the supplier i' to supply, as stated in constraint (7).

$$\sum_{k \in K} VS_{k,i,i',j} \leq \left(C_{i,j}^{station} + s_j^{station} \right) M \quad k \in K, i \in I, i = i', j \in J \tag{6}$$

$$VS_{k,i,i',j} \leq \sum_{j' \in I \cup J} (B_{i',j,j'} + s_{i',j,j'}^{pipe}) M \quad k \in K, j \in J i, i' \in I \tag{7}$$

As stated in constraint (8), the purchased natural gas volume of the supplier i in the season k should be less than the maximum purchased volume of the supplier $i(vpr_{k,i}^{max})$. The summation of the total volume of supplier i for each node through its

own pipeline and the supply through other suppliers' pipelines ($\sum_{k \in K} \sum_{i' \in I} \sum_{j \in J} VS_{k,i,i',j}$) should not be greater than the maximum supply of supplier i (i.e., (9)).

$$V_{k,i}^{purchase} \leq vpr_{k,i}^{\max} \quad k \in K, i \in I \tag{8}$$

$$\sum_{k \in K} \sum_{i' \in I} \sum_{j \in J} VS_{k,i,i',j} \leq sv_i^{\max} + \sum_k V_{k,i}^{purchase} \quad i \in I \tag{9}$$

Only if there is a storage in node j ($C_{i,j}^{storage} + s_j^{storage} = 1$) can the storage be operated for injection or export (i.e., (10)). As stated in constraint (11), the peak shaving volume ($VTF_{k,i,j}$) should be within the allowable volume of the underground gas storage. The demand of node j in the season $k(d_{k,j})$ equals to the summation of the volume supplied by each supplier through its own pipeline and other suppliers' pipelines ($\sum_{i \in I} \sum_{i' \in I} VS_{k,i,i',j}$) and the peak shaving volume of storage ($\sum_{i \in I} VTF_{k,i,j} n^{season}$) (i.e., (12)).

$$|VTF_{k,i,j} n^{season}| \leq (C_{i,j}^{storage} + s_j^{storage})M \quad k \in K, i \in I, j \in J \tag{10}$$

$$vos_j^{\min}(C_{i,j}^{storage} + s_j^{storage}) \leq vtf_j^{initial} - \sum_{k \in K} VTF_{k,i,j} n^{season} \leq vos_j^{\max} \quad i \in I, j \in J \tag{11}$$

$$d_{k,j} = \sum_{i \in I} VTF_{k,i,j} n^{season} + \sum_{i \in I} \sum_{i' \in I} VS_{k,i,i',j} \quad k \in K, j \in J \tag{12}$$

3.4 Flowrate Constraints

The transportation of natural gas considered in the model is unidirectional transportation. If there is no pipeline between the two nodes, the pipeline flowrate will equal to 0(i.e., (13)). Or else, the flowrate must have boundaries for economic concerns (i.e., (14)). The flowrate out of the gas source is calculated through Eq. (15). Equation (16) indicates that the volume into node j minus the delivery volume is equal to the volume out of node j.

$$Q_{k,j,j'} \leq fl_{j,j'}M \quad k \in K j, j' \in I \cup J \tag{13}$$

$$(fl_{j,j'} - 1)M + q_{j,j'}^{\min} \leq Q_{k,j,j'} \leq (1 - fl_{j,j'})M + q_{j,j'}^{\max} \quad k \in K j, j' \in I \cup J \tag{14}$$

$$\sum_{j' \in I \cup J} Q_{k,i,j'} = \sum_{i' \in I} \sum_{j' \in I \cup J} VS_{k,i,i',j}/24 n^{season} \quad k \in K, i \in I, j \in J \tag{15}$$

$$\sum_{j' \in I \cup J} Q_{k,j',j} - (\sum_{i \in I} V_{k,i,j} + \sum_{i \in I} \sum_{i' \in I} VS_{k,i,i',j})/24n^{season} = \sum_{j' \in I \cup J} Q_{k,j,j'} \quad k \in K, j \in J$$

(16)

3.5 Station and Storage Construction Constraints

If the station has been constructed at node j, the supplier i cannot construct more. To avoid the repeated construction, only one supplier can construct station at node j(i.e., (17)–(18)). Only the city that satisfies special geological conditions can construct the storage (i.e., (19)). Similar to the station construction, constraints (20)–(21) limit the storage construction of supplier i.

$$C_{i,j}^{station} \le 1 - s_j^{station} \quad i \in I, j \in J$$

(17)

$$\sum_{i \in I} C_{i,j}^{station} \le 1 \quad j \in J$$

(18)

$$C_{i,j}^{storage} \le sc_j^{storage} \quad i \in I, j \in J$$

(19)

$$C_{i,j}^{storage} \le 1 - s_j^{storage} \quad i \in I, j \in J$$

(20)

$$\sum_{i \in I} C_{i,j}^{storage} \le 1 \quad j \in J$$

(21)

4 Methodology

This paper imposes particle swarm optimization (PSO) algorithm to solve the model. PSO method is one of the intelligent algorithms, which has been widely applied to many engineering optimization problems since it was proposed [37, 38]. The calculation process of the method is shown as follow. Firstly, the number of suppliers should be determined, that is, the number of elements in set I. For different suppliers, we can divide the universal MILP model into several MILP models. Each MILP model corresponds to one supplier, and the model can be decomposed into a pipeline layout model and a gas allocation model. Finally, PSO algorithm is utilized to randomly generate the initial supply of each supplier ($i = 2, \ldots \ldots, i_{max}$) for demand cities, and the objective function of MILP model of supplier i_{max} is set as the fitness function (max $F_{i_{max}} = f_{i_{max},1} - (f_{i_{max},2} + f_{i_{max},3} + f_{i_{max},4} + f_{i_{max},5} + f_{i_{max},6} + f_{i_{max},7} + f_{i_{max},8})$). Through the iterative solution, the optimal infrastructure construction scheme and gas supply scheme can be obtained. The overall structure of the proposed method is shown as Fig. 2.
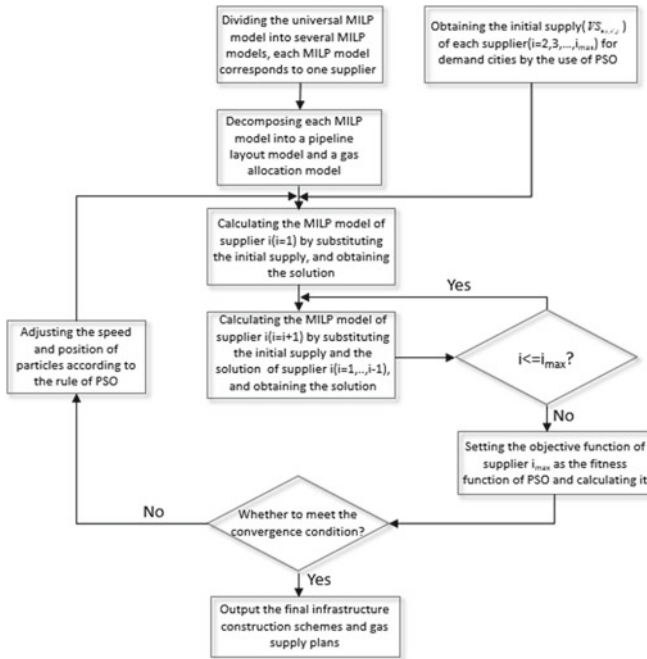
Fig. 2 Framework of the proposed method

5 Results and Discussion

5.1 Basic Data

This paper optimizes a real natural gas supply chain system in China. The Positions and numberings of all nodes are shown in Fig. 3. This system includes two gas fields S1 and S2, ten demand cities from D1 to D10. In this system, the demand of each city mainly is supplied by the two gas fields S1 and S2 that belong to supplier 1 and supplier 2 respectively. The seasonal demand of each city is shown in Fig. 4. Currently, pipelines from S1 to D1 and from D1 to D2 have already existed, and stations of these nodes have been constructed. Considering the geological conditions of each city, D2, D3, D8, and D9 are chosen to build underground natural gas storages. The maximum of underground gas storage capacity is 10×10^8 m^3 of which the base load is 20%. A year is divided into four seasons with equal duration. In the paper, the depreciation rate of each equipment is considered as 7.1%. The construction cost of infrastructure is shown in Table 1. The maximum annual production and purchased volume of each supplier are shown in Table 2. The case is programmed by Matlab R2014a and solved by Gurobi 7.5.1 MILP solver.
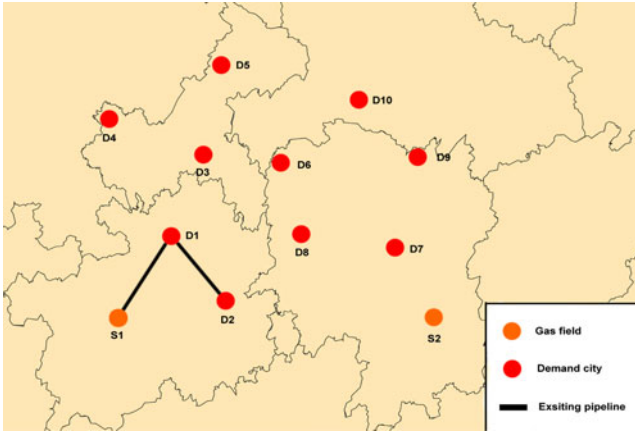
Fig. 3 Positions and numberings of nodes

Table 1 Construction cost of infrastructure

Name	Unit	Price
Station construction cost ($co^{station}$)	CNY	5×10^8
Storage construction cost ($co^{storage}$)	CNY	10^8
Unit price of pipeline construction (co^{pipe})	CNY/km	10^7
Unite price of pipeline maintenance ($co^{maintenace}$)	CNY/km	15,000
Unit price of energy consumption for pipeline transportation (co^{energy})	CNY/ (km·m³/h)	0.0168

Table 2 Basic parameters of each source

Source	Maximum supply volume (10^6m³)	Maximum purchased volume (10^6m³)
S1	9,000	200
S2	13,000	300

Fig. 4 Demands of cities
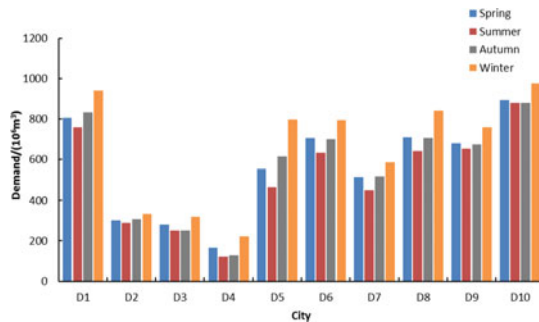
5.2 Case Study

The convergence curve of PSO algorithm is shown in Fig. 5. As can be seen from the figure, the algorithm basically starts to converge at the 12th iteration, and then reaches the minimum value. Table 3 shows the calculation results of model.

The optimized structure of studied supply chain is shown in Fig. 6. It can be seen that supplier 1 constructs pipelines from D2 to D8, D6 to D3, D3 to D4. Pipelines from S2 to D7, D7 to D8, D8 to D6, D6 to D5, D6 to D10, D10 to D9 are constructed by supplier 2. Pipelines constructed by each supplier can connect into a network, improving the flexibility of the integral transportation system. The detailed construction scheme of each supplier is shown in Table 4. As shown in table, supplier 1 constructs stations at D3 and D4, underground gas storages at D2 and D3, and supplier 2 constructs stations at D5–D10.

Figure 7 shows the purchased schemes of supplier 1 and supplier 2 at different seasons. Due to the limitation of gas field production, each supplier also needs to import natural gas from abroad or purchase gas from other land fields, so as to satisfy downstream demand. It can be seen that each city has high demand for natural gas in the winter. To meet downstream demand, supplier 1 purchases natural gas with volume of $150 \times 10^6 m^3$ and $200 \times 10^6 m^3$ in the autumn and the winter

Fig. 5 Convergence curve of PSO method
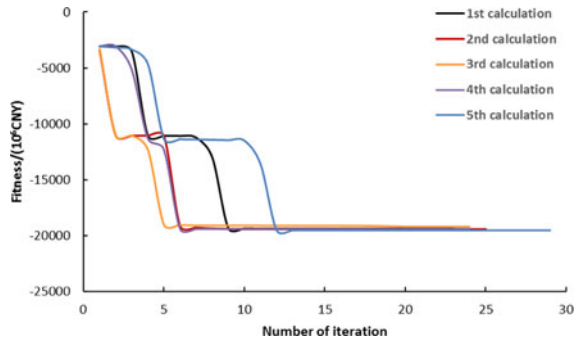
Table 3 Calculation results

Supplier	1	2
F/(10^7CNY)	1324.55	1936.13
f_1/(10^7CNY)	1598.67	2285.36
f_2/(10^7CNY)	7.14	21.43
f_3/(10^7CNY)	59.00	135.79
f_4/(10^7CNY)	0.089	0.20
f_5/(10^7CNY)	1.58	11.81
f_6/(10^7CNY)	1.43	–
f_7/(10^7CNY)	42.00	180.00
f_8/(10^7CNY)	162.88	–

respectively, while supplier 2 purchases natural gas with volume of $300 \times 10^6 m^3$ in each season. The gas supply schemes of supplier 1 and supplier 2 are shown in Fig. 8. As can be seen from Fig. 6, supplier 1 transports gas to D1-D4 through its own pipelines and supplies gas to D6 and D8 through the pipelines constructed by supplier 2, while supplier 2 only supplies gas to D5-D10 by its own pipelines. The total volume supplied by each supplier can satisfy the demand of the downstream cities.
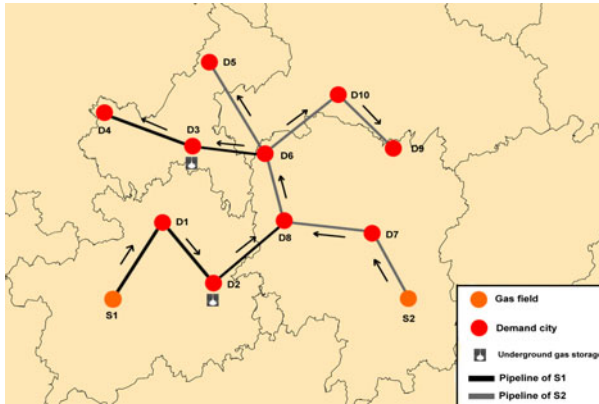
Fig. 6 Construction scheme of each supplier

Table 4 Construction scheme of each supplier

Supplier	1	2
Length of pipeline (km)	826	1901
Number of station	2	6
Number of underground gas storage	2	–

Fig. 7 Purchase schemes of suppliers
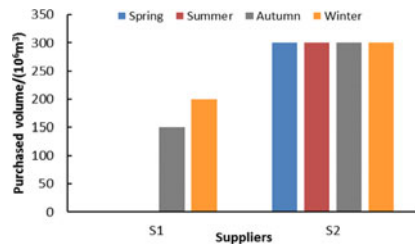
The optimization result indicates that supplier 1 builds underground gas storages at D2 and D3 for peak shaving in winter. The storage schemes of underground gas storages D2 and D3 are shown in Fig. 9, where the negative value represents the injection scheme and the positive value represents the production scheme. It can be seen that there will be shifts of storage plans in different seasons. In the summer, amounts of natural gas will be injected into storage and stored for use in the winter or next spring. As can be seen from Fig. 8 and Fig. 9, part of the supply of D2 and D3 from supplier 1 is injected into the storage in the summer, with the volume being $10.41 \times 10^6 m^3/d$ and $6.46 \times 10^6 m^3/d$, respectively. During the spring and the winter, the storages at D2 and D3 export natural gas with the volume being $3.72 \times 10^6 m^3/d$ and $3.12 \times 10^6 m^3/d$, respectively, to meet the demand of D2 and D3 in these two seasons.
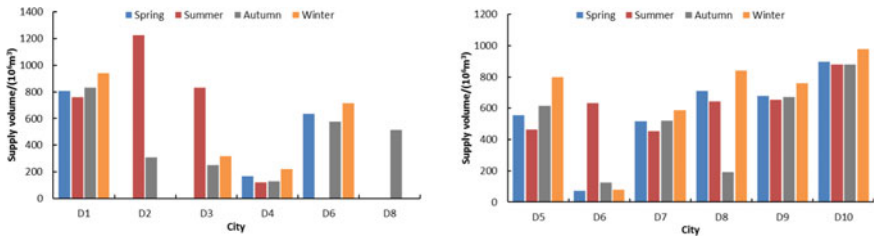
Fig. 8 Supply schemes of suppliers

Fig. 9 Storage schemes of underground gas storages
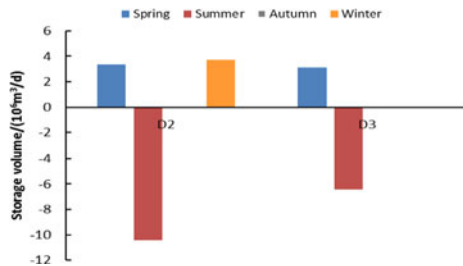
6   Conclusion

This paper puts forward a universal MILP model for the design and operation of natural gas supply chain under multi-source. Infrastructure construction cost and transportation cost are considered in the model. Based on the above cost, this model takes the maximum profit of each supplier as the objective function, and considers the constraints of balance between supply and demand, pipeline construction, station construction, underground reservoir construction and so on. In the paper, PSO algorithm is imposed to solve the model. Finally, a real case is presented to illustrate the applicability of the proposed model. The result shows that the solution can converge to a better value in an acceptable time. For the future works, the authors intend to take the uncertainties of supply chain into consideration.

References

1. Liang, T., et al.: Refined analysis and prediction of natural gas consumption in China. J. Manag. Sci. Eng. **4**(2), 91–104 (2019)
2. Wang, T., Lin, B.: China's natural gas consumption peak and factors analysis: a regional perspective. J. Clean. Prod. **142**, 548–564 (2017)
3. Wang, T., Lin, B.: China's natural gas consumption and subsidies—from a sector perspective. Energy Policy **65**, 541–551 (2014)
4. Wang, B., et al.: Multi-objective site selection optimization of the gas-gathering station using NSGA-II. Process Saf. Environ. Prot. **119**, 350–359 (2018)
5. Wang, B., et al.: Sustainable refined products supply chain: a reliability assessment for demand-side management in primary distribution processes. Energy Sci. Eng. **8**, 1029–1049 (2019)
6. Yuan, M., et al.: Downstream oil supply security in China: policy implications from quantifying the impact of oil import disruption. Energy Policy **136**, 111077 (2020)
7. Zhou, X., et al.: Future scenario of China's downstream oil supply chain: low carbon-oriented optimization for the design of planned multi-product pipelines. J. Clean. Prod. **244**, 118866 (2020)
8. Su, H., et al.: A method for the multi-objective optimization of the operation of natural gas pipeline networks considering supply reliability and operation efficiency. Comput. Chem. Eng. **131**, 106584 (2019)
9. Zarei, J., Amin-Naseri, M.R.: An integrated optimization model for natural gas supply chain. Energy **185**, 1114–1130 (2019)
10. Yu, W., et al.: A methodology to quantify the gas supply capacity of natural gas transmission pipeline system using reliability theory. Reliab. Eng. Syst. Saf. **175**, 128–141 (2018)

11. Rioux, B., et al.: The economic impact of price controls on China's natural gas supply chain. Energy Econ. **80**, 394–410 (2019)
12. Geng, J.-B., Ji, Q., Fan, Y.: The behaviour mechanism analysis of regional natural gas prices: a multi-scale perspective. Energy **101**, 266–277 (2016)
13. Hamedi, M., et al.: A distribution planning model for natural gas supply chain: a case study. Energy Policy **37**(3), 799–812 (2009)
14. Zhang, H., et al.: Optimal design and operation for supply chain system of multi-state natural gas under uncertainties of demand and purchase price. Comput. Ind. Eng. **131**, 115–130 (2019)
15. An, J., Peng, S.: Layout optimization of natural gas network planning: synchronizing minimum risk loss with total cost. J. Nat. Gas Sci. Eng. **33**, 255–263 (2016)
16. An, J., Peng, S.: Prediction and verification of risk loss cost for improved natural gas network layout optimization. Energy **148**, 1181–1190 (2018)
17. Zhang, W.W., et al.: Multi-period operational optimization of natural gas treating, blending, compressing, long-distance transmission, and supply network. In: Eden, M.R., Ierapetritou, M.G., Towler, G.P., (eds.) Computer Aided Chemical Engineering, pp. 1249–1254. Elsevier (2018)
18. Azadeh, A., Raoofi, Z., Zarrin, M.: A multi-objective fuzzy linear programming model for optimization of natural gas supply chain through a greenhouse gas reduction approach. J. Nat. Gas Sci. Eng. **26**, 702–710 (2015)
19. Jokinen, R., Pettersson, F., Saxén, H.: An MILP model for optimization of a small-scale LNG supply chain along a coastline. Appl. Energy **138**, 423–431 (2015)
20. Zhang, H., et al.: A three-stage stochastic programming method for LNG supply system infrastructure development and inventory routing in demanding countries. Energy **133**, 424–442 (2017)
21. Massol, O., Tchung-Ming, S.: Cooperation among liquefied natural gas suppliers: is rationalization the sole objective? Energy Econ. **32**, 933–947 (2010)
22. Csercsik, D., et al.: Modeling transfer profits as externalities in a cooperative game-theoretic model of natural gas networks. Energy Econ. **80**, 355–365 (2019)
23. Gong, C., et al.: An optimal time-of-use pricing for urban gas: a study with a multi-agent evolutionary game-theoretic perspective. Appl. Energy **163**, 283–294 (2016)
24. Nagayama, D., Horita, M.: A network game analysis of strategic interactions in the international trade of Russian natural gas through Ukraine and Belarus. Energy Econ. **43**, 89–101 (2014)
25. Castillo, L., Dorao, C.: Decision-Making on Liquefied Natural Gas (LNG) projects using game theory (2011)
26. Gabriel, S.A., Zhuang, J., Kiet, S.: A large-scale linear complementarity model of the North American natural gas market. Energy Econ. **27**(4), 639–665 (2005)
27. Crow, D.J.G., Giarola, S., Hawkes, A.D.: A dynamic model of global natural gas supply. Appl. Energy **218**, 452–469 (2018)
28. Guo, Y., Hawkes, A.: Simulating the game-theoretic market equilibrium and contract-driven investment in global gas trade using an agent-based method. Energy **160**, 820–834 (2018)
29. Zhang, W., et al.: A stochastic linear programming method for the reliable oil products supply chain system with hub disruption. IEEE Access **7**, 124329–124340 (2019)
30. Zhou, X., et al.: A two-stage stochastic programming model for the optimal planning of a coal-to-liquids supply chain under demand uncertainty. J. Clean. Prod. **228**, 10–28 (2019)
31. Yu, W., et al.: Gas supply reliability analysis of a natural gas pipeline system considering the effects of underground gas storages. Appl. Energy **252**, 113418 (2019)

32. Yuan, M., et al.: Future scenario of China's downstream oil supply chain: an energy, economy and environment analysis for impacts of pipeline network reform. J. Clean. Prod. **232**, 1513–1528 (2019)
33. Wang, B., et al.: An MILP model for the reformation of natural gas pipeline networks with hydrogen injection. Int. J. Hydrogen Energy **43**(33), 16141–16153 (2018)
34. Wang, B., et al.: Optimisation of a downstream oil supply chain with new pipeline route planning. Chem. Eng. Res. Des. **145**, 300–313 (2019)
35. Liu, E., Changjun, L., Yang, Y.: Optimal energy consumption analysis of natural gas pipeline. Sci. World J. **2014**, 506138 (2014)
36. Ríos-Mercado, R.Z., Borraz-Sánchez, C.: Optimization problems in natural gas transportation systems: a state-of-the-art review. Appl. Energy **147**, 536–555 (2015)
37. Zhang, H., et al.: An improved PSO method for optimal design of subsea oil pipelines. Ocean Eng. **141**, 154–163 (2017)
38. Zhang, H., et al.: A risk assessment based optimization method for route selection of hazardous liquid railway network. Saf. Sci. **110**, 217–229 (2018)

Encrypted Malware Traffic Detection Using TLS Features and Random Forest

Weilong Lee and Shuyuan Jin

Abstract Encrypted malware traffic detection has been an essential issue in network security. With more and more malwares using TLS to encrypt their traffic, most payload-based methods can't work well as before. Several works have been conducted in this field. To improve performance, this paper analyzes and proposes several new TLS features to analyze the characteristics of encrypted malwares and further detects their traffic. The analysis shows that during the TLS sessions, most malwares use less TLS extensions and cipher suites and often use self-signed certificates. Based on the result of analysis, this paper proposes to use random forest to identify TLS encrypted malware traffic. Experimental results show that when combined with proposed features, the TLS features can achieve higher recall rate.

Keywords Malware traffic detection · Traffic classification · Random Forest

1 Introduction

With the development of internet and information security technology, malwares have become an essential threat on the internet. Accordingly, many intrusion detection systems (IDS) have been built to resist them. However, to protect their communication, many applications start to encrypt their traffic. For example, until 2018, about 80% of Google web traffic are transported on TLS [1]. Meanwhile, malwares also start to utilize TLS to encrypt their traffic. Moreover, according to Z-scaler report, more and more malwares adopt TLS to encrypt their traffic, for example, about 60% of banking trojans leverage TLS to encrypt their traffic [2]. Since most IDSs inspect the content of packets to recognize malicious flows, they

W. Lee · S. Jin (✉)
School of Computer Science and Engineering, Sun Yat-Sen University, Guangdong, China
e-mail: jinshuyuan@mail.sysu.edu.cn

W. Lee
e-mail: liwl23@mail2.sysu.edu.cn

will fail when meeting encrypted traffic. New methods are required to detect encrypted malicious traffic.

Although the use of TLS makes payload-based detection fail, it also brings some observable features that can be used in traffic detection. In this paper, we research features extracted from TLS sessions and analyze their effectiveness in malicious traffic detection. The contributions of this paper are as follows:

1. This paper proposes several new TLS features and TCP features for encrypted malware traffic detection. Like existing work, TLS features are mainly extracted from the packets in handshake process and TCP features are extracted from the first 10 packets after handshake. This paper analyzes the distribution of TLS features and find that the differences are quite obvious.
2. This paper proposes an encrypted malware traffic detection method based on random forest. To explore the effectiveness of proposed method, we conduct several experiments on MFCP dataset.

The remainder of this paper is organized as follows Sect. 2 reviews related work on malware traffic detection. Section 3 introduces features used in this paper. Section 4 introduces random forest algorithm and feature encoding method. Section 5 presents experiment results. Section 6 discusses limitations of our work and points out the future work. Finally, Sect. 7 summarizes this paper and gives the conclusion.

2   Related Work

Malware traffic detection has been studied for many years. The goal of malware traffic detection is to identify traffic flows generated by malwares. This problem has profound significance in network security. With the power of identifying malware traffic, we can protect a network from many network-related malwares, such as botnet, worms, and so on.

Malware traffic detection methods are mainly inspired from traffic classification which identifies the applications that generate the traffic. There used to be two kind of popular methods in traffic detection, port-based method and payload-based method. Port-based methods utilized those well-known port arranged by IANA, such as 23 (telnet), 80 (http) to associate port number with specific application. However, as applications use dynamic port negotiation which allows applications use any port they want, port-based method can't work well. In literature, Thomas Karagiannis et al. found that using port number to recognize p2p traffic is much inaccuracy [3]. Duffield et al. also proposed several limitations of port-based methods [4]. So, port-based methods are not suitable for traffic classification. To address the drawbacks of port-based methods, payload-based methods were pro-posed. In these methods, packet payloads are inspected to check whether they match any signatures of known applications. Many studies have shown that these

approaches work very well. But when encountering TLS encrypted traffic, payload-based method fails, so researchers start to find new solutions for encrypted traffic. There are two solutions for TLS encrypted traffic classification. One is to use TLS proxy to hijack the TLS sessions, another is using flow statistics information. In the first solution, the TLS encrypted traffic is decrypted by TLS proxy and payload-based methods can be applied. While this solution can be successful, there are some shortcomings. Decrypting TLS message does not only invade the privacy of users, but also causes heavy overhead.

Most current methods utilize flow statistics information. These methods extract features from TCP flows, such as inter arrival time and use machine learning to learn from them. Many methods have been proposed based on variant traffic classification and achieved perspective result. Li Jun, et al. use flow features and random forest to identify Skype traffic, and get accuracy of 96.22% [5]. Wei Wang, et al. classify 12 kinds of different traffic with one-dimensional convolutional neural network [6, 7]. Although the result shows the accuracy is lower than other method, neural network can generate feature automatically. Chao Wang et al. try to improve random forest, and their method can get much better result [8]. Although these flow-based methods can achieve high accuracy, some features, such as average packet length must be calculated when the flow finished. This can also result in high complexity. To solve the problem, Yang Liu et al. proposed to use the information from the first N packets in a flow to classify traffic [9]. They use a slide-window to extract features. The experiment result showed that their accuracy can achieve 96% and the method has lower time cost. There are many other studies using different machine learning algorithm, such as, SVM [13, 16], Bayesian analysis [10], clustering algorithm [11, 12], Markov chain [14] and so on [17, 18]. Although flow statistics perform well on traffic classification, TLS features are also a choice for research. The state of art method in malware detection is from Anderson Blake et al. In their work, they use TLS handshake information to detect malware and get accuracy of 99.6% and when they combine TLS features with flow contextual flow data, they get about accuracy of over 99.98% [15]. Although their precision is so high, the calculations of some features such as SPLT (sequence of packet lengths and times) and BD (byte distribution), are time-consuming, and the dimension of their features is very high. In this paper, we try to use some lightweight features from TLS sessions to identify TLS malware traffic.

3   Features

In this section, this paper introduces features used in experiments. These features can be classified into two categories, TLS handshake features and TCP flow features.

3.1 TLS Features

TLS features are extracted from TLS handshake. As traffic during TLS handshake is not encrypted, many information, like certificate, cipher suites can be extracted easily.

Table 1 lists all the TLS features used in this paper. Among them, six features are proposed in [15], and seven features are proposed by this paper. To analysis these features more detailly, this paper classifies them into three categories TLS Client features, TLS Server feature and TLS Certificate features.

Table 1 TLS features used in experiment

	Feature	Descriptions
Features used in [15]	Offered cipher suites	Cipher suites offered by clients in client hello message
	Offered extensions	TLS extensions offered by client in client hello message
	Selected cipher suites	Cipher suites selected by server in server hello message
	Selected extensions	TLS extensions selected by server in server hello message
	Cert validation time	The total validation time of server certificate
	#san DNS	The number of domain name in DNS field of Subject Alternative Name extension
Features proposed in this paper	Highest TLS version	The highest TLS version supported by client
	#offered cipher suites	The amount of cipher suites offered by client
	#offered extension	The amount of TLS extensions offered by client
	Selected TLS version	TLS version selected by server
	#selected extensions	The amount of TLS extensions selected by server
	Cert path length	The length of certificate chain offered by server
	Self-signed cert	Whether the certificate of server is self-signed
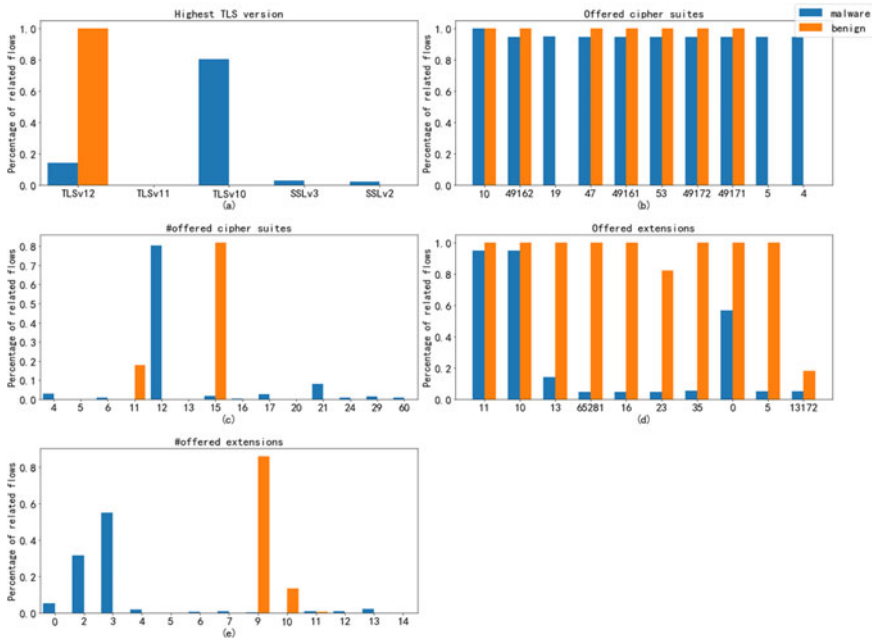
Fig. 1 The distribution of client features

TLS Client Features

TLS client features are extracted from client hello message. These features mainly contain communication parameters provided by client. Figure 1 illustrates the differences between malware sessions and normal sessions on client features.

As we can see in Fig. 1(a), nearly 100% of benign clients offer TLS 1.2, while about 80% clients of malware clients offer TLS 1.0. TLS 1.0 has been obsolete and has been considered as unsafe, and TLS 1.2 is the latest version. The security of TLS 1.2 is higher than TLS 1.0, so this suggests malwares are prone to use unsafe version.

As for cipher suites supported by client, in Fig. 1(b), we can see, about 90% of malware sessions offered cipher suites: 4 (TLS-RSA-WITH-RC4-128-MD5), 5 (TLS-RSA-WITH-RC4-128-SHA) and 19 (TLS-DHE-DSS-WITH-3DES-EDE-CBC-SHA), while few benign sessions offered them. The reason may be that the digest algorithms of them are weak. The distribution of cipher suites is similar to [15]. In addition to the types, there is evident difference in the number of offered cipher suites. As illustrated in Fig. 1(c), most malware clients offer no more than 13 cipher suites, while 80% benign clients offer more than 13 cipher suites.
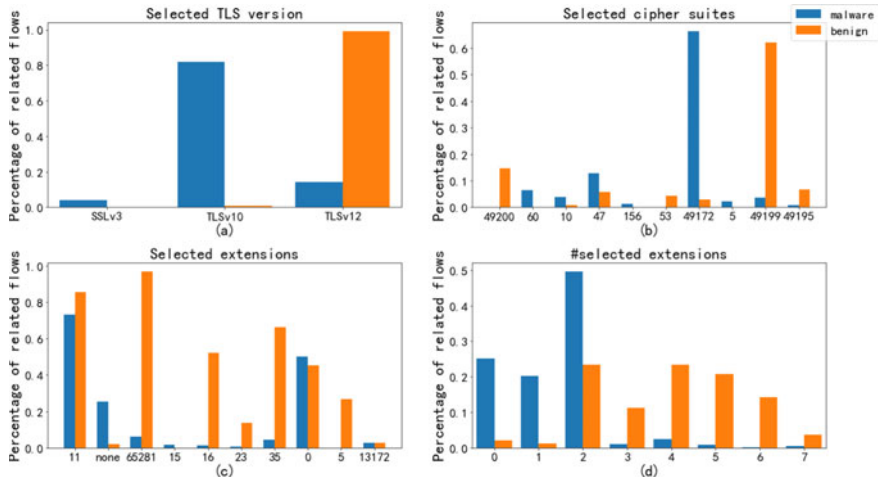
Fig. 2 The distribution of server features.

When it comes to TLS extensions, the differences are more evident. On the one hand, more than 80% of benign software offer extensions 13 (signature-algorithms), 65,281 (renegotiation-info), 16 (application-layer-protocol-negotiation), 23 (extended-master-secret), 5 (status-request) and 35 (session-ticket), while few malware clients support them. We also find that the distribution of these TLS extensions is a little different from that in [15] where the extensions mostly used by benign are 10 (supported-groups), 11 (ec-point-formats), 13 (signature-algorithms), 23 (extended-master-secret), 65,281 (renegotiation-info). But there is a common fact that malwares tend to support less extensions. As illustrated in Fig. 1(a), most benign software offers 9 extensions while most malwares offer 3.

In summary, most malwares use obsolete TLS and support less extensions and cipher suites than benign software, which suggests that the confidentiality of malwares is not as strong as benign software.

TLS Server Features

TLS server features are extracted from server hello message. Figure 2 illustrates the difference on distribution of TLS server features between malicious sessions and begin sessions

As illustrated in Fig. 2(a), malware servers tend to choose TLS 1.0, while almost 100% benign servers choose TLS 1.2. During the analysis, we observed an interesting phenomenon that in some sessions, the highest version supported by client is SSL 2.0, while the server chooses SSL 3.0. This violates the standard of TLS. We inspect related sessions and find that all these sessions are generated by malwares.

There are evident differences in the selected TLS extensions. On the one hand, most benign servers select 65,281, while just 5% of malware servers choose it. On the other hand, the number of extensions selected by benign servers is much bigger.
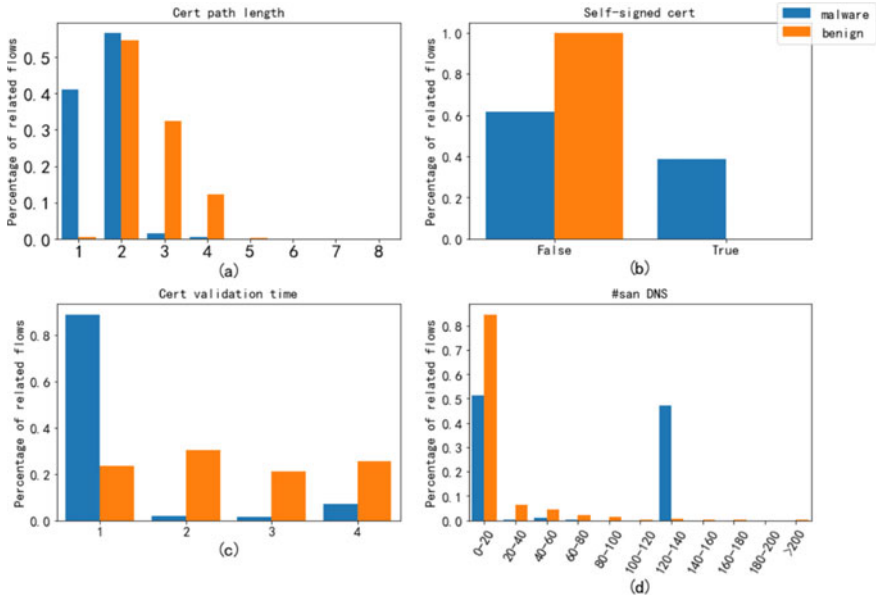
Fig. 3 The distribution of certificate features

As we can see, nearly 90% malware servers select no more than two extensions, while 90% benign servers select more than two.

When considering selected cipher suites, the differences are also evident. Most malware server selected 49,199 (TLS-ECDHE-RSA-WITH-AES128-GCM-SHA 256), 49,200 (TLS-ECDHE-RSA-WITH-AES256-GCM-SHA384) and 49,195 (TLS-ECDHE-ECDSA-WITH-AES128-GCM-SHA256), while benign software selects 49,172 (TLS-ECDHE-RSA-WITH-AES-256-CBC-SHA), 49,171 (TLS-EC DHE-RSA-WITH-AES-128-CBC-SHA) and 47 (TLS-RSA-WITH-AES128-CBC-SHA). The main difference is that the cipher suites selected by normal server use GCM mode while cipher suites selected by malware server use CBC mode. Compared with CBC mode, GCM mode can provide extra authority. This may be the reason why normal server prefer GCM mode. In addition to the mode, the digest algorithm is another difference. Malware servers like to use SHA-1, while normal servers prefer SHA-2. As SHA-1 has been considered as unsafe and was cracked [22], most applications have abandoned it and choose SHA-2.

In summary, malware servers tend to choose obsolete TLS version, unsafe cipher suite and less TLS extensions.

Server Certificate Features

Certificate features are extracted from certificates in server certificates message. These features show the identity of the TLS server. Figure 3 illustrates the differences between the certificates of benign servers and that of malware servers.

As illustrated in Fig. 3(a), the certificate chain length of benign software are generally greater than that of malwares. For malware sessions, the length of certificate chains is no more than 2, while that of most benign server is no less than 2 and about 45% benign servers offer certificate chains with length over 3.

The second feature is self-signed certificate. Using a self-signed certificate means the identity of server is not authorized by CA (certificate authority), which means the server is not trustable. As we can see, almost no benign servers use self-signed certificates, while about 40% malware use self-signed certificate.

The third feature is the total validation time of certificate. As for this feature, we can see evident differences. The validation time of about 90% malware certificates is less than 1 year, while that of about 80% normal certificate is over 1 year.

As for alternative name number, the alternative name numbers of nearly 100% normal certificates are less than 100, while those of about 50% malware certificates are over 100.

In summary, there are differences on certificate features. The length of certificate chain of malwares is smaller and the validation time of malware certificates is shorter.

3.2   TCP Features

TCP features mainly consist of flow statistics such as packet number and byte number during TCP sessions. Table 2 lists all the TCP features used in this paper. The first four features are derived from the whole flow. They have been used in many researches. As these features are generated from the whole flow, we can't get them until the TLS session ends. To overcome this issue, we proposed to use some lightweight features, the statistics of the first 10 packets in applications messages in TLS sessions. They can be calculated in the early stage of TLS sessions.

In experiment, we extract statistics of first 10 packets after TLS handshake. Similar features have been studied in many researches, most researchers use features during the whole TCP flows and get promising result. But in this paper, we try to choose light-weight features, so we choose the features of the first 10 packets after TLS handshake. The effectiveness of first N packets has been demonstrate in [9]. The reason why we choose the first 10 packets after TLS handshake are as followed. First, we have extracted many features from TLS handshake, so there is no need to consider the packets appearing in TLS handshake. Second, as said in [9], applications may do some initialization in the first 10 packets, so the statistics of the first 10 packets after TLS handshake may contain initialization information of the application.

Table 2 TCP features used in this paper

	Features	Description
Features proposed in [15]	in_bytes	Total number of inbound bytes during the TLS session
	out_bytes	Total number of outbound bytes during the TLS session
	in_pkt	Total number of inbound packets during the TLS session
	out_pkt	Total number of outbound packets during the TLS session
Features proposed in this paper	avg_arvl	Average inter-arrival time of the first 10 packets after TLS handshake
	std_arvl	Standard deviation of inter-arrival time of the first 10 packets after TLS handshake
	max_arvl	Max inter-arrival time of the first 10 packets
	min_arvl	Min inter-arrival time of the first 10 packets after TLS handshake
	avg_len	Average length of the first 10 packets after TLS handshake
	std_len	Standard deviation of length of the first 10 packets after TLS handshake
	max_len	Max length of the first 10 packets after TLS handshake
	min_len	Min length of the first 10 packets after TLS handshake

4 Classifier and Feature Encoding

This section introduces the machine learning algorithm, Random Forest (RF) and the encoding method for features.

4.1 Random Forest

Random forest is an ensemble learning method for classification and regression [19]. In random forest, a multitude of decision trees are built and are trained with random subset of features and samples. The mode of classes of each tree will be the result of the random forest. Random forest has been researched in traffic classification for many years, and demonstrates promising performance. The reasons why we choose random forest are as follows:

1. Random forests have high stability against noise in dataset.
2. Random forests have low bias and variance and it is hard to overfit.
3. Random can evaluate the importance of each feature.
4. The time cost of training is relatively low.

In this paper, we employ the random forest implementation in scikit-learn [23] and use Gini impurity to measure the quality of a split in each decision tree.

$$allciphers :< cipher1, cipher2, \ldots cipher10 >$$
$$< cipher5, cipher7, cipher10 > \xrightarrow{Encode} (0, 0, 0, 0, 1, 0, 1, 0, 0, 1)$$

Fig. 4 An example for encoding cipher suites list

4.2 Feature Encoding

Among features used in this paper, some are numerical and others are discrete features. For numerical features, we feed them into classifier directly. For discrete features, we first apply bag-of-word to encoding them before further processing.

Concretely, features like the list of offered cipher suites, are discrete and we have to convert them to numeric. For those lists, we represent them by a binary vector, each bit in the vector is associate with an element and is set to 1 when related element appears in the list. For example, in Fig. 4, we have 10 ciphers in total. Then we get a list of three cipher suites <cipher5, cipher7, cipher10> , so the encoding result is a bit vector with bit 5, bit 7 and bit 10 set.

After feature encoding, all features can be represented as real numbers, and can also be further processed.

5   Evaluation

To evaluate the effectiveness of proposed TLS features, we conduct experiments on MCFP dataset. We first evaluate different combinations of TLS features, certificate features and flow features and then explore the optimal value for hyper-parameters.

5.1 Dataset

The dataset used in this paper comes from a public malware traffic dataset MCFP [20] (Malware Traffic Project). This dataset has been collecting malware traffic since 2012. And until now, this dataset has collected 384 malware traffic captures and 18 normal traffic captures in the form of PCAP files. As our experiment only needs traffic encrypted by TLS, we only choose those captures with TLS sessions. Finally, we get 41 malicious captures and 18 normal captures.

After that, Bro [21] is adopted to extract features of each TLS sessions in these captures. Bro is an open-source, effective network monitor, its log framework and

script engine provide convenience for traffic analysis. With the help of Bro, we filter those captures and get 282,174 TLS sessions.

Among the 282,174 sessions, there are 215,167 sessions generated by malwares and 67,007 normal sessions. As 103,969 sessions are established by session resume and don't contain TLS handshake information, we just discard them. At last, we get 135,933 malicious sessions and 42,272 benign sessions for our experiment.

5.2 Cross Validation

Based on random forest and the features extracted, we can build a classifier to identify malware traffic. In this section, we will demonstrate the effectiveness of the classifier by using 10-fold cross validation. The performance metrices used in experiment are precision and recall. In these equations, TP means true positive, FN means false negative and FP means false positive.

$$recall = \frac{TP}{TP + FN} \qquad (1)$$

$$precision = \frac{TP}{FP + TP} \qquad (2)$$

To validate the effectiveness of our features, different feature combinations are experimented and compared. We divide all the features into 4 categories, TLS feature proposed in [15], TLS features that we proposed, TCP features that are used in 15 and TCP features proposed in this paper, and we name them TLS, newTLS, TCP, newTCP respectively. Then we experiment with different combinations of these categories. The combinations used in this paper are as follows, TLS, TLS +newTLS, TCP, TCP+newTCP, TLS+TCP, newTLS+newTCP, ALL. For each feature combination, we perform 10-fold cross validation on the dataset and calculate the average of precision and recall in the 10 folds. The experimental results are listed in Table 3.

In the first comparison experiment, we compare the effectiveness between TLS and TLS+newTLS. As we can see, when only using TLS handshake features, we can get average precision of 99.64% and recall of 95.78%. When combining TLS with newTLS, the average recall increases 2.46%.

In the second comparison experiment, we compare the effectiveness between TCP and TCP+newTCP. As we can see, when only using TCP, we can get precision of 92.07% and recall of 93.56%. When combining TCP and newTCP, the precision and recall get evident increasement, with average precision increasing 3.43% and average recall 1.2%.

Table 3 Result of 10-fold cross validation

		1	2	3	4	5	6	7	8	9	10	Avg
TLS	Precision (%)	99.96	99.54	99.72	99.95	99.95	99.93	99.79	98.38	99.24	99.93	99.64
	Recall (%)	95.22	96.22	95.67	95.45	95.52	96.34	95.61	96.48	95.88	95.41	95.78
TLS +newTLS	Precision (%)	99.97	99.95	99.98	99.95	99.96	99.94	99.97	99.95	99.97	99.93	99.96
	Recall (%)	99.42	96.26	99.66	95.74	99.36	99.60	99.60	99.43	96.49	96.80	98.24
TCP	Precision (%)	92.33	91.90	92.00	92.23	91.66	92.10	92.11	91.97	91.97	92.35	92.07
	Recall (%)	93.27	94.08	93.23	93.36	93.29	94.04	93.51	93.40	94.09	93.31	93.56
TCP +newTCP	Precision (%)	96.21	94.75	95.09	96.39	96.30	94.11	95.54	95.55	95.21	95.84	95.50
	Recall (%)	94.24	94.82	95.05	94.37	94.00	95.07	95.24	94.48	95.86	94.44	94.76
TLS+TCP	Precision (%)	99.63	99.74	99.16	99.66	99.60	99.48	99.57	99.70	99.75	99.82	99.61
	Recall (%)	97.86	98.74	98.49	96.67	97.67	95.91	97.24	95.72	96.50	94.95	96.97
newTLS +newTCP	Precision (%)	99.67	99.63	99.55	99.61	99.57	99.70	99.54	99.57	99.55	99.61	99.60
	Recall (%)	98.88	98.96	98.72	98.91	98.85	98.84	98.78	98.69	99.00	98.89	98.85
ALL	Precision (%)	99.95	99.96	99.99	99.96	99.95	99.93	99.98	99.95	99.98	99.93	99.96
	Recall (%)	91.69	99.67	99.60	99.19	99.40	99.75	99.67	99.49	99.40	99.27	99.51

In the last comparison experiment, we compared TLS+TCP, newTLS+newTCP and ALL. As we can see, when using all the features, we can get the best performance. And the combination of our new features can achieve precision of 99.60%. Although the performance is not as good as features in 15, the feature vector is much shorter.

In this section, we compare the performance of different feature combinations. The result shows that, when combined with old features, proposed TLS features can improve the recall. And when just using those new features, although the performance is a little worse than the old features, the number of old features is much less. Because in old features, there are many lists such as the list of cipher suits, these features are encoded to long vector, which results in large amounts of features. In new features, we just use a number for each list and get less features.

5.3 Hyper-parameter Search

There are many hyper-parameters in random forest and different settlement may result in different performance. Table 4 lists several hyper-parameters in random forest. To find the optimal value of them, we explore their impact on performance. In this section, we use ROC AUC as the performance metric. For each hyper-parameter, we set other parameters as default and try it with different values. Figure 5 illustrates the effect of 4 hyper-parameters on the performance of random forest.

Table 4 Four hyper-parameters of random forest in scikit-learn

Parameter	Description	Optimal
n_estimators	The number of decision trees in the forest	12
max_depth	The maximum depth of each decision tree	7
min_samples_leaf	The minimum number of samples required to be at a leaf node	0.002
min_samples_split	The minimum number of samples required to split an internal node	0.002
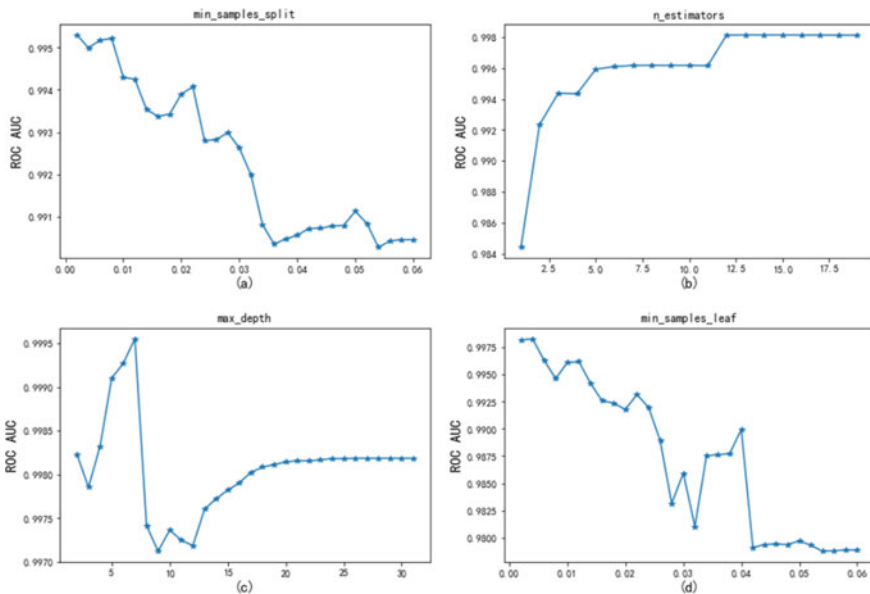
Fig. 5 The effect of each parameter on ROC AUC

Figure 5(a) illustrates the impact of min_samples_split on the performance. As we can see, the larger this hyper-parameter is set, the poorer performance we get. When it increases, the AUC ROC declines gradually. When it is set to 0.002, the best performance is achieved. Although it is possible that smaller value can achieve higher AUC, the AUC is high enough. So, we choose 0.002 as its optimal value.

Figure 5(b) illustrates the impact of n_estimators on the performance. It can be seen that more estimators can result in higher AUC. When n_estimators is less than 13, increasing its value can improve ROC AUC. While n_estimators is larger than 13, increasing it doesn't have obvious effects. So, we choose 13 as its optimal value.

Figure 5(c) illustrates the impact of max_depth. As we can see, when max_depth is less than 7, increasing it may promote AUC, but when it is larger than 7, increasing it will lead AUC firstly decline and then increase. But no matter how we increase it, the AUC never exceeds 0.9995. So, we choose 7 as its optimal value.

Figure 5(d) illustrates the impact of min_samples_leaf. It is clear that there is also a negative correlation between min_samples_leaf and AUC. When min_samples_leaf increases, the ROC AUC decreases. So, we choose 0.002 as its optimal value.

Apart from these 4 parameters, we also utilize the same method to search for the optimal value of other parameters. As their effect is not evident, we don't report them.

6   Discussion

This section discusses the limitations and points out several work directions.

6.1   Limitations

Although the experiments achieve promising results, all the features in this paper are easy to change. Most features are extracted from TLS handshake, and depend on the TLS libraries used by malwares. If malware authors are aware that the TLS version of their malware is obsolete, they can change their code and rebuild malwares with latest TLS libraries. Then distribution of most features will become similar to benign software and our method cannot identify their traffic. After all, the detection and antidetection are a process of confrontation.

6.2   Future Work

During experiments, we encounter several problems and remain them as future work.

The first one is about dataset. Although researches on malwares have been going on for decades, there is no public traffic dataset that is large enough. Here, the large means large number of malware samples and malware kinds. On the one hand, the traffic should be generated by many malware samples. On the other hand, the traffic should be generated by malware of many kinds. Without large dataset, performance evaluation may be not precise enough.

The second one is about certificate features. Digital certificate usually contains many information, like domain name. With information and some techniques, much more information can be digested. For example, for domain names, we can consider

their *whois* information and other information. These kinds of information have been used in malicious domain detection and may help to detect malicious certificates.

The last one is about adversarial examples. As we have discussed in the last subsection, most features in this paper are easy to forge. In the field of machine learning, those forged samples that are aimed at bypassing detection are named adversarial examples. With adversarial examples, researchers can evaluate the robustness of their method and build more effective method. So, how to generate adversarial examples in malware traffic is also a work direction.

7 Conclusions

With more and more malwares encrypting their traffic with TLS, traditional payload-based methods cannot detect malware traffic effectively. Noticing the use of TLS introduces new features that can be used to detect malware, this paper proposes new features from TLS sessions. With the analysis on them, it can be found that malware sessions are prone to use obsolete TLS version, fewer extensions and fewer cipher suites. To validate the effectiveness of proposed features, we conduct experiment on MCFP dataset with random forest by 10-fold cross validation. The experiment results show that when combined with old features, our features can achieve higher precision and recall, which identifies the effectiveness of proposed features.

References

1. Google. HTTPS encryption on the web [EB/OL]. https://transparencyreport.google.com/https/overview?hl=en. Accessed 3 Apr 2018
2. Deepen Desai. SSL/TLS-based malware attacks [EB/OL]. https://www.zscaler.com/blogs/research/ssltls-based-malware-attacks. Accessed 08 Aug 2017/05 Apr 2018
3. Karagiannis, T., Broido, A., Brownlee, N., et al.: Is P2P dying or just hiding? [P2P traffic measurement]. In: Global Telecommunications Conference, 2004. GLOBECOM 2004, vol. 3, pp. 532–1538. IEEE. (2004)
4. Duffield, N.G., et al.: Statistical, signature-based approach to IP traffic classification. U.S. Patent No. 7,660,248. 9 February 2010
5. Li, J., Zhang, S., Ye, X., et al.: Identifying skype traffic by random forest. In: International Conference on Wireless Communications, NETWORKING and Mobile Computing, pp. 2841–2844. IEEE (2007)
6. Wang, W., et al.: End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In: IEEE International Conference on Intelligence and Security Informatics, pp. 43–48. IEEE (2017)

7. Wang, W., et al.: Malware traffic classification using convolutional neural network for representation learning. In: 2017 International Conference on Information Networking (ICOIN). IEEE (2017)
8. Wang, C., Xu, T., Qin, X.: Network traffic classification with improved random forest. In: International Conference on Computational Intelligence and Security, pp. 78–81. IEEE (2016)
9. Liu, Y., et al.: A novel algorithm for encrypted traffic classification based on sliding window of flow's first N packets. In: 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA). IEEE (2017)
10. Moore, A.W., Zuev, D.: Internet traffic classification using Bayesian analysis techniques. ACM SIGMETRICS Perform. Eval. Rev. 33(1). ACM (2005)
11. Bernaille, L., et al.: Traffic classification on the fly. ACM SIGCOMM Comput. Commun. Rev. 36(2), 23–26 (2006)
12. Erman, J., Arlitt, M., Mahanti, A.: Traffic classification using clustering algorithms. In: Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data. ACM (2006)
13. Nelms, T.L., Hobson, A., Ward, J.: Systems and methods for traffic classification. U.S. Patent No. 9,571,511. 14 February 2017
14. Shen, M., et al.: Certificate-aware encrypted traffic classification using Second-Order Markov Chain. In: IEEE/ACM, International Symposium on Quality of Service, pp. 1–10. IEEE (2016)
15. Anderson, B., McGrew, D.: Identifying encrypted malware traffic with contextual flow data. In: Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. ACM (2016)
16. Kondo, S., Sato, N.: Botnet traffic detection techniques by C&C session classification using SVM, pp. 91–104. In: International Workshop on Security. Springer, Heidelberg (2007)
17. Homayoun, S., et al.: BoTShark: a deep learning approach for botnet traffic detection. Cyber Threat Intell. 137–153 (2018)
18. Zulkifli, A., et al.: Android malware detection based on network traffic using decision tree algorithm. In: International Conference on Soft Computing and Data Mining. Springer, Cham (2018)
19. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forest for classification, regression, density estimation, manifold learning and semi-supervised learning. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/decisionForests_MSR_TR_2011_114.pdf. Accessed 02 2016/03 Apr 2018
20. Garcia, S., Grill, M., Stiborek, J., Zunino, A.: Malware Capture Facility Project [DB/MT]. https://www.stratosphereips.org/datasets-overview/. Accessed 28 Mar 2018/01 Apr 2018
21. Paxson, V.: Bro: a system for detecting network intruders in real-time. Comput. Netw. 31(23–24), 2435–2463 (1999)
22. Stevens, M., et al.: The first collision for full SHA-1. In: Annual International Cryptology Conference. Springer, Cham (2017)
23. Wang, X., Yu, H.: How to Break MD5 and Other Hash Functions (2005)
24. Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)

Determination of the Radiative Properties of the TiO$_2$ Porous Thin Films in the UV-VIS Spectral Range

Rui Qi, Rong Chen, Junming Zhao, Qiang Liao, Xun Zhu, and Dingding Ye

Abstract TiO$_2$ porous thin film has been widely used as the photoanode of the photocatalytic fuel cell. The distribution of the light intensity in this type of the photoanode greatly affects its performance, which can be obtained by solving the radiative transfer equation. Under this circumstance, the determination of the radiative properties of the TiO$_2$ porous thin films is crucial. In this work, based on the measured thickness, the normal-hemispherical reflectance and transmittance of the porous thin films, the absorption coefficient and scattering coefficient were determined by using the spectral element method. It was found that the TiO$_2$ porous thin films exhibited high scattering coefficient and relatively low absorption coefficient. The absorption coefficient sharply decreased in the UV region with increasing the wavelength and then was almost unchanged in the visible-light range. The scattering coefficient firstly increased and then decreased. There existed a maximum scattering coefficient at a wavelength of about 378 nm. Moreover, the effect of the porosity was investigated by adding the pore former of PMMA in the porous thin film. The results indicated that the scattering coefficient remarkably increased with increasing the porosity, while the absorption coefficient was insensitive to the porosity variation.

Keywords TiO$_2$ porous thin films · Porosity · Reflectance and transmittance · Absorption and scattering coefficients · Spectral element method

R. Qi · R. Chen (✉) · Q. Liao · X. Zhu · D. Ye
Key Laboratory of Low-Grade Energy Utilization Technologies and Systems, Chongqing University, Ministry of Education, Chongqing 400030, China
e-mail: rchen@cqu.edu.cn

R. Qi · R. Chen · Q. Liao · X. Zhu · D. Ye
Institute of Engineering Thermophysics, School of Energy and Power Engineering, Chongqing University, Chongqing 400030, China

J. Zhao (✉)
School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China
e-mail: jmzhao@hit.edu.cn

1   Introduction

In recent years, photocatalytic fuel cells that combine the advantages of both the photocatalytic and fuel cell technologies, have received increasing attention, because they can directly convert chemical energy stored in wastewater into electricity via the photoelectrochemical reactions [1–4]. This new technology has exhibited significant potentials in the sewage treatment and the production of clean energy simultaneously. Although photocatalytic fuel cells have been widely investigated, there are still many problems to be solved, such as low photonic efficiency, serious electron-hole recombination, low photocurrent density, and low power output, etc. These problems are clearly related to the photoanode so that extensive efforts have been devoted to the development of high-performance photoanode. Actually, the transport characteristics in the photoanode also greatly affect its performance because the reactant/product, light, ion, electron and hole simultaneously transport in the photoanode and are coupled with the photoelectrochemical reactions. Unfortunately, the transport mechanism still remains unclear, such as the light transfer, which restricts further development of the photoanode.

One of the approaches to understand the transport characteristics in the photoanode is the simulation method. In particular, the light transport, which is a primary factor for initiating the photoelectrochemical reaction in the photoanode, significantly affects the accuracy of the simulation. In conventional, an empirical formula has been widely used to simulate the light transport in the photoanode [5–7]. Although the empirical formula is simple and easy to be incorporated into the model, it is unable to objectively describe the light transport in the photoanode because the influence of light scattering is ignored but considerably strong in the nano-porous structure [8]. Hence, to develop a more accurate transport model for the photoanode, it is essential to accurately describe the distribution of light intensity in the photoanode. Solving the radiative transport equation (RTE) is one of the best methods. However, in order to solve this equation, two important radiative parameters must be obtained, namely the absorption coefficient and the scattering coefficient of the photoanode. Currently, these two parameters of the photoanode are unavailable, limiting the development of the transport model.

In fact, there exist many methods to determine the radiative parameters of porous media. In 1984, Makino et al. [9] obtained the absorption coefficients and scattering coefficients of ceramics materials using the Kubelka–Munk method. Subsequently, Molenaar et al. [10] applied this method to other porous media. Meanwhile, Pickering et al. and Cheong et al. [11, 12] used the inverse adding-doubling (IAD) method to estimate the optical properties of various biological tissues by the experimentally determined reflectance and transmittance. At present, this method has been widely used in the calculation of optical parameters of chaotic media. In 2006, Petrasch et al. [13] calculated the radiative properties such as the extinction coefficient and scattering phase functions of porous media by Monte Carlo method. Xie et al. [14] also utilized this method to investigate the radiative properties of ZnO particles and ZnO-Au composite particles in suspension.

In addition, Morales et al. [15] used the spectral element method to determine the reflection and transmission coefficients of collimated light in a multiple slabs. Overall, although there are many works on the determination of radiative parameters of porous media, far too little attention has been paid to the radiative properties of the TiO_2 porous thin films.

In this work, therefore, we prepared typical porous thin films using the commercial TiO_2 nanoparticles, which have been widely used in the photoanode. The normal-hemispherical reflectance, normal-hemispherical transmittance and thicknesses of the TiO_2 porous thin films were measured, which were then employed to retrieve the radiative properties, i.e., the absorption coefficient and the scattering coefficient of the porous thin film. During the retrieving, the RTE was solved using the spectral element method [16] and combined with the Levenberg-Marquardt algorithm. On the other hand, the effect of porosity on the variation of the spectral absorption and scattering coefficients was also examined. The achieved results are helpful for further experimental and theoretical research on the transport characteristics of the TiO_2 porous thin films with the consideration of the microstructural effect on the photoanode.

2 Experimental Section

2.1 Preparation of Typical TiO₂ Porous Thin Films

In this work, the typical TiO_2 porous thin films were prepared on the glass slides by the wet spraying method with the TiO_2 colloid [17]. To prepare the colloid, 12 g of TiO_2 commercial nanoparticles (Aeroxide P25, Acros, Belgium) were firstly dispersed to 120 mL of deionized water before adding 0.4 mL acetylacetone (Kelong Chemical Reagent Factory, China). After that, 0.2 mL of Triton X-100 (Solibao Technology, China) and 2.4 g of polyethylene glycol (Kelong Chemical Reagent Factory, China) were added to the mixed solution and continuously stirred in a magnetic stirrer for 12 h. The well-prepared TiO_2 colloid was diluted with ethanol and sprayed on the glass slide by an air spray gun. Finally, the samples were calcined at 550 °C for 2 h to obtain the TiO_2 porous thin films with different loadings, about 0.5 mg/cm^2, 1.0 mg/cm^2, 2.0 mg/cm^2, and 3.0 mg/cm^2, respectively.

2.2 Preparation of TiO₂ Porous Thin Films with Pore Formers

The TiO_2 porous thin films with pore formers were prepared by using TiO_2 colloid and adding the pore formers of PMMA microspheres (Ruige Technology, China) with 1 μm in diameter as templates [18]. First, the aqueous solution of polyvinyl

alcohol (1 wt%) (Kelong Chemical Reagent Factory, China) was dissolved at 98 °C, before 1 mL of the solution was added to 10 mL of the suspension containing 2.3 wt % of PMMA microsphere particles. Next, the prepared PMMA suspension was added to the TiO_2 colloid in terms of the TiO_2/PMMA mass ratios of 1: 0.1, 1: 0.2, 1: 0.4, 1: 0.6, respectively. After stirring for 12 h at room temperature, the mixed solution was sprayed on a glass slide. At last, the samples were calcined at 550 °C for 2 h, and the loadings of the TiO_2 porous thin films with pore formers (termed as TiO_2/PMMA porous thin film) were the same. Each mass ratio had two loadings of about 1.0 mg/cm^2 and 0.4 mg/cm^2. The former was used to measure the microstructural parameters while the later one was used to obtain the normal-hemispherical reflectance and transmittance, because the normal-hemispherical transmittance can only be measured when the thickness of the porous thin film is enough small. Note that given a catalyst loading, the change of the TiO_2/PMMA mass ratio can lead to the variation of the porosity. The larger mass ratio, the smaller porosity.

2.3 Measurements

After the preparation of the porous thin films, their optical and microstructure measurements were performed. The spectral normal-hemispherical reflectance and transmittance of the porous thin films were directly measured by ultraviolet-visible light (UV-VIS) spectrophotometer (PerkinElmer Lambda 950, USA) equipped with a 150-mm integrating sphere. All the TiO_2 porous thin films were illuminated by normal incident beam. The scanning wavelength range was from 300 nm to 800 nm and a blank glass slide was used as the baseline to eliminate the influence of the glass substrate. Every measurement was repeated three times at least to obtain the reliable data. Additionally, the field emission scanning electron microscope (TESCAN MIRA3) was used to obtain the cross-sectional morphologies and microstructures of the prepared porous thin films.

3 Results and Discussion

3.1 Morphologies of the Porous Thin Films

The cross-sectional morphologies of the prepared TiO_2 porous thin films are shown in Fig. 1. It can be seen that although the TiO_2 loadings were different, these porous thin films had similar porous structure with the pore size ranging from several nanometers to hundreds of nanometers (see Fig. 1). Using the cross-sectional views of these porous thin films, we also measured the thicknesses at different locations to estimate the average thickness. In this study, the thicknesses were about 5.3 μm,

9.9 μm, 20.7 μm and 31.0 μm corresponding to the TiO_2 loadings of about 0.5 mg, 1.0 mg, 2.0 mg and 3.0 mg, respectively. It can be found that the thickness of the porous thin film was almost linear with the TiO_2 loading. Therefore, it can be inferred that the porosity and the refractive index remain unchanged for these porous thin films. Based on the TiO_2 density and the measured thickness, the porosity of the TiO_2 porous thin film could be estimated to be about 74.3%. According to the relationship between the equivalent refractive index and the porosity of the porous media [19], the equivalent refractive index was estimated to be 1.572.

Figure 2 shows the morphologies of the cross-sectional views of the TiO_2/ PMMA porous thin films with the loading of about 1.0 mg at various TiO_2/PMMA mass ratios. As can be seen from Fig. 2, after the PMMA particles were added, the pores with about 1 μm in diameter were randomly distributed inside the thin films,

(a)

(b)

(c)

(d)
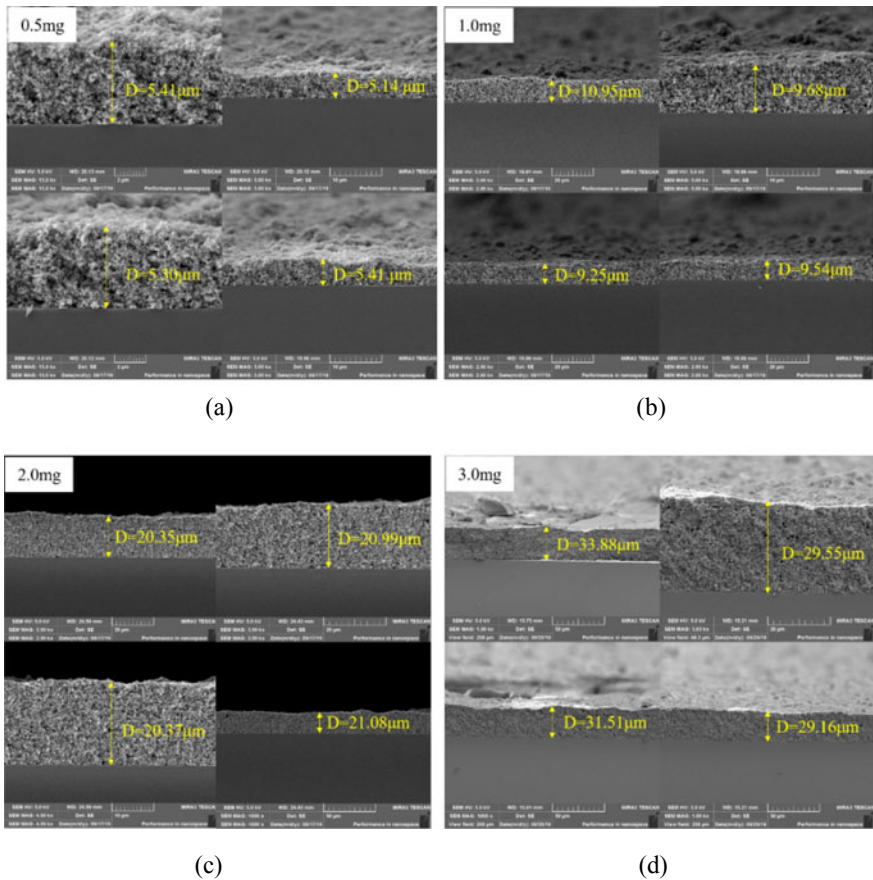
Fig. 1 Morphologies of the cross-sectional views of the TiO_2 porous thin films with different loadings: (a) 0.5 mg; (b) 1.0 mg; (c) 2.0 mg; (d) 3.0 mg
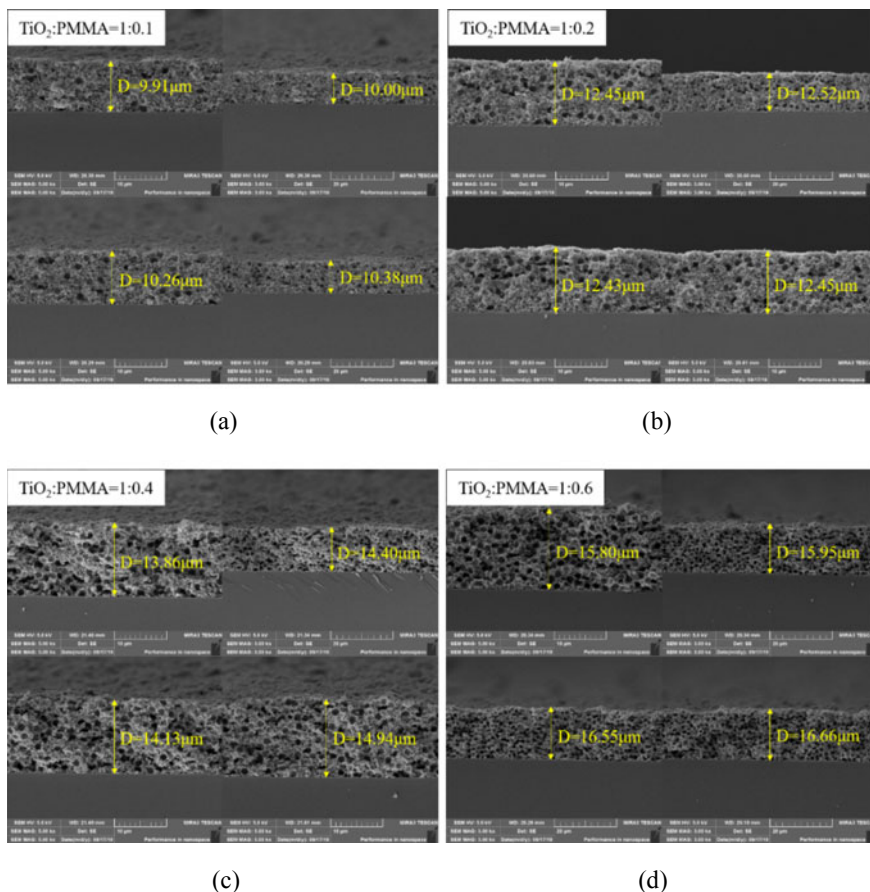
Fig. 2 Morphologies of the cross-sectional views of the TiO_2/PMMA porous thin films with different mass ratios: (a) TiO_2:PMMA = 1:0.1; (b) TiO_2:PMMA = 1:0.2; (c) TiO_2:PMMA = 1:0.4; (d) TiO_2:PMMA = 1:0.6

which were consistent with the diameter of the PMMA particles. This fact indicates that the PMMA particles can be successfully used as the pore former for the preparation of the TiO_2 porous thin films. Moreover, the pore density increased linearly with decreasing the TiO_2/PMMA mass ratio. This is because the decrease of the TiO_2/PMMA mass ratio led to more pore formers to be removed during the calcination, which then created more pores in the porous thin film. According to the cross-sectional images, the thicknesses of the TiO_2/PMMA porous thin films with the loading of about 1.0 mg were estimated to be about 10.1 μm, 12.5 μm, 14.3 μm, 16.2 μm, corresponding to the TiO_2/PMMA mass ratios of 1 : 0.1, 1 : 0.2, 1 : 0.4, 1 : 0.6, respectively. This result indicated that the porosity and thickness of the porous thin film rose with decreasing the mass ratio due to the increased pore volume. Finally, the thicknesses, porosities and equivalent refraction index of

various TiO_2/PMMA thin films were calculated and the results are listed in the Table 1. It can be seen that the equivalent refractive index of the TiO_2/PMMA porous thin film decreases with increasing the porosity.

3.2 Reflectance and Transmittance of the Porous Thin Films

Figure 3 presents the variations in the normal-hemispherical reflectance and transmittance of the TiO_2 porous thin film with the TiO_2 loading. It can be seen that in the ultraviolet spectrum, the normal-hemispherical reflectance of the TiO_2 porous thin film apparently decreased with decreasing the wavelength (see Fig. 3a). There was almost no reflection when the wavelength was below 350 nm. Similarly, the normal-hemispherical transmittance decreased with the decrease of the wavelength, dropping to zero at nearly 350 nm (see Fig. 3b). The reason is that TiO_2 has strong ability to absorb UV light. As the wavelength decreases and the photon energy gradually goes up, the TiO_2 porous thin films are able to absorb more photons so that the transmitted and reflected light significantly reduces. Moreover, the light attenuation increases with the thickness caused by the increased TiO_2 loading, so that the normal-hemispherical transmittance declines, as shown in Fig. 3b.

In the visible-light spectrum (400 nm−800 nm), the normal-hemispherical reflectance decreased whereas the normal-hemispherical transmittance increased with increasing the wavelength, as shown in Fig. 3a and 3b. This result indicates that the intensities of scattering and back-scattering decline when the wavelength increases. In addition, the light scattering in porous media increases due to the increased thickness. In this case, the directly transmitted light is weakened and the back scattering is enhanced. Therefore, the normal-hemispherical transmittance decreases while the normal-hemispherical reflectance increases.

Figure 4 shows the variations in the normal-hemispherical reflectance and transmittance of the TiO_2/PMMA porous thin film with the TiO_2:PMMA mass ratio. It can be seen that the variation trend of the normal-hemispherical reflectance and transmittance shows no significant difference between the TiO_2 and TiO_2/PMMA porous thin films, which can be confirmed by comparing Fig. 4 with Fig. 3.

Another finding is that the normal-hemispherical reflectance of the TiO_2/PMMA porous thin films with different mass ratios (i.e. various porosities, small mass ratio represents large porosity) almost remained identical in the UV spectrum (<390 nm), but increased with increasing the porosity in the visible-light spectrum. With the same TiO_2 loading, the normal-hemispherical transmittance of TiO_2/PMMA porous thin films decreased with increasing the porosity in the whole spectrum. The increase of the porosity could enhance the light scattering in the porous thin film in the visible spectrum, leading to the increase of the back-scattering. As a result, the reflectance increased due to the increase of back-scattering while the transmittance decreased.

Table 1 Microstructural parameters of the TiO$_2$/PMMA porous thin films

Sample	Average thickness (μm)	Pore volume (mm^3/mg)	Porosity	Refraction index
TiO$_2$	9.9	0.736	0.743	1.572
TiO$_2$:PMMA = 1:0.1	10.1	0.756	0.748	1.562
TiO$_2$:PMMA = 1:0.2	12.5	0.981	0.794	1.476
TiO$_2$:PMMA = 1:0.4	14.3	1.179	0.822	1.420
TiO$_2$:PMMA = 1:0.6	16.2	1.370	0.843	1.377

(a)   (b)
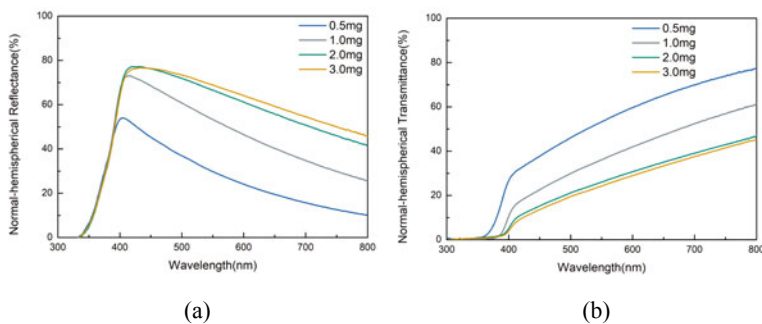
Fig. 3 Variations in (a) the normal-hemispherical reflectance and (b) the normal-hemispherical transmittance of the TiO$_2$ porous thin film with the TiO$_2$ loading

(a)   (b)
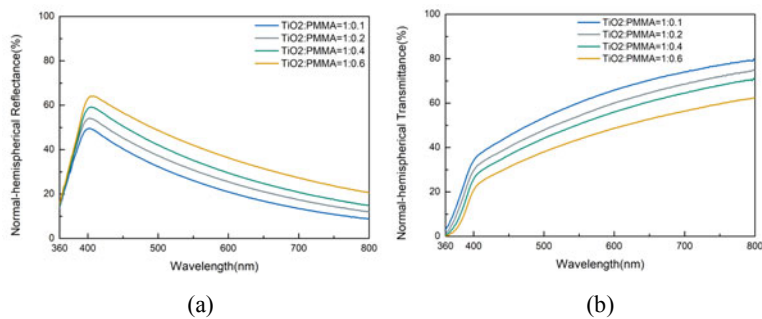
Fig. 4 Variations in (a) the normal-hemispherical reflectance and (b) the normal-hemispherical transmittance of the TiO$_2$/PMMA porous thin film with the TiO$_2$:PMMA mass ratio

3.3 Absorption and Scattering Coefficients

Based on the porosity, equivalent refractive index and the measured normal-hemispherical reflectance and transmittance, the absorption coefficient and the scattering coefficient of the TiO_2 porous thin film could be retrieved by using the spectral element method [16]. It is worth noting that the thickness of the TiO_2 porous thin films should be enough thin to obtain the normal-hemispherical transmittance, so that the TiO_2 porous thin film with a loading of 0.5 mg was only calculated. The result is shown in Fig. 5. It is found that the absorption coefficient reduced sharply from 2703.5 cm^{-1} at 360 nm to 91.4 cm^{-1} at 400 nm and then becomes unchanged in the visible-light region. The reason is that TiO_2, as a semiconductor, whose band gap is 3.2 eV, can only absorb UV light. In the meantime, the absorption ability will increase with decreasing the wavelength. However, this is not the case for the scattering coefficient, which rose to the highest value of 9742.8 cm^{-1} at the wavelength of 378 nm and then dropped exponentially with an increase in the wavelength. It is noticeable that the absorption coefficient was 1692.3 cm^{-1} and the scattering coefficient was 8466.0 cm^{-1} at 365 nm, which are typically used in the photocatalytic fuel cells with TiO_2 as the photoanode.

Using the same approach, the radiative properties of the TiO_2/PMMA porous thin films can also be achieved, as shown in Fig. 6. It can be seen that as the wavelength increased, the trend of scattering coefficient and absorption coefficient of the TiO_2/PMMA porous thin films is the same as those of the TiO_2 porous thin films. For the absorption coefficient, although the increase of the porosity could slightly increase the absorption coefficient (see the inserted figure in Fig. 6a) in the ultraviolet spectrum (<390 nm), a noticeable increase for the scattering coefficient of the TiO_2/PMMA porous thin film could be observed with increasing the porosity. This phenomenon can be attributed to the increased surface due to the increase of pore volume, which effectively strengthens the surface reflection inside the porous

(a)   (b)
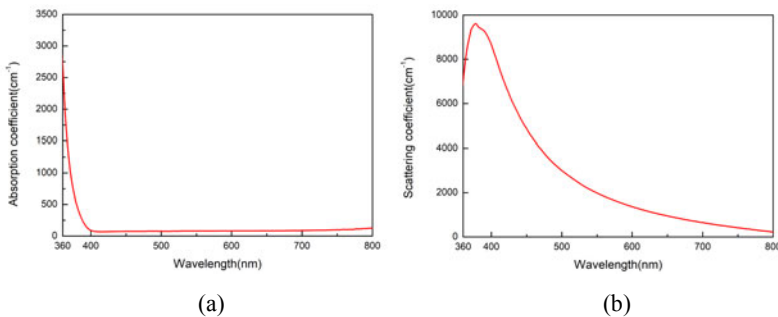
Fig. 5 Radiative properties of the TiO_2 porous thin film: (a) absorption coefficient and (b) scattering coefficient
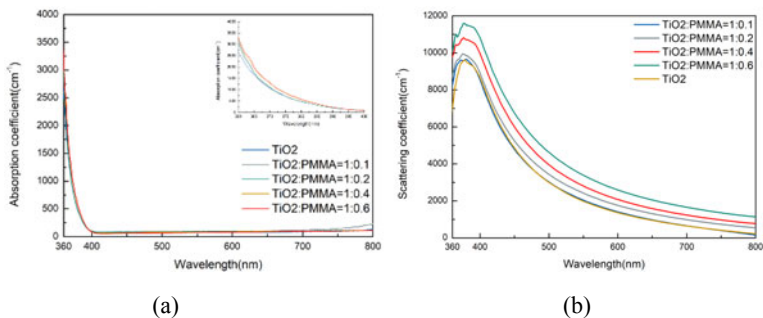
Fig. 6 Radiative properties of the TiO$_2$/PMMA porous thin film: (a) absorption coefficient and (b) scattering coefficient

Fig. 7 Radiative properties of the TiO$_2$ and TiO$_2$/PMMA porous thin films at 365 nm
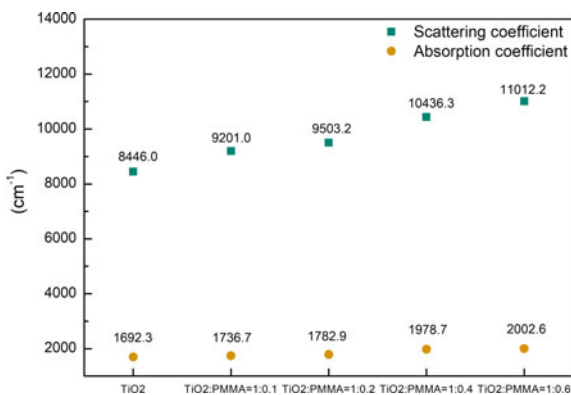
thin film. In this case, the increased scattering coefficient slightly enhanced the absorption. However, because the absorption coefficient is mainly dependent on the nature of TiO$_2$ properties, there was no significant difference between them.

Because TiO$_2$ can only respond to ultraviolet light and is usually illuminated at 365 nm for the TiO$_2$ photoanode, we particularly compared the absorption coefficient and scattering coefficient of the TiO$_2$ and TiO$_2$/PMMA porous thin films at this wavelength to examine the effect of the porosity, as shown in Fig. 7. It can be found that the increased porosity could efficiently enhance the light scattering in the TiO$_2$ porous thin film. Moreover, the scattering coefficient is five times more than the absorption coefficient. Therefore, proper design of the porous structure of the TiO$_2$ photoanode is critically important to improve the performance of the photocatalytic fuel cell.

4 Conclusions

In this paper, the radiative properties of the TiO_2 porous thin films were investigated in the ultraviolet spectrum and visible-light spectrum. We prepared the photoanodes with various loadings and porosities by using TiO_2 nanoparticles and the pore former of PMMA. The cross-sectional views of the porous thin films were characterized, by which the porosity and equivalent refractive index could be estimated. We also measured the normal-hemispherical reflectance and transmittance of the porous thin films. It was found that in the ultraviolet spectrum, the normal-hemispherical reflectance and transmittance of the porous thin film apparently decreased with decreasing the wavelength. The normal-hemispherical reflectance of the porous thin films was similar while the normal-hemispherical transmittance reduced as the TiO_2 loading and the porosity increased (the TiO_2:PMMA mass ratio decreased). In the visible-light spectrum, the normal-hemispherical reflectance decreased whereas the normal-hemispherical transmittance increased as the wavelength increased.

Based on the experimental results, the absorption coefficient and the scattering coefficient could be retrieved by the spectral element method. It is found that the absorption coefficient sharply decreased in the UV light region and then was almost unchanged. For the scattering coefficient, it firstly increased and then dropped. There existed a maximum scattering coefficient at the wavelength of about 378 nm. Moreover, the increase of porosity (the decrease of the TiO_2:PMMA mass ratio) obviously increased the scattering coefficient of the porous thin films, while slight rise occurs for the absorption coefficient. The obtained radiative parameters of the TiO_2 porous thin film can be used to develop more accurate transport model for the photoanode.

References

1. Antoniadou, M., Vaiano, V., Sannino, D., Lianos, P.: Photocatalytic oxidation of ethanol using undoped and Ru-doped titania: acetaldehyde, hydrogen or electricity generation. Chem. Eng. J. **224**, 144–148 (2013)
2. Du, Z., Li, H., Gu, T.: A state of the art review on microbial fuel cells: a promising technology for wastewater treatment and bioenergy. Biotechnol. Adv. **25**(5), 464–482 (2007)
3. Logan, B.E.: Exoelectrogenic bacteria that power microbial fuel cells. Nat. Rev. Microbiol. **7** (5), 375–381 (2009)
4. Liu, H., Cheng, S., Logan, B.E.: Production of electricity from acetate or butyrate using a single-chamber microbial fuel cell. Environ. Sci. Technol. **39**(2), 658–662 (2005)
5. Padoin, N., Soares, C.: An explicit correlation for optimal TiO_2 film thickness in immobilized photocatalytic reaction systems. Chem. Eng. J. **310**, 381–388 (2017)

6. Chang, H.T., Wu, N.M., Zhu, F.: A kinetic model for photocatalytic degradation of organic contaminants in a thin-film TiO_2 catalyst. Water Res. **34**(2), 407–416 (2000)

7. Khataee, A.R., Fathinia, M., Aber, S.: Kinetic modeling of liquid phase photocatalysis on supported TiO_2 nanoparticles in a rectangular flat-plate photoreactor. Ind. Eng. Chem. Res. **49**(24), 12358–12364 (2010)

8. Tournat, V., Pagneux, V., Lafarge, D., Jaouen, L.: Multiple scattering of acoustic waves and porous absorbing media. Phys. Rev. E **70**(2), 026609–026619 (2004)

9. Makino, T., Kunitomo, T., Sakai, I., Kinoshita, H.: Thermal radiation properties of ceramic materials. Heat Transfer-Japanese Res. **13**, 33–50 (1984)

10. Molenaar, R., Jaap, J., Zijp, J.R.: Determination of Kubelka-Munk scattering and absorption coefficients by diffuse illumination. Appl. Opt. **38**(10), 2068–2077 (1999)

11. Pickering, J.W., Prahl, S.A., Van Wieringen, N., Beek, J.F., Sterenborg, H.J., Van Gemert, M.J.: Double-integrating-sphere system for measuring the optical properties of tissue. Appl. Opt. **32**(4), 399–410 (1993)

12. Cheong, W.F., Prahl, S.A., Welch, A.J.: A review of the optical properties of biological tissues. IEEE J. Quantum Electron. **26**(12), 2166–2185 (1990)

13. Petrasch, J., Wyss, P., Steinfeld, A.: Tomography-based Monte Carlo determination of radiative properties of reticulate porous ceramics. J. Quant. Spectr. Radiative Transf. **105**(2), 180–197 (2007)

14. Xie, B.W., Ma, L.X., Zhao, J.M., Liu, L.H., Wang, X.Z., He, Y.R.: Experimental study of the radiative properties of hedgehog-like ZnO–Au composite particles. J. Quant. Spectr. Radiative Transf. **232**, 93–103 (2019)

15. Morales, J.J., Nuevo, M.J.: A general method for calculating reflection and transmission coefficients in multiple slabs. Am. J. Phys. **59**(12), 1140–1143 (1991)

16. Zhao, J.M., Liu, L.H.: Least-squares spectral element method for radiative heat transfer in semitransparent media. Numer. Heat Transf. Part B Fundam. **50**(5), 473–489 (2006)

17. Li, L., Wang, G., Chen, R., Zhu, X., Wang, H., Liao, Q., Yu, Y.: Optofluidics based micro-photocatalytic fuel cell for efficient wastewater treatment and electricity generation. Lab Chip **14**(17), 3368–3375 (2014)

18. Li, L., Chen, R., Zhu, X., Liao, Q., Wang, H., An, L., Zhang, M.: A cascading gradient pore microstructured photoanode with enhanced photoelectrochemical and photocatalytic activities. J. Catal. **344**, 411–419 (2016)

19. Yang, Z., Zhu, D., Lu, D.: The relationship between porous ratio and refractive index in nanoporous film. Acta Optica Sinica **23**(11), 1366–1369 (2003)

Mechanical Behavior of Drillstring with Oscillator During Sliding Drilling

Xiaolei Shi, Wenjun Huang, and Deli Gao

Abstract In the process of drilling operation, the mechanical behavior of the drill string is complex and the failure problem is prominent, and friction is an important factor restricting the extension of the borehole. Based on the vibration mechanism and basic knowledge of drill string mechanics, this paper introduces the action of oscillator comprehensively considers the vibration function form of excitation force excited by oscillator, drill string friction model, drill string end constraint conditions and continuity conditions, establishes a drill string dynamic model with oscillator and uses finite difference method to solve it. Then, the above model is verified by an example, and the drag reduction effects with and without oscillators are analyzed and compared. Finally, sensitivity analysis is carried out to evaluate the drag reduction effects of various parameters. The results show that the fluctuation are more obvious for the displacement of drill string and the hook load with oscillator, and the effect of reducing the average friction coefficient is significant, which is more conducive to the transfer of axial force. When the energy excited by the vibrator is the same, the vibration function form of large amplitude and small frequency excitation force is beneficial to load transfer and can reduce the average friction coefficient of the drill string. When other parameters are constant, increasing the amplitude or increasing the number of oscillators is conducive to improving the drag reduction efficiency of drill string, increasing the penetration rate and increasing the borehole extension distance. This study has important guiding significance for the design optimization of drill string and safe operation on site.

Keywords Drillstring · Vibration mechanism · Oscillator · Friction reduction · Load transfer

X. Shi · W. Huang · D. Gao (✉)
MOE Key Laboratory of Petroleum Engineering, China University of Petroleum, Beijing 102249, China
e-mail: gaodeli@cup.edu.cn

113

1   Introduction

In recent years, with the rapid development of drilling technology, the depth of drilling has been continuously extended. More and more complex structure wells such as horizontal wells and extended reach wells have become the focus of exploration and development. Compared with conventional well types, these new complex structure wells are facing many technical problems, especially in the long horizontal section where the friction torque is relatively large, which is prone to serious propping up and sticking, thus causing serious drilling accidents and affecting the normal drilling of oil and gas wells. In order to effectively reduce the friction torque between the drill string and the borehole wall and reduce the propping pressure of the drill string, the average friction force of the whole drill string should not only be reduced, but also the static friction force of the drill string should be changed into dynamic friction force. Starting from the vibration mechanism of the drill string, it is of great significance to develop suitable vibration tools and establish a reasonable drill string vibration drag reduction model for in-depth theoretical analysis to improve the extension limit of the well and the drilling velocity.

At present, the vibration drag reduction tools abroad are mainly mechanical vibration type and hydraulic type, mechanical vibration type is represented by Agitator hydraulic oscillator, and hydraulic type is represented by FDR tool. The Agitator hydraulic oscillator [1, 2] can generate an impact force of 3.45–4.14 MPa according to the difference of frequency, amplitude, outer diameter and length. Through axial vibration, the sliding friction of the drilling tool can be reduced by 75%–80%. On the premise of obtaining the same mechanical drilling tool, the required weight on bit with the tool is about 60% of that without the tool, and all MWD will not be disturbed by Agitator tool, which can ensure the smooth transfer of MWD signals. Meanwhile, axial vibration generated by Agitator tool will not damage the bearings and teeth of drill bit. Agitator hydraulic oscillator has been applied in many oil fields around the world and has achieved good results. It can greatly reduce the friction of sliding drilling and improve the penetration rate. The FDR tool [3] was developed by RF-Rogaland Research. The tool is mainly composed of a two-way moving hydraulic cylinder with an inner valve and an outer valve. It has been comprehensively tested in a large displacement well with a diameter of 177.8 mm, and has been tested in a coiled tubing drilling with an outer diameter of 44.5 mm and a length of 600 m. The results show that the friction resistance of coiled tubing can be reduced by 90% and the drag reduction effect is obvious. In order to solve the problem of axial force transfer in extended reach wells, Sinopec has developed a hydraulic oscillator and developed a SLZDDQ172 hydraulic pulse oscillation tool [4]. The tool has an outer diameter of 172 mm, a working frequency of 15 Hz, a working pressure consumption of 3–4 MPa, and a displacement of 10–30 L/s. It has good field application effect, can increase the penetration rate by more than 20%, and has good adaptability to PDC bits, cone bits, etc. However, the research and development of the new turbine hydraulic

oscillator [5] is still at the laboratory test stage, and no on-site experiment has been carried out. Most of the other vibration drag reduction tools introduced by others [6–9] still learn from the principle of Agitator tools and make improvements.

The research on theoretical model is relatively mature at present. Johancsik et al. [10] established a soft rope model using model analysis method and predicted the pulling force of directional well drill string by computer programming method, which provided theoretical basis for well trajectory design and so on. It was obtained that the friction coefficient of drilling was between 0.25 and 0.4 when sliding drilling was performed. Ho [11] established a new stiffness model that considered the three factors of drill string rigidity, centralizer position and annulus clearance was established and revised the original model considering the influence of bottom hole assembly. Considered the influence of joints on the basis of predecessors, Gao et al. [12] calculated the critical force of helical buckling of tubular string in inclined straight section, the contact force between tubular string and borehole wall distributed along helical section and the average lateral contact force of post-buckling and proposed a new method to calculate the critical force of sinusoidal buckling, proving that when the drill string undergoes helical buckling, there will be a large friction between the drill string and borehole wall. Miska et al. [13] established an improved soft rod dynamic model based on the movement of drill pipe in 2D and 3D boreholes. The model can better calculate the stress of moving drill pipe or casing. The actual design calculation can be solved by using numerical calculation method and computer code programming. Wang et al. [14] proposed a new friction reduction technology called "earthworm-like drilling" based on "soft-string" model, which can improve the load transfer and extended-limit. Wang et al. [15] considered the transformation and decomposition mechanism of static friction force and dynamic friction force, and established a drag reduction model which was solved by second-order finite difference method and analyzed the vibration behavior and parameters under different axial load transfer and tool surfaces. In order to investigate the friction reduction mechanism of the hydraulic axial vibratory tool in coiled tubing drilling, Zhang et al. [16] proposed a dynamic model by introducing the Coulomb friction term to the wave equation. According to the wave equation, Liu et al. [17] established a FEA model by considering the coupling of the drillstring-wellbore contract and the action of AOT, which was verified analytically and experimentally. The results show that the friction-reduction is mainly influenced by drillstring material properties and wellbore parameters. Mahjoub et al. [18] proposed a new modeling approach to accurately model the effect of the AOT on drilling operations without the need to carry out resource-intensive and time-consuming dynamic computation. The results show that once the influence length and the maximum displacement are calculated, an effective friction coefficient is estimated as the mean value of the instantaneous friction coefficient and used in a stiff-string torque and drag model. Omojuwa et al. [19] presented an analytical model to predict the behavior of axial oscillation-supported drillstings working under downhole and surface conditions. The model is useful to perform parametric studies, downhole data correlation and

placement study of AOTs along the drillstring in horizontal and extended-reach wells.

Through investigation, it can be found that previous studies have independently analyzed vibration drag reduction tools or drag reduction models based on the vibration mechanism of drill string, ignoring the friction between the drill string and the borehole wall, or replacing the friction with Coulomb friction. Few people have carried out mechanical analysis on drill string with oscillator and studied its drag reduction effect. Starting from the vibration mechanism of the drill string, this paper analyzes the process of the drill string from static to motion. Considering the action of the oscillator on the drill string, a bilinear hysteretic friction model is adopted to establish the dynamic model of the drill string with the oscillator. The new model is verified by field data. Sensitivity analysis of the key parameters of the drill string or oscillator can accurately evaluate the drag reduction effect of the oscillator.

2   Theoretical Model

Before establishing the model, the following assumptions need to be made on the model:

(1) Both the drill string and the inner wall of the borehole are rigid, and the cross-sectional area of the drill string is always circular.
(2) The drill string is in uniform contact with the borehole wall without rotation.
(3) Only consider axial vibration and ignore the lateral and torsional vibration of the drill string.
(4) The bilinear hysteretic restoring force model is adopted to depict the friction between tubular string and wellbore.

A micro-element segment with an arc length of ds is selected as the research unit on the entire drill string. The unit is connected to the adjacent micro-element segment through a spring-like oscillator, and θ is the inclination angle of the drill string. To study its vibration response during the drilling process, establish the schematic diagram shown in Fig. 1.

2.1   Differential Equation

As shown in Fig. 1 above, the drill string is composed of multiple micro-element segments, and the force analysis of one of the micro-element bodies is performed. Under static conditions, the drill string lies flat on the borehole wall, relying on the weight of the drill string, static friction and bottom weight to maintain balance. The force equation of the micro-element body of this section of the drill string can be written as:
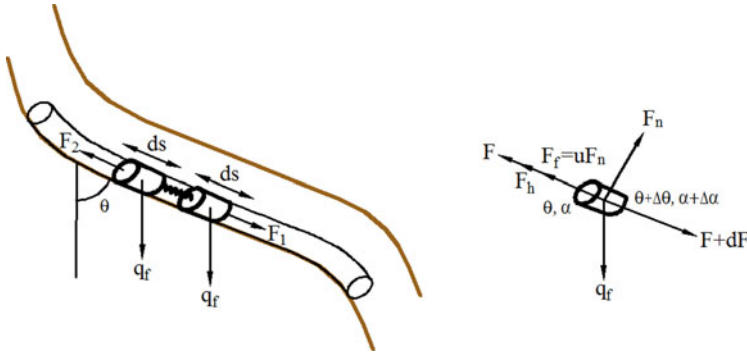
Fig. 1 Stress state of differential element of drillstring while drilling

$$F + \mathrm{d}F + \rho gA \cos \theta ds - F - \mu \rho gA \sin \theta ds = 0 \qquad (1)$$

During the vibration process, the oscillator can apply regular excitation force to the drill string. The drill string also needs to overcome the viscous force of drilling fluid. At this time, the static friction force between the drill string and the borehole wall is converted into dynamic friction force. Under the action of the weight, friction force, weight of bottom and viscous force of the drill string, the force equation of the drill string micro-element body in this section can be written as follows:

$$F + \mathrm{d}F + \rho gA \cos \theta ds - F - \pi DC\frac{\partial U}{\partial t}ds - F_f ds = \rho Ads \frac{\partial^2 U}{\partial t^2} \qquad (2)$$

Where U, u, ρ, D, A are axial displacement function, friction coefficient, density, annulus hydraulic diameter and cross-sectional area of drillstring, respectively; F_f is the friction force between drillstring and borehole wall (casing); s is the distance from any point of the drillstring to the top of the section; C is the viscosity coefficient of drilling fluid; t is time; g is gravitational acceleration; θ is angle of inclination.

2.2 Friction Model

As we all know, vibration is a natural phenomenon in nature. If divided according to the nature of the motion equation in the vibration process, the contact between the drill string and the borehole wall will produce nonlinear vibration, considering the elastic nature of the contact surface and the dynamic response characteristics of the external disturbance excitation to the system. Iwan et al. [20] put forward a bilinear hysteretic restoring force model, which equates the dry friction surface with

a spring and an ideal Coulomb friction pair connected in series. the nonlinear restoring force F_f with memory characteristics is approximately described by a double fold line model. The constitutive relation in incremental form can be written as follows:

$$
\begin{cases}
dF_f = \frac{k_s}{2}\left[1 + \mathrm{sgn}\left(F_s - \left|F_f\right|\right)\right]dx \\
k_s = \frac{F_s}{x_s}
\end{cases}
\tag{3}
$$

Where x_s is the limit value of elastic deformation when the contact surface of the dihedral slides macroscopically; F_s is the memory restoring force when slipping; x is the relative displacement deformation of the two ends of the hysteresis link; k_s is the linear stiffness before the dry friction link slips.

3   Calculation Method

3.1   Finite Difference Method

The finite difference method can be used to solve the above problems. In the process of solving differential equations by finite difference method, independent variables take discrete values. We transform the differential equations of continuous variables into difference equations in discrete variable grids, and determine the function values at each grid point with initial values or boundary values. The specific method is to discretize the differential equations first, and then solve the difference equations.

Since, $F = EA\frac{\partial U}{\partial s}$ the explicit central difference scheme of Eq. 1 and Eq. 2 can be expressed as:

$$
\frac{U_{i+1}^j - 2U_i^j + U_{i-1}^j}{\Delta s_i^2} = \frac{\mu_i^j \rho g \sin\theta - \rho g \cos\theta}{E}
\tag{4}
$$

$$
\frac{E}{\rho}\frac{U_{i+1}^j - 2U_i^j + U_{i-1}^j}{\Delta s_i^2} + g\cos\theta - \frac{\pi DC}{\rho A}\frac{U_i^{j+1} - U_i^j}{\Delta t} - \frac{F_{f_i}}{\rho A} = \frac{U_i^{j+1} - 2U_i^j + U_i^{j-1}}{\Delta t^2}
\tag{5}
$$

Where E is elastic modulus; Δt is the time interval, the superscript "j" in U_i^j represents the i-th time point.

3.2   Initial Condition

In general, the calculation of finite difference equations requires initial conditions. In another word, when the right side of Eq. 1 and Eq. 2 is set to zero, the initial displacement needs to satisfy the equations. The discretization scheme of the initial displacement can be expressed as:

$$U_i^1 = u_{initial} \tag{6}$$

The values of $u_{initial}$ are obtained by solving Eq. (4).
The discretized scheme of initial velocity condition is expressed as:

$$\frac{U_i^2 - U_i^0}{2\Delta t} = v_{initial} \tag{7}$$

Note that, the term U_j^0 in Eq. 5 can be eliminated by combing Eq. 5 and Eq. 7 while j = 1.

3.3   Boundary Condition

The top of the drillstring is tied to hook, so the axial displacement of the top of drillstring is equal to the vertical displacement of hook. Then, the top boundary condition is expressed as:

$$U_1^j = u_{hook} \tag{8}$$

For the drillstring installed with oscillator, an additional excitation force will be applied to the drillstring. Therefore, the boundary conditions at the upper and lower ends of the oscillator can be expressed as the axial force as:

$$(EA)_{i-1}\frac{U_i^j - U_{i-2}^j}{2\Delta s_{i-1}} = F_t + FF(t), (EA)_i\frac{U_{i+1}^j - U_{i-1}^j}{2\Delta s_i} = F_t \tag{9}$$

Where F_t is the axial force at the joint of oscillator and drillstring; $FF(t)$ is the excitation force applied by the vibrator with time.

In the sliding drilling process, the axial force on bit is determined with bit-rock interaction model. Here, the axial force on the bottom of drillstring is equal to drill bit, namely

$$(EA)_n\frac{U_{n+1}^j - U_{n-1}^j}{2\Delta s_n} = Wob(t) \tag{10}$$

Where *Wob(t)* is the weight-on-bit which changes with time. Note that, the term U_{n+1}^{j} in Eq. 10 can be eliminated by combing Eq. 5 and Eq. 10 while i = n.

3.4 Continuous Condition

When two or more than two kinds of drillstrings are adopted, the relevant parameters such as tubular diameter, weight, etc. are discrepant for different drilling strings. Then the continuous conditions on the node connecting to different tubular strings should be satisfied. To be specific, the axial displacement and equivalent axial forces on adjacent segments are continuous, namely

$$U_{i-1}^{j} = U_{i}^{j} \tag{11}$$

$$(EA)_{i-1}\frac{U_{i}^{j} - U_{i-2}^{j}}{2\Delta s_{i-1}} = (EA)_{i}\frac{U_{i+1}^{j} - U_{i-1}^{j}}{2\Delta s_{i}} \tag{12}$$

4 Results and Discussions

On the basis of the above model and calculation method, the mechanical behaviors of drillstring in sliding drilling operation are studied. A horizontal well is drilled of which the depth of kick off point is 500 m, inclination angle of the horizontal section is π/2 and well depth is 3000 m. A 5-1/2″drill pipe is adopted of which drilling string weight in air per unit length is 28.03 kg/m, Bingham drill fluid is used of which the density is 1.2 g/cm³. The rotation speed is 50 rad/s, the rate of penetration is 15 m/h, and four oscillators are arranged below the kick off point. The excitation frequency of the oscillator is 5 Hz, and the excitation load changes in sine function with respect to time.

In the numerical simulation, to reveal the transition between sticking friction and sliding friction, the time interval should be very small. The time interval is set to 4e −4 s, segment length is set to 3 m, operation time is set to 30 s. It takes about 30 s to obtain the results.

4.1 Vibration Behaviors

As shown in Fig. 2, during the sliding drilling operation, when the drill bit breaks rock at the set drilling speed, the axial displacement of the drillstring increases. When the drillstring starts to move, the displacement increases more rapidly. As the
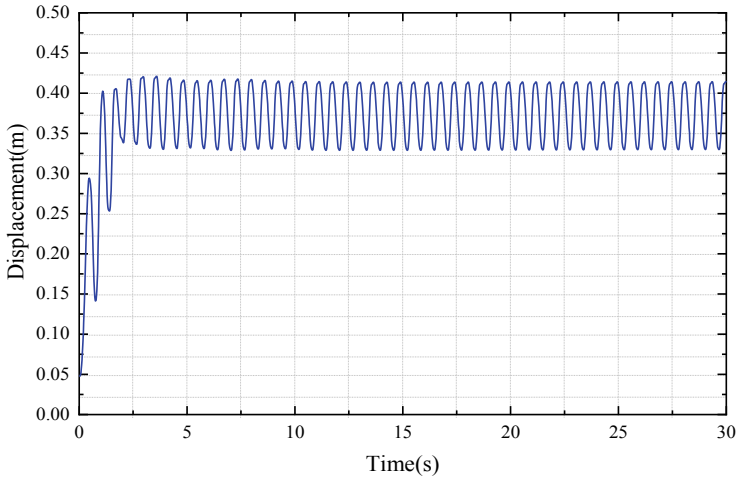
Fig. 2 Displacement change of drillstring at 600 m hole depth in sliding drilling operations

time increases, the movement of the drillstring enters a stable state, the displacement oscillates around its average value.

As shown in Fig. 3, during the sliding drilling operation, the friction coefficient of the drillstring always shows a regular change. Generally speaking, because the vibration of the drillstring is nonlinear, and the friction coefficient is opposite to the direction of the movement of the drillstring, the change of the friction coefficient is not constant. When the drillstring moves in the positive direction, it always transitions from the static friction coefficient to the dynamic friction coefficient in the form of a parabola, and then returns to the static friction coefficient with the same law, and finally changes the direction of movement and reciprocates with the same law. In other words, the friction coefficient first decreases and then increases, and the direction of change continues to decrease and then increases, and reciprocates.

As shown in Fig. 4, the velocity of the drillstring presents a relatively regular movement pattern, the drillstring is in an unstable state at the beginning of the movement, and the fluctuation of the drillstring velocity is not stable. With the increase of time, the velocity change amplitude gradually stabilizes, and the velocity always reciprocates around the zero point.

As shown in Fig. 5, since the top of the drillstring is always controlled by the hook, the load change at the hook can reflect the working conditions of the bottom hole drillstring and bit in real time. At the beginning of drilling, the hook load is relatively small. With the increase of time, combined with the action of the downhole oscillator, the hook load increases rapidly and reaches a stable state. The load always oscillates around an average value. It can be seen that during the sliding drilling process, the increase in the hook load is more obvious, indicating that axial friction is decreased a lot and sliding drilling process is more smooth.
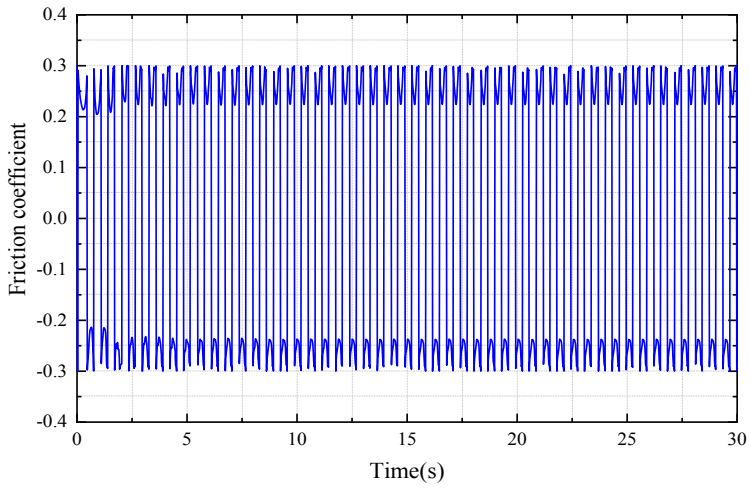
Fig. 3 Friction coefficient change of drillstring at 600 m hole depth in sliding drilling operations
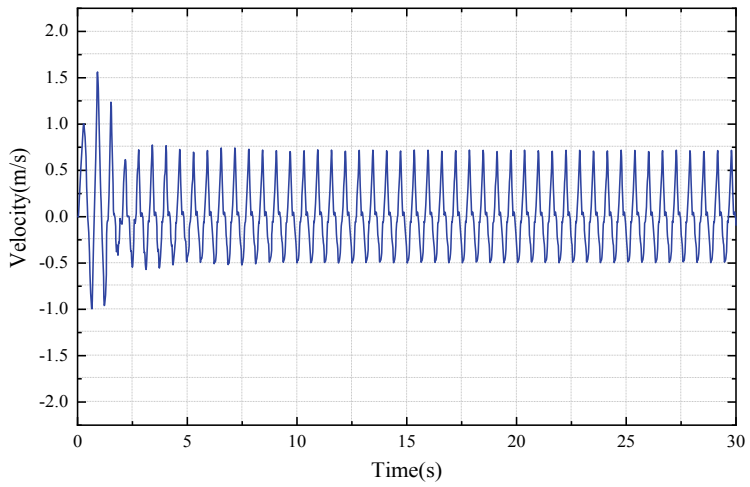
Fig. 4 Velocity change of drillstring at 600 m hole depth in sliding drilling operations
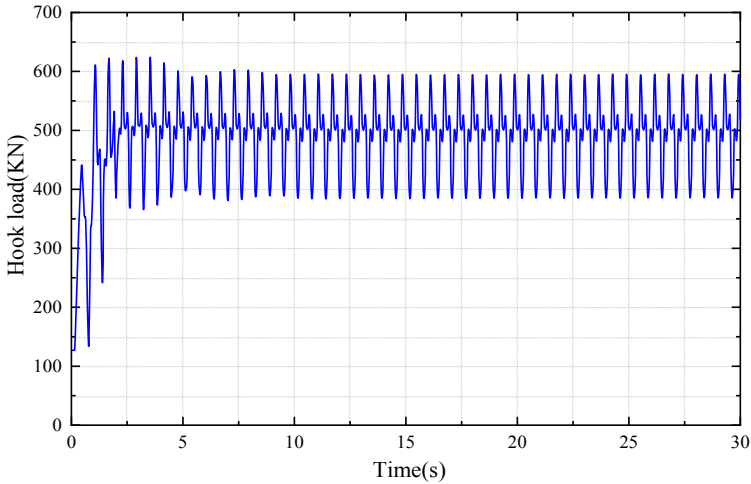
Fig. 5 Change of hook load in sliding drilling operations

4.2 Drag Reduction of Oscillator

Before optimizing the parameters, the mechanical behaviors of the drillstring with and without oscillator are first analyzed. To simulate the no oscillator case, the excitation force of the oscillator is set to zero in the above model.

The calculation results with or without oscillator are shown in Figs. 6, 7, 8. Figure 6 and Fig. 7 show similar change trends. When the oscillator is installed on the drillstring, the displacement of the drillstring and the hook load increase greatly, showing a more violent fluctuation pattern. But the displacement of the drillstring and the hook load without the oscillator basically have no change, only a slight jump at the beginning of the movement, and then it is stable. This indicates that the drillstring with oscillator is excited by the excitation force, the load transfer is more obvious, and the drillstring has been in motion; The comparison between the average friction coefficients of the drillstring with and without the oscillator are given in Fig. 8. The average friction coefficient of the drillstring with the oscillator will be greatly reduced, especially at the installation location of the oscillator. The decrease in the average friction coefficient of the drillstring is very obvious. The analysis of the above results shows that the load transfer on drillstring with oscillator is much better than that without oscillator, and the average friction coefficient is greatly reduced.
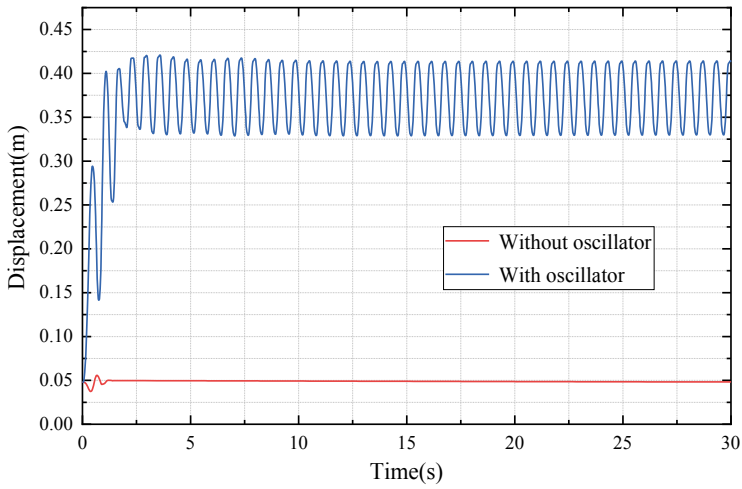
Fig. 6 Change of drillstring displacement in sliding drilling operations
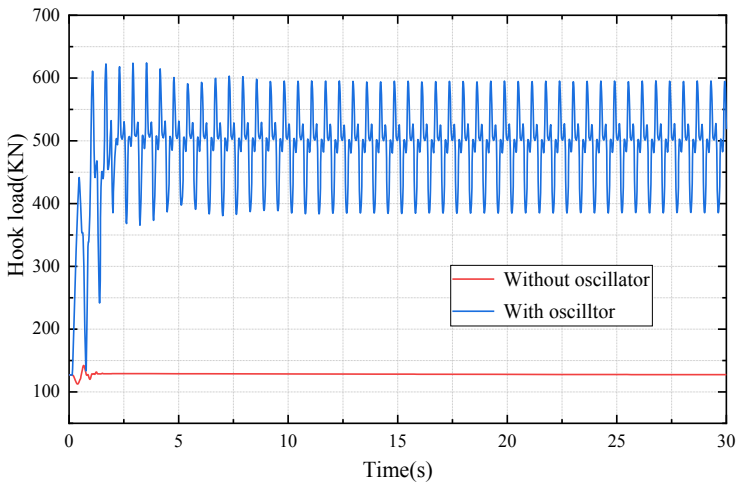
Fig. 7 Change of hook load under sliding drilling operation
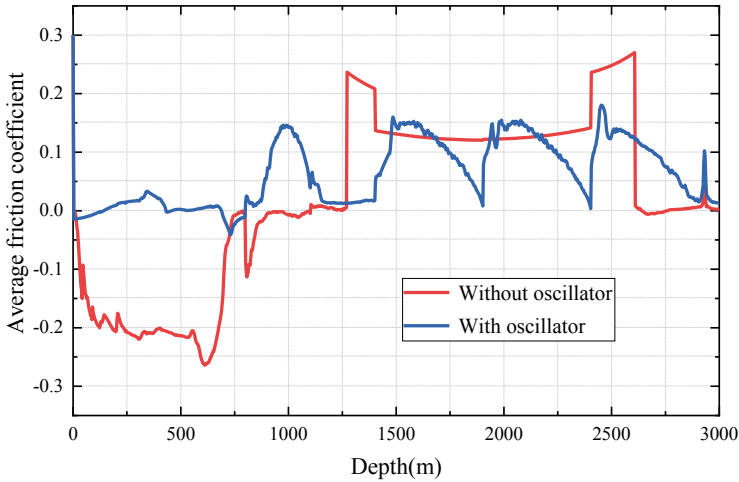
Fig. 8 Change of average friction coefficient of drillstring in sliding drilling operations

4.3 Oscillator Optimization

Through the above analysis, we have known that the drill string with a oscillator can play a friction reduction effect during the drilling process, but the acquisition of the relevant indicators of the oscillator is not clear, so the following parameters will be optimized to maximize the effect of friction reduction from the form of excitation force generated by oscillator, amplitude variation and oscillator number.

The evaluation of the friction reduction effects under different vibration forms of the excitation force is shown in Figs. 9, 10, 11. As shown in Fig. 9, oscillators with different amplitudes and vibration frequencies generate the same energy in one cycle to ensure that the oscillator can provide the same load in the same cycle. As shown in Fig. 10, the vibration force of larger amplitude has a greater impact on the hook load. The vibration function form has basically no effect on the average hook load, because the oscillator can generate the same vibration energy in the same period. As shown in Fig. 11, when the vibration energy is the same, the vibration function form with larger amplitude and smaller frequency has the most obvious effect on reducing the average friction coefficient. But when the amplitude increases and the frequency decreases to a certain degree, the drag reduction effect stabilizes in a good state.

The evaluation of the amplitude reduction effect on friction reduction is shown in Fig. 12 and Fig. 13. As shown in Fig. 12, the larger the vibration force amplitude, the larger the average hook load and the fluctuation amplitude of hook load, indicating that large vibration force amplitude is the key factor improving the axial force transfer along the tubular string; as shown in Fig. 13, the amplitude increases, the average friction coefficient decreases, indicating that the large vibration force amplitude has a significant effect on reducing the average friction coefficient.
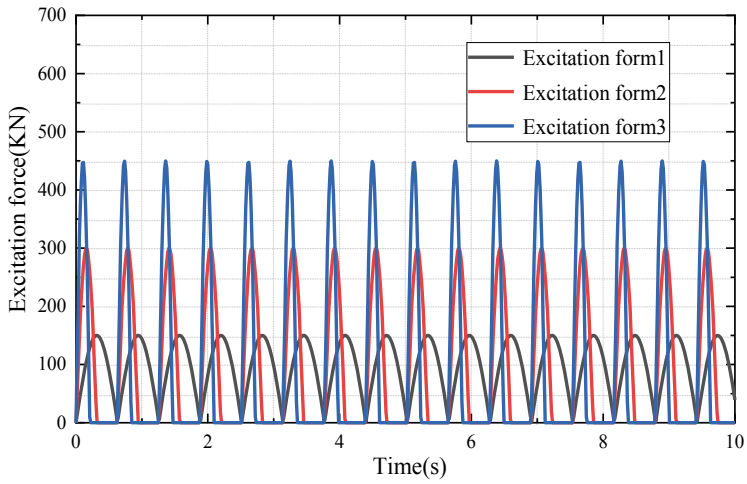
Fig. 9 The action form of the oscillator with excitation force
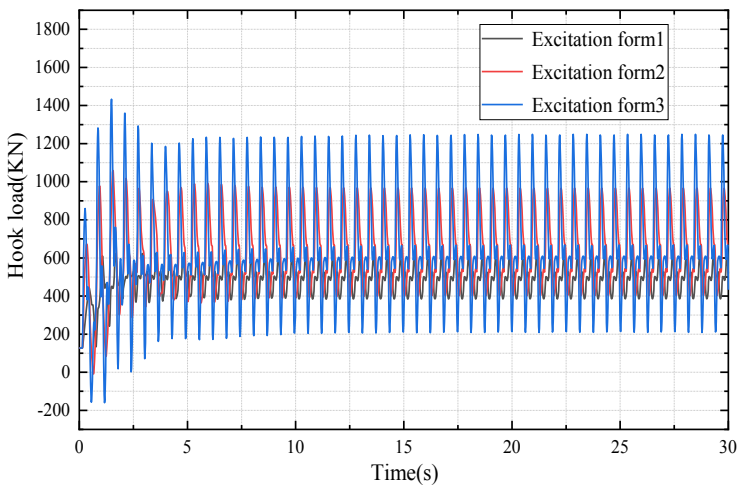
Fig. 10 Effect of the excitation force form on hook load
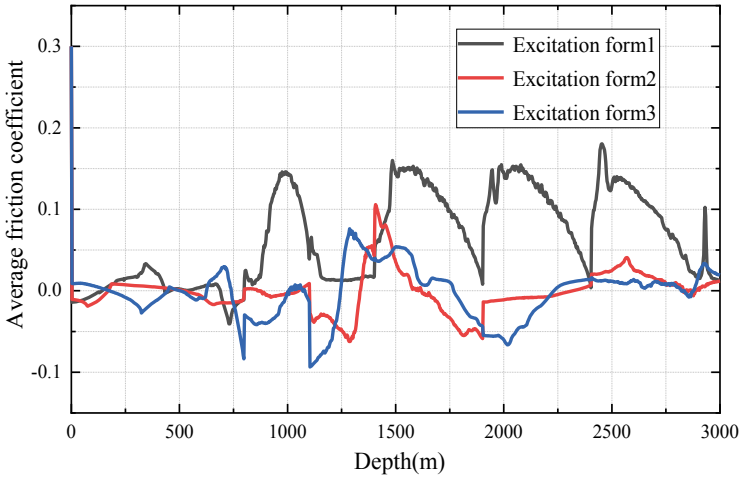
Fig. 11 Effect of the excitation force form on the average friction coefficient
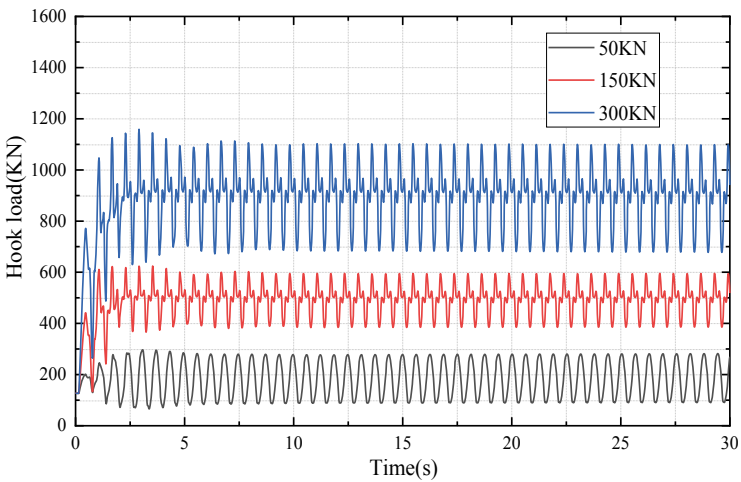
Fig. 12 Effect of vibration amplitude on hook load

However, if the amplitude is too large, the drillstring is subjected to alternating loads, which may lead to the risk of slippage of the drill pipe joint and fatigue damage of the drillstring.

The evaluation of the number of oscillators on friction reduction effect is shown in Fig. 14 and Fig. 15. As shown in Fig. 14, the more oscillators there are, the more violent the hook load fluctuation is, and the average hook load increases significantly, indicating that the number of oscillators is crucial to the transmission of
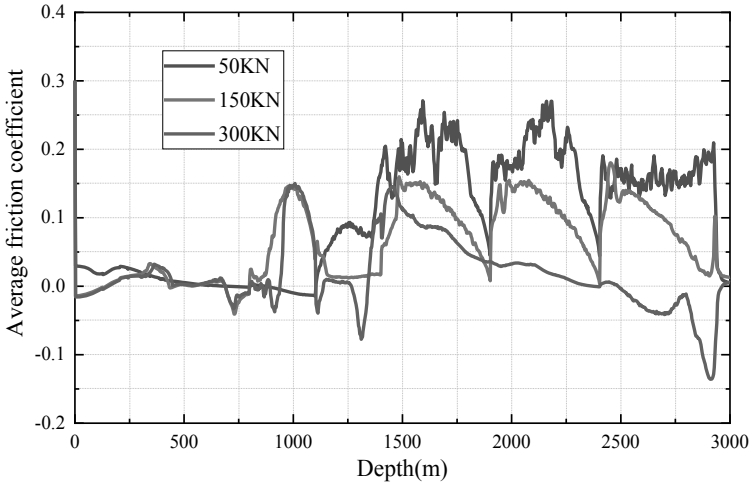
Fig. 13 Effect of vibration amplitude on average friction coefficient
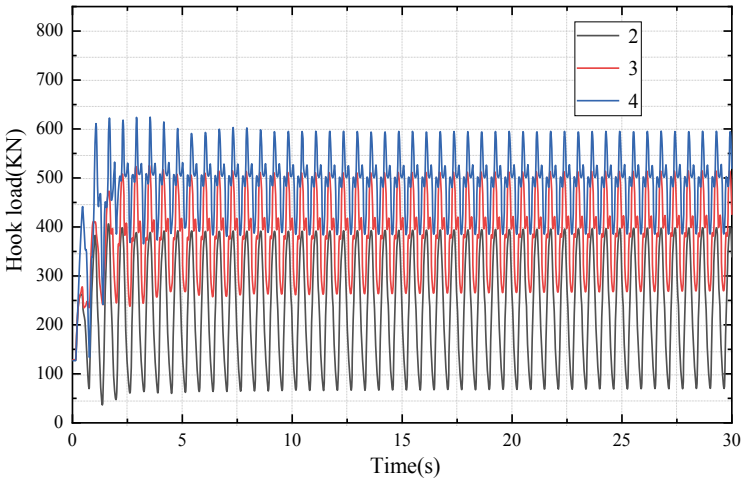
Fig. 14 Effect of the oscillator number on hook load

axial force of the drillstring. As shown in Fig. 15, the more oscillators, the lower the average friction coefficient of the drillstring, indicating that multiple oscillators can significantly decrease the average friction coefficient. But if too many oscillators are installed, may lead to large hydraulic power loss and high risk of tubular fatigue.
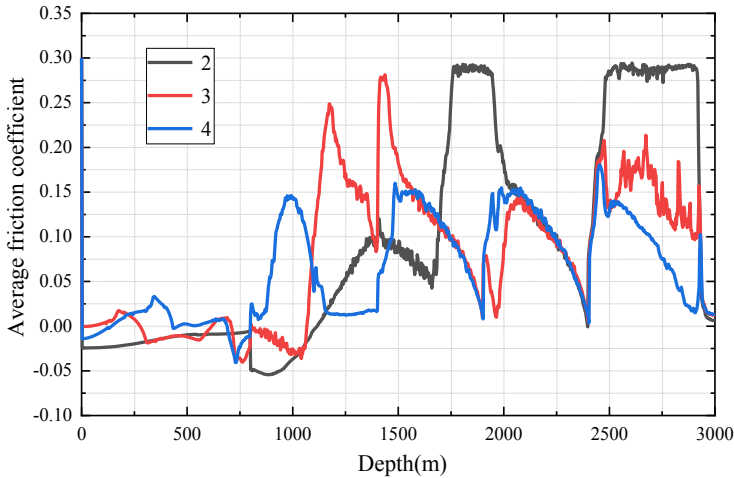
Fig. 15 Effect of the oscillator number on average friction coefficient

5 Conclusions

From above analyses we can see obtain the following conclusions:

(1) A dynamic model of drillstring with drag reduction oscillators is established based on the bilinear hysteretic restoring force model, which provides the basis for mechanical analysis and optimal design of drillstring with drag reduction oscillators.

(2) Comparing the results of drill string with oscillator and without oscillator, the displacement of the drill string with oscillator, the hook load fluctuation is more obvious, and the effect of reducing the average friction coefficient is significant, which indicates that installing oscillator in the drilling process is beneficial to the transmission of axial force.

(3) With the increase of drag reduction oscillator number and their vibration force amplitudes, the friction factor and axial friction are both decreased. However, too large vibration force amplitude will lead to drillstring fatigue and the hydraulic energy loss increasing with the drag reduction oscillator number. Therefore, there are optimal values predicted for vibration force amplitude and drag reduction oscillator number.

References

1. Newman, K., Burnett, T., Pursell, J., Gouasmia, O.: Modeling the affect of a downhole vibrator. In: SPE/ICoTA Colied Tubing and Well Intervention Conference, Woodlands, Texas, USA, 5–6 October. SPE121752 (2009)
2. Skyles, L.P., Amiraslani, Y.A., Wilhoit, J.E.: Converting static friction to kinetic friction to drill further and faster in directional holes. In: SPE Drilling Conference and Exhibition, San Diego, California, 6–8 March. SPE151221 (2012)
3. Sola, K.-I., Lund, B.: New downhole tool for coiled tubing extended reach. In: SPE/ICoTA Colied Tubing Roundtable, Houston, Texas, USA, 5–6 April. SPE60701 (2000)
4. Zhang, H., Yu, W.T., Chen, Z.S., Cai, W.J.: Development of hydropulse axial-oscillation friction-reduce tool. Oil Field Equip. **43**(7), 73–76 (2014)
5. Wang, J., Xia, C.Y., Feng, D., Yu, C.B.: Design and experimental study on a new type of turbine driven hydraulic oscillator. Chinese J. Eng. Des. **23**(4), 391–395 (2016)
6. Li, B.: Development and pilot test of hydro-oscillator. Petrol. Drill. Tech. **42**(1), 111–113 (2014)
7. Livescu, S., Watkins, T.: Water hammer modelling in extended reach wells. SPE/ICoTA Well Intervention and Coiled Tubing Conference and Exhibition, Woodlands, Texas, USA, 25–26 March. SPE-168297-MS (2014)
8. Livescu, S., Craig, S.H.: A critical review of the coiled tubing friction-reducing technologies in extended-reach wells. Part 1: Lubricants. J. Pet. Sci. Eng., 747–759 (2017)
9. Wang, P., Ni, H., Wang, R.: A novel vibration drilling tool used for reducing friction and improve the penetration rate of petroleum drilling. J. Pet. Sci. Eng., 436–443 (2018)
10. Johancsik, C.A., Friesen, D.B., Rapier, D.: Torque and drag in directional wells-prediction and measurement. J. Petrol. Technol. **36**(6), 987–992 (1984)
11. Ho, H.S.: An improved modeling program for computing the torque and drag in directional and deep wells. Soc. Petrol. Eng. (1988). https://doi.org/10.2118/18047-ms
12. Gao, D.L., Gao, B.K.: Effects of tubular buckling on torque and drag in horizontal well. J. Univ. Petrol. China **24**(2), 1–3 (2000)
13. Miska, S.Z., Zamanipour, Z., Merlo, A., Porche, M.N.: Dynamic soft string model and its practical application. In: SPE/IADC Drilling Conference and Exhibition, London, England, UK, 17–19 March. SPE-173084 (2015)
14. Wang, P., Ni, H.J., Wang, X.Y., et al.: Research on the characteristics of earthworm-like vibration drilling. J. Petrol. Sci. Eng. **160**, 60–71 (2018)
15. Wang, P., Ni, H.J., Wang, X.Y., Wang, R.H.: Modelling the load transfer and tool surface for friction reduction drilling by vibration drill-string. J. Petrol. Sci. Eng. **164**, 333–343 (2018)
16. Zhang, W.P., Shi, H.Z., Li, G.S., et al.: Mechanism analysis of friction reduction in coiled tubing drilling with axial vibratory tool. J. Petrol. Sci. Eng. **175**, 324–337 (2018)
17. Liu, J.X., Zheng, H.L., Kuang, Y.C., et al.: A feasible model for friction-reduction investigation of drillstrings in long-horizontal wells with axial oscillation tools and analysis of key influencing factors. J. Petrol. Sci. Eng. **185**. (2019). https://doi.org/10.1016/j.petrol.2019.106643
18. Mahjoub, M., Dao, N.-H., Menand, S.: Drillscan: modeling the effect of axial oscillation tools in torque and drag computations. In: SPE/IADC Drilling International Conference and Exhibition, 5–7 March, Hague, Netherlands, SPE/IADC-194133-MS (2019)
19. Omojuwa, E., Ahmed, R.: Analytical modeling of axial oscillation-supported drillstrings in high-angle wells. J. Petrol. Sci. Eng. **191**, 107139 (2020). https://doi.org/10.1016/j.petrol.2020
20. Iwan, W.D., Caughey: The Dynamic Response of Bilinear Hysteretic Systems. California Institute of Technology (1961)

Study on Preparation of TiO$_2$ Nanotube Arrays and Their Degradation Mechanism on Oilfield Wastewater Containing Polyacrylamide

Yongtao Fan, Feng Qian, Xiangji Kong, Yuan Gao, Ping Li, and Hang Yin

Abstract With the use of multi-functional organic polyacrylamide additives in various oilfield operations, the amount of polyacrylamide-containing wastewater remaining in surface and groundwater is gradually increasing, resulting in potential environmental pressure and impact. In order to study the effect of a recyclable photocatalyst on the degradation of polyacrylamide (PAM) pollutants, the TiO$_2$ nanotube arrays were prepared by electrochemical anodization method. The microstructure and phase of the samples were characterized. The effects of the main parameters of anodization on the morphology of the nanotubes were discussed. Taking polyacrylamide wastewater with initial concentration of 98 mg·L^{-1} as the target degradation product, the effects of heat treatment temperature, pipe diameter, pipe length and usage times on the degradation rate were studied. The results show that the photocatalytic activity of nanotubes is the highest after heat treatment at 600 °C. The degradation rate of polyacrylamide increases with the increase of the diameter of nanotubes. When the diameter of nanotubes is 80–120 nm, the degradation rate exceeds 80%, while the pipe length has little effect on the degradation rate. During 8 times of recycling, the photocatalytic activity of nanotubes is still high, and the degradation rate is about 80% in 5 h. Through the research in this paper, the theoretical basis and experimental basis for the application of TiO$_2$ nanotube arrays to degrade the sewage in lots of oil and gas fields are provided.

Y. Fan (✉)
Well Logging Technology Research Institute, China Petroleum Logging Co., Ltd., Beijing 102206, China

Y. Fan · Y. Gao · P. Li
CNPC Bohai Drilling Engineering Institute,
Tanggu District, Tianjin 300457, China

F. Qian · X. Kong
CNODC, Xicheng District, Beijing 100034, China

H. Yin
CNPC Greatwall Drilling Company, Beijing 124010, China

1 Introduction

Since the 1960s, the major oilfields in China have been exploited in large quantities for a long time, and now they have entered the production tail - the development phase of high water cut and extra high water cut. In order to improve crude oil production and the economy, convenience and safety during production process, the researchers used organic polyacrylamide (PAM) as a multi-functional additive in mining operations. Because of the high activity amido group in the side chain of the polyacrylamide molecule, the additive has the characteristics of thickening, stabilization, oil displacement and lubrication, etc., and it has been widely used in major oilfields. However, polyacrylamide is not easy to degrade under natural conditions. With the increase of the amount of polyacrylamide, long-term retentions in ground and groundwater is liable to cause environmental pollution. Therefore, an efficient and environmentally friendly degradation method [1, 2] is urgently needed. At present, the degradation of polyacrylamide includes mechanical, biodegradable, thermal and chemical degradation methods. As a chemical degradation method, the photocatalytic degradation method using nano-TiO$_2$ as photocatalyst can utilize the oxidation ability of photogenerated holes and hydroxyl radicals on the surface under ultraviolet or natural light conditions to decompose most organic pollutants [3–5] without causing secondary pollution, thus making it a potential method for treating oilfield wastewater containing polyacrylamide.

The application form of nano-TiO$_2$ as photocatalyst is mainly divided into powder and film, but the application of powder photocatalyst in industry is limited by its difficulty in recycling. Therefore, researchers pay more attention to film photocatalyst with load and easy to be recycled. In 2001, Grimes, an American scientist, prepared TiO$_2$ nanotube array film materials [6, 7] by electrochemical anodic oxidation process, which attracted widespread attention. The prepared TiO$_2$ nanotube arrays have very high ordered structure. They are firmly connected with the titanium substrate by Schottky barrier. Compared with TiO$_2$ in other forms, TiO$_2$ nanotube arrays have larger surface to volume ratio and stronger adsorptive capacity, exhibit higher photocatalytic activity and photoelectric conversion efficiency [8], and have been applied in the fields of gas sensing materials, solar cells, catalyst carriers and supercapacitors, etc. [9, 10].

Because of the high activity and recyclability of TiO$_2$ nanotube arrays grown by Ti-based anodic oxidation, the photocatalyst of TiO$_2$ nanotube arrays was prepared by anodic oxidation process. The effects of the length and diameter of TiO$_2$ nanotube arrays on the degradation of polyacrylamide pollutants in oilfields were studied.

2 Experimental Study

2.1 *Preparation of TiO$_2$ Nanotube Arrays*

During the anodic oxidation process, the industrial titanium plate with purity of no less than 99% and specifications of 50 mm * 50 mm * 0.1 mm was used as the anode, and the high purity graphite plate with purity of 50 mm * 50 mm * 8 mm was used as the cathode, and the 0–30 V DC regulated power supply was used.

Firstly, the surface of titanium plate was polished with 400 mesh metallographic sandpaper, and finishing polished with buffing machine. Then, the titanium plate was cleaned with acetone and deionized water for 20 min, and then polished for 25 s in a mixture with volume ratio of $V_{HF}:V_{HNO3}:V_{H2O} = 1:3.5:5.5$ to remove oxides. Finally, the surface of the titanium plate was cleaned with deionized water and then dried.

A 400 ml electrolyte consisting of 1 wt% hydrofluoric acid and 0.5 wt% glacial acetic acid was disposed in the beaker. Titanium plate and graphite plate were connected with the positive and negative poles of the regulated power supply respectively. Then they were immersed in the electrolyte, and the spacing was adjusted to 5 cm. Turned on the power supply and adjusted the voltage within 5–30 V. The anodizing time was selected to be 10–100 min. After the reaction, the titanium plate samples were taken out, cleaned and dried with deionized water, and then heated to the assigned temperature (300– 700 °C) in the muffle furnace and kept for 2 h to prepare crystalline TiO$_2$ nanotube arrays.

2.2 *Design and Characterization of Photodegradation of Polyacrylamide*

Polyacrylamide-containing wastewater from Shengli Oilfield Wastewater Treatment Plant of Dongying City was selected in this experiment. After removal of dissolved solid matter by sedimentation and filtration, the concentration of polyacrylamide in the wastewater was determined to be 98 mg·L^{-1}. After heat treatment, the samples of TiO$_2$ nanotube arrays were placed in a beaker with an initial concentration of 98 mg·L^{-1} polyacrylamide solution of 150 ml. The 40 W ultraviolet lamp was used as a light source and placed 8 cm away from the liquid surface. At the same time, the magnetic stirrer under the beaker was turned on and the magnetic rotor was continuously stirred to make the reaction uniform. After reaching the time to be measured, some samples were taken and the viscosity of polyacrylamide solution was measured by Ukrainian viscometer. The degradation rate was calculated according to the following formula:

$$\eta = \frac{P_0 - P_t}{P_0} \times 100\%$$

In which, P_t—Solution viscosity after photodegradation for t hours;
P_0—Initial solution viscosity.

JSM-6510 scanning electron microscope (SEM) was used to characterize the morphology of TiO_2 nanotube arrays, and Bruker D8 Advance X-ray diffractometre was used to analyze the phase of the materials.

3   Results and Discussions

3.1   Effect of Test Parameters on the Morphology and Structures of Tio2 Nanotube Arrays

3.1.1   Relationship Between Oxidation Voltage and Pipe Diameter

Figure 1 shows the apparent morphology of TiO_2 nanotube arrays with anodic oxidation voltage of 10–30 V and reaction time of 30 min. It can be seen from the figure that when the voltage is 10 V, the inner diameter of the nanotube is 30–50 nm and the wall thickness is about 10 nm (shown in Fig. 1(a)). When the voltage is 20 V, the diameter of the nanotube increases to 80–120 nm, and the wall thickness is basically unchanged (shown in Fig. 1(b)). That is, after the formation of the initial nanotube, the diameter of the nanotube is mainly affected by the voltage in a certain range, and increases with the increase of voltage. When the voltage is 30 V, the wall of the nanotube is damaged and the tubular morphology of the nanotube disappears (shown in Fig. 1(c)).

(a) 10V   (b) 20V   (c) 30V
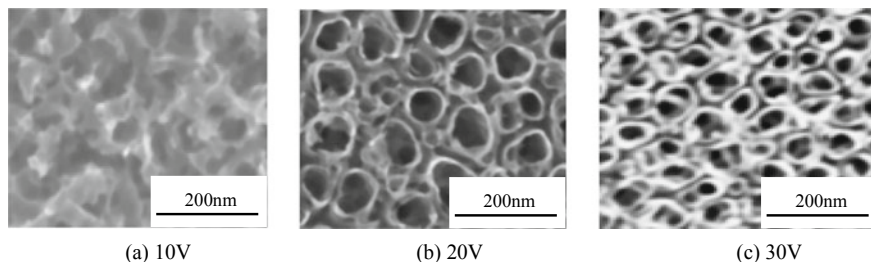
Fig. 1   SEM photographs of TiO_2 nanotube arrays after 10–30 V anodic oxidation

3.1.2 Relationship Between Oxidation Time and Pipe Length

Figure 2 are SEM photographs of TiO$_2$ nanotube arrays after 10–100 min of oxidation (anode voltage is 20 V). After the reaction of titanium plate in electrolyte for 10 min, a clear tubular array with a pipe length of about 230 nm (shown in Fig. 2 (a)) appears. With the increase of oxidation time, the barrier layer gradually advances, and the pipe length increases rapidly in a short time. After 30 min of oxidation, the length reaches about 410 nm (shown in Fig. 2(b)). When the oxidation time continues to increase, the increase of pipe length tends to decrease and is basically stable. After 60 min of oxidation, the pipe length only increases to about 500 nm (shown in Fig. 2(c)). After 100 min of oxidation, the pipe length remains at about 500 nm (shown in Fig. 2(d)). It can be seen that the pipe length increases gradually with the prolongation of oxidation time after the anode voltage is determined. The results show that when the pipe length reaches its maximum, it will not change with time.

(a)10min

(b) 30min

(c) 60min
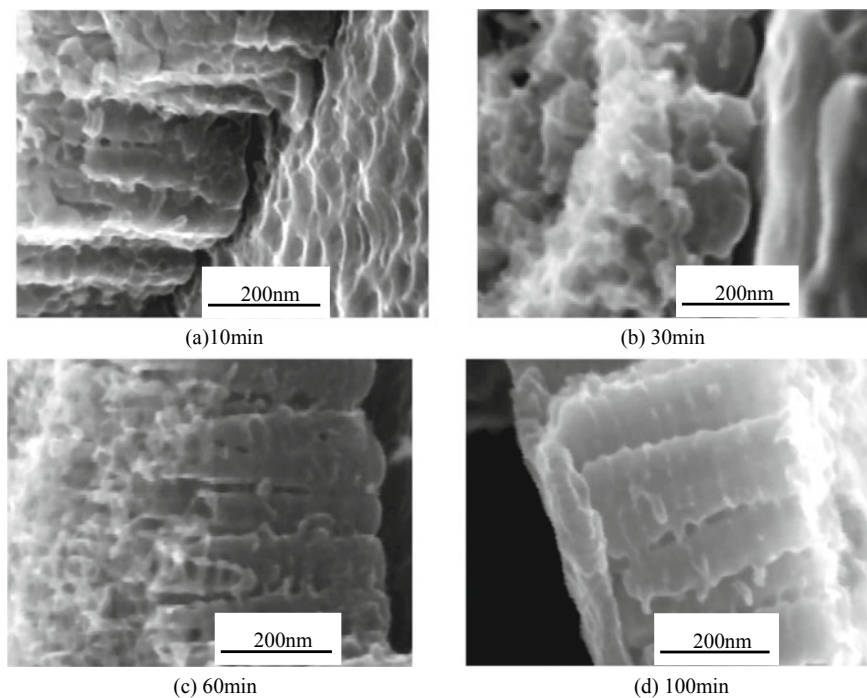
(d) 100min

Fig. 2 SEM photographs of TiO$_2$ nanotube arrays after 10–100 min oxidation

(a) Non heat-treated   (b) 600°C ×2h   (c) 700°C ×2h
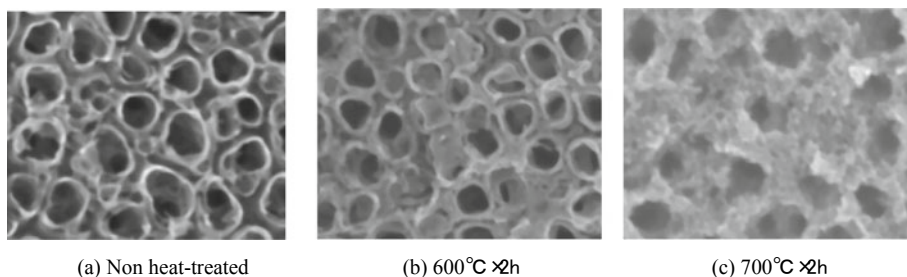
Fig. 3 SEM photographs of heat-treated TiO2 nanotube arrays

3.1.3   Relationship Between Heat Treatment Temperature and Morphology of Nanotube Arrays

Figure 3 shows the morphology of heat treated TiO_2 nanotube arrays after 20 V anodic oxidation. It can be seen from the figure that the morphology of nanotubes after heat treatment at 600 °C is basically unchanged compared with that of micro-heated samples, except that the wall thickness of the nanotubes increases slightly (shown in Fig. 3(b)), which may be related to the merger and growth of particles during sintering. However, when the heat treatment temperature increases to 700 °C, the TiO_2 particles that make up the nanotubes form a serious agglomeration, the wall thickness of the nanotubes increases obviously, and the tubular structure is incomplete (shown in Fig. 3(c)).

3.1.4   Phase Analysis of Heat Treated Nanotube Arrays

Figure 4 is the XRD spectra of anodized TiO_2 nanotube arrays before and after heat treatment. It can be seen from the figure that the non heat-treated TiO_2 nanotube arrays have only the diffraction peak of metallic titanium, which is uncrystalline (amorphous). After heat treatment at 500 °C, the diffraction peak attributable to anatase TiO_2 (101) crystal plane appears at the diffraction angle 2θ of 25.3°, which indicates that the nanotubes are transformed into anatase crystal structure. After the heat treatment temperature rises to 600 °C, the diffraction peaks of rutile (110) and other crystal planes appear at the diffraction angle 2θ of 27.5°, which indicates that the anatase crystal structure continue to transform into the mixed crystal structure of anatase and rutile phases. It is generally believed that the photocatalytic activity of anatase phase and rutile mixed crystal structure is higher.
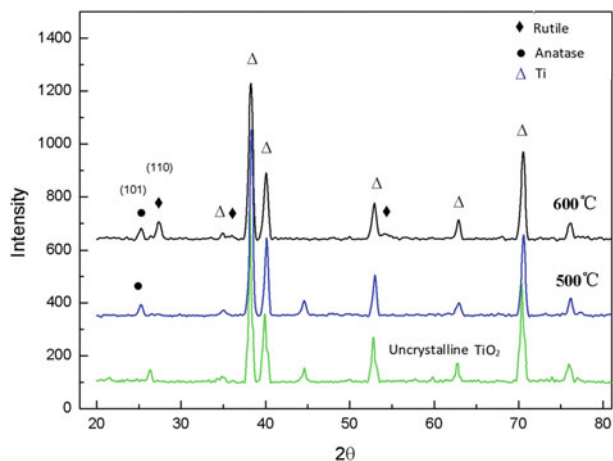
Fig. 4 XRD spectra of TiO$_2$ nanotube arrays before and after heat treatment

3.2 Influencing Factors of Photocatalytic Degradation of Polyacrylamide Solution

3.2.1 Effect of Heat Treatment Temperature

Figure 5 shows the effect of nanotube samples with anode voltage of 20 V and heat treatment at 400–600 °C for 2 h on photocatalytic degradation of polyacrylamide. It can be seen that the degradation rate of polyacrylamide is the highest when the heat treatment temperature is 600 °C, which reaches 82% after 5 h, while the degradation rate of the sample heat treated at 500 °C is only about 50%. Combining with Fig. 4, it can be seen that the TiO$_2$ nanotube arrays are anatase and rutile mixed crystals at 600 °C. When they are combined, overlapping electron levels can be generated. When the valence band electrons are excited by photons, they may transit between energy levels to form a narrow forbidden band, which is more conducive to the absorption of low energy visible light, so the photocatalytic activity of the mixed crystals is higher than that of the single component anatase.

3.2.2 Effect of Pipe Diameter

Figure 5 shows the effect of nanotube samples with diameters of 30–50 nm and 80–120 nm on photocatalytic degradation of polyacrylamide. It can be found that the degradation rate of samples with diameter of 30–50 nm is about 60% in 5 h, and for samples with diameter of 80–120 nm after the same reaction time, the degradation

rate has exceeded 80% and increased by about 33%. This indicates that the degradation rate of polyacrylamide increases with the increase of pipe diameter under the condition that the tubular morphology remains intact. The reason may be that more polyacrylamide molecules to be degraded are adsorbed in the nanotube samples with larger diameter, and the products after degradation are easier to desorb (Fig. 6).
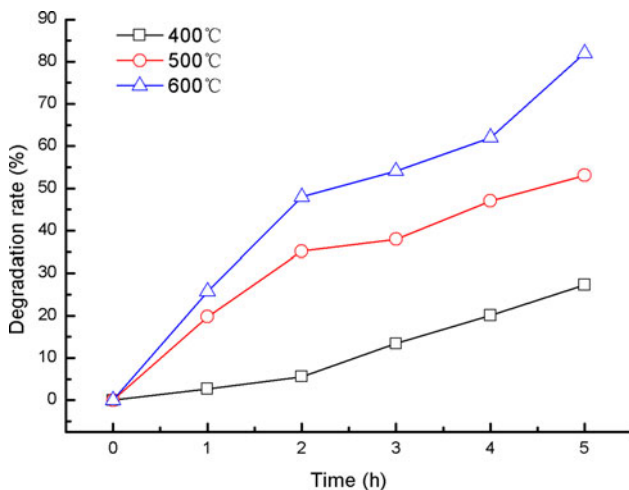
Fig. 5 Effect of heat treatment temperature on degradation rate of polyacrylamide
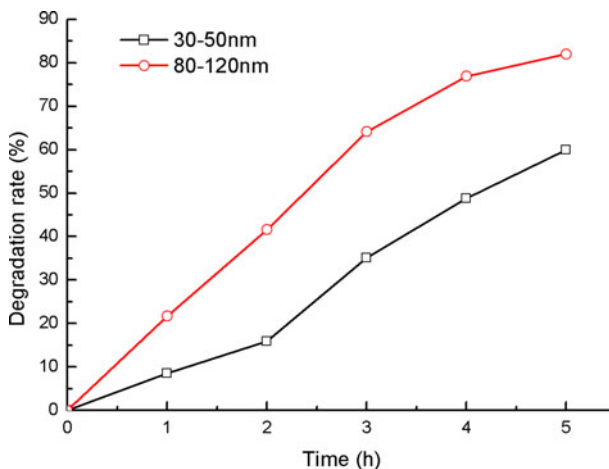
Fig. 6 Effect of pipe diameter on degradation rate of polyacrylamide
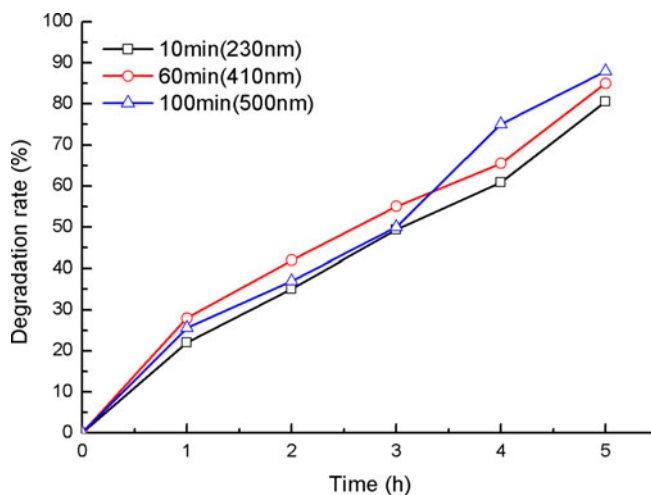
Fig. 7 Effect of pipe length on degradation rate of polyacrylamide

3.2.3 Effect of Pipe Length

Figure 7 shows the effect of nanotube samples with length of 230 nm, 410 nm and 500 nm on photocatalytic degradation of polyacrylamide (20 V anode voltage, corresponding oxidation time are 10 min, 60 min and 100 min, respectively). It can be seen from the figure that the degradation rate of nanotube samples with length of 410 nm and 500 nm is slightly higher than that nanotube samples with length of 230 nm in different time intervals, which indicates that the pipe length has little influence on the degradation rate. This may be due to the fact that the adsorption of the polyacrylamide to be degraded changes slightly with the increase of pipe length under the condition of the same diameter (the anode voltage remains unchanged), so the effect of pipe length on the degradation rate is also small.

3.2.4 Effect of Usage Times on Photocatalysis

Figure 8 shows the effect of photocatalytic degradation of polyacrylamide by nanotube samples with diameter of 80–120 nm after 5 h of reaction under repeated use. It can be found that the degradation rate is about 80% after 5 h during the 8 times of recycling, which indicates that the catalytic activity of the TiO$_2$ nanotube arrays to the target degradation products can be maintained for a long time after the degradation of polyacrylamide, and because the nanotubes grow on the titanium substrate in situ, it is convenient for the recovery and multiple use of the catalysts.
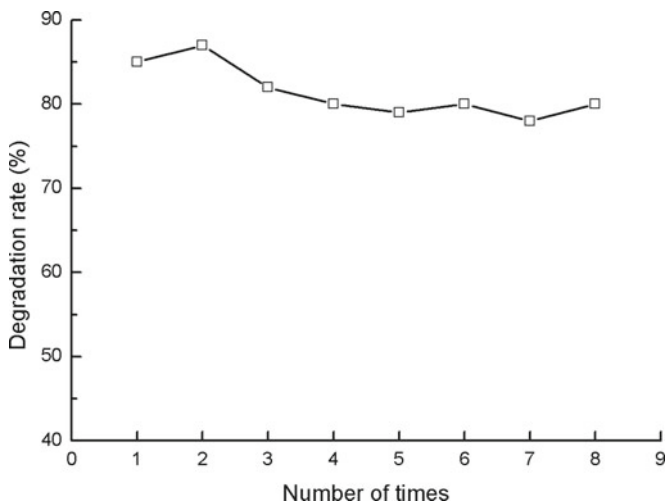
Fig. 8 Effect of usage times on degradation rate of polyacrylamide

4   Conclusion

(1) In order to study the effect of nanotube structure on the degradation of oilfield polyacrylamide wastewater, different diameter and length of TiO_2 nanotube array films were prepared by changing the anode voltage and oxidation time, and the phase change of TiO_2 nanotube arrays was studied by changing the heat treatment temperature.

(2) Taking polyacrylamide wastewater with initial concentration of 98 mg·L^{-1} as the target degradation product, the effects of heat treatment temperature, pipe diameter, pipe length and usage times on the degradation rate were studied. The results show that the photocatalytic activity of nanotubes is the highest after heat treatment at 600°C. The degradation rate of polyacrylamide increases with the increase of the diameter of nanotubes. When the diameter of nanotubes is 80–120 nm, the degradation rate exceeds 80%, while the length of nanotubes has little effect on the degradation rate. During 8 times of recycling, the photocatalytic activity of nanotubes is still high, and the degradation rate is about 80% in 5 h.

(3) Through the study of preparation methods of TiO_2 nanotube arrays and the analysis of factors affecting the degradation of polyacrylamide solution, the theoretical basis and experimental basis for the application of TiO_2 nanotubes to degrade the sewage in lots of oil and gas fields are provided, with a view to the popularization and application of this research in the degradation of polyacrylamide-containing wastewater in oil and gas fields.

Data Availability The data used to support the findings of this study are included within the article.

Conflicts of Interest The authors declare that they have no conflicts of interest.

References

1. Yin, G.: Preparation and Application of TiO$_2$ Nanomaterials in Oilfield. China University of Geosciences, Beijing (2012)
2. Zhao, H., Dai, C., Liang, L., Wang, X., et al.: Research on nonionic polyacrylamide zirconium gel fracturing fluids in coalbed methane gas wells. Petrol. Drill. Tech. **4**(1), 64–68 (2012)
3. Shu, B.: Preparation of Polymetallic Loaded Particle Electrode and Its Application to Simulated Wastewater Treatment of Polyacrylamide. Southwest Petroleum University (2017)
4. Song, W., Nie, Y., Li, S., et al.: preparation and photocatalytic properties of coated TiO$_2$ photocatalytic material. J. Funct. Mater. **49**(10), 10006–10011 (2018)
5. Wu, W.: A study on acidizing flowback water treatment. Chem. Eng. Oil Gas **46**(3), 110–114 (2017)
6. Dawei, G., Grimes, C.A., Varghese, O.K., et al.: Research development of wear-resistance coating alloy surface. Mater. Res. Soc **16**(1), 3331–3334 (2001)
7. Yang, S., Yin, L., Yang, S., Gong, Q., et al.: Preparation of TiO$_2$/ZnO nanotubes and their photocatalytic activity. Nanotechnol. Precis. Eng. **9**(4), 322–328 (2011)
8. Adachi, M., Murata, Y., Okada, I., Yoshikawa, Y.: Formation of Titania nanotubes and applications for dye-sensitized solar ceils. J. Electrochem. Soc. **150**(8), 488–493 (2003)
9. Uchida, S., Chiba, R., Tomiha, M., Masaki, N., Shirai, M.: Application of Titania nanotubes to a dye-sensitized solar cell. E-lectrochemistry **70**(6), 418–420 (2002)
10. Varghese, O.K., Gong, D., Paulose, M., Ong, K.G., Grimes, C.A., et al.: Hydrogen sensing using titania nanotubes. Sens. Actuators B **93**, 338–344 (2003)

Experimental Study on Anisotropy Characteristics of Shale Mechanical Properties

Honglin Huang, Jun Li, and Cheng Zhang

Abstract Shale is a layered sedimentary rock and its bedding planes is developed. In order to study the influence of shale bedding planes on rock mechanics properties, the compressive, tensile and shear tests of Longmaxi formation shale in south Sichuan, China were carried out. The results show that: (1) the bedding planes is the weak part of the shale structure, and its mechanical properties show obvious anisotropic characteristics. (2) With the increase of the included Angle of the bedding planes, the compressive and shear strength showed a "U" pattern of change, which first decreased and then increased, while the tensile strength and elastic modulus gradually decreased, and Poisson's ratio first decreased and then increased. (3) With the increase of confining pressure or normal stress, the compressive strength, shear strength and elastic modulus gradually increase, while Poisson's ratio presents two ways of increasing and decreasing, and the rate of change gradually decreases. (4) The increase of confining pressure or normal stress can reduce the influence of bedding planes on shale mechanical properties, that is, it has "confining pressure effect". The research results provide technical reference for wellbore stability and hydraulic fracturing design of shale gas in Longmaxi Formation, and theoretical basis for efficient and economic exploitation of shale gas.

H. Huang · J. Li (✉) · C. Zhang
China University of Petroleum-Beijing, Changping, Beijng 102200, China

J. Li
China University of Petroleum-Beijing at Karamay, Karamay 834000, China

1   Introduction

Shale gas mainly occurs in the dark mud shale or high carbon mud shale with low porosity and low permeability and rich in organic matter. Shale gas has a high amount of geological resources and huge economic benefits [1, 2]. However, the horizontal section of the shale gas horizontal well is generally long, and the bedding planes of reservoir rock is developed, and some of them have high and steep inclination Angle. The surrounding rock of the wellbore is prone to collapse during the drilling process, and the drilling fluid loss is serious, which severely limits the safe and efficient drilling of the shale gas well [3]. Shale gas reservoirs have extremely low porosity and permeability, and their airflow resistance is much larger than that of conventional natural gas reservoirs. Therefore, economic exploitation can only be realized after reservoir transformation such as acidification and fracturing [4–7].

At present, the anisotropy of layered rocks has been studied extensively by many scholars. Chen et al. [8] studied the mechanical properties of sandstone; Liu et al. [9, 10] studied the mechanical properties of slate based on different theoretical criteria. Tan et al. [11] studied the mechanical properties of non-mean gneiss. All the above studies show that the tensile strength has a strong correlation with the bedding planes. Moreover, Du et al. [12] conducted the Brazilian disk splitting test on the carbonaceous shale and found that the strength gradually increased with the increase of the Angle between foliation and loading direction. Cho et al. [13] studied the anisotropy of elastic parameters and strength of gneiss, shale and schist through uniaxial compression and Brazilian splitting test from different angles. Niandou et al. [14] studied Tournemire shale anisotropy based on triaxial compression test and divided its fracture mode into shear failure and tensile failure. Kovrizhnykh et al. [15–17], based on the triaxial compression test, conducted a systematic study on the anisogenies of mechanical properties of layered rock masses such as siltstone, schist, Himalayan schist and sand Slate, and the test results all showed that the compressive strength varied with the Angle between foliation surface/schist surface/bedding surface and the loading direction in a "U" shape. In addition, based on experimental tests, Heng et al. [18] tested the anisotropy characteristics of shale direct shear strength. Therefore, although there are many researches on the anisotropy of layered rocks, the experimental research on the influence of bedding planes on the shale mechanical properties is relatively weak systematically, and the mechanical properties of reservoir rocks are of great significance to the wellbore stability analysis of horizontal well and hydraulic fracturing construction design. Therefore, it is particularly necessary to study the anisotropy of shale's mechanical properties.

Sichuan area in China based on shale of Longmaxi Formation as an example, based on the different Angle of bedding planes shale tensile, shear and compressive strength test, analysis of the bedding planes under the influence of the anisotropic characteristics of the mechanical properties of shale, for shale gas horizontal wellbore stability and the design of hydraulic fracturing construction to provide

Table 1 Experimental equipment and load control methods for each test item

Test project	Laboratory equipment	Load control method
The compressive test	TAW-1000 Deep water pore pressure servo experimental system	Axial displacement control (0.04 mm/min)
The tensile test	MTS-815 Rock mechanics testing system	Horizontal shear displacement control (0.002 mm/s)
The shear test	RMT–150C Digital controlled electro-hydraulic servo testing machine	Axial displacement control (0.005 mm/s)

technical reference, to supply a basis for high efficient, economic exploitation of shale gas.

2 Test Preparation

The samples were taken from Longmaxi Formation in south Sichuan, China, and buried at a depth of 2560–2600 m. The formation is black - dark black carbonaceous shale. The size error of the processed sample is ±0.5 mm, and the parallelism of the end face is ±0.02 mm. Three samples were made for each test group and the average value was taken to ensure the reliability of the test results.

The tensile test is a single load test, while the confining pressure and shear stress used in the compressive test are all 0, 25, 40 and 50 MPa. The equipment and load control methods used in each test item are shown in Table 1.

3 Test Results and Analysis

3.1 Mechanical Properties of Shale

The Tensile Strength. The tensile strength was measured by The Brazilian disk splitting method. The included Angle between the loading direction and the bedding planes was 0, 30, 60 and 90°, respectively (see Fig. 1).
According to Eq. (1), the tensile strength of different bedding angles is calculated.

$$S_t = \frac{2P}{\pi DH} \tag{1}$$

Fig. 1 The schematic
diagram of the Angle between
load direction and bedding
planes
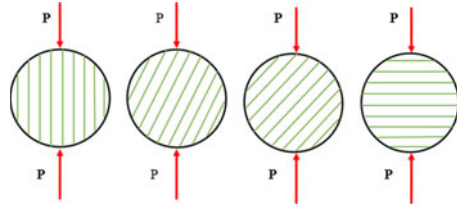
Fig. 2 The test results for
tensile strength
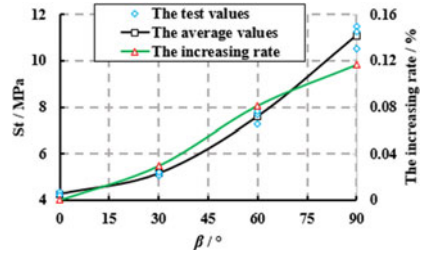
Where, S_t is the tensile strength of the shale sample, MPa; P is the size of the applied concentrated load, KN; D is the diameter of the sample, mm; H is the thickness of the sample, mm.

According to the test and calculation results (see Fig. 2), the tensile strength of shale ranges from 4.25 to 11.50 MPa. With the increase of the included Angle, the tensile strength of shale increases gradually, and the increase rate increases gradually. When the included Angle increased from 0 to 30°, the tensile strength of shale increased by 0.9–1.02 MPa, with a small increase. When increasing from 30 to 60°, the increment of tensile strength is 2.19–2.61 MPa; when increasing from 60 to 90°, the increment of compressive strength is 3.27–3.79 MPa. When the included Angle is small, the cohesive force of the bedding plane is mainly overcome. Therefore, the shale at 0° has the lowest tensile strength. With the increase of the included Angle, not only the cohesive force of the bedding plane but also the friction force of the bedding plane should be overcome, and the influence degree of the rock matrix body also increases gradually. When the included Angle is 90°, the tensile strength of shale is mainly affected by the nature of the matrix body, so the tensile strength is the highest.

The Shear Strength. Direct shear test is the simplest and most direct way to determine the shear strength of a rock (see Fig. 3).

According to the Eq. (2), the shear strength of shale with different included angles under different normal stress conditions is calculated.

$$\begin{cases} \sigma_n = \frac{N}{A} \\ \tau = \frac{Q}{A} \end{cases} \tag{2}$$

Fig. 3 The schematic
drawing of shear test
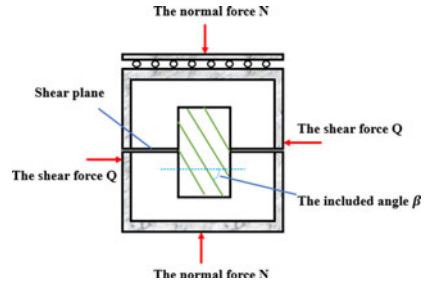
Fig. 4 The test results for
shear strength
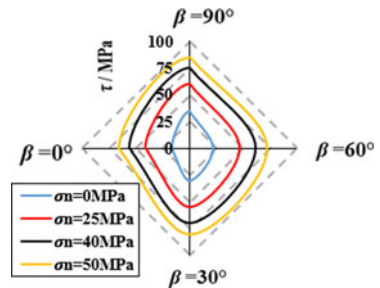
Where, N is the normal force, KN; Q is the shear force received by the sample, KN; A is the effective shear area of the sample along the shear direction, m^2.

As can be seen from the test and calculation results (see Fig. 4), when the included Angle is 0°, the shear strength of shale is the shear mechanical property of bedding surface, and the shear strength of shale is the lowest. When the included Angle is 90°, the shear strength is the shear mechanical property of matrix body, and the shear strength of shale is the largest. Shale samples with an Angle of 30 and 60° are influenced by the shear mechanical properties of matrix body and bedding surface, and the shear strength of shale samples with an Angle of 30° is greater. When the normal stress increases, the shear strength of shale also increases gradually. When the normal stress increases to 25 MPa or above, the shear strength of the shale samples with an Angle of 0° is greater than that of the shale samples with an Angle of 0° when the normal stress is 0 MPa, indicating that with the increase of the normal stress, the influence of the anti-shear strength of the bedding weakens. It may be due to the compaction of micro-cracks and micro-pores between the bedding planes and matrix body, and the enhancement of the adhesion and friction force between the particles, thus enhancing the overall strength of shale samples.

The Compressive Strength. Uniaxial and triaxial compression tests were carried out on the shale with different bedding angles in order to further study the influence of bedding planes on the shale mechanical properties. The tested confining pressure is transferred by liquid, and the side of the sample is sealed by soft rubber sleeve to avoid the influence of liquid on the sample (see Fig. 5).

Fig. 5 The schematic
diagram of compressive test
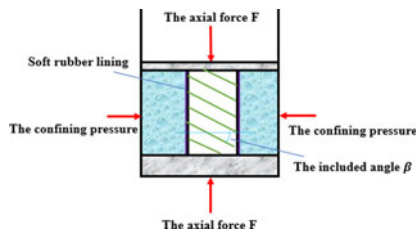
According to the Eq. (3), the compressive strength of shale samples with different bedding angles is calculated.

$$\sigma_c = \frac{F_{max}}{A} \tag{3}$$

Where, σ_c is the compressive strength of the shale sample, MPa; F_{max} is the maximum failure load of rock specimen, KN; A is the effective compression area of the specimen, m^2.

As can be seen from the test and calculation results (see Fig. 6), with the increase of confining pressure, the compressive strength of shale with the same included Angle increases. At low confining pressures (0, 25 MPa), the compressive strength in the direction of parallel bedding plane is greater than that in the direction of vertical bedding plane, while at high confining pressures (40, 50 MPa), the compressive strength in the direction of parallel bedding plane is less than that in the direction of vertical bedding plane. With the increase of the included Angle, the compressive strength of shale under the same confining pressure firstly decreases and then increases, and the minimum value is about 30°. When drilling in the horizontal section of shale reservoir, if the Angle between the wellbore axis and the surrounding rock strata is a certain Angle, the overspeed of drilling tool lowering, the disturbed drilling tool colliding with the wellbore wall, and the excessive ECD at the bottom of the wellbore can easily lead to falling block and collapse of surrounding rock.

The Elastic Modulus. According to the results of tests, the elastic modulus of shale under different confining pressure is calculated by Eq. (4).

$$E = \frac{\sigma_c - \sigma_3}{\varepsilon_1 - \varepsilon_3} \tag{4}$$

Where, E is the elastic modulus of shale sample, GPa; σ_3 is the lateral stress, MPa; ε_1 is the axial dependent variable. ε_3 is the lateral stress variable when the axial stress is equal to the lateral stress.

According to the test and calculation results (see Fig. 7), as the confining pressure increases, the elastic modulus of the sample also increases. During uniaxial compression, the elastic modulus of shale decreases first and then increases with the increase of Angle. Under the same confining pressure, the elastic modulus decreases

Fig. 6 The test results for compressive strength
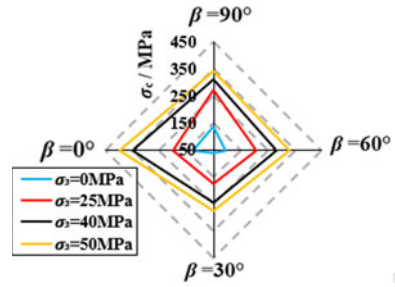
Fig. 7 The results of tests for elastic modulus
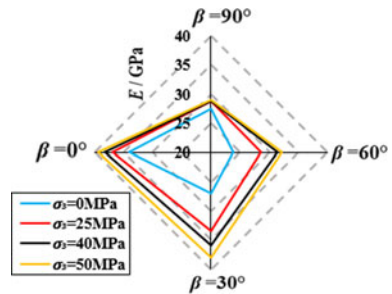
with the increase of Angle, and the decreasing velocity first increases and then decreases. Confining pressure increases, the elastic modulus increases rate size comparison: < 0° 90° <= 30° to 60°. And the elastic modulus of sample at 90° is basically unchanged, which may be caused by the suppression of lateral deformation due to the compaction effect of confining pressure on the micro-pores and micro-cracks of the bedding surface, while in the vertical direction of the bedding planes, the compaction effect of confining pressure on the micro-pores and micro-cracks of the bedding surface is relatively weak, so it has little influence on the axial deformation capacity of the sample, making the elastic modulus basically unchanged.

The Poisson's Rate. According to the test results, Poisson's ratio of shale is calculated by Eq. (5), and the influence of various confining pressures on the Poisson's ratio of shale samples is discussed.

$$\mu = \frac{\varepsilon_2 - \varepsilon_a}{\varepsilon_1 - \varepsilon_a} \tag{5}$$

Where, μ is the elastic modulus of shale sample; ε_2 is the lateral strain at the failure of the shale sample, MPa; The ε_1 is the axial stress variable when the shale sample is broken. ε_3 is the lateral stress variable when the axial stress is equal to the lateral stress.

Fig. 8 The results of tests for Poisson's rate
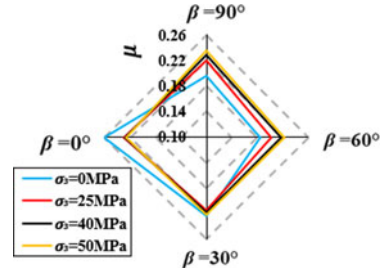
It can be seen from the test and calculation results (as shown in Fig. 8) that under the same confining pressure, Poisson's ratio of the sample generally decreases first and then increases with the increase of Angle, and its minimum value is around 60°. However, when the confining pressure is 0 MPa, Poisson's ratio of the sample at all angles changes greatly, which may be caused by the disintegration of pores and micro-cracks between the layers. When the confining pressures are 25, 40 and 50 MPa, Poisson's ratio changes closely with the Angle, and its size does not change much, which may be caused by the suppression of disintegration under high confining pressures.

3.2 Mechanical Properties of Shale

In order to express more intuitively that the mechanical properties of shale are affected by bedding planes, the anisotropy δ_i is introduced, and its calculation method is as follows:

$$\delta_i = \sum \frac{|x_\beta - x_0|}{x_0} \tag{6}$$

Where, x_β is the mechanical parameter of shale when the included Angle is; x_0 is the shale mechanical parameter when $\beta = 0$.

The calculation results of anisotropy of mechanical parameters under different confining pressures are shown in Fig. 9. In the figure, δ_1 is the anisotropy of the compressive strength, δ_2 is the anisotropy of the shear strength, δ_3 is the anisotropy of the elastic modulus, and δ_4 is the anisotropy of Poisson's ratio. From the definition of δ_i it can be seen that the greater the value of δ_i is, the greater the numerical fluctuation is, indicating that the mechanical property is more affected by the bedding planes. It can be seen from the calculation results that, on the whole, with the increase of confining pressure (or normal stress), the anisotropy of various mechanical parameters of shale sample decreases gradually. This indicates that with the increase of confining pressure (or normal stress), the shale sample is strengthened by compaction, and the micro-fractures and micro-pores of bedding surface

Fig. 9 The calculation results of anisotropy of mechanical parameters
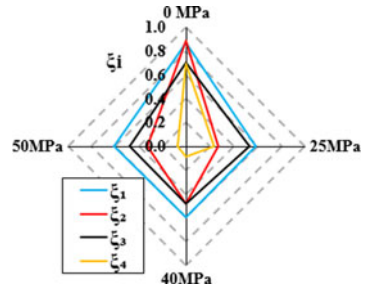
are compacted, thus reducing the influence of bedding planes on various shale mechanical properties, that is, it has the "confining pressure (or normal stress) effect", and Poisson's ratio is more sensitive to the change of confining pressure (or normal stress).

To sum up, the stratified surface is the weak part of the shale structure, and its mechanical properties show obvious anisotropic characteristics. Moreover, the size of micro-pores and micro-fractures on the bedding surface belongs to the category of nanometer, and the formation of the capillary force is much big, In the process of drilling, drilling fluid easily along the bedding surface invades wellbore wall rock, thus the strength decrease (and the drilling fluid immerses the longer, the higher the extent of damage), collision of drilling tool, the drag of drilling fluid with high viscosity, etc. all can cause damage of wellbore wall rock, which are easy to cause some accidents, such as drop a block and get stuck, wellbore collapse; In the process of hydraulic fracturing, with the injection of fracturing fluid, the fracturing fluid first penetrates into the reservoir rock along the weak bedding surface, that is, the fractures will expand along the weak bedding surface, and it is difficult to form a network fracture structure, which seriously affects the fracturing effect and fails to meet the expected production requirements.

4 Conclusion

In this paper, tensile, shear and compressive tests have been carried out on the stratified shale, and the effects of the bedding planes on tensile, shear, compressive, elastic modulus and Poisson's ratio have been analyzed, and the anisotropic characteristics of the shale have been studied.

(1) bedding surface is the structural weak surface of shale, and its mechanical properties show obvious anisotropic characteristics.
(2) With the increase of the included Angle of the bedding, the compressive and shear strength showed a "U" pattern of change, which first decreased and then increased, while the tensile strength and elastic modulus gradually decreased, and Poisson's ratio first decreased and then increased.

(3) With the increase of confining pressure, the compressive strength, shear strength and elastic modulus gradually increase, while Poisson's ratio presents two ways of increasing and decreasing, and the rate of change gradually decreases. The increase of confining pressure can reduce the influence of bedding planes on the mechanical properties of shale rock, that is, it has "confining pressure effect".

(4) The layered sedimentary structure of shale and the weak cementation between layers are the main reasons for the anisotropy of mechanical properties. And the friction on the bedding surface is relatively small and the strength is relatively low, so sliding shear failure is easy to occur along the bedding planes, causing rock fragments to fall off, and in serious cases, complicated accidents such as tool jamming and wall collapse will occur.

References

1. Zou, C.N., Dong, D.Z., Wang, S.J., Li, J.Z., Li, X.J., Wang, Y.M., Li, D.H., Cheng, K.M.: Geological characteristics, formation mechanism and resource potential of shale gas in China. Petrol. Explor. Dev. **7**(06), 641–653 (2010)
2. Dong, D.Z., Zou, C.N., Li, J.Z., Wang, S.J., Li, X.J., Wang, Y.M., Li, D.H., Huang, J.L.: Resource potential, exploration and development prospect of shale gas in the whole world. Geol. Bull. China **30**(Z1), 324–336 (2011)
3. Zhao, K., Yuan, J.L., Deng, J.G., Tan, Q., Yu, B.H., Wang, M.C.: Effect of bedding plane occurrence on horizontal shale gas wellbore stability. Sci. Technol. Eng. **13**(03), 580–583 (2013)
4. Tang, Y., Tang, X., Wang, G.Y., Zhang, Q.: Summary of hydraulic fracturing technology in shale gas development. Geol. Bull. China **30**(Z1), 393–399 (2011)
5. Wang, Y.K.: Overview of shale gas exploration and development technology. Petrol. Knowl. **02**, 42–43 (2019)
6. Loucks, R.G., Reed, R.M., Ruppel, S.C., et al.: Morphology, genesis, and distribution of nanometer scale . J. Sediment. Res. **79**(12), 848–861 (2009)
7. Brannon, H.D., Kendrick, D.E., Luckey, E., et al.: Multistage fracturing of horizontal Wells using ninety - five quality foam provided improved shale gas production. SPE **124**, 767 (2009)
8. Chen, C.S., Pane, E., Amadei, B.: Determination of deformability and tensile strength of anisotropic rock using Brazilian tests. Int. J. Rock Mech. Min. Sci. **35**(1), 43–61 (1998)
9. Liu, Y.S., Fu, H.L., Rao, J.Y., Dong, H., Cao, Q.: Research on Brazilian disc splitting tests for anisotropy of slate under influence of different bedding orientations. Chin. J. Rock Mech. Eng. **31**(04), 785–791 (2012)
10. Liu, Y.S., Fu, H.L., Wu, Y.M., He, X.B., Dong, H.: Study on Brazilian splitting test for slate based on single weak plane theory. J. China Coal Soc. **38**(10), 1775–1780 (2013)

11. Tan, X., Heinz, K.: Brazilian split tests and numerical simulation by discrete element method for heterogeneous gneiss with bedding structure. Chin. J. Rock Mech. Eng. **33**(05), 938–946 (2014)
12. Du, M.P., Pen, P.Z., Ji, W.W.: Study on the spatial and temporal characteristics of the failure process of carbonaceous shale under Brazilian splitting load. Rock Soil Mech. **37**(12), 3437–3446 (2016)
13. Cho, J.-W., Kim, H., Jeon, S., Min, K.-B.: Deformation and strength anisotropy of Asan gneiss, Boryeong shale, and Yeoncheon schist. Int. J. Rock Mech. Min. Sci. **50**, 158–169 (2012)
14. Niandou, H., Shao, J.F., Henry, J.P., et al.: Laboratory investigation of the mechanical behavior of Tournemire shale. Int. J. Rock Mech. Min. Sci. **4**(1), 3–16 (1997)
15. Kovrizhnykh, A.M., Usol'tseva, O.M., Kovrizhnykh, S.A., Tsoi, P.A., Semenov, V.N.: Investigation of strength of anisotropic rocks under axial compression and lateral pressure. J. Min. Sci. **53**(5), 831–836 (2017)
16. Kuila, U., Dewhurst, D.N., Siggins, A.F., Raven, M.D.: Stress anisotropy and velocity anisotropy in low porosity shale. Tectonophysics **503**(1–2), 34–44 (2011)
17. Loucks, R.G., Reed, R.M., Ruppel, S.C., Jarvie, D.M.: Morphology, genesis, and distribution of nanometer-scale pores in siliceous mudstones of the Mississippian Barnett Shale. J. Sediment. Res. **79**(12), 848–861 (2009)
18. Heng, S., Yang, C.H., Zeng, Y.J., Zhang, B.P., Guo, Y.T., Wang, L., Wei, Y.L.: Anisotropy of shear strength of shale based on direct shear test. Chin. J. Rock Mech. Eng. **33**(05), 874–883 (2014)

Interactive Web Based Drilling 3D Visualization Software for RTOC of Engineering Technology Internet of Things

Shanshan Liu, Minglei Li, Chuanjie Jin, Lun Shi, Xianlu Cai, Jinchao Wang, and Pengyin Yan

Abstract The objective of this paper is to propose a brand new drilling 3D visualization solution based on WebGL and describe the architecture and application of the cross-platform drilling 3D Visualization Software. The software is part of the Engineering Technology Internet of Things (IoT) System, interfacing with other software and also capable of direct hardware interfacing for data retrieval and system control. It is fully web-based, real time and used in RTOC (Real-Time Operating Center) of IoT system which is a software system for drilling process remote monitor and decision. The data acquired from rig sites is remote transmitted to the data center and analysed in the RTOC, the exports in RTOC use many software do the monitoring and analyzing, if there's some problems, they can quickly give advices to rig sites. It can put borehole trajectories, LWD, MWD, mud logging, well logging, seismic slice, and formation tops all into a 3D scene, by rotating and zooming, thus experts can get realistic underground scenes. It provides a convenient way to viewing real-time 3D drilling scene in the mainstream browsers on difference systems and devices. The web-based 3D visualization of drilling, LWD, MWD and geological data greatly improves the efficiency of drilling process monitoring and decision-making. Contrast with traditional desktop software, this solution does not need install and can supports all mainstream operating systems, such as Windows, Linux, macOS, iOS and Android, it's a more lightweight and economical solution.

Keywords Internet of Things · Cross-platform · Drilling 3D visualization · WebGL · Lightweight data service

S. Liu (✉) · L. Shi · X. Cai · J. Wang · P. Yan
College of Petroleum Engineering, China University of Petroleum, Beijing 102249, China

M. Li
China Petroleum Materials Procurement Center, China National Petroleum Corporation, Beijing 100029, China

C. Jin
School of Mechanical Science and Engineering, Northeast Petroleum University, Da qing 163318, China

1 Introduction

In November 2005, the International Telecommunication Union (ITU) released a report entitled "the Internet of things", which formally put forward the Internet of things (IOT), which attracted the attention of governments and industries. Petroleum industry is an important industry of IOT application, which has been applied in some enterprises. For example, geological and engineering integration work ideas and workflow are used to solve practical problems encountered in production. Exploring the remote support platform of engineering technology integrated service information can provide the best customer value. The engineering technology Internet of things system uses the Internet of things technology to establish data automatic acquisition and transmission platform, data integrated storage platform, data comprehensive application and analysis platform. The construction of remote operation support center and decision-making center can realize the automatic collection, integrated storage and remote transmission of field data, real-time data detection, early warning and analysis optimization, and remote technical support and decision-making.

The Internet of things system of engineering technology includes the automatic data acquisition, remote data transmission, data center and remote operation support center (RTOC). Through the remote operation support center, technical experts can view the data and video information of the well pad on the computer of the center for remote analysis and decision-making.

Access to the display of drilling, MWD, LWD and geological data available in 3D way on web greatly enhances the monitoring of drilling process and production command efficiency. RTOC (Real-Time Operations Center) aims at building a data acquisition and application, multi-professional collaborative service platform to improve the integrated service capabilities, realize intelligent operation support, the integration of engineering and technology wellbore business and information technology [1–3]. It provides a powerful guarantee for making and transmitting drilling decision quickly and effectively, as is shown in Fig. 1.
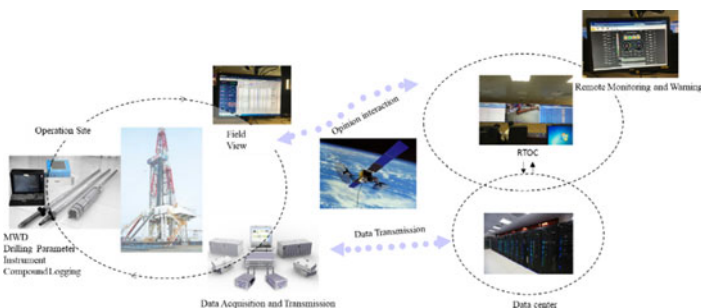
Fig. 1 Real-time operations center architecture

The remote technology required by the engineering Internet of things system puts forward higher requirements for relevant software, It should be convenient to support all mainstream operating systems, browsers and mobile devices.

At present, some companies built such visualization products: developing a 3D visualization software tool to enable drilling engineers, geophysical engineers, geological engineers and reservoir engineers to work with the same data in the same data volume environment, breaking the restrictions of specialty and region [4–9]. Most of the existing 3D visualization technologies adopt client–server (C/S) mode, such as DecisionSpace® and WellViz3D are released as standalone software, which must be installed on local OS. Discovery Web of Kongsberg adopt B/S mode, However, based on ActiveX plug-in technology, special plug-ins need to be installed, which can only support IE browser on PC Windows and cannot support mobile devices (Android, IOS) [10]. Undoubtedly, web-based drilling 3D visualization system is a more convenient way to perform real-time monitoring of drilling. However, there is no 3D visualization technology of drilling based on standard browser at present.

This article introduced a systematic method to design a cross-platform drilling 3D visualization software, which supports real-time data and reservoir geological model. It also proposed an innovative way of drilling engineering design. Experts do not have to sit in front a dedicated workstation in the office all the time and install any app, just open a carried device, such as laptop, iPad, smart phone that can connect the online system, open an WebGL enabled HTML 5 browser, then they can view it.

2   Data of Visualization

The 3D visualization of oil field drilling engineering and geological information is mainly about underground geology, oil reservoir, oil well, etc. 3D visualization of drilling trajectory and other 3D information can accurately establish the virtual reality loop of underground oil field. In order to improve the digitalization of oil field drilling engineering and free the designers and builders from low-level drawing and construction. It is convenient for a series of production processes such as petroleum exploration, design and drilling development, so as to accelerate the speed of petroleum development and production.

To realize engineering geology visualization, the main data include engineering and geology. Engineering data include: design trajectory (target), drilling trajectory, logging curve, logging curve, complex accident information. Geological data include: geological structure information, reservoir information (porosity, permeability, saturation).

The trajectory data include: measuring depth, deviation angle and azimuth angle. Through these three parameters, the vertical depth, North migration, East migration, closure orientation, closure distance and full angle change rate (dogleg degree) of each measuring point can be calculated, so as to determine the 3D trajectory of the

borehole. The software provides many common trajectory calculation methods, including minimum curvature method, curvature radius method, average angle method, balanced tangent method, correction average angle method and chord step method.

Logging curves include real-time LWD and non-real-time conventional logging curves. Generally, they are sampled according to depth. At intervals of a certain depth, there will be a set of logging curve parameters, commonly used parameters include natural gamma ray, resistivity and so on.

In order to improve the displaying efficiency of geological data, in the software, a grid optimization algorithm for reservoir geological model is proposed. Through the optimization algorithm, the requirement of computer hardware for 3D visualization can be reduced and the display efficiency can be improved. The complex geological body data (from Petrel Geological Modeling Data) are displayed by the mesh elimination algorithm, which can also reduce the amount of data and ensure the operation efficiency of the software.

3   3D Visualization Technology on the Web

RTOC enabled the experts to monitor the drilling process with a 3D real-time scene. In the past few years, modern browsers and standards have made great progress [11, 12]. Cross-browser and cross-platform standardization greatly simplifies the design and implementation of rich interactive Web applications. A powerful set of Web standards (commonly known as HTML5) has evolved into a framework on which real browser-based and cross-platform visualization tools can be developed. Special interest in drilling and geological visualization is HTML5, HTML5 canvas, WebGL.

4   System Architecture

Overall Framework: The system all build on HTML 5 standards and open source frameworks.

Under the environment of HTML5, the system adopts B/S architecture building a network platform using Typescript development language. With the help of the language's full object-oriented and strong typing characteristics, the code has strong expansibility, reusability, security, robustness and stability. The system uses object-oriented design method, builds a scalable 3D visualization class library, and encapsulates it as an Angular module, which has strong reusability and inheritance.

The software realizes the unified management of visual objects, displays the coordinates of 3D scene and performs the fundamental functions including zooming, rotating and translating. Figure 2 illustrates software architecture including data, service and view layer.

The data layer is Geological Engineering Warehouse, it uses real-time and relational databases to receive and preserve static and dynamic drilling geological data. Static data include oilfield information, adjacent well data, drilling design data, formation tops, seismic slices and other geological engineering data. Dynamic data include drilling, MWD, LWD and mud logging data.

The service layer based on Node.js and Restify framework, provides a RESTful API. It is the data interface between front-end application and the back-end data warehouse. Communications between the data warehouse and the visualization program will also be performed trough the servers.

The view layer is a single page Web application (SPA), based on HTML5 standards and TypeScript language, using Angular front-end framework and Bootstrap UI library to realize responsive SPA. Through the WebGL interface of browser, 3D visualization components are implemented to realize web-based cross platform interactive 3D visualization application for drilling and geological.

Technology Stack: This software uses a series of open source technologies to build an economical and efficient technology stack including: TypeScript, Angular, Responsive web design, Three.js, Restify.

Fig. 2 System architecture
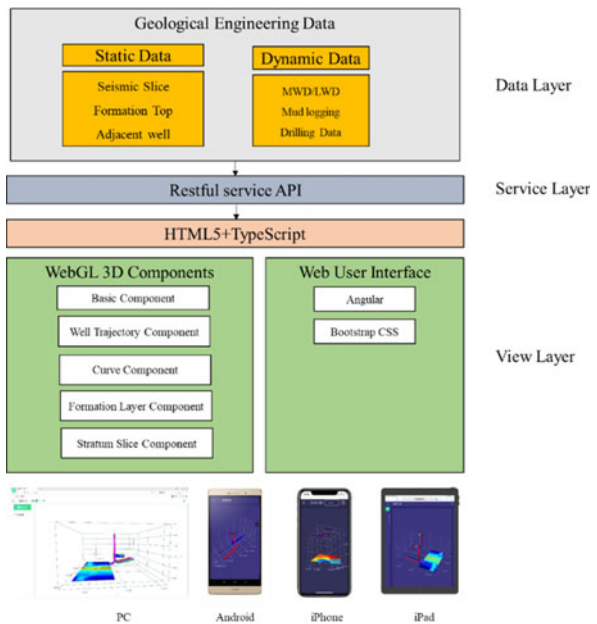
5   3D Visualization Components

According to the type of information to be visualized, corresponding components are developed to realize detailed display of well, borehole, real-time data and geological model, which makes visualization more useful for drilling experts. For example, the Gamma-ray curve related to formation lithology, the curve component can show the curve along the wellbore, and the change of formation lithology can be intuitively displayed, as is shown in Fig. 2.

The main 3D components include 3D visualization fundamental component, well trajectory component, curve component, formation tops component, seismic slice component and reservoir model component, as shown in Fig. 3.

6   RESTful Data Service

It is necessary to develop corresponding background data services to provide data interface support for front-end application. The concept of REST was first proposed by Dr. Roy Thomas Fielding (Fielding and Taylor, 2002), as a method of establishing communication between computer systems. The scheme adopts micro-service architecture, chooses Node.js—a lightweight development platform suitable for cloud platform [13].

Through this data service, an efficient data access channel between front-end application and back-end data warehouse is established, as is shown in Fig. 4. With the help of the static type and object-oriented features of Typescript language, the reliability and maintainability of the service are greatly improved.

Fig. 3   Class diagram of 3D visualization components for drilling

Fig. 4 REST request basic process
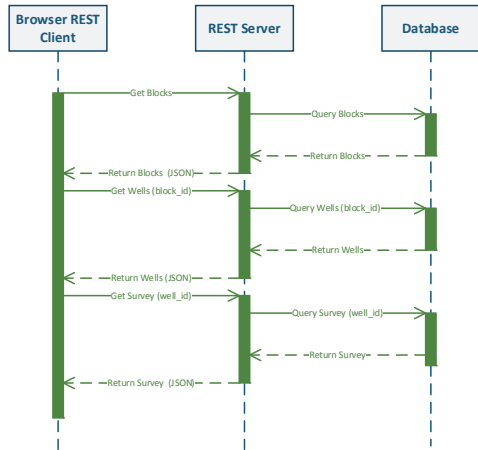
7 Real-Time Data Transmission Scheme

WITSML (Wellsite Information Transfer Standard Markup Language) is the standard for well site information transmission based on XML, which can realize data exchange between rig sites and RTOC [1], but it is huge and the cost of implementing a WITSML server is high. Referring to some design ideas of WITSML, this paper designs a set of real-time data transmission scheme. Data transmission adopts HTTP REST mode and data objects are encapsulated in JSON format. The scheme is simple to implement and can make up for the shortcomings of WITS0.

8 Data Service API

Data Type Definition: Visual drilling data are defined below, including well basic data, drilling trajectory data, logging data, layer interface data and geological slice data and reservoir grid data.

Service Interface Design: According to the definition of drilling visualization data, the corresponding data interface is designed. To access data services, you only need to use HTTP interface address. The address format of the interface is //host: port/api/interface/{parameter}, For example, the host is localhost, the port is 4000, "the block list", and the interface name is "block", no parameters, the interface address is //localhost:4000/blocks. The result is in JSON format. Interface definitions are as shown in Table 1.

The design has strong expansibility and can easily expand support for other drilling and completion and geological data by adding new data interface.

Table 1 Service interface definition

Interface name	Interface address
Block list	Blocks
Well list	wells/{block_id}
Formation tops list	surfaces/{block_id}
Seismic slice list	slices/{block_id}
Trajectory data	survey/{well_id}
Curve list	curves/{well_id}
Events list	events/{well_id}
Curve data	curve/{curve_id}
Formation tops data	surface/{surface_id}

Cross Domain Problem: JSONP (JSON with Padding) is a "usage mode" of JSON. For security reasons, browsers have a homologous policy. The service program is designed to support both JSON and JSONP. It only needs to add "? Callback = jsonp" after the interface address. The program returns JavaScript code in JSONP mode, otherwise it returns JSON data directly. The web page can get JavaScript code generated dynamically from other sources with the open strategy of <script> element. Through ingenious design, executing this code can generate JSON data dynamically required by users. This usage pattern is called JSONP.

Service Deployment: Service program can be deployed conveniently on the cloud server using PM2, NodeJS, and the update and execution of the program can be monitored to ensure the uninterrupted operation and hot update of the program through the powerful process management function of PM2 [14, 15]. Through JSONP, cross-domain access is realized. PM2 is used to deploy the service program. The drilling 3D visualization system developed by this service has the characteristics of high reliability, reliability and easy expansion.

9 Applications

The software is now installed/used in the RTOC, is as shown in Fig. 5, Fig. 6 and Fig. 7. An example—block 291, which contains 11 wells of carbonate rock, including 4 vertical wells, 1 directional well and 6 horizontal wells, among which W291-H9 is the latest horizontal well and the target formation is Ordovician. The drilling trajectory, logging data, complex & accidents and seismic slice data along the well trajectory of the target formation are loaded into the database.

The software can display the complex & accidents of drilling in the block, the seismic characteristics of the target layer, the tops of the target formation, the drilling trajectories and LWD data, and can be used for auxiliary analysis and decision-making. The drilling and geological experts can not only use the workstation of the Office, but also use the mobile equipment to view the data and graphics.
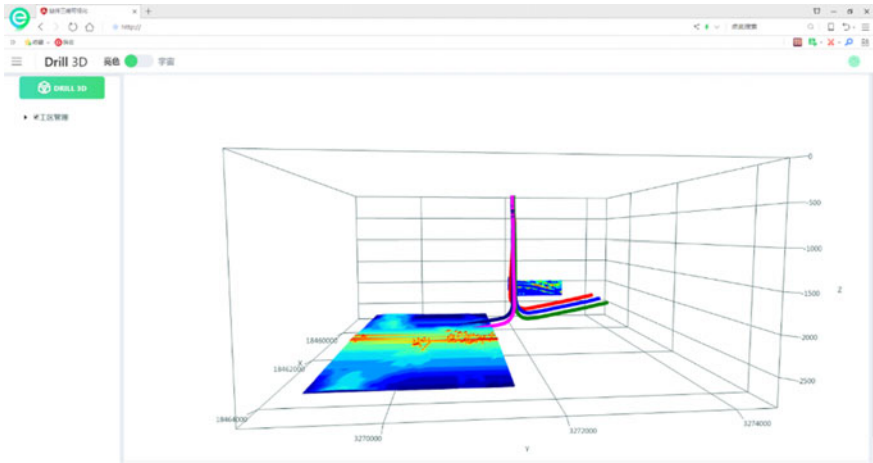
Fig. 5. Drilling 3D scene: well tracks and surface on windows
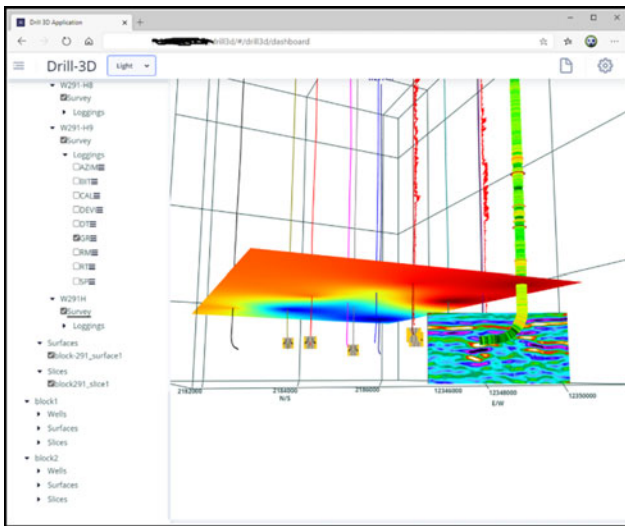
Fig. 6. Drilling 3D scene: well tracks, surface and loggings, trouble tips on windows

Due to the consistency and efficiency of WebGL on various platforms, the software has good performance and consistency on the main platforms, including windows, Linux, macOS, Android and iOS, and the responsive layout can adapt well for the different resolution of desktop devices and mobile devices, for small mobile screen, the software will automatically hide the left tree menu, making the software more practical. The goal of the development is achieved.
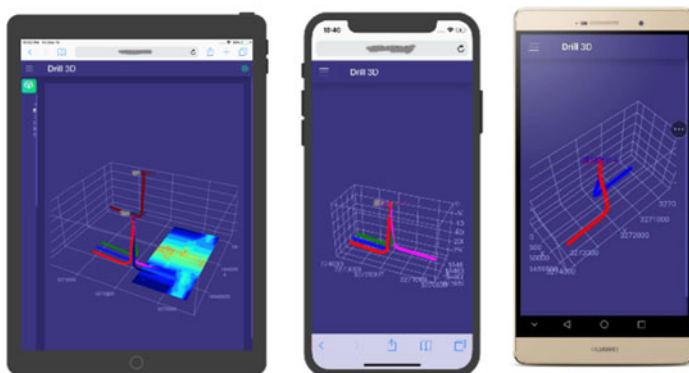
Fig. 7. Drilling 3D scene on iPad, android and iPhone

10   Conclusions

This paper introduces the structure and development of 3D visualization software for drilling engineering and geology. The software is based on WebGL and developed with typescript, an open source framework. Combined with the real-time transmission scheme proposed, a set of low-cost and practical real-time 3D visualization monitoring software is provided for RTOC. The conclusions can be summarized as follows:

(1) The use of native Web technologies allows detailed renderings of the drilling and geological data with no additional software to install. 3D drilling scene renderings are processed entirely within the modern Web browser without a significant tradeoff in performance. The system, termed as WebDrillingViz, meets the three basic requirements of lightweight, cross platform, openness and economic. It is now installed in the RTOC, interfacing with other software system, will enable experts to better understand the status of the drilling, to make better and faster decisions.

(2) WebDrillingViz integrates responsive design and high-performance back-end services, using the latest technologies, such as NodeJS, Angular and Bootstrap frameworks with excellent compatibility. The front-end components can also quickly be embedding into other web applications and the back-end efficient service based on NodeJS platform. The software is completely implemented in Typescript language on both sides, the whole development cycle reduced significantly because using same language. The REST-style data service is a lightweight technology which is more suitable for cloud platform at present. Through JSONP, cross-domain access is realized. PM2 is used to deploy the service program. The service has the characteristics of high reliability, reliability and easy expansion.

References

1. Kaminiski, D.H., Pellerin, N.M., Williams, J.H.: A new data integration and work process system for providing online real-time drilling collaboration. In: Proceedings of the European Petroleum Conference (2002). https://doi.org/10.2523/78343-ms
2. Madaleno, A.L.F., Neto, S.L.S., Dos Santos, L.A., De Oliveira, C.A.L.: Operational and safety improvements of applying real-time analytics in a drilling contractor RTOC. In: Proceedings of the Annual Offshore Technology Conference (2018). https://doi.org/10.4043/28824-ms
3. Ursem, L.J., Williams, J.H., Pellerin, N.M., Kaminski, D.H.: Real time operations centers; the people aspects of drilling decision making. In: Proceedings of the Drilling Conference (2003). https://doi.org/10.2523/79893-ms
4. Dorn, G.A., Touysinhthiphonexay, K., Bradley, J., Jamieson, A.: Immersive 3-D visualization applied to drilling planning. Lead. Edge (Tulsa, OK) (2001). https://doi.org/10.1190/1.1487271.
5. Holt, J., Wright, W.J., Nicholson, H., Kuhn-De-Chizelle, A., Ramshorn, C.: Mungo field : Improved communication through 3D visualization of drilling problems. In: SPE/AAPG Western Regional Meetings (2000). https://doi.org/10.2118/0101-0025-jpt.
6. Rommetveit, R., Bjørkevoll, K.S., Ødegård, S.I., Herbert, M., Halsey, G.W.: Automatic real-time drilling supervision, simulation, 3D visualization and diagnosis on ekofisk. In: SPE/IADC Drilling Conference, Proceedings (2008)
7. Sanstrom, W.C., Hawkins, M.J.: Perceiving drilling learning through visualization. In: Proceedings of the IADC/SPE Asia Pacific Drilling Technology Conference, APDT (2000). https://doi.org/10.2523/62759-ms.
8. Song, P., Shi, L., Zhou, Y., Zhao, Q., Jiang, H., Cui, P.: Study and applications on integrated drilling engineering software for drilling engineering design and real-time optimization, in: Society of Petroleum Engineers - SPE Annual Caspian Technical Conference and Exhibition (2014). https://doi.org/10.2118/172288-ru.
9. Rommetveit, R., Bjørkevoll, K.S., Ødegård, S.I., Sandve, O., Larsen, B., Herbert, M.: EDrilling: linking advanced models and 3D visualization to drilling control systems in real time. In: Offshore Mediterranean Conference and Exhibition 2007, OMC 200 (2007).
10. Phillips, W.: Interactive web-based 3D wellbore viewer enables collaborative analysis, in: SPE Production and Operations (2018). https://doi.org/10.2118/174024-pa.
11. Liu, X., Xie, N., Tang, K., Jia, J.: Lightweighting for Web3D visualization of large-scale BIM scenes in real-time. Graph. Models. (2016) https://doi.org/10.1016/j.gmod.2016.06.001
12. Morozov, I., Chubak, G., Blyth, S.: Interactive 3D/2D visualization for geophysical data processing and interpretation. Comput. Geosci. (2009). https://doi.org/10.1016/j.cageo.2008.10.005
13. Cantelon, M., Harter, M., Holowaychuk, T.J., Rajlich, N.: Node.js in Action. American: Manning Publications (2013)
14. Chenjie, X., Xiang, Z., Xin, P., Wenyun, Z.: Microservice system oriented runtime deployment optimization. Comput. Appl. Softw. **35**(10), 85–93 (2019)
15. Yongyi, L.: Research on cross-domain request scheme for large distributed web system. J. Chang. Univ. **35**(02), 54–56 (2018)

Fracture Pressure Prediction Model for Vertical Wells Considering Temperature and Natural Fractures

S. L. Xia, J. Li, and H. L. Huang

Abstract Accurate prediction of formation fracture pressure before drilling can effectively reduce the occurrence of complex situations in the drilling process. When drilling in high temperature formation with natural fracture development, the current prediction model of fracture pressure is not very applicable. Therefore, by analyzing the stress state of borehole wall affected by temperature and natural fractures and combining with the fracture criterion of different borehole wall rocks, the author established the fracture pressure prediction model of high-temperature formation suitable for the development of natural fractures, and analyzed the influence of fracture strike, fracture dip Angle and borehole wall temperature changes on the fracture pressure. The research results show that: the fracture pressure increases gradually with the increase of crack inclination angle. As the angle between the fracture strike and the horizontal maximum principal stress increases, the fracture pressure first increases, then remains unchanged, and finally decreases. With the decrease of wall temperature, the fracture pressure decreases. The model was verified with field data, and the error was 2.9%, and the accuracy met the engineering requirements. The research results can provide a theoretical reference for on-site borehole stability prediction and fracturing parameter design.

Keywords Vertical well · Thermal stress · Natural fracture · Fracture pressure

S. L. Xia · J. Li (✉) · H. L. Huang
China University of Petroleum-Beijing, Changping, Beijng 102200, China
e-mail: lijun446@vip.163.com

J. Li
China University of Petroleum-Beijing at Karamay, Karamay 834000, China

1   Introduction

In drilling activities, formation fracture pressure is one of the important reference data for well casing structure design, well control design and drilling fluid design. It represents the minimum mud column pressure in the well that causes fracture of open hole formation or re-opening of original fractures.

At present, the formation fracture pressure prediction models widely used in the field include Eaton method, Stephen method, Anderson method, Huang Rongzun method, etc. These models mainly consider the influence of overburden pressure, tectonic stress and pore pressure on fracture pressure, but do not consider the influence of temperature on fracture pressure [1–7]. The study by Maury [8] et al. showed that: in deep wells, it is common for the temperature of the well wall to vary between 25 and 50 °C, which can cause a thermal stress of 25 to 50 MPa on the well wall. Such a large thermal stress will have a great impact on the wellbore fracture pressure. To predict the fracture pressure of high-temperature Wells, Li [9] et al. established a new model for the calculation of fracture pressure of coupled temperature by studying the change of formation temperature caused by drilling fluid circulation and formation heat exchange and the resulting thermal stress. Deng [10] et al. proposed the calculation method of additional fracture pressure caused by temperature and natural fracture, and established a new model for calculation of fracture pressure of high-temperature and high-pressure formations with comprehensive consideration of the influence of temperature and natural fracture. However, with the increasing efforts of petroleum exploration and development in China, many new complex well conditions have emerged [11–13], and the effect of temperature on fracture pressure alone cannot meet the engineering requirements of field drilling. For example, when drilling in high-temperature formation with natural fractures, the wellbore path intersects the fracture surface, and the fracture pressure is affected not only by temperature, but also by natural fractures to a greater extent.

Aiming at the fracture pressure and prediction of high temperature formation with natural fracture development, based on the full consideration of wellbore thermal stress and the analysis of fracture stress state of wellbore fracture surface and fracture criterion under different fracture forms of wellbore, a straight well fracture pressure prediction model for high temperature formation with natural fracture development is established in this paper.

2   Thermal Stress Model of Borehole Wall

In general, the straight well ophthalmology model can be simplified as a two-dimensional plane strain model. In polar coordinates when the ground stress, pore pressure and drilling fluid column pressure are considered, the stress on the borehole wall can be expressed as [14]:

$$\sigma_{r1} = p - \delta\phi(p - p_p) \tag{1}$$

$$\sigma_{\theta 1} = (1 - 2\cos 2\theta)\sigma_H + (1 + 2\cos 2\theta)\sigma_h - p + A(p - p_p) \tag{2}$$

$$\sigma_{z1} = \sigma_v - 2v(\sigma_H - \sigma_h)\cos 2\theta + A(p - p_p) \tag{3}$$

Where: σ_{r1}, $\sigma_{\theta 1}$ and σ_{z1} are respectively radial, circumferential and axial stress, MPa; p is wellbore pressure, MPa; δ is the permeability coefficient of the shaft wall, when permeability occurs, $\delta = 1$; when permeability does not occur, $\delta = 0$; ϕ is formation porosity; P_p is the original formation pore pressure, MPa; v is Poisson's ratio of borehole wall rock, dimensionless; $A = \delta\left[\frac{\alpha(1-2v)}{1-v} - \phi\right]$, α is the effective stress coefficient.

In drilling operations, the circulation of drilling fluid usually causes the lower formation sidewall temperature to decrease and the upper formation sidewall temperature to increase. The surrounding rock will produce thermal strain due to temperature change, which will produce a heating stress field around the borehole. For homogeneous isotropic and porous thermos elastic formation, the thermal stress field around the well can be expressed by the heat conduction theory and thermos elastic mechanics as follows [15]:

$$\sigma_{rT} = \frac{E\lambda}{3(1 - v)r^2}\int_{r_w}^{r} T^f(r,t)rdr \tag{4}$$

$$\sigma_{\theta T} = -\frac{E\lambda}{3(1 - v)}[\frac{1}{r^2}\int_{r_w}^{r} T^f(r,t)rdr - T^f(r,t)] \tag{5}$$

$$\sigma_{zT} = \frac{E\lambda}{3(1 - v)}T^f(r,t) \tag{6}$$

Where, σ_{r1}, $\sigma_{\theta 1}$ and σ_{z1} are respectively radial, circumferential and axial thermal stress, MPa; E is the formation elastic modulus, MPa; λ is the thermal expansion coefficient of rock, 1/°C; $T^f(r,t) = T(r,t) - T_0$, $T(r,t)$ is the distribution function of formation temperature around the well, and at the borehole wall is T_w, °C; T_0 is the original formation temperature, °C; r is the radial distance, m.

When $r = R$, the thermal stress expression at the borehole wall can be obtained:

$$\sigma_{rT} = 0 \tag{7}$$

$$\sigma_{\theta T} = \sigma_{zT} = \frac{E\lambda}{3(1 - v)}(T_w - T_0) \tag{8}$$

It can be seen that the change of borehole wall temperature only produces thermal stress in circumferential and vertical directions, and the thermal stress is positively correlated with the change of borehole wall temperature. According to the superposition criterion of small deformation stress of pore elasticity, the expression of wellbore stress distribution at high temperature can be obtained by the comprehensive Eq. (1)–(8) as follows:

$$\sigma_r = p - \delta\phi(p - p_p) \tag{9}$$

$$\sigma_\theta = (1 - 2\cos 2\theta)\sigma_H + (1 + 2\cos 2\theta)\sigma_h - p + A(p - p_p) + \frac{E\lambda}{3(1 - v)}(T_w - T_0) \tag{10}$$

$$\sigma_z = \sigma_v - 2v(\sigma_H - \sigma_h)\cos 2\theta + A(p - p_p) + \frac{E\lambda}{3(1 - v)}(T_w - T_0) \tag{11}$$

3   Calculation Model of Fracture Pressure Considering the Influence of Temperature and Natural Fracture Property

When the natural fractures are not completely closed or the conductivity is strong, the pressure of the liquid column in the wellbore is slightly greater than the pore pressure of the formation, which will cause leakage. Such fractures are beyond the scope of this article. This paper mainly studies the natural cracks with completely closed crack surfaces and certain cementation strength. When drilling against such natural fractures, excessively high wellbore liquid column pressure will lead to three forms of wellbore failure [16]: tensile failure of wellbore rock, shear failure along natural fractures, and tensile failure along natural fractures.

The fracture pressure prediction models for the three failure forms are respectively established as follows:

3.1   Tensile Failure of Borehole Wall Rock

It can be seen from Eq. (10) that when the wellbore pressure increases gradually, it will gradually decrease. When the wellbore pressure is large enough, σ_θ will decrease to a negative value. At this time, the wellbore rock is subjected to tensile action. When the tensile stress on the rock on the wellbore wall reaches its tensile strength S_t, the rock will undergo tensile failure, that is, the rock failure condition is as follows [17]:

$$\sigma_\theta = -S_t \tag{12}$$

Normally, vertical borehole wall fracture occurs at the point of minimum σ_θ. According to Eq. (10), when $\theta = 0°$ or $\theta = 0°$, σ_θ take the minimum value [18], that is:

$$\sigma_\theta = 3\sigma_h - \sigma_H - p + A(p - p_p) + \frac{E\lambda}{3(1 - v)}(T_w - T_0) \tag{13}$$

In actual working conditions, there is pore pressure in the formation, and the above formula is rewritten as:

$$\sigma_\theta = 3\sigma_h - \sigma_H - p + A(p - p_p) + \frac{E\lambda}{3(1 - v)}(T_w - T_0) - \alpha p_p \tag{14}$$

Combine Eq. (14) and Eq. (12), the fracture pressure expression of rock tensile failure when thermal stress is considered is obtained:

$$p_{m1} = \frac{3(1 - v)(3\sigma_h - \sigma_H - (A + \alpha)p_p + S_t) + E\lambda(T_w - T_0)}{3(1 - A)(1 - v)} \tag{15}$$

3.2 Shear Failure along Natural Fractures

The shear failure of the borehole wall usually occurs in the natural fracture surface and other places where the cement strength is weak. The weak surface failure criterion can be used to study the fracture pressure. The weak surface failure criterion is [19]:

$$\sigma_1 - \sigma_3 = \frac{2(C_w + \mu_w \sigma_3)}{(1 - \mu_w \cot \beta) \sin 2\beta} \tag{16}$$

Where, σ_1 and σ_2 are the maximum and minimum principal stress, respectively, MPa; C_w is the adhesion force of weak surface, MPa; μ_w is the tangent value of the internal friction angle on the weak surface, $\mu_w = \tan\varphi_w$; β is the angle between the normal direction of the weak plane and the maximum principal stress.

It can be seen from the weak surface failure criterion that when β is close to or 90°, $\sigma_1 - \sigma_2$ approaches infinity. Therefore, only when β is greater than φ_w and less than 90°, the well wall may be sheared along the fracture. In general, when natural fractures in vertical wells undergo shear failure, the rock stress of the borehole wall satisfies the relationship [20]: $\sigma_z > \sigma_r > \sigma_\theta$, that is $\sigma_1 = \sigma_z$, $\sigma_3 = \sigma_\theta$. At this time, the fracture pressure expression is:

$$p_{m2} = \frac{3(1-v)(b-a-e \cdot C_w) - e \cdot \mu_w(3(1-v)(a-A \cdot p_p) + E\lambda(T_w - T_0))}{3(1-v)(e \cdot \mu_w(A-1)-1)}$$

$$(17)$$

Where,

$$a = (1 - 2\cos 2\theta)\sigma_H + (1 + 2\cos 2\theta)\sigma_h$$

$$b = \sigma_v - 2v(\sigma_H - \sigma_h)\cos 2\theta$$

$$e = \frac{2}{(1 - \mu_w \cot \beta)}\sin 2\beta$$

3.3 Tensile Failure along Natural Fractures

Assuming that the strike of the natural fracture in the reservoir is TR (in the geodetic coordinate system), the dip angle is θ_{DIP}. The angle between the natural fracture and the wellbore wall is θ_1, and the horizontal maximum ground stress azimuth is θ_H. The normal stress on the surface of the natural crack is expressed as [21]:

$$\sigma_n = \sigma_z l_1^2 + \sigma_\theta l_2^2 + \sigma_r l_3^2 \tag{18}$$

where,

$$l_1 = \sin \theta_{DIP}$$

$$l_2 = \cos \theta_1 \cos \theta_{DIP} \sin(TR - \theta_H) - \sin \theta_1 \cos \theta_{DIP} \cos(TR - \theta_H)$$

$$l_3 = \cos \theta_1 \cos \theta_{DIP} \cos(TR - \theta_H) + \sin \theta_1 \cos \theta_{DIP} \sin(TR - \theta_H)$$

When the wellbore pressure is equal to the sum of the normal stress and tensile strength of the fracture plane, tensile failure of natural fractures will occur. Fracture criteria are as follows:

$$p_{m3} = \sigma_n + \sigma_t - \alpha p_p \tag{19}$$

Substituting Eq. (9), (10), (11), (18) into Eq. (19) to obtain the calculation formula of rupture pressure:

$$p_{m3} = \frac{f \cdot l_1^2 + g \cdot l_2^2 + \delta\phi p_p \cdot l_3^2 - \alpha p_p + \sigma_t}{h} \tag{20}$$

Where,

$$f = \sigma_v - 2v(\sigma_H - \sigma_h)\cos 2\theta - A \cdot p_p + \frac{E\lambda}{3(1-v)}(T_w - T_0)$$

$$g = (1 - 2\cos 2\theta)\sigma_H + (1 + 2\cos 2\theta)\sigma_h - A \cdot p_p + \frac{E\lambda}{3(1-v)}(T_w - T_0)$$

$$h = 1 - A \cdot l_1^2 - (A - 1)l_2^2 - (1 - \delta\phi)l_3^2$$

3.4 Calculation Model of Fracture Pressure

In actual working conditions, for different formations, borehole wall rupture can only be one of the three types. Therefore, the minimum value of the fracture pressure under the three fracture types is the fracture pressure under the actual working conditions. The comprehensive calculation model is the Eq. (21)

$$p_f = \min\{p_{m1}, p_{m2}, p_{m3}\} \tag{21}$$

Where,

$$p_{m1} = \frac{3(1-v)(3\sigma_h - \sigma_H - (A+\alpha)p_p + S_t) + E\lambda(T_w - T_0)}{3(1-A)(1-v)}$$

$$p_{m2} = \frac{3(1-v)(b - a - e \cdot C_w) - e \cdot \mu_w(3(1-v)(a - A \cdot p_p) + E\lambda(T_w - T_0))}{3(1-v)(e \cdot \mu_w(A-1) - 1)}$$

$$p_{m3} = \frac{f \cdot l_1^2 + g \cdot l_2^2 + \lambda\phi p_p \cdot l_3^2 - \alpha p_p + \sigma_t}{h}$$

4 Analysis of Factors Affecting Fracture Pressure

4.1 The Influence of Natural Fracture Inclination on Fracture Pressure

Taking the fracture inclination angle from 5° to 90°, the trend of fracture pressure with the fracture inclination angle is obtained (Fig. 1). With the increase of the crack inclination angle, p_{m1}(the fracture pressure of the borehole wall rock tensile failure) remains unchanged; p_{m2}(the fracture pressure of shear failure along natural fractures) decreases rapidly, but the rate of decrease gradually decreases; p_{m3} (the rupture pressure along the tensile failure of natural fractures) gradually increases, and the increase rate is small. When the crack inclination is equal to 90°, p_{m1} and p_{m3} are similar in size. This is due to the small angle between the fracture strike and the horizontal maximum principal stress selected in the calculation. When the inclination angle of the fracture surface is 90° (the fracture surface is parallel to the axis of the wellbore), p_{m1} and p_{m3} mainly depend on the circumferential stress of the well wall.

4.2 The Influence of Fracture Strike on Fracture Pressure

Taking the fracture strike from 79° to 169°, the change trend of the fracture pressure with θ_2 (the angle between the fracture trend and the maximum principal stress) is obtained (Fig. 2). As θ_2 increasing, p_{m1} remains unchanged; p_{m2} increases, and the rate of increase first increases and then decreases; p_{m3} first increases and then decreases, and the maximum value is taken at a certain angle, and the maximum value is greater than p_{m1}. Combining Eq. (21), it can be seen that the change of θ_2 will cause the change of the failure form of the borehole wall. In the calculation

Fig. 1 Trend diagram of fracture pressure changing with fracture dip Angle
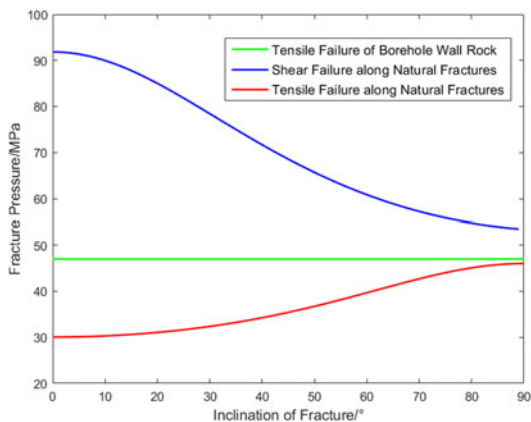
Fig. 2 Trend chart of burst pressure changing with θ_2
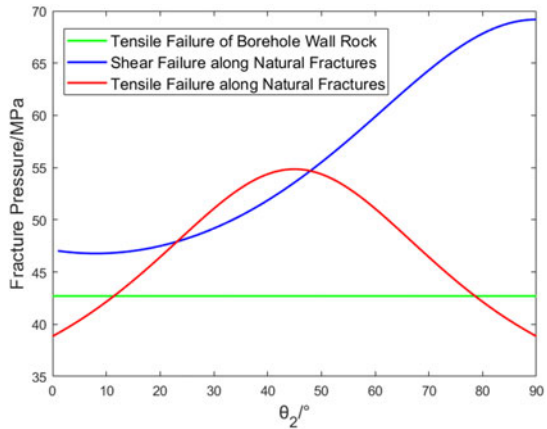
example in this paper, as θ_2 increases, the borehole wall fracture changes from the fracture surface tensile failure form to the rock tensile failure form, and after θ_2 increases to a certain extent, it becomes the fracture surface tensile failure form.

4.3 Effect of Temperature on Fracture Pressure

Taking the temperature change of the borehole wall form −50 °C to 0 °C, the trend of the change of the fracture pressure with the temperature of the well wall can be obtained (Fig. 3). As the borehole wall temperature decreases, p_{m1}, p_{m1} and p_{m3} all decrease. The fracture pressure changes linearly with the temperature of the well wall. The decrease of the well wall temperature will cause the fracture pressure to decrease, but the failure form of the borehole wall will not change.

5 Example Calculation

In a fractured formation, the natural fracture inclination is 53° and the strike is 12°. The azimuth of the maximum principal stress of the formation is 68° and the magnitude is 181.4 MPa. The minimum ground stress is 145.3 MPa, and the overburden pressure is 179.8 MPa, the pore pressure is 120.1 MPa, the rock tensile strength is 9.1 MPa. Poisson's ratio is 0.254, effective stress coefficient is 0.91, porosity is 0.132, fractured rock cohesion is 8.95, weak surface friction coefficient is 0.491, formation temperature is 150 °C. The fracture pressure measured on site is 101.4 MPa.

Use the well data to calculate the fracture pressure. The calculation results are shown in Table 1. When ignoring the influence of temperature and natural fractures, the calculated fracture pressure is 188.4 MPa, which is quite different from the
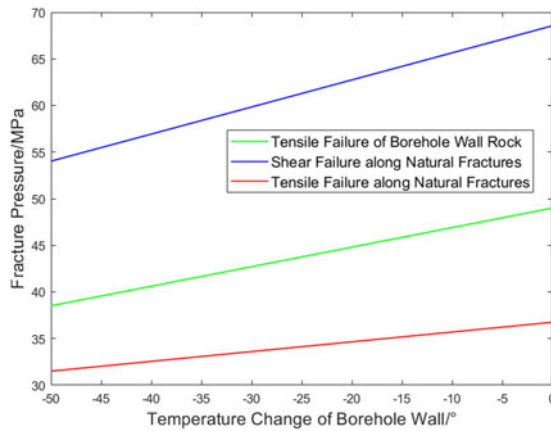
Fig. 3 The relationship between the fracture pressure and the temperature of the borehole wall

Table 1 Calculation error table of fracture pressure

Calculation conditions	Calculated value /MPa	Measured value /MPa	error
Ignore temperature and natural fracture effects	188.4	101.4	85.8%
Consider only natural fracture	123.5		21.8%
Consider natural fracture and temperature effects (models in this article)	104.3		2.9%

actual measured value, with an error of 85.8%; When only considering the influence of natural fractures, the calculated fracture pressure is 123.5 MPa, which is less different from the actual measured value, but the error is still large; When considering the effects of natural fractures and temperature on the fracture pressure at the same time, the calculated fracture pressure value is 104.3 MPa, which is close to the measured fracture pressure value of 101.4, and the error is only 2.9%. It can be seen that the error calculated by the model established in this paper is the smallest, and it is suitable for the prediction of fracture pressure in high-temperature formations with natural fractures.

6 Conclusion

In this paper, a prediction model of rupture pressure considering temperature and natural fractures is established, and a series of analyses are carried out, and the following conclusions are obtained:

(1) The thermal stress distribution around the borehole caused by the circulation of drilling fluid has nothing to do with the well circumference angle. The vertical and circumferential thermal stresses at the well wall are equal. The radial thermal stress is zero.

(2) The factors affecting the fracture pressure of high-temperature formations with natural fractures are analyzed, and the analysis results show that: as the fracture inclination angle increases, the fracture pressure gradually increases; as the angle between the fracture direction and the horizontal maximum principal stress increases, the fracture pressure first increases, then remains unchanged, and finally decreases; as the borehole wall temperature decreases, the fracture pressure decreases.

(3) Using the model in this paper to calculate the example well, the error of the calculation result is 2.9%, which meets the field requirements. Therefore, the fracture pressure prediction model established in this paper can provide a theoretical reference for on-site borehole stability prediction and fracturing parameter design.

References

1. Feng, Q.H., Cheng, Y.F., Zhang, J.G.: Elastic & plastic model for borehole stability and its application. Pet. Drilling Tech. **04**, 9–11 (2000)
2. Zhao, J.Z., Ren, L., Hu, Y.Q., Wang, L.: A calculation model of breakdown pressure for perforated wells in fractured formations. Acta Petrolei Sin. **33**(05), 841–845 (2012)
3. Li, M., Lian, Z.H., Chen, S.C., Wei, C.X., Lin, T.J., Li, X.J.: Rock mechanical parametric experiments and the research of formation fracture pressure prediction. Oil Drilling Prod. Technol. **31**(05), 15–18 (2009)
4. Zhou, N.Y., Yang, Z.Z.: Overview on pressure prediction of formation fracture. J. Chongqing Univ. Sci. Technol. (Nat. Sci. Ed.) **13**(01), 36–39 (2011)
5. Nie, C.J., Zhao, J., Xia, H.Q.: Study on statistical model of predicting formation fracture pressure using logging data. Nat. Gas Geosci. **06**, 633–636 (2004)
6. Guo, Y.B.: Predicting formation fracture pressure on high temperature high pressure wells. China University of Petroleum (2010)
7. Wang, L., Zhang, Y.C., Yan, Z.M.: Application of formation pressure prediction and analysis method in an oil field in Qinhuangdao. Technol. Innov. Appl. **02**, 161–162 (2018)
8. Maury, V.: Practical advantages of mud cooling systems for drilling. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts, **33**(2): A82 (1996)
9. Li, S.G., Deng, J.G., Wei, B.H., Yu, L.J.: Formation fracture pressure calculation in high temperatures wells. Chin. J. Rock Mechan. Eng. **S2**, 5669–5673 (2005)
10. Deng, J.G., Liu, Y., Wei, B.H.: Formation fracture pressure prediction method in high temperature and high pressure formations. Pet. Drilling Tech. **37**(05), 43–46 (2009)
11. Wang, H.G., Liu, Y.S., Yang, L.P.: Effect of temperature and pressure on drilling fluid density in HTHP wells. Drill. Prod. Technol. **01**, 58–62 (2000)

12. Lu, Y.H., Xiao, X.H., Zhao, L., Jin, Y., Chen, M.: The effect of temperature on stability of borehole wall in ultra-deep fractured formation. Drilling Fluid Completion Fluid **37**(02), 160–167 (2020)
13. Wang, X., Chen, Z.H., Shang, W.T.: The leakage law of drilling fluid in fractured formations. Chem. Enterp. Manage. **19**, 89–90 (2020)
14. Zhang, W.D., Chang, L., Gao, J.J.: Stress analysis at the borehole wall in transverse isotropic formations. Eng. Mech. **32**(11), 243–250 (2015)
15. Tang, L., Luo, P.: The effect of the thermal stress on wellbore stability. Soc. Pet. Eng. **3**(1) (1993)
16. Jin, Y., Chen, M., Zhang, X.D.: Hydraulic fracturing initiation pressure models for directional wells in naturally fractured formation. Acta Petrolei Sinica **05**, 124–126 (2006)
17. Huang, R.Z.: Discussion on the prediction model of formation fracture pressure. J. China Univ. Pet. (Ed. Nat. Sci.) **04**, 335–347 (1984)
18. Huang, R.Z., Zhang, J.J.: A model for prediction formation fracture pressure. Oil Drilling Prod. Technol. **03**, 1–14 (1986)
19. Li, K. Ch.: Study on fracture mechanism and wellbore initiation pressure of ultra-deep fractured sandstone in Tarim Oilfield. China University of Petroleum (2018)
20. Jin, Y., Chen, M., Zhang, X. D.: Initiation pressure models for hydraulic fracturing of vertical wells in naturally fractured formation. Acta Petrolei Sinica (06), 113–114+118 (2005)
21. Song, B.: Study on the mechanism of fracture and extension of hydraulic fracture in fractured reservoir. Yangtze University (2018)

Study on Enhanced Oil Recovery by CO_2 Huff and Puff of the Tight Oil Reservoir in the Lucaogou Formation of Jimsar Sag

Furong Yang and Yongsheng An

Abstract Aiming at the problems such as high viscosity of tight oil, poor pressure conductivity, large seepage resistance and difficulty in starting crude oil in the Lucaogou Formation of Jimsar Sag, CO_2 injection for enhanced recovery was studied. The CO_2 huff and puff component model of fractured horizontal well in tight reservoir is established, and the numerical simulation of CO_2 huff and puff is carried out by using ECLIPSE software. The enhanced oil recovery value and oil change rate are selected as evaluation indexes to analyze the sensitivity of relevant parameters. The results show that there is an optimal bottom-hole flow pressure, soaking time, gas injection rate, gas injection time and production time. The optimal scheme is selected by orthogonal test. Compared with depleted production, the EOR of CO_2 huff and puff method can reach 6.26%, which has a significant increase in production.

Keywords Tight oil reservoir · CO_2 huff and puff · Numerical simulation · Sensitivity analysis · Enhanced oil recovery · Orthogonal test

1 Introduction

CO_2 huff and puff is a technical means to improve the recovery of tight oil reservoir by injecting liquid CO_2 into the formation and then shutting in and soaking for a period of time before opening the well for production, so that the CO_2 gas in the formation mixed with the crude oil can be produced at the wellhead. The mechanism mainly includes reducing crude oil viscosity, crude oil volume expansion, wettability and relative permeability reversal, extraction of light components, gas flow carrying effect, increasing formation pressure, acidizing and plugging removal of rocks, and gas resistance effect [1]. In general, the above mechanisms work at the

F. Yang · Y. An (✉)
China University of Petroleum (Beijing),
18 Fuxue Road, Changping District, Beijing Municipality 102249, China
e-mail: anyongsheng@cup.edu.cn

same time during CO_2 huff and puff production [2], but the specific role of specific mechanism and which mechanism or mechanisms are dominant are different under different formation, fluid characteristics and injection production characteristics, which depends on the overall characteristics of reservoir production system [3].

In terms of CO_2 huff and puff research, scholars at home and abroad have conducted a lot of research: Zhong et al. [4] of Daqing Petroleum University studied the volume changes of CO_2 and crude oil in contact with each other under constant pressure and variable pressure through the gas–liquid equilibrium experiment of sand filling. It showed that CO_2 had an extraction effect on crude oil, and the higher the injection pressure, the more obvious the extraction effect. Wang [5] of Shengli Oilfield determined the applicable scope of CO_2 huff and puff through indoor research on the stimulation mechanism of it. Li [6] conducted in-depth research on the enhanced recovery of CO_2 huff and puff from horizontal wells in complex fault-block reservoirs from the perspectives of minimum miscible pressure and numerical model calculations. He believed that the main mechanisms of CO_2 enhanced recovery were light hydrocarbon extraction, crude oil expansion and crude oil viscosity reduction. Yang et al. [7] conducted an in-depth study on CO_2 huff and puff for enhancing oil recovery of staged fractured horizontal wells in tight reservoirs by means of indoor simulation experiments. He believed that indoor physical experiments can simulate the staged fracturing law of tight reservoirs well. Through indoor experimental simulation studies, it is found that CO_2 huff and puff can effectively improve the production degree and oil recovery of tight reservoirs. Yi et al. [8] optimized injection parameters and successfully increased the on-site construction rate to 82% by combining field experiment with adaptability study on minimum miscible pressure of CO_2 in Shengli Oilfield. Wu et al. [9] studied the influence of CO_2 injection rate, bottom-hole flow pressure and huff and puff cycle on the development effect, and found that the volume of formation crude oil could be expanded by 41.2% after CO_2 dissolved in the formation, which proves CO_2 huff and puff is an economic and effective method for the production of low-permeability reservoirs.

The Lucaogou Formation in Jimsar Sag has a large area of oil whose layers are distributed in thin interbeds vertically. The pore types are various, the pore throat structure is complex, and the micro-nano pore throat system is mainly developed. The reservoir space is dominated by intergranular and intragranular dissolved pores with fine channels and undeveloped fractures. The pore radius is distributed between 100–150 μm, the throat radius is mainly below 0.1 μm, the average permeability is 0.01 mD, and the average porosity is 11%. The upper "dessert body" crude oil has a density of 0.882 g/cm^3 and a 50 °C viscosity of 50.27 mPa·s; the lower "dessert body" crude oil has a density of 0.903 g/cm^3 and a 50 °C viscosity of 123.23 mPa·s. The purpose of this study is to analyze the effect of different injection and production parameters on oil production, and to provide theoretical basis and guidance for the optimization scheme of CO_2 huff and puff production in tight reservoirs.

2 Model Establishment

Based on the actual situation, the ECLIPSE software is used to establish a single well numerical simulation model of fractured horizontal well in tight reservoir. The component model with corner grid is adopted, and the grid number in X, Y, Z direction is 33, 11 and 7 respectively. The well is about 1,400 m long and has 15 artificial fractures, with a length of 300 m and a gap of 100 m. The combination of local mesh refinement and equivalent fracture is used to treat artificial fracture.

3 Production History Matching

The reason why the well only produces water and no oil at the initial stage is that the fracturing fluid is not completely reversed. Therefore, advanced water injection is used to simulate the large amount of fracturing fluid remaining in the formation. The results show that the production history is well fitted by using the method of constant liquid production to fit the bottom-hole flow pressure and other parameters, which can meet the requirements of numerical simulation of CO_2 huff and puff development indicators. The history matching results are shown in the Fig. 1.

4 Sensitive Factor Analysis

4.1 Bottom-Hole Flow Pressure

With other parameters unchanged, the production effect under different bottom-hole flow pressures is predicted as shown in Fig. 2. Lower bottom-hole flow pressures can improve the production speed and daily oil production, thus reducing the time for CO_2 separation from formation crude oil, which is conducive to the oil-increasing effect of CO_2 huff and puff. However, the lower the bottom-hole flow pressure is not the better. CO_2 dissolved in the crude oil is more likely to overflow, which reduces the mobility of the crude oil. At the same time, if the bottom-hole flow pressure is too low, the oil will not be produced out of the wellhead.

4.2 Daily Gas-Injection Rate

With other parameters unchanged, the production effect under different daily gas-injection rate is predicted as shown in Fig. 3. With the increase of daily gas-injection rate, the EOR value increases rapidly; but when the daily gas-injection rate reaches a certain value, the EOR trend slows down, presumably because the
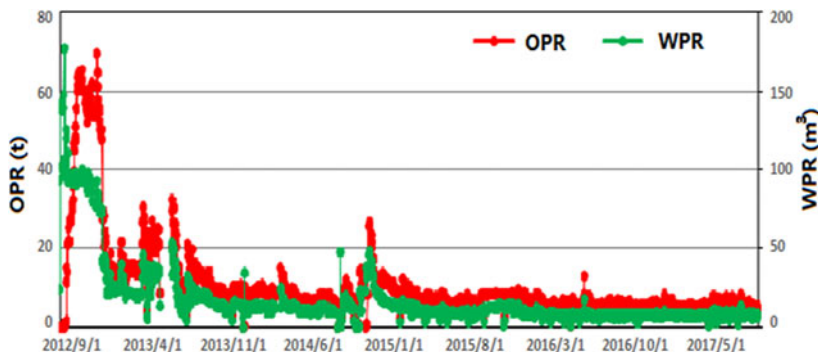
Fig. 1 Matching results of production history
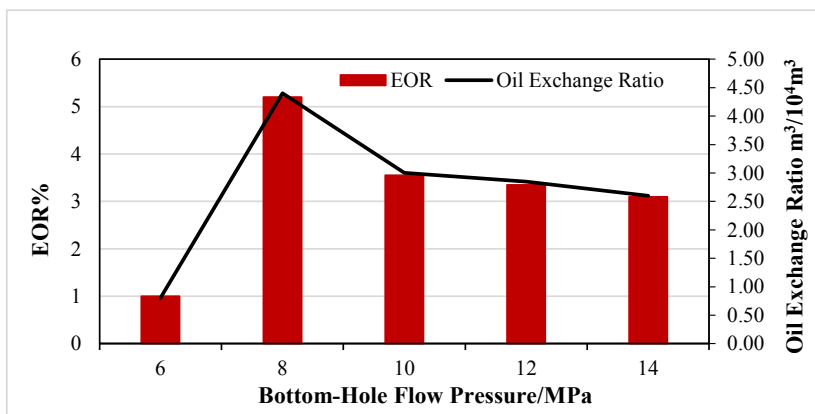
Fig. 2 Evaluation index of different bottom-hole flow pressure

solubility of CO_2 in crude oil has reached the limit value. Therefore, there is an optimal daily gas-injection rate. Below this value, the oil production and oil change rate will decrease. If it is higher than this value, the oil production will increase slowly, the oil change rate will be greatly reduced, and the economic benefit will be low.

4.3 Gas Injection Time

With other parameters unchanged, the production effect under different gas injection time is predicted as shown in Fig. 4. There is an optimal gas injection time. Too low gas injection time is not conducive to the sweep of CO_2 and cannot

maintain a high average formation pressure. However, the high gas injection time will make it difficult for the gas swept to the far distance to drive the crude oil out during the "puff" stage.

4.4 Soaking Time

With other parameters unchanged, the production effect under different soaking time is predicted as shown in Fig. 5. There is an optimal soaking time. The analysis is that a longer soaking time provides more opportunities for CO_2 and crude oil to dissolve, and the formation pressure is relatively high. With the increase of soaking time, the dissolved gas drive effect of CO_2 is more obvious, and when the soaking time is too long, the convection and diffusion of CO_2 dissolved in crude oil will occur, which weakens the driving effect on crude oil.

4.5 Production time

The predicted production effect under different production time is shown in Fig. 6. There is an optimal production time. Under the same huff and puff cycles, the recovery increases with the increase of single cycle production time. However, too long production time will easily lead to lower formation pressure, which is not conducive to maintaining good huff and puff effect.
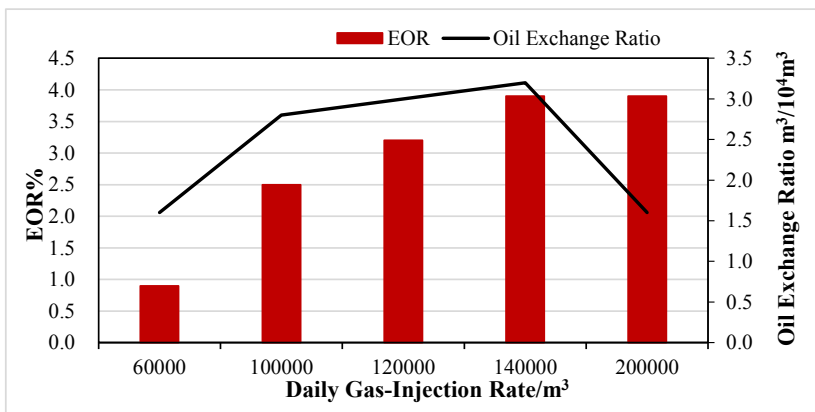
Fig. 3 Evaluation index of different daily gas-injection rate
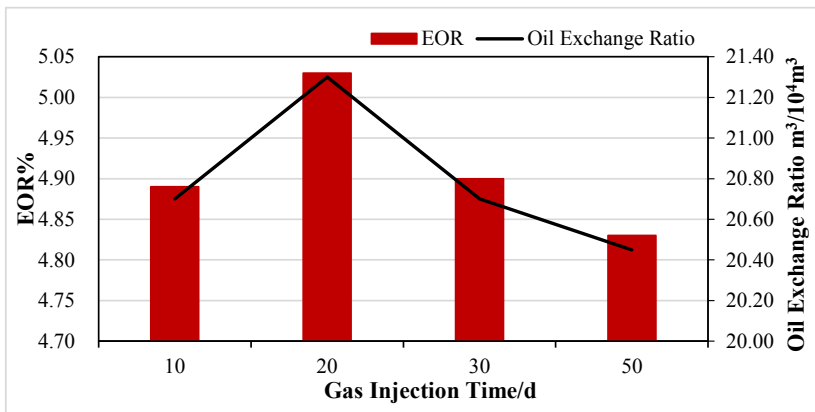
Fig. 4 Evaluation index of different gas injection time
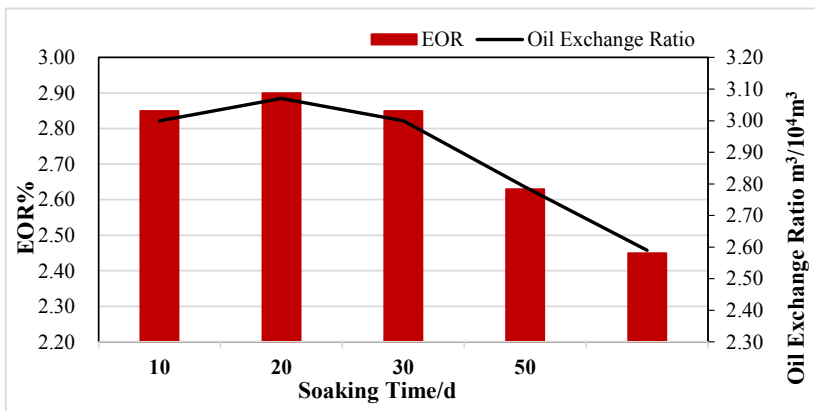
Fig. 5 Evaluation index of different soaking time
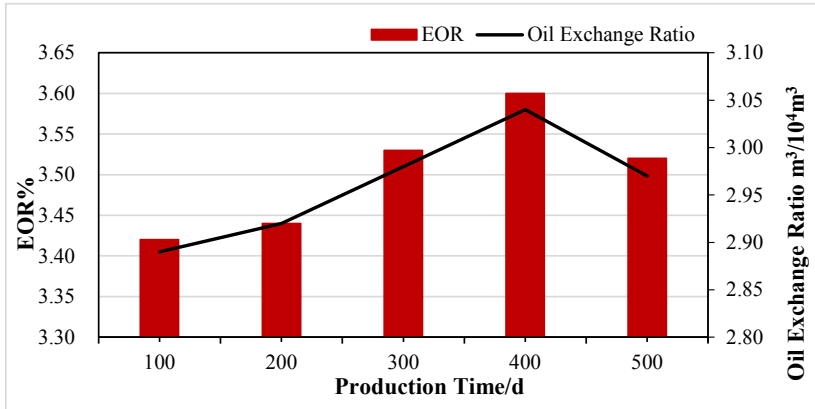
Fig. 6 Evaluation index of different production time

5 Scheme Optimization

Carry out orthogonal test to optimize scheme. This experiment is based on 4 factors and 3 levels, with a constant bottom-hole flow pressure of 8 MPa for production. With reference to the principle of orthogonal table selection [10], orthogonal table L9 (3^4) is selected for orthogonal experiment design. A total of 9 tests are conducted. The test results are shown in Table 1.

The results of orthogonal test are analyzed by range analysis and comprehensive balance method. The optimal scheme is 140000 cubic meters of gas injection per day, 20 days of gas injection, 5 days of soaking time, 400 days of production time and 8 MPa of bottom-hole flow pressure (Fig. 7).

The enhanced oil recovery and oil exchange rate of the optimal scheme are both higher than those of the 9 trials. Compared with the depletion production under the same production time, the cumulative net oil increase of six rounds of CO_2 huff and puff is 28002.87 m^3, and the enhanced oil recovery value is 6.26%, which has a significant oil increase effect.

Table 1 L9 (3^4) orthogonal design and test results

Test Number	Gas Injection Rate/ $10^4 m^3$	Gas Injection Time/d	Soaking Time/d	Production Time/d	EOR /%	Oil Exchange Ratio /m^3/ $10^4 m^3$
1	13	15	3	350	5.2620538	4.2699488
2	13	20	5	400	5.2620536	4.2699479
3	13	25	8	450	5.2611485	4.2668858
4	14	15	3	450	5.5518898	4.3754670
5	14	20	5	300	5.4533814	4.0977281
6	14	25	8	350	5.5011286	4.2323486
7	15	15	3	400	5.7687115	4.2743860
8	15	20	5	450	5.7766303	4.2935232
9	15	25	8	350	5.6646426	4.0228863
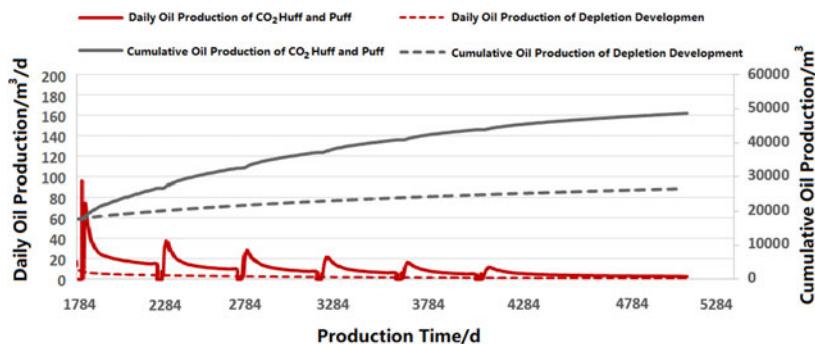
Fig. 7 Comparison of oil production curves between CO_2 huff and puff and depletion production

6 Conclusion

(1) For CO_2 huff and puff in tight reservoirs, there are optimal injection and production parameters, which lead to the highest EOR and oil exchange rates.
(2) The optimal scheme can be selected by orthogonal test. The optimal scheme of CO_2 huff and puff in the Lucaogou Formation of Jimsar Sag is 140000 cubic meters of gas injection per day, 20 days of gas injection, 5 days of soaking time, 400 days of production time and 8 MPa of bottom-hole flow pressure. Compared with depletion production, the oil-increasing effect of the optimal scheme is obvious.

References

1. Yu, W., Sepehrnoori, K.: Optimization of well spacing for Bakken tight oil reservoirs. In: Unconventional Resources Technology Conference (2014)
2. Zhang, Y., Yu, W., Li, Z., et al.: Simulation study of factors affecting CO_2 huff- n -puff process in tight oil reservoirs. J. Pet. Sci. Eng. (2018)
3. Zhan, M., Yang, S., Lu, W., et al.: Investigation on numerical simulation of CO_2 huff and puff in tight oil reservoirs. J. Petrochem. Univ. **29**(6), 39–42 (2016)
4. Zhong, L., Han, D., Li, L., et al.: Simulation of carbon dioxide throughput in ultra-low permeability reservoirs. J. Daqing Pet. Inst. **33**(4), 120–124 (2009)
5. Wang, S., Sun, B., Wang, L., et al.: Indoor research and application of CO_2 stimulation yield increase mechanism. Drill. Prod. Technol. **27**(1), 91–94 (2004)
6. Li, G., Ye, S., Feng, J., et al.: Carbon dioxide stimulation water control and oil enhancement technology for horizontal wells in complex fault block reservoirs and its application. Pet. Geol. Recov. Effi. **19**(4), 62–65 (2012)
7. Yang, Z., Liu, X., Zhang, Z., et al.: Physical simulation of Carbon dioxide throughput in horizontal wells of tight reservoirs. Editorial Office of Acta Petrolei Sinica **36**(6), 724–729 (2015)
8. Yin, Z., Tian, Z., Shan, J., et al.: Research and application of carbon dioxide huff and puff in complex fault block reservoirs. Henan Pet. **17**(4), 26–28 (2003)
9. Wu, W., Zhang, L.: Feasibility study on improving the development effect of low permeability oil field by CO_2 huff and puff. Daqing Pet. Geol. Dev. **20**(6), 51–53 (2001)
10. Pu, W., Wei, B., Jin, F., et al.: Experimental investigation of CO_2 huff-n-puff process for enhancing oil recovery in tight reservoirs. Chem. Eng. Res. Des. **111**, 269–276 (2016)

A New Approach of Flange Leakage Check for Piping Design Based on ASME PCC-1 and Equivalent Pressure

Yong Zuo, Yue Song, and Kai Feng

Abstract This paper analyzes the common practices of flange leakage check in piping design, compares the advantages and disadvantages of these methods, and propose a new engineering method considering bolt pre-tightening stress; gasket stress and piping stress, which based on flange installation requirements of ASME PCC-1 and equivalent Pressure. The key steps of this method and the matters need attention in engineering application are introduced by calculation example. The new method is scientific, reasonable, simple and easy to use approved by engineering practice. The method has reference significance for flange leakage check and flange installation in piping engineering.

Keywords Flange leakage check · Equivalent Pressure · Bolt pre-tightening stress · Gasket stress

1 Introduction

Flange, as a kind of connector, is indispensable in petroleum, chemical engineering and municipal water supply and drainage facilities. It is widely used in the connection of pipes, equipments, valves and other components. If flange leak happens during operation, it will not only cause fluid or energy loss, but also affect the normal production. In the petrochemical industry, because of the dangers of the medium, such as flammability, explosion and toxicity, once the flange leaks, it will not only affect the normal production, but also pollute the environment and even cause explosion, causing significant losses [1]. Therefore, in order to prevent flange leakage and ensure the safety and reliability of the flange throughout its life cycle, it is necessary to analyze the principle of flange leakage and sealing, carry out flange leakage check in the engineering design phase, carry out correct pre-tightening

Y. Zuo (✉) · Y. Song · K. Feng
China Petroleum Pipeline Engineering Co., Ltd.,
HePing Road 146#, LangFang City 065000, China
e-mail: zuoyong@cnpc.com.cn

installation in the construction phase, and perform correct maintenance in the operation phase. For standard flanges, the main task in the design phase is to check whether the flange is at risk of leakage and provide guidance on the pre-tightening force of the bolt. This paper will study these two issues and provide a new approach for flange check and provide matching flange bolt pre-tightening force for the installation stage.

2 Flange Leakage Types and Influence Factors

Flange sealing system is composed of flange, gasket, and bolt [2], shown in Fig. 1. Flange sealing is achieved by tightening the bolts to increase the pressure between the flange and gasket to reduce the leakage gap. When the resistance of the medium pass through the sealing is greater than the pressure difference between the inside and outside of the flange sealing, the medium is sealed and the sealing purpose is achieved.

Flange leakage is usually divided into two types: seepage leakage and interface leakage.

2.1 Flange Leakage Types

Seepage Leakage: Asbestos, graphite, polytetrafluoroethylene are commonly used gaskets, but due to its loose structure and poor compactness, there are micro gaps between fibers. The medium will penetrate to the low pressure side through the gaps. Seepage leakage is closely related to sealing materials, pressure of the medium and physical characteristics of the medium. Therefore, seepage leakage can be avoided as long as suitable gasket are selected according to the pressure and medium.

Fig.1 Flange structure diagram
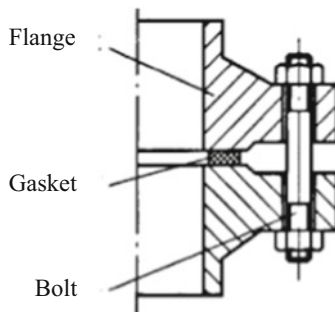
Flange

Gasket

Bolt

Interface Leakage: Insufficient sealing pressure of gasket, improper surface finish of flange, thermal displacement of piping, etc. will cause loose between gasket and flange. In addition, the bolt deformation and elongation, and excessive pressure of gasket after flange installation, plastic deformation, reduction of resilience, material aging, cracking, deterioration of gasket after long-term service, will also cause loose between gasket and flange. This kind of leakage occurs at the interface between flange surface and gasket is called "interface leakage". Interface leakage is the main leakage of flange and is closely related to four main factors, which are gasket, bolt pre-tightening force, operating conditions and flange [3].

2.2 Influence Factors

Gasket: The elastic deformation, plastic deformation and compression resistance of gasket are three key performances for flange sealing. Elastic deformation has rebound ability. It enables the gasket to better compensate the separation of the flange surface and adapt to the fluctuation of operating pressure and temperature. Plastic deformation enables the gasket to better fill the micro unevenness on flange surface to achieve the initial seal between the flanges. Compression resistance ensures that the gasket will not be damaged under a certain pressure and will not lose its sealing performance.

Bolt Pre-tightening Force: Under installation and operation conditions, the gasket shall be compressed by flange with bolt pre-tightening force. Generally, the larger the pre-tightening force, the better, if the gaskets and flanges are not damaged and the bolts are not overloaded.

Operating Conditions: Operating conditions include temperature, pressure, external force (moment) and medium. Among them, temperature has the greatest impact on sealing performance. High temperature will reduce the performance of gasket and bolt, and enhance the permeability and corrosiveness of the medium. Operation pressure will increase leakage rate and make it difficult to compress the gasket on the flange surface. Flange external force (moment) will reduce gasket compression and increase bolt tension stress. The medium is also a very important factor, its viscosity and corrosion shall be considered when selecting gasket.

Flange: Misalignment of flanges during installation, such as centerline high/low, parallelism, and two rotating holes will affect the sealing ability. Flanges with insufficient rigidity of bolted connections will be excessively warped, resulting in uneven pressure on the gaskets and leading to leakage. A good seal can only be achieved when the flange surface has a good finish and matches the selected gasket.

Based on the above factors, for selected flange system (flange, bolt and gasket), gasket stress is the key point for flange leakage check. The bolt pre-tightening force shall meet requirements of gasket initial stress and operating stress, and the gasket,

bolt and flange shall not be damaged under the bolt pre-tightening force and external force (moment) caused by piping connection.

3 Common Method for Flange Leakage Check

At present, frequently used flange leakage checking methods include Kellogg equivalent pressure method, ASME NC-3658.3 method, EN1591 method, Taylor-Waters method, and ASME PCC-1 method. Different factors and principles are considered in the above methods, and the results of Flange Leakage Check are different, which brings great controversy and trouble to the flange leakage check in engineering design phase.

3.1 Kellogg Equivalent Pressure Method

This method considers that the force and moment on the flange are equivalently acting on the gasket, and the equivalent pressure is obtained by summing the operating pressure of the flange, which is compared with the maximum allowable working pressure of the flange.

The calculation formula is as follows:

$$Pe = P + 4F/\pi D_G^2 + 16M/\pi D_G^3 \tag{1}$$

Where Pe is the equivalent pressure, F is the axial force acting on the flange, M is the bending moment acting on the flange, D_G is the gasket effective working diameter, and P is the flange operating pressure [4].

This method is conservative to a certain degree. It is almost impossible to check when the operating pressure is close to the maximum allowable working pressure of the flange. Especially for blind flanges, when superimposed with the operating pressure the equivalent pressure will greatly exceed the maximum allowable working pressure. When this method being used for flange leakage check in piping engineering design, flanges always need to be upgraded, resulting in uneconomical engineering.

3.2 ASME NC-3658.3 Method

ASME NC-3658.3 is the rules for analysis of ASME B16.5 flanged joints with high-strength bolting in nuclear facility components. This method is to apply the force and moment of the flange to the bolt, and then compare it with the yield strength of flange material. The calculation formula is as follows:

$$250M_{fs}/21.7CA_b \leq S_y \tag{2}$$

$$250M_{fd}/43.4CA_b \leq S_y \tag{3}$$

Where A_b is total cross-sectional area of bolts at root of thread; C is bolt circle diameter; M_{fs} is bending or torsional moment applied to the joint due to weight, thermal expansion of the piping, sustained anchor movements, relief valve steady state thrust, and other sustained mechanical loads applied to the flanged joint during the design or service condition; M_{fd} is bending or torsional moment as defined for M fs but including dynamic loadings applied to the flanged joint during the design or service condition; S_y is yield strength of flange material at design temperature.

This method is very radical. Almost all working conditions will meet the requirements. Even if the results of flange leakage check met its requirements, the risk of leakage is very high.

3.3 EN1591 Method [5]

This method considers the characteristics of the entire flange-bolt-gasket system, calculates the integrity and tightness of the flange and its connections under the pressure, temperature, external force (moment) and other loads to ensure the integrity of the structure. It also controls the leakage of the connection joints, allowing the design calculation to be based on any given leakage level, which more comprehensively reflects the mechanical and sealing performance of the flange joint, and the calculation is more in line with operation condition.

However, the calculation process is too complicated, because hundreds of formulas are involved, too much input data is needed, and these data are not all easy to obtain. It may be a good method for flange design and verification in key parts. However, there are hundreds of flanges in the engineering project, massive calculation makes it difficult to apply to the flange leakage check in the engineering design phase [6].

3.4 Taylor-Waters Method

This method is based on the conditions of design pressure, gasket factor, gasket sealing ratio pressure, effective sealing width, etc. By calculating the required compression force of the gasket in the pre-tightened state and the operating state, the required bolt load and area are calculated and the flange strength is checked by the bolt load. In this method, the bolts are checked and pre-tightened according to the allowable stress.

3.5 ASME PCC-1Method [7]

This method is used to determine the bolt installation stress and can also be used for flange leakage check. The bolt pre-tightening force is determined according to the bolt material. Under this pre-tightening force, the pressure on flange gasket is big enough to maintain the seal in the installation and operating conditions, and the gasket is not crushed. This method comprehensively considers the sealing performance of bolts, flanges, and gaskets, which is consistent with the performance of on-site flange sealing [8]. The calculation formula is as follows:

$$b_{sel} = max(Sb_{min}, Sb_{sel}) \tag{4}$$

$$b_{sel} = min(Sb_{max}, Sb_{sel}) \tag{5}$$

$$b_{sel} = min(Sf_{max}, Sb_{sel}) \tag{6}$$

$$Sb_{sel} \geq Sg_{min-S}\left[A_g/(A_b n_b)\right] \tag{7}$$

$$Sb_{sel} \geq \left[Sg_{min-O}A_g + (\pi/4)P_{max}G_{ID}^2\right]/(\Phi g A_b n_b) \tag{8}$$

$$Sb_{sel} \leq Sg_{max}\left[A_g/(A_b n_b)\right] \tag{9}$$

$$Sb_{sel} \leq Sf_{max}(\theta g_{max}/\theta f_{max}) \tag{10}$$

where Sb_{sel} is bolt pre-tightening stress (related to the bolt material), Sb_{min} is the minimum bolt pre-tightening stress (140 MPa–245 MPa), Sb_{max} is the maximum allowable stress of the bolt (related to the bolt material, it is recommended to be 0.7 times the bolt yield strength), Sf_{max} is the maximum allowable stress of the bolt considering that the flange is not damaged, Sg_{min-S} is the minimum sealing pressure of the gasket in the installed state (provided by the supplier, or consult the relevant standard), Sg_{min-O} is the minimum sealing pressure of the gasket under operating conditions (provided by the supplier, or check related standards), Sg_{max} is the maximum allowable pressure of the gasket (provided by the supplier, or check related standards), A_g is the effective area of the gasket, A_b is the area of the bolt root, n_b is the number of flange bolts, G_{ID} is the effective sealing inner diameter of the gasket, G_{OD} is the effective sealing outer diameter of the gasket, Φ_g is the gasket relaxation coefficient (provided by the supplier, if not available, 0.7), θf_{max} is the rotation angle of the flange under the maximum stress, θg_{max} is the maximum allowable rotation angle of the flange (usually the value 1), and P_{max} is the maximum design pressure [9].

3.6 Advantages and Disadvantages of Flange Leakage Check Methods

The advantages and disadvantages of each flange leakage check method are listed in the following Table 1.

4 A New Flange Leakage Check Method Based on ASME PCC-1

4.1 Introduction

From the above flange leakage check methods, it can be seen that the ASME PCC-1 method is more suitable for flange leakage check because of its accuracy and operability, and ASME PCC-1 method can provide the bolt pre-tightening stess while checking the flange leakage, closely connect the design calculation and construction, and can play a guiding role. However, this method does not consider

Table 1 Advantages and disadvantages of flange leakage check methods

Methods	Advantages	Disadvantages	Related standard
Kellogg equivalent pressure	Simple to apply	Conservative, cannot be used to check blind flanges, cannot provide suggestions for bolt pre-tightening	GB/T20801 HG/T20645
NC-3658.3	Simple to apply	Radical, high risk of leakage, cannot provide suggestions for bolt pre-tightening	ASME III
EN1591	In line with actual condition, can provide accurate recommendations for the bolt pre-tightening stress	Very complicated, difficult to obtain gasket parameters, difficult to apply	EN13445, GB/T17186-2
Taylor-Waters	Accurate, in line with actual condition, can provide accurate recommendations for the bolt pre-tightening stress	Complicated, allowable bolt pre-tightening force is small, and the crush of the flange gasket is not considered	ASMEVIII, GB150, GB/T17186-1
ASME PCC-1	Accurate, in line with actual condition, can provide accurate recommendations for the bolt pre-tightening stress	Moderate complexity	ASME B31.3 API 660

the influence of the force (moment) of the pipe on the flange that causes the change of gasket sealing pressure and the bolt tensile stress. Therefore, in this paper, the equivalent pressure P_e, which combines the axial force and bending moment of the flange, is used instead of the maximum design pressure P_{max} to check the gasket sealing pressure of the flange in the operating state. The tensile stress of the bolt under the equivalent pressure is also checked to ensure the safety of the bolts under operating conditions, the corrected flange leakage check steps are as follows:

$b_{sel} \geq Sb_{min}$Confirm that the bolt pre-tightening stress is bigger than the minimum recommended pre-tightening stress.

$b_{sel} \leq Sb_{max}$ Confirm that the bolt pre-tightening stress is smaller than the maximum allowable stress.

$b_{sel} \leq Sf_{max}$ Confirm that the bolt pre-tightening stress is smaller than the maximum stress of the bolt without flange damage.

$Sb_{sel} \geq Sg_{min-S}[A_g/(A_b n_b)]$ Confirm that the bolt pre-tightening stress is bigger than the minimum sealing pressure of the gasket under the installed state.

$Sb_{sel} \geq [Sg_{min-O}A_g + (\pi/4)P_e G_{ID}^2]/(\Phi g A_b n_b)$ Confirm that the bolt pre-tightening stress is bigger than the minimum sealing pressure of the gasket under the operating state.

$Sb_{sel} \leq Sg_{max}[A_g/(A_b n_b)]$ Confirm that the bolt pre-tightening stress is smaller than the gasket crushing sealing pressure.

$Sb_{sel} \leq Sf_{max}(\theta g_{max}/\theta f_{max})$ Confirm that the bolt pre-tightening stress is smaller than the limit of flange torsion.

$P_e \leq 4 * (\Phi g A_b n_b Sb_{sel} - Sg_{min-O}A_g)/\pi G_{ID}^2$ Confirm that the bolt pre-tightening stress is bigger than the flange allowable equivalent pressure under the minimum sealing pressure of the gasket in the operating state.

$Sb_{sel} \leq Sb_y - \pi P_e G_{ID}^2/4A_b n_b$ Confirm that the bolt tensile stress under the operating state is smaller than the bolt yield strength.

4.2 Calculation Example

Take the convex face welding flange with the following properties as an example for checking analysis: the design pressure is 9.8 MPa, the design temperature is 60 °C, the flange standard is ASME B16.5, the specification is Class 600 16″, the material is A105, and the gasket is metal graphite spiral wound gasket with inner and outer rings, the bolt follows ASTM B193 B7.

The flange, bolt and gasket parameters required for the calculation are determined at first:

Effective gasket sealing inner diameter G_{ID} = 412.75 mm.

Effective gasket sealing outer diameter G_{OD} = 463.55 mm.

Effective gasket area A_g = 34,945.1 mm^2.

Number of flange bolts n_b = 20.

Bolt root area A_b = 907mm^2.

Minimum bolt pre-tightening stress Sb_{min} = 140 MPa.

Bolt yield strength Sb_y = 720 MPa.

Maximum allowable stress of the bolt without damage to the flange Sf_{max} = 471 MPa.

Minimum gasket sealing pressure under installed state Sg_{min-S} = 80 MPa.

Minimum gasket sealing pressure under operating condition Sg_{min-O} = 40 MPa.

Maximum allowable pressure of gasket Sg_{max} = 311 MPa.

Gasket relaxation coefficient Φg = 0.7 $\Phi_g = 0.7$

Angle of flange under maximum stress θf_{max} = 0.39.

Maximum allowable angle of flange θg_{max} = 1.

Confirm the bolt pre-tightening force and check flange leakage according to the following steps:

Step 1: Initially determine the bolt pre-tightening force Sb_{sel} = 350 MPa.

Step 2: Confirm that the bolt pre-tightening stress is bigger than the minimum recommended bolt pre-tightening stress b_{sel} = max (Sb_{min}, Sb_{sel}) = 350 MPa.

Step 3: Confirm that the bolt pre-tightening stress is smaller than the bolt maximum allowable stress b_{sel} = min(Sb_{max}, Sb_{sel}) = 350 MPa.

Step 4: Confirm that the pre-tightening stress of the bolt is smaller than the bolt maximum stress without damage to the flange b_{sel} = min(Sf_{max}, Sb_{sel}) = 350 MPa.

Step 5: Confirm that the bolt pre-tightening stress is bigger than the minimum sealing pressure of the gasket under the installed state $Sb_{sel} \geq Sg_{min-S}$ [$A_g/(A_b n_b)$] = 154.1 MPa.

Step 6: Confirm that the bolt pre-tightening stress is smaller than the gasket crushing sealing pressure $Sb_{sel} \leq Sg_{max}[A_g/(A_b n_b)]$ = 599.1 MPa.

Step 7: Confirm that the bolt pre-tightening stress is smaller than the limit of flange torsion $Sb_{sel} \leq Sf_{max}(\theta g_{max}/\theta f_{max}) = 1207.7\,\text{MPa}$.

Step 8: Determine the allowable equivalent pressure of the flange $P_e \leq 4 * (\Phi g A_b n_b Sb_{sel} - Sg_{min-O}A_g)/\pi G_{ID}^2 = 22.8\,\text{MPa}$.

Step 9: Confirm that the bolt tensile stress under operating state is smaller than the bolt yield strength $Sb_{sel} \leq Sb_y - \pi P_e G_{ID}^2/4A_b n_b = 552\,\text{MPa}$.

Step 10: Check whether the equivalent pressure of the flange under operating state is smaller than the allowable equivalent pressure. If it is satisfied, the check is completed. If not, the piping layout shall be changed and check until satisfied.

4.3   Method Application

When applying this method for flange leakage check, proper bolt pre-tightening force should be selected. Although higher bolt pre-tightening force can increase the gasket pressure and makes the sealing effect better, it will massively increase the gasket pressure in the large diameter and high-pressure flange. Therefore, it is recommended to use a higher bolt pre-tightening force for small-diameter low-pressure flanges, and a lower bolt pre-tightening force for large-diameter high-pressure flanges. Then calculate the allowable equivalent pressure and the corresponding bolt pre-tightening force table for a series of commonly used flanges, gaskets and bolts. Compare the flange equivalent stress in project with the stress calculated by the software to complete the flange leakage check and the provide guidance of bolt pre-tightening force.

Apply this method in some projects to check flange leakage and help flange installation, and it was found to be in good agreement with engineering practice. No leakage occurred during the pressure test and operation stage of the flange and flange blind plate which proves that the method is reliable.

5   Conclusion

Flange connections are widely used in pressure vessels and pipelines. Ensuring the sealing performance and correct installation of flange are of great significance to the safe operation of pressure vessels and pipelines.Although there are many flange leakage check methods, they all have certain limitations and the conclusions are different, which brings great controversy and inconvenience to flange leakage check in the engineering design phase.Base on the ASME PCC-1 method and consider the the influence of the force (moment) of the pipe on the flange that causes the change of gasket sealing pressure and the bolt tensile stress, this paper proposes a new method, which can better ensure the flange sealing and bolt safety under operating conditions.What's more, the allowable equivalent pressure and corresponding bolt pre-tightening force are calculated for commonly used flanges, gaskets and bolts, which are used for flange check in the design phase and flange installation in the construction phase. It has achieved good application effects in engineering practice, and hasbetter accuracy and maneuverability.

References

1. Guo, Y.W., Qing, X.H.: Analysis and check of flange leakage in engineering design stage. Chem. Fertilizer Des. **56**(5), 18–21 (2018)
2. Su, G.: Study on mechanical properties of pressed loose steel pipe flange (Master Thesis). Jiangnan University, pp. 5–9 (2009)
3. Xing, G.P., Luo, G.H., Zheng, X.B.: Cause Analysis of Flange Leakage Petrol-Chemical Equipment Technology **29**(2), 63–66 (2008)
4. Zhang, X.: Discussion on flange checking method in pressure piping design. Petrol-Chem. Equip. **68**(8), 5–9 (2010)
5. Flanges and Their Joint -Design Rules for Gasketed Circular Flange Connections-Part 1: Calculation Method EN 1591–1–2011, Berlin (2011)
6. Ji, Y.D., Zhang, L.Z.: Test and analysis of gasket factors based on leakage rate. Lubr. Eng. **44**(8), 99–113 (2019)
7. Guidelines for Pressure Boundary Bolted Flange Joint Assembly ASME PCC-1–2019, New York, pp. 73–76 (2019)
8. Noble, R.: Guidelines for managing the integrity of bolted joints-developments in european standards. In: Proceedings of the ASME 2013 Pressure Vessels and Piping Conference, Paris, July 2013
9. Guan, K.S., Wang, Z.W., Cai, R.L.: Resolution of ASME PCC - 1–2010 guidelines for pressure boundary bolted flange joint assembly. Press. Vessel Technol. **28**(12), 41–48 (2011)

Calibration of Numerical Models of Railway Vehicles Based on Dynamic Tests

Diogo Ribeiro, Rui Calçada, and Pedro Montenegro

Abstract The dynamic analysis of the vehicle-bridge system is a complex interaction problem which typically involves the development of advanced finite element (FE) models of the vehicles, track and bridge subsystems and their interfaces. The experimental calibration of the FE models of these subsystems is usually performed based on modal parameters, namely, the natural frequencies and mode shapes. This article describes the experimental calibration of 3D FE numerical models of railway vehicles, one for passenger and one for freight, based on modal parameters. The BBN vehicle is a tourist class vehicle that belongs to Alfa Pendular train and the vehicle Kbs is a freight railway wagon for transport of timber, both operating in the Portuguese railway network. The 3D FE numerical model of the vehicles consider the flexibility of the carbody/base floor and includes the suspensions and axles. The dynamic tests of the vehicles allowed the determination of the frequencies and modal configurations of several vibration modes involving rigid body and structural movements. The calibration was successfully performed through an iterative method based on a genetic algorithm, which demonstrated the stability of a significant number of numerical parameters estimates, and above all, a very good agreement between experimental and numerical modal parameters.

Keywords Railways vehicles · FE model · Dynamic tests · Calibration · Genetic algorithm

D. Ribeiro (✉)
CONSTRUCT-LESE, School of Engineering of Polytechnic of Porto, Porto, Portugal
e-mail: drr@isep.ipp.pt

R. Calçada · P. Montenegro
CONSTRUCT-LESE, Faculty of Engineering of University of Porto, Porto, Portugal

1 Introduction

In recent years the railway traffic in Europe is rapidly developing, and an increasing number of railway operators are planning to increase the axle loads and speed of trains, in order to make the passenger and freight traffic on rails more competitive.

When interacting with the railway track, moving trains induce vibrations that can affect the structural stability of the infrastructure and rolling stock components. In case of bridges, this can lead to damages and can affect its performance and durability.

The analysis of the dynamic behaviour of the train-bridge system can be carried out through different methodologies, mainly analytical and numerical [1, 2]. For complex problems the use of numerical methodologies including the train-bridge interaction are common, and therefore there is the need of modeling the train, and inherently the knowledge of its geometrical and mechanical parameters is required [3–7].

Currently few works have been published about the experimental calibration of numerical models of train vehicles. From the works identified it should be highlighted the studies performed by Ribeiro et al. [3] and Jung-Seok et al. [8], for passenger vehicles, and Harak et al. [9] and Ahmadian [10] for freight vehicles.

In this paper some results of a calibration of the numerical model of a passenger vehicle and a freight wagon based on modal parameters are presented. The 3D FE numerical model of both vehicles considers the flexibility of the carbody and includes the suspensions and axles. The modal parameters of the vehicles were determined based on dedicated dynamic tests performed in rest position. The calibration methodology involved a sensitivity analysis and an optimization based on a genetic algorithm. The results revealed a very good correlation between experimental and numerical after updating modal parameters.

2 Calibration Methodology

The computational implementation of an iterative method based on a genetic algorithm involved the use of three software packages: Ansys, Matlab and OptiSlang [11, 12]. Figure 1 shows a flowchart that illustrates the computational implementation of the method, indicating the softwares involved in the different phases.

In ANSYS environment the FE numerical model is developed based on a set of initial parameter values $\theta_1, \theta_2, \ldots, \theta_k$, where k is the number of individuals in each generation. The pre-selection of the calibration parameters is performed based on global sensitivity analysis. The sets of parameter values of generation 1 are randomly generated in OptiSlang software by applying the Latin Hypercube method. The modal analysis is performed as well as the export of the numerical modal parameters.
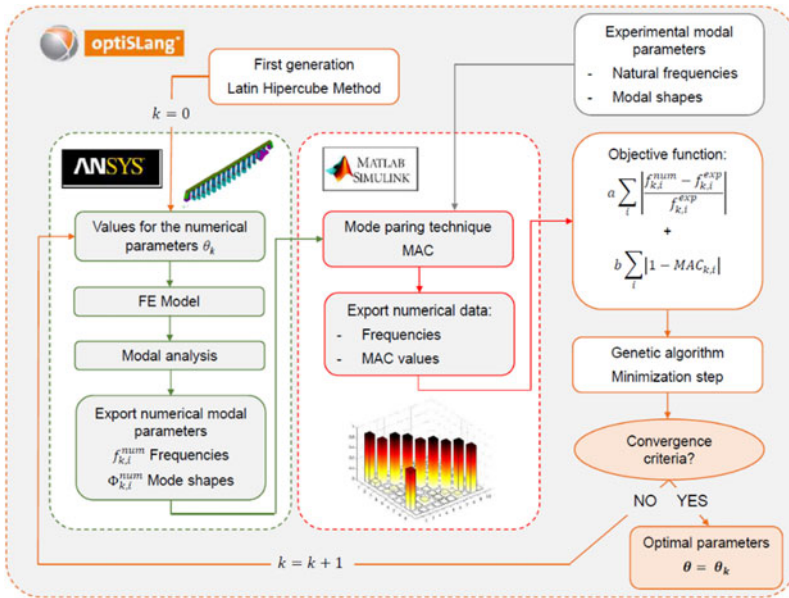
Fig. 1 Flowchart of the calibration methodology

In Matlab software, based on the experimental modal information, the mode pairing between numerical and experimental modes based on MAC (Modal Assurance Criterion) parameter is performed. The values of the natural frequencies and the corresponding MAC values are exported in text format.

Finally, the OptiSlang software, based on an objective function and on the application of an optimization technique supported by a genetic algorithm, estimates a new set of parameters focused on the minimization of the objective function residuals. The objective function includes two terms, one relative to the residuals of the frequencies of vibration and other related to the residuals of modal configurations. This procedure is repeated iteratively until the maximum number of generations is reached.

3 Passenger Railway Vehicle

3.1 Description and Numerical Model

The BBN vehicle, Fig. 2a, is a tourist class vehicle of the Alfa Pendular train [3]. The total mass of the vehicle, including the bogies, achieve 52.2 t, in running order, and 55.0 t, in normal load conditions. The length of the vehicle is 25.9 m, the distance between the bogie pivots of the vehicle equals 19.0 m and the wheelbase
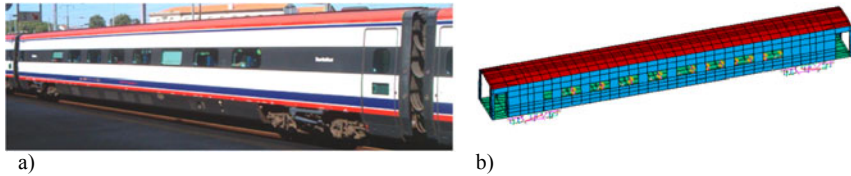
Fig. 2 Passenger railway vehicle BBN: a) perspective view, b) FE numerical model

equals 2.7 m. The carbody is formed by a tubular structure made of aluminum alloy consisting of 20 alveolar extruded panels longitudinally welded.

The modal analysis of the BBN vehicle was performed using a three-dimensional finite element model developed in the ANSYS software. The use of a finite element formulation allows considering the influence of the deformability of the carbody, bogies and axles. Figure 2b presents a perspective of the numerical model.

The carbody was modeled by shell finite elements while the bogies were modeled by beam finite elements, with the exception of the suspensions, the connecting rods and the tilting system which were modeled by spring-damper assemblies. Additionally, the passenger-seat system was modeled, in a simplified manner, by a 1-DOF system composed of a mass over a spring-damper assembly.

3.2 Experimental Tests

The dynamic test of the carbody was conducted at EMEF facilities, in Porto, in order to identify the natural frequencies and vibration modes of the carbody involving rigid body and structural movements. The carbody was instrumented with 14 piezoelectric accelerometers (A0 to A13), PCB's model 393A03, in the vertical direction, and placed on the base of the carbody next to the intersection with the side walls, according to the scheme of Fig. 3a. The accelerometers located within the area of the seats were positioned by means of metallic angles fixed to the structure of the seat with magnetic disks. Attending to the high damping of the vibration modes of the vehicle, it was necessary to use an external excitation, through people jumping, in order to increase the vibration levels. The modal identification was carried out through the application of the stochastic subspace identification method based on the time series of acceleration (SSI-DATA). Figure 3b shows the stabilization diagrams that were estimated based on state models of order between 1 and 160. The alignments highlighted in the figure correspond to the five vibration modes of the carbody with natural frequencies ranging from 1.01 Hz to 12.26 Hz. These modes are associated with rigid body (modes 1 to 3), torsion (mode 4) and bending (mode 5) movements of the carbody.

3.3 Calibration

The calibration process based on a genetic algorithm involved 7 design variables, particularly, the stiffness of secondary suspension—front (K_{S1}), stiffness of secondary suspension—rear (K_{S2}), stiffness of base panel (RMI_b), stiffness of side walls panel (RMI_p), additional mass—base (ΔM_b), additional mass—side walls (ΔM_p) and additional mass—cover (ΔM_c). Ten modal responses were considered, in particular 5 natural frequencies and 5 MAC parameters. Figure 4 presents the ratios of the values of four parameters of the model, in relation to the predefined limits of variation, as well as the optimized values, for four independent optimization runs (GC1 to GC4). A 0% ratio means that the parameter coincides with the lower limit. A ratio of 100% means that it coincides with the upper limit. The analyzed parameters present a very good stability with variations below 10%.
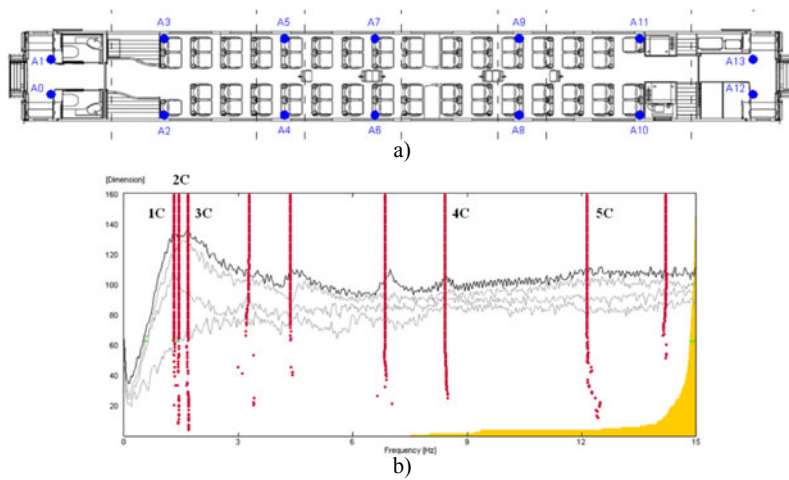
Fig. 3 Dynamic test: a) experimental setup (plan view), b) SSI-DATA: stabilization diagrams
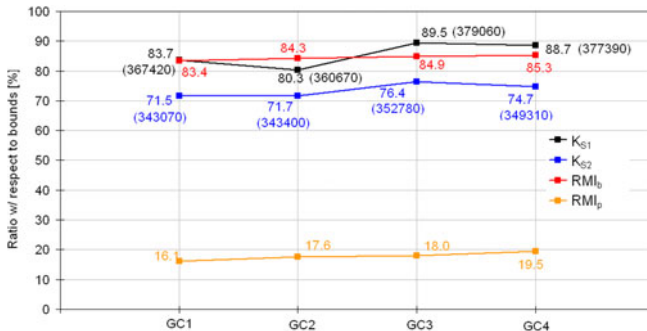
Fig. 4 Values of four numerical parameters for optimization cases GC1 to GC4
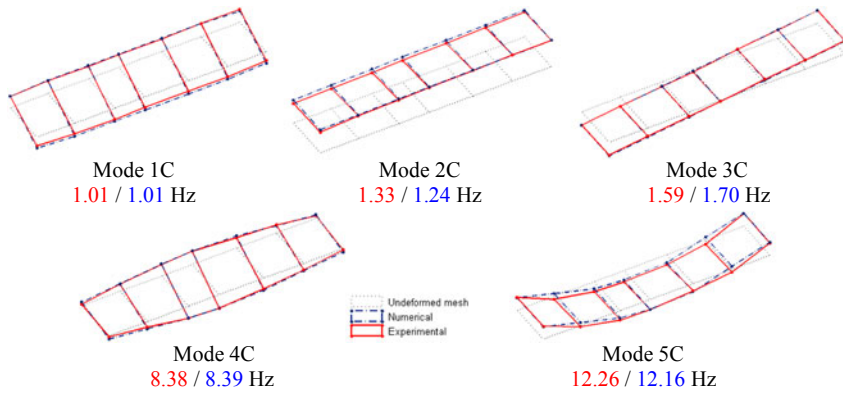
Fig. 5 Comparison between experimental (red) and numerical after updating (blue) modal parameters of the passenger vehicle

Figure 5 shows the comparison between experimental and numerical after updating modal parameters. The average error of the frequencies decreased from 20.3% before calibration to 2.9% after calibration. In turn, the average value of the MAC parameter increased from 0.927 before calibration to 0.937 after calibration.

4 Freight Railway Wagon

4.1 Description and Numerical Model

The vehicle Kbs, Fig. 6a, is a freight railway wagon for transport of timber (usually eucalyptus type) that operates in the Portuguese railway network. The wagon has a length of 12.8 m with a width of 2.7 m and the axles spacing is equal to 8 m. The wagon's structure is formed by a grid of steel beams and suspensions of the type UIC standard. The maximum axle load is equal to 23.7 t and its average tare weight is 16.3 t.

The 3D FE model of the wagon Kbs developed in ANSYS software is presented in Fig. 6b. The model includes beam elements for simulating the grid of girders of the base floor and lateral shafts, and shell elements, for simulating the base platform and extremities plates of the vehicle. The axles were modelled using rigid bars and

are connected to the base platform by spring elements simulating the primary suspension system. The wheel-rail contact was represented by Hertzian spring element. The mass of on-board equipment was simulated based on mass elements. The towed load was also represented by means solid elements.

4.2 Dynamic Test

The dynamic test of the freight car allows the identification of the natural frequencies and modes of vibration of the freight car involving rigid body and structural movements. The test was performed in two different configurations, loaded and unloaded, in order to consider, the possible non-linear behaviour of the suspensions. The freight car was instrumented with 16 piezoelectric accelerometers, PCB's model 393B12. The response was evaluated in terms of accelerations in transversal (y) and vertical (z) directions in only one setup (Fig. 7a). The accelerometers are attached to the base platform using metallic plates fixed to the vehicle with magnetic disks. The identification of the vehicle modal parameters was carried out through the application of Enhanced Frequency Domain Decomposition (EFDD) method (Fig. 7b). Five modes of vibration with natural frequencies ranging from 5.10 Hz to 8.11 Hz were identified. These modes are associated with bending (modes 1 and 2), rigid body (modes 3 and 4) and torsional (mode 5) movements of the vehicle's platform.

a)   b)
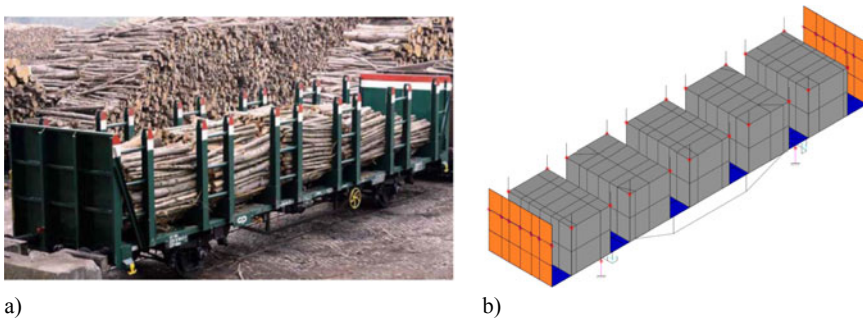
Fig. 6 Freight railway wagon Kbs: a) perspective view, b) FE numerical model

4.3 Calibration

The calibration process based on a genetic algorithm involved 7 design variables, particularly, the vertical stiffness of the suspensions (K_1 to K_4), modulus of elasticity of steel (E_s) and additional masses on the lateral shafts (M_{lat}) and extremity plates (M_{ext}). Five modal responses were considered, particularly 5 natural frequencies and 5 MAC parameters. Figure 8 presents the ratios of the values of each parameter of the model, in relation to the predefined limits of variation, as well as the optimized values, for four independent optimization runs (GA1 to GA4). The analyzed parameters present a good stability with variations below 10%, except for the modulus of elasticity of steel. Precisely, this is one of the parameters that the sensitivity analysis has shown to have a smaller influence over the numerical responses.

Figure 9 shows the comparison between experimental and numerical after updating modal parameters. The average error of the frequencies decreased from 8.5% before calibration to 3.2% after calibration. In turn, the average value of the MAC parameter increased from 0.911 before calibration to 0.950 after calibration.
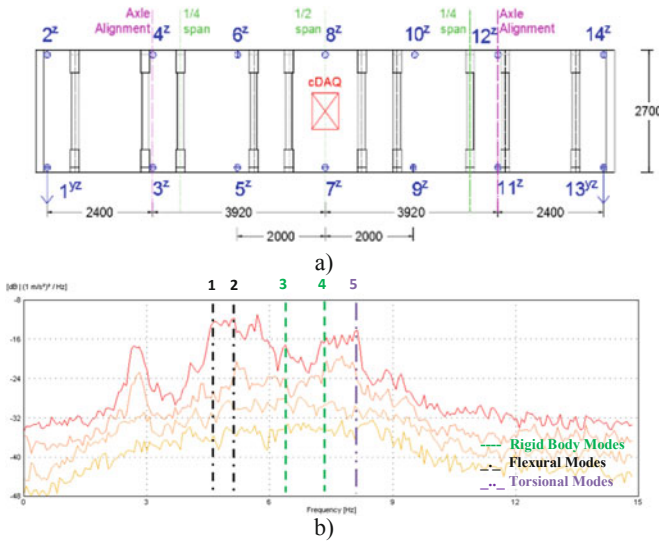
Fig. 7 Dynamic test: a) experimental setup (plan view), b) EFDD method: average and normalized singular values of the spectral matrix
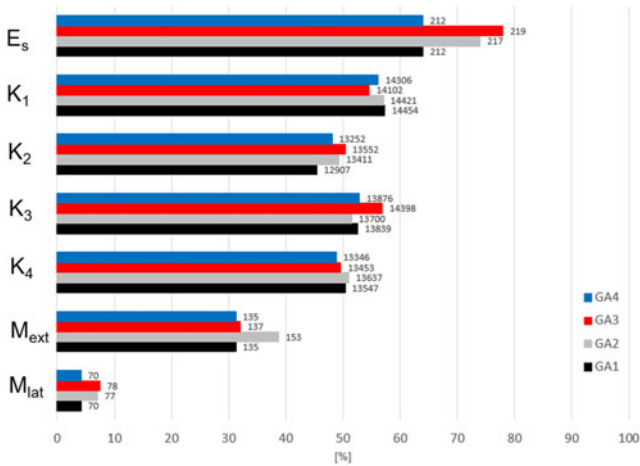
Fig. 8 Values of numerical parameters for optimization cases GA1 to GA4

Mode 1	Mode 2	Mode 3
5.10 / 5.00 Hz	5.69 / 5.39 Hz	6.38 / 6.41 Hz

Mode 4	Mode 5
7.56 / 7.25 Hz	8.11 / 8.46 Hz
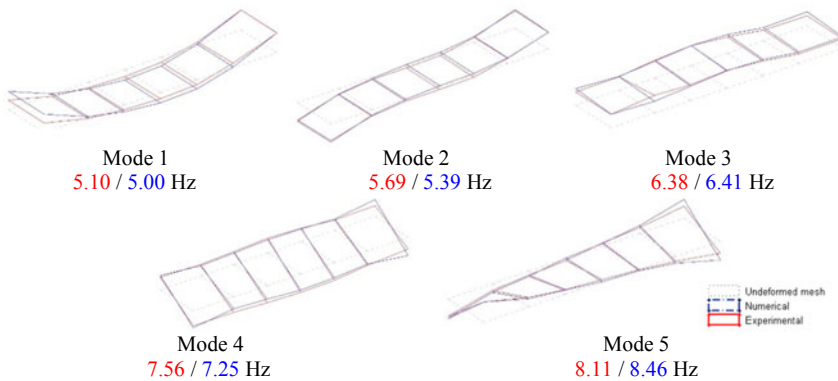
Fig. 9 Comparison between experimental (red) and numerical after updating (blue) modal parameters of the freight vehicle

5 Conclusions

This paper described the experimental calibration of the numerical models of a passenger and freight vehicles based on modal parameters. The dynamic tests allowed the identification of frequencies and modal configurations of several vibration modes involving rigid body and structural movements. The results show important differences in the dynamic properties of the passenger vehicles in comparison to the freight wagon, particular in what concerns the order of the rigid body and flexural modes of vibration.

The calibration of the numerical model was conducted through an iterative methodology based a genetic algorithm. The calibration process was successful which is demonstrated by the stability and consistence of the numerical parameters estimates and by the very good agreement between experimental and numerical modal parameters. In future studies, the calibrated models of the railway vehicles will be used for the dynamic analysis of the train-tack-bridge system in order to evaluate the integrity and safety of the railway infrastructures.

References

1. Ticona Melo, L., Malveiro, J., Ribeiro, D., Calçada, R., Bittencourt, T.: Dynamic analysis of the train-bridge system considering the non-linear behaviour of the track-deck interface. Eng. Struct. **220**, 110980 (2020). https://doi.org/10.1016/j.engstruct.2020.110980
2. Neves, S., Montenegro, P., Jorge, P., Calçada, R., Azevedo, A.: Modelling and analysis of the dynamic response of a railway viaduct using an accurate and efficient algorithm. Eng. Struct. **226**, 111308 (2021). https://doi.org/10.1016/j.engstruct.2020.111308
3. Ribeiro, D., Calçada, R., Delgado, R., Brehm, M., Zabel, V.: Finite-element model calibration of a railway vehicle based on experimental modal parameters. Veh. Syst. Dyn. **51**, 821–856 (2013). https://doi.org/10.1080/00423114.2013.778416
4. Ticona Melo, L., Ribeiro, D., Calçada, R., Bittencourt, T.: Validation of a vertical train–track–bridge dynamic interaction model based on limited experimental data. Struct. Infrastruct. Eng. **16**(1), 181–201 (2020). https://doi.org/10.1080/15732479.2019.1605394
5. Meixedo, A., Ribeiro, D., Calçada, R., Delgado, R.: Global and Local dynamic effects on a railway viaduct with precast deck. In: Pombo, J. (ed.) Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance, Civil-Comp Press, Stirlingshire, UK, Paper 77 (2014). https://doi.org/10.4203/ccp.104.77.
6. Montenegro, P., Barbosa, D., Carvalho, H., Calçada, R.: Dynamic effects on a train-bridge system caused by stochastically generated turbulent wind fields. Eng. Struct. **211**(110430), 1–16 (2020). https://doi.org/10.1016/j.engstruct.2020.110430
7. Montenegro, P., Heleno, R., Carvalho, H., Calçada, R., Baker, C.: A comparative study on the running safety of trains subjected to crosswinds simulated with different wind models. J. Wind Eng. Ind. Aerodyn. **207**(104398), 1–15 (2020). https://doi.org/10.1016/j.jweia.2020.104398
8. Jung-Seok, K., Jong-Cheoi, J.: Natural frequency evaluation of a composite train carbody with length of 23m. Compos. Sci. Technol. **66**(13), 2272–2283 (2006). https://doi.org/10.1016/j.compscitech.2005.11.036
9. Harak, S., Sharma, S., Harsha, S.: Structural dynamic analysis of freight railway wagon using finite element method. Proc. Mater. Sci. **6**, 1891–1898 (2014). https://doi.org/10.1016/j.mspro.2014.07.221
10. Ahmadian, H.: Extracting modes of a railway vehicle from measured responses. In: Proceedings of 1st International Operational Modal Analysis Conference (2004)

11. Alves, V., Meixedo, A., Ribeiro, D., Calçada, R., Cury, A.: Evaluation of the performance of different damage indicators in railway bridges. Proc. Eng. **114**, 746–753 (2015). https://doi.org/10.1016/j.proeng.2015.08.020
12. Costa, C., Ribeiro, D., Jorge, P., Silva, R., Calçada, R., Arêde, A.: Calibration of the numerical model of a short-span masonry railway bridge based on experimental modal parameters. Proc. Eng. **114**, 846–853 (2015). https://doi.org/10.1016/j.proeng.2015.08.038

Numerical Simulation of Crack Evolution Under Hydraulic Fracturing of Medium-Hard Rock

Jialiang Liu, Jinyang Li, Yujie Zhu, Dongping Zhou, Hua He, Junjie Zhou, and Kai Wang

Abstract The mechanical excavation technology, as a commonly used method in tunnel excavation of hard rock, has many shortcomings, such as poor geological adaptability and intense wear of cutter head and so on, affecting the excavation efficiency. The hydraulic fracturing technology can effectively evade these issues and achieve the efficient excavation by pre-fracturing rock. Therefore, taking medium-hard rock as an example, this paper established the numerical models of single-borehole and double-borehole hydraulic fracturing medium-hard rock based on the extended finite element method (XFEM), and explored the evolution law of cracks and the damage dissipation energy variation of the rock. Results indicated that under single-borehole hydraulic fracturing, the propagation length and width of the crack experienced the rapid growth period and slow growth period, and the damage dissipation energy of rock mainly experienced the step-like growth period. By the comparative analyses of crack propagation length and width evolution as well as the damage dissipation energy of rock under the single-borehole and double-borehole hydraulic fracturing, it is found that under the double-borehole hydraulic fracturing the propagation rate and damage dissipation energy of single prefabricated fracture are greatly in-creased, and the fracturing efficiency is greatly improved, compared with the single-borehole hydraulic fracturing. Results can provide a theoretical reference for improving the application level of hydraulic fracturing in tunnel excavation.

D. Zhou (✉) · J. Zhou · K. Wang
Chongqing Energy Investment Group Science & Technology Co., LTD.,
Chongqing 400061, People's Republic of China
e-mail: dpzcq2001@163.com

J. Liu · J. Li · Y. Zhu
Civil Engineering College, Chongqing Jiaotong University,
Chongqing 400074, People's Republic of China

H. He
Chongqing Energy Investment Group Co., LTD.,
Chongqing 401121, People's Republic of China

1   Introduction

The exploitation and excavation of the various underground resources usually
encounter many problems, especially the hard rock layer problems such as high
density, poor breaking conditions and so on. Usually, the mechanical drilling tool is
used in the tunneling excavation, but the geological adaptability is poor and the
wear of the hard rock to the drilling tool causes the frequent replacement of cutting
head, thereby engendering the high cost and low efficiency. Therefore, it is urgent
to take efficient measures to achieve the efficient excavation of hard rock [1].
Hydraulic fracturing technology was widely applied in the field of geotechnical
engineering [2–4]. The technology can reduce the hardness and integrity of the rock
by crack initiation and expansion, which greatly reduces the loss of the tool head
and promotes the mechanical tunneling of rock.

Recently, many scholars have carried out a large number of experiments and
numerical simulation studies on the hydraulic fracturing. Nasehi [5] et al. performed
the numerical simulation of hydraulic fracturing rock and analyzed the effects of
in-situ stress and strength parameters of intact rock on the fracturing properties
based on the discrete element method. By the extended finite element method
(XFEM), Tian [6] et al. simulated the processes of continuous fracturing, the
alternative fracturing and the improved zipper hydraulic fracturing, and studied the
effects of the in-situ stress difference and the crack spacing on crack propagation.
Shimizu et al. [7] discussed the influence of fluid viscosity and rock granularity on
hydraulic fracturing by means of discrete element method. Francisco et al. [8] made
a significant prediction concerning the coupling crushing behavior of the hydraulic
fracturing rock based on the extended finite element method, and analyzed the
interaction between the hydraulic crack and the natural crack. By the Cohesive
units, Xue et al. [9] studied the relationship between the fluid pressure on the crack
surface and the crack propagation as well as fluid percolation. The effect of fluid
pressure on the width variation of crack during the crack propagation was studied
by Ji et al. [10]. Shilova [11] et al. explored the influence of wellbore and crack
radius on initiation pressure of crack and crack propagation by the hydraulic
fracturing experiment of rock.

The above-mentioned scholars had made great progress in the effects of the
in-situ stress difference, rock strength and natural cracks on the crack propagation
of rock under hydraulic fracturing, but there are few influence researches of the
borehole and its arrangement on the crack propagation. The borehole has a sig-
nificant effect on the energy dissipation in rock and crack propagation under
hydraulic fracturing, which influences the rock crushing efficiency. Hence, the
numerical models of hydraulic fracturing medium-hard rock with single and double
boreholes were established based on the XFEM. And it systematically explored the

evolution law of crack propagation length, width and rate as well as damage dissipation energy under the working conditions of single-borehole and double-borehole. The study is expected to more deeply reveal the crack propagation law under porous fracturing of medium-hard rock and provide a potent reference for the actual hydraulic fracturing construction in tunnel excavation.

2 Numerical Simulation

2.1 Basic Hypotheses and Theories

In the process of hydraulic fracturing, the basic assumptions for the establishment of the two-dimensional numerical model are as follows [12]: 1) the model is isotropic and homogeneous limestone; 2) only the hydraulic effect in the two-dimensional plane is considered, ignoring the vertical flow of the fluid; 3) the fluid in the model is single-phase and incompressible; 4) the model of rock is incompressible; 5) the influence of gravity is not considered.

For the two-dimensional model, the Griffith strength theory under biaxial stress can be expressed as:

$$P_t = \frac{(\sigma_1 - \sigma_3)^2}{8(\sigma_1 + \sigma_3)}, \frac{\sigma_3}{\sigma_1} \geq -\frac{1}{3} \tag{1}$$

$$P_t = -\sigma_3 \quad \frac{\sigma_3}{\sigma_1} \leq -\frac{1}{3} \tag{2}$$

In the formula, σ_1 and σ_3 are the maximum and minimum principle stress respectively, P_t is the compressive strength.

2.2 Model Establishment

Taking limestone as the simulation model of medium-hard rock, the two-dimensional models were established to simulate the actual conditions (Fig. 1). The mechanical parameters of the models are shown in Table 1. In the models, prefabricated fractures were simulate by inserting spread function, and the boreholes of fracturing fluid on the prefabricated fractures were selected. The maximum circumferential stress criterion of crack initiation and damage evolution criterion were introduced to determine the crack initiation and damage evolution of element [13]. Therefore, two analysis steps were established in the process of single-borehole hydraulic fracturing: the one was the in-situ stress balance stage, the other was the hydraulic fracturing stage. For the double-borehole hydraulic fracturing, it was distributed to four analysis steps: the first one was the in-situ stress

balance stage, the second one was the hydraulic fracturing stage of borehole I, the third one was the unload stage of borehole I, and the fourth one was the hydraulic fracturing stage of borehole II.

The minimum and maximum principal in-situ stress of the rock model were 6 MPa and 15 MPa respectively, and the initial void ratio was 0.1. The viscosity of water as fracturing liquid was 0.001 Pa•s. The unit weight of water was 9800 N/m^3. The constraint boundary conditions were set to fix the dis-placement in X and Y directions. The initial pore pressure was 0, with the perforation angle of 60°. The initial time step at the hydraulic injection stage was 0.1 s, with the injection time of 100 s. And, the water pressure of hydraulic fracturing was 30 MPa.

2.3 Control Equation

The crack growth under hydraulic fracturing is caused by the expansion of fluid pressure and shear displacement effect, so the crack under hydraulic fracturing is composite crack. The failure cohesion model was established by XFEM method to study the crack initiation and propagation criterion of the composite crack, and the main analyzed contents contained the functional response relationship between the interface tensile stress and the relative displacement of the interface, as well as the relationship between the interface energy in the failure process.

In the process of hydraulic fracturing, as the injection of the fracturing fluid, the seepage pressure of the fluid acting on the crack surface increased, which led to the increase of fluid loss to the rock and the change of the stress state in the crack. Whereas the stress change will inevitably cause the change of the parameters such as the porosity of the medium-hard rock and the seepage velocity of the fluid. Conversely, the stress change has the important influence on the seepage pressure
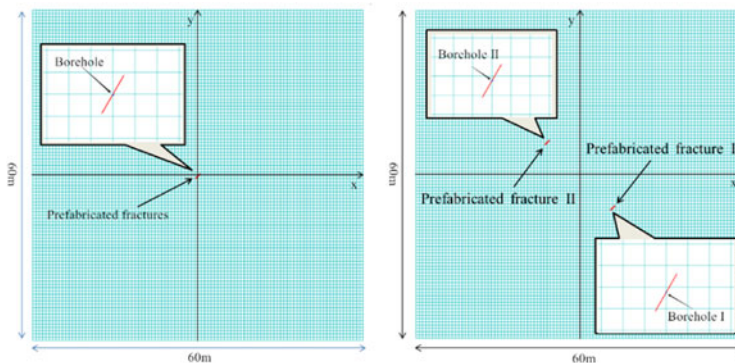
Fig. 1 Two-dimensional geometric model of hydraulic fracturing limestone

Table 1 Mechanical parameters of limestone model

Elastic Modulus / GPa	Poisson's ratio	Porosity ratio	Permeability coefficient	Filtration coefficient	Tensile strength/MPa
15	0.25	0.1	1×10^{-7}	1×10^{-14}	6

of the fluid. The interaction between the fluid seepage and the rock deformation in medium-hard rock is called the seepage-stress coupling [14].

The equilibrium equation of solid rock [9]:

$$\int_V (\overline{\sigma} - PwI) : \delta_\varepsilon dV = \int_S \hat{T} \delta u dS + \int_V \hat{f} \delta u dV \tag{3}$$

The continuity equation of seepage liquid:

$$\frac{d}{dt}\left(\int_V n dV\right) = -\int_S n \overrightarrow{n} v_w dS \tag{4}$$

Critical stress criterion of element damage:

$$\max\left\{\frac{\langle \sigma_n \rangle}{\sigma_n^0}, \frac{\sigma_s}{\sigma_s^0}, \frac{\sigma_t}{\sigma_t^0}\right\} = 1 \tag{5}$$

Linear degradation criterion of elastic modulus is represented as follows.

$$E = (1 - D)E_0 \tag{6}$$

$$D = \frac{\delta_m^f(\delta_m^{\max} - \delta_m^0)}{\delta_m^{\max}(\delta_m^f - \delta_m^0)} \tag{7}$$

In these formulas, $\delta_\varepsilon, \overline{\sigma}, \hat{T}, \hat{f}$ and δu are the virtual strain, effective stress, surface force, unit volume force and displacement except gravity, respectively. $v_w, \overrightarrow{n}$ and n are the seepage velocity of the fluid in the solid, the unit vector in the normal direction of the surface S and the porosity, respectively. σ_n is the normal stress, σ_s and σ_t are the tangential stress (in two dimensions, $\sigma_t = 0$), σ_n^0 is the threshold stress of normal damage, σ_s^0 and σ_t^0 are the threshold stresses of tangential damage. E_0, E, D, $\delta_m^{\max}$ and δ_m^0 are the Young's modulus of the undamaged and damaged element, the damage factor, the maximum displacement of the element in the loading process, and the opening displacement when it reaches the damage.

3   Results and Discussion

3.1   *Propagation Analysis of Prefabricated Fracture Under Single-Borehole Hydraulic Fracturing*

With the continuous injection of high pressure fluid into limestone, the internal pressure of the limestone increased gradually, and the concentrated cluster formed in the perforation, which caused the crack initiation along the direction of the prefabricated fractures. When the pressure increased to a certain value, the prefabricated fracture evolved into the wide fracture. As the high-pressure fluid entered into the fracture, the fluid reached the fracture tip and the fracture continued to expand. Simultaneously, the continuous expansion of the fracture caused the decrease of the stress concentration at the crack tip. The crack propagation versus time is shown in Fig. 2. There was an energy accumulation period of crack propagation, and the prefabricated fracture propagated at $t = 5$ s. The crack extended along both positive and negative directions of Y-axis. Due to the restriction of the maximum and minimum horizontal in-situ stresses, the crack expansion length in the positive negative directions of Y-axis was larger than that in the negative directions of Y-axis.

The change curves of crack length and the maximum width with time are shown in Fig. 3. Before $t = 5$ s, the crack length and width have no obvious change. The stage from $t = 5$ s to $t = 20$ s was the rapid growth period of the crack propagation length. The stage from $t = 20$ s to $t = 100$ s was the slow growth period of the crack propagation length, and the propagation length tended to be stable at the terminal stage. For the crack propagation width, the stage from $t = 5$ s to $t = 25$ s was the rapid growth period of crack propagation width, and the stage from $t = 25$ s to $t = 100$ s was the slow growth period of crack propagation width. The detailed process of crack propagation is as follows: From $t = 5$ s to $t = 20$ s, the propagation length of the crack increased from 1.58 m to 9.56 m at an average rate of 0.550 m/s, and from 20 to 100 s, the propagation length of the crack increased from 9.56 m to 22.02 m at an average rate of 0.156 m/s; from 5.5 s to 25 s the propagation width of the crack increased from 1.152 mm to 2.223 mm at an average rate of 0.055 mm/s, and from 25 to 100 s, the propagation width of the crack increased from 2.223 mm to 3.355 mm at an average rate of 0.015 mm/s.

It is obvious that both the transverse and longitudinal propagation rates of the prefabricated fracture first sharply increased and then slowly increased. This is because that as the high pressure fluid enters the prefabricated fracture, the fluid pressure in the crack begins to increase. And the propagation rate of the crack increases at a sharp speed due to the stress concentration at the crack tips, hence the expansion rate of the crack increased rapidly. Whereas, as the crack length expanded to a certain length, the fluid pressure reached the peak value of 57.22 MPa, and the pressure began to decrease, finally it dropped to 23.74 MPa at $t = 100$ s (Fig. 4), thereby the expansion rate of the crack increased slowly.

(a) *t*=5 s   (b) *t* = 20 s   (c) *t* = 25 s   (d) *t* = 100 s

Fig. 2 The prefabricated fracture propagation nephogram under single-borehole hydraulic fracturing at different time

(a) Propagation length   (b) Maximum width

Fig. 3 The change curve of propagation length and maximum width of prefabricated fracture under single-borehole hydraulic fracturing with time

The higher the damage dissipation energy of rock is, the larger the fragmentation degree of the rock is [15, 16]. The energy dissipation process reflects the evolution of rock damage under hydraulic fracturing. By the change curve of damage dissipation energy of rock with time in Fig. 5, the stage from $t = 0$ s to $t = 5$ s was the energy accumulation period, the damage dissipation energy of rock was 0. The stage from $t = 5$ s to $t = 22$ s was the rapid energy development period where the damage dissipation energy of rock increased rapidly and reached a certain threshold value. And the stage from $t = 22$ s to $t = 100$ s was the step-like growth period of energy where the damage dissipation energy increased as the time increased. The maximum damage dissipation energy of rock was 46.1 kJ in the whole fracturing process.

3.2 Propagation Analysis of Prefabricated Fractures Under Double-Borehole Hydraulic Fracturing

Under double-borehole hydraulic fracturing, the spacing of the two prefabricated fractures was set as 15 mm, which can effectively improve the fracturing effect of rock [17]. The numerical simulation of double-borehole hydraulic fracturing was divided into two stages: the first one was the hydraulic fracturing of borehole I for 100 s, the second one was the hydraulic fracturing of borehole II for 100 s. The crack propagation process is shown in Fig. 6. The prefabricated fracture I first initiated and expanded, and the prefabricated fracture II had simultaneously occurred crack initiation and propagation due to the influence of propagation stress field of prefabricated fracture I. The propagation characteristics of prefabricated fracture I under double-borehole hydraulic fracturing were basically same as that under single-borehole fracturing. For the prefabricated fracture II, it initiated and expanded at $t = 5$ s. From $t = 5$ s to $t = 30$ s, the propagation length of the crack increased from 1.27 m to 10.92 m at an average rate of 0.386 m/s, and the propagation width of the crack increased from 1.09 mm to 3.53 mm at an average rate of 0.098 mm/s. From 30 to 100 s, the propagation length of the crack increased from 10.92 m to 17.60 m at an average rate of 0.095 m/s, and the propagation width of the crack increased from 3.53 mm to 5.45 mm at an average rate of 0.027 mm/s.

By the statistical analysis of the simulation results, the change laws of the propagation length and maximum width of the prefabricated fracture II under double-borehole hydraulic fracturing with time were obtained. According to Fig. 7, the maximum propagation length of the prefabricated fracture II was 17.6 m at $t = 100$ s, which decreased by 25.1% than that of single-borehole hydraulic fracturing. The maximum propagation width of the prefabricated fracture II was 5.453 mm, which increased by 62.5% than that of hydraulic fracturing with single-borehole hydraulic fracturing. Due to the stress interference of the

(a) Nephogram of maximum fluid pressure   (b) Nephogram of fluid pressure at $t = 100$s
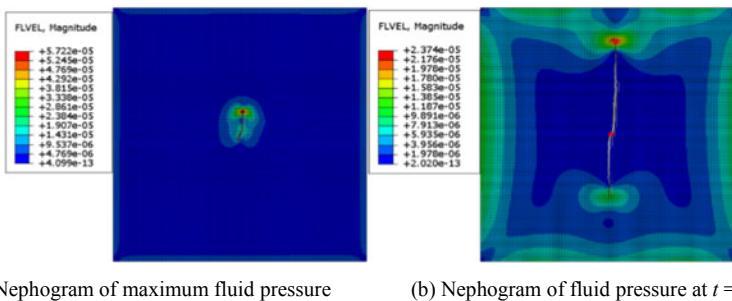
Fig. 4 Fluid pressure distribution at the prefabricated fracture tip under single-borehole hydraulic fracturing

prefabricated fracture I to prefabricated fracture II, the propagation rate of pre-fabricated fracture II under double-borehole hydraulic fracturing is increased by 10.49%, compared with that under single-borehole hydraulic fracturing.

According to the variation law of damage dissipation energy of rock with time in Fig. 8, the stage before $t = 7$ s was the energy accumulation period of prefabricated fractures I and II. The stage from $t = 7$ s to $t = 90$ s was the step-like growth period of energy. And the stage from 90 to 100 s was the steady development period where the energy was almost constant with the increase of time. The maximum damage dissipation energy is 163.1 kJ in the whole process, which is at least 3.5 times of that under double-borehole hydraulic fracturing, thereby the fragmentation

Fig. 5 Change curve of damage dissipation energy of limestone under single-borehole hydraulic fracturing with time

(a) $t = 5$ s   (b) $t = 30$ s   (c) $t = 100$ s

(d) $t = 5$ s   (e) $t = 30$ s   (f) $t = 100$ s
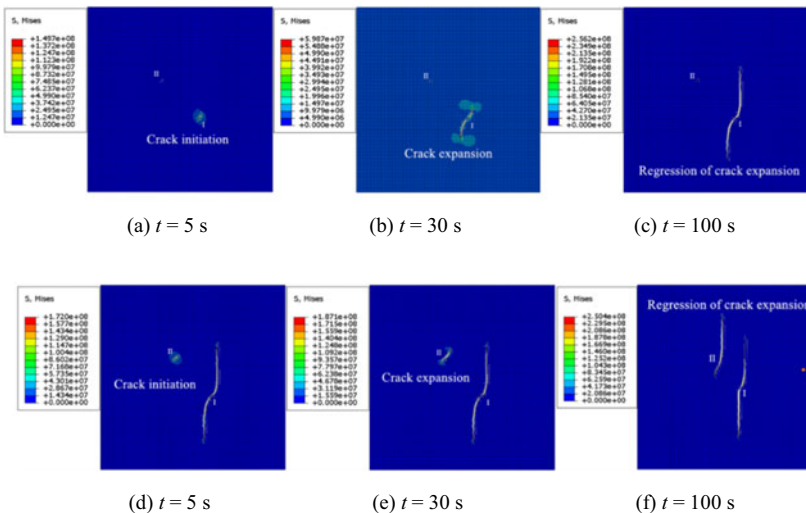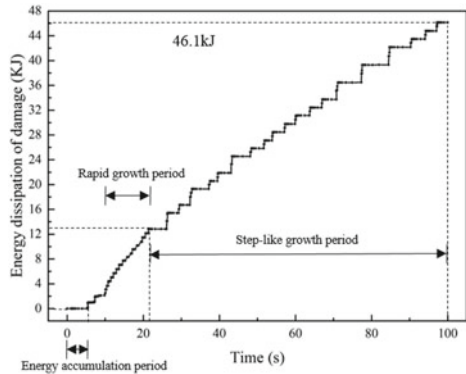
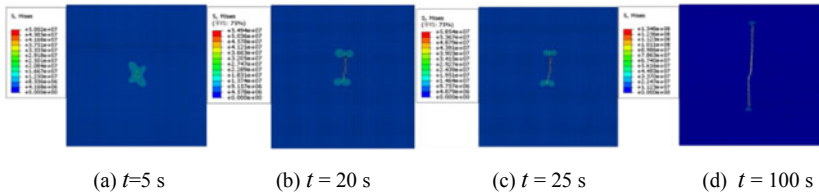**Fig. 6** The prefabricated fracture propagation nephogram under double-borehole hydraulic fracturing at different time. (The Fig. 6(a), Fig. 6(b) and Fig. 6(c) are the nephogram at the first stage, the Fig. 6(d), Fig. 6(e) and Fig. 6(f) are the nephogram at the second stage)

degree of rock under double-borehole hydraulic fracturing is larger than that under single-hole hydraulic fracturing. This is because the hydraulic fracturing of borehole I has the promoting effect to the fracturing of borehole II. In the process of double-borehole hydraulic fracturing, due to the stress interference between the two cracks, there will be a stress interaction zone between the two cracks, which benefits the prefabricated fractures to propagate in different directions and different degrees. Therefore, under the double-borehole hydraulic fracturing, the damage dissipation energy of prefabricated fracture II are greatly increased, and the fracturing efficiency is greatly improved under double-borehole hydraulic fracturing, compared with the single-borehole hydraulic fracturing.

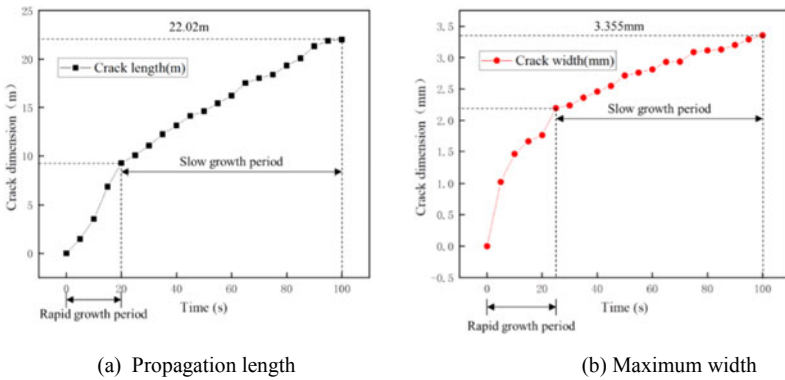(a) Propagation length   (b) Maximum width
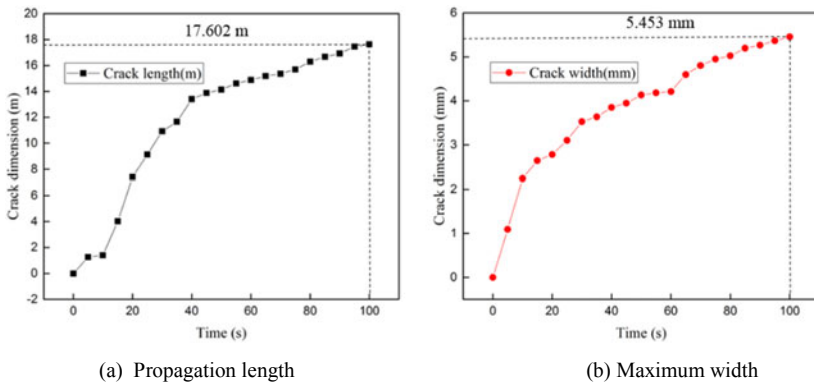
Fig. 7 The change curve of propagation length and maximum width of prefabricated fracture II under double-borehole hydraulic fracturing with time

Fig. 8 Change curve of damage dissipation energy of rock with double-borehole prefabricated fracture with time
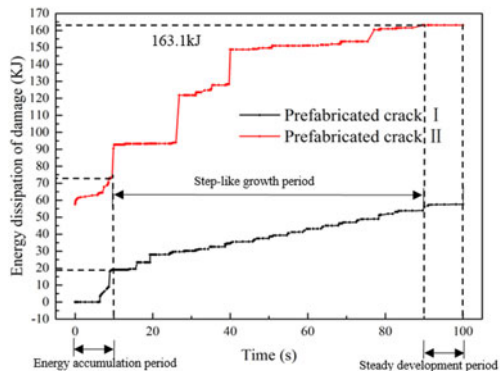
4   Conclusion

On the basis of XFEM, this paper established the numerical models of single-borehole and double-borehole hydraulic fracturing medium-hard rock, and comparatively analyzed the evolution law of crack and the damage dissipation energy of rock under the single-borehole and double-borehole hydraulic fracturing. The main conclusion is concluded as follows:

(1) Under the single-borehole and double-borehole hydraulic fracturing, the propagation lengths and widths of the prefabricated fracture as well as the damage dissipation energy of rock increase with the increase of time. And the crack propagation length and width experience both the rapid growth period and the slow growth period, and the damage dissipation energy of rock mainly appears the step-like increase continuously.

(2) Because of the promotion effect of early borehole hydraulic fracturing to later borehole hydraulic fracturing, under the double-borehole hydraulic fracturing, the propagation length of single prefabricated fracture is reduced by 25.1%, and the propagation width is increased by 62.5%, compared with single-borehole hydraulic fracturing. By the comparative analysis of damage dissipation energy of rock under hydraulic fracturing of the single and double boreholes, the damage dissipation energy of single prefabricated fracture under the double-borehole hydraulic fracturing is increased by at least 3.5 times than that of single-borehole hydraulic fracturing. It can be concluded that the fracturing efficiency of single prefabricated fracture under the double-borehole hydraulic fracturing is higher than that under single-hole hydraulic fracturing.

References

1. Muther, T., Khan, M.J., Chachar, M.H., et al.: A Study on designing appropriate hydraulic fracturing treatment with proper material selection and optimized fracture half-length in tight multilayered formation sequence. SN Appl. Sci. **2**(5), 1–12 (2020)
2. Legarth, B., Huenges, E., Zimmermann, G., et al.: Hydraulic fracturing in a sedimentary geothermal reservoir: Results and implications. Int. J. Rock Mech. Min. Sci. **42**, 1028–1041 (2005)
3. Fomin, S., Hashida, T., Shimizu, A., et al.: Fractal concept in numerical simulation of hydraulic fracturing of the hot dry rock geothermal reservoir. Hydrol. ProcHydrological Processesrnesses **17**(14), 1–15 (2003)
4. Damjanac, B., Cundall, P.: Application of distinct element methods to simulation of hydraulic fracturing in naturally fractured reservoirs. Comput. Geotech. **71**, 283–294 (2016)
5. Nasehi, M.J., Mortazavi, A.: Effects of in-situ stress regime and intact rock strength parameters on the hydraulic fracturing. J. Petrol. Sci. Eng. **108**, 211–221 (2013)
6. Wei, T., Peichao, L., Yan, et al.: Numerical simulation of sequential, alternate and modified zipper hydraulic fracturing in horizontal wells using XFEM. J. Petrol. Sci. Eng. (183), 106251 (2019)

7. Shimizu, H., Murata, S., Ishida, T.: The distinct element analysis for hydraulic fracturing in hard rock considering fluid viscosity and particle size distribution. Int. J. Rock Mech. Min. Sci. **48**(5), 712–727 (2011)

8. Francisco, C., Deane, R., do Amaral Vargas Jr., E.: An XFEM implementation in Abaqus to model intersections between fractures in porous rocks. Comput. Geotech. (112), 135–146 (2019)

9. Xue, B., Zhang, G., Wu, H., et al.: Three-dimensional numerical simulation of hydraulic fracture in oil wells. J. Univ. Sci. Technol. China **38**(11), 1322–1325 (2008). 1347. (in Chinese)

10. Ji, S.-H., Koh, Y.-K., Kuhlman, K.L., et al.: Influence of pressure change during hydraulic tests on fracture aperture. Ground Water (51), 298–304 (2013)

11. Tatjana, S., Andrey, P., Leonid, R., et al.: Development of the impermeable membranes using directional hydraulic fracturing. In: Symposium of the International Society for Rock Mechanics, no. 191, pp. 520–524 (2017)

12. Xing, J., Hou, Li., Zhang, T.: Numerical simulation of the whole process of hydraulic fracturing under different directional fractures. Water Conservancy Technical Supervision (6), 175–179 (2019). (in Chinese)

13. Liang, B., Yue, L., Sun, W.: The influence of shale mineral composition on crack growth: numerical simulation. Marine Petrol. Geol. 24(4), 97–101(2019). (in Chinese)

14. Zhang, R., Wang, Q., Zhang, Z., et al.: Research of ABAQUS numerical simulation of 3D fracture propagation in hydraulic fracturing process. Oil Drill. Prod. Technol. **34**(6), 69–72 (2012). (in Chinese)

15. Wang, L., Gao, Q.: Fragmentation distribution prediction of rock based on damage energy dissipation. Chin. J. Rock Mech. Eng. **26**(6), 1202–1211 (2007). (in Chinese)

16. Xia, C., Ju, Y., Xie, H.: Numerical analysis of propagation of explosion wave and energy dissipation in tunnel and surrounding rockmass. J. Ballistics **17**(4), 1–5 (2005). (in Chinese)

17. Li, S., Ren, Y., Fan, C., et al.: Influence of spacing between boreholes on hydraulic fracturing to enhance coal seam gas extraction. J. Safety Sci. Technol. **14**(1), 70–76 (2018). (in Chinese)

The Use of Reverse Design Tool Version 1.0 in the Design of Five-Section Directional Wellbore Track

Minglei Li

Abstract Directional well trajectory design is an essential part of a drilling plan process and requires rigorous mathematics to ensure proper execution of the drill path that would accommodate safety of the tools being used to drill the planned well. Various well trajectory design applications have been developed such as the Halliburton COMPASS which aid in constructing well trajectory designs. In the early stage of drilling design for directional wells and horizontal wells, geological departments only provide drilling target coordinates according to geological research results. It is necessary to calculate well position coordinates in reverse according to target points and other geological engineering constraints. Due to the common track design software such as Compass does not have the function of reverse calculation of drilling track, therefore, it is necessary to carry out trial calculation repeatedly to determine reasonable wellhead coordinates, hence as a result of this research a computer application was developed adopting Minimum Curvature Method, named WellTRDT (Well Trajectory Reverse Design Tool) version 1.0, which could be used to enlighten engineers further on the Design of Five-Section Directional Wellbore Track and realize the optimization calculation of fast well position coordinates and drilling trajectory and improves the work efficiency. Combining this software with trajectory design software can improve work efficiency. A survey data ran by this software and COMPASS was used to validate the ability of the WellTRDT v. 1.0 application, of which results from both applications fairly matched.

Keywords Five-stage directional wells · Borehole trajectory design · Minimum curvature method

Abbreviations

R_z	Curvature Radius of the First Increasing Slope, m
R_{zz}	Curvature Radius of the Second Increasing Slope, m

M. Li (✉)
Engineering, Procurement and Equipment Management Department, China National Petroleum Corporation, Beijing 100007, China
e-mail: lminglei@cnpc.com.cn

γ_1 Slope Building Rate of the First Increasing Slope Sectio, deg/30 m
γ_2 Slope Building Rate of the Second Increasing Slope Section, deg/30 m
α_a Well deviation angle of the first stable section, deg
α_b Well deviation angle of the second stable section, deg
D_a Vertical Depth of Target A, m
D_b Vertical Depth of Target B, m
ΔD_{mm} Distance of Targets A and B, m
(cx_2, cy_2) Centroid coordinates of the second inclined segment, m
H_L Stable slope length, m
D_L Depth of limestone roof, m
D_{kop1} Inclination point 1 depth, m
D_{kop2} Inclination point 2 depth, m
S_t Pre-target displacement, m

1 Introduction

The design process of directional well and horizontal well is as follows: firstly, in the early stage of design, the geological department provides the coordinates of drilling target points and geological constraints according to the geological research results, points are marked to determine the likely course a well might take. (such as the requirement of Tarim Oilfield that the second opening should be completed at the top of limestone 10 m). Secondly, drilling engineering design department needs to calculate well position coordinates in reverse according to targets and geological constraints and engineering constraints provided by geological department, such as dogleg degree, lithology of tilting point, etc. In the industry, one of the major applications used for Borehole trajectory design is the Halliburton's Landmark COMPASS. It is a very robust application, but it does not have the function of calculating well position coordinates in reverse according to targets. Therefore, in practice, it is often necessary to make repeated trial calculations to determine the reasonable wellhead coordinates by drawing the possible route with a pencil. From the sketches, idea of the wellhead coordinates to be tried are calculated for to provide the design parameter. Once the target or other conditions are adjusted, such work has to be repeated, the dense calculations involved with providing the appropriate directional well trajectory is very time-consuming and laborious, and often the results are not optimal. Therefore, the special applications are developed for determining and drawing out records for a pre-planned well trajectory.

There are many studies on directional well and horizontal well trajectory design [1, 2] and mature methods for calculating the common trajectory parameters. But the existing methods basically take the well position coordinates as known parameters, and there is no design tool for the unknown well position coordinates.

Aiming at this practical problem, this paper firstly analyses the required drilling trajectory according to the design method of drilling trajectory. For the most common dual target design, the double-incremental drilling trajectory is adopted to establish the corresponding reverse calculation method. Based on this method, the software WellTRDT (Well Trajectory Reverse Design Tool l) v1.0 was designed for the sole purpose to assist in reverse calculating well position coordinates, key control points (deviation point, dogleg degree, well deviation angle) for 2D Five-stage directional well trajectory plan and the reverse optimization design are realized. By inputting the design results into the COMPASS software for forward design, the design results are completely consistent, which verifies the correctness of the software. The practical application in Tarim Oilfield proves that the method and software have high accuracy and practicability.

2 Theory and Definitions

Purpose of economics was the main reason why wells were drilled directionally in the early days of its practice. Though as time went on the advance of directional drilling technology justified more for its purpose other the economical fact. At present a directional well has to be properly planned to ensure it fits in to the available or level of technology being practiced, economical (minimum expenditure to be met), entirely safe for operating personnel and the environment, and must be an efficient conduit for maximum hydrocarbon recovery. Few practical reasons why drilling directionally is so as;

- To sidetrack out of an existing well
- Develop Multilateral wells using one drill pad
- To use an single well string to cut across pockets of layered reservoirs
- To avoid existing well collision
- To avoid unnecessary exposure to unstable formations
- To explore reservoir beneath a surface obstruction.

Directional drilling is the process of directing the wellbore along some trajectory to a predetermined target [3].

Several programmed well planning tools are used to estimate, forecast and predict the best well choice or pattern to consider for the actual plan. The industry leader within well planning software is the Halliburton's landmark COMPASS.

Key points need to be known when using COMPASS design well trajectory (KOP, dogleg degree, well deviation angle), usually geological design gives only target information which not enough to apply COMPASS to design. Through inverse calculation tool developed in this paper (taking the 2D Five-Section Track Design as an Example), the key parameters are obtained, then by inputting the design results into the COMPASS software for forward design, the target information obtained is consistent with known information and specific trajectory design also obtain with COMPASS.

We shall briefly define some important key points needing inverse computation associated with directional well trajectory:

1. **Survey Station**: This is could be any point along the well path that is used as a reference point to determine the direction and distance of the well path from that particular point. There are several survey stations along any well path. Survey stations could be iterated of equal intervals or otherwise across the entire well distance.
2. **Kick-Off point** (KOP): This is a vertical distance from the surface when the well path begins to deviate in an inclined manner from the original path [4]. Highlighted the necessity for proper kick off points for wells, considering that;
a. Is easier to kick-off in shallow formations than in deep formation
b. The more distant targets have the shallowest kick-off point
c. Kick-off points are initiated in stable formations that would not cause drilling problems.
   In actual practice, kick-off points for directional wells are initiated between 2500ft to 3000ft from the surface.
3. **Well Target**: This is a predetermined target whereby the proposed well plan path hits, reaches or ends. Proper well survey calculations are often utilized to ensure high degree of closeness or accuracy to the target. Such survey calculations [5–8] could be;
4. **KOP**: Depth of build-up point, Two wellhead coordinates (XY), Constraint conditions limestone top depth

The key parameters are calculated by Minimum Curvature Method in Software. The minimum curvature method has emerged as the accepted industry standard for the calculation of 2D directional surveys. Using this model, the well's trajectory is represented by a series of circular arcs and straight lines. Collections of other points, lines, and planes can be used to represent features such as adjacent wells, lease lines, geological targets, and faults. The relationships between these objects have simple geometrical interpretations, making them amenable to mathematical treatment.

3   Description of Computer Model (WellTRDT version 1.0)

Reverse Design Model of Five-Stage Borehole Trajectory
WellTRDT was developed for the purpose of solving reverse design key points parameters of directional well trajectories.

Dual incremental wellbore trajectory is a common five-stage wellbore trajectory, which is widely used in Tazhong Oilfield. The commonly used dual incremental wellbore trajectory is shown in Fig. 1. Target A and target B, the depth of limestone roof are given, it is required that the second opening drilling completion ten meters
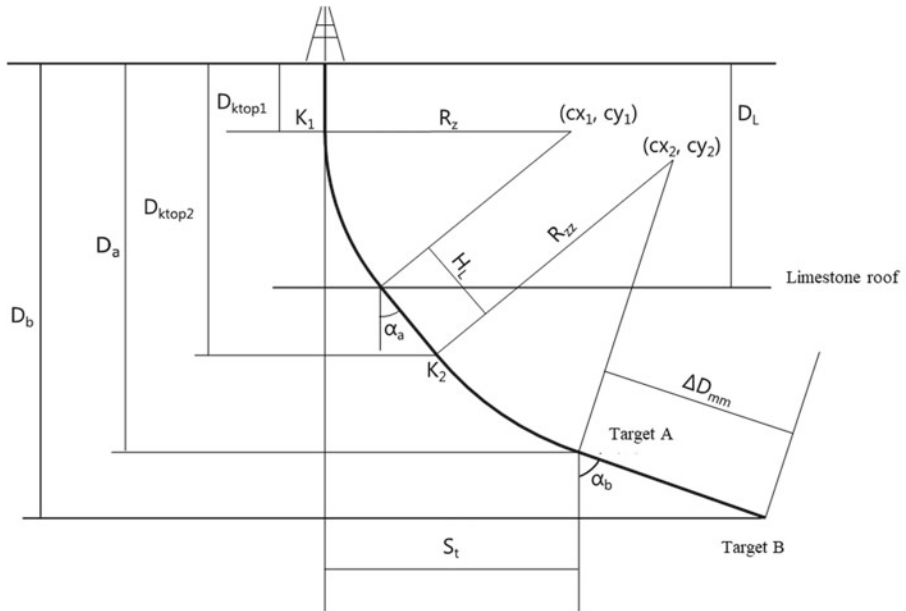
Fig. 1 Five-Section Track Diagram

below the top of limestone and steady inclination 20 m after triple opening of casing shoes, then start to increase tilt, finally stabilize tilt from target A to target B. Based on the above information, it needs iInverse calculate two depth of KOP1 KOP2 (γ_1, γ_2), well deviation (α_a) build-up rate (D_{kop^1}, D_{kop^2}), wellhead coordinates (X,Y).

Previous dual-incremental trajectory calculation models need well coordinates to be known, which is not suitable for this problem. Following is the calculation model of well position coordinates inversely calculated from lower target data using the minimum curvature method, as shown in Formula 1–8.

$$R_z = \frac{30 \times 180}{\pi\gamma_1} \tag{1}$$

$$R_{zz} = \frac{30 \times 180}{\pi\gamma_2} \tag{2}$$

$$\alpha_b = \arcsin\left(\frac{D_b - D_a}{\Delta D_{mm}}\right) \tag{3}$$

$$(cx_2, cy_2) = (R_{zz}\cos(\alpha_b), D_a - R_{zz}\sin(\alpha_b)) \tag{4}$$

$$\alpha_a = 2 \arctan \left(\frac{-R_{zz} + \sqrt{R_{zz}^2 + H_L^2 - (D_L - cy_2)^2}}{H_L - D_L + cx_2} \right) \tag{5}$$

$$D_{kop1} = D_L - R_z \sin(\alpha_a) \tag{6}$$

$$D_{kop2} = D_L + H_L \cos(\alpha_a) \tag{7}$$

$$S_t = R_{zz} \cos \alpha_a + H_L \sin \alpha_a + R_z (1 - \cos \alpha_a) \tag{8}$$

WellTRDT Development

The flow chart of WellTRDT is illustrated in Fig. 2 and further explanations are provided.

Fig. 2 Computing flow chart
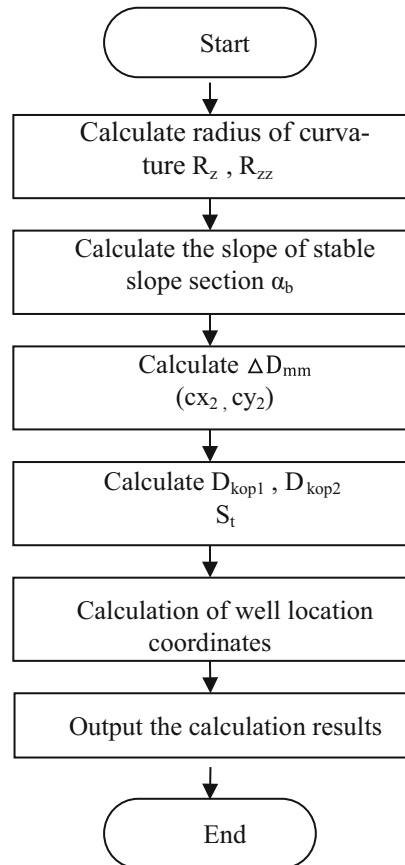
4 Validation of WellTRDT Program

In this section, we took an example trajectory design processed by WellTRDT (adopts Metric units) and COMPASS, Firstly, by inputting parameters design targets (see Table 1) into the software, borehole trajectory parameters can be calculated immediately, the calculation results are as follows in Table 1. Inputting the final key point parameters done by WellTRDT into COMPASS to Validate trajectory design, the calculation results are as follows in Fig. 2. Comparing the above calculation results, we can see that except for rounding reasons, the second decimal point is slightly different, and the two software calculation results are completely consistent, which shows that the design results of this software are correct. Table 2 shows the detailed design results furthermore by using COMPASS.

The software has been applied in Tarim Oilfield. In practical application, the workload of manual repeated calculation has been greatly reduced, and the optimized well coordinates can be quickly calculated, which has great practical value (Figs. 4 and 5).

Table 1 Target parameters

Target	X(m)	Y(m)	Target depth (m)	Limestone top depth(m)
Target A	9633	6577	6215	6075
Target B	9600	6417	6230	6080

Note: Target coordinates are relative coordinates

Table 2 COMPASS forward calculation results

Desc	MD (m)	CL (m)	Inc (°)	Azi (°)	TVD (m)	NS (m)	EW (m)	Dogleg (°/30 m)	C.Azi (°)	C.Dist (m)	Map N (m)	Map E (m)
Well	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9705.85	6930.23
Kop1	5857.56	5857.56	0.00	0.00	5857.56	0.00	0.00	0.00	0.00	0.00	9705.85	6930.23
Hold1	6104.45	246.89	49.38	258.35	6075.00	−20.19	−97.90	6.00	258.35	99.96	9685.66	6832.33
Kop2	6194.45	90.00	49.38	258.35	6133.60	−33.99	−164.81	0.00	258.35	168.28	9671.86	6765.42
Target A	6406.70	212.26	84.75	258.35	6215.00	−72.85	−353.23	5.00	258.35	360.66	9633.00	6577.00
Target B	6570.76	164.05	84.74	258.35	6230.00	−105.85	−513.23	0.00	258.35	524.03	9600.00	6417.00
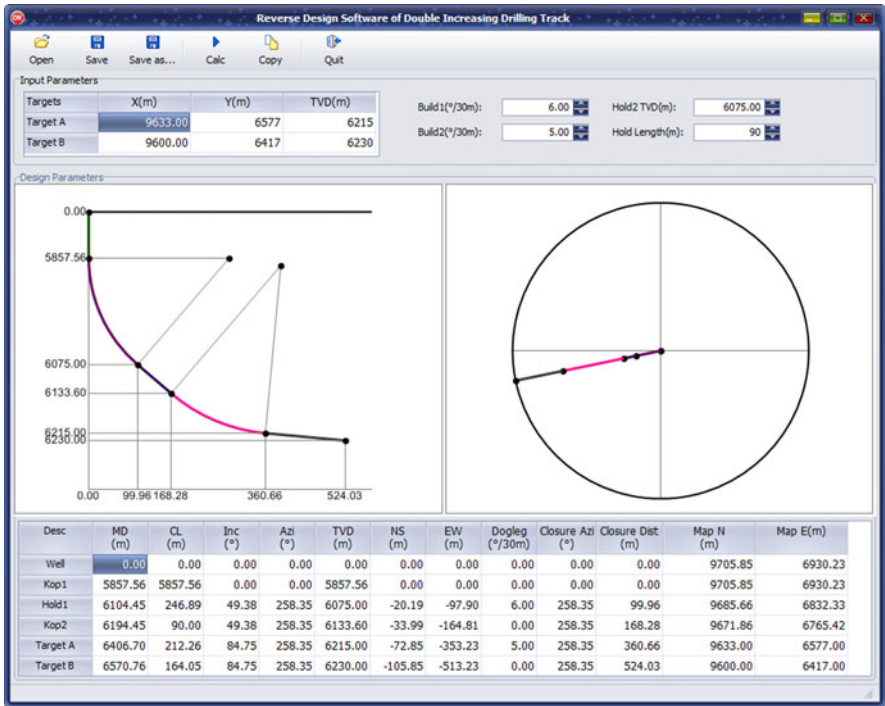
Fig. 3 WellTRDT running interface
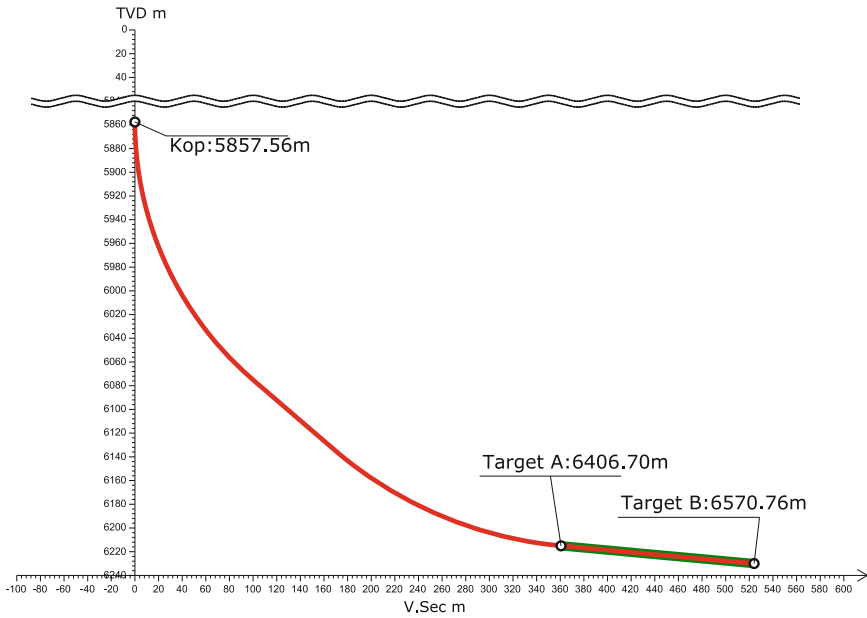
Fig. 4 Vertical projection
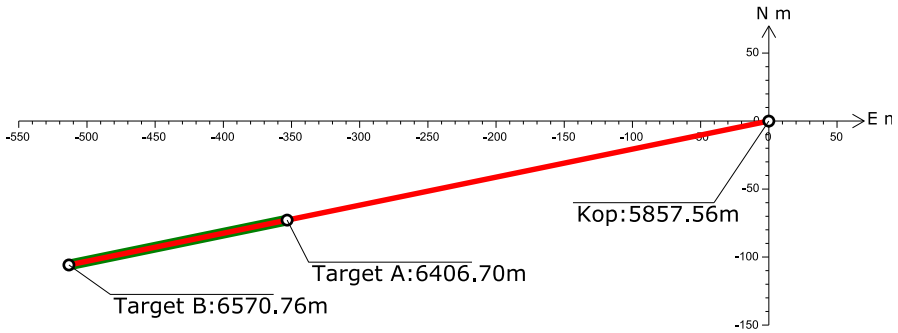
Fig. 5 Horizontal projection

5 Conclusions

Having developed and tested the WellTRDT version 1.0 application on double-increment borehole trajectory reverse designs, as shown Fig. 3, highlights its ability which is useful in double-increment borehole trajectory reverse design. WellTRDT survey calculation makes use of "Minimum Curvature Method" method. By comparing with the authoritative design software COMPASS, the results are completely consistent, which proves the correctness of the software. Through the practical application in Tarim Oilfield, the software greatly reduces the workload of calculating well position and wellbore trajectory design, improves the efficiency of design work, and has great practical value. This paper only establishes the reverse calculation model of dual-increment wellbore trajectory. WellTRDT is still under further development and in future time will consist of more robust and complex mathematical models and calculations.

References

1. Mason, C.M., Taylor, H.L.: A Systematic Approach to Well Surveying Calculations, SPEJ **12** (06), 474-488 (1971)
2. Sawaryn, S.J., Sanstrom, W., McColpin, G.: The management of drilling engineering and well services software as safety critical systems. In: Paper SPE 73893 Presented at the 2002 SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Kuala Lumpur, 20–22 March (2002)
3. Bourgoyne Adam Jr., T., Millheim Keith, K., Chenevert Martin, E., Young Jr., F.S.: Applied Drilling Engineering. Society of Petroleum Engineers, Richardson, Texas, pp. 351–66 (1986)

4. Farah, F.O.: Directional well design, Trajectory and survey calculations, with a case study in Fiale, Asalrift, Djibouti. Geothermal Training Programme (2013)
5. Craig Jr., J.T., and Randall, B.V.: Directional drilling survey calculation. Pet. Eng. **73**, 103081 (1976)
6. Taylor, H.l., Mason, C.M.: A systematic approach to well survey calculations. Soc. Pet. Eng. J. **12**(06), 474–88 (1972)
7. Wilson, G.J.: An improved method for computing directional survey. J. Pet. Technol. **20**(08), 871–876 (1968)
8. Amorin, R., Broni-Bediako, E.O.: Application of minimum curvature method to well path calculations. Res. J. Appl. Sci. Eng. Technol. **2**(7), 679–686 (2010)

Three-Dimensional Simulation of Ice Crystal Trajectory with State Change Around Rotor Blade of Axial Fan

Koichiro Hirose, Koji Fukudome, and Makoto Yamamoto

Abstract Ice crystal icing occurs in the compressors of jet engines at a static mainstream temperature of approximately 30 °C. Therefore, it is important to consider the melting of the ice crystals (i.e., state change) because ice crystals fly in a high-temperature environment. In this study, to accurately clarify how ice crystals impinge on a rotor blade, we conducted a trajectory simulation considering the state change of ice crystals flying around the rotor blade of an axial fan. The ice crystal trajectories were tracked using a Lagrangian approach with a one-way coupling. The state change of the ice crystals was evaluated from the energy balance of the heat received from the surrounding fluid, the heat passing through the ice crystals through heat conduction, and the heat of the evaporation or condensation on the surface. The results show that the longer the ice crystals fly, the more they are affected by the phase transition, which becomes more remarkable as the sizes of the ice crystals decrease. In addition, the water content of ice crystals impinging near the tip side is particularly high in the entire rotor blade.

Keywords Ice crystal icing · Particle trajectory · Ice particle melting · Axial fan

1 Introduction

In general, aircraft icing is a phenomenon in which super-cooled droplets in an atmosphere impinge on a solid surface to form an ice layer. In particular, the icing phenomenon in a jet engine is a threat to the safe navigation of an aircraft because it

K. Hirose
Graduate School of Mechanical Engineering, Tokyo University of Science,
6-3-1, Niijuku, Katsushika-ku, Tokyo, Japan

K. Fukudome · M. Yamamoto (✉)
Department of Mechanical Engineering, Tokyo University of Science,
6-3-1, Niijuku, Katsushika-ku, Tokyo, Japan
e-mail: yamamoto@rs.tus.ac.jp

deteriorates the aerodynamic performance by changing the blade/vane shapes, and can cause serious accidents owing to a detachment of the ice blocks [1].

In recent years, ice crystal icing has also been reported [2]. This is an icing phenomenon that occurs on low-pressure compressor vanes by ice crystals with a mainstream static temperature of approximately 30 °C. Therefore, the icing factor and the environment are significantly different from an ordinary icing by super-cooled droplets. In addition, it is difficult to reproduce ice crystal icing experimentally in terms of the safety and costs because such event occurs inside the jet engine during flight, and it is therefore necessary to reproduce and predict this phenomenon through a numerical simulation.

In previous studies [3], ice crystal icing simulations have been conducted on a two-dimensional vane. However, the impingement behavior of ice crystals in the span direction for a three-dimensional geometry have yet to be clarified. In addition, to improve the accuracy of the ice crystal icing prediction, it is necessary to clarify the behavior of the ice crystals before the impingement on the rotor blades as well as stator vanes.

The purpose of this study is to develop an ice crystal trajectory code corresponding to the three-dimensional flow field around a rotor blade. We conducted an ice crystal trajectory simulation for the rotor blade of an axial fan. The state change of the ice crystals was considered to accurately reproduce the actual phenomenon.

2   Numerical Method

The numerical simulation consists of three steps: the generation of computational grids, the computation of the flow field, and the computation of the ice crystal trajectories and the acquisition of the impingement characteristics of the ice crystals on the rotor blade surface. The details of these steps are briefly described below.

2.1   Flow Field Computation

The flow field is assumed to be three-dimensional compressible and turbulent. The governing equations are the continuity, the Navier–Stokes, and the energy equations with the Kato-Launder k-ε model [4]. The second-order upwind TVD scheme [5] is applied to the inviscid terms, and the second-order central difference scheme is applied to the viscous terms. The LU-ADI scheme [6] is used for the time integration.

2.2 Ice Crystal Trajectory Computation

The ice crystal trajectories are tracked using the Lagrangian approach. As a pre-condition, the ice crystal shape is assumed to be a sphere, and the break-up and collisions of the ice crystals are not considered before impingement on the rotor blade. In addition, the ice crystals do not affect the flow field (i.e., one-way coupling) because of the low volume fraction of ice crystals in the atmosphere. Based on the above, we employ the simplified Basset-Boussinesq-Oseen (B-B-O) equation as the governing equation, which assumes that only the drag and centrifugal and Coriolis forces act on the ice crystals.

2.3 State Change Computation

Flying ice crystals are affected by the surrounding fluid, and melt, evaporate, and condense. In this study, as shown in Fig. 1, we assumed a model [7] in which the melted liquid phase covers the surface of the ice particles uniformly and concentrically, and evaporation and condensation occurs on the surface of the ice crystals. The water film radius r_w, the ice particle radius r_i, and the water film surface temperature T_w in Fig. 1 are evaluated using Eqs. (1)–(3). The state of the ice crystals is based on the energy balance of the heat received from the surrounding fluid Q_1 [8], the heat passing through an ice crystal Q_2 [9], and the heat of evaporation or condensation on the surface Q_3 [9].

$$\frac{dr_w}{dt} = \frac{D_c M_w}{\rho_w R} \left(\frac{P_{s,\infty}}{T_\infty} RH - \frac{P_{s,w}}{T_w} \right) \frac{1}{r_w} \tag{1}$$

$$\frac{dr_i}{dt} = -\left\{ k_\infty (T_\infty - T_w) + \frac{D_c M_w}{R} \left(\frac{P_{s,\infty}}{T_\infty} RH - \frac{P_{s,w}}{T_w} \right) L_{ev} \right\} \frac{r_w}{r_i^2 \rho_i L_m} \tag{2}$$

$$k_w r_i \frac{T_w - T_0}{r_w - r_i} = k_\infty (T_\infty - T_w) + \frac{D_c M_w}{R} \left(\frac{P_{s,\infty}}{T_\infty} RH - \frac{P_{s,w}}{T_w} \right) L_{ev} \tag{3}$$

Fig. 1 Schematic of the ice crystal model
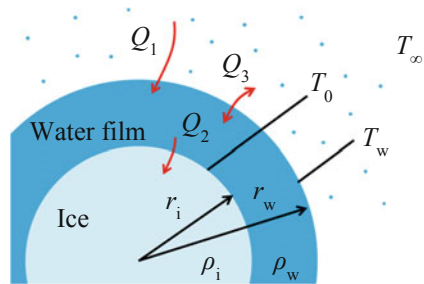
Here, T_∞ is the temperature of the flow fluid, D_c is the diffusion coefficient, M_w is the mass of water, R is the universal gas constant, RH is the relative humidity, k_∞ and k_w are the heat conductivity of the air and water, L_{ev} and L_m are the latent heat of evaporation and melting, and ρ_i and ρ_w are the densities of ice and water, respectively. In addition, $P_{s,\infty}$ and $P_{s,w}$ are the saturated vapor pressures at T_∞ and T_w.

3 Computational Target and Conditions

In this study, we conducted an ice crystal trajectory simulation for a commercial axial fan (Showa Denki Co., Kairyu series A2D6H-411), as shown in Fig. 2. The computations were conducted for only one blade in consideration of the periodicity. The computational grids are generated using the overset-grid method to reproduce the impingement characteristics on the rotor blade in detail, and the total number of grid points is approximately 6 million. For the boundary conditions of the inlet, the total temperature, the total pressure, the flow angle, and turbulent quantities are fixed, and the Mach number is extrapolated. The periodic boundary condition is applied in the rotational direction. At the outlet, the static pressure is specified and other variables are extrapolated. The computational conditions are listed in Table 1. The rotational speed and the mass flow rate are set according to the experiment conducted by Murooka et al. [10], and the other parameters are set to an environment in which the ice crystal icing occurs. In this study, we evaluated the impingement characteristics of the ice crystals for the rotor blade with and without the state change in four cases of median volume diameters (MVDs) of 25, 50, 100, and 200 μm.
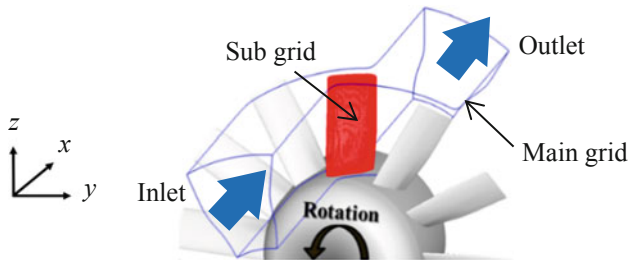
Fig. 2 Schematic of computational domain of axial fan

Table 1 Computational conditions

Rotational speed	[rpm]	1800
Mass flow rate	[kg/s]	5.21
Inlet static temperature	[K]	303.15
Partial pressure of water	[Pa]	2000
Ice water content	[g/m^3]	7.0
Median volume diameter	[μm]	25, 50, 100, 200

4 Results and Discussion

4.1 Collection Efficiency

Figure 3 shows the collection efficiency of the ice of the ice crystal on the pressure surface with and without the state change for MVD of 25 and 200 μm. In this study, we defined the collection efficiency as the impingement mass per unit time and area. The leading and trailing edges are indicated as "L.E." and "T.E.," respectively. For MVD of 25 μm, when the state change is not considered (Fig. 3 (a)), a large amount of ice impinges with the tip side, but when the state change is considered (Fig. 3 (b)), a high impingement distribution on the tip side is not confirmed. However, when the state change is considered, the number of the ice crystals impinging with the rotor blade is 1.33% higher, and the mass of the ice impinging with the tip side decreases. This is because the ice crystals melt under the influence of the flow field during flight. By contrast, for MVD of 200 μm (Fig. 2 (c) and (d)), no significant difference appears between with or without the state change.

4.2 Impinging Mass

Figure 4 shows the mass and water content of the ice crystals impinging on the entire rotor blade per unit of time. When the state change is considered, the mass of the ice and the water film of the ice crystals are also included. The water content is defined as the ratio of the mass of the water film to the mass of the ice crystals. The mass of the impinging ice crystals is smaller as the sizes of the ice crystals become smaller regardless of whether the state change occurs. Considering the state change, the mass of the impinging ice crystals is slightly larger in the cases of MVD of 25 and 50 μm because the moisture in the atmosphere is condensed and taken into the water film of the ice crystals. By contrast, the mass of the impinging water film increases as the sizes of the ice crystals decrease because the amount of heat per unit of mass obtained from the flow field increases as the ice crystals become smaller. In addition, the water content also increases as the ice crystals decrease in size, and the water content at MVD of 25 μm is significantly higher than that of the other ice crystal sizes.
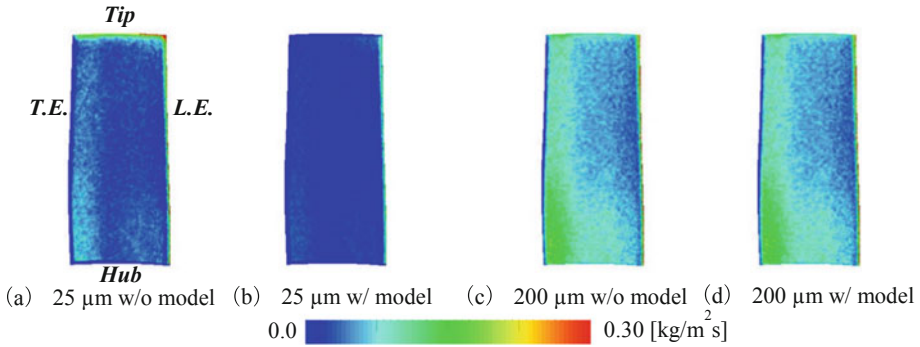
Fig. 3 Collection efficiency on the pressure surface under different MVD conditions
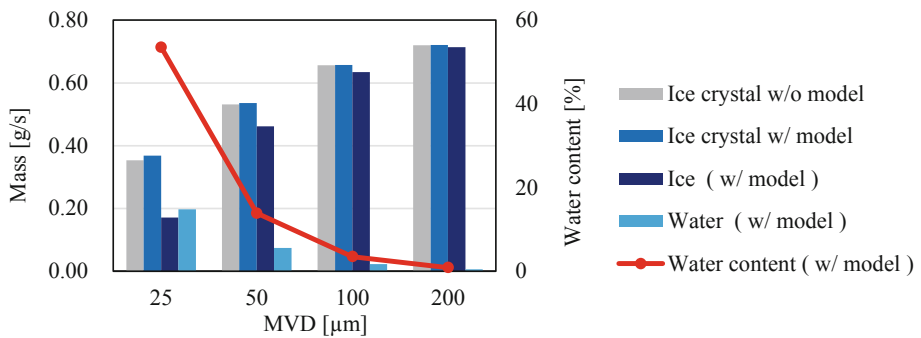
Fig. 4 Impinging ice crystal mass and water content on the blade under different MVD conditions

4.3 Water Content

Figure 5 shows the water content of the ice crystals impinging with the pressure surface in the cases of MVD of 25, 50, 100, and 200 μm. For the color contour range, the water content of the ice crystals that impinge with a 90% span or less of the rotor blade in each cases are set to the minimum and maximum values. The water content increases toward the trailing edge from the leading edge for all ice crystal sizes. This is because the flight time of the ice crystals before impinging on the rotor blade is longer for the ice crystals impinging on the trailing edge, i.e., the melting time of the ice crystal increases. The deviation in the water content between the leading edge and the trailing edge increases as the ice crystal sizes decrease, and is 0.24% at MVD of 200 μm and 13.4% at MVD of 25 μm. However, when compared with the same chord position, the water content on the tip side is higher, which is remarkable at MVD of 25 μm and MVD of 50 μm, as discussed in the following sections.

4.4 Impingement Characteristics at the Leading Edge of the Rotor Blade

Figure 6 shows the time variation of the static temperature of the flow field around the flying ice crystals with the state change at MVD of 25 and 200 μm. The ice crystals are injected from five different span positions of 15 to 99%, as shown in Fig. 6, all of which impinge with the leading edge of the rotor blade. In each case, the static temperature increases sharply before the impingement around the stagnation point, where the initial ice crystal temperature is approximately 303 K regardless of the span position. Focusing on the flight time until the impingement, for MVD of 25 μm, the ice crystal injected at a 99% span location takes about twice as long as the other ice crystals. By contrast, for MVD of 200 μm, no differences are shown between the ice crystals injected from the 99% span and the other ice crystals. Therefore, the reason why MVD of 25 μm shows a higher water content on the tip side is considered to be due to the long flight time and the ease of melt of the ice crystals.

4.5 Ice Crystal Trajectory Near the Tip Side

Figure 7 shows the ice crystal trajectories within the 99% span. The blue and red lines indicate the trajectory for MVD of 25 and 200 μm, respectively. Figure 7(a) shows the axial velocity distribution in the 1/5 pitch section from the periodic boundary, and Fig. 7(b) shows the relative velocity vectors of the flow field within the 99% span. As shown in Fig. 7(a), both ice crystals fly in a region where the velocity is relatively low owing to the boundary layer on the casing wall of the axial fan. Therefore, the flow field near the tip side is dominated by the circumferential velocity. By contrast, from Fig. 7(b), comparing the direction of the velocity vectors and the ice crystal trajectories, MVD of 25 μm results in a small pitch spiral following the flow, and MVD of 200 μm results in a large pitch spiral. This is because ice crystals with MVD of 25 μm have a smaller particle Reynolds number than that of MVD of 200 μm, and the drag received from the surrounding flow field has a significant effect on the ice crystal trajectory at MVD of 25 μm. Therefore, the flight time at MVD of 25 μm is longer. The tendency at MVD of 50 μm is similar to that of MVD of 25 μm, and the higher water content on the tip side, as shown in Fig. 5, is due to the difference in the flight time.

(a) 25 μm   (b) 50 μm   (c) 100 μm   (d) 200 μm

42.6 ▮▮▮ 67.6 [%]   11.1 ▮▮▮ 16.9 [%]   2.85 ▮▮▮ 4.36 [%]   0.72 ▮▮▮ 1.10 [%]
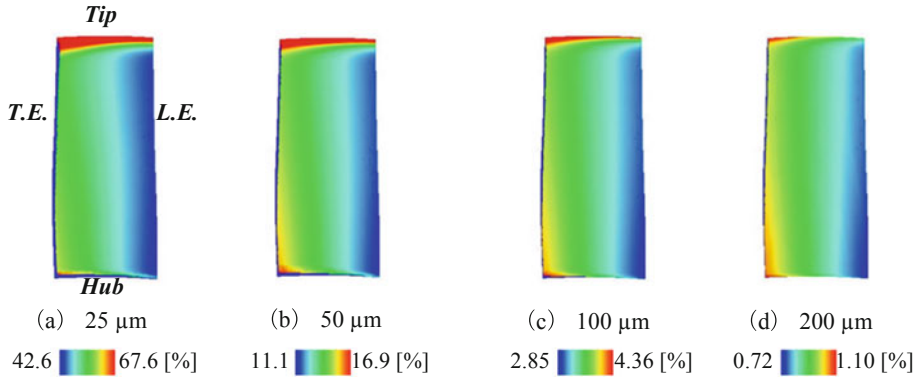
Fig. 5 Water content on the pressure surface under different MVD conditions

(a) 25 μm   (b) 200 μm
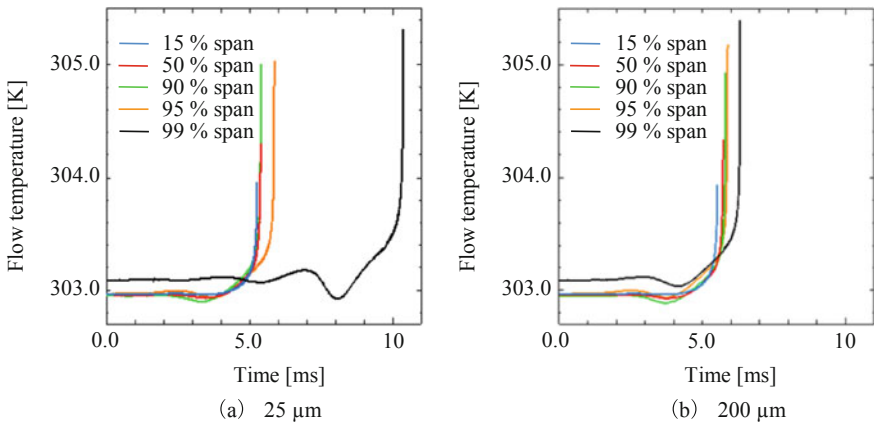
Fig. 6 Flow temperature surrounding the ice crystals impinging on the leading edge

(a) Axial velocity （Casing view）   (b) Velocity vectors （Top view）
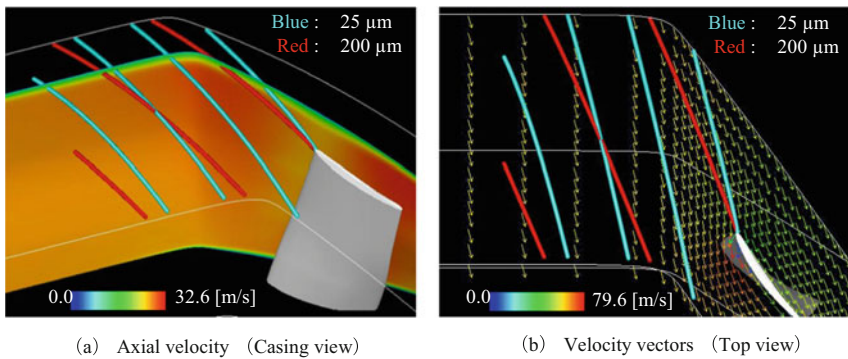
Fig. 7 Ice crystal trajectories within the 99% span

5 Conclusion

We conducted ice crystal trajectory simulations with four types of ice crystal sizes on the flow field around an axial fan blade to investigate the effect of the consideration of the state change. The following insights were obtained.

- The water content of ice crystals that impinge on the entire rotor blade increases as the ice crystals become smaller, which is 0.81% at MVD of 200 μm and 53.5% at MVD of 25 μm.
- The water content of the ice crystals at the time of impingement on the rotor blade increases from the leading edge to the trailing edge, which is 0.24% at MVD of 200 μm and 13.4% at MVD of 25 μm in midspan.
- The water content of the ice crystals that impinge near the tip side shows a particularly high value in the entire rotor blade as the ice crystal size decreases. This is because the smaller the ice crystal size is, the more it follows the flow field owing to the influence of the boundary layer on the casing of the axial fan, and the flight time of the ice crystal impinging near the tip side becomes longer than that of the other ice crystals.

Reference

1. Struk, P.M., Broeren, A.P., Tsao, J.-C., Vargas, M., Wrigh, W.B., Currie, T., Knezevici, D., Fuleki, D.: Fundamental ice crystal accretion physics studies. SAE Technical paper, pp. 1–3 (2011)
2. Mason, J.G., Strapp, J.W., Chow, P.: The ice particle threat to engines in flight. In: 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, pp. 1–21. AIAA (2006)
3. Iwago, M., Fukudome, K., Mamori, H., Fukushima, N., Yamamoto, M.: Fundamental investigation to predict ice crystal icing in jet engine. In: Recent Asian Research on Thermal and Fluid Sciences, pp. 305–317 (2020)
4. Kato, M., Launder, B.E.: The modeling of turbulent flow around stationary and vibrating square cylinders. In: 9th Symposium on Turbulent Shear Flows, Kyoto, pp. 10-4-1–10-4-6. Springer (1993)
5. Yee, H.C.: Upwind and symmetric shock-capturing schemes. NASA Technical reports server, pp. 4–127 (1987)
6. Fujii, K., Obayashi, S.: Practical applications of new LU-ADI scheme for the three-dimensional Navier-stokes computation of transonic viscous flows. In: 24th AIAA Aerospace Sciences Meeting, AIAA-86-0513, Reno. AIAA (1986)
7. Lou, D., Hammond, D.W.: Heat and mass transfer for ice particle ingestion inside aero-engine. J. Turbomach. **133**(3), 031021-1–031021-5 (2011)
8. Penner, S.S.: On maximum evaporation rates of liquid droplets in rocket motors. J. Am. Rocket Soc. **23**(2), 85–88 (1953)
9. Pruppacher, H.R., Klett, J.D.: Microphysics of Clouds and Precipitation, vol. 18, pp. 502–507. Springer, Heidelberg (2010)
10. Murooka, T., Shishido, S., Hiramoto, R., Minoya, T.: Surface coating effect on protection of icing for axial fan blade. SAE Technical paper, 2011-38-0009 (2011)

Train Running Safety Analysis Against Stochastically Generated Crosswinds

P. A. Montenegro , **D. Barbosa, H. Carvalho** , **D. Ribeiro** , **and R. Calçada**

Abstract This article presents a methodology for assessing the train running safety on bridges subjected to strong crosswinds. The methodology is based on an in-house numerical tool called VSI - Vehicle-Structure Interaction Analysis, which computes the dynamic response of the coupled system formed by the train and the bridge. By imposing the wind load, based on a turbulent wind model, to both the bridge and the vehicle, it is possible to evaluate the wheel-rail contact forces and, consequently, the running safety criteria, namely the Nadal, Prud'homme and unloading indexes. A case study is presented in this paper, which consists of a high-speed (HS) train running over the new Volga River railway bridge belonging to the future Russian HS network. The critical wind speeds for which the train's safety can be jeopardize are identified for a vast range of running speeds. At the end, the results obtained with the stochastic turbulent model are compared with those obtained using the discrete gust model proposed by the European Norms. A comparative critical analysis between the realistic turbulent wind model and the simplified discrete gust model is carried out.

P. A. Montenegro (✉) · D. Barbosa · R. Calçada
CONSTRUCT-LESE, Faculty of Engineering of University of Porto, Porto, Portugal
e-mail: paires@fe.up.pt

H. Carvalho
Department of Structural Engineering, Federal University of Minas Gerais, Belo Horizonte, Brazil

D. Ribeiro
CONSTRUCT-LESE, School of Engineering of Polytechnic of Porto, Porto, Portugal

1   Introduction

The fast development in the last decades of several high-speed rail networks around the globe made it necessary to build new railway lines that would meet the strict design requirements of this type of transport. Thus, the necessity to ensure smoother tracks with larger curve radius resulted in new railway lines with a high percentage of viaducts and bridges. Some countries in Asia, for example, such as China, Japan and Taiwan, have a highly developed high-speed railway network in which some of the lines have more than 75% of viaducts [1–3]. This reality led to an increase in the probability of a train being over a bridge during the occurrence of natural hazards, such as earthquakes or crosswinds, that might compromise its running safety. The latter is particular important in bridges that situated in regions with deep valleys, where the crosswinds tend to be extremely severe to the train's safety. As a result, lateral winds acting over a moving vehicle may significantly increase the risk of overturning derailments, particularly due to the development of ever faster [4, 5] and lighter trains [6]. Hence, the assessment of threshold values for the wind speed above which the train should lower the speed to decrease the derailment risk is a topic of the utmost importance and that deserves to be comprehensively studied.

Several researchers have been working in the topic of bridge dynamics using the Finite Element Method (FEM) [7–9] and vehicle dynamics using Multibody Simulation (MBS) techniques [10, 11], but to evaluate the train running safety against crosswinds, it is necessary to develop a coupled vehicle-structure interaction model. These models have to be able to accurately simulate the wheel-rail contact phenomenon, since it is this interface that the derailment is triggered. Several researchers have been developing this type of models mainly in the last two decades [6, 12–14] and a comprehensive literature review on this topic can be found in [15].

Some studies regarding the train running safety on bridges subjected to crosswinds have been published in the last years. Xia et al. [12] evaluated the running safety of a train running on a suspended bridge in China and concluded that the critical mean wind speed when the train cross the bridge at the design speed of 160 km/h was 25 m/s. Guo et al. [16] studied the derailment risk in a wind prone stretch of the Lanzhou-Xinjiang HS railway line composed by 103 viaducts with a total length of almost 52 km. The authors concluded that the derailment overturning risk could be substantially reduced with the implementation of wind barriers. Olmos and Astiz [6] also conducted numerical simulations on a high pier viaduct belonging to the Spanish HSR network and analyzed the influence of wind barriers and tuned mass dampers (TMDs) in the train's running safety. The TMDs installed at the top of the piers significantly decreased the bridge lateral response and, consequently, increased the train safety.

All the publications mentioned before simulate the wind velocity field with stochastic turbulent wind models that take into consideration the coherence between the wind time-histories in the different generation points. However, no comparisons with other simplified wind models can be found in the literature. The present work aims to fill this gap, by comparing the results in terms of train running safety

obtained with a turbulent wind model with those obtained with the simplified CEN discrete gust model stipulated in the European norm EN 14,067-6 [17]. The case study adopted to perform this comparison is the Volga River bridge, which belongs to the future Russian HS railway network. The calculations are carried out with a numerical tool called VSI - Vehicle-Structure Interaction Analysis, which evaluates the coupling dynamic response of train-bridge system and computes the wheel-rail contact forces used to assess the running safety indexes. This tool has been previously developed and validated by part of authors of the present paper and successfully used recently in other crosswind safety studies [18–21].

2 Wind Models

2.1 Stochastic Turbulent Wind Model

The wind velocity field in the Stochastic Turbulent Wind (STW) model is generated taking into account the power spectral density function S for each direction. According to Cao et al. [22], if the elevation is constant along the wind field and the distance between any successive wind generation points is the same, which can be perfectly adopted for the majority of the studies regarding running safety over bridges, where the deck is at a constant height, the time-histories of the horizontal u_j and vertical v_j fluctuating components of the wind in the jth generation point can be simulated by

$$u_j(t) = \sqrt{2\Delta\omega} \sum_{m=1}^{j} \sum_{f=1}^{N} \sqrt{S_u(\omega_{mf})} G_{jm}(\omega_{mf}) \cos(\omega_{mf}t + \phi_{mf}) \tag{1}$$

$$w_j(t) = \sqrt{2\Delta\omega} \sum_{m=1}^{j} \sum_{f=1}^{N} \sqrt{S_w(\omega_{mf})} G_{jm}(\omega_{mf}) \cos(\omega_{mf}t + \phi_{mf}) \tag{2}$$

where, ω_{mf} is the wind frequency, $\Delta\omega$ is the frequency increment, ϕ_{mf} is a random variable phase angle uniformly distributed between 0 and 2π, N is the number of wind frequencies, $G_{jm}(\omega_{mf})$ is an element of the coefficient matrix $\mathbf{G}(\omega_{mf})$ that represents the coherence between the wind generation points (see [19, 21] for details) and $S_u(\omega_{mf})$ and $S_w(\omega_{mf})$ are the horizontal [23] and vertical [24] wind spectra defined by

$$S_u(\omega_{mf}) = \frac{u_*^2}{\omega_{mf}} \frac{200f_z}{(1+50f_z)^{5/3}} \tag{3}$$

$$S_w(\omega_{mf}) = \frac{u_*^2}{\omega_{mf}} \frac{3.36f_z}{(1+10f_z)^{5/3}} \tag{4}$$

Fig. 1 Example of the horizontal component of the wind velocity field at different generation points for $\overline{U}_z = 20\,\text{m/s}$
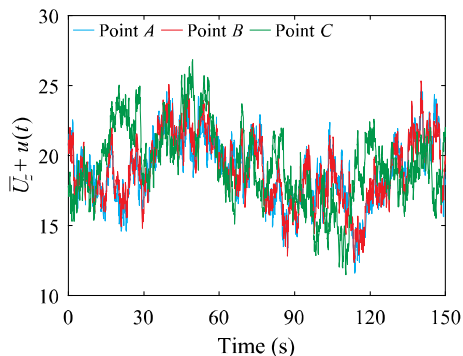
in which f_z is the normalized frequency dependent on the height of the generation point relative to the ground z and on the mean wind speed $\overline{U}_z$ at that height given by

$$f_z = \frac{\omega_{mf} z}{2\pi \overline{U}_z} \tag{5}$$

and u_* is the friction velocity of the wind flow expressed as

$$u_* = \frac{k\overline{U}_z}{\ln\left(\frac{z}{z_0}\right)} \tag{6}$$

where k is the von Karman constant equal to 0.4 [17] and z_0 is the roughness length.

Figure 1 shows three time histories of the horizontal component of the wind velocity field generated in different points. In this figure, points A and B are 1 m apart, while point C is located 180 m away. As expected, the coherence between points A and B, which are very close to each other, is very high, but it decreases for the point that is located far away from them.

2.2 CEN Discrete Gust Model

According to EN 14,067–6 [25], the gust generated in the CEN discrete gust model (CEN) has a fixed amplitude and a probability level of exceedance of 50% for the gust duration. The following simplifications are assumed:

- The bi-exponencial model corresponds to an approximation of the average of a random process in the vicinity of a local maximum.

- The mean wind is assumed to be horizontal, since this component represents the prominent part of wind fluctuations.
- The variations of the wind direction during the gust are neglected.
- The wind gust is fixed in space and the train speed is constant, which means that the temporal and spatial distributions of the wind velocity are linearly related to each other.
- The height above the ground is fixed in 4 m, leading to an integral scale turbulence of $L_u = 96$ m.

The mean wind speed $\overline{U}_z$ at the height z is calculated from the maximum wind speed U_{max} according to the following equation

$$\overline{U}_z = \frac{U_{\max}}{1.6946} \tag{7}$$

which corresponds to turbulent intensity of $I_u = 24.46\%$. The horizontal component of the wind speed defined by the CEN model is based on the von Karman wind spectrum defined as

$$S_u(n) = \frac{4f_u \sigma_u^2}{\left[1 + 70.7 f_u^2\right]^{5/6}} \cdot \frac{1}{n} \tag{8}$$

where the frequency range is limited to $n = [1/300$ Hz; 1 Hz$]$, σ_u is the standard deviation of the turbulence and f_u represents the normalized frequency given by

$$f_u = \frac{n L_u}{\overline{U}_z} \tag{9}$$

With some mathematical manipulation (see [17] and [20] for details), the wind speed normal to the vehicle at height z U_z along the track is calculated from

$$U_z(x) = \overline{U}_z + 2.84 \sigma_u e^{\frac{-16fx}{\overline{U}_z}} \tag{10}$$

where x is the longitudinal distance towards the position of the gust peak. The mathematical description only refers to the second half of the CEN model, where the wind decays from U_{max} to $\overline{U}_z$, therefore, mirroring is necessary in the y axis for the complete representation of the model. Figure 2 shows an example of the spatial distribution of the wind velocity U_z using the CEN model (positions x_i represent the points in which the wind speed changes). The wind gust shall be low-pass filtered using a moving spatial average based on a window size equal to the vehicle length and a step size smaller than 0.5 m.

Fig. 2 Example of the spatial distribution of the wind using the CH model (adapted from [17])
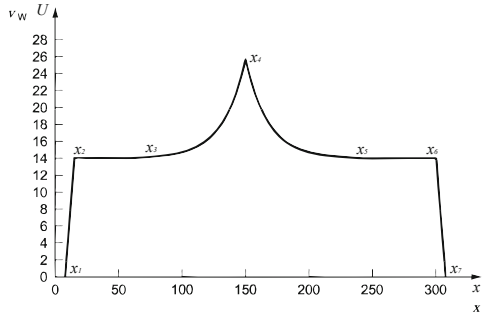
3 Wind Aerodynamic Forces

3.1 Wind Loads on the Bridge

The wind loads acting on the bridge are divided into the steady-state and buffeting components. While the former is induced by the mean wind velocity, the latter is caused by the wind fluctuating flow in the horizontal $u(t)$ and vertical directions $w(t)$. Hence, and according to the quasi-steady theory, the drag, $F_{d,b}$, and lift, $F_{l,b}$, wind forces per unit length of the deck (see Fig. 1) can be expressed as

$$F_{d,b}(t) = \frac{1}{2}\rho V_{w,b}(t)^2 C_{d,b}(\alpha)H_b \tag{11}$$

$$F_{l,b}(t) = \frac{1}{2}\rho V_{w,b}(t)^2 C_{l,b}(\alpha)B_b \tag{12}$$

where $C_{d,b}(\alpha)$ and $C_{l,b}(\alpha)$ are the drag and lift wind force aerodynamic coefficients of the bridge dependent on the angle of incidence α, H_b and B_b are the height and width of the deck exposed area, respectively, ρ is the air density and $V_{w,b}(t)$ is the resultant wind velocity given by

$$V_{w,b}(t) = \sqrt{\left[\overline{U}_z + u(t)\right]^2 + w(t)^2} \tag{13}$$

leading to an incidence angle α of

$$\alpha(t) = \tan^{-1}\left(\frac{w(t)}{\overline{U}_z + u(t)}\right) \tag{14}$$

The moment $M_{m,b}$ acting on the deck (see Fig. 3) arises from the vertical eccentricity e_z between the deck gravity center, C_g, and the pressure center of the lateral windward area, C_p, and from the horizontal eccentricity e_y, taken as $B_b/4$ as

Fig. 3 Resulting wind velocity and wind loads acting on the bridge
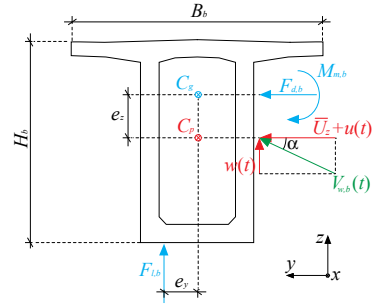
stated in EN 1991-1-4 [26]. In the present work, the drag and lift aerodynamic coefficients have been calculated according to the procedure defined in Sect. 8.3 from [26].

3.2 Loads on Railway Vehicles

The wind loads acting on the vehicle are calculated by a procedure similar to that used in the bridge. Considering that the main stream flow acts perpendicular to the direction of the vehicle movement, the resultant wind velocity $V_{w,v}(t)$ is given by the same expression defined in Eq. (13). Therefore, the relative wind velocity to the vehicle $V_r(t)$ can be defined as (see Fig. 2)

$$V_r(t) = \sqrt{V_v^2 + V_{w,v}^2} = \sqrt{V_v^2 + \left[\overline{U}_z + u(t)\right]^2 + w(t)^2} \tag{15}$$

while the yaw angle β is given by

$$\beta(t) = \tan^{-1}\left(\frac{V_{w,v}}{V_v}\right) = \tan^{-1}\left(\frac{\sqrt{\left[\overline{U}_z + u(t)\right]^2 + w(t)^2}}{V_v}\right) \tag{16}$$

where V_v is the vehicle's running speed.

Finally, the forces acting at the center of gravity C_g of the carbody of each car of the train, namely the drag, $F_{d,v}(t)$, and lift, $F_{l,v}(t)$, forces and the rolling moment $M_{m,v}(t)$ are depicted in Fig. 2 and can be expressed by the following equations

$$F_{d,v}(t) = \frac{1}{2}\rho V_r(t)^2 C_{d,v}(\alpha, \beta) A_v \tag{17}$$

$$F_{l,v}(t) = \frac{1}{2}\rho V_r(t)^2 C_{l,v}(\alpha, \beta) A_v \tag{18}$$
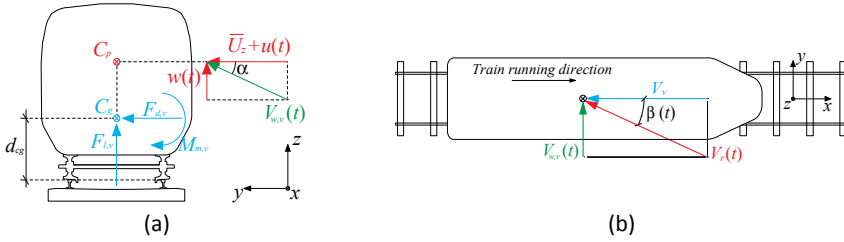
Fig. 4 Relative wind velocity and wind loads acting on the vehicle: (a) front and (b) plan views

$$M_{m,v}(t) = \frac{1}{2}\rho V_r(t)^2 C_{m,v}(\alpha, \beta) A_v H_v - F_{d,v} d_{gc} \tag{19}$$

where $C_{d,v}$ (α,β), $C_{l,v}$ (α,β) and $C_{m,v}$ (α,β) are the aerodynamic drag, lift and moment coefficients of the vehicle, respectively, that depend on the incidence and yaw angles, and A_v and Hv are the windward exposed area and height of the vehicle, respectively. In the present work, and due to the similarity of the train used in this study, the aerodynamic coefficients given by the Annex C of [17] for the ICE3 train have been adopted for all the calculations. In this normative, the rolling moment aerodynamic coefficient refers to the moment along the longitudinal axis centered with the track and situated at the top of the rails. Therefore, since the moment has to be applied at the center of gravity of the carbody, it is necessary to remove the portion of the moment that arises from the eccentricity d_{gc} between this point and the top of the rails (see Fig. 4).

4 Train-Bridge Interaction Model

4.1 Bridge Model

The Volga River HS railway bridge has four prestressed concrete continuous spans of $98.55 + 2 \times 190 + 98.55$ m, as shown in Fig. 5. The deck consists of a cast-in-place single-cell box girder with 6200 mm width and height ranging from 5000 mm at the middle of each span to 12,500 mm over the main piers. The middle pier is monolithically connected to the deck, while the remaining ones are con- nected through bearing supports. The top slab is 13800 mm width and supports two ballasted tracks.

The FEM numerical model of the bridge has been developed in ANSYS® [27]. The deck, pier and rails have been modeled using beam finite elements, while ballast and pads have been simulated through linear spring dampers. Mass point elements have been also used to model the ballast mass and other non-structural elements. The connections between the top of the middle pier and the deck and
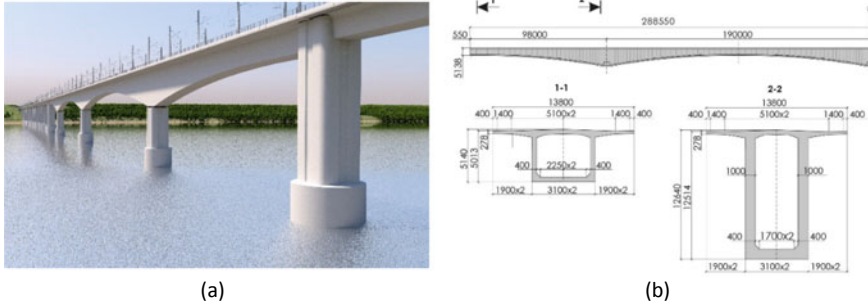
Fig. 5 Volga River HS bridge: (a) photomontage and (b) elevation view and cross sections
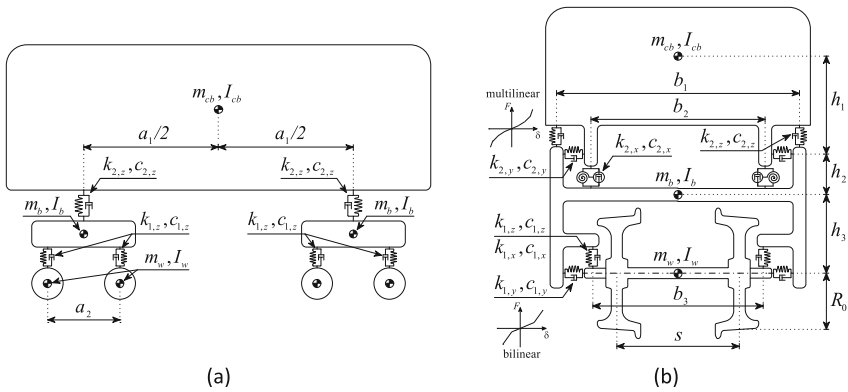
Fig. 6 Dynamic model of the railway vehicle: (a) lateral and (b) front views

between the deck and the track have been accomplished using rigid frame elements. The mechanical and dynamic properties of the bridge can be consulted in [21].

4.2 Train Model

Due to lack of information regarding the Russian HS train, the vehicle used in this work consists of a Japanese Shinaknsen train whose properties were known. The train model is developed in ANSYS® [27] and consists of ten independent cars with 25 m length each and with an average static axle load of 110 kN. The dynamic model is represented in Fig. 6, where k, c, m and I represent stiffness, damping, concentrated mass and rotary inertia, respectively, while the subscripts cb, b and w denote carbody, bogie and wheelset, respectively. The suspensions are modeled with spring-damper elements, while the carbody, bogies and wheelsets are modeled with beam rigid elements.

4.3 Train-Bridge Dynamic Coupling

The dynamic train-track interaction coupling system is solved by using the software VSI - Vehicle-Structure Interaction Analysis developed in MATLAB® [28]. This tool is based on the direct method, in which the constraint equations, that relate the displacements of the train with the nodal displacements of the track, are added to the equilibrium equations, forming a single system that defines the coupling between the two sub-structures. Neves et al. [29] and Montenegro et al. [30] extended the original formulation developed by [31] to deal with train structure detachment. Later, Montenegro et al. [32] reformulated the method to take into account the lateral interaction by considering the geometry of the wheel and rail and the contact model between them. This last upgrade made it possible to compute both vertical and lateral contact forces, allowing the model to deal with any type of scenarios with lateral loads, such as in the presence of crosswinds. A detailed description of the vehicle track coupling model may be found in [32].

5 Results of the Running Safety Analysis

5.1 Running Safety Indexes

The running safety of the train subjected to crosswinds is evaluated by three safety indexes, namely Nadal, Prud'homme and Unloading. These indexes ξ are computed according to the following equations

$$\xi_N = \frac{Y}{Q} \tag{20}$$

$$\xi_P = \frac{\sum_{ws} Y}{10 + \frac{2Q_0}{3} [\text{kN}]} \tag{21}$$

$$\xi_U = 1 - \frac{Q_i + Q_j}{2Q_0} \tag{22}$$

where the subscripts N, P and U refer to the Nadal, Prud'homme and Unloading indexes, respectively, Y and Q are the dynamic lateral and vertical contact forces per wheel, respectively, Q_0 is the vertical static load per wheel, Q_i and Q_j are the vertical dynamic contact forces of the windward wheels from the wheelsets i and j belonging to the same bogie and $\sum_{ws} Y$ represents the sum of the lateral contact forces from the two wheels from the same wheelset. According to [33], the Nadal and Prud'homme indexes are limited to 0.8 and 1.0, respectively, while the Unloading index should not exceed 0.9, as stipulated in [17].

5.2 Results Obtained with the STW Model

The maximum safety indexes obtained for each combination of train and wind speeds are depicted in Fig. 7. In these graphs, $\overline{U}_z$ refers to the mean wind velocity at height $z = 48$ m, which corresponds to the height of the Volga River bridge deck. It is possible to observe that the train can safely cross the bridge at any speed if the wind mean velocity does not exceed 15 m/s. Even for wind loads characterized by mean velocities of 20 m/s, the safety is put at risk only for very high train speeds (higher than 350 km/h), based on the Prud'homme criteria. However, when the mean wind speed reaches 25 m/s, the unloading criterion allowance is exceeded for typical train speeds practiced in HS railway networks (around 300 km/h). In this scenario, the train would need to reduce its speed to avoid safety problems related to overturning derailment.

Finally, it is important to notice that the train's stability is not related to any specific critical location in the bridge, since the maximum values of the derailment factors are related to the instants in which the wind speed, and consequently the
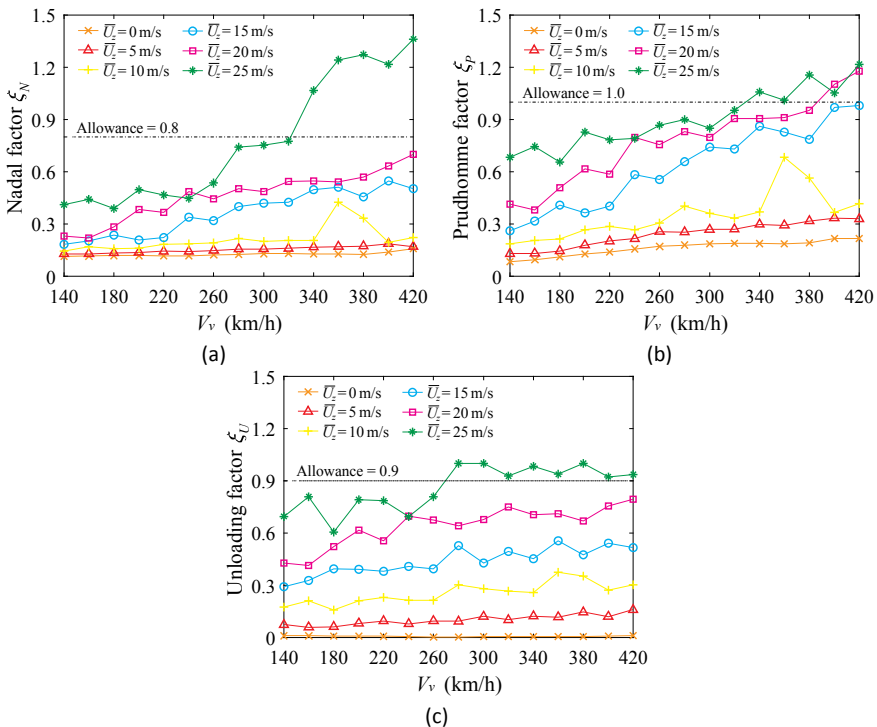
Fig. 7 Maximum safety indexes: (a) Nadal, (b) Prud'homme and (c) unloading criteria
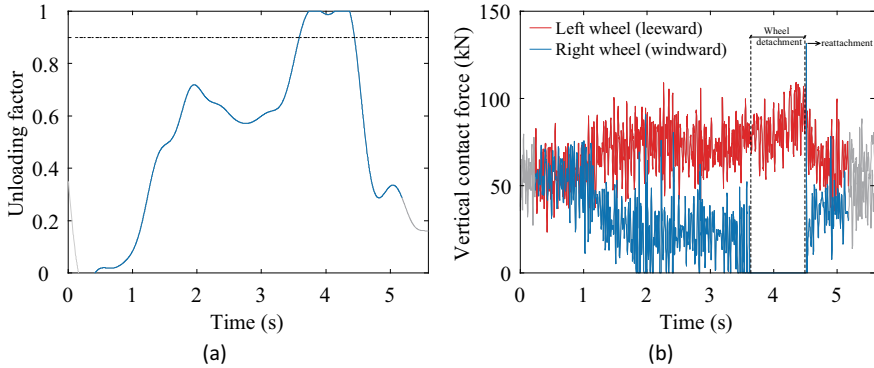
Fig. 8 Extreme scenario of a train travelling at 420 km/h subjected to a wind load corresponding to a mean wind speed of 20 m/s simulated with the CEN model: (a) unloading criterion and (b) vertical contact forces in the first wheelset

wind load, is higher. Therefore, depending on the random generated stochastic wind field used in a specific simulation, the maximum values of the derailment factors may occur in different positions of the bridge.

5.3 Results Obtained with the CEN Model

Figure 8 depicts the results obtained in an scenario corresponding to a train travelling at the design speed (420 km/h) during the occurrence of a crosswind simulated with the CEN discrete gust model. The mean wind speed is 20 m/s, which, according to Eq. (7), corresponds to a gust peak of approximately 34 m/s. Under this extreme scenario, it is possible to observe that train reaches a limit situation where the windward wheels end up even losing contact with the rail, leading to a probable overturning derailment. It can be observed in Fig. 8b that the vertical contact force of the right wheel (windward side) drops to zero during the gust peak but it suddenly reaches 130 kN on $t \approx 4.5$ s due to the reattachment impact.

By observing Fig. 9, it is interesting to notice that the windward wheel jumps around 50 mm from the rail, while the 6 mm gap that exists between the rail and the flange of the leeward wheel closes between $t \approx 1.7$ s and $t \approx 4.7$ s. During this time interval, the rail is subjected to an almost constant impact with the wheel flange. This extreme scenario clearly demonstrates the importance of limiting the unloading criterion to 0.9 in order to avoid wheel detachments that can lead to overturning derailments during the occurrence of crosswinds.
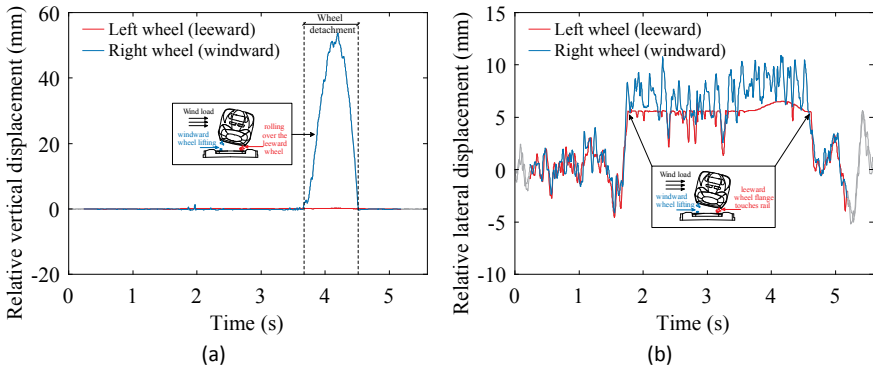
Fig. 9 Relative displacements between wheel and rail during the extreme scenario of a train travelling at 420 km/h subjected to a wind load corresponding to a mean wind speed of 20 m/s simulated with the CEN model: (a) vertical and (b) lateral displacements

5.4   Results Comparison

The present section aims to compare the results in terms of train running safety obtained for the same scenarios but considering the two distinct wind models presented in Sect. 2.

The Prud'homme and unloading safety indexes obtained with the CEN and STW models for different train and wind speeds are plotted in Fig. 10 (the subscripts *st* and *dg* indicates stochastic and discrete gust models, respectively, while $\overline{U}_z$ is the mean wind velocity at height $z = 48$ m). It is clear that the CEN model is more conservative for all the scenarios, leading to higher safety index values. Such results are due to the fact that the CEN model imposes a height above the terrain of $z = 4$ m with an associated turbulence intensity of 24.46% (see Sect. 2.2). The decks of the bridges, however, may be located at higher elevations, resulting in much lower turbulence intensity and, consequently, in a lower ratio between peak and mean wind velocity. In the present case, since the deck is located at $z = 48$ m, the turbulence intensity is only 14.56%, leading to lower peak wind speed when using the STW model. Thus, for the same mean wind velocity $\overline{U}_z$ at height $z = 48$ m, the CEN model leads to higher peak wind velocities and, generally, to more unfavorable safety indexes.

Figure 11 depicts the critical wind speeds that the vehicle can withstand, for each analyzed running speed, before exceeding any of the safety criteria allowances. It is clear that the results obtained when using the CEN model to define the wind action is evident, leading to critical mean wind velocities 5 to 6 m/s lower than those obtained with the STW model. Hence, although more complex to implement, the models based on turbulent wind velocity fields are an interesting option when investigating the derailment of trains running over bridges. These
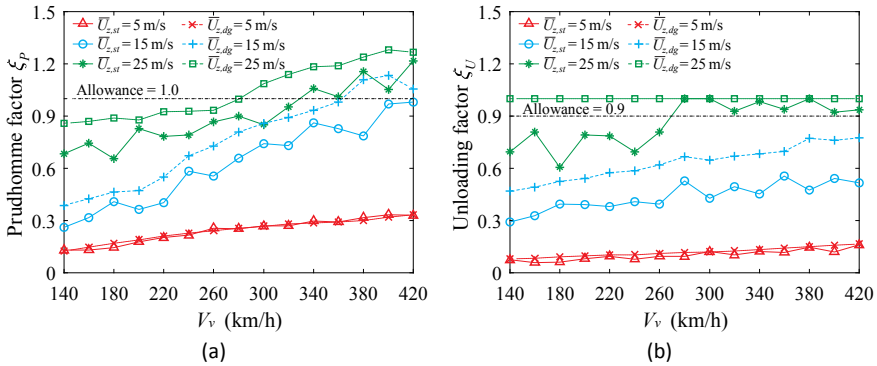
Fig. 10 Comparison of the maximum safety indexes for several wind velocities obtained with the STW and CEN models: (a) Prud'homme and (b) unloading criteria
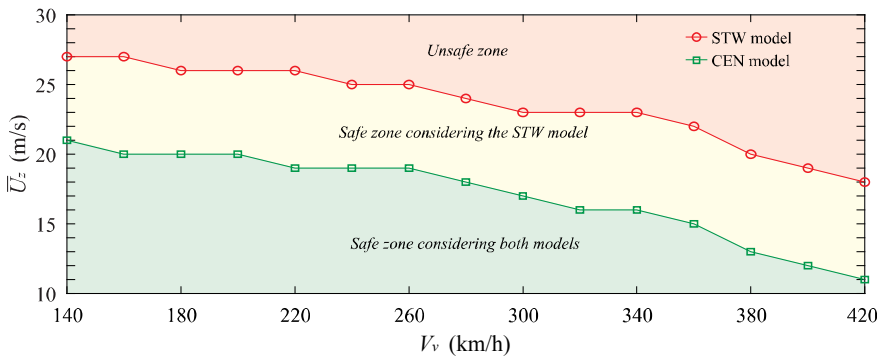
Fig. 11 Running safety curve obtained by the STW and CEN models

conclusions are of the utmost importance to the current state of the art on this field since there is currently no information available in the literature that compares both models in terms of safety indexes.

6 Conclusions

The present paper presents a running safety study of a train moving over the new Volga River HS railway bridge, located in Russia, subjected to crosswinds. The wind is taken into account through two distinct models, namely the stochastic turbulent wind (STW) model and the CEN discrete gust (CEN) model. While the first represents the wind in a realistic way, the second represents it in a simplified way through a single gust fixed in space and it has the advantage of being easily to define and presents a much higher computational efficiency. However, the results

are not so accurate and can be extremely conservative for situations where the train runs on decks located at high elevations. Based on this, the following conclusions can be drawn:

- When using the STW model. The results show that, for wind loads characterized by mean wind speeds up to 15 m/s, the train can always safely cross the bridge. However, for wind speeds of 25 m/s, the unloading criterion allowance is exceeded for typical train speeds practiced in HS railway networks (around 300 km/h), which would imply reducing the train speed to avoid safety problems related to overturning derailments.
- The CEN model proved to be extremely conservative for being used in scenarios where the train runs over decks situated at high elevations. Such results are due to the fact that the CEN model imposes a height above the terrain of $z = 4$ m with an associated turbulence intensity of 24.46%, while the STW model uses the correct turbulence intensity depending on the height. Thus, as the elevation increases, the turbulence intensity decreases and, consequently, the ratio between peak and mean wind speed is lower. For the Volga River bridge, since its deck is located at $z = 48$ m, the turbulence intensity is only 14.56%, the STW model leads to lower peak wind speeds when compared to those obtained with the CEN model. Hence, for the same mean wind velocity at height $z = 48$ m, the CEN model leads to higher peak wind velocities and, consequently, to more unfavorable safety indexes.
- The critical wind speeds obtained with the CEN mode are, generally, 5 to 6 m/s lower than those obtained with the STW model.

References

1. Kao, T.C., Lin, C.K.: Taiwan high speed rail & its impact to regional development. In: 4th CECAR - Civil Engineering Conference in the Asian Region, Taipei, Taiwan (2007)
2. Ishibashi, T. Shinkansen structures in Japan. In: Workshop on Bridges for High-Speed Railways, Porto, Portugal (2004)
3. Dai, G., Hu, N., Liu, W.: The recent improvement of high-speed railway bridges in China. In: 34th IABSE Symposium for Bridge and Structural Engineering: Large Structures and Infrastructures for Environmentally Constrained and Urbanised Areas, Venice, Italy (2010)
4. Giappino, S., Rocchi, D., Schito, P., Tomasini, G.: Cross wind and rollover risk on light weight railway vehicles. J. Wind Eng. Ind. Aerodyn. **153**, 106–112 (2016). https://doi.org/10.1016/j.jweia.2016.03.013
5. Xu, L., Zhai, W.: Cross wind effects on vehicle–track interactions: a methodology for dynamic model construction. J. Comput. Nonlinear Dyn. **14**(3), 1–11 (2019). https://doi.org/10.1115/1.4042142

6. Olmos, J.M., Astiz, M.A.: Improvement of the lateral dynamic response of a high pier viaduct under turbulent wind during the high-speed train travel. Eng. Struct. **165**, 368–385 (2018). https://doi.org/10.1016/j.engstruct.2018.03.054

7. Peixer, M.A., Carvalho, H., Montenegro, P.A., Correia, J.A.F.O., Bittencourt, T., Calçada, R., Guo, T.: Influence of the double composite action solution in the behavior of a high-speed railway viaduct. J. Bridge Eng. **25**(7), 05020002: 1–13 (2020). https://doi.org/10.1061/(ASCE)BE.1943-5592.0001563

8. Moliner, E., Martínez-Rodrigo, M.D., Museros, P.: Dynamic performance of existing double track railway bridges at resonance with the increase of the operational line speed. Eng. Struct. **132**, 98–109 (2017). https://doi.org/10.1016/j.engstruct.2016.11.031

9. Meixedo, A., Ribeiro, D., Calçada, R., Delgado, R.: Global and local dynamic effects on a railway viaduct with precast deck. In: Civil-Comp Proceedings of the The Second International Conference on Railway Technology: Research, Development and Maintenance, Ajaccio, France (2014). https://doi.org/10.4203/ccp.104.77

10. Shabana, A., Zaazaa, K.E., Sugiyama, H.: Railroad Vehicle Dynamics: A Computational Approach. CRC Press - Taylor & Francis Group, Boca Raton (2008)

11. Magalhães, H., Marques, F., Liu, B., Antunes, P., Pombo, J., Flores, P., Ambrósio, J., Piotrowski, J., Bruni, S.: Implementation of a non-Hertzian contact model for railway dynamic application. Multibody Syst. Dyn. **48**, 41–78 (2020). https://doi.org/10.1007/s11044-019-09688-y

12. Xia, H., Guo, W.W., Zhang, N., Sun, G.J.: Dynamic analysis of a train–bridge system under wind action. Comput. Struct. **86**(19–20), 1845–1855 (2008). https://doi.org/10.1016/j.compstruc.2008.04.007

13. Montenegro, P.A., Calçada, R., Vila-Pouca, N., Tanabe, M.: Running safety assessment of trains moving over bridges subjected to moderate earthquakes. Earthq. Eng. Struct. Dyn. **45**, 483–504 (2016). https://doi.org/10.1002/eqe.2673

14. Ticona Melo, L., Malveiro, J., Ribeiro, D., Calçada, R., Bittencourt, T.: Dynamic analysis of the train-bridge system considering the non-linear behaviour of the track-deck interface. Eng. Struct. 110980: 1–21 (2020). https://doi.org/10.1016/j.engstruct.2020.110980

15. Zhai, W.M., Han, Z., Chen, Z., Ling, L., Zhu, S.: Train–track–bridge dynamic interaction: a state-of-the-art review. Veh. Syst. Dyn. **57**(7), 984–1027 (2019). https://doi.org/10.1080/00423114.2019.1605085

16. Guo, W.W., Xia, H., Karoumi, R., Zhang, T., Li, X.: Aerodynamic effect of wind barriers and running safety of trains on high-speed railway bridges under cross winds. Wind Struct. **20**(2), 213–236 (2015). https://doi.org/10.12989/was.2015.20.2.213

17. EN 14067-6. Railway applications - Aerodynamics - Part 6: Requirements and test procedures for cross wind assessment, European Committee for Standarization (CEN), Brussels (2016)

18. Neto, J., Montenegro, P.A., Vale, C., Calçada, R.: Evaluation of the train running safety under crosswinds - a numerical study on the influence of the wind speed and orientation considering the normative Chinese hat model. Int. J. Rail Transp. (2020). https://doi.org/10.1080/23248378.2020.1780965

19. Montenegro, P.A., Heleno, R., Carvalho, H., Calçada, R., Baker, C.J.: A comparative study on the running safety of trains subjected to crosswinds simulated with different wind models. J. Wind Eng. Ind. Aerodyn. **207**, 104398 (2020). https://doi.org/10.1016/j.jweia.2020.104398

20. Montenegro, P.A., Calçada, R., Carvalho, H., Bolkovoy, A., Chebykin, I.: Stability of a train running over the Volga river high-speed railway bridge during crosswinds. Struct. Infrastruct. Eng. **16**(8), 1121–1137 (2020). https://doi.org/10.1080/15732479.2019.1684956

21. Montenegro, P.A., Barbosa, D., Carvalho, H., Calçada, R.: Dynamic effects on a train-bridge system caused by stochastically generated turbulent wind fields. Eng. Struct. **211**, 110430 (2020). https://doi.org/10.1016/j.engstruct.2020.110430

22. Cao, Y., Xiang, H., Zhou, Y.: Simulation of stochastic wind velocity field on long-span bridges. J. Eng. Mech. **126**(1), 1–6 (2000). https://doi.org/10.1061/(ASCE)0733-9399(2000)126:1(1)

23. Kaimal, J.C., Wyngaard, J.C., Izumi, Y., Coté, O.R.: Spectral characteristics of surface-layer turbulence. Q. J. Roy. Meteorol. Soc. **98**, 563–589 (1972). https://doi.org/10.1002/qj.49709841707

24. Lumley, J.L., Panofsky, H.A.: The Structure of Atmospheric Turbulence. Interscience Publishers, New York (1964)

25. EN 14067-6. Railway applications - Aerodynamics - Part 6: Requirements and test procedures for cross wind assessment, European Committee for Standarization (CEN), Brussels (2010)

26. EN 1991-1-4. Eurocode 1: Actions on structures - Part 1–4: General actions - Wind actions, European Committee for Standarization (CEN), Brussels (2010)

27. ANSYS®. Academic Research, Release 19.2, ANSYS Inc., Canonsburg, PA, USA (2018)

28. MATLAB®. Release R2018a, The MathWorks Inc., Natick, MA, USA (2018)

29. Neves, S.G.M., Montenegro, P.A., Azevedo, A.F.M., Calçada, R.: A direct method for analyzing the nonlinear vehicle–structure interaction. Eng. Struct. **69**, 83–89 (2014). https://doi.org/10.1016/j.engstruct.2014.02.027

30. Montenegro, P.A., Neves, S.G.M., Azevedo, A.F.M., Calçada, R.: A nonlinear vehicle-structure interaction methodology with wheel-rail detachment and reattachment. In: COMPDYN 2013 - 4th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Kos, Greece (2013)

31. Neves, S.G.M., Azevedo, A.F.M., Calçada, R.: A direct method for analyzing the vertical vehicle–structure interaction. Eng. Struct. **34**, 414–420 (2012). https://doi.org/10.1016/j.engstruct.2011.10.010

32. Montenegro, P.A., Neves, S.G.M., Calçada, R., Tanabe, M., Sogabe, M.: Wheel-rail contact formulation for analyzing the lateral train-structure dynamic interaction. Comput. Struct. **152**, 200–214 (2015). https://doi.org/10.1016/j.compstruc.2015.01.004

33. TSI. Technical specification for interoperability relating to the infrastructure subsystem of the trans-European high-speed rail system. Official Journal of the European Union, Brussels (2002)

Application of High Efficiency Speed-Up Drilling Technology in Well W60

Yudong Tian, Jun Li, Gonghui Liu, Yue Qi, Baishan Chen, and Jianfeng Mu

Abstract There are two sets of faults in the eastern slope belt of WN subsag in HLE basin, with large dip angle and interbedded glutenite of NT formation and TBM formation. Therefore, there are some difficult problems in the process of drilling, such as easy well deviation, well leakage, difficult deviation control, easy bit wear, low base fracture pressure, difficult cementing and so on. The above problems have been effectively solved by optimizing wellbore structure and drilling fluid system, optimizing drilling tool assembly and optimizing drilling parameters. The torque impact speed increasing tool is applied in well W60, equipped with PDC bit and preferred SMD537X cone bit to enhance the adaptability of the bit. The single-trip drilling footage is 449.76 m, and the ROP is 5.57 m/h, which is 109% higher than that of the adjacent well. The speed-up effect is outstanding, realizing anti-deviation and fast drilling. The industrial oil flow of 5.50t/d was obtained in well W60 after fracturing, leading to find a high-yield and oil-rich zone for the east slope of WN sub-sag and achieve the purpose of exploration and discovery.

Keywords Speed-up tools · Efficient drilling · Fast anti-deviation · Practical application

Y. Tian · J. Li (✉) · G. Liu · Y. Qi
China University of Petroleum-Beijing, Changping, Beijing 102249, China
e-mail: lijun446@vip.163.com

Y. Tian · Y. Qi · B. Chen
Drilling Engineering Technology Research Institute of Daqing Drilling Engineering Company, Daqing 163413, China

J. Li
China University of Petroleum-Beijing at Karamay, Karamay 834000, China

J. Mu
China Petroleum Logging Co., Ltd., Daqing Branch, Daqing 163411, China

1 Introduction

In order to further understand the hydrocarbon condition of the TBM group in the eastern slope zone of the WN sub-sag in the WEX sag of the BEH depression in the HLE basin, explore the basement, NT and DMGH groups, and look for new hydrocarbon enrichment zones, targeting the shielding effect of the W38-5 fault block reservoir, combined with the oil–water layer at the low part of the fault block reservoir and the possible reservoir characteristics of the pure oil layer at the high part, and it is preferable to deploy the exploratory well—W60 as a favorable target. This well is a pre-exploration vertical well with a designed well depth of 2385 m. The purpose of drilling is to explore the hydrocarbon condition of the TBM group in the east slope of the WN sub-concave, determine the type of reservoir, and find new hydrocarbon-rich zones. The target formation is the TBM group, the basement, NT and DMGH groups are explored at the same time, and the final TD is the basement. By analyzing the drilling and completion data of adjacent wells, the wellbore structure and drilling fluid system are optimized, the drilling tool assembly is optimized, and the drilling parameters are optimized. The application of high-efficiency drilling technology such as torsion impact speed-up tools has enhanced the adaptability of the drill bit, and the speed-up effect is outstanding, realizing anti-deviation and fast drilling.

2 Operation Difficulties and Risk Analysis

2.1 Well Deviation and Lost Circulation

The eastern slope belt of the WN sub-sag shows a structural pattern of high east and low west as a whole. It is cut by a number of NNE and NEE faults, forming many reverse fault blocks within the structural belt. From the seismic data of adjacent wells, this well will encounter faults at about 1540 m in the lower part of the NT 2 section and about 2240 m in the lower part of the TBM group. The fault distances are about 10 m and 40–50 m, respectively. During the drilling operation, attention should be paid to prevent the occurrence of engineering accidents such as lost circulation, well collapse, and well deviation [2, 3].

2.2 Difficult to Control Well Deviation

The formation dip angle encountered in Well W60 is 4.6° in YM group, 7.3° in DMGH group, 8.0° in NT group, and 11.9° in TBM group (see Table 1). The formation dip angle is large, and it is difficult to control the well deviation. Therefore, attention should be paid to preventing deflection during drilling.

Table 1. Statistics of formation dip angle encountered in Well W60

Vertical depth, m	Stratum	Inclination direction, °	Inclination angle, °
390	QYG group	265	0.8
865	Y third and second stage	215	3.2
1175	Y first stage	205	4.6
1335	DMGH second stage	235	7.3
1740	NT first stage	225	8.0
2285	TBM group	104	11.9

2.3 Drill Bit is Easy to Wear

The second spud drilling of NT Formation is interbedded with dark gray mudstone, fine sandstone and glutenite, and interbedded with gray tuff. The middle and upper part of the TBM group is an interbedded silty mudstone and sandy conglomerate, with a section length of 739 m. The lower part of the TBM group and the basement are drilled into large sections of gray, variegated sandy conglomerate, and tuffaceous sandy conglomerate, with drillability grades of 5–6, and the drill bit is easy to wear. The use of PDC bits is limited, and it is very challenging to increase the ROP.

2.4 Low Basement Fracture Pressure Makes Cementing Difficult

Combined with the data from adjacent wells, it can be seen from the three pressure prediction graphs of the formation that the fracture pressure gradient of the gray-green mudstone schist at the base of 2380 m is 1.42 (see Fig. 1). Therefore, leakage may occur during the cementing process, and conventional cementing is difficult.

3 Technical Measures and Countermeasures

As mentioned above, Well W60 has developed faults, large stratum dip angle, easy well deviation, lost circulation, difficult well deviation control, NT group and TBM group glutenite interbedded and long section, the drill bit is easy to wear, the basement fracture pressure is low, and the cementing is difficult. Based on the situation of adjacent wells, the following optimization and guarantee technical measures and countermeasures were carried out on the drilling engineering design of Well W60.
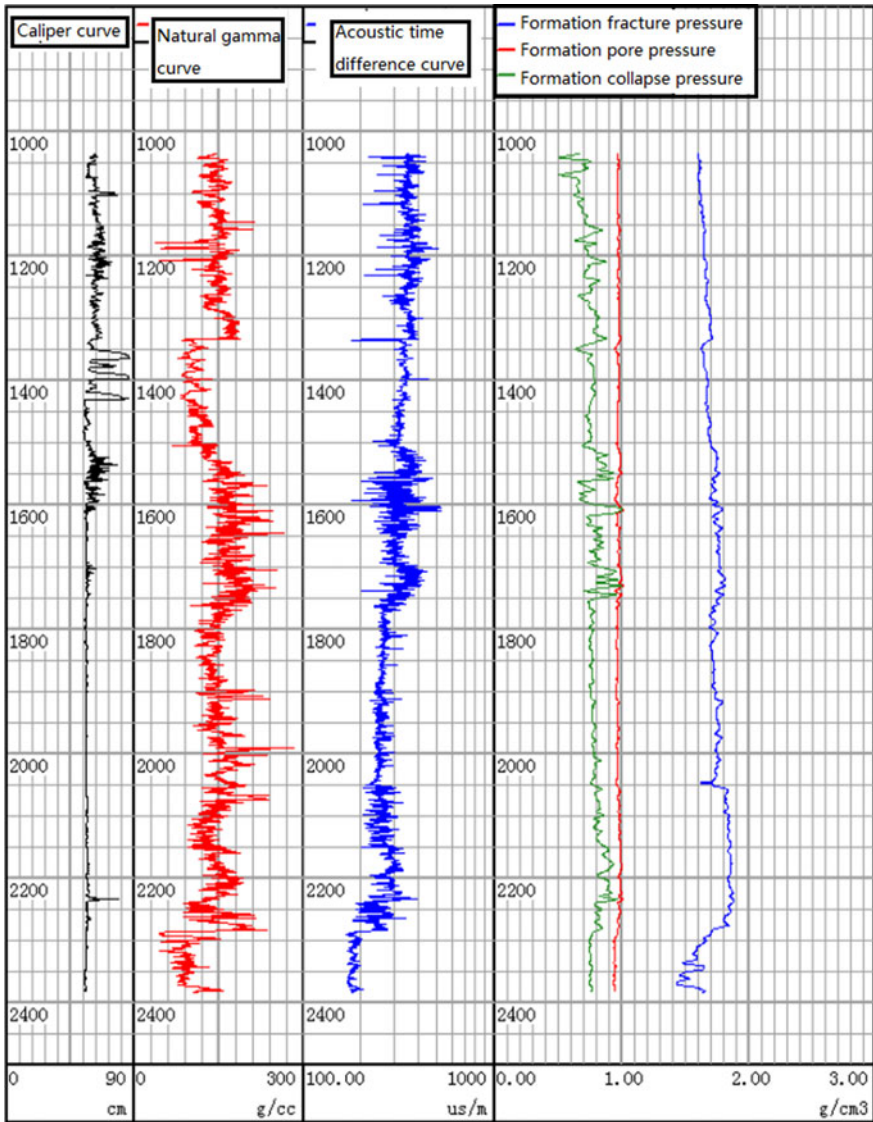
Fig. 1. The three pressure prediction diagrams of the formation in Well W60

3.1 Optimize the Well Structure Design and Promote the Two-Layer Casing Structure

The pore pressure of the formation in the HLE block is normal, and the two-layer well structure design is adopted (see Fig. 2). The surface casing goes down to the
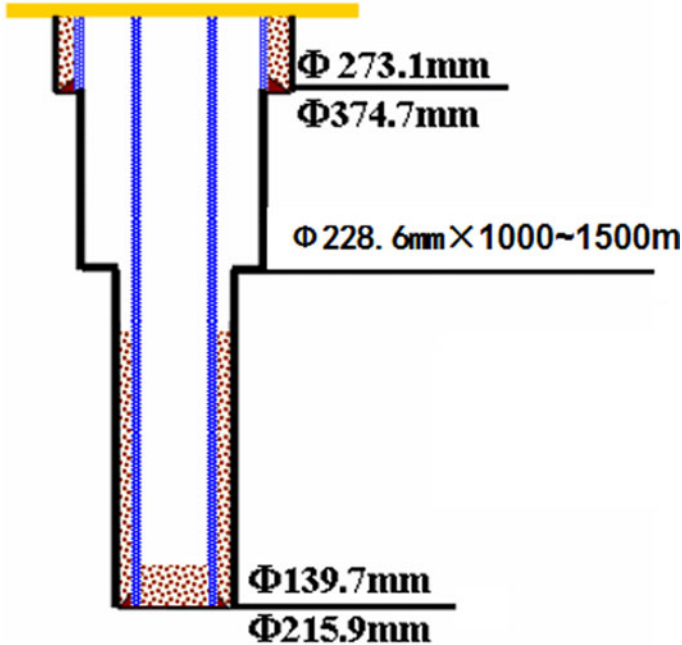
Fig. 2. Wellbore structure design drawing of W60 well

stable mudstone layer of the second and third stages of 400M in Yi 2 and the third section to protect shallow surface water while preventing tight hole, and meet the requirements of safe drilling and pressure test of the second spud. With a two-layer casing structure, the Φ228.6 mm borehole is drilled to the lower part of the Yimin Formation to prevent tight hole and pressure holding, which is beneficial to shorten the drilling cycle.

3.2 Optimizing the Drilling Tool Assembly and Optimizing the Drilling Parameters

W60 well encounters faults and the formation dip angle is large. Through wellplan software analysis, the second opening optimizes the full-bore drilling tool combination and the double pendulum anti-deflection drilling tool combination. The upper part drilling is carried out with a small weight-on-bit (see Table 2), and inclination measurement is carried out every 150 m of drilling in the normal formation. Technical measures such as further control of the weight-on-bit and intensification of the inclination measurement where the fault may be encountered [2–4].

Table 2. Recommended drilling parameters for the second spud of Well W60

Drilling sequence	Drilling formation	Well section, m	Drilling parameters			
			Weight on bit, kN	Rotating speed, r/min	Displacement, l/s	Pump pressure, MPa
Second opening drilling	YM Group	400–1150	20–60	100–180	34–38	10.00
	JD Group	1150–2600	100–180	60–90	32–36	12.00

3.3 Drilling Fluid System Optimization

Potassium salt copolymer drilling fluid system is applied in the second spud interval, and a low-concentration aqueous solution of composite inhibitor and composite ammonium salt are applied to maintain the drilling fluid performance. According to the characteristics of the formation, the density of the drilling fluid is optimized to 1.10–1.15 g/cm^3, which is conducive to exploration discovery and reservoir protection.

3.4 Design Low-Density Cementing to Ensure Cementing Safety

According to the three pressure prediction maps of the formation, the fracture pressure gradient at 2380 m in the basement is 1.42. The second spud design 1.60 g/cm^3 low-density cementing, return To 300 m above the target layer. Perform formation pressure test before cementing construction to ensure cementing safety.

4   Application Effect

4.1 Engineering Effect

Well W60 TBM group Section 1793.24-2243 m, application of torque impact speed-up tool + matching PDC bit (see Fig. 3), single-trip drilling footage 449.76 m, ROP is 5.57 m/h. Compared with adjacent wells, the single-trip drilling footage has been increased by 329.90 m, the ROP increased by 109%, and the speed-up effect has been outstanding. The lower part of the TBM group and the basement are drilled into a large section of tuffaceous sandy conglomerate, the rock drillability is 5–6, the bit is easy to wear, the PDC adaptability is poor, and the footage of a single bit is less. For this design, the high-speed SMD537X roller cone

Fig. 3. Torque impact
speed-up tool + matching
PDC bit

bit is preferred. Enhance the adaptability of the drill bit, the average ROP is 1.63 m/h, an increase of 15.6% year-on-year (see Table 3).

Final TD of Well W60 is 2600 m (intensified 215 m), an average ROP of 6.07 m/h, and a drilling cycle of 32.42d. Compared with the ROP of 5.24 m/h in the adjacent well, the average ROP of the whole well has increased by 15.84%, and the average drilling cycle has been shortened by 15.70% in Well W60. Through the integrated application of speed-up technology such as packed hole assembly, torsion impact speed-up tool, optimized high-speed SMD537X cone bit, and enhanced drilling parameters, the field operation overcomes the problems of faults encountered in the lower part of the TBM group, large inclination angles, and difficult well deviation control. The adaptability of the drill bit is enhanced, the drilling efficiency is improved, and the well deviation is effectively controlled. The maximum well inclination of the whole well is 3.36°/2225.00 m, the average hole diameter expansion rate of the second spud is 7.47%, and the low density cementing of 1.60 g/cm^3 is used, and the cementing quality is qualified after returning to 300 m above the target layer.

4.2 Exploration Effect

W60 well uses the optimized plan of drilling engineering design, and the field operation is safe and efficient. TBM group sandy conglomerate drilled well oil and

Table 3. Speed-up comparison table of torque impact speed-up tools

Well	Size mm	Speed-up type	Well section m	Strata	Footage m	Total drilling time h	Average ROP m/h
W60	Φ215.9	Torque impact tool	1793.24–2243	TBM Group	449.76	80.75	5.57
W58	Φ215.9	/	1355.57–1475.43	NE Stage	119.86	44.92	2.67
Comparison					+329.90	+35.83	+109%
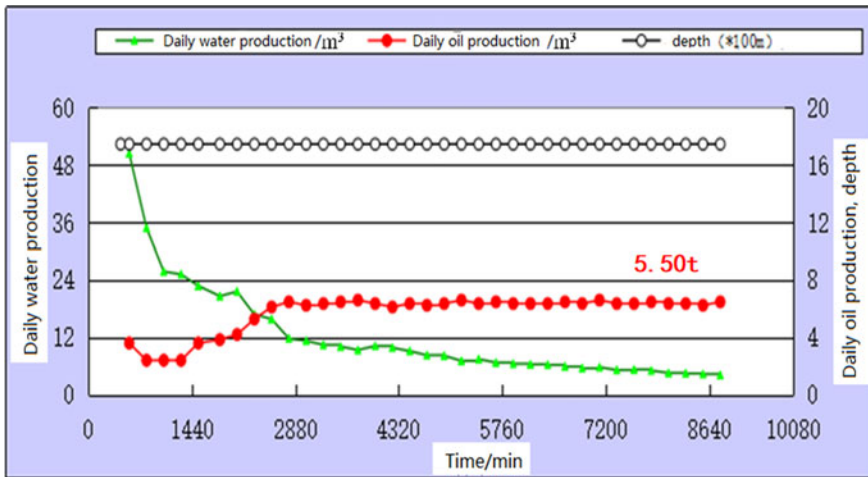
Fig. 4. Production curve of well W60 production layer m^3

gas shows (2 layers of 18.2 m in oil layer, 29.2 m in 8 layers of poor oil layer). According to the statistics of oil test production, the daily oil production of W60 TBM group was 5.50t (see Fig. 4), and industrial oil flow discovery was obtained, which found a high-yield oil-rich zone for the WN sub-concave eastern slope. This exploration was achieved. The purpose of exploration discovery has achieved great exploration results.

5 Conclusions

(1) Well W60 encounters faults and has a large formation dip. By optimizing the packed hole assembly, optimizing drilling parameters, and infilling inclination measurement, the well deviation is well controlled and the wellbore quality is ensured.

(2) The successful drilling of Well W60 verifies the feasibility of the WN sub-sag with a two-layer well structure and a drilling fluid density of 1.10–1.15 g/cm^3, which is conducive to exploration discovery and reservoir protection.

(3) Well W60, through the integrated application of torque impact speed-up tools and optimization of high-speed SMD537X cone bits, enhanced drilling parameters and other speed-up technologies, compared with adjacent wells, the average ROP increased by 15.84%, and the drilling cycle was shortened by 15.70%, achieving better results. Speed up and improve efficiency.

(4) The on-site operation of Well W60 is safe and effective. The sandy conglomerate of the TBM group has encountered good oil and gas. The daily output after compaction is 5.50t, and industrial oil flow discoveries have been obtained. It has found a high-yield oil-rich zone for the east slope of the WN sub concave. For the purpose of geological exploration, it has good exploration and economic benefits.

References

1. Liu, G.H., Li, Y.M., Li, J., et al.: New technology with composite percussion drilling and rock breaking. Petrol. Drilling Tech. **44**(5), 10–15 (2016)
2. Chen, L.-L.: Research and application of anti-deviation and fast drilling technology in easily inclined formation of Hailar. Explor. Eng. (Rock Soil Drill. Tunn. Eng.) **40**(2), 21–25 (2015)
3. Zha, C.Q., Liu, G.H., Li, J., et al.: The rock breaking mechanism of the compound percussive-rotary drilling method with a PDC bit. Petrol. Drilling Tech. **45**(2), 20–24 (2017)
4. Zha, C.Q., Liu, G.H., Li, J., et al.: Analysis of torsional vibration characteristics of PDC bit and design of the damping tool. China Petrol. Mach. **45**(3), 1–5 (2017)

Optimization and Application of Drilling Period Prediction Method for Deep Exploration Wells in Daqing

Yue Qi, Jun Li, Gonghui Liu, Yudong Tian, Xin Zhang, and Yong Yang

Abstract With the rising production cost of drilling and the needs of market expansion at home and abroad, the scientific budget of the cost of deep exploration wells is required, and the accurate and reasonable prediction of drilling period is becoming more and more important. The deep exploration wells in AD area of Daqing SL Basin mainly explore the gas reservoirs of YC formation and SHZ formation, which are characterized by high geothermal gradient, high rock hardness and many complex situations. At the same time, the distribution of volcanic breccia and tuff, the great difference of lithology in different strata and well sections and other factors seriously affect the prediction of drilling period. In this paper, the influence of the above factors on drilling period prediction is analyzed, and the characteristics of drilling production and geological stratification in different areas of AD area of Daqing Oilfield are taken into account. Based on well history data, a set of period prediction methods suitable for deep exploration wells in AD area are selected by using many period prediction methods such as learning curve and periodic quota method. these methods are used to guide the drilling period prediction of deep exploration wells in Daqing Oilfield. The field practice shows that the error range of the improved drilling period prediction is less than 10%. This study has a certain reference significance for the drilling period prediction in oil field.

Keywords Drilling period prediction · Learning curve · Periodic quota method · Deep exploration well

Y. Qi · J. Li (✉) · G. Liu · Y. Tian · X. Zhang
China University of Petroleum-Beijing, Changping, Beijing 102249, China
e-mail: lijun446@vip.163.com

Y. Qi · Y. Tian · Y. Yang
Daqing Drilling Engineering Company, Honggang District, Daqing 163511, China

J. Li
China University of Petroleum-Beijing at Karamay, Karamay 834000, China

1   Introduction

The drilling period prediction of deep exploration wells in Daqing Oilfield is affected by the lack of geological data, concentrated man-made geographical factors, and different drilling purposes. It has always been a difficult problem in the exploration well industry. At the same time, the problem also includes the diversity of drilling technology, existing equipment and technical capabilities. There are many prediction methods and the scope of prediction is limited. So far there is no scientific method for predict the drilling period of deep exploration wells. The usual drilling period prediction generally include: historical average method, typical case method, and process man-hour method. For Daqing Oilfield, the period prediction error is larger, the limitation is larger, and the scientific basis is less. Therefore, we can only learn from similar adjacent wells, which are in same formation or have the same casing program. According to statistics, of the 16 wells with the same casing program in block AD, 10 of them have a relative error of over 15% in the drilling period. The period design is greatly influenced by man-made, and the prediction process lacks scientific basis. The prediction method is relatively simple, and the accuracy is not high. Therefore, in order to achieve the standardization of period prediction and the maximization of expected benefits, new scientific drilling period prediction methods are urgently needed.

2   Technical Status

Currently, the commonly used drilling design period prediction methods include: historical average method, typical case method, and process man-hour method [1–4]. See Table 1 for Specific application range and characteristics. The historical average method is widely used in the mid-shallow well period prediction of a large number of repeated production in the later stage of the oilfield exploration, this method has limitation and high accuracy without combined with other methods. The typical case method and the process man-hour accumulation method refer to the well data of other blocks for calculation, and the methods rely on the adjacent well experiences to some extent. The process man-hour method is very suitable for factory drilling and offshore drilling, and the reference well cannot have large lithology and process changes. The periodic quota method is also called the weighted periodic quota method. The drilling and completion data that needs to be referenced are very recent. The more the data, the more accurate it is. The wells selected under the same conditions can also be divided and filtered by period length to improve prediction Accuracy. The learning curve method takes into account the ability of the co operation party during the entire time period. In the time span from the first well to the last well from the beginning of the project, all the factors that have an impact on the exploratory well period are transformed into effective and regular data, the standard for sample selection of abnormal data points is established, and the deep

Table 1 The scope and characteristics of the main period prediction methods

Method	Features and limitations
Historical average	Simple calculation, use when there are few data, limited to adjacent wells, recent and the same type of well data reference is more accurate, ≥ 2 well average calculation method
Typical case	Use the same type of well period in other blocks for reference, measure geological factors less, ≥ 1 wells, subject to subjective consciousness, and have large errors
Process man-hour	Under the premise of being familiar with geology, engineering process and operation process, irregularity, cumbersome calculation, less consideration, ≥ 1 well, subjective tendency, large error
Rock drillability	As long as the adjacent well has rock seismic or logging data, it can be predicted, and the pure drilling time prediction is accurate. It is accurate without considering other strokes, drilling speed, process changes, and time span changes. Only rock factors can be considered
Period quota	There are many well completion data that can be referred to. In the recent repeated operation of similar processes, one influencing factor can be considered temporarily, and ≥ 3 wells can be used
learning curve	Including learning ability, small calculation errors, increasing data, unlimited reference factors, and screening of abnormal data, the more accurate the prediction. Currently only used in newly opened blocks and well sections, ≥ 3 wells can be predicted

exploratory well oil layer in AD area is drilled. There is still a lot of room for improvement in reservoir drilling, and the learning curve can predict the periodic change of exploratory wells according to the learning and improvement ability of the operation party in this block [2–5]. A currently popular rock drillability prediction method requires logging or seismic data from adjacent wells with the same casing program, infers the accurate ROP of different rocks, and calculates the pure drilling time. According to the pure drilling time, set an artificial ratio to introduce the interval or whole well ROP.

3 Difficulties in Period Prediction of Deep Exploratory Wells in AD Area

(1) The data needs to go through a large number of statistics, quantitative analysis and multi-level screening, and the calculation formula is complicated. It takes a relatively long-term process to realize the standard, simple use and software intelligence.

(2) There are many new formations encountered in exploration well, and the lithology and technology are complex and changeable. It is very difficult to reduce the period prediction errors by innovating a new method using upper lithologic drillability data.

(3) There are many period prediction methods, combining multiple prediction methods to standardize different data usage and limitation, and to innovate a set of template or standard for exploratory well period prediction in AD area.

4 Analysis of the Main Influencing Factors of Period

The main influencing factors of an exploratory well period are: geographic location, geology and lithology, drilling technology, drilling tools and bits, drilling fluid technology, non-production time, complex accidents, additional change time, and the operation party's equipment and personnel management and technology level. These influencing factors have a greater impact on time, stratum lithology, operation plan, and even the management ability of the operation arty.

It is mainly divided into two parts, similar drilling and completion in the near future, and casing drilling, which still has much room for improvement. There are also drilling cores with small proportions that have not changed much for many years. Irregularities are difficult to predict. Items such as non-production time, change and measurement. These period calculations need to be subtracted in advance, and the changes of the first two major parts are predicted and then added later.

5 Optimization and Improvement of Period Prediction Method

The period prediction of different intervals of exploratory wells needs to consider technology, equipment and drilling tools, rock lithology, and shorten the period generated by technological progress across different ages. At the same time, consider avoiding the use of unreliable man-made data and expanding the general scope of existing data. However, there are too many factors influencing the drilling process of exploratory well, and the calculation of period quota method and learning curve method are complicated and need to be improved.

The upper well section and casing completion in AD area have high repetition, recent similarity is high, and drilling data are more and more accurate. The 3.5.2 weighted period quota method commonly used by PetroChina is adopted. The oil layer uses the data of all wells with similar casing program in the AD area from the first well in history to the last well, and the prediction period span of the well section is easy to use the learning curve method to follow the law of scientific progress (see Table 2).

Table 2 Optimization of drilling period prediction methods

Method	Instructions
Moving, completion	Periodic quota method and additional process plus
Surface and intermediate casing drilling	Periodic quota method
Production casing drilling	Learning curve
Complex accidents, waiting	Irregular, no prediction
Drilling coring	Average number of single tube coring

5.1 Improved Weighted Period Quota Method

At present, the "3, 5, 2" weighted period quota method is mainly adopted by China Petroleum Drilling Man-hour Quotas, which means that the fastest period weight accounts for 30%. In accordance with the principle of average advancement, the weight of the intermediate period is 50%, and the weight of the slowest period is 20%. The obtained period value is slightly lower than the average of the recent period, reflecting the trend of continuous shortening of the period.

"3, 5, 2" weighted period quota prediction method, the Surface drilling (L_d), intermediate casing (L_u) drilling section "golden section" period quota calculation formula is:

$$L_d = 0.618\left(\sum_{i=1}^{n} T_i / \sum_{i=1}^{n} H_i\right) \tag{1}$$

$$L_u = 1.382\left(\sum_{i=1}^{n} T_i / \sum_{i=1}^{n} H_i\right) \tag{2}$$

Where T is the working-hour quota under certain wellbore size or section conditions, h/m; T_i is the periodic quota under certain wellbore size or section conditions, h/m; n is a certain wellbore size or The number of wellbore statistics under well section conditions, n = 1, 2, 3, ……, i; Hi is the actual section length of a single well under certain wellbore size or section conditions, m.

Due to the actual drilling situation of the upper strata with a relatively fast ROP, only the well depth is introduced as a necessary reference condition and cannot be used as a basis for period prediction. The trip takes up a large proportion of the drilling time in the actual operation. The time spent is basically the same, and the number of drill bits used is basically proportional to the period. Therefore, in addition to the depth of the well as the necessary reference for each well, the number of drill bits can be used as the main reference.

The "3.5.2" weighted period quota method has been improved to apply to the drill section of the surface and intermediate casing. The "golden section" selection method in Formulas 1 and 2 is modified.

$$L_d = 0.618\left(\sum_{i=1}^{n} T_i \Big/ \left(\sum_{i=1}^{n} H_i \times \sum_{i=1}^{n} Z_i / n\right)\right) \tag{3}$$

$$L_d = 1.382\left(\sum_{i=1}^{n} T_i \Big/ \left(\sum_{i=1}^{n} H_i \times \sum_{i=1}^{n} Z_i / n\right)\right) \tag{4}$$

5.2 Improved Weighted Period Quota Method

5.2.1 Drilling Learning Curve Method

The current drilling learning curve method uses the modified learning curve method [1–5], based on the learning curve equation originally proposed by Mr. T.P.Wright:

$$T = r^{log_2 n} \tag{5}$$

where: n is the cumulative number of wells produced; r is the learning rate; T is the drilling period of the first well.

Liu Chaoquan introduced the well depth H as a reference, $T = e^{aH + c} r^{log_2 n}$, applied to 20 wells in Xijiang Oilfield, and the prediction period was accurate. However, the application in many other blocks has large errors. The main reason is that although there is a certain connection between well depth and period [1–5], the main factor is formation lithology.

5.2.2 Improvement of Drilling Learning Curve Method

The main influencing factors of the drilling period are stratum lithology, technology level, drilling tools and bit selection, solid content of mud, equipment management and maintenance, management operation level and casing program design. Many of these interact synergistically to affect the trend of the period of exploratory wells, and their period influence weight coefficients are obtained by using the learning curve method and the multiple equation regression method. Accurate field data is obtained from the drilling summary, after sorting and collecting, the trend curve analysis is performed with the popular Eview software to find out the key influencing factors of the period prediction.

(1) Drilling time sequence.

In a certain event period, the same type of drilling production is bound to follow a certain development and change law as long as it is a continuously improving production unit. The drilling team and process production methods in the SL Basin have been stable for a long time. The standard casing program in AD area changes with the drilling sequence and the exploratory well period decreases as shown in Fig. 3.

It can be seen from Fig. 1 that with the time sequence from far to near, the period has a trend of shrinking continuously, which roughly conforms to the law of the one-variable linear curve. The learning curve can well represent the decreasing trend of drilling period caused by drilling tools and bit selection, technological progress, well design optimization, equipment update and personnel management improvement during the drilling process, and follow the learning curve formula.

(2) Rock lithology

The number of rock lithology directly affects the pure drilling time of the drilling period. In the past, the learning curve only considered the time change trend, and only wells with similar blocks in the same block could be selected. Effective means to predict the period. In order to increase the scope of application, the learning curve exploration well period prediction needs to introduce lithological data.

The factors that have an impact on the period are the length of the well section and the rock hardness. The rock hardness data is obtained from the deep and full-stratum rock drillability prediction research project. The research processes the geological and lithological data of the SL Basin and divides the entire oil interval into i For a 0.125 m long micro well section, the total footage hardness formula for the entire well section is:

$$X_3 = \sum_1^i L \times H_i \tag{6}$$

L is the unit section length, 0.125 m; Hi is the rock hardness of the i-th section length, MPa; X3 is the total footage hardness of the oil zone drilled into the entire well section, km·MPa.

The trend of the standard well structure exploratory well period with the footage hardness in AD area is shown in Fig. 2.

3) Drilling assistance

General exploratory well period design estimates the interval drilling speed time based on pure drilling time, such as drilling auxiliary time, tripping time, pipe connection time. From the data in the completion data, only There is a great relationship between the drilling auxiliary time and the number of drill bits and the depth of the drilling. The relationship between the period and them is shown in Fig. 3 and Fig. 4.

Fig. 1 Trend chart of period and drilling sequence curve
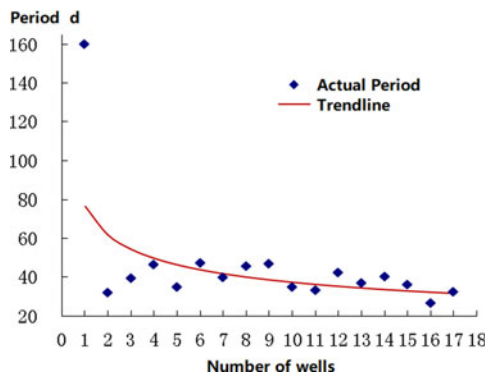
Fig. 2 Period and trend chart of rock meter hardness curve
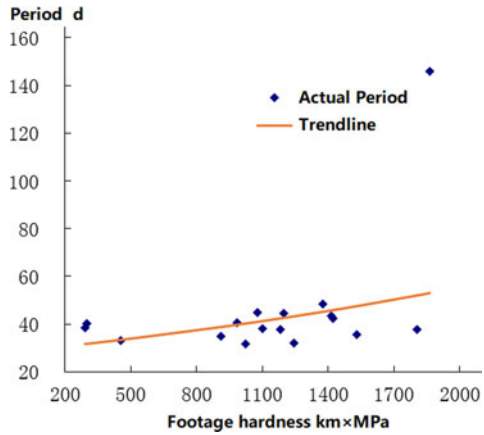
Fig. 3 Period and the number of drill bits used curve trend chart
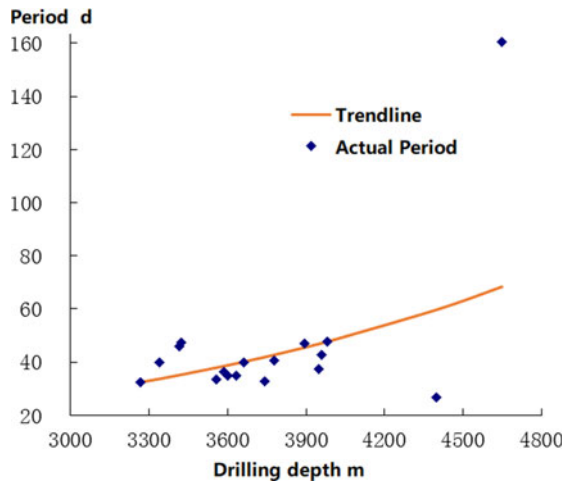
4) Solid phase content

The solid phase content of fine particles has a great influence on the drilling speed, but the solid phase content of fine particles cannot be obtained in the completion data. The density of the drilling fluid for exploratory wells is low, and most of the solid phase is formed by the useless solid phase dispersion of the well wall. The low-density solid phase content LGS can be transformed by density. The formula is:

$$LGS = \frac{(\rho_i - \rho_w) \times 100}{(\rho_s - \rho_w)} \qquad (7)$$

Fig. 4 Trend chart of period and completion depth curve
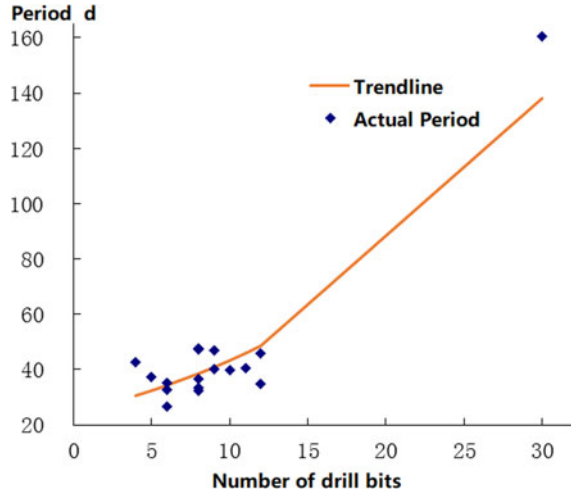
Fig. 5 Period and low-density solid content curve trend graph
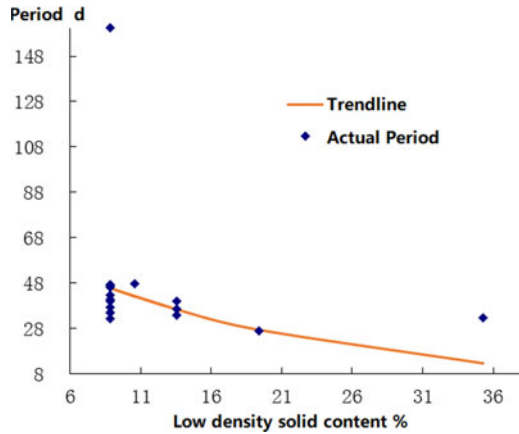
LGS is the low-density solid content, %(v/v); ρ_i is the maximum drilling density of the oil interval of the i-th well; ρ_w is the density of water and 1.0 g/cm^3; ρ_s is the density of low-density solids in the drilling fluid Take 2.7 g/cm^3.

It can be seen from Fig. 5 that the period decreases in the opposite direction with the increase of the low-density solid content concentration, which does not match the actual situation. This shows that the influencing factor of low-density solid content does not conform to the periodic change trend in AD regions.

5.3   Software Screening

After preliminary screening by conditions, only from the visual point of view of the trend chart, the time sequence, the number of drill bits used, the rock hardness and the completion depth are all in line with the trend direction. The regression software is used to calculate the X1, X2 and X3 The F-statistic Prob test value is low. The Prob. value of X5 is as high as 0.6696, which is higher than 0.5. Try to filter out and see the changes in the test value.

After screening out the influencing factor of completion depth, the Prob. value of other factors dropped to 0.5 overall, and the effect was obvious after deletion, indicating that the completion depth is used as an influencing factor to interfere with the effect of period prediction, and rock hardness can be completely replaced Well depth as an influencing factor.

5.4   Selection of Sample Abnormal Points

Periodic statistical data will have abnormal data beyond the expected range, it is inevitable that there will be irregular data interference in drilling production. In addition, artificial tendencies and deviations are inevitable in the data statistics process. In order to achieve a certain expected law of the learning curve, it is necessary to screen data samples for abnormal points.

The learning curve abnormal point screening method uses the relative error and absolute value of the prediction period and the actual period of multiple samples to take the average value, and the "golden section" is used to set the upper limit filter, as shown in formula 8. The absolute value of the relative error between the predicted period and the actual period exceeds Lu's completion data without consideration (Fig. 6).
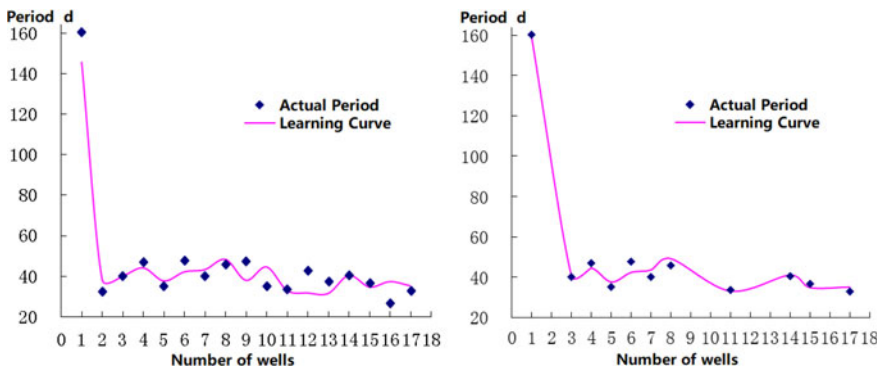
Fig. 6   Effect diagram of 17 wells in AD area after sample screening

$$Lu = 1.382 \times \left(\sum_1^n \left| T_{pi} - T_{ai} \right| / T_{ai} \right) / n \tag{8}$$

6 Field Application Situation

According to the drilling and completion data, the period quota method selects samples from 2007 to 2018. The learning curve method requires selection from the first well drilled in the AD area to 2018, and predicts DS28 and DS34 For the well period, select all completion data with three-layer standard wellbore structure, vertical wells, casing perforation completion methods, and water-based drilling fluid system conditions for compliance calculations.

Well DS28 has a standard three-casing well structure with a intermediate casing section of 2000 m. conventional water-based drilling fluid system is applied in the entire well section. One drill bit is applied in the intermediate casing interval, and the oil layer casing is perforated. The actual well operation period is 83.35 days. There are 6 drill bits used in the oil zone, the rock hardness encountered is 1531.525 km·MPa, and the oil zone length is 1504 m. According to the sequence, it is the 18th well of the ø215.9 wellbore in the AD area. The drilling time of the oil zone is estimated by Eviews software to be 34.99d. The drilling period after accident removal, drilling, waiting and coring is 30.54d. the prediction period of Well DS28 is 81.11 days, the actual operation period is 81.20 days, and the relative error is −0.11% (see Table 3).

Well DS34 has a standard three-casing well structure with a intermediate casing section of 1928 m. conventional water-based drilling fluid system is applied in the entire well section. Two drill bits are applied in the intermediate casing interval, and the oil layer casing is perforated. The actual well operation period is 85.31 days. There are 4 drill bits used in the oil zone, the rock hardness encountered is 1198.167 km·MPa, and the oil zone length is 1258 m. According to the sequence, it is the 19th well of the ø215.9 wellbore in the AD area. The drilling time of the oil zone is estimated by Eviews software to be 25.74d. The drilling period after accident removal, drilling, waiting and coring is 17.10d. the prediction period of Well DS34 is 78.05 days, the actual operation period is 72.51 days, and the relative error is 7.64% (see Table 3).

Table 3 Compliance with drilling period prediction in AD area

Name	Well DS28			Well DS34		
	Actual time d	Prediction time d	Relative error %	Actual days d	Prediction time d	Relative error %
Period	81.20	81.11	−0.11	72.51	78.05	7.64

7   Conclusions

(1) The learning curve and the "3, 5, 2" weighted period quota method in period prediction are the basis of period prediction in the drilling industry, which can fully reflect the trend of exploratory well period changes. After innovative and improved formulas and screening criteria, it can adapt to the drilling production characteristics of Daqing AD area. This has certain reference significance for the on-site drilling period prediction of Daqing Oilfield.

(2) The learning curve method is a scientific data statistical analysis method. The combination of screening with the Eviews software and then screening with the "golden section" makes the period prediction of each process have scientific basis. The creation of footage rock hardness and the improvement of the calculation process are the key to the technology. The newly added rock lithology data promotes the integrated utilization of drilling engineering and geological lithology data.

(3) The research is based on the statistical analysis of drilling and completion data, and there are accurate drilling and completion data and lithology data. After conditions, software and sample screening, unimportant influencing factors and abnormal data interference are eliminated, and the relative error of the improved drilling period prediction is within 10%. The actual prediction data tends to curve development law, which is a new scientific and reasonable drilling period method.

References

1. Zangl, G., Oberwinkler, C.P.: Schlumberger. SPE Annual Technical Conference and Exhibition, Houston, Texas, 26–29 September 2004 (2004)
2. Huang, W., Zheng, P.: The establishment of a new method for drilling cycle prediction. Nat. Gas. Ind. **26**(9), 93–95 (2006)
3. Zhang, W., Zhu, D., Zhang, Y., et al.: New method for drilling cycle prediction. Nat. Gas Ind. **33**(3), 68–71 (2013)
4. Liu, C., Cai, J., Zhai, S., et al.: Using the modified learning curve to measure the drilling time of a well section. J. Univ. Pet. (Nat. Sci. Ed.) **22**(4), 41–42 (1998)
5. Sheng, L., Yang, W., Tang, H., et al.: Deepwater drilling cycle learning curve analysis. Pet. Drilling Technol. **34**(9), 91–94 (2012)

Influence of the Position of Flex Sub on the Build-Up Capacity of the Static Push-The-Bit Rotary Steering Tool

Yantao Bi, Gonghui Liu, Guanghua Dong, and Jun Li

Abstract As one of the most widely used rotary steering tools in China, static push-the-bit rotary steering tool is widely used in shale gas, offshore oil and gas development. The application of this tool not only shortens the drilling cycle but also improves the drilling safety. In order to improve the build-up capability of rotary steering tools, the flex sub is always contained in the bottom hole assembly (BHA) of static push-the-bit rotary steering tools. In this paper, it establishes the mechanical model for the BHA in the static push-the-bit rotary steering tool with flex sub, analyzes and calculates the influence of flex sub's position on the build-up capacity of static push-the-bit rotary steering tool under various BHAs. The analysis and calculation show that the position of flex sub has different influences on the build-up capacity of rotary steering tool under different BHAs. The rotary steering tool equips the flex sub alone that provides the maximum build-up capacity.

Keywords Static push-the-bit rotary steering · Flex sub · Bottom hole assembly (BHA) · Build-up capacity

1 Introduction

With the increasing difficulty of oil and gas development, the well bore design has become more and more complex. The traditional sliding steering drilling tools have not meet the operation requirements of modern directional drilling, especially in extended reach wells and long-distance horizontal wells [1, 2]. Thus, the rotary

Y. Bi (✉) · G. Liu · J. Li
College of Petroleum Engineering, China University of Petroleum-Beijing, Changping, Beijing 102203, China

Y. Bi · G. Dong
Drilling Technology Research Institute of Shengli Petroleum Engineering Company, SINOPEC, Dongying 257000, Shandong, China

G. Liu
Beijing University of Technology, Chaoyang, Beijing 100023, China

285

steering drilling technology emerges as the times require. With its rotary drilling feature, the rotary steering drilling tool overcomes the large friction resistance in the drilling of sliding steering drilling tool, and also reduces the tripping operations, and causes the borehole quality maintain smooth, so as to improve the development efficiency and economic benefits of oil and gas wells. It has gradually developed into one of the most core technologies in modern drilling engineering and brought oil drilling a new technological revolution [3–6].

At present, the most mature commercial application of rotary steering tools is mainly concentrated in the three major international petroleum technology service companies, including Baker Hughes' Auto Trak, Schlumberger's Power Drive and Halliburton's GeoPilot [7–9]. Among the above systems, the Auto Trak is a static push-the-bit rotary steering tool with strong build-up capacity and good stability, which has realized large-scale application.

The steering ability of rotary steering drilling tools can be expressed by their build-up capacity which can be characterized by the build-up rate. The basic factors affecting the build-up rate of rotary steering drilling tools mainly include the parameters of: bit characteristic, BHA structure, rotary steering tools' characteristics, borehole geometry, drilling process and formation characteristics [10, 11]. In the design and application of BHA in the static push-the-bit rotary steering tool will include the flex sub. The design of flex sub is directly related to the mechanical characteristics of BHA, which then affects the realization of build-up function of rotary steering drilling tool. Therefore, the design and application of flex sub is one of the priorities in the design of BHA in the rotary steering drilling tool.

Through the static analysis of the BHA in static push-it-bit rotary steering system, this paper studies the influence of the flex sub's position on the build-up capacity of the rotary steering tool with different BHAs, and evaluates the function of the flex sub.

2   Calculation Model of Build-up Capacity in the Static Push-the-bit Rotary Steering System

The mechanical analysis model of rotary steering variable cross-section is expressed by the following form [12, 13], and the schematic diagram is shown in Fig. 1

$$
\begin{cases}
E_i I_i \frac{d^4 X_i}{dl^4} = -M_{ii} \frac{dY_i}{dl^3} + (q_i l \cos \alpha_i - Z_i) \frac{d^2 X_i}{dl^2} + q_i \frac{d^2 X_i}{dl} \cos \alpha_i + q_i \sin \alpha_i \\
E_i I_i \frac{d^4 Y_i}{dl^4} = M_{ii} \frac{dX_i}{dl^3} + (q_i l \cos \alpha_i - Z_i) \frac{d^2 Y_i}{dl^2} + q_i \frac{d^2 Y_i}{dl} \cos \alpha_i
\end{cases}
\tag{1}
$$
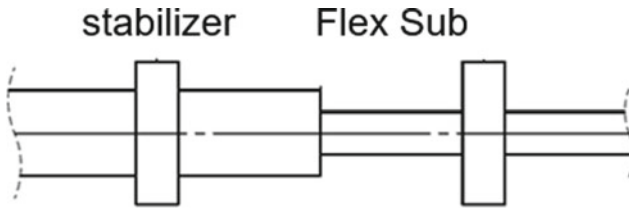
Fig. 1 Schematic diagram of stress on the variable cross-section in BHA with flex sub

where:

$$Z_i = Z_1 - \sum_{j=1}^{i=1} (q_i L_j \cos \alpha_j - N_j f_a) \tag{2}$$

$$M_{ii} = M_{i1} - \frac{f_r D_w}{2} \times \sum_{j=1}^{i=1} N_j \tag{3}$$

$$f_n = \frac{2vf}{\sqrt{4v^2 + (\omega d_w)^2}} \tag{4}$$

$$f_t = \frac{\omega D_m f}{\sqrt{4v^2 + (\omega d_w)^2}} \tag{5}$$

L(m) = curve coordinate along the axis of drilling string, starting from the bottom side of section-i in the drilling string; E_i(MPa) = elastic modulus of drilling string in section-i; I_i(m4) = moment of inertia of section-i drill string; M_{ii}(N·m) = the torque of section-i, also as the bit torque when i = 1; q_i(N/m) = weight of section-i of the drill string in drilling fluid; α_i(radian) = well angle of section-i; X_i(m) = displacement or coordinate of section-i in X direction; Y_i(m) = displacement or coordinate of section-i in Y direction; Z_i(kN) = pressure in Z direction at lower part of section-i, the pressure is positive, and it is the drilling pressure when i = 1; f = friction coefficient between centralizer and well wall; v(m/h) = rate of penetration; w (r/min) = rotational angular velocity of the drilling string; d_w(m) = well diameter.

3 Types and Calculation Parameters of BHA in Static Push-the-bit Rotary Steering System

The naming rules of BHA model in rotary steering system with flex sub are as follows:

FS-A-(B1, B2, … Bn), (Bn ≦A).

where

FS = flex sub;

A = the number of centralizers (the rotary steering tool is already equipped with centralizers);

BN = the No.BN centralizer with a flex sub at the back of itself.

Table 1 Analysis and schematic diagram on the guidance force of the BHAs with flex sub

Model	sketch map
FS-1-(1)	
FS-2-(1)	
FS-2-(2)	
FS-3-(1)	
FS-3-(2)	
FS-3-(3)	
FS-4-(1)	
FS-4-(2)	
FS-4-(3)	
FS-4-(4)	

According to the possible location of flex sub, BHAs with flex sub can be divided into five types (as showed in Table 1 which lists the corresponding BHA structure diagram).

The parameters of flex sub are: length = 1.2 (m), outer diameter = 122 (mm), inner diameter = 71.4 (mm), weight of the drill string = 588.3 (n/m), elastic modulus = 205,800 × 106 (Pa).

Other parameters are as follows: the diameter of the drill bit is 215.9 (mm); the outer diameter(near the bit) of the drill collar is 165 (mm), and 172 (mm) in the rest; the inner diameter of the collar is 71.4 (ms); the weight near the bit is 122672 (N/m), the rest is 1472.17 (N/m); the elastic modulus is 205800 × 106 (Pa); the deviation angle is 60° and the bit pressure is 80 (kN); the outer diameter of the conventional centralizer is 212 (mm).

4 Influence of the Position of Flex Sub in BHA on the Guidance Force

While the static push-the-bit rotary steering tool is under the modes of angle building and dropping (the eccentricity was 2 mm), the influence of the positions of flex sub on the build-up capacity was calculated. When the tool is under the angle building mode, the deflection angle is 180°, while it is under the dropping mode, the deflection angle is 0°.

4.1 The Flex Sub is Placed Between the First and Second Centralizers

When the steering tool is under the angle building and the flex sub is placed at different positions between the first and second centralizers, the variation of the lateral force on the bit is showed in the Fig. 2. The Fig. 2 shows that whether the BHA equips double centralizers, three centralizers or four centralizers, when the flex sub nipple is placed between the first and the second centralizers, the build-up capacity forms a "increase-steady-increase" trend due to the backward moving of the flex sub. The variation range basically keeps the same. Besides, FS-3-(1) owns the largest build-up capacity, while the FS-2-(1) owns smaller capacity. As the steering tool maintaining in the angle building, the azimuth forces of the three types are almost zero.

Figure 3 shows the variation of the lateral force on the bit when the steering tool is in the dropping mode. It can be seen from Fig. 3 that when the flex sub is placed between the first and the second centralizers, whether the BHA equips double

Fig. 2 Variation of the guidance force according to the distance between the flex sub and the first centralizer in the building mode
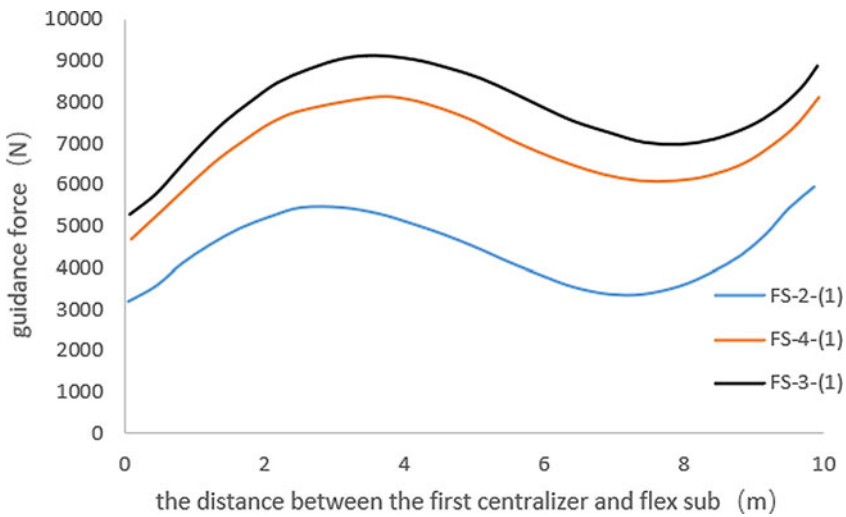
Fig. 3 Variation of the guiding force with the distance between the flexible nipple and the first centralizer in the downhill mode

centralizers or three centralizers, the building-up capacity forms an "increase–decrease—increase" quasi-sinusoidal curve which follows backward moving of the flex sub. The variation range is basically the same. Similarly, FS-3-(1) employs the largest building-up capacity, while the FS-2-(1) employs a smaller capacity.

4.2 The Flex Sub is Placed Between the Second and Third Centralizers

When the steering tool is under the angle building mode and the flex sub is placed at different positions between the second and third centralizers, the variation of the lateral force on the bit is showed in the Fig. 4. Whether the BHA equips three centralizers or four centralizers, the building-up capacity changes as "decrease-slightly increase–decrease" as following the backward moving of the flex sub. It can be seen that the capacity of FS-3-(2) is larger than that of FS-4-(2) as a whole.

Figure 5 shows the variation of the lateral force on the bit due to the position changes of the flex sub when the steering tool is in the angle dropping mode. It can be seen that whether the BHA equips three centralizers or four centralizers, the build-up capacity changes with the backward movement of the centralizer with the trend of "decrease-slightly increase–decrease". Among them, the build-up capacity maintains positive with the FS-3-(2) in a large range (0.5–9 m), while it becomes negative when it is in the FS-4 -(2) where the distance between the flex sub and the former centralizer is more than 1.5 m.
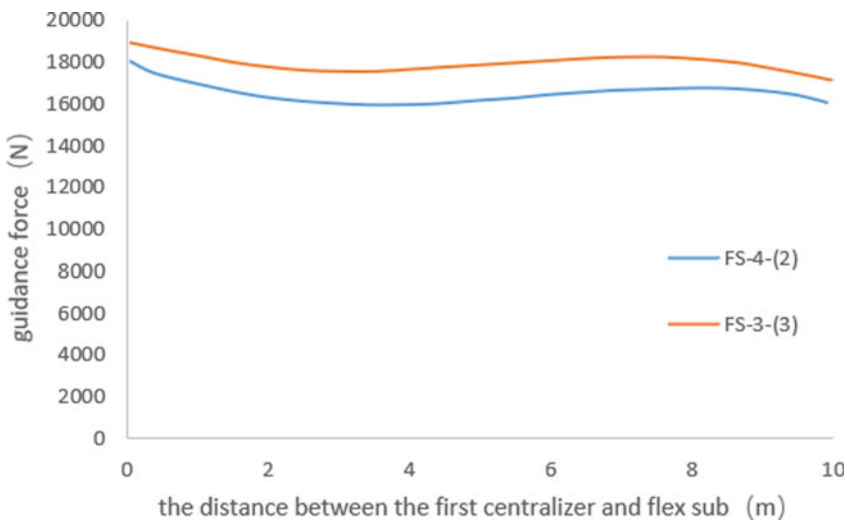
Fig. 4 The variation of the guidance force with the distance between the flex sub and the second centralizer under the angle building mode
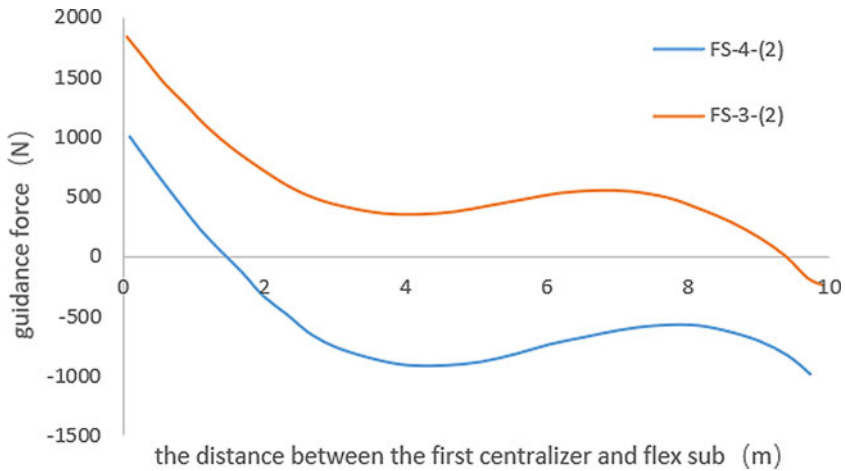
Fig. 5 The variation of the guidance force with distance between the flex sub and the second centralizer under the angle dropping mode

4.3 The Flex Sub is Placed at the Last Centralizer

When the steering tool is under the angle building mode and the flex sub is placed at different positions in the last centralizers, the variation of the lateral force on the bit is showed in the Fig. 6. It can be seen from the Fig. 6 that the impact of flex sub on the bit's building-up capacity becomes smaller and smaller with the increase of centralizers. When the bit equips with one centralizer, the variation range of its building-up capacity is up to 15 kn, while that of the four centralizers is almost unchanged.

Figure 7 shows the variation of the lateral force on the bit due to the position changes of the flex sub when the steering tool is in the angle dropping mode. It shows that the impact of flex sub on the building-up capacity of bit becomes smaller and smaller with the increase of centralizers. When the eccentricity is 2 mm, only FS-2-(2) shows the characteristics of dropping force.

4.4 Comparison of Guidance Forces

It can be seen from the above calculations that the BHA of FS-1–1 has the largest building-up capacity. The ranking relations of the above models are shown in Table 2.

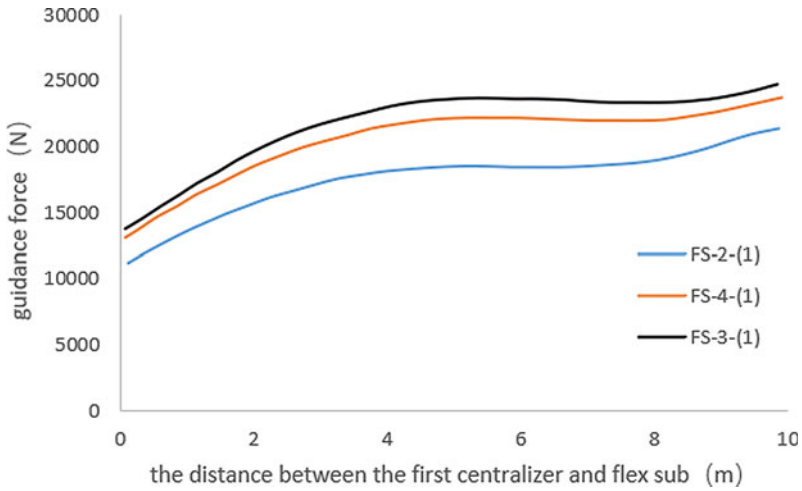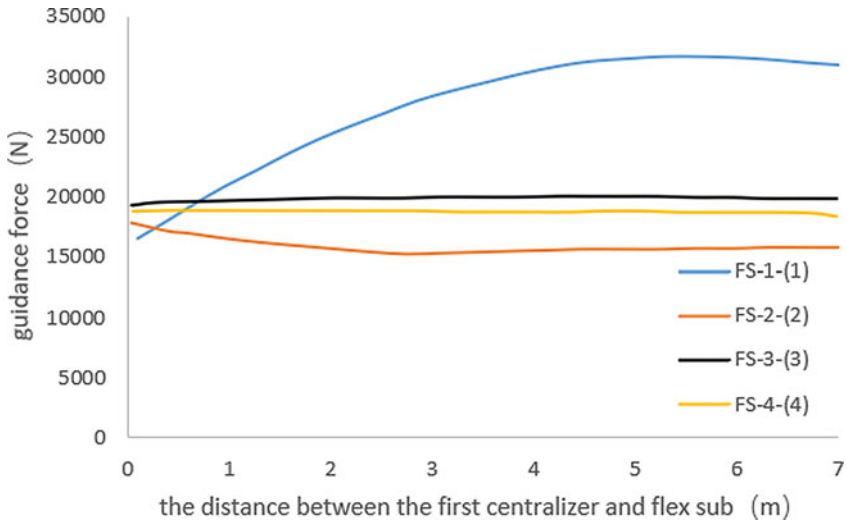**Fig. 6** Variation of guidance force with the distance between flex sub and the last centralizer in angle building mode
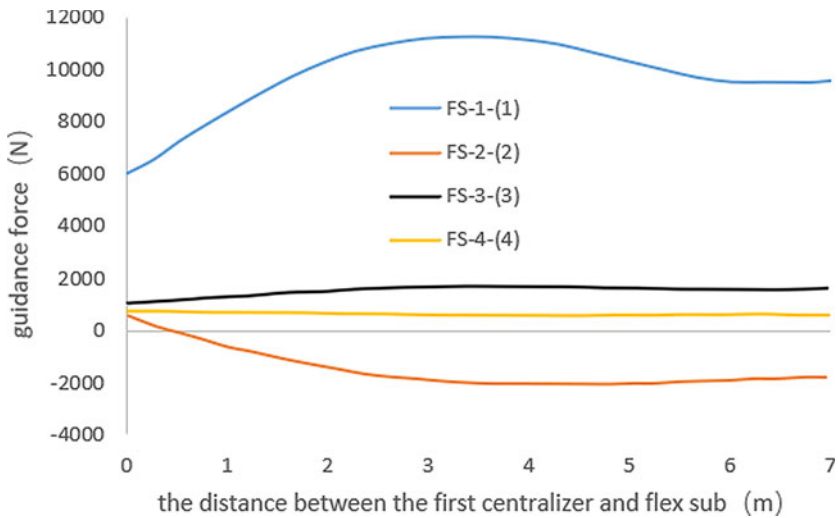
Fig. 7 Variation of the guidance force with the distance between the flex sub and the last centralizer in the dropping mode

Table 2 The sequence of the building-up capacity for different BHA types of static push-the-bit rotary steering system (from large to small)

NO	Model	NO	Model
1	FS-1-(1)	6	FS-2-(1)
2	FS-3-(1)	7	FS-4-(4)
3	FS-4-(1)	8	FS-4-(2)
4	FS-3-(3)	9	FS-2-(2)
5	FS-3-(2)	10	FS-4-(3)

5   Conclusion

(1) With different BHAs, the position of flex sub causes different influences on the static rotary steering tool.
(2) Among different BHAs, the maximum building-up capacity can be provided when the rotary steering tool equips with the unique flex sub.

References

1. Zhang, S.: New progress and development direction of modern steering drilling techniques. Acta Petrolei Sinica **24**(3), 82–83 (2003)
2. Pater, D.C., Mark, C.J., Zoback, D., et al.: Complications with stress tests- insights from a fracture experiment in the ultra-deep KTB borehole. SPE 36437 (1996)
3. Buker, M.: Advancements in rotary steerable technology. Petroleum Society of Canada. https://doi.org/10.2118/2001-040-EA
4. Al Bardisi, T., Akhmetov, R., Sanderson, M., et al.: Hybrid high build rate RSS improves challenging directional control in a soft abrasive drilling environment in oman drilling operation. SPE (2014). https://doi.org/10.2118/170019-MS
5. Xue, Q., Shan, D.Q., Leilei, H.: The latest progress and development trend of rotary steering drilling technology. China Petrol. Mach. **41**(07), 1–6 (2013)
6. Wang, P., Sheng, L., Dou, X., et al.: Newest development and tendency of foreign rotary steerable systems. Drill. Prod. Technol. **36**(06), 32–35 (2013)
7. Junichi, S., Adam, B., Richard, H., et al.: Downhole steering automation and new survey measurement method significantly improves high-dogleg rotary steerable system performance. In: SPE, p. 166165 (2013)
8. Richard, H., Steve, J., James, O., et al.: Design development and field testing of a high-doglegslim-hole rotary steerable system. In: SPE/IADC, p. 163472 (2013)
9. Xiao, R., Liang, Z.: Development status and prospect of rotary steering drilling technology. China Petrol. Mach. **34**(4) (2006)
10. Li, J., Li, D., Zhang, H., Wang, H.: The influencing factors of the inclination ability of push-the-bitrotary guiding tool. Oil Drill. Prod. Technol. **41**(4), 460–466 (2019)
11. Zheng, D., Gao, D., Feng, J., et al.: Study on deflection performance of backup and directional types of rotary steering tools. Oil Drill. Prod. Technol. **33**(6), 10–13 (2011)
12. Wang, M., Di, Q., Wang, W.: Influences of flexible joints on guiding force characteristics of RSBHA. Drill. Prod. Technol. **35**(1), 52–55 (2012)
13. Li, Z.: Mechanics and Application of Rod and Tubing in Oil and Gas Wells, pp. 127-140. Petroleum Industry Press, Beijing (2008)

Numerical Analysis of Crack Propagation by Using the P-version Finite Element Method and Contour Integral Method

Jianming Zhang, Wensheng Yang, and Yong He

Abstract The stress intensity factors (SIFs) are extracted by applying the p-version finite element method (P-FEM) and contour integral method (CIM) for the two dimensional linear elastic fracture mechanics models. Crack propagation paths are analyzed in details. Two benchmark numerical models, SIFs of a finite plate with a central crack under uniaxial tension and prediction of crack propagation path in a rectangle plate with oblique central crack, are used to verify the effectiveness and accuracy of the proposed approaches. Numerical results show higher accuracy and better stability compared with those results obtained by the numerical manifold method (NMM) in the literature available. Crack propagation paths are in good agreement with those results in the literature and experiments.

Keywords Crack propagation · P-version finite element method · Contour integral method · Numerical simulation

1 Introduction

The extraction of the stress intensity factors (SIFs) and analysis of crack propagation path are very important in engineering structures and materials, a great amount of research shows that the structural instability is closely related to internal or exterior crack propagation in the structures and materials. At present, many numerical methods have been developed to extract the stress intensity factors and simulate crack propagation, for example, the boundary element method (BEM) [1], finite element method (FEM) [2, 3], numerical manifold method (NMM) [4, 5],

J. Zhang (✉) · W. Yang
Department of Engineering Mechanics, Kunming University of Science and Technology, Kunming 650500, China
e-mail: jmzhang@kmust.edu.cn

Y. He
Yunnan Institute of Water and Hydropower Engineering Investigation, Design and Research, Kunming 650032, China

meshless method (MLM) [6] and generalized or extended finite element method (GFEM or XFEM) [7–9]. These numerical methods show different characteristics in solving crack and crack propagation problems and have achieved fruitful research results on fracture and crack propagation of complex engineering structures and materials. Generally speaking, the boundary element method needs to know the fundamental solution of control equation. The finite element method and numerical manifold method often need to divide fine meshes around the crack tip, which will lead to complex preprocessing and unsatisfied accuracy. The calculations of the meshless method cost greatly. The extended finite element method (XFEM) was first proposed by Belytschko in 1999 [2, 7, 10]. XFEM employs the partition of unity (PU) concept to simulate fracture and crack propagation by using the enrichment function of the crack surface and the crack tip asymptotic displacement function reflecting the local characteristics of the crack tip and avoids remeshing of meshes. The extended finite element method is one of the most effective numerical methods for solving discontinuous problems such as structural cracking and crack propagation. But the extended finite element method uses mostly low order interpolation (interpolation polynomial order $p \leq 2$), its convergence rate is slower and accuracy is not satisfied. In order to demonstrate the accuracy and effectiveness of the proposed approaches, SIFs are calculated directly for a finite plate with a central crack under uniaxial tension and crack propagation path in a rectangular plate with oblique central crack are analyzed by applying P-FEM combined CIM in this paper.

2 The P-version Finite Element Method and the Contour Integral Method

It is well known that the p-version finite element method (P-FEM) was rapidly developed and widely used in different engineering fields in the past three decades. In the P-FEM, the number of finite element mesh is fixed and the degree p of interpolation polynomials is increased uniformly or selectively to achieve the higher accuracy. The solution of P-FEM converges exponentially to smooth solution, and it converges twice as fast as the classical FEM if the solution has singularity of r^{γ}-type [11–14]. The high convergence rates of the P-FEM are very important for analysts to validate mathematical models of engineering structures and materials such as cracks and crack growth in cracking specimens, see [15–17].

The contour integral method (CIM) is a super convergent technique for the calculation of SIFs, In the framework of the P-FEM, Szabó and Babuška firstly proposed the contour integral method (CIM) for the extraction of stress intensity factors [18], Pereira and Duarte [19] generalized the contour integral method for the case of cracks with tractions applied to their faces and extracted stress intensity factors for loaded cracks, Cui et al. calculated the SIFs for two dimensional cracks by using contour integral approaches and displacement discontinuity method [20].

The convergence rate of this method is at least faster than the strain energy. The CIM can be also used to extract SIFs of Mode I, II and III from single and mixed mode crack problems. For the discussions in detail, please see [19, 21].

3 Numerical Examples

3.1 Central Cracked Plate Under Uniaxial Tension

To illustrate the accuracy and effectiveness of the P-FEM and CIM, the SIFs are extracted for a finite plate with a central crack of length $2a$ under uniaxial tension (shown in Fig. 1). The physical dimensions for this model in Fig. 1 are $h = 2$ m, $b = 1$ m. The plate is subjected to uniform traction of $\sigma = 3$ MPa in y direction. The material is linearly elastic and material parameters are considered for Young's modulus $E = 3.0 \times 10^4$ MPa and Poisson's ratio $v = 0.3$. The example is solved under plane stress condition.

SIFs K_I and relative error obtained by the p-version FEM for $a/b = 0.5$ are listed in Table 1, it is easy to see that the p-version FEM has faster convergent rate and smaller error from Table 1. With the increasing of interpolation polynomial degree p, convergence rate of K_I is very fast.
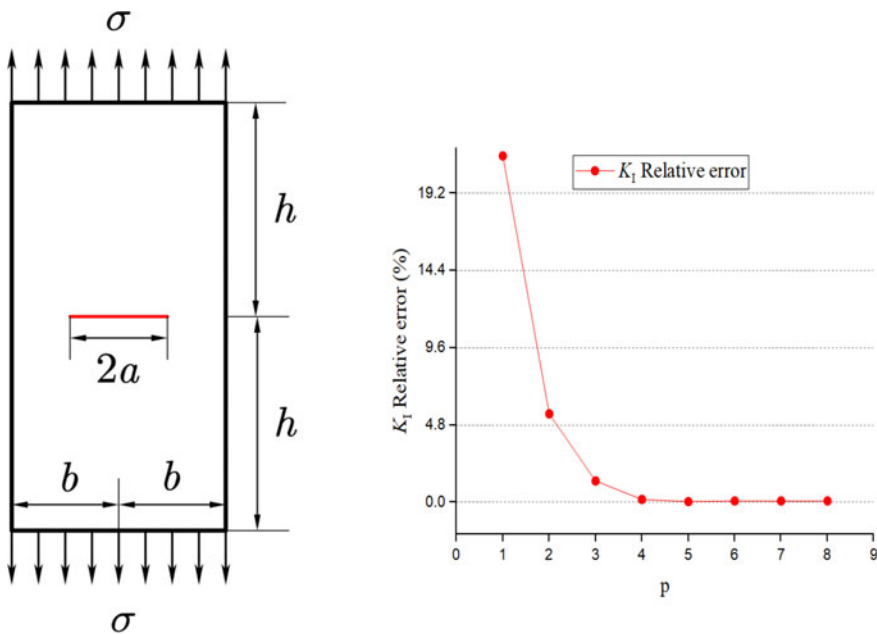
Fig. 1 Model of central cracked plate

Table 1 The relative error of K_I obtained by the p-version FEM for $a/b = 0.5$

p	K_I (Present)	Analytical	Relative error
1	3.500	4.4602	21.53%
2	4.215	4.4602	5.50%
3	4.404	4.4602	1.33%
4	4.452	4.4602	0.18%
5	4.462	4.4602	0.04%
6	4.464	4.4602	0.09%
7	4.464	4.4602	0.09%
8	4.464	4.4602	0.09%

Table 2 Comparison of the SIFs K_I obtained for different values of a/b

a/b	Analytical [22]	Cai et al. [5]	Present	Relative error Cai et al. [5]	Relative error Present
0.2	2.4362	2.4313	2.4370	0.20%	0.033%
0.4	3.7297	3.6863	3.7330	1.16%	0.088%
0.6	5.3658	5.2952	5.3710	1.32%	0.097%

The computational results of the SIFs K_I by the present method and NMM are listed in Table 2. In the computation of this model, the maximum number of meshes we used is 200.

3.2 Crack Propagation in Titanium Ti-6Al-4V Plate

In this section, the crack trajectory predicted by the P-FEM and CIM is compared with the parametric meshless Galerkin method (X-PMGM) and the experimental data provided in [23, 24]. An isotropic titanium Ti-6Al-4V plate with oblique central crack of length $2a = 13.5$ mm at $\beta = 43°$ is shown in Fig. 2, the titanium Ti-6Al-4V plate is subjected to a uniaxial tension load $\sigma^\infty = 207$ MPa. The dimensions of the titanium Ti-6Al-4V plate are $2H = 304.8$ mm and $2W = 76.2$ mm. Material properties considered for Young's modulus $E = 110$ GPa and Poisson's ratio $v = 0.29$.

The predicted crack trajectories obtained by the P-FEM and CIM are compared with the X-PMGM and experimental results [23, 24] (shown in Fig. 3). Obviously, the predicted crack trajectories obtained by present method are closer experimental results.

(a)   (b)
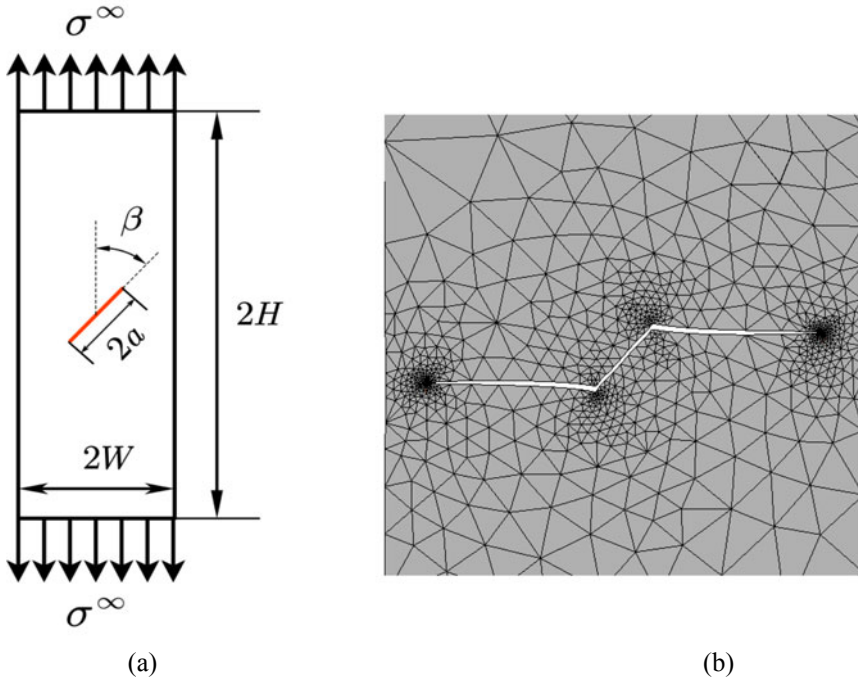
Fig. 2 (a) Geometry and loading of the cracked plate, (b) Division of mesh and crack propagation paths in plate

Fig. 3 Comparison of the predicted crack trajectories among the present method and X-PMGM & experimental results for the titanium Ti-6Al-4V plate
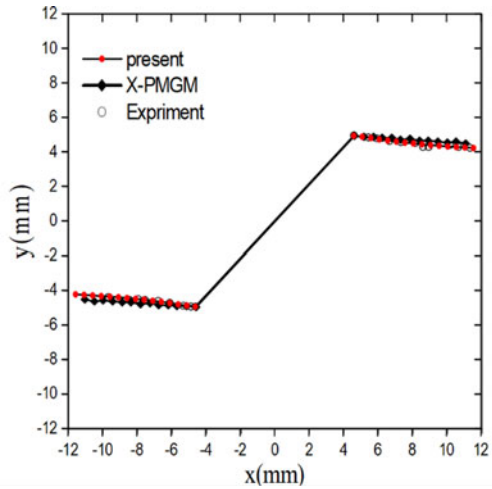
4   Conclusions

In the paper, the SIFs of central cracked plate under uniaxial tension are calculated by the P-FEM combined CIM and compared with analytical solution and those results available in the literature, numerical results of present method show higher accuracy, faster convergent rate and less preprocessing. Furthermore, the crack propagation paths in a rectangular titanium Ti-6Al-4V plate with oblique central crack are simulated numerically by applying P-FEM combined CIM, and compared with the X-PMGM and experimental results. Results show that P-FEM and CIM is effectiveness and feasible for the simulation of crack propagation paths for the model chosen. The present work can also be extended to analyze three dimensional crack problems, this will be future work.

References

1. Wen, P.H., Aliabadi, M.H.: Dual boundary element method for modelling curved crack paths. Int. J. Fract. **176**(1), 127–133 (2012)
2. Belytschko, T., Black, T.: Elastic crack growth in finite elements with minimal remeshing. Int. J. Numer. Meth. Eng. **45**(5), 601–620 (1999)
3. Branco, R., Antunes, F.V.: Finite element modelling and analysis of crack shape evolution in mode-I fatigue middle cracked tension specimens. Eng. Fract. Mech. **75**(10), 3020–3037 (2008)
4. Wu, Z.J., Wong, L.N.Y.: Frictional crack initiation and propagation analysis using the numerical manifold method. Comput. Geotech. **39**, 38–53 (2012)
5. Cai, Y.C., Wu, J., Atluri, S.N.: A new implementation of the numerical manifold method (NMM) for the modeling of non-collinear and intersecting cracks. Comput. Model. Eng. Sci. **92**(1), 63–85 (2013)
6. Wu, S., Fang, S.: Modeling cohesive cracks with meshless method. Int. J. Damage Mech. **18**(8), 721–737 (2009)
7. Daux, C., Möes, N., Dolbow, J., Sukumar, N., Belytschko, T.: Arbitrary branched and intersecting cracks with the extended finite element method. Int. J. Nume. Meth. Eng. **48**, 1741–1760 (2000)
8. Fries, T.P., Belytschko, T.: The extended/generalized finite element method: an overview of the method and its applications. Int. J. Numer. Meth. Eng. **84**(3), 253–304 (2010)
9. Song, G., Waisman, H., Lan, M.Y., Harari, I.: Extraction of stress intensity factors from Irwin's integral using high-order XFEM on triangular meshes. Int. J. Numer. Meth. Eng. **102**(3–4), 528–550 (2015)
10. Möes, N., Dolbow, J., Belytschko, T.: A finite element method for crack growth without remeshing. Int. J. Numer. Meth. Eng. **46**(1), 131–150 (1999)
11. Babuška, I., Szabó, B.A., Katz, I.N.: The p-version of the finite element method. SIAM J. Numer. Anal. **18**(3), 515–545 (1981)
12. Babuška, I., Suri, M.: The optimal convergence rate of the p-version of the finite element method. SIAM J. Numer. Anal. **24**(4), 750–776 (1987)

13. Guo, B.Q.: Approximation theory of the p-version of the finite element method in three dimensions, Part 2: Convergence of the p-version of FEM. SIAM J. Numer. Anal. **47**(4), 2578–2611 (2009)
14. Szabó, B.A., Babuška, I.: Introduction to Finite Element Analysis. Wiley, New York (2011)
15. Rahulkumar, P., Saigal, S., Yunus, S.: Singular p-version finite elements for stress intensity factor computation. Int. J. Numer. Meth. Eng. **40**(6), 1091–1114 (1997)
16. Wowk, D., Alousis, L., Underhill, R.: An adaptive remeshing technique for predicting the growth of irregular crack fronts using p-version finite element analysis. Eng. Fract. Mech. **207**, 36–47 (2019)
17. Wu, Y., Xing, Y.F., Liu, B.: Hierarchical p-version C^1 finite elements on quadrilateral and triangular domains with curved boundaries and their applications to Kirchhoff plates. Int. J. Numer. Meth. Eng. **119**(3), 177–207 (2019)
18. Szabó, B.A., Babuška, I.: Computation of the amplitude of stress singular terms for cracks and reentrant corners. In: Cruse, T.A. (ed.) Fracture Mechanics: Nineteenth Symposium, ASTM STP 969, pp. 101–124 (1988).
19. Pereira, J.P., Duarte, C.A.: The contour integral method for loaded cracks. Commun. Numer. Methods Eng. **22**(5), 421–432 (2006)
20. Cui, X., Li, H., Cheng, G.W., Tang, C., Gao, X.: Contour integral approaches for the evaluation of stress intensity factors using displacement discontinuity method. Eng. Anal. Bound. Elem. **82**, 119–129 (2017)
21. Garzon, J., Duarte, C.A., Pereira, J.P.: Extraction of stress intensity factors for the simulation of 3-D crack growth with the generalized finite element method. Key Eng. Mater. **560**, 1–36 (2013)
22. Tada, H., Paris, P.C., Irwin, G.R.: The Stress Analysis of Cracks Handbook, 2nd edn. Paris Productions, St. Louis (1985)
23. Pustejovsky, M.A.: Fatigue crack propagation in titanium under general in-plane loading–I: experiments. Eng. Fract. Mech. **11**, 9–15 (1979)
24. Pustejovsky, M.A.: Fatigue crack propagation in titanium under general in-plane loading–II: analysis. Eng. Fract. Mech. **11**, 17–31 (1979)

Timoshenko Elastic and Electroelastic Beam Models Incorporating the Local Mass Displacement Effect

Olha Hrytsyna (ORCID)

Abstract The objective of the paper is to model the size-dependent mechanical behaviors of elastic and electroelastic micro-/nanosize beams. To this end, the relations of local gradient theory of polarized continua are utilized. This higher-grade theory is based on taking account of non-diffusive and non-convective mass flux related to the changes in the material microstructure. The above mass flux is associated with the process of local mass displacement. In this paper, the general non-classical governing equations and associated boundary conditions for linear local gradient theory of piezoelectric media are obtained by making use of a variational approach. The local gradient elastic and electroelastic bending theories for plane-strain beams are developed as well. The variational equation for the thin beam bending problem was formulated by considering the contribution of the local mass displacement to the free energy density. The kinematic hypotheses of Timoshenko beam theory were used to obtain the coupled ordinary differential equations governing the states of elastic and dielectric beams. The problem of a thin elastic cantilever beam under a point load at the free end was solved to illustrate the theory efficiency. The effect of the piezoelectricity on the beam behavior is studied as well. It is shown that the local mass displacement and material piezoelectric properties being taken into account stiffens the nanocantilever beam. It is found that the size effect induced by the local mass displacement is significant when the beam thickness is comparable in size to the material length scale parameter.

O. Hrytsyna (✉)

Institute of Construction and Architecture, Slovak Academy of Sciences, Dúbravská cesta 9 45, 84503 Bratislava, Slovakia

e-mail: usarolha@savba.sk

303

1   Introduction

During recent years, micro-/nanosize beams and fibers are widely utilized in nanodevices, such as nanoactuators, nanosensors, nanoresonators, etc. In this regard, the studies of electromechanical properties of small-scale structures are of great importance. It is known that solid structures at nanoscale demonstrate specific electromechanical properties. Experimental validation of this phenomenon is presented in [8, 18, 20, 25]. The classical theory that used the local constitutive equations is incapable of describing the mentioned size effect. Therefore, new non-classical theories have been intensively developed in order to describe the specific size-dependent behavior of solids at the nanoscale. Among various modified continuum theories, the strain gradient theory of dielectrics [9, 17], coupled stress theory [29], theories of dielectric media with polarization gradient [4, 26] and electric field gradient [19, 33], as well as the nonlocal theory of piezoelectricity [10, 11] have been effectively used to study the electromechanical behavior of small-scale structures. For the last few decades, these theories were extended to elastic and piezoelectric beam mathematical models. The most famous among them are the strain gradient beam models [21, 32, 34] and the coupled stress beam model [28]. Some authors implement the nonlocal elasticity theory [2, 23], the polarization gradient theory [1], and the electric field gradient theory [22] to study the elastic and electroelastic beam behavior at the nanoscale. A review of generalized mathematical beam models was presented by Thai et al. [30], Behera and Chakraverty [2].

Despite the big number of studies of the size-dependent behavior of micro-/nanosize beams using various advanced continuum theories, the available literature lacks the mathematical model of shear deformable beams that takes account of the non-diffusive and non-convective mass fluxes due to the changes in the material microstructure. Note that non-diffusive and non-convective mass flux was considered within the local gradient theories of thermoelastic [12] and electrothermoelastic [13] continua where such flux was associated with the process of the local mass displacement. These theories have been successfully used to analyze a number of problems. For instance, it has been utilized to analyze the high-frequency dispersion of longitudinal elastic waves [13] and to study electromagnetic emission that accompanies the process of formation of new surfaces in the body [5]. The local gradient theory found various applications in the electromechanics of nanoscale solids. In particular, it was applied to describe a near-surface inhomogeneity of electromechanical fields [6] and to predict the size dependence of shear module in thin strips [16]. The local gradient theory was used to describe the nonlinear dependence of the inverse capacitance of a plane capacitor on its thickness [7]. Recently, this theory was extended to the Bernoulli–Euler elastic and piezoelectric beam model [14, 15]. Within this theory, it has been shown that the local mass displacement being taken into account stiffens the nanocantilever beam [14, 15]. However, it is known that for a large ratio of beam thickness to its length, the beam theory based on the Bernoulli–Euler kinematic hypotheses provides an inaccurate explanation of the beam behavior. Therefore, the purpose of the present

paper is to formulate a variational principle for an anisotropic three-dimensional local gradient electroelastic media and to use the proposed variational approach to get the equations of local gradient Timoshenko beam theory.

This paper is organized as follows. The theoretical background of the linear local gradient electroelasticity is presented in Sect. 2. This section includes the field equations, kinematic and gradient-type constitutive relations. In Sect. 3, the local gradient theory of polarized, deformable media is derived from fundamental principles, using the method of variational calculus. In Sect. 4, the variational equation is utilized to derive the linear governing equations of the local gradient Timoshenko elastic beam theory and the corresponding boundary conditions. In the next section, this theory is extended to the piezoelectric beams. The deflections and rotations of the cantilever beam loaded by an end-point force that verifies the correctness of the local gradient elastic and piezoelectric beam models are given in Sects. 4 and 5. The numerical results for both elastic and electroelastic beam models are discussed in Sect. 6. Conclusions are drawn in the final Sect. 7.

2 Theoretical Background for Local Gradient Electroelasticity

The general theory of local gradient electrothermoelasticity was presented in a series of papers and was recently generalized in the book [13]. The equations of this theory were obtained by making use of the continuum-thermodynamic approach which was proposed in [5, 6]. This approach is based on the fundamental principles of non-equilibrium thermodynamics, continuum mechanics and electrodynamics. The local gradient theory takes into consideration the local mass displacement and its coupling with the mechanical deformation, heat conduction, and polarization. Within this higher-grade theory, the local mass displacement is associated with the changes in the material microstructure and is described by the new physical quantities for which the additional balance equation of induced mass is formulated [13]. It should be noted that the above changes in the material microstructure can be observed for instance in the near-surface regions of solids (see Fig. 1).

For isothermal approximation, the consideration of the local mass displacement yields the gradient-type constitutive equations [15]:

$$\hat{\boldsymbol{\sigma}} = \rho_0 \frac{\partial G}{\partial \hat{\boldsymbol{\varepsilon}}}, \ \ \boldsymbol{\pi}_e = -\frac{\partial G}{\partial \mathbf{E}}, \ \ \rho_m = -\frac{\partial G}{\partial \mu_\pi'}, \ \ \boldsymbol{\pi}_m = \frac{\partial G}{\partial (\nabla \mu_\pi')}. \tag{1}$$

The above constitutive equations should be completed by the following relation for the electric displacement vector $\mathbf{D}$
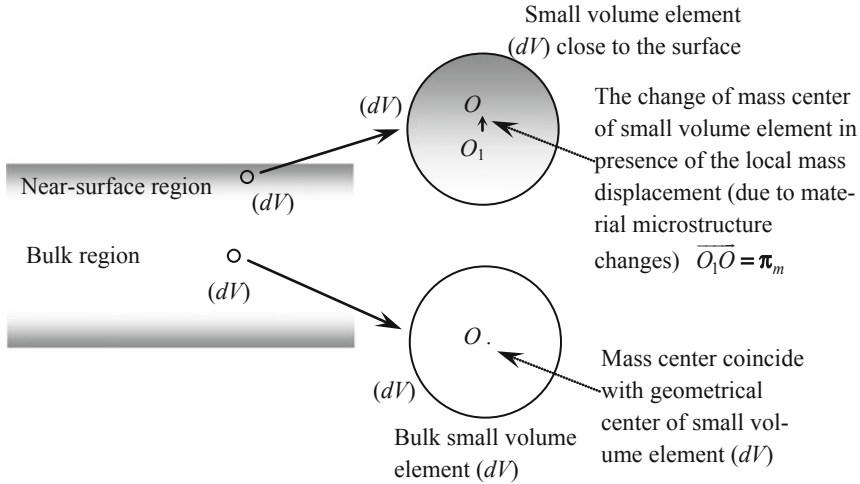
Fig. 1 Schematic showing the cross-section of a free infinite layer that shows the different material properties close to the body surface and bulk region. The change of mass center of a small volume element close to the surface is described with the non-diffusive and non-convective mass flux due to material microstructure changes in the vicinity of the body surface

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \rho_0 \boldsymbol{\pi}_e. \tag{2}$$

In the above equations, $\hat{\sigma}$ is the stress tensor, ρ_0 denotes the mass density, $\hat{\varepsilon}$ is the strain tensor, $\mathbf{E}$ and $\boldsymbol{\pi}_e$ are electric field and specific polarization vectors, respectively, ρ_m is the specific density of the induced mass (see [13] for more details), $\mu'_\pi = \mu_\pi - \mu$, μ denotes a chemical potential, μ_π is the energy measure of the effect of local mass displacement on the internal energy, $\boldsymbol{\pi}_m$ is the specific vector of the local mass displacement, ε_0 is the permittivity of vacuum, G represents the free energy density function generalized for polarized media with local mass displacement, and ∇ is the nabla operator. Here and in what follows, bold symbols stand for vectors and bold symbols with caps denote the second and higher-rank tensors.

As it follows from Eq. (1), within the local gradient theory two additional pairs of constitutive parameters, namely (ρ_m, μ'_π) and $(\boldsymbol{\pi}_m, \nabla \mu'_\pi)$, are introduced into the space of constitutive parameters alongside the classical pairs $(\hat{\sigma}, \hat{\varepsilon})$ and $(\mathbf{E}, \boldsymbol{\pi}_e)$. The inclusion of these additional constitutive parameters enables the new mathematical model to describe the surface, size, and other microstructure-dependent effects [13].

Besides the constitutive relations (1) and (2), the linear local gradient electroelasticity includes the equation of motion, the Gauss law and the Maxwell-Faraday equation, as well as the induced mass balance equation:

$$\nabla \cdot \hat{\boldsymbol{\sigma}} + \rho_0 \mathbf{F} = \rho_0 \ddot{\mathbf{u}}, \tag{3}$$

$$\nabla \cdot \mathbf{D} = \rho_e, \quad \nabla \times \mathbf{E} = 0, \tag{4}$$

$$\nabla \cdot \boldsymbol{\pi}_m + \rho_m = 0. \tag{5}$$

Here, $\mathbf{F}$ represents the mechanical mass force, $\mathbf{u}$ is the displacement vector, ρ_e is the density of free electric charge, the dot and $\times$ denote the scalar and vector products, respectively, and a superimposed dot means differentiation with respect to time.

This set of equations should be also completed by the kinematic relations:

$$\hat{\boldsymbol{\varepsilon}} = \frac{1}{2}\left[\nabla \otimes \mathbf{u} + (\nabla \otimes \mathbf{u})^{\mathrm{T}}\right], \tag{6}$$

$$\mathbf{E} = -\nabla \varphi, \tag{7}$$

where φ is the electric potential, $\otimes$ denotes the dyadic product, and superscript T stands for "transposed".

According to the linear theory of local gradient electroelasticity, the free energy G can be expanded as a quadratic function of the strain tensor $\hat{\varepsilon}$, electric field vector $\mathbf{E}$, modified chemical potential μ'_π and its gradient $\nabla \mu'_\pi$. Assuming $\varepsilon_{ij} \ll 1$, $E_i \ll 1$, $\tilde{\mu}'_\pi \ll 1$, and $\nabla \tilde{\mu}'_\pi \ll 1$, where $\tilde{\mu}'_\pi = \mu'_\pi - \mu'_{\pi 0}$, for an anisotropic continuum, we can write [15]:

$$\begin{aligned}
G = G_0 &+ \frac{1}{2\rho_0}\hat{\mathbf{C}}^{(4)} :: (\hat{\boldsymbol{\varepsilon}} \otimes \hat{\boldsymbol{\varepsilon}}) - \frac{1}{2}\hat{\chi}^E : (\mathbf{E} \otimes \mathbf{E}) - \frac{1}{2}d_\mu \tilde{\mu}'^2_\pi - \frac{1}{2}\hat{\chi}^m : \left(\nabla \tilde{\mu}'_\pi \otimes \nabla \tilde{\mu}'_\pi\right) \\
&- \frac{1}{\rho_0}\hat{e}^{(3)} : \cdot(\hat{\boldsymbol{\varepsilon}} \otimes \mathbf{E}) - \frac{1}{\rho_0}\hat{\alpha}^\mu : \hat{\boldsymbol{\varepsilon}}\ \tilde{\mu}'_\pi - \frac{1}{\rho_0}\hat{g}^{(3)} : \cdot\left(\hat{\boldsymbol{\varepsilon}} \otimes \nabla \tilde{\mu}'_\pi\right) - \left(\boldsymbol{\gamma}^\mu \cdot \nabla \tilde{\mu}'_\pi\right) \tilde{\mu}'_\pi \\
&+ \boldsymbol{\gamma}^{E\pi} \cdot \mathbf{E}\ \tilde{\mu}'_\pi + \hat{\chi}^{Em} : \left(\mathbf{E} \otimes \nabla \tilde{\mu}'_\pi\right).
\end{aligned} \tag{8}$$

Here, $\hat{\mathbf{C}}^{(4)}$ is the elastic constant tensor, $\hat{e}^{(3)}$ is the third-rank tensor of piezo-electric coefficients, $\hat{\chi}^E$ is the dielectric susceptibility tensor, $\hat{g}^{(3)}$, $\hat{\chi}^m$, $\hat{\chi}^{Em}$, $\hat{\alpha}^\mu$, $\boldsymbol{\gamma}^\mu$, $\boldsymbol{\gamma}^{E\pi}$, and d_μ are non-classical (higher-grade) material constants which introduce the local mass displacement effects. Hereinafter, the superscript in parentheses to the right from the bold symbol indicates the tensor rank.

Making use of formulae (1) and (8), the constitutive equations for polarized elastic media can be obtained as

$$\hat{\boldsymbol{\sigma}} = \hat{C}^{(4)} : \hat{\boldsymbol{\varepsilon}} - \mathbf{E} \cdot \hat{e}^{(3)} - \hat{\alpha}^\mu\ \tilde{\mu}'_\pi - \nabla \tilde{\mu}'_\pi \cdot \hat{g}^{(3)}, \tag{9}$$

$$\boldsymbol{\pi}_e = \frac{1}{\rho_0}\hat{e}^{(3)} : \hat{\boldsymbol{\varepsilon}} + \hat{\chi}^E \cdot \mathbf{E} - \gamma^{E\pi}\ \tilde{\mu}'_\pi - \nabla \tilde{\mu}'_\pi \cdot \hat{\chi}^{Em}, \tag{10}$$

$$\rho_m = \frac{1}{\rho_0}\hat{\alpha}^{\mu} \; : \; \hat{\varepsilon} - \boldsymbol{\gamma}^{E\pi} \cdot \mathbf{E} + d_{\mu}\tilde{\mu}_{\pi}' + \boldsymbol{\gamma}^{\mu} \cdot \nabla \tilde{\mu}_{\pi}', \tag{11}$$

$$\boldsymbol{\pi}_m = -\frac{1}{\rho_0}\hat{\mathbf{g}}^{(3)} \; : \; \hat{\varepsilon} + \hat{\chi}^{Em} \cdot \mathbf{E} - \boldsymbol{\gamma}^{\mu} \; \tilde{\mu}_{\pi}' - \hat{\chi}^{m} \cdot \nabla \tilde{\mu}_{\pi}'. \tag{12}$$

Field Eqs. (3)–(5), kinematic and constitutive relations (6), (7), (9)-(12) form a complete set of equations of linear local gradient electroelasticity.

3 Variational Approach

In this section, the governing three-dimensional equations and associated boundary conditions of the local gradient linear electroelasticity are derived based on the Hamilton variational principle.

Let us assume that the free energy density G exists for the elastic nonferromagnetic polarized continuum introduced in the previous section. To incorporate the local mass displacement into the mathematical model, we assume the free energy G to be a C^2-continuous function of the strain tensor $\hat{\varepsilon}$, electric field vector $\mathbf{E}$, modified chemical potential $\tilde{\mu}_{\pi}'$ and its space gradient $\nabla\tilde{\mu}_{\pi}'$, that is: $G = G(\hat{\varepsilon}, \mathbf{E}, \tilde{\mu}_{\pi}', \nabla\tilde{\mu}_{\pi}')$. Assume also that constitutive equations are defined by formulae (1) and (2).

Consider the elastic, polarized, deformable body that occupies the domain (V) bounded by a smooth surface (Σ) which separates it from an outer vacuum (domain (V')). Suppose that body is subjected to the external mechanical loads. For the considered continuum, the variational principle can be expressed by formula:

$$\delta \int_{t_1}^{t_2} (K - L)dt + \delta W = 0. \tag{13}$$

Here, K is the total kinematic energy, L denotes the total electric enthalpy energy, and W is the work done by the external forces. The variation of the external work is assumed as:

$$\delta W = \int_{t_1}^{t_2} \int_{(V)} \rho_0 \mathbf{F} \cdot \delta\mathbf{u}\, dVdt + \int_{t_1}^{t_2} \int_{(\Sigma)} \left(\boldsymbol{\sigma}_s \cdot \delta\mathbf{u} + \rho_0 \pi_{ms}\delta\tilde{\mu}_{\pi}' \right) d\Sigma dt, \tag{14}$$

where, $\boldsymbol{\sigma}_s = \{\sigma_{si}\}$ and π_{ms} are prescribed functions on (Σ) (i.e., the given surface tractions, and the surface value of the normal component of the local mass displacement vector).

The total kinematic energy K and electric enthalpy energy L can be presented as follows:

$$K = \frac{1}{2} \int\limits_{(V)} \rho_0 \dot{\mathbf{u}} \cdot \dot{\mathbf{u}}\, dV, \quad L = \int\limits_{(V_*)} \rho_0 H dV, \tag{15}$$

where $(V_*) = (V) \cup (V')$. Following [27], we can write:

$$\int\limits_{t_1}^{t_2} \delta K dt = \int\limits_{t_1}^{t_2} \int\limits_{(V)} \rho_0 \dot{\mathbf{u}} \cdot \delta \dot{\mathbf{u}}\, dV dt = - \int\limits_{t_1}^{t_2} \int\limits_{(V)} \rho_0 \ddot{\mathbf{u}} \cdot \delta \mathbf{u}\, dV dt \tag{16}$$

For electroelastic body that occupies the domain (V), through the relation

$$H = G(\hat{\varepsilon}, \mathbf{E}, \tilde{\mu}'_\pi, \nabla \tilde{\mu}'_\pi) - \frac{\varepsilon_0}{2\rho_0} \mathbf{E} \cdot \mathbf{E}, \tag{17}$$

the density of generalized electric enthalpy energy H can be introduced [15]. In view of relations (1), (15) and (17), we get

$$L = \delta \int\limits_{(V)} \rho_0 H\, dV = \int\limits_{(V)} \rho_0 \left(\frac{\partial G}{\partial \hat{\varepsilon}} : \delta \hat{\varepsilon} + \frac{\partial G}{\partial \mathbf{E}} \cdot \delta \mathbf{E} + \frac{\partial G}{\partial \tilde{\mu}'_\pi} \delta \tilde{\mu}'_\pi + \frac{\partial G}{\partial \nabla \tilde{\mu}'_\pi} \cdot \delta \nabla \tilde{\mu}'_\pi - \frac{\varepsilon_0}{\rho_0} \mathbf{E} \cdot \delta \mathbf{E} \right) dV$$

$$= \int\limits_{(V)} \left(\hat{\boldsymbol{\sigma}} : \delta \hat{\varepsilon} - (\varepsilon_0 \mathbf{E} + \rho_0 \boldsymbol{\pi}_e) \cdot \delta \mathbf{E} - \rho_0 \rho_m \delta \tilde{\mu}'_\pi + \rho_0 \boldsymbol{\pi}_m \cdot \delta \nabla \tilde{\mu}'_\pi \right) dV.$$

By substituting kinematic relations (6), (7), and constitutive Eq. (2) into the above formula, after some mathematical manipulations, one obtains

$$\delta \int\limits_{(V)} \rho_0 H\, dV = \int\limits_{(V)} [\nabla \cdot (\hat{\boldsymbol{\sigma}} \cdot \delta \mathbf{u}) - \nabla \cdot (\mathbf{D}\delta\varphi) + \rho_0 \nabla \cdot (\boldsymbol{\pi}_m \delta \tilde{\mu}'_\pi)$$
$$-\nabla \cdot \hat{\boldsymbol{\sigma}} \cdot \delta \mathbf{u} + \nabla \cdot \mathbf{D}\, \delta\varphi - \rho_0 (\nabla \cdot \boldsymbol{\pi}_m + \rho_m) \delta \tilde{\mu}'_\pi] dV. \tag{18}$$

Functions $\mathbf{u}$, $\boldsymbol{\pi}_e$, $\boldsymbol{\pi}_m$, and ρ_m are zero in vacuum. Thus, proceeding in a similar manner, for the domain (V'), we get

$$\delta \int\limits_{(V')} \rho_0 H\, dV = \varepsilon_0 \int\limits_{(V')} [\nabla \cdot (\nabla\varphi\delta\varphi) - \Delta\varphi\, \delta\varphi]\, dV. \tag{19}$$

By substituting Eqs. (14)–(19) into variational Eq. (13) and applying the divergence theorem, after some mathematical transformations, the following formula can be obtained:

$$\int\limits_{t_1}^{t_2} \int\limits_{(V)} \left[(\nabla \cdot \hat{\boldsymbol{\sigma}} + \rho_0 \mathbf{F} - \rho_0 \ddot{\mathbf{u}}) \cdot \delta \mathbf{u} + (\varepsilon_0 \Delta \varphi - \rho_0 \nabla \cdot \boldsymbol{\pi}_e) \delta \varphi + \rho_0 (\nabla \cdot \boldsymbol{\pi}_m + \rho_m) \delta \tilde{\mu}'_\pi \right] dV dt$$

$$+ \varepsilon_0 \int\limits_{t_1}^{t_2} \int\limits_{(V')} \Delta \varphi \, \delta \varphi \, dV dt - \int\limits_{t_1}^{t_2} \int\limits_{(\Sigma)} \left(\varepsilon_0 |\nabla \varphi|_n - \rho_0 \pi_{en} \right) \delta \varphi d\Sigma dt \qquad (20)$$

$$- \int\limits_{t_1}^{t_2} \int\limits_{(\Sigma)} \left[(\boldsymbol{\sigma}_n - \boldsymbol{\sigma}_{ns}) \cdot \delta \mathbf{u} + \rho_0 (\pi_{mn} - \pi_{ms}) \delta \overline{\mu}'_\pi \right] d\Sigma dt = 0,$$

where $\boldsymbol{\sigma}_n = \mathbf{n} \cdot \hat{\sigma}$, $\pi_{mn} = \mathbf{n} \cdot \boldsymbol{\pi}_m$, $\pi_{en} = \mathbf{n} \cdot \boldsymbol{\pi}_e$, $|\nabla \varphi|_n = |\nabla \varphi| \cdot \mathbf{n} = |\nabla_i \varphi|_{n_i}$ is the jump of the function $\nabla \varphi$ across the body surface (Σ), and $\mathbf{n}$ is the outward unit vector normal to the body surface (Σ).

Due to the arbitrariness of $\delta \mathbf{u}$, $\delta \varphi$, and $\delta \tilde{\mu}'_\pi$, the field equations, boundary and jump conditions can be formulated from variational relation (20) as follows:

$$\nabla \cdot \hat{\boldsymbol{\sigma}} + \rho_0 \mathbf{F} = \rho_0 \ddot{\mathbf{u}}, \quad \varepsilon_0 \Delta \varphi - \rho_0 \nabla \cdot \boldsymbol{\pi}_e = 0, \quad \nabla \cdot \boldsymbol{\pi}_m + \rho_m = 0, \, \forall \mathbf{r} \in (V), \quad (21)$$

$$\Delta \varphi = 0, \quad \forall \mathbf{r} \in (V'), \qquad (22)$$

$$(\boldsymbol{\sigma}_n - \boldsymbol{\sigma}_{ns}) \cdot \delta \mathbf{u} = 0, \quad \left(\varepsilon_0 |\nabla \varphi|_n - \rho_0 \pi_{en} \right) \delta \varphi = 0, \quad (\pi_{mn} - \pi_{ms}) \delta \tilde{\mu}'_\pi = 0, \, \forall \mathbf{r} \in (\Sigma).$$
$$(23)$$

Equations (21), (22), along with the boundary and jump conditions (23) constitute the boundary-value problems for linear local gradient electroelasticity.

It should be noted that the proposed variational approach can be used not only to establish the foundations of the local gradient theory but also to develop numerical procedures for the study of static and dynamic responses of microstructure-dependent dielectric bodies having complex shapes.

A simple application of variational methods in the stationary boundary-value problems for elastic and electroelastic cantilever Timoshenko beams is discussed in the next section.

4   Timoshenko Elastic Local Gradient Beam Model. Cantilever Beam Bending Problem

In this section, making use of the gradient-type constitutive equations of the local gradient elasticity [14], the stationary balance equations and the corresponding boundary conditions for the cantilever elastic beam subjected to the end-point loading are derived from the variational principle.

Note that for elastic continua, relations (9)–(12) read as [24]:

$$\hat{\boldsymbol{\sigma}} = \hat{C}^{(4)} : \hat{\boldsymbol{\varepsilon}} - \hat{\alpha}^\mu \tilde{\mu}'_\pi - \nabla \tilde{\mu}'_\pi \cdot \hat{\mathbf{g}}^{(3)}, \qquad (24)$$

$$\rho_m = \frac{1}{\rho_0} \hat{\alpha}^\mu : \hat{\boldsymbol{\varepsilon}} + d_\mu \tilde{\mu}'_\pi + \boldsymbol{\gamma}^\mu \cdot \nabla \tilde{\mu}'_\pi, \qquad (25)$$

$$\boldsymbol{\pi}_m = -\frac{1}{\rho_0}\hat{\mathbf{g}}^{(3)} : \hat{\boldsymbol{\varepsilon}} - \boldsymbol{\gamma}^\mu \tilde{\mu}'_\pi - \hat{\chi}^m \cdot \nabla \tilde{\mu}'_\pi. \tag{26}$$

For simplicity, in the above equations we will neglect the influence of the gradient of modified chemical potential on the stress distribution and on the induced mass density ρ_m. Within this assumption, constitutive Eqs. (24)–(26) of the local gradient elasticity and the strain energy density G can be written as [14]:

$$\hat{\boldsymbol{\sigma}} = \hat{\mathbf{C}}^{(4)} : \hat{\boldsymbol{\varepsilon}} - \hat{\alpha}^\mu \tilde{\mu}'_\pi, \tag{27}$$

$$\rho_m = d_\mu \tilde{\mu}'_\pi + \frac{1}{\rho_0}\hat{\alpha}^\mu : \hat{\boldsymbol{\varepsilon}}, \tag{28}$$

$$\boldsymbol{\pi}_m = -\hat{\chi}^m \cdot \nabla \tilde{\mu}'_\pi, \tag{29}$$

$$G = G_0 + \frac{1}{2\rho_0}\hat{\mathbf{C}}^{(4)} :: (\hat{\boldsymbol{\varepsilon}} \otimes \hat{\boldsymbol{\varepsilon}}) - \frac{1}{2}d_\mu\tilde{\mu}'^2_\pi - \frac{1}{2}\hat{\chi}^m : \left(\nabla\tilde{\mu}'_\pi \otimes \nabla\tilde{\mu}'_\pi\right)$$
$$- \frac{1}{\rho_0}\hat{\alpha}^\mu : \hat{\boldsymbol{\varepsilon}}\,\tilde{\mu}'_\pi. \tag{30}$$

Let us consider a thin rectangular elastic cantilever beam that in Cartesian coordinate system (x, y, z) occupies the region (V) defined as $0 \leq x \leq L$, $0 \leq y \leq b$, and $|z| \leq h/2$. The cantilever length is assumed to prevail over the other two dimensions, i. e., $L \gg h$ and $L \gg b$. The beam carries a point load $\mathbf{P}$ at the free end $x = L$ whereas its another end $x = 0$ is fixed, as shown in Fig. 2. In this case, the plane stress conditions are realized in the body and the displacement vector $\mathbf{u} = (u_1, u_2, u_3)$ and the modified chemical potential $\tilde{\mu}'_\pi$ are functions of x and z coordinate only.
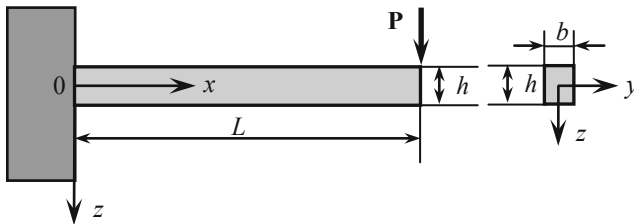
Fig. 2 The cantilever Timoshenko beam with rectangular cross section under a point load at the free end

To study the mechanical response of the beam, the kinematic hypotheses of Timoshenko are assumed to be valid for the displacement vector. Thus, we represent the displacement vector components as follows:

$$u_1 = -z\phi(x), \quad u_2 = 0, \quad u_3 = w(x). \tag{31}$$

Here, u_1 is the longitudinal displacement, $w(x)$ is the transverse displacement of the neutral axis (the beam deflection), $\phi(x)$ is the rotation of the beam cross-section, and the comma indicates differentiation with respect to the spatial variables. Using formulae (31) and the strain–displacement relations (6), the kinematic equations are obtained as:

$$\varepsilon_{11} = -z\phi_{,x} \quad \varepsilon_{13} = \frac{1}{2}(-\phi + w_{,x}). \tag{32}$$

Here, comma stands for partial differentiation with respect to the indicated space coordinate. Similarly to formulae (31), we express the modified chemical potential $\tilde{\mu}'_\pi$ as the linear function of z-coordinate, namely:

$$\tilde{\mu}'_\pi(x, z) = zm(x), \tag{33}$$

where $m(x)$ is the unknown function of x-coordinate.

The material property matrices for orthorhombic continuum can be represented in the matrix form as follows:

$$\begin{bmatrix} C_{11} & C_{13} & 0 \\ C_{13} & C_{33} & 0 \\ 0 & 0 & C_{44} \end{bmatrix}, \quad \begin{bmatrix} \alpha_1^\mu \\ \alpha_3^\mu \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -\chi_1^m & 0 \\ 0 & -\chi_3^m \end{bmatrix}.$$

Here, the Voigt notation is employed, that is: $C_{ijkl} \to C_{pr}$, $\alpha_{ij}^\mu \to \alpha_p^\mu$, $\chi_{ij}^m \to \chi_p^m$ $(p, r = \{1, 2, \ldots 6\})$, where $11 \to 1$, $22 \to 2$, $33 \to 3$, $23 \to 4$, $13 \to 5$, and $12 \to 6$.

Under the assumptions adopted, the three-dimensional constitutive Eqs. (27)–(29) take the form:

$$\sigma_{11}(x, z) = -zC_{11}\phi_{,x} - \alpha_1^\mu zm = -z(C_{11}\phi_{,x} + \alpha_1^\mu m) = z\sigma_1'(x), \tag{34}$$

$$\sigma_{33}(x, z) = -z(C_{13}\phi_{,x} + \alpha_3^\mu m) = z\sigma_3'(x), \tag{35}$$

$$\sigma_{31}(x) = C_{44}(-\phi + w_{,x}), \tag{36}$$

$$\rho_m(x, z) = z\left(d_\mu m - \frac{\alpha_1^\mu}{\rho_0}\phi_{,x}\right) = z\rho_m'(x), \tag{37}$$

$$\pi_m^1(x,z) = -\chi_1^m z\, m_{,x} = z\,\pi_m'^1(x),\tag{38}$$

$$\pi_m^3(x) = -\chi_3^m\, m.\tag{39}$$

To obtain the governing equations and boundary conditions for local gradient Timoshenko beam model, let us consider the variation of the strain energy density G. In view of relations (30), (34)–(39), we get the following form for this energy:

$$G - G_0 = \frac{1}{\rho_0}\sigma_{11}\varepsilon_{11} + \frac{2}{\rho_0}\sigma_{13}\varepsilon_{13} - \rho_m\tilde{\mu}_\pi' + \pi_m^1\,\tilde{\mu}_{\pi,x}' + \pi_m^3\, d\tilde{\mu}_{\pi,z}'\,.$$

By means of the variation of the above expression, the following result is received:

$$\delta\int_{(V)} G\,dV = \int_{(V)}\left(\tfrac{1}{\rho_0}\sigma_{11}\delta\varepsilon_{11} + \tfrac{2}{\rho_0}\sigma_{13}\delta\varepsilon_{13} - \rho_m\,\delta\tilde{\mu}_\pi' + \pi_m^1\,\delta\tilde{\mu}_{\pi,x}' + \pi_m^3\,\delta\tilde{\mu}_{\pi,z}'\right)dV$$

$$= \int_0^L\int_{-h/2}^{-h/2}\int_0^b\left[-\tfrac{1}{\rho_0}\sigma_{11}z\delta(\phi_{,x}) - \tfrac{1}{\rho_0}\sigma_{13}\delta\phi + \tfrac{1}{\rho_0}\sigma_{13}\delta(w_{,x})\right.$$
$$\left. -\left(\rho_m + \pi_{m,x}^1 + \pi_{m,z}^3\right)\delta\tilde{\mu}_\pi' + \left(\pi_m^1\,\delta\tilde{\mu}_\pi'\right)_{,x} + \left(\pi_m^3\,\delta\tilde{\mu}_\pi'\right)_{,z}\right]dxdydz.$$

This relation can be reformulated in the form:

$$\delta\int_{(V)} G\,dV = -\frac{b}{\rho_0}\int_0^L\int_{-h/2}^{h/2}\left[\sigma_{11}z\delta(\phi_{,x}) + \sigma_{13}\delta\phi - \sigma_{13}\delta(w_{,x}) + \rho_0\left(\rho_m + \pi_{m,x}^1\right)\delta\tilde{\mu}_\pi'\right]dxdz$$

$$+ b\int_0^L\left(\pi_m^3\,\delta\tilde{\mu}_\pi'\right)\big|_{z=-\frac{h}{2}}^{z=\frac{h}{2}}dx + b\int_{-h/2}^{h/2}\left(\pi_m^1\,\delta\tilde{\mu}_\pi'\right)\big|_{x=0}^{x=L}dz.$$

In view of the constitutive relations and formulae $\sigma_{11}\delta(\phi_{,x}) = (\sigma_{11}\delta\phi)_{,x} - \sigma_{11,x}\,\delta\phi$, $\sigma_{13}\delta(w_{,x}) = (\sigma_{13}\delta w)_{,x} - \sigma_{13,x}\,\delta w$, for the variation of the free energy density, the following result is obtained:

$$\delta\int_{(V)} G\,dV = -\int_0^L\left[-\tfrac{1}{\rho_0}\left(I\,\sigma_{1,x}' - S\sigma_{13}\right)\delta\phi + \tfrac{S}{\rho_0}\sigma_{13,x}\,\delta w + \left[I\left(\rho_m' + \pi_{m,x}'^1\right) - S\pi_m^3\right]\delta m\right]dx$$

$$-\tfrac{I}{\rho_0}\left(\sigma_1'\delta\phi\right)\big|_{x=0}^{x=L} + \tfrac{S}{\rho_0}\left(\sigma_{13}\delta w\right)\big|_{x=0}^{x=L} + I\left(\pi_m'^1\,\delta m\right)\big|_{x=0}^{x=L}.$$

Here, $S = bh$ is the cross-section area and $I = bh^3/12$ is the second moment of the cross-section area.

Due to the arbitrariness of $\delta\phi$, δw and δm, from the above expression we get the governing equations

$$M_{,x} - Q = 0, \quad Q_{,x} = 0, \quad I\left(\rho'_m + \pi'^1_{m,x}\right) - S\pi^3_m = 0, \tag{40}$$

with boundary conditions that defined the following functions:

$$M \text{ or } \phi, \quad Q \text{ or } w, \quad I\pi'^1_m \text{ or } m \quad \text{at } x = 0 \text{ and } x = L.$$

In Eq. (40), the integral characteristics Q and M are defined as:

$$(Q, M) = \int\limits_{-h/2}^{h/2} \int\limits_0^b (\sigma_{13}, z\sigma_{11}) dz dy. \tag{41}$$

In view of Eqs. (34), (36) and (41), for transverse shear force and bending moment, the following relations are found:

$$Q = C_{44}S(-\phi + w_{,x}), \quad M = -I\left(C_{11}\phi_{,x} + \alpha_1^\mu m\right).$$

Finally, by substituting the above formulae into Eq. (40), to determine the unknown functions $\phi(x)$, $w(x)$ and $m(x)$, the stationary non-classical governing equations can be derived:

$$C_{11}I\phi_{,xx} - C_{44}S\phi + C_{44}Sw_{,x} + \alpha_1^\mu Im_{,x} = 0, \tag{42}$$

$$(-\phi + w_{,x})_{,x} = 0, \tag{43}$$

$$m_{,xx} - \lambda_0^2(1 + M_h)m + \frac{\alpha_1^\mu}{\rho_0 \chi_1^m}\phi_{,x} = 0, \tag{44}$$

where $M_h = S\chi_3^m/Id_\mu$ and $\lambda_0 = \sqrt{d_\mu/\chi_1^m}$.

For cantilever Timoshenko beam, the boundary conditions are:

$$x = 0: \quad w = 0, \quad \phi = 0, \quad m = 0, \tag{45}$$

$$x = L: \quad M = 0, \quad Q = -P, \quad \pi'^1_m = 0. \tag{46}$$

The solution to the formulated boundary-value problem (42)–(46) is as follows:

$$w(x) = \frac{P}{6C_{11}I(1+\Omega)}\left\{x^3 - 3Lx^2 - \frac{6IC_{11}(1+\Omega)}{SC_{44}}x\right.$$
$$\left. + \frac{6L\Omega}{\lambda^2}\left[1 + \frac{\left(1 - \lambda Le^{-\lambda L}\right)\left(e^{\lambda x} - \lambda x\right) - \left(1 + \lambda Le^{\lambda L}\right)\left(e^{-\lambda x} + \lambda x\right)}{\lambda L(e^{\lambda L} + e^{-\lambda L})}\right]\right\} \tag{47}$$

$$\phi(x) = \frac{P}{C_{11}I(1+\Omega)}\left\{\frac{x^2}{2} - Lx\right. \tag{48}$$
$$\left. + \frac{\Omega\left[\left(1-\lambda Le^{-\lambda L}\right)\left(e^{\lambda x}-1\right)+\left(1+\lambda Le^{\lambda L}\right)\left(e^{-\lambda x}-1\right)\right]}{\lambda^2\left(e^{\lambda L}+e^{-\lambda L}\right)}\right\},$$

$$m(x) = \frac{P\Omega}{\alpha_1^{\mu}I(1+\Omega)}\left[x - L + \frac{\left(1+\lambda Le^{\lambda L}\right)e^{-\lambda x} - \left(1 - \lambda Le^{-\lambda L}\right)e^{\lambda x}}{\lambda\left(e^{\lambda L}+e^{-\lambda L}\right)}\right]. \tag{49}$$

Here, $\lambda = \lambda_0\sqrt{1+M+M_h}$, $\Omega = M/(1+M_h)$, and $M = \left(\alpha_1^{\mu}\right)^2\big/\rho_0 d_{\mu}C_{11}$ is the coupling factor between the mechanical fields and the local mass displacement. Note that assuming $\Omega = 0$, Eqs. (47)–(49) can be reduced to the classical solution.

5 Timoshenko Electroelastic Local Gradient Beam Model

In a similar manner, the relations of the local gradient Timoshenko model for non-ferromagnetic dielectric beam can be formulated in this section. To this end, now we consider piezoelectric beam that contacts with the vacuum. In this case, the constitutive equations of local gradient piezoelectricity (9)–(12) are used as the basis for the construction of the mathematical model of the beam.

To demonstrate the principal aspects of the developed theory and for the sake of simplicity, we neglect the effect of the gradient of the modified chemical potential on the stress distribution, polarization and induced mass density. Assume also that the modified chemical potential does not influence the body polarization and the local mass displacement vector. We neglect the contributions of the electric field to the density of the induced mass. Finally, accept that the local mass displacement vector is a function of the gradient of the modified chemical potential only. It is shown below that such assumptions can greatly simplify the solution of the considered boundary-value problem and still allow us to obtain the physically sound results.

Under the above simplifications, constitutive Eqs. (9)–(12) reduced to:

$$\hat{\boldsymbol{\sigma}} = \hat{C}^{(4)} : \hat{\boldsymbol{\varepsilon}} - \mathbf{E} \cdot \hat{\boldsymbol{e}}^{(3)} - \hat{\alpha}^{\mu}\,\tilde{\mu}_{\pi}', \tag{50}$$

$$\boldsymbol{\pi}_e = \hat{\chi}^{E} \cdot \mathbf{E} + \frac{1}{\rho_0}\hat{\boldsymbol{e}}^{(3)} : \hat{\boldsymbol{\varepsilon}}, \tag{51}$$

$$\rho_m = d_{\mu}\tilde{\mu}_{\pi}' + \frac{1}{\rho_0}\hat{\alpha}^{\mu} : \hat{\boldsymbol{\varepsilon}}, \tag{52}$$

$$\boldsymbol{\pi}_m = -\hat{\chi}^{m} \cdot \nabla\tilde{\mu}_{\pi}'. \tag{53}$$

In the state of a plane strain, constitutive relations (50)–(53) for the Timoshenko beam are obtained to be:

$$\sigma_{11} = C_{11}\varepsilon_{11} - e_{31}E_3 - \alpha_1^\mu \tilde{\mu}_\pi', \tag{54}$$

$$\sigma_{33} = C_{13}\varepsilon_{11} - e_{33}E_3 - \alpha_3^\mu \tilde{\mu}_\pi', \tag{55}$$

$$\sigma_{31} = 2C_{44}\varepsilon_{31} - e_{15}E_1, \tag{56}$$

$$\pi_e^1 = \chi_1^E E_1 + \frac{2}{\rho_0} e_{15}\varepsilon_{31}, \tag{57}$$

$$\pi_e^3 = \chi_3^E E_3 + \frac{1}{\rho_0} e_{31}\varepsilon_{11}, \tag{58}$$

$$\rho_m = d_\mu \tilde{\mu}_\pi' + \frac{\alpha_1^\mu}{\rho_0}\varepsilon_{11}, \tag{59}$$

$$\pi_m^1 = -\chi_1^m \tilde{\mu}_{\pi,x}', \tag{60}$$

$$\pi_m^3 = -\chi_3^m \tilde{\mu}_{\pi,z}'. \tag{61}$$

For a thin piezoelectric beam with a large ratio of length to thickness, the component D_1 of the electric induction vector $\mathbf{D}$ can be neglected [31, 34]. Thus, in the absence of free charges, from the Gauss law it follows that $D_{3,z} = 0$, that is $D_3 = const$. Under the open circuit condition, the electric displacement is equal to zero on the beam surfaces and thus the electric induction component D_3 is zero within the body. Since components D_1 and D_3 vanish, the electric field components E_1 and E_3 can be determined by making use of constitutive Eqs. (2), (57), and (58):

$$E_1 = -\frac{2e_{15}}{a_{11}}\varepsilon_{31}, \quad E_3 = -\frac{e_{31}}{a_{33}}\varepsilon_{11}, \tag{62}$$

where $a_{33} = \varepsilon_0 + \rho_0 \chi_3^E$, and $a_{11} = \varepsilon_0 + \rho_0 \chi_1^E$.

By substituting formulae (62) into the constitutive Eqs. (54)–(56), the stress components can be rewritten as:

$$\sigma_{11} = \overline{C}_{11}\varepsilon_{11} - \alpha_1^\mu \tilde{\mu}_\pi', \tag{63}$$

$$\sigma_{33} = \overline{C}_{13}\varepsilon_{11} - \alpha_3^\mu \tilde{\mu}_\pi', \tag{64}$$

$$\sigma_{31} = 2\overline{C}_{44}\varepsilon_{31}. \tag{65}$$

Here, $\overline{C}_{11} = C_{11} + e_{31}^2/a_{33}$, $\overline{C}_{13} = C_{13} + e_{33}e_{31}/a_{33}$, and $\overline{C}_{44} = C_{44} + e_{15}^2/a_{11}$.

Thus, for the considered case with vanishing x- and z-component of the electric induction, the Hamilton principle can be written as follows:

$$\int\limits_{0}^{L} \int\limits_{(\Sigma)} \left(\frac{1}{\rho_0} \sigma_{11} \delta\varepsilon_{11} + \frac{2}{\rho_0} \sigma_{13} \delta\varepsilon_{13} - \rho_m \, \delta\tilde{\mu}'_\pi + \pi_m^1 \, \delta\tilde{\mu}'_{\pi,x} + \pi_m^3 \, \delta\tilde{\mu}'_{\pi,z} \right) d\Sigma dx = 0.$$

It should be noted that both D_1 and D_3 do not contribute to the variation of the electric field energy $D_i \delta E_i$ of the beam (though the mechanical response is affected by the electric field). Employing the same process as in the previous section, the governing equations for this boundary-value problem can be obtained. It is easy to show that in view of constitutive Eqs. (59)–(61), (63)–(65), the solution to the considered boundary-value problem can be obtained directly from Eqs. (42)–(44), (47)–(49) by replacing $C_{11} \rightarrow \overline{C}_{11}$, $C_{13} \rightarrow \overline{C}_{13}$, $C_{44} \rightarrow \overline{C}_{44}$, $\lambda \rightarrow \Lambda$, and $\Omega \rightarrow \overline{\Omega}$, where

$$\Lambda^2 = \lambda_0^2 \left(1 + M_h + \frac{M}{1 + \delta_e} \right), \quad \overline{\Omega} = \frac{M}{(1 + M_h)(1 + \delta_e)}, \quad \delta_e = \frac{e_{31}^2}{C_{11} a_{33}}.$$

6 Numerical Calculations

In this section, the analytical solutions to the formulated problems are used for the validation of the size-dependent behavior of the elastic and electroelastic nanocantilever beams. The derived solutions are compared to the results predicted by the classical Timoshenko beam model.

To illustrate size effect, the deflection and rotation of cantilever beam was plotted in Figs. 3, 4, 5, 6 and 7. The numerical results are presented for PZT-5H with the following material constants [24, 34]:

$$C_{11} = 12.6 \times 10^{10} \text{ Pa}, \ C_{13} = 5.3 \times 10^{10} \text{ Pa}, \ C_{33} = 11.7 \times 10^{10} \text{ Pa}, \ C_{44} = 3.53 \times 10^{10} \text{ Pa},$$
$$e_{31} = -6.5 \text{ C/m}^2, \ e_{33} = 23.3 \text{ C/m}^2, \ e_{15} = 17.0 \text{ C/m}^2,$$
$$a_{11} = 15.1 \times 10^{-9} \text{ C}^2/(\text{N} \cdot \text{m}^2), \ a_{33} = 13.0 \times 10^{-9} \text{ C}^2/(\text{N} \cdot \text{m}^2).$$

To study the mechanical response of the elastic beam to the external load, the higher-order material parameters α_1^μ, χ_1^m, χ_3^m, and d_μ should be known. However, such data are not yet available in the literature. Thus, the material properties related to the local mass displacement adopted here are: $l_* = \sqrt{\chi_1^m / d_\mu} = 3 \text{ nm}$, and $\chi_3^m / \chi_1^m = 0.2$.

The nanobeam deflection and rotation predicted by the classical Timoshenko beam model (CTBM), the local gradient elastic beam model (LGTBM) and the local gradient piezoelectric beam model (LGPTBM) are plotted in Figs. 3 and 4. In Figs. 3 and 4, dash-dotted lines correspond to the classical problem statement solution while the solid and dashed lines refer to the case of accounting for the local
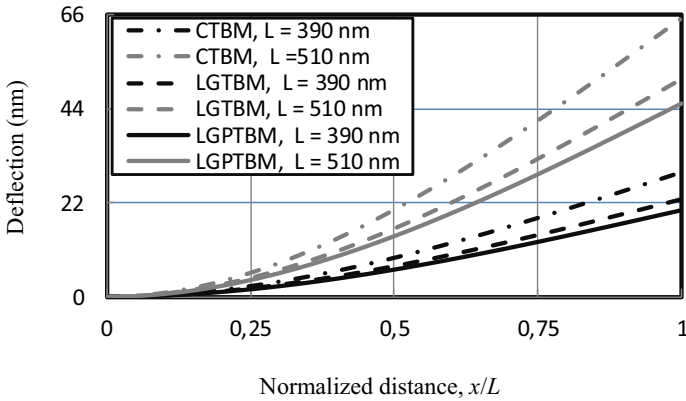
Fig. 3 Deflection predicted by the classical Timoshenko beam model, local gradient Timoshenko beam model with and without the effect of piezoelectricity
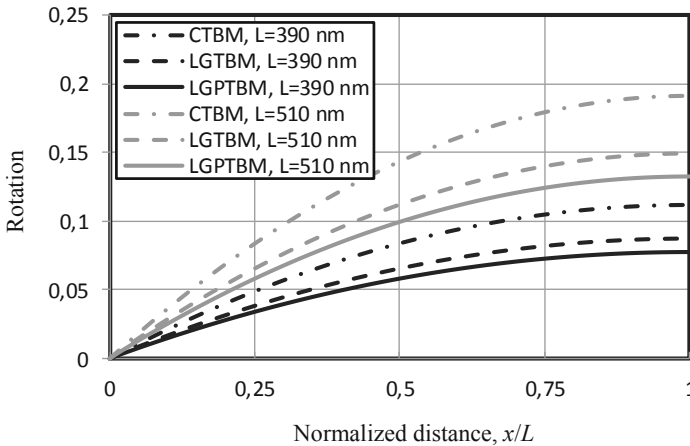
Fig. 4 Cross-section rotation predicted by the classical Timoshenko beam model, local gradient Timoshenko beam model with and without the effect of piezoelectricity

mass displacement in the beam with and without the effect of piezoelectricity. The concentrated force is 1 nN, $M = 0.6$, the length of the beam is 510 nm (gray lines), 390 nm (black lines), the thickness, h, is 20 nm, and the width, b, is 10 nm. The deflection and cross-section rotation predicted by the local gradient theory of the dielectric beam are smaller than that provided by the classical theory and local gradient elastic beam theory. Such a result agrees well with experimental studies and with the local gradient Bernoulli–Euler beam model [14, 15]. It should be noted that the mechanical response of piezoelectric nanobeam is reduced due to both the local mass displacement and the piezoelectric effect.
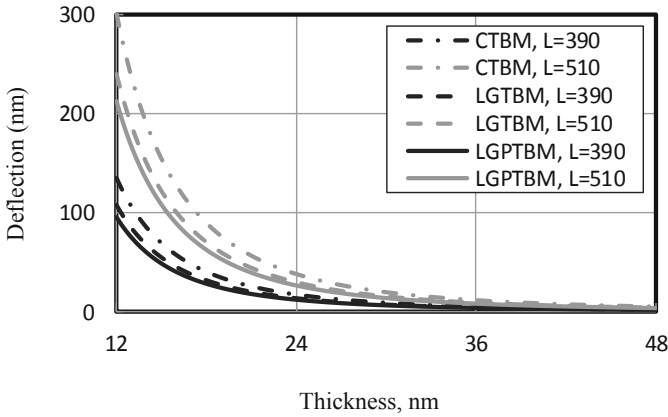
Fig. 5 Variation of the end point beam deflection $w(L)$ with the beam thickness h. Comparison between the classical Timoshenko beam model (CTBM), local gradient Timoshenko beam model (LGTBM), and local gradient Timoshenko piezoelectric beam model (LGPTBM)
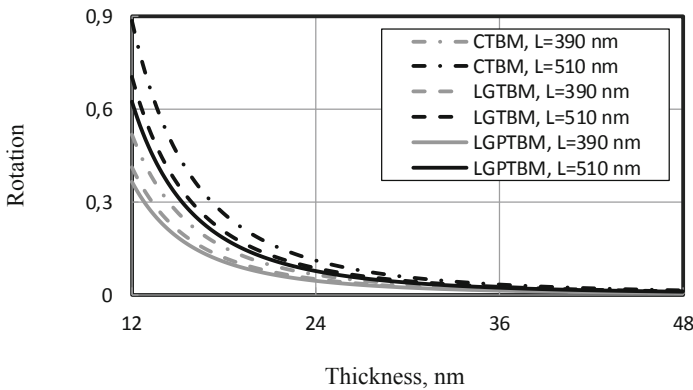
Fig. 6 Variation of the end point rotation $\phi(L)$ with the beam thickness h. Comparison between the classical Timoshenko beam model (CTBM), local gradient Timoshenko beam model (LGTBM), and local gradient Timoshenko piezoelectric beam model (LGPTBM)

Numerical results showing a variation of the end-point deflection $w(L)$ and rotation $\phi(L)$ of the cantilever beam with the beam thickness h are presented in Figs. 5 and 6, respectively. Within the local gradient elastic and dielectric beam models, the deflection increases if the beam thickness decreases. This effect became more pronounced when the beam thickness is close to the material length scale parameter. Since the beam stiffness is directly related to the deflection, the stiffness predicted by the local gradient piezoelectricity is higher than that provided by the
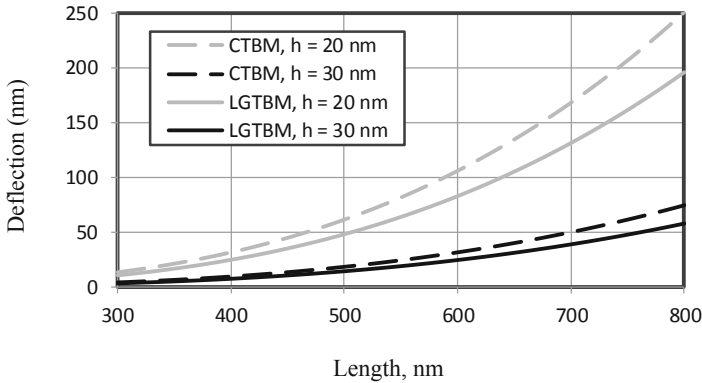
Fig. 7 Variation of the end point beam deflection $w(L)$ with the beam length L. Comparison between the classical Timoshenko beam model (CTBM) and local gradient Timoshenko beam model (LGTBM) for $M = 0.3$, $h = 20$ nm (gray lines) and $h = 30$ nm (black lines)

classical theory. This means that the local mass displacement being taken into account stiffens the piezoelectric nanobeam.

The deflection of the end-point of the cantilever beam versus the different beam lengths is shown in Fig. 7. Here, dashed and solid lines correspond to the solutions obtained in the framework of classical and local gradient problem statements. For small beam thicknesses, the difference between the beam deflections predicted by classical and local gradient theories became more pronounced when the beam length increases.

The presented results indicate that the effect of the local mass displacement is significant in studying the electroelastic behavior of small-scale structures.

7 Conclusion

Based on the gradient-type constitutive relations of the local gradient electrome-chanics, the governing equations and the corresponding boundary conditions for nonferromagnetic piezoelectric solids are derived from the variational principle. The above variational principle was utilized to get an analytical solution to the cantilevered elastic and dielectric beams subjected to the end-point loading. These solutions can be used for testing the numerical methods which should be developed within the local gradient elasticity and electroelasticity.

The main conclusions of the considered boundary-value problems are summarized as follows:

- The deflection and cross-section rotation calculated within the local gradient theory are smaller than those predicted by the classical theory in the same besam geometry and material constants. Thus, the local mass displacement being taken

into account stiffens the nanocantilever beam. The piezoelectric properties of the material being taken into consideration essentially enhance this effect.

- The size effect induced by the local mass displacement is pronounced when the beam thickness is comparable in size to the material length scale parameter. In this case, the differences in the beam deflection and rotation predicted by the local gradient beam theory and classical one may be large. Thus, for nanosize elements, the presented local gradient models of elastic and electroelastic beams should be used in order to obtain accurate results.
- In the case of neglecting the local mass displacement, the local gradient elastic and dielectric beam theories reduced to their classical counterparts.

The presented results indicate that the effect of local mass displacement is significant in studying the small-scale structures. Thus, these theories should be useful for the prediction of mechanical and electromechanical behavior of small-scale structures that have many practical applications in nanoengineering industries.

References

1. Barati, M.R.: On non-linear vibrations of flexoelectric nanobeams. Int. J. Eng. Sci. **121**, 143–153 (2017)
2. Behera, L., Chakraverty, S.: Recent researches on nonlocal elasticity theory in the vibration of carbon nanotubes using beam models: a review. Arch. Comput. Methods Eng. **24**, 481–494 (2017)
3. Bredov, M.M., Rumyantsev, V.V., Toptyhin, I.N.: Classic Electrodynamics (Классическая электродинамика). Nauka, Moscow (1985). (in Russian)
4. Burak, Y.: The equations of electroelasticity of isotropic dielectrics in electrostatic field. Phis. Chim. Mech. Mater. **2**(1), 51–57 (1966). in Russian
5. Burak, Y.I., Kondrat, V.F., Hrytsyna, O.R.: Subsurface mechanoelectromagnetic phenomena in thermoelastic porized bodies in the case of local displacements of mass. Mater. Sci. **43**(4), 449–463 (2007)
6. Burak, Y., Kondrat, V., Hrytsyna, O.: An introduction of the local displacements of mass and electric charge phenomena into the model of the mechanics of polarized electromagnetic solids. J. Mech. Mat. Struct. **3**(6), 1037–1046 (2008)
7. Chapla, Ye., Kondrat, S., Hrytsyna, O., Kondrat, V.: On electromechanical phenomena in thin dielectric films. Task Q. **13**(1–2), 145–154 (2009)
8. Cuenot, S., Fretigny, C., Demoustier-Champagne, S., Nysten, B.: Surface tension effect on the mechanical properties of nanomaterials measured by atomic force microscopy. Phys. Rev. B **69**, 165410(5) (2004)
9. Enakoutsa, K., Della Corte, A., Giorgio, I.: A model for elastic flexoelectric materials including strain gradient effects. Math. Mech. Solids **21**, 242–254 (2016)
10. Eringen, A.C.: Theory of nonlocal piezoelectricity. J. Math. Phys. **25**(3), 717–727 (1984)
11. Eringen, A.C.: Nonlocal Continuum Field Theories. Springer, New York (2002)
12. Hrytsyna, O., Kondrat, V.: Thermomechanics of condensed systems taking the local mass displacement into account: I. Theory Fundamentals. "Rastr – 7", Lviv (2017). (in Ukrainian)

13. Hrytsyna, O., Kondrat, V.: Local Gradient Theory for Dielectrics: Fundamentals and Applications, 1st edn. Jenny Stanford Publishing Pte. Ltd., Singapore (2020)
14. Hrytsyna, O.: Bernoulli-Euler beam model based on local gradient theory of elasticity. J. Mech. Mat. Struct. **15**(4), 471–487 (2020)
15. Hrytsyna, O.: Local gradient Bernoulli-Euler beam model for dielectrics: effect of local mass displacement on coupled field. Math. Mech. Solids (2020). https://journals.sagepub.com/eprint/3KT3SIZHNJMUN2QAGAID/full
16. Hrytsyna, O.: Applications of the local gradient elasticity to the description of the size effect of shear modulus. SN Appl. Sci. **2**, 1453(9) (2020)
17. Kogan, S.M.: Piezoelectric effect during inhomogeneous deformation and acoustic scattering of carriers in crystals. Sov. Phys. Solid State **5**, 2069–2070 (1964)
18. Lam, D.C.C., Yang, F., Chong, A.C.M., Wang, J., Tong, P.: Experiments and theory in strain gradient elasticity. J. Mech. Phys. Solids **51**(8), 1477–1508 (2003)
19. Landau, L.D., Lifshitz, E.M.: Electrodynamics of Continuum Media, 2nd edn. Butterworth-Heinemann, Oxford (1984)
20. Li, X.-F., Wang, B.-L., Lee, K.Y.: Size effect in the mechanical response of nanobeams. J. Adv. Res. Mech. Eng. **1**(1), 4–16 (2010)
21. Liang, X., Hu, S., Shen, S.: Bernoulli–Euler dielectric beam model based on strain-gradient effect. J. Appl. Mech. **80**, 044502(6) (2013)
22. Liang, X., Hu, S., Shen, S.: Size-dependent buckling and vibration behaviors of piezoelectric nanostructures due to flexoelectricity. Smart Mater. Struct. **24**, 105012(13) (2015)
23. Liu, C., Ke, L.L., Wang, Y.S., Yang, J., Kitipornchai, S.: Buckling and post-buckling of size-dependent piezoelectric Timoshenko nanobeams subject to thermo-electro-mechanical loadings. Int. J. Struct. Stab. Dyn. **14**(03), 1350067 (2014)
24. Ma, W.H., Cross, L.E.: Flexoelectricity of barium titanate. Appl. Phys. Lett. **88**(23), 232902 (2006)
25. McFarland, A.W., Colton, J.S.: Role of material microstructure in plate stiffness with relevance to microcantilever sensors. J. Micromech. Microeng. **15**(5), 1060–1067 (2005)
26. Mindlin, R.D.: Polarization gradient in elastic dielectrics. Int. J. Solids Struct. **4**, 637–642 (1968)
27. Nowacki, W.: Teoria sprężystości, Państwowe Wydawnictwo Naukowe, Warszawa (1970). (in Polish)
28. Park, S.K., Gao, X.-L.: Bernoulli-Euler beam model based on a modified couple stress theory. J. Micromech. Microeng. **16**, 2355–2359 (2006)
29. Singh, M., Verma, P.D.S.: Nonlinear couple stress theory of elastic dielectrics with applications to dynamic deformations. J. Elasticity **13**, 379–393 (1983)
30. Thai, H.-T., Vo, T.P., Nguyen, T.-K., Kim, S.-E.: A review of continuum mechanics models for size-dependent analysis of beams and plates. Compos. Struct. **177**, 196–219 (2017)
31. Weinberg, M.S.: Working equations for piezoelectric actuators and sensors. J. Microelectromech. Syst. **8**(4), 529–533 (1999)
32. Yan, Z., Jiang, L.: Flexoelectric effect on the electroelastic responses of bending piezoelectric nanobeams. J. Appl. Phys. **113**, 194102 (2013)
33. Yang, X.M., Hu, Y.T., Yang, J.S.: Electric field gradient effects in antiplane problems of polarized ceramics. Int. J. Solids Struct. **41**, 6801–6811 (2004)
34. Zhang, R., Liang, X., Shen, S.: A Timoshenko dielectric beam model with flexoelectric effect. Meccanica **51**, 1181–1188 (2016)

Bending of Piezoelectric FGM Plates Under Thermal Loads

Ladislav Sator, Vladimir Sladek, and Jan Sladek

Abstract In this paper we derived a formulation for bending of thermo-piezoelectric FGM plates with taking into account the assumptions of the Kirchhoff–Love plate bending theory within the classical theory of thermoelasticity. Uncoupled thermoelasticity is considered; thus, the heat conduction problem is analyzed independently of the mechanical fields. The functional gradation of material coefficients complies the rule of mixture. For numerical solution of complex boundary value problems a strong form meshless method is developed with using the Moving Least Square approximation for spatial variations of field variables. The inaccuracy of solution due to the high order derivatives of field variables is overcome by decomposing the original governing partial differential equations (PDE) into the system of PDEs with lower order derivatives. Several numerical experiments are presented for illustration of coupling effects in bending of FGPM plates under stationary thermal loading.

Keywords FGPM plates · Kirchoff-Love bending theory · Classical thermoelasticity · Meshless strong formulation · MLS approximation

1   Introduction

Nowadays, the selfmonitoring of the health of structures [1, 2] is the important part of engineering practice. For this purpose several types of sensors and actuators were developed. The smart materials and structures, such as piezoelectric ones are commonly used like sensors and actuators in environment with variable temperature, where the thermal effect on the piezoelectric devices is not negligible [3]. These materials are usually made as laminated composites from ceramic slices. It is well known the discontinuities on interfaces between two layers in such a composites lead to concentration of gradient fields, which is the significant factor of failure of laminated structures. This negative phenomena of delamination can be

L. Sator (✉) · V. Sladek · J. Sladek
Institute of Construction and Architecture SAS, Dubravska cesta 9, Bratislava, Slovakia
e-mail: ladislav.sator@savba.sk

overcome by replacing laminated piezoelectric plate structures by functionally graded ones, if possible. Therefore the investigation of the response of functionally graded piezoelectric (FGPM) plates subjected to thermal loadings has become attractive, what can be confirmed by the large number of papers [4–7] on this topic. In this paper, the bending of thin FGPM plates with transversal gradation of material coefficients under stationary thermal loading is considered within the classical thermoelasticity and Kirchhoff–Love's theory of plates. In order to overcome the inaccuracy of approximation of high order derivatives of field variables the governing equations which are given by the 4th order PDE are decomposed into the 2nd order PDEs. The strong form meshless formulations in combination with Moving Least Squares (MLS) approximation scheme [8] are developed for solution of boundary value problems for FGPM plates. The influence of transversal gradations of various material coefficients on various physical fields of FGPM plate is presented by the results of several numerical experiments.

2 Governing Equations for Piezoelectric FGM Plates

2.1 Formulation of Boundary Value Problem

Let us consider a thermo-piezoelectric plate with spatially variable material properties. Then the constitutive equations of thermo- piezoelectric solids are given as

$$\sigma_{ij}(\mathrm{x}, x_3) = c_{ijkl}(\mathrm{x}, x_3)s_{kl}(\mathrm{x}, x_3) - e_{kij}(\mathrm{x}, x_3)E_k(\mathrm{x}, x_3) - \gamma_{ij}(\mathrm{x}, x_3)\theta(\mathrm{x}, x_3),$$

$$D_k(\mathrm{x}, x_3) = e_{kij}(\mathrm{x}, x_3)s_{ij}(\mathrm{x}, x_3) + \chi_{ki}(\mathrm{x}, x_3)E_i(\mathrm{x}, x_3) + p_j(\mathrm{x}, x_3)\theta(\mathrm{x}, x_3), \quad (1)$$

with the third coordinate axis being orthogonal to the plate plane. The elastic strains s_{ij} and electric field E_k being the natural variables, which are expressed in terms of primary field variables, displacements v_i and electric potential ϕ, as $s_{ij} = \frac{1}{2}(v_{i,j} + v_{j,i}), E_i = -\phi_{,i}$.

Many piezoelectric ceramic materials have transversely isotropic elastic properties with hexagonal *mm6* class of symmetry. In this case the poling direction is parallel to the x_3 axis and the plane perpendicular to poling direction being the isotropic plane. Then the Voigt notation is used for elastic material coefficients in matrix form of constitutive equations:

$$
\begin{pmatrix}
\sigma_{11} \\
\sigma_{22} \\
\sigma_{33} \\
\sigma_{23} \\
\sigma_{13} \\
\sigma_{12} \\
D_1 \\
D_2 \\
D_3
\end{pmatrix}
=
\begin{pmatrix}
c_{11}^E & c_{12}^E & c_{13}^E & 0 & 0 & 0 & 0 & 0 & -e_{31} \\
c_{12}^E & c_{11}^E & c_{13}^E & 0 & 0 & 0 & 0 & 0 & -e_{31} \\
c_{13}^E & c_{13}^E & c_{33}^E & 0 & 0 & 0 & 0 & 0 & -e_{33} \\
0 & 0 & 0 & c_{44}^E & 0 & 0 & 0 & -e_{15} & 0 \\
0 & 0 & 0 & 0 & c_{44}^E & 0 & -e_{15} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & c_{66}^E & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & e_{15} & 0 & \chi_{11}^\varepsilon & 0 & 0 \\
0 & 0 & 0 & e_{15} & 0 & 0 & 0 & \chi_{11}^\varepsilon & 0 \\
e_{31} & e_{31} & e_{33} & 0 & 0 & 0 & 0 & 0 & \chi_{33}^\varepsilon
\end{pmatrix}
\begin{pmatrix}
s_{11} \\
s_{22} \\
s_{33} \\
s_{23} \\
s_{13} \\
s_{12} \\
E_1 \\
E_2 \\
E_3
\end{pmatrix}
$$

$$
-
\begin{pmatrix}
c_{11}^E & c_{12}^E & c_{13}^E & 0 & 0 & 0 & 0 & 0 \\
c_{12}^E & c_{11}^E & c_{13}^E & 0 & 0 & 0 & 0 & 0 \\
c_{13}^E & c_{13}^E & c_{33}^E & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\alpha_{11} \\
\alpha_{11} \\
\alpha_{33} \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}
\theta +
\begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
p_1 \\
p_1 \\
p_3
\end{pmatrix}
\theta, \qquad (2)
$$

where $\quad c_{66}^E = \frac{1}{2}(c_{11}^E - c_{12}^E)$, $\quad \begin{pmatrix} \gamma_{11} \\ \gamma_{11} \\ \gamma_{33} \end{pmatrix} = \begin{pmatrix} c_{11}^E & c_{12}^E & c_{13}^E \\ c_{12}^E & c_{11}^E & c_{13}^E \\ c_{13}^E & c_{13}^E & c_{33}^E \end{pmatrix} \begin{pmatrix} \alpha_{11} \\ \alpha_{11} \\ \alpha_{33} \end{pmatrix}$, and the superscript variables denote that those variables are kept constant during the variation of other variables.

In stationary problems, the distribution of temperature field θ is independent on the mechanical and electric fields, and is governed by the heat conduction equation

$$
\left(k_{ij}\theta_{,j}\right)_{,i} = w, \qquad
\begin{pmatrix}
k_{11} & 0 & 0 \\
0 & k_{11} & 0 \\
0 & 0 & k_{33}
\end{pmatrix}. \qquad (3)
$$

where the full thermal conductivity tensor reduces to the diagonal form as shown above, and w is the density of heat sources. The governing equation for elastic fields in stationary theory is given by the force of equilibrium as

$$
\sigma_{ij,j} + f_i = 0, \qquad (4)
$$

while governing equation for electric fields is by Maxwell's equation as

$$
D_{j,j} = 0. \qquad (5)
$$

2.2 Derivation of Unified Formulation for FGM Plates

Let us consider a transversally isotropic FGPM plates with the power-law transversal gradation of elastic, piezoelectric, dielectric material coefficients and mass density as:

$$c_{ab}(\mathbf{x}, x_3) = c_{ab(0)} c_V^E(x_3), \qquad c_V^E(x_3) = 1 + \zeta \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^p$$
$$a, b = \{1, 2, 3, 4, 5, 6\}$$

$$e_{ab}(\mathbf{x}, x_3) = e_{ab(0)} e_V(x_3), \qquad e_V(x_3) = 1 + \xi \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^r$$

$$\chi_{ij}^{\varepsilon}(\mathbf{x}, x_3) = \chi_{ij(0)}^{\varepsilon} \chi_V^{\varepsilon}(x_3), \chi_V^{\varepsilon}(x_3) = 1 + \lambda \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^s, \quad i, j = \{1, 2, 3\}$$

$$\alpha_{ij}(\mathbf{x}, x_3) = \alpha_{ij(0)} \alpha_V(x_3), \qquad \alpha_V(x_3) = 1 + \psi \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^l$$

$$k_{ij}(\mathbf{x}, x_3) = k_{ij(0)} k_V(x_3), \qquad k_V(x_3) = 1 + \omega \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^n$$

$$p_j(\mathbf{x}, x_3) = p_{j(0)} p_V(x_3), \qquad p_V(x_3) = 1 + \upsilon \left(\frac{1}{2} \pm \frac{x_3}{h} \right)^q, \tag{6}$$

where $\zeta, \xi, \lambda, \psi, \omega, \upsilon$ and p, r, s, l, n, q are the values of the level and exponent of the power-law gradation, respectively.

Before writing of the governing equations in Kirchhoff – Love (classical) theory (KLT) of thin plates for thermo-piezoelectric boundary value problems (BVP) we define the dimensionless coordinates as

$$x_{\beta}^* := \frac{x_{\beta}}{L}, x_3^* := \frac{x_3}{h_0} = h^*(\mathbf{x}) z, \quad h(\mathbf{x}) = h_0 h^*(\mathbf{x}), \tag{7}$$

with L, h_0 being the characteristic in-plane dimension and plate thickness. The dimensionless field variables in the KLT are the in-plane displacements, deflections, electric potential and temperature, which are introduced as

$$u_{\beta}^*(\mathbf{x}) := \frac{u_{\beta}(\mathbf{x})}{h_0}, w^*(\mathbf{x}, t) := \frac{w(\mathbf{x})}{h_0}, \phi_0^*(\mathbf{x}) := \frac{\phi_0(\mathbf{x})}{\Phi_0} \Phi^{+*}(\mathbf{x}) := \frac{\Phi^+(\mathbf{x})}{\Phi_0},$$

$$\Phi^{-*}(\mathbf{x}) := \frac{\Phi^-(\mathbf{x})}{\Phi_0}, \vartheta_0^*(\mathbf{x}) := \frac{\vartheta_0(\mathbf{x})}{\theta_0}, \Theta^{+*}(\mathbf{x}) := \frac{\Theta^+(\mathbf{x})}{\theta_0}, \Theta^{-*}(\mathbf{x}) := \frac{\Theta^-(\mathbf{x})}{\theta_0}, \tag{8}$$

where we have omitted the superscript $*$ in coordinates.

The truncated expansions of the electric potential and temperature fields w.r.t. the z-coordinate are expressed as: $\phi^*(\mathbf{x}, z) \approx (1 - 4z^2)\phi_0^*(\mathbf{x}) + z\Phi^{*-}(\mathbf{x}) + 2z^2\Phi^{*+}(\mathbf{x})$, $\theta^*(\mathbf{x}, z) \approx (1 - 4z^2)\vartheta_0^*(\mathbf{x}) + z\Theta^{*-}(\mathbf{x}) + 2z^2\Theta^{*+}(\mathbf{x})$, respectively. The expressions $\Phi^{\pm}(\mathbf{x})$ and $\Theta^{\pm}(\mathbf{x})$ are staying for $\Phi^{\pm}(\mathbf{x}) = \phi^+(\mathbf{x}) \pm \phi^-(\mathbf{x})$ and $\Theta^{\pm}(\mathbf{x}) = \theta^+(\mathbf{x}) \pm \theta^-(\mathbf{x})$, where $\phi^+(\mathbf{x}), \theta^+(\mathbf{x})$ and $\phi^-(\mathbf{x})$ $\theta^-(\mathbf{x})$ are the electric potential and temperature prescribed on the top and bottom surface of the plate, respectively.

Combination of the variational formulation for thermo-piezoelectric problems with assumptions of the KLT for thin plates yields the strong form of the governing equations in view of the PDE

$$\left(G^{(\theta)}(\mathbf{x})\vartheta_{0,\beta}^*(\mathbf{x})\right)_{,\beta} - C^{(\theta)}(\mathbf{x})\vartheta_0^*(\mathbf{x})$$
$$= C^{(\theta)}(\mathbf{x}) + R^-(\mathbf{x}) + R^+(\mathbf{x}) - \left(R_\beta^-(\mathbf{x}) - R_\beta^+(\mathbf{x})\right)_{,\beta} \tag{9a}$$

$$T_{\alpha\beta,\beta}^*(\mathbf{x}) = 0 \tag{9b}$$

$$M_{\alpha\beta,\alpha\beta}^{*(w)}(\mathbf{x}) = -\bar{t}_3^*(\mathbf{x}) \tag{9c}$$

$$B_{\alpha,\alpha}^*(\mathbf{x}) - B_3^*(\mathbf{x}) = 0, \tag{9d}$$

where the semi-integral fields and dimensionless coefficients $C^{(\theta)}(\mathbf{x}), G^{(\theta)}(\mathbf{x})$, and $R^{\pm}(\mathbf{x})$ are defined in the Appendix. The dimensionless heat flux and transversal load are defined as $\bar{q}^* = \left(L/\theta_0 k_{11(0)}\right)\bar{q}, \bar{t}_3^*(\mathbf{x}) := \left(L^4/D_0 h_0\right)\bar{t}_3(\mathbf{x})$, respectively, where $D_0 = c_{11(0)}^E(h_0)^3$ is the bending stiffness of the homogeneous plate.

The most frequently applied boundary conditions for mechanical fields on the boundary edge of the plate are given as follows:

(i) on clamped edge:

$$u_\alpha^*(\mathbf{x})\big|_{\partial\Omega} = 0, \quad w^*(\mathbf{x})\big|_{\partial\Omega} = 0, \quad \frac{\partial w^*}{\partial \mathbf{n}}(\mathbf{x})\bigg|_{\partial\Omega} = 0, \tag{10a}$$

(ii) on simply supported edge:

$$n_\beta(\mathbf{x})T_{\alpha\beta}^*(\mathbf{x})\bigg|_{\partial\Omega} = 0, \quad w^*(\mathbf{x})\big|_{\partial\Omega} = 0, \quad n_\alpha(\mathbf{x})n_\beta(\mathbf{x})M_{\alpha\beta}^{*(w)}(\mathbf{x})\bigg|_{\partial\Omega} = 0, \tag{10b}$$

while for electric potential and temperature field on the boundary edge $\partial\Omega$ we name the following boundary conditions.

(i) Dirichlet type:

$$\text{electric potential}: \phi_0^*(\mathbf{x})\big|_{\partial\Omega} = \overline{\phi}_0(\mathbf{x}) \tag{11a}$$

$$\text{temperature}: \vartheta_0^*(\mathbf{x})\big|_{\partial\Omega} = \overline{\theta}^*(\mathbf{x}) - \theta_0 \tag{11b}$$

(ii) Neuman type:

$$\text{electric potential}: n_\alpha(\mathbf{x})B_\alpha^*(\mathbf{x})\big|_{\partial\Omega} = 0, \tag{11c}$$

$$\text{temperature}: n_\beta(\mathbf{x})\tilde{G}^{(\theta)}(\mathbf{x})\vartheta_{0,\beta}(\mathbf{x})\big|_{\partial\Omega} = -\overline{q}(\mathbf{x})\tilde{C}^{(q)}(\mathbf{x}) - n_\beta(\mathbf{x})\left[\tilde{R}_\beta^-(\mathbf{x}) + \tilde{R}_\beta^+(\mathbf{x})\right], \tag{11d}$$

The semi-integral fields in Eq. (9) include higher order derivatives of primary field variables, which lead to increasing inaccuracy of the solution. To overcome this problem, we decompose original 4th order governing PDEs into coupled set of 2nd order PDEs by introducing new field variables as

$$m^*(\mathbf{x}) := \nabla^2 w^*(\mathbf{x}), \qquad s_\alpha^*(\mathbf{x}) := \nabla^2 u_\alpha^*(\mathbf{x}) \tag{12}$$

Then, the complet set of governing equations for primary field variables $\left\{ w^*(\mathbf{x}), m^*(\mathbf{x}), u_\beta^*(\mathbf{x}), s_\beta^*(\mathbf{x}), \phi_0^*(\mathbf{x}), \vartheta_0^*(\mathbf{x}) \right\}$ is given by Eqs. (9) and (12).

3   Numerical Implementation

In this paper, we have used the strong formulation with employing the meshless approximation for primary field variables by Central Approximation Node (CAN) concept of Moving Least Square (MLS) approximation technique [9, 10].

Without going into details [11], the approximation of field variable $f(\mathbf{x})$ around the central approximation node $\mathbf{x}^q$ can be expressed by

$$f(\mathbf{x}, t) \approx \sum_{a=1}^{N^q} \hat{f}^{\overline{a}}(t)\Psi^{(q,a)}(\mathbf{x}), \qquad \overline{a} = n(q, a), \tag{13}$$

where $\overline{a}$ is the global number of the a-th node from the influence domain of $\mathbf{x}^q$, N^q is the number of nodal points in the influence domain, and $\Psi^{(q,a)}(\mathbf{x})$ is the shape function associated with the node $n(q, a)$. The CAN can be selected as the nearest node to the field point $\mathbf{x}$.

The derivatives of the field variable $f(\mathbf{x})$ can be approximated by derivatives of approximated fields [11], i.e.

$$f_{,ijk...l}(\mathbf{x}, t) \approx \sum_{a=1}^{N^q} \hat{f}^{\bar{a}}(t) \Psi_{,ijk...l}^{(q,a)}(\mathbf{x}). \tag{14}$$

4 Numerical Experiments

Within this paper we shall investigate the bending response of thin FGPM plates under stationary thermal loads within the KLT, we consider a thin square plate $L \times L$ with, $L = 1$ m, and transversal gradations of material coefficients. Poisson's ratio is assumed to be constant $v = 0.3$. The material coefficients corresponding to BaTiO$_3$ [12]. The stationary thermal loading is applied as uniformly distributed temperatures on the top and bottom surfaces $\Theta^{\pm *}(\mathbf{x}) := \theta^{+*}(\mathbf{x}) - \theta^{-*}(\mathbf{x}) = const = 20$, while $\bar{t}_3(\mathbf{x}) = 0$. The plate edges are clamped (CE); while the electric boundary conditions are assumed to be Dirichlet (DE) electric boundary conditions and the heat fluxes are vanishing on all the edges.

In all numerical computations, we have used a uniform distribution of nodal points (36×36 nodes) with δ being the distance between two neighbour nodes. The other parameters in the MLS-approximation have been taken as: radius of the interpolation domain $\rho^a = 3.001\delta$, shape function parameter $c^a = \delta$, and cubic polynomial basis.

First of all we shall present the in-plane distribution of temperature (Fig. 1). It can be seen that the temperature field in FGPM plates is affected only by the transversal gradation of the heat conduction coefficient.

In the following, we shall present the of numerical results for in-plane displacements of the FGPM plates. The influence of transversal gradation of various material coefficient on in-plane displacements can be observed on Fig. 2. The transversal power-law gradation of elastic coefficient, the heat conduction coefficient and the thermal expansion coefficient play the crucial role for the in-plane displacement responses of FGPM plates subjected to stationary thermal load. The gradation of dielectric, piezoelectric and pyroelectric coefficients have no influence on the distribution of in-plane displacements.

The in-plane distribution of deflections of the FGPM plates are illustrated on Fig. 3. It is shown the gradation of material coefficients has significant influence on the maximum value of deflections. It is observable, the transversal gradation of elastic and dielectric coefficients lead to the reduction of maximum deflections, while transversal gradation of piezoelectric and pyroelectric coefficients have an opposite effect. Also, the transversal functional gradation of heat conduction and thermal expansion coefficient has no influence on the deflection response of FGPM plate.

Finally, Fig. 4. shows that the influence of the transversal gradation of the piezoelectric, dielectric and pyroelectric coefficent on the in-plane distribution of electric potential field is significant, while the effect of the gradation of other

KLT

$c_v^*(z)=1+\zeta(1/2+z)^p$ $\alpha_v^*(z)=1+\psi(1/2+z)^l$

$e_v^*(z)=1+\xi(1/2+z)^r$ $k_v^*(z)=1+\omega(1/2+z)^n$

$\chi_v^*(z)=1+\lambda(1/2+z)^s$ $p_v^*(z)=1+\upsilon(1/2+z)^q$

$p=r=s=l=n=q=1$

Legend:
- homog.
- $\zeta=1, \xi=0, \lambda=0, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=1, \lambda=0, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=1, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=1, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=0, \omega=1, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=0, \omega=0, \upsilon=1$

$L/h_0=50$

y-axis: $\vartheta_0(x_1, x_2=0.5)$ x-axis: x_1
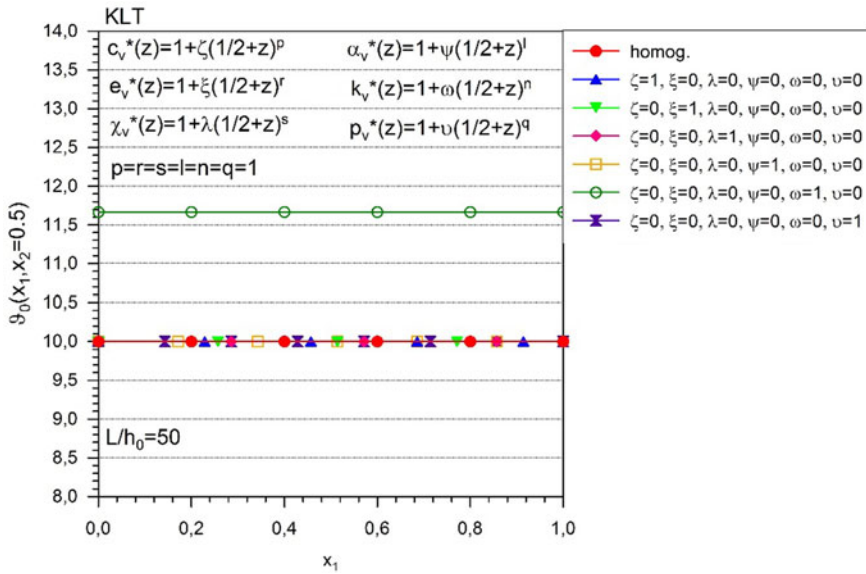
Fig. 1 In-plane distribution of temperature in homogeneous and/or FGPM plates under stationary thermal loading

KLT

$c_v^*(z)=1+\zeta(1/2+z)^p$ $p=r=s=l=n=q=1$

$e_v^*(z)=1+\xi(1/2+z)^r$

$\chi_v^*(z)=1+\lambda(1/2+z)^s$

$\alpha_v^*(z)=1+\psi(1/2+z)^l$

$k_v^*(z)=1+\omega(1/2+z)^n$

$p_v^*(z)=1+\upsilon(1/2+z)^q$

Legend:
- homog.
- $\zeta=1, \xi=0, \lambda=0, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=1, \lambda=0, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=1, \psi=0, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=1, \omega=0, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=0, \omega=1, \upsilon=0$
- $\zeta=0, \xi=0, \lambda=0, \psi=0, \omega=0, \upsilon=1$

$L/h_0=50$

y-axis: $u_1^*(x_1, x_2=0.5)$ x-axis: x_1
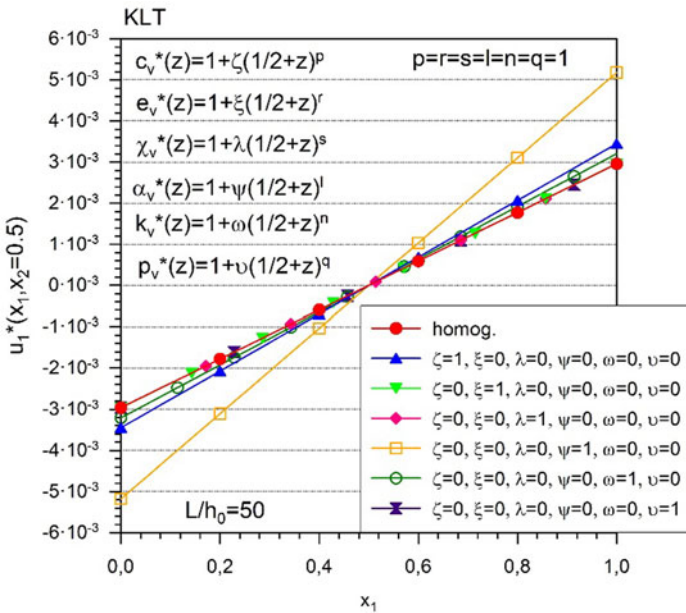
Fig. 2 In-plane distribution of in-plane displacements in homogeneous and/or FGPM plates under stationary thermal loading
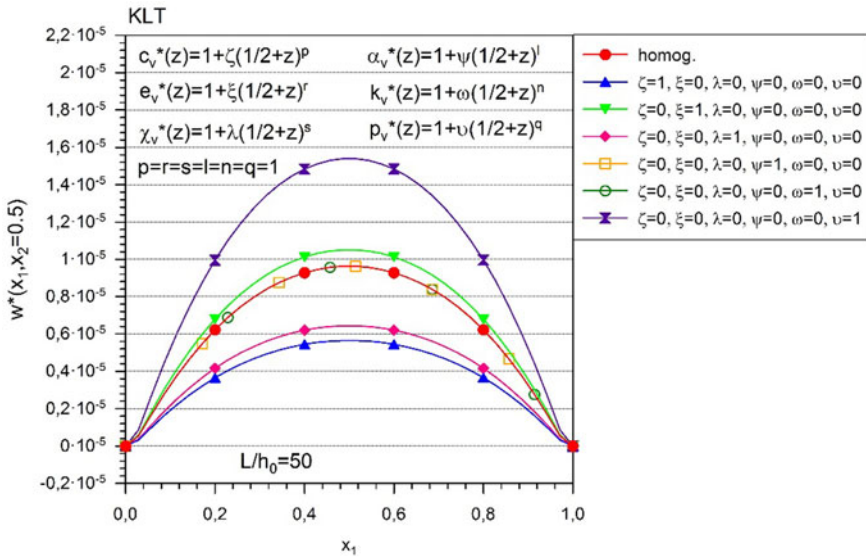
Fig. 3 In-plane distribution of deflections in homogeneous and/or FGPM plates under stationary thermal loading

material coefficients is negligible. The transversal gradation of piezoelectric and dielectric coefficients lead to the decreasing of electric potential, while transversal gradation of pyroelectric coefficient leads to increasing of the value of the electric potential.

5 Conclusions

The governing equations have been developed for FGPM plate bending problems in Kirchhoff–Love (KLT) plate bending theory. The transversal functional gradation of material coefficients is considered, which lead to coupling effects between mechanical fields. The strong form meshless method combined with MLS approximation scheme were developed for solution of boundary value problems of FGPM plates. Numerical examples have been performed in order to investigate the effects of transversal gradations of material coefficients within FGPM plates subjected to stationary thermal loadings.
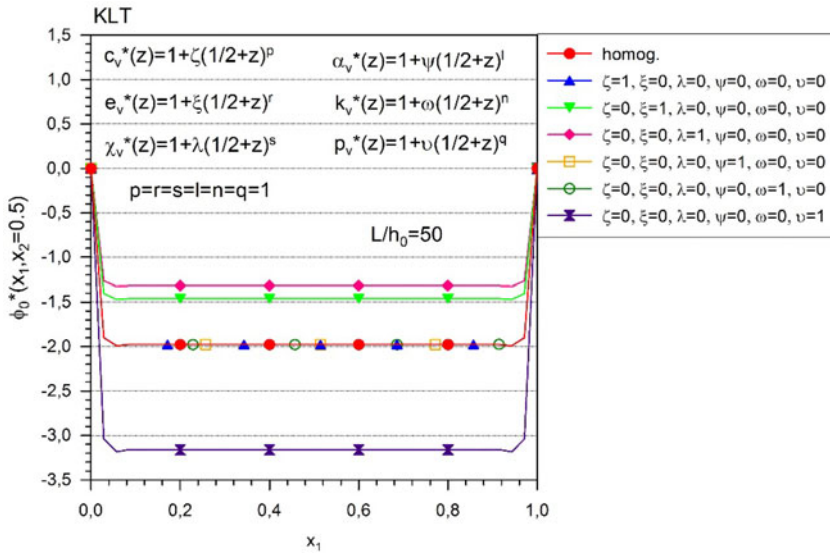
Fig. 4 In-plane distribution of electric potential in homogeneous and/or FGPM plates under stationary thermal loading

Appendix

The integrations prescribed in Eq. (14) can be performed in closed form and expressed in terms of two kinds of integrals

$$d_{(0)a} := \int_{-1/2}^{1/2} z^a dz = \frac{1+(-1)^a}{(a+1)2^{a+1}} \tag{A.1}$$

$$d_{(p)a} := \int_{-1/2}^{1/2} z^a \left(\frac{1}{2}\pm z\right)^p dz = (\pm 1)^a \int_0^1 \left(y-\frac{1}{2}\right)^a y^p dy$$

$$= (\pm 1)^a \sum_{k=0}^a \binom{a}{k}\left(-\frac{1}{2}\right)^k \frac{1}{p+1+a-k}$$

The coefficients occurring in thermal problem in Eq. (28a) are given as

$$C^{(\theta)}(\mathrm{x}) = 64\left(d_{(0)2} + \omega d_{(n)2}\right)\frac{k_{33(0)}k_H(\mathrm{x})}{k_{11(0)}h^*(\mathrm{x})}$$

$$G^{(\theta)}(\mathbf{x}) = \left[d_{(0)0} - 8d_{(0)2} + 16d_{(0)4} + \omega\left(d_{(n)0} - 8d_{(n)2} + 16d_{(n)4}\right) \right] \left(\frac{h_0}{L}\right)^2 k_H(\mathbf{x})h^*(\mathbf{x})R^+(\mathbf{x})$$

$$= -\frac{1}{2}C^{(\theta)}(\mathbf{x})\Theta^{+*}(\mathbf{x})$$

$$R^-(\mathbf{x}) = -8\omega d_{(n)1}\frac{k_{33(0)}k_H(\mathbf{x})}{k_{11(0)}h^*(\mathbf{x})}\Theta^{-*}(\mathbf{x})$$

$$R_\beta^+(\mathbf{x}) = 2\left(\frac{h_0}{L}\right)^2 \left[d_{(0)2} - 4d_{(0)4} + \omega\left(d_{(n)2} - 4d_{(n)4}\right)\right]k_H(\mathbf{x})h^*(\mathbf{x})\Theta_{,\beta}^{+*}(\mathbf{x})$$

$$R_\beta^-(\mathbf{x}) = \left(\frac{h_0}{L}\right)^2 \omega\left(d_{(n)1} - 4d_{(n)3}\right)k_H(\mathbf{x})h^*(\mathbf{x})\Theta_{,\beta}^{-*}(\mathbf{x})$$

$$C^{(q)}(\mathbf{x}) = \left(\frac{h_0}{L}\right)^2 \left(d_{(0)0} - 4d_{(0)2}\right)h^*(\mathbf{x}), \qquad q^* = \frac{L}{\theta_0 k_{11(0)}}q, \qquad \text{(A.2)}$$

The dimensionless semi-integral fields in Eqs. (28b)–(28e) are given as

$$\begin{aligned}
T_{11}^*(\mathbf{x}) =\, & C_{11}^{(uu)}(\mathbf{x})u_{1,1}^*(\mathbf{x}) + C_{12}^{(uu)}(\mathbf{x})u_{2,2}^*(\mathbf{x}) + C_{11}^{(uw)}(\mathbf{x})w_{,11}^*(\mathbf{x}) + C_{12}^{(uw)}(\mathbf{x})w_{,22}^*(\mathbf{x}) \\
& + C_{31}^{(u\Phi^-)}(\mathbf{x})\Phi^{-*}(\mathbf{x}) + C_{31}^{(u\Phi^+)}(\mathbf{x})\Phi^{+*}(\mathbf{x}) + C_{31}^{(u\phi_0)}(\mathbf{x})\phi_0^*(\mathbf{x}) + C_{11}^{(u\Theta^-)}(\mathbf{x})\Theta^{-*}(\mathbf{x}) \\
& + C_{11}^{(u\Theta^+)}(\mathbf{x})\left(\Theta^{+*}(\mathbf{x}) - 2\right) + \left(C_{11}^{(u\vartheta_0)}(\mathbf{x}) - 2C_{11}^{(u\Theta^+)}(\mathbf{x})\right)\vartheta_0^*(\mathbf{x})
\end{aligned}$$

$$\begin{aligned}
T_{22}^*(\mathbf{x}) =\, & C_{12}^{(uu)}(\mathbf{x})u_{1,1}^*(\mathbf{x}) + C_{11}^{(uu)}(\mathbf{x})u_{2,2}^*(\mathbf{x}) + C_{12}^{(uw)}(\mathbf{x})w_{,11}^*(\mathbf{x}) + C_{11}^{(uw)}(\mathbf{x})w_{,22}^*(\mathbf{x}) \\
& + C_{31}^{(u\Phi^-)}(\mathbf{x})\Phi^{-*}(\mathbf{x}) + C_{31}^{(u\Phi^+)}(\mathbf{x})\Phi^{+*}(\mathbf{x},t) + C_{31}^{(u\phi_0)}(\mathbf{x})\phi_0^*(\mathbf{x},t) + C_{11}^{(u\Theta^-)}(\mathbf{x})\Theta^{-*}(\mathbf{x}) \\
& + C_{11}^{(u\Theta^+)}(\mathbf{x})\left(\Theta^{+*}(\mathbf{x}) - 2\right) + \left(C_{11}^{(u\vartheta_0)}(\mathbf{x}) - 2C_{11}^{(u\Theta^+)}(\mathbf{x})\right)\vartheta_0^*(\mathbf{x})
\end{aligned}$$

$$T_{12}^*(\mathbf{x},t) = C_{66}^{(uu)}(\mathbf{x})\left(u_{1,2}^*(\mathbf{x},t) + u_{2,1}^*(\mathbf{x},t)\right) + 2C_{66}^{(uw)}(\mathbf{x})w_{,12}^*(\mathbf{x},t)$$

$$\text{(A.3)}$$

$$C_{ab}^{(uu)}(\mathbf{x}) = \frac{L(h_0)^2}{D_0}\left(d_{(0)0} + \zeta d_{(p)0}\right)c_{ab(0)}^E c_{(H)}^E(\mathbf{x})h^*(\mathbf{x})$$

$$C_{ab}^{(uw)}(\mathbf{x}) = \frac{(h_0)^3}{D_0}\zeta\left((c_1 - 1)d_{(p)1} - \frac{4}{3}c_1 c_2 d_{(p)3}\right)c_{ab(0)}^E c_{(H)}^E(\mathbf{x})h^{*2}(\mathbf{x})$$

$$C_{3\alpha}^{(u\Phi^-)}(x) = \frac{L^2\Phi_0}{D_0}\left(d_{(0)0} + \xi d_{(r)0}\right)e_{3\alpha(0)}e_{(H)}(x)$$

$$C_{3\alpha}^{(u\Phi^+)}(x) = \frac{L^2\Phi_0}{D_0}4\xi d_{(r)1}e_{3\alpha(0)}e_{(H)}(x)$$

$$C_{3\alpha}^{(u\phi_0)}(x) = -2C_{3\alpha}^{(u\Phi^+)}(x)$$

$$C_{\alpha\beta}^{(u\Theta^-)}(x) = -\frac{L^2 h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left(\zeta d_{(p)1} + \psi d_{(l)1} + \zeta\psi d_{(p+l)1}\right)B_{1H}(x)$$

$$C_{\alpha\beta}^{(u\Theta^+)}(x) = -2\frac{L^2 h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left(d_{(0)2} + \zeta d_{(p)2} + \psi d_{(l)2} + \zeta\psi d_{(p+l)2}\right)B_{1H}(x)$$

$$C_{\alpha\beta}^{(u\vartheta_0)}(x) = -\frac{L^2 h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left(d_{(0)0} + \zeta d_{(p)0} + \psi d_{(l)0} + \zeta\psi d_{(p+l)0}\right)B_{1H}(x)$$

$$\begin{aligned}M_{11}^{(w)*}(x) = {}& C_{11}^{(wu)}(x)u_{1,1}^*(x) + C_{12}^{(wu)}(x)u_{2,2}^*(x) + C_{11}^{(ww)}(x)w_{,11}^*(x) + C_{12}^{(ww)}(x)w_{,22}^*(x) \\ & + C_{31}^{(w\Phi^-)}(x)\Phi^{-*}(x) + C_{31}^{(w\Phi^+)}(x)\Phi^{+*}(x) + C_{31}^{(w\phi_0)}(x)\phi_0^*(x) + C_{11}^{(w\Theta^-)}(x)\Theta^{-*}(x) \\ & + C_{11}^{(w\Theta^+)}(x)\left(\Theta^{+*}(x) - 2\right) + \left(C_{11}^{(w\vartheta_0)}(x) - 2C_{11}^{(w\Theta^+)}(x)\right)\vartheta_0^*(x)\end{aligned}$$

$$\begin{aligned}M_{22}^{(w)*}(x) = {}& C_{12}^{(wu)}(x)u_{1,1}^*(x) + C_{11}^{(wu)}(x)u_{2,2}^*(x) + C_{12}^{(ww)}(x)w_{,11}^*(x) + C_{11}^{(ww)}(x)w_{,22}^*(x) \\ & + C_{31}^{(w\Phi^-)}(x)\Phi^{-*}(x) + C_{31}^{(w\Phi^+)}(x)\Phi^{+*}(x) + C_{31}^{(w\phi_0)}(x)\phi_0^*(x) + C_{11}^{(w\Theta^-)}(x)\Theta^{-*}(x) \\ & + C_{11}^{(w\Theta^+)}(x)\left(\Theta^{+*}(x) - 2\right) + \left(C_{11}^{(w\vartheta_0)}(x) - 2C_{11}^{(w\Theta^+)}(x)\right)\vartheta_0^*(x)\end{aligned}$$

$$M_{12}^{(w)*}(x,t) = C_{66}^{(wu)}(x)\left(u_{1,2}^*(x,t) + u_{2,1}^*(x,t)\right) + 2C_{66}^{(ww)}(x)w_{,12}^*(x,t)$$

$$\text{(A.5)}$$

$$C_{ab}^{(wu)}(x) = \frac{L(h_0)^2}{D_0}\zeta\left[(1-c_1)d_{(p)1} + \frac{4}{3}c_1 c_2 d_{(p)3}\right]c_{ab(0)}^E c_{(H)}^E(x)h^{*2}(x)$$

$$\begin{aligned}C_{ab}^{(ww)}(x) = {}& \frac{(h_0)^3}{D_0}\left\{(c_1-1)d_{(0)2} - \frac{16}{9}c_1 c_2 d_{(0)6}\right. \\ & \left. + \zeta\left[(c_1-1)d_{(p)2} - \frac{16}{9}c_1 c_2 d_{(p)6}\right]\right\}c_{ab(0)}^E c_{(H)}^E(x)h^{*3}(x)\end{aligned}$$

$$C_{3\alpha}^{(w\Phi^-)}(x) = \frac{L^2\Phi_0}{D_0}\xi\left((1-c_1)d_{(r)1} + \frac{4}{3}c_1 c_2 d_{(r)3}\right)e_{3\alpha(0)}e_{(H)}(x)h^*(x)$$

$$C_{3\alpha}^{(w\Phi^+)}(\mathbf{x}) = \frac{4L^2\Phi_0}{D_0}\left[(1-c_1)d_{(0)2} + c_1c_2\frac{4}{3}d_{(0)4}\right.$$
$$\left. + \xi\left((1-c_1)d_{(r)2} + c_1c_2\frac{4}{3}d_{(r)4}\right)\right]e_{3\alpha(0)}e_{(H)}(\mathbf{x})h^*(\mathbf{x})$$

$$C_{3\alpha}^{(w\phi_0)}(\mathbf{x}) = -2C_{3\alpha}^{(w\Phi^+)}(\mathbf{x})$$

$$C_{\alpha\beta}^{(w\Theta^-)}(\mathbf{x}) = -\frac{L^2h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left\{(1-c_1)d_{(0)2} + \frac{4}{3}c_1c_2d_{(0)4} + \zeta\left[(1-c_1)d_{(p)2} + \frac{4}{3}c_1c_2d_{(p)4}\right]\right.$$
$$\left. + \psi\left[(1-c_1)d_{(l)2} + \frac{4}{3}c_1c_2d_{(l)4}\right] + \zeta\psi\left[(1-c_1)d_{(p+l)2} + \frac{4}{3}c_1c_2d_{(p+l)4}\right]\right\}B_{2H}(\mathbf{x})$$

$$C_{\alpha\beta}^{(w\Theta^+)}(\mathbf{x}) = -2\frac{L^2h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left\{\zeta\left[(1-c_1)d_{(p)3} + \frac{4}{3}c_1c_2d_{(p)5}\right]\right.$$
$$\left. \psi\left[(1-c_1)d_{(l)3} + \frac{4}{3}c_1c_2d_{(l)5}\right] + \zeta\psi\left[(1-c_1)d_{(p+l)3} + \frac{4}{3}c_1c_2d_{(p+l)5}\right]\right\}B_{2H}(\mathbf{x})$$

$$C_{\alpha\beta}^{(w\vartheta_0)}(\mathbf{x}) = -\frac{L^2h_0\theta_0}{D_0}\gamma_{\alpha\beta(0)}\left\{\zeta\left[(1-c_1)d_{(p)1} + \frac{4}{3}c_1c_2d_{(p)3}\right]\right.$$
$$\left. \psi\left[(1-c_1)d_{(l)1} + \frac{4}{3}c_1c_2d_{(l)3}\right] + \zeta\psi\left[(1-c_1)d_{(p+l)1} + \frac{4}{3}c_1c_2d_{(p+l)3}\right]\right\}B_{2H}(\mathbf{x})$$

in which $B_{jH}(\mathbf{x}) := (h^*(\mathbf{x}))^j c_H(\mathbf{x})\alpha_H(\mathbf{x})$.

$$B_1^*(\mathbf{x}) := C_{11}^{(D\Phi^-)}(\mathbf{x})\Phi_{,1}^{-*}(\mathbf{x}) + C_{11}^{(D\Phi^+)}(\mathbf{x})\Phi_{,1}^{+*}(\mathbf{x}) + C_{11}^{(D\phi_0)}(\mathbf{x})\phi_{0,1}^*(\mathbf{x})$$
$$+ C_1^{(D\Theta^-)}(\mathbf{x})\Theta^{-*}(\mathbf{x}) + C_1^{(D\Theta^+)}(\mathbf{x})\left(\Theta^{+*}(\mathbf{x}) - 2\right) + C_1^{(D\vartheta_0)}(\mathbf{x})\vartheta_0^*(\mathbf{x})$$

$$B_2^*(\mathbf{x}) := C_{11}^{(D\Phi^-)}(\mathbf{x})\Phi_{,2}^{-*}(\mathbf{x}) + C_{11}^{(D\Phi^+)}(\mathbf{x})\Phi_{,2}^{+*}(\mathbf{x}) + C_{11}^{(D\phi_0)}(\mathbf{x})\phi_{0,2}^*(\mathbf{x})$$
$$+ C_1^{(D\Theta^-)}(\mathbf{x})\Theta^{-*}(\mathbf{x}) + C_1^{(D\Theta^+)}(\mathbf{x})\left(\Theta^{+*}(\mathbf{x}) - 2\right) + C_1^{(D\vartheta_0)}(\mathbf{x})\vartheta_0^*(\mathbf{x}) \tag{A.7}$$

$$C_{\alpha\beta}^{(D\Phi^-)}(\mathbf{x}) = -\frac{1}{\chi_{11(0)}^\varepsilon}\lambda\left(d_{(s)1} - 4d_{(s)3}\right)\chi_{\alpha\beta(0)}^\varepsilon\chi_{(H)}^\varepsilon(\mathbf{x})h^*(\mathbf{x})$$

$$C_{\alpha\beta}^{(D\Phi^+)}(\mathbf{x})$$
$$= -\frac{2}{\chi_{11(0)}^\varepsilon}\left[d_{(0)2} - 4d_{(0)4} + \lambda\left(d_{(s)2} - 4d_{(0s)4}\right)\right]\chi_{\alpha\beta(0)}^\varepsilon\chi_{(H)}^\varepsilon(\mathbf{x})h^*(\mathbf{x})$$

$$C_{\alpha\beta}^{(D\phi_0)}(\mathbf{x})$$
$$= -\frac{1}{\chi_{11(0)}^\varepsilon}\left[d_{(0)0} - 8d_{(0)2} + 16d_{(0)4} + \lambda\left(d_{(s)0} - 8d_{(s)2} + 16d_{(s)4}\right)\right]\chi_{\alpha\beta(0)}^\varepsilon\chi_{(H)}^\varepsilon(\mathbf{x})h^*(\mathbf{x})$$

$$C_\alpha^{(D\Theta^-)}(\mathbf{x}) = \frac{L\theta_0}{\Phi_0\chi_{11(0)}^\varepsilon}\upsilon\left(d_{(q)1} - 4d_{(q)3}\right)p_{\alpha(0)}p_H(\mathbf{x})h^*(\mathbf{x})$$

$$C_\alpha^{(D\Theta^+)}(\mathbf{x}) = \frac{2L\theta_0}{\Phi_0\chi_{11(0)}^\varepsilon}\big[d_{(0)2} - 4d_{(0)4} + \upsilon\big(d_{(q)2} - 4d_{(q)4}\big)\big]p_{\alpha(0)}p_H(\mathbf{x})h^*(\mathbf{x})$$

$$C_\alpha^{(D\vartheta_0)}(\mathbf{x}) = \frac{L\theta_0}{\Phi_0\chi_{11(0)}^\varepsilon}\big\{d_{(0)0} - 8d_{(0)2} + 16d_{(0)4} + \upsilon\big(d_{(q)0} - 8d_{(q)2} + 16d_{(q)4}\big)\big\}p_{\alpha(0)}p_H(\mathbf{x})h^*(\mathbf{x})$$

$$\begin{aligned}
B_3^*(\mathbf{x}) :=\ & C_{31}^{(Du)}(\mathbf{x})u_{\alpha,\alpha}^*(\mathbf{x}) + C_{31}^{(Dw)}(\mathbf{x})\nabla^2 w^*(\mathbf{x}) + C_{33}^{(D\Phi^-)}(\mathbf{x})\Phi^{-*}(\mathbf{x}) \\
& + C_{33}^{(D\Phi^+)}(\mathbf{x})\Phi^{+*}(\mathbf{x}) + C_{33}^{(D\phi_0)}(\mathbf{x})\phi_0^*(\mathbf{x}) + C_3^{(D\Theta^-)}(\mathbf{x})\Theta^{-*}(\mathbf{x}) \\
& + C_3^{(D\Theta^+)}(\mathbf{x})\big(\Theta^{+*}(\mathbf{x}) - 2\big) + C_3^{(D\vartheta_0)}(\mathbf{x})\vartheta_0^*(\mathbf{x})
\end{aligned}$$

$$\text{(A.8)}$$

$$C_{3\alpha}^{(Du)}(\mathbf{x}) = -\frac{L}{\Phi_0\chi_{11(0)}^\varepsilon}8\xi d_{(r)1}e_{3\alpha(0)}e_{(H)}(\mathbf{x})$$

$$\begin{aligned}
C_{3\alpha}^{(Dw)}(\mathbf{x}) = -\frac{h_0}{\Phi_0\chi_{11(0)}^\varepsilon}8\bigg[&(c_1 - 1)d_{(0)2} - c_1 c_2\frac{4}{3}d_{(0)4} \\
& + \xi\bigg((c_1 - 1)d_{(r)2} - c_1 c_2\frac{4}{3}d_{(r)4}\bigg)\bigg]e_{3\alpha(0)}e_{(H)}(\mathbf{x})h^*(\mathbf{x})
\end{aligned}$$

$$C_{33}^{(D\Phi^-)}(\mathbf{x}) = \left(\frac{L}{h_0}\right)^2\frac{8}{\chi_{11(0)}^\varepsilon h^*(\mathbf{x})}\lambda d_{(s)1}\chi_{33(0)}^\varepsilon\chi_{(H)}^\varepsilon(\mathbf{x})$$

$$C_{33}^{(D\Phi^+)}(\mathbf{x}) = \left(\frac{L}{h_0}\right)^2\frac{32}{\chi_{11(0)}^\varepsilon h^*(\mathbf{x})}\big(d_{(0)2} + \lambda d_{(s)2}\big)\chi_{33(0)}^\varepsilon\chi_{(H)}^\varepsilon(\mathbf{x}),$$

$$C_{33}^{(D\phi_0)}(\mathbf{x}) = -2C_{33}^{(D\Phi^+)}(\mathbf{x})$$

$$C_3^{(D\Theta^-)}(\mathbf{x}) = -\frac{8L^2\theta_0}{\Phi_0 h_0\chi_{11(0)}^\varepsilon}\big(d_{(0)2} + \upsilon d_{(q)2}\big)p_{3(0)}p_H(\mathbf{x})$$

$$C_3^{(D\Theta^+)}(\mathbf{x}) = -\frac{16L^2\theta_0}{\Phi_0 h_0\chi_{11(0)}^\varepsilon}\upsilon d_{(q)3}p_{3(0)}p_H(\mathbf{x}),$$

$$C_3^{(D\vartheta_0)}(\mathbf{x}) = -\frac{8L^2\theta_0}{\Phi_0 h_0\chi_{11(0)}^\varepsilon}\upsilon\big(d_{(q)1} - 4d_{(q)3}\big)p_{3(0)}p_H(\mathbf{x})$$

References

1. Rao, S.S., Sunar, M.: Piezoelectricity and its use in disturbance sensing and control of flex-ible structures: a survey. Appl. Mech. Rev. **47**, 113–123 (1994)
2. Wang, X., Zhou, W., Wu, Z., Zhang, X.: Integrated design of laminated composite structures with piezocomposite actuators for active shape control. Comput. Str. **215**, 166–177 (2019)
3. Zhang, S.J., Yu, F.P.: Piezoelectric materials for high temperature sensors. J. Am. Ceram. Soc. **94**, 3153–3170 (2011)
4. Chen, X.L, Zhao, Z.Y., Liew, K.M: Stability of piezoelectric FGM rectangular plates subjected to non-uniformly distributed load, heat and voltage. Advances in Eng. Software **39**, 121–131 (2008)
5. Komijani, M., Gracie, R.: Nonlienar thermo-electro-mechanical behaviour of FGPM beams. Comput. Str. **150**, 208–218 (2016)
6. Adineh, M., Kadkhodayan, M.: Three-dimensional thermo-elastic analysis and dynamic response of a multi-directional functionally graded skew plate on elastic foundation. Compos. B Eng. **125**, 227–240 (2017)
7. Yang, J., Shen, H.S.: Vibration characteristics and transient response of shear deformable functionally graded plates in thermal environments. J. Sound Vib. **255**, 579–602 (2002)
8. Lancaster, P., Salkauskas, K.: Surfaces generated by moving least square methods. Math Comput. **37**, 141–158 (1981)
9. Sladek, V., Sladek, J., Zhang, Ch.: Computation of stresses in non-homogeneous elastic solids by local integral equation method: a comparative study. Comput. Mech. **41**, 827–845 (2008)
10. Sladek, V., Sladek, J.: Local integral equations implemented by MLS approximation and analytical integrations. Eng. Anal. Bound Elem. **34**, 904–913 (2010)
11. Sladek, V., Sladek, J., Sator, L.: Physical decomposition of thin plate bending problems and their solution by mesh-free methods. Eng. Anal. Boundary Elem. **37**, 348–365 (2013)
12. Yoshihiro Ootao, Y., Ishihara, M.: Exact solution of transient thermal stress problem of a multilayered magneto-electro-thermoelastic hollow sphere. Appl. Math. Model. **36**, 1431–1443 (2012)

Accounting for Size Dependence on the meso- or on the Micro-scale in Polycrystalline Plasticity. A Comparative Study for Different Grain Size Distributions

Mathieu Calvat, Baptiste Flipon, Clément Keller, and Fabrice Barbe

Abstract This paper presents a full-field analysis of size dependent crystal plasticity according to two robust and efficient ways to account for microstructural internal length effects in the constitutive laws. Both consist in introducing a Hall–Petch term in the definition of the resolved shear stress. They differ by the nature of the internal length and thus, intrinsically, on the scale at which this size dependency is introduced. In the meso-scale approach, it is directly the size of the considered grain. In the micro-scale approach, it is the distance to the grain boundary. Analyses show that meso- and micro-scale differences increase as the grain size distribution evolves from a unimodal to a bimodal one.

1 Introduction

The Hall–Petch effect is the well-known phenomenon according to which the yield stress of a metal alloy increases proportionally to the inverse of the square root of the Grain Size (GS). This relation, although empirical, is respected in most of the observations for grain sizes ranging above 1 µm on a large variety of metals. It has motivated the development of nanostructured metals for enhancing their yield stress and by extension their strength. The counterpart of such a grain downsizing is a dramatic reduction of the elongation to fracture, which limits the domains of applications of such materials. A simple concept to enhance the ductility consists in introducing Coarse Grains (CG) in a matrix of Fine Grains (FG), which results in

M. Calvat · C. Keller · F. Barbe (✉)
Normandie Univ, INSA Rouen, UNIROUEN, CNRS, GPM, 76000 Rouen, France
e-mail: fabrice.barbe@insa-rouen.fr

B. Flipon
Mines PSL, CNRS, CEMEF, UMR 7635, 06904 Sophia Antipolis, France

polycrystalline materials with a bimodal GS distribution. Experimental analyses conducted on 316L stainless steel have shown that, provided small enough FG and large enough contrast in the characteristic sizes of FG and CG, CG could act as stress relaxers and enabled to extend the elongation to fracture whilst not affecting significantly the yield stress [1].

Full-field crystal plasticity finite element modelling has been concurrently developed to predict the effective macroscale plasticity of bimodal polycrystal, to guide their experimental development, and to explain the specific interactions taking place in these bimodal polycrystals [2]. In the simulations, a pragmatic approach of the size-dependent crystal plasticity modelling has purposely been adopted, in order to enable the exploration of a large range of microstructural metrics (viz. FG size, CG size and CG volume fraction). This approach consists in introducing a microstructural internal length in the constitutive behavior. This paper presents a study on the effect of the choice of this internal length. It is particularly shown that significant effects can take place on the local scale of grains and that this effect is more pronounced in the case of a bimodal GS distribution.

2 Full-Field Modelling

2.1 Microstructure Generation

The polycrystalline microstructures are represented by Laguerre-Voronoi mosaics generated with the Neper software [3, 4]. A comprehensive review of Voronoi-based tessellations for the micromechanics of complex polycrystalline microstructures can be found in [5]. Whereas the unimodal distributions are simply obtained from a Poisson-Voronoi tessellation, leading to a log-normal GS distribution, a Laguerre tessellation with a control on the definition of the nuclei is used for the generation of polycrystals with a bimodal GS distribution (Fig. 1).

2.2 Crystal Plasticity

The crystal plasticity model corresponds to the model of Méric-Cailletaud [6] in which a Hall–Petch type GS dependence has been introduced in the definition of hardening of the slip systems. As defined in Eq. 1, slip γ takes place on a given slip system when the resolved shear stress τ reaches the critical shear stress r which itself hardens according to Eq. 2. In these equations, (K, n, Q, b, R_0) are material parameters originally presented in the Méric-Cailletaud model. h is the matrix of interactions between slip systems, whose values are taken from [7]. Size dependence is introduced on the yield shear stress R_0 by a Hall–Petch relation to the internal length d, as defined in Eq. 3.

$$\dot{\gamma} = \left(\frac{|\tau| - r}{K}\right)^n sign(\tau) \qquad (1)$$

$$r = R_0 + Q \sum h(1 - \exp(-b|\dot{\gamma}|)) \qquad (2)$$

$$R_0 = r_0 + \frac{k}{\sqrt{d}} \qquad (3)$$

In the approach of the *meso-scale* internal length, which has been widely used in size dependent crystal plasticity, the internal length d is the size of the grain to which the considered finite element belongs (Fig. 2a). It thus corresponds to a constitutive parameter which is uniform inside a grain but which varies from grain to grain. In the extended approach of the *micro-scale* internal length, which has been the subject of few studies [8], it is the distance of the considered element to the grain boundary (Fig. 2b). It then corresponds to a local constitutive parameter which varies inside each grain. All the simulations with this constitutive modelling have been performed with the Zset software [9].

2.3 Parameters Identification

For the identification of the material parameters, tensile tests have been performed on isotropic polycrystals containing 343 grains and corresponding to different mean GSs. The resulting stress–strain curves have been fitted to experimental ones corresponding to the same GSs. The obtained set of parameters is presented in Table 1. SEGS and SEGB refer to the cases where the internal length d (or the Size Effect) is respectively the Grain Size (SEGS) or the distance to the Grain Boundary (SEGB).
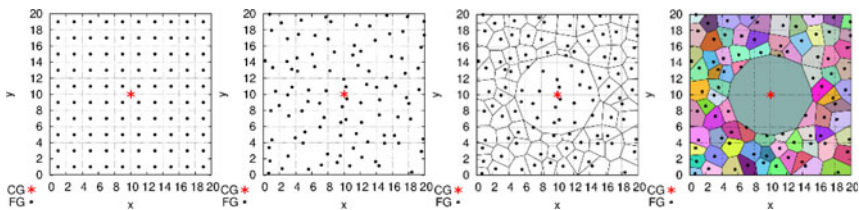

Fig. 1 Generation of a bimodal microstructure by Laguerre tessellation. From left to right: initial positioning of the nuclei, randomization of nuclei positions, repulsion of nuclei about the nucleus of the central coarse grain, result of the tessellation
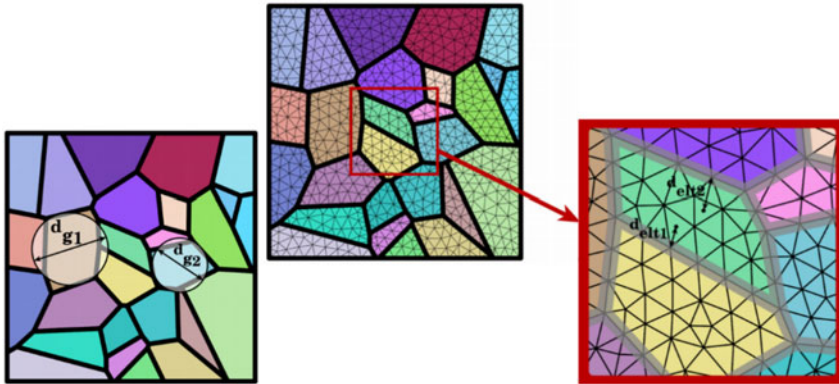
Fig. 2 Definitions of the microstructural internal lengths: (left) grain size, (right) shortest distance of an element to the grain boundary

Fig. 3 Representation of a 2D extruded polycrystal and the imposed boundary conditions for a simple tensile test to 5% in strain. Upper and lower faces are imposed vertical displacements (along z). The back face is imposed a null displacement along x and the front face is constrained to have constant uniform along x
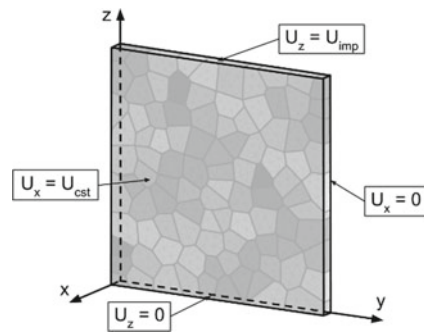

Table 1 The identified parameters for the crystal plasticity model

E (GPa)	υ	Q (MPa)	b	K (MPa.s$^{1/n}$)	n	SEGS r_0 (MPa)	SEGS k (MPa)	SEGB r_0 (MPa)	SEGB k (MPa)
200	0.3	10	9.1	9	24	34.3	200	34.3	70

3 Analyses

All the simulations correspond to simple tensile tests with imposed displacements on polycrystals represented in a 2D geometry extruded in the third direction. This choice of reducing the geometry to a 2D-extruded microstructure is motivated by the important gain in computational resources that this configuration provides as compared to the simulation of a 3D polycrystal in a cubic domain. The boundary conditions on the lateral faces of the microstructure enable to obtain the same

effective (macroscopic) properties as with a 3D Representative Volume element [2] (Fig. 3).

The analyses deal with unimodal and bimodal polycrystals for two families of GS: 0.5 µm for FG and 5 µm for CG. The considered microstructures have been shown in [2] to be representative volume elements concerning the macroscopic properties. For the three considered polycrystals, a same tensile test is performed with the internal length d being the GS (SEGS) and the distance to the GB (SEGB). Comparisons are made on different scales: macroscopic one, scale of the grains (mean behavior of a grain), and local scale (integration points).

3.1 Macro-scale

Figure 4 presents the macroscopic tensile responses of the three considered polycrystals according to each considered internal length. Whatever the GS, there is no significant effect of the choice of the internal length in the unimodal and bimodal cases. In the unimodal case, this observation validates the possibility to predict macroscopic size dependent properties with the internal length scale being indifferently the GS or the distance to the GB. Besides, it can also be observed that the bimodal polycrystal behavior lies in between those of the unimodal FG and unimodal CG. It was shown in [2] that a correct first approximation of the bimodal polycrystal's elastoplastic tensile properties could be obtained from a linear mixture (extension of the Voigt bound) of the properties of the unimodal polycrystals with respectively FGs and CGs of the same characteristic grain sizes. Here, in the SEGS case, given the volume fraction of CG in the bimodal polycrystal (17%), the linear mixture rule leads to 780 MPa at 5% strain which is very close to the simulated stress.

Fig. 4 Macroscopic stress–strain curves of three polycrystals (Unimodal FG, Unimodal CG, bimodal) with the two means to introduce Size Effect (SE): GS or distance to the GB
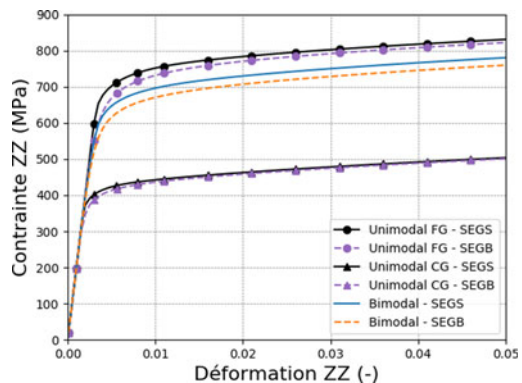
3.2 Micro-scale

Unimodal Case

Figure 5 presents the axial strain and stress fields on the unimodal FG polycrystal. Similar trends are obtained with the unimodal CG polycrystal. One can note that localization occurs on the intra-granular scale of the polycrystal. It is characterized by bands at 45° to the tensile direction concerning strain and by grain-to-grain heterogeneity concerning stress. This is typical of the localization patterns in polycrystals, as has been characterized statistically in [10]. Yet, the model SEGB presents intensified localization patterns, which conforms to the relation established between the distance to the grain boundary and the constitutive behavior: grain cores along the path of a strain localization band are more deformed whereas grain boundary feature more intense stresses.

Bimodal Case

Figure 6 presents the axial strain and stress fields on the bimodal polycrystal with isolated CG. Observations are similar to those made from the unimodal case: the choice of the internal length does not affect significantly the local fields from the general qualitative point of view. Once again the characteristic localization patterns do not differ much from the SEGS to the SEGB case. This confirms that the
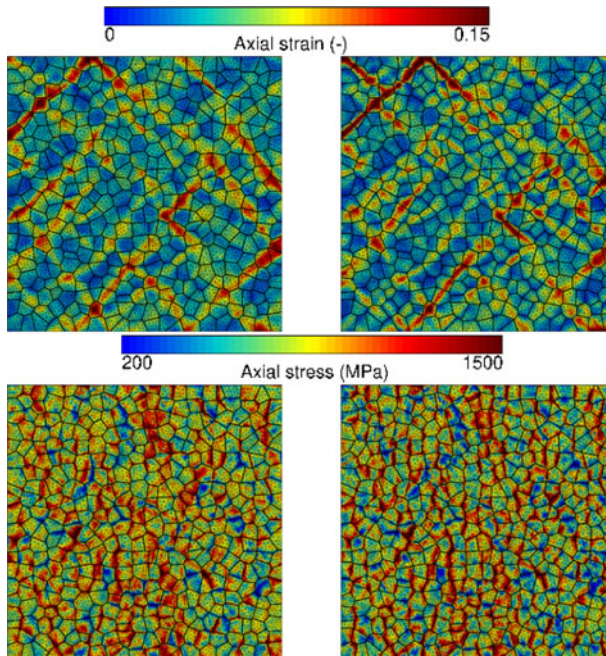

Fig. 5 Axial strain (top raw) and stress (bottom raw) fields for the unimodal FG polycrystal. Left: SEGS. Right: SEGB

formation of intergranular heterogeneities is prominently controlled by the spatial arrangement of grains with their specific crystallographic orientations.

3.3 meso-scale

We now consider the mean response per grain for each polycrystal case: unimodal FG, unimodal CG, bimodal with isolated CG and bimodal with clustered CG. For a given microstructure and a given grain, the mean strain (resp. stress) of the considered grain obtained from the SEGS model is divided by the mean grain strain (resp. stress) as given by the SEGB model. The corresponding SEGS/SEGB strain ratio is plotted against the SEGS/SEGB stress ratio.

The effect of the internal length appears to be more pronounced in the bimodal than in the unimodal polycrystals: whereas the dispersion of the points over the central reference point (1,1) is of 15% and 5% in strain and stress respectively for unimodal polycrystals, whatever the mean GS, it reaches 30% and 10% in the bimodal cases (Fig. 7).
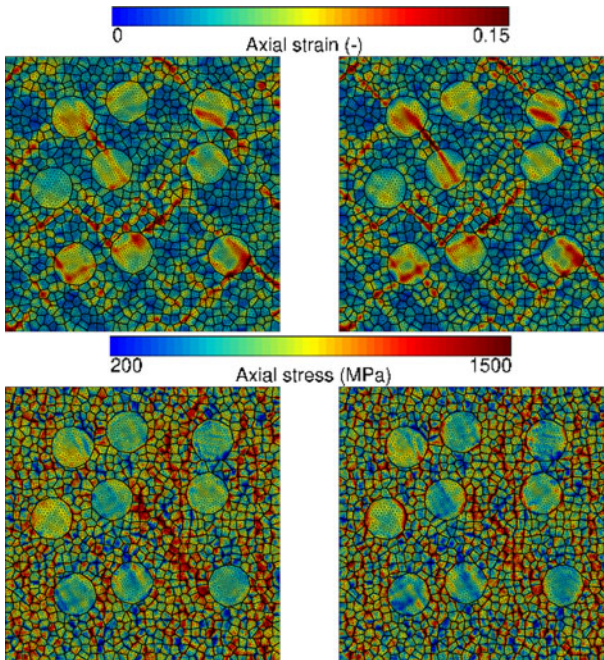

Fig. 6 Axial strain (top raw) and stress (bottom raw) fields for the bimodal polycrystal with isolated CG. Left: SEGS. Right: SEGB
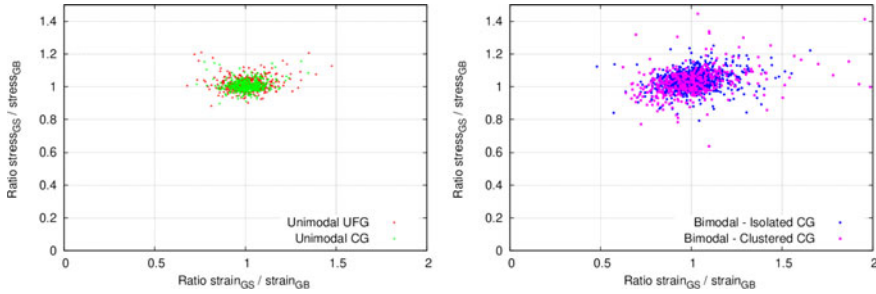
Fig. 7 Synthetic representation of the effect of internal length on the mean behavior of grains: the strain provided from the SEGS model is divided by that of the SEGB model, idem for stress, and the obtained ratios are represented in a stress–strain diagram. (Left) unimodal polycrystals, (Right) bimodal polycrystals

4   Conclusion

A methodology to pragmatically introduce microstructural internal length effect on the elastoplastic behavior of polycrystals has been presented. It is based on the use of either the Grain Size (GS) or the distance to the grain boundary to determine the resolved shear stress on a slip system. Whereas similar methodologies have been presented in the literature, none addressed the question of the effects related to the choice of the internal length scale. In accordance with the construction of the size-dependent crystal plasticity model and with the identification process, there is no effect of the choice of internal length on the macro-scale of the polycrystals, whatever the GS distribution (unimodal or bimodal). But as mean-grain-scale responses are considered, discrepancies of the order of 10% appear in the unimodal polycrystal, whatever the mean GS (fine or coarse grains), and are increased by a factor 2 in the bimodal polycrystal. Micro-scale contour fields provide an explanation for this effect: whereas there is indeed no significant effect of the chosen internal length on the meso-scale stress and strain localization patterns, significant differences can appear on the local intragranular scale and these differences are enhanced if a contrast in properties is introduced in the polycrystal by combining fine and coarse grains.

References

1. Flipon, B., Keller, C., Garcia de la Cruz, L., Hug, E., Barbe, F.: Tensile properties of spark plasma sintered AISI 316L stainless steel with unimodal and bimodal grain size distributions. Mater. Sci. Eng. A **729**, 248–256 (2018)
2. Flipon, B., Keller, C., Quey, R., Barbe, F.: A full-field crystal-plasticity analysis of bimodal polycrystals. Int. J. Solids Struct. **184**, 178–192 (2020)

3. Quey, R., Dawson, P.R., Barbe, F.: Large-scale 3D random polycrystals for the finite element method: generation, meshing and remeshing. Comput. Meth. Appl. Mech. Engng **200**(17–20), 1729–1745 (2011)

4. Neper: Polycrystal generation and meshing. https://neper.sourceforge.net (2020)

5. Flipon, B., Milhem, L., Keller, C., Quey, R., Barbe, F.: A Modelling of polycrystals using well controlled Voronoi-type tessellations and its applications to micromechanical analyses. In: Willot, F., Forest, S. (eds.) Physics and Mechanics of Random Media: From Morphology to Material Properties, chapter 23, pp. 187–198. Presse des Mines, Paris (2018)

6. Méric, L., Poubanne, P., Cailletaud, G.: Single crystal modeling for structural calculations: part 1- model presentation. J. Engng Mater. Technol. **113**(1), 162–170 (1991)

7. Méric, L., Cailletaud, G., Gaspérini. M.F.E.: calculations of copper bicrystal specimens submitted to tension-compression tests. Acta Metall. Mater. **42**(3), 921–935 (1994)

8. Lefebvre, S.: Étude expérimentale et simulation numérique du comportement mécanique de structures sub-micrométriques de cuivre: application aux interconnexions dans les circuits intégrés. Ph.D. thesis, École Centrale Paris (2006)

9. Z-set: Non-linear material & structure analysis suite. https://www.zset-software.com (2020)

10. Barbe, F., Quey, R., Musienko, A., Cailletaud, G.: Three-dimensional characterization of strain localization bands in high resolution elastoplastic polycrystals. Mech. Res. Commun. **36**, 762–768 (2009)

Study on Dynamic Wavy Distribution of Sand Bed by a Transient Two-Layer Modeling

Na Zhu⬤, **Wenjun Huang**⬤, **and Deli Gao**

Abstract This paper investigates a wavy distribution of the cuttings bed in extended-reach drilling. First, a transient two-layer model is established, considering the buoyancy in the deviated section. And a diffusion equation is integrated to obtain entrainment and deposition rates. Next, a Stability Enhancing Two-Step (SETS) method is applied to solve the strong nonlinear equations. Finally, the wavy distribution of cuttings bed, effects of the buoyancy, *ROP*, and the ratio of drilling to flushing time (*R*), circulation stoppage, etc. are studied. The results show that, in the iterations of drilling and washing, the cuttings bed is no longer of equal height but distributed in waves. The bed height decreases with *R*. For the indicator *Risk_10%* (the ratio of the intervals higher than 10% wellbore diameter), the lower limits of *R* under different flow rates are around 0.4–0.6, and the recommended upper limit is 1.4–2.5. And for *Risk_5%*, the lower limits of *R* are around 1.5–3, and the upper limits are 3.5–7. Besides, when considering the buoyancy, the calculated bed height is lower than that without considering it for the deviated section. Thus, the transit section (inclination of 30–60°) is not necessarily the most challenging area to clean if adjusting the drilling fluid's density and buoyancy properly. Furthermore, the cuttings bed slides downwards after pump-off, amassing on the lower side of the buildup Section (70–90°), which poses a considerable risk of pipe sticking. The results provide an essential guide for the optimization design and safety control of extended-reach drilling.

Keywords Cutting transport · Two-layer model · Pipe sticking

N. Zhu · W. Huang · D. Gao (✉)
MOE Key Laboratory of Petroleum Engineering, China University of Petroleum,
Beijing 102249, China
e-mail: gaodeli@cup.edu.cn

W. Huang
e-mail: huangwenjun1986@126.com

349

1 Introduction

Among the stuck pipe issues, 1/3 is due to inefficient hole cleaning [1], which will lead to operational difficulties such as bridging, pack-off, hole fill, excessive torque and drag, and more severely, stuck pipe, lost circulation, and even loss of well control [2]. The research on the hole cleaning has been more than 60 years [3], which can be devided into two groups: experimental and theoretical works.

One of the methods is via experiments, which studies several variables to show the influences on hole cleaning [4–8]. Another approach is through mechanistic and semi-mechanistic models, which seems to be appealing over the experimental method in many ways, such as being applicable for a broader range of operational conditions [9].

Mechanistic and semi-mechanistic hole-cleaning models mainly follow two routes from microscopy and macroscopy to modeling the cuttings transport. The former focuses on the force analysis of particles, such as fluid-hydrodynamic force, buoyancy, and adhesion forces [10]. And the latter follows a layer-modeling approach, including two-layer or three-layer modeling [11–14]. Regardless of variations in those different models, most of them are based on mass balance equations for solids and liquid together with momentum balance equations. And the momentum balance equations involve the pressure and shear force on the interface between solids and liquid phases and annuli surface.

However, despite the continuous development of layer-modeling for so many years, the buoyance for the cutting bed is overlooked, which is essential in the inclined zone and should be considered in the mechanistic model. Failing to consider the buoyance force for the layer-modeling in the inclined wellbore will lead to overestimating cuttings bed height, and unnecessary treatment, such as higher flow rate, drilling fluid density, and lower ROP.

On the other hand, the lay-modeling is a one-dimensional case along the axial direction (along with the wellbore), and two or three-dimensional modeling will take much more calculation time. To simulate the cuttings transport of radial direction (perpendicular to the wellbore), the additional closure relationship for the mass exchange due to deposition and entrainment of cuttings between the adjacent layers is developed, which can be regarded as a simplified supplement for the particles' movement in the radial direction.

Naganawa and Nomura [15] simulated the cuttings transport over the whole trajectory by two-layer model and defined the mass transfer rates between two adjacent layers as:

$$\Gamma_{dep,k} = C_{k1}\rho_k S_I v_{dep}(C_C/C_B)$$
$$\Gamma_{ent,k} = C_{k1}\rho_k S_I v_{ent}$$
$$\tag{1}$$

where, v_{dep} is the cuttings deposition rate, which is the function of the single-particle terminal velocity; v_{ent} is the entrainment rate; k = c, f for solid and

liquid phases, separately; S_I is the wetted perimeter of interface, m;C_{k1} is the concentration, dimensionless.

Zhang et al. [16] also considered the mass transfer in the three-layer model, and the mass exchange is calculated as:

$$\Delta\Gamma = \Gamma_{ent} - \Gamma_{dep} = (v_{ent} - v_{dep})S_I \tag{2}$$

The diffusion equation also illustrates the diffusion and deposition phenomenon [12], and has been employed to the steady two-layer model [12, 17] to obtain the mass exchange between two layers. While, it has not been applied to transient layer-model.

Hence, a fully transient cuttings-transport two-layer model will be developed with improvements: considering the buoyance force and integrating the diffusion equation into the model to calculate mass exchange between two layers. And the simulation will be conducted for the whole well trajectory to predict the cuttings-bed buildup or removal that evolves over time.

2 Transient Two-Layer Mode

2.1 Basic Assumption

A schematic of the transient two-layer model is shown in Fig. 1. This model has three parts: the fluid and solid phases in the suspension layer and the solid bed layer.
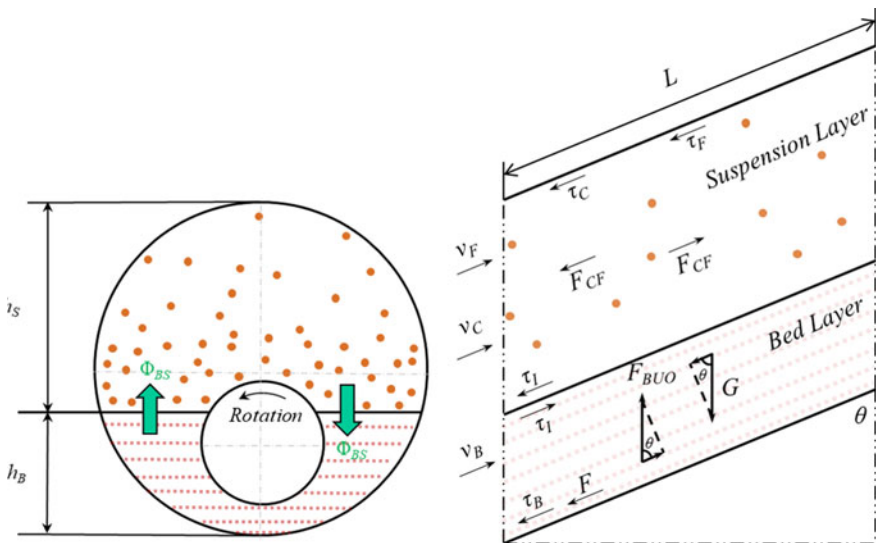
The basic assumptions are as follows [18]:

Fig. 1 Cuttings bed height of drilling [6]

(1) The bed layer represents the cuttings bed, in which a fixed solids concentration is assumed;
(2) The suspension layer contains the carrier fluid and sand particles;
(3) The solid-liquid system is incompressible and its rheological parameters are the same of the fluid;
(4) The height of the interface between the two layers is constant in the annular section studied and, consequently, a hydrostatic distribution of pressure in a cross-section is assumed;
(5) The solids are characterized by average diameter and by their sphericity.

2.2 A Transient Two-Layer Model

The mass conservation of the fluid and solid phase in the suspension layer can be represented by:

$$\frac{\partial(\rho_F C_F A_S)}{\partial t} + \frac{\partial(\rho_F C_F A_S v_F)}{\partial x} = \Phi_{SF} \tag{3}$$

$$\frac{\partial(\rho_C C_C A_S)}{\partial t} + \frac{\partial(\rho_C C_C A_S v_C)}{\partial x} = \Phi_{SC} \tag{4}$$

And, the mass conservation of the bed layer is:

$$\frac{\partial(\rho_B A_B)}{\partial t} + \frac{\partial(\rho_B A_B v_B)}{\partial x} = -\Phi_{BS} \tag{5}$$

The momentum conservation for the two layers can be represented by Eqs. (6–8), separately. And the momentum conservation for the upper fluid and solid phases can be written as:

$$\frac{\partial v_F}{\partial t} + v_F \frac{\partial v_F}{\partial x} = -\frac{1}{\rho_F} \frac{\partial p}{\partial x} - \zeta_F g \cos\theta - \frac{\tau_F S_S}{\rho_F A_S} - \frac{\tau_I S_I}{\rho_F A_S} - \frac{F_{CF}}{\rho_F C_F A_S} + \frac{v_F \Phi_{SF}}{\rho_F C_F A_S} \tag{6}$$

$$\frac{\partial v_C}{\partial t} + v_C \frac{\partial v_C}{\partial x} = -\frac{1}{\rho_C} \frac{\partial p}{\partial x} - \zeta_C g \cos\theta - \frac{\tau_C S_S}{\rho_C A_S} - \frac{\tau_I S_I}{\rho_C A_S} + \frac{F_{CF}}{\rho_C C_C A_S} + \frac{v_C \Phi_{SC}}{\rho_C C_C A_S} \tag{7}$$

And the momentum conservation for solid bed layer is:

$$\frac{\partial v_B}{\partial t} + v_B \frac{\partial v_B}{\partial x} = -\frac{1}{\rho_B} \frac{\partial p}{\partial x} - \zeta_B g \cos\theta - \frac{\tau_B S_B}{\rho_B A_B} + \frac{\tau_I S_I}{\rho_B A_B} - \frac{F}{\rho_B A_B} - \frac{v_B \Phi_{BS}}{\rho_B A_B} \tag{8}$$

where, A is the cross-section area, m^2; C is the concentration, dimensionless; v is velocity, g/cm^3; ρ is density, g/cm^3; p is the pressure, Pa; τ is shear stress, Pa; F is dry friction force, N; F_{CF} is Multiparticle drag forces between cuttings and drilling fluid, N; θ is inclination angle, °; Φ_{SC}, Φ_{SF} is deposition mass flux of cuttings and drilling fluid per unit length, kg/s/m; Φ_{BS} is the total entrainment mass flux of cuttings and drilling fluid per unit length, kg/s/m. The subscript S represents the upper layer, B represents the solid bed layer, I represents the interface between the bed layer and the upper layer, F represents the fluid phase in the upper layer, C represents the solid phase in the upper layer. And ζ is the modified factor considering the effect of buoyancy, which will be [6]:

$$\zeta_B = 1 - \frac{\rho_L}{\rho_B} C_B \tag{9}$$

$$\zeta_C = 1 - \frac{\rho_L}{\rho_C} \tag{10}$$

$$\zeta_F = 1 + \frac{\rho_L}{\rho_F} C_B \tag{11}$$

where, ζ_B is the modified factor of the bed layer, ζ_C is the modified factor of the solid phase of the suspension layer, ζ_F is he modified factor of the liquid phase of the suspension layer.

2.3 A Mass Exchange Between Two Layers

The closure relationships will contain the cuttings deposition and entrainment between two layers, contributing to the mass exchange of both solid and liquid phase in the radial direction (y-axis) [15, 16]. The deposition and entrainment phenomenon can be represented by the well-known diffusion equation [17]:

$$\varepsilon_p \frac{\partial^2 C(y)}{\partial y^2} + v_{dep} \frac{\partial C(y)}{\partial y} = 0 \tag{12}$$

Where, $C(y)$ is the local volumetric concentration in the upper layer; v_{dep} is the particles' terminal setting velocity, ε_p is the local diffusion coefficient.

Then by integration in the radial direction, the increase of solids concentration in the suspension layer per unit time can be expressed by [6]:

$$\Delta C_T = (C_T^{n+1} - C_T^n)/dt \tag{13}$$

Where, C_T is the solids volumetric concentration in the upper layer; if ΔC_T is positive, it means the entrainment rate is dominant, and the solids concentration increases, i.e., upward movement, and vice versa.

Thus, the mass exchange between the two layers will be:

$$\Phi_{SC} = \Delta C_T \rho_C A_S \tag{14}$$

$$\Phi_{SF} = \Delta C_T \rho_F A_S (1 - C_B)/C_B \tag{15}$$

$$\Phi_{BS} = \Delta C_T \rho_B A_S/C_B \tag{16}$$

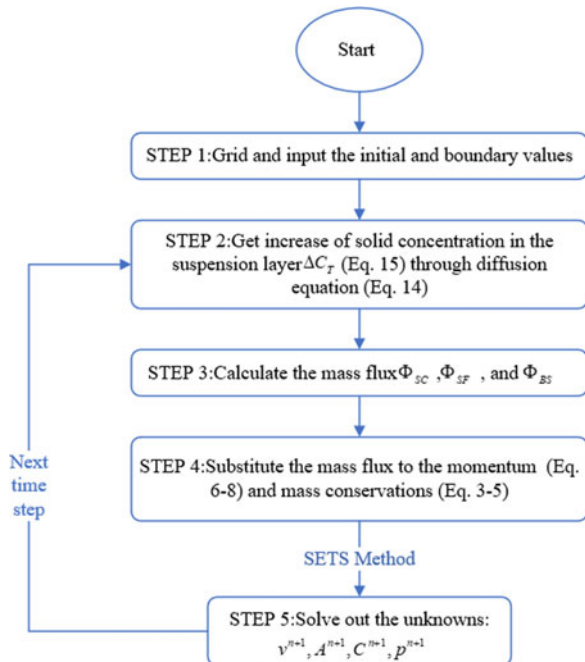### 2.3.1 New Indicators of Hole Cleaning for Whole Well Trajectory

Generally, the hole condition is deemed to be safe for drilling if cuttings bed height less than 10% of d_o. This is an indicator of the stable and uniform state. Since this two-layer model can stimulate the cuttings distribution along the whole trajectory dynamically, two new indicators, i.e. *Risk_10%* and *Risk_5%*, are defined as the length ratios of the interval which bed height is more than 10% and 5% of d_o to the whole well length:

$$Risk\_10\% = L\_10\%/L \times 100\% \tag{17}$$

$$Risk\_5\% = L\_5\%/L \times 100\% \tag{18}$$

where, *L_10%* and *L_5%* are the total lengths of the well interval of which the bed height is more than 10% and 5% of d_o, m; L is the whole well length, m.

Fig. 2 Workflow of the calculating process

STEP 1:Grid and input the initial and boundary values

STEP 2:Get increase of solid concentration in the suspension layerΔC_T (Eq. 15) through diffusion equation (Eq. 14)

STEP 3:Calculate the mass fluxΦ_{SC},Φ_{SF}, and Φ_{BS}

Next time step

STEP 4:Substitute the mass flux to the momentum (Eq. 6-8) and mass conservations (Eq. 3-5)

SETS Method

STEP 5:Solve out the unknowns:
$v^{n+1}, A^{n+1}, C^{n+1}, p^{n+1}$

If the new indicator *Risk* _10% > 0, there are intervals of which the bed height is more than 10%, and conversely, the cuttings bed is less than 10% for the whole well. Therefore, these transient indicators can predict the hole cleaning effect more precisely.

A detailed workflow of the calculating process is presented in Fig. 2. The prediction and modified step of equations are obtained via a Stability Enhancing Two-Step (SETS) method in Guo et al.'s study [7].

3   Results and Discussion

A transient cuttings-transport model makes it possible to better predict the downhole conditions because it can represent phenomena that evolve over time, such as cuttings-bed buildup or removal [19]. To clearly understand the cuttings bed distribution along the entire wellbore, an ERW well in the South China Sea is studied, which is the length of 5200 m, and the parameters are listed in Table 1 and 2. The refined mesh was implemented in the inclined section as the hole angle varies rapidly.

Table 1 Data table of well structure

Casing program	Measured depth/m	TVD/m	Inclination angle/°
Vertical section	427	426.9	0.14
Intermediate section	1026	767	0.14–83.6
Horizontal section	5200	1168	84

Table 2 List of input data for modeling

Variables	Symbols	Values	Units
Liquidity index	n	0.68	Dimensionless
Consistency coefficient	K	0.5	$Pa \cdot s^n$
Drilling fluid density	ρ_F	1180	kg/m^3
Cuttings density	ρ_C	2510	kg/m^3
Drill string rotation speed	N	60	r/min
Rate of penetration	ROP	30	m/h
Hole angle	θ	0–90	°
Wellbore diameter	d_o	0.31	m
Pipe outer diameter	d_i	0.14	m
Cuttings diameter	d_p	0.005	m

The typically standard procedure is to execute additional mud circulations for approximately 1 h each time during pipe connections after drilling a three-joint stand of drill pipe [20]. Accordingly, the stimulated operation is set as an alternate combination of drilling and well flushing process, for instance, drilling for 1 h first, performing well flushing for 1 h, and then continuing the above iteration process.

3.1   Basic Calculation

The basic calculation illustrates the wavy distribution of cuttings transport. Based on the simulated results, Figs. 3 and 4 depict the cuttings bed transport's evolution under the first cycle of drilling and well flushing process. It can be seen from Fig. 3 that a 7% d_o is formed after drilling for 1 h, with *ROP* of 30 m/h, which is matched with the calculation of Martin [18]. Then, in the well flushing process (Fig. 4), the cuttings bed moves upwards from the bottom of the hole, and one dune has been formed. At around 4000 s (1.1 h), the cuttings dune has been transported to the wellhead, and thus most of the cuttings are cleaned out.

Following the above exiting condition, the second iteration process of drilling and well flushing is conducted, and two giant dunes are developed (Figs. 5 and 6 (a)). Each dune is approximately 2000 m wide, and the maximum height is 6% of d_o. The dune's height and length are related to the *ROP*, and the transport distance depends on the flow rate and well flushing time, which will be discussed later.
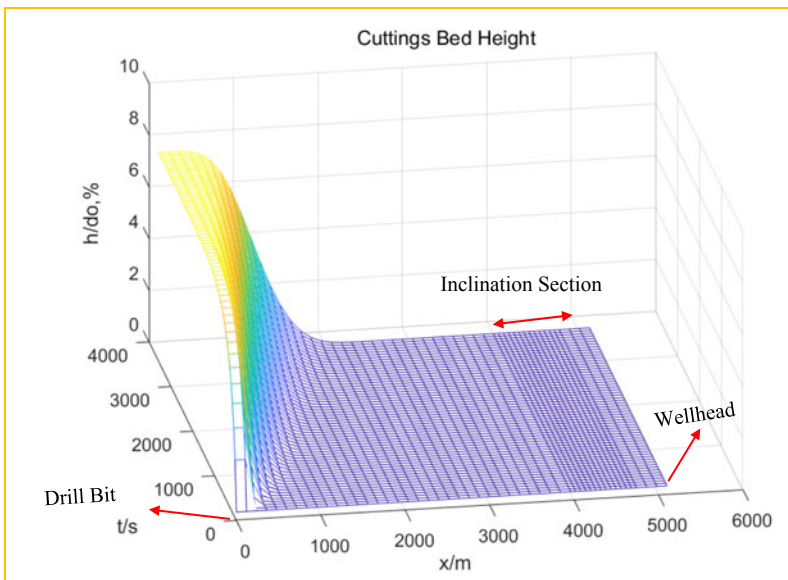

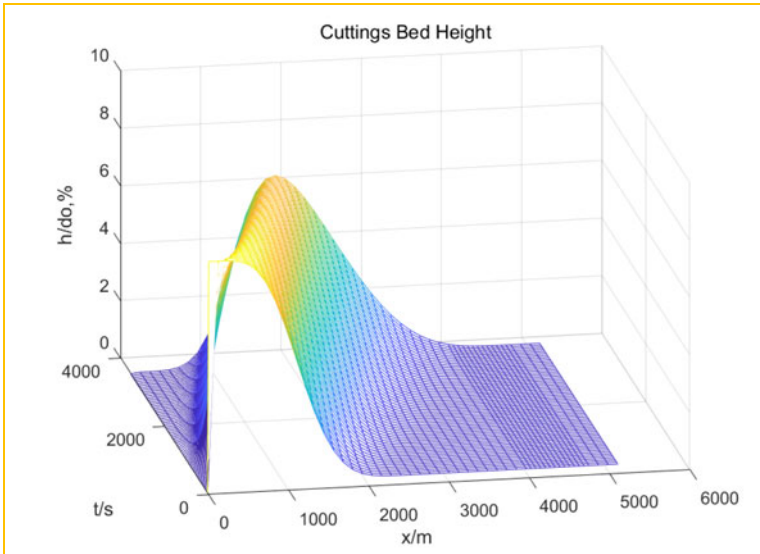**Fig. 3** Cuttings bed height of drilling

Fig. 4 Cuttings bed height of well flushing for 1 h after drilling, Q = 50 L/s
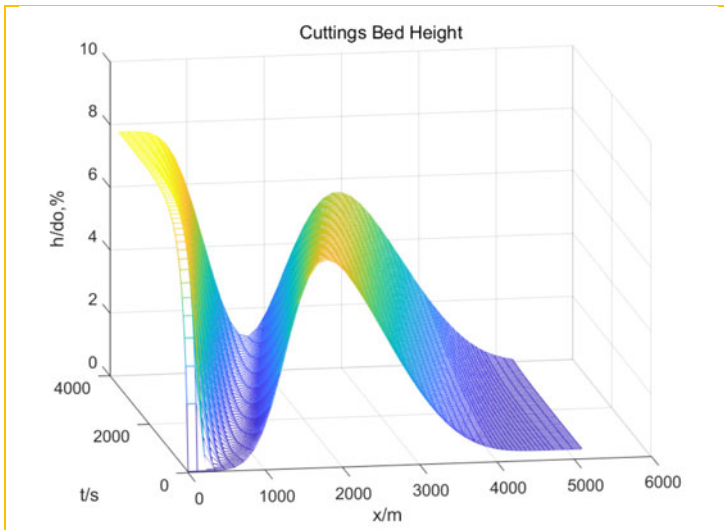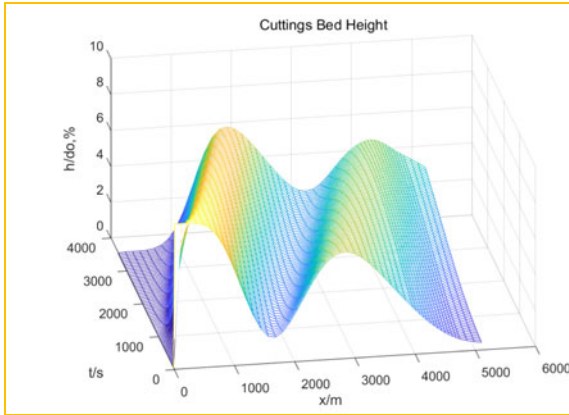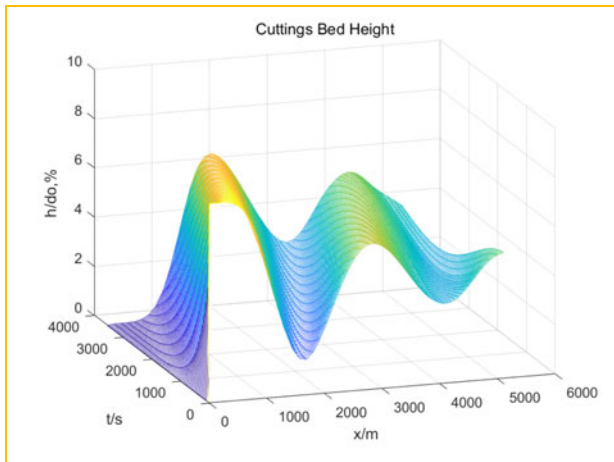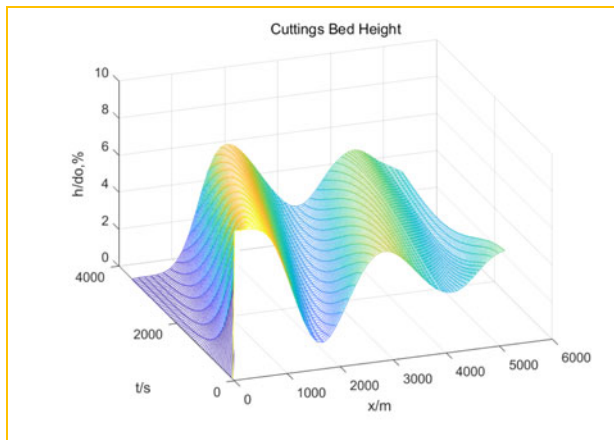

Fig. 5 Cuttings bed height of continue drilling for 1 h based on Fig. 4, *ROP* = 30 m/h

(a). Cuttings bed height of continuing well flushing for 1h based on Fig 5, *ROP*=30 m/h.

(b). Cuttings bed height of the third cycle.

(c). Cuttings bed height of the fourth cycle.

Fig. 6 Cuttings bed height of the second, third and fourth cycles
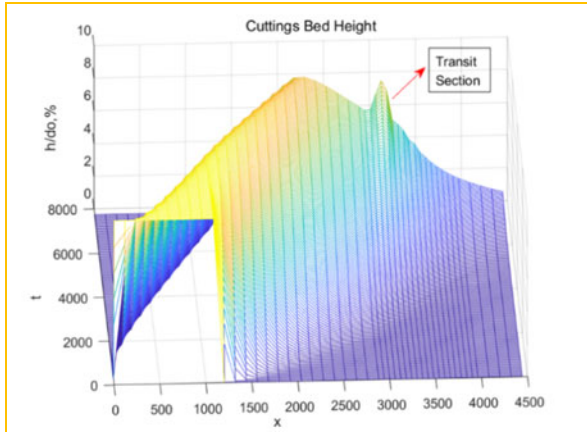
Fig. 7 Cuttings bed height without considering the effect of buoyance
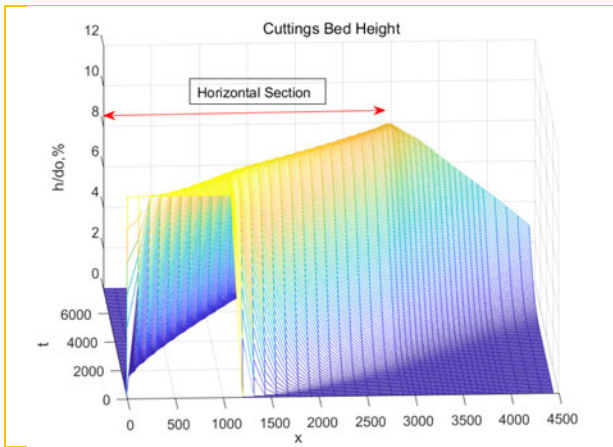
Fig. 8 Cuttings bed height considering the effect of buoyance

Similarly, with the drilling and flushing continues, more dunes will be formed, which indicates the wavy distribution of cuttings consists of several dunes, associated with cuttings bed buildup and removal, instead of a constant height in the stable model. Thus, the transient cuttings-transport model can better predict the downhole conditions that evolve over time.

However, as can be seen in Figs. 6(b) and (c), the number of the fourth cycle's dunes is the same as the third cycle, which suggests the dune's number is stable and does not increase indefinitely. This is because with the iteration process continues, the dunes formed in the early stage have been already transported out the wellbore and cleaned, the dunes number become dynamically stable.

3.2   Sensitivity Analysis

The effect of buoyance force, *ROP*, time ratio of washing to drilling ($R = t_{wash}/t_{drill}$) and the influence of pumping stops will be investigated in this paper.

Impact of Buoyance. As mentioned in Sect. 2, the buoyance force plays a vital role in its lifting ability to transport the cuttings upward along the wellbore. Figures 7 and 8 presents the cuttings bed height with and without considering the effect of buoyance. At the inclination of around 65°, the cuttings bed height reaches the maximum (Fig. 7) at 8% of d_o in the inclined section. However, when considering the buoyance forces, it is lower than the horizontal section (Fig. 8). This phenomenon is matched with both experimental results and CFD simulations [22–24]. Thus, we can infer that the transit section is not necessarily the most difficult area to clean if adjusting the drilling fluid's density and buoyance force.

Impact of ROP. Figure 9 shows that when the time interval of drilling and well flushing is constant, with the increase of *ROP*, the cuttings bed height is higher, which is easy to understand because higher *ROP* produces more cuttings.

Impact of the Ratio of Drilling to Well Flushing Time. Figure 10 vividly presents the cuttings bed heights with different values of washing to drilling time (*R*). And four groups of iteration progress (conducting drilling and well flushing for once, twice, third and fourth, separately) are studied, of which the flow rates are 30, 40, 50, and 60 L/s.

As shown in Fig. 10(a), with R increases, the *Risk*_10% decrease rapidly first and then declines more gradually. And the cuttings' bed heights show a great difference between four groups (Fig. 10(a)) when the ratio is less than 0.6, indicating the sand cuttings are not cleaned enough in the previous cycle, and seriously affect the next one and make more accumulation. When the ratio R is more than 0.6, the value of *Risk*_10% is the same for the twice, third, and fourth iteration cycle, which means the cuttings bed can maintain a relatively stable height, regardless of the operation time or the cycle number of the drilling and well flushing processes. Moreover, the *Risk*_10% is zero when the ratio is 2.5 under the flow rate of 30 L/s, which means the cuttings bed height is less than 10%, and the hole condition is safe for the whole trajectory. So, the recommended range of R is from 0.6 to 2.5 under this condition, because if R less than 0.6, the hole condition is not clean enough, and when the ratio is less than 2.5, the cuttings bed height can meet the requirement of being less than 10% for the entire wellbore, and additional circulations can be saved economically.

Figure 10(c), (e), and (g) depict the percent of cuttings bed height over 10% of the do with R at the flow rates of 40, 50, and 60 L/s. A similar trend can be found that the *Risk*_10% decrease rapidly first, and then decline more gradually. The lower limits of R under different flow rates are very close, around 0.4–0.6. While

Fig. 9 Cuttings bed height with different *ROP*s
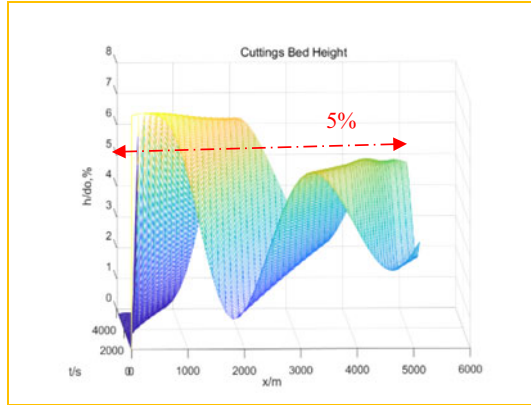
(a). *ROP* is 30 m/h.
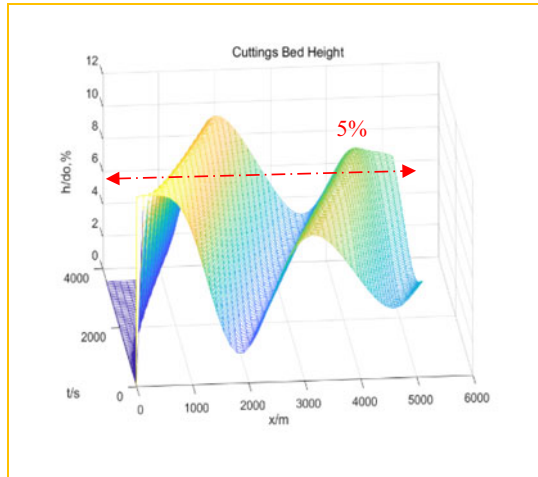
(b). *ROP* is 60 m/h.
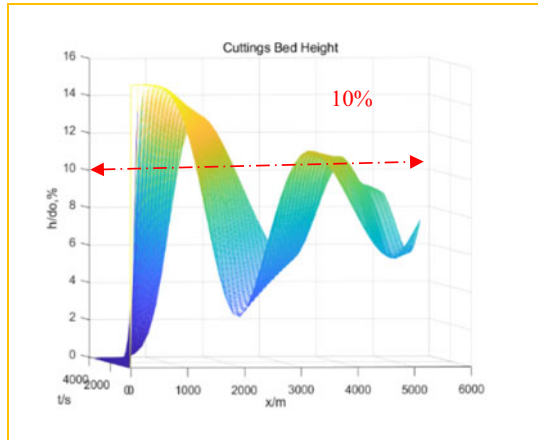
(c). *ROP* is 90 m/h.
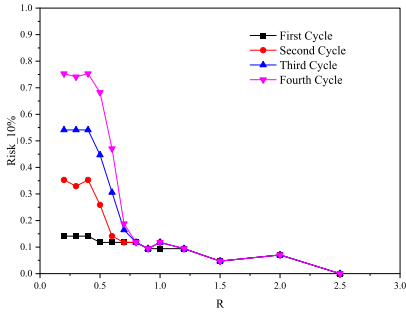
(a). *Risk_10%* with *R*, *Q*=30L/s
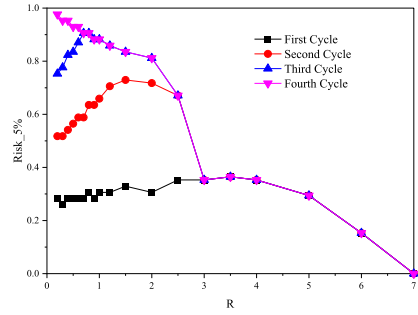
(b). *Risk_5%* with *R*, *Q*=30L/s
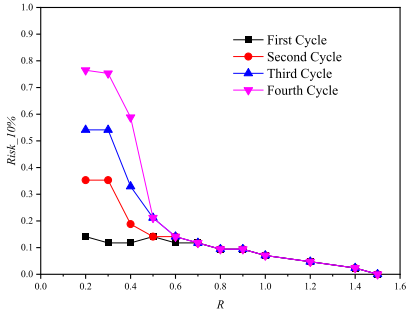
(c). *Risk_10%* with *R*, *Q*=40L/s
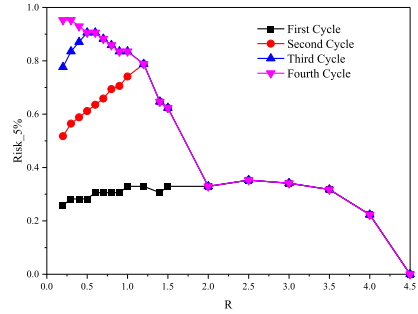
(d). *Risk_5%* with *R*, *Q*=40L/s
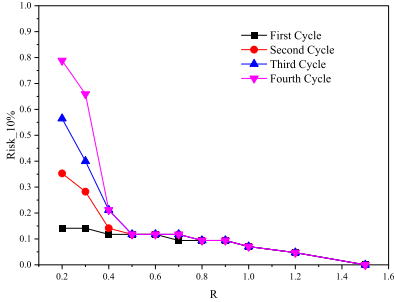
(e). *Risk_10%* with *R*, *Q*=50L/s
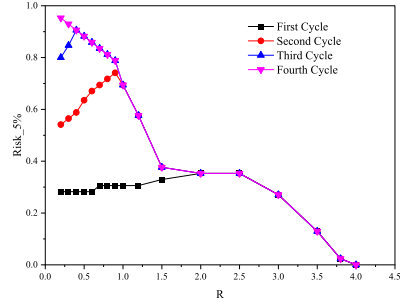
(f). *Risk_5%* with *R*, *Q*=50L/s
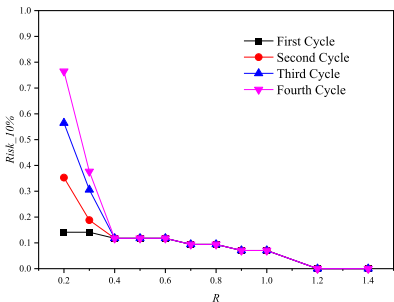
(g). *Risk_10%* with *R*, *Q*=60L/s
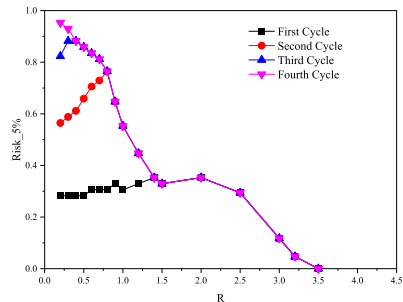
(h). *Risk_5%* with *R*, *Q*=60L/s

Fig. 10 Length percent of cuttings bed height over 10% of d_o (*Risk*_10%) and over 5% (*Risk*_5%) with different ratios of well flushing to drilling time (*R*), *Q* = 30, 40, 50, and 60 L/s

the recommended upper limit is 1.4–2.5, indicating the recommended upper limits are different a lot and decrease with the increase of the flow rate.

Figure 9(b), (d), (f), and (h) present the percent of cuttings bed height over 5% of the do with R at the flow rates of 30, 40, 50, and 60 L/s. A similar trend can be found with $Risk\_10\%$. As can be seen that the lower limits of R under different flow rates are around 1.5–3, and the recommended upper limits are 3.5–7.

Impact of Stopping Pump. Figure 11 shows the cuttings bed sliding downwards after stopping the pump (flow rate is zero), the initial height is 5% d_o along the whole well. At the inclination below 65° (around the angle of repose), the sliding speed is high and become slower to zero. As a result, the sand cuttings transport to the angle of 70°–90°, accumulating in the lower side of the deflecting interval, which poses a huge risk of pipe sticking.

Fig. 11 Cuttings bed height considering stopping pump
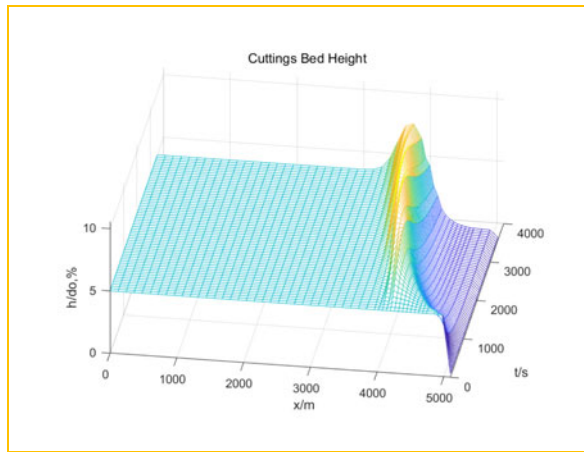
(b). Front view before stopping pump
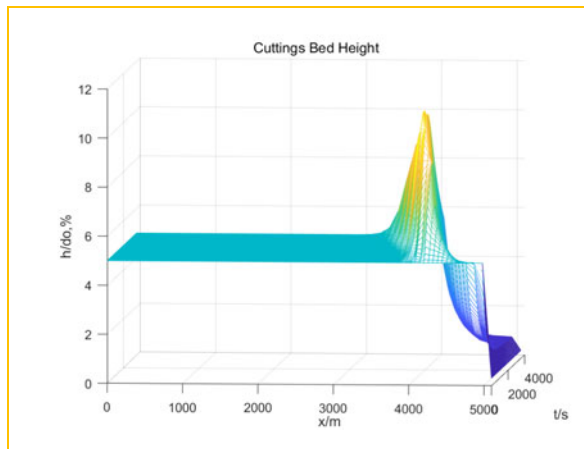
(d). Front view after stopping pump

The reason for this phenomenon is because the shear stress of the drilling fluid becomes smaller when the flow rate is zero, thus driving forces are not big enough to overcome the gravity force of the sand cuttings when the inclination is smaller than the angle of repose, so the cuttings begin to fall downwards. With the increase of hole angle, the gravity force's axial component becomes smaller, and at the angle of repose, the forces will get balanced [25], the cuttings bed velocity becomes slower, until stop slipping, and then accumulating on the lower side.

4   Conclusions

(1) With the iteration of the drilling and well flushing process, the wavy distribution of cuttings bed is formed by several dunes, associated with cuttings bed buildup and removal. Thus, the transient cuttings-transport model can better predict the downhole condition that evolves over time and the tight spots.

(2) The time ratio of washing to drilling (R) is a controllable parameter to adjust the cuttings bed height. With R increases, the $Risk\_10\%$ decrease rapidly first, and then decline more gradually. For the indicator $Risk\_10\%$, the lower limits of R under different flow rates are very close, which are around 0.4–0.6. While the recommended upper limit is 1.4–2.5. And for the indicator $Risk\_5\%$, the lower limits of R are around 1.5–3, and the upper limits are 3.5–7. When R reaches the upper limit, the $Risk\_10\%$ or $Risk\_5\%$ is zero, which means the cuttings bed height is less than 10% or 5%.

(3) The calculated cuttings bed height, considering the buoyance force, is lower than ignoring it for the deviated section. Thus, if adjusting the drilling fluid's density and buoyance force, the transit section is not necessarily the most challenging area to clean anymore.

(4) The cuttings bed will slide downwards after stopping the pump, accumulating in the lower side of the buildup Section (70–90°), which poses a massive risk of pipe sticking.

References

1. Zhang, H., et al.: Study on the influence of the rotation of drill string on cuttings transportation in horizontal well. Sci. Technol. Eng. **16**(2), 125–130 (2016)
2. Bizhani, M., Corredor, F.E.R., Kuru, E.: Quantitative evaluation of critical conditions required for effective hole cleaning in coiled-tubing drilling of horizontal wells. SPE Drill. Complet. **31**(3), 188–199 (2016)

3. Kelessidis, V.C., Bandelis, G.E.: Flow patterns and minimum suspension velocity for efficient cuttings transport in horizontal and deviated wells in coiled-tubing drilling . SPE Drill Complet. **19**, 213–227 (2004)
4. Pandya, S., Ahmed, R., Shah, S.: Experimental study on wellbore cleanout in horizontal wells. J. Petrol. Sci. Eng. **177**, 466–478 (2019)
5. Song, X., et al.: Experimental Study on the Wellbore-Cleaning Efficiency of Microhole-Horizontal-Well Drilling. SPE J. **22**(04), 1189–1200 (2017)
6. Zhu, N., Huang, W., Gao, D.: Dynamic wavy distribution of cuttings bed in extended reach drilling. J. Petrol. Sci. Eng. **198**, 108171 (2021)
7. Guo, X.L., Wang, Z.M., Long, Z.H.: Study on three-layer unsteady model of cuttings transport for extended-reach well. J. Petrol. Sci. Eng. **73**(1–2), 171–180 (2010)
8. Naganawa, S., et al.: Cuttings transport in directional and horizontal wells while aerated mud drilling. In: IADC/SPE Asia Pacific Drilling Technology, Jakarta, Indonesia, 8–11 September 2002 (2002)
9. Bizhani, M., Kuru, E.: Critical review of mechanistic and empirical (semimechanistic) models for particle removal from sandbed deposits in horizontal annuli with water. SPE J. **23**(02), 237–255 (2018)
10. Ramadan, A., Skalle, P., Johansen, S.T.: A mechanistic model to determine the critical flow velocity required to initiate the movement of spherical bed particles in inclined channels. Chem. Eng. Sci. **58**(10), 2153–2163 (2003)
11. Cho, H., Shah, S.N., Osisanya, S.O.: A three-layer modeling for cuttings transport with coiled tubing horizontal drilling. In: SPE Annual Technical Conference and Exhibition, Dallas, Texas, 1–4 October 2000 (2000)
12. Cho, H., Shan, S.N., Osisanya, S.O.: A three-segment hydraulic model for cuttings transport in coiled tubing horizontal and deviated drilling. J. Can. Petrol. Technol. **42**, 32–39 (2002)
13. Li, J., Luft, B.: Overview solids transport study and application in oil-gas industry-theoretical work. In: International Petroleum Technology Conference, Kuala Lumpur, Malaysia, 10–12 December 2014 (2014)
14. Nguyen, D., Rahman, S.S.: A three-layer hydraulic program for effective cuttings transport and hole cleaning in highly deviated and horizontal wells. SPE Drill. Complet. **13**(03), 182–189 (1998)
15. Naganawa, S., Nomura, T.: Simulating transient behavior of cuttings transport over whole trajectory of extended reach well. In: IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition, Bangkok, Thailand, 13–15 November 2006 (2006)
16. Zhang, F., et al.: A unified transient solid-liquid two-phase flow model for cuttings transport-modelling part. J. Petrol. Sci. Eng. **166**, 146–156 (2018)
17. Doron, P., Granica, D., Barnea, D.: Slurry flow in horizontal pipes—experimental and modeling. Int. J. Multiph. Flow **13**(4), 535–547 (1987)
18. Martins, A.L., et al.: Evaluating the transport of solids generated by shale instabilities in ERW drilling. In: SPE International Conference on Horizontal Well Technology, Calgary, Alberta, Canada, 1–4 November 1998 (1998)
19. Cayeux, E., et al.: Real-time evaluation of hole-cleaning conditions with a transient cuttings-transport model. SPE Drill. Complet. **29**(01), 5–21 (2014)
20. Naganawa, S., Kudo, H., Matsubuchi, H.: Simulation study on influences of wellbore tortuosity on hole cleaning in extended-reach drilling. In: Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 7–10 November 2016 (2016)
21. Li, J., Luft, B.: Overview of solids transport studies and applications in oil and gas industry—experimental work. In: SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, 14–16 October 2014 (2014)

22. Tomren, P.H., Iyoho, A.W., Azar, J.J.: Experimental study of cuttings transport in directional wells. SPE Drill. Eng. **1**(01), 43–56 (1986)
23. Pandya, S., Ahmed, R., Shah, S.: Effects of particle density on hole cleanout operation in horizontal and inclined wellbores. In: SPE/ICoTA Well Intervention Conference and Exhibition, The Woodlands, Texas, USA, 26–27 March 2019 (2019)
24. Akhshik, S., Behzad, M., Rajabi, M.: CFD-DEM simulation of the hole cleaning process in a deviated well drilling: the effects of particle shape. Particuology **25**, 72–82 (2016)
25. Ni, X.: Study on transient migration law of cuttings in horizontal well section. Doctor thesis, Northeast Petroleum University (2017)
26. Wang, W.: Research on the cuttings migration law and carrying technology in highly deviated well and horizontal well. Doctor thesis, China University of Petroleum, Beijing (2008)

Element-Free Discretization Method with Moving Finite Element Approximation

Vladimir Sladek and Jan Sladek

Abstract The popularity of standard Finite Element Method (FEM) results from its simplicity and wide applicability in engineering computations. The simplicity consists in polynomial interpolation of field variables within finite elements, while the wide applicability in rather close similarity of engineering problems based mostly on variational formulations. On the other hand, the standard FEM suffers from limited continuity on element intersections, global meshing difficulty in analysis of complex geometry, and global equations may not be parallelized efficiently. All these shortcomings can be removed in the mesh free formulations, where the arbitrary net of nodal points is used for discretization instead of subdivision of the analyzed domain into finite elements. The spatial approximations based on cluster of nodes are not expressed in terms of elementary functions and evaluation of the shape functions prolongs the computation. To get rid of disadvantages of the standard FEM and mesh free formulations, the moving finite element (MFE) method has been proposed. In this paper, we pay attention to explanation of the basis of the MFE approximation and to its combination with the strong as well as local weak formulations of boundary value problems for stationary heat conduction in solids with functionally graded heat conduction coefficient. Finally, in numerical test examples the reliability (accuracy and numerical stability) and computational efficiency are compared for three variants of the MFE method and the standard FEM.

Keywords Element-free method · Polynomial interpolation · Lagrange finite element · Strong/weak formulation · Numerical experiments

V. Sladek (✉) · J. Sladek
Institute of Construction and Architecture, Slovak Academy of Sciences, Bratislava 845 03, Slovakia
e-mail: vladimir.sladek@savba.sk

1 Introduction

In engineering computations, there have been developed huge amount of computational methods, which can be classified according to various criterions. For instance, they can be split into two large groups such as: (i) strong formulations (mathematical model—governing equations and boundary conditions—are required to be satisfied), (ii) weak formulations (mathematical model equations are satisfied in certain integral sense). In general, an approximation of field variables is needed in order to decrease the amount of degrees of freedom (DOF) and to be able to solve the problem numerically. According to the discretization we can distinguish three main groups utilizing: (i) grid of nodes in the analyzed domain; (ii) mesh of finite elements; (iii) randomly distributed nodes (mesh-free approximations). It is not the aim to discuss here details of many computational methods. We concern only some of them, which have a relationship with the technique presented in this paper. The main advantage of weak formulations is reduction of the order of derivatives of field variables occurring in numerical computation and rather high numerical stability. On the other hand, though the strong formulations do not require any integration, they suffer from lower precision of approximation of high order derivatives of field variables as well as from failure of conditioning of discretized governing equations when the density of nodes is increasing extremely.

The most frequently used method in engineering computations, the finite element method (FEM) [1, 2] is based on subdivision of the analyzed domain using a mesh of finite elements (FE). Although the approximations of field variables are considered as local polynomial interpolations within particular FE, the approximation is made to be global by stitching the FE in global weak formulation on the whole analyzed domain. The finite volume method (FVM) [3, 4] belongs to weak formulation methods, but in contrast to FEM the balance equations are considered in a local sense on local subdomains (control volumes) around nodal points. The approximations in FVM can be considered like in finite difference method (FDM) or in FEM. One of the most criticized shortcoming of the standard FEM is limited continuity on element intersections. The C^0 continuity is insufficient for the unique calculation of gradients of field variables on element intersections as well as for numerical solution of problems with governing equations of higher than 2nd order partial differential equations (PDE). Moreover, the analysis of complex geometry is making the global meshing difficult and construction of the global equations may not be parallelized efficiently [5]. Such difficulties have become strong motivation for development of mesh-free method (MFM) during the last three decades [6–11]. The most distinct advantage of MFM over FEM is that no elements are needed for approximations of field variables. In MFM, a cluster of nodes around the computational node are locally selected for computing the shape function in global coordinate space. Thus, the combination of the mesh-free approximation with the strong formulation leads to truly meshless method, while in the case of weak formulation it is possible, if it is considered on local subdomains. Without restricting the applicability as compared with the standard FEM, the MFM

has become rapidly popular. The most critical point of MFM is prolongation of computation because of time consuming evaluation of the shape functions, since these are not expressed in terms of elementary functions. Moreover, the shape function in some mesh-free approximations do not satisfy the Konecker-delta property, which yields the difficulty to impose essential boundary conditions directly at boundary nodes.

It is natural to combine advantages of polynomial interpolations and element-free formulations. Yagawa et al. [5, 11–13] proposed the free mesh method (FMM), where the local subdomain is created from triangular elements which are constructed around a central node using satellite nodes and the local weak formulation is applied like in FVM. Similar idea has been proposed with using quadrilateral finite elements [14–16]. Recently, this method has been elaborated and called the Moving Finite Element (MFE) method [17–19]. To each nodal point, we associate one finite element which is moving over the net of nodal points and constructed according with the position of the nodal point in the analyzed domain. Both the strong and local weak formulations are applicable in combination with the MFE approximations of field variables. The local subdomain is inside the FE associated with the collocation node. Since only one FE is needed for approximation at the collocation node or within the local subdomain around this node, we obtain extremely narrow bandwidth of the system of discretized equations. Gao et al. [20, 21] proposed the Element Differential Method (EDM), where the strong formulation of the governing equations is combined with the approximation at a collocation node using the Lagrange isoparametric FE involving this node. Similar strong-form techniques called Finite Block Method (FBM) [22, 23] and Strong Formulation finite Element Method (SFEM) [24–26] are based on collocation approach in strong sense and mapping technique applied to isoparametric element-like blocks in FBM or subdomains in SFEM. Recently, Gao et al. [27] proposed the Free Element Collocation Method (FECM) as modified version of the EDM with replacing fixed elements by one element associated with the collocation point in the same way as it is done in the MFE approximation.

In this paper, we concern with the development of the MFE approximation, strong and local weak formulations for heat conduction problems in solid media with the functionally graded heat conduction coefficient, and special attention is devoted to study the reliability (accuracy and numerical stability) as well as computational efficiency of three variants of the MFE method in comparison with the standard FEM.

2 Moving Finite Element Method

2.1 Moving Finite Element Approximation

Let us start with approximation of field variables in standard finite element method (FEM). The analyzed domain Ω and its boundary $\partial\Omega$ are covered by finite elements E^s in such a way that $\Omega \cup \partial\Omega = \bigcup_s E^s$ and $E^s \cap E^b = \varnothing$ if $s \neq b$. If the finite element (FE) E^s is formed with n nodal points $\mathbf{x}^{(\gamma,s)}$, where (γ, s) is the global number of the node with local number γ on E^s, the approximation of a field variable is given by polynomial interpolation

$$u(\mathbf{x})|_{E^s} \approx \sum_{\gamma=1}^{n} u^{(\gamma,s)} N^\gamma(\boldsymbol{\xi}), \ u^{(\gamma,s)} = u(\mathbf{x}^{(\gamma,s)})\big|_{E^s} \tag{1}$$

in which $N^\gamma(\boldsymbol{\xi})$ are the shape functions [2] constructed from the Lagrange interpolation polynomials of intrinsic coordinates $\xi_k \in [-1, \ 1]$, $(k = 1, \ \ldots, \ d)$, with d being the dimensionality of the problem. The transformation between the global Cartesian coordinates x_i and intrinsic coordinates on E^s is usually given by using the same interpolation (isoparametric elements) as for field variables, i.e.

$$x_i|_{E^s} \approx \sum_{\gamma=1}^{n} x_i^{(\gamma,s)} N^\gamma(\boldsymbol{\xi}) \tag{2}$$

In case of Moving Finite Element (MFE) approximation, we shall use the same interpolations as in standard FEM, but instead of utilization of net of FE, only the net of nodal points is employed. Our aim is to use the MFE approximation for numerical solution of boundary value problems in both the strong and the weak sense. To achieve the element-free formulation, the weak formulation is considered on local subdomains $\Omega^a \subset \Omega \cup \partial\Omega$ defined arbitrarily around each nodal point $\mathbf{x}^a$ with including only one nodal point in each subdomain. The finite element is constructed individually for each nodal point with taking into account its position in the analyzed domain $\Omega \cup \partial\Omega$. In this paper, we shall consider 2D problems and two kinds of FE, namely the bi-quadratic quadrilateral Lagrange element (with 9 nodes) and the bi-cubic quadrilateral Lagrange element (with 16 nodes). The construction of the bi-quadratic Lagrange elements is illustrated in Fig. 1 according to the following rules: (i) if $\mathbf{x}^a \in \Omega$, $\mathbf{x}^a$ is the center of the FE; (ii) if $\mathbf{x}^b \in \partial\Omega$ and $\partial\Omega$ is smooth at the node $\mathbf{x}^b$, this node is the mid-side node of the FE; (iii) if $\mathbf{x}^b$ lies at a corner on the boundary edge $\partial\Omega$, the FE is constructed uniquely.
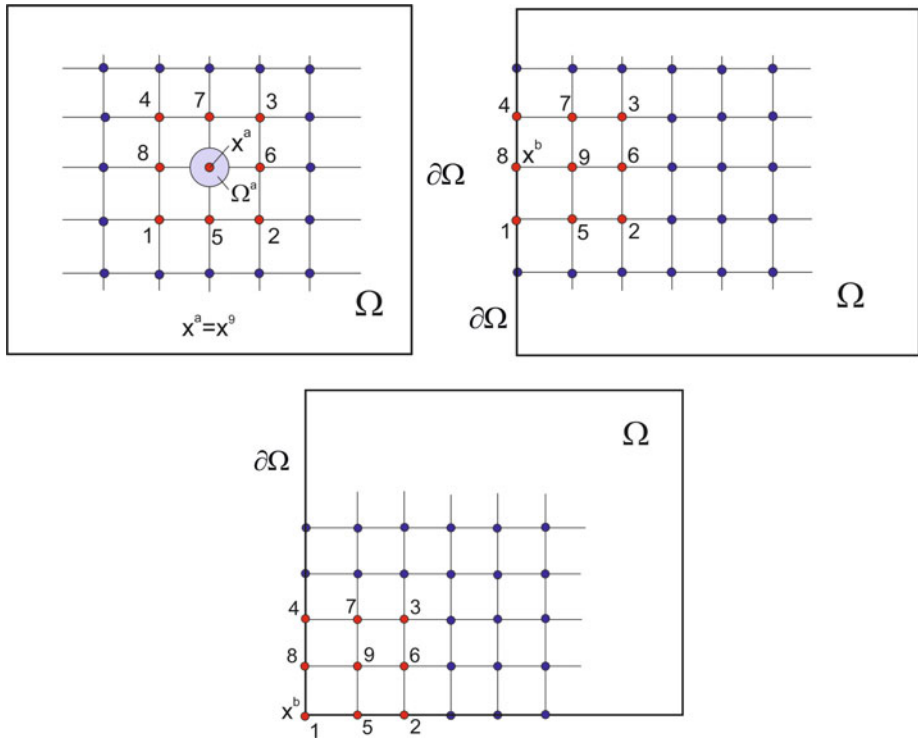
Fig. 1 Illustration of construction of MFE around the interior $\mathbf{x}^a$ and boundary $\mathbf{x}^b$ nodes

A uniform mesh of nodes appears to be the most appropriate for automatic construction of moving FE. After attribution of all needed nodes to the moving finite elements E^a associated with the nodal point $\mathbf{x}^a$, $a = 1, 2, \ldots, N$ (i.e. after creation of moving finite element for each of nodal points), one can generate a non-uniform mesh of nodes by modifying their positions. Since moving FE are associated with nodal points (not with elements), it is appropriate to rewrite the FE interpolations (1) and (2) as

$$u(\mathbf{x})\big|_{E^a} \approx \sum_{\gamma=1}^{n} u^{a_\gamma} N^\gamma(\boldsymbol{\xi}), \quad x_i\big|_{E^a} \approx \sum_{\gamma=1}^{n} x_i^{a_\gamma} N^\gamma(\boldsymbol{\xi}) \tag{3}$$

where a_γ is the global number of the node whose local number on the element E^a is γ. In order to get the approximation of the 2$^{\text{nd}}$ order derivatives of field variables (entering the 2$^{\text{nd}}$ order partial differential equations), we need to use at least the bi-quadratic FE. Bearing in mind the approximations (3), the gradient of the field variable is approximated within the moving FE as [14]

$$u_{,i}(\mathbf{x})\big|_{E^a} = \frac{\partial u(\mathbf{x})}{\partial x_i}\bigg|_{E^a} = \frac{\partial u(\mathbf{x})}{\partial \xi_k}\frac{\partial \xi_k}{\partial x_i}\bigg|_{E^a} \approx \sum_{\gamma=1}^{n} u^{a_\gamma} b_i^{\gamma a}(\xi), \tag{4}$$

$$b_i^{\gamma a}(\xi) = Y_{ik}^a \frac{\partial N^\gamma(\xi)}{\partial \xi_k}, \quad Y_{ik}^a = \frac{\partial \xi_k}{\partial x_i}\bigg|_{E^a}$$

with $[Y^a] = [J^a]^{-1}$ and

$$J_{ki}^a = \frac{\partial x_i}{\partial \xi_k}\bigg|_{E^a} = \sum_{\gamma=1}^{n} x_i^{a_\gamma}\frac{\partial N^\gamma}{\partial \xi_k}, \quad Y_{ij}^a = \frac{\varepsilon_{ik3}\varepsilon_{jl3}}{|J^a|}J_{lk}^a, \quad |J^a| = \varepsilon_{mn3}J_{1m}^a J_{2n}^a \tag{5}$$

where $|J^a|$ is the Jacobian of the transformation (3_2).

Applying the same differentiation procedure to (4), we obtain the approximation for the 2$^{\text{nd}}$ order derivatives of the field variable

$$u_{,ij}(\mathbf{x})\big|_{E^a} = \frac{\partial^2 u(\mathbf{x})}{\partial x_i \partial x_j}\bigg|_{E^a} \approx \sum_{\gamma=1}^{n} u^{a_\gamma}\frac{\partial \xi_m}{\partial x_j}\frac{\partial}{\partial \xi_m}b_i^{\gamma a}(\xi)$$

$$= \sum_{\gamma=1}^{n} u^{a_\gamma}Y_{jm}^a\frac{\partial}{\partial \xi_m}\left(Y_{ik}^a N_{,k}^\gamma(\xi)\right) = \sum_{\gamma=1}^{n} u^{a_\gamma}Y_{jm}^a\left[Y_{ik}^a N_{,km}^\gamma(\xi) + Y_{ik,m}^a N_{,k}^\gamma(\xi)\right]$$

$$= \sum_{\gamma=1}^{n} u^{a_\gamma}Y_{jm}^a\left[Y_{ik}^a N_{,km}^\gamma(\xi) - Y_{is}^a J_{sl,m}^a Y_{lk}^a N_{,k}^\gamma(\xi)\right]$$

$$= \sum_{\gamma=1}^{n} u^{a_\gamma}Y_{jm}^a\left[Y_{ik}^a N_{,km}^\gamma(\xi) - Y_{is}^a J_{sl,m}^a b_l^{\gamma a}(\xi)\right] = \sum_{\gamma=1}^{n} u^{a_\gamma}b_{ij}^{\gamma a}(\xi) \tag{6}$$

where $J_{sl,m}^a = \sum_{\gamma=1}^{n} x_l^{a_\gamma}\frac{\partial^2 N^\gamma}{\partial \xi_s \partial \xi_m} = J_{ml,s}^a$

$$b_{ij}^{\gamma a}(\xi) := Y_{jm}^a\left[Y_{ik}^a N_{,km}^\gamma(\xi) - Y_{is}^a J_{sl,m}^a b_l^{\gamma a}(\xi)\right] = b_{ji}^{\gamma a}(\xi). \tag{7}$$

2.2 Strong Formulation for Stationary Heat Conduction

The stationary heat conduction in 2D domain of isotropic solid is governed by the PDE

$$\left(\lambda u_{,i}\right)_{,i} = 0 \text{ in } \Omega \tag{8}$$

and the boundary conditions

$$u(\mathbf{x}) = T(\mathbf{x}) \text{ on } \partial\Omega_T; \ -\lambda(\mathbf{x})n_i(\mathbf{x})u_{,i}(\mathbf{x}) = f(\mathbf{x}) \text{ on } \partial\Omega_f; \ \partial\Omega = \partial\Omega_T \cup \partial\Omega_f. \tag{9}$$

Making use the approximations for derivatives shown in the previous subsection, we can directly collocate the governing Eq. (7) as

$$\lambda(\mathbf{x}^c) \sum_{\gamma=1}^{n} u^{c_\gamma} b_{ii}^{\gamma c}(\boldsymbol{\xi}^{\alpha_c}) + \lambda_{,i}(\mathbf{x}^c) \sum_{\gamma=1}^{n} u^{c_\gamma} b_i^{\gamma c}(\boldsymbol{\xi}^{\alpha_c}) = 0 \text{ for each } \mathbf{x}^c \in \Omega \tag{10a}$$

and the discretized boundary conditions become

$$u(\mathbf{x}^b) = T(\mathbf{x}^b) \text{ for } \mathbf{x}^b \in \partial\Omega_T, \ \lambda n_i(\mathbf{x}^b) \sum_{\gamma=1}^{n} u^{b_\gamma} b_i^{\gamma b}(\boldsymbol{\xi}^{\alpha_b}) = -f(\mathbf{x}^b) \text{ for } \mathbf{x}^b \in \partial\Omega_f \tag{10b}$$

where $\xi_1^{\alpha_c}$, $\xi_2^{\alpha_c}$ are intrinsic coordinates of the global node $\mathbf{x}^c$ on the associated finite element E^c. Note that the heat flux is not uniquely defined at a corner point and the natural condition requires a special treatment as will be shown later.

2.3 Local Weal Formulation for Stationary Heat Conduction

The weak form of the governing equation is given as

$$\int_{\Omega} \left(\lambda(\mathbf{x})u_{,i}(\mathbf{x})\right)_{,i} \omega(\mathbf{x})d\Omega = 0. \tag{11}$$

Selecting the weight function $\omega(\mathbf{x})$ as compactly supported function

$$\omega^a(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \bar{\Omega}^a \\ 0, & \text{otherwise} \end{cases}$$

in which $\bar{\Omega}^a = \Omega^a \cup \partial\Omega^a$ is a subdomain around a nodal point $\mathbf{x}^a \in \bar{\Omega}^a \subset \bar{\Omega}$ of the analyzed domain $\bar{\Omega}$, the local weak form becomes

$$\int\limits_{\partial\Omega^a} n_i(\mathbf{x})\lambda(\boldsymbol{x})u_{,i}(\mathbf{x})d\Gamma = 0 \tag{12}$$

where $n_i(\mathbf{x})$ is the unit outward normal vector at $\mathbf{x} \in \partial\Omega^a$.

Bearing in mind the approximation, the integration is performed in the intrinsic space. Therefore, it is appropriate to select the subdomain of the circular shape in the intrinsic space, i.e. $\tilde{\Omega}^a = \tilde{C}^a \cap \tilde{E}^a = \{\forall\, \boldsymbol{\xi} \in \tilde{E}^a;\ |\boldsymbol{\xi} - \boldsymbol{\xi}^{\alpha_a}| < \xi^a\}$ with $\tilde{C}^a = \{\forall \boldsymbol{\xi},\ |\boldsymbol{\xi} - \boldsymbol{\xi}^{\alpha_a}| < \xi^a\}$, $\xi^a < 1$ and $\tilde{\Omega}^a \ni \boldsymbol{\xi} \rightarrow \mathbf{x} \in \Omega^a$. If $\boldsymbol{\xi}^{\alpha_a}$ is an interior node of the finite element $\tilde{E}^a$, then $\tilde{\Omega}^a = \tilde{C}^a$ and $\partial\tilde{\Omega}^a$ is the circle $\partial\tilde{C}^a$. If $\boldsymbol{\xi}^{\alpha_a}$ is on the boundary of the finite element $\partial\tilde{E}^a$, then $\tilde{\Omega}^a \subset \tilde{C}^a$ and $\partial\tilde{\Omega}^a = \tilde{\Gamma}^a \cup \tilde{L}^a$, where $\tilde{L}^a := \partial\tilde{C}^a \cap \tilde{E}^a$ is the circular portion of the boundary $\partial\tilde{\Omega}^a$ and $\tilde{\Gamma}^a := \tilde{C}^a \cap \partial\tilde{E}^a$. Anyway

$$\tilde{L}^a = \{\forall\, \boldsymbol{\xi} \in \tilde{E}^a;\ \xi_k = \xi_k^{\alpha_a} + \xi^a(\delta_{k1}\cos\varphi + \delta_{k2}\sin\varphi)\} \text{ and } d\Gamma|_{L^a} = \tau_i dx_i|_{L^a}$$
$$= g|_{L^a}\xi^a d\varphi$$

with

$$g_i^{L^a} = \frac{1}{\xi^a}\frac{\partial x_i}{\partial\xi_j}\frac{\partial\xi_j}{d\varphi}\Big|_{L^a} = \sum_{\gamma=1}^{n} x_i^{a_\gamma}\left(N_{,2}^{\gamma}(\boldsymbol{\xi}^{L^a})\cos\varphi - N_{,1}^{\gamma}(\boldsymbol{\xi}^{L^a})\sin\varphi\right),\ n_i|_{L^a} = \varepsilon_{ik3}g_k^{L^a}/g^{L^a}.$$

The discretized weak form of the governing Eq. (12) at interior nodes $\mathbf{x}^a \in \Omega$ become

$$\varepsilon_{ik3}\xi^a \sum_{\gamma=1}^{n} u^{a_\gamma} \int\limits_0^{2\pi} \lambda(\mathbf{x}^{L^a})b_i^{\gamma a}(\boldsymbol{\xi}^{L^a})g_k^{L^a}d\varphi = 0 \tag{13}$$

with $x_i^{L^a} := x_i|_{L^a} = \sum_{\gamma=1}^{n} x_i^{a_\gamma}N^{\gamma}(\boldsymbol{\xi}^{L^a})$.

If the natural boundary condition is prescribed on both sides of a corner node $\mathbf{x}^c \in \partial\Omega$, this boundary condition can be taken into account in the weak sense. Then, $\partial\Omega^c = L^c \cup \Gamma^c$ (see Fig. 2) and the local weak form becomes

$$0 = \int\limits_{\partial\Omega^c} n_i(\mathbf{x})\lambda(\mathbf{x})u_{,i}(\mathbf{x})d\Gamma = \int\limits_{L^c} n_i(\mathbf{x})\lambda(\mathbf{x})u_{,i}(\mathbf{x})d\Gamma - \int\limits_{\Gamma^c} f(\mathbf{x})d\Gamma \tag{14}$$

Fig. 2 Illustration of boundary contours $\Gamma^c = \Gamma^{c1} \cup \Gamma^{c2}$ and L^c associated with a corner point $\mathbf{x}^c$
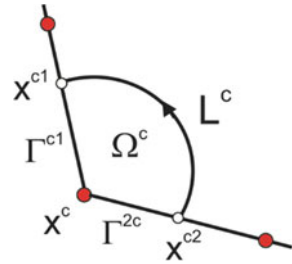
The discretization of Eq. (14) results in

$$\varepsilon_{ik3}\xi^c \sum_{\gamma=1}^{n} u^{c_\gamma} \int_{\varphi^s}^{\varphi^f} \lambda(\mathbf{x}^{L^c}) b_i^{\gamma c}(\boldsymbol{\xi}^{L^c}) g_k^{L^c} \, d\varphi = \int_{0}^{\xi^c} \left(f(\mathbf{x}^{\Gamma^{c2}}) g^{\Gamma^{c2}} + f(\mathbf{x}^{\Gamma^{c1}}) g^{\Gamma^{c1}} \right) ds \quad (15)$$

with $g_i|_{\Gamma^c} = \frac{\partial x_i}{\partial \xi_j} \frac{\partial \xi_j}{\partial s}\Big|_{\Gamma^c} = \frac{\partial x_i}{\partial \xi_j} \tilde{\tau}_j^{c\cdot}\Big|_{\Gamma^c} = \sum_{\gamma=1}^{n} x_i^{c_\gamma} \frac{\partial N^\gamma}{\partial \xi_j} \tilde{\tau}_j^{c\cdot}, \quad x_i^{\Gamma^c} = \sum_{\gamma=1}^{n} x_i^{c_\gamma} N^\gamma(\boldsymbol{\xi}|_{\Gamma^c})$

3 Numerical Examples

In order to test the reliability (accuracy and numerical stability) and computational efficiency of the proposed method based on MFE approximation, we consider simple examples for stationary heat conduction in solids with constant and/or functionally graded heat conduction coefficient when the exact solutions can be used as benchmark. If we consider the square domain $L \times L = 3 \times 3$ with the mixed boundary conditions: $u(x_1, x_2 = 0) = u_0 = 0$, $u(x_1, x_2 = L) = u_L = 100$, $n_i u_{,i}|_{x_1=0} = 0 = n_i u_{,i}|_{x_1=L}$, the following exact solutions can be used as the reference solutions:

(i) *homogeneous body* $\lambda = const$:

$$u(x_1, x_2) = u_0 + (u_L - u_0)x_2/L, \ u_{,i}(x_1, x_2) = \delta_{i2}(u_L - u_0)/L, \ u_{,ij}(x_1, x_2) = 0$$

$$(16)$$
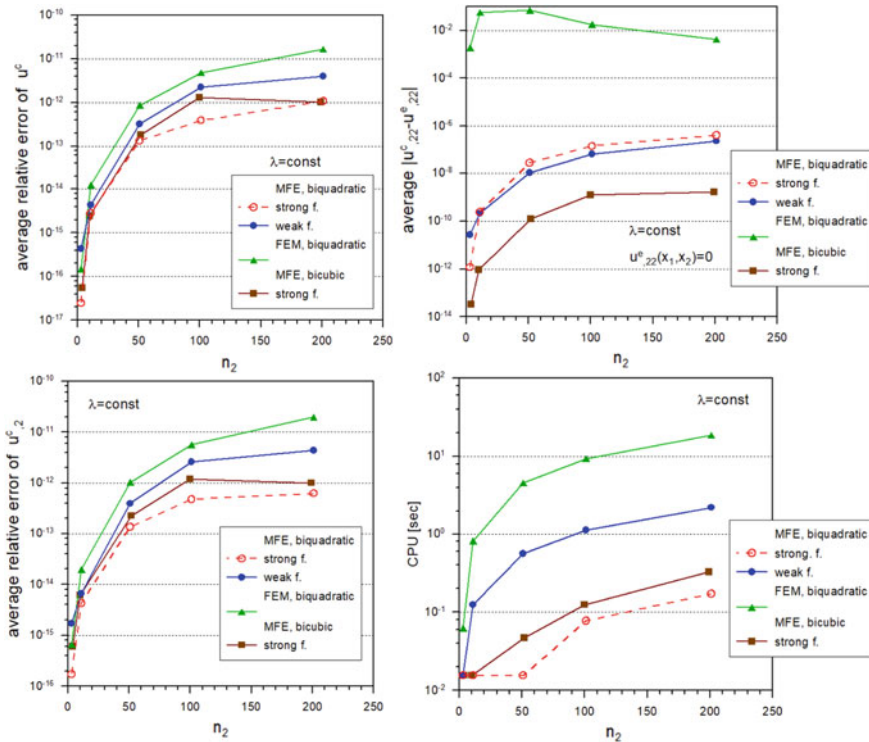
Fig. 3 Accuracy and computational efficiency of considered methods vs number of nodes in example with homogeneous body

(ii) *FGM body with power-law gradation* $\lambda(x_1, x_2) = \lambda_0\left(1 + \zeta\frac{x_2}{L}\right)^p, p \neq 1$:

$$u(x_1, x_2) = u_0 + \frac{u_L - u_0}{(1+\zeta)^{1-p} - 1}\left[(1 + \zeta x_2/L)^{1-p} - 1\right],$$

$$u_{,i}(x_1, x_2) = \delta_{i2}\frac{u_L - u_0}{(1+\zeta)^{1-p} - 1}\frac{1-p}{L}\frac{\zeta}{(1+\zeta x_2/L)^p} \qquad (17)$$

$$u_{,ij}(x_1, x_2) = \delta_{i2}\delta_{j2}\frac{u_L - u_0}{(1+\zeta)^{1-p} - 1}\frac{p(1-p)(\zeta/L)^2}{(1+\zeta x_2/L)^{p+1}}$$

(iii) *FGM body with exponential gradation* $\lambda(x_1, x_2) = \lambda_0\exp(\zeta x_2/L)$:

$$u(x_1, x_2) = u_0 + \frac{u_L - u_0}{1 - e^{-\zeta}}\left(1 - e^{-\zeta x_2/L}\right), \ u_{,i}(x_1, x_2) = \delta_{i2}\frac{u_L - u_0}{1 - e^{-\zeta}}\frac{\zeta}{L}e^{-\zeta x_2/L}$$
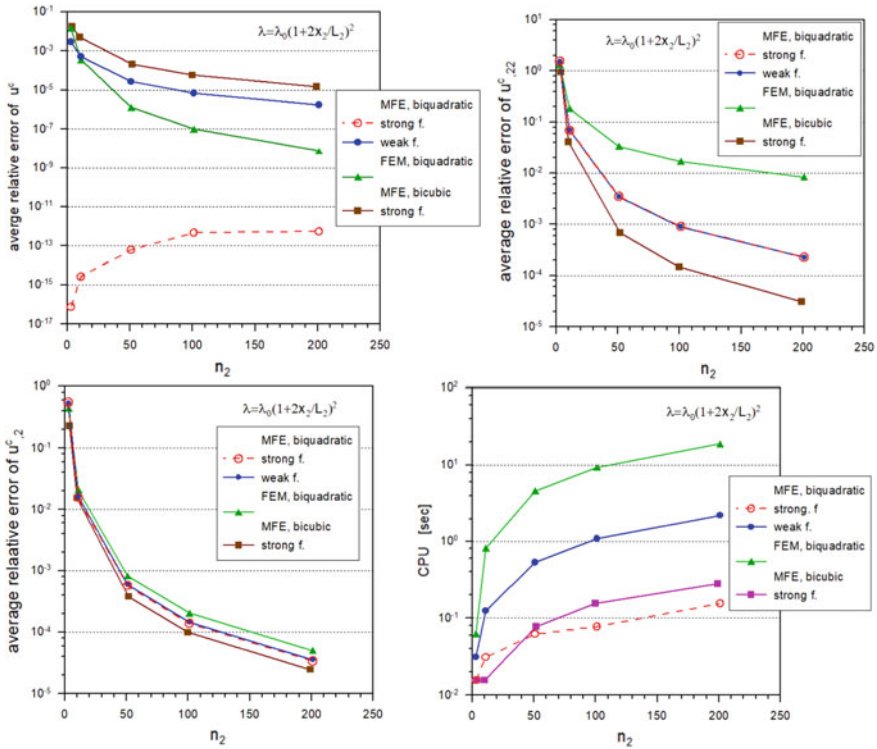
Fig. 4 Accuracy and computational efficiency of considered methods vs number of nodes in example with power-law gradation of heat conduction coefficient λ

$$u_{,ij}(x_1,\, x_2) = \delta_{i2}\delta_{j2}\frac{u_L - u_0}{1 - e^{-\zeta}}\left(\frac{\zeta}{L}\right)^2 e^{-\zeta x_2/L} \tag{18}$$

In numerical tests, we shall compare the exact solutions with the numerical results obtained by the following formulations:

(a) mesh-free strong formulation with MFE approximation using bi-quadratic Lagrange FE
(b) mesh-free local weak formulation with MFE approximation using bi-quadratic Lagrange FE
(c) mesh-free strong formulation with MFE approximation using bi-cubic Lagrange FE
(d) standard FEM using bi-quadratic Lagrange FE
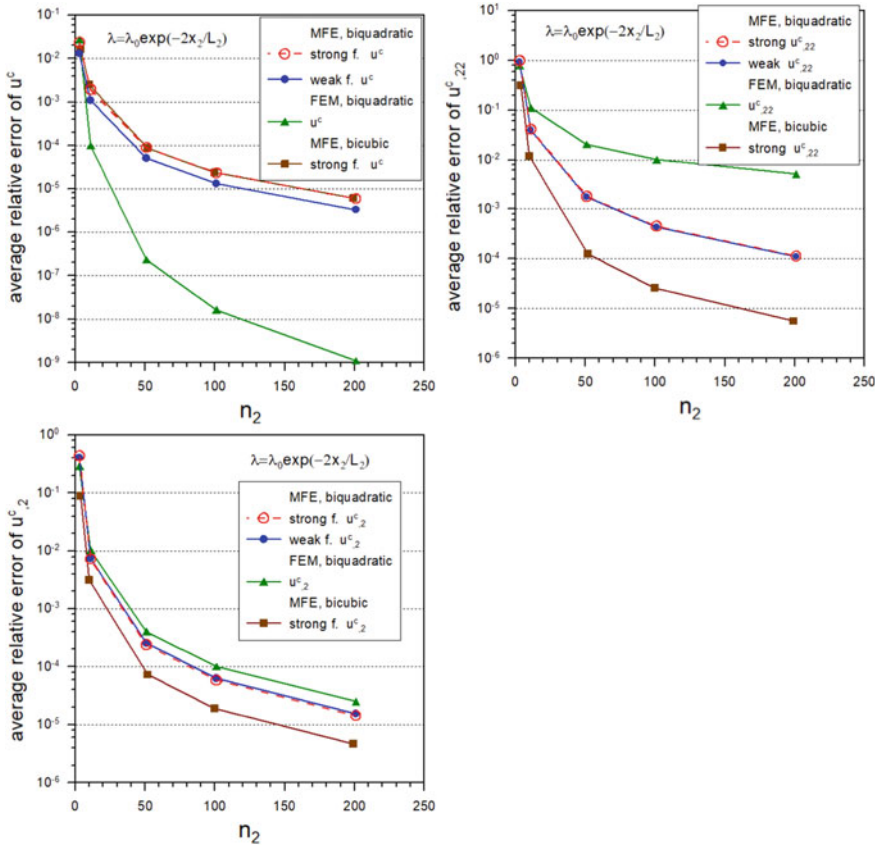
Fig. 5 Accuracy and computational efficiency of considered methods vs number of nodes in example with exponential gradation of heat conduction coefficient λ

The average relative error defined as

$$\frac{1}{N}\sum_{a=1}^{N}\left|\frac{f^c(\mathbf{x}^a)}{f^{ex}(\mathbf{x}^a)}-1\right|$$

is used for assessment of global accuracy, with $f^c(\mathbf{x}^a)$ and $f^{ex}(\mathbf{x}^a)$ being the numerically computed and exact results, respectively. The total number of nodes $N = n_1 \times n_2$ with n_1 being equal to 3 or 4 for the bi-quadratic and bi-cubic FE, respectively, in all examples except the numerical stability study.
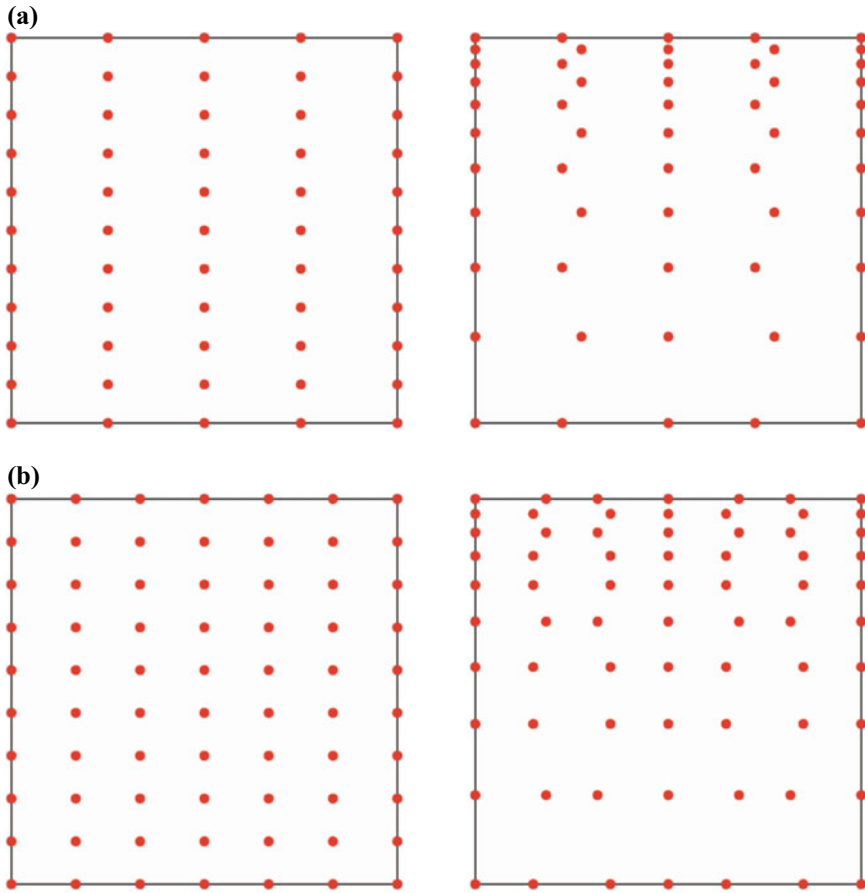
Fig. 6 Net of $n_1 \times n_2$ nodes used for: (a) bi-quadratic FE ($n_1 = 5$, $n_2 = 11$); (b) bi-cubic FE ($n_1 = 7$, $n_2 = 10$); Left – uniform distribution of nodes, Right – non-uniform distribution

The linear dependence of temperature on x_2 coordinate in case of homogeneous body, is achieved very accurately by each of the considered numerical approach (Fig. 3). Obviously, the accuracy in this example is decreasing with increasing the number of DOF (increasing n_2 under fixed n_1). Mesh-free approaches with MFE approximations give very good accuracy also for derivatives of temperature, while the accuracy for $u_{,22}(\mathbf{x}) \equiv 0$ by the standard FEM is significantly worse than by other employed methods.

As expected, the computational efficiency of weak formulation approaches is worse than that of strong formulations and the efficiency of standard FEM is the worst.
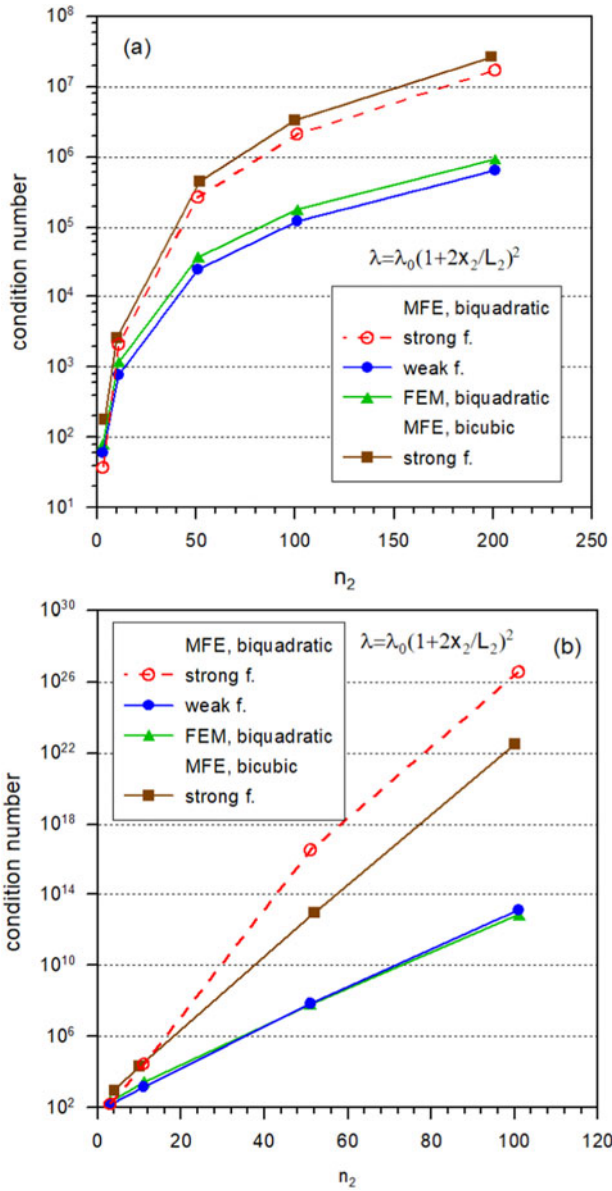
Fig. 7 Dependence of the condition numbers on n_2: (a) uniform distribution of nodes; (b) non-uniform distribution of nodes
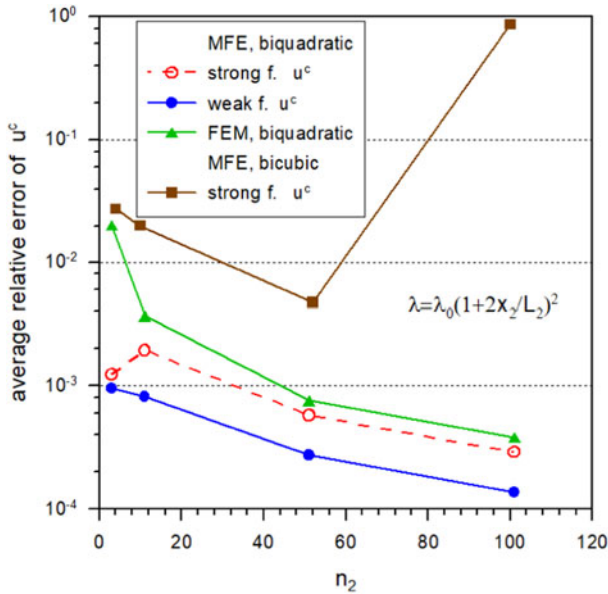
Fig. 8 Dependence of accuracy of temperature on the number of non-uniformly distributed nodes

The results for problems with power-law graded and exponentially graded heat conduction coefficient are shown in Fig. 4 and Fig. 5, respectively.

Surprisingly high accuracy of temperature results is achieved by strong formulation with MFE approximation using bi-quadratic Lagrange FE. All other numerical results converge to exact solution with accuracy for derivatives of temperature by mesh-free approaches combined with MFE approximation being better than that by standard FEM. As regards the computational efficiency, it exhibits similar behavior as in the case of homogeneous medium.

Similar behavior have been confirmed also in the problem with exponential gradation of the heat conduction coefficient, though surprisingly high rate of convergence of temperature results is obtained by standard FEM.

The stability of numerical solution of the systems of discretized equations in considered approaches have been studied via the dependence of condition numbers on the number of DOF for the uniform and non-uniform distributions of nodes as shown in Fig. 6. We present the numerical results for the case of power-law gradation of heat conduction coefficient.

In case of weak formulations, the condition numbers are smaller than those in case of strong formulations and disturbance of the regularity of distribution of nodes yields rapid increase of condition numbers (Fig. 7). Nevertheless, the accuracy is still acceptable also in case of non-uniform distributions of nodes (Fig. 8). Only the convergence of numerical results by using bi-cubic FE fails, when the density of nodes is very high.

4   Conclusions

The paper presents new element-free computational method, where instead of discretization of the analyzed domain into finite elements only the net of nodal points is needed. The spatial variations of field variables are approximated by polynomial interpolation within so-called moving finite elements which are created around particular nodes automatically depending on the position of node in the analyzed domain. Since there are no element interfaces, the known continuity difficulties of standard FEM are avoided. Moreover, utilization of polynomial interpolations for approximations brings better computational efficiency in comparison with other meshless approximations, where the shape functions are not expressed in terms of elementary functions. The MFE approximation is applicable in both the strong and weak formulations of boundary value problems, with the latter being taken on local subdomains involving only one node. The bandwidth of the system of discretized equations is very narrow, since only nodes of one element contribute to the discretized equation. The consideration of weak formulation in local sense enables truly meshless formulation in contrast to the global weak formulation, where the background mesh of elements is needed for integration over the analyzed domain. The proposed method takes advantages of both the mesh-free methods and the polynomial interpolation.

The performed numerical tests shown that the present MFE method gives better accuracy (especially for derivatives of primary field variables), exhibits better numerical stability with respect to irregularity of distribution of nodal points, and yields significantly better computational efficiency than the standard FEM.

References

1. Zienkiewicz, O.C., Taylor, R.I.: The Finite Element Method, 6th edn. Butterworth-Heinemann, Oxford (2005)
2. Hughes, T.J.R.: The Finite Element Method, Linear Static and Dynamic Finite Element Analysis. Prentice-Hall Inc., Englewood Clifs (1987)
3. Kolditz, O.: Computational Methods in Environmental Fluid Mechanics. Springer, Heidelberg (2002)
4. LeVeque, R.J.: Finite-Volume Methods for Hyperbolic Problems. Cambridge University Press, Cambridge (2002)
5. Yagawa, G., Furukawa, T.: Recent developments of free mesh method. Int. Jour. Num. Meth. Engng. **47**, 1419–1443 (2000)
6. Atluri, S.N., Shen, S.: The Meshless Local Petrov-Galerkin (MLPG) Method. Tech Science Press, Encino (2002)
7. Atluri, S.N.: The Meshless Method (MLPG) for Domain & BIE Discretizations. Tech Science Press, Forsyth (2004)

8. Liu, G.R.: Mesh Free Methods, Moving Beyond the Finite Element Method. CRC Press, Boca Raton (2003)
9. Liu, G.R.: An overview on meshfree methods: for computational solid mechanics. Int. J. Comput. Meth. **13**, 1630001 (2016)
10. Nguyen, V.P., Rabczuk, T., Bordas, S., Duflot, M.: Meshless methods: a review and computer implementation aspects. Math. Comput. Simul. **79**, 763–813 (2008)
11. Sladek, J., Stanak, P., Han, Z.D., Sladek, V., Atluri, S.N.: Applications of the MLPG method in engineering & sciences: a review. CMES **92**, 423–475 (2013)
12. Yagawa, G., Yamada, T.: Free mesh method: a new meshless finite element method. Comput. Mech. **18**, 383–386 (1996)
13. Kobayashi, Y., Shioya, R., Yagawa, G.: Parallel eigen frequency analysis using enriched free mesh method. Key Eng. Mater. **462–463**, 628–633 (2011)
14. Sladek, V., Sladek, J., Zhang, Ch.: Local integro-differential equations with domain elements for the numerical solution of partial differential equations with variable coefficients. Jour. Eng. Math. **51**, 261–282 (2005)
15. Sladek, V., Sladek, J., Zhang, Ch.: Computation of stresses in non-homogeneous elastic solids by local integral equation method. Comput. Mech. **41**, 827-845 (2008)
16. Sladek, V., Sladek, J., Zhang, Ch.: The use of finite elements for approximation of field variables on local subdomains in a mesh-free way. Chapter 6 In: Kompis, V. (ed.) Composites with Micro- and Nano-Structure, pp. 87–106. Springer, Heidelberg (2008)
17. Sladek, V., Repka, M., Sladek, J.: Moving finite element method. WIT Trans. Eng. Sci. **122**, 119–129 (2018)
18. Sladek, V., Sladek, J., Repka, M.: Mesh-free analysis of plate bending problems by Moving finite element approximation. WIT Trans. Eng. Sci. **126**, 211–223 (2019)
19. Sladek, V., Sladek, J., Repka, M., Sator, L.: FGM micro/nano-plates within modified couple stress elasticity. Compos. Struct. **245**, 112294 (2020)
20. Gao, X.W., Huang, S.Z., Cui, M., et al.: Element differential method for solving general heat conduction problems. Int. J. Heat Mass transfer **115**, 882–894 (2017)
21. Gao, X.W., Li, Z.Y., Yang, K., et al.: Element differential method and its application in thermal-mechanical problems. Int. J. Num. Meth. Eng. **113**, 82–108 (2018)
22. Wen, P.H., Cao, P., Korakianitis, T.: Finite block method in elasticity. Eng. Anal. Bound. Elem. **46**, 116–125 (2014)
23. Li, M., Wen, P.H.: Finite block method for transient heat conduction analysis in functionally graded media. Int. J. Num. Meth. Eng. **99**, 372–390 (2014)
24. Fantuzzi, N., Tornabene, F.: Strong formulation finite element method for arbitrarily shaped laminated plates—part I. Theoretical analysis. Adv. Aircr. Spacecr. Sci. **1**, 125–143 (2014)
25. Fantuzzi, N.: New insight into the strong formulation finite element method for solving elastostatic and elastodynamic problems. Curved Layer Struct. **1**, 93–126 (2014)
26. Tornabene, F., Fantuzzi, N., Ubertini, F., Viola, E.: Strong formulation finite element method based on differential quadrature: a survey. Appl. Mech. Rev. **1**, 145–175 (2015)
27. Gao, X.W., Gao, L.F., Zhang, Y., Cui, M., Lv, J.: Free element collocation method: a new method combining advantages of finite element and mesh free methods. Comput. Struct. **215**, 10–26 (2019)

The MLPG Method in Multiphysics and Scale Dependent Problems

Jan Sladek, Vladimir Sladek, and Miroslav Repka

Abstract Two multiphysical and scale dependent problems in flexoelectricity and thermoelectricity are analysed by the meshless Petrov-Galerkin (MLPG) method. The size-effect is considered in constitutive equations by the strain- and electric field-gradients in the flexoelectricity. The thermal conductivity can be reduced significantly in nano-sized structures and higher-grade heat flux theory is developed in the thermoelectricity. The governing equations are derived for both problems by the variational principle. The coupled governing partial differential equations (PDE) are satisfied in a local weak-form on small fictitious subdomains. A simple test function is applied there. The moving least-squares (MLS) approximation is used for all physical fields. It is convenient approach for problems with higher order derivatives occurred in gradient theories. Applying the spatial approximations in local integral equations a system of algebraic equations is obtained for the nodal values.

Keywords Flexoelectricity · Thermoelectricity · Gradient theories · Governing equations · MLS approximation

1 Introduction

The recent miniaturized trend of electromechanical systems leads to a tremendous growth in the scale of strain gradient in nano-sized structures. The physical properties can be significantly improved in nanocomposites. It is an increasing demand for reliable and accurate theories and numerical methods for modelling of nano-sized structures. The size-effect phenomena are experimentally observed in structures, if the intrinsic length parameter of structure is comparable with the size of the analyzed body [1–5]. The direct flexoelectricity describes an electromechanical coupling effect between the electric polarization and mechanical strain

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
S. N. Atluri and I. Vušanović (eds.), *Computational and Experimental Simulations in Engineering*, Mechanisms and Machine Science 98,
https://doi.org/10.1007/978-3-030-67090-0_31

gradient [6–8]. It can be viewed as a higher order effect [9] with respect to piezoelectricity. Also the electric intensity vector gradients are very large in nano-sized dielectrics. In the converse flexoelectricity [10–12] the stresses are proportional to the gradients of electric intensity vector. Thermoelectric properties can be also significantly improved by nanotechnology [13]. Nano-sized structures have a potential to convert waste heat directly into electricity effectively. The thermal conductivity is reduced without an influence on the high electrical conductivity [14].

It is an increasing demand for advanced continuum models to incorporate the intrinsic length parameter (characteristic of material microstructure) into consideration [15–17]. Former gradient theory in elasticity with many length scales as material parameters were simplified by Aifantis [18], where only one length parameter is occurred. These theories are utilized in the present model for flexoelectricity. Similarly, the size effect should be considered in the heat conduction problems in nano-sized structures with large thermal gradients [19]. The integro-differential form of the constitutive law in nonlocal theory can be transformed into a more convenient differential form if the weight function satisfies the Helmholtz equation. In both flexoelectric and thermoelectric problems, characteristic lengths of material structure are considered in constitutive equations and corresponding governing equations contain higher order of derivatives than in the classical theory. Both problems are also multiphysical, since it is there interaction of more physical fields.

We need a reliable and accurate computational method to solve the above mentioned problems with size effects. Due to higher order derivatives it is needed to have the C^1-continuous elements for spatial approximations. The mixed formulation in the finite element method (FEM) [20, 21] are preferred in literature. It is well known that desired continuity for our problems can be received very easily by the Moving Least-square (MLS) approximation. The Meshless Local Petrov-Galerkin method (MLPG) has been successfully applied for many engineering problems in macro-sized structures [22–24]. The first efforts to apply the MLPG to nano-sized structures authors have developed recently in works [25, 26].

The meshless method based on the MLPG weak-form of multiphysical problems in dielectric and thermoelectric solids is developed in the present paper on more general base than in earlier work [26]. The converse flexoelectricity for large electric intensity gradients is considered here. To the best of the authors' knowledge, no work has been carried out so far to develop the MLPG for converse flexoelectricity and gradient theory of thermoelectricity.

2   The MLPG for Flexoelectricity

2.1   *The Direct and Converse Flexoelectricity*

The constitutive equations for piezoelectric solids with strain and electric intensity vector gradients are given by [27, 28]

$$
\begin{aligned}
\sigma_{ij} &= \frac{\partial H}{\partial \varepsilon_{ij}} = c_{ijkl}\varepsilon_{kl} - e_{kij}E_k - b_{klij}E_{k,l}, \\
\tau_{jkl} &= \frac{\partial H}{\partial \eta_{jkl}} = -f_{ijkl}E_i + g_{jklmni}\eta_{nmi}, \\
D_i &= -\frac{\partial H}{\partial E_i} = a_{ij}E_j + e_{ijk}\varepsilon_{jk} + f_{ijkl}\eta_{jkl} \\
Q_{ij} &= -\frac{\partial H}{\partial E_{i,j}} = b_{ijkl}\varepsilon_{kl} + h_{ijkl}E_{k,l}
\end{aligned}
\tag{1}
$$

where σ_{ij}, D_k, τ_{jkl} and Q_{ij} are the stress tensor, electric displacements, higher order stress and electric quadrupole, respectively. Symbols **a** and **c** are used for the second-order permittivity and the fourth-order elastic constant tensors, respectively. The piezoelectric and flexoelectric coefficients are denoted by **e** and **f**. Symbol **g** is used for the higher order elastic coefficients. The symbols **b** and **h** denote the converse flexoelectric coefficients and higher-order electric parameters, respectively.

The strain ε_{ij} and the electric intensity vector E_j are related to the mechanical displacements u_i and the electric potential ϕ by

$$
\varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right), \quad E_j = -\phi_{,j}.
\tag{2}
$$

The strain-gradients are given as

$$
\eta_{ijk} = \varepsilon_{ij,k} = \frac{1}{2}\left(u_{i,jk} + u_{j,ik}\right).
\tag{3}
$$

In the Aifantis' simplified gradient theory of elasticity [18, 29, 30] only one internal length-scale parameter l is occurred. According to this theory the higher-order elastic parameters g_{jklmni} are proportional to elastic stiffness coefficients c_{klmn} and the intrinsic length parameter l [31, 32] as

$$
g_{jklmni} = l^2 c_{jkmn}\delta_{li},
\tag{4}
$$

with δ_{li} being the Kronecker delta.

A similar approach is utilized for the higher-order electric parameters

$$h_{ijkl} = q^2 a_{ik}\delta_{jl}, \tag{5}$$

where q is another length-scale parameter.

Only two independent parameters f_1 and f_2 for the direct flexoelectric coefficient f_{ijkl}, $f_{ijkl} = f_1\delta_{jk}\delta_{il} + f_2(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl})$ are used in the simplified theory [20].

Also the number of independent converse flexoelectric coefficients b_{ijkl} can be reduced. The stresses in the orthotropic piezoelectric material with poling direction along the transversal isotropy x_3-axis are proportional to the electric intensity vector $\sigma_{11} = e_{31}E_3$, $\sigma_{33} = e_{33}E_3$, $\sigma_{13} = e_{15}E_1$,

with

$$e_{kij} = (e_{31}\delta_{i1}\delta_{j1} + e_{33}\delta_{i3}\delta_{j3})\delta_{k3} + e_{15}(\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1})\delta_{k1} \tag{6}$$

where standard Voight notation is applied [25].

Similarly, stresses are proportional to gradients of electric intensity vectors in the converse flexoelectricity

$$\sigma_{ij} = \delta_{ij}b_1(E_{1,1} + E_{3,3}), \sigma_{13} = \sigma_{31} = b_2E_{1,3} + b_3E_{3,1} \tag{7}$$

where three independent converse flexoelectric coefficients b_1, b_2 and b_3 are used $b_{klij} = b_1\delta_{ij}\delta_{kl} + (\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1})(b_2\delta_{k1}\delta_{l3} + b_3\delta_{k3}\delta_{l1})$.

It follows from above assumptions that constitutive Eqs. (1) for orthotropic materials can be rewritten into a matrix form as

$$
\begin{bmatrix} \sigma_{11} \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{13} & 0 \\ c_{13} & c_{33} & 0 \\ 0 & 0 & c_{44} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{33} \\ 2\varepsilon_{13} \end{bmatrix} - \begin{bmatrix} 0 & e_{31} \\ 0 & e_{33} \\ e_{15} & 0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_3 \end{bmatrix} - \begin{bmatrix} b_1 & 0 \\ b_1 & 0 \\ 0 & b_3 \end{bmatrix} \begin{bmatrix} E_{1,1} \\ E_{3,1} \end{bmatrix} - \begin{bmatrix} 0 & b_1 \\ 0 & b_1 \\ b_2 & 0 \end{bmatrix} \begin{bmatrix} E_{1,3} \\ E_{3,3} \end{bmatrix}
$$
$$
= \mathbf{C} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{33} \\ 2\varepsilon_{13} \end{bmatrix} - \boldsymbol{\Lambda} \begin{bmatrix} E_1 \\ E_3 \end{bmatrix} - \boldsymbol{\Phi}_1 \begin{bmatrix} E_{1,1} \\ E_{3,1} \end{bmatrix} - \boldsymbol{\Phi}_3 \begin{bmatrix} E_{1,3} \\ E_{3,3} \end{bmatrix}, \tag{8}
$$

$$
\begin{bmatrix} D_1 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & e_{15} \\ e_{31} & e_{33} & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{33} \\ 2\varepsilon_{13} \end{bmatrix} + \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} E_1 \\ E_3 \end{bmatrix}
$$
$$
+ \begin{bmatrix} f_1 + 2f_2 & f_1 & 0 \\ 0 & 0 & f_2 \end{bmatrix} \begin{bmatrix} \varepsilon_{11,1} \\ \varepsilon_{33,1} \\ 2\varepsilon_{13,1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & f_2 \\ f_1 & f_1 + 2f_2 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{11,3} \\ \varepsilon_{33,3} \\ 2\varepsilon_{13,3} \end{bmatrix}, \tag{9}
$$
$$
= \boldsymbol{\Lambda}^T \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{33} \\ 2\varepsilon_{13} \end{bmatrix} + \mathbf{A} \begin{bmatrix} E_1 \\ E_3 \end{bmatrix} + \mathbf{F}_1 \begin{bmatrix} \varepsilon_{11,1} \\ \varepsilon_{33,1} \\ 2\varepsilon_{13,1} \end{bmatrix} + \mathbf{F}_3 \begin{bmatrix} \varepsilon_{11,3} \\ \varepsilon_{33,3} \\ 2\varepsilon_{13,3} \end{bmatrix}
$$

$$
\begin{bmatrix} \tau_{11k} \\ \tau_{33k} \\ \tau_{13k} \end{bmatrix} = -\mathbf{F}_k^T \begin{bmatrix} E_1 \\ E_3 \end{bmatrix} + l^2 \mathbf{C} \begin{bmatrix} \varepsilon_{11,k} \\ \varepsilon_{33,k} \\ 2\varepsilon_{13,k} \end{bmatrix}, \tag{10}
$$

$$
\begin{bmatrix} Q_{1k} \\ Q_{3k} \end{bmatrix} = \mathbf{\Phi}_k^T \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{33} \\ 2\varepsilon_{13} \end{bmatrix} + q^2 \mathbf{A} \begin{bmatrix} E_{1,k} \\ E_{3,k} \end{bmatrix}. \tag{11}
$$

The governing equations for the considered constitutive equations are derived in work [33]

$$
\begin{aligned}
\sigma_{ij,j}(\mathbf{x}) - \tau_{ijk,jk}(\mathbf{x}) &= 0, \\
D_{i,i}(\mathbf{x}) - Q_{ij,ji}(\mathbf{x}) &= 0.
\end{aligned} \tag{12}
$$

For a unique formulation of boundary value problems the above governing equations have to be supplied by essential or natural boundary conditions (b.c.):

1.) Essential b.c.:

$$
\begin{aligned}
u_i(\mathbf{x}) &= \bar{u}_i(\mathbf{x}) \text{ on } \Gamma_u, \Gamma_u \subset \Gamma \\
s_i(\mathbf{x}) &= \bar{s}_i(\mathbf{x}) \text{ on } \Gamma_s, \Gamma_s \subset \Gamma \\
\phi(\mathbf{x}) &= \bar{\phi}(\mathbf{x}) \text{ on } \Gamma_\phi, \Gamma_\phi \subset \Gamma \\
p(\mathbf{x}) &= \frac{\partial \phi}{\partial n} \bar{p}(\mathbf{x}) \text{ on } \Gamma_p, \Gamma_p \subset \Gamma
\end{aligned} \tag{13}
$$

2.) Natural b.c.:

$$
\begin{aligned}
t_i(\mathbf{x}) &= \bar{t}_i(\mathbf{x}) \text{ on } \Gamma_t, \Gamma_t \cup \Gamma_u = \Gamma, \Gamma_t \cap \Gamma_u = \varnothing \\
R_i(\mathbf{x}) &= \bar{R}_i(\mathbf{x}) \text{ on } \Gamma_R, \Gamma_R \cup \Gamma_s = \Gamma, \Gamma_R \cap \Gamma_s = \varnothing \\
S(\mathbf{x}) &= \bar{S}(\mathbf{x}) \text{ on } \Gamma_S, \Gamma_S \cup \Gamma_\phi = \Gamma, \Gamma_S \cap \Gamma_\phi = \varnothing \\
Z(\mathbf{x}) &= \bar{Z}(\mathbf{x}) \text{ on } \Gamma_Z, \Gamma_Z \cup \Gamma_p = \Gamma, \Gamma_Z \cap \Gamma_p = \varnothing,
\end{aligned} \tag{14}
$$

where

$$
s_i := \frac{\partial u_i}{\partial \mathbf{n}}, p := \frac{\partial \phi}{\partial \mathbf{n}}, R_i := n_k n_j \tau_{ijk}, Z := n_i n_j Q_{ij}, \tag{15}
$$

and the traction vector, and the electric charge are given by

$$
t_i := n_j (\sigma_{ij} - \tau_{ijk,k}) - \frac{\partial \rho_i}{\partial \pi} + \sum_c \|\rho_i(\mathbf{x}^c)\| \delta(\mathbf{x} - \mathbf{x}^c), \tag{16}
$$

$$S := n_k\big(D_k - Q_{kj,j}\big) - \frac{\partial \alpha}{\partial \boldsymbol{\pi}} + \sum_c \|\alpha(\mathbf{x}^c)\|\delta(\mathbf{x} - \mathbf{x}^c), \tag{17}$$

with $\rho_i := n_k\pi_j\tau_{ijk}$, $\alpha := n_i\pi_j Q_{ij}$, $\delta(\mathbf{x})$ being the Dirac delta function and π_i is the Cartesian component of the unit tangent vector on Γ.

The jump at a corner ($\mathbf{x}^c$) on the oriented boundary contour Γ is defined as

$$\|\rho_i(\mathbf{x}^c)\| := \rho_i(\mathbf{x}^c - 0) - \rho_i(\mathbf{x}^c + 0), \tag{18}$$

$$\|\alpha(\mathbf{x}^c)\| := \alpha(\mathbf{x}^c - 0) - \alpha(\mathbf{x}^c + 0). \tag{19}$$

2.2 The MLPG Formulation

Equations for boundary value problems for direct and converse flexoelextricity are formulated in earlier paragraph. Recently, authors have developed the mixed FEM for nano-structured solar cells [34]. The MLPG with the MLS approximation is convenient approach for problems described by governing equations with higher order derivatives, since gradients of primary fields are continuous here [22–24].

The MLPG method is based on the local weak-form on local fictitious subdomains Ω^q. These subdomains are constructed for each node $\mathbf{x}^q$ inside the analysed domain. The local weak-form of the mechanical governing Eq. (12) can be written as

$$\int_{\Omega^q} \big[\sigma_{ij,j}(x) - \tau_{ijk,jk}(\mathbf{x})\big]\, u_{im}^*(\mathbf{x})d\Omega = 0, \tag{20}$$

where $u_{im}^*(\mathbf{x})$ is a test function.

It is convenient to select a Heaviside step function as the test function $u_{ik}^*(\mathbf{x})$

$$u_{ik}^*(\mathbf{x}) = \begin{cases} \delta_{ik} & \text{at } \mathbf{x} \in \Omega^q \cup \partial\Omega^q \\ 0 & \text{at } \mathbf{x} \notin \Omega^q \cup \partial\Omega^q \end{cases}. \tag{21}$$

Then, the local weak-form (20) is transformed into the following local integral equation

$$\int_{L^q + \Gamma_u^q} n_j\big(\sigma_{ij} - \tau_{ijk,k}\big)d\Gamma + \rho_i(\mathbf{x}_t^f) - \rho_i(\mathbf{x}_t^s) = -\int_{\Gamma_t^q} \bar{t}_i d\Gamma, \tag{22}$$

where $\partial\Omega^q$ is the boundary of the local subdomain which consists of three parts $\partial\Omega^q = L^q \cup \Gamma_t^q \cup \Gamma_u^q$, in general (see Fig. 1) and $\mathbf{x}_t^f$, $\mathbf{x}_t^s$ stand for the final and starting points on Γ_t^q.

Similarly, one can derive the local integral equation (LIE) for the second governing Eq. (12)

$$\int\limits_{L^q+\Gamma_\phi^q} n_j\left(D_j - Q_{ij,i}\right)d\Gamma + \alpha(\mathbf{x}_S^f) - \alpha(\mathbf{x}_S^s) = -\int\limits_{\Gamma_S^q} \bar{S}d\Gamma. \tag{23}$$

The MLS approximation of trial functions is applied for numerical solution of integral equations. The mechanical displacements and electric potential are approximated by [23]

$$\mathbf{u}^h(\mathbf{x}) = \mathbf{N}^T(\mathbf{x}) \cdot \hat{\mathbf{u}} = \sum_{a=1}^{n} N^a(\mathbf{x})\hat{\mathbf{u}}^a, \qquad \phi^h(\mathbf{x}) = \sum_{a=1}^{n} N^a(\mathbf{x})\hat{\phi}^a, \tag{24}$$

where $\hat{\mathbf{u}}^a = \left(\hat{u}_1^a, \ \hat{u}_3^a\right)^T$ and $\hat{\phi}^a$ are fictitious nodal parameters for displacements and electric potential, respectively. Shape functions are denoted by $N^a(\mathbf{x})$. The size of support domain of the weight function $w^a(\mathbf{x})$ has influence on the number of nodes n used for the spatial approximation.
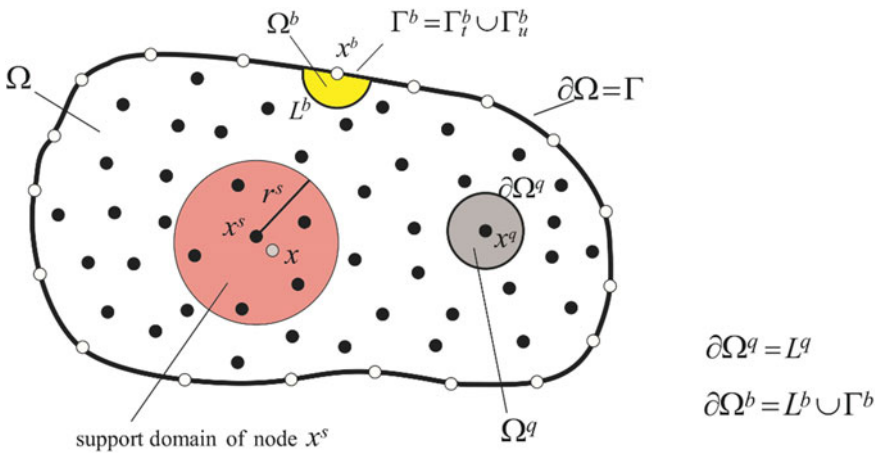

Fig. 1 Local subdomains Ω^q and Ω^b with their boundaries for MLS approximation of the trial function; support domain of weight function around node $\mathbf{x}^s$

The strains and electric intensity vector are approximated by

$$\boldsymbol{\varepsilon}^h(\mathbf{x}) = \begin{bmatrix} \varepsilon_{11}^h \\ \varepsilon_{33}^h \\ 2\varepsilon_{13}^h \end{bmatrix} = \sum_{a=1}^{n} \mathbf{B}^a(\mathbf{x})\hat{\mathbf{u}}^a, \quad \mathbf{E}^h(\mathbf{x}) = \begin{bmatrix} E_1^h \\ E_3^h \end{bmatrix} = -\sum_{a=1}^{n} \mathbf{P}^a(\mathbf{x})\hat{\phi}^a,$$

$$\boldsymbol{\varepsilon}_{,k}^h(\mathbf{x}) = \begin{bmatrix} \varepsilon_{11,k}^h \\ \varepsilon_{33,k}^h \\ 2\varepsilon_{13,k}^h \end{bmatrix} = \sum_{a=1}^{n} \mathbf{B}_{,k}^a(\mathbf{x})\hat{\mathbf{u}}^a, \quad \mathbf{E}_{,k}^h(\mathbf{x}) = \begin{bmatrix} E_{1,k}^h \\ E_{3,k}^h \end{bmatrix} = -\sum_{a=1}^{n} \mathbf{P}_{,k}^a(\mathbf{x})\hat{\phi}^a$$

$$(25)$$

where

$$\mathbf{B}^a(\mathbf{x}) = \begin{bmatrix} N_{,1}^a & 0 \\ 0 & N_{,3}^a \\ N_{,3}^a & N_{,1}^a \end{bmatrix}, \quad \mathbf{P}^a(\mathbf{x}) = \begin{bmatrix} N_{,1}^a \\ N_{,3}^a \end{bmatrix}, \quad \mathbf{B}_{,k}^a(\mathbf{x})$$

$$= \begin{bmatrix} N_{,1k}^a & 0 \\ 0 & N_{,3k}^a \\ N_{,3k}^a & N_{,1k}^a \end{bmatrix}, \quad \mathbf{P}_{,k}^a(\mathbf{x}) = \begin{bmatrix} N_{,1k}^a \\ N_{,3k}^a \end{bmatrix}. \quad (26)$$

The system of algebraic equations for unknown nodal quantities is obtained from the local boundary-domain integral Eqs. (22) and (23) if the MLS-approximation are substituted there.

2.3 Numerical Examples

A cantilever beam (Fig. 2) is analyzed in this example. The beam is made of zinc oxide (ZnO) with material properties [36]:

$$c_{11} = 209.7 \cdot 10^9 \,\mathrm{Nm}^{-2}, \, c_{12} = 105 \cdot 10^9 \,\mathrm{Nm}^{-2}, \, c_{13} = 105 \cdot 10^9 \,\mathrm{Nm}^{-2},$$

$$c_{33} = 211 \cdot 10^9 \,\mathrm{Nm}^{-2}, \, c_{44} = 42.37 \cdot 10^9 \,\mathrm{Nm}^{-2},$$

$$e_{15} = -0.48 \,\mathrm{Cm}^{-2}, \, e_{31} = -0.567 \,\mathrm{Cm}^{-2}, \, e_{33} = 1.32 \,\mathrm{Cm}^{-2},$$

$$a_0 = 8.854 \cdot 10^{-12} C(Vm)^{-1}, \, a_{11} = a_{33} = 8.544 a_0.$$

The direct flexoelectricity coefficients are assumed to be $f_1 = 1.0 \times 10^{-7}$ C/m and $f_2 = 0$. The electric displacement flux vanishes on all surfaces except for the bottom surface where the electric potential is zero.

The length of the beam is $L = 500$ nm and height is $H = 100$ nm. The internal material size parameter is selected as $l = 10$ nm. Results are presented for the end load $F = 100$ nN.
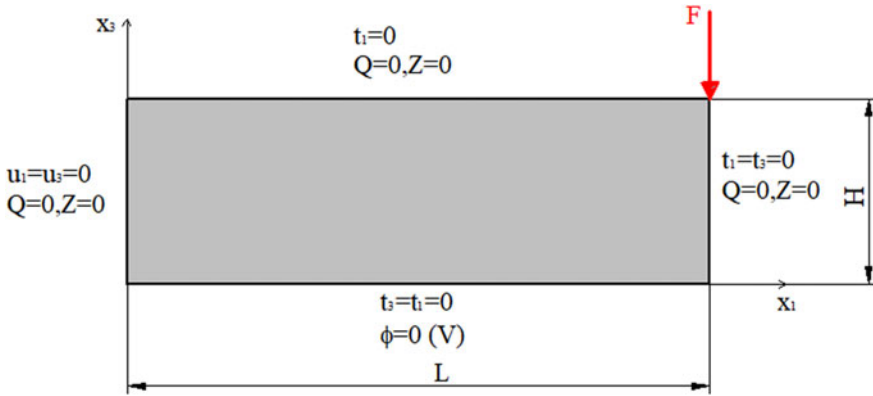
Fig. 2 Cantilever beam with shear force load

This problem was analyzed in our previous papers by the FEM [31]. The FEM and MLPG results are in a good agreement. One can observe from Fig. 3 that, the beam deflection of the neutral axis decreases with increasing value of the flexoelectric coefficient. This is in agreement with our earlier results.

The induced electric potential on the upper surface is presented in Fig. 4. The electric potential is significantly growing for larger flexoelectric coefficients. This example illustrates that nanotechnology can enhance the piezoelectricity effect.

3 Gradient Theory in Thermoelectricity

In nano-sized structures it is needed to develop a theory of heat conduction, where size effect is considered. Similarly to the elasticity problems it is possible to consider the size effect here through the nonlocal heat transport [19]. The heat flux vector in nonlocal theory is given by

$$\lambda_i(\mathbf{x}) = - \int_V \alpha(\mathbf{x} - \mathbf{y}) \kappa_{ij}(\mathbf{y}) \theta_{,j}(\mathbf{y}) dV(\mathbf{y}) , \qquad (27)$$

where the temperature differences are denoted by $\theta = T - T_0$ with the reference temperature T_0. The thermal conductivity and the nonlocal kernel function are denoted by and $\alpha(\mathbf{x} - \mathbf{y})$, respectively.

For a special weight function satisfying the Helmholtz equation

$$\left(1 - l^2 \nabla^2\right) \alpha(|\mathbf{x} - \mathbf{y}| = \delta(\mathbf{x} - \mathbf{y}), \qquad (28)$$
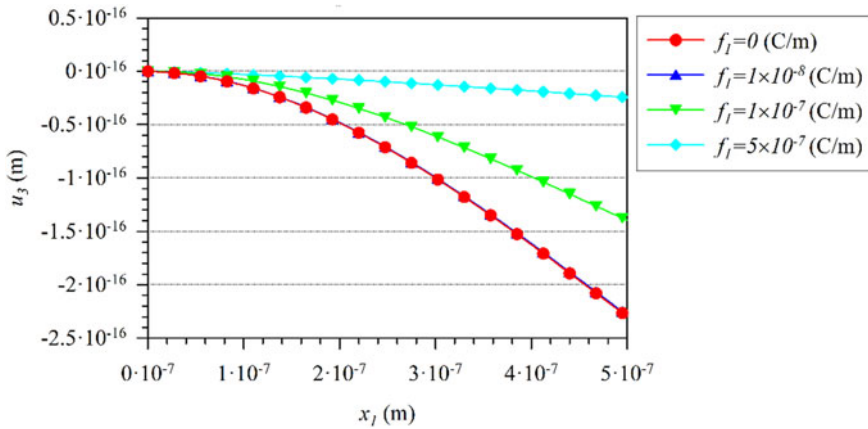
Fig. 3 Variation of the beam deflection u_3 along neutral axis x_1 for various flexoelectric coefficients
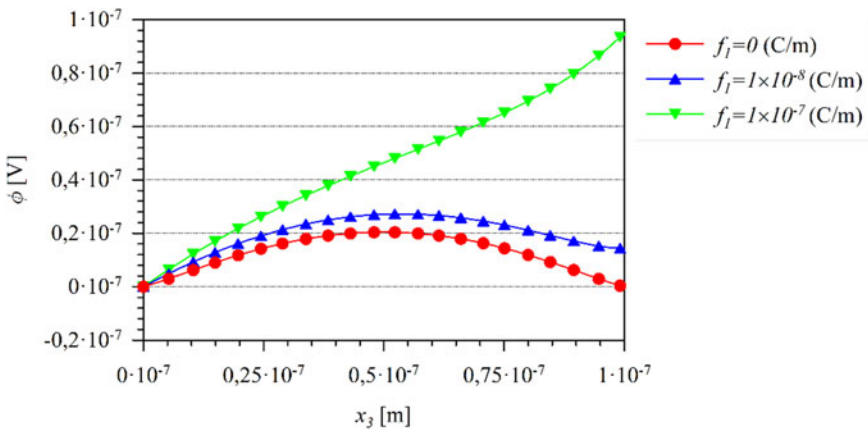
Fig. 4 Variation of the induced electric potential along neutral axis x_1 for various flexoelectric coefficients

it is possible to derive the heat conduction equation with the size effect

$$\left(1 - l^2\nabla^2\right) \Lambda_i = -\kappa_{ij}\theta_{,j} \text{ or } \left(1 - l^2\nabla^2\right) \Lambda_{i,i} = w \tag{29}$$

where w is the volume density of heat source, $\delta(\mathbf{x} - \mathbf{y})$ is the Dirac function and l is a characteristic length material parameter.

Formally, the same governing equation as given in Eq. (29), can be obtained also in the higher-grade theory with including the higher-grade heat flux m_{ik} (i.e.

canonically conjugated fields with $\theta_{,jk}$) into constitutive equations in addition to the classical heat flux λ_i as

$$\lambda_i = -\kappa_{ij}\theta_{,j} , \tag{30}$$

$$m_{ik} = -l^2\kappa_{ij}\theta_{,jk} . \tag{31}$$

Applying this theory one can rewrite the classical constitutive equations for thermoelectric materials into the following form

$$
\begin{aligned}
\lambda_i &= -\,\bar{\kappa}_{ij}\theta_{,j} + \bar{\zeta}_{ij}E_j. \\
J_i &= s_{ij}E_j - \zeta_{ij}\theta_{,j}, \\
m_{ik} &= -\,l^2\kappa_{ij}\theta_{,jk} ,
\end{aligned}
\tag{32}
$$

where the electric current density is denoted by J_i. The electrical conductivity, Seebeck and Peltier coefficients are denoted by s_{ij}, ζ_{ij} and $\bar{\zeta}_{ij}$, respectively. Both Seebeck and Peltier coefficients are correlated via the absolute temperature T as $\bar{\zeta}_{ij} = \zeta_{ij}T$, with $T = T_0 + \theta$, where T_0 is the reference temperature. The total heat conduction coefficient $\bar{\kappa}_{ij} = (\kappa_{ij} + \kappa_{ij}^e)$ is composed of the heat conduction κ_{ij} measured when $J_i = 0$ and contribution to heat conduction κ_{ij}^e because of electric current [35]. The Seebeck coefficient is proportional to the electric current conductivity $\zeta_{ij} = \alpha s_{ij}$, $\bar{\zeta}_{ij} = \alpha s_{ij}T$ and $\bar{\kappa}_{ij} = (\kappa_{ij} + \alpha^2 s_{ij}T)$.

The matrix form of constitutive Eqs. (32) for 2-d problems are given by

$$
\begin{pmatrix} J_1 \\ J_2 \end{pmatrix} = \begin{pmatrix} s_{11} & 0 \\ 0 & s_{22} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} - \begin{pmatrix} \zeta_{11} & 0 \\ 0 & \zeta_{22} \end{pmatrix} \begin{pmatrix} \theta_{,1} \\ \theta_{,2} \end{pmatrix} = [\mathbf{S}]\{\mathbf{E}\} - [\mathbf{Z}]\begin{pmatrix} \theta_{,1} \\ \theta_{,2} \end{pmatrix},
\tag{33}
$$

$$
\begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix} = \begin{pmatrix} \bar{\zeta}_{11} & 0 \\ 0 & \bar{\zeta}_{22} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} - \begin{pmatrix} \bar{\kappa}_{11} & 0 \\ 0 & \bar{\kappa}_{22} \end{pmatrix} \begin{pmatrix} \theta_{,1} \\ \theta_{,2} \end{pmatrix} = [\bar{\mathbf{Z}}]\{\mathbf{E}\} - [\bar{\kappa}]\begin{pmatrix} \theta_{,1} \\ \theta_{,2} \end{pmatrix},
\tag{34}
$$

$$
\begin{pmatrix} m_{1k} \\ m_{2k} \end{pmatrix} = -l^2 \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_{,1k} \\ \theta_{,2k} \end{pmatrix} = -l^2[\mathbf{\kappa}]\begin{pmatrix} \theta_{,1k} \\ \theta_{,2k} \end{pmatrix}.
\tag{35}
$$

The weak form for this problem can be written as

$$
\int_V \left(J_i \delta\phi_{,i} + \lambda_i \delta\theta_{,i} + m_{ik}\delta\theta_{,ik}\right)dV + \int_V w\delta\theta dV
$$

$$
= -\int_V \left[J_{i,i}\delta\phi + \left(\lambda_{i,i} - w\right)\delta\theta + m_{ik,k}\delta\theta_{,i}\right]dV
$$

$$
+ \int_{\partial V} \left(n_i J_i \delta\phi + n_i \lambda_i \delta\theta + n_k m_{ik}\delta\theta_{,i}\right)d\Gamma
$$

$$
- \int_V \left[J_{i,i}\delta\phi + \left(\lambda_{i,i} - m_{ik,ik} - w\right)\delta\theta\right]dV \tag{36}
$$

$$
+ \int_{\partial V} \left[n_i J_i \delta\phi + n_i \left(\lambda_i - m_{ik,k}\right)\delta\theta + n_k m_{ik}\delta\theta_{,i}\right]d\Gamma
$$

$$
= -\int_V \left[J_{i,i}\delta\phi + \left(\lambda_{i,i} - m_{ik,ik} - w\right)\delta\theta\right]dV
$$

$$
- \int_{\partial V} \left(\Lambda\delta\theta + P\delta p + Q\delta\phi\right)d\Gamma
$$

where P, Q and Λ are independent boundary densities conjugated with $p = \partial\theta/\partial\mathbf{n}$, ϕ and θ, respectively, and given as $P = n_k n_i m_{ik}$, $Q = n_k J_k$

$$
\Lambda = n_j \left(\Lambda_i - m_{ik,k}\right) - \frac{\partial\mu}{\partial\boldsymbol{\pi}} + \sum_c [\![\mu(\mathbf{x}^c)]\!]\delta(\mathbf{x} - \mathbf{x}^c) \tag{37}
$$

$$
\mu = n_k \pi_i m_{ik} \tag{38}
$$

with Λ being the heat flux, n_i and π_i are the Cartesian component of the unit tangent vector on Γ, and the jump at a corner on the oriented boundary contour Γ is defined as $[\![\mu(\mathbf{x}^c)]\!] := \mu(\mathbf{x}^c - 0) - \mu(\mathbf{x}^c + 0)$.

Since variations of the electric potential and temperature are independent and arbitrary under the vrespectively, are approximatedolume integral in (36) we get the following governing equations

$$
\begin{aligned}
\lambda_{i,i} - m_{ik,ik} &= w \\
J_{i,i} &= 0.
\end{aligned} \tag{39}
$$

The rate of work of the external "forces" $(\bar{\Lambda}, \bar{P}, \bar{Q})$ and body source is given by

$$
\delta W = \int_{\Gamma_\Lambda} \bar{\Lambda}\delta\theta d\Gamma + \int_{\Gamma_P} \bar{P}\delta p d\Gamma + \int_{\Gamma_Q} \bar{Q}\delta\phi d\Gamma + \int_V w\delta\theta dV. \tag{40}
$$

Next, only the Joule heating plays the role of heat sources,

$$\delta W = \int_{\Gamma_A} \bar{\Lambda}\delta\theta d\Gamma + \int_{\Gamma_P} \bar{P}\delta p d\Gamma + \int_{\Gamma_Q} \bar{Q}\delta\phi d\Gamma + \int_V E_i J_i \delta\theta dV, \qquad (41)$$

and the governing Eqs. (39) become

$$\lambda_{i,i}(\mathbf{x}) - m_{ik,ik}(\mathbf{x}) - E_i(\mathbf{x})J_i(\mathbf{x}) = 0, J_{i,i}(\mathbf{x}) = 0. \qquad (42)$$

From the weak form (36) one can also deduce the boundary conditions for this thermoelectric problem:

essential b.c.:

$$\begin{aligned}
\theta(\mathbf{x}) &= \bar{\theta}(\mathbf{x}) \text{ on } \Gamma_\theta, \Gamma_\theta \subset \Gamma \\
p(\mathbf{x}) &= \bar{p}(\mathbf{x}) \text{ on } \Gamma_p, \Gamma_p \subset \Gamma \\
\phi(\mathbf{x}) &= \bar{\phi}(\mathbf{x}) \text{ on } \Gamma_\phi, \Gamma_\phi \subset \Gamma.
\end{aligned} \qquad (43)$$

natural b.c:

$$\begin{aligned}
\Lambda(\mathbf{x}) &= \bar{\Lambda}(\mathbf{x}) \text{ on } \Gamma_\Lambda, \Gamma_\Lambda \cup \Gamma_\theta = \Gamma, \Gamma_\Lambda \cap \Gamma_\theta = \varnothing \\
P(\mathbf{x}) &= \bar{P}(\mathbf{x}) \text{ on } \Gamma_P, \Gamma_P \cup \Gamma_p = \Gamma, \Gamma_P \cap \Gamma_p = \varnothing \\
Q(\mathbf{x}) &= \bar{Q}(\mathbf{x}) \text{ on } \Gamma_Q, \Gamma_Q \cup \Gamma_\phi = \Gamma, \Gamma_Q \cap \Gamma_\phi = \varnothing.
\end{aligned} \qquad (44)$$

Substituting the constitutive equations into the governing ones (42), the non-linear system of the partial differential equations for primary fields θ and ϕ is obtained

$$\kappa_{ij}(1 - l^2\nabla^2)\theta_{,ij} + \zeta_{ij}T(\phi_{,ij} + \alpha\theta_{,ij}) + \zeta_{ij}\theta_{,i}(\phi_{,j} + \alpha\theta_{,j}) + \phi_{,i}(s_{ij}\phi_{,j} + \zeta_{ij}\theta_{,j}) = 0,$$
$$s_{ij}\phi_{,ij} + \zeta_{ij}\theta_{,ij} = 0. \qquad (45)$$

3.1 The MLPG Formulation in Thermoelectricity

The local weak-form of the first governing Eq. (42) is given as

$$\int_{\Omega^q} [\lambda_{i,i}(\mathbf{x}) - m_{ik,ik}(\mathbf{x}) - E_i J_i] u^*(\mathbf{x}) d\Omega = 0, \qquad (46)$$

where $u^*(\mathbf{x})$ is a test function.

Again, the test function can be arbitrary and we have selected a Heaviside step function

$$u^*(\mathbf{x}) = \begin{cases} 1 & \text{at} \ \ \mathbf{x} \in \Omega^q \cup \partial\Omega^q \\ 0 & \text{at} \ \ \mathbf{x} \notin \Omega^q \cup \partial\Omega^q \end{cases}. \tag{47}$$

Applying the Gauss divergence theorem to two domain integrals in (46) and considering test function (47) the local integral equation is obtained

$$\int\limits_{L^q + \Gamma_\theta^q} n_i\big(\Lambda_i - m_{ik,k}\big) d\Gamma + \mu(\mathbf{x}_\Lambda^f) - \mu(\mathbf{x}_\Lambda^s) - \int\limits_{\Omega^q} E_i J_i \, d\Omega = -\int\limits_{\Gamma_\Lambda^q} \bar{\Lambda} d\Gamma \tag{48}$$

where $\mathbf{x}_\Lambda^f$, $\mathbf{x}_\Lambda^s$ stand for the final and starting points on Γ_Λ^q with prescribed heat flux.

The local integral equation for the second governing Eq. (42) is given as

$$\int\limits_{L^q + \Gamma_\phi^q} n_i J_i d\Gamma = -\int\limits_{\Gamma_Q^q} \bar{Q} d\Gamma. \tag{49}$$

The MLS approximation is applied to numerical solution of the above local integral equations. The spatial approximation of temperature and electric potential are given by [23]

$$\theta^h(\mathbf{x}) = \sum_{a=1}^n N^a(\mathbf{x})\hat{\theta}^a, \qquad \phi^h(\mathbf{x}) = \sum_{a=1}^n N^a(\mathbf{x})\hat{\phi}^a, \tag{50}$$

where $\hat{\theta}^a$ and $\hat{\phi}^a$ are fictitious nodal parameters for temperature and electric potential, respectively.

The heat flux λ_i and the higher-grade heat flux m_{ik} defined in (30) and (31), respectively, are approximated by

$$\boldsymbol{\lambda}^h(\mathbf{x}) = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}^h = -\bar{\kappa}\sum_{a=1}^n \mathbf{P}^a(\mathbf{x})\hat{\theta}^a - \bar{\mathbf{Z}}\sum_{a=1}^n \mathbf{P}^a(\mathbf{x})\hat{\phi}^a,$$

$$\mathbf{m}_k^h(\mathbf{x}) = -l^2\boldsymbol{\kappa}\sum_{a=1}^n \mathbf{P}_{,k}^a(\mathbf{x})\hat{\theta}^a \quad . \tag{51}$$

where

$$\mathbf{P}^a(\mathbf{x}) = \begin{bmatrix} N_{,1}^a \\ N_{,2}^a \end{bmatrix}, \ \mathbf{P}_{,k}^a(\mathbf{x}) = \begin{bmatrix} N_{,1k}^a \\ N_{,2k}^a \end{bmatrix}. \tag{52}$$

Finally, the electric current density and intensity of electric field are approximated by

$$\mathbf{J}^h(\mathbf{x}) = -\mathbf{S}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\phi}^a - \mathbf{Z}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\theta}^a, \mathbf{E}^T = -\sum_{c=1}^{n}\mathbf{P}^{cT}(\mathbf{x})\hat{\theta}^c. \quad (53)$$

Substituting the above MLS-approximation into the LIEs (48) and (49), we obtain the nonlinear system of algebraic equations

$$\int\limits_{L^q+\Gamma_\theta^q}\left[\left(l^2\mathbf{F}_\theta(\mathbf{x})\sum_{a=1}^{n}\mathbf{P}_{,kk}^a(\mathbf{x}) - \bar{\mathbf{F}}_\theta(\mathbf{x})\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\right)\hat{\theta}^a - \bar{\mathbf{F}}_\phi(\mathbf{x})\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\phi}^a\right]d\Gamma$$

$$+ l^2\left(n_k(\mathbf{x}_A^s)\mathbf{F}_\mu(\mathbf{x}_A^s)\sum_{a=1}^{n}\mathbf{P}_{,k}^a(\mathbf{x}_A^s) - n_k(\mathbf{x}_A^f)\mathbf{F}_\mu(\mathbf{x}_A^f)\sum_{a=1}^{n}\mathbf{P}_{,k}^a(\mathbf{x}_A^f)\right)\hat{\theta}^a$$

$$- \int\limits_{\Omega^q}\sum_{c=1}^{n}\hat{\theta}^c\mathbf{P}^{cT}(\mathbf{x})\left(\mathbf{S}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\phi}^a + \mathbf{Z}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\theta}^a\right)d\Omega = -\int\limits_{\Gamma_A^q}\bar{A}d\Gamma,$$

$$(54)$$

$$\int\limits_{L^q+\Gamma_\phi^q}\mathbf{N}(\mathbf{x})\left(\mathbf{S}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\phi}^a + \mathbf{Z}\sum_{a=1}^{n}\mathbf{P}^a(\mathbf{x})\hat{\theta}^a\right)d\Gamma = \int\limits_{\Gamma_Q^q}\bar{Q}d\Gamma. \quad (55)$$

This nonlinear system has to be solved iteratively.

Fig. 5 Geometry and boundary conditions

θ=400 [K]
φ=0 [V]

θ=300 [K]
φ=0 [V]

r1

r2
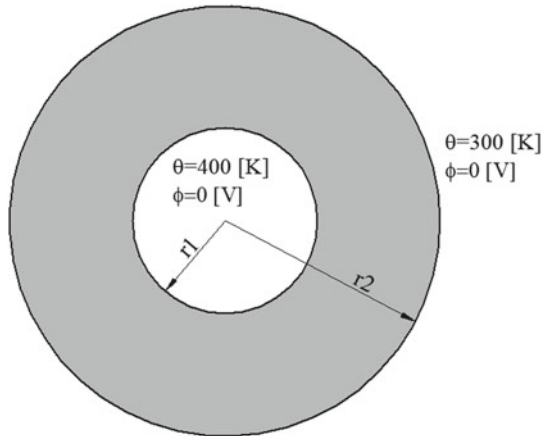
3.2 Numerical Examples

A hollow cylinder as axially symmetric problem is analysed (see Fig. 5). The thermoelectric material Bi_2Te_3, is considered [35]:

$$s = 1.1 \times 10^5 \text{ Am/V}, \alpha = \zeta/s = 2 \times 10^{-4} \text{ V}^2/\text{KAm}, \kappa = 1.6 \text{ W/Km}. \quad (56)$$

Characteristic length for the selected material structure is $l = 5 \times 10^{-9}$ m.

The hollow cylinder has following geometry: internal radius r_1 = 1e−7 m and external radius r_2 = 2.5e−7 m. The electric potentials are prescribed on both inner and external surfaces with vanishing value. Temperature gradients on both surfaces are vanishing too.

The influence of the hollow cylinder thickness $L = r_2 - r_1$ on temperature and induced electric potential is investigated. Variations of the temperature and electric potential along the non dimensional r/L radius in hollow cylinder are presented in Figs. 6 and 7. One can observe that induced electric potential grows with increasing the value l/L.

In classical thermoelectricity, $l = 0$, it is possible to get analytical solution. It is convenient to introduce the dimensionless radial coordinate $r^* = r/L$. Then, the governing equations become

$$\left(\partial_r^{*2} + \frac{1}{r^*} \partial_r^* \right)(\phi + \alpha\theta) = 0, \tilde{J}_r = -s\partial_r^*(\phi + \alpha\theta) = LJ_r, \partial_r^* = \partial/\partial r^*$$

$$\tilde{J}_r \partial_r^*(\phi + \alpha\theta) - \kappa \left(\partial_r^{*2} + \frac{1}{r^*} \partial_r^* \right)\theta = 0. \quad (57)$$

Following boundary conditions are considered

$$\phi(r_1^*) = \phi_1, \phi(r_2^*) = \phi_2, \theta(r_1^*) = T_1 - T_0, \theta(r_2^*) = T_0 - T_0 = 0.$$

The analytical solution for the classical thermoelectricity is given by

$$\phi(r^*) = a_0 + a_1 \ln r^* - \alpha\theta(r^*)$$

$$\theta(r^*) = -\frac{sa_1^2}{2\kappa} (\ln r^*)^2 + b_1 \ln r^* + b_0, \quad (58)$$
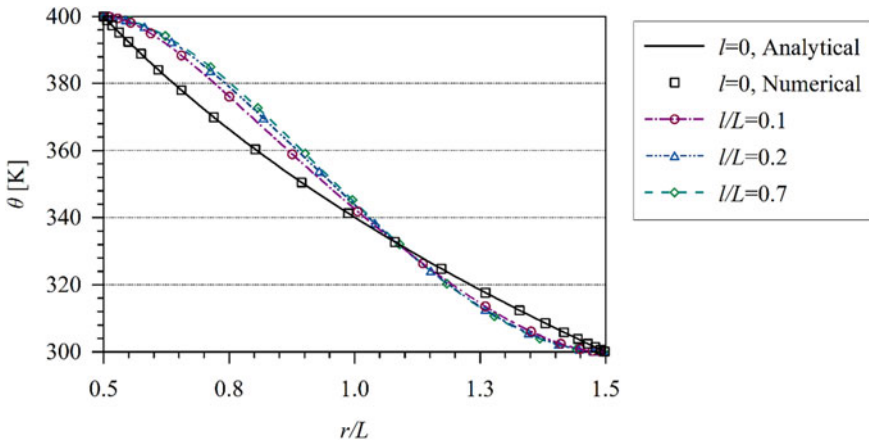
Fig. 6 Temperature variation along non dimensional *x/L* coordinate in hollow cylinder
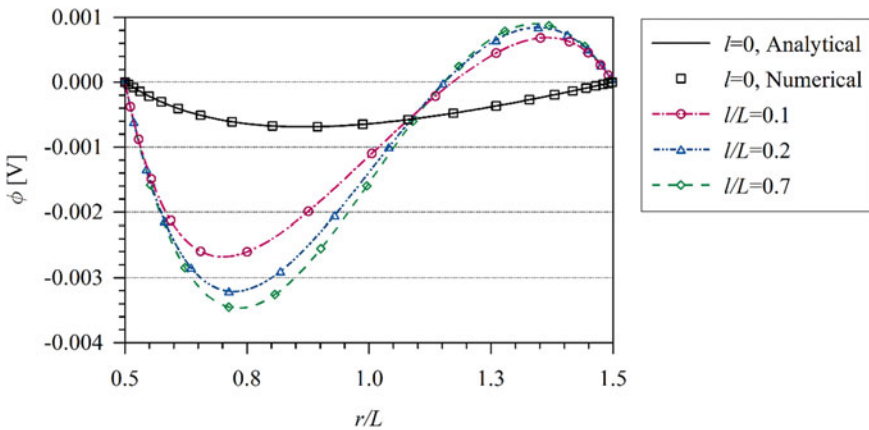
Fig. 7 The electric potential variation along non dimensional x/L coordinate in hollow cylinder

where

$$a_0 = \left\{ [\phi_1 + \alpha(T_1 - T_0)] \ln r_2^* - \phi \ln r_1^* \right\} / ln\left(r_2^*/r_1^*\right),$$

$$a_1 = -\frac{\phi_1 - \phi_2 + \alpha(T_1 - T_0)}{ln\left(r_2^*/r_1^*\right)},$$

$$b_0 = \left[-\frac{sa_1^2}{2\kappa} \ln r_1^* + \frac{T_1 - T_0}{\ln(r_2^*/r_1^*)} \right] \ln r_2^*, \quad b_1 = \frac{T_0 - T_1 + \frac{sa_1^2}{2\kappa}\left[\left(\ln r_2^*\right)^2 - \left(\ln r_1^*\right)^2\right]}{\ln(r_2^*/r_1^*)}.$$

A very good agreement of results is received for this special case.

4   Conclusions

The meshless Petrov-Galerkin method has been developed for general 2-d boundary value problems in flexoelectricity and thermoelectricity described by gradient theories. The size effect is considered by strain and electric intensity vector gradients in direct and converse flexoelectricity. Similarly, the novel gradient theory for thermoelectric materials is developed here. Both gradient theories allow more realistic description of physical processes in nano-sized structures. Governing equations in advanced continuum models contain higher order derivatives of physical fields. Application of standard domain discretization methods, like the FEM, to corresponding boundary value problems brings some difficulties with continuity of approximated higher derivatives of field variables.

The continuity of gradients of primary fields in the MLS approximation is satisfied very easily. Therefore, the proposed MLPG computational method with the MLS approximation of fields is very convenient to solve problems with high-order derivatives. The present computational method is promising to be applied to multiphysical problems described by gradient theories.

References

1. Buhlmann, S., Dwir, B., Baborowski, J., Muralt, P.: Size effects in mesoscopic epitaxial ferroelectric structures: increase of piezoelectric response with decreasing feature-size. Appl. Phys. Lett. **80**, 3195–3197 (2002)
2. Cross, L.E.: Flexoelectric effects: charge separation in insulating solids subjected to elastic strain gradients. J. Mater. Sci. **41**, 53–63 (2006)
3. Harden, J., Mbanga, B., Eber, N., Fodor-Csorba, K., Sprunt, S., Gleeson, J.T., Jakli, A.: Giant flexoelectricity of bent-core nematic liquid crystals. Phys. Rev. Lett. **97**, 157802 (2006)
4. Zhu, W., Fu, JY., Li, N., Cross, LE.: Piezoelectric composite based on the enhanced flexoelectric effects. Appl. Phys. Lett. **89**, 192904 (2006)
5. Catalan, G., Lubk, A., Vlooswijk, AHG., Snoeck, E., Magen, C., Janssens, A., Rispens, G., Rijnders, G., Blank, D.H.A., Noheda, B.: Flexoelectric rotation of polarization in ferroelectric thin films. Nat. Mater. **10**, 963–967 (2011)
6. Kogan, S.M.: Piezoelectric effect during inhomogeneous deformation and acoustic scattering of carriers in crystals. Sov. Phys. Solid State **5**, 2069–2070 (1964)
7. Meyer, R.B.: Piezoelectric effects in liquid crystals. Phys. Rev. Lett. **22**, 918–921 (1969)
8. Sharma, P., Maranganti, R., Sharma, N.D.: On the possibility of piezoelectric nanocomposites without using piezoelectric materials. J. Mech. Phys. Solids **55**, 2328–2350 (2007)
9. Deng, Q., Liu, L., Sharma, P.: Flexoelectricity in soft materials and biological membranes. J. Mech. Phys. Solids **62**, 209–227 (2014)
10. Landau, LD., Lifshitz, EM.: Electrodynamics of Continuous Media, pp. 358–371, 2nd edn. Butterworth-Heinemann (1984)
11. Yang, J.S.: Effects of electric field gradient on an antiplane crack in piezoelectric ceramics. Int. J Fracture **127**, L111–L116 (2004)

12. Yang, X.M., Hu, Y.T., Yang, J.S.: Electric field gradient effects in anti-plane problems of polarized ceramics. Int. J. Solids Struct. **41**, 6801–6811 (2004)
13. Hicks, L., Dresselhaus, M.S.: Thermoelectric figure of merit of a one-dimensional conductor. Phys. Rev. B **47**, 16631 (1993)
14. Minnich, A.J., Dresselhaus, M.S., Ren, Z.F., Chen, G.: Bulk nanostructured thermoelectric materials: current research and future prospects. Energy Env. Sci. **2**, 466–479 (2009)
15. Eringen, A.C.: Non-local polar field theory. In: Eringen, A.C. (ed.) Continuum Physics, vol. 4. Academic Press, New York (1976)
16. Mindlin, RD.: Micro-structure in linear elasticity. Arch. Ration Mech. Anal. **16**, 51–78 (1964)
17. Mindlin, R.D.: Second gradient of strain and surface-tension in linear elasticity. Int. J. Solids Struct. **1**, 417–438 (1965)
18. Aifantis, E.: On the microstructural origin of certain inelastic models. ASME J. Eng. Mater. Technol. **106**, 326–330 (1984)
19. Allen, P.B.: Size effects in thermal conduction by phonons. Physical Rev. B **90**, 054301 (2014)
20. Deng, F., Deng, Q., Yu, W., Shen, S.: Mixed finite elements for flexoelectric solids. J. Appl. Mech. **84**, 081004 (2017)
21. Dong, L., Atluri, S.N.: A simple procedure to develop efficient & stable hybrid/mixed elements, and Voronoi cell finite elements for macro- & micromechanics. CMC Comput. Mater. Continua **24**, 61–104 (2011)
22. Atluri, S.N.: The Meshless Method (MLPG) for Domain and BIE Discretizations. Tech. Science Press, Forsyth (2004)
23. Sladek, J., Stanak, P., Han, Z.D., Sladek, V., Atluri, S.N.: Applications of the MLPG method in engineering & Sciences: a review. CMES Comput. Model. Eng. Sci. **92**, 423–475 (2013)
24. Sladek, J., Sladek, V., Repka, M., Kasala, J., Bishay, P.: Evaluation of effective material properties in magneto-electro-elastic composite materials. Compos. Struct. **109**, 176–186 (2017)
25. Sladek, J., Sladek, V., Jus, M.: The MLPG for crack analyses in composites with flexoelectric effects. Compos. Struct. **204**, 105–113 (2018)
26. Sladek, J., Sladek, V., Wen, P.H.: The meshless analysis of scale-dependent problems for coupled fields. Materials **13**, 2527 (2020)
27. Maranganti, R., Sharma, N.D., Sharma, P.: Electromechanical coupling in nonpiezoelectric materials due to nanoscale nonlocal size effects: Green's function solutions and embedded inclusions. Phys. Rev. B **74**, 014110 (2006)
28. Hu, SL., Shen, SP.: Electric field gradient theory with surface effect for nano-dielectrics. CMC Comput. Mater. Continua **13**, 63–87 (2009)
29. Altan, S., Aifantis, E.: On the structure of the mode III crack-tip in gradient elasticity. Scripta Metall. Mater. **26**, 319–324 (1992)
30. Askes, H., Aifantis, E.: Gradient elasticity in statics and dynamics: An overview of formulations, length scale identification procedures, finite element implementations and new results. Int. J. Solids Struct. **48**, 1962–1990 (2011)
31. Gitman, I., Askes, H., Kuhl, E., Aifantis, E.: Stress concentrations in fractured compact bone simulated with a special class of anisotropic gradient elasticity. Int. J. Solids Struct. **47**, 1099–1107 (2010)
32. Yaghoubi, S.T., Mousavi, S.M., Paavola, J.: Buckling of centrosymmetric anisotropic beam structures within strain gradient elasticity. Int. J. Solids Struct. **109**, 84–92 (2017)
33. Sladek, J., Sladek, V., Wunsche, M., Zhang, Ch.: Effects of electric field and strain gradients on cracks in piezoelectric solids. Eur. J. Mech./A Solids **71**, 187–198 (2018)
34. Sladek, J., Sladek, V., Repka, M., Schmauder, S.: Mixed FEM for quantum nanostructured solar cells. Compos. Struct. **229**, 111460 (2019)
35. Yang, Y., Gao, C., Li, J.: The effective thermoelectric properties of core-shell composites. Acta Mech. **225**, 1211–1222 (2014)
36. Auld, B.A.: Acoustic Fields and Waves in Solids, pp. 357–382. Wiley, New York (1973)

Rupture Failure of Tubing and Casing with a Crack in Mixed Modes

Xuelin Dong and Zhiyin Duan

Abstract Extreme downhole conditions often apply a high pressure on tubing or casing that poses a severe challenge to well integrity. Imperfections in these tubular elements are very prone to cause rupture under high pressures. The American Petroleum Institute (API) proposed a method to calculate the rupture pressure for the tubing or casing with radial cracks perpendicular to their inner surface. Stress analysis indicates that an inclined crack would propagate in mixed modes. This study presents an investigation on the stress intensity factor at an oblique fracture initiating from the inner surface of tubing based on finite element methods. The simulation shows that the shear mode would dominate the crack extension as the incline angle increases. Both stress intensity factors in traction and shear modes rise with the pressure inside the tubing. Thicker tubing would mitigate the stress intensity factor. A conservative criterion for crack propagation in mixed modes provides the minimum pressure under which the inclined crack would grow. This criterion also estimates a burst strength factor used for rupture pressure prediction. Results show that the burst strength factor increases firstly and then decreases with the incline angle. Its maximum value reaches 3.2 and occurs at about 20°, which is much higher than the default value by API. This work demonstrates the necessity of considering the crack propagation in mixed modes in rupture failure of tubing.

Keywords Rupture failure · Stress intensity factor · Mixed mode

X. Dong (✉)
Key Laboratory of Petroleum Engineering, China University of Petroleum,
Beijing 102249, China
e-mail: dongxl@cup.edu.cn

Z. Duan
Beijing Key Lab of Heating, Gas Supply, Ventilating and Air Conditioning Engineering,
Beijing University of Civil Engineering and Architecture, Beijing 100044, China

1 Introduction

Most unconventional oil and gas resources are associated with extreme downhole conditions such as high pressure and high temperature (HPHT) [1, 2]. According to the American Petroleum Institution (API), an HPHT well has a subsurface condition with pressure or temperature greater than 103 MPa or 177 °C, respectively [1]. This harsh environment severely challenges the safety and efficiency of the well barriers including tubing, casing, and cement. Therefore, well integrity is crucial to the successful exploitation of unconventional resources.

Tubing or casing is inevitable to contain imperfections like small cracks. These cracks could extend to lead rupture in pipes under extreme downhole conditions. API has provided a mechanical method to predict the rupture pressure for the pipes with imperfections [3]. However, it still needs elaborate considerations to calculate the rupture pressure for pipes. In this work, we investigate the propagation of an inclined crack inside the tubing. Using finite element methods, we estimate the stress intensity factors at the crack tip both in traction and shear modes. This study would help engineers design a better well construction.

2 Rupture Pressure Based on Fracture Mechanics

2.1 Stress State in a Cylinder Under Pressure

For a thick-walled cylinder such as tubing or casing, it is conventional to apply the Lamé equation to calculate stress components in the cylinder under pressure [4]. Figure 1(a) shows the distribution of the hoop stress (σ_θ) in the cross-section of tubing or casing under inner pressure. D is the cylinder's outer diameter, and t is its thickness. The highest tensile hoop stress occurs on the inner surface of the cylinder. This stress distribution makes a crack initiating at the inner surface most
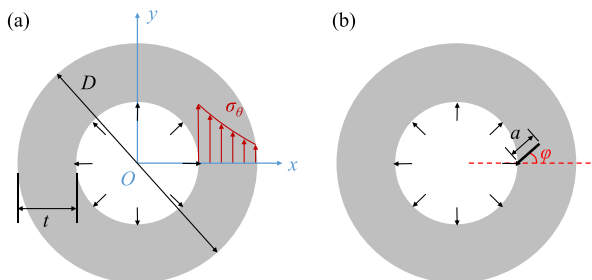

Fig. 1 A cylinder under inner pressure: (a) the distribution of the hoop stress, (b) an inclined crack initiating from the inner surface

likely to extend as shown in Fig. 1(b). The crack length is denoted by a, and the angle between the crack and the horizontal line is φ.

2.2 API Recommendation

The API technical report recommends a method to predict the rupture pressure as [3]:

$$\frac{p_{ir}}{\sigma_{tm}} = \frac{2k_h\left(k_w - k_c\frac{a}{t}\right)}{\frac{D}{t} - \left(k_w - k_c\frac{a}{t}\right)} \tag{1}$$

where p_{ir} represents the minimum inner pressure that could lead to burst, σ_{tm} is the minimum tensile strength of tubing or casing, k_h is a correction factor related to pipe deformation and material strain hardening, k_w is a coefficient accounting for the manufacturing tolerance of the pipe wall, and k_c is a factor caused by the existent of imperfection. API refers k_c as to the burst strength factor and suggests a default value of 2.0 without measurement. Figure 2 shows the variation of the rupture pressure with k_c for different crack sizes. The rupture pressure decreases with k_c and drops nearly 50% when k_c increases from 0 to 3.0 for $a/t = 0.125$.

Equation 1 demonstrates that k_c has a strong influence on the rupture pressure. API also provides an analytical method to estimate the value of k_c. Reference [3] suggests constructing a finite element model of pipe with three crack depths of a/t: 0, 5%, and 12.5%. Calculate the J-integral in the model with inner pressure until it reaches the experimentally measured value of J_{Ic}. Set the ratio of the pressure for 5% crack to the pressure with no crack equal to $(1-k_c a/t)$ and solve the equation to obtain k_c. Then repeat the same procedure for the 12.5% crack. At last, k_c is the average of these two values.

Fig. 2 The effect of the burst strength factor on the rupture pressure
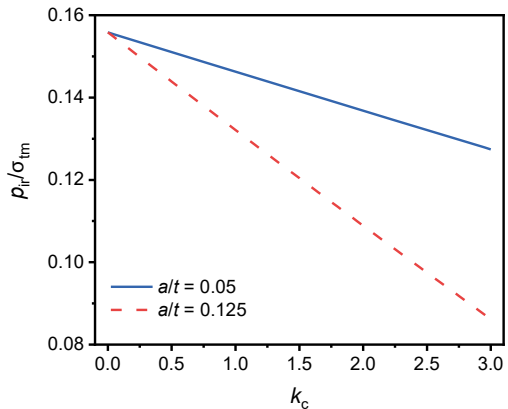
Fig. 3 A schematic of criteria for crack propagation in mixed modes
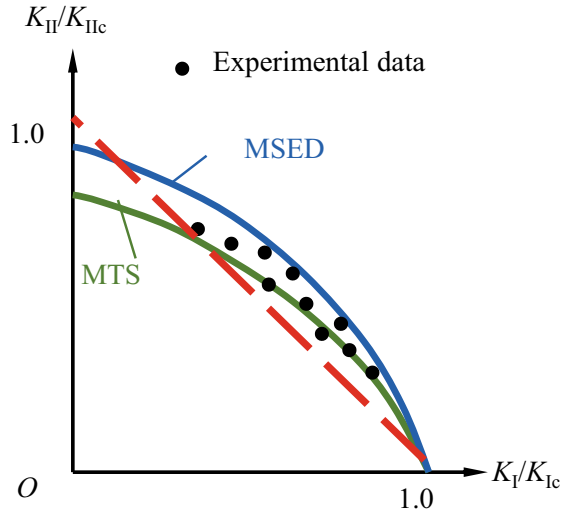
2.3 Crack Propagation in Mixed Modes

The above estimation for k_c assumes that the crack is perpendicular to the pipe
surface, i.e. $\varphi = 0$. With a non-zero inclined angle, the crack would propagate in
mixed modes. There are several criteria to characterize crack extension in mixed
modes. The minimum strain energy density (MSED) criterion and the maximum
tangential stress (MTS) criterion are two prevalent criteria based on linear elastic
fracture mechanics (LEFM) [5]. MSED supposes the crack would extend along the
angle at which the strain energy density factor (S) is minimum [5]:

$$\left.\frac{\partial S}{\partial \beta}\right|_{\beta=\beta_f} = 0, \left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta=\beta_f} > 0 \tag{2}$$

where β represents the deviation from the original crack, and β_f the is crack
extension angle. For a crack in mixed I-II loading (I and II denote traction and
shear, respectively), S is expressed as [5]:

$$S = a_{11}K_I^2 + a_{12}K_I K_{II} + a_{22}K_{II}^2 \tag{3}$$

where K_I and K_{II} represent the stress intensity factors in Mode-I and Mode-II,
respectively, and a_{11}, a_{22}, and a_{12} are functions related to β.

Fig. 4 The variation of $K_I + K_{II}$ with the inner pressure: (a) $a/t = 0.05$, (b) $a/t = 0.125$
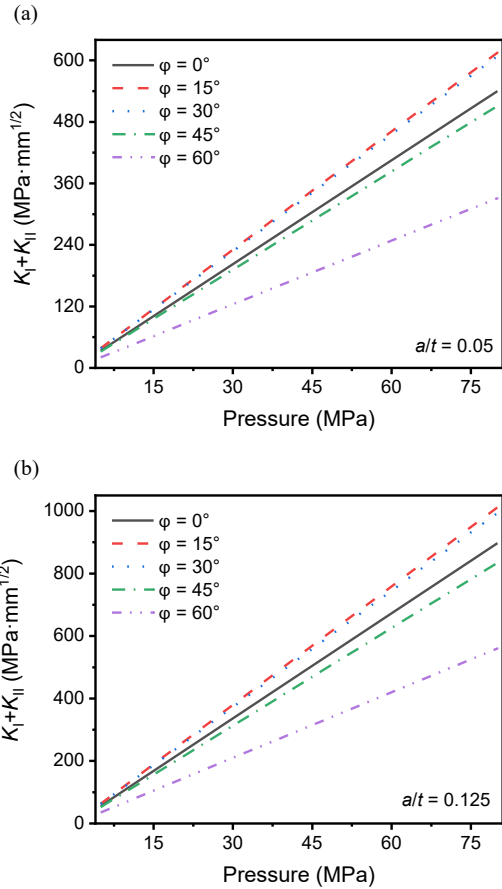
MTS criterion assumes that the crack will propagate along with the maximum hoop stress as [5]:

$$\left.\frac{\partial \sigma_\beta}{\partial \beta}\right|_{\beta=\beta_0} = 0, \left.\frac{\partial^2 \sigma_\beta}{\partial \beta^2}\right|_{\beta=\beta_0} < 0 \tag{4}$$

where σ_β is the local hoop stress near the crack tip and β_0 is the angle at which σ_β increases fastest. Equation (4) requires the combination of the stress intensity factors to meet K_{Ic} (the critical stress intensity factor in Mode-I, and is equivalent to J_{Ic} under linear elastic conditions) [5]:

$$\cos \frac{\beta_0}{2}\left[K_I \cos^2 \frac{\beta_0}{2} - \frac{3}{2}K_{II} \sin \beta_0\right] = K_{IC} \tag{5}$$

Figure 3 shows a schematic of the two criteria in which K_{IIc} is the critical stress intensity factor in Mode-II. It indicates that both criteria have deviated from experimental data [6]. Hence, it is conservative to draw a straight line under both criteria (the red dash line in Fig. 3). This criterion has the form of:

$$K_I + K_{II} = K_{IC} \tag{6}$$

We will use Eq. (6) to estimate the rupture pressure for an oblique crack in the pipe under inner pressure and calculate the burst strength factor with different inclined angles.

3 Results

In this section, we present the stress intensity factors of an inclined crack with different pressures and thicknesses using finite element methods. The outer diameter of the pipe is 88.9 mm, which is a typical size for tubing. During the simulation, we change the pressure or the tubing thickness to investigate the mechanical behavior of the inclined crack under inner pressure. To estimate the burst strength factor, the rupture pressure is obtained by fitting the pressure as a function of $K_I + K_{II}$.

From LEFM [6], the stress intensity factor is proportional to the applied load, as shown in Fig. 4. It is worth noting that the values of $K_I + K_{II}$ at 15° and 30° are higher than that of a straight crack. Figure 4(b) infers that a longer fracture would induce a higher stress intensity factor, but its trend with the crack angle is the same as Fig. 4(a).

Figure 5 presents the change of $K_I + K_{II}$ with the tubing thickness. Intuitively, the stress intensity factor drops as the tubing becomes thicker (D/t decreases). Unlike the pressure, lines in Fig. 5 are non-linear.

Finally, we present the burst strength factor with different crack angles, as illustrated in Fig. 6. k_c increases first and then decreases as φ rises, which is consistent with Figs. 4 and 5. The most notable is that k_c would reach a maximum value of 3.2 when φ is about 20°. From Fig. 2, this maximum value of k_c could reduce the rupture pressure by 50% nearly. It is necessary to consider crack propagation in mixed modes to predict the rupture pressure during well construction design.

Fig. 5 The change of $K_{I} + K_{II}$ with the tubing thickness: (a) $a/t = 0.05$, (b) $a/t = 0.125$
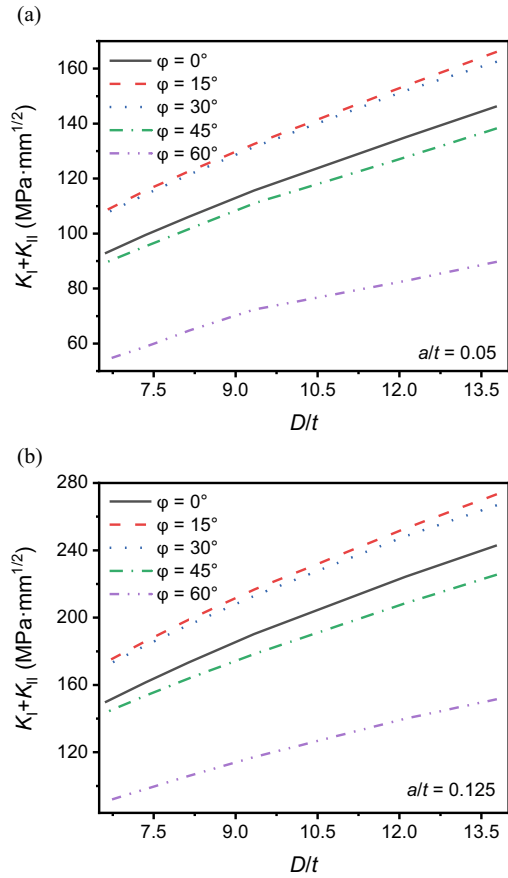
Fig. 6 The estimation of k_c with different crack angles
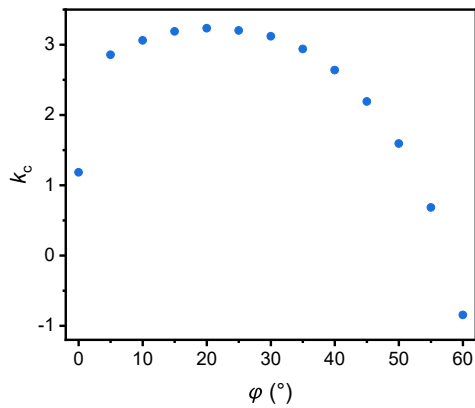
4   Conclusions

In this study, we present an analysis of the stress intensity factor at an inclined crack in the pipe under inner pressure. Based on finite element methods, we show that both the stress intensity factors in traction and shear modes increase linearly with the applied pressure. The estimated burst strength factor based on a conservative criterion for crack propagation in mixed modes fluctuates with the crack angle. A 20° of the crack angle would produce the maximum factor of 3.2. This study is conducive to maintain well integrity under extreme downhole conditions.

References

1. Wilcox, B., Oyeneyin, B., Islam, S.: HPHT well integrity and cement failure. In: SPE Nigeria Annual International Conference and Exhibition, 2–4 August, Lagos, Nigeria (2016)
2. Phi, T., Ramadan, M.A., Ahmed, R., Teodoriu, C.: Well integrity issues: Extreme high-pressure high-temperature wells and geothermal wells a review. In: SPE Thermal Well Integrity and Design Symposium, 19–21 November, Banff, Alberta, Canada (2019)
3. Calculating Performance Properties of Pipe Used as Casing or Tubing. API Technical Report 5C3. 7th edn. (2018)
4. Sadd, M.: Elasticity Theory Applications and Numerics. Academic Press, Burlington (2009)
5. Wang, Y., Wang, W., Zhang, B., Li, C.-Q.: A review on mixed mode fracture of metals. Eng. Fract. Mech. **235**, 107126 (2020)
6. Perez, N.: Fracture Mechanics, 2nd edn. Springer, Switzerland (2017)

Analysis for the Coupled Thermo-Hydro-Mechanical Process of Hydraulic Fracturing in Shale Gas Extraction

Haiyang Wen

Abstract Shale gas, a kind of natural gas from unconventional reservoirs, can be economically produced due to the application of hydraulic fracturing. This stimulation provides a technically feasible solution to improve the fracture conductivity in most shale formations. The degree of crack extension is a fundamental controlling factor in the efficient exploitation of shale gas. The shale can be treated as porous media. This paper presents a coupled thermo-hydro-mechanical (THM) model to simulate the intact rock and the cracked rock with a single joint respectively. The existence of the joint has a significant effect on the performance of hydraulic fracturing. The governing equations are derived on the basis of the momentum, mass and energy conservation after the mechanism of the three-field coupling is analyzed. The investigation is carried out for the variations of maximum principal stress, seepage pressure and temperature under the designed operational conditions. The numerical results between the intact and cracked rock are compared.

1 Introduction

As a new source of energy, shale gas is the unconventional natural gas mainly occurring within or extracted from dark or high-carbon shale formations. The efficient exploitation of shale gas increases energy supply, but also has important significance for the sustainable development of energy. However, low porosity and low permeability are both the physical properties of shale gas reservoirs [1–4] where the gas flow resistance is usually greater than that of conventional reservoirs. Only a few shale gas wells with well-developed cracks may be put into production

H. Wen (✉)
Department of Geotechnical Engineering, Tongji University, Shanghai 200092, China
e-mail: 1410258@tongji.edu.cn

directly, and crucial stimulations (e.g. hydraulic fracturing) should be taken to improve the fracture conductivity in most shale formations. Therefore, the degree of crack extension is one of the primary controlling factors in the economical exploitation of shale gas. The hydraulic fracturing method [2, 5–12] is that the high-pressure fluid containing various additives is injected into the formations to extend natural or induced fracture networks, with proppant holding cracks open. This treatment can be regarded as a coupled THM process [4, 7, 12–19] involving rock deformation, crack propagation, seepage in rock mass, ground temperature variation, etc.

In this work, three fields (i.e. mechanical, hydraulic and thermal fields) are considered to simulate the hydraulic fracturing technique of shale gas extraction. Section 2 describes the governing equations for the momentum, mass and energy conservation. In Sect. 3, the coupled THM model is applied to investigate the intact and cracked rock performances. In Sect. 4, conclusions are given.

2 Governing Equations of the Three-Field Coupling

The interaction between the mechanical field and the hydraulic field is strong. The hydraulic field has a great effect on the thermal field, and the influence of temperature on the physical, mechanical and thermal properties of groundwater cannot be ignored. The coupling between the thermal field and the mechanical field is basically unidirectional, that is, the mechanical field may be strongly affected by the thermal field, whereas the influence of the mechanical field on the thermal field is very weak and can even be neglected within the range of elastic deformation. In general, the full THM coupling is not necessary from an engineering perspective. Hence, the following assumptions in the present three-field coupling are proposed.

a. The rock mass is treated as the small-deformation, homogeneous, isotropic and elastic porous media; the mechanical parameters of the rock mass are independent of temperature.
b. The solid particles are incompressible; the pores are compressible.
c. The pores are saturated by the single-phase fluid (i.e. groundwater).
d. The seepage satisfies Darcy's law; the heat conduction follows Fourier's law.
e. The heat capacity and conductivity coefficients of all the media are independent of temperature and pressure.

The governing equations in next subsections will be derived based on the above assumptions.

2.1 Mechanical Field: Momentum Conservation

The static equilibrium of the porous media is expressed as

$$\nabla \cdot \sigma_{ij} + \rho f_i = 0, \tag{1}$$

$$\rho = (1 - n)\rho_s + n\rho_l, \tag{2}$$

where ∇ is the Hamiltonian operator, σ_{ij} is the component of the stress tensor, ρ is the equivalent density of the rock mass, n is the pore ratio of the rock mass, ρ_s is the density of the rock skeleton, ρ_l is the density of groundwater, and f_i is the component of the body force vector. According to the principle of effective stress, we can get

$$\sigma_{ij} = \sigma'_{ij} + \alpha P \delta_{ij}, \tag{3}$$

where σ'_{ij} is the component of the effective stress tensor, α is the Biot's consolidation coefficient, P is the pore pressure, and δ_{ij} is the Kronecker symbol. By substituting Eq. (3) into Eq. (1), we can get

$$\nabla \cdot \left(\sigma'_{ij} + \alpha P \delta_{ij} \right) + \rho f_i = 0. \tag{4}$$

By considering the influence of temperature, we can adapt the relationship between effective stress and strain as

$$\sigma'_{ij} = 2\mu\varepsilon_{ij} + \lambda\delta_{ij}\varepsilon_{kk} - \beta\delta_{ij}(T_s - T_{s0}), \tag{5}$$

$$\varepsilon_{kk} = \varepsilon_x + \varepsilon_y + \varepsilon_z, \tag{6}$$

$$\beta = (2\mu + 3\lambda)\beta_s, \tag{7}$$

where λ and μ are the Lamé constants, ε_{ij} is the component of the strain tensor, β is the thermal stress coefficient, and ε_{kk}, β_s, T_s and T_{s0} are the volumetric strain, linear thermal expansion coefficient, temperature and initial temperature of the rock skeleton. Substituting Eq. (5) into Eq. (4) leads to the governing equation of the mechanical field as

$$\nabla \cdot \left[2\mu\varepsilon_{ij} + \lambda\delta_{ij}\varepsilon_{kk} - \beta\delta_{ij}(T_s - T_{s0}) + \alpha P \delta_{ij} \right] + \rho f_i = 0. \tag{8}$$

2.2 *Hydraulic Field: Mass Conservation*

The continuity equation of the groundwater seepage is described by

$$\frac{\partial(\rho_l n)}{\partial t} + \nabla \cdot (\rho_l n \mathbf{V}_l) = Q_m, \tag{9}$$

$$\mathbf{V}_l = \mathbf{V}_r + \mathbf{V}_s, \tag{10}$$

$$\mathbf{q} = n\mathbf{V}_r, \tag{11}$$

where Q_m is the mass source, $\mathbf{V}_l$ is the velocity vector of the fluid, $\mathbf{V}_s$ is the velocity vector of the rock skeleton, $\mathbf{V}_r$ is the relative velocity vector, and $\mathbf{q}$ is the seepage velocity vector.

Equation (9) can be further calculated as

$$\rho_l \frac{\partial n}{\partial t} + n\frac{\partial \rho_l}{\partial t} + \nabla \cdot (\rho_l n \mathbf{V}_r) + \rho_l n \nabla \cdot \mathbf{V}_s = \rho_l \frac{\partial n}{\partial t} + n\frac{\partial \rho_l}{\partial t} + \nabla \cdot (\rho_l \mathbf{q}) + \rho_l n \nabla \cdot \mathbf{V}_s = Q_m. \tag{12}$$

On the other hand, the continuity equation of the rock skeleton is described by

$$\frac{\partial[\rho_s(1-n)]}{\partial t} + \nabla \cdot [\rho_s(1-n)\mathbf{V}_s] = 0. \tag{13}$$

Equation (13) can be also further calculated as

$$\frac{\partial n}{\partial t} = (1-n)\nabla \cdot \mathbf{V}_s = (1-n)\frac{\partial}{\partial t}(\nabla \cdot \mathbf{U}_s) = (1-n)\frac{\partial}{\partial t}(\delta_{ij}\varepsilon_{ij}) = (1-n)\frac{\partial \varepsilon_{kk}}{\partial t}, \tag{14}$$

where $\mathbf{U}_s$ is the displacement vector of the rock skeleton. By substituting Eq. (14) into Eq. (12), we can get

$$\rho_l(1-n)\frac{\partial \varepsilon_{kk}}{\partial t} + n\frac{\partial \rho_l}{\partial t} + \nabla \cdot (\rho_l \mathbf{q}) + \rho_l n \frac{\partial \varepsilon_{kk}}{\partial t} = \rho_l \frac{\partial \varepsilon_{kk}}{\partial t} + n\frac{\partial \rho_l}{\partial t} + \nabla \cdot (\rho_l \mathbf{q})$$
$$= Q_m. \tag{15}$$

The state equation of the groundwater density is described by

$$\rho_l = \rho_{l0} + \rho_{l0}c_l(P - P_0) - \rho_{l0}\beta_l(T_l - T_{l0}), \tag{16}$$

where ρ_{l0}, c_l, P_0, β_l, T_l, and T_{l0} are the initial density, volumetric compression coefficient, initial pressure, thermal expansion coefficient, temperature and initial temperature of groundwater. From Darcy's law, we know

$$\mathbf{q} = -\frac{k}{\eta}\nabla P, \tag{17}$$

where k is the permeability coefficient, and η is the viscosity coefficient. Substituting Eq. (16) and (17) into Eq. (15) leads to the governing equation of the hydraulic field as

$$\rho_l\frac{\partial\varepsilon_{kk}}{\partial t} + n\rho_{l0}\left(c_l\frac{\partial P}{\partial t} - \beta_l\frac{\partial T_l}{\partial t}\right) - \nabla\cdot\left(\rho_l\frac{k}{\eta}\nabla P\right) = Q_m. \tag{18}$$

2.3 Thermal Field: Energy Conservation

The thermal equilibrium of the rock skeleton is expressed as

$$(1-n)\rho_s C_s\frac{\partial T_s}{\partial t} = \nabla\cdot[(1-n)\lambda_s\nabla T_s] - K_T(T_s - T_l), \tag{19}$$

where C_s and λ_s are the specific heat and thermal conductivity of the rock skeleton, and K_T is the heat exchange coefficient. The thermal equilibrium of groundwater is expressed as

$$n\rho_l C_l\frac{\partial T_l}{\partial t} = \nabla\cdot(n\lambda_l\nabla T_l) - \rho_l C_l\mathbf{q}\cdot\nabla T_l + K_T(T_s - T_l), \tag{20}$$

where C_l and λ_l are the specific heat and thermal conductivity of groundwater.

By combining Eq. (19) with Eq. (20), the energy conservation equation (i.e. the governing equation of the thermal field) can be obtained by

$$n\rho_l C_l\frac{\partial T_l}{\partial t} + (1-n)\rho_s C_s\frac{\partial T_s}{\partial t} = \nabla\cdot[n\lambda_l\nabla T_l + (1-n)\lambda_s\nabla T_s] - \rho_l C_l\mathbf{q}\cdot\nabla T_l + Q, \tag{21}$$

where Q is the heat source. Moreover, the local thermal equilibrium between the rock skeleton and groundwater is assumed, that is, they are at the same temperature T.

Equation (21) can be simplified as

$$(\rho C)_{eq}\frac{\partial T}{\partial t} = \nabla \cdot \left(\lambda_{eq}\nabla T\right) - \rho_l C_l \mathbf{q} \cdot \nabla T + Q, \tag{22}$$

$$(\rho C)_{eq} = n\rho_l C_l + (1 - n)\rho_s C_s, \tag{23}$$

$$\lambda_{eq} = n\lambda_l + (1 - n)\lambda_s, \tag{24}$$

where $(\rho C)_{eq}$ is the equivalent specific heat capacity, and λ_{eq} is the equivalent thermal conductivity coefficient.

3 Numerical Simulation of Hydraulic Fracturing

As shown in Fig. 1, the numerical model is 60 m long and 30 m wide; the horizontal well is excavated in the middle with a length of 20 m and a width of 2 m; the vertical joint is 4 m in length and 0.003 m in width. The excavation of the well is modeled for changing the elastic modulus from 30 GPa to 0.0001 GPa, which ensures the continuity of the field function at the interface. The joint is modeled as a discrete crack, that is, the internal boundaries form a channel to allow the fluid to flow inside. Seven points (i.e. P1-P7) marked in Fig. 2 are used to illustrate the simulation results.

In this study, the fluid injection is expressed as the term Q_m (the mass source) of Eq. (18). The maximum injection rate is 1.104 m^3/min. Then we can get

$$Q_{m\_max} = \frac{1.104}{60}\,m^3/s \times \frac{1}{20 \times 2 \times 1\,m^3} \times 1000\,kg/m^3 = 0.46\,kg/\left(m^3 \cdot s\right). \tag{25}$$

The hydraulic fracturing process lasts 1800 s; in the first 300 s, the injection rate reaches 0.2208 m^3/min linearly and keeps it for 300 s; then after 600 s injection, the rate attains the maximum and maintains it for 600 s. At the same time, the fluid is injected at the temperature of 333.15 K (60 °C), and the term Q (the heat source) of Eq. (21) is obtained by

$$Q = Q_m \times (333.15\,K - T) \times 4200\,J/(kg \cdot K). \tag{26}$$

The trend with respect to time of the above two sources can be described through the piecewise function plotted in Fig. 3. The turning points are 'softened' for the continuous second derivative. In addition, the upper surface of the domain is located at the depth of 500 m, and the initial ground-stress balance has been set. The excavation steps are simulated under the designed operational conditions (e.g. standard atmospheric pressure and ground temperature of 293.15 K), and the parameters used in this research are listed in Table 1.

Figure 4 shows the variations of the seepage pressure at P4, P5 and P6 in the intact rock. It can be seen that after the injection is accelerated at the time of 600 s,
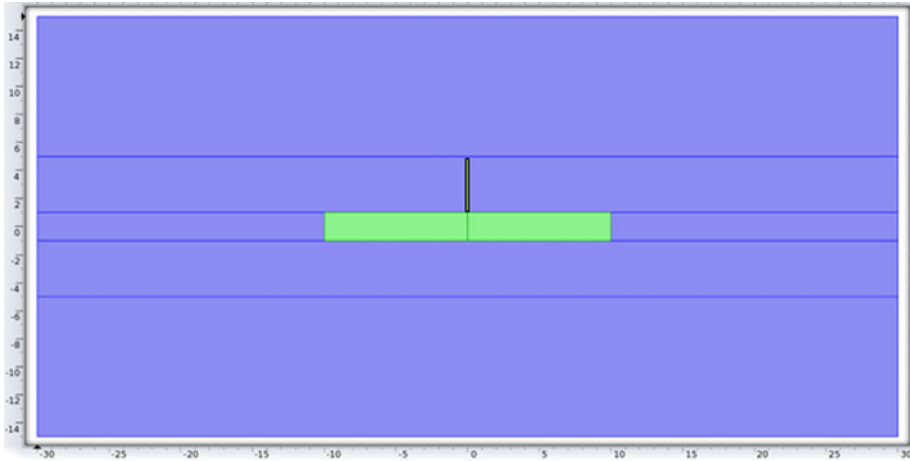
Fig. 1 The schematic diagram of the model
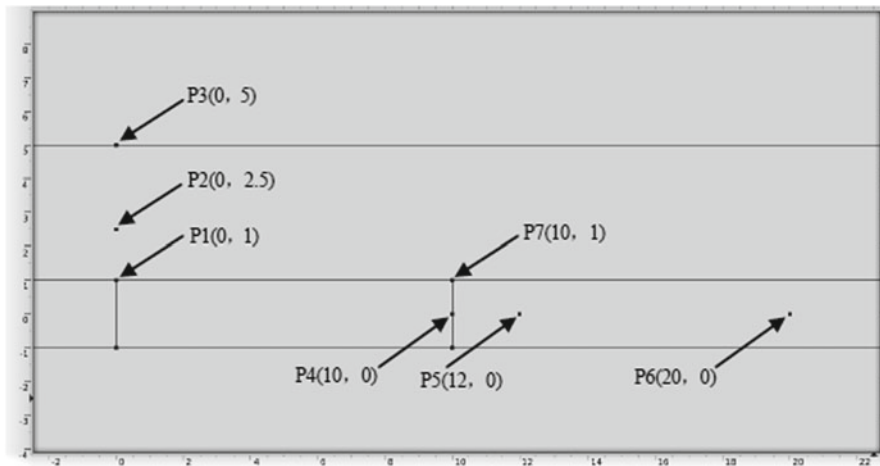
Fig. 2 Points of interest

the seepage pressure of P4 increases significantly and reaches 42 MPa in the end. In comparison, the seepage pressure of P5 is 1.3 MPa while that of P6 is basically zero. The observation is that the seepage pressure drops by 96.9% at about 2 m away from the horizontal well. On the other hand, the similar trends of the temperature at P4, P5 and P6 in the intact rock are shown in Fig. 5. It can be calculated that the final temperature growth of P5 (i.e. 0.3 K) is only 4.3% of that of P4 (i.e. 7 K), which indicates the heat conduction from the fluid to the rock is almost restricted within the same zone (i.e. 2 m around the well). It can also be found that there is a smaller change range of the temperature compared with the seepage pressure, that is, the heat source is 'consumed' faster than the mass source.
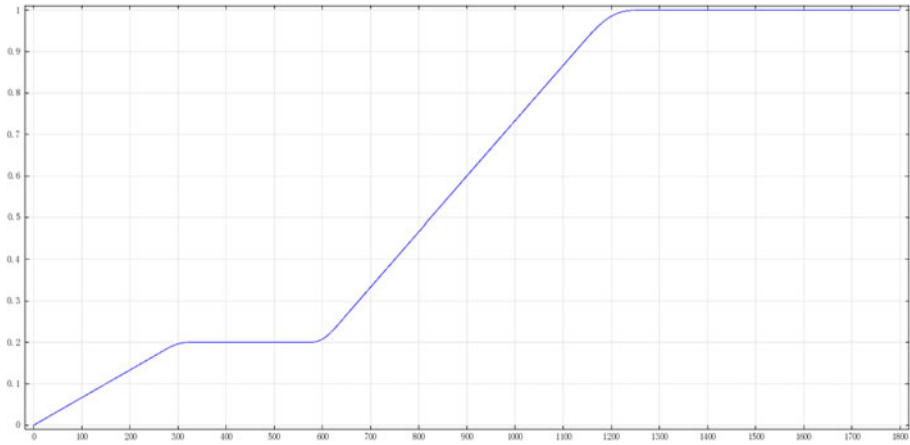
Fig. 3 The piecewise function with respect to time

Table 1 Parameters used in this research

Parameters	The horizontal well	The surrounding rock
Density of the rock skeleton $\rho_s(\text{kg}/\text{m}^3)$	2500	2500
Density of groundwater $\rho_l(\text{kg}/\text{m}^3)$	1000	1000
Elastic modulus before excavation $E(\text{GPa})$	30	30
Elastic modulus after excavation $E(\text{GPa})$	0.0001	30
Poisson's ratio ν	0.22	0.22
Pore ratio n	5%	5%
Permeability coefficient $k(\text{m}^2)$	5×10^{-10}	5×10^{-17}
Viscosity coefficient $\eta(\text{Pa} \cdot \text{s})$	0.005	0.005
Biot's consolidation coefficient α	0.9	0.9
Linear thermal expansion coefficient $\beta_s(1/\text{K})$	5×10^{-6}	5×10^{-6}
Specific heat of the rock skeleton $C_s(\text{J}/(\text{kg} \cdot \text{K}))$	2000	2000
Specific heat of groundwater $C_l(\text{J}/(\text{kg} \cdot \text{K}))$	4200	4200
Thermal conductivity of the rock skeleton $\lambda_s(\text{W}/(\text{m} \cdot \text{K}))$	2.4	2.4
Thermal conductivity of groundwater $\lambda_l(\text{W}/(\text{m} \cdot \text{K}))$	1000	1000
Joint width $d_f(\text{m})$	—	0.003
Initial pressure $P_0(\text{atm})$	1	1
Initial temperature $T_0(\text{K})$	293.15	293.15
Fluid injection temperature $T_l(\text{K})$	333.15	—
Load on the upper surface $F_0(\text{MPa})$	—	12.25

Fig. 4 The seepage pressure with respect to time

Fig. 5 The temperature variation with respect to time

Fig. 6 The comparison of the maximum principal stress between the cracked and intact rock
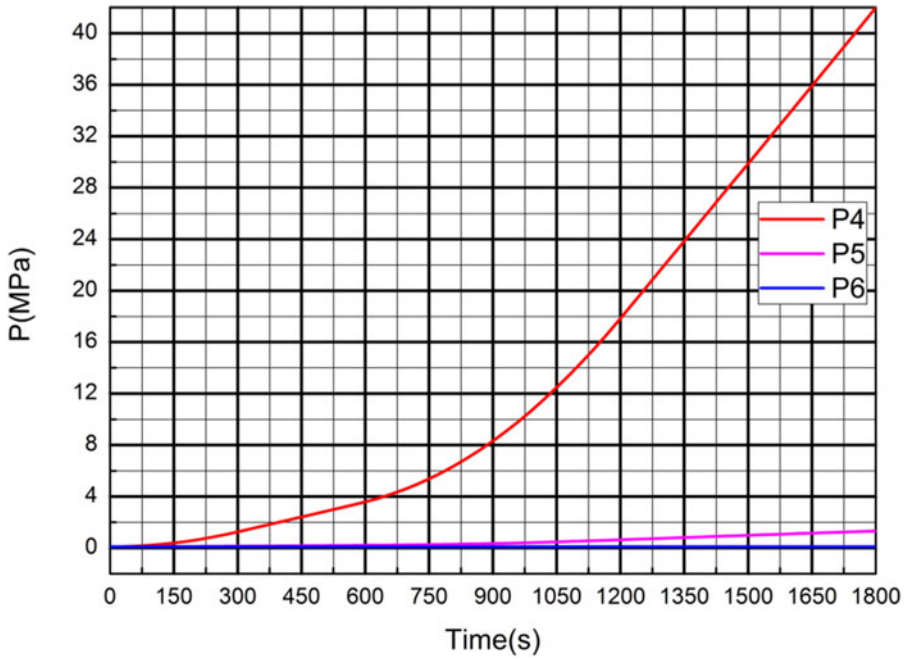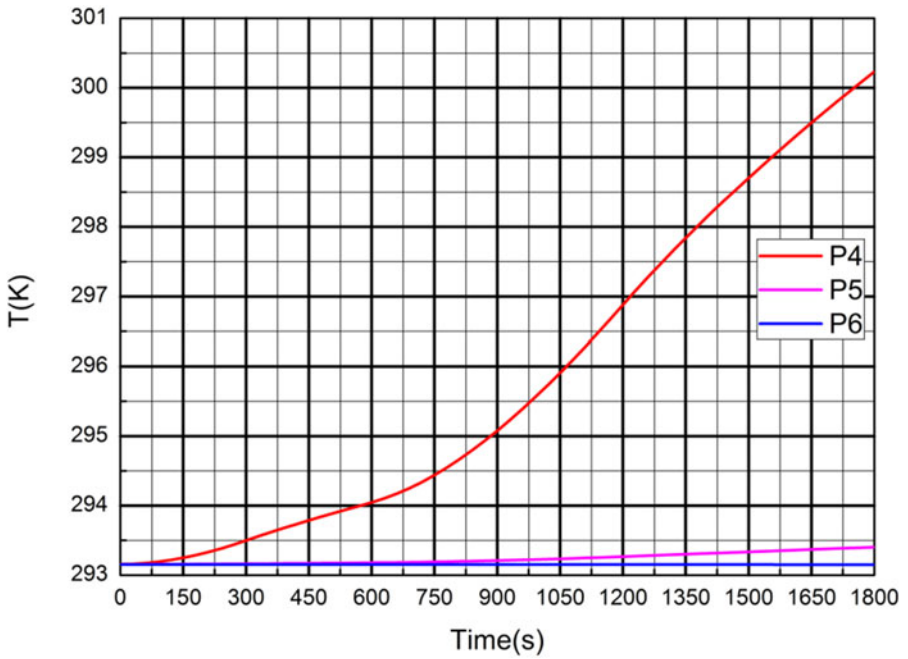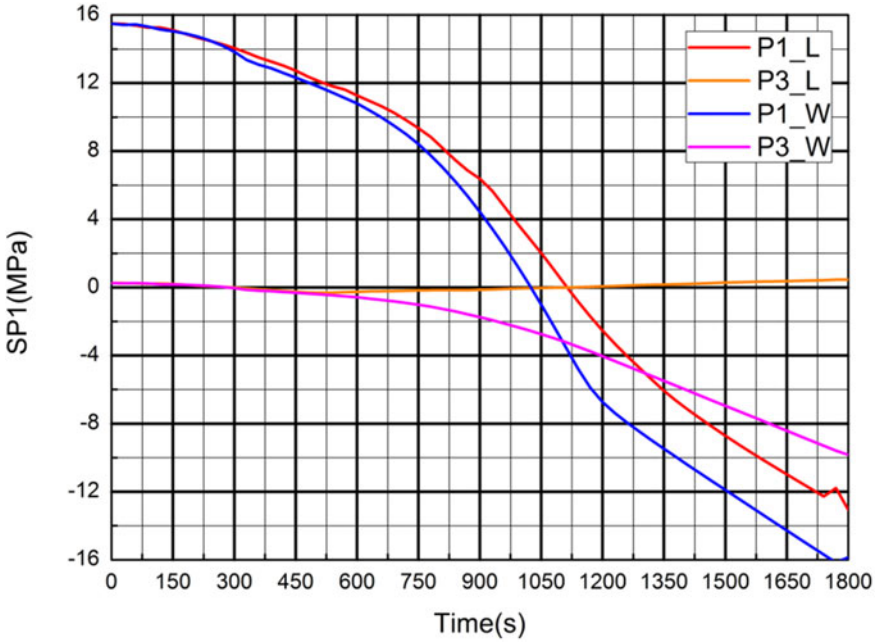

The cracked rock   The intact rock
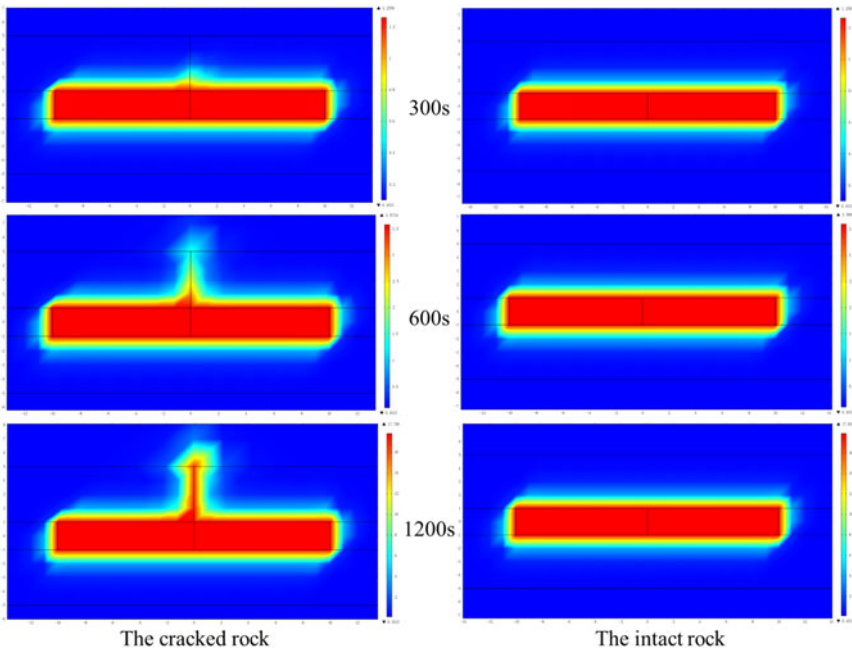
Fig. 7 The seepage pressure distribution in the cracked and intact rock
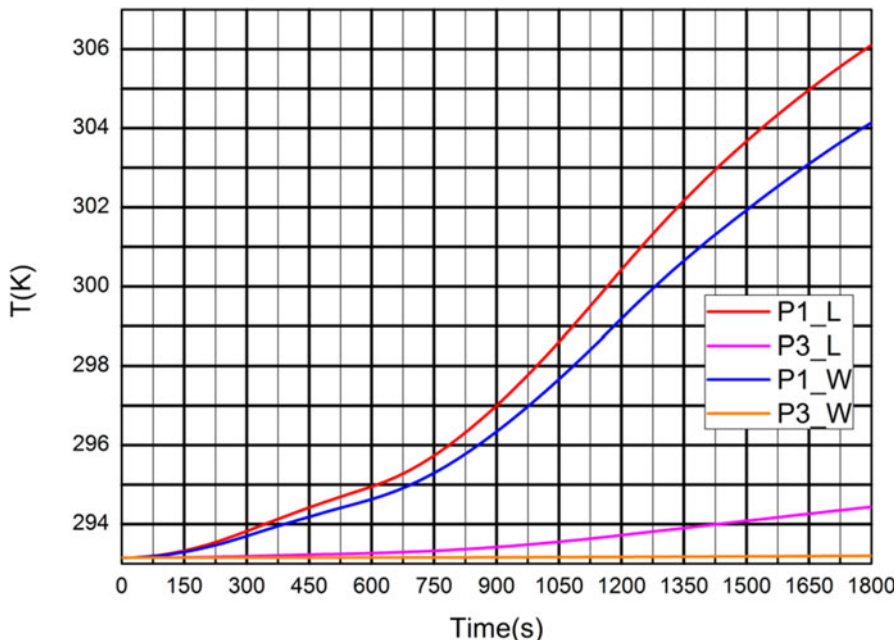
Fig. 8 The comparison of the temperature between the cracked and intact rock

This paper also presents a cracked-rock model with a single joint; in following figures, '_L' represents the cracked rock, and '_W' represents the intact rock. For the mechanical field, the differences of the maximum principal stress (i.e. SP1) at P1 and P3 between the two conditions are compared in Fig. 6. It can be observed that the maximum principal stress of the intact rock at P1 decreases more 4 MPa than that of the cracked rock at the same point, and the maximum principal stress of the cracked rock at P3 even increases instead of decreasing. The existence of the joint makes the hydro-mechanical coupling more obvious. For the hydraulic field, the seepage pressure distribution reveals the influence of the joint in the hydraulic fracturing process as shown in Fig. 7. It can be seen that at the time of 300 s, that is the first phase of the injection acceleration is completed, the seepage pressure 'highlights' the intersection of the joint and the horizontal well; then at the time of 1200 s, that is the second phase of the injection acceleration is completed, the distribution profile of the seepage pressure is relatively fixed. The existence of the joint expands the influence scope of hydraulic fracturing. For the thermal field, the comparison of the temperature between the cracked and intact rock is illustrated in Fig. 8. It can be found that the temperature increase of P3 is from 0 K in the intact rock to 1.3 K in the cracked rock. The indication is that the fluid can flow further along the joint of which the tip is a concentration point.

4 Conclusions

In this work, the coupled THM model of hydraulic fracturing in shale gas extraction is investigated. The numerical results of maximum principal stress, seepage pressure and temperature between the intact rock and the cracked rock with a single joint are compared. It can be concluded that the existence of the joint has a great effect on all the three fields, especially the mechanical and hydraulic fields. Meanwhile, the existence of the joint also makes the hydro-mechanical coupling more obvious. The variation trends of seepage pressure and temperature in the intact rock are similar. The seepage pressure decreases by 96.9% at about 2 m away from the horizontal well, and the heat conduction from the fluid to the rock is almost restricted within the same zone (i.e. 2 m around the well). The seepage pressure 'highlights' the intersection of the joint and the horizontal well when the first phase of the injection acceleration is completed; then the distribution profile of the seepage pressure is relatively fixed when the second phase of the injection acceleration is completed. The existence of the joint expands the influence scope of hydraulic fracturing, and the fluid can flow further along the joint of which the tip is a concentration point.

In this study, there is only a vertical crack, and the crack propagation is not discussed. The gas-phase influence is not considered either. The above topics can be the research subjects in the future.

References

1. Curtis, J.B.: Fractured shale-gas systems. AAPG Bull. **86**(11), 1921–1938 (2002)
2. Rahm, D.: Regulating hydraulic fracturing in shale gas plays: the case of Texas. Energy Policy **39**(5), 2974–2981 (2011)
3. Wang, H., Marongiu-Porcu, M.: Impact of shale-gas apparent permeability on production: combined effects of non-Darcy flow/gas slippage, desorption, and geomechanics. SPE Reserv. Eval. Eng. **18**(4), 495–507 (2015)
4. Braun, P., et al.: Determination of multiple thermo-hydro-mechanical rock properties in a single transient experiment: application to shales. Rock Mech. Rock Eng. **52**(7), 2023–2038 (2019)
5. Ishida, T., et al.: Influence of fluid viscosity on the hydraulic fracturing mechanism. J. Energy Resour. Technol.-Trans. ASME **126**(3), 190–200 (2004)
6. Adachi, A., et al.: Computer simulation of hydraulic fractures. Int. J. Rock Mech. Min. Sci. **44**(5), 739–757 (2007)
7. Dean, R.H., Schmidt, J.H.: Hydraulic-fracture predictions with a fully coupled geomechanical reservoir simulator. SPE J. **14**(4), 707–714 (2009)
8. Shimizu, H., et al.: The distinct element analysis for hydraulic fracturing in hard rock considering fluid viscosity and particle size distribution. Int. J. Rock Mech. Min. Sci. **48**(5), 712–727 (2011)
9. Yao, Y.: Linear elastic and cohesive fracture analysis to model hydraulic fracture in brittle and ductile rocks. Rock Mech. Rock Eng. **45**(3), 375–387 (2012)
10. Eshiet, K.I., et al.: Microscopic modelling of the hydraulic fracturing process. Environ. Earth Sci. **68**(4), 1169–1186 (2013)

11. Shiozawa, S., et al.: The effect of stress boundary conditions on fluid-driven fracture propagation in porous media using a phase-field modeling approach. Int. J. Numer. Anal. Methods Geomech. **43**(6), 1316–1340 (2019)
12. Gao, Q., Ghassemi, A.: Three-dimensional thermo-poroelastic modeling and analysis of flow, heat transport and deformation in fractured rock with applications to a lab-scale geothermal system. Rock Mech. Rock Eng. **53**(4), 1565–1586 (2020)
13. Obeid, W., et al.: Mathematical formulation of thermo-hygro-mechanical coupling problem in non-saturated porous media. Comput. Methods Appl. Mech. Eng. **190**(39), 5105–5122 (2001)
14. Rutqvist, J., et al.: A modeling approach for analysis of coupled multiphase fluid flow, heat transfer, and deformation in fractured porous rock. Int. J. Rock Mech. Min. Sci. **39**(4), 429–442 (2002)
15. Nguyen, T.S., et al.: Modelling the FEBEX THM experiment using a state surface approach. Int. J. Rock Mech. Min. Sci. **42**(5–6), 639–651 (2005)
16. Sheng, J., et al.: Stress analysis of a borehole in saturated rocks under in situ mechanical, hydrological and thermal interactions. Energy Sources Part A-Recovery Util. Environ. Effects **30**(2), 157–169 (2008)
17. Tong, F., et al.: A fully coupled thermo-hydro-mechanical model for simulating multiphase flow, deformation and heat transfer in buffer material and rock masses. Int. J. Rock Mech. Min. Sci. **47**(2), 205–217 (2010)
18. Zhang, Z., Cheng, X.: A fully coupled THM model based on a non-equilibrium thermodynamic approach and its application. Int. J. Numer. Anal. Methods Geomech. **41**(4), 527–554 (2017)
19. Kwon, S., Lee, C.: THM analysis for an in situ experiment using FLAC3D-TOUGH2 and an artificial neural network. Geomech. Eng. **16**(4), 363–373 (2018)

Research Progress on the Build-Up Rate of Rotary Steerable System Tools

Ge Wang, Wenjun Huang, and Deli Gao

Abstract Rotary steerable system (RSS) tools technology has dramatically ameliorated the control accuracy of wellbore trajectory, it can be applied to the needs of well completion such as ultra-deep well, extended-reach well, clusters well and horizontal branch well, etc. Nevertheless, it is vulnerable to a series of restrictive factors for the use of the RSS under the practical drilling conditions, such as drill tool structure, drill bit capacity, drilling parameters, and stratum properties, which are resulting in inaccurate prediction of the build-up rate. Push-the-bit RSS refers to a steerable drilling system that controls well trajectory through adjustment of stretch length of pushing block. Based on this feature, the mechanical during the drilling process for RSS is analyzed, a three-dimensional mechanical of RSS is established. Meanwhile, the model of interaction between the drill bit and the formation is considered, a prediction model for the drilling trajectory of RSS is established. Next, by utilizing the trajectory prediction model, it needs to be further considered the impact of the actual drilling process to RSS, using the criterion of zero lateral rate of penetration (ROP), a three-dimensional calculation model of build-up rate for RSS is established, and a conversion coefficient is introduced to reflect the ratio of actual build-up rate to theoretical build-up rate. Finally, the field experience chart is drawn as the main basis for the subsequent optimization of the RSS tools and drilling parameters.

Keywords Rotary steerable system · Build-up rate · Pushing force · Trajectory control

Citations

Rotary steering drilling technology is a new cutting-edge automated drilling technology that emerged in the 1990s. Its appearance is a technological revolution in world drilling engineering. Compared with the conventional drilling tool assembly,

G. Wang · W. Huang · D. Gao (✉)
MOE Key Laboratory of Petroleum Engineering, China University of Petroleum,
Beijing 102249, China
e-mail: gaodeli@cup.edu.cn

RSS can flexibly adjust the offset mechanism according to the actual formation conditions to achieve the well deviation and azimuth control, which greatly improves the drilling efficiency and working safety. With the development of RSS technology, the theoretical research on wellbore trajectory prediction has been relatively complete [1–4].

Due to the complex geological structure of unconventional oil and gas resources and the late start of the technology in China, there are still many problems with the use of RSS tools, which has been restricted its popularization. Nowadays, the research hotspot of RSS technology is mainly aimed at the design research of build-up rate prediction and tool optimization [5, 6].

More than 40% of the directional wells in the world are drilled using RSS. In order to achieve RSS tools drilling through the target formations distributed in the "three-dimensional" area, it is necessary to solve the potential problems of rotary steering and form more precise RSS [7–9]. The calculation method of drilling tool build-up rate, based on this, considers the various factors to give full play to its potential performance. Therefore, the following work needs to be carried out:

1. Use tubular mechanics algorithm to study the mechanical properties of RSS tools;
2. Use the mechanical model of the RSS tools and the bit-formation interaction model to study the prediction method of wellbore trajectory;
3. Under consideration of the performance of the RSS tools, drilling parameters, and formation factors, conduct the research on the calculation method of the RSS tools build-up rate, the optimization design method of the RSS tools and drilling parameters.

1   Trajectory Model of Rotary Steerable System Tool

The trajectory prediction model of RSS is the basis of the build-up rate prediction model. Based on the effect of the wellbore trajectory under the situations of anisotropic formation, wellbore shape, BHA multiple loads, drilling process, offset mechanism, bit drift, etc., the analysis is mainly focus on the following two aspects, as shown in Fig. 1:

(1) Aimed at function of pushing block, the mechanical during the drilling process for RSS is analyzed, a three-dimensional mechanical of RSS is established.
(2) The model of interaction between the drill bit and the formation is need to take into consideration, a prediction model for the drilling trajectory of RSS is established.
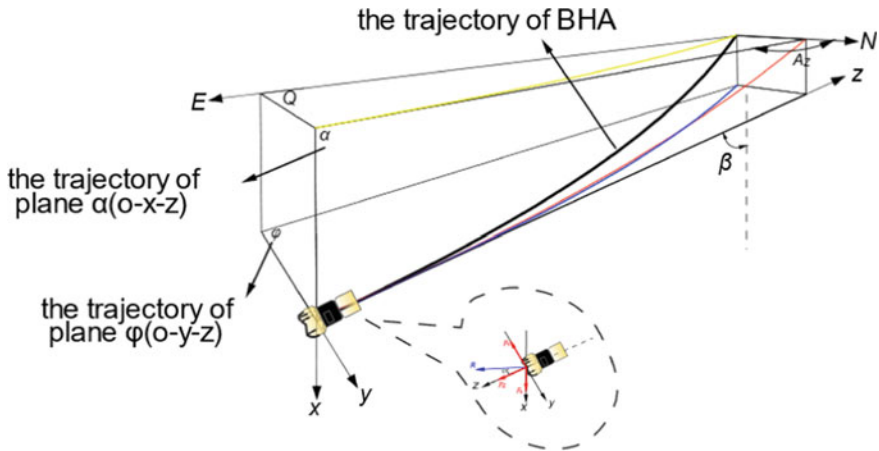
Fig. 1 RSS trajectory prediction model during the drilling process
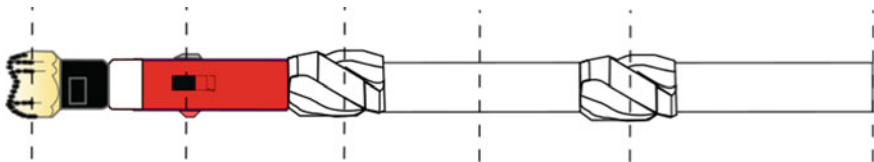
Fig. 2 The overall schematic diagram of RSS unit

1.1 Rotary Steerable drilling Trajectory Model

For RSS tools, the offset mechanism is an imperative factor affecting its offset mechanical analysis. During the drilling process, the reaction force of the offset mechanism on the borehole in different directions may change the contact state of RSS tools assembly and borehole, which will cause the drill string to be stressed under nonlinear conditions. The situation has become more complicated. Therefore, it is necessary to tackle the steering drilling tool as a vertical and horizontal bending beam and column for analysis and research [10, 11], and to consider the possible contact between the drill bit, the offset mechanism, each span stabilizer, and the pipe string and the well wall The positions are all set as nodes, and the unit is divided into the steering drilling tool to find its movement law, that is, the overall unit is shown in Fig. 2, and the unit division of the contact between each node and the borehole wall is shown in Fig. 3.
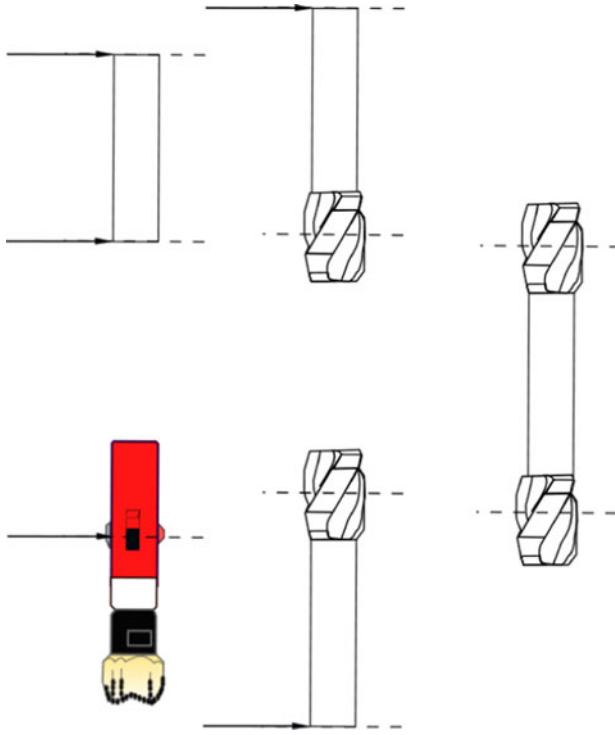
Fig. 3 Schematic diagram of unit split of RSS
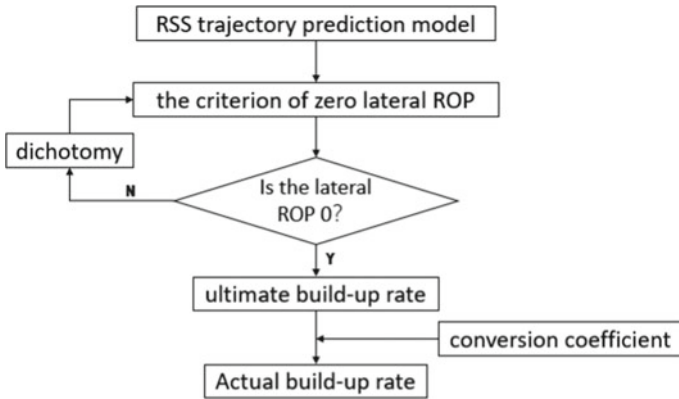
Fig. 4 Forecasting method of build-up rate

Compared with the finite element method and the finite difference method, the weighted residual method is more efficient in dealing with mechanical problems, and can obtain more accurate numerical solutions, which can greatly reduce the workload of the program code, so the weighted residual method will be selected Solve the mechanical model of RSS tool. The essence of the weighted residual method is to first assume a trial function as the approximate solution of the control differential equation. This function has both a certain trial function term and a coefficient term to be solved. The trial function can be brought into the control differential of the BHA in the equation, the approximate solution of the three-dimensional mechanical model is obtained. According to different conditions, different weight functions are selected to eliminate its internal residual value and obtain a more accurate numerical solution, and then establish a three-dimensional mechanical model for the prediction of the drilling trajectory of the RSS tool.

1.2 The Model of Bit and Formation Interaction

For push-to-type rotary steering, the larger lateral force will make the borehole trajectory fluctuate greatly during the steering process, and it is also easy to produce larger dog legs and well diameters. While the drill bit is subjected to greater axial and lateral force, due to the difference in the cutting ability of the drill bit in the axial and lateral (radial) direction, under the action of the lateral force, the lateral cutting of the drill bit causes lateral displacement. Therefore, the displacement direction of the drill bit is not consistent with the direction of its mechanical force during the drilling process, resulting in the deviation of the wellbore trajectory.

However, borehole deflection is not only related to the difference in the cutting ability of the bit, but also formation factors are another essential reason for the deflection of the borehole trajectory. According to the theory of well deviation control. Generally, when drilling and sloping formations, when sloping formations angle is less than 45°, drill bit tends to drill in the direction of the vertical plane, if sloping formations angle exceeds 60°, drill bit will tend to drill in the direction of the parallel plane. Due to the inhomogeneity of the drill ability of the formation, when the formation is anisotropic, it will cause asymmetric cutting of the bottom hole by the drill bit.

Therefore, in the wellbore trajectory prediction and control calculation, in addition to the analysis and research on the mechanical properties of BHA, it is also necessary to analyze the anisotropic drilling characteristics of the bit and the formation. The anisotropy of the formation can be based on empirical data and experimental data. Bit anisotropy can be calculated based on the concept of effective drilling force [12].

$$\begin{bmatrix} R_a \\ R_x \\ R_y \end{bmatrix} = D_n \cdot \mathbf{S}^T \cdot \begin{bmatrix} 1 & & \\ & I_b & \\ & & I_b \end{bmatrix} \cdot \mathbf{S} \cdot \begin{bmatrix} W_b \\ F_x \\ F_y \end{bmatrix} \quad (1)$$

Among them, R_a, R_x, and R_y are the components in the drilling penetration along the borehole axis direction, well inclination, and the lateral direction of azimuth plane, D_n is the standard (normal) drilling efficiency, S is the bottom hole coordinate and the bit Coordinate conversion relationship, I_b is the anisotropy index of the bit, W_b, F_x and F_y are the mechanical load on the bit.

2 Prediction of Build-Up Rate

By utilizing the trajectory prediction model, it needs to be further considered the impact of the actual drilling process to RSS, using the criterion of zero lateral rate of penetration (ROP), a three-dimensional calculation model of build-up rate for RSS is established. Secondly, the conversion coefficient is added and the practical data is used for inversion calculation to reach a more precise method of the build-up rate, as shown in Fig. 2. Finally, the field experience chart is drawn as the main basis for the subsequent optimization of the RSS tool and drilling parameters.

The expression of the build-up rate can be obtained by Eq. (2):

$$\kappa_A = \lambda \cdot \kappa_L \quad (2)$$

Among them, the conversion coefficient λ is the conversion coefficient, K_A is the actual build rate, (degree)/30 m; KL is the limit build rate (degree)/30 m. The conversion factor is related to the drilling process, and its values ranges from 0.4 to 0.7.

3 Case Study of RSS Prediction

Table 1 shows that the drilling tool parameters of RSS of a well in Tanggu, including drill bit, guide head, the flexible sub, gamma, MWD, stabilizer, BCPM, collar, and drill pipe.

According to the actual drilling trajectory data, RSS tool parameters, and working condition parameters of a well in Tanggu, the deflection rate of the well is predicted. The details are shown in Fig. 5.

Among them, the actual drilling parameters provided by the well are as follows: weight on bit 5 tons; mud density 1.16 g/cm^3; bit anisotropy index 0.03; borehole expansion rate 0.03 m.
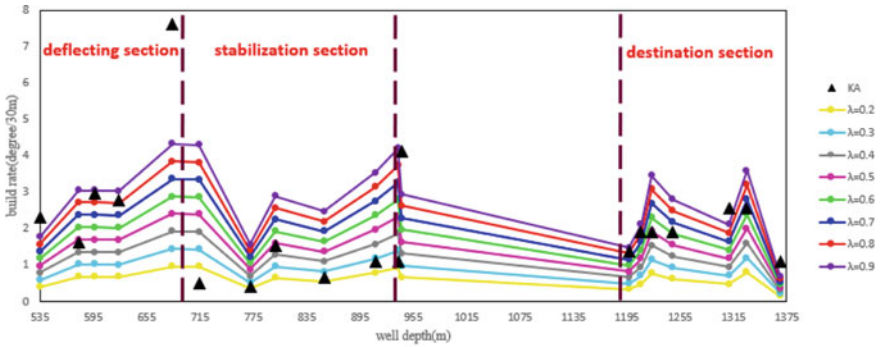
(a) Results of slope building rate under different well depths
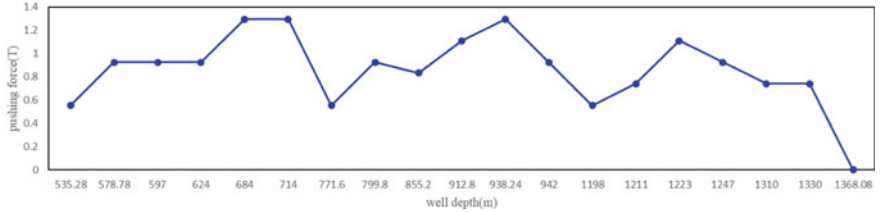
（b）Pushing force values under different well depths

Fig. 5 Inversion chart of conversion coefficient of building slope

Table 1 Size of RSS

Name	Diameter (in)	Length (ft)
Bit	8.5	1.312
Offset mechanism	8.188	0.334
Stabilizer	8.366	0.492
Flexible sub	5	3.163
Gamma	7.244	3.478
MWD	7	9.33
BCPM	7.244	4.396
Collar	7	125.391
Drilling string	5	6541.679

According to the inversion of different conversion coefficients, a conversion coefficient chart is obtained, and the relative error method is adopted to verify the validity of the model.

$$\varepsilon = \left| \frac{\kappa_{real} - \kappa_A}{\kappa_{real}} \right| \times 100\% \tag{3}$$

Among them, ε is the relative error. As shown in Fig. 5(a), the measured value of the build rate is mainly distributed between the conversion coefficient value of 0.6–0.7. When the value of the conversion coefficient is 0.65, the average relative error of the build rate is 9.45%. The calculated value agrees well.

3.1 Effect of Pushing Force Magnitude

Figure 6 shows the condition of the change of pushing force with the build-up rate. As the pushing force is increased, the build-rate of the steerable tool during the drilling process will increase for the pushing effect increases. And the build-up rate changes significantly under the influence of the pushing effect.

Fig. 6 Effect of pushing force magnitude
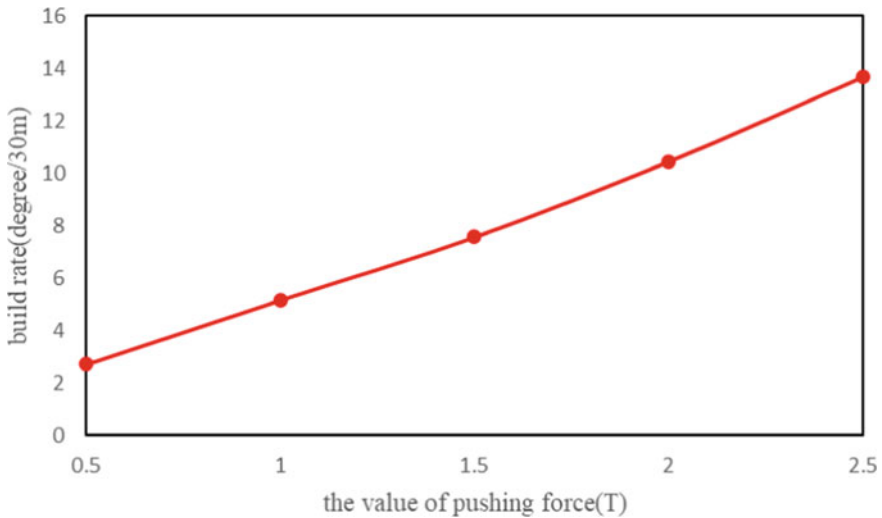
(a) effect of weight on bit   (b) effect of inclination angle
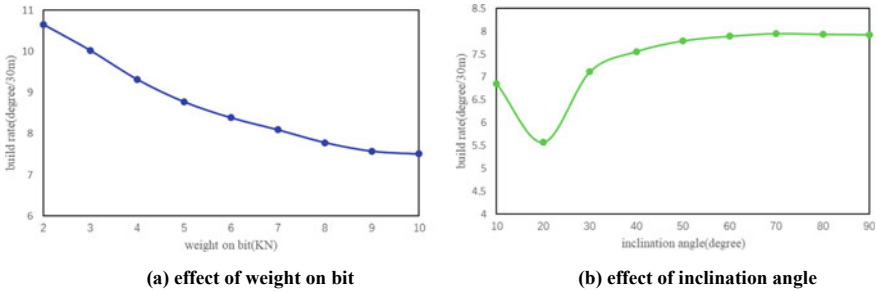
Fig. 7 Effect of drilling parameters

3.2 Effect of Drilling Parameters

Figure 7 reveals the regular of influence of drilling parameters with the change of build rate. As shown in Fig. 7(a), with the increase of weight on bit, the ratio of thrust force to weight on bit decreases, and the relative effect of thrust force on the drilling trajectory decreases, that is, the overall trend of deflection rate decreases. According to Fig. 7(b), as the inclination angle increases, the pendulum effect and leverage effect will gradually increase. At first, the leverage effect increases at a faster rate, and the subsequent pendulum effect increases at a faster rate. The build rate shows a trend of first increasing and then decreasing.

3.3 Effect of Bit Anisotropy

Generally, the bit anisotropy index (I_b) can be calculated by inversion of time drilling data. As I_b increases, the lateral cutting ability of the drill increases. However, the bit anisotropy index is difficult to determine under normal circumstances, so this article assumes isotropic formation conditions and incorporates the formation conditions into the conversion coefficient, which can more truly reflect the impact of the bit anisotropy index on the RSS build rate. According to Fig. 8, the larger the anisotropy coefficient of the drill bit, the comprehensive effect of various effects will be enlarged, thus the build-up rate will be increased.
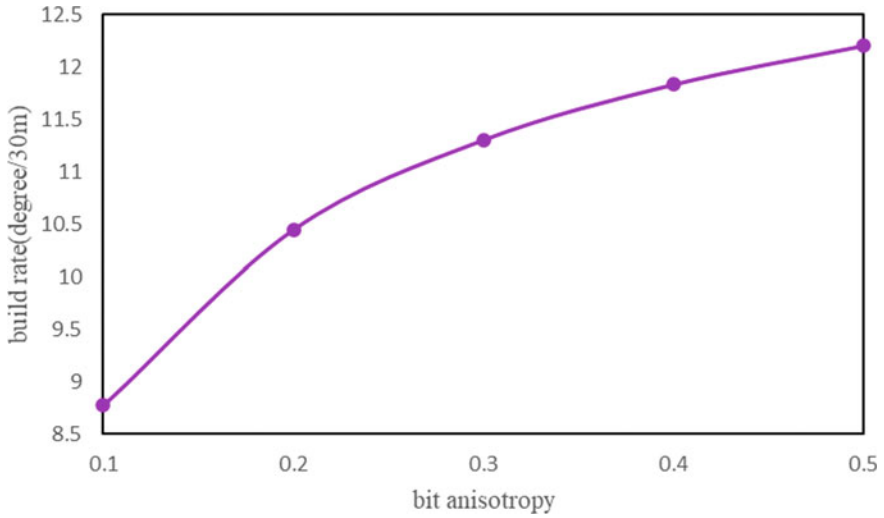
Fig. 8 Effect of bit anisotropy

Fig. 9 Effect of flexible sub length

3.4 *Effect of RSS Structure*

Figure 9 reveals the regular of influence of flexible sub length with the change of build rate. As the length of the flexible sub-joint increases, the influence of the rigidity of the upper drilling tool decreases, which is more conducive to the play of the pushing effect, that is, the increase in build rate.

4 Results, Observations, Conclusions

The result show that the calculated value is in good agreement with the measured value by utilizing the criterion of zero lateral ROP. This method can provide a basis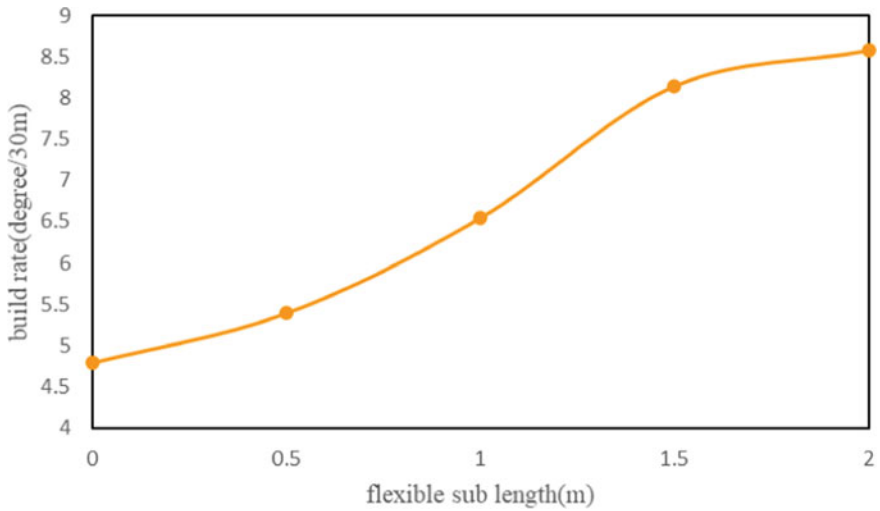 for subsequent theoretical research and practical application. Adding the conversion coefficient to the calculation model of the build-up rate, on the one hand, drill bit anisotropy index has been made easier to invert so that the efficiency of calculation of the bit-formation interaction has increased. On the other hand, the conversion coefficient inverted by actual drilling data comprehensively considers the error caused by tool structure, bit anisotropy index inversion, bit drift, formation, etc., which can more accurately predict the build-up rate.

Build-up rate is closely related to tool structure parameters (flexible sub-joints), pushing force, drill bit capacity (anisotropy index, etc.), drilling parameters (weight on bit, etc.). In order to control the wellbore trajectory more accurately, on the one hand, the drill tool assembly and matching bit should be reasonably selected in the design stage, and on the other hand, the drilling parameters and the pushing force should be reasonably controlled during the actual drilling process.

References

1. Gao, D.L.: Optimized design and control techniques for drilling & completion of complex-structure wells. China University of Petroleum Press (2011)
2. Hu, W.D.: The present and future of low permeability oil and gas in China. China National Petroleum Corporation **11**(8), 29–37 (2009)
3. Jiang, W., Jiang, S.Q., Shen, L.M., Fu, X.S., Chen, P.: Research on rotary navigation drilling tools and its application. Oil Drilling Prod. Technol. **05**, 21–24 (2008)
4. HO, H.S.: Prediction of drilling trajectory in directional wells via a new rock-bit interaction model. SPE 16658 (1987)
5. Jiang, W., Jiang, S.Q., Shen, L.M., Fu, X.S., et al.: Application of rotary steering drilling technology and its research progress. Nat. Gas. Ind. **33**(4), 75–79 (2013)
6. Zhu, X.H., Liu, Q.Y., Li, H.Y., Tong, H.: Study on theory and method of borehole trajectory prediction. Natural Gas Industry, (04), 38–40+6–7 (2004)
7. Jerez, H., Tilley, J.: Advancements in powered rotary steerable technologies result in record-breaking runs. SPE 169348 (2014)
8. Biscaro, E., D. Alessandro, J.D., Moreno, A., Hahn, M., Lamhorn, R., Al, Naabi. M.H., Bowser, A.C.: New Rotary Steerable Drilling System Delivers Extensive Formation Evaluation for High Build Rate Wells. SPE 174016 (2015)
9. Wang, P., Shen, L.M., Dou, X.R., et al.: Ne-west development and tendency of foreign rotary steerable systems. Drill. Prod. Technol. **36**(6), 32–35 (2013)
10. Gao, D.L., Liu, X.S., XU, B.Y.: Prediction and Control of Wellbore Trajectory. China University of Petroleum Press, Dongying (1994)
11. Bai, J.ZH., Su, Y.N.: Theory and practice of well deviation control. Petroleum Industry Press (1990)
12. Gao, D.L., Nie, C.P., Liu, X.S.: A Study on anisotropic drilling characteristic of drilling bit, J. Univ. Petroleum. China (Edition of Natural Science) **6**, 28–31 (1991)

Generalized Finite Difference Method for Solving Viscoelastic Problems

Check for updates

Jian Li and Tao Zhang

Abstract In this paper, the generalized finite difference method (GFDM) combined with the implicit Euler method is developed to solve the viscoelastic problem. The mathematical description of the viscoelastic problem is a time-dependent boundary value problem, governed by a second-order partial differential equation and non-linear boundary conditions. To solve the time-dependent differential governing equation and boundary conditions of viscoelasticity, the implicit Euler method and GFDM are employed for the temporal discretization and the spatial discretization respectively. GFDM is a newly developed meshless method, which avoids time-consuming mesh generation and numerical integration. The basic idea of the GFDM originates from the moving least squares method to transform the spatial derivatives at each node into linear summation of nearby node function values with different weighting coefficients. Two numerical examples are presented to illustrate the accuracy, stability and efficiency of GFDM, including the creep and stress relaxation of viscoelastic materials with single connected domains and double connected domains.

Keywords Generalized finite difference method · Viscoelastic problem

1 Introduction

Viscoelastic problems are very common in engineering, such as the creep of clay or rock, which can be regarded as a class of time-dependent boundary value problems. For the simple viscoelastic problem, the analytic solutions can be obtained. But for the complex engineering problems, numerical methods are inevitably required.

Finite element method (FEM) [1, 2] and boundary element method (BEM) [3, 4] were initially used to solve the viscoelastic problem, while the element would be distorted in the process of calculation and this reduced the accuracy and led to the

J. Li · T. Zhang (✉)
Beijing Institute of Technology, Beijing 100000, China
e-mail: taozhang@bit.edu.cn

failure of calculation. The meshless method which was based only on node information could partially or completely avoid element constraints. So it was widely used to solve the viscoelastic problems [5–8].

The GFDM, which preserves the accuracy and simplicity of the traditional finite difference method and avoids the time-consuming meshing generation and numerical integration, is a developable meshless method proposed by Benito et al. [9] in 2001. GFDM is based on the moving-least squares method to transform the partial derivative at each node into the linear summation of the nearby node function values with different weight function. Each internal node and boundary node is forced to respectively satisfy the governing equation and boundary conditions. Then the partial differential equations will be transformed into linear algebraic equations that can be written as sparse matrix and solved easily. The GFDM preserves the physical conservation of the original equation (including the conservation of mass, momentum, energy, etc.) and other important characteristics in the subdomain. It is an important discrete method for solving the partial differential equations. Moreover, GFDM can be used to accurately and efficiently solve various problems, such as dynamic propagation of crack, impact and collision, large deformation and complex high-dimensional geometric problems and so on.

In this paper, the GFDM is extended for the first time to solve the viscoelastic problem. The outline of this paper is as follows: In Sect. 2, The temporal and special discretization of governing equation and boundary condition by implicit Euler method and GFDM is introduced. In Sect. 3, Two examples are presented to illustrate the stability, accuracy and efficiency of the proposed method. In Sect. 4, a brief summary of this paper is provided.

2 The Generalized Finite Difference Method for Viscoelastic Problem

2.1 Basic Equation of Viscoelastic Problem

The three-parameter viscoelastic model is used to descript the viscoelasticity behavior. The constitutive equations of this model are

$$\varepsilon(t) = \frac{1}{E_2} G\sigma(t), \quad t = 0 \tag{1}$$

$$q_0\varepsilon(t) + q_1\frac{d\varepsilon(t)}{dt} = G\left(\sigma(t) + p_1\frac{d\sigma(t)}{dt}\right), \quad t > 0 \tag{2}$$

where

$$p_1 = \frac{\eta_1}{E_1 + E_2}, q_0 = \frac{E_1 E_2}{E_1 + E_2}, q_1 = \frac{E_2 \eta_1}{E_1 + E_2}, \tag{3}$$

$$G = \begin{pmatrix} G_{11} & G_{12} & 0 \\ G_{12} & G_{22} & 0 \\ 0 & 0 & G_{33} \end{pmatrix}. \tag{4}$$

For the plane stress problem,

$$G_{11} = G_{22} = 1, G_{33} = 2(1+v), G_{12} = -v, \tag{5}$$

where v is Poisson's ratio.

The strain–displacement relationship can be written as

$$\varepsilon_{ij} = \left(u_{i,j} + u_{j,i} \right)/2. \tag{6}$$

The equilibrium equation of viscoelasticity is

$$\sigma_{ij,j} + b_i = 0, \tag{7}$$

and the boundary conditions are specified by:

$$u = \tilde{u} \quad \text{on } \Gamma_u, \tag{8}$$

$$p = \tilde{p} \quad \text{on } \Gamma_t, \tag{9}$$

where σ and ε are the stress and strain respectively; b is the body force; u is the displacement; p is the traction; $\tilde{u}$ and $\tilde{p}$ are the prescribed values of u and p on the boundary.

2.2 Generalized Finite Difference Method for Viscoelastic Problem

After the time term of the governing equation and boundary conditions are discretized by implicit Euler method, The GFDM is used to discretize the space terms of governing equations and boundary conditions. At the beginning, the interesting domain is discretized into regularly or randomly distributed boundary nodes and internal nodes. A node xi is selected as the center node, the m nodes x_j (j = 1, 2, …, m) around the x_i will be searched according to the principle of nearest distance, the node x_i and m nearest nodes are used to form a star, which is shown in Fig. 1. Assume the u_i is the displacement at point x_i and u_j is the displacement at point x_j. Using the Taylor series expansion, we can expand the u_j as:

Fig. 1 The circular shape of
the star
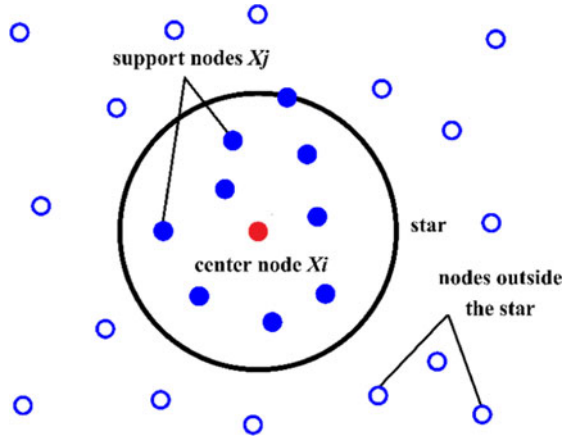
$$u_j = u_i + h_{j,i}\frac{\partial u_i}{\partial x} + k_{j,i}\frac{\partial u_i}{\partial y} + \frac{1}{2}h_{j,i}^2\frac{\partial^2 u_i}{\partial x^2} + \frac{1}{2}k_{j,i}^2\frac{\partial^2 u_i}{\partial y^2} + h_{j,i}k_{j,i}\frac{\partial^2 u_i}{\partial x\partial y} + \cdots\cdots, \quad (10)$$

where $h_{j,i} = x_j - x_i, k_{j,i} = y_j - y_i$ are the distances between x_i and x_j in x and y directions respectively. The accurate of u_j will be higher if the order of Taylor series increase. Here, the Eq. (10) is truncated after the second-order derivatives. A residual function is defined as:

$$B(u) = \sum_{j=1}^{m}\left[\left(\begin{array}{c}u_i - u_j + h_{j,i}\frac{\partial u_i}{\partial x} + k_{j,i}\frac{\partial u_i}{\partial y} + \frac{1}{2}h_{j,i}^2\frac{\partial^2 u_i}{\partial x^2}\\ + \frac{1}{2}k_{j,i}^2\frac{\partial^2 u_i}{\partial y^2} + h_{j,i}k_{j,i}\frac{\partial^2 u_i}{\partial x\partial y}\end{array}\right)w\left(h_{j,i}, k_{j,i}\right)\right]^2, \quad (11)$$

where $w\left(h_{j,i}, k_{j,i}\right)$ is the weighting function at x_i. There are many choices for $w\left(h_{j,i}, k_{j,i}\right)$, such as potential function, exponential function, cubic spline, quartic spline, etc. The quartic spline is chosen as the weighting function in this study:

$$w(d_{ij}) = \begin{cases} 1 - 6\left(\frac{d_{ij}}{dm_i}\right)^2 + 8\left(\frac{d_{ij}}{dm_i}\right)^3 - 3\left(\frac{d_{ij}}{dm_i}\right)^4, d_{ij} \leq dm_i \\ 0, d_{ij} > dm_i \end{cases}, \quad (12)$$

where d_{ij} denotes the distance between nodes x_i and x_j, dm_i denotes the distance between the farthest node and the center node in the star.

By minimizing the above function $B(u)$ with respect to $D_u = \left\{\frac{\partial u_i}{\partial x}, \frac{\partial u_i}{\partial y}, \frac{\partial^2 u_i}{\partial x^2}, \frac{\partial^2 u_i}{\partial y^2}, \frac{\partial^2 u_i}{\partial x\partial y}\right\}^T$, a linear equation system is obtained by:

$$AD_u = b, \quad (13)$$

The vector b in Eq. (13) can be reformulated as the following form

$$b = BQ, \tag{14}$$

where $Q = \begin{bmatrix} u_i & u_1^i & u_2^i & \cdots & u_m^i \end{bmatrix}^T$ are the displacement of the total nodes inside the star. Therefore, D_u can be written as:

$$D_u = \begin{bmatrix} \frac{\partial u_i}{\partial x} \\ \frac{\partial u_i}{\partial y} \\ \frac{\partial^2 u_i}{\partial x^2} \\ \frac{\partial^2 u_i}{\partial y^2} \\ \frac{\partial^2 u_i}{\partial x \partial y} \end{bmatrix} = A^{-1}b = A^{-1}BQ = DQ = D \begin{bmatrix} u_i \\ u_1^i \\ u_2^i \\ \cdots \\ u_m^i \end{bmatrix}. \tag{15}$$

From the Eq. (15), we can obtain the partial derivatives of u_i at x_i by a linear combination of u_j at its supporting nodes x_j. The partial derivatives can be written as follows:

$$\left. \frac{\partial u}{\partial x} \right|_i = wx_0^i u_i + \sum_{j=1}^{m} wx_j^i u_j^i, \tag{16}$$

$$\left. \frac{\partial u}{\partial y} \right|_i = wy_0^i u_i + \sum_{j=1}^{m} wy_j^i u_j^i, \tag{17}$$

$$\left. \frac{\partial^2 u}{\partial x^2} \right|_i = wxx_0^i u_i + \sum_{j=1}^{m} wxx_j^i u_j^i, \tag{18}$$

$$\left. \frac{\partial^2 u}{\partial y^2} \right|_i = wyy_0^i u_i + \sum_{j=1}^{m} wyy_j^i u_j^i, \tag{19}$$

$$\left. \frac{\partial^2 u}{\partial x \partial y} \right|_i = wxy_0^i u_i + \sum_{j=1}^{m} wxy_j^i u_j^i, \tag{20}$$

where $\left\{ wx_j^i \right\}_{j=0}^{m}, \left\{ wy_j^i \right\}_{j=0}^{m}, \left\{ wxx_j^i \right\}_{j=0}^{m}, \left\{ wyy_j^i \right\}_{j=0}^{m}, \left\{ wxy_j^i \right\}_{j=0}^{m}$ are weight coefficients of the center node x_i.

The implementing procedure at each node inside the computational domain can be carried out as described above. And the derivatives at each node will be replaced with the linear combination of the approximate nodal values. Each interior node is forced to satisfy the governing equation and each boundary node is forced to satisfy the boundary condition. The partial differential equations will be transformed into linear (or nonlinear) algebraic equations.

3   Numerical Examples

3.1   Example 1: Creep Analysis of a Rectangular Plate

The first example is a rectangular viscoelastic plate subjected to a uniform traction loading $P = 1\,\text{N/m}^2$ at the right end $x = L$. The left and lower ends are fixed in the x-direction and y-direction respectively and the other boundary is free, as shown in Fig. 2. The material parameters are $E_1 = 1000\,\text{N/m}^2, E_2 = 2000\,\text{N/m}^2$, $\eta = 1000\,\text{N/m}^2$, and $v = 0.3$. The length of the plate is $L = 2\,\text{m}$ and the width is $w = 1\,\text{m}$. The analytical solutions of displacements are given as follows:

$$u_x = Px\left(\frac{1}{E_2} + \frac{1}{E_1}(1 - e^{-\frac{\eta_1}{E_1}t})\right), \tag{21}$$

$$u_y = -vPy\left(\frac{1}{E_2} + \frac{1}{E_1}(1 - e^{-\frac{\eta_1}{E_1}t})\right). \tag{22}$$

In this example, 60 boundary nodes and 171 internal nodes are used, see Fig. 3. In Fig. 4 and Fig. 5, The numerical results for the variation of displacement from t = 0 s to t = 10 s at nodes A (2, 0.5) and B (1, 1) obtained by the present method are compared with analytical results. We can observe a very good agreement of both results.
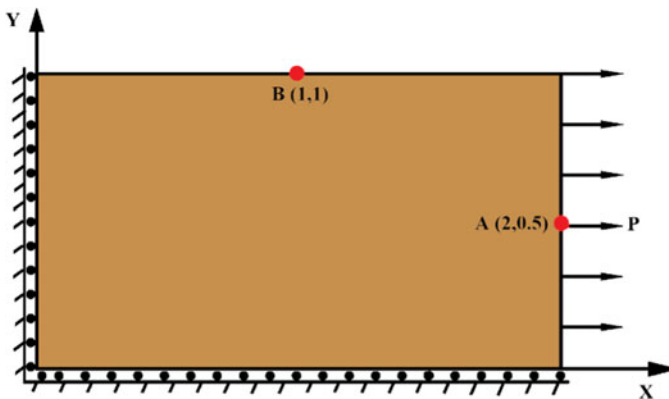
Fig. 2 The geometry for example 1

Fig. 3 The nodes distribution for example 1
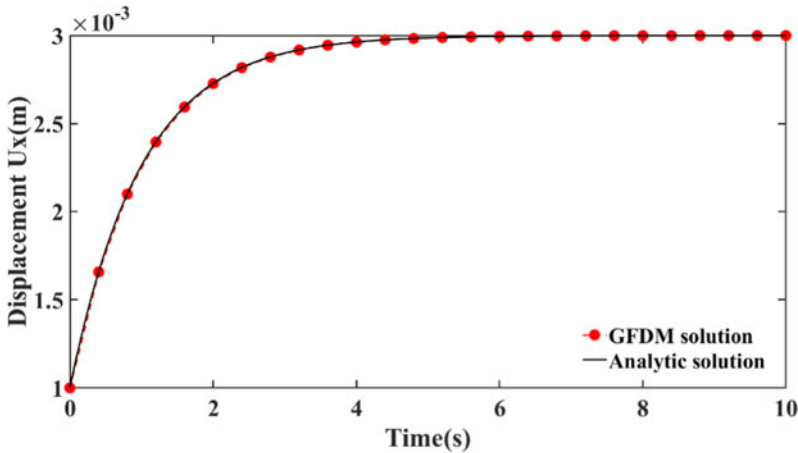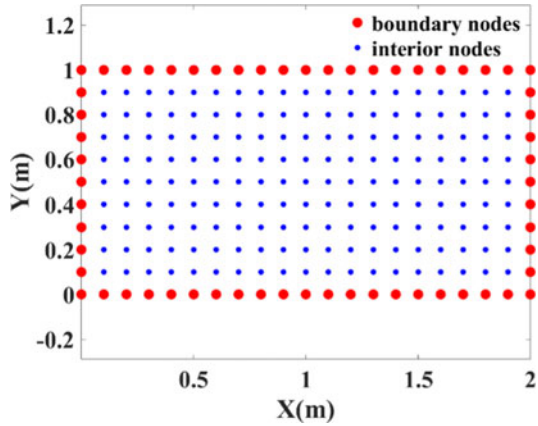
Fig. 4 The variation of displacement at point A (2, 0.5) from t = 0 s to t = 10 s

3.2 Example 2: Relaxation Analysis of a Square Plate with a Circular Hole

The second example is a square plate with a circular hole at the center subjected to a uniform displacement $u_0 = 5 \times 10^{-4}$ m, as shown in Fig. 6. The side length is $W = 20$ m, and the radius of circular hole is $R = 5$ m. The material parameters are given as: $E_1 = 1000\,\text{N/m}^2, E_2 = 2000\,\text{N/m}^2, \eta_1 = 1000\,\text{N/m}^2$, and $v = 0.3$.

Due to the symmetry of the model, only one quarter of the plate is considered as shown in Fig. 7. The total number of nodes is 813 as shown in Fig. 8. As there is no analytical solution, a convergent ABAQUS viscoelastic solution with 98,102 element nodes is used as reference. In the ABAQUS model, only quadrilateral element is used.
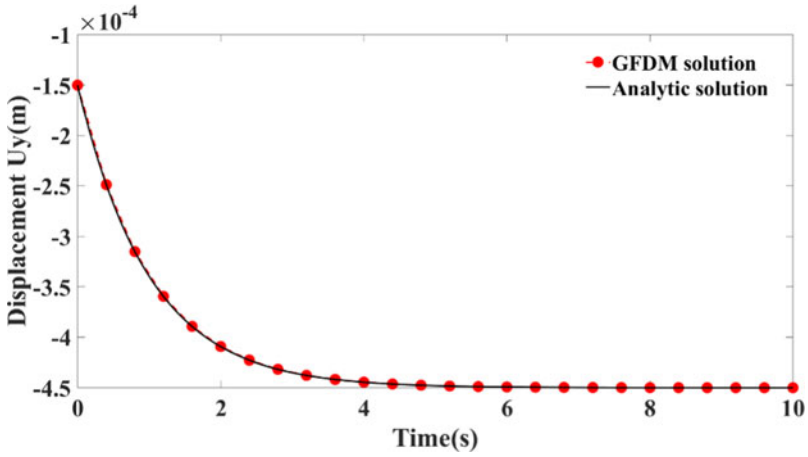
Fig. 5 The variation of displacement at point B (1, 1) from t = 0 s to t = 10 s
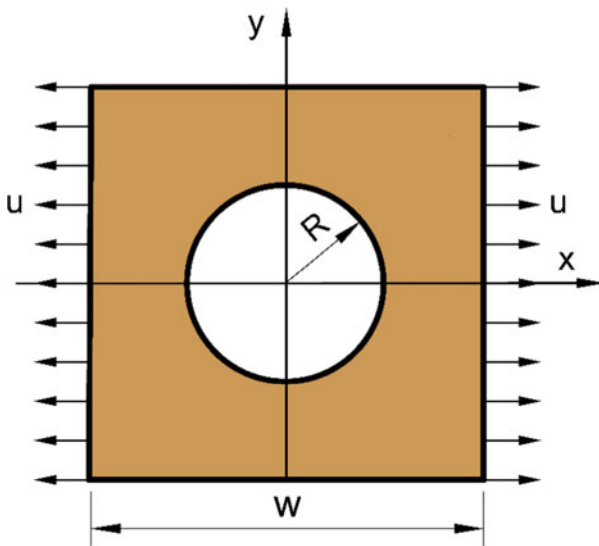
Fig. 6 Geometry model for example 2

Since there is stress concentration in the circular hole of the plate, it is difficult to obtain the accurate results of stress in the circular hole. The fourth-order Taylor series is used to obtain the accurate numerical results in the numerical procedures of

Fig. 7 Simplified model for example 2
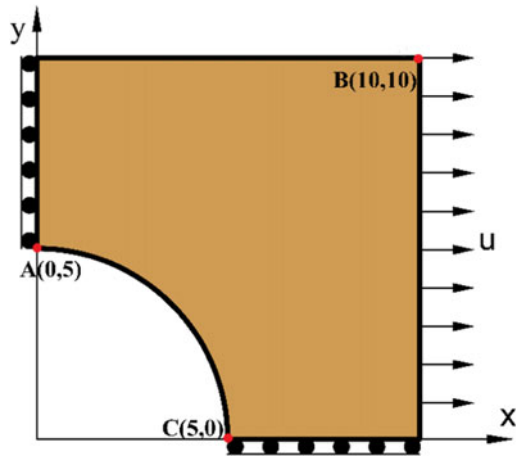
Fig. 8 The nodes distribution for example 2
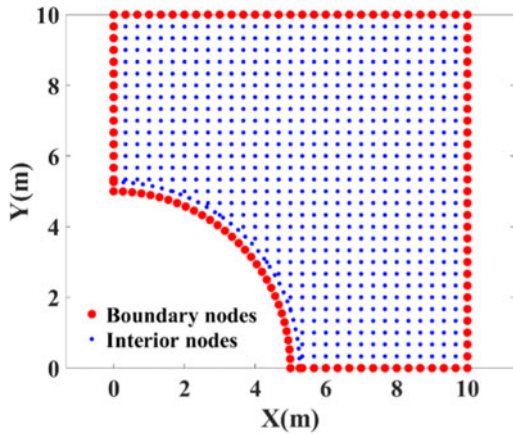
GFDM. In Fig. 9, Fig. 10 and Fig. 11, the variation of stress from t = 0 s to t = 2 s at nodes A (0, 5), B (10, 10) and C (5, 0) obtained by the present method are compared with the reference results. We can observe that the results obtained using GFDM are in good agreement with the ABAQUS solution.
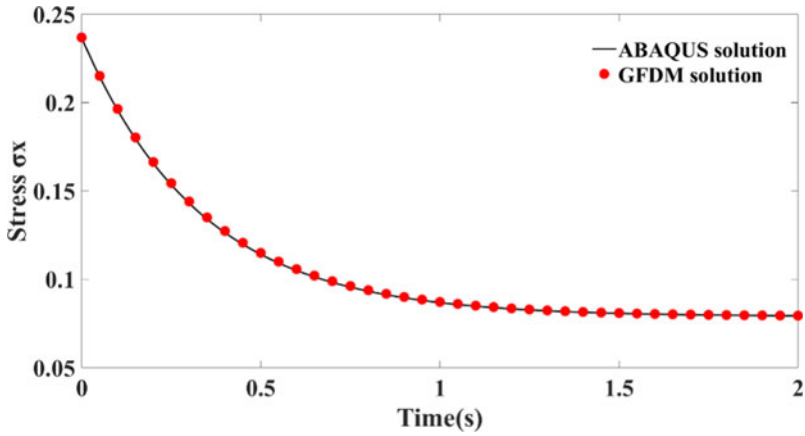
Fig. 9 The variation of stress at points A (0, 5) from t = 0 s to t = 2 s
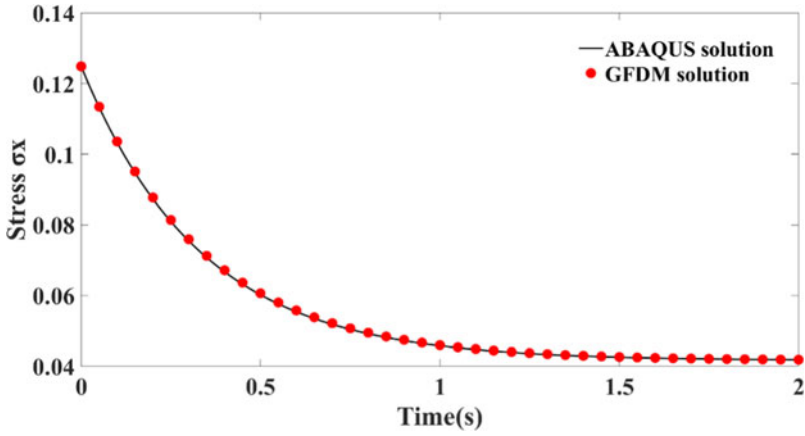
Fig. 10 The variation of stress at points B (10, 10) from t = 0 s to t = 2 s
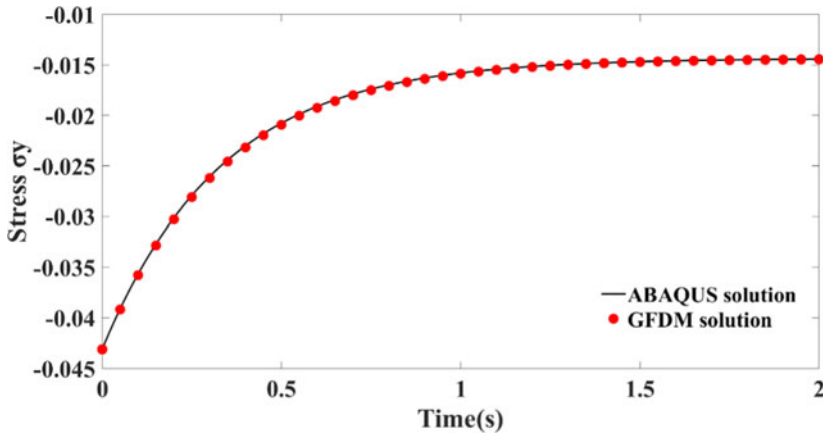
Fig. 11 The variation of stress at points C (5, 0) from t = 0 s to t = 2 s

4 Conclusions

In this paper, the GFDM and the implicit Euler method are adopted to discretized spatial and temporal domain of viscoelastic problem respectively. As the viscoelastic material are often accompanied by continuous shape deformation, the meshless method is a better choice. GFDM is a newly-developed domain-type meshless method, which really avoids time-consuming mesh generation and numerical integration. So it has a simple form and high calculation efficiency. Two examples are proposed to prove the simplicity and efficiency of the proposed method for viscoelastic problems.

References

1. Chen, W.H., Lin, T.C.: Dynamic analysis of viscoelastic structures using incremental finite element method. Eng. Struct. **4**(4), 271–276 (1982)
2. Mesquita, A.D., Coda, H.B.: Alternative Kelvin viscoelastic procedure for finite elements. Appl. Math. Model. **26**(4), 501–516 (2002)
3. Yang, H., Guo, X.: Perturbation boundary–finite element combined method for solving the linear creep problem. Int. J. Solids Struct. **37**(15), 2167–2183 (2000)
4. Mesquita, A.D., Coda, H.B., Venturini, W.S.: Alternative time marching process for BEM and FEM viscoelastic analysis. Int. J. Numer. Method Eng. **51**, 1157–1173 (2001)
5. Haitian, Y., Yan, L.: A combined approach of EFGM and precise algorithm in time domain solving viscoelasticity problems. Int. J. Solids Struct. **40**(3), 701–714 (2003)

6. Sladek, J., Sladek, V., Zhang, C., et al.: Meshless local Petrov-Galerkin method for continuously nonhomogeneous linear viscoelastic solids. Comput. Mech. **37**(3), 279–289 (2006)
7. Miao-Juan, P., Qian, L.: Improved complex variable element-free Galerkin method for viscoelasticity problems. Acta Physica Sinica **63**(18), 180203–180203 (2014)
8. Peng, M.J., Li, R.X., Cheng, Y.M.: Analyzing three-dimensional viscoelasticity problems via the improved element-free Galerkin (IEFG) method. Eng. Anal. Bound. Elem. **40**, 104–113 (2014)
9. Benito, J.J., Urena, F., Gavete, L.: Influence of several factors in the generalized finite difference method. Appl. Math. Model. **25**(12), 1039–1053 (2001)

Application of Near Infrared Spectroscopy in Mixed Spectrum Detection

Jianing Jiang and Jun Liu

Abstract The samples of lemon yellow and carmine cream in cream were analyzed by near infrared spectroscopy combined with machine learning model. Fourier transform infrared spectrometer InfraLUM FT-12 was used to sample the spectral data of lemon yellow and carmine cream samples, and the spectral data of lemon yellow and carmine were made into a mixed sample set. The data label of lemon-yellow sample set was marked as 0, and the label of carmine cream sample set was marked as 1. The best preprocessing of NIR spectral data is achieved by means of Mean Centralization (MC), Standard Variable Transformation (SNV) and Savitzky-Golay convolution smoothing (SG). K-S algorithm is used to divide the preprocessed spectral data into training data set and test data set. Based on XGBoost algorithm, the feature matrix PCA model is used to reduce the spectral dimension. Nine kinds of machine learning methods are used to establish the hybrid spectral NIR identification model and judge the performance of the model. This experiment shows that the best result can be obtained by using principal component analysis based on XGBoost. The F1 value of the model reaches 83.33%, Cohen's kappa coefficient reaches 76.47%, and AUC value reaches 93.33%.

Keywords NIRS · XGBoost · Random forest

1 Introduction

With the progress of society and the improvement of living standards, people have higher and higher requirements for the quality and safety of food. Food safety data contains a lot of information. How to use data analysis and mining technology to solve the problems faced by food safety industry has become the focus of current

J. Jiang · J. Liu (✉)
Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China
e-mail: 2746780322@qq.com

J. Jiang · J. Liu
School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China

research. With the development of food safety field, the large-scale, multi-type and high-dimensional characteristics of data have seriously affected the processing efficiency of relevant technologies. In recent years, near-infrared spectroscopy (NIR) technology has developed rapidly. NIR spectroscopy can reflect the frequency doubling and frequency combination of molecular vibration of organic molecules containing H groups (C-H, O-H, N-H, etc.) and other groups (such as $C = O$, etc.), so as to obtain complex spectral information in the wavelength range of 4000–14,000 cm^{-1}, and can quickly determine the content and characteristics of chemical components in samples. Compared with traditional chemical analysis methods, near-infrared spectroscopy has the advantages of fast, simple, sensitive, green and non-destructive. It has been successfully applied in food industry, agriculture [1], biochemical industry [2] and medicine [3]. It has become an indispensable rapid, efficient and non-destructive detection technology in the industrial production process, and has broad application prospects. Many scholars at home and abroad have used near-infrared spectroscopy technology to identify material varieties, such as molecular detection [4], pills rapid determination [5], and fresh milk. Although the advantages of spectrum technology in variety identification are outstanding, near-infrared technology also has its limitations. The spectral information includes the spectral response caused by the temperature, surface texture, density and uneven distribution of internal components of the measured object, resulting in overlapping of spectral information, complex spectral bands and low absorption intensity. The established prediction model has high complexity and low accuracy. Therefore, the criteria can be determined according to the specific purpose, and a smaller feature subset can be selected from all the features of the data set to minimize the prediction error of the model. At present, the methods of modeling spectral data and extracting spectral feature subset are compared with interval partial least squares (iPLS) and principal component analysis (PCA), However, the use of PCA and iPLS requires a high amount of calculation, and is very sensitive to the selection of initial values, interval selection and evaluation criteria, with strong randomness. In recent years, some researchers have used ant colony optimization algorithm to select the characteristic wavelength, but it takes a long time for the ant colony algorithm to find the optimal path, and the model complexity is high. Therefore, in this paper, principal component analysis and ensemble learning XGBoost were used to extract spectral feature subsets, and the feasibility of non-destructive identification of lemon yellow and carmine cream samples was discussed.

2 Material and Methods

2.1 Data Preprocessing

Preprocessing can help to reduce the influence of non-target factors on spectral data, retain effective information and reduce the complexity of the model. Therefore, the original data are preprocessed by means centralization, multiple scattering correction and SG smoothing. Mean centralization is to subtract the average spectrum of each sample from each spectral data. Where n is the total number of samples, i = 1, 2,…, n samples, k = 1, 2,…, m wavelength points. The formula is as follows:

$$x_{\text{centered}} = x - \frac{\sum_{i=1}^{n} x_{i,k}}{n} \tag{1}$$

Multiple scattering correction (MSC) can eliminate the influence of optical path change or sample dilution on spectral response. The calculation method is as follows:

(1) Calculate the average value of each column of data

$$\overline{A}_{i,j} = \frac{\sum_{i=1}^{n} A_{i,j}}{n} \tag{2}$$

(2) The least square method is used to calculate m_i, b_i

$$A_i = m_i \overline{A} + b_i \tag{3}$$

(3) Organized

$$A_{i(MSC)} = \frac{(A_i - b_i)}{m_i} \tag{4}$$

SG smoothing is a convolution smoothing method proposed by savitzky and Golay.SG smoothing effect is achieved by selecting a window with a fixed width. The wavelength of the sub variable in the center of the window is k, the wavelength of the independent variable on the left side of the window is k−w, and the wavelength of the independent variable on the right side of the window is k+w. there are 2 * w + 1 independent variables in the window width range h_i is the smoothing coefficient at each corresponding wavelength point. The result of each corresponding wavelength point is multiplied by the average value of corresponding smoothing coefficient.The formula is as follows:

$$x_{k,\text{smooth}} = \overline{x}_k = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+i} h_i \tag{5}$$

2.2 Sample Set Division

Kennard-selection algorithm (K-S algorithm) is used to treat all spectral data after preprocessing as candidate samples of training set. In order to select the samples into the training set, the representative samples with large spectral difference are selected into the training set, while the other samples with similar spectrum are selected into the test set. To a certain extent, the uneven distribution of samples in the training set is avoided, and the applicability and accuracy of the model are ensured. The formula is as follows, where Xp and Xq represent two different samples, and N represents the number of spectral wave points of the sample.

$$d_x(\mathrm{p},\, \mathrm{q}) = \sqrt{\sum\nolimits_{\mathrm{j}=1}^{\mathrm{N}} \left[x_{\mathrm{p}}(\mathrm{j}) - x_{\mathrm{q}}(\mathrm{j})\right]^2}; \mathrm{p}, \mathrm{q} \in [1, \mathrm{N}] \tag{6}$$

2.3 Feature Extraction Based on XGBoost Model

XGBoost is an integrated algorithm optimized on the basis of AdaBoost (Adaptive boosting) and GDBT (Gradient boosting). In order to prevent over fitting learning, XGBoost uses the quadratic Taylor expansion of the loss function and adds a regular term to control the complexity of the model, which represents the complexity of the generated regression tree. At the same time, in the process of model optimization, XGBoost uses the second-order approximation of loss function $L(\phi)$ to improve the operation efficiency, which is different from the gradient lifting decision tree model only using the first derivative. Similar to the gradient lifting decision tree model, the model with k trees can be expressed as, where i = 1, 2,…, n samples and m features:

$$\widetilde{y_i} = \emptyset(X_i) = \sum\nolimits_{k=1}^{K} f_k(X_i), f_k \in F \tag{7}$$

Among them, $F = \left\{f(x) = W_{q(x)}\right\}$(q: $R^m \to T, \omega \in R^T$) F is the function space composed of trees, q is the mapping of input feature X to leaf nodes, and ω is the leaf weight. The loss function is as follows: $L(\phi) = \sum_i l(\widetilde{y_i}, y_i) + \sum_k \Omega(f_k)$ among them, $\Omega(\mathrm{f}) = \Upsilon T + \frac{1}{2}\lambda\|\omega\|^2$, T represents the number of leaf nodes.

2.4 Principal Component Analysis

PCA is an unsupervised learning algorithm, which uses orthogonal transformation to transform the observed data represented by linearly correlated variables into a few data represented by linearly independent variables, and the linearly independent

variables are called principal components. The number of principal components is usually less than the number of original variables, so principal component analysis belongs to dimension reduction method, which is used for preprocessing of other machine learning methods. It can also be used for data de-noising. By denoising, the principal component data in the original data can be retained, and the interference data can be removed to improve the accuracy and stability of the model.

2.5 Model Evaluation Criteria

The commonly used evaluation indexes are Accuracy, Precision (P), Recall (R), F1 value and Cohen's kappa coefficient, AUC value and ROC curve. For the binary classification problem, the samples are divided into four cases according to the combination of real class and learner prediction category: True case TP, False counterexample FN, False positive case FP, True counterexample TN. obviously, TP + FP + FN + TN = total sample number. The higher the Precision ratio, the lower the Recall ratio; on the contrary, the higher the Recall ratio, the lower the Precision ratio. F1 value combines Precision and Recall, and is defined based on harmonic average of Precision and Recall. When the Precision ratio and Recall ratio are very high, the F1 value will be particularly high. When one value is particularly low and the other is particularly high, the F1 value obtained is particularly low. The closer the evaluation index is to 1, the better the model performance is. The formula is as follows:

$$Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \tag{8}$$

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

In Cohen's kappa coefficient, p_0 represents the overall classification accuracy, the sum of the number of correctly classified samples divided by the total number of samples. $p_e = \frac{a1 \times b1 + a2 \times b2}{n \times n}$, a1 and a2 are the real samples of lemon yellow and carmine cream, b1 and b2 are the predicted number of lemon yellow and carmine cream samples, and n is the total number of samples. The formula is as follows:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{12}$$

The vertical axis of ROC curve is TPR, and the horizontal axis is FPR. AUC is defined as the area enclosed by the coordinate axis under the ROC curve. The value range of AUC is between 0.5 and 1. The closer AUC is to 1.0, the better the model performance is; when it is equal to 0.5, the model performance is the lowest and has no application value. TPR and FPR are defined as:

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

$$FPR = \frac{TP}{TN + FP} \tag{14}$$

3 Experiment

3.1 Sample Collection and Instrument

The experimental instruments include fresh cream, edible pigments (lemon yellow and carmine), refrigerator, cream beater, Sartorius CP224S electronic balance with weighing accuracy of 0.0001 g, electronic wet thermometer, InfraLUM FT-12 Fourier transform infrared spectrometer of Russian LUMEX company. Sample preparation: weigh the pigment, record, add the whipped cream, weigh, record, stir evenly, put it into a 40 cm^3 sample container, and scan it in an infrared spectrometer. Each sample was scanned 3 times to obtain its reflectance spectrum in the range of $4000–14000 \text{ cm}^{-1}$, with a spectral resolution of 8 cm^{-1}. And use the average spectrum of these three reflection spectra as the spectrum of the sample. During the entire sample scanning process, the room temperature was kept at 20–25 °C.

3.2 Raw Spectral Data

The original spectral lines of the near-infrared spectra of the two pigments (lemon yellow and carmine) and their mixed sample set are shown in Fig. 1. It can be seen from the figure that in the range of $4000–14000 \text{ cm}^{-1}$, samples with different concentrations of the same pigment have similar spectral curves, with only slightly different relative absorption intensities between samples. Therefore, the absorbance intensity of the same pigment sample depends on the pigment concentration. The absorbance range of different types of pigments is different. It can be seen from
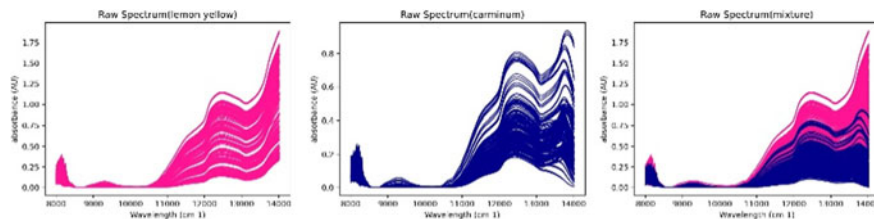
Fig. 1 Original spectrum

Fig. 1 that the absorbance range of lemon yellow is between 0–1.75, and the absorbance range of carmine is between 0–0.8. Mixing two pigments for qualitative analysis will increase spectral complexity.

4 Results and Discussion

4.1 Spectral Characteristics and Spectral Pretreatment

At the same time, a variety of different pretreatment analysis methods were performed on 192 samples to determine the appropriate pretreatment method. The experimental results are shown in Table 1 When no preprocessing operation is performed at all, the F1-score, Cohen's Kappa coefficient, and AUC values of the model are only 0.5581, 0.3391, 0.6625, and the model performance is poor. The combination of different data preprocessing methods can effectively remove the data noise information, and the modeling effect is the best. As shown in Table 1, the F1-score reaches 0.8292 and the Cohen's Kappa coefficient reaches 0.7522 after the MSC, SG smoothing and centralized preprocessing methods are shown in Table 1, AUC value reaches 0.8542. Therefore, the spectral data smoothed and centralized by MSC and SG are used as research data.

4.2 K-S Algorithm

Figure 2 shows the KS algorithm from three perspectives. The training set and the test set are divided into sample sets according to 2:1. The PCA algorithm is used to extract all the spectral data to the third principal component number. Figure 2 shows the effect of K-S algorithm from three perspectives. The training set and test set are divided into 2:1 sample set. PCA algorithm is used to extract all spectral data to the third principal component. The lavender circle represents 192 samples, and the red five pointed star represents 128 samples selected for calibration. It can be seen from the graph that K-S algorithm takes all the representative samples with

Table 1 Qualitative analysis under different pretreatments

Pretreatments	f1_score	Cohen_kappa_score	AUC
None	0.5581	0.3391	0.6625
MSC	0.6	0.4182	0.7091
MSC+MC	0.7273	0.6327	0.8104
MC+SG	0.7381	0.6489	0.7925
MSC+MC +SG	0.8292	0.7522	0.8542

Fig. 2 Partition of K-S algorithm sample set

3D Visualization of K-S algorithm
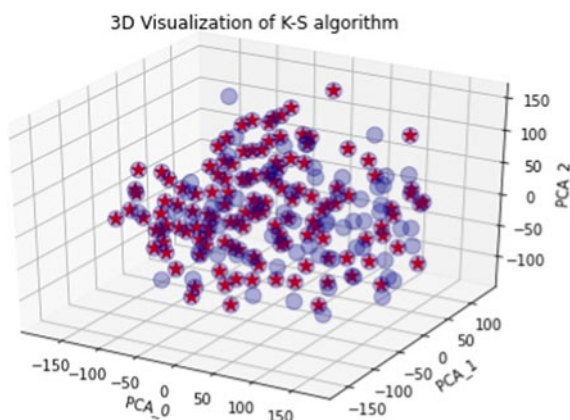
large spectral difference into the calibration set, while the other samples with similar spectra enter the verification set, which avoids the uneven distribution of samples in the calibration set to a certain extent, and improves the applicability and accuracy of the model.

4.3 Modeling Method and Model Evaluation Index

In the experiment, the training data set was used to train KNN, GradientBoostingClassifier, LogisticRegression, MLP, SVC, SGD, CART, AdaBoostClassifier and RandomForestClassifier models, and the trained model was used to make predictions. In order to ensure reasonable experimental comparison results. Each method uses the same training set and test set, and in order to ensure the objectivity of the overall result analysis conclusion, multiple evaluation indicators are used likes accuracy, precision, recall, f1_score, cohen_kappa_score, AUC. It can be clearly seen from Table 2 that compared with KNN, GradientBoostingClassifier,

Table 2 Recognition results of different models combined with principal component analysis and integrated learning XGBoost method

Data type	Accuracy	P	R	f1_score	Cohen's kappa score	AUC
KNN	0.85	1.0	0.4	0.5714	0.5	0.7
GDBT	0.8	0.6	0.6	0.6	0.4667	0.7333
LogisticRegression	0.85	0.75	0.6	0.6667	0.5714	0.7667
MLP	0.8	0.5714	0.8	0.6667	0.5294	0.8
SVC	0.9	1.0	0.6	0.75	0.6923	0.8
SGD	0.85	0.6667	0.8	0.7273	0.625	0.8333
CRAT	0.8	0.5556	1.0	0.7143	0.5789	0.8667
AdaBoost	0.85	0.625	1.0	0.7692	0.6667	0.9
RandomForest	0.9	0.7143	1.0	0.8333	0.7647	0.9333

Fig. 3 ROC curve of random forest
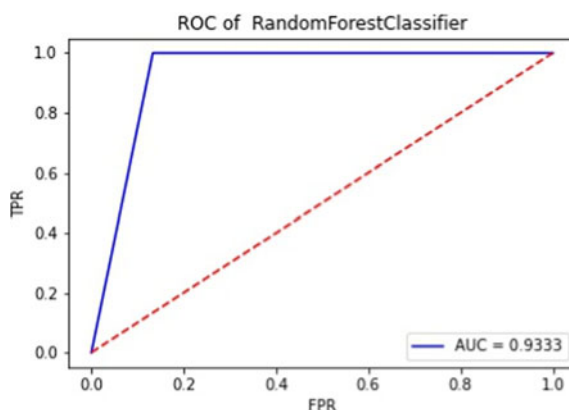
LogisticRegression, MLP, SVC, SGD, CART, AdaBoostClassifier, the evaluation indicators of the RandomForestClassifier model: accuracy rate reaches 0.9, precision rate reaches 0.7143, recall rate reached 1.0, F1 value reached 0.8333, Cohen's Kappa coefficient reached 0.7647, AUC value reached 0.9333, better than other models. The research results show that the application of spectral preprocessing methods (Mean Centralization, standard variable transformation, Savitzky-Golay convolution smoothing), KS algorithm, XGBoost-PCA algorithm and random forest machine learning model have a good qualitative analysis of mixed pigments in cream samples effect and feasibility.

When FPR = 0 and TPR = 1, that is, the point (0, 1) in the figure, the closer the ROC curve is to the point (0, 1) and the more it deviates from the 45-degree diagonal, the better the classification effect of the model. It can be clearly seen from Fig. 3 that the ROC curve and AUC value of the random forest algorithm model have achieved very good classification results.

5   Conclusion

In this paper, we proposed Principal Component Analysis (PCA) dimension reduction algorithm based on XGBoost. XGBoost algorithm is used for feature selection, and PCA algorithm is used to reduce the dimension of the processed data matrix. The proposed algorithm is compared with different models to verify the accuracy of classification in cream sample set detection. The results show that the method based on XGBoost-PCA can effectively enhance the stability and accuracy of characteristic wavelength selection of mixed spectral data of lemon yellow and carmine cream. In the experiment, the prediction accuracy value of the random forest classification model is the highest by comparing the prediction and evaluation indicators of different models.Using near infrared spectroscopy technology combined with XGBoost-PCA-random forest model can meet the needs of practical application. In the follow-up research tasks of the experiment, we will further study the algorithm improvement of the model mentioned in this paper to reduce the algorithm complexity of the model and reduce the consumption of computing resources.

References

1. Tan, Z.L., Wu, M.C., Fu, T.D.: Application review of near infrared spectroscopy technology on quality analyzing of agricultural products. Chin. J. Oil Crop Sci. **34**(04), 455–460 (2012)
2. Chen, H., Tan, C., Lin, Z.: Quantitative determination of the fiber components in textiles by near-infrared spectroscopy and extreme learning machine. Anal. Lett. **53**(6), 844–857 (2020)
3. Chen, H., Lin, Z., Tan, C.: Nondestructive discrimination of pharmaceutical preparations using near-infrared spectroscopy and partial least-squares discriminant analysis. Anal. Lett. **51**(4), 564–574 (2018)
4. Mes, S.W., Heideman, D.A.M., Bloemena, E., et al.: Development and validation of a novel and rapid molecular detection method for high-risk human papillomavirus in formalin-fixed. Paraffin-Embed. Tumor Tissue **22**(2), 262–271 (2020)
5. Bai, Y., Gong, H.Y., Li, X.Q., et al.: Rapid determination of paeoniflorin and moisture in Xiaoyao pills (concentrated) by near-infrared spectroscopy. Adv. Mater. Res. **2695**, 1972–1977 (2013)

Production Line Vacuum Robot Based on Image Processing and Fuzzy PID Control

Qian Guo and Baocheng Yu

Abstract Aiming at the problems such as low efficiency of manual cleaning and easy wear of production line, a production line vacuuming robot based on image processing and fuzzy PID control was studied and designed. First of all, the dust image of the conveyor belt is collected by the camera. After the original image is preprocessed, the low-frequency image and the high-frequency image are obtained by using wavelet transform to decompose the two-dimensional dust image of the conveyor belt. The grayscale symbiosis matrix of the low-frequency subband is constructed. Then, the speed of DC motor is controlled by fuzzy PID algorithm. Through the experiment to test the reliability of the design, implementation along the conveyor belt production line robot automatic vacuuming, combining wavelet transform and gray level co-occurrence matrix algorithm described the conveyor dust image texture features, has good recognition effect, and the simulation results show that compared with the traditional PID control, fuzzy PID control dc motor speed control precision is higher, quantitative evaluation production line vacuuming robots vacuuming effect, dust collection efficiency can reach above 93.5%.

Keywords Production line vacuum robot · Fuzzy PID control · Image processing · Dust detection · Gray level co-occurrence matrix · Wavelet transform

1 Introduction

The world has entered the era of the fourth Industrial Revolution. Based on Industry 4.0, the main goal is to realize unmanned chemical plants and intelligent industrial equipment. After working for a long time, a large amount of dust and debris will

Q. Guo (✉)
School of Computer Science and Engineering, Wuhan University of Technology, Wuhan, Hubei, China
e-mail: 2413128731@qq.com

B. Yu
Intelligent Robot Hubei Key Laboratory, Wuhan, Hubei, China

accumulate on the transformer detection line in the actual factory, which will cause the damage of the detection line. Traditionally, manual cleaning is difficult, wastes human and material resources, and causes serious abrasion of the verification line. In order to improve the detection accuracy, wavelet transform and GLCM are combined to extract image texture features [1, 2]. In addition, traditional PID control algorithm cannot meet the requirements of dc motor starting and running, which is mainly characterized by weak stability, slow response speed, and difficult control of overshoot. Many scholars have carried out a large number of studies and combined fuzzy control and PID control to control the DC motor, achieving good results [3]. In this paper, a production line vacuuming robot based on image processing and fuzzy PID control is developed. It can clean the dust on the verification line online without causing damage to the verification line, thus improving the work efficiency of the production line vacuuming robot and saving energy.

2 Hardware System Design

2.1 Overall Design

The main task of the intelligent dust cleaning robot control system of the production line is to match the template according to the images collected by the camera, and then control the dust collection strength of the robot and complete various control actions. The intelligent dust cleaning robot of production line is controlled by STM32 single chip microcomputer and arm cortex is used for STM32 single chip microcomputer Gamma-M3 processor, as the core of the control system of dust cleaning robot, has extremely high computing ability and strong interrupt response ability [4]. STM32 MCU control system includes camera module, motor drive module, wireless communication module, power module, LED indicator and LCD module.

2.2 Motor Module

The motor used in this paper is a brush DC motor, which is driven by H-bridge driver chip. The motor is used in two different positions, one for a centrifugal fan to suck in dust and send it to the dust collection chamber. The other is used to sweep the dust and paper scraps away from the conveyor belt and suspended in the air. The single chip microcomputer sends out the motor steering control signal, the motor speed control signal, the motor brake control signal, the logic circuit carries on the logic operation, the driving signal is amplified and connected to the H-bridge circuit, and then controls the start stop, positive and reverse rotation and speed of

the motor, and is fed back to the single chip microcomputer after photoelectric coding.

2.3 Sensor Module

In order to facilitate the calculation of dust concentration at the inlet and outlet of dust collection, the power supply uses dust sensor to detect the measured value of dust and displays it on the display. The principle of dust sensor dust collection is: the sensor center emits LED light directionally, detects the light after dust refraction, and judges the dust content at the inlet and outlet of the production line dust collection robot. The dust concentration is converted into light signal, and then converted into analog signal. Because the single-chip microcomputer on the intelligent robot can not recognize the analog signal, it is necessary to use A/D converter to convert the analog signal into digital signal.

3 Dust Treatment of Conveyor Belt Surface and Motor Control

3.1 Image Preprocessing

After the hardware structure of the robot is designed and the dust collection function of the production line intelligent dust cleaning robot is realized, the image processing is also needed, so that the production line intelligent dust cleaning robot can control the dust collection intensity according to the image similarity, further improve the efficiency and save energy. The quality of the image will affect the analysis of the image, therefore, before the image analysis, it is necessary to preprocess the image. In this paper, the reason of image preprocessing is image graying, image denoising and image sharpening.

In this paper, the reason for image preprocessing consists of three parts. Firstly, the image is grayed, and the collected image is transformed into gray image to reduce the amount of data to be processed. In this paper, the average method is used to grayscale the image, where R, G and B represent the components of the color image. The formula of average method is as follows:

$$f(x, y) = (R(x, y) + G(x, y) + B(x, y))/3 \tag{1}$$

Secondly, there will inevitably be some noise in image acquisition and image transmission. Therefore, it is necessary to remove image noise. In this paper, median filtering is used to remove image noise. All pixels in the window with odd number of points in a pixel point are arranged into a sequence, in which, is the

median of the sequence, and their median value is calculated by the following formula.

$$g_i = Med\{f_1, f_2, f_3, \dots, f_{2n+1}\} \tag{2}$$

Finally, for the problem of unclear image edge, we need to use image sharpening algorithm to deal with it. After comparing various methods, it is found that Sobel operator method has better sharpening effect. Therefore, this paper uses Sobel operator method for image sharpening. The calculation formula of this method is as follows:

$$\Delta_x f = [f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)] - [f(x-1, y-1) \\ + 2f(x, y-1) + f(x+1, y-1)] \tag{3}$$

$$\Delta_y f = [f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)] \\ - [f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)] \tag{4}$$

$$g(x, y) = |\Delta_x f| + |\Delta_y f| \tag{5}$$

Where g(x, y) is the gray value function of two-dimensional graphics, is the horizontal convolution, is the vertical convolution.

The camera collects the dust image on the conveyor belt and takes the dust image on the conveyor belt as the object. The original image is shown in Fig. 1.

After graying the original image collected by the camera, the grayed image is shown in Fig. 2.

Then, the gray-scale image is denoised, and the noise on the image is basically eliminated. The denoised image is shown in Fig. 3.

The denoised image is sharpened, and the sharpened image is shown in Fig. 4.

3.2 Wavelet Transform

After image preprocessing, the next step is to analyze the image, and finally evaluate the dust collection effect of the production line vacuum robot. In this paper, wavelet transform and gray level co-occurrence matrix are used to extract the
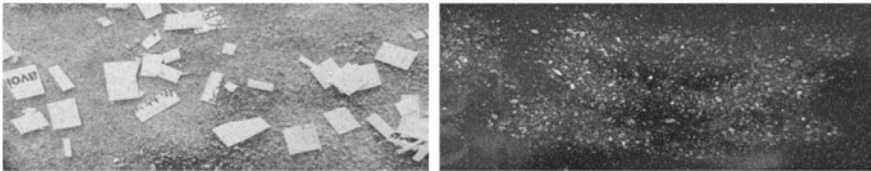
Fig. 1 Original image
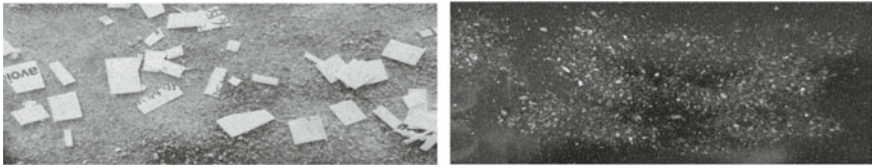
Fig. 2 Image after graying
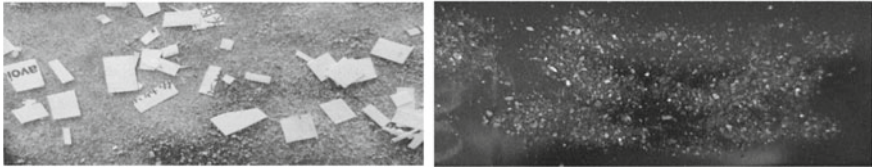
Fig. 3 Image after denoising
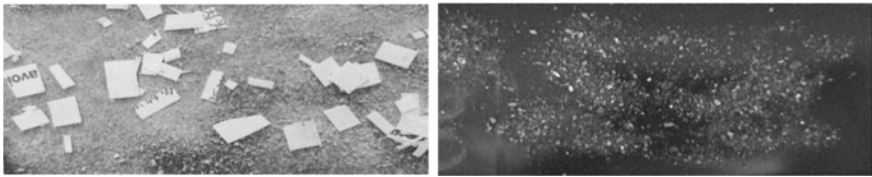
Fig. 4 Image after sharpening

texture features of dust image on conveyor belt. Wavelet transform can describe the local characteristics of the image in time and frequency domain at the same time. The dust image on the conveyor belt is processed by two-dimensional discrete wavelet transform DWT. DWT uses high pass filter and low pass filter to filter image. After DWT decomposition, the preprocessed image is decomposed into one low-frequency sub-band and three high-frequency sub-band. The low-frequency sub-band contains most of the information of the dust image of the conveyor belt, and the high-frequency word band includes the high-frequency information in the horizontal, vertical and diagonal directions. The sub images of the preprocessed image after DWT decomposition are shown in Fig. 5.

3.3 GLCM Algorithm

GLCM is a method to describe texture features by calculating the spatial correlation of gray level. Suppose the image size is M × N, i, j are the gray value of pixels, GLCM is used to calculate the probability and frequency of the occurrence of pixel

(a) Low frequency subband

(b) Horizontal direction

(c) Vertical direction

(d) Diagonal direction
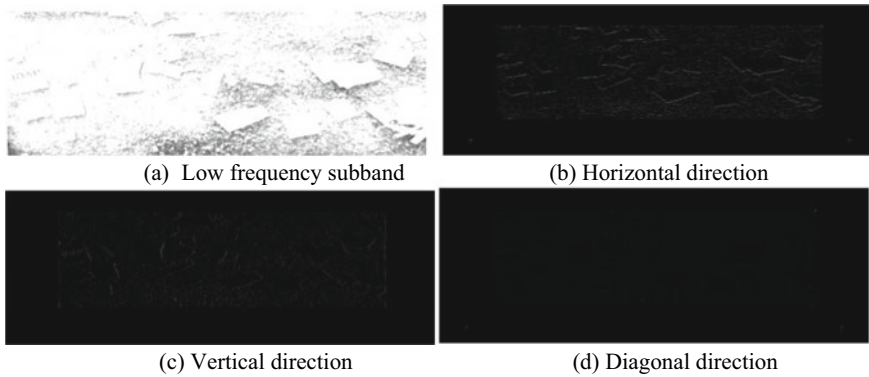
Fig. 5 Low frequency sub image and high frequency sub image

(a) Similarity between template image and dust-free image

(b) Similarity between template image and dust image
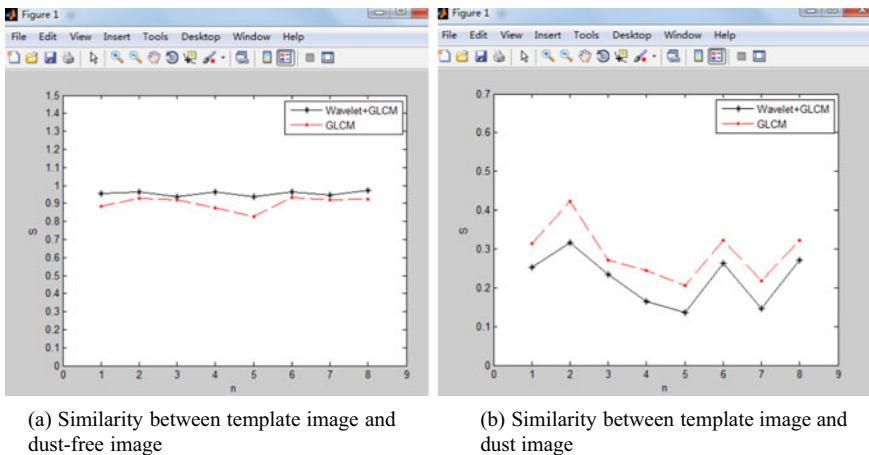
Fig. 6 Similarity between real-time image and template image

pairs, and θ(0°, 45°, 90°, 135°) is the direction between pixel pairs, and the distance between pixel pairs is d. the probability matrix p(i,j,d,θ) can be obtained by traversing all the pixels in the dust image, The gray level co-occurrence matrix can be expressed as follows:

$$P(i,j,d,\theta) = \{[(x,y),(x+\Delta x, y+\Delta y)|f(x,y) = i, f(x+\Delta x, y+\Delta y) = j]\} \quad (6)$$

Where x, y are the coordinates of the pixel, $1 \leq x \leq M$, $1 \leq y \leq N$, And X and y are integers; Δx, Δy are offset, And $\Delta x = d\cos\theta$, $\Delta y = d\sin\theta$; f(x, y) represents the gray value of the gray image at the pixel (x, y).

Then, the similarity of gray level co-occurrence matrix is calculated.

$$S = 1 - \frac{1}{2} \frac{1}{(M-1)(N-1)} \sum_{i=1}^{0} \sum_{j=1}^{0} A_m(i,j) - A_b(i,j) \tag{7}$$

Where A_m is the co-occurrence matrix of the real-time dust image of the conveyor belt, A_b is the co-occurrence matrix of the template image, and M, N are the number of pixels in the row and column.

3.4 Motor Control of Vacuum Cleaning Robot in Production Line

In this paper, fuzzy PID is used to control the speed of DC motor, and then the wind speed of dust suction port is adjusted. Fuzzy PID control combines traditional PID control with fuzzy control to control DC motor. Traditional PID control is weak in anti-interference and can not realize accurate control of nonlinear system, so it is difficult to adjust PID parameters in real time. Fuzzy PID control uses fuzzy reasoning to adjust PID parameters according to the dynamic changes of E and EC [5]. Firstly, the initial PID parameters Δk_p, Δk_i and Δk_d are determined, Then, according to the speed deviation E and the change rate EC of speed deviation, the fuzzy controller makes fuzzy inference to get the correction values Δk_p, Δk_i and Δk_d, After adding with the initial value, the PID parameters can be adjusted dynamically. The deviation E and the deviation change rate EC are (8) and (9).The fuzzy controller inputs Δk_p, Δk_i, Δk_d into PID controller, and adds the initial value to get the output K_p, K_i, K_d,The calculation results of PID parameters K_p, K_i and K_d are shown in formula (10).

$$e(t) = r(t) - c(t) \tag{8}$$

$$u(t) = k_p[e(t) + \frac{1}{k_i}\int_0^t e(t)dt + k_d\frac{de(t)}{dt}] \tag{9}$$

$$\begin{cases} K_p = K_{po} + \Delta K_p \\ K_i = K_{i0} + \Delta K_i \\ K_d = K_{d0} + \Delta K_d \end{cases} \tag{10}$$

4 Experiment and Simulation

In this paper, the production line vacuum robot is developed to remove dust on the conveyor belt. In order to improve the efficiency of production line intelligent robot and save energy, a camera is installed on the intelligent dust collection line to

collect the dust picture on the conveyor belt. After the image is processed, the dust absorption strength of the production line dust suction robot is controlled according to the similarity between the real-time image and the template image. The level of dust collection is shown in Table 1.

The similarity values of template image, dust free image and dust image are shown in Fig. 6.

Analysis of Fig. 6(a) shows that the similarity between the dust-free image and the template image calculated by this algorithm is greater than that of the GLCM algorithm, and it is closer to 1. Compared with the GLCM algorithm only, the algorithm in this paper has higher accuracy in detecting the similarity value of the dust-free image. From the analysis of Fig. 6(b), it can be seen that the similarity between the dust image and the template image calculated by this algorithm is less than that of the GLCM algorithm. Compared with the GLCM algorithm only, the detection accuracy of the proposed algorithm is higher than that of the GLCM algorithm when detecting the similarity value of the dust image.

In this paper, Matlab/Simulink software is used to simulate the system. In MATLAB software environment, the traditional PID algorithm and fuzzy PID algorithm are used to control the speed of DC motor. The simulation effect of scope is shown in the figure below. Figure 7 shows the simulation results of DC motor control by traditional PID algorithm, in which PID (proportional, integral and derivative) parameters are set to 10, 100 and 1 respectively. It is observed that the overshoot of the simulation results of the traditional PID control DC motor is not easy to control, the stability is not strong, and the operation fluctuates greatly. Figure 8 shows the simulation results of DC motor controlled by fuzzy PID algorithm and traditional PID algorithm. Through comparison, it can be seen that compared with traditional PID algorithm, the fuzzy PID control algorithm has small overshoot, strong stability and smooth operation. When the fluctuation occurs, compared with the traditional PID algorithm, the fuzzy PID algorithm adopted in this paper has faster response speed and stronger anti-interference.

Through the use of dust sensor to detect the dust concentration, the data is processed and transmitted to STM32 MCU, and then the measured value is displayed on the display. After the sensor detection, the concentration of dust particles at the dust suction inlet is, the concentration of dust and debris at the outlet is, and the dust collection efficiency of the intelligent dust suction robot of the production line is:

$$\eta = \left(1 - \frac{C_c}{C_v}\right) \times 100\%$$

(11).

Table 1 Level of dust collection

Similarity S	$0.9 < S \leq 1$	$0.7 < S \leq 0.9$	$0.5 < S \leq 0.7$	$S \leq 0.5$
Intensity of dust collection	Weak	Little Strong	Strong	Very strong

Fig. 7 Simulation results of traditional PID control.
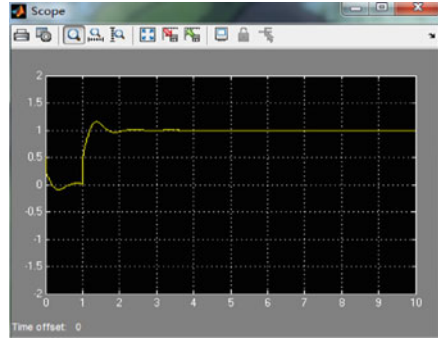
Fig. 8 Simulation results of scope

After calculation, the dust collection efficiency of the robot is shown in the table below:

It can be seen from Table 2 that the camera monitors the dust image on the transformer detection line in real time and transmits the image information to STM32 single chip microcomputer for processing. After image preprocessing, wavelet transform and gray level co-occurrence matrix algorithm are used to retrieve the texture features of detection line dust image. After wavelet transform decomposition, low-frequency and high-frequency sub-band of the image are obtained. Gray level co-occurrence matrix (GLCM) of low-frequency sub-band in four directions is constructed. The similarity S between real-time image and template image is calculated, and template matching is carried out. Using traditional PID algorithm to control motor speed has the disadvantages of high overshoot and poor stability. Therefore, the fuzzy PID algorithm is proposed to control the speed of DC motor, which can effectively solve these problems. The intelligent dust cleaning robot designed in this paper can accurately match the image similarity, accurately adjust the motor speed, and reasonably control the dust suction strength of the robot. After the actual production line test, the dust collection efficiency of the production line intelligent dust cleaning robot can reach more than 93.5%.

Table 2 Dust efficiency table

Experiment number	1	2	3	4	5	6	7	8
Dust collection efficiency	0.957	0.963	0.936	0.965	0.939	0.953	0.976	0.981

(a)   (b)   (c)
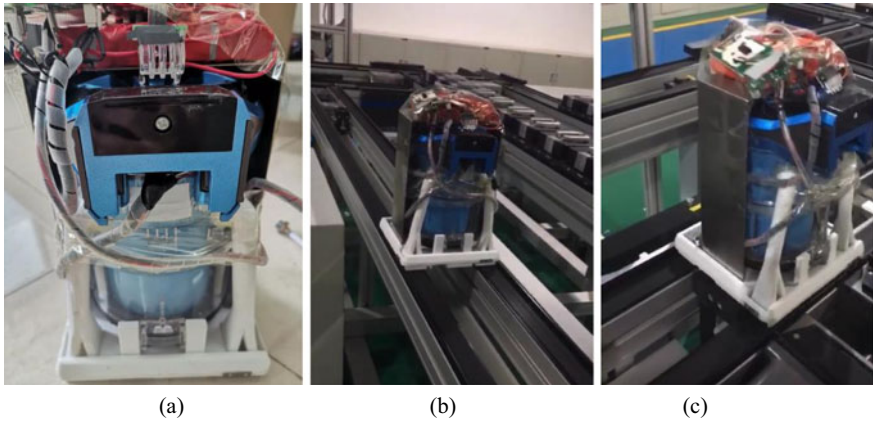
Fig. 9 Test chart of production line robot and actual production line. (a) Product line vacuum cleaning robot. (b) Dust collection test chart of actual production line. (c) Dust collection test chart of actual production line

Part of the components are made by 3D printing technology, and the actual picture of the intelligent dust cleaning robot on the production line is shown in Fig. 9(a). After realizing the function of intelligent dust cleaning robot of production line, vacuum cleaning test of actual production line is carried out for the robot. The test environment is shown in Fig. 9(a)(b).

5 Epilogue

The experimental results show that after image preprocessing, image information can be extracted by combining wavelet transform and gray level co-occurrence matrix algorithm, which can improve the accuracy of similarity calculation. In view of the shortcomings of the traditional PID control system, such as slow response speed and poor anti-interference, a fuzzy PID control algorithm is proposed to control the DC motor. When the speed of the DC motor rises suddenly, compared with the traditional PID algorithm, the fuzzy PID algorithm adopted in this paper has faster response speed, stronger anti-interference, smaller overshoot, stronger robustness and higher stability. The validity of fuzzy PID control algorithm is verified by Simulink

experimental platform. After the actual production line environment test, the robot along the conveyor belt dust removal effect is good, has a relatively high dust collection efficiency, has broad application prospects in industry.

Reference

1. Tudavekar, G., Patil, S.R., Saraf, S.S.: Dual-tree complex wavelet transform and super-resolution based video inpainting application to object removal and error concealment. CAAI Trans. Intell. Technol. **5**(4), 314–319 (2020)
2. Gupta, Y., Lama, R.K., Lee, S.W., Kwon, G.R.: An MRI brain disease classification system using PDFB-CT and GLCM with kernel-SVM for medical decision support. Multimedia Tools Appl. **79**(43–44), 32195–32224 (2020)
3. Wang, S., Shi, Y., Feng, Z.: Research on control method based on Fuzzy PID controller. Mech. Sci. Technol. **30**(01), 166–172 (2011)
4. Hayat, A.A., Karthikeyan, P., Vega-Heredia, M., Elara, M.R.: Modeling and assessing of self-reconfigurable cleaning robot hTetro based on energy consumption. Energies **12**(21), 4112 (2019)
5. Wang, Y.W., Zhang, W.A., Dong, H., Yu, L.: A LADRC based fuzzy PID approach to contour error control of networked motion control system with time-varying delays. Asian J. Control **22** (5), 1973–1985 (2020)

Research on the Prediction of Advertisement Click-Through Rate Based on Logistic Regression Model

Xinzhe Zhang, Wei Qi, Yedan Li, and Yuanyuan Fu

Abstract Search engines use data such as keywords submitted by users and users' information to estimate the click-through rate of search ads, and display search ads to users in a particular order. The accuracy of prediction of search ads click-through rate is related to advertisers' and advertising media's revenue and closely related to advertising audiences' user experience. Therefore, it is critical to improving the accuracy of prediction of search ads click-through rate. This paper uses a logistic regression model to estimate CTR through historical data, evaluates the model's accuracy, and obtains a useful prediction model.

1 Research Background

As a new form of advertising, online advertising shows excellent market potential and commercial value. Compared with traditional advertising, online advertising has certain advantages, such as convenient bidding, fast delivery, and quantifiable effects, especially when the amount of data becomes more and more abundant, online advertising can be delivered more accurately. At present, search advertising is the most extensive online advertising form [1]. It displays related advertisements based on keywords searched by users, and it is also one of the primary sources of income in the Internet industry. In search advertising, the key technology is to predict the click-through rate of advertisements [2]. The prediction accuracy of

search ads click-through rate is related to advertisers' and advertising media's rev-

X. Zhang (✉) · W. Qi · Y. Li · Y. Fu
Harbin Institute of Technology, No. 92, Xidazhi Street, Nangang District, Harbin City, Heilongjiang Province, China

W. Qi
e-mail: qiwei@hit.edu.cn

473

enue and the user experience of advertising audiences, which shows the significance of using historical data to predict the click-through rate of search ads effectively.

2 Model Introduction

2.1 Sigmoid Function

A vital function in the logistic regression model [3] is the sigmoid function, also known as the logistic function. The formula of this function is:

$$g(z) = \frac{1}{1 + \exp(-z)}. \tag{1}$$

For multiple logistic regression, the classification is as follows:

$$P(y = 1|x, \theta) = \frac{1}{1 + e^{-\theta^T x}}. \tag{2}$$

$$P(y = 0|x, \theta) = 1 - P(y = 1|x, \theta) = P(y = 1|x, -\theta). \tag{3}$$

For the training set's data, for example, the classification data $x = \{x_1, x_2, ..., x_m\}$ corresponds to the feature data $y = \{y_1, y_2, ..., y_m\}$. Build a logistic regression model $f(\theta)$, in which the typical construction method is maximum likelihood estimation, and the posterior probability for a single sample is:

$$P(y|x, \theta) = (h_\theta(x))^y (1 - h_\theta(x))^{1-y}, \text{ in which } y = 1(or\ 0) \tag{4}$$

Then the log maximum likelihood function can be obtained as:

$$l(\theta) = \log(L(\theta|\mathrm{x}, y)) = \sum_{i-1}^{m} y^{(i)} \log(h(x^{(i)})) + (1 - y^{(i)}) \log(1 - h(x^{(i)})). \tag{5}$$

3 Data Analysis

This data set is data from a CTR estimation contest jointly organized by Criteo and Kaggle in July 2014. This data set includes two files, the training set is about 10G, and the test set is about 1G. The training set's data volume is about 40 million, and the data volume of the test set is about 6 million. Among them, the training set is Criteo's advertising display data for 7 consecutive days, which includes click and non-click data. Since the positive and negative samples are sampled at different sampling rates every day, the overall positive and negative sample rates are not too different. The test set is the advertising display data of the day after the training set, and the sampling method of the test set is the same as the training set. As the

information is confidential, the company has encrypted all variables to a certain extent. The data features are as follows (Fig. 1):

Among them, the Label column shows whether the advertisement has been clicked or not. I1 to I13 are 13 continuous numerical variables, and C1 to C26 are 26 non-numerical categorical variables.

3.1 Data Preprocessing

This article uses a random sampling method to extract a part of the data from the training set as samples for subsequent experiments. The data after extraction is about 30,000. As shown in Fig. 2, about 25% of the advertisements are clicked before and after sampling. Therefore, it shows no significant impact on data distribution. After random sampling, the blank lines in the data were removed to obtain an initial sample set. In this model, the threshold of the proportion of missing variables is set as 15%, and the missing value processing eliminates 12 variables. The extreme value processing deletes numerical variables that are not within the range of plus or minus three times the standard deviation, as well as nominal variables and ordinal variables with a frequency of less than 1%. The categorical variable base reduction processing eliminates variables with a classification level of more than 2000, and the final experiment leaves 16 variables.

3.2 Variable Selection

In this data set, there are some discrete variables. Since such variables are challenging to participate in the model calculating, the WOE transformation is introduced to make the discrete variables continuous. WOE is the abbreviation of "Weight of Evidence", which represents the significance of the variable.The specific formula is as follows:

Label	I1	I2	⋯	I13	C1	C2	⋯	C26
1	3	20	⋯	2741	68fd1e64	80e26c9b	⋯	4cf72387
0	7	91	⋯	1157	3516f6e6	cfc86806	⋯	796a1a2e
0	12	73	⋯	1844	05db9164	38a947a1	⋯	5d93f8ab
⋮								
?	9	62	⋯	1457	68fd1e64	cfc86806	⋯	cf59444f

Fig. 1 Data characteristics

Fig. 2 Proportion of training set before and after sampling
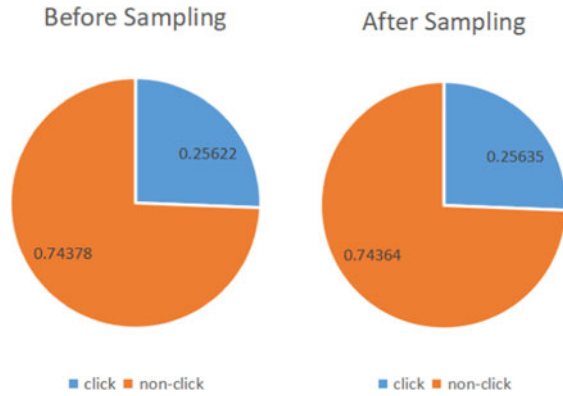
Table 1 WOE conversion variables

Serial number	Variable name	Variable name after WOE conversion	Number of categories
1	F20	WOE_f20	9
2	F23	WOE_f23	3
3	F28	WOE_f28	26
4	F31	WOE_f31	10
5	F37	WOE_f37	14
6	F15	WOE_f15_new2	6
7	F16	WOE_f16_new2	6
8	F19	WOE_f19_new2	6
9	F22	WOE_f22_new2	6
10	F32	WOE_f32_new2	6

$$WOE(x_i) = \ln\left[\frac{n_{i0}/n_0}{n_{i1}/n_0}\right]. \tag{6}$$

Among them: WOE(Xi) represents the WOE value corresponding to the value of the i-th type of variable X.

n_{i0} represents the frequency of the dependent variable taking the value of 0 in the i-th type of variable X.

n_{i1} represents the frequency of the dependent variable taking the value of 1 in the i-th type of variable X.

n_0 represents the frequency of the dependent variable taking the value of 0 in the overall.

n_1 represents the frequency of the dependent variable taking the value of 1 in the overall.

The variables of this model for WOE transformation are shown in Table 1:

In the data analysis, too many variables have a minimal effect on the model's accuracy. However, it will greatly affect the efficiency of data processing, so this paper uses independence screening and IV screening to obtain the key independent variables into the final modelling.

In this data set, the 16 independent variables obtained in the above processing are used to perform logistic regression analysis with the dependent variables, and independent variables independent of dependent variables are screened out to complete the independence screening.

IV is the abbreviation of "Information Value", which is a measure of the predictive power of the independent variable. The IV is calculated based on the WOE value.

$$IV_i = (p_{y_i} - p_{y_n}) \times WOE_i. \tag{7}$$

In this paper, the critical value of IV is positioned as 0.1. If the variable's IV is less than 0.1, it will be eliminated to complete the IV screening. Finally, six variables are retained. The specific variables are shown in Table 2:

3.3 Data Set Division

This article uses 80% samples as the training set for modelling and 20% samples as the validation set to evaluate the model's accuracy and generalization ability. The data set is divided by the principle of random sampling, and the final data set distribution is shown in Table 3:

3.4 Analysis of Modeling Results

This article uses the stepwise regression [4] in logistic regression to model and analyze the training set to get the regression results shown in Table 4. The stepwise regression method refers to introducing each variable into the model one by one for calculation. After a new variable is introduced, an F test is required, and the t-test is performed on the selected variables one by one. If the previously introduced variable becomes insignificant after adding a new variable, delete the previously introduced variable. The above process is repeatedly calculated until each time a

Table 2 Modeling variables

Serial number	Variable name
1	WOE_f31
2	WOE_f16_new2
3	WOE_f32_new2
4	f6
5	f8
6	f12

Table 3 Data Set Distribution

Data set	Sample size	Clicks(f1 = 1)	Non-Clicks(f1 = 0)
Training Set	17,425(75%)	4419	13,006
Validation Set	7467 (25%)	1893	5574
Total	24,892 (100%)	6312	18,580

Table 4 Maximum Likelihood Estimation

Parameter	Degree of Freedom	Estimate	Standard error	Wald 2	P value	Standardized estimate
Intercept	1	−2.2532	0.0480	2199.5922	<0.0001	
WOE_f31	1	0.7204	0.0544	175.1416	<0.0001	0.1656
WOE_f32_new2	1	0.9354	0.0271	1190.4986	<0.0001	0.5644
f6	1	−3.42E-6	5.116E-7	44.6348	<0.0001	−0.1293
f8	1	−0.00156	0.000551	7.9784	0.0047	−0.0418
f12	1	0.0465	0.00570	66.6232	<0.0001	0.1262

new variable is added, it is significant, to ensure that the final model has the best effect. The final selected model include f6, f8, f12, f31, f32, among which f6 and f8 are negatively correlating with the ads click-through rate, that is, when f6 and f8 are smaller, the probability of clicking the ads is greater. f12 is positively correlating with the ads click-through rate, that is, the larger the f12, the more likely it is to click on the advertisement. WOE_f31 and WOE_f32_new2 are positively correlating with the ads click-through rate, that is, the customer click-through rate of the category corresponding to the maximum value of WOE_f31 and WOE_f32_new2 is the highest. Therefore, when search engines arrange advertisements according to the possibility of click-through rate, they should place the advertisements with the smaller f6 and f8, the larger f12, and the categories corresponding to the larger values of WOE_f31 and WOE_f32_new2 in more attractive places. Since each variable's specific meaning cannot be known for confidentiality reasons in this data set, the actual meaning of the result cannot be explored, and only the degree of influence of an individual variable can be analyzed.

3.5 Model Accuracy Evaluation

KS Curve
KS is mainly to verify the model's ability to distinguish between two-class objects. By calculating the sensitivity and specificity under different thresholds, and plotting the two on the graph, the KS curve of the training set is shown in Fig. 3. The larger the difference between the two curves in the figure, the better its effect. The KS value is equal to 0.33, indicating that the model has a good classification effect on the training set.

Fig. 3 KS Diagram of
Training Set
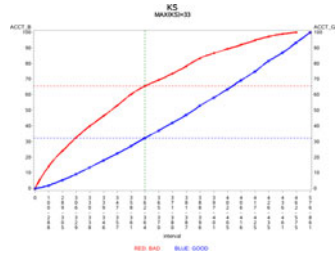
Fig. 4 KS Diagram of
Validation Set
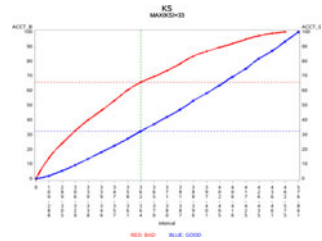
Next, we will analyze the model's generalization ability by viewing the classification effect of the model on the validation set. The validation set's classification effect still passes the KS curve, as shown in Fig. 4, where KS is greater than 0.3. It indicates that the classification effect is good. With fluctuations in normal range, it shows that the model'ýs stability is good.

ROC [5, 6] Curve
The area under the ROC curve is used to evaluate the classification of the model. The larger the value, the better the classification effect of the model. Figure 5 shows that the ROC area is 0.743, indicating that the model has a reasonable degree of discrimination.

The classification effect of the validation set is shown in Fig. 6. It can be seen that the ROC curve area is more significant than 0.73, indicating that the classification effect is good and the model is stable.

Log-Loss Value
The loss function is a function to measure the degree of loss and error. The smaller the value, the better the model predicts. The log-loss values of the training set and the validation set are calculated, as shown in the following Table 5, indicating that the accuracy of the click-through rate estimation model in this article is high.

Fig. 5 ROC Curve of
Training Set
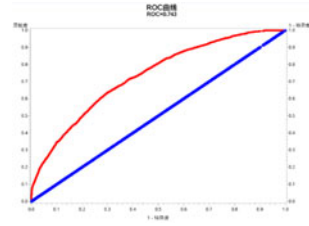
Fig. 6 ROC Curve of
Validation Set
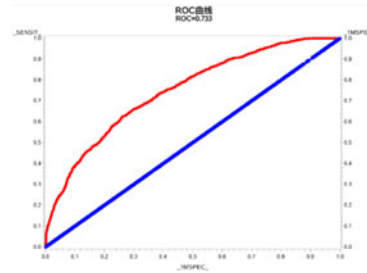
Table 5 Log-loss value

Data set	Log-loss value
Training Set	0.29513
Validation Set	0.28669

4 Conclusion

In search advertising, the key technology is to predict the click-through rate of advertisements. Accurate prediction of the click-through rate of search advertising can increase the revenue of advertisers and advertising media and can also improve the user experience of advertising audiences. Based on logistic regression model, this article studies the prediction model of search advertisement click-through rate and gets a good prediction effect. This paper's data set is data from a CTR estimation contest jointly organized by Criteo and Kaggle in July 2014. This paper has completed the random extraction of data, missing value processing, extreme value processing, and categorical variable base reduction. Based on independence screening and IV screening, the excess variables are eliminated. Finally, six main variables are introduced into the model. By stepwise regression, the final selected model includes f6, f8, f12, f31, f32, among which f6 and f8 are negatively correlating with the ads click-through rate, that is, when f6 and f8 are smaller, the probability of clicking the ads is greater. f12 is positively correlating with the ads click-through rate, that is, the larger the f12, the more likely it is to click on the advertisement. WOE_f31 and WOE_f32_new2 are positively correlating with the ads click-through rate, that is, the customer click-through rate of the category

corresponding to the maximum value of WOE_f31 and WOE_f32_new2 is the highest. Therefore, when search engines arrange advertisements according to the possibility of click-through rate, they should place the advertisements with the smaller f6 and f8, the larger f12, and the categories corresponding to the larger values of WOE_f31 and WOE_f32_new2 in more attractive places. By evaluating three assessing indicators, the conclusion is reached that the model can predict the click-through rate of search ads very well. Due to the difficulty of obtaining data such as search ads click-through rate, few Internet companies will announce such data to the public. For confidentiality reasons, the published data will undergo some transformation processing. The future studies will break through those limitations caused by the inability to conduct specific and practical analysis based on the variables in final modelling, and make more deeply analysis and accurate prediction.

References

1. https://morketing.baijia.baidu.com/article/456770
2. Yang, S., Ghose, A.: Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence. Mark. Sci. **29**(4), 602–623 (2010)
3. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression, 2nd end. Wiley, New York; Chichester (2000). ISBN 0–471–35632–8
4. https://baike.baidu.com/link?url=xXf4MopZ2Z0t0zTkIN_ejpjpGZJLn3Ri8eMX2-Etcv91l9b-CT14Z9fVDBo4uHv75swO5zIlTNuuUKiYIg8Toiad_jxa4MntV2qs-w28CXDhw2dt6MdcQ-KTLLwwU1X1wI
5. Fawcett, T.: ROC graphs: Notes and practical considerations for researchers. (2004)
6. Shuai. J.: Researches on Performance Evaluation of Classifier Based on AUC. Jilin University (2016)
7. Broder, A.: Computational Advertising. SODA, 992–992 (2008)
8. Menon, A.K., Chitrapura, K.-P., Garg, A., Agarwal, D., Kota, N.: Response prediction using collaborative filtering with hierarchies and side-information. In: Conference on Knowledge Discovery and Data Mining (KDD 2011), New York, NY, USA, 2011:141–149. ACM (2011)
9. Dembczynski, K.,Kotlowski, W., Weiss, D.: Predicting ads'click-through rate with decision rules. In: WWW 2008 Beijing (2008)
10. Gupta, M.: Predicting click through rate for job listings. In: WWW 2009 Madrid (2009)
11. Zhang, Y., Jansen, B.J., Spink, A.: Identification of factors predicting clickthrough in web searching using neural network analysis. J. Am. Soc. Inform. Sci. Technol. **60**, 1–14 (2008)
12. Richardson, M., Dominowska, E., Ragno, R.: Predicting clicks: estimating the click-through rate for new ads. In: WWW 2007 Banff, Alberta, Canada (2007)
13. Xinran et al.: Practical lessons from predicting clicks on ads at Facebook. ADKDD (2014)
14. Agarwal, D.: LASER: a scalable response prediction platform for online. Advert. ACM Tran. Intell. Syst. Technol. (2014)

Analysis of the Impact of the US Presidential Election on the US Economy Based on Natural Language Processing and Big Data

Mingzhen Li and Xiangdong Liu

Abstract Based on the background of the 2020 US presidential election, this paper analyzes the impact of the two candidates' different policy propositions on the US economy. Firstly, using web crawler technology to obtain the public information, then using Natural Language Processing to generate the theme of the document and corresponding high frequency keywords. Secondly, build a mathematical model. GDP, CPI and unemployment are selected as the macro indicator system to measure economic changes. In the stage of data collection and preprocessing, which also spent a lot of time to ensure that the quality of data can be applied to the establishment of subsequent models. In the model part, this paper uses the multiple linear regression model. Based on this, it objectively gives significant indicators to measure the size and direction of the influence of the two candidates' positions on the American economy. Finally, based on the time series data, this paper uses the ARIMA Model to predict the next four periods, and gives the economic trend of the United States in the next year if President Trump continues to be in office.

Keywords The impact of political events on the American economy · Natural language processing · ARIMA model

1 Introduction

As is known to all, presidential elections are held every four years in the United States. 2020 is the year of US presidential election, with Republican candidate Donald Trump and Democratic counterpart Joe Biden running for president. The candidates of both parties have different political stands and administrative programs in finance and trade, economic and financial governance, and some other

M. Li (✉) · X. Liu
Department of Statistics, Jinan University, Guangzhou 511400, Guangdong, China
e-mail: lmz177@163.com

X. Liu
e-mail: tliuxd@jnu.edu.cn

different key development areas. The election of different candidates will shape different strategic patterns of global economic and financial development, and have a greater impact on the U.S. economy and the global economy. Therefore, we will analyze the impact of the US presidential election on the economic and financial development of the US by analyzing the election of Republican presidential candidate Donald Trump and Democratic presidential candidate Joe Biden.

2   LDA Document Topic Generation Model on Two Candidates

In the field of Natural Language Processing, a large amount of data is structured, which is difficult to directly obtained from the information relevant and desired information. Subject model (LDA) is a kind of method based on Bayes Theory. It able to recognize in the document, dig out the theme of the hidden information. And it has a wide range of uses in topics aggregation, extracting information from unstructured text and feature selection and other scenarios.

LDA model is usually composed of three layers: Document (d), Theme (z) and Word (W), it is also called three-layer Bayesian probability model. Assume that any document is a mixture of topics, generated from the Dirichlet distribution by sampling.

First, the model assumes that each policy text contains a variety of topics, and topics are randomly combined in accordance with a mixed proportion of multinomial distribution:

$$Z|\theta = \text{Multionomial}(\theta) \tag{1}$$

At the same time, any topic contains multiple words, and the words are mixed according to a certain proportion of multinomial distribution:

$$W|Z, \varphi = Multionomial(\varphi) \tag{2}$$

The probability of generating word w_i under the condition of document d_j is expressed as:

$$P(w_i|d_j) = \sum_{s=1}^{k} P(w_i|z = s) \times P(z = s|d_j) \tag{3}$$

$P(w_i|z = s)$—The probability of word w_i belongs to the theme s;
$P(z = s|d_j)$—The probability of the theme s belongs to document d_j.

The LDA model uses Gibbs sampling in Metropolis-Hastings algorithm to approximate the parameters θ, φ, The formula is as follows:

$$P(z_i = s | Z_{-i}, W) \propto (n_{s,-i} + \beta_i)/(\sum_{i=1}^{V} n_{s,-i} + \beta_i) \times (n_{s,-j} + \alpha_s) \qquad (4)$$

$z_i = s$—The probability of word w_i in the theme s.

According to the above formula can obtained $\varphi_{s,i}$ and $\theta_{j,s}$. The parameter estimation of word w_i in theme z_s, and the parameter estimation of document z_s in document d_j, denoted as:

$$\varphi_{s,i} = (n_{s,i} + \beta_i)/(\sum_{i=1}^{V} n_{s,i} + \beta_i) \qquad (5)$$

$$\theta_{j,s} = (n_{j,s} + \alpha_s)/(\sum_{j=1}^{V} n_{j,s} + \alpha_s) \qquad (6)$$

$n_{s,i}$—the number of occurrences of w_i in theme z_s;

$n_{j,s}$—the number of items contained in topic z_s by document d_j.

Before applying the model algorithm, the three variables of prior parameters, and the number of subjects K in the model should be set as the best values. The first two are set as empirical values, $\alpha = 0.1$, $\beta = 0.02$. In this text analysis, set the number of themes k = 8.

Through text mining on the political propositions of the two candidates, a total of eight themes were generated, including education, tax, medical care, COVID-19 fighting measures, employment and environmental protection, and high-frequency words were generated under each theme to facilitate us to have a further understanding of the policy propositions of the two candidates. At the same time, this paper also makes a word cloud map to intuitively display the policy and propositions, and makes a table for intuitive comparison. The results are shown in Table 1 (Fig. 1).

Fig. 1 The word cloud of
two candidates about
covid-19 fighting measures

Table 1 Keywords and document topic of two candidates' policy by LDA Model

Trump	Keywords		Biden
Education	Founding, school, parents, teachers, authority	Universities, students, families, technology, investment	Education
Covid-19 fighting measures	States, epidemic, crown, production, pneumonia	Prevention, production, epidemic, masks, security	Covid-19 fighting measures
Immigration	Border, immigration, wall, illegal, refugee, apprehensions	Immigrants, comprehensive, reform, restrictions, bans	Immigration
Medical care	Republican, reform, bill, abolish, plan, related, representative	Public, insurance, epidemic, vaccine, Medicaid	Medical care
Tax	Tax, bill, rate, passed, reduced, corporate, progressive	Income, precent, rich, taxes, business, plan, families, business	Tax
Infrastructure	Border, federal, government, investment, security, wall	Billion, roads, energy, bridges, emissions, transportation, carbon	Infrastructure
Employment	Workers, opportunities, programs, workforce, education, apprenticeships, disabilities	Minimum wage, tax, loan, government procurement, promote	Employment
Environmental protection	Endangered, marine, species, promotes, debris, executive	Emissions, climate, plan, states, achievement, carbon, revolution	Environmental protection

3 Preparation of the Analysis Model

Autoregressive Integrated Moving Average Model (ARIMA Model)
Differential integrated moving average autoregressive model, also known as integrated moving average autoregressive model, is one of the predictive analysis methods of time series. $ARIMA(p, d, q)$ Where, AR is "autoregressive" and P is the number of autoregressive terms; MA is the 'sliding average', Q is the number of sliding average terms, and D is the difference order made to make it into a stationary series.

$ARIMA(p, d, q)$ is an extension of a model. $ARMA(p, q)ARIMA(p, d, q)$ The model can be expressed as:

$$\left(1 - \sum_{i=1}^{P} \varphi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t \tag{7}$$

Among them, L is Lag Operator. The methods and steps for establishing ARIMA model are as follows:

1. Acquisition of time series Acquisition of time series can be obtained through experimental analysis or statistical data of relevant departments. For the data obtained, first check for the presence of abrupt points, and analyze whether these points exist due to human negligence or other reasons.
2. Preprocessing of Time series Preprocessing of time series includes two tests, stationarity test and white noise test. The time series that can be analyzed and predicted by ARMA model must satisfy the condition of stationary non-white noise series. If there is an increasing or decreasing trend in the non-stationary time series, it is necessary to carry out difference processing and then conduct stationarity test until stationary. Generally, the difference of the order number less than 2.

Table 2 Selection and classification of indicator variables

	Index	Explanation
Response variable	GDP_speed	Growth rate of Gross domestic product
	CPI	Growth rate of Consumer price Index
	Unemployment	Us unemployment rate
Stimulus variable	Fiscal	US government fiscal
	PMI	Purchasing Managers index
	Trade	The trade deficit
	Custom	Consumer Confidence Index
	Home-sales	Sales of previously owned homes
	Dow-Jones	Dow Jones Industrial Average

Note: GDP below refers to GDP growth rate

3. Model recognition means to select a model that is consistent with the given time series process from the known models. Then the BIC criterion method can be used to determine the order.

4   Analysis of the Impaction

4.1   The Construction of Index System

Fuyi Li [1] article on the study of the America's economic future, to the American economy make predictions about the future, select the level of nine more index data, including the supply side, employment, currency, prices, trade, finance and so on, at the same time, combined with the topic given, both candidates are different in different areas of the claim, we select a few core areas of important index respectively to multiple regression fitting, explain the economic change and make predictions about the future (Table 2).

We treat GDP as follows:

$$GDP\_speed = \frac{GDP_t - GDP_{t-1}}{GDP_t} \tag{8}$$

4.2   Preprocessing of Data

For the above selected indicators, we searched relevant data through relevant databases. These data need to be processed in the later period.

Stationarity Test. The stationarity test and stationary processing of time series are important prerequisites for the establishment of the model.

Time series data often have the property of autoregression. The P-order autoregression can be expressed as:

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \ldots + \rho_p X_{t-p} + \varepsilon_t \tag{9}$$

Table 3 The result of ADF unit root test ($\alpha = 0.05$)

	Dickey-Fuller	Prob.	Exiting unit root	Disposing
GDP	0.8617	0.9409	T	First difference
CPI	4.4037	0.0100	T	Steady
Unemployment	2.5515	0.3547	F	First difference
Fiscal	1.8795	0.6183	T	First difference
PMI	2.6947	0.2977	T	First difference
PPI	4.1772	0.0108	F	Steady
Trade	1.2330	0.8789	T	First difference
Custom	1.6390	0.7175	T	First difference
Home-sales	2.3968	0.4162	T	First difference
Dow-Jones	2.8077	0.2528	T	First difference
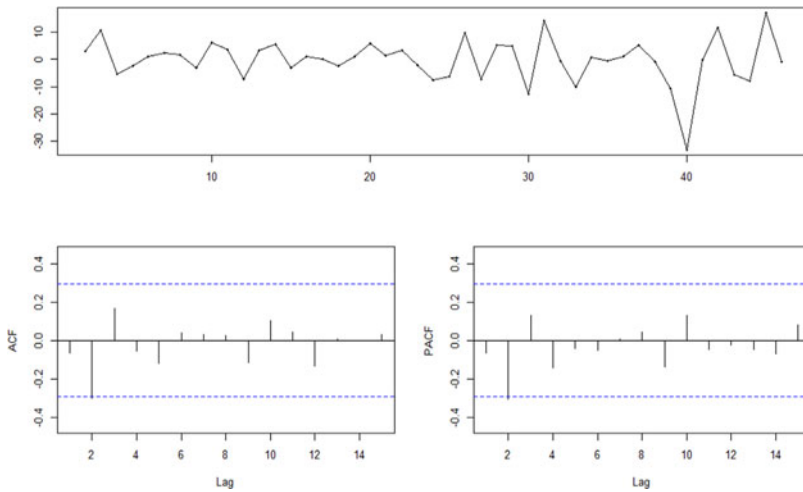
Fig. 2 Sequence, autocorrelation and partial correlation of consumer confidence data after first difference

Among them, ε_t is the random error term that meets the basic requirements of regression model, which is called white Noise error term.

The ADF unit root test results are shown in the table below (Table 3):

Except for CPI and PMI index data, the ADF unit root text of the other index data is not significant and fails to pass the test. Consider differential processing.

Through the situation of trailing and truncating, difference processing is carried out.

$$\Delta X_t = X_t - X_{t-1} \tag{10}$$

Table 4 ADF unit root test after first different ($\alpha = 0.05$)

	Dickey-Fuller	p-value		Dickey-Fuller	p-value
GDP	2.2374	0.4819	Trade	3.5468	0.0482
Fiscal	1.9750	0.6038	Custom	4.0712	0.0154
Unemployment	3.8130	0.0269	Home-sales	3.7495	0.0319
PMI	4.6136	0.0100	Dow-jones	4.4244	0.0100

Table 5 The result of white nose test

	X squared	Prob.		X squared	Prob.
GDP	1.6537	0.1985	PPI	0.8252	0.3637
CPI	6.1993	0.0128	Trade	5.9860	0.0142
Fiscal	0.3415	0.5590	Custom	0.2019	0.6532
Unemployment	30.789	2.876e-08	Home-sales	0.5856	0.4441
PMI	0.8252	0.3637	Dow-Jones	0.0007	0.9779

In order to maintain the balance of the data volume, for the small sample data, we continue to retain the first data value, which is called Prince-Winston Transformation. The ADF Unit root test was conducted again, and the results all passed the test with good significance (Fig. 2).

The ADF test results after data difference processing are shown in the Table 4.

Residual White Noise Test. When white noise applied to time series, it means that in the case of autocorrelation, the random interference term in the correlation between time series data and the previous period data obeys the classical regression assumption [2].

$$E(\varepsilon) = 0, D(\varepsilon) = \sigma^2, cov(\varepsilon_t, \varepsilon_{t-p}) = 0 \tag{11}$$

Ljung - Box test is commonly used for white noise testing. And the results after difference is shown in the following table (Table 5):

It can be seen from the table that, except a few time series fail to pass the test, the rest of the series are all white noise series. It can be used with the ARIMA model.

4.3 The Establishment of Multiple Regression Model

GDP. Assume that the multiple regression equation affecting GDP is [3]

$$\widehat{GDP} = \hat{\beta}_0 + \hat{\beta}_1 PMI + \hat{\beta}_2 custom + \hat{\beta}_3 homs + \hat{\beta}_4 trade + \hat{\beta}_5 DowJ + \hat{\beta}_6 fiscal + \hat{\beta}_7 PPI$$

The adjusted quarterly data was input into the software, and after stepwise regression using AIC maximum criterion, multiple linear regression was established to fit. Four variables were entered into the equation, and the regression results were output as shown in the table below (Table 6):

The goodness of fit of the linear regression R2 = 0.887, good fitting degree. In conclusion, the regression equation is as follows:

$$\begin{aligned}
\widehat{GDP} &= \hat{\beta}_0 + \hat{\beta}_1 PMI + \hat{\beta}_2 custom + \hat{\beta}_3 homs + \hat{\beta}_6 fiscal \\
&= -0.473 + 0.026 PMI - 0.008 custom + 0.004 homs + 0.001 fiscal
\end{aligned} \tag{12}$$

Unemployment. Assume that the multiple regression equation affecting Unemployment is (Table 7):

Table 6 Coefficients of model

variable	Coefficient	Std.Error	Beta	t-Statistic	Prob.
(intercept)	3.473	0.516		6.734	0.000
Homesales	0.004	0.001	0.742	6.319	0.000
fiscal	0.000	0.000	0.398	4.166	0.002
custom	0.008	0.002	0.555	4.427	0.001
PMI	0.026	0.010	0.383	2.603	0.026
R	0.959		Std.Error		0.064
R_square	0.919		F_statistic		28.479
Adjust R_square	0.887		Prob.		0.000

Table 7 Coefficients of model

variable	Coefficient	Std.Error	Beta	t-Statistic	Prob.
(intercept)	0.265	0.016		16.719	0.000
custom	0.002	0.000	0.862	13.039	0.000
PPI	4.080	0.887	0.304	4.598	0.001
R	0.975		Std.Error		0.006
R_square	0.950		F_statistic		114.159
Adjust R_square	0.942		Prob.		0.000

$$\widehat{Une} = \widehat{\beta}_0 + \widehat{\beta}_1 PMI + \widehat{\beta}_2 custom + \widehat{\beta}_3 homs + \widehat{\beta}_4 trade + \widehat{\beta}_5 DowJ + \widehat{\beta}_6 fiscal + \widehat{\beta}_7 PPI$$

The goodness of fit of the linear regression equation $R^2 = 0.942$, good fitting degree. In conclusion, the regression equation is as follows (Table 8):

$$\begin{aligned} \widehat{Une} &= \hat{\beta}_0 + \hat{\beta}_2 custom + \hat{\beta}_7 PPI \\ &= 0.265 - 4.080 PPI - 0.002 Custom \end{aligned} \tag{13}$$

CPI. Assume that the multiple regression equation affecting CPI is

$$\widehat{CPI} = \widehat{\beta}_0 + \widehat{\beta}_1 PMI + \widehat{\beta}_2 custom + \widehat{\beta}_3 homs + \widehat{\beta}_4 trade + \widehat{\beta}_5 DowJ + \widehat{\beta}_6 fisc + \widehat{\beta}_7 PPI$$

The goodness of fit of the linear regression equation $R^2 = 0.5875$. The fitting degree is general, which may be the reason for variable selection, or the problem of data processing, because CPI is a year-on-year or sequential index, the fluctuation may be reduced in this process. In conclusion, the regression equation is as follows:

Table 8 Coefficients of model

variable	Coefficient		Std.Error	t-Statistic	Prob.
(Intercept)	0.0034		3.040 e-03	1.119	0.26995
PMI	2.780 e-04		8.205 e-05	3.388	0.00159
PPI	0.2131		9.912 e-02	2.150	0.03763
trade	0.4773		2.026 e-01	2.356	0.02346
DowJ	1.821 e-07		1.202 e-07	1.515	0.13764
R		0.7906	Std.Error		0.0016
R_square		0.6250	F_statistic		16.66
Adjust R_square		0.5875	Prob.		4.091 e-0. 8

$$\widehat{\mathrm{CPI}} = \hat{\beta}_0 + \hat{\beta}_1 PMI + \hat{\beta}_4 trade + \hat{\beta}_5 DowJ + \hat{\beta}_7 PPI$$
$$= -0.0034 + 0.0003\mathrm{PMI} + 0.4773\mathrm{trade} + 1.821e - 07\mathrm{DowJ} + 0.2131 \ \mathrm{PPI}$$

$$(14)$$

Table 9 Predictive value of ARIMA Model

Point forecast	GDP_speed	Unemployment	CPI
2020Q4	0	0.069	0.00075
2021Q1	0	0.069	0.00154
2021Q2	0	0.069	0.00154
2021Q3	0	0.069	0.00154

Fig. 3 The ARIMA model for U.S. GDP, CPI and Unemployment in the coming year
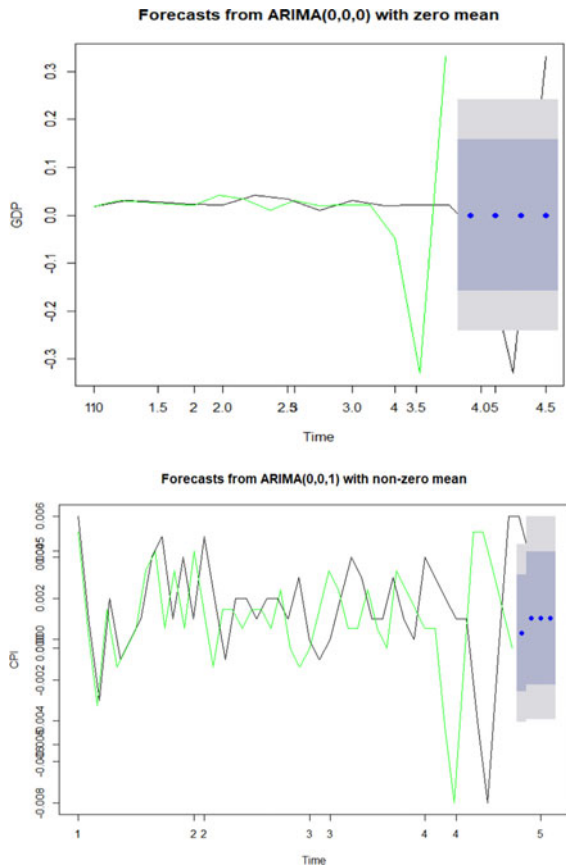
4.4   Predictions of the ARIMA Model

On the basis of the previous stationary white noise time series, we use R language to automatically establish ARIMA model to simulate the forecast of GDP, CPI and unemployment rate in the next four periods after President Trump remains in office (Fig. 3 and Table 9).

5   Conclusion

For US macroeconomic indicators, the variables significantly associated with GDP growth are investment, fiscal, consumer confidence and PMI. The correlation between investment and GDP growth is the most significant. For the US unemployment rate index, the significantly correlated variables are the consumer confidence index and PPI. Government policies affect the expectations of consumers and producers on the future through macro-control, affect output to a certain extent, and then affect household consumption. Finally, for the US CPI index, the variables that are mainly related to CPI fluctuation are PMI, trade and PPI of US, indicating that CPI mainly has a strong influence on the supply side, and all show a positive correlation.

The ARIMA model gives a picture of economic volatility in the year ahead if Trump remains in office. But with the aftermath of the COVID-19 outbreak and the return of the winter outbreak, health care and COVID-19 policies will also be at the center of a battle between the two candidates that will impact the U.S. economy from the outside.

By comparing the policy propositions of the two candidates in different fields and combining the results of multiple regression fitting, the policy fields that have a greater impact on various macroeconomic variables are obtained. Combined with the summary and comparison in Table 1, the policy propositions of the two candidates can be used to make a qualitative analysis of their respective impacts on the US economy.

References

1. Li, F.: The short-term downward pressure on the US economy highlights that it will remain the main engine of world economic growth in the medium to long term. Institute of Foreign Economics, China Academy of Macroeconomics (2018)
2. Kabacoff, R.I.: R in action. Posts & Telecom Press (2020)
3. Gujarati, D.N.: Fundamentals of Econometrics. RENMING University of China Press (2018)

Development of Driving Cycle for Light Vehicle Based on the AP Clustering Method

Jiarui Chen and Baoqin Chen

Abstract In order to establish the light vehicle driving cycle of Fuzhou city, a driving cycle construction method was proposed based on t-SNE combined with AP clustering algorithm, which can well reflects non-linear relationship between parameters and obtain more stable clustering effect. Based on the micro-trip theory, 2011 micro-fragments are identified from 496464 effective speed information and an evaluation criterion consisting of 15 characteristic parameters was founded. Then, t-SNE was utilized to reduce the dimensionality of characteristic parameters and all micro-fragments were classified into different categories by AP clustering. Finally, according to the relative error of six characteristic parameters and the SAFD, the target driving cycle is compared with the other driving cycle obtained by the method of principal component analysis combined with k-means clustering, which verified the accuracy of the proposed method.

Keywords Driving cycles · T-SNE · AP clustering

1 Introduction

A driving cycle can be used to describe the relation of vehicle velocity versus time which reflects ideally the real-life driving behavior in a city. In general, it can be used to generate more accurate fuel consumption and emissions ratings from vehicles for emission testing and certification [1, 2].

Therefore, more and more cities begin to develop new and representative driving cycles for themselves. For example, Hong Kong [3], California [4], and Berlin [5] have constructed their local driving cycles for research on energy conservation and pollution evaluation.

J. Chen · B. Chen (✉)
Faculty of Mathematics and Computer Science,
Guangdong Ocean University, Zhanjiang 524088, China

495

With the development of automotive industry, the traditional methods of developing driving cycles are principal component analysis (PCA) [6, 7] and clustering analysis [8, 10, 11]. The k-means clustering approach is one of the most common used machine learning algorithm to solve clustering problems. Moreover, recent studies by some scholars have used Markov chain to establish cycles [9]. In 2013, Fotouhi [12] gathered driving data from the city of Tehran and used k-means clustering to build driving cycles while comparing with others standard cycles. In 2016, Galgamuwa [13] reconstructed a driving cycle for Sri Lanka with the modal cycle construction method which is basing on the Markov theory.

However, PCA uses orthogonal transformation to transform a series of linearly related variables into a set of new linearly uncorrelated variables, so it can not solve the nonlinear relationships of between driving parameters such as speed, duration and distance [14]. What's more, the k-means clustering algorithm is simple and fast, but it is quite sensitive to the initial value and it will appear locally optimal.

To address these issues, the main purpose of this paper is to propose a approach for constructing a representative driving cycle of light-duty vehicles in Fuzhou based on the t-distributed stochastic neighbor embedding (t-SNE) approach combined with the affinity propagation (AP) clustering method. What's more, the Silhouette equation is used to assess the selection of the number of clusters to reduce the error of artificial selection. Eventually, a comparison between the proposed the t-SNE combined with AP clustering method and the PCA combined with k-means clustering method is performed.

2   Data Preparation

The sampling frequency of vehicle is 1 Hz in Fuzhou area and the vehicle driving routes shown as Fig. 1, total of collecting 496464 valid speed information and including many driving parameters, such as velocity and time, etc.

Moreover, the acceleration recorded by the sensor greater than or equal to 4 m/s^2 should be deleted during processing data. After data statistical analysis, the frequency of speed-acceleration distribution as depicted in Fig. 2.

Fig. 1   Coverage map of light vehicle routes

3 Micro-trips Fragment

In real-world, the vehicle will stop and start many times when moving from the original point to the destination. The process from one engine idling state to another engine idling state is described as a micro-trip fragment. The driving cycle can be considered as a combination of micro-trips [15]. The four driving states that make up a fragment are defined as follows [16]:

(1) Idling: a continuous process in which the vehicle stops moving but the engine keeps running at the lowest speed.
(2) Acceleration: a continuous process with vehicle acceleration greater than $0.1 \ \text{m/s}^2$.
(3) Cruising: a continuous process where the absolute value of vehicle acceleration is less than $0.1 \ \text{m/s}^2$ at non-idling.
(4) Deceleration: a continuous process with vehicle acceleration less than $-0.1 \ \text{m/s}^2$.

A micro-trip segment can be described with multiple driving parameters. As shown in Table 1, nineteen driving parameters will be used to describe these fragments and the first fifteen are used for driving feature extraction, and clustering of the driving data.

In this paper, Python is used to identify 2111 fragments from the sample data. Many characteristic parameters were extracted from each fragment, demonstrated in Table 2.

Table 1 Driving parameters of micro-trip

Parameter	Definition	Unit
T	Segment duration	s
T_a	Acceleration time	s
T_d	Deceleration time	s
T_c	Cruise time	s
T_i	Idle time	s
S	Running distance	km
V_{max}	Maximum speed	km/h
V_m	Average speed	km/h
V_{mr}	Average run speed	km/h
V_{sd}	Standard deviation of speed	km/h
a_{max}	Maximum acceleration	m/s^2
a_a	Average acceleration of the acceleration section	m/s^2
a_{min}	Maximum deceleration	m/s^2
a_d	Average deceleration of the deceleration section	m/s^2
a_{sd}	Standard deviation of acceleration	m/s^2
P_a	Acceleration time ratio	%
P_d	Deceleration time ratio	%
P_c	Cruise time ratio	%
P_i	Idle time ratio	%

Fig. 2 The frequency of
speed-acceleration
distribution
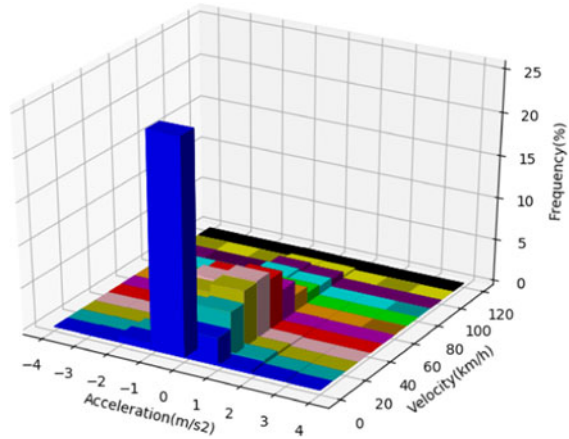
4   Methodology for Developing the Driving Cycle

In the process of building the driving cycle, characteristic parameters should be reduced dimension to relieve computational burden. What's more, the micro-trips should be classified into two or more classes for constructing a driving cycle. The following two methods are used to construct cycle and compare them, including proposed approach and traditional approach.

4.1   Proposed Method: t-SNE and AP Clustering

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear dimensionality reduction algorithm which developed by Maaten and Hinton [17], representing the local structure of original high-dimensional data in a low-dimensional space and visualizing the data in 2D or 3D. And the problem of information overlap and correlation between characteristic parameters can be solve by the t-SNE method.

In 2007, Frey [18] put forward Affinity Propagation(AP) clustering method which takes as input a collection of real-valued similarities between data points. Moreover, the AP clustering don't need to specify the number of classes in advance.

The t-SNE method are utilized to reduce the dimensionality of micro-fragments characteristic parameters, then AP clustering are used to deal with the reduced-dimensional samples, and finally we selects the appropriate segment in each cluster to construct driving cycle.

Table 2 Fifteen characteristic parameters determined for each fragment

Number	T/s	T_a/s	T_d/s	…	a_{min}/(m/s^2)	a_d/(m/s^2)	a_{sd}/(m/s^2)
1	68	26	18	…	−1.8611	−0.5448	0.4920
2	370	111	86	…	−2.1944	−0.4942	0.4017
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2010	21	8	6	…	−1.1111	−0.5417	0.5017
2011	109	41	22	…	−1.6944	−0.6427	0.4455

Remark 1. We call the method above as TSNE-AP (t-SNE and AP).

Firstly, the sample data needs to be standardized because the unit of each parameter in Table 2 is different. Secondly, t-SNE was used to reduce the normalized characteristic parameters to 2 dimensions. To acquire a more accurate number of clusters, AP clustering is used to divide fragments into two categories, three categories and four categories separately. To obtain the best clustering result, the equation of silhouette will be applied in this paper. The silhouette value is used to describe the distance between each point in adjacent categories and decides whether the classification of each fragment is rational [18]. The equation is as follows:

$$L(i) = \frac{\min(\mathbf{b}) - a}{\max[a, \min(\mathbf{b})]}. \tag{1}$$

where a is the average distance between fragments and other fragments in the same class. $\mathbf{b}$ is a vector which is the average distance between the i fragments and other fragments of different category.

The value range of $L(i)$ is from −1 to 1. When $L(i)$ is closer to 1, it means that the farther the fragments is from other categories, the better the clustering result. When $L(i)$ is equal to 0, it cannot be distinguished which category the fragments i belongs to. When $L(i)$ is less than 0, the fragments i is likely to be classified into the wrong category.

The silhouette values of different fragments in different class are displayed in Fig. 3. And from scatter plot we can see that each cluster of fragment is clearly distinguished.

As Fig. 3a displays, the silhouette value of almost each fragment was greater than 0 when the number of clusters is 2, indicating that most fragment s were well clustered. As Fig. 3b displays, the silhouette value of a few micro-trips appear negative when the number of clusters is 3. As Fig. 3c shows, the number of micro-trips with negative in the silhouette value starts to increase when micro-fragments are clustered into 4 categories. In summary, many micro-trips are incorrectly classified when the number of categories is greater than 2. Therefore, the optimal number of categories is selected as two based on the analysis of silhouette value.
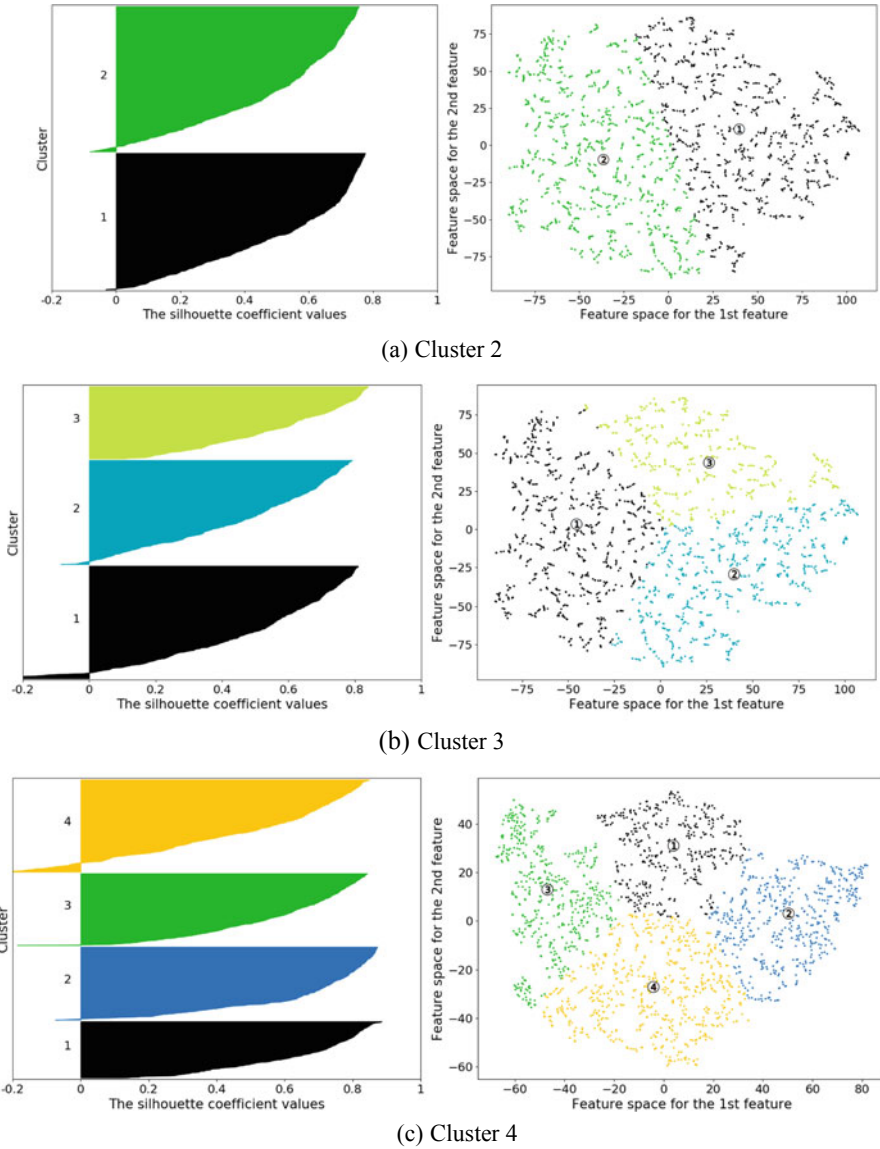
(a) Cluster 2

(b) Cluster 3

(c) Cluster 4

Fig. 3 Evaluation values of different AP clustering numbers. The numbers in the scatter plot represent the cluster centers

4.2 Traditional Method: PCA and k-means Clustering

As for the traditional construction method of driving cycle, the principal component analysis of the normalized data is shown in Table 3. It can be known from Table 3 that the cumulative contribution rate of the first four principal components has exceeded 86% and can already represent most of the feature of the original 15 driving parameters.

So the first 4 principal components are selected for k-means clustering method, and then the silhouette equation was used to compare the classification effect under different clusters, as shown in Fig. 4.

Remark 2. We call the method above as PCA-KM (PCA and k-means).
It can be seen from Fig. 4 that the wrong micro trips of cluster 3 is distinctly less than cluster 2 or cluster 4. Therefore, cluster 3 are selected as the final result of k-means clustering.

4.3 Result

The comparison between Fig. 3 and Fig. 4 shows that it is more the wrong micro trips in Fig. 4, and the micro trips is evenly distributed in different class in Fig. 3. Therefore, we can infer that the TSNE-AP method proposed in this paper is better than the commonly used PCA-KM method in clustering effect.

According to the clustering results of the two methods, the time proportion of four driving status, including acceleration, deceleration, idling and cruising are respectively calculated, as shown in Fig. 5.

Table 3 Contribution rate of principal components and the cumulative contribution rate

Principal component	Variance	Contribution rate (%)	Accumulation rate (%)
1	7.378	49.184	49.184
2	3.111	20.742	69.926
3	1.304	8.693	78.619
4	1.181	7.873	86.492
5	0.775	5.165	91.657
…	…	…	…
15	3.25e−05	2.17e−06	100.000

(a) Cluster 2

(b) Cluster 3
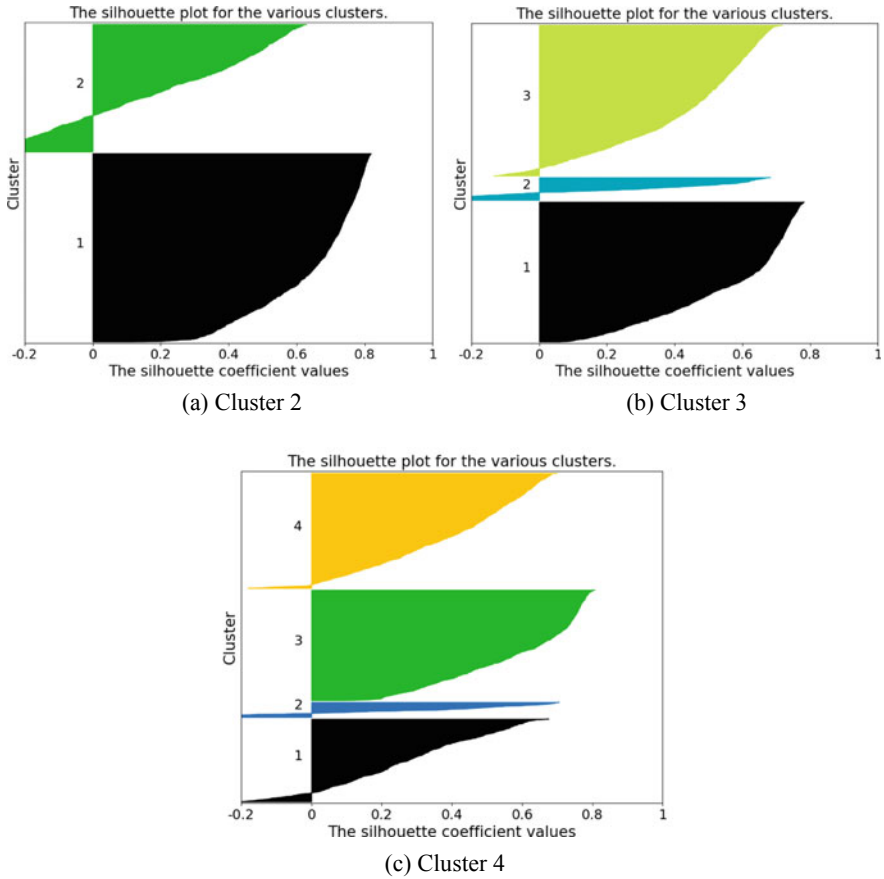
(c) Cluster 4

Fig. 4 Evaluation values of different k-means clustering numbers

As Fig. 5a shows, in the TSNE-AP method, the idling ratio of the first class is 28.11% and the acceleration ratio is 27.59%, implying that the vehicle drives in a congested traffic status. While the idling ratio of the second class is obviously lower and the acceleration ratio is higher, implying that the vehicle drives in a smooth traffic status. As Fig. 5b shows, in the PCA-KM method, the first category has the highest idling ratio, meaning that it is low velocity cluster. The four driving status in second category are almost average distributed, meaning that it is medium velocity cluster. The third category has the highest acceleration ratio, reaching 40.18%, meaning that it is high velocity cluster.
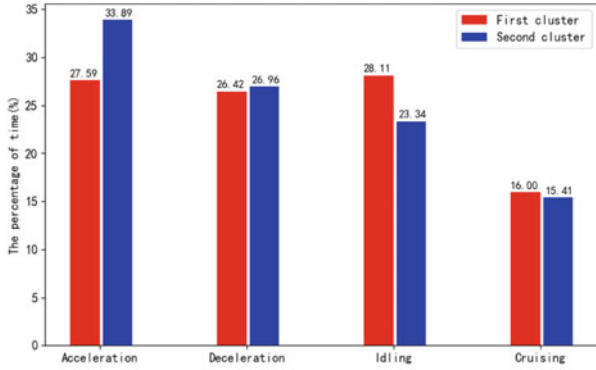
(a) TSNE-AP method
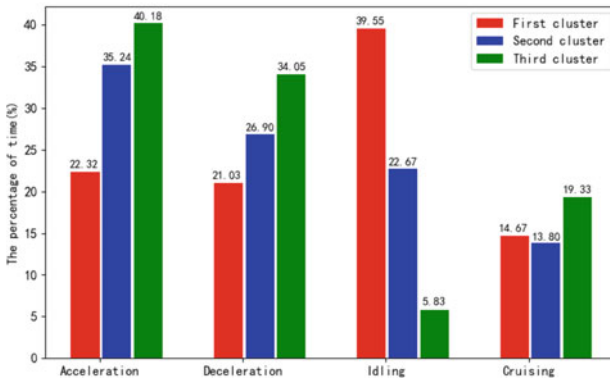
(b) PCA-KM method

Fig. 5 Comparison of time proportion of four driving states under different clusters
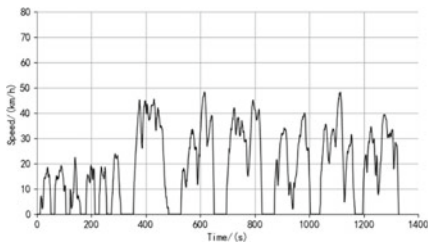
(a) TSNE-AP result
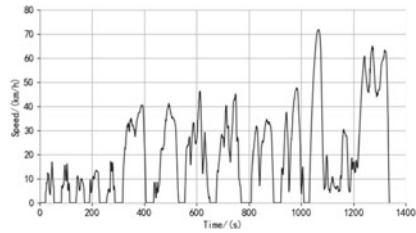
(b) PCA-KM result

Fig. 6 Road driving cycle of light-duty vehicles in Fuzhou

The driving cycles are constructed based on proposed two methods. First, we calculate the Euclidean distance between each fragments in the same class and the average characteristic parameter value of that class. Next, according to the size of the Euclidean distance, the fragments of each class are selected from small to large to establish two types of driving cycles over 1200 s. Finally, the duration of each traffic condition in the target driving cycle is determined by the time proportion of each category in the sample data. It is defined by:

$$t_i = \frac{t_{dc}}{t_{all}} \sum_{j=1}^{n_i} t_{i,j} \tag{2}$$

Where t_i is total time of category number i in the target driving cycle. t_{dc} is total time of the target driving cycle. t_{all} is total time of all segments. $t_{i,j}$ is the time of segments number j in category number i, and n_i is the total number of segments in category number i.

According to the result of AP clustering, we extracts 6 congested segments and 6 unblocked segments to make up driving cycle 1330 s in length, as shown in Fig. 6a. At the same time, we also selects 5 low speed segments, 6 medium speed segments and 1 high speed segments from the result of k-means clustering to establish driving cycle 1338 s in length, as shown in Fig. 6b.

5   Verification and Comparison

There are many ways to decide the target driving cycle. In this study, we use the difference of speed-acceleration frequency distribution(SAFD), and the six driving characteristic parameters to select the representative driving cycle.

The SAFD difference between the whole sample data and the two types of driving cycle is demonstrated in Fig. 7. In Fig. 7a, we can know that the SAFD difference between sample and the outcome of TSNE-AP approach is within ±4%. While the SAFD difference in PCA-KM approach is within ±7%, as presented in Fig. 7b.

The assessment of the driving parameters between sample data and the two types of driving cycle is demonstrated in Table 4, and shows that the relative errors between the driving cycle constructed by the TSNEAP method and the sample data is within the range of ±5%, which means the constructed driving cycle embody the driving condition of the real-world at good. Furthermore, the relative errors obvious larger between the driving cycle constructed by the PCA-KM method and the sample data, the error of the cruising ratio reaches ±15.51%. In contrast to PCA, the t-SNE can retain more original sample information, and well express the non-linear relationship between the characteristic parameters such as speed, time, and distance. And AP clustering can achieve better clustering results. Therefore, in this study, we choose the driving cycle established by the TSNE-AP approach as the final driving cycle.
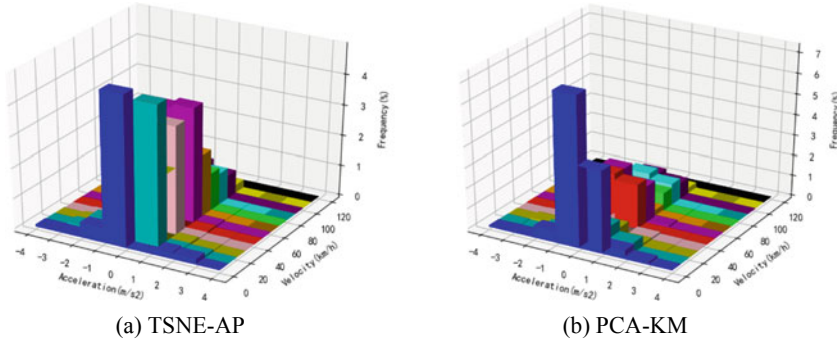
(a) TSNE-AP   (b) PCA-KM

Fig. 7 Speed-acceleration frequency distribution of two driving cycles

Table 4 Evaluation and Comparison of overall sample and constructed driving cycles constructed by 2 methods

Characteristic parameter	Sample data	TSNE-AP	Variance (%)	PCA-KM	Variance (%)
T_a	0.298	0.307	2.94	0.303	1.34
T_c	0.173	0.157	9.27	0.146	15.51
T_i	0.262	0.257	1.95	0.283	7.86
V_a	15.12	15.04	0.55	14.52	3.97
a_a	0.509	0.482	5.23	0.476	6.52
a_{sd}	0.517	0.503	2.71	0.493	4.61

6 Conclusion

In this research, a new method based on AP clustering approach combined with t-SNE to construct a representative driving cycle of light vehicle is proposed, and compares with the PCA combined with k-means clustering approach. The results show that the t-SNE approach can well solve the non-linear relationship between the characteristic parameters and obtain a more accurate dimensionality reduction effect. At the same time, AP clustering can ensure convergence to the global optimum, and obtain more stable clustering results than k-means. In addition, the six characteristic parameters and the SAFD difference are used to comprehensively evaluate the representative driving cycle and overall sample. The experimental results demonstrated that the relative error between the Fuzhou light vehicle driving cycle constructed by this approach and the sample data is within a reasonable range, confirming the accuracy and effectiveness of the proposed approach.

References

1. Nyberg, P., Frisk, E., Nielsen, L.: Driving cycle adaption and design based on mean tractive force. IFAC Proc. Vol. **7**, 689–694 (2013)
2. Ho, S.-H., Wong, Y.-D., Chang, V.W.-C.: Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions. Atmos. Environ. **97**, 353–362 (2014)
3. Lina, J., Niemeier, D.A.: An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle. Atmos. Environ. **36**(38), 5759–5770 (2002)
4. Tong, H.Y.: Development of a driving cycle for a supercapacitor electric bus route in Hong Kong. Sustain. Cities Soc. **48**, 101588 (2019)
5. Brady, J., O'Mahony, M.: Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas. Appl. Energy **177**, 165–178 (2017)
6. Yang, Y., Li, T., Hu, H., Zhang, T., Cai, X., Chen, S., Qiao, F.: Development and emissions performance analysis of local driving cycle for small-sized passenger cars in Nanjing, China. Atmos. Pollut. Res. **10**(5), 1514–1523 (2019)
7. Mi, J., Jiang, J., Wei, S., Jia, X.: Synthesis of representative driving cycles for energy storage based urban rail vehicles. In: DEStech Transactions on Environment, Energy and Earth Sciences (2018)
8. Berzi, L., Delogu, M., Pierini, M.: Development of driving cycles for electric vehicles in the context of the city of Florence. Transp. Res. Part D: Transp. Environ. **47**, 299–322 (2016)
9. Nyberg, P., Frisk, E., Nielsen, L.: Generation of equivalent driving cycles using Markov chains and mean tractive force components. IFAC Proc. Vol. (IFAC-PapersOnline) **19**, 8787–8792 (2014)
10. Ji, Y., Li, C., Xie, J., Wang, Y., Guo, W.: Bus driving cycle construct based on principal component analysis for Lanzhou City. In: DEStech Transactions on Computer Science and Engineering (2018). https://doi.org/10.12783/dtcse/wicom2018/26289
11. Abas, M.A., Rajoo, S., Abidin, S.F.Z.: Development of Malaysian urban drive cycle using vehicle and engine parameters. Transp. Res. Part D: Transp. Environ. **63**, 388–403 (2018)
12. Fotouhi, A., Montazeri-Gh, M.: Tehran driving cycle development using the k-means clustering method. Scientia Iranica **20**(2), 286–293 (2013)
13. Galgamuwa, U., Perera, L., Bandara, S.: Development of a driving cycle for Colombo, Sri Lanka: an economical approach for developing countries. J. Adv. Transp. **50**(7), 1520–1530 (2016)
14. Andre, M.: The ARTEMIS European driving cycles for measuring car pollutant emissions. Sci. Total Environ. **334–335**, 73–84 (2014)
15. Xiao, Z., Dui-Jia, Z., Jun-Min, S.: A synthesis of methodologies and practices for developing driving cycles. International materials science society. In: Proceedings of 2012 International Conference on Future Energy, Environment, and Materials(V16 (Part C)), pp. 474–479 (2012)
16. Knez, M., Muneer, T., Jereb, B., Cullinane, K.: The estimation of a driving cycle for Celje and a comparison to other European cities. Sustain. Cities Soc. **11**, 56–60 (2014)
17. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2625 (2008)
18. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)

A Survey on an Artificial Intelligence Approach to Maintenance of Solar Photovoltaic Modules

Oscar Famous Darteh, Qi Liu, Collins Oduro, Xiaodong Liu, and Charity Oseiwah Adjei

Abstract This paper focuses on appropriate technology to improve solar photovoltaic module maintenance. Literature were reviewed on existing traditional approaches provided by PV module manufacturers, solar energy institutional boards as well as the related works done by researchers. An artificial intelligence approach to the maintenance of Solar PV modules was proposed. The approach is to eliminate huge maintenance line during maintenance, reduce accidents during maintenance and reduce work hours during maintenance. A theoretical prototype of the robotic control system was studied. The system is designed to have abilities to initiate maintenance process on condition of timing and environmental factors. A block diagram consisting of mechatronic devices, electronic sensors and timing circuits were used. The system detects targets by pattern, feature, and colour of the solar photovoltaic modules in array. The proximity technology used was the inductive type that detects magnetic loss due to eddy currents that are generated on a conductive surface by an external magnetic field.

Keywords Artificial Intelligence · Solar photovoltaic (PV) module · Solar PV cell · Robotic system

O. F. Darteh (✉) · C. O. Adjei
School of Computer and Software,
Nanjing University of Information Science and Technology, Nanjing, China

Q. Liu (✉)
Shandong Beiming Medical Technology Ltd., Jinan, China

C. Oduro
Research Institute for History of Science and Technology,
Nanjing University of Information Science and Technology, Nanjing, China

X. Liu
School of Computing, Napier University, Edinburgh, UK

1   Introduction

Renewable energy use has been one of the main drives of mankind, especially in the eighteenth century, where solar energy also played a remarkable role. In 1767, a scientist from Switzerland, Horace de Saussure started developing a solar energy collector which was popularly known as the hot box, and this was the first in the entire world. In 1839, French scientist Edmond Becquerel also discovered the photovoltaic effect while experimenting with an electrolytic cell. This advancement improved in 1860 when the French mathematician Auguste Mouchout developed the very first motor that uses solar energy. A steam engine that operated on solar power was invented by him and was eventually used to make ice when connected to a refrigerator. A research by William Grylls and Richard day in 1876 also showed how selenium can produce electricity when exposed to light. It was discovered that the selenium cells even though, not too thorough shows that electricity can be generated without heat of light or moving parts [1]. It is undeniable that energy produce by the sun is so enormous and sustainable energy source among the others on the earth. The sun is estimated to have provided more than 150,000 TW the earth's power. It was also projected that the small amount of energy which reaches the surface of the earth is able to meet energy demand globally. It was also revealed that one of lucrative scientific opportunity is to devise thorough ways to track, convert, and store and make use of solar energy [2].

One of the many countries making good use of the sun's abundant energy is China. According to a report by Bloomberg New Energy Finance in 2017, solar power installation in China was predicted to reach 54 GW as compared to the 34.5 GW predicted in 2016 [3].

In 2016, a solar power generating Chinese owned company, BXC installed 20 mega-watts of energy to the Ghana national electricity grid [4]. Amidst this great intervention by solar energy in the energy mix, there have been concerns about how tedious and unfavourable the maintenance of the solar PV modules which are the main component in the solar energy system is and how it has impacted negatively on return on investment.

This paper seeks to identify various solar energy operation and maintenance methods used by industry players in Ghana, access the advantages and disadvantages, and propose an artificial intelligence approach to the maintenance of the solar photovoltaic modules.

2   Problem and Observation of PV Module Maintenance

Generally, water from bore wells and treated standpipes are used as sources for periodic cleaning of panels. Otherwise, water needs to be procured from tankers. Maintenance of this nature is controllable for smaller solar PV installations, but expensive, time-wasting and problematic for biggest solar PV installations, where

PV's are in-stalled across over 20 acres of land and/or on high rise buildings. The tendency of growth of vegetation around solar PV plants and number of manpower and equipment needed raises much concern about this method of maintenance. It is against this backdrop that this paper seeks to design a system prototype of an artificial intelligence method to the maintenance of solar photovoltaic modules.

The establishment of an artificial intelligence method to the maintenance of solar photovoltaic modules is essential for maintaining accuracy. It has been observed that the current traditional method employed in solar PV maintenance exposes maintenance personnel to a lot of safety hazards, requires a sizeable maintenance line, time-wasting, and is expensive. Human error can also not be eliminated with the use of this traditional method.

The artificial intelligence approach will among other things ensure:

- A lean maintenance line during maintenance since an automated robot has the ability to be more productive in a shorter time than a human worker and also can work at a constant speed without breaks, pausing, sleep, or vacations.
- Reduce accidents, health issues and work-related stress among employees working at height and dangerous environments.

This paper focuses mainly on an artificial intelligence approach to the maintenance of solar PV modules in grid-connected PV systems; and explicitly does not cover maintenance for inverters, and other related equipment.

2.1 Existing Approaches for PV Module Maintenance

Mostly, environmental conditions contributes to loss of solar PV arrays' output power. Accumulation of dirt on the surface of the PV module shown in Fig. 1 reduces solar PV arrays' output power which finally leads to localized hot spot flops if the dirt is uneven.

Currently, most solar PV maintenance companies rely on maintenance guidelines provided by solar PV module manufacturers and approved PV System Operations and Maintenance institutions as listed below:

Jinko Solar Manufacturers' Manual. Jinko Solar manufacturers' manual stipulates the following guidelines for cleaning:

1. The output energy is affected by dirt on the module and dust build-up. Cleaning the panel every year or less, preferably with soft cloth or damp depending on the conditions is highly recommended. However, water with high mineral content will leave deposits on the glass surface and is not advised.
2. Abrasive material is not advisable to use under any circumstances in cleaning.
3. Risk of electric shock should be avoided by carrying out maintenance or cleaning of PV modules in the morning or during low-temperature hours.

Fig. 1 Soiling of solar PV module

4. Moreover, it is highly recommended to prevent potential shock risk by avoiding PV modules with broken glasses or exposed wires since water is a good conductor of electricity.

Solar America Board for Codes and Standards. A technical report written by Josh Haney and Adam Burstein on PV System Operations and Maintenance Fundamentals supported by the Solar America Board for Codes and Standards (Solar ABCs) published on Solar ABCs website further stipulates the following procedures for PV module cleaning:

1. Demineralized water is recommended to prevent spotting and calcium build-up on the module.
2. The use of considerable pressure washer that sprays water between 3.5 to 6 kg/cm^2.
3. However, it is advisable to keep the washer some distance away from the PV module under cleaning when using a high-pressure washer to prevent cause of damage to the PV modules.
4. Sprinkle water on the PV module.
5. Use a soft brush to get debris or dirt off of the module.
6. Ensure the soap being used for cleaning will not cause damage to the solar PV modules.
7. When necessary use extensions with the tools to be able to reach extended distances.
8. The modules should be cleaned dry after cleaning to aid testing of post maintenance output power [5, 6].

It was identified that the above maintenance approaches had faced serious set-backs in water conservation, work speed, maintenance period prediction. A 63 Kilo-watts-Hour peak solar PV systems installed in Elmina, Ghana covering 520 m^2 area uses over 26000 L of water for maintenance of the solar PV modules. Additionally, there is no reliable system to detect energy output problems of PV modules caused by environmental factors such as soil, bed droppings and icing.

PV Cleaner Robot V 1.0. The most current approach is a research paper published by Mark Anderson and colleagues on robotic cleaning system for solar PV modules. The designed PV Cleaner Robot V 1.0 is made up of a head that moves across the modules by a pair of motorized trolleys in the drive system which rides along the top and bottom edges of the array as schematically shown in Fig. 2. In terms of safety of the module, the PV Cleaner Robot V1.0 provides the safest means of cleaning and ensures the module is scratch-free. Moreover, in terms of the functionality of the system the PV Cleaner Robot V1.0 can be further expanded to provide more useful functions. For example, the robotic system can adapt easily by interchanging belts depending on the number of panel rows [7].

The drive system has inductive proximity sensors attached to the frame of the trolley and is responsible for tracking the cleaning head when it approaches the module edge and slows the motion of cleaning head to enable gap crossing or stop the cleaning head when it gets to the end of the array. A prototype of the PV Cleaner Robot V1.0. The Fig. 3 left below shows how thoroughly the module was
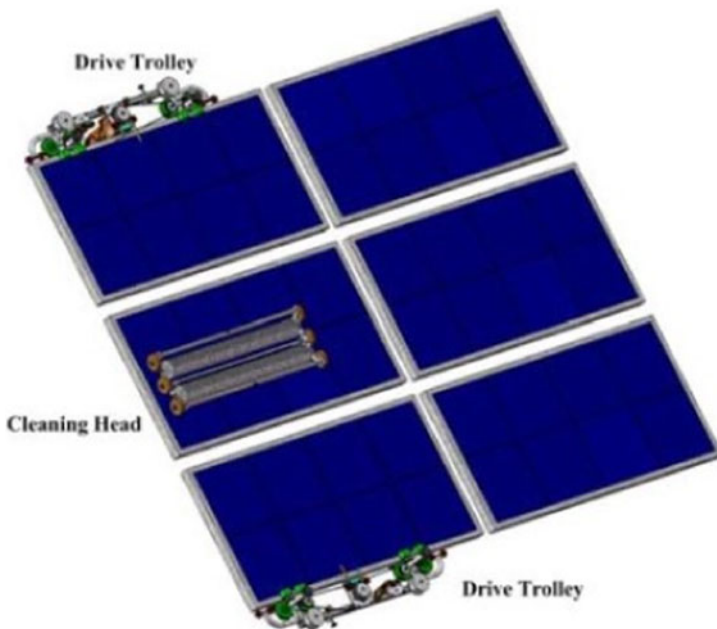
Fig. 2 Schematic of a PV cleaner robot V1.0

cleaned after passing the cleaner head once. Figure 3 right is the module at its dirty state. Sawdust, dirt, dry wet clayey soil was successfully cleaned after the first cleaning process.

The cleaning width for the PV Cleaner Robot is between 2 m to 5 m, which makes implementation of the system very cumbersome as a sizeable number of drive trolley will be required for bulky system installation covering huge solar farms. Implementation in solar PV power installations where over ten (10) numbers of a one (1) meter solar PV modules are connected in arrays is limited for this approach. The designer, however, guaranteed 90% cleaning of the modules [7]. However, any minimal debris of sand can lead to loss of substantial power output.

T. Mobin of National Institute of Technology, Rourkela-India also designed and developed a solar module cleaning system. The system consists of a locomotion Unit and Cleaning Unit. The central spline houses the cleaning unit. The cleaning unit moves into and fro motion. On the cleaning unit is the cylindrical brush that moves only in a clockwise direction. The movement of the central spline at the bottom of the module causes the rotating brushes mounted on the cleaning unit to move along the modules. On each module route, the central spline pushes the dirt in the movement direction of the cleaning unit. The cleaning unit then expels the dirt off the module. The cleaning unit returns to its position when it approaches the central spline's lower end. The cleaning unit also seizes movement when the approaches the top of the spline. At this stage, the locomotive units discharge suction cup and engage the system in rest. The discharge of the suction cup causes movement of the wheels parallel to the end of the module that has not been cleaned. To make the system still again, the suction cups are reengaged. The engagement
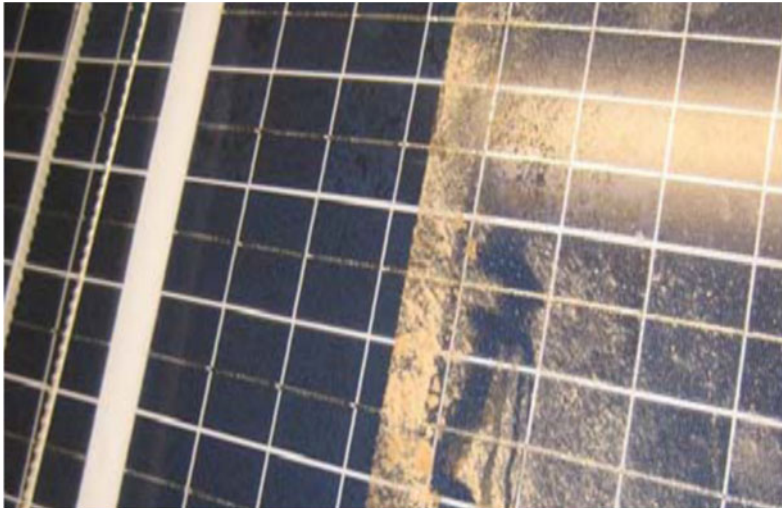
Fig. 3 Dirty panel after single pass with the PV cleaner robot V1.0 on the left while the right is the PV module at its dirty state

and reengagement of the cups continue in process until the entire array system is serviced [14] (Fig. 4).

3 An Artificial Intelligence Approach

A Solar PV module consists of several numbers of solar cells arranged in a block. A solar cell can be simply likened to as the fundamental building component of a solar PV module. One solar cell generates close to 0.5 V, and the charging voltage of a solar PV module is dependent on the number of solar cells making up the PV module. The voltage of batteries 36 cell solar PV modules can charge, for instance, is normally 12 V. In this paper, a photovoltaic cell (PV cell) is defined as a semiconductor material which changes sunlight into electricity in the direct current form. The solar array, however, is the arrangement of modules in either series or parallel to produce the desired charging voltage.

Data that was used for this paper came from the solar power installation and maintenance companies based in Accra, Ghana. The data were collected via phone conversation due to the limited resources available. The questionnaire was centred on the existing maintenance approach and proposal for a new maintenances approach. The sample size was twenty (20). Amongst these companies, sixteen (16) could not remember the last date maintenance was carried out on the PV systems installed for clients due to clients' resistance to maintenance schedules as a result of the high cost of maintenance. The companies could not also tell if the PV systems were still functioning. The remaining four companies rely on maintenance
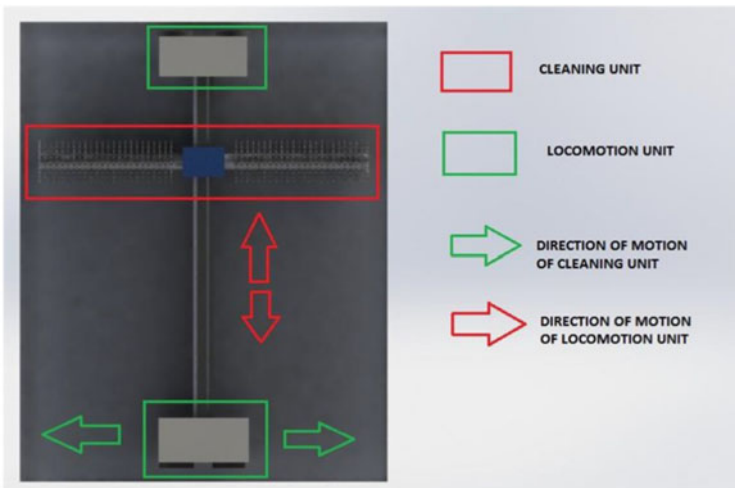
Fig. 4 Operation of the solar module cleaning system

guidelines provided in Sect. 2.1. All the 20 companies welcomed the idea of an artificial intelligence approach to the maintenance of the PV modules.

The paper defines artificial intelligence as the act of improving the functions of machines to perform human-like intelligence abilities such as listening, adapting, learning and auto-correction [8]. The paper seeks to propose an artificial intelligence approach that employs an automated maintenance solution that will solve the problems in traditional maintenance methods. The most important benefit of an artificial intelligence approach for maintenance of solar PV modules is the great speed, efficiency, and recognition the system exhibits during operation in comparison to human methods and existing approaches.

The approach is to provide a design of a theoretical prototype of a robotic control system with built-in auto inspection and detection sensors to perform the following functions:

- Monitor maintenance schedule via timing circuit.
- Detect the darkness at the installed location of the PV module.
- Control movement and proximity of robotic cleaning system.

The system will be similar to a robot system designed for a pipe welding by Huang et al. 2008. The designed robot pipe welding system uses vision sensor which insusceptible to noise but has a very high pattern recognition ability. The vision sensor is precise in identifying dimensions; tracks welded seams and control movement of robot [9].

The approach in this paper proposes the use of a robotic system control system block comprising of an electronic timer switch, a 555-dark detector, an actuating device, optical vision and proximity sensing and a robotic control system as shown in Fig. 5. The red termination slots are listed 1–5 according to the process flow.
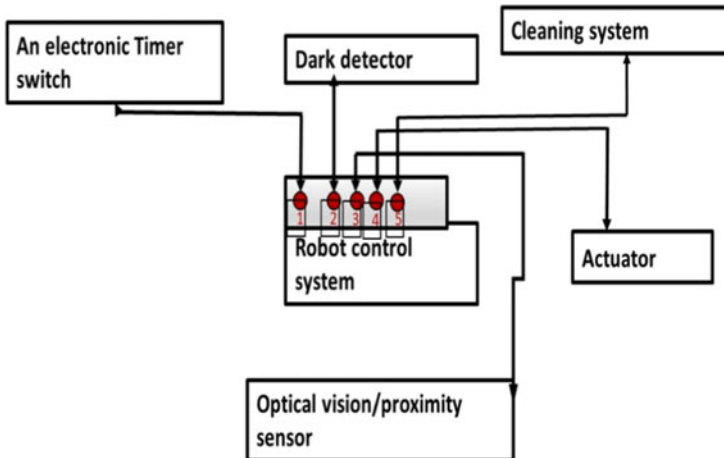
Fig. 5 Theoretical prototype of the robotic control system block

3.1   Theoretical Prototype for a Robotic Control System

Timer Block.   Timers are used to provide logic when a circuit turns on or off. The contacts are provided both normally open (NO) and normally closed (NC) for each type of timer [10]. The purpose of the timer in the block is to close the Dark detector circuit after every 3 months (In this case, cleaning of the PV module is to be carried out in every three months). In NO position, the timer disconnects the electrical connection to the dark detector. In the NC position, the timer makes electrical connection to the dark detector.

Dark Detector Block (555).   The 555 dark detector block will check the presence of sunlight. It has an electrical connection with the robotic control system and the actuator. If there's sunlight, it will need actuator will not send an activation message to the robot control system. The reason for this action is to ensure the safety of the PV system. The PV modules produce electricity with the slightest sun, so using of water during this period is dangerous. The dark detector will only activate the robot control system when there is dark. The detector to be used is similar to a regular cadmium-sulfide light dependent Resistor that senses the unavailability of light for operation via an output [11].

Vision Sensor Block.   Vision sensors show numerous attributes which qualify them to be thorough in industrial automation. Critically, vision sensors are able to manipulate several inspections on a single instruction. Vision sensors are also able recognize patterns, features and colour based on target instructed. In a very wide working environments, they can even detect specific components within the specific time instructed [12]. In this paper, the vision sensor will be electrically connected to the actuating system and will perform functions of capturing and inspecting the pattern, feature and colour of the modules. The sensor consists of the following components: lights, lenses, image capture, control unit, communications units and vision and an inductive proximity sensing component. The lighting component provides light to illuminate the modules and allow its feature to stand out for the camera's viewing. The lens records the module's image and presents the message to the vision sensors for capture. The purpose of the image sensor here is to convert the illuminated light into a digital image to be processed by the image processor. Vision processor then processes and rearranges the image for analysis; evaluate the image, extracts required information; and then uses the algorithms to simulate the needed inspection and makes the decision required. The communication protocols then send information to the actuating system to perform a task of either starting process or ending process. A metallic object will be fixed on edges of the PV arrays as the proximity. The work of the inductive proximity sensor is to detect the presence of the metallic object fixed at edges of every solar PV modules.

Actuator Block.   Actuators are devices that control the operation of mechanisms. An actuator turns a control signal into mechanical action such as an electric motor [13]. The proposed actuator for this is electric/mechanical driven by software. Its

function is to activate the robotic control system to initiate movement, reverse the movement and end working process. This is done based on information from the vision sensor. For in-stance, the actuator will activate the robot control system reverse movement when there is a proximity object.

Robot Control System. The robot control system is the mother of all blocks in the system. The communication, processing and executing are all done in this block. This block houses the board for software installation. An Arduino board is best suited. All electrical connection from all blocks is made through this system.

Cleaning System Block. The cleaning system will consist of the robot, the cleaning agents like detergent, water sprinkler and a soft cloth. It is controlled and monitored by the robot control system.

4   Conclusion

Based on the functions of each block, in the theoretical prototype of the robotic control system, the system after experiment and design should function. Moving away from the traditional method of Solar PV module maintenance is imminent. Further research should subject this paper to laboratory experiments and design.

References

1. Tomar, A., Jain, L., Batra, P.: Solar energy-finding new ways. Int. J. Res. Advent Technol. **1** (5), 1–3 (2013)
2. Camacho, E.F., Berenguel, M., Alvarado, I., Limon, D.: Control of solar power systems: a survey. In: Proceedings of the 9th International Symposium on Dynamics and Control of Process Systems, pp. 809–810. International Federation of Automatic Control, Leuven, Belgium DYCOPS (2010)
3. Buckley, T., Nicholas, S., Brown, M.: China 2017 review: world's second biggest economy continues to drive global trends in energy investment. Institute of energy economics and financial analysis. http://ieefa.org/wp-content/uploads/2018/01/China-Review-2017.pdf. Accessed 06 Nov 2019
4. National energy statistics 2007–2016, Ghana Energy Commission, Strategic Planning and Policy Directorate. http://energycom.gov.gh/files/ENEERGY_STATISTICS_2017_Revised. pdf
5. Jinko Solar Photovoltaic Module Installation Manual. https://jinkosolar.com/ftp/TUV%20DV %20Manual%ef%bc%88D-V%ef%b-c%8920-16-07.pdf. Accessed 30 Oct 2019

6. Haney, J., Burstein, A.: Solar America board for codes and standards report: "PV System Operations and Maintenance Fundamentals", p. 40. http://www.solarabcs.org/about/public-ations/reports/operations-maintenance/p-dfs/SolarABCs-35-2013.pdf. Accessed 18 Oct 2019

7. Anderson, M., et al.: Robotic device for cleaning photovoltaic panel arrays. In: International Conference on Climbing and Walking Robots (CLAWAR 2009), pp. 1–10. Northeastern University (2014)

8. Kok, J.N., et al.: Artificial intelligence: definition, trends, techniques and cases. https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf. Accessed 25 Oct 2019

9. Huang, J., et al.: Study on a pipe welding robot based on laser vision sensing. In: IEEE Conference on Robotics, Automation and Mechatronics 2008, Beijing, China, pp. 720–723 (2008)

10. Anonymous: Timers, Counters and T/C Applications. http://www.en-g.utoledo.edu/∼wevans/chap7_S.pdf. Accessed 23 Oct 2019

11. Corporation SIc TMachine TMc M, The 555 Timer IC. http://www.electronics.dit.ie/staff/mtully/555%20folder/555%20timer.htm. Accessed 05 Nov 2019

12. Benoit, B.: Introduction to vision sensors: the case for automation with machine vision. https://files.constantcontact.com/f9f40497001/97189b78-5d4e-4658-8545-d9e088b896f3.pdf. Accessed 29 Oct 2019

13. Anonymous: What does actuator really mean? https://www.tech-opedia.com/definition/17043/actuator. Accessed 29 Oct 2019

14. Mobin, S.T., Pnada, S.K.: Design and development of solar panel cleaning system. Int. J. Adv. Eng. Res. Dev. **4**(4) (2017). https://doi.org/10.21090/ijaerd.me035

Author Index