Journal of Applied Mechanics

Published Bimonthly by ASME

VOLUME 74 • NUMBER 6 • NOVEMBER 2007

Transactions of the ASME, Journal of Applied Mechanics (ISSN 0021-8936) is published bimonthly (Jan., Mar., May, July, Sept., Nov.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Applied Mechanics, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300.

CHANGES OF ADDRESS must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address.

Canadian Goods & Services Tax Registration #126148048.

TECHNICAL PAPERS

(Contents continued on inside back cover)

This journal is printed on acid-free paper, which exceeds the ANSI Z39.48-1992 specification for permanence of paper and library materials. ♾™
85% recycled content, including 10% post-consumer fibers.

(**Contents continued**)

Journal of Applied Mechanics **Volume 74, Number 6** **NOVEMBER 2007**

TECHNICAL BRIEFS

The ASME Journal of Applied Mechanics is abstracted and indexed in the following:

Alloys Index, Aluminum Industry Abstracts, Applied Science & Technology Index, Ceramic Abstracts, Chemical Abstracts, Civil Engineering Abstracts, Compendex (The electronic equivalent of Engineering Index), Computer & Information Systems Abstracts, Corrosion Abstracts, Current Contents, EEA (Earthquake Engineering Abstracts Database), Electronics & Communications Abstracts Journal, Engineered Materials Abstracts, Engineering Index, Environmental Engineering Abstracts, Environmental Science and Pollution Management, Fluidex, Fuel & Energy Abstracts, GeoRef, Geotechnical Abstracts, INSPEC, International Aerospace Abstracts, Journal of Ferrocement, Materials Science Citation Index, Mechanical Engineering Abstracts, METADEX (The electronic equivalent of Metals Abstracts and Alloys Index), Metals Abstracts, Nonferrous Metals Alert, Polymers Ceramics Composites Alert, Referativnyi Zhurnal, Science Citation Index, SciSearch (Electronic equivalent of Science Citation Index), Shock and Vibration Digest, Solid State and Superconductivity Abstracts, Steels Alert, Zentralblatt MATH

Experiment and Theory on the Nonlinear Vibration of a Shallow Arch Under Harmonic Excitation at the End

Jen-San Chen
Professor
e-mail: jschen@ccms.ntu.edu.tw

Cheng-Han Yang

Department of Mechanical Engineering,
National Taiwan University,
Taipei, Taiwan 10617

In this paper we study, both theoretically and experimentally, the nonlinear vibration of a shallow arch with one end attached to an electro-mechanical shaker. In the experiment we generate harmonic magnetic force on the central core of the shaker by controlling the electric current flowing into the shaker. The end motion of the arch is in general not harmonic, especially when the amplitude of lateral vibration is large. In the case when the excitation frequency is close to the nth *natural frequency of the arch, we found that geometrical imperfection is the key for the* nth *mode to be excited. Analytical formula relating the amplitude of the steady state response and the geometrical imperfection can be derived via a multiple scale analysis. In the case when the excitation frequency is close to two times of the* nth *natural frequency two stable steady state responses can exist simultaneously. As a consequence jump phenomenon is observed when the excitation frequency sweeps upward. The effect of geometrical imperfection on the steady state response is minimal in this case. The multiple scale analysis not only predicts the amplitudes and phases of both the stable and unstable solutions, but also predicts analytically the frequency at which jump phenomenon occurs.* [DOI: 10.1115/1.2165231]

1 Introduction

Historically, the interest in shallow arch research is primarily on the snap-through buckling when the arch is under some kind of lateral loading. The first theoretical prediction on the static critical load was conducted by Timoshenko in 1935 [1], in which a pinned sinusoidal arch was subjected to a uniformly distributed load. Timoshenko's pioneering work was followed and extended by many other researchers on various kinds of topics, including the snap-through phenomenon under dynamic load. The first theoretical prediction of dynamic buckling load was conducted by Hoff and Bruce in 1954 [2], in which they studied the stability of a sinusoidal arch under unit step loading and ideal impulsive loading. A good introduction and more references on the static and dynamic snap-through buckling can be found in the two books by Simitses [3,4]. A more up-to-date reference list on these subjects can be found in a recently published paper by Chen and Liao [5].

Besides the quasi-static, impulsive, and step loadings discussed above, the dynamic response of a shallow arch under periodic excitation has also been studied by some researchers. Previous research in this regard may be roughly divided into two groups. The first group considers the case when the vibration amplitude is so large that dynamic snap-through occurs. The critical load amplitude and frequency of an arch under harmonic excitation is harder to predict compared to the problem of step or impulsive loading. Huang [6] used a cycle-averaging approach to predict the critical load when the excitation frequency is high. Plaut and Hsieh [7] used a one-term approximation to numerically study the critical load when the arch is subject to a two-frequency excitation. Blair et al. [8] simplified the arch with a two-rigid-link model and used harmonic balance method to study its dynamic response when it is under harmonic excitation.

The second group focuses on smaller excitation and vibration amplitude with emphasis on the nonlinear response such as superharmonic, subharmonic, and internal resonances. Thompson [9] studied the chaotic vibration of a circular high arch loaded by a harmonic force at its crown, a problem initiated by Boloton in his book [10]. Tien and Sri Namachchivaya [11,12] and Bi and Dai [13] used a two-term approximation to study the internal resonance of a shallow arch under lateral loading with both averaging method and direct integration method. With the same two-term approximation, Malhotra and Sri Namachchivaya [14,15] also investigated the possibility of chaotic response when the shallow arch is under 1:1 and 1:2 resonances.

From the above literature review we notice that while some theoretical investigations on the dynamic response of a periodically excited shallow arch exist, experimental investigation is relatively rare. Generally speaking, theoretical investigations tend to oversimplify the reality. For instance, it is not an easy matter to realize a pulsating load with constant amplitude in the laboratory because the structure will change the characteristics of the loading mechanism, especially when the vibration amplitude becomes large. For the internal resonance phenomenon discussed in Refs. [11–15], it can occur only when special relations among the natural frequencies of the arch happen to exist. In the real world, this situation occurs only accidentally.

One of the values of experimental work is that it can sometimes inspire new research ideas and help us establish more realistic model for analysis. In this paper we set out to design an experimental setup aiming to observe the nonlinear vibration of a shallow arch with an electro-mechanical shaker attached to one end. Special attention is focused on the case when the excitation frequency is close to the *n*th natural frequency (coupling resonance) and two times of the *n*th natural frequency (parametric resonance). Although the vibration phenomenon is nonlinear in nature the vibration amplitude of the arch is limited to be small enough so that no snap-through buckling will occur. The experimental observation is then compared to the numerical simulation based on the theoretical model. In order to capture the physical essence of the nonlinear vibration observed in the experiment and numerical simulation, we also develop an analytical technique based on the

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received May 21, 2005; final manuscript received November 7, 2005. Review conducted by N. Sri Namachchivaya.

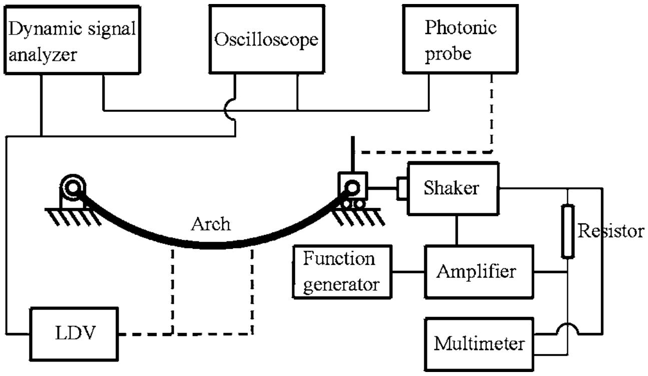


Fig. 1 Schematic diagram of the experimental setup

multiple scale method to formulate the closed-form solution of the steady state response of the arch. The theoretical and experimental results agree reasonably well.

2 Experimental Setup

Figure 1 is a schematic diagram of the experimental setup. The arch is made of copper strip with Young's modulus 101 GPa and mass density 8840 kg/m^3. The length L of the arch is 44 cm and the cross section is 25 mm×1.2 mm. Both ends of the curved beam are attached to roller bearings to simulate pinned condition. One end of the arch is attached to an electro-magnetic shaker via a stinger. The magnetic force on the central core of the shaker is controlled by the current flowing into the shaker. A power amplifier connected to the shaker is responsible for pumping electric current proportional to the harmonic voltage signal from a function generator. The current can be monitored by measuring the voltage across a high-power low-ohm resistor. With this arrangement, the magnetic force on the central core of the shaker can be estimated accurately. Linear bearing is installed to reduce the friction when the end attachment slides on the guiding rods. The total mass of the bearing installation at the end is measured as $m_e = 0.8$ kg. The axial motion of the sliding end is monitored by a photonic probe (MTI 2000). The lateral speeds at various locations of the arch are measured by a two-channel LDV system. The speed signals from the LDV system can be integrated to obtain the displacement history.

Although the arch is designed to match a half-sinusoidal curve as closely as possible, some minor deviation can still be detected. This deviation is the geometrical imperfection. It will be shown that these minor imperfections are of paramount importance when the excitation frequency of the shaker is close to one of the natural frequencies of the arch. We assume that the initial shape of the arch y_0, measured from the base line passing through the two ends, can be expanded in a Fourier series

$$y_0 = \sum_{n=1}^{\infty} h_n \sin \frac{n\pi x}{L} \tag{1}$$

To determine the coefficients of various Fourier components in Eq. (1), we measure the deviation of the arch to the designed half-sinusoidal curve at 40 equidistant locations. By employing a least-squares method we can estimate the first eight harmonic components $h_1 - h_8$, as listed in Table 1. The second row of Table 1 lists the physical height h_n. The third row lists the dimensionless height h_n^*, whose definition can be found in Eq. (5) later. It is noted that the initial shape of the specimen is indeed very close to the half-sinusoidal shape with h_1 being the dominant coefficient. Most of the coefficients corresponding to the imperfection (h_n with $n \geqslant 2$) are on the order of less than 1% of h_1, except h_3 which amounts to 6.4% of h_1.

Table 1 Coefficients of Fourier expansion of the initial shape. The second row lists the physical height h_n. The third row lists the dimensionless height h_n^*.

n	1	2	3	4	5	6	7	8
h_n (mm)	16.10	−0.19	−1.03	0.14	0.06	−0.01	−0.005	0.003
h_n^*	46.5	−0.55	−2.98	0.41	0.18	−0.04	−0.014	0.01

As a first approximation, the central core and suspension of the shaker is modeled as a one degree-of-freedom mass-spring oscillator. The natural frequency of this central core-suspension system is measured at 114.88 Hz. The spring constant of the suspension is estimated as $k = 22253$ N/m by measuring the displacements (with power off) of the central core when different forces (weights) are applied. The effective mass of the oscillator is then calculated as $m_s = 0.043$ kg. We are not interested in the damping of the shaker at this stage because we are more interested in the damping of the whole system when the shaker is attached to the movable end of the arch. When the central core is connected to the movable end of the arch in series, the total mass $m = 0.843$ kg will be the combination of m_e and m_s. The theoretical model of our experimental setup is shown in Fig. 2. It is noted that while we can control the force on the end mass m by controlling the electric current flowing into the shaker, we cannot control the motion e of the end mass. Therefore, the experimental setup is under load control.

3 Equations of Motion

Consider the theoretical model shown in Fig. 2. The two pinned ends of the elastic shallow arch are originally separated by a distance L. The arch is free of lateral loading. At one end, the arch is connected to a sliding mass m, which is restrained by a spring k. The mass is subjected to a harmonic excitation force $2f \cos \gamma t$, where $2f$ and γ are the amplitude and the frequency of the force. The initial and deformed shapes of the arch are y_0 and y, both measured from the same baseline. The equation of motion of the arch can be written as

$$\rho A y_{,tt} = -EI(y - y_0)_{,xxxx} + p y_{,xx} \tag{2}$$

The parameters E, ρ, A, and I are Young's modulus, mass density, area, and area moment of inertia of the cross section of the arch. p is the axial force in the deformed arch. The force balance of the attached end mass gives an additional equation

$$m\ddot{e} = -p - ke + 2f \cos \gamma t \tag{3}$$

The relation between the end motion e and the shape change of the arch can be established from the elastic extensibility of the arch

$$p = \frac{EA}{L}\left\{ e + \frac{1}{2}\int_0^L [(y_{,x})^2 - (y_{0,x})^2]dx \right\} \tag{4}$$

After replacing the axial force p in Eqs. (2) and (3) with Eq. (4), we can obtain the two equations of motion governing the deformed shape y and end motion e of the arch. It is noted that

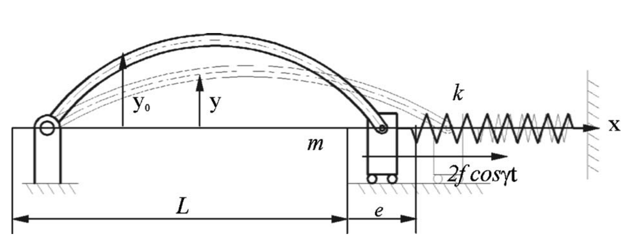


Fig. 2 Theoretical model of the arch-shaker assembly

Mettle formulated similar equations for a straight beam with end mass and under prescribed oscillating axial load in 1962 [16].

Equations (2)–(4) can be nondimensionalized by introducing the following dimensionless parameters (with asterisks)

$$(y^*,y_0^*,h^*)=\frac{1}{r}(y,y_0,h)\quad x^*=\frac{\pi x}{L}\quad t^*=\frac{\pi^2 t}{L^2}\sqrt{\frac{EI}{A\rho}}$$

$$(\gamma^*,\omega^*)=\frac{L^2}{\pi^2}\sqrt{\frac{A\rho}{EI}}(\gamma,\omega)\quad (p^*,f^*)=\frac{L^2}{\pi^2 EI}(p,f)\quad e^*=\frac{Le}{\pi^2 r^2}$$

$$m^*=\frac{I\pi^4 m}{L^3A^2\rho}\quad k^*=\frac{Lk}{EA}\quad \mu^*=\frac{\mu L^2}{\pi^2\rho Ar}\sqrt{\frac{\rho}{E}} \tag{5}$$

The parameters μ and ω are the damping and natural frequency of the arch-shaker system, which will be discussed later. r is the radius of gyration of the cross section of the arch. After substituting the above relations into Eqs. (2)–(4), and dropping all the superposed asterisks thereafter for simplicity, we obtain the two dimensionless equations of motion

$$y_{,tt}=-(y-y_0)_{,xxxx}+\left(e+\frac{1}{2\pi}\int_0^\pi[(y_{,x})^2-(y_{0,x})^2]\right)y_{,xx} \tag{6}$$

$$m\ddot{e}=-\frac{1}{2\pi}\int_0^\pi[(y_{,x})^2-(y_{0,x})^2]dx-(1+k)e+2f\cos\gamma t \tag{7}$$

The boundary conditions for y at $x=0$ and π are

$$y(0)-y_0(0)=y_{,xx}(0)-y_{0,xx}(0)=y(\pi)-y_0(\pi)=y_{,xx}(\pi)-y_{0,xx}(\pi)=0 \tag{8}$$

The dimensionless version of Eq. (1), the initial shape, can be written as

$$y_0=\sum_{n=1}^\infty h_n\sin nx \tag{9}$$

It is assumed that the shape of the loaded arch can be expanded as

$$y(t)=y_0+\sum_{n=1}^\infty \alpha_n(t)\sin nx \tag{10}$$

After substituting Eqs. (9) and (10) into Eqs. (6) and (7), multiplying Eq. (6) by $\sin nx$ and integrating from $x=0$ to π (Galerkin's procedure), we obtain the equations governing α_n and e

$$\ddot{\alpha}_n=-n^4\alpha_n-n^2(h_n+\alpha_n)\left[e+\frac{1}{4}\sum_{j=1}^\infty j^2(\alpha_j^2+2h_j\alpha_j)\right]\qquad n=1,2,3,\ldots \tag{11}$$

$$m\ddot{e}=-e(1+k)+2f\cos\gamma t-\frac{1}{4}\sum_{j=1}^\infty j^2(\alpha_j^2+2h_j\alpha_j) \tag{12}$$

4 Natural Frequencies of the Assembly

The linearized version of Eqs. (11) and (12), without excitation, are

$$\ddot{\alpha}_n=-n^4\alpha_n-n^2\left(h_ne+\frac{1}{2}\sum_{j=1}^\infty j^2h_nh_j\alpha_j\right)\quad n=1,2,3,\ldots \tag{13}$$

$$m\ddot{e}=-e(1+k)-\frac{1}{2}\sum_{j=1}^\infty j^2h_j\alpha_j \tag{14}$$

Table 2 The first eight natural frequencies of the assembly. The second row (dimensionless) and the third row (dimensional) are calculated from Eqs. (13) and (14). The fourth row is measured from experiment.

i	1	2	3	4	5	6	7	8
ω_i^*	1.744	3.999	8.578	15.95	24.96	35.99	48.99	64.02
ω_i (Hz)	16.56	37.99	81.49	151.56	237.13	341.94	465.46	608.20
ω_i (Hz)	17	38	78	138	214	319	461	616

From these two equations we can calculate the first eight natural frequencies $\omega_1-\omega_8$ of the assembly, both dimensionless (ω_i^*, second row) and dimensional (ω_i, third row), as listed in Table 2. It is noted that the dimensionless natural frequency ω_i^* for $i\geqslant 2$ is very close to i^2, the ith natural frequency of the perfectly sinusoidal arch. The small deviation of the calculated natural frequencies from i^2 is caused by the geometrical imperfections.

One way to check whether the theoretical model for connection between the shaker and the arch is correct is to compare the above calculated natural frequencies of the arch-shaker assembly with those observed experimentally. Figure 3 shows the measured power spectrum of the arch itself without shaker (upper graph) and the arch-shaker assembly (lower graph). It is noted that the effect of the shaker attachment is to raise the first natural frequency from 10 to 17 Hz. The measured natural frequencies of the arch-shaker assembly are also recorded on the fourth row of Table 2. From the comparison of the third row and the fourth row of Table 2, we are confident that the mechanical model described in Fig. 2 and the estimated physical parameters for the shaker are not far from the truth.

5 Estimate of Damping

To simulate the motion of the assembly numerically, we still have to estimate the damping of the arch-shaker assembly. The dissipating mechanism of the system comes from the friction in the moving parts and the material damping in the arch. In order to accommodate the damping effect in the numerical simulation, we modify Eq. (11) by adding a damping parameter 2μ

$$\ddot{\alpha}_n=-2\mu\dot{\alpha}_n-n^4\alpha_n-n^2(h_n+\alpha_n)\left(e+\frac{1}{4}\sum_{j=1}^\infty j^2(\alpha_j^2+2h_j\alpha_j)\right)\qquad n=1,2,3,\ldots$$

In order to estimate the damping parameter of the arch-shaker assembly, we displace by hand the movable end of the arch a distance $e=-108$ (-0.29 mm) with the shaker attached, and then release it. The measured displacement history $y-y_0$ at the middle point of the arch is recorded as solid line in Fig. 4. For convenient reference, we present the measured results with both dimensionless parameters (left and bottom sides) and the physical ones (right and top sides). The same labeling style is adopted in all the figures involving experimental measurement. Since the oscillation frequency (17 Hz) corresponds to the first natural frequency of the arch-shaker assembly, we may assume that the first mode is dominant in the dynamic response in Fig. 4. The damping factor may be estimated from the decaying rate of the two peaks as signified by black dots. The heights of the two peaks are measured at $y_1=2.77$ (0.96 mm) and $y_2=1.90$ (0.66 mm). The ratio of the damping of the system μ to a critical damping μ_c is [17]

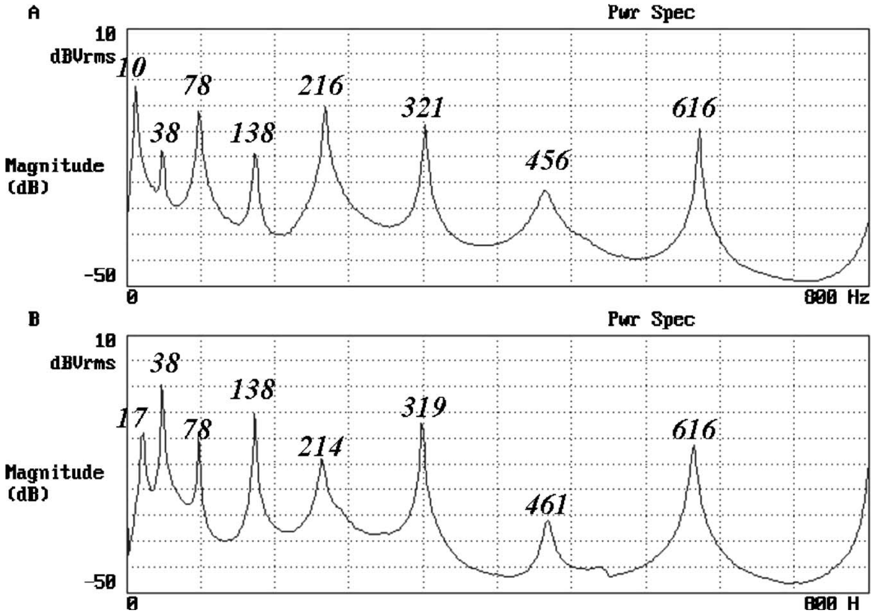


Fig. 3 Power spectrums of the arch itself (upper graph) and the arch-shaker assembly (lower graph)

$$\frac{\mu}{\mu_c} = \frac{\ln(y_1/y_2)}{\{(2\pi)^2 + [\ln(y_1/y_2)]^2\}^{1/2}} \tag{16}$$

The damping ratio in Eq. (16) is calculated as 0.06. The critical damping μ_c is estimated numerically by adjusting the damping parameter μ in Eq. (15) (using only one-mode approximation) until the response α_1 ceases to oscillate following an impulsive excitation. In this way the dimensionless critical damping μ_c is estimated as 3.0. As a consequence the damping of the assembly is calculated as $\mu=0.18$. The numerical result obtained by integrating Eqs. (12) and (15) based on this estimated damping is plotted in Fig. 4 as dashed line. Although this estimate may appear somewhat engineering-oriented, it is believed that the damping factor in our experimental setup is about this order.

6 Coupling Resonance

In the case when the arch is subjected to a lateral force with the excitation frequency close to a natural frequency of the arch, primary resonance will occur. Although in this section we also adjust the excitation frequency close to one of the natural frequencies ω_n ($n \geqslant 2$) of the arch-shaker assembly, the arch is excited in the axial direction. In this situation, the nth mode of the assembly will not be excited directly by the excitation force. Instead, the nth mode will be excited indirectly by the axial mode after the axial mode is excited directly by the excitation force. It will be shown that the geometrical imperfection serves as the coupling factor between the lateral vibration and the axial excitation. Therefore, we choose to use the term "coupling resonance" here instead of "primary resonance." More physical insight regarding this matter can be found in the next section. In Fig. 5 we record the measured steady state amplitudes of the arch at 17 evenly spaced locations with closed dots "•" when the excitation frequency γ is close to the four natural frequencies ($\omega_2-\omega_5$) of the arch, i.e., (*a*) 38 Hz, (*b*) 78 Hz, (*c*) 138 Hz, and (*d*) 216 Hz. The amplitude of excitation forces $2f$ in these four experiments are (*a*) 2.00 N, (*b*) 4.04 N, (*c*) 8.56 N, and (*d*) 10.68 N. The solid lines represent the numerical predictions from the complete theoretical model ($\alpha_1-\alpha_8$) including the small imperfections. The dashed lines are the simplified model neglecting all the imperfections h_n with $n>1$.

Several comments can be made regarding Fig. 5. (1) By comparing the closed dots and the solid lines, we observe that the measured amplitudes agree fairly well with the complete-model prediction in all four cases. The worst agreement is observed in Fig. 5(*d*). (2) The amplitude profile is symmetric with respect to the midpoint when γ is close to the natural frequency of a symmetric mode, such as ω_3 and ω_5. On the other hand, the amplitude profile is asymmetric when γ is close to the natural frequency of an asymmetric mode, such as ω_2 and ω_4. (3) In all four cases the steady state vibrations contain two dominant components, i.e., α_1 and α_n. (4) If the small imperfections h_n with $n>1$ are neglected in the numerical simulation, all the corresponding components α_n will be suppressed to zero. This simplified model gives erroneous prediction as can be seen from the comparison of solid and dashed lines. The above observations indicate that in the case when the excitation frequency is close to the nth natural frequency, the corresponding small geometrical imperfection h_n should not be ignored.

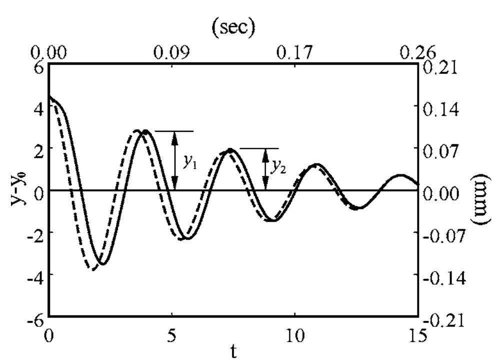


Fig. 4 Solid line is the measured lateral displacement history at the middle point of the arch after an initial displacement at the end of the arch-shaker assembly. Dashed line is the calculated response based on the estimated damping.

7 Multiple Scale Analysis for Coupling Resonance

To predict analytically the amplitude of vibration when the excitation frequency is close to ω_n, we retain only the coordinates α_1

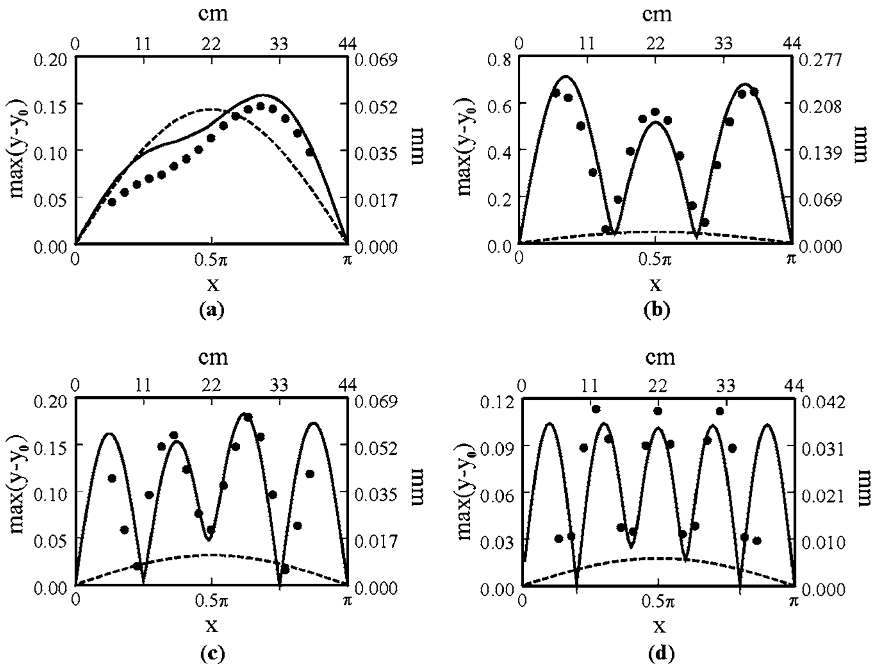


Fig. 5 Steady state amplitude profiles when the assembly is excited at (*a*) $\omega_1=38$ Hz, (*b*) $\omega_2=78$ Hz, (*c*) $\omega_3=138$ Hz, and (*d*) $\omega_4=216$ Hz. Closed dots represent experimental measurements. Solid and dashed lines are the theoretical predictions including and excluding geometrical imperfections, respectively.

and α_n and neglect all other components in Eqs. (12) and (15). The three nontrivial equations of motion can be rewritten in the following form

$$\ddot{\alpha}_1+\left(1+\frac{h_1^2}{2}\right)\alpha_1=-2\mu\dot{\alpha}_1-\frac{1}{2}n^2h_1h_n\alpha_n-h_1e-\frac{3}{4}h_1\alpha_1^2-\frac{1}{4}n^2h_1\alpha_n^2-\frac{1}{2}n^2h_n\alpha_1\alpha_n-\alpha_1e-\frac{1}{4}\alpha_1^3-\frac{1}{4}n^2\alpha_1\alpha_n^2 \tag{17}$$

$$\ddot{\alpha}_n+n^4\left(1+\frac{h_n^2}{2}\right)\alpha_n=-2\mu\dot{\alpha}_n-\frac{1}{2}n^2h_1h_n\alpha_1-n^2h_ne-\frac{1}{4}n^2h_n\alpha_1^2-\frac{3}{4}n^4h_n\alpha_n^2-\frac{1}{2}n^2h_1\alpha_1\alpha_n-n^2\alpha_ne-\frac{1}{4}n^2\alpha_1^2\alpha_n-\frac{1}{4}n^4\alpha_n^3 \tag{18}$$

$$m\ddot{e}+(1+k)e=-\tfrac{1}{2}h_1\alpha_1-\tfrac{1}{2}n^2h_n\alpha_n-\tfrac{1}{4}\alpha_1^2-\tfrac{1}{4}n^2\alpha_n^2+2f\cos\gamma t \tag{19}$$

The numerical simulation based on Eqs. (17)–(19) is confirmed to agree well with the complete equations (12) and (15) when excitation frequency γ is close to ω_n. In the following we will use multiple scale method to study analytically the steady state amplitude and phase of α_n based on the simplified three-mode equations.

We first rescale Eqs. (17)–(19) by defining

$$(h_{n(n\geqslant2)},\mu,f,\alpha_1,\alpha_{n(n\geqslant2)},e)=\varepsilon(\hat{h}_{n(n\geqslant2)},\hat{\mu},\hat{f},\hat{\alpha}_1,\hat{\alpha}_{n(n\geqslant2)},\hat{e}) \tag{20}$$

ε is an artificial scale used to define the order of magnitude of various parameters. Equation (20) assumes that the three variables α_1, α_n, and e are of the same order of magnitude. After substituting relation (20) into Eqs. (17)–(19), the equations can be rewritten in the following form

$$\ddot{\hat{\alpha}}_1+\left(1+\frac{h_1^2}{2}\right)\hat{\alpha}_1+h_1\hat{e}=-\varepsilon\left(2\hat{\mu}\dot{\hat{\alpha}}_1+\frac{1}{2}n^2h_1\hat{h}_n\hat{\alpha}_n+\frac{3}{4}h_1\hat{\alpha}_1^2+\frac{1}{4}n^2h_1\hat{\alpha}_n^2+\hat{\alpha}_1\hat{e}\right)-\varepsilon^2\left(\frac{1}{2}n^2\hat{h}_n\hat{\alpha}_1\hat{\alpha}_n+\frac{1}{4}\hat{\alpha}_1^3+\frac{1}{4}n^2\hat{\alpha}_1\hat{\alpha}_n^2\right) \tag{21}$$

$$\ddot{\hat{\alpha}}_n+n^4\hat{\alpha}_n=-\varepsilon\left(2\hat{\mu}\dot{\hat{\alpha}}_n+\tfrac{1}{2}n^2h_1\hat{h}_n\hat{\alpha}_1+n^2\hat{h}_n\hat{e}+\tfrac{1}{2}n^2h_1\hat{\alpha}_1\hat{\alpha}_n+n^2\hat{\alpha}_n\hat{e}\right)-\varepsilon^2\left(\tfrac{1}{2}n^4\hat{h}_n^2\hat{\alpha}_n+\tfrac{1}{4}n^2\hat{h}_n\hat{\alpha}_1^2+\tfrac{3}{4}n^4\hat{h}_n\hat{\alpha}_n^2+\tfrac{1}{4}n^2\hat{\alpha}_1^2\hat{\alpha}_n+\tfrac{1}{4}n^4\hat{\alpha}_n^3\right) \tag{22}$$

$$m\ddot{\hat{e}}+(1+k)\hat{e}+\tfrac{1}{2}h_1\hat{\alpha}_1=-\varepsilon\left(\tfrac{1}{2}n^2\hat{h}_n\hat{\alpha}_n+\tfrac{1}{4}\hat{\alpha}_1^2+\tfrac{1}{4}n^2\hat{\alpha}_n^2\right)+2\hat{f}\cos\gamma t \tag{23}$$

We now assume the following expansions for $\hat{\alpha}_1$, $\hat{\alpha}_n$, and $\hat{e}$,

$$\hat{\alpha}_1=\alpha_{10}(T_0,T_1)+\varepsilon\alpha_{11}(T_0,T_1) \tag{24}$$

$$\hat{\alpha}_n=\alpha_{n0}(T_0,T_1)+\varepsilon\alpha_{n1}(T_0,T_1) \tag{25}$$

$$\hat{e}=e_0(T_0,T_1)+\varepsilon e_1(T_0,T_1) \tag{26}$$

where $T_n=\varepsilon^n t$. Substituting Eqs. (24)–(26) into Eqs. (21)–(23) and equating coefficients of like powers of ε yields ε^0

$$D_0^2\alpha_{10}+\left(1+\frac{h_1^2}{2}\right)\alpha_{10}+h_1e_0=0 \tag{27}$$

$$mD_0^2e_0+(1+k)e_0+\tfrac{1}{2}h_1\alpha_{10}=\hat{f}e^{i\gamma T_0}+\hat{f}e^{-i\gamma T_0} \tag{28}$$

$$D_0^2\alpha_{n0}+n^4\alpha_{n0}=0 \tag{29}$$

ε^1

$$D_0^2\alpha_{11}+\left(1+\frac{h_1^2}{2}\right)\alpha_{11}+h_1e_1$$

$$=-2D_0D_1\alpha_{10}-2\hat{\mu}D_0\alpha_{10}$$

$$-\frac{1}{2}n^2h_1\hat{h}_n\alpha_{n0}-\frac{3}{4}h_1\alpha_{10}^2-\frac{1}{4}n^2h_1\alpha_{n0}^2-\alpha_{10}e_0 \quad (30)$$

$$mD_0^2e_1+(1+k)e_1+\tfrac{1}{2}h_1\alpha_{11}=-2mD_0D_1e_0-\tfrac{1}{2}n^2\hat{h}_n\alpha_{n0}-\tfrac{1}{4}\alpha_{10}^2 -\tfrac{1}{4}n^2\alpha_{n0}^2 \quad (31)$$

$$D_0^2\alpha_{n1}+n^4\alpha_{n1}=-2D_0D_1\alpha_{n0}-2\hat{\mu}D_0\alpha_{n0}-\tfrac{1}{2}n^2h_1\hat{h}_n\alpha_{10} -\tfrac{1}{2}n^2h_1\alpha_{10}\alpha_{n0}-n^2\hat{h}_ne_0-n^2\alpha_{n0}e_0 \quad (32)$$

where $D_n\equiv\partial/\partial T_n$. It is noted that Eqs. (27) and (28) are coupled equations of α_{10} and e_0, while Eq. (29) is an independent homogeneous equation of α_{n0}. The general solutions of Eqs. (27)–(29) can be written as

$$\alpha_{10}=A_1(T_1)e^{i\eta_1T_0}+B_1(T_1)e^{i\eta_2T_0}+\Lambda_1e^{i\gamma T_0}+\mathrm{cc} \quad (33)$$

$$e_0=A_1(T_1)u_1e^{i\eta_1T_0}+B_1(T_1)u_2e^{i\eta_2T_0}+\Lambda_2e^{i\gamma T_0}+\mathrm{cc} \quad (34)$$

$$\alpha_{n0}=A_n(T_1)e^{in^2T_0}+\mathrm{cc} \quad (35)$$

where cc stands for the complex conjugates of the preceding terms. The parameters η_1 and η_2 are the two eigenvalues of Eqs. (27) and (28) (excluding the forcing term), with $(1,u_1)$ and $(1,u_2)$ being the corresponding eigenvectors. These parameters are calculated as

$$\eta_1=\pm\left(\frac{\Psi_1-\sqrt{\Psi_2+\Psi_1^2}}{4m}\right)^{1/2}, \quad \eta_2=\pm\left(\frac{\Psi_1+\sqrt{\Psi_2+\Psi_1^2}}{4m}\right)^{1/2} \quad (36)$$

$$u_1=\frac{\Psi_3-\sqrt{\Psi_2+\Psi_1^2}}{4h_1m}, \quad u_2=\frac{\Psi_3+\sqrt{\Psi_2+\Psi_1^2}}{4h_1m} \quad (37)$$

The constants Ψ_1, Ψ_2, and Ψ_3 are defined as

$$\Psi_1=2+2k+2m+h_1^2m, \quad \Psi_2=-8m(2+2k+h_1^2k),$$

$$\Psi_3=2+2k-2m-h_1^2m$$

The amplitudes Λ_1 and Λ_2 of the particular solutions are

$$\Lambda_1=\frac{-h_1}{\Phi(\gamma)}\hat{f}, \quad \Lambda_2=\frac{2+h_1^2-2\gamma^2}{2\Phi(\gamma)}\hat{f} \quad (38)$$

where

$$\Phi(\gamma)=\left(1+\frac{h_1^2}{2}-\gamma^2\right)(1+k-m\gamma^2)-\frac{h_1^2}{2} \quad (39)$$

After substituting Eqs. (33)–(35) into Eqs. (30) and (31), following a solvability analysis we can conclude that $A_1\to0$ and $B_1\to0$ as $T_1\to\infty$ [18]. Therefore, the terms containing A_1 and B_1 in Eqs. (33) and (34) can be dropped. We now substitute Eqs. (33)–(35) into Eq. (32) and conclude that the secular term of Eq. (32) can be eliminated if

$$2iD_1A_ne^{in^2T_0}+2i\hat{\mu}A_ne^{in^2T_0}+\tfrac{1}{2}h_1\hat{h}_n\Lambda_1e^{i\gamma T_0}+\hat{h}_n\Lambda_2e^{i\gamma T_0}=0 \quad (40)$$

We assume that the excitation frequency γ is close to $\omega_n=n^2$, and can be expressed as

$$\gamma=n^2+\Delta\gamma=n^2+\varepsilon\sigma \quad (41)$$

where $\Delta\gamma$ is a small deviation from $\omega_n=n^2$, and is assumed to be of the order ε. σ is a detuning parameter. After using the above relation, Eq. (40) can be rearranged into the form

$$2iD_1A_n+2i\hat{\mu}A_n+\left(\tfrac{1}{2}h_1\hat{h}_n\Lambda_1+\hat{h}_n\Lambda_2\right)e^{i\sigma T_1}=0 \quad (42)$$

To solve A_n from Eq. (42), we express the complex variable A_n as

$$A_n=\tfrac{1}{2}a_ne^{i\theta_n} \quad (43)$$

where a_n and θ_n are real variables. Substituting Eq. (43) into Eq. (42) and equating the real part and imaginary part of Eq. (42) to zero, we obtain

$$a_n'+\hat{\mu}a_n=-\left(\tfrac{1}{2}h_1\hat{h}_n\Lambda_1+\hat{h}_n\Lambda_2\right)\sin\beta_n \quad (44)$$

$$\theta_n'a_n=\left(\tfrac{1}{2}h_1\hat{h}_n\Lambda_1+\hat{h}_n\Lambda_2\right)\cos\beta_n \quad (45)$$

where $\beta_n=\sigma T_1-\theta_n$. The superposed prime denotes the differentiation with respect to T_1. By assuming the existence of the steady state solution, we can let $a_n'=0$ and $\beta_n'=0$. As a consequence, the amplitude and phase of the steady state solution A_n can be solved from the following two algebraic equations

$$\hat{\mu}a_n=\left(\tfrac{1}{2}h_1\hat{h}_n\Lambda_1+\hat{h}_n\Lambda_2\right)\sin\beta_n \quad (46)$$

$$\sigma a_n=\left(\tfrac{1}{2}h_1\hat{h}_n\Lambda_1+\hat{h}_n\Lambda_2\right)\cos\beta_n \quad (47)$$

From Eqs. (46) and (47) we obtain the amplitude a_n and phase β_n as

$$a_n=\frac{\hat{h}_n(1-\gamma^2)}{\Phi(\gamma)\sqrt{\hat{\mu}^2+\sigma^2}}\hat{f} \quad (48)$$

$$\beta_n=\tan^{-1}\left(\frac{-\hat{\mu}}{\sigma}\right) \quad (49)$$

The final expressions of the steady state response can then be expressed in the following forms

$$\alpha_1=\frac{-2h_1}{\Phi(\gamma)}f\cos\gamma t \quad (50)$$

$$\alpha_n=\frac{h_n(1-\gamma^2)}{\Phi(\gamma)\sqrt{\mu^2+(\Delta\gamma)^2}}f\cos(\gamma t-\beta_n) \quad (51)$$

$$e=\frac{2+h_1^2-2\gamma^2}{\Phi(\gamma)}f\cos\gamma t \quad (52)$$

Therefore, to the order of our analysis, the amplitude of the mode α_n is proportional to the imperfection h_n. For a perfect sinusoidal arch with $h_n=0$, the mode α_n will never be excited even though the excitation frequency γ is equal to the natural frequency $\omega_n=n^2$.

The closed-form solutions for the amplitude and phase of α_n predicted by multiple scale analysis can be verified by numerically integrating the complete equations of motion Eqs. (12) and (15). The solid lines in Fig. 6 show the variation of (*a*) amplitude and (*b*) phase of α_4 as functions of frequency deviation parameter $\Delta\gamma$ predicted from the multiple scale analysis. The closed dots represent the amplitude and phase predicted from numerically integrating Eqs. (12) and (15). All the parameters (h_1, h_4, and f) used in the calculation correspond to the experiment described in Fig. 5(*c*). The analytical predictions via multiple scale analysis agree with the numerical integration quite well.

The above analysis also reveals the nature of the coupling resonance. In the last section we suggest that the resulted vibration should not be called "primary resonance" even though the excitation frequency of the axial force is close to the *n*th natural frequency of the arch. Our analysis in this section shows that the axial force excites the modes α_1 and e in a "nonresonant" manner (see Eqs. (33) and (34)). These two modes then excite the α_n-mode internally. However, it does not seem right to call it internal resonance because internal resonance is usually referred

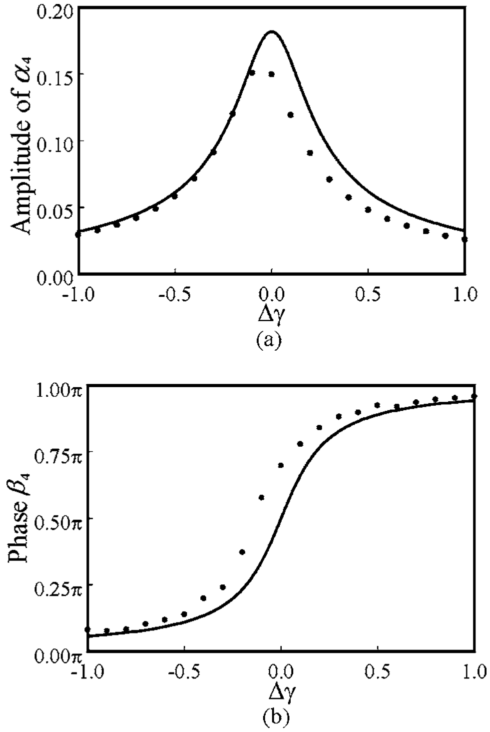


Fig. 6 (*a*) Amplitude and (*b*) phase of α_4 as functions of the frequency deviation $\Delta\gamma$. The solid lines are from multiple scale analysis. The closed dots are from numerically integrating the complete equations of motion.

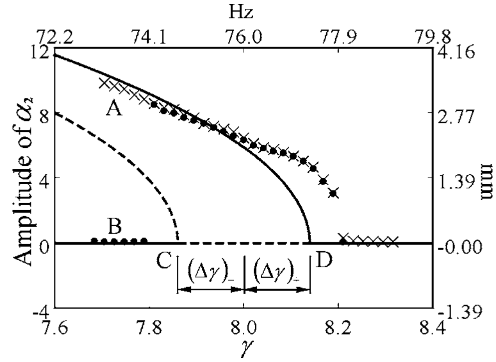


Fig. 7 Steady state amplitude of component α_2. "•" represents the measurement during the sweeping-down process. "×" represents the measurement during the sweeping-up process. Solid (stable) and dashed (unstable) lines are multiple scale predictions.

to the situation when the natural frequencies involved possess certain relations [11–13]. The key for the resonance discussed in this section to occur is the geometrical imperfection h_n, which couples the lateral vibration with the axial excitation. Therefore, we choose the term "coupling resonance." The indirectly excited mode α_n is found to be of the same order of magnitude of the directly excited modes α_1 and e.

8 Parametric Resonance

In this section we direct our attention to the case when the excitation frequency is close to $2\omega_2$ (76 Hz). The amplitude of the excitation force $2f$ in this experiment is fixed at 40.06 N. We record the steady state time history at two locations symmetric with respect to and 11 cm (a quarter of the total length) away from the midpoint simultaneously and subtract one from the other. By this procedure all the signals from the symmetric modes (α_1 and α_3 particularly) will be canceled out. Therefore, the resulted signal contains mainly the α_2-mode, whose amplitude is recorded in Fig. 7. We first sweep the excitation frequency from a little higher than 76 Hz to a little lower than 76 Hz. The amplitude of the resulted signal is denoted by the symbol "×." It is observed that in this "sweeping-down" process, the amplitude of α_2 changes from very small to a noticeable amount when the excitation frequency is at 78 Hz, slightly above $2\omega_2$. It is noted that 78 Hz happens to be the natural frequency ω_3 of the arch. Therefore, the overall vibration amplitude is large, as have been demonstrated in Fig. 5(*b*). However, the net signal after subtraction is still relatively small. The amplitude of α_2 continues to increase when the excitation frequency decreases. In another experiment we sweep the excitation frequency from 73 Hz upward. The measured amplitude is recorded by the symbol "•." In the lower end of this "sweeping-up" process the measured amplitude is very small (about 0.1 mm) until the excitation frequency reaches 74.2 Hz, at which the measured amplitude experiences an obvious jump to 2.77 mm. After the jump, the measured amplitude decreases gradually as the excitation frequency continues to sweep up, following more or less the locus recorded during the previous "sweeping-down" experiment. Obviously at frequency below 74.2 Hz, there exist two steady state solutions simultaneously.

To investigate these two steady state responses further, we measured the overall amplitude profiles at frequency 73.4 Hz, the points A and B in Fig. 7. We can easily switch the arch vibration between states A and B by disturbing the arch with hands. Figure 8 shows the amplitude profiles of these two steady states. Symbols "×" are for state A, while symbols "•" are for state B. In Fig. 9 we record the time history of these two responses at location $x = 11$ cm. We found that in state A the dominant frequency is 36.7 Hz, while in response B the dominant frequency appears to

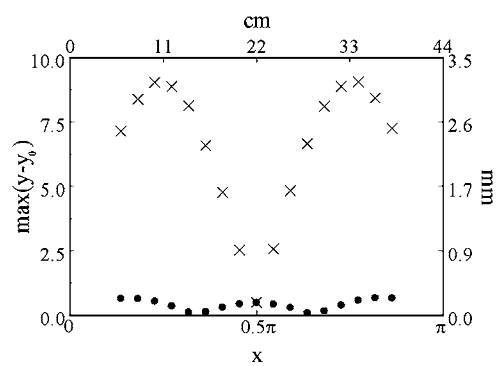


Fig. 8 Amplitude profiles corresponding to the two steady states *A* and *B* in Fig. 7

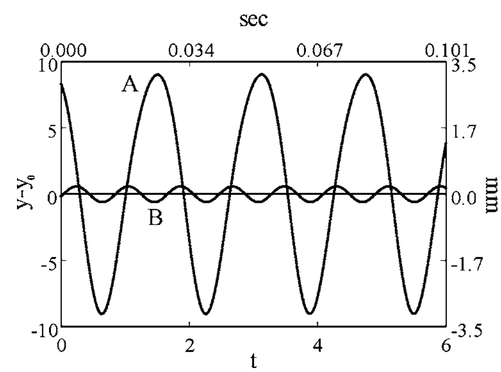


Fig. 9 Time history of the two states *A* and *B* in Fig. 7 measured at location 11 cm from the midpoint

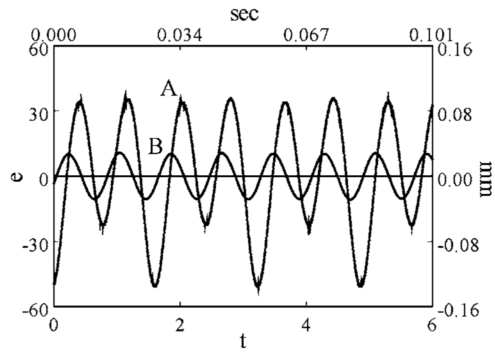


Fig. 10 Time history of the end motion of the two states *A* and *B* in Fig. 7

be two times of the state A. By inspecting Figs. 8 and 9 we can determine that the dominant component in the amplitude profile corresponding to state A is apparently $\sin 2x$. We can also identify the minor components in the response A as $\sin x$ and $\sin 3x$. To determine that the response contains the components $\sin x$ and $\sin 3x$, we record the time history at two points symmetrical with respect to the midpoint simultaneously and add them up (instead of subtracting one from another as described in Fig. 7). This adding process will eliminate the dominant component $\sin 2x$. By repeating this process for eight pairs of locations we obtain an amplitude profile that contains mostly the $\sin x$ and $\sin 3x$ modes. For the amplitude profile corresponding to state B the dominant component is $\sin 3x$. The reason why the mode $\sin 3x$ is present in both responses A and B is because the excitation frequency $2\omega_2$ happens to be close to ω_3.

It is noted that the vibration amplitude of steady state A is almost ten times larger than that of state B, as seen from Fig. 9. We also record the time history of the end motion for these two states in Fig. 10. It is noted that for state A with large lateral vibration the end motion is no longer harmonic. On the other hand, the end motion of state B is still quite close to harmonic.

9 Multiple Scale Analysis for Parametric Resonance

As discussed in Sec. 8, the reason why the mode $\sin 3x$ is present in the measured response is because the excitation frequency $2\omega_2$ happens to be close to ω_3. If we exclude this complicating factor, we may conclude that α_1 and α_2 will be two dominant components when the excitation frequency γ is close to $2\omega_2$. Further numerical calculation shows that in this case the geometrical imperfection has only minimal effect on the final steady state response. To analytically predict the steady state solutions and gain more physical insight into the jump phenomenon observed in the experiment, we employ again the multiple scale analysis on this case.

As in Sec. 7 we start with the three-mode equations (17)–(19), except that the imperfection h_n is set to be zero for simplicity. The numerical simulation based on the three-mode equations with $h_n=0$ is found to agree quite well with the complete Eqs. (12) and (15) when excitation frequency γ is close to $2\omega_n$. In this case we rescale the parameters in a somewhat different manner

$$\alpha_n = \varepsilon\hat{\alpha}_n \tag{53}$$

$$(\mu, f, \alpha_1, e) = \varepsilon^2(\hat{\mu}, \hat{f}, \hat{\alpha}_1, \hat{e}) \tag{54}$$

In other words, we assume that α_n is one order of magnitude larger than those parameters listed in Eq. (54). This assumption is based mainly on the experimental observation. After substituting Eqs. (53) and (54) into Eqs. (17)–(19) the equations can be rewritten in the following form

$$\ddot{\hat{\alpha}}_1 + \left(1+\frac{h_1^2}{2}\right)\hat{\alpha}_1 + h_1\hat{e} = -\frac{1}{4}n^2h_1\hat{\alpha}_n^2 - \varepsilon^2\left(2\hat{\mu}\dot{\hat{\alpha}}_1 + \hat{\alpha}_1\hat{e} + \frac{3}{4}h_1\hat{\alpha}_1^2\right) \tag{55}$$

$$\ddot{\hat{\alpha}}_n + n^4\hat{\alpha}_n = -\varepsilon^2\left(2\hat{\mu}\dot{\hat{\alpha}}_n + \tfrac{1}{2}n^2h_1\hat{\alpha}_1\hat{\alpha}_n + n^2\hat{\alpha}_n\hat{e} + \tfrac{1}{4}n^4\hat{\alpha}_n^3\right) \tag{56}$$

$$m\ddot{\hat{e}} + (1+k)\hat{e} + \tfrac{1}{2}h_1\hat{\alpha}_1 = 2\hat{f}\cos\gamma t - \tfrac{1}{4}n^2\hat{\alpha}_n^2 - \varepsilon^2\left(\tfrac{1}{4}\hat{\alpha}_1^2\right) \tag{57}$$

We now assume the following expansions for $\hat{\alpha}_1$, $\hat{\alpha}_n$, and $\hat{e}$

$$\hat{\alpha}_1 = \alpha_{10}(T_0,T_2) + \varepsilon^2\alpha_{12}(T_0,T_2) \tag{58}$$

$$\hat{\alpha}_n = \alpha_{n0}(T_0,T_2) + \varepsilon^2\alpha_{n2}(T_0,T_2) \tag{59}$$

$$\hat{e} = e_0(T_0,T_2) + \varepsilon^2 e_2(T_0,T_2) \tag{60}$$

Following a similar procedure as described in Sec. 7, and equating coefficients of like powers of ε yields ε^1

$$D_0^2\alpha_{n0} + n^4\alpha_{n0} = 0 \tag{61}$$

ε^2

$$D_0^2\alpha_{10} + \left(1+\frac{h_1^2}{2}\right)\alpha_{10} + h_1e_0 = -\frac{n^2}{4}h_1\alpha_{n0}^2 \tag{62}$$

$$mD_0^2e_0 + (1+k)e_0 + \frac{1}{2}h_1\alpha_{10} = \hat{f}e^{i\gamma T_0} + \hat{f}e^{-i\gamma T_0} - \frac{n^2}{4}\alpha_{n0}^2 \tag{63}$$

ε^3

$$D_0^2\alpha_{n2} + n^4\alpha_{n2} = -2D_0D_2\alpha_{n0} - 2\hat{\mu}D_0\alpha_{n0} - n^2\alpha_{n0}e_0 - \frac{n^2}{2}h_1\alpha_{10}\alpha_{n0} - \frac{n^4}{4}\alpha_{n0}^3 \tag{64}$$

The general solution of the homogeneous equation (61) can be written as

$$\alpha_{n0} = A_n(T_2)e^{in^2T_0} + \text{cc} \tag{65}$$

After substituting Eq. (65) into Eqs. (62) and (63) we can solve for α_{10} and e_0 as

$$\alpha_{10} = A_1e^{i\eta_1T_0} + B_1e^{i\eta_2T_0} + \Lambda_{11}e^{i\gamma T_0} + \Lambda_{12}e^{2in^2T_0} + \Lambda_{13} + \text{cc} \tag{66}$$

$$e_0 = u_1A_1e^{i\eta_1T_0} + u_2B_1e^{i\eta_2T_0} + \Lambda_{21}e^{i\gamma T_0} + \Lambda_{22}e^{2in^2T_0} + \Lambda_{23} + \text{cc} \tag{67}$$

The parameters η_1, η_2, u_1, and u_2 have been defined in Eqs. (36) and (37). Λ_{ij} are defined as

$$\Lambda_{11} = -\frac{h_1\hat{f}}{\Phi(\gamma)}, \quad \Lambda_{21} = \frac{(2+h_1^2-2\gamma^2)\hat{f}}{2\Phi(\gamma)}$$

$$\Lambda_{12} = -\frac{n^2h_1(k-4mn^4)}{4\Phi(2n^2)}A_n^2, \quad \Lambda_{22} = -\frac{n^2(1-4n^4)}{4\Phi(2n^2)}A_n^2$$

$$\Lambda_{13} = -\frac{n^2h_1k}{4\Phi(0)}A_n\bar{A}_n, \quad \Lambda_{23} = -\frac{n^2}{4\Phi(0)}A_n\bar{A}_n$$

where the function $\Phi(\cdot)$ has been defined in Eq. (39). We assume that the excitation frequency is close to $2\omega_n$, i.e.

$$\gamma = 2n^2 + \Delta\gamma = 2n^2 + \varepsilon^2\sigma \tag{68}$$

Substituting Eqs. (65)–(68) into Eq. (64), the solvability condition for α_{n2} can be obtained as

$$A_n' + C_1A_n + iC_2\bar{A}_ne^{i\sigma T_2} + iC_3A_n^2\bar{A}_n = 0 \tag{69}$$

where the coefficients C_i are

$$C_1 = \hat{\mu}, \quad C_2 = \frac{(\gamma^2-1)}{2\Phi(\gamma)}\hat{f},$$

$$C_3 = \frac{n^2}{16}\left(\frac{h_1^2(k-4mn^4)+2(1-4n^4)}{\Phi(2n^2)} + \frac{2kh_1^2+4}{\Phi(0)} - 6\right)$$

The superposed prime in Eq. (69) represents derivative with respect to T_2. To solve the homogeneous equation (69) we express A_n in the same form as in Eq. (43). After substituting Eq. (43) into Eq. (69) we obtain the two equations for the amplitude a_n and θ_n

$$a_n' + C_1 a_n - C_2 a_n \sin\beta_n = 0 \tag{70}$$

$$a_n(4\theta_n' + 4C_2\cos\beta_n + C_3 a_n^2) = 0 \tag{71}$$

where $\beta_n = \sigma T_2 - 2\theta_n$. The amplitude and phase of the steady state solutions can then be solved from

$$a_n(C_1 - C_2\sin\beta_n) = 0 \tag{72}$$

$$a_n(2\sigma + 4C_2\cos\beta_n + C_3 a_n^2) = 0 \tag{73}$$

Two types of solutions can be found from Eqs. (72) and (73). The first is the trivial solution $a_n = 0$. In this case the mode α_n will not be excited. The second solution is nontrivial, where

$$\beta_n = \sin^{-1}\left(\frac{C_1}{C_2}\right) \tag{74}$$

$$a_n = \left(\frac{-2\sigma \pm 4\sqrt{C_2^2 - C_1^2}}{C_3}\right)^{1/2} \tag{75}$$

To determine the stability of the above trivial and nontrivial steady state solutions, we express A_n in Eq. (69) as the superposition of the steady state solution A_{ns} and a small disturbance $\tilde{A}_n$

$$A_n = A_{ns} + \tilde{A}_n \tag{76}$$

After substituting Eq. (76) into Eq. (69) and linearizing the equation with respect to $\tilde{A}_n$, we obtain the linear equation for $\tilde{A}_n$

$$\tilde{A}_n' + C_1\tilde{A}_n + iC_2\bar{\tilde{A}}_n e^{i\sigma T_2} + iC_3(2A_{ns}\bar{A}_{ns}\tilde{A}_n + A_{ns}^2\bar{\tilde{A}}_n) = 0 \tag{77}$$

A_{ns} in Eq. (77) is the trivial and nontrivial solutions solved previously. After expressing $\tilde{A}_n$ in the form $\tilde{A}_n = (a_{nR} + i a_{nI})e^{(1/2)i\sigma T_2}$, Eq. (77) can be rewritten as two real equations

$$a_{nR}' + \left(C_1 + \tfrac{1}{4}C_3 a_n^2\sin\beta_n\right)a_{nR} + \left(C_2 - \tfrac{1}{2}\sigma - \tfrac{1}{4}C_3 a_n^2(2-\cos\beta_n)\right)a_{nI} = 0 \tag{78}$$

$$\left(C_2 + \frac{1}{2}\sigma + \frac{1}{4}C_3 a_n^2(2+\cos\beta_n)\right)a_{nR} + a_{nI}' + \left(C_1 - \frac{1}{4}C_3 a_n^2\sin\beta_n\right)a_{nI} = 0 \tag{79}$$

a_n and phase β_n in Eqs. (78) and (79) are the amplitude and phase expressed in Eqs. (74) and (75). By solving the eigenvalues of Eqs. (78) and (79) we can determine the stability of the steady state solutions. The final expressions of the steady state response of α_n are expressed in the following form

$$\alpha_n = \left(\frac{-2\Delta\gamma \pm 4\sqrt{\left(\frac{(\gamma^2-1)}{2\Phi(\gamma)}f\right)^2 - \mu^2}}{C_3}\right)^{1/2}\cos\left(\frac{\gamma - \beta_n}{2}t\right) \tag{80}$$

After substituting the physical parameters corresponding to the experiment shown in Fig. 7, we can plot the calculated amplitude of the steady state solution α_2 as solid (stable) and dashed (unstable) lines. Two bifurcation points can be observed in Fig. 7. Point C is a subcritical pitch-fork bifurcation point, while D is a supercritical pitch-fork bifurcation point. The two bifurcation frequencies are $2n^2 + (\Delta\gamma)_+$ and $2n^2 + (\Delta\gamma)_-$, (recall that $n = 2$ in this case) where frequency deviations $(\Delta\gamma)_+$ and $(\Delta\gamma)_-$ can be calculated as

$$(\Delta\gamma)_\pm = \pm 2\sqrt{\left[\frac{(\gamma^2-1)}{2\Phi(\gamma)}f\right]^2 - \mu^2} \tag{81}$$

In evaluating $(\Delta\gamma)_\pm$ from Eq. (81) we can replace γ by $2n^2$ for simplicity. The resulting error in doing so is less than 0.5%. It is noted that fairly good agreement in amplitudes between the experimental results and analytical prediction can be observed. Although the bifurcation points are off somewhat, the multiple scale analysis predicts the jump phenomenon quite well.

10 Conclusions

In this paper we study the nonlinear vibration of a shallow arch with one end attached to an electro-mechanical shaker. The arch is excited in the axial direction. In the experiment we can control the excitation force on the end mass but not the end motion of the arch. Attention is focused on two cases, i.e., the coupling resonance and parametric resonance. In the case of coupling resonance the excitation frequency is close to the natural frequency of the nth mode. In the parametric resonance case, the excitation frequency is two times the nth natural frequency. Experimental, numerical, and analytical methods are used to look into the physical insight of these complicated nonlinear oscillation phenomena. Several conclusions can be summarized in the following.

(1) Geometrical imperfection is the key for the coupling resonance to occur. For a perfect sinusoidal arch, the nth mode will not be excited even when the excitation frequency is close to its corresponding natural frequency. When geometrical imperfection exists, which is almost inevitable in the real world, the first mode and the axial mode will be excited directly by the excitation in a "nonresonant" manner, and then these modes in turn excite the nth mode internally.

(2) In the case of parametric resonance, two stable steady state solutions can exist simultaneously when the excitation frequency is slightly lower than two times the nth natural frequency. As a consequence jump phenomenon is observed when the excitation frequency sweeps upward. The effect of geometrical imperfection on the steady state response is minimal in this case. While only the stable steady states can be realized in the experiment, the multiple scale method can predict both the stable and unstable solutions.

Acknowledgment

The results presented here were obtained in the course of research supported by a grant from the National Science Council of the Republic of China.

References

[1] Timoshenko, S. P., 1935, "Buckling of Flat Curved Bars and Slightly Curved Plates," ASME J. Appl. Mech., **2**, pp. 17–20.
[2] Hoff, N. J., and Bruce, V. G., 1954, "Dynamic Analysis of the Buckling of Laterally Loaded Flat Arches," J. Math. Phys., **32**, pp. 276–288.
[3] Simitses, G. J., 1986, *Elastic Stability of Structures*, Krieger, Malabar, FL, Chap. 7.
[4] Simitses, G. J., 1990, *Dynamic Stability of Suddenly Loaded Structures*, Springer, New York.
[5] Chen, J. S., and Liao, C. Y., 2005, "Experiment and Analysis on the Free Dynamics of a Shallow Arch After an Impact Load at the End," ASME J. Appl. Mech., **72**, pp. 54–61.
[6] Huang, N. C., 1972, "Dynamic Buckling of Some Elastic Shallow Structures Subject to Periodic Loading with High Frequency," Int. J. Solids Struct., **8**, pp. 315–326.
[7] Plaut, R. H., and Hsieh, J. C., 1985, "Oscillations and Instability of a Shallow Arch Under Two-Frequency Excitation," J. Sound Vib., **102**, pp. 189–201.
[8] Blair, K. B., Krousgrill, C. M., and Farris, T. N., 1996, "Non-Linear Dynamic Response of Shallow Arches to Harmonic Forcing," J. Sound Vib., **194**, pp.

353–367.
[9] Thomsen, J. J., 1992, "Chaotic Vibrations of Non-Shallow Arches," J. Sound Vib., **153**, pp. 239–258.
[10] Bolotin, V. V., 1964, *The Dynamic Stability of Elastic Systems*, Holden-Day, San Francisco.
[11] Tien, W. M., Sri Namachchivaya, N., and Bajaj, A. K., 1994, "Non-Linear Dynamics of a Shallow Arch under Periodic Excitation—I. 1:1 Internal Resonance," Int. J. Non-Linear Mech., **29**, pp. 367–386.
[12] Tien, W. M., Sri Namachchivaya, N., and Bajaj, A. K., 1994, "Non-Linear Dynamics of a Shallow Arch under Periodic Excitation—II. 1:2 Internal Resonance," Int. J. Non-Linear Mech., **28**, pp. 349–366.
[13] Bi, Q., and Dai, H. H., 2000, "Analysis of Nonlinear Dynamics and Bifurcations of a Shallow Arch Subjected to Periodic Excitation With Internal Resonance," J. Sound Vib., **233**, pp. 557–571.
[14] Malhotra, N., and Sri Namachchivaya, N., 1997, "Chaotic Dynamics of Shallow Arch Structures Under 1:1 Resonance," J. Eng. Mech., **123**, pp. 620–627.
[15] Malhotra, N., and Sri Namachchivaya, N., 1997, "Chaotic Dynamics of Shallow Arch Structures under 1:2 Resonance," J. Eng. Mech., **123**, pp. 612–619.
[16] Mettler, E., 1962, "Dynamic Buckling," in *Handbook of Engineering Mechanics*, W. Flugge, ed., McGraw Hill, New York, Chap. 62.
[17] Rao, S. S., 1995, *Mechanical Vibrations*, 3rd ed., Addision-Wesley, Reading, MA.
[18] Yang, C.-H., 2005, "Nonlinear Vibration of a Shallow Arch Under Periodic Excitation at One End," MS thesis, Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan.

M. Belubekyan
Professor

K. Ghazaryan
Professor

Institute of Mechanics,
National Academy of Sciences of Armenia,
Yerevan, Armenia

P. Marzocca
Assistant Professor
e-mail: pmarzocc@clarkson.edu

C. Cormier
Student

Mechanical and Aeronautical Engineering Department,
Clarkson University,
Potsdam, NY 13699-5725

Localized Magnetoelastic Bending Vibration of an Electroconductive Elastic Plate

The study of the magnetoelastic vibrations of a flat plate immersed in a uniform applied external magnetic field is presented. Kirchhoff's plate theory and the model of a perfect conductive medium are used. The conditions for the existence of localized bending vibrations in the vicinity of the free edge of the plate are established. It is shown that the localized vibrations can be detected and eventually can be eliminated by means of an applied magnetic field. [DOI: 10.1115/1.2424469]

1 Introduction

Health monitoring of modern structures poses new challenges with regard to increased safety and operational reliability. Different schemes or algorithms for damage detection, remote monitoring, and continuous real time evaluation are increasingly needed. Of particular importance is the detection of material defects using nondestructive techniques, such as ultrasonic void and crack detection. For a comprehensive literature review on damage identification and health monitoring, the reader is referred to Ref. [1] and the references therein.

Presented in this paper is a theoretical contribution toward the development of a novel structural health monitoring technique (SHMT) based on vibration characteristics. Of critical importance to this form of structural health monitoring is the fact that the presence of defects, cracks, and dynamic changes in elastic, electrically conductive, structures may be identified through measurement of an electromagnetic field at a known distance from the elastic body. This is a comparative method, meaning a baseline measurement must be acquired prior to evaluation of the target structure. Other comparative techniques include piezoelectric sensors and ultrasonic imaging. However, the magnetoelastic technique presented in this paper differs from other nondestructive (ND) SHMT's.

Using an external magnetic field to create induced magnetic fields around defects in vibrating, electrically conductive materials allows detection to occur on a global scale. Since the magnetic field penetrates the entire structure, any defects may be observed by measurement of the magnetic field produced by vibrating the structure, referred to as vibration induced magnetic response (VIMR). Piezoelectric devices only measure defects within the range of their installation locale. Also, whereas the use of piezoelectric sensors requires installation within the matrix of the structure, potentially initiating defects in construction, VIMR can be measured passively and does not require sensors to be installed within the matrix. Also, unlike ultrasonic techniques, the magnetoelastic technique may be used remotely.

The physical model of VMIR is easily described by the difference between vibration frequencies close to and far from defects. These localized bending waves are usually located where cracks and defects are located and they produce a unique magnetic response in the presence of an externally applied magnetic field since the frequency of the localized waves is usually much lower than other bending waves. Thus, by measuring the electromagnetic field intensity of electro-active elastic structures, it is theoretically possible to passively identify the location of cracks or defects.

Localized bending plate vibrations were first investigated by Konenkov [2]. Subsequently, the model presented in Ref. [2] was widely developed in Refs. [3–8] and then in Refs. [9,10] extended to the study of vibrations in a simply modeled, perfectly electroconductive plate. Considering the hypothesis of magnetoelasticity of thin bodies [11–15], the spatial problem of magnetoelastic vibration can be reduced to two dimensions. Application of simple models depends on the direction of the external magnetic field (longitudinal or transverse) as well as the character of the considered problem (planar or transverse vibrations) [16,17]. A review of investigations in the field of electromagneto mechanics of thin plates and shells is given in Ref. [13].

A number of basic linear problems of nonlocalized vibrations of conductive plates and shells of various configurations were solved in the case of arbitrary orientation of an external magnetic field [11–17]. In addition, in the framework of the main assumptions of the magnetoelasticity hypothesis, the nonlocalized nonlinear magnetoelastic vibrations of electroconductive plates are studied in Refs. [18,19]. The basic field equations and boundary conditions necessary for the dynamic approach of electromagnetically conducting flat plates subjected to an external magnetic field are derived in Ref. [18], while investigation of the interacting effects among the magnetic, thermal, and elastic fields in orthotropic thin plates has been presented in Ref. [19].

In Sec. 2 the mathematical model will be presented while in Sec. 3 two typical cases will be illustrated Sec. 3.1 presents the application of a magnetic field perpendicular to the axis along which the wave propagates. Sec. 3.2 presents the application of a magnetic field parallel to the axis along which the wave propagates. In Sec. 4, pertinent conclusions will be illustrated.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 20, 2005; final manuscript received May 19, 2006. Review conducted by Sridhar Krishnaswamy.

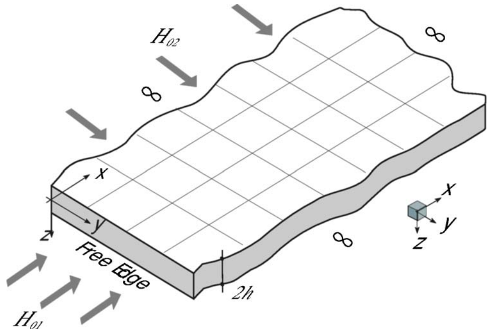


Fig. 1 Model of a plate immersed in a magnetic field in *x* and *y* directions

2 Mathematical Model

To illustrate the proposed method we will consider the case of an elastic electroconductive plate in the Cartesian frame $x_1=x$, $x_2=y$, $x_3=z$. The plate is immersed in the external longitudinal magnetic field parallel to the (x,y) plane, as shown in Fig. 1. That is $\vec{H}_0=H_{01}\vec{i}_x+H_{02}\vec{i}_y$, H_{01}=const, and H_{02}=const. The three-dimensional equations of the plate vibrations can be written as [12,20]

$$\frac{\partial \sigma_{ik}}{\partial x_k}+F_i=\rho\frac{\partial^2 u_i}{\partial t^2}, \quad (i=1,2,3) \tag{1}$$

where u_i, F_i, σ_{ij} are the components of the elastic plate displacement vector $\vec{u}$, electromagnetic bulk force vector $\vec{F}$, and elastic stress tensor $\hat{\sigma}$, respectively; and ρ is the density of the plate material. In Eq. (1) and thereafter, the rule of summation, (indices k) are assumed with respect to dummy indices. Considering the plate's perturbed state, the vector of electromagnetic bulk force is represented by

$$\vec{F}=\frac{1}{c}(\vec{j}\times\vec{H}_0) \tag{2}$$

Here $\vec{j}$ is the density vector of the induced electric current; and c is the electrodynamic constant equal to the velocity of light in a vacuum. The Gaussian system of electromagnetic units is used. By means of the Maxwell electromagnetic tensor $\hat{T}$, F_i can be written as [12,21]

$$F_i=\frac{\partial T_{ik}}{\partial x_k} \tag{3}$$

where T_{ik} are the components of Maxwell electromagnetic tensor $\hat{T}$. The components of the linearized Maxwell electromagnetic tensor can expressed as

$$T_{ik}=\frac{1}{4\pi}\left(H_{0k}h_i+H_{0i}h_k-\frac{\delta_{ik}}{2}H_{0s}h_s\right) \tag{4}$$

where the indices s indicate dummy indices.

The current–density vector $\vec{j}$ of the perturbed plate can be expressed as

$$\vec{j}=\sigma\left(\vec{e}+\frac{1}{c}\frac{\partial\vec{u}}{\partial t}\times\vec{H}_0\right) \tag{5}$$

In Eqs. (4) and (5), $\vec{e}$ is the induced electrical field vector; $\vec{h}$ is the induced magnetic field vector; h_i are the components of the vector $\vec{h}$; while σ is the electro-conductivity of plate material, and δ_{ik} is Kronecer's delta symbols.

In the plate body the perturbed electrical and magnetic fields (vectors $\vec{e}$ and $\vec{h}$) satisfy the Maxwell electrodynamics quasi-stationary equations for deformable bodies [12,21]

$$\operatorname{rot}\vec{e}+\frac{1}{c}\frac{\partial\vec{h}}{\partial t}=0 \tag{6a}$$

$$\operatorname{rot}\vec{h}=\frac{4\pi}{c}\vec{j} \tag{6b}$$

$$\operatorname{div}\vec{h}=0 \tag{6c}$$

$$\operatorname{div}\vec{e}=4\pi\rho_e \tag{6d}$$

where ρ_e is the density of electrical charges. On the other hand, in the vacuum outside of the plate, the following Maxwell electrodynamics equations with respect to the perturbed electrical and magnetic fields (vectors $\vec{e}_{(e)}$ and $\vec{h}_{(e)}$) are satisfied

$$\operatorname{rot}\vec{e}_{(e)}+\frac{1}{c}\frac{\partial\vec{h}_{(e)}}{\partial t}=0 \tag{7a}$$

$$\operatorname{rot}\vec{h}_{(e)}-\frac{1}{c}\frac{\partial\vec{h}_{(e)}}{\partial t}=0 \tag{7b}$$

$$\operatorname{div}\vec{h}_{(e)}=0 \tag{7c}$$

$$\operatorname{div}\vec{e}_{(e)}=0 \tag{7d}$$

where "e" identifies quantities associated to the outer plate domain (i.e., of the vacuum). On the plate surface Ω the following boundary conditions have to be fulfilled

$$(\sigma_{ik}+T_{ik}-T_{ik}^{(e)})n_k=0 \tag{8a}$$

Herein $T_{ik}^{(e)}$ are the components of the corresponding Maxwell electromagnetic tensor $\hat{T}^{(e)}$ of the outside media; and n_k are the components of the outward normal vector $\vec{n}$ at the plate surface Ω. The vectors $\vec{e}$, $\vec{h}$, $\vec{e}_{(e)}$, $\vec{h}_{(e)}$ are coupled through the boundary conditions at Ω

$$(\vec{h}-\vec{h}_{(e)})\cdot\vec{n}=0, \quad (\vec{e}-\vec{e}_{(e)})\times\vec{n}=0 \tag{8b}$$

The equations and boundary conditions are three dimensional. Based on the analysis of the exact solutions of the specific problems at hand and the asymptotic behavior of the general solutions of the three-dimensional problem of magnetoelasticity of thin bodies [12], the hypothesis of the magnetoelasticity of electroconductive thin bodies can be formulated. Besides assumptions of Kirchhoff's plate theory [12,13], the main assumptions of this hypothesis consist of tangential components of induced electrical field vector and normal component of induced magnetic field vector in the body of elastic plate which remain unchanged along the plate thickness. This hypothesis enables one to reduce the three-dimensional coupled equations to two-dimensional equations. Considering the problem of the plate free vibration in a longitudinal magnetic field, it has been found that the assumptions of the hypothesis of the magnetoelasticity of electroconductive thin bodies and the model of perfectly conducting media [12] achieve the same results with sufficient accuracy.

According to the model of perfect conducting medium ($\sigma\to\infty$) the perturbed electrical field $\vec{e}$ from Eq. (5) can be recast as [10]

$$\vec{e}=-\frac{1}{c}\frac{\partial\vec{u}}{\partial t}\times\vec{H}_0 \tag{9a}$$

Based on Eqs. (6), the perturbed magnetic field $\vec{h}$ and the induced electric current $\vec{j}$ are defined as follows

$$\vec{h}=\operatorname{rot}(\vec{u}\times\vec{H}_0) \tag{9b}$$

$$\vec{j} = \frac{c}{4\pi}\text{rot}\,\vec{h} \tag{9c}$$

According to the Kirchhoff's plate theory [12,13] the following assumptions for displacements and main stresses are taken into account

$$u_1(x,y,z) = u(x,y) - z\frac{\partial w}{\partial x} \tag{10a}$$

$$u_2(x,y,z) = v(x,y) - z\frac{\partial w}{\partial y} \tag{10b}$$

$$u_3 = w(x,y) \tag{10c}$$

$$\sigma_{11} = \frac{E}{1-v^2}\left[\frac{\partial u}{\partial x} + \nu\frac{\partial v}{\partial y} - z\left(\frac{\partial^2 w}{\partial x^2} + \nu\frac{\partial^2 w}{\partial y^2}\right)\right] \tag{11a}$$

$$\sigma_{22} = \frac{E}{1-v^2}\left[\frac{\partial v}{\partial y} + \nu\frac{\partial u}{\partial x} - z\left(\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2}\right)\right] \tag{11b}$$

$$\sigma_{12} = \frac{E}{2(1+v)}\left[\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) - 2z\frac{\partial^2 w}{\partial x\partial y}\right] \tag{11c}$$

Here $u(x,y)$, $v(x,y)$ are tangential displacements; $w(x,y)$ is the normal displacement of the plate's middle plane; E is the modulus of elasticity; and ν is Poisson's ratio.

Based on Eqs. (9b), the components of the magnetic field vector can be defined as

$$h_x = H_{02}\frac{\partial u}{\partial y} - H_{01}\frac{\partial v}{\partial y} - z\left(H_{02}\frac{\partial^2 w}{\partial x\partial y} - H_{01}\frac{\partial^2 w}{\partial y^2}\right) \tag{12a}$$

$$h_y = H_{01}\frac{\partial v}{\partial x} - H_{02}\frac{\partial u}{\partial x} - z\left(H_{01}\frac{\partial^2 w}{\partial x\partial y} - H_{02}\frac{\partial^2 w}{\partial x^2}\right) \tag{12b}$$

$$h_z = H_{01}\frac{\partial w}{\partial x} + H_{02}\frac{\partial w}{\partial y} \tag{12c}$$

The components of the Maxwell tensor can be written as

$$T_{xx} = \frac{1}{4\pi}(H_{01}h_x - H_{02}h_y) = \frac{1}{4\pi}\left[z\left(H_{01}^2\frac{\partial^2 w}{\partial y^2} - H_{02}^2\frac{\partial^2 w}{\partial x^2}\right) + H_{02}^2\frac{\partial u}{\partial x} - H_{01}^2\frac{\partial v}{\partial y} + H_{01}H_{02}\left(\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\right)\right] \tag{13a}$$

$$T_{yy} = -T_{xx} \tag{13b}$$

$$T_{xy} = \frac{1}{4\pi}(H_{01}h_y + H_{02}h_x) = \frac{1}{4\pi}z_{01}\left[H_{01}H_{02}\Delta w - (H_{01}^2 + H_{02}^2)\frac{\partial^2 w}{\partial x\partial y}\right] + \frac{1}{4\pi}\left[H_{02}^2\frac{\partial u}{\partial y} + H_{01}^2\frac{\partial v}{\partial x} - H_{01}H_{02}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\right] \tag{13c}$$

$$T_{xz} = \frac{1}{4\pi}H_{01}h_z = \frac{1}{4\pi}\left(H_{01}^2\frac{\partial w}{\partial x} + H_{01}H_{02}\frac{\partial w}{\partial y}\right) \tag{13d}$$

$$T_{yz} = \frac{1}{4\pi}H_{02}h_z = \frac{1}{4\pi}\left(H_{02}^2\frac{\partial w}{\partial y} + H_{01}H_{02}\frac{\partial w}{\partial x}\right) \tag{13e}$$

$$T_{zz} = -\frac{1}{4\pi}(H_{01}h_x + H_{02}h_y) = \frac{1}{4\pi}z\left[2H_{01}H_{02}\frac{\partial^2 w}{\partial x\partial y} - \left(H_{01}^2\frac{\partial^2 w}{\partial y^2} + H_{02}^2\frac{\partial^2 w}{\partial x^2}\right)\right] + \frac{1}{4\pi}\left[H_{02}^2\frac{\partial u}{\partial x} + H_{01}^2\frac{\partial v}{\partial y} - H_{01}H_{02}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right] \tag{13f}$$

Averaging the three-dimensional equations along coordinate z, in the plate thickness direction, the two-dimensional equations of plate vibration can be cast as

$$\frac{\partial Q_1}{\partial x} + \frac{\partial S}{\partial y} + \int_{-h}^{h}\left(\frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{xy}}{\partial y}\right)dz + \sigma_{xz}(h) - \sigma_{xz}(-h) + T_{xz}(h) - T_{xz}(-h) = 2\rho h\frac{\partial^2 u}{\partial t^2} \tag{14a}$$

$$\frac{\partial Q_2}{\partial y} + \frac{\partial S}{\partial x} + \int_{-h}^{h}\left(\frac{\partial T_{xy}}{\partial y} + \frac{\partial T_{xy}}{\partial x}\right)dz + \sigma_{yz}(h) - \sigma_{yz}(-h) + T_{yz}(h) - T_{yz}(-h) = 2\rho h\frac{\partial^2 v}{\partial t^2} \tag{14b}$$

$$\frac{\partial N_1}{\partial x} + \frac{\partial N_2}{\partial y} + \int_{-h}^{h}\left(\frac{\partial T_{xz}}{\partial x} + \frac{\partial T_{yz}}{\partial y}\right)dz + \sigma_{zz}(h) - \sigma_{zz}(-h) + T_{zz}(h) - T_{zz}(-h) = 2\rho h\frac{\partial^2 w}{\partial t^2} \tag{14c}$$

$$\frac{\partial M_1}{\partial x} + \frac{\partial M_{12}}{\partial y} + \int_{-h}^{h}z\left(\frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{xy}}{\partial y} + \frac{\partial T_{xz}}{\partial z}\right)dz + h[\sigma_{xz}(h) + \sigma_{xz}(-h)] = N_1 \tag{14d}$$

$$\frac{\partial M_{12}}{\partial x} + \frac{\partial M_2}{\partial y} + \int_{-h}^{h}z\left(\frac{\partial T_{yy}}{\partial y} + \frac{\partial T_{xy}}{\partial x} + \frac{\partial T_{yz}}{\partial z}\right)dz + h[\sigma_{yz}(h) + \sigma_{yz}(-h)] = N_2 \tag{14e}$$

Here

$$Q_1 = \int_{-h}^{+}\sigma_{xx}\,dz = \frac{2Eh}{1-v^2}\left(\frac{\partial u}{\partial x} + \nu\frac{\partial v}{\partial y}\right) \tag{15a}$$

$$Q_2 = \int_{-h}^{+}\sigma_{yy}\,dz = \frac{2hE}{1-v^2}\left(\frac{\partial v}{\partial y} + \nu\frac{\partial u}{\partial x}\right) \tag{15b}$$

$$S = \int_{-h}^{+}\sigma_{xy}\,dz = \frac{Eh}{(1+v)}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \tag{15c}$$

$$M_1 = \int_{-h}^{h}z\sigma_{xx}\,dz = -D\left(\frac{\partial^2 w}{\partial x^2} + \nu\frac{\partial^2 w}{\partial y^2}\right) \tag{15d}$$

$$M_2 = \int_{-h}^{h}z\sigma_{yy}\,dz = -D\left(\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2}\right) \tag{15e}$$

$$M_{12} = \int_{-h}^{h}z\sigma_{xy}\,dz = -D(1-\nu)\frac{\partial^2 w}{\partial x\partial y} \tag{15f}$$

$$D \equiv \frac{2Eh^3}{3(1-\nu^2)} \tag{15g}$$

are expressions for the stress resultants $Q_i(x,y)$, $S(x,y)$, and stress couples $M_i(x,y)$, where D is the stiffness of the plate material; and N_1 and N_2 are shear force resultants.

Based on Eq. (8a), the following boundary conditions are set on the plate faces ($z = \pm h$)

$$\sigma_{zi} + T_{zi} = T_{zi}^{(e)} \quad (i = x,y,z) \tag{16}$$

where T_{zj}, $T_{zj}^{(e)}$ are the components of the Maxwell tensor of the internal region occupied by the plate and the external region outside of the plate, respectively.

Notice that, from the boundary conditions and Eq. (15), the plate vibration equations must be solved jointly with the electrodynamic equations for the media surrounding the plate. This situation essentially complicates the problem of magnetoelastic vibrations of the plate. However solutions can be obtained when the magnetic field is parallel to the planes bounding the plate. In such a case, the conditions of continuity of the normal components of perturbed induction of the magnetic field ($h_z = h_{z(e)}$) should be satisfied on the $z = \pm h$ planes. Based on this continuity condition $h_z(\pm h) = h_{z(e)}(\pm h)$ we have

$$T_{zx}(\pm h) - T_{zx}^{(e)}(\pm h) = 0 \tag{17a}$$

$$T_{zy}(\pm h) - T_{zy}^{(e)}(\pm h) = 0 \tag{17b}$$

and therefore

$$\sigma_{xz}(\pm h) = 0, \quad \sigma_{yz}(\pm h) = 0 \tag{18}$$

Neglecting $T_{zz}^{(e)}$ in comparison with T_{zz}, since there is discontinuity of the tangential component of the perturbed magnetic field, the boundary conditions of Eq. (8a) are being replaced with the following

$$\sigma_{zz}(\pm h) + T_{zz}(\pm h) = 0 \tag{19}$$

Substituting Eqs. (17)–(19), into Eq. (14), the equations for the plate planar and bending vibrations can be obtained as follows

1. Planar vibration

$$\Delta u + \theta \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{H_{02}}{4\pi G}\Delta u - \frac{H_{01}H_{02}}{4\pi G}\Delta v = \frac{\rho}{G}\frac{\partial^2 u}{\partial t^2} \tag{20a}$$

$$\Delta v + \theta \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{H_{01}}{4\pi G}\Delta v - \frac{H_{01}H_{02}}{4\pi G}\Delta u = \frac{\rho}{G}\frac{\partial^2 v}{\partial t^2} \tag{20b}$$

Herein the following notations are used

$$\theta = \frac{1+\nu}{1-\nu}, \quad G = \frac{E}{2(1+\nu)}, \quad \Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

2. Bending vibration

$$D\Delta^2 w - \frac{h}{2\pi}\left(H_{01}^2\frac{\partial^2 w}{\partial x^2} + H_{02}^2\frac{\partial^2 w}{\partial y^2} + 2H_{01}H_{02}\frac{\partial^2 w}{\partial x \partial y}\right) + 2\rho h\frac{\partial^2 w}{\partial t^2} + \underline{\frac{h^3}{6\pi}\left[\left(H_{02}^2\frac{\partial^2}{\partial x^2} + H_{01}^2\frac{\partial^2}{\partial y^2} - 2H_{01}H_{02}\frac{\partial^2}{\partial x \partial y}\right)\Delta w\right] - \frac{2\rho h^3}{3}\frac{\partial^2}{\partial t^2}\Delta w} = 0 \tag{21}$$

Notice that Eq. (21), governing the plate bending vibration, is decoupled from Eqs. (20a) and (20b) of in-plane vibrations. In addition, the underlined expression is negligible as compared to the other terms, and therefore this term will not be taken into account.

The boundary conditions on the plate's x=const and y=const edges are obtained by averaging the boundary conditions of the spatial problem of the theory of magnetoelasticity for a perfect conductive medium

$$\int_{-h}^{+h} u_i \, dz = 0 \int_{-h}^{+h} z u_i \, dz = 0 \tag{22a}$$

$$\int_{-h}^{+h} (\sigma_{xi} + T_{xi}) dz = \int_{-h}^{+h} T_{xi}^{(e)} \, dz \tag{22b}$$

$$\int_{-h}^{+h} z(\sigma_{xi} + T_{xi}) dz = \int_{-h}^{+h} z T_{xi}^{(e)} \, dz \tag{22c}$$

Clearly, the boundary conditions for the fixed, hinged edges are the same as in the common plate theory.

We have common boundary conditions with regard to normal deflection, $w(x,y)$, with the exception of the case of the free edge. In this case, the boundary conditions at the x=const edge can be written as

$$M_1 = 0 \tag{23a}$$

$$N_1 + \frac{\partial M_{12}}{\partial x} + \frac{h}{2\pi}\left(H_{01}^2\frac{\partial w}{\partial x} + H_{01}H_{02}\frac{\partial w}{\partial y}\right) + \frac{h^3}{6\pi}\left(H_{01}H_{02}\frac{\partial}{\partial y} - H_{02}^2\frac{\partial}{\partial x}\right)\Delta w = 2hT_{xz}^{(e)} \tag{23b}$$

Substituting the expressions for N_1, M_1, M_{12}, from Eqs. (14) and (15) into Eq. (23), and neglecting $T_{xz}^{(e)}$ in comparison with T_{xz} due to the discontinuity of the tangential (to x=const edge) component of the perturbed magnetic field h_z, the following boundary conditions at x=const are assumed

$$\left(\frac{\partial^2 w}{\partial x^2} + \nu\frac{\partial^2 w}{\partial y^2}\right) = 0 \tag{24a}$$

$$D\frac{\partial}{\partial x}\left[\frac{\partial^2 w}{\partial x^2} + (2-\nu)\frac{\partial^2 w}{\partial y^2}\right] - \frac{h}{2\pi}\left(H_{01}^2\frac{\partial w}{\partial x} + H_{01}H_{02}\frac{\partial w}{\partial y}\right) \underline{- \frac{h^3}{6\pi}\left(H_{01}H_{02}\frac{\partial}{\partial y} - H_{02}^2\frac{\partial}{\partial x}\right)\Delta w} = 0 \tag{24b}$$

Also in this case, the underlined expression is negligible as compared to the other terms and later will not be taken into account.

3 Selected Cases

Let us consider the localized bending vibrations of a semi-infinite plate which occupies the space $0 \leqslant x < \infty$, $-\infty < y < \infty$, $-h \leqslant z \leqslant h$ (see Fig. 1).

It is assumed that the localized waves are propagating along the y axis. Two cases will be considered: (a) a magnetic field that is perpendicular to the axis along which waves are propagating, $H_{01} \neq 0$, $H_{02} = 0$; and (b) a magnetic field that is parallel to the axis along which waves are propagating, $H_{01} = 0$, $H_{02} \neq 0$.

3.1 Case A: Plate Immersed in a Magnetic Field That is Perpendicular to the Axis Along Which the Waves are Propagating, $H_{01} \neq 0$, $H_{02}=0$. In Case A, in the region $0 \leqslant x < \infty$; $-\infty < y < \infty$ the governing equation of bending vibration can be cast as

$$D\Delta^2 w - \frac{hH_{01}^2}{2\pi}\frac{\partial^2 w}{\partial x^2} + 2\rho h \frac{\partial^2 w}{\partial t^2} = 0 \tag{25}$$

and the associated boundary conditions at $x=0$ are

$$\left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2}\right) = 0 \tag{26a}$$

$$D\frac{\partial}{\partial x}\left[\frac{\partial^2 w}{\partial x^2} + (2-\nu)\frac{\partial^2 w}{\partial y^2}\right] - \frac{hH_{01}^2}{2\pi}\frac{\partial w}{\partial x} = 0 \tag{26b}$$

At $x \to \infty$ the vibration damps out, implying that

$$\lim_{x\to\infty} w = 0 \tag{27}$$

The solution of Eq. (25) satisfying the condition from Eq. (27) can be cast as

$$w(x,y) = w_0(C_1 e^{-kpx} + C_2 e^{-kqx})\exp i(\omega t - ky) \tag{28a}$$

where k is a wave number; and ω is a frequency of vibration:

$$p = \sqrt{(1 + \chi + \sqrt{\eta^2 + 2\chi + \chi^2})} \tag{28b}$$

$$q = \sqrt{(1 + \chi - \sqrt{\eta^2 + 2\chi + \chi^2})} \tag{28c}$$

and

$$\eta^2 = \frac{2\rho h \omega^2}{Dk^4} \tag{28d}$$

$$\chi = \frac{hH_0^2}{4\pi D k^2} \tag{28e}$$

$$= \frac{3H_0^2(1-v^2)}{8\pi E k^2 h^2} \tag{28f}$$

The nondimensional η parameter in Eq. (28c) defines the frequency of localized vibration, and according to the condition of damping, Eq. (27), it should satisfy the following inequalities

$$0 < \eta^2 < 1 \tag{29}$$

Substituting the solution, Eq. (28a), into the boundary conditions, Eqs. (26), a homogeneous system of equations, with respect to the arbitrary C_1, C_2 constants, are obtained. The dispersion equation can be obtained by equating the determinant of the simultaneous set of equations to zero, yielding

$$K(\eta) \equiv q(p^2 - v)[q^2 - (2 - v - 2\chi)] - p(q^2 - v)[p^2 - (2 - v - 2\chi)] = 0 \tag{30a}$$

or in a compact form

$$K(\eta) \equiv (p-q)K_1(\eta) \tag{30b}$$

where

$$K_1(\eta) \equiv (pq)^2 + 2(1 - v - \chi)pq - v^2 \tag{30c}$$

Since $p \neq q$, the frequency of the vibrations can be obtained from the equation

$$K_1(\eta) = 0 \tag{31}$$

Noting that

$$\frac{dK_1}{d\eta} > 0 \quad \text{at } 0 \leqslant \eta \leqslant 1 \tag{32a}$$

$$K_1(0) = (1-v)(3+v) - 2\chi, \quad K_1(1) = -v^2 < 0 \tag{32b}$$

one can conclude that Eq. (30) has only one root in the interval $0 \leqslant \eta \leqslant 1$ when $K_1(0) > 0$. Therefore, the sufficient and necessary conditions of localized wave existence are

$$\chi < (3+v)(1-v)/2, \quad v \neq 0 \tag{33}$$

The root of Eq. (31) is found in the following way

$$\eta^2 = 1 + 2(1 - v - \chi)\sqrt{(1 - v - \chi)^2 + v^2} - 2(1 - v - \chi)^2 - v^2 \tag{34}$$

From Eq. (33), it follows that in satisfying the condition

$$H_{01}^2 > 4\pi E(kh)^2(3+\nu)/3(1+\nu) \tag{35}$$

the localized vibrations in the vicinity of the free edge of the plate are eliminated.

In particular, for a metal plate ($E=110$ GPa; $v=0.3$) with relative wavelength of $kh=0.02$ and $kh=0.01$ the induction of the intensity of the magnetic field ($\chi_0=1.15$) necessary for eliminating the localized vibrations is on the order of $H_{01} \sim 1.5T$ and $H_{01} \sim 0.37T$, respectively

3.2 Case B: Plate Immersed in a Magnetic Field That is Parallel to the Axis Along Which the Waves are Propagating, $H_{01}=0$, $H_{02} \neq 0$. In this case we have the following boundary value problem

$$D\Delta^2 w - \frac{hH_{02}^2}{2\pi}\frac{\partial^2 w}{\partial y^2} + 2\rho h \frac{\partial^2 w}{\partial t^2} = 0 \quad 0 \leqslant x < \infty; \quad -\infty < y < \infty \tag{36}$$

and boundary conditions at $x=0$

$$\left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2}\right) = 0 \tag{37a}$$

$$D\frac{\partial}{\partial x}\left[\frac{\partial^2 w}{\partial x^2} + (2-\nu)\frac{\partial^2 w}{\partial y^2}\right] = 0 \tag{37b}$$

At $x \to \infty$, the vibration damps out, and Eq. (27) is valid.

The solution of Eq. (36) satisfying the condition Eq. (27), can be written as

$$w(x,y) = w_0(C_1 e^{-k\tilde{p}x} + C_2 e^{-k\tilde{q}x})\exp i(\omega t - ky) \tag{38a}$$

where

$$\tilde{p} = \sqrt{1 + \sqrt{\eta^2 - 2\chi}}, \tag{38b}$$

$$\tilde{q} = \sqrt{1 - \sqrt{\eta^2 - 2\chi}} \tag{38c}$$

The notations for the η^2 and χ are the same as in Eqs. (28c) and (28d).

The condition of the existence of the localized vibrations is formulated in the following way

$$2\chi < \eta^2 < 1 + 2\chi \tag{39}$$

The dispersion equation is obtained from the requirement that the solution, Eq. (38), should fulfill the boundary conditions from Eq. (37) as

$$K_2(\eta) \equiv (\tilde{p}\tilde{q})^2 + 2(1-v)(\tilde{p}\tilde{q}) - v^2 = 0 \tag{40}$$

The function $K_2(\eta)$ at the ends of the interval of Eq. (39) results in

$$K_2(2\chi) = (1-v)(3+v)/2 > 0, \quad K_2(1+2\chi) = -v^2 < 0 \tag{41}$$

Therefore, the localized vibrations exist ($v \neq 0$), regardless of the magnitude of intensity of the magnetic field. Notice that the influence of the magnetic field on the localized vibrations in this case is weaker than in the previous case.

Table 1 Dimensionless localized vibration frequencies for selected magnetic field and Poisson's ratio

χ		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
$\nu=0.1$	$\eta_1=0.99$	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.90	0.78	0.58	—
	$\eta_2=0.99$	1.09	1.18	1.26	1.34	1.41	1.48	1.54	1.61	1.67	1.73	1.78	1.84	1.89	1.94
$\nu=0.2$	$\eta_1=0.99$	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.94	0.87	0.75	0.53	—	—
	$\eta_2=0.99$	1.09	1.18	1.26	1.34	1.41	1.48	1.54	1.61	1.67	1.73	1.78	1.84	1.89	1.94
$\nu=0.3$	$\eta_1=0.99$	0.99	0.99	0.99	0.99	0.98	0.97	0.95	0.90	0.83	0.68	0.43	—	—	—
	$\eta_2=0.99$	1.09	1.18	1.26	1.34	1.41	1.48	1.54	1.61	1.67	1.73	1.78	1.84	1.89	1.94
$\nu=0.4$	$\eta_1=0.99$	0.99	0.98	0.97	0.96	0.94	0.85	0.76	0.6	0.25	—	—	—	—	—
	$\eta_2=0.99$	1.08	1.17	1.25	1.33	1.40	1.47	1.54	1.60	1.66	1.72	1.78	1.83	1.89	1.94
$\nu=0.5$	$\eta_1=0.98$	0.97	0.95	0.94	0.91	0.86	0.79	0.67	0.46	—	—	—	—	—	—
	$\eta_2=0.98$	1.07	1.16	1.24	1.32	1.39	1.46	1.53	1.59	1.66	1.71	1.77	1.83	1.88	1.91

4 Results

In Table 1 the localized vibration frequencies η_1 and η_2 are given as a function of the magnetic field parameter χ and Poisson's ratio υ. The frequencies $\eta_1(\chi,\upsilon)$ correspond to the case of the magnetic field H_{01} acting in a direction perpendicular to the axis along which the waves are propagating, while the frequencies $\eta_2(\chi,\upsilon)$ correspond to the case of the magnetic field H_{02} acting in a direction that is parallel to the axis along which the waves are propagating. Dashes in the table correspond to the lack of localized vibration. For the first case, application of H_{01}, we have a decrease in the localized vibration frequencies when magnetic field increases up to critical value $\chi_0=\frac{1}{2}(3+\upsilon)(1-\upsilon)$, beyond which the localized vibration vanishes. For the second case, application of H_{02}, it is evident that there is an increase in the localized vibration frequencies for increasing magnetic field intensity, implying that there are always localized vibrations. The data presented in Table 1 are obtained from Eqs. (31) and (40), respectively. These cases have also been reported in Figs. 2 and 3, for Cases A and B, respectively.

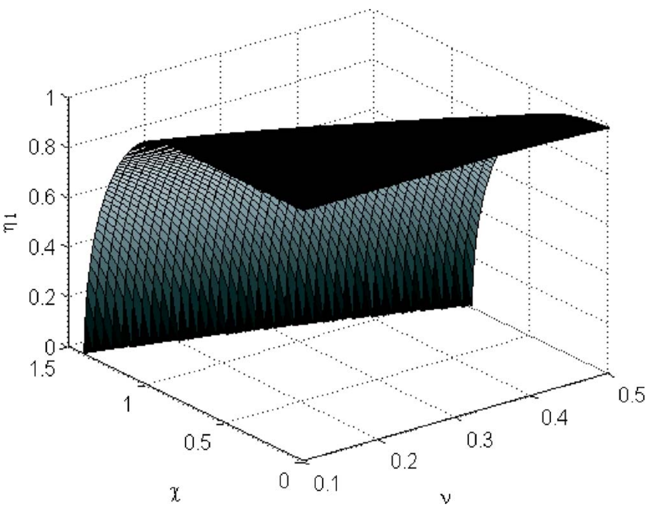


Fig. 2 Frequency spectrum from Case A

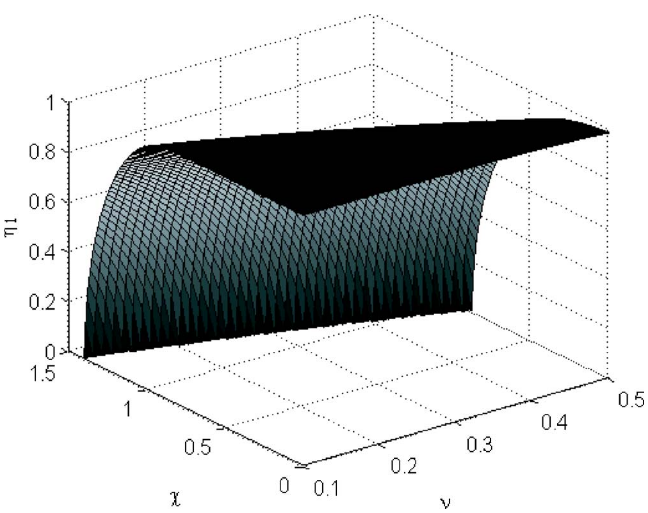


Fig. 3 Frequency spectrum from Case B

5 Conclusions

The study of planar and bending magnetoelastic vibrations of a perfectly conductive flat plate immersed in a uniform external magnetic field is presented. For this system, this paper demonstrates the presence of localized bending vibrations in the vicinity of the plate free edge and shows that localized vibrations can be detected and eliminated by means of the magnetic field.

Acknowledgment

The authors would like to acknowledge the partial funding support of the National Science Foundation, NSF Grant Nos. OISE-0532683 and NSF CMS-0521217. The authors would like to dedicate this paper to Professor Liviu Librescu of Virginia Tech, that lost his life on April 16, 2007. He will be remembered and deeply missed.

Nomenclature

c = electrodynamic constant, speed light in a vacuum
C_1, C_2 = arbitrary constants Eq. (28a)
D = stiffness of plate material
$\vec{e}$ = induced electric field vector
E = modulus of elasticity of plate material
$\vec{F}, F_1$ = electromagnetic bulk force vector and components, respectively
$2h$ = plate thickness
$\vec{h}, h_i$ = induced magnetic field vector and components, respectively
$\vec{H}, H_{0.1}, H_{0.2}$ = magnetic field vector and tensor coordinate in the x and y dimensions, respectively
$\vec{j}$ = induced current vector
k = wave number
M_i = stress couples Eqs. (15d)–(15f)
$\vec{n}, n_k$ = outward normal vector vector to plate surface and components, respectively
Q_i = stress resultants Eqs. (15a) and (15b)
N_1, N_2 = shear force resultants Eqs. (14c).
S = stress resultant Eq. (15c)

t = time
$\hat{T}, T_{ik}$ = Maxwell electromagnetic tensor and components, respectively
$\vec{u}, u_i$ = elastic plate displacement vector
u, v, w = displacements of plate's middle plane
x_1, x_2, x_3 = vector reference frame
x, y, z = Cartesian reference frame
χ = dimensionless parameter of magnetic field intensity
η = dimensionless frequency of localized vibration
ν = Poisson ratio of plate material
ω = frequency of vibration
Ω = plate surface
ρ = density of plate material
ρ_e = density of electrical charges
$\hat{\sigma}, \sigma_{ik}$ = elastic stress tensor and components, respectively
σ = electric conductivity of plate material
δ_{ik} = Kronecer's delta symbol

References

[1] Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W., 1996, "Damage Identification and Health Monitoring of Structural and Mechanical Systems From Changes in Their Vibration Characteristics: A Literature Review," Los Alamos National Laboratory, Los Alamos, NM, Technical Rep. No. LA-13070-MS.
[2] Konenkov, Y. K., 1960, "On Rayleigh Type Bending Waves," Akust. Zh., **6**(1), pp. 124–126 (in Russian).
[3] Belubekyan, M. V., and Engibaryan, J. A., 1996, "Waves Localized Along the Free Edge of the Plate With Cubic Symmetry," Proceedings of RAS, MTT, Springer, New York, Vol. 16, pp. 139–143 (in Russian).
[4] Ambartsumian, S. A., and Belubekyan, M. V., 1994, "On Bending Waves Localized Along the Edge of a Plate," Int. Appl. Mech., **30**(2), pp. 135–140.
[5] Mkrtchyan, H. P., 2003, "Localized Bending Waves in an Elastic Orthotropic Plate," Proceedings of NAS Armenia, Mechanics, Vol. 56, No. 4, pp. 66–68.
[6] McKenna, J., Boyd, G. D., and Thurston, R. N., 1974, "Plate Theory Solution for Guided Flexural Acoustic Waves Along the Tip of a Wedge," IEEE Trans. Sonics Ultrason., **21**, pp. 178–186.
[7] Norris, A. N., Krylov, V. V., and Abrahams, I. D., 1998, "Flexural Edge Waves and Comments on a New Bending Wave Solution for the Classical Plate Equation," J. Acoust. Soc. Am., **104**, pp. 2220–2222.
[8] Norris, A. N., 1994, "Flexural Edge Waves," J. Sound Vib., **171**, pp. 571–573.
[9] Belubekyan, M. V., 2003, "Magnetoelastic Vibrations Localized in the Vicinity of the Free Edge of a Thin Plate," Proceedings of Armenian National Academy of Sciences, Mechanics, Yerevan, Armenia, Vol. 56, No. 2.
[10] Kaliski, S., 1962, "Magnetoelastic Vibration of a Perfectly Conducting Plates and Bars Assuming the Principle of Plane Sections," Proc. Vib. Probl., **3**(14), pp. 225–234.
[11] Ambartsumian, S. A., Bagdasaryan, G. E., and Belubekian, M. V., 1971, "On the Three-Dimensional Problem of the Plate Magnetoelastic Vibration," Prikladnaia Matematica i Mekhanika, **35**(12), pp. 216–228.
[12] Ambartsumian, S. A., Bagdasaryan, G. E., and Belubekian, M. V., 1977, *The Magnetoelasticity of Thin Shells and Plates*, Nauka, Moscow, pp. 272 (in Russian).
[13] Ambartsumian, S. A., 2002, "Nontraditional Theories of Shells and Plates," Appl. Mech. Rev., **5**, pp. R35–R44.
[14] Rudnicki, M., 2002, "Eigenvalue Solutions for Motion of Electroconductive Plate in Magnetic Field," Int. J. Eng. Sci., **40**, pp. 93–107.
[15] Ambartsumian, S. A., Belubekyan, M. V., and Minassyan, M. M., 1983, "On the Problem of Non-Linear Elastic Electroconductive Plates in Transverse and Longitudinal Magnetic Fields," Int. J. Non-Linear Mech., **19**(2), pp. 141–149.
[16] Ambartsumian, S. A., and Belubekian, M. V., 1979, "On the Approximate Methods in the Problems of the Plate Magnetoelastic Vibration." *Thermal Stresses in the Elements of Constructions*, Institute of Mechanics, Academy of Sciences, Ukrainian SSR, Kiev, Vol. 19, pp. 3–6 (in Russian).
[17] Ambartsumian, S. A., and Bagdasaryan, G. E., 1996, "Conductive Plates and Shells in Magnetic Fields." M.: Phys-Math. Lit., pp. 288 (in Russian).
[18] Librescu, L., Hasanyan, D., and Ambur, D. R., 2004, "Electromagnetically Conducting Elastic Plates in a Magnetic Field: Modeling and Dynamic Implications," Int. J. Non-Linear Mech., **39**, pp. 723–739.
[19] Librescu, L., Hasanyan, D., Qin, Z., and Ambur, D. R., 2003, "Nonlinear Magnetothermoelasticity of Anisotropic Plates Immersed in a Magnetic Field," J. Therm. Stresses, **26**, pp. 1277–1304.
[20] Soedel, W., 1993, *Vibrations of Plates and Shells*, Marcel Dekker, New York.
[21] Moon, F. C., 1984, *Magneto-Solid Mechanics*, Wiley, New York.

L. I. Manevitch
N. N. Semenov Institute for Chemical Physics,
Russian Academy of Sciences,
ul. Kosygina 4,
119991 Moscow, Russia

E. Gourdon[1]
e-mail: gourdon@entpe.fr

C. H. Lamarque

Ecole Nationale des Travaux Publics de l'Etat,
LGM,
URA CNRS 1652,
Rue Maurice Audin,
F-69518 Vaulx en Velin Cedex, France

Towards the Design of an Optimal Energetic Sink in a Strongly Inhomogeneous Two-Degree-of-Freedom System

Analytical, numerical, and experimental results of energy pumping in a strongly inhomogeneous two-degree-of-freedom system are to be presented in this study. The latter is based both on efficient analytical solution and comparative analysis for various types of energetic sinks. Considering the efficient pumping process as damped beating with strong energy transfer, it is shown that we can design the sinks with amplitude-phase variables which provide the most efficient result. In this study, the main types of energetic sinks are to be compared. Computer simulation has confirmed the analytical predictions which had been obtained. Experimental verification of the analytical prediction is considered for a particular type of sink. [DOI: 10.1115/1.2711221]

Keywords: energy pumping, multiple scales, complex variables, building model

1 Introduction

The problem of passive irreversible transfer of mechanical energy (referred to as energy pumping) in oscillatory systems is now a subject of growing interest [1–9]. This paper tackles a nonlinear energy sink (NES). The NES is a passive isolation device which is basically composed of a nonlinear and weakly damped oscillator. This nonlinear oscillator is coupled to the primary system to absorb energy. The first papers in this field [1–5] were devoted to the demonstration of the phenomenon itself, but only on one condition: both the main subsystem and the energetic sink have to be of an equal mass. In Ref. [1], the redistribution of energy was considered after impulsive inputs had been applied to a linear primary system coupled to a NES. This redistribution of energy was referred to as "energy pumping." It occurred in a one-way and irreversible fashion: the energy, once pumped from the linear to the nonlinear oscillator, does not return back into the linear system. Energy pumping was attributed to 1:1 resonance capture [3]. References [3,4] demonstrated that the NES's damped dynamics is dependent on its undamped dynamics. To be more precise, it was shown that the undamped system possesses a nonlinear normal mode (NNM) which has a large amount of energy localized in the nonlinear oscillator. When weak damping is added, this NNM is transformed into a damped NNM, and when the latter is sufficiently excited, energy pumping occurs. A recent study [7] took into account a possible strong asymmetry of the masses. Plus, it was assumed that such a sink was not only coupled with the main subsystem but was also linked to the ground with an anchor spring. As this assumption does not correspond to realistic conditions required for applications in engineering, such a restriction is to be removed from this paper. Instead, we will consider a "free" attachment, that is to say a strong nonlinear attachment (strongly inhomogeneous two-degree-of-freedom system) coupled to a linear primary system. Although there are numerous papers devoted to the study of the NES, most attempt to explain the phenomenon which is responsible for energy pumping, but none was found to design optimal energetic sink parameters for given primary system specifications.

As the phenomenon of energy pumping in a strongly nonlinear damped system [1–9] clearly manifests regularities of transitional dynamical processes in the vicinity of internal resonances [10], a series of significant problems can be formulated, and among them is the following question: What are the possibilities of optimization of energetic sink parameters? The efficient analytical description of the pumping process in a strongly inhomogeneous two-degree-of-freedom system proposed in Ref. [9] turned out to be appropriated to the resolution of the optimization problem if applied to a "cubic"-type sink which has widely been studied in previous papers [14]. In parallel, some other strongly nonlinear sinks have also been discussed within the framework of energy pumping problems. However, isolated numerical estimations as well as the absence of criteria on sinks efficiency make their comparison difficult. Therefore, such a criterion has been proposed in Ref. [14] by considering the energy pumping process as a damped beating, and is now being developed in the present paper. Together with this study, experimental verifications are to be carried out using a reduced-scale building [13].

2 Analytical Study

The following system of coupled oscillators is considered

$$\begin{aligned} &M\frac{d^2x_1}{dt^2}+\tilde{\mu}_1\frac{dx_1}{dt}+\tilde{\eta}\left(\frac{dx_1}{dt}-\frac{dx_2}{dt}\right)+k_1x_1 \\ &\quad +k_3(x_1-x_2)^{2n-1}\pm D(x_1-x_2)=0 \\ &m\frac{d^2x_2}{dt^2}-\tilde{\eta}\left(\frac{dx_1}{dt}-\frac{dx_2}{dt}\right)-k_3(x_1-x_2)^{2n-1} \\ &\quad \pm D(x_2-x_1)=0, \end{aligned} \qquad n \geqslant 2 \tag{1}$$

Nonlinear coupling with multiple states of equilibrium corresponds to the sign "−" in the terms $\pm D(x_2-x_1)$. The linear primary structure is excited by an impulse, so we consider free oscillations of structures with initial conditions: $x_1(t=0)=x_2(t=0)=dx_2/dt(t=0)=0$, $dx_1/dt(t=0)=C_I$.

System (1) can be analyzed by using the perturbation theory. The following change of variables $\tilde{U}_1=x_1$, $\tilde{U}_2=x_2-x_1$ is considered. Then system (1) looks like

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received September 7, 2005; final manuscript received August 2, 2006. Review conducted by Oliver M. O'Reilly.

$$(M+m)\frac{d^2\tilde{U}_1}{dt^2}+m\frac{d^2\tilde{U}_2}{dt^2}+\tilde{\mu}_1\frac{d\tilde{U}_1}{dt}+k_1\tilde{U}_1=0$$
$$m\frac{d^2\tilde{U}_2}{dt^2}+m\frac{d^2\tilde{U}_1}{dt^2}+\tilde{\eta}\frac{d\tilde{U}_2}{dt}+k_3\tilde{U}_2^{2n-1}\pm D\tilde{U}_2=0, \quad n\geqslant 2$$

To clarify the equations, dimensionless coefficients and displacements are used and previous equations are rewritten as

$$(1+\varepsilon)\frac{d^2U_1}{d\tau^2}+\varepsilon\frac{d^2U_2}{d\tau^2}+\varepsilon\mu_1\frac{dU_1}{d\tau}+U_1=0$$
$$\varepsilon\frac{d^2U_2}{d\tau^2}+\varepsilon\frac{d^2U_1}{d\tau^2}+\varepsilon\eta\frac{dU_2}{d\tau}+cU_2^{2n-1}\pm\varepsilon\alpha U_2=0, \quad n\geqslant 2 \quad (2)$$

where

$$\omega=\sqrt{\frac{k_1}{M}},\quad U_1=\frac{\omega}{C_I}\tilde{U}_1,\quad U_2=\frac{\omega}{C_I}\tilde{U}_2,$$

$$\varepsilon=\frac{m}{M},\quad \tau=\omega t,\quad \varepsilon\mu_1=\sqrt{\frac{1}{k_1M}}\tilde{\mu}_1,$$

$$\varepsilon\eta=\sqrt{\frac{1}{k_1M}}\tilde{\eta},\quad \varepsilon\alpha=\sqrt{\frac{1}{k_1M}}D$$

and

$$c=\frac{C_I^{2n-2}k_3}{\omega^{2n-2}k_1}$$

ε is a small parameter, representing a mass ratio which has to be very small.

The infinitesimal order of the nonlinear terms in the second equation of System (2) is less than the infinitesimal order of linear terms in the same equation. The following change of variables $u_1=\varepsilon^{-1/(2n-2)}U_1$, $u_2=\varepsilon^{-1/(2n-2)}U_2$ is introduced. Then Eq. (2) looks like

$$(1+\varepsilon)\frac{d^2u_1}{d\tau^2}+(1+\varepsilon)u_1+\varepsilon\left(\frac{d^2u_2}{d\tau^2}+\mu_1\frac{du_1}{d\tau}-u_1\right)=0$$
$$\frac{d^2u_2}{d\tau^2}+u_2+\varepsilon\left(-\delta u_2+\delta\frac{d^2u_1}{d\tau^2}+\delta\eta\frac{du_2}{d\tau}+\delta cu_2^{2n-1}\pm\delta\alpha u_2\right)=0 \quad (3)$$

where $\delta=1/\varepsilon$.

Then we assume that the oscillations occur next to the resonance on frequency ω. Indeed, as shown in a lot of numerical evidence [2,8] and analytical results [5,6,14], when energy pumping occurs, u_1 and u_2 are oscillating with the same frequency. So, to validate this point we should suppose that

$$\delta\left(-u_2+\frac{d^2u_1}{d\tau^2}+\eta\frac{du_2}{d\tau}+cu_2^{2n-1}\pm\alpha u_2\right)\sim O(1) \quad (4)$$

Such a procedure is fully justified by detailed numerical analysis and by numerous previous papers. Therefore expression in brackets should be of order ε (but each term in the sum is not necessarily small). It is rather natural, since it describes slow modulation and damping of the vibrations with frequency close to unity. Then Eq. (3) appears as follows

$$(1+\varepsilon)\frac{d^2u_1}{d\tau^2}+(1+\varepsilon)u_1+\varepsilon\left(\frac{d^2u_2}{d\tau^2}+\mu_1\frac{du_1}{d\tau}-u_1\right)=0$$
$$\frac{d^2u_2}{d\tau^2}+u_2+\varepsilon\delta\left(-u_2+\frac{d^2u_1}{d\tau^2}+\eta\frac{du_2}{d\tau}+cu_2^{2n-1}\pm\alpha u_2\right)=0 \quad (5)$$

Introducing the change of variables [11]

$$\varphi_1=e^{-i\tau}\left(\frac{du_1}{d\tau}+iu_1\right)\quad \varphi_1^*=e^{i\tau}\left(\frac{du_1}{d\tau}-iu_1\right)$$
$$\varphi_2=e^{-i\tau}\left(\frac{du_2}{d\tau}+iu_2\right)\quad \varphi_2^*=e^{i\tau}\left(\frac{du_2}{d\tau}-iu_2\right) \quad (6)$$

and performing multiple scale analysis

$$\tau_0=\tau,\quad \tau_1=\varepsilon\tau,\quad \tau_2=\varepsilon^2\tau,\ldots \quad (7)$$

$$\varphi_1=\varphi_{10}+\varepsilon\varphi_{11}+\varepsilon^2\varphi_{12}+\ldots \quad (8a)$$

$$\varphi_2=\varphi_{20}+\varepsilon\varphi_{21}+\varepsilon^2\varphi_{22}+\ldots \quad (8b)$$

leads to the equations

$$\frac{\partial\varphi_{10}}{\partial\tau_1}+\frac{i}{2}(\varphi_{20}+\varphi_{10})+\frac{\mu_1}{2}\varphi_{10}=0 \quad (9)$$

and

$$\frac{\partial\varphi_{20}}{\partial\tau_1}+\delta\left[\frac{i}{2}(\varphi_{20}+\varphi_{10})+\frac{\eta}{2}\varphi_{20}-\frac{C_{2n-1}^{n-1}ic}{2^{2n-1}}\varphi_{20}|\varphi_{20}|^{2n-2}\pm\frac{i\alpha}{2}\varphi_{20}\right]=0 \quad (10)$$

Equations (9) and (10) are multiplied by φ_{10}^* and φ_{20}^*, respectively. Then, by combining these equations using the complex conjugates, the following equation is obtained

$$\frac{\partial|\varphi_{20}|^2}{\partial\tau_1}+\delta\frac{\partial|\varphi_{10}|^2}{\partial\tau_1}+\eta\delta|\varphi_{20}|^2+\delta\mu_1|\varphi_{20}|^2=0 \quad (11)$$

If there is no damping in the system Eqs. (1), i.e., $\eta=\mu_1=0$, then Eq. (11) is the conservation law of quantity $H=|\varphi_{20}|^2+\delta|\varphi_{10}|^2$ relative to time τ_1. One can consider relation (11) as an ordinary differential equation with respect to function $|\varphi_{20}|^2$, the term $\delta\partial|\varphi_{10}|^2/\partial\tau_1+\delta\mu_1|\varphi_{10}|^2$ being a right-hand member. Applying direct Laplace transformation to Eq. (11), we obtain its solution in the form

$$\Psi(s)=\frac{G(s)+|\varphi_{20}|^2(0)}{s+\delta\eta}$$

where $\Psi(s)$ is a Laplace representation of function $|\varphi_{20}|^2(\tau_1)$; and $G(s)$ is a Laplace representation of function $-\delta(\partial|\varphi_{10}|^2/\partial\tau_1)-\delta\mu_1|\varphi_{10}|^2$. After application of inverse Laplace transformation to this equation, we can find the following representation for function $H(\tau_1)$

$$H(\tau_1)=e^{-\delta\eta\tau_1}\left[H(0)+\delta(\delta\eta-\mu_1)\int_0^{\tau_1}|\varphi_{10}|^2(p)\exp(\delta\eta p)dp\right] \quad (12)$$

To find a solution, we expand the integral in the right-hand member of Eq. (12) in a Taylor series in the vicinity of point $\tau_1=0$. It allows us to calculate function $H(\tau_1)$ avoiding solutions of Eqs. (9) and (10) or initial Eqs. (1). Then Eq. (12) looks like

$$H(\tau_1)=e^{-\delta\eta\tau_1}\left\{H(0)+\delta(\delta\eta-\mu_1)\left[\tau_1|\varphi_{10}|^2(0)+\frac{\tau_1^2}{2}\left(2|\phi_{10}|\times(0)\frac{\partial|\varphi_{10}|(0)}{\partial\tau_1}+\delta\eta|\varphi_{10}|^2(0)\right)+\ldots\right]\right\} \quad (13)$$

The quantities $H(0)$ and $|\varphi_{10}|^2(0)$ are known from the initial conditions. The derivative $\partial|\varphi_{10}|/\partial\tau_1$ and higher-order derivatives of function $|\varphi_{10}|$ at the same point $\tau_1=0$ can be found from the initial conditions and equations of motion (1). In all the next numerical examples, values of the system parameters are the same and are given in Table 1.

We should notice that the treatment is based on computing the Taylor series for solution of the averaged system. So, the number

Table 1 Values of system parameters

Parameters	Values
n	2
ω	1
μ_1	0
ε	0.1
η	0.5
c	0.8
α	0.2
$dx_1/dt(t=0)$	0.3

of terms taken into account in the series must be sufficient to have good convergence of the series. Moreover, we must keep in mind that the interest here is to obtain a complete analytical solution to be able to design and optimize the nonlinear attachment. That is why a "minimum" number of terms (with good approximation) is taken into account. In all the next examples, the analytical approximation (13) is used taking into account the Taylor series up to the terms of the fifth order on τ_1. We have computed the solution numerically (and we have done error evaluation) to be sure that the number of terms we pick is sufficient For example, by taking the values of system parameters indicated in Table 1 and sign "+" in Eqs. (1), we can compare the analytical approximation (13) by considering different orders on τ_1 for the Taylor series expansion with the numerical solution Eq. (12) (the integral in Eq. (12) is expanded in Taylor series). Thus, Fig. 1 shows that the fifth order is sufficient to have a good approximation (the maximum errors between Eqs. (12) and (13) are the following: order 1): 0.5155; order 2: 0.3158; order 3: 0.2395; order 4: 0.008792; order 5: 0.0049731; order 6: 0.0047692; and order 7: 0.0047474).

When energy pumping occurs, the analytical approximation (13) is suitable, as shown in Fig. 2. In this figure, the analytical solution H of Eq. (13) (the fifth order for the Taylor series is considered) is compared with the numerical integration of System (1). The system's parameters are indicated in Table 1 and the + sign is considered in Eqs. (1).

In this case, energy pumping occurs as shown in Fig. 3 where the numerical solutions of System (1) have been plotted with and without coupling.

Not only is the analytical approximation (13) suitable, but the different φ_{10}, φ_{20} introduced are also a good approximation, as shown in Fig. 4 where those analytical approximations are compared with results of integrating initial System (1). The system's parameters are indicated in Table 1 and the "+" sign remains in Eqs. (1).

If the "−" sign is considered in Eqs. (1), then the analytical expression (13) is also suitable, as shown in Fig. 5 where the system's parameters are indicated in Table 1 by taking into account the Taylor series up to the terms of the fifth order on τ_1, inclusively.

So, it is now possible to attempt to design the optimal energy sink owing to the calculation of H. Indeed, we can see that if the "+" sign is considered in Eqs. (1), then energy pumping appears to be more efficient since the decrease of energy H is more abrupt. The energy decreases are faster with the "+" sign in Eqs. (1) than with the "−" sign (if all other parameters are fixed), as shown in Fig. 6 where the system's parameters are indicated in Table 1.

Thus, energy pumping is more efficient when the "+" sign is considered in Eqs. (1), as shown in Fig. 7 with numerical integration of System (1) with the same values of parameters as shown

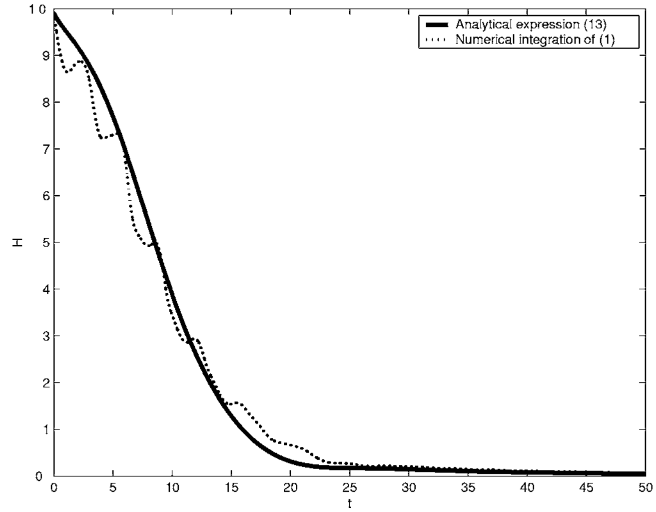


Fig. 2 Function $H(t)$. Solid line depicts solution Eq. (13) taking into account the Taylor series up to the terms of the fifth order on τ_1, inclusive. Dash line depicts the numerical solution of System (1).

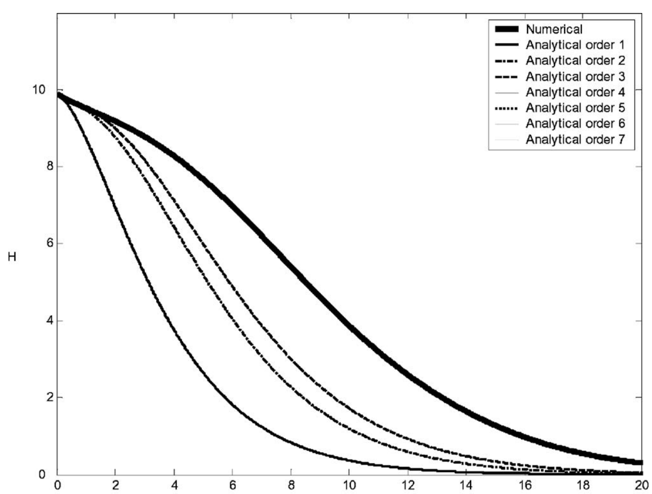


Fig. 1 Function $H(t)$. Comparison between the numerical solution Eq. (12) and the analytical expression Eq. (13) for different orders of the Taylor series.

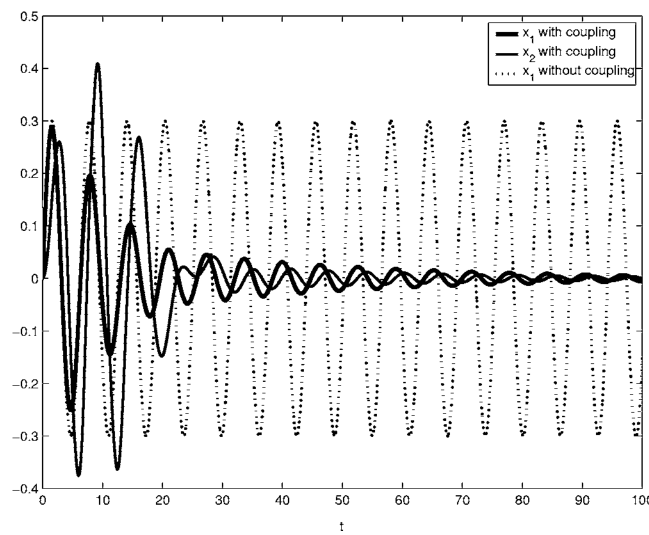


Fig. 3 Responses with numerical integration of Eq. (1) with and without coupling.

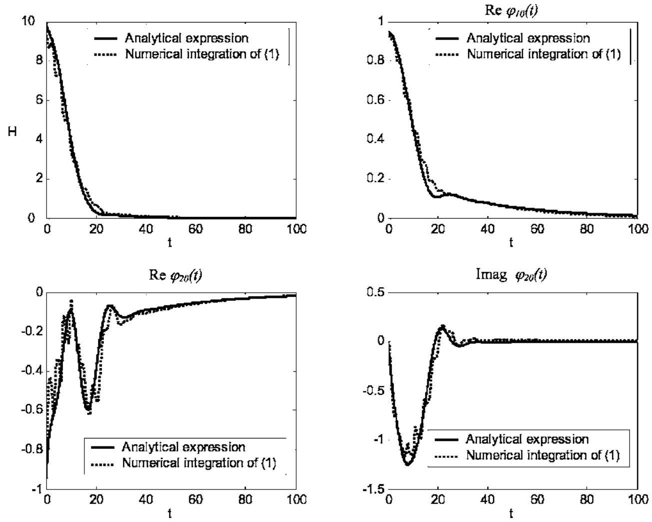


Fig. 4 Function $H(t)$, Imag $\varphi_{10}(t)$, Re $\varphi_{10}(t)$, and Re $\varphi_{20}(t)$, compared with numerical integration of System (1)

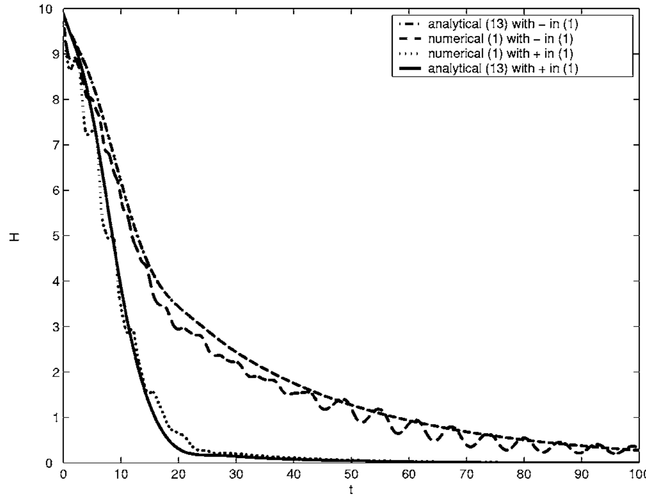


Fig. 6 Comparison of function $H(t)$: consideration of sign + or− in Eqs. (1)

previously. In this figure, it clearly appears that the vibrations are almost completely attenuated at $t=20$ s when the "+" sign is considered in Eqs. (1).

Moreover, we can also consider the influence of the degree n of the nonlinearity on the efficiency of the sink. Indeed, for a given set of parameters, an optimal value of n can be found for which the efficiency of energy pumping is optimal. For this study, we now consider the case of the "+" sign in Eqs. (1) since the efficiency in this case seems better. For example, $dx_1/dt(t=0)=0.4$, $dx_2/dt(t=0)=x_1(t=0)=x_2(t=0)=0$, $n=2/3/4$, $\eta=0.2$ are taken, all other parameters are similar to those indicated in Table 1 and the "+" sign is considered in Eqs. (1). Then, the optimal value of n is 3 (the degree of the nonlinearity is 5), as shown in Fig. 8. In this figure, we can also see that for $n=4$ the analytical approximation is not very suitable after $t=30$ s. Indeed, after $t=30$ s, energy pumping does not occur and there is no longer any resonance.

As we will see, important information about sink efficiency can be extracted from analysis of the corresponding conservative system. Indeed, energy pumping can take place only in the damped system, and is caused by a 1:1 resonance capture of the dynamics on a 1:1 resonant manifold of the system [3]. Reference [15] pointed out a paradoxical fact, namely that although energy pumping takes place only in the damped system, the dynamics governing this phenomenon is influenced by the structure of the NNMs, e.g., the free and synchronous periodic motions of the underlying undamped and unforced system [16]. In the following analysis, we study the bifurcation structure of the NNMs of the undamped

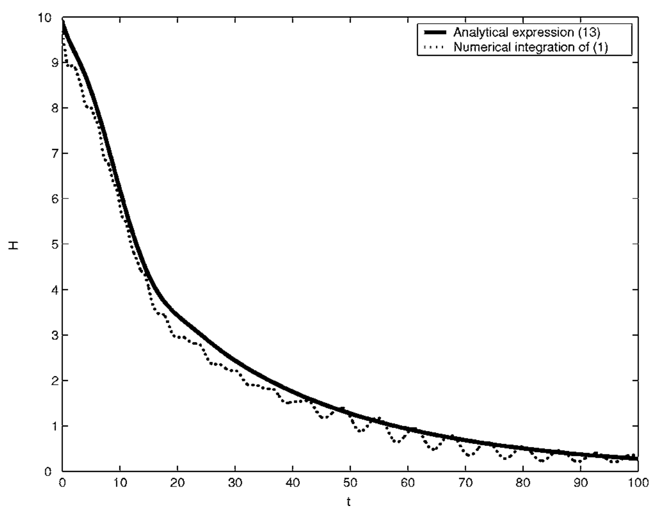


Fig. 5 Function $H(t)$. Solid line depicts solution Eq. (13) taking into account the Taylor series up to the terms of the fifth order on τ_1, inclusive. Dash line depicts the numerical solution of System (1).

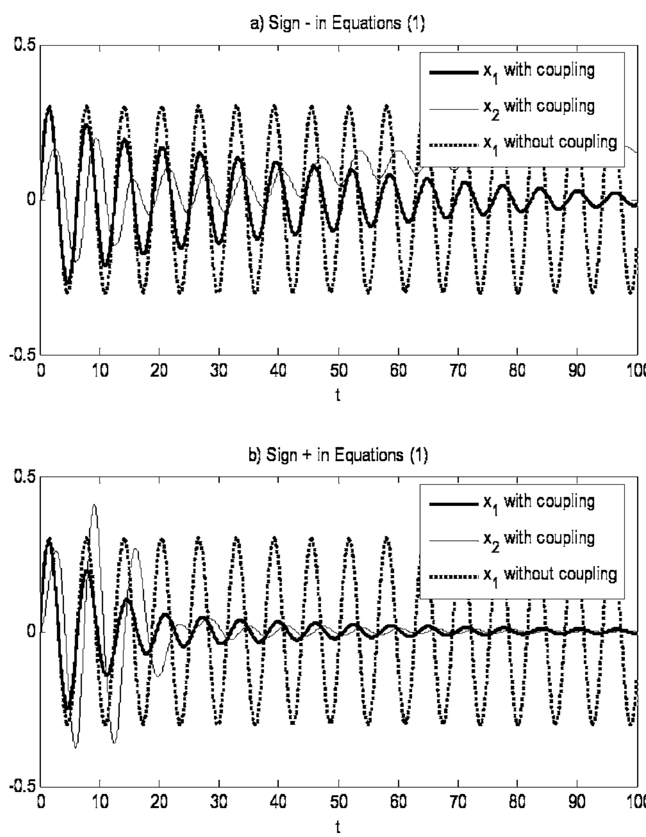


Fig. 7 Comparison of responses owing to numerical integration of (1): consideration of sign + or − in Eqs. (1)

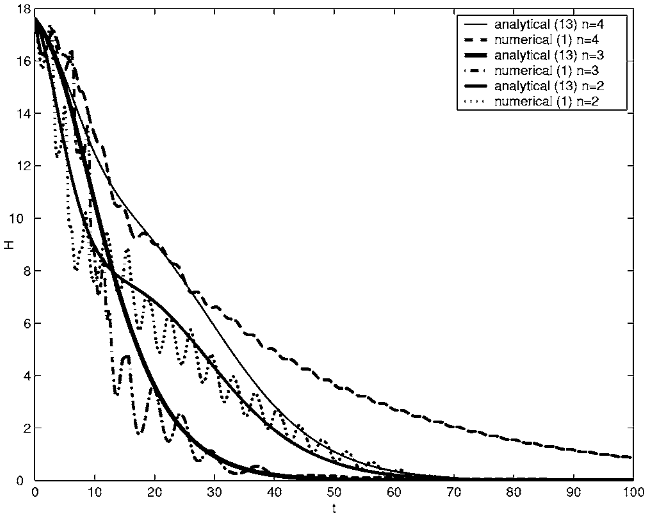


Fig. 8 Comparison of $H(t)$ for different values of n

system of coupled oscillators. As shown in Ref. [17], this NNM bifurcation provides the necessary conditions for the occurrence of nonlinear energy pumping in the corresponding damped system.

If $\eta=\mu 1=0$, Eqs. (9) and (10) look like

$$\frac{\partial \varphi_{10}}{\partial \tau_1}+\frac{i}{2}(\varphi_{20}+\varphi_{10})=0 \tag{14}$$

and

$$\frac{\partial \varphi_{20}}{\partial \tau_1}+\delta\left[\frac{i}{2}(\varphi_{20}+\varphi_{10})-\frac{C_{2n-1}^{n-1}ic}{2^{2n-1}}\varphi_{20}|\varphi_{20}|^{2n-2}\pm\frac{i\alpha}{2}\varphi_{20}\right]=0 \tag{15}$$

By introducing the change of variables

$$\varphi_{10}=f_1, \quad \varphi_{20}=\sqrt{\delta}f_2 \tag{16}$$

they can be written as follows

$$\frac{\partial f_1}{\partial \tau_1}+\frac{i}{2}(f_1+\sqrt{\delta}f_2)=0 \tag{17}$$

and

$$\frac{\partial f_2}{\partial \tau_1}+\sqrt{\delta}\left[\frac{i}{2}(f_1+\sqrt{\delta}f_2)-\frac{C_{2n-1}^{n-1}ic}{2^{2n-1}}\delta^{(2n-1)/2}f_2|f_2|^{2n-2}\pm\frac{i\alpha}{2}\delta^{1/2}f_2\right]=0 \tag{18}$$

The system is now completely integrable with the two first integrals of motion

$$H_1=-\frac{i}{2}(|f_1|^2+\delta|f_2|^2)-\frac{i}{2}(f_2f_1^*+f_1f_2^*)+\frac{C_{2n-1}^{n-1}ic}{2^{2n-1}}\delta^n|f_2|^{2n}\pm\frac{i\alpha}{2}\delta|f_2|^2 \tag{19}$$

and

$$N=|f_1|^2+|f_2|^2 \tag{20}$$

So we can introduce the following change of variables

$$f_1=\sqrt{N}\cos\theta e^{i\delta_1}, \quad f_2=\sqrt{N}\sin\theta e^{i\delta_2} \tag{21}$$

and

$$\Delta=\delta_1-\delta_2 \tag{22}$$

Finally we obtain

$$\frac{\partial\theta}{\partial\tau_1}-\frac{\sqrt{\delta}}{2}\sin\Delta=0 \tag{23}$$

and

$$\frac{\partial\Delta}{\partial\tau_1}-\frac{\delta-1}{2}\pm\frac{\alpha\delta}{2}-\sqrt{\delta}\cos\Delta\,\mathrm{cotg}\,2\theta+\frac{C_{2n-1}^{n-1}c}{2^{2n-1}}\delta^nN^{n-1}\sin^{2n-2}\theta=0 \tag{24}$$

The free periodic solutions (NNMs) of the system (5) (with no damping) correspond to stationary points of the slow flow Eqs. (23) and (24), obtained by setting the derivatives equal to zero

$$\frac{\partial\theta}{\partial\tau_1}=0\Rightarrow\Delta=0,\pi \tag{25}$$

$$\frac{\partial\Delta}{\partial\tau_1}=0\Rightarrow\frac{\delta-1}{2}\pm\frac{\alpha\delta}{2}\pm\sqrt{\delta}\,\mathrm{cotg}\,2\theta+\frac{C_{2n-1}^{n-1}c}{2^{2n-1}}\delta^nN^{n-1}\sin^{2n-2}\theta=0 \tag{26}$$

It turns out that without loss of generality, we can consider only one of the two stationary values of Δ; hence, in the following analysis we assign $\Delta=0$ and the + sign before $\sqrt{\delta}\,\mathrm{cotg}\,2\theta$ in Eq. (26). Then, the stationary solutions of Eq. (26) can be calculated. Thus, for different values of N (with values of parameters indicated in Table 1 except $\eta=0$) it is possible to plot the stationary solutions of Eq. (26), as shown in Fig. 9 where the bifurcation diagrams with alpha as the varying parameter have been plotted. Thus, under a certain value of N, only two NNMs exist (i.e., for a fixed value of alpha, there are only two solutions for theta), as described in Figs. 9(*a*) and 9(*b*), whereas four NNMs can exist for a larger value of N, as underlined in Figs. 9(*b*)–9(*e*) (i.e., for a fixed value of alpha, four solutions for theta exist).

Moreover, by studying Eq. (26) more precisely, the following conclusions can be drawn (two cases appear):

1. First, if $-[(\delta-1)/2]\pm\alpha\delta/2>0$ then for all values of N, only two NNMs exist (i.e., for a fixed value of alpha, there are only two solutions for theta), as shown in Fig. 9; and
2. Second, if $-[(\delta-1)/2]\pm\alpha\delta/2<0$ then under a certain value of N two NNMs exist, as described in Fig. 9, and above a certain value of N, four NNMs can be found (i.e., for a fixed value of alpha, there are four solutions for theta).

In the case where four NNMS appear, the energy pumping phenomenon will occur when damping will be added. Indeed this point can be explained by the fact that a beating phenomenon appears when the damping is not present (this point has been investigated in Ref. [17] and is described further on in this study).

This nonlinear beating can also be seen in the two displacements x_1 and x_2 as shown in Fig. 10, with the same values as previously (the damping is not taken into account in this figure). Indeed, looking at the responses, it is obvious that there is a transfer of energy taking place back and forth between the two masses. There is a resonance of the nonlinear oscillator (x_2). In Fig. 10, x_1 is not trending toward zero since the natural damping has not been taken into account ($\eta=0$) and it is necessary for energy pumping activation. By taking into account the damping (the NNMs are calculated with no damping), energy pumping will occur (i.e., one-way irreversible transfer of energy from one mass to another), as shown in Fig. 11 where damping has been taken into account ($\eta\neq 0, \eta=0.15$), and all other parameters were identical to those taken into account in Fig. 10. This point has been justified by Ref. [1]: the conditions of resonance of x_2 are searching without damping. This NNMs approach is common in energy pumping studies [1–7]. References [1,3,5] demonstrated the dependence of the system's damped dynamics on its undamped dynamics. In particular,

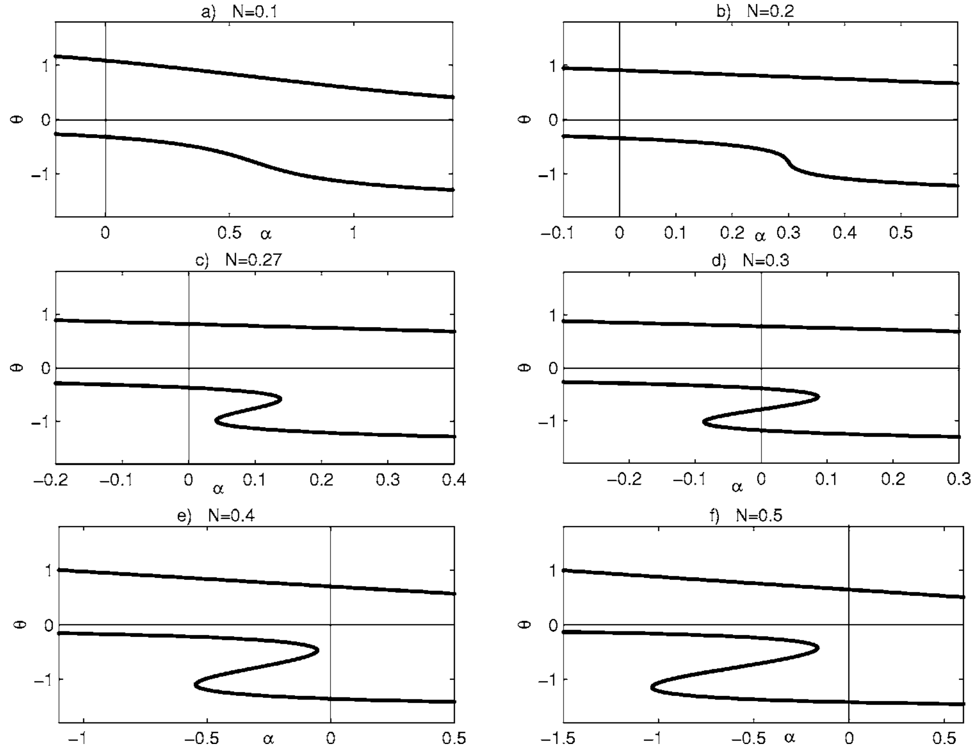


Fig. 9 Bifurcation diagrams: stationary solutions of Eq. (26)

it was shown that the undamped system possesses a nonlinear normal mode which has a large amount of energy localized in the nonlinear oscillator. When weak damping is added, this NNM is transformed into a damped NNM, and when it is sufficiently excited, energy pumping occurs. In our case, the beating phenomenon (when no damping is taken into account) is responsible for energy pumping activation when damping is added. When natural damping is added, all the responses are trending toward zero. That is the reason why it is not justified to make a bifurcation diagram by taking into account damping since energy pumping is a transient phenomenon (the stationary solutions are zero for each initial condition even if there is no energy pumping). Therefore, the

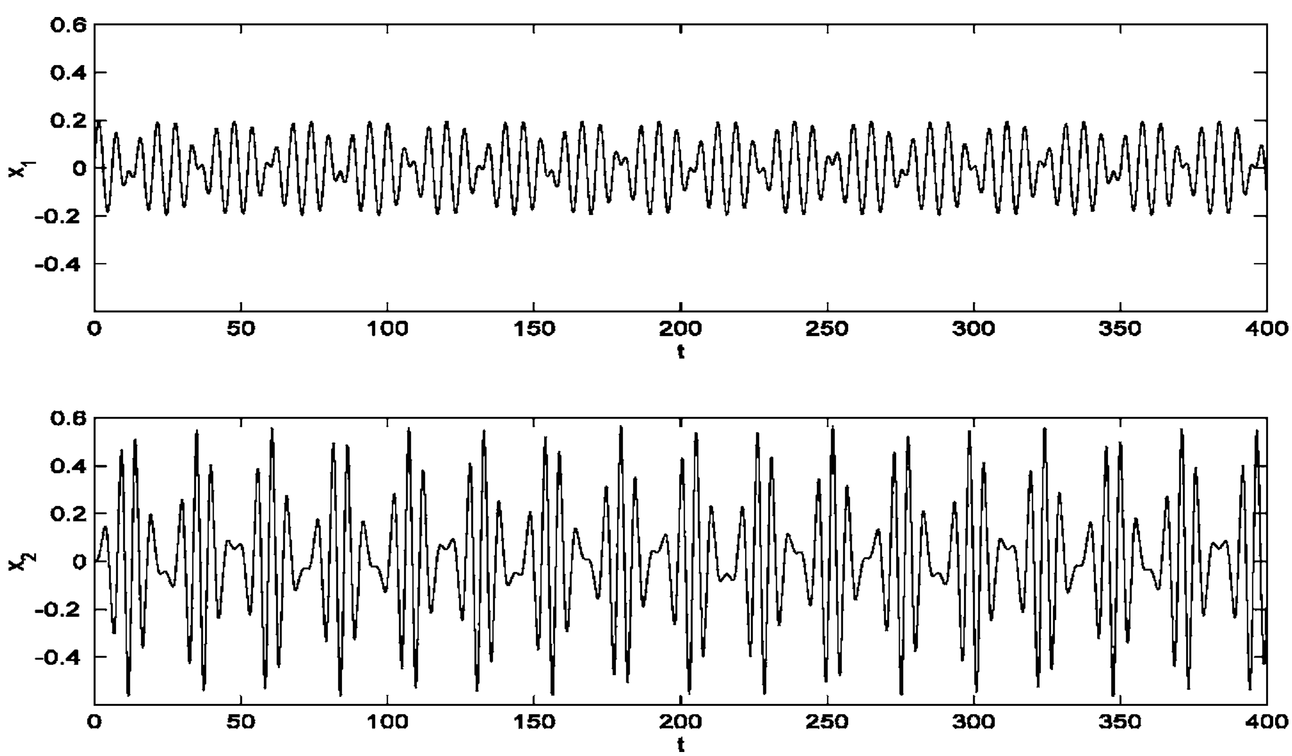


Fig. 10 Nonlinear beating with no damping

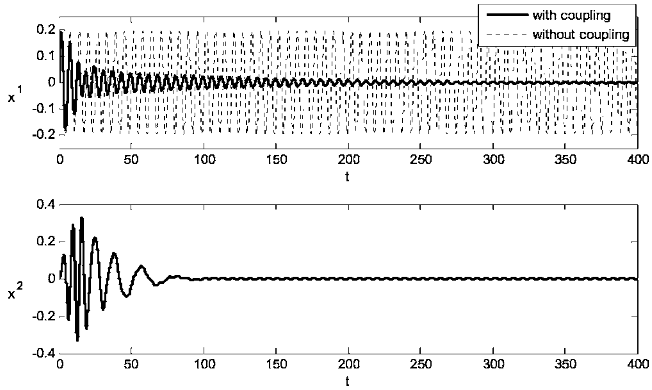


Fig. 11 Energy pumping activation with damping

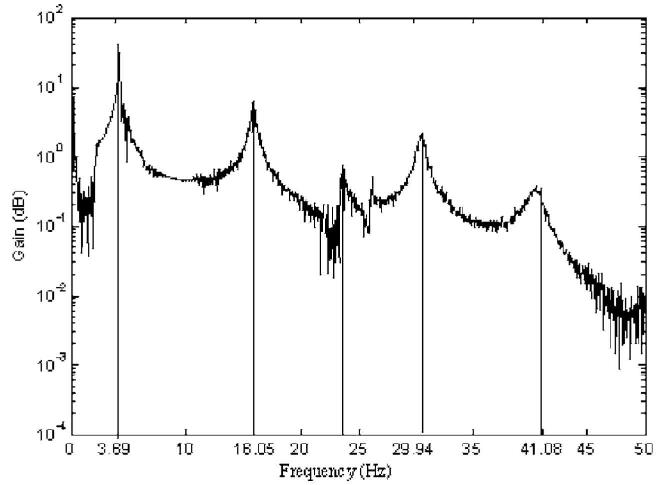


Fig. 13 Linear identification

efficiency of the sink is measured owing to bifurcation diagrams shown in Fig. 9 in terms of undamped NNMs (resonance of the nonlinear oscillator).

3 Experimental Verification

An experimental study is achieved on a small scaled four-story building to verify the previous results about energy pumping phenomenon in a realistic case study. The experimental system is shown in Fig. 12.

The four-story building, namely the linear master structure, was manufactured by welding stainless-steel columns and supporting beams. Steel plates were also welded on each story to simulate realistic mixed steel–concrete slabs. The system is clamped on a plexiglass plate mounted on a shaking table driven by a Linmot electromagnetic linear motor. The latter is controlled by a Linmot E1000 MT controller which is characterized by the ability to efficiently stream almost any excitation profile (sine dwells, sine sweeps, random noise, pulses, earthquakes). Data acquisition of six PCB piezotronics accelerometers is performed by using a HP 3566A/67 analyzer up to a sampling frequency of 12,800 Hz. Linear eigenfrequencies and related specific damping of the main structure are identified when analyzing averaged FrF response curves obtained for random white noise or sine sweep excitations as illustrated in Fig. 13.

Modal identification is performed by a pole-residual technique using Matlab structural dynamic toolbox SDtools. Generally speaking, eigenmodes are shear modes because of the low stiffness of the columns and high plates. As a result accelerometers placed on each story of the structure indeed record actual horizontal components and not rotated ones. The secondary mass of the absorber can slide along a rail fixed to the top of the simple building. x_1 and x_2 represent, respectively, absolute displacements of the primary structure and of the added mass.

For the primary structure, we consider only the first mode. The reduction of the model building to its first mode (mass M) is reasonable since as underlined in Ref. [8], only the first mode here is responsible for energy pumping and other modes of the linear structure are just simple damped oscillators. m denotes the mass of the second added structure. In this experiment, the first modal idealized viscous damping coefficient between the primary mass and the support is c_1 and between the primary mass and the secondary mass is c_2. This model was designed, built, and tested at the LGM ("Laboratoire Géo-Matériaux") Laboratory (Ecole Nationale des Travaux Publics de l'Etat). As underlined in Ref. [12] a cubic nonlinearity is made geometrically with two linear springs (k and l are the stiffness and length of each linear spring). However, when this spring is experimentally implemented, the linear term appears since a small pre-stress force exists in the spring. That is why even if the pre-stress force is kept to a minimum linear coupling also appears.

An impulse is considered at the top of the primary structure. Thus, the building and nonlinear absorber can be idealized by the model displayed in Fig. 14.

The complete system is given by the following equations

$$M\frac{d^2x_1}{dt^2} + \tilde{\mu}_1\frac{dx_1}{dt} + \tilde{\eta}\left(\frac{dx_1}{dt} - \frac{dx_2}{dt}\right) + k_1x_1 + D(x_1-x_2) + k_3(x_1-x_2)^3 = 0$$

$$m\frac{d^2x_2}{dt^2} - \tilde{\eta}\left(\frac{dx_1}{dt} - \frac{dx_2}{dt}\right) - D(x_1-x_2) - k_3(x_1-x_2)^3 = 0 \quad (27)$$

where $k_3 = k/l^2$. Equation (25) is similar to Eqs. (1) studied in the previous parts with consideration of the "+" sign in front of the terms $\pm D(x_1 - x_2)$ and $n=2$. The experimental parameters are $m = 0.121$ kg, $M = 1.677$, so $\varepsilon = 0.07215$. A modal analysis and experimental dynamic analysis of the structures give $k_1 = 900.3\ \mathrm{N\,m^{-1}}$, $\tilde{\mu}_1 = 0.995\ \mathrm{N\,s\,m^{-1}}$, $\tilde{\eta} = 1.452\ \mathrm{N\,s\,m^{-1}}$, $D = 30\ \mathrm{N\,m^{-1}}$, and $k_3 = 1.48.10^6\ \mathrm{N\,m^{-3}}$. The natural frequency of the linear oscillator is 3.69 Hz. By considering an impulse at the first

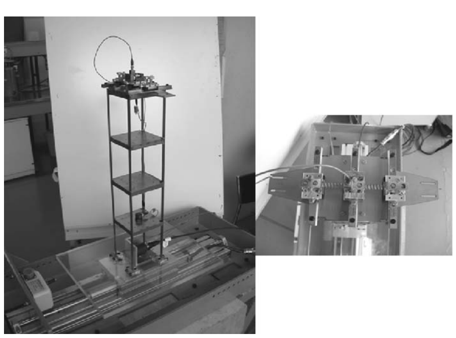


Fig. 12 Experimental system

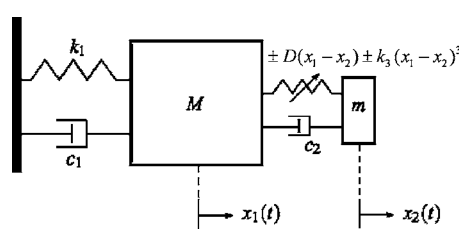


Fig. 14 Considered system with 2 DOF

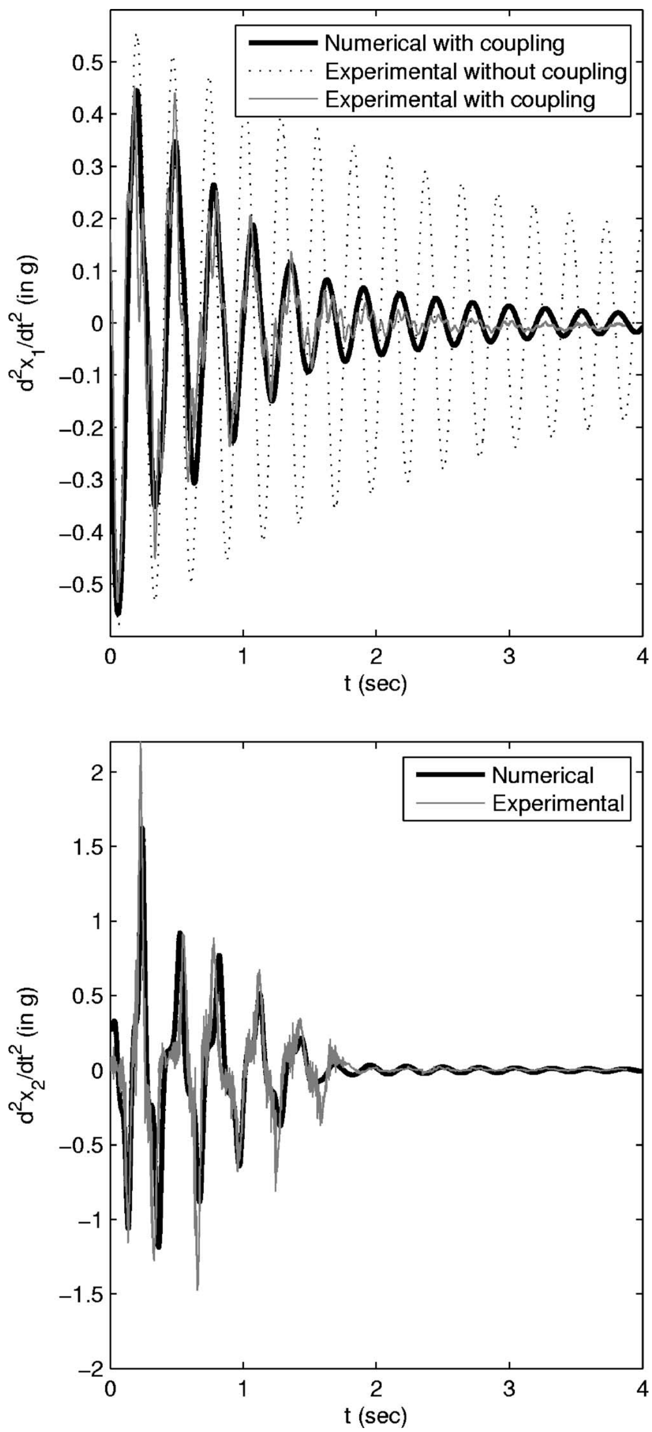


Fig. 15 Comparison of experimental and numerical results

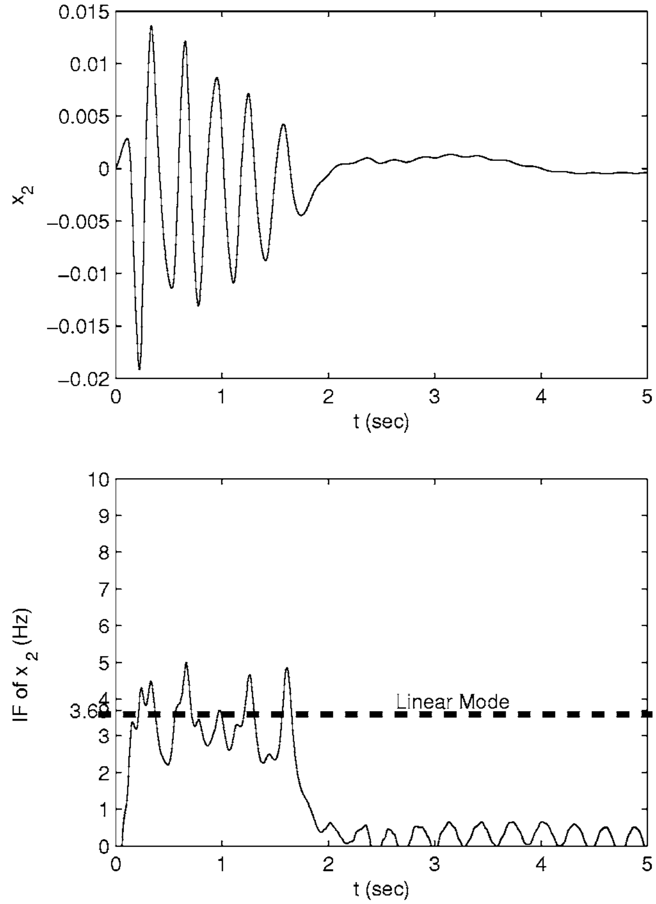


Fig. 16 Resonance capture phenomenon

primary mass $(x_1(t=0)=x_2(t=0)=dx_2/dt(t=0)=0, dx_1/dt(t=0)=0.25)$ with the help of a hammer (the load has been applied to the top floor of the model building), accelerations of free oscillations are measured and plotted, as shown in Fig. 15. In this figure energy pumping occurs. Indeed, there is attenuation of the vibrations of the first mass (compared to the case with no coupling) owing to the resonance of the nonlinear one when coupling is considered.

The experimental results are in good agreement with the numerical integration of System (1) with the previous parameters (by taking only one mode into account in the numerical analysis), as shown in Fig. 15.

We can clearly see that during energy pumping phenomenon, only the first mode is responsible for the total response. It can be underlined that the other modes slightly change the response of x_1 but, as underlined in Ref. [8], only the first mode here is responsible for energy pumping, and other modes of the linear structure are only damped oscillators. Then, when energy pumping occurs, it appears that a resonant capture occurs with the nonlinear oscillator, as shown in Fig. 16. In this figure the instantaneous frequency of experimental signal x_2 has been calculated with the classical Hilbert transform. To obtain x_2, the experimental acceleration has been integrated twice with special filters.

In Fig. 16, the instantaneous frequency of x_2 becomes identical to the instantaneous frequency of the linear mode (3.69 Hz): energy transfer occurs. After $t=2$ s, the nonlinear normal mode is totally destroyed (brutal change of frequency of x_2) resulting in quasi-destruction of vibrations.

Then the experimental function $H(t)$ can be calculated owing to the experimental displacements x_1 and x_2 to see if there is good agreement between the analytical expression (13) and experimental results. Thus experimental, analytical, and numerical $H(t)$ can be compared, as shown in Fig. 17 where the values of the parameters are the same as previously.

Figure 17 shows that the experimental results confirm the analytical expression and numerical simulations since approximation appears to be quite good. Moreover, as shown in Fig. 17, solutions $\varphi_{10}(t)$ and $\varphi_{20}(t)$, which had been introduced to obtain analytical expression (13), are verified experimentally.

4 Conclusion

In the present study, an analytical description of pumping process has been achieved. This analytical description turns out to be

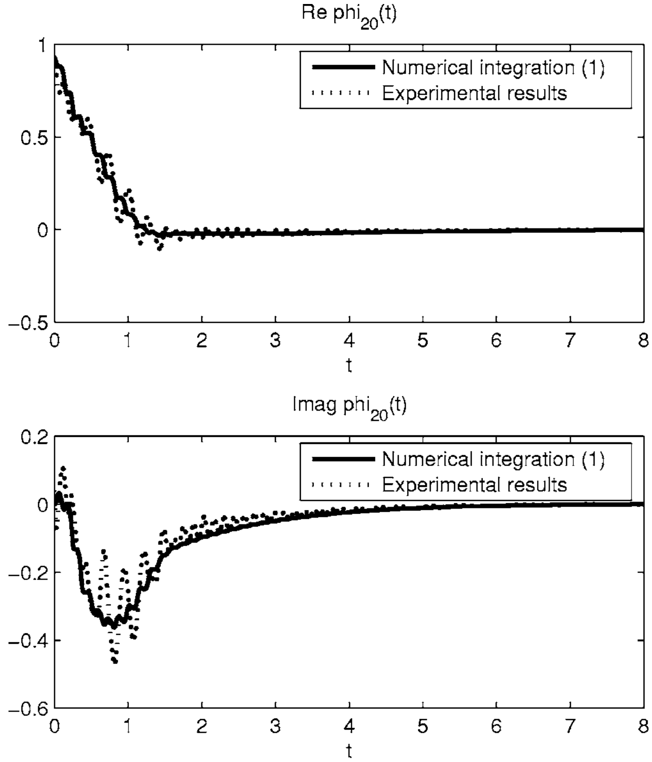


Fig. 17 Comparison of experimental result with the numerical integration Eqs. (1)

efficient for various types of energetic sinks. In spite of similar topological structure, the "optimal" amplitude-phase structures for various types of sink demonstrate significant differences. It must be underlined that design of desirable amplitude–phase curves enables to provide the optimal sink parameters. Then, as the analytical expression seems to be quite good when energy pumping occurs, it is possible to analyze the influence of parameters on efficiency of energy pumping. The aim is to optimize the parameters to obtain the best efficiency. Parameters of the linear primary structure are often given. The aim is to design parameters of the nonlinear added structure to obtain the best attenuation of vibrations of the primary structure. That is why the role of parameters can be analyzed. Since $H(t)$ represents the energy it is possible to observe and analyze the behavior of $H(t)$ if parameters change. In the pumping region analytical results are in very good accordance with the data of numerical simulation (contrary to other types of motion). So, such a coincidence may be considered as a reliable sign of closeness to optimal design. Moreover, computer simulation has confirmed the analytical predictions which have been obtained. Experimental verification has confirmed the analytical expression with good accuracy.

References

[1] Gendelman, O., 2001, "Transition of Energy to a Nonlinear Localized Mode in a Highly Asymmetric System of Two Oscillators," Nonlinear Dyn., **25**, pp. 237–253.

[2] Vakakis, A. F., 2001, "Inducing Passive Nonlinear Energy Sinks in Linear Vibrating Systems," ASME J. Vibr. Acoust., **123**, pp. 324–332.

[3] Vakakis, A. F., and Gendelman, O., 2001, "Energy Pumping in Nonlinear Mechanical Oscillators II: Resonance Capture," ASME J. Appl. Mech., **68**, pp. 42–48.

[4] Gendelman, O., Manevitch, L. I., Vakakis, A. F., and M'Closkey, R., 2001, "Energy Pumping in Nonlinear Mechanical Oscillators I: Dynamics of the Underlying Hamiltonian Systems," ASME J. Appl. Mech., **68**, pp. 34–41.

[5] Manevitch, L. I., Gendelman, O., Musienko, A. I., Vakakis, A. F., and Bergman, L., 2003, "Dynamic Interaction of a Semi-infinite Linear Chain of Coupled Oscillators with a Strongly Nonlinear End Attachment," Physica D, **178**, pp. 1–18.

[6] Vakakis, A. F., Manevitch, L. I., Musienko, A. I., Kerschen, G., and Bergman, L. A., 2005, "Transient Dynamics of a Dispersive Elastic Wave Guide Weakly Coupled to an Essentially Nonlinear End Attachment," Wave Motion, **41**, pp. 109–132.

[7] Gendelman, O. V., Gorlov, D. V., Manevitch, L. I., and Musienko, A. I., 2005, "Dynamics of Coupled Linear and Essentially Nonlinear Oscillators with Substantially Different Masses," J. Sound Vib., **286**, pp. 1–19.

[8] Gourdon, E., and Lamarque, C. H., 2005, "Energy Pumping with Various Nonlinear Structures: Numerical Evidences," Nonlinear Dyn., **40**, pp. 281–307.

[9] Manevitch, L. I., Musienko, A. I., and Lamarque, C. H., 2006, "A New Analytical Approach to Energy Pumping in Strongly Nonhomogeneous 2DOF Systems," Meccanica, 41(5).

[10] Manevitch, L. I., 1999, "Complex Representation of Dynamics of Coupled Nonlinear Oscillators," *Mathematical Models of Nonlinear Excitations, Transfer, Dynamics, and Control in Condensed Systems and Other Media*, Kluwer Academic, Plenum, New York, pp. 269–300.

[11] Manevitch, L. I., 2001, "The Description of Localized Normal Modes in a Chain of Nonlinear Coupled Oscillators Using Complex Variables," Nonlinear Dyn., **25**, pp. 95–109.

[12] McFarland, D., Bergman, L., and Vakakis, A., 2005, "Experimental Study of Non-Linear Energy Pumping Occurring at a Single Fast Frequency," Int. J. Non-Linear Mech., **40**, pp. 891–899.

[13] Gourdon, E., Coutel, S., Lamarque, C. H., and Pernot, S., 2005, "Nonlinear Energy Pumping with Strongly Nonlinear Coupling: Identification of Resonance Captures in Numerical and Experimental Results," *Proceedings of IDETC/CIE 2005*, September 24–28, Long Beach, CA, Paper No. DETC2005-84233.

[14] Manevitch, L. I., Lamarque, C. H., and Gourdon, E., 2007, "Parameters Optimization for Energy Pumping in Strongly Nonhomogeneous 2DOF System," Chaos, Solitons Fractals, **31**(4), pp. 900–911.

[15] Vakakis, A. F., Manevitch, L. I., Gendelman, O., and Bergman, L., 2003, "Dynamics of Linear Discrete Systems Connected to Local, Essentially Nonlinear Attachments," J. Sound Vib., **264**, pp. 559–577.

[16] Vakakis, A. F., Manevitch, L. I., Mikhlin, Yu. V., Pilipchuk, V. N., and Zevin, A. A., 1996, *Normal Modes and Localization in Nonlinear Systems*, Wiley Interscience, New York.

[17] Gendelman, O., Manevitch, L. I., Vakakis, A. F., and Bergman, L., 2003, "A Degenerate Bifurcation Structure in the Dynamics of Coupled Oscillators with Essential Stiffness Nonlinearities," Nonlinear Dyn., **33**, pp. 1–10.

Vibration Characteristics of Multiwalled Carbon Nanotubes Embedded in Elastic Media by a Nonlocal Elastic Shell Model

Renfu Li
Post-Doctoral Fellow

George A. Kardomateas
Professor of Aerospace Engineering
Fellow ASME

Georgia Institute of Technology,
Atlanta, GA 30332-0150

In this paper, the vibrational behavior of the multiwalled carbon nanotubes (MWCNTs) embedded in elastic media is investigated by a nonlocal shell model. The nonlocal shell model is formulated by considering the small length scales effects, the interaction of van der Waals forces between two adjacent tubes and the reaction from the surrounding media, and a set of governing equations of motion for the MWCNTs are accordingly derived. In contrast to the beam models in the literature, which would only predict the resonant frequencies of bending vibrational modes by taking the MWCNT as a whole beam, the current shell model can find the resonant frequencies of three modes being classified as radial, axial, and circumferential for each nanotube of a MWCNT. Big influences from the small length scales and the van der Waals' forces are observed. Among these, noteworthy is the reduction in the radial frequencies due to the van der Waals' force interaction between two adjacent nanotubes. The numerical results also show that when the spring constant k_0 of the surrounding elastic medium reaches a certain value, the lowest resonant frequency of the double walled carbon nanotube drops dramatically. [DOI: 10.1115/1.2722305]

Keywords: vibration, multiwalled, carbon nanotubes, nonlocal shell, small length scale, van der Waals, resonant frequencies

1 Introduction

Carbon nanotubes [1], cylindrical-shaped tubes of seamless graphite with extraordinary electrical and mechanical properties [2–8] potentially have remarkable applications for novel materials or structures such as carbon-nanotube-reinforced composites [9,10] or as individual elements of nanometer-scale devices and sensors [11–13], and have attracted considerable attention worldwide [14–22]. Usually, the properties of carbon nanotubes are evaluated via experiments [2,8] or atomistic and molecular dynamics simulations [14,15]. As has been pointed out in the literature, these experiments are extremely difficult to conduct and control and the molecular dynamics simulations are very time-consuming for large systems [16,17] because of the involvement of internal nanolength scales. Therefore, many authors have made great efforts to extend the classical continuum mechanics to large-sized atomistic systems. Such kind of models can yield reasonable results when the nanotubes are large enough to be viewed as a homogenized material system [17,18]. However, the size of a nanotube is usually very small, maybe a few atoms in diameter, hence it may not be viewed as a continuum medium. Therefore, the small-scale may call the direct application of the classical continuum mechanics model into question [17]. Having realized the limitations of classical continuum models in the study of nanotechnology, some researchers started to use a nonlocal elastic model in their studies [17,18,21,22]. This nonlocal elastic model cannot only include the merits of the classical continuum model but also take the internal small-length scale into account.

In the classical (local) theory of elasticity, the stress at a reference point $x=(x_1,x_2,x_3)$ can be uniquely determined by the strains at that point. However, the nonlocal elasticity [23–25] postulates that the stress at a reference point x in a body not only depends on the strains at x but also depends on the strains of all other points x within the body considered. The stress-strain relations can be written as $\sigma_{ij}(x)=\int_v \alpha(|x'-x|,\tau)c_{ijkl}\epsilon_{kl}(x')dv(x')$, where $\alpha(|x'-x|,\tau)$ is the nonlocal moduli; $\tau=e_0a/l$ with a an internal characteristic length (e.g., lattice parameter, granular distance), l is an external characteristic length (e.g., crack length, wavelength), and e_0 is a constant appropriate to each material; c_{ijkl} is the elastic moduli tensor. For two-dimensional nonlocal elasticity, there exists a differential form for the stress-strain relation,

$$(1-\tau^2 l^2 \nabla^2)\sigma_{ij}=c_{ijkl}\epsilon_{kl} \tag{1}$$

where the operator ∇^2 is the Laplacian operator and has the form $(\partial^2/x_1^2)+(\partial^2/x_2^2)$ in a rectangular coordinate system. Notice that in the nonlocal elasticity the effect of small length scale is considered by incorporating the internal parameter length into the constitutive equation. One may also see that when the internal characteristic length a is neglected, i.e., the particles of a medium are considered to be continuously distributed, then $\tau=0$, and Eq. (1) reduces to the constitutive equation of classical elasticity. Also, it should be noted that, through Eq. (1), the l is cancelled from the rest of the analysis, leaving a and e_0 as the internal characteristic constants.

The vibration of nanotubes is a important subject in the study of nanotechnology since it relates to the electronic and optical properties of MWCNT [26–28]. However, the models in the literature to-date have been exclusively based on beam theory such as these referring to thermal vibration and resonant frequencies [2,8,19,28]. But the topological structure of a nanotube can be viewed as a cylindrical shell, therefore when the multiwalled nanotubes vibrate, not one but three resonant frequencies (radial, axial, and circumferential) can be activated for each nanotube of the multiwalled carbon-nanotube assembly. Since the nanotubes of a MWCNT are coupled via the van der Waals' forces, these three resonant frequencies of each of the nanotubes may be dif-

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received March 11, 2006; final manuscript received January 2, 2007. Review conducted by Yonggang Huang.

ferent from the ones predicted based on isolated nanotubes. The differences in resonant frequencies among the nanotubes of a MWCNT can further affect the electronic and optical properties of the nanostructure. Therefore, the study of the vibration of the MWCNT from a shell-type topological viewpoint has technological significance and in the current research, the vibration of a MWCNT is analyzed by a nonlocal elastic shell model.

We note again that although the internal characteristic length, a, may be on the same scale order as the C–C bond length, this does not mean that the nonlocal elastic theory follows each atom. This important internal characteristic parameter, a, enters into the constitutive relations to reflect the effects of the discrete character [23–25] of the nanostructures when extending the continuum theory to deal with such materials. As has been shown in the literature, experimental and molecular-dynamics simulation methods are often used in most of the studies in understanding the material properties and applications of carbon nanotubes. However, the experiments at the nano-scale are often hard to control and the simulations by molecular-dynamics are difficult to accurately formulate and quite expensive for large-scale atomic systems. Therefore, researchers have attempted to expand the classical continuum mechanics approach to the atomic or molecular-based discrete systems. The classical continuum models are efficient and accurate in computation for a material system in large length scales. But the length scales at nanometers such as in carbon nanotubes are not big enough to homogenize the discrete structure into a continuum. But by using the nonlocal theory, one would harvest the efficiency of classical continuum models and take the nanoscale effects into account at the same time, thus obtaining a satisfactory approximation [24,17,18], etc. These are the advantages of the current theory when comparing with molecular mechanics.

Therefore, in this paper, a nonlocal multiple shell model is developed to investigate the vibration characteristics of multiwalled carbon nanotubes. In this model, not only the terms concerning the van der Waals forces between adjacent nanotubes are incorporated into the Donnell shell model, but also the full nonlocal constitutive relationship is adopted in the derivation of the formulas. Therefore, this model includes both the interactions from the van der Waals forces and the effects from the internal small scales of the nanodevices. Compared with some nonlocal models in the literature for nanotubes subjected to mechanical loading, our model is a comprehensive nonlocal elastic model in the sense that no approximation has been made in the use of the nonlocal elastic constitutive equations and each tube is treated as a shell, not as a one-dimensional column. The work presented in this paper is organized as follows: in Sec. 2 we present the development of a nonlocal elastic shell model for the motion of multiwalled nanotubes; in Sec. 3 we present the analysis for the vibration of DWCNTs under simply-supported end boundary conditions and the derivation of the characteristic equation for the natural frequencies; in Sec. 4 we present numerical results and associated discussions on the vibration behavior of double-walled carbon nanotubes embedded in an elastic medium; finally, conclusions are given in Sec. 5.

2 The Nonlocal Elastic Shell Model of Multiwalled Carbon Nanotubes

Let x, θ, z be the axial, circumferential, and radial coordinates of the nanotube (Fig. 1), respectively. In terms of the the axial, circumferential, and radial displacements of mid-surface, u, v, w, respectively, the strains and displacements of a nanotube have the following relations:

$$\epsilon_{xx}=u_{,x}-z\chi_x \quad \epsilon_{\theta\theta}=\frac{1}{R}v_{,\theta}+\frac{w}{R}-z\chi_\theta$$

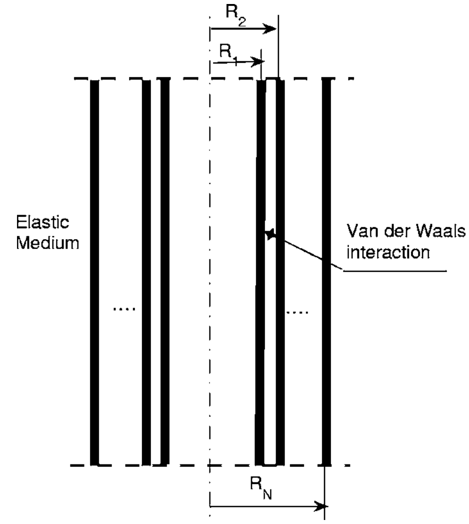


Fig. 1 A shell model of multiwalled nanotubes in an elastic medium

$$\gamma_{x_\theta}=\frac{1}{R}u_{,\theta}+v_{,x}-2z\chi_{x\theta} \tag{2}$$

where R is the mid-surface radius; χ_x, χ_θ, and $\chi_{x\theta}$ are curvatures; the comma denotes differentiation with respect to the corresponding coordinates. We would like to emphasize that our work is a shell theory and not a 3D elasticity solution. Therefore, although three coordinates x, θ, and z are involved in Eq. (2), only two variables x and θ enter into the operators and the problem becomes a 2D problem. Similar handling of the z-direction effects can be also found in Refs. [17,18]. Hence, this problem becomes a 2D problem and the nonlocal theory is then applied onto this 2D problem.

From Eq. (1), the nonlocal stress-strain relations are written as

$$(1-\tau^2l^2\nabla^2)\sigma_{xx}=\frac{E}{1-\nu^2}(\epsilon_{xx}+\nu\epsilon_{\theta\theta}) \tag{3a}$$

$$(1-\tau^2l^2\nabla^2)\sigma_{\theta\theta}=\frac{E}{1-\nu^2}(\epsilon_{\theta\theta}+\nu\epsilon_{xx}) \tag{3b}$$

$$(1-\tau^2l^2\nabla^2)\sigma_{x\theta}=\frac{E}{1+\nu}\gamma_{x\theta} \tag{3c}$$

in which $\nabla^2=(\partial^2/\partial x^2)+(1/R^2)(\partial^2/\partial\theta^2)$, and E, ν, are the elastic modulus and Poisson's ratio, respectively.

From Eqs. (2) and (3), one can, respectively, write the resultant forces

$$(1-\tau^2l^2\nabla^2)N_x=K\left(u_{,x}+\frac{\nu}{R}v_{,\theta}-\nu\frac{w}{R}\right) \tag{4a}$$

$$(1-\tau^2l^2\nabla^2)N_\theta=K\left(\nu u_{,x}+\frac{1}{R}v_{,\theta}-\frac{w}{R}\right) \tag{4b}$$

$$(1-\tau^2l^2\nabla^2)N_{x\theta}=\frac{1-\nu}{2}K\left(\frac{1}{R}u_{,\theta}+v_{,x}\right) \tag{4c}$$

and resultant moments,

$$(1-\tau^2l^2\nabla^2)M_x=D(\chi_x+\nu\chi_\theta) \tag{5a}$$

$$(1-\tau^2l^2\nabla^2)M_\theta=D(\chi_\theta+\nu\chi_x) \tag{5b}$$

$$(1-\tau^2 l^2\nabla^2)M_{x\theta}=(1-\nu)D\chi_{x\theta} \tag{5c}$$

where h is the thickness of the nanotube, $K=Eh/(1-\nu^2)$, and $D=Eh^3/12(1-\nu^2)$.

The governing equations in terms of resultant forces and moments may read as follows:

$$RN_{x,x}+N_{\theta x,\theta}-\rho Rh\ddot{u}=0 \tag{6a}$$

$$N_{\theta,\theta}+RN_{\theta x,x}+Q_\theta-\rho Rh\ddot{v}=0 \tag{6b}$$

$$RQ_{x,x}+Q_{\theta,\theta}+N_\theta+Rp(x,\theta)-\rho Rh\ddot{w}=0 \tag{6c}$$

$$RM_{x,x}+M_{x\theta,\theta}-RQ_x=0 \tag{6d}$$

$$RM_{x\theta,x}-M_{\theta,\theta}+RQ_\theta=0 \tag{6e}$$

If the Donnell assumptions are adopted, then substitution of Eqs. (4) and (5) into Eq. (6) leads to the following nonlocal elastic shell model of a nanotube:

$$L_1(u,v,w)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{u} \tag{7a}$$

$$L_2(u,v,w)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{v} \tag{7b}$$

$$L_3(u,v,w)=(1-\tau^2 l^2\nabla^2)[\hat{\rho}\ddot{w}-p(x,\theta)/K] \tag{7c}$$

where, $k^2=h^2/12R^2$, $\hat{\rho}=\rho(1-\nu^2)/E$, and the operators $L_j(u,v,w)(j=1,2,3)$ are defined as

$$L_1(u,v,w)=u_{,xx}+\frac{1-\nu}{2R^2}u_{,\theta\theta}+\frac{1+\nu}{2R}v_{,x\theta}-\frac{\nu}{R}w_{,x} \tag{8a}$$

$$L_2(u,v,w)=\frac{1+\nu}{2R}u_{,x\theta}+\frac{1-\nu}{2}v_{,xx}+\frac{1}{R^2}v_{,\theta\theta}-\frac{1}{R^2}w_{,\theta} \tag{8b}$$

$$L_3(u,v,w)=\frac{\nu}{R}u_{,x}+\frac{1}{R^2}v_{,\theta}-\frac{1}{R^2}w-k^2R^2\nabla^4 w \tag{8c}$$

The set of governing Eqs. (7) and (8) forms the basis of the nonlocal elastic shell model for the study of the vibration behavior of nanotubes. It is worthy to mention that the applied loading $p(x,\theta)$ plays a very important role in the study of multiwalled nanotubes. This loading usually simulates the van der Waals interactions between two adjacent nanotubes. One may readily see that the effects of the internal characteristic parameter are included in this model as reflected by the terms in the right-hand side of Eqs. (7). When the parameter τ is zero, this model returns to the classical elastic shell model.

Applying Eq. (7) to each of the multiwalled nanotubes, we have for the first nanotube,

$$L_1(u_1,v_1,w_1)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{u}_1 \tag{9a}$$

$$L_2(u_1,v_1,w_1)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{v}_1 \tag{9b}$$

$$L_3(u_1,v_1,w_1)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{w}_1-(1-\tau^2 l^2\nabla^2)\frac{1}{K}p_{12}(x,\theta) \tag{9c}$$

and for the jth wall, $j=2,\ldots,(N-1)$,

$$L_1(u_j,v_j,w_j)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{u}_j \tag{9d}$$

$$L_2(u_j,v_j,w_j)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{v}_j \tag{9e}$$

$$L_3(u_j,v_j,w_j)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{w}_j-(1-\tau^2 l^2\nabla^2)\frac{1}{K}\left[p_{j(j+1)}(x,\theta)-\frac{R_{j-1}}{R_j}p_{(j-1)j}(x,\theta)\right] \tag{9f}$$

and for the Nth wall,

$$L_1(u_N,v_N,w_N)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{u}_N \tag{9g}$$

$$L_2(u_N,v_N,w_N)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{v}_N \tag{9h}$$

$$L_3(u_N,v_N,w_N)=(1-\tau^2 l^2\nabla^2)\hat{\rho}\ddot{w}_N-(1-\tau^2 l^2\nabla^2)\frac{1}{K}\left[p_N(x,\theta)-\frac{R_{N-1}}{R_N}p_{(N-1)N}(x,\theta)\right] \tag{9i}$$

where

$$p_{j(j+1)}(x,\theta)=c[w_{j+1}(x,\theta)-w_j(x,\theta)] \quad j=1,2,\ldots,(N-1) \tag{10a}$$

$$p_N(x,\theta)=-k_0 w_N(x,\theta) \tag{10b}$$

in which, $p_{j(j+1)}(x,s)$ is the interaction pressure exerted on the tube j from the tube $j+1$, while $p_{(j+1)j}(x,s)$ is the interaction pressure exerted on the tube $j+1$ from the tube j; they have the following relationship:

$$R_j p_{j(j+1)}(x,\theta)=-R_{j+1}p_{(j+1)j}(x,\theta) \quad j=1,2,\ldots,N-1; \tag{11}$$

and p_N is the interaction pressure between the outmost tube and the surrounding elastic medium; k_0 is the spring constant of the surrounding elastic medium; and c is the van der Waals interaction coefficient and can be estimated as [22]

$$c=\frac{200}{0.16\pi d^2}\ \text{erg/cm}^2 \quad d=0.142\ \text{nm} \tag{12}$$

where d is a parameter related to the C–C bond length. One may realize that here we are dealing with a linear dynamic problem, so the van der Waals interaction and interaction between the outer tube and elastic surrounding media can be estimated from a linear function of the deflection jump at two points, and the interactions in the tangential direction can be neglected, as discussed in Refs. [18,19]. But for a nonlinear dynamic behavior, the nonlinear higher order terms and effects from the tangential force should be included in these interaction expressions.

3 Double-Walled Carbon Nanotubes

Double-walled carbon nanotubes (DWCNTs) are considered in this section to demonstrate how the nonlocal model can be used to study the dynamics of multiwalled nanotubes. For the tubes, a solution must be periodic in θ $(=s/R)$. Therefore, we can set for $j=1,2$:

$$u_j(x,\theta,t)=\sum_n u_{jn}(x,t)\cos n\pi\theta \tag{13a}$$

$$v_j(x,\theta,t)=\sum_n v_{jn}(x,t)\sin n\pi\theta \tag{13b}$$

$$w_j(x,\theta,t)=\sum_n w_{jn}(x,t)\cos n\pi\theta \tag{13c}$$

Substitution of Eq. (13) into (9) yields the following differential equations with respect to the variables x and t, for $j=1,2$,

$$\tilde{L}_1(u_{jn},v_{jn},w_{jn})=\left[1+\tau^2 l^2\left(\frac{n\pi}{R_j}\right)^2\right]\hat{\rho}\ddot{u}_{jn}-\tau^2 l^2\hat{\rho}\ddot{u}''_{jn} \tag{14a}$$

$$\widetilde{L}_2(u_{jn},v_{jn},w_{jn}) = \left[1+\tau^2 l^2\left(\frac{n\pi}{R_j}\right)^2\right]\hat{\rho}\ddot{v}_{jn} - \tau^2 l^2\hat{\rho}\ddot{v}''_{jn} \quad (14b)$$

$$\widetilde{L}_3(u_{jn},v_{jn},w_{jn}) = \left[1+\tau^2 l^2\left(\frac{n\pi}{R_j}\right)^2\right]\hat{\rho}\ddot{w}_{jn} - \tau^2 l^2\hat{\rho}\ddot{w}''_{jn} + F_j(w_{1n},w_{2n}) \quad (14c)$$

where

$$F_1(w_{1n},w_{2n}) = \frac{c\tau^2 l^2}{K}(w''_{2n}-w''_{1n}) - \frac{c}{K}\left[1+\tau^2 l^2\left(\frac{n\pi}{R_1}\right)^2\right](w_{2n}-w_{1n}) \quad (14d)$$

$$F_2(w_{1n},w_{2n}) = \frac{\tau^2 l^2}{K}\left[k_0 w''_{2n} + c\frac{R_1}{R_2}(w''_{2n}-w''_{1n})\right] - \frac{1}{K}\left[1 + \tau^2 l^2\left(\frac{n\pi}{R_2}\right)^2\right]\left[k_0 w_{2n} + c\frac{R_1}{R_2}(w_{2n}-w_{1n})\right] \quad (14e)$$

where, the superscript $'$ denotes differentiation with respect to the variable x and the operators $\widetilde{L}_1$ $\widetilde{L}_2$ and $\widetilde{L}_3$ are defined as

$$\widetilde{L}_1(u_{jn},v_{jn},w_{jn}) = u''_{jn} - \frac{1-\nu}{2}\left(\frac{n\pi}{R_j}\right)^2 u_{jn} + \frac{1+\nu}{2}\left(\frac{n\pi}{R_j}\right)v'_{jn} - \frac{\nu}{R_j}w'_{jn} \quad (15a)$$

$$\widetilde{L}_2(u_{jn},v_{jn},w_{jn}) = -\frac{1+\nu}{2}\left(\frac{n\pi}{R_j}\right)u'_{jn} + \frac{1-\nu}{2}v''_{jn} - \left(\frac{n\pi}{R_j}\right)^2 v_{jn} + \frac{1}{R_{jn}}\left(\frac{n\pi}{R_j}\right)w_{jn} \quad (15b)$$

$$\widetilde{L}_3(u_{jn},v_{jn},w_{jn}) = \frac{\nu}{R}u'_{jn} + \frac{1}{R_j}\left(\frac{n\pi}{R_j}\right)v_{jn} - k_j^2 R_j^2\left\{w''''_{jn} - 2\left(\frac{n\pi}{R_j}\right)^2 w''_{jn} + \left[\left(\frac{n\pi}{R_j}\right)^4 + \frac{1}{k_j^2 R_j^4}\right]w_{jn}\right\} \quad (15c)$$

3.1 Solution for Simply Supported Tow Ends. For a simple edge-supported DWCNT, the boundary conditions read

$$v = w = N_x = M_x = 0 \text{ at } x=0 \text{ and } x=L \quad (16)$$

or, by using $(4)_1$ and $(5)_1$,

$$v=w=0 \quad u_{,x} + \nu v_{,s} + \nu\frac{w}{R} = 0 \quad w_{,xx} + \nu w_{,ss} - \frac{\nu}{R}v_{,s} = 0 \text{ at } x=0,L \quad (17)$$

A solution satisfying Eqs. (17) may be expressed as

$$u_{jn} = A_1^j \cos\lambda x \cos\omega t \;\; v_{jn} = A_2^j \sin\lambda x\cos\omega t \;\; w_{jn} = A_3^j \sin\lambda x \cos\omega t \;\; j=1,2 \quad (18)$$

in which $\lambda = m\pi/L$.

By substituting Eqs. (18) into the set of Eqs. (14), we obtain a system of six algebraic equations for the unknown constants A_i^j ($i=1,2,3$, $j=1,2$). For nontrivial values of A_i^j, the determinant of the algebraic equations must vanish, which leads to the following characteristic equation for a double-walled carbon nanotubes:

$$\det\begin{bmatrix} \beta_1\hat{\rho}\omega^2 - a_{11} & \frac{1+\nu}{2}n\pi\hat{\lambda}_1 & -\nu\hat{\lambda}_1 & 0 & 0 & 0 \\ \frac{1+\nu}{2}n\pi\hat{\lambda}_1 & \beta_1\hat{\rho}\omega^2 - a_{22} & n\pi & 0 & 0 & 0 \\ -\nu\hat{\lambda}_1 & n\pi & \beta_1\hat{\rho}\omega^2 - a_{33} & 0 & 0 & -\frac{c}{K}\beta_1 \\ 0 & 0 & 0 & \beta_2\hat{\rho}\omega^2 - a_{44} & \frac{1+\nu}{2}n\pi\hat{\lambda}_2 & -\nu\hat{\lambda}_2 \\ 0 & 0 & 0 & \frac{1+\nu}{2}n\pi\hat{\lambda}_2 & \beta_2\hat{\rho}\omega^2 - a_{55} & n\pi \\ 0 & 0 & -\frac{k_0+cR_1/R_2}{K}\beta_2 & -\nu\hat{\lambda}_2 & n\pi & \beta_2\hat{\rho}\omega^2 - a_{66} \end{bmatrix} = 0 \quad (19)$$

where

$$\hat{\lambda}_i = m\pi\frac{R_i}{L} \quad \beta_i = R_i^2 + \tau^2 l^2[\hat{\lambda}_i^2 + (n\pi)^2] \quad i=1,2 \quad (20a)$$

$$a_{11} = \hat{\lambda}_1^2 + \frac{1-\nu}{2}(n\pi)^2 \quad a_{22} = \frac{1-\nu}{2}\hat{\lambda}_1^2 + (n\pi)^2 \quad (20b)$$

$$a_{33} = 1 + k_1^2[\hat{\lambda}_1^2 + (n\pi)^2]^2 - \frac{c}{K}\beta_1 \quad a_{44} = \hat{\lambda}_2^2 + \frac{1-\nu}{2}(n\pi)^2 \quad (20c)$$

$$a_{55} = \frac{1-\nu}{2}\hat{\lambda}_2^2 + (n\pi)^2 \quad a_{66} = 1 + k_2^2[\hat{\lambda}_2^2 + (n\pi)^2]^2 - \frac{k_0 + cR_1/R2}{K}\beta_2 \quad (20d)$$

The characteristic equation for a multiwalled carbon nanotube ($N>2$) can be formulated by employing an analogous procedure. As far as a single-walled carbon nanotube, the corresponding solution is presented in the Appendix.

It is evident that Eq. (19) represents a sixth order frequency equation for the unknown ω^2, and that it has six positive, real roots for a given DWCNT. Hence, each tube of the tow tubes of a DWCNT may vibrate in three different vibrational modes, namely, radial (or bending), longitudinal (or axial), or circumferential (or

Table 1 Comparison of frequencies for a DWCNT and the corresponding to each of the nanotubes SWCNT ω/ω_{10} ($m=n=1$; $L=10R_2$)

SWCNTs	1.0000000, 1.3358769, 2.2041474 $R=0.35$ nm	0.9258549, 1.2368282, 2.0407208 $R=0.79$ nm
DWCNT	0.99834007, 1.3358729, 2.2040182 (inner tube $R_1=0.35$ nm)	0.41628983, 1.243199, 2.0236288 (outer tube $R_2=0.79$ nm)

torsional). The two lowest eigenfrequencies may primarily relate to the radial motions of the inner and outer tube, respectively. Because of the van der Waals' interaction between the inner and outer tubes, one can expect that the values of the frequencies pertaining to each tube would be different than the ones corresponding to a single-walled carbon nanotube with the same geometric sizes and boundary conditions, as illustrated in Table 1. More details will be discussed in the next section.

4 Numerical Results and Discussions

Results of the resonant frequencies of carbon nanotubes (Fig. 1) are presented in this section. The influence of the internal characteristic parameter on the vibration of nanotubes is also demonstrated. Unless otherwise specified, the following properties of the carbon nanotubes are used: the length of a C–C bond is $a=1.42$ nm [18]; $E=742$ GPa, $\nu=0.17$ [9]; the density $\rho=2150$ kg/m^{-3} [2]; the radius of the inner carbon tube $R_1=R_1^o=0.35$ nm and the radius of the outer carbon tube $R_2=R_2^o=0.79$ nm, the thickness of each nanotube $h=0.495(R_2-R_1)$, the length of the nanotubes $L=10\times R_2$; the van der Waals interaction coefficient c can be obtained from Eq. (12); the spring constant from the surrounding elastic medium $k_0=0.01\times c$. Also in the following discussion, Ψ_{I} and Ψ_{VI} are defined, respectively, as the lowest and highest frequencies, normalized by the corresponding values with no consideration of the inner characteristic parameter a.

Listed in Table 1 are the natural frequencies of two SWCNTs and one MWCNT for both ends simply supported. Each of the SWCNT has three natural frequencies, the lowest one corresponding to radial vibration. There are six frequencies for the MWCNT, three for the inner tube vibration and the other three for the outer tube. All values in Table 1 are normalized with the lowest value, ω_{11} of the SWCNT with $R=0.35$ nm. Geometrically, the parameters of the inner tube of the MWCNT are exactly the same as these of the SWCNT with $R=0.35$ nm while those of the outer tube of the MWCNT correspond to a SWCNT with $R=0.79$ nm. But these two nanotubes in the MWCNT are coupled via the van der Waals' force. The coupling interaction definitely has an influence on the vibration of the MWCNT, as reflected in Table 1, the radial natural frequency of the outer tube of the MWCNT reduces 55%, from 0.9258549 to 0.41628983. One can also see that variations of natural frequencies corresponding to the axial and the circumferential vibration exist but not too much because the van der Waals interactions are proportional to the displacement differences in the radial direction. The lowest is usually of primary interest in vibration, therefore, it can be concluded that the van der Waals interaction has a significant influence on the vibration analysis of multiwalled carbon nanotubes. This observation would also suggest that if a model for the vibration of MWCNTs neglects, or inadequately handles the van de Waals interaction between two adjacent nanotubes, it would significantly compromise the accuracy in the predictions of MWCNT properties.

The variation of frequencies with the ratio L/R_2 is shown in Fig. 2 and can be used to justify the current nonlocal shell model for the study of nanodynamics. One can observe that when the ratio L/R_2 is larger than 10, the results of Ψ_{I} and Ψ_{VI} are convergent. This observation is in good agreement with the classical thin shell theory. Of course, for the study of a nanostructure with the geometric parameter $L/R_2<10$, the nonlocal thick shell theory may be needed, and can be derived by a procedure similar to the one employed in the current paper.

Figure 3 shows the results of the variation of frequencies with the ratio $h/(R_2-R_1)$, i.e., the ratio of thickness over the distance in radial direction between the two nanotubes. One can see that the value of Ψ_{VI}, corresponding to the axial frequency of the inner nanotubes is not affected by the varying value of $h/(R_2-R_1)$ and the radial frequency of the outer nanotube, Ψ_{I} does not change either when the ratio $h/(R_2-R_1)$ reaches a certain value such as 0.3 in the current example.

However, the results of the radial natural frequency Ψ_{I} exhibit

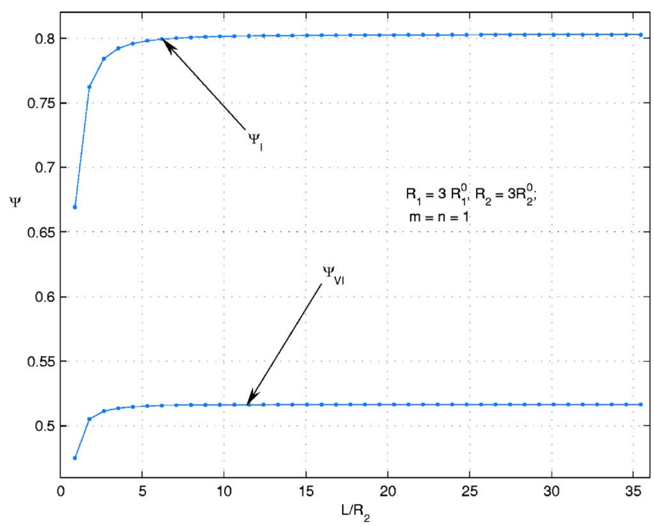


Fig. 2 Variation of frequencies with the aspect ratio, L/R_2

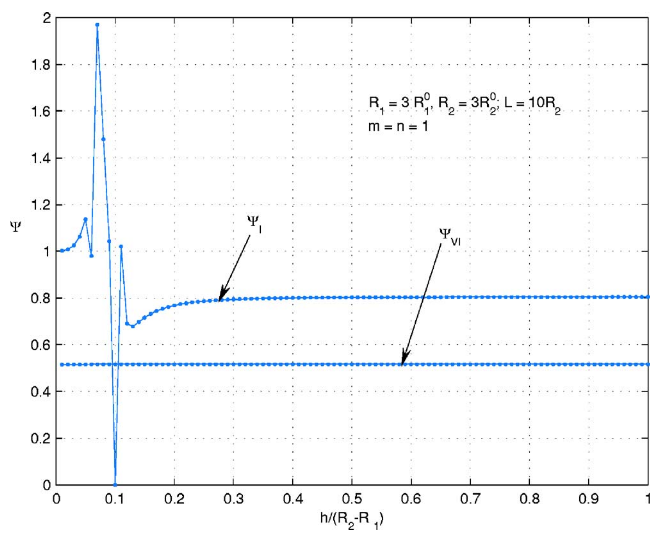


Fig. 3 Variation of frequencies with the ratio of thickness over distance between the nanotubes, $h/(R_2-R_1)$

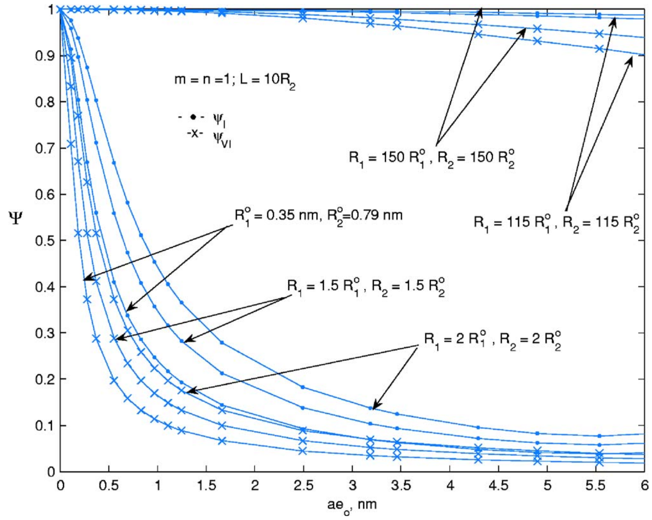


Fig. 4 The influence of the internal characteristic parameter, ae_o of the DWCNTs

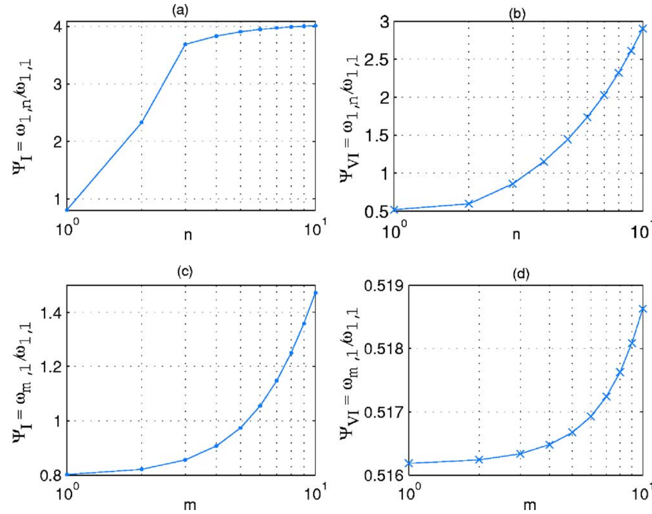


Fig. 5 The variation of the frequencies versus (m,n)

an oscillation for a value of the ratio $h/(R_2-R_1)$ of 0.1. One can further observe that at this point the scale of the nanotube thickness (around 1.3×10^{-11} m) falls into the range of molecular order. But beyond this point, the thickness of the tube enters into the range of nanoscale and the results are shown to be convergent. Therefore, this figure demonstrates that the current shell model would give a reasonable prediction in the study of nanotubes.

The curves in Fig. 4 are used to show the influences of the inner characteristic parameter a and the material constant e_0 on the natural frequencies of this simply edge-supported MWCNT. Two interesting phenomena can be observed in this figure. First, both the Ψ_{I} and the Ψ_{VI} decrease as the ae_0 increases. The increase of ae_0 means that the length of the C–C bond increases for a given nanomedium. Large values of the length of the C–C bond imply a big discontinuity. This observation suggests that a big error could be created by directly applying a classical continuum elastic model in nanostructures. Second, both the Ψ_{I} and the Ψ_{VI} increase as the geometric sizes of the MWCNT increase. Indeed, the Ψ_{I} and Ψ_{VI} almost reach 1 when the sizes of MWCNT reach a certain value, for example $R_1\geqslant0.5\times10^{-8}$ m, $R_2\geqslant1.2\times10^{-8}$ m, and $L\geqslant1.2\times10^{-7}$ m, a range beyond the nanometer order $(1\times10^{-9}\ \mathrm{m}<L<1\times10^{-8}\ \mathrm{m}$ [17]). This observation means that the classical continuum elastic model can give a good prediction if the geometric size of the nanotubes is large enough that the whole structure can be homogenized as a continuum. These two observations conclude that in the range of nanometer order $1\times10^{-9}\ \mathrm{m}<L<1\times10^{-8}$ m, the inner characteristic parameter a, though small, has a significant effect on the natural frequencies predicted by the elastic model and cannot be ignored in the study of nanostructural behavior.

Presented in Fig. 5 are the variations of frequencies Ψ_{I} and Ψ_{VI} versus the modal numbers (m,n). Figures 5(a) and 5(b) are for the case of $m=1$ and varying n, while Figs. 5(c) and 5(d) are for $n=1$ and varying m. Here, the value of m refers to the axial modal number, and n to the circumferential modal number. One can see that the values of Ψ_{I} and Ψ_{VI} are more sensitive to n than to m. Since the beam model theory does not take the circumferential mode into account, this observation suggests that the beam model for the study of the dynamics of nanotubes may be inadequate and may not yield proper results.

Figure 6 shows the effects of the surrounding elastic medium on the frequencies of the MWCNTs. It can be seen that the values of Ψ_{VI} for each mode (for example for $n=1$, $m=1-5$) in the axial direction of vibration, almost do not vary with the variation of the elastic constant $\log[k_0/c]$ of the medium (Fig. 6(b)). But the values of Ψ_{I} corresponding to the basic radial vibration mode $(m=1,n=1)$ and higher modes in the axial direction such as $(m=2,\ldots,5,n=1)$ of the outer nanotube decrease as the value of $\log[k_0/c]$ increases, especially after $\log[k_0/c]\geqslant1$ (Fig. 6(a)). This observation is in good agreement with the results obtained by a beam model simulation [19]. The elastic constant effects on the higher modes in the circumferential direction such as $n=2$ are shown in Fig. 7, in which a tendency similar to the one in Fig. 6 can also been found. The variation of frequencies for higher modes of $n\geqslant2$ can only be predicted by the current shell model. Combining the results in Figs. 6 and 7, we can see that for each higher mode in the radial direction, such as $m=1$, $n=1$, 2, Ψ_{I} decreases while Ψ_{VI} does not change as the $\log[k_0/c]$ increases. It is obvious that the bigger the $\log[k_o/c]$ value, the stiffer the surrounding media. As a limiting case, if the surrounding elastic media is rigid, then no relative motions are possible for the outer nanotube. Therefore, the predictions of a decrease in Ψ_{I} in Figs. 6 and 7 are expected.

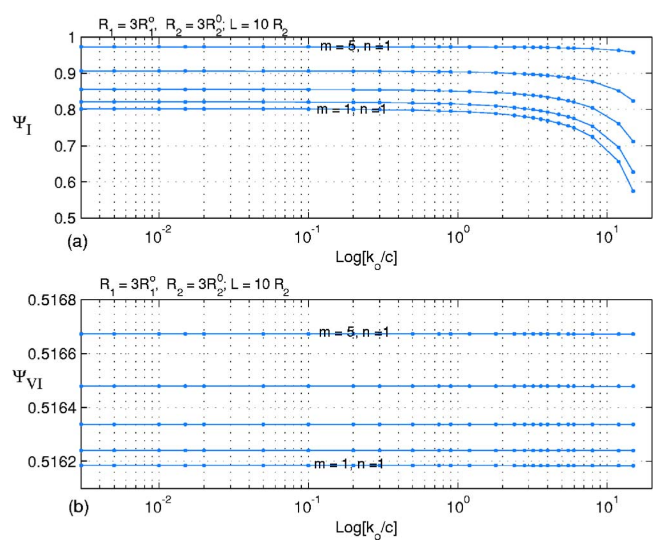


Fig. 6 The influence on the frequencies of the stiffness of the surrounding medium, k_0 versus $(m,n=1)$

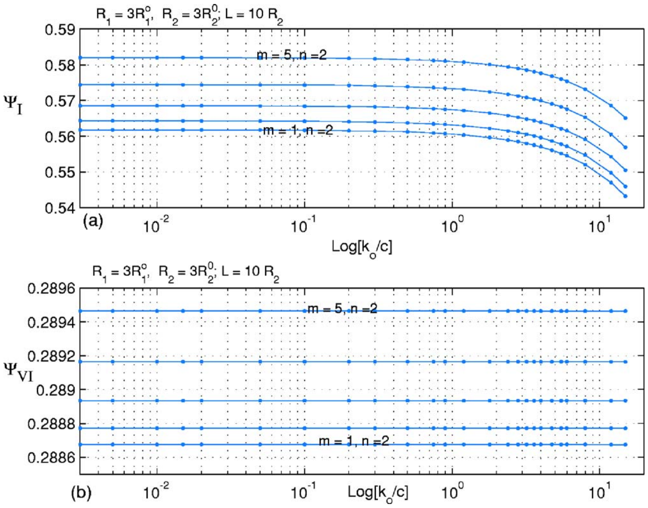


Fig. 7 The influence on the frequencies of the stiffness of the surrounding medium, k_0 versus $(m, n=2)$

5 Conclusions

In the present paper, the dynamic behavior of multiwalled carbon nanotubes embedded in elastic media is studied by a nonlocal shell model. In this model, small nanoscale parameters and the van de Waals' force between two adjacent nanotubes are included. A closed-form solution for the simply supported case is presented. The influences of the small internal parameters of the carbon nanotubes on the natural frequencies are investigated. The effects of the elastic constant of the surrounding medium are also addressed. Compared to a beam model, this shell model has the ability to capture the higher vibration modes in the radial direction. Moreover, the validation of this model is discussed. The observations in this study suggest the following specific conclusions: (1) The small internal parameters of the nanotubes have a significant influence on the natural frequencies of the MWCNTs when the structures are in the order of nanometers. These influences diminish as the geometric sizes of the structures increase. When the order of the geometric size of the structures is beyond the nanometer range, the influence from the small scale parameters could be neglected. (2) The van der Waals' interaction has a significant effect on the radial modal natural frequencies of the outer tubes of the MWCNT. Since this radial frequency is of primary interest in the vibration study of a structure, the van der Waals' force should not neglected in the study of nanodynamics. (3) When the relative stiffness, the ratio of the elastic modulus k_0 of the surrounding media and the van der Waals interaction coefficient c increases, the radial frequencies decrease significantly. This result can provide a guidance for the design of nanocomposites in order to obtain a desirable vibration behavior.

Acknowledgment

The financial support of the Office of Naval Research, Ship Structures and Systems, S&T Division, Grants Nos. N00014-90-J-1995 and N00014-0010323, and the interest and encouragement of the Grant Monitor, Dr. Y.D.S. Rajapakse, are both gratefully acknowledged. We are also grateful to Professor Z.L. Wang of the School of Materials Engineering of the Georgia Institute of Technology for his helpful discussions.

Appendix: A Single-Walled Carbon Nanotube

The solution for the vibration of a single-walled carbon nanotube (SWCNT), simply supported at the ends, can be easily obtained from the results in Sec. 3. The external force considered is only the force from the surrounding elastic medium and it reads

$$p(x,\theta) = -k_0 w(x,\theta) \tag{A1}$$

Following a similar procedure as in Sec. 2, the governing equations can be written as

$$\tilde{L}_1(u_n, v_n, w_n) = \left[1 + \tau^2 l^2 \left(\frac{n\pi}{R}\right)^2\right]\hat{\rho}\ddot{u}_n - \tau^2 l^2 \hat{\rho}\ddot{u}_n'' \tag{A2a}$$

$$\tilde{L}_2(u_n, v_n, w_n) = \left[1 + \tau^2 l^2 \left(\frac{n\pi}{R}\right)^2\right]\hat{\rho}\ddot{v}_n - \tau^2 l^2 \hat{\rho}\ddot{v}_n'' \tag{A2b}$$

$$\tilde{L}_3(u_n, v_n, w_n) = \left[1 + \tau^2 l^2 \left(\frac{n\pi}{R}\right)^2\right]\hat{\rho}\ddot{w}_n - \tau^2 l^2 \hat{\rho} w_n'' + \frac{\tau^2 l^2}{K} k_0 w_n'' - \frac{1}{K}\left[1 + \tau^2 l^2 \left(\frac{n\pi}{R}\right)^2\right] k_0 w_n \tag{A2c}$$

The characteristic equation for natural frequencies of the SWCNT can be according derived as

$$\det\begin{bmatrix} \beta\hat{\rho}\omega^2 - a_1 & \frac{1+\nu}{2} n\pi\hat{\lambda} & -\nu\hat{\lambda} \\ \frac{1+\nu}{2} n\pi\hat{\lambda} & \beta\hat{\rho}\omega^2 - a_2 & n\pi \\ -\nu\hat{\lambda} & n\pi & \beta\hat{\rho}\omega^2 - a_3 \end{bmatrix} = 0 \tag{A3}$$

where

$$\hat{\lambda} = m\pi\frac{R}{L} \quad \beta = R^2 + \tau^2 l^2[\hat{\lambda}^2 + (n\pi)^2] \quad a_1 = \hat{\lambda}^2 + \frac{1-\nu}{2}(n\pi)^2 \tag{A4a}$$

$$a_2 = \frac{1-\nu}{2}\hat{\lambda}^2 + (n\pi)^2 \quad a_3 = 1 + k^2[\hat{\lambda}^2 + (n\pi)^2]^2 - \frac{k_0}{K}\beta \tag{A4b}$$

This model is the accurate solution for single-walled carbon nanotube reinforced composites.

References

[1] Ijima, S., 1991, "Helical Micro-Tubes of Graphite Carbon," Nature (London), **354**, pp. 56–58.
[2] Treacy, M. M. J., Ebbesen, T. W., and Gibson, J. M., 1996, "Exceptionally High Young's Modulus Observed for Individual Carbon Nanotubes," Nature (London), **381**, pp. 678–680.
[3] Falvo, M. R., Clary, G. J., Taylor II, R. M., Chi, V., Brooks Jr., F. P., Washburn, S., and Superfine, R., 1997, "Bending and Buckling of Carbon Nanotubes Under Large Strain," Nature (London), **389**, pp. 582–584.
[4] Roberston, D. H., Brenner, D. W., and Mintimore, J. W., 1992, "Energetics of Nanoscale Graphitic Tubules," Phys. Rev. B, **45**, pp. 12592–12595.
[5] Ruoff, R. S., and Lorents, D. C., 1995, "Mechanical and Thermal Properties of Carbon Nanotubes," Carbon, **33**(7), pp. 925–930.
[6] Sawada, S., and Hamada, N., 1992, "Energetics of Carbon Nano-Tubes," Solid State Commun., **83**, pp. 917–919.
[7] Wong, E. W., Sheehan, P. E., and Lieber, C. M., 1997, "Nanobeam Mechanics: Elasticity, Strength and Toughness of Nanorods and Nanotubes," Science, **277**, pp. 1971–1974.
[8] Poncharal, P., Wang, Z. L., Ugarte, D., and de, Heer W. A., 1999, "Electrostatic Deflections and Electromechanical Resonances of Carbon Nanotubes," Science, **283**, pp. 1513–1516.
[9] Luo, J., and Daniel, I. M., 2003, "Characterization and Modeling of Mechanical Behavior of Polymer/Clay Nano Composites," Compos. Sci. Technol., **63**, pp. 1607–1616.
[10] Frankland, S. J. V., Harik, V. M., Odegard, G. M., Brenner, D. W., and Gates, T. S., 2003, "The Stress-Strain Behavior of Polymer-Nanotube Composites From Molecular Dynamics Simulation," Compos. Sci. Technol., **63**, pp. 1655–1661.
[11] Dai, H., Hafner, J. H., Rinzler, A. G., Colber, T. D., and Smalley, R. E., 1996, "Nanotubes as Nanoprobes In Scanning Probe Microscopy," Nature (London), **384**, pp. 147–150.
[12] Rueckers, T., Kim, K., Joselevich, E., Tseng, G. T., Cheung, C. L., and Lieber, C. M., 2000, "Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing," Science, **289**, pp. 94–97.
[13] Dercke, V., Martel, R., Appendzeller, J., and Avouris, P., 2001, "Carbon Nanotube Inter and Intramolecular Logic Gates," Nano Lett., **1**, pp. 453–456.
[14] Saito, R., Matsuo, R., Kimura, T., Dresselhaus, G., and Dresselhaus, M. S.,

2001, “Anomalous Potential Barrier of Double-Wall Carbon Nanotubes,” Chem. Phys. Lett., **348**(9), pp. 187–193.
[15] Yakobson, B. I., Brabec, C. J., and Bernholc, J., 1996, “Nanomechanics of Carbon Tubes: Instabilities Beyond Linear Response,” Phys. Rev. Lett., **76**, pp. 2511–2514.
[16] Harik, V. M., 2001, “Ranges of Applicability for the Continuum Beam Model in the Mechanics of Carbon Nanotubes and Nanorods,” Solid State Commun., **120**, pp. 331–335.
[17] Peddieson, J., Buchanan, R., and McNitt, R. P., 2003, “Application of Nonlocal Continuum Models to Nanotechnology,” Int. J. Eng. Sci., **41**, pp. 305–312.
[18] Sudak, L. J., 2003, “Column Buckling of Multiwalled Carbon Nanotubes Using Nonlocal Continuum Mechanics,” J. Appl. Phys., **94**, pp. 7281–7287.
[19] Yoon, J., Ru, C. Q., and Mioduchowski, A., 2003, “Vibration of an Embedded Multiwall Carbon Nanotube,” Compos. Sci. Technol., **63**, pp. 1533–1542.
[20] Kim, Y. A., Muramatsu, H., Hayashi, T., Endo, M., Terrones, M., and Dresselhaus, M. S., 2004, “Thermal Stability and Structural Changes of Double-Walled Carbon Nanotubes by Heat Treatment,” Chem. Phys. Lett., **398**, pp. 87–92.
[21] Wang, L., and Hu, H., 2005, “Flexural Wave Propagation in Single-Walled Carbon Nanotubes,” Phys. Rev. B, **71**, p. 195412.
[22] Li, R., and Kardomateas, G. A., 2006, “Thermal Buckling of Multi-Walled Carbon Nanotubes by Nonlocal Elasticity,” ASME J. Appl. Mech., **74**(3), pp. 399–405.
[23] Eringen, A. C., 1972, “Non-Local Polar Elastic Continua,” Int. J. Eng. Sci., **10**, pp. 1–16.
[24] Eringen, A. C., and Edelen, D. G. B., 1972, “On Nonlocal Elasticity,” Int. J. Eng. Sci., **10**, pp. 233–248.
[25] Eringen, A. C., 1983, “On Differential Equations of Nonlocal Elasticity and Solutions of Screw Dislocation and Surface Waves,” J. Appl. Phys., **54**, pp. 4703–4710.
[26] Miyamoto, Y., Saito, S., and Tomanek, D., 2001, “Electronic Interwall Interactions and Charge Redistribution in Multiwall Nanotubes,” Phys. Rev. B, **65**(04), p. 041402.
[27] Lu, W., Dong, J., and Li, Z., 2000, “Optical Properties of Aligned Carbon Nanotube Systems Studied by the Effective-Medium Approximation Method,” Phys. Rev. B, **63**, p. 033401.
[28] Chopra, N. G., and Zettl, A., 1998, “Measurement of the Elastic Modulus of a Multi-Wall Boron Nitride Nanotube,” Solid State Commun., **105**(5), pp. 297–300.

Vladimir Rovenski
Department of Mathematics,
University of Haifa,
Haifa, 31905, Israel
e-mail: rovenski@math.haifa.ac.il

Eugene Harash
e-mail: eharash@techunix.technion.ac.il

Haim Abramovich
e-mail: haim@aerodyne.technion.ac.il

Faculty of Aerospace Engineering,
Technion-I.I.T.,
Haifa, 32000, Israel

Saint-Venant's Problem for Homogeneous Piezoelectric Beams

This paper is devoted to the linear analysis of a slender homogeneous piezoelectric beam that undergoes tip loading. The solution of the Saint-Venant's problem presented in this paper generalizes the known solution for a homogeneous elastic beam. The analytical approach in this study is based on the Saint-Venant's semi-inverse method generalized to electroelasticity, where the stress, strain, and (electrical) displacement components are presented as a set of initially assumed expressions that contain tip parameters, six unknown coefficients, and three pairs of auxiliary (torsion/bending) functions in two variables. These pairs of functions satisfy the so-called coupled Neumann problem (CNP) in the cross-sectional domain. In the limit "elastic" case the CNP transforms to the Neumann problem, for a beam made of a poled piezoceramics the CNP is decomposed into two Neumann problems. The paper develops concepts of the torsion/bending functions, the torsional rigidity and shear center, the tip coupling matrix for a piezoelectric beam. Examples of exact and numerical solutions for elliptical and rectangular beams are presented. [DOI: 10.1115/1.2722315]

1 Introduction

This paper develops the linear analysis of homogeneous piezoelectric beams that undergo tip loading (i.e., the *Saint-Venant's problem*, see [1–5]). The analytical approach in this work is based on the *Saint-Venant's principle and semi-inverse method of solution*, see [6–12], etc., generalized to electroelasticity, see [13–15]. The initially assumed expressions for the stress, strain, and (electrical) displacement components ("solution hypothesis") contain twelve integral tip parameters, six unknown coefficients, and three pairs of torsion/bending functions in two variables (accordingly to two basic cases: *actuating response or sensor response*).

The Solution of the "elastic" Saint-Venant's problem includes three auxiliary (torsion/bending) functions satisfying the Neumann problem in a cross-sectional domain, see [8], pp. 263–264. Analytical solutions of the "piezoelectric" Saint-Venant's problem presented in the paper generalizes the solution for an elastic beam. Namely, the displacement and strain expressions coincide with those of the "elastic" solution, but involved three pairs of auxiliary (torsion/bending) functions satisfying more complicated so-called coupled Neumann problems (CNP). In the limit case the CNP transforms to the "elastic" Neumann problem. For a beam made of poled piezoceramics the CNP is decomposed into two Neumann problems.

The piezoelectric tip loads are divided into two dual groups (each of them contains six tip forces and tip moments) related by the tip coupling matrix.

The present study develops concepts of the torsion/bending functions, torsional rigidity, and the shear center for a monoclinic piezoelectric beam. The entire analytical derivation and solution expressions are symbolically proved by suitable computerized codes, see also [9].

The structure of the work is as follows: Sec. 2 reviews the basic definitions and equations for piezoelectric beam derivation. Section 3 surveys the matrix presentation of anisotropic piezoelectric materials properties: monoclinic, orthotropic, and transtropic. Section 4 introduces a two-dimensional boundary value problem, CNP, playing a central role in the work. Section 5 presents the superposition of all solution components and calculates the six parameters to satisfy the tip conditions. By employing the governing equations (equilibrium, compatibility, existence conditions, etc.), the field equations and the boundary conditions for the unknown six auxiliary functions are established, and the involved functions are determined (and hence, the validity of the preassumed expressions is proved). Section 6 contains illustrative examples of exact solutions for elliptical monoclinic beams and numerical solutions for piezoelectric beams made of transtropic material.

2 Basic Equations for Piezoelectric Beams

To clarify the derivation, the present analysis is applied to anisotropic piezoelectric beams of homogeneous and simply connected cross sections.

Following Saint-Venant, we assume that the principal effects on the elastic field are caused by the force resultant at the beam's ends (Saint-Venant Principle), while the exact dissipation of tractions at the beam's ends are of secondary importance. Hence the tip loads are introduced in an integral manner, and likewise the root boundary conditions. The discussion presented in what follows is adequate mainly for small and finite deformations.

A model of a slender, uniform cylindrical beam with a homogeneous simply connected cross section Ω, its circumference $\partial\Omega$, and the circumferential arc-length coordinate, s, is shown in Figs. 1 and 2. The angle cosines between the normal to $\partial\Omega$ and the x-, y-axes are $\cos(\bar{\mathbf{n}},x)=dx/dn$, $\cos(\bar{\mathbf{n}},y)=dy/dn$, and the following identity holds: $\cos(\bar{\mathbf{n}},x)^2+\cos(\bar{\mathbf{n}},y)^2=1$.

We shall assume that the origin of the coordinate system is placed at the center of the cross-section area and that the x- and y-axes are directed along the principal axes of inertia of the cross section. The above may be written together with the definition of the cross-section area, S_Ω, and the cross-sectional moments of inertia I_y and I_x about the x- and y-axes, respectively, as

$$\int\int_\Omega \{1,x,y,x^2,xy,y^2\}=\{S_\Omega,0,0,I_y,0,I_x\} \tag{1}$$

For Cartesian coordinates x, y, z we denote u_i the *displacement* components, ε_{ij} and σ_{ij} are the *strain* and *stress tensors* components. We also use engineering notation for the stress and strain

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received August 10, 2006; final manuscript received December 19, 2006. Review conducted by Edmundo Corona.

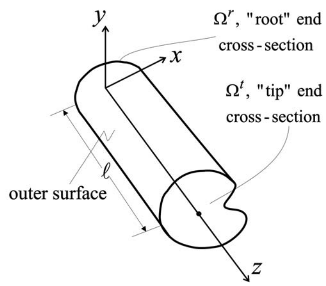


Fig. 1 Notation for a slender beam

tensors $\boldsymbol{\sigma}=(\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5,\sigma_6)$ and $\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\varepsilon_5,\varepsilon_6)$, and displacement vector $\mathbf{u}=(u_1,u_2,u_3)=(u,v,w)$. In the small deformation case, the strain tensor is defined by the Cauchy relations $2\varepsilon_{ij}=u_{i,j}+u_{j,i}$.

One may first derive a separate analysis for a clamped-free piezoelectric beam of the length l (with z-axis) for each of the tip loads, and then summarize all tip effects. The collection of these solutions is referred to as the solutions for the "Saint-Venant's problem." It is shown that, the internal energy stored in the portion of the bar which is beyond a distance z from the loaded end decreases exponentially with the distance z [13].

The most common version of integral clamping conditions that will be denoted as "clamped-I," or "engineering" clamping conditions, is $u_0=v_0=w_0=u_{0,z}=v_{0,z}=\omega_{z0}=0$ at $z=0$. Yet, another possible version for such an integral approximation that will be denoted as "clamped-II" conditions, is $u_0=v_0=w_0=\omega_{x0}=\omega_{y0}=\omega_{z0}=0$ at $z=0$. To study "*sensor response*" the beam is assumed to be acted upon over the tip end cross-section, by three *tip forces* P_x, P_y, P_z and three *tip moments* M_x, M_y, M_z in the coordinate directions, where

$$\{P_z,M_x,-M_y,P_x,M_z\}=\int\int_{\Omega^t}\{\sigma_z,y\sigma_z,x\sigma_z,\sigma_5,\sigma_4,x\sigma_4-y\sigma_5\} \tag{2}$$

To study "*actuating response*," we define the *piezoelectric tip forces* and the *piezoelectric tip moments*,

$$\{P_3^D,M_1^D,-M_2^D,P_1^D,P_2^D,M_3^D\}=\int\int_{\Omega^t}\{D_3,yD_3,xD_3,D_1,D_2,xD_2-yD_1\} \tag{3}$$

The two groups of tip parameters, see Eqs. (2) and (3), are linearly related by the tip coupling matrix that will be defined in what follows.

The linear constitutive equations for a *direct* piezoelectric effect (called *piezoelectric stress equations*) are

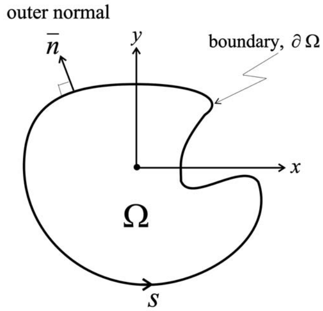


Fig. 2 A cross section

$$\sigma_{ij}=c^e_{ijkl}\varepsilon_{kl}-e_{ijm}E_m \tag{4a}$$

$$D_i=e_{imn}\varepsilon_{mn}+\epsilon^s_{ij}E_j \tag{4b}$$

where $e_{imn}=-e_{inm}$, $\epsilon^s_{ij}=\epsilon^s_{ji}$, (see [16,7,17], etc). The behavior of piezoelectric materials may be also governed by the *converse* piezoelectric effect

$$\varepsilon_{ij}=C^e_{ijkl}\sigma_{kl}+d_{ijm}E_m \tag{5a}$$

$$D_i=d_{imn}\sigma_{mn}+\epsilon^T_{ij}E_j \tag{5b}$$

Here $\mathbf{E}=(E_1,E_2,E_3)$ is the *electric field strength* and $\mathbf{D}=(D_1,D_2,D_3)$ is the *electric displacement*, c^e_{ijkl} (C^e_{ijkl}) is a *stiffness (compliance) tensor at constant electric field*, ϵ^s_{ij} (ϵ^T_{ij}) is the *permittivity at constant strain (stress)*, e_{imn} is a *piezoelectric stress constant*. The third rank tensors d_{mij}, e_{mij} may be simplified using matrix notation to d_{ij},e_{ij} (Voigt's "piezoelectric constants and moduli"), the second rank tensors. The planar isotropy of poled ceramics is expressed in their piezoelectric constants by the equalities $d_{32}=d_{31}$ and $d_{24}=d_{15}$.

For a quasistatic case (and without the body-force components) we obtain three *equilibrium equations*,

$$\sigma_{x,x}+\tau_{xy,y}+\tau_{xz,z}=0 \tag{6a}$$

$$\tau_{xy,x}+\sigma_{y,y}+\tau_{yz,z}=0 \tag{6b}$$

$$\tau_{xz,x}+\tau_{yz,y}+\sigma_{z,z}=0 \tag{6c}$$

Maxwell's equations are reduced to the static dielectric state [7]

$$\operatorname{curl}\mathbf{E}=0\Rightarrow\mathbf{E}=-\operatorname{grad}\Phi \tag{7}$$

$$\operatorname{div}\mathbf{D}=D_{1,x}+D_{2,y}+D_{3,z}=0 \tag{8}$$

where Φ is the *electrical potential*. A unique solution of PDE (6*a*)–(6*c*) and (8) will be found if the initial and boundary conditions of the problem are specified.

The boundary conditions for stress (without surface loads) and electric displacement (the dielectric permittivity of the external medium is much less than of the piezoelectric body) are $\sigma_n=0$ and $D_n=0$ on $\partial\Omega$, or, equivalently,

$$\sigma_x\cos(\bar{\mathbf{n}},x)+\tau_{xy}\cos(\bar{\mathbf{n}},y)=0 \tag{9a}$$

$$\tau_{xy}\cos(\bar{\mathbf{n}},x)+\sigma_y\cos(\bar{\mathbf{n}},y)=0 \tag{9b}$$

$$\tau_{xz}\cos(\bar{\mathbf{n}},x)+\tau_{yz}\cos(\bar{\mathbf{n}},y)=0 \tag{9c}$$

$$D_1\cos(\bar{\mathbf{n}},x)+D_2\cos(\bar{\mathbf{n}},y)=0 \tag{9d}$$

3 Matrix Presentation of Piezoelectric Materials

The independent constants characterizing the electrical and mechanical properties of a piezoelectric material may be considerable reduced in number if the symmetry of the material increased.

For the *monoclinic* type of elastic materials (denoted MON13z [8]) the stiffness matrix $\mathbf{A}=\mathbf{c}^e$ and compliance matrix $\mathbf{a}=\mathbf{C}^e=\mathbf{A}^{-1}$ are

$$\mathbf{A}=\begin{pmatrix} A_{11} & A_{12} & A_{13} & 0 & 0 & A_{16}\\ & A_{22} & A_{23} & 0 & 0 & A_{26}\\ & & A_{33} & 0 & 0 & A_{36}\\ & & & A_{44} & A_{45} & 0\\ & \text{Sym.} & & & A_{55} & 0\\ & & & & & A_{66}\end{pmatrix}$$

$$
\mathbf{a}=\begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & a_{16}\\ & a_{22} & a_{23} & 0 & 0 & a_{26}\\ & & a_{33} & 0 & 0 & a_{36}\\ & & & a_{44} & a_{45} & 0\\ & \text{Sym.} & & & a_{55} & 0\\ & & & & & a_{66} \end{pmatrix}
$$

and the constitutive properties (4*a*) and (4*b*) can be decomposed into two subsystems

$$
\begin{pmatrix} D_1\\ D_2\\ \sigma_4\\ \sigma_5 \end{pmatrix}=\begin{pmatrix} \epsilon^s_{11} & \epsilon^s_{12} & e_{14} & e_{15}\\ \epsilon^s_{12} & \epsilon^s_{22} & e_{24} & e_{25}\\ -e_{14} & -e_{24} & A_{44} & A_{45}\\ -e_{15} & -e_{25} & A_{45} & A_{55} \end{pmatrix}\begin{pmatrix} E_1\\ E_2\\ \varepsilon_4\\ \varepsilon_5 \end{pmatrix}
$$

$$
\begin{pmatrix} D_3\\ \sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_6 \end{pmatrix}=\begin{pmatrix} \epsilon^s_{33} & e_{31} & e_{32} & e_{33} & e_{36}\\ -e_{31} & A_{11} & A_{12} & A_{13} & A_{16}\\ -e_{32} & A_{12} & A_{22} & A_{23} & A_{26}\\ -e_{33} & A_{13} & A_{23} & A_{33} & A_{36}\\ -e_{36} & A_{16} & A_{26} & A_{36} & A_{66} \end{pmatrix}\begin{pmatrix} E_3\\ \varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_6 \end{pmatrix}
$$

Since the **e** constants and *clamped dielectric matrix* $\boldsymbol{\epsilon}^s$ are sometimes unavailable, they can be expressed in terms of *piezoelectric constants* **d** (a second rank tensor) and a *free dielectric matrix* $\boldsymbol{\epsilon}^T$ as

$$
\mathbf{e}=\mathbf{d}\cdot\mathbf{A}\qquad \boldsymbol{\epsilon}^s=\boldsymbol{\epsilon}^T-\mathbf{d}\cdot\mathbf{A}\cdot\mathbf{d}^* \tag{10}
$$

For the monoclinic system: Class 3, *digonal polar*, C_2 [16], the matrices of piezoelectric constants and piezoelectric stress constants have the form

$$
\mathbf{d}=\begin{pmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0\\ 0 & 0 & 0 & d_{24} & d_{25} & 0\\ d_{31} & d_{32} & d_{33} & 0 & 0 & d_{36} \end{pmatrix}
$$

$$
\mathbf{e}=\begin{pmatrix} 0 & 0 & 0 & e_{14} & e_{15} & 0\\ 0 & 0 & 0 & e_{24} & e_{25} & 0\\ e_{31} & e_{32} & e_{33} & 0 & 0 & e_{36} \end{pmatrix} \tag{11}
$$

and the clamped dielectric matrix has the form

$$
\boldsymbol{\epsilon}^s=\begin{pmatrix} \epsilon^s_{11} & \epsilon^s_{12} & 0\\ \epsilon^s_{12} & \epsilon^s_{22} & 0\\ 0 & 0 & \epsilon^s_{33} \end{pmatrix} \tag{12}
$$

Beams of more general anisotropy have less practical use and will not considered in this work.

A simpler class of materials known as *orthotropic materials* (i.e., $A_{45}=A_{16}=A_{26}=A_{36}=0$), is defined by nine independent moduli only so-called *engineering constants* E_{ii}, G_{ij}, and ν_{ij}, as

$$
\mathbf{a}=\begin{pmatrix} \frac{1}{E_{11}} & -\frac{\nu_{21}}{E_{22}} & -\frac{\nu_{31}}{E_{33}} & 0 & 0 & 0\\ -\frac{\nu_{12}}{E_{11}} & \frac{1}{E_{22}} & -\frac{\nu_{32}}{E_{33}} & 0 & 0 & 0\\ -\frac{\nu_{13}}{E_{11}} & -\frac{\nu_{23}}{E_{22}} & \frac{1}{E_{33}} & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{1}{G_{23}} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{G_{13}} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{12}} \end{pmatrix} \tag{13}
$$

For the orthotropic system: class 3, *didigonal polar*, C_{2v} [16], i.e., the piezoelectric constants and piezoelectric stress constants additionally to (11) satisfy $d_{14}=d_{25}=d_{36}=0$, $e_{14}=e_{25}=e_{36}=0$.

Tetragonal material denoted by TRG6z, may be defined by the *six* independent parameters E, E', G, G', ν, ν', so that $E_{11}=E_{22}=E$, $E_{33}=E'$, $G_{23}=G_{13}=G'$, $G_{12}=G$, $\nu_{12}=\nu_{21}=\nu$, $\nu_{13}=\nu_{23}=\nu(E/E')$, $\nu_{31}=\nu_{32}=\nu'$.

Many natural and manmade materials are classified as *transversely isotropic* (*transtropic* for short). The compliance matrix of this transtropic material TI5z is identical to that of the tetragonal material TRG6z, while in addition, a dependency of G on E and ν should be introduced $G=E/[2(1+\nu)]\Leftrightarrow a_{66}=2(a_{11}-a_{12})$. In this case one may use the *ditetragonal polar*, C_{4v}, type of material [16], i.e., (11) with additional restrictions $d_{31}=d_{32}$, $d_{24}=d_{15}$, $e_{31}=e_{32}$, $e_{24}=e_{15}$, $e_{14}=e_{25}=d_{14}=d_{25}=0$. The clamped dielectric matrix has the diagonal view, with $\epsilon^s_{22}=\epsilon^s_{11}$, $\epsilon^s_{12}=0$. Among poled piezoceramics examples of transtropic materials are ceramic B, PZT-2, PZT-4, PZT-5A, PZT-5H, PZT-8, etc. Note that a poled piezoelectric ceramic has only one type of piezoelectric matrix.

Example 1 (transtropic material). In examples of Sec. 4.2 we deal with a beam, made of transtropic material (e.g., a poled piezoelectric ceramic PZT-5A) with characteristics, presented in rational numbers view $E=100/1641\cdot10^{12}\approx0.061\cdot10^{12}$, $E'=10/188\cdot10^{12}\approx0.174\cdot10^{12}$, $\nu=574/1641\approx0.35$, $\nu'=722/1880\approx3.84$, $G'=10/475\cdot10^{12}\approx0.021\cdot10^{12}$, $d_{31}=-171\cdot10^{-12}$, $d_{33}=374\cdot10^{-12}$, $d_{15}=584\cdot10^{-12}$, $\epsilon^T_{11}=1730\varepsilon_0$, $\epsilon^T_{33}=1700\varepsilon_0$ where $\varepsilon_0\approx885/100\cdot10^{-12}$ F/m is the permittivity of vacuum. The system of constitutive relations is decomposed into two subsystems

$$
\begin{pmatrix} D_1\\ D_2\\ \sigma_4\\ \sigma_5 \end{pmatrix}=10^{-12}\begin{pmatrix} 8130 & 0 & 0 & 12.3\\ 0 & 8130 & 12.3 & 0\\ 0 & -12.3 & 0.021 & 0\\ -12.3 & 0 & 0 & 0.021 \end{pmatrix}\begin{pmatrix} E_1\\ E_2\\ \varepsilon_4\\ \varepsilon_5 \end{pmatrix}
$$

$$
\begin{pmatrix} D_3\\ \sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_6 \end{pmatrix}=10^{-12}\begin{pmatrix} 7312 & -5.34 & -5.34 & 15.8 & 0\\ 5.34 & .12 & .075 & .075 & 0\\ 5.34 & .075 & .12 & .075 & 0\\ -15.8 & .075 & .075 & .11 & 0\\ 0 & 0 & 0 & 0 & .026 \end{pmatrix}\begin{pmatrix} E_3\\ \varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_6 \end{pmatrix}
$$

4 A Coupled Neumann Problem

This section contains the formulation and investigation of a two-dimensional boundary value problem that plays a central role in the present study.

We shall first list the elliptical differential operators used in the paper,

$$
\nabla^{(2)}_3=a_{44}\frac{\partial^2}{\partial x^2}-2a_{45}\frac{\partial^2}{\partial x\partial y}+a_{55}\frac{\partial^2}{\partial y^2}\ \text{(see [8])} \tag{14a}
$$

$$
\nabla^{(2)}_e=e_{15}\frac{\partial^2}{\partial x^2}+(e_{25}+e_{14})\frac{\partial^2}{\partial x\partial y}+e_{24}\frac{\partial^2}{\partial y^2} \tag{14b}
$$

$$
\nabla^{(2)}_\epsilon=\epsilon^s_{11}\frac{\partial^2}{\partial x^2}+2\epsilon^s_{12}\frac{\partial^2}{\partial x\partial y}+\epsilon^s_{22}\frac{\partial^2}{\partial y^2} \tag{14c}
$$

The ellipticity of these second order operators follows from the inequalities $a_{44}a_{55}-a^2_{45}>0$, $e_{15}e_{24}-\frac{1}{4}(e_{25}+e_{14})^2>0$, $\epsilon^s_{11}\epsilon^s_{22}-\epsilon^2_{12}>0$. The *boundary differential operators* used in what follows are

$$
D^n_1=(a_{44}\partial_x-a_{45}\partial_y)\cos(\bar{\mathbf{n}},x)+(a_{55}\partial_y-a_{45}\partial_x)\cos(\bar{\mathbf{n}},y)
$$

$$
D^n_\epsilon=(\epsilon^s_{11}\partial_x+\epsilon^s_{12}\partial_y)\cos(\bar{\mathbf{n}},x)+(\epsilon^s_{12}\partial_x+\epsilon^s_{22}\partial_y)\cos(\bar{\mathbf{n}},y)
$$

$$
D^n_e=(e_{15}\partial_x+e_{25}\partial_y)\cos(\bar{\mathbf{n}},x)+(e_{14}\partial_x+e_{24}\partial_y)\cos(\bar{\mathbf{n}},y)
$$

$$\bar{D}_e^n = (e_{15}\partial_x + e_{14}\partial_y)\cos(\bar{\mathbf{n}},x) + (e_{25}\partial_x + e_{24}\partial_y)\cos(\bar{\mathbf{n}},y)$$

Note that $\bar{D}_e^n = D_e^n$ when $e_{14} = e_{25}$.

Let P^Λ, Q^Λ, P_e^φ, Q_e^φ and F^Λ, F_e^Λ be continuous functions in a finite simply connected domain $\Omega \subset \mathbb{R}^2$ with a piecewise smooth boundary $\partial\Omega$. The following boundary value problem plays the central role in our work:

$$\nabla_3^{(2)}\Lambda - a_0 \cdot \nabla_e^{(2)}\Lambda_e = F^\Lambda \quad \text{over } \Omega \tag{15a}$$

$$\nabla_e^{(2)}\Lambda + \nabla_\epsilon^{(2)}\Lambda_e = F_e^\Lambda \quad \text{over } \Omega \tag{15b}$$

$$D_1^n\Lambda - a_0 \cdot D_e^n\Lambda_e = P^\Lambda \cos(\bar{\mathbf{n}},x) + Q^\Lambda \cos(\bar{\mathbf{n}},y) \quad \text{on } \partial\Omega \tag{15c}$$

$$\bar{D}_e^n\Lambda + D_\epsilon^n\Lambda_e = P_e^\Lambda \cos(\bar{\mathbf{n}},x) + Q_e^\Lambda \cos(\bar{\mathbf{n}},y) \quad \text{on } \partial\Omega \tag{15d}$$

We also assume the initial values for two unknown functions Λ, Λ_e are

$$\Lambda(0,0) = \Lambda_e(0,0) = 0 \tag{16}$$

We call this boundary value problem a *coupled Neumann problem* (CNP) in Ω for the functions Λ, Λ_e. The CNP generalizes Neumann problem that serves for monoclinic elastic beam, i.e., $e_{ij} = \epsilon_{ij}^s = 0$, see [8]

$$\nabla_3^{(2)}\Lambda = F^\Lambda \quad \text{over } \Omega \tag{17a}$$

$$D_1^n\Lambda = P^\Lambda \cos(\bar{\mathbf{n}},x) + Q^\Lambda \cos(\bar{\mathbf{n}},y) \quad \text{on } \partial\Omega \tag{17b}$$

The next statement is similar to the necessary condition for a classical Neumann problem.

Proposition 2. *The necessary conditions for the solution existence of the CNP (15a)–(15d) and (16) are*

$$\int\int_\Omega (P_{,x}^\Lambda + Q_{,y}^\Lambda - F^\Lambda) = \int\int_\Omega (P_{e,x}^\Lambda + Q_{e,y}^\Lambda - F_e^\Lambda) = 0 \tag{18}$$

The proof of Proposition 2 is based on the following identities for a smooth function Λ:

$$\int\int_\Omega \{\nabla_3^{(2)}, \nabla_e^{(2)}, \nabla_e^{(2)}, \nabla_\epsilon^{(2)}\}\Lambda = \oint_{\partial\Omega} \{D_1^n, D_e^n, \bar{D}_e^n, D_\epsilon^n\}\Lambda$$

Following the classical scheme (of Neumann problem) one may prove that condition (18) is SUFFICIENT for the existence of a solution of a CNP, and the solution is unique.

Remark 3. In the case of the transtropic piezoelectric material, the differential operators above are

$$\nabla_3^{(2)} = a_{44} \cdot \nabla^{(2)} \quad \nabla_e^{(2)} = e_{15} \cdot \nabla^{(2)} \quad \nabla_\epsilon^{(2)} = \epsilon_{11}^s \cdot \nabla^{(2)}$$

$$D_1^n = a_{44} \cdot \frac{d}{dn} \quad D_e^n = \bar{D}_e^n = e_{15} \cdot \frac{d}{dn} \quad D_\epsilon^n = \epsilon_{11}^s \cdot \frac{d}{dn}$$

where $d/dn = (\partial/\partial x)\cos(\bar{\mathbf{n}},x) + (\partial/\partial y)\cos(\bar{\mathbf{n}},y)$ is the *(geometrical) normal derivative* and $\nabla^{(2)} = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$ is the *Laplace operator*. Hence the CNP (15*a*)–(15*d*) and (16) is reduced to a simpler system. Namely,

$$\Lambda = \frac{1}{\Delta}(\epsilon_{11}^s\tilde{\Lambda} + d_{15}\tilde{\Lambda}_e) \quad \Lambda_e = \frac{1}{\Delta}(-e_{15}\tilde{\Lambda} + \tilde{\Lambda}_e)$$

where $\Delta = \epsilon_{11}^s + d_{15}e_{15} > 0$ and the functions $\tilde{\Lambda}$ and $\tilde{\Lambda}_e$ are the (unique) solutions of the Neumann problems

$$\nabla^{(2)}\tilde{\Lambda} = \frac{1}{a_{44}}F^\Lambda \quad \text{over } \Omega$$

$$\frac{d}{dn}\tilde{\Lambda} = \frac{1}{a_{44}}[P^\Lambda \cos(\bar{\mathbf{n}},x) + Q^\Lambda \cos(\bar{\mathbf{n}},y)] \quad \text{on } \partial\Omega$$

$$\tilde{\Lambda}(0,0) = 0,$$

$$\nabla^{(2)}\tilde{\Lambda}_e = F_e^\Lambda \quad \text{over } \Omega$$

$$\frac{d}{dn}\tilde{\Lambda}_e = P_e^\Lambda \cos(\bar{\mathbf{n}},x) + Q_e^\Lambda \cos(\bar{\mathbf{n}},y) \quad \text{on } \partial\Omega$$

$$\tilde{\Lambda}_e(0,0) = 0$$

5 The Tip Loading Effects

5.1 Torsional Moment. The analytical solution originates from the displacement hypothesis

$$u = -\alpha_6 yz \quad v = \alpha_6 xz \quad w = \alpha_6\varphi(x,y)$$

Hence we assume that the rotation $\omega_z = \alpha_6 z$ is a linear function of z. The warping, $w(x,y)$ is a product of the beam twist $\theta = \alpha_6$ (the warping amplitude), and an unknown function, $\varphi(x,y)$, that physically represents the shape of the out-of-plane warping (usually termed as the *torsion function*). The constant α_6 (not known as yet) plays the role of the piezoelectric beam *torsional rigidity* (i.e., the factor by which we divide a torsional moment of a beam to obtain the twist per unit length). Assume that electrical potential function does not depend on z, i.e., $\Phi = -\alpha_6\varphi_6(x,y)$, $E_3(x,y) = 0$, and

$$E_1(x,y) = \alpha_6\varphi_{6,x}(x,y) \quad E_2(x,y) = \alpha_6\varphi_{6,y}(x,y)$$

The above displacement assumption yields a state where all strain components vanish (e.g., $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_6$) except for

$$\varepsilon_4 = \alpha_6(\varphi_{,y} + x) \quad \varepsilon_5 = \alpha_6(\varphi_{,x} - y)$$

Let $a_0 = a_{44}a_{55} - a_{45}^2 > 0$. Hence the stress components are $\sigma_i = 0$ $(i = 1,2,3,6)$, and

$$\frac{1}{\alpha_6}\sigma_4 = -\frac{1}{a_0}[Q^\varphi + a_{45}\varphi_{,x} - a_{55}\varphi_{,y}] - e_{14}\varphi_{6,x} - e_{24}\varphi_{6,y}$$

$$\frac{1}{\alpha_6}\sigma_5 = -\frac{1}{a_0}[P^\varphi - a_{44}\varphi_{,x} + a_{45}\varphi_{,y}] - e_{15}\varphi_{6,x} - e_{25}\varphi_{6,y}$$

where $P^\varphi = a_{45}x + a_{44}y$, $Q^\varphi = -a_{55}x - a_{45}y$, see [8]. The electrical displacement components are $D_3 = 0$ and

$$\frac{1}{\alpha_6}D_1 = \epsilon_{11}^s\varphi_{6,x} + \epsilon_{12}^s\varphi_{6,y} + [e_{14}(\varphi_{,y} + x) + e_{15}(\varphi_{,x} - y)]$$

$$\frac{1}{\alpha_6}D_2 = \epsilon_{12}^s\varphi_{6,x} + \epsilon_{22}^s\varphi_{6,y} + [e_{24}(\varphi_{,y} + x) + e_{25}(\varphi_{,x} - y)]$$

The first two equilibrium equations (6*a*) and (6*b*) are satisfied identically as well, while the third one and the equation div $\mathbf{D} = 0$, yields (15*a*) and (15*b*) with $\Lambda = \varphi$, $\Lambda_e = \varphi_6$ and

$$F^\varphi = 0 \quad F_e^\varphi = e_{25} - e_{14}$$

From (9*c*) and (9*d*) we obtain Neumann-type boundary conditions (15*c*) and (15*d*) where $P_e^\varphi = e_{15}y - e_{14}x$, $Q_e^\varphi = e_{25}y - e_{24}x$.

The other cases of tip loading are considered similarly.

5.2 Summarizing Solution Components. We summarize here the so-called "Saint-Venant's problem," namely, the effect of the *tip forces* P_x, P_y, P_z, the *tip moments* M_x, M_y, M_z, the *piezoelectric tip forces* P_1^D, P_2^D, P_3^D, and the *piezoelectric tip moments* M_1^D, M_2^D, M_3^D.

It should be noted that the solution expressions are "symmetric" under the following parameter interchange: $1 \leftrightarrow 2$, $4 \leftrightarrow 5$, $x \leftrightarrow y$, $u \leftrightarrow v$, $P \leftrightarrow Q$, etc. One may obtain symmetric expressions using operation "Sym." For example, $u \leftrightarrow v$ means $u = \mathrm{Sym}(v)$ and v

$=\mathrm{Sym}(u)$. Also, $Q^{\chi_1} \leftrightarrow P^{\chi_2}$ and $D_2=\mathrm{Sym}(D_1)$.

The stress components are $\sigma_j=0$, $j=1,2,6$, and

$$\sigma_3 = \frac{1}{a_{33}}[\alpha_3 - \alpha_1 x - \alpha_2 y - (l-z)(\alpha_4 x + \alpha_5 y)]$$

$$\sigma_4 = \frac{\alpha_4}{a_0}(Q^{\chi_1} + a_{45}\chi_{1,x} - a_{55}\chi_{1,y}) + \frac{\alpha_5}{a_0}(Q^{\chi_2} + a_{45}\chi_{2,x} - a_{55}\chi_{2,y})$$
$$- \frac{\alpha_6}{a_0}(Q^{\varphi} + a_{45}\varphi_{,x} - a_{55}\varphi_{,y}) - \alpha_6(e_{14}\varphi_{6,x} + e_{24}\varphi_{6,y})$$
$$+ \alpha_4(e_{14}\varphi_{4,x} + e_{24}\varphi_{4,y}) + \alpha_5(e_{14}\varphi_{5,x} + e_{24}\varphi_{5,y})$$

$$\sigma_5 = \mathrm{Sym}(\sigma_4) \tag{19}$$

These expressions are based on six nontrivial functions: the electrical potential functions $\varphi_i(x,y)$ $(i=4,5,6)$, the torsion function $\varphi(x,y)$, and the bending functions $\chi_1(x,y)$ and $\chi_2(x,y)$, that should be solved for each specific nonhomogeneous cross-section geometry. Note that for the MON13z material under discussion, the in-plane stress components $(\sigma_1,\sigma_2,\sigma_6)$ vanish for all tip loads. The strain components are obtained from the stress-strain relationships (are similar to the solution for elastic beam in [8], pp. 263–264)

$$\varepsilon_i = \frac{a_{i3}}{a_{33}}[\alpha_3 - \alpha_1 x - \alpha_2 y - (l-z)(\alpha_4 x + \alpha_5 y)] \; i=1,2,3,6$$

$$\varepsilon_4 = \alpha_4\left(\frac{a_{36}x^2 + 2(a_{23}-2a_{33})xy}{2a_{33}} - \chi_{1,y}\right)$$
$$+ \alpha_5\left(\frac{a_{23}y^2 - (a_{13}+2a_{33})x^2}{2a_{33}} - \chi_{2,y}\right) + \alpha_6(\varphi_{,y} + x)$$

$$\varepsilon_5 = \mathrm{Sym}(\varepsilon_4) \tag{20}$$

The corresponding displacements are (see [8] for an elastic beam)

$$u = \alpha_1\left(\frac{a_{23}}{2a_{33}}y^2 - \frac{a_{13}}{2a_{33}}x^2 + \frac{1}{2}z^2\right) - \alpha_2\left(\frac{a_{13}}{a_{33}}xy + \frac{a_{36}}{2a_{33}}y^2\right)$$
$$+ \alpha_3\left(\frac{a_{13}}{a_{33}}x + \frac{a_{36}}{2a_{33}}y\right) - \alpha_6 yz + \alpha_4\left[(l-z)\left(-\frac{a_{13}}{2a_{33}}x^2 + \frac{a_{23}}{2a_{33}}y^2\right)\right.$$
$$\left.+ \frac{z^2}{2}\left(l - \frac{z}{3}\right)\right] - \alpha_5(l-z)\left(\frac{a_{13}}{a_{33}}xy + \frac{a_{36}}{2a_{33}}y^2\right)$$
$$+ u^0 - \omega_z^0 y + \omega_y^0 z$$

$$v = \mathrm{Sym}(u)$$

$$w = (\alpha_3 - \alpha_1 x - \alpha_2 y)z - \alpha_4\left[\left(l - \frac{z}{2}\right)xz + xy^2 + \chi_1\right]$$
$$- \alpha_5\left[\left(l - \frac{z}{2}\right)yz + x^2 y + \chi_2\right] + \alpha_6\varphi + w^0 - \omega_y^0 x + \omega_x^0 y$$

where u^0, v^0, w^0 are *rigid displacements*, while $\omega_x^0, \omega_y^0, \omega_z^0$ are *rigid rotations*. For a clamped-free beam with clamped-I type of geometrical conditions at its root we obtain $u^0=v^0=w^0=\omega_x^0=\omega_y^0=\omega_z^0=0$. For a clamped-free beam with clamped-II type of geometrical conditions at its root we apply

$$u^0 = v^0 = w^0 = \omega_z^0 = 0 \;\; \omega_x^0 = \mathrm{Sym}(\omega_y^0)$$

$$2\omega_y^0 = -\alpha_4\chi_{1,x}(0,0) - \alpha_5\chi_{2,x}(0,0) - \alpha_6\varphi_{,x}(0,0)$$

The electrical potential function and the electric field strength are

$$\Phi = \alpha_4\varphi_4(x,y) + \alpha_5\varphi_5(x,y) - \alpha_6\varphi_6(x,y) + \mathrm{const}$$

$$E_1 = \alpha_6\varphi_{6,x} - \alpha_4\varphi_{4,x} - \alpha_5\varphi_{5,x}$$

$$E_2 = \alpha_6\varphi_{6,y} - \alpha_4\varphi_{4,y} - \alpha_5\varphi_{5,y}, E_3 = 0$$

The electrical displacement components are

$$D_1 = \alpha_4(P_4^e - e_{14}\chi_{1,y} - e_{15}\chi_{1,x}) + \alpha_5(P_5^e - e_{15}\chi_{2,x} - e_{14}\chi_{2,y})$$
$$+ \alpha_6[e_{14}(\varphi_{,y} + x) + e_{15}(\varphi_{,x} - y)] + \alpha_6(\epsilon_{11}^s\varphi_{6,x} + \epsilon_{12}^s\varphi_{6,y})$$
$$- \alpha_4(\epsilon_{11}^s\varphi_{4,x} + \epsilon_{12}^s\varphi_{4,y}) - \alpha_5(\epsilon_{11}^s\varphi_{5,x} + \epsilon_{12}^s\varphi_{5,y})$$

$$D_3 = [\alpha_3 - \alpha_1 x - \alpha_2 y - (\alpha_4 x + \alpha_5 y)(l-z)]\frac{d_{33}}{a_{33}}$$

while $D_2=\mathrm{Sym}(D_1)$. Recall that $d_{33}=\Sigma_i e_{3i}a_{i3}$. The formulas for α_i, calculated in what follows from the tip conditions, are similar to those for an elastic beam in [8].

5.3 Verification of Solution Hypothesis. As was shown, the pair of functions, (φ,φ_6) satisfies a CNP (15*a*)–(15*d*). The other pairs of functions, (χ_1,φ_4) and (χ_2,φ_5), also satisfy a CNP (15*a*)–(15*d*) and (16), where

$$F^{\chi_1} = \frac{a_0 + a_{13}a_{44} + a_{23}a_{55} - a_{36}a_{45} - 2a_{33}a_{55}}{a_{33}}x + 4a_{45}y$$

$$F_4^e = \frac{1}{a_{33}}[y(e_{14}(a_{23} - 2a_{33}) - e_{25}(a_{23} + 2a_{33})) + x(d_{33} + e_{14}a_{36}$$
$$+ e_{15}a_{13} + e_{24}(a_{23} - 2a_{33}))]$$

$$F^{\chi_2} = \mathrm{Sym}(F^{\chi_1}) \quad F_5^e = \mathrm{Sym}(F_4^e)$$

$$P^{\chi_1} = \frac{a_{44}a_{13} - a_{36}a_{45}}{2a_{33}}x^2 + \frac{a_{45}(2a_{33} - a_{23})}{a_{33}}xy - \frac{a_{44}(a_{23} + 2a_{33})}{2a_{33}}y^2$$

$$Q^{\chi_1} = \frac{a_{55}a_{36} - a_{45}a_{13}}{2a_{33}}x^2 + \frac{a_{55}(a_{23} - 2a_{33})}{a_{33}}xy + \frac{a_{45}(a_{23} + 2a_{33})}{2a_{33}}y^2$$

$$P_4^e = e_{14}\frac{a_{36}x^2 + 2(a_{23} - 2a_{33})xy}{2a_{33}} + e_{15}\frac{a_{13}x^2 - (a_{23} + 2a_{33})y^2}{2a_{33}}$$

$$Q_4^e = e_{24}\frac{a_{36}x^2 + 2(a_{23} - 2a_{33})xy}{2a_{33}} + e_{25}\frac{a_{13}x^2 - (a_{23} + 2a_{33})y^2}{2a_{33}}$$

and $P^{\chi_2}=\mathrm{Sym}(Q^{\chi_1})$, $Q^{\chi_2}=\mathrm{Sym}(P^{\chi_1})$, $Q_5^e=\mathrm{Sym}(P_4^e)$, $P_5^e=\mathrm{Sym}(Q_4^e)$. In view of

$$P_{,x}^{\varphi} + Q_{,y}^{\varphi} = 0 \quad P_{e,x}^{\varphi} + Q_{e,y}^{\varphi} = e_{25} - e_{14} = F_e^{\varphi}$$

the solution existence conditions (18) of the CNP for the pair (φ,φ_6) are satisfied. Because

$$P_{,x}^{\chi_1} + Q_{,y}^{\chi_1} - F^{\chi_1} \quad P_{4,x}^e + Q_{4,y}^e - F_4^e$$

$$P_{,x}^{\chi_2} + Q_{,y}^{\chi_2} - F^{\chi_2} \quad P_{5,x}^e + Q_{5,y}^e - F_5^e$$

are linear functions of x,y, the solution existence conditions (18) of the CNP for the functions (χ_1,φ_4) and (χ_2,φ_5) are satisfied.

To gain a global look at the above presented solutions for the M_z, P_x, and P_y tip loads, one may adopt the stress expressions, see Eqs. (19), as the initial stress hypothesis and carry out a verification procedure that proves that these proposed expressions constitute a valid solution of the Saint-Venant's problem. This procedure may be summarized as follows:

(a) We first note that the equilibrium Eqs. (6*a*) and (6*b*), are satisfied identically. From the third equilibrium Eq. (6*c*), we deduce the field equations of the type (15*a*) (in Ω) for the functions φ, φ_i, χ_1, and χ_2. From Eqs. (8) we deduce the field equations of the type (15*b*) (in Ω) for the functions φ, φ_i, χ_1, and χ_2.

(b) Similar to the above, the boundary conditions, Eqs. (9*a*) and (9*b*), are satisfied identically (since $E_3=0$). From the boundary conditions, Eqs. (9*c*) and (9*f*), we deduce the boundary conditions of the type (15*c*) and (15*d*) (on the contour $\partial\Omega$) for the harmonic functions φ, φ_i, χ_1, χ_2.

(c) By employing Eqs. (20) one may verify that all compatibility equations, are satisfied identically.

Remark 4. In the case of transtropic piezoelectric material, the functions $\varphi - d_{24}\varphi_6$, $\varphi + (\epsilon^s_{11}/e_{15})\varphi_6$ satisfy the same Neumann problem, see Remark 3:

$$\nabla^{(2)}\tilde{\varphi} = 0 \quad \text{over } \Omega$$

$$\frac{d}{dn}\tilde{\varphi} = y\cos(\bar{\mathbf{n}},x) - x\cos(\bar{\mathbf{n}},y) \quad \text{on } \partial\Omega$$

$$\tilde{\varphi}(0,0) = 0$$

From the uniqueness of the solution for Neumann problem, we conclude that the functions $\varphi - d_{24}\varphi_6$ and $\varphi + (\epsilon^s_{11}/e_{15})\varphi_6$ coincide, i.e., $(\epsilon^s_{11} + d_{24}e_{15})\varphi_6 = 0$. Since the factor $\epsilon^s_{11} + d_{24}e_{15}$ is positive, we get $\varphi_6 = 0$.

The CNP for χ_1, φ_4 is decomposed into two independent Neumann problems for the functions $\tilde{\chi}_1 = \chi_1 - d_{24}\cdot\varphi_4$ and $\tilde{\varphi}_4 = e_{15}\chi_1 + \epsilon^s_{11}\cdot\varphi_4$, respectively. Similarly for χ_2, φ_5.

5.4 Fulfilling the Tip Conditions.

The following integral quantities will be shown to play the role of the piezoelectric beam *torsional rigidity* of a beam,

$$D_\varphi = \int\int_\Omega \frac{1}{a_0}[(yP^\varphi - xQ^\varphi) - (P^\varphi\varphi_{,x} + Q^\varphi\varphi_{,y})] \tag{21}$$

$$D^e_\varphi = \int\int_\Omega [(e_{24}x - e_{14}y)(\varphi_{,y} + x) + (e_{25}x - e_{15}y)(\varphi_{,x} - y)] \tag{22}$$

$$D^e_i = \int\int_\Omega [(e_{14}x - e_{15}y)\varphi_{i,x} + (e_{24}x - e_{25}y)\varphi_{i,y}] \tag{23}$$

$$D^\epsilon_i = -\int\int_\Omega [(\epsilon^s_{12}x - \epsilon^s_{11}y)\varphi_{i,x} + (\epsilon^s_{22}x - \epsilon^s_{12}y)\varphi_{i,y}] \tag{24}$$

where $i=4,5,6$. In addition, we have

$$D_{\chi_i} = \int\int_\Omega \frac{1}{a_0}[yP^{\chi_i} - xQ^{\chi_i} - (a_{45}x + a_{44}y)\chi_{i,x} + (a_{55}x + a_{45}y)\chi_{i,y}] \; i=1,2 \tag{25}$$

$$D^e_{\chi_1} = \int\int_\Omega \Bigg[(e_{24}x - e_{14}y)\left(\frac{a_{36}x^2 + 2(a_{23} - 2a_{33})xy}{2a_{33}} - \chi_{1,y}\right) + (e_{25}x - e_{15}y)\left(\frac{a_{13}x^2 - (a_{23} + 2a_{33})y^2}{2a_{33}} - \chi_{1,x}\right)\Bigg] \tag{26}$$

The tip integrals P_x, P_y, M_z and P^D_1, P^D_2, M^D_3 are presented as

$$\begin{pmatrix} P_x \\ P_y \\ M_z \end{pmatrix} = \begin{pmatrix} \dfrac{I_y}{a_{33}} & 0 & 0 \\ 0 & \dfrac{I_x}{a_{33}} & 0 \\ D^e_4 - D_{\chi_1} & D^e_5 - D_{\chi_2} & D_\varphi - D^e_6 \end{pmatrix} \begin{pmatrix} \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix}$$

$$\begin{pmatrix} P^D_1 \\ P^D_2 \\ M^D_3 \end{pmatrix} = \begin{pmatrix} \dfrac{d_{33}I_y}{a_{33}} & 0 & 0 \\ 0 & \dfrac{d_{33}I_x}{a_{33}} & 0 \\ D^\epsilon_4 + D^e_{\chi_1} & D^\epsilon_5 + D^e_{\chi_2} & D^e_\varphi - D^\epsilon_6 \end{pmatrix} \begin{pmatrix} \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix}$$

From the tip integrals P_x, P_y we get $\alpha_4 = (P_x/I_y)a_{33}$, $\alpha_5 = (P_y/I_x)a_{33}$. Denote the matrices above by $\mathbf{Q}_1$, $\mathbf{Q}_2$, respectively. Hence the fundamental relations between 12 tip quantities are expressed by a *tip coupling matrix* $\mathbf{Q} = \mathbf{Q}_2\mathbf{Q}_1^{-1}$: $(P^D_1, P^D_2, M^D_3) = \mathbf{Q}\cdot(P_x, P_y, M_z)$. This matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{pmatrix} d_{33} & 0 & 0 \\ 0 & d_{33} & 0 \\ Q_{31} & Q_{32} & \dfrac{D^e_\varphi - D^\epsilon_6}{D_\varphi - D^e_6} \end{pmatrix}$$

where

$$Q_{31} = \frac{a_{33}}{I_y}\cdot\frac{(D^\epsilon_4 + D^e_{\chi_1})(D_\varphi - D^e_6) + (D^e_\varphi - D^\epsilon_6)(D_{\chi_1} - D^e_4)}{D_\varphi - D^e_6}$$

and $Q_{32} = \mathrm{Sym}(Q_{31})$. From the tip integrals P_z, M_x, M_y we obtain $\alpha_1 = (M_y/I_y)a_{33}$, $\alpha_2 = -(M_x/I_x)a_{33}$, $\alpha_3 = a_{33}(P_z/S_\Omega)$. This means $M^D_3 = Q_{31}P_x + Q_{32}P_y + (D^e_\varphi - D^\epsilon_6)/(D_\varphi - D^e_6)M_z$ and

$$\frac{M^D_1}{M_x} = \frac{M^D_2}{M_y} = \frac{P^D_3}{P_z} = \frac{P^D_2}{P_y} = \frac{P^D_1}{P_x} = d_{33}$$

Remark 5 (shear centers). The problem of the bending of an isotropic cantilever and of a cantilever having anisotropy of a special kind was studied by Saint-Venant [11]. In order to avoid the torsional load due to a force applied at the tip of a beam, the force must be applied at the elastic shear center of Ω (the location where transverse bending induces no twist, see [10,8]). The coordinates of the shear center of a piezoelectric beam are

$$x_{sc} = a_{33}\frac{D_{\chi_2} - D^e_5}{I_x} \quad y_{sc} = -a_{33}\frac{D_{\chi_1} - D^e_4}{I_y}$$

For electrical tip load we define the *electric shear center* as *the location where transverse P^D_i-bending induces no M^D_3-tip moment*. We obtain

$$x^e_{sc} = a_{33}\frac{D^e_{\chi_2} + D^\epsilon_5}{d_{33}I_x} \quad y^e_{sc} = a_{33}\frac{D^e_{\chi_1} + D^\epsilon_4}{d_{33}I_y}$$

Remark 6 (torsional rigidity). The rigorous theory of pure torsion was developed by Saint-Venant [12]. To establish an upper bound for the torsional rigidity, we estimate the difference of integrals, $D_\varphi - D^e_6$,

$$D_{\varphi}-D_{6}^{e}\geqslant-\frac{1}{4}\int\int_{\Omega}(a_{55}\beta_{1}^{2}+2a_{45}\beta_{1}\beta_{2}+a_{44}\beta_{2}^{2})$$

$$\geqslant-\frac{1}{4}\max\{a_{44},a_{55}\}\int\int_{\Omega}(\beta_{1}^{2}+\beta_{2}^{2})$$

where for short we use the notation

$$\beta_{1}=e_{15}\varphi_{6,x}+e_{25}\varphi_{6,y}\quad\beta_{2}=e_{14}\varphi_{6,x}+e_{24}\varphi_{6,y}$$

and equality holds if, and only if,

$$\varphi_{,y}=\frac{1}{2}(a_{45}\beta_{1}+a_{44}\beta_{2})-x\quad\varphi_{,x}=\frac{1}{2}(a_{55}\beta_{1}+a_{45}\beta_{2})+y$$

Note that for a purely elastic beam, $\beta_1=\beta_2=0$, the equality is not reachable, [8], since $\varphi_{,yx}=-1\neq 1=\varphi_{,xy}$. We subsequently conclude that

$$D_{\varphi}-D_{6}^{e}\leqslant\frac{a_{44}I_{x}+a_{55}I_{y}}{a_{0}}+\int\int_{\Omega}\left[\frac{1}{4}(a_{55}\beta_{1}^{2}+2a_{45}\beta_{1}\beta_{2}+a_{44}\beta_{2}^{2})+x\beta_{2}-y\beta_{1}\right]$$

When the inequality above is taken as equality the functions $\tilde{\varphi}$, $\tilde{\varphi}_6$ are polynomials in x, y of degree $\leqslant 2$, their expressions are very long. In extremal case $D_{\varphi}-D_6^e$ obtains a maximal value on a domain with elliptical boundary. Analogous estimates can be obtained for the difference of integrals $D_{\varphi}^{e}-D_6^{\epsilon}$ to establish an upper bound for the electric component of torsional rigidity.

5.5 Solution Procedure

1. **Extension and bending**. To determine the effect of the P_z, M_x, and M_y (or P_3^D, M_1^D, and M_2^D) triad, the system origin location may be selected arbitrarily and the following procedure is adopted: (*a*) calculate α_1, α_2, α_3, (*b*) calculate the stress, strain, and (electrical) displacement components (while $\alpha_4=\alpha_5=\alpha_6=0$).
2. **Shear and torsion**. To determine the effect of the P_x, P_y, M_z (or P_1^D, P_2^D, M_3^D) triad, the following procedure is adopted: (a) select system orientation (i.e., the axes direction in the x, y-plane), (b) solve three CNP for the functions φ, φ_i, χ_1, χ_2, (c) calculate D_{φ}, D_{φ}^e, D_i^e, D_i^{ϵ}, D_{χ_i}, $D_{\chi_i}^e$ of Eqs. (21)–(26), respectively, and then the tip coupling matrix $\mathbf{Q}$, (d) calculate α_4, α_5, α_6, (e) calculate the stress, strain, and (electrical) displacement components (while $\alpha_1=\alpha_2=\alpha_3=0$).

6 Examples of Solutions

6.1 Exact Solution for an Elliptical Beam. Consider a clamped elliptical beam, made of MON13z type material. The cosines of the angles between the normal to the ellipse (of semi-axes $\tilde{a}\geqslant\tilde{b}>0$) and the x- and y-axes are given as $\cos(\bar{\mathbf{n}},x)=(\tilde{b}/\tilde{\lambda})(x/\tilde{a})$, $\cos(\bar{\mathbf{n}},y)=(\tilde{a}/\tilde{\lambda})(y/\tilde{b})$, where $\tilde{\lambda}=\sqrt{(\tilde{b}^2/\tilde{a}^2)x^2+(\tilde{a}^2/\tilde{b}^2)y^2}$ is the "*parametrization velocity*." The cross-section area, and the moments (of inertia) are $S_{\Omega}=\pi\tilde{a}\tilde{b}$ and $I_y=\frac{1}{4}\pi\tilde{a}^3\tilde{b}$, $I_x=\frac{1}{4}\pi\tilde{a}\tilde{b}^3$.

Proposition 7. *Let F_{Λ}, F_{Λ}^{e} and P^{Λ}, Q^{Λ}, P_e^{Λ}, Q_e^{Λ} be general polynomials in x, y, satisfying conditions, Eqs. (18), in ellipse $\Omega:(x^2/a^2)+(y^2/b^2)\leqslant 1$. Then the solution functions Λ, Λ_e of a CNP defined in Ω by Eqs. (15a)–(15d) are also polynomials of degree $k_{\Lambda}=\max\{k_1+1,k_2+2\}$, where $k_1=\max\{\deg P^{\Lambda},\deg Q^{\Lambda},\deg P_e^{\Lambda},\deg Q_e^{\Lambda}\}$ and $k_2=\max\{\deg F_{\Lambda},\deg F_{\Lambda}^{e}\}$.*

This proposition plays a key role for producing test examples in what follows. Based on Proposition 7 one may show (similarly as in [8]) that the torsion and bending functions φ, χ_1, χ_2 and their piezoelectric companions φ_i ($i=4,5,6$) are polynomials of x, y coordinates in an ellipse. Namely, for the torsion warping function $\varphi(x,y)$ and its electrical representative $\varphi_6^e(x,y)$ we arrive at the second-order polynomials

$$\varphi=H_{20}^{\varphi}x^{2}+H_{11}^{\varphi}xy+H_{02}^{\varphi}y^{2}\quad\varphi_{6}^{e}=H_{20}^{\varphi_6}x^{2}+H_{11}^{\varphi_6}xy+H_{02}^{\varphi_6}y^{2}$$

The bending functions $\chi_1(x,y)$, $\chi_2(x,y)$ and their electrical representatives $\varphi_4^e(x,y)$, $\varphi_5^e(x,y)$ for an elliptical cross section are sums of third-degree and linear homogeneous polynomials of x, y

$$\chi_{i}=H_{30}^{\chi_i}x^{3}+H_{21}^{\chi_i}x^{2}y+H_{12}^{\chi_i}xy^{2}+H_{03}^{\chi_i}y^{3}+H_{10}^{\chi_i}x+H_{01}^{\chi_i}y$$

$$\varphi_{j}=H_{30}^{\varphi_j}x^{3}+H_{21}^{\varphi_j}x^{2}y+H_{12}^{\varphi_j}xy^{2}+H_{03}^{\varphi_j}y^{3}+H_{10}^{\varphi_j}x+H_{01}^{\varphi_j}y$$

where $i=1,2$, $j=4,5$ Hence for elliptical beam the solution functions (stress, strain, etc.) of Saint-Venant's problem are polynomials in x, y, z.

Example 8. (orthotropic beam). Consider a piezoelectric elliptical beam made from *orthotropic* material, $a_{45}=0$, and let $\epsilon_{12}^{s}=0$ and $e_{14}=e_{25}=0$. Note that $d_{15}=e_{15}a_{55}$, $d_{24}=e_{24}a_{44}$. Then

$$\varphi=\frac{a_{44}a_{55}(e_{15}^{2}\tilde{b}^{4}-e_{24}^{2}\tilde{a}^{4})+(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(a_{44}\tilde{b}^{2}-a_{55}\tilde{a}^{2})}{(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(a_{55}\tilde{a}^{2}+a_{44}\tilde{b}^{2})+a_{44}a_{55}(e_{24}\tilde{a}^{2}+e_{15}\tilde{b}^{2})^{2}}xy$$

$$\varphi_{6}=-\frac{2\tilde{a}^{2}\tilde{b}^{2}(e_{24}a_{44}-a_{55}e_{15})xy}{(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(a_{55}\tilde{a}^{2}+a_{44}\tilde{b}^{2})+a_{44}a_{55}(e_{15}\tilde{b}^{2}+e_{24}\tilde{a}^{2})^{2}}$$

From Eqs. (22) and (24) follow zero values of some tip integrals: $D_{\chi_j}^{e}=D_{\chi_j}=D_4^{\epsilon}=D_5^{\epsilon}=D_4^{e}=D_5^{e}=0$. Substituting $\varphi(x,y)$ and $\varphi_6(x,y)$ in Eqs. (21) and (23) yields the following expressions for the integrals:

$$D_{\varphi}=\frac{\frac{\pi}{2}\tilde{a}^{3}\tilde{b}^{3}[(d_{15}+d_{24})(e_{24}\tilde{a}^{2}+e_{15}\tilde{b}^{2})+2(\epsilon_{11}^{s}\tilde{b}^{2}+\epsilon_{22}^{s}\tilde{a}^{2})]}{(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(a_{55}\tilde{a}^{2}+a_{44}\tilde{b}^{2})+a_{44}a_{55}(e_{15}\tilde{b}^{2}+e_{24}\tilde{a}^{2})^{2}}$$

$$D_{6}^{e}=\frac{\frac{\pi}{2}\tilde{a}^{3}\tilde{b}^{3}(e_{24}\tilde{a}^{2}-e_{15}\tilde{b}^{2})(d_{15}-d_{24})}{(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(a_{55}\tilde{a}^{2}+a_{44}\tilde{b}^{2})+a_{44}a_{55}(e_{15}\tilde{b}^{2}+e_{24}\tilde{a}^{2})^{2}}$$

$$D_{\varphi}^{e}=\frac{\pi}{2}\tilde{a}^{3}\tilde{b}^{3}\cdot\frac{2d_{24}d_{15}(e_{24}\tilde{a}^{2}+e_{15}\tilde{b}^{2})+(\epsilon_{22}^{s}\tilde{a}^{2}+\epsilon_{11}^{s}\tilde{b}^{2})(d_{24}+d_{15})}{(\epsilon_{22}^{s}a_{44}+\epsilon_{11}^{s}a_{55})\tilde{a}^{2}\tilde{b}^{2}+\epsilon_{22}^{s}a_{55}\tilde{a}^{4}+\epsilon_{11}^{s}a_{44}\tilde{b}^{4}+a_{44}a_{55}(e_{15}\tilde{b}^{2}+e_{24}\tilde{a}^{2})^{2}}$$

$$D_{6}^{\epsilon}=\frac{\pi}{2}\tilde{a}^{3}\tilde{b}^{3}\cdot\frac{(d_{24}-d_{15})(\epsilon_{22}^{s}\tilde{a}^{2}-\epsilon_{11}^{s}\tilde{b}^{2})}{(\epsilon_{22}^{s}a_{44}+\epsilon_{11}^{s}a_{55})\tilde{a}^{2}\tilde{b}^{2}+\epsilon_{22}^{s}a_{55}\tilde{a}^{4}+\epsilon_{11}^{s}a_{44}\tilde{b}^{4}+a_{44}a_{55}(e_{15}\tilde{b}^{2}+e_{24}\tilde{a}^{2})^{2}}$$

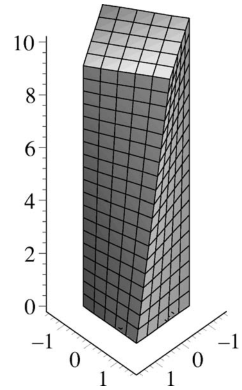


Fig. 3 Piezoelectric beam deformation for M_z=1

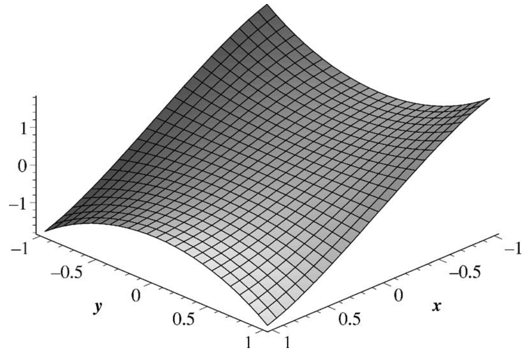


Fig. 5 Bending function χ_1

Hence the matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$ are

$$\mathbf{Q}_1=\begin{pmatrix} \dfrac{I_y}{a_{33}} & 0 & 0 \\ 0 & \dfrac{I_x}{a_{33}} & 0 \\ 0 & 0 & D_\varphi-D_6^e \end{pmatrix} \mathbf{Q}_2=\begin{pmatrix} \dfrac{d_{33}I_y}{a_{33}} & 0 & 0 \\ 0 & \dfrac{d_{33}I_x}{a_{33}} & 0 \\ 0 & 0 & D_\varphi^e-D_6^\epsilon \end{pmatrix}$$

The tip coupling matrix $\mathbf{Q}$ has a diagonal view with

$$\mathbf{Q}_{11}=\mathbf{Q}_{22}=d_{33} \quad \mathbf{Q}_{33}=\frac{D_\varphi^e-D_6^\epsilon}{D_\varphi-D_6^e}$$

$$=\frac{d_{15}\epsilon_{22}^s\tilde{a}^2+d_{24}\epsilon_{11}^s\tilde{b}^2+d_{15}d_{24}(e_{24}\tilde{a}^2+e_{15}\tilde{b}^2)}{(e_{24}d_{24}+\epsilon_{22}^s)\tilde{a}^2+(e_{15}d_{15}+\epsilon_{11}^s)\tilde{b}^2}$$

For the case $d_{24}=d_{15}$ we have $\varphi_6=0$, $D_6^\epsilon=D_6^e=0$, $D_\varphi^e=\pi\tilde{a}^3\tilde{b}^3e_{24}/\tilde{a}^2+\tilde{b}^2$, and expressions for φ, D_φ are simplified to known formulas in elasticity theory, see [8],

$$\varphi=\frac{a_{44}\tilde{b}^2-a_{55}\tilde{a}^2}{a_{55}\tilde{a}^2+a_{44}\tilde{b}^2}xy \quad D_\varphi=\frac{\pi\tilde{a}^3\tilde{b}^3}{a_{55}\tilde{a}^2+a_{44}\tilde{b}^2}$$

Additionally, for the case $\tilde{b}^2/\tilde{a}^2=\epsilon_{22}^s/\epsilon_{11}^s$ we have $D_6^\epsilon=0$, and for the case $\tilde{b}^2/\tilde{a}^2=e_{24}/e_{15}$ we have $D_6^e=0$.

Example 9 (transtropic circular beam). Consider a piezoelectric circular beam ($\tilde{a}=\tilde{b}$, radius of a cross section) made from a poled *piezoelectric ceramic*, $a_{55}=a_{44}$, $\epsilon_{11}^s=\epsilon_{22}^s, e_{15}=e_{24}$, $a_{45}=\epsilon_{12}^s=e_{14}=e_{25}=0$. Then $\varphi=\varphi_6=0$, see Eqs. (27), the other four auxiliary functions are nonzero,

$$\chi_1=\frac{x}{8a_{33}(e_{24}d_{24}+\epsilon_{11}^s)}\{[(2a_{13}+a_{44})\epsilon_{11}^s+(d_{33}+2e_{24}a_{13})d_{24}]x^2$$
$$+[(2a_{13}-8a_{33}+a_{44})\epsilon_{11}^s+d_{24}(2(a_{13}-4a_{33})e_{24}+d_{33})]y^2$$
$$-\tilde{a}^2[(2a_{13}+3a_{44})\epsilon_{11}^s+(3d_{33}+2e_{24}a_{13})d_{24}]\}$$

$$\varphi_4=\frac{d_{33}-d_{24}}{8a_{33}(e_{24}d_{24}+\epsilon_{11}^s)}(x^2+y^2-3\tilde{a}^2)x$$

and $\varphi_5=\mathrm{Sym}(\varphi_4)$, $\chi_2=\mathrm{Sym}(\chi_1)$. The electrical potential is

$$\Phi_e=\frac{d_{33}-d_{24}}{8(e_{24}d_{24}+\epsilon_{11}^s)}(x^2+y^2-3\tilde{a}^2)\left(\frac{P_x}{I_y}x+\frac{P_y}{I_x}y\right)$$

The tip integrals are $D_\varphi=\pi\tilde{a}^4/2a_{44}$, $D_\varphi^e=\pi\tilde{a}^4e_{24}/2$. Finally, we obtain $(D_\varphi^e-D_6^\epsilon)/(D_\varphi-D_6^e)=d_{24}$.

6.2 Numerical Solutions. Examples of what follows deal with a clamped beam made of transtropic material PZT-5A. The length of a beam is $l=10$. All expression in what follows are given approximately. One may derive $d_{33}\approx 374\cdot 10^{-12}$.

Example 10 (elliptical piezoelectric beam). Consider a piezoelectric beam made of piezoceramic PZT-5A with elliptical cross section, semi axes $\tilde{a}=2$, $\tilde{b}=1$, that undergoes a tip moment $M_z=1$ only. We obtain $D_\varphi\approx 0.106\cdot 10^{12}$, $D_\varphi^e\approx 61.8$, $D_6^e=D_6^\epsilon=0$. The tip coupling matrix $\mathbf{Q}$ has a diagonal view with

$$\mathbf{Q}_{11}=\mathbf{Q}_{22}\approx 374\cdot 10^{-12} \quad \mathbf{Q}_{33}\approx 584\cdot 10^{-12} \qquad (27)$$

The electric tip moment is $M_3^D=(D_\varphi^e-D_6^\epsilon)/(D_\varphi-D_6^e)\approx 584\cdot 10^{-12}$. The α-parameters are

$$\alpha_1\approx 3.0\cdot 10^{-12}M_y \quad \alpha_2\approx -11.97\cdot 10^{-12}M_x$$

$$\alpha_3\approx 3.0\cdot 10^{-12}P_z \quad \alpha_4\approx 3.0\cdot 10^{-12}P_x$$

$$\alpha_5\approx 12.0\cdot 10^{-12}P_y \quad \alpha_6\approx 9.45\cdot 10^{-12}M_z$$

The auxiliary functions are $\varphi\approx -0.60xy$, $\varphi_6=0$, and

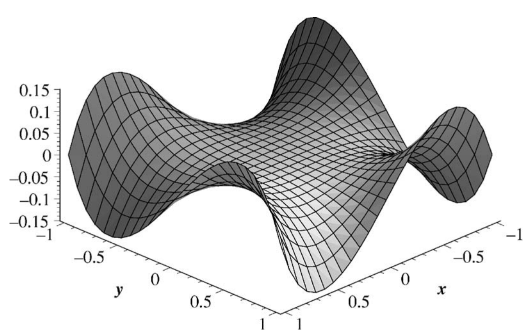


Fig. 4 Torsional function φ

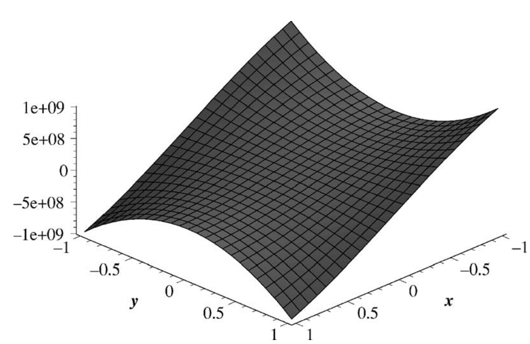


Fig. 6 Electrical bending function φ_4

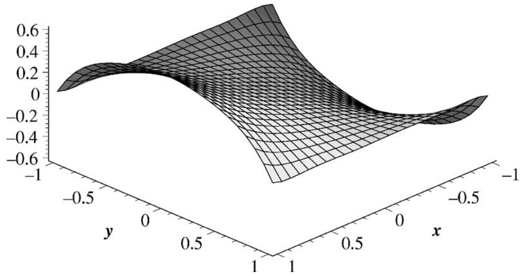


Fig. 7 Stress component σ_4 (the tip load M_z=1)

$$\chi_1 \approx 0.17x^3 - 0.84xy^2 - 2.8x$$

$$\chi_2 \approx 0.16y^3 - 0.82x^2y - 0.68y$$

$$\varphi_4 \approx (1.01x - 0.0842x^3 - 0.112xy^2) \cdot 10^9$$

$$\varphi_5 \approx (0.313y - 0.0521x^2y - 0.104y^3) \cdot 10^9$$

The nonzero stress and strain components are $\sigma_4 \approx 0.08x$, $\sigma_4 \approx -0.32y$, and $\varepsilon_4 \approx 3.718 \cdot 10^{-12}x$, $\varepsilon_5 \approx -15.12 \cdot 10^{-12}y$. The displacement is $u \approx -9.45 \cdot 10^{-12}yz$, $v \approx 9.45 \cdot 10^{-12}xz$, $w \approx -5.66 \cdot 10^{-12}xy$. The electrical potential is $\Phi_e=0$, and the electrical displacements components are $D_1 \approx -0.19 \cdot 10^{-9}y$, $D_2 \approx 0.048 \cdot 10^{-9}x$.

Example 11 (rectangular piezoelectric beam). Let a beam with a square cross section, $d=h=1$, made of piezoceramic PZT-5A, undergo a tip moment $M_z=1$ only, Fig. 3. The cross-section area, and the moments (of inertia) are $S_\Omega=4hd=4$ and $I_y=\frac{4}{3}d^3h=\frac{4}{3}$, $I_x=\frac{4}{3}dh^3=\frac{4}{3}$. As was shown in Remark 4, $\varphi_6=0$. We obtain $D_\varphi \approx 0.047 \cdot 10^{12}$, $D_\varphi^e \approx 27.6$, $D_6^e=D_6^\epsilon=0$. The induced electric tip moment is $M_3^D=584 \cdot 10^{-12}$. The tip coupling matrix is given in Eq. (27). We find approximate solutions of the CNPs presenting auxiliary torsion and bending functions φ, χ_1, χ_2 and φ_4, φ_5 by polynomials of the 12th degree, see Figs. 4–6 (the graphs of the symmetric expressions are omitted). The relative error in this case is $\approx 10^{-2}$. Then the stress components $\sigma_i=0$ $(i=1,2,3,6)$, and displacement components are $u \approx -21 \cdot 10^{-12}yz$, $v \approx -21 \cdot 10^{-12}xz$. Other functions of the solution have long polynomial expressions, and they are presented in Figs. 7–9. The weight coefficients are $\alpha_1=\alpha_2=\alpha_3=\alpha_4=\alpha_5=0$ and $\alpha_6 \approx 18 \cdot 10^{-12}$.

7 Conclusions

(1) Analytical solution of the Saint-Venant's problem for a homogeneous piezoelectric beam is presented. It generalizes the known "elastic" solution, see [8].

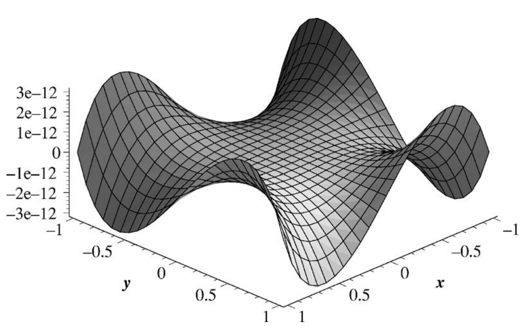


Fig. 8 Torsional warping function *w*

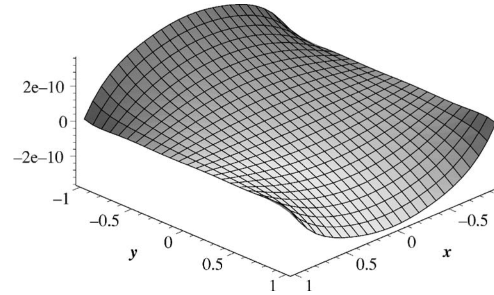


Fig. 9 Electric displacement D_1 (the tip load M_z=1)

(2) Three pairs of auxiliary (torsion and bending) functions for a piezoelectric beam are introduced. They satisfy the coupled Neumann problem (CNP). In the limit case CNP transforms to the "elastic" Neumann problem.

(3) The concepts of the torsion/bending functions, torsional rigidity and shear center, tip coupling matrix for a piezoelectric beam are developed.

(4) Examples of exact solutions for elliptical/circular beam and numerical solution for a rectangular beam are presented.

Acknowledgment

The authors would like to thank Mrs. O. Rand and A. Kozhevnikov for useful discussions.

References

[1] Bisegna, P., 1998, "The Saint-Venant Problem in the Linear Theory of Piezoelectricity," Atti Convegni Lincei, Accad. Naz. Lincei, Rome, **140**, pp. 151–165.
[2] Davh, F., 1996, "Saint-Venant's Problem for Linear Piezoelectric Bodies," J. Elast., **43**, pp. 227–245.
[3] dell'Isola, F., and Rosa, L., 1996, "Saint-Venant's Problem in Linear Piezoelectricity," Mathematics and Control in Smart Structures, **2715**, pp. 399–409.
[4] Iesan, D., 1989, "On Saint-Venant's Problem for Elastic Dielectrics," J. Elast., **21**, pp. 101–115.
[5] Zanfir, S., 1985, "Saint-Venant's Problem for Heterogeneous Anisotropic Dipolar Elastic Solids. I, II," Bull. Math. Soc. Sci. Math. Repub. Soc. Roum., **29**(77), pp. 177–187; **29**(77), pp. 371–376.
[6] Michell, J. H., 1901, "The Theory of Uniformly Loaded Beams," Journal of Mathematics, **32**, pp. 28–42.
[7] Kudryavtsev, B. A., Parton, V. Z., and Senik, N. A., 1990, "Electromagnetoelasticity," *Applied Mechanics: Soviet Reviews*, Vol. 2, Hemisphere, Washington, DC, pp. 1–230.
[8] Rand, O., and Rovenski, V., 2005, *Analytical Methods in Anisotropic Elasticity With Symbolic Computational Tools*, Birkhauser, Boston.
[9] Rovenski, V., Harash, E., and Abramovich, H., 2006, "St. Venant's Problem for Homogeneous Piezoelastic Beams," TAE Report No. 967, 1–100.
[10] Ruchadze, A. K., 1975, "On One Problem of Elastic Equilibrium of Homogeneous Isotropic Prismatic Bar (in Russian)," Transactions of Georgian Polytechnical Institute, **3**(176), pp. 208–218.
[11] St. Venant, B., 1856, "Memoire Sur la Flexion des Prismes," J. Math. Pures Appl., **1**, pp. 89–189.
[12] St. Venant, B., 1856, "Memoire Sur la Torsion des Prismes," Memoires Presentes par Divers Savants a l'academie des Sciences, Sciences math, et phys. Paris, Vol. **14**, pp. 233–560.
[13] Batra, R. C., and Yang, J. S., 1995, "Saint-Venant's Principle in Linear Piezoelectricity," J. Elast., **38**(2), pp. 209–218.
[14] Horgan, C. O., and Payne, L. E., 1997, "Saint-Venant's Principle in Linear Isotropic Elasticity for Incompressible or Nearly Incompressible Materials," J. Elast., **46**(1), pp. 43–52.
[15] Xu, X. S., Zhong, W. X., and Zhang, H. W., 1997, "The Saint-Venant Problem and Principle in Elasticity," Internat. J. Solids Structures, **34**(22), pp. 2815–2827.
[16] Cady, W. G., 1964, *Piezoelectricity—An Introduction to the Theory and Applications of Electromechanical Phenomena in Crystals*, Vol. 1, Dover.
[17] Yang, J., 2006, *The Mechanics of Piezoelectric Structures*, World Scientific.

Modeling Helicopter Blade Sailing: Dynamic Formulation in the Planar Case

A. S. Wall
e-mail: awall2@connect.carleton.ca

R. G. Langlois

F. F. Afagh

Department of Mechanical & Aerospace Engineering,
Carleton University,
1125 Colonel By Drive,
Ottawa, ON K1S 5B6, Canada

As part of a research project aimed at simulating rotor dynamic response during shipboard rotor startup and shutdown operations, a dynamic model of the ship–helicopter–rotor system that is appropriate for use in predicting rotor elastic response was developed. This planar model consists of a series of rigid bodies connected by rotational stiffness and damping elements that allow motion in the flapwise direction. The rotors were partitioned into an arbitrary number of rigid beam segments having the inertial and geometrical properties of a typical rotor. Helicopter suspension flexibility and damping were also modeled, although the helicopter was otherwise considered as a rigid body. Lagrange's equation was used to derive the governing dynamic equations for the helicopter–rotor model. The effect of ship motion on blade deflection was also considered. The ship motion supplied as input to the model included representative frigate flight deck motion in three dimensions corresponding to an actual sea spectrum, ship particulars and ship operating conditions. This paper is intended to detail the dynamic approach adopted for this blade sailing study, and its conceptual validation in the planar case. The methodologies that have been developed lend themselves to easy expansion into three dimensions, and into torsion and lead/lag modeling. The amount of blade motion induced by ship motion on nonrotating helicopter blades is included. Although aerodynamic loads are a major contributor to blade sailing, this paper focuses on the dynamics aspect of the problem, and thus does not include aerodynamic effects. [DOI: 10.1115/1.2722766]

Keywords: blade sailing, tunnel strike, shipboard helicopter, multi-body dynamics, Lagrange's equation, embedded Riemann sums, dynamic interface analysis, rotor engage/disengage

Introduction

During startup and shutdown operations on ship decks, helicopter rotor elastic response is of concern for flight safety reasons. While the rotors are engaged or disengaged, they turn at low speeds and therefore can be subjected to high wind-induced aerodynamic forces without the benefit of the centrifugal stiffening present at operating speeds. This excitation, combined with ship deck motion during all but the most benign sea and wind conditions, can cause excessive deflection of rotor blades, which, as a result, can come into contact with the fuselage or tailboom of the helicopter. This phenomenon, called "tunnel strike" or "tailboom strike," compromises the safety of flight crews, results in airframe damage, and may bring the airworthiness of the helicopter into question. Figure 1 shows a Canadian patrol frigate with a landing helicopter. Blade sailing, which is the excessive motion of helicopter blades during engage/disengage, could be experienced by this or any other shipboard helicopter upon landing and prior to launch.

Current guidelines for safe engage/disengage operations are based largely on experience and some understanding of the ship motion and flight deck aerodynamics for different ship headings and speeds as well as sea and wind conditions. As such, blade sailing has been a relevant research topic since the 1960s. Literature relating to the dynamics of blade sailing has been well summarized by several researchers, including Newman [1] and Keller [2]. The studies cited in these works cover the modeling of semirigid rotors [3] and articulated rotors, using finite element methods [4–9] and multibody methods [10,11]. For articulated rotors, the effect of droop and flap stop impacts [12] have been discussed. They also include various validation experiments [13–15] that have been compared with the models. Due to the complex nature of the helicopter–ship interface and the variety of contributors that are believed to affect blade motion, further research is required to develop tools suitable for comprehensively defining safe engage/disengage operational envelopes for many different types of helicopters. The purpose of this research is to develop a simulation to model the complete ship–helicopter–rotor system, including a potentially computationally efficient blade modeling concept that has not previously been applied in the blade sailing context. This paper lays the mathematical groundwork for the dynamics of this problem.

Model Description

The helicopter model used in this research is based on the philosophy that a continuously flexible body can be modeled by a series of rigid segments connected by flexible elements. Provided that the geometric nonlinearities that result from large blade deflections are preserved in the equations of motion, this method has been shown to approximate flexible body behavior as well as geometrically nonlinear finite element analysis and has the potential for decreased solution times [16,17].

The helicopter model, with four blade segments per blade as an example, is shown in Fig. 2.

The number of degrees of freedom, n_{dof}, required to completely define the helicopter motion is given by

$$n_{\text{dof}} = 3 + n_1 + n_2 \tag{1}$$

where the constant 3 comes from the two translational and one rotational degrees of freedom provided by the helicopter suspension, and n_1 and n_2 are the number of segments in the port and

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received October 12, 2004; final manuscript received January 4, 2007. Review conducted by Marc P. Mignolet.

Fig. 1 A Canadian patrol frigate

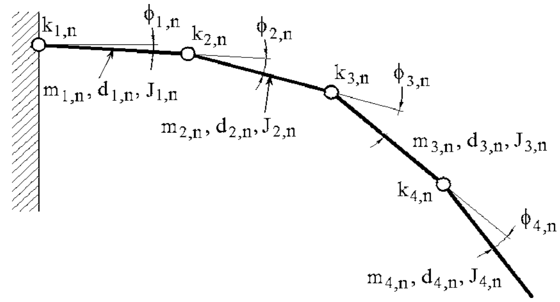


Fig. 3 A helicopter blade modeled using rigid blade segments

starboard blades, respectively. Ship motion is a function of time and as a result does not have associated degrees of freedom.

The orientation of each blade segment is defined by the angle, $\phi_{i,n}$, between the ith segment and the $(i-1)$th segment as shown in Fig. 3. This means that the rotational spring force in each joint depends on only a single degree of freedom, but that the position and velocity of the mass associated with the ith segment of the nth blade depends on angles θ, which is the orientation of the helicopter body, and $\phi_{1,n}$ to $\phi_{i,n}$. This results in a cascading angle effect that is inherent in the position and velocity expressions for each mass and therefore a major characteristic of the equations of motion for this system.

In this model, the generalized coordinates corresponding to the degrees of freedom are

$$\{\mathbf{q}\}^T = \left\{ Y_C \quad Z_C \quad \theta \quad \phi_{1,1} \cdots \phi_{n_1,1} \quad \phi_{1,2} \cdots \phi_{n_2,2} \right\}^T \tag{2}$$

where Y_C, Z_C, and θ are the horizontal, vertical, and angular positions of the equivalent suspension attachment point on the helicopter in the stationary global reference frame. The angles, $\phi_{i,n}$, are the relative joint angles of the ith segment on the nth blade. The model is completely general, in that the number of segments into which the blade is partitioned and the properties for each individual segment can be user defined. Each blade segment is defined with properties $m_{i,n}$, the segment mass which is assumed concentrated at the segment center (although the model is theoretically general enough to allow it to be concentrated at any point in space relative to the blade segment); $d_{i,n}$, the segment length; and $J_{i,n}$, the segment mass moment of inertia. Each blade segment is connected to the last with a rotational element of stiffness $k_{i,n}$ and viscous rotational damping $c_{i,n}$. Since the properties can be individually assigned to each segment, they can be tuned to closely approximate a nonuniform blade. As with variable finite element gridding, smaller segments can be used in areas of higher flexibility, while longer segments can be used in stiffer areas.

The distinction between semi-rigid and articulated blades is an important one in helicopter blade modeling. The blade model shown in Fig. 3 can be used to simulate either, provided the correct parameters are included in the model. The semi-rigid rotor requires one representative stiffness parameter at the root. The articulated rotor is more complicated, in that the joint stiffness can be modeled as a function of joint angle, where the stiffness is zero while the blade flaps in the hinge range, and finite while the blade is interacting with the droop and flap stops. In the planar case discussed in this paper, with the rotational and aerodynamic effects not included, the articulated blade behaves very much like a semi-rigid blade in that the blade does not lift off the droop stops. As such, validation of the stop effect on blade behavior is not discussed.

Figure 4(a) shows a schematic of a typical frigate, and indicates where the helicopter is typically located during startup and shutdown operations. Since the model exists in the vertical plane, there are two orthogonal vertical planes in which the helicopter can exist on the ship deck. Figure 4(b) shows the helicopter in the roll plane, (Y,Z), and in the pitch plane, (X,Z). All coordinate systems used in the dynamic formulation are right handed.

The effects of ship motion on the helicopter in both planes were examined separately. Note that most of the paper discusses the roll plane, (Y,Z), during equation development, since the equations for the pitch plane are identical if all quantities in Y and y are replaced with equivalent quantities in X and x, respectively.

Equations of Motion

Because the dynamics are based on a rigid body model, it is possible to derive a set of completely general equations of motion which need only be evaluated for any set of user-defined parameters. While many methods exist [18] that can be used to derive the equations of motion for a system of rigid bodies, the applications of these in a sense where the number of rigid bodies is arbitrary is not trivial. The derivation and application of the equations of motion are detailed here.

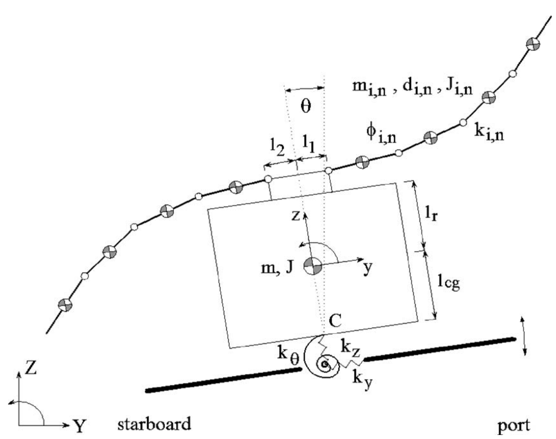


Fig. 2 A planar helicopter model with four blade segments per blade

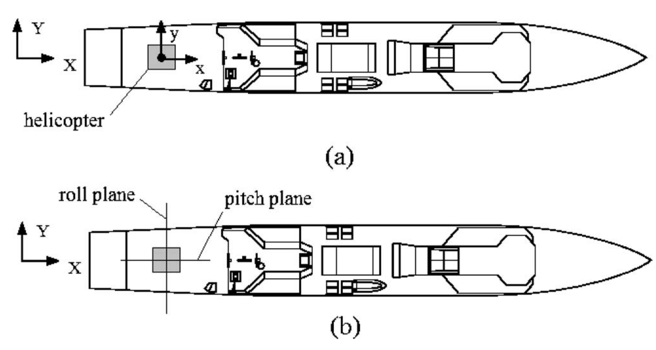


Fig. 4 Ship and helicopter schematic

Equation Derivation. The equations of motion for the ship–helicopter–rotor system were derived using Lagrange's equation [19]

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_p}\right) - \frac{\partial T}{\partial q_p} + \frac{\partial U}{\partial q_p} = Q_p \tag{3}$$

where T is the system kinetic energy; U is the system potential energy; and Q_p is the total nonconservative applied force associated with the pth degree of freedom. This method is useful because the kinetic and potential energy expressions are scalar and therefore relatively straightforward to determine.

The kinetic energy for the ship–helicopter–rotor model comes from two major sources: translational and rotational energy. Each rigid body is affected by kinetic energy from both sources. The kinetic energy is given by

$$T = T_1 + T_2 + \sum_{i=1}^{n_1}(T_3 + T_4) + \sum_{i=1}^{n_2}(T_5 + T_6) \tag{4}$$

where T_1=body translational energy; T_2=body rotational energy; T_3= port segment translational energy; T_4= port segment rotational energy; T_5= starboard segment translational energy; and T_6= starboard segment rotational energy.

The potential energy comes from two sources as well: spring and gravitational energy. The expressions are very similar to the kinetic energy. The potential energy is given by

$$U = U_1 + U_2 + \sum_{i=1}^{n_1}(U_3 + U_4) + \sum_{i=1}^{n_2}(U_5 + U_6) \tag{5}$$

where U_1= suspension spring energy; U_2=body gravitational energy; U_3=port segment spring energy; U_4=port segment gravitational energy; U_5=starboard segment spring energy; and U_6=starboard segment gravitational energy.

A complete expansion including all the terms in both the kinetic and potential energy expressions can be found in Eqs. (A1)–(A16) in the Appendix.

In order to obtain a complete set of equations of motion, Lagrange's equation was applied once for each degree of freedom, with p varying from 1 to n_{dof}. Each application resulted in an expression of the form

$$\alpha_{1,p}\ddot{Y}_C + \alpha_{2,p}\ddot{Z}_C + \alpha_{3,p}\ddot{\theta} + \sum_{i=1}^{n_1}\beta_{i,p}\ddot{\phi}_{i,1} + \sum_{i=1}^{n_2}\gamma_{i,p}\ddot{\phi}_{i,2} + \epsilon_p = Q_p \tag{6}$$

where $\alpha_{1,p}$, $\alpha_{2,p}$, $\alpha_{3,p}$, $\beta_{k,p}$, $\gamma_{k,p}$, ϵ_p, and Q_p are all functions of the generalized coordinates and their first time derivatives.

Since the solution was advanced through time using the Runge–Kutta–Fehlberg integration method, the values of each of the coordinates and the corresponding first time derivatives (generalized velocities) are known at each time step, and the second time derivatives (generalized accelerations) of each coordinate are unknown. Therefore, Eq. (6) can be rearranged to

$$\alpha_{1,p}\ddot{Y}_C + \alpha_{2,p}\ddot{Z}_C + \alpha_{3,p}\ddot{\theta} + \sum_{i=1}^{n_1}\beta_{i,p}\ddot{\phi}_{i,1} + \sum_{i=1}^{n_2}\gamma_{i,p}\ddot{\phi}_{i,2} = Q_p - \epsilon_p \tag{7}$$

which is in the form of Newton's second law and is therefore a set of linear equations with the accelerations being the unknown quantities.

The resulting system of equations can be written in the matrix form $[\mathbf{M}]\{\ddot{\mathbf{q}}\}=\{\mathbf{Q}\}$. The equivalent mass matrix

$$[\mathbf{M}] = \begin{bmatrix} \dfrac{\partial}{\partial \ddot{Y}_C}\left(\dfrac{d}{dt}\left(\dfrac{\partial T}{\partial \dot{Y}_C}\right)\right) & \cdots & \dfrac{\partial}{\partial \ddot{\phi}_{n_2,2}}\left(\dfrac{d}{dt}\left(\dfrac{\partial T}{\partial \dot{Y}_C}\right)\right) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial}{\partial \ddot{Y}_C}\left(\dfrac{d}{dt}\left(\dfrac{\partial T}{\partial \dot{\phi}_{n_2,2}}\right)\right) & \cdots & \dfrac{\partial}{\partial \ddot{\phi}_{n_2,2}}\left(\dfrac{d}{dt}\left(\dfrac{\partial T}{\partial \dot{\phi}_{n_2,2}}\right)\right) \end{bmatrix} \tag{8}$$

can be constructed if the derivative of Eq. (7) is taken with respect to the second time derivative of each degree of freedom sequentially to extract the corresponding matrix elements.

The equivalent force vector is

$$\{\mathbf{Q}\} = \begin{Bmatrix} Q_1 & - & \epsilon_1 \\ \vdots & - & \vdots \\ Q_{n_{\mathrm{dof}}} & - & \epsilon_{n_{\mathrm{dof}}} \end{Bmatrix} \tag{9}$$

where the terms contained in ϵ_p come from the left-hand side of Lagrange's equation, and are largely the result of the geometrically nonlinear centripetal and Coriolis forces. The terms in Q_p come from externally applied forces, that are discussed in a subsequent section.

Once the quantities $[\mathbf{M}]$ and $\{\mathbf{Q}\}$ have been determined, the system can be converted to first order form as shown

$$\begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{M} \end{bmatrix}\begin{Bmatrix} \dot{\mathbf{q}} \\ \ddot{\mathbf{q}} \end{Bmatrix} = \begin{Bmatrix} \dot{\mathbf{q}} \\ \mathbf{Q} \end{Bmatrix} \tag{10}$$

This equation is of the familiar form, $[\mathbf{A}]\{\dot{\mathbf{x}}\}=\{\mathbf{b}\}$ thereby allowing the dynamic solution to be advanced numerically through time by an adaptive time step Runge–Kutta–Fehlberg integrator.

Both $[\mathbf{M}]$ and $\{\mathbf{Q}\}$, which were extracted from Lagrange's equation symbolically and depend on $\{\mathbf{q}\}$ and $\{\dot{\mathbf{q}}\}$, are easily reevaluated at each time step.

Associated Challenges. The method for deriving the equations of motion described above is straightforward, however differentiating the completely general expressions presents challenges that are worth mentioning.

The following equation

$$\begin{aligned} S_1 = \frac{1}{2}\sum_{k=1}^{n_1} m_{k,1}\Bigg(\Bigg(-\sum_{i=1}^{k}\left(\sin\left(\theta + \sum_{j=1}^{i}\phi_{j,1}\right)\right)\left(\dot{\theta} + \sum_{j=1}^{i}\dot{\phi}_{j,1}\right)d_{i,1} \\ - v\dot{\theta}\sin(\theta + a) + \dot{Y}_C + \frac{1}{2}\left(\sin\left(\theta + \sum_{j=1}^{k}\phi_{j,1}\right)\right) \\ \times\left(\dot{\theta} + \sum_{j=1}^{k}\dot{\phi}_{j,1}\right)d_{k,1}\Bigg)^2\Bigg) \end{aligned} \tag{11}$$

which is the kinetic energy from the horizontal component of velocity of the port blade, is used to illustrate the challenges.

One characteristic of this equation is three layers of embedded sums. The innermost layer, with index j, comes from the fact that the position and orientation of each blade segment depends on the angle of that segment and all others inboard of it. This cascading angle effect appears frequently in the kinetic and potential energy expressions, as well as in the final equations of motion. The middle layer, with index i, comes from the fact that the velocity of each segment is the sum of that body's velocity relative to the other bodies plus the other bodies' velocities. The outermost layer, with index k, results from the fact that the energy from each segment must be summed to obtain a complete expression.

The challenge in differentiation arises from the fact that n_1, which is both the upper summation limit on the first sum and the number of segments in the port blade, is arbitrary. This means that all three layers of embedded sums are arbitrary including the innermost angle summation, which contains the degree of freedom

$\phi_{l,1}$, with respect to which the derivative is required.

If both n_1 and the index l with respect to which the derivative is required are selected prior to differentiation, then all three layers of sums become finite and defined, and the derivatives can easily be taken using a symbolic mathematics package such as Maple [20]. However, for a general solution, these quantities must remain unknown. It is impossible, then, for the software to evaluate $\partial S/\partial\phi_{l,1}$ because there is no way to indicate that $\phi_{l,1}$ is simply any one of the $(\phi_{j,1})$.

To this end, a rule for differentiating the expressions by hand was developed. The problem lies in differentiating

$$S_2 = \sum_{j=1}^{k} \phi_{j,n} \tag{12}$$

with respect to an arbitrary $\phi_{l,n}$. By inspection, it is clearly seen that the derivative is

$$\frac{\partial S_2}{\partial \phi_{l,n}} = 1 \tag{13}$$

Furthermore, the derivative of

$$S_3 = \sum_{k=1}^{n_1} \sum_{j=1}^{k} \phi_{j,n} \tag{14}$$

is

$$\frac{\partial S_3}{\partial \phi_{l,n}} = \sum_{k=l}^{n_1} 1 = (n_1 - l)(1) \tag{15}$$

which is perhaps not so obvious, since the lower index on the summation must be changed to l to obtain the correct expression. This is because any inner sums differentiated for $k<l$ do not contain the important quantity $\phi_{l,n}$, and therefore do not contribute to the final sum.

This phenomenon can be generalized in the following rule

$$\frac{\partial}{\partial \phi_{l_2,n}}\left(\sum_{i=l_1}^{n_1} G_i\right) = \sum_{i=l_{max}}^{n_1} \frac{\partial G_i}{\partial \phi_{l_2,n}}$$

$$\text{where } l_{\max} = \begin{cases} \max(l_1,l_2) & \text{if } G_i \text{ contains } \sum_{j=1}^{k} \phi_{j,n} \\ l_1 & \text{otherwise} \end{cases} \tag{16}$$

which shows how the indices of embedded sums must be updated depending on the form of G_i. As is consistent with the rules of differentiation, the chain rule applies when differentiating G_i.

This rule can be quickly illustrated with the example

$$S_4 = \sum_{i=1}^{k} \left(d_{i,n} \sum_{j=1}^{i} \phi_{j,n} \right) \tag{17}$$

If $k=3$, clearly

$$S_4 = d_{1,n}(\phi_{1,n}) + d_{2,n}(\phi_{1,n} + \phi_{2,n}) + d_{3,n}(\phi_{1,n} + \phi_{2,n} + \phi_{3,n}) \tag{18}$$

The derivative with respect to the second angle is

$$\frac{\partial S_4}{\partial \phi_{2,n}} = d_{2,n} + d_{3,n} = \sum_{i=2}^{k} d_{i,n} \tag{19}$$

which can also be obtained from Eq. (16).

Applying the rule shown, the derivative of Eq. (11) with respect to an arbitrary blade segment angle, is given by

$$\begin{aligned}\frac{\partial S_1}{\partial \dot{\phi}_{l,1}} = \sum_{k=l}^{n_1} m_{k,1}\Bigg(\Bigg(&-\sum_{i=1}^{k}\left(\sin\left(\theta + \sum_{j=1}^{i}\phi_{j,1}\right)\right)\left(\dot{\theta} + \sum_{j=1}^{i}\dot{\phi}_{j,1}\right)d_{i,1} \\ &- v\,\dot{\theta}\sin(\theta + a) + \dot{Y}_C + \frac{1}{2}\left(\sin\left(\theta + \sum_{j=1}^{k}\phi_{j,1}\right)\right) \\ &\times\left(\dot{\theta} + \sum_{j=1}^{k}\dot{\phi}_{j,1}\right)d_{k,1}\Bigg)\Bigg(-\sum_{i=l}^{k}\left(\sin\left(\theta + \sum_{j=1}^{i}\phi_{j,1}\right)\right)d_{i,1} \\ &+ \frac{1}{2}\left(\sin\left(\theta + \sum_{j=1}^{k}\phi_{j,1}\right)\right)d_{k,1}\Bigg)\Bigg)\end{aligned} \tag{20}$$

The rule given in Eq. (16) was used to generate, by hand, the mass matrix, $[\mathbf{M}]$, and parts of the forcing vector, $\{\mathbf{Q}\}$. These expressions were carefully checked by selecting a finite number of blade segments for each blade, n_1 and n_2, and differentiation segment, l, and comparing the Maple-differentiated [20] expressions with the hand derived expressions for the same finite values. In all cases, the expressions were equal, indicating that the stated rule works for every expression encountered in this study.

External Forces

External forces, applied to the system as part of $\{\mathbf{Q}\}$, come from two major sources: ship motion and aerodynamics.

Ship Motion. Ship motion is believed to be an important component of a blade sailing study [21]. Ship motion depends on the condition of the sea, including wave spectrum characterized by significant wave height and modal period, and on the way the ship responds to the sea. The latter is dependent on ship geometrical and inertial properties and the operating conditions of the ship including heading relative to the principal wave direction and ship speed. Real sea profiles are made up of wave components having many different amplitudes and frequencies, and can therefore be assumed to be the sum of an infinite number of sine waves. If wave amplitude is plotted in the frequency domain, it can be seen that the sea is typically represented by a gamma distribution.

When exposed to some sea conditions, a ship will respond with motion in six degrees of freedom: three translational motions, surge, sway, and heave; and three angular motions, roll, pitch, and yaw. These motions can also be shown to have gamma distributions of amplitude in the frequency domain. Thus, ship motion is often obtained by multiplying the spectrum of ship response, called response amplitude operators (RAOs) [22], by the incoming wave spectrum [23]. This yields a representation of the ship motion in the frequency domain, which can be generated to give motion at the ship center of mass, or at some other point such as the flight deck. A time history representation of the ship motion in each degree of freedom can be obtained by summing a finite number of frequency components to arrive at a realistic approximation of the actual motion. In general, 40 frequency components are considered sufficient to approximate motion for each degree of freedom.

Since the amplitudes and frequencies used to approximate ship motion have been extensively measured and researched [24], the displacement of each degree of freedom, k, can be determined computationally by summing sine waves of 40 different frequencies using

$$D_{\text{ship}k} = \sum_{j=1}^{40} (A_{k,j}\sin(\omega_{k,j}t + \psi_{k,j})) \tag{21}$$

where $A_{k,j}$, $\omega_{k,j}$, and $\psi_{k,j}$ are the amplitude, frequency, and random phase for the kth ship degree of freedom and the jth frequency component. The amplitudes and frequencies are found using simulated models of ship behavior, and the phase is generated ran-

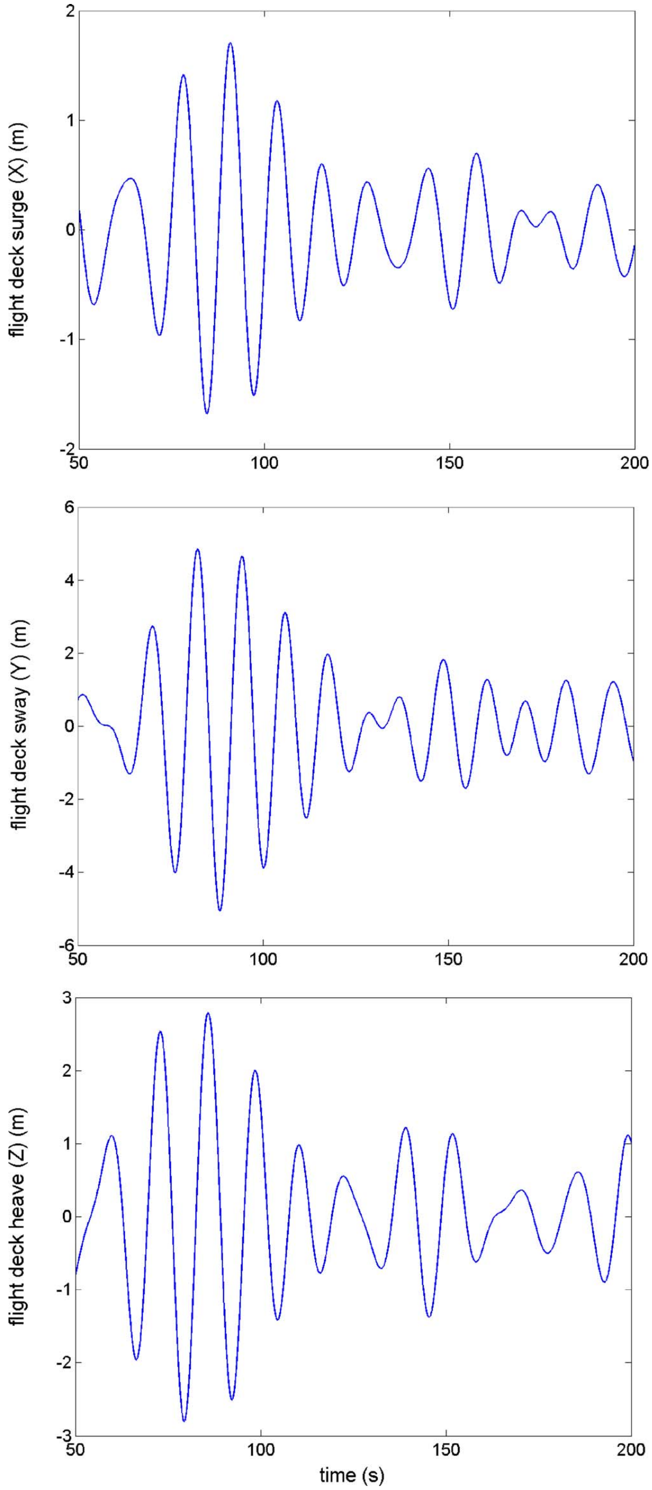


Fig. 5 Ship surge, sway, and heave

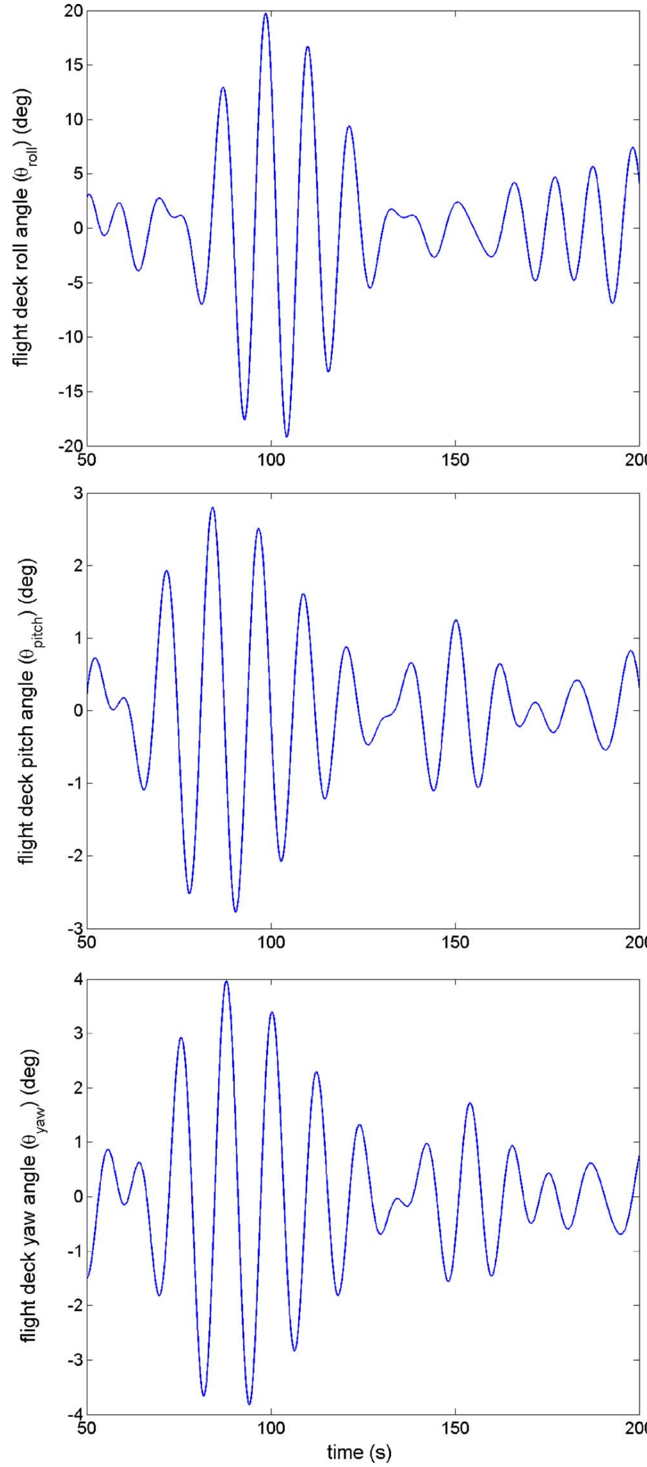


Fig. 6 Ship roll, pitch, and yaw

domly to enhance simulation fidelity [25,26].

Representative ship motions at the flight deck of a Canadian patrol frigate in sea state 5 are shown in Figs. 5 and 6. This ship has a length of 134 m, a width of 16 m, and a displacement of 4770 t. Sea state 5 is characterized by a significant wave height of 4 m.

The velocity of each degree of freedom can be calculated similarly, by differentiating Eq. (21) with respect to time, t

$$V_{\mathrm{ship}k} = \sum_{j=1}^{40} (A_{k,j}\omega_{k,j} \cos(\omega_{k,j}t + \psi_{k,j})) \qquad (22)$$

Only three components of the six degree-of-freedom motion affect the planar model. In the roll plane, (Y,Z), only sway, heave, and roll apply. In the pitch plane, (X,Z), only surge, heave, and pitch apply.

Once the ship motions have been determined, they can be used along with the relative displacements and velocities of the helicopter body to calculate the forces that act through the suspension.

The fact that the suspension properties are defined in the helicopter reference frame while the displacements and velocities are defined in the global reference frame must be considered.

Therefore, the relative displacements and velocities between the ship and helicopter must be transformed to the helicopter-fixed reference frame for suspension force evaluation, and the resulting forces transformed back to the global system. The applied suspension forces, $F_{\text{ship}Y}$, $F_{\text{ship}Z}$, and the applied suspension moment, $M_{\text{ship}\theta}$ are calculated using

$$\begin{Bmatrix} F_{\text{ship}Y} \\ F_{\text{ship}Z} \\ M_{\text{ship}\theta} \end{Bmatrix} = [\mathbf{R}] \begin{bmatrix} k_y & 0 & 0 \\ 0 & k_z & 0 \\ 0 & 0 & k_\theta \end{bmatrix} [\mathbf{R}]^T \begin{Bmatrix} Y_{\text{ship}} - Y_C \\ Z_{\text{ship}} - Z_C \\ \theta_{\text{ship}} - \theta \end{Bmatrix} + [\mathbf{R}] \begin{bmatrix} c_y & 0 & 0 \\ 0 & c_z & 0 \\ 0 & 0 & c_\theta \end{bmatrix} [\mathbf{R}]^T \begin{Bmatrix} \dot{Y}_{\text{ship}} - \dot{Y}_C \\ \dot{Z}_{\text{ship}} - \dot{Z}_C \\ \dot{\theta}_{\text{ship}} - \dot{\theta} \end{Bmatrix} \tag{23}$$

where

$$[\mathbf{R}] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{24}$$

The rotational matrix $[\mathbf{R}]$ is used to perform the coordinate transformations.

Aerodynamics. The aerodynamic forces, which come from a highly unsteady flow field, are a major contributor to blade sailing. In order to accurately estimate the aerodynamic forces, the airwake in the vertical-streamwise plane, (Y,Z), for a typical frigate in beam winds has been characterized through wind tunnel tests. The tests were completed in the vertical wind tunnel facility at the National Research Council of Canada using hot wire X probes and scaled models of a typical frigate flight deck at a number of different ship roll angles. The airwake model, obtained from the experiments, describes variations in mean wind speed, direction, and turbulence intensity with position in the (Y,Z) plane and with ship roll angle.

The aerodynamic forces on the blade are calculated from the model based on the wind tunnel experiments in the classical manner using angle of attack, wind velocity, and the appropriate coefficient of lift [27]. Drag does not act in the plane of the model. The collective and cyclic blade angle inputs supplied by the pilot are also known to affect blade sailing, and can also be considered.

Although a blade sailing study is incomplete without aerodynamic effects, they have been excluded from this paper in order to focus on the governing dynamics.

Model Validation

Since this model is intended for use in helicopter blade sailing, the helicopter and ship motion parameters were chosen based on the properties of a representative maritime helicopter and frigate obtained from manufacturers, experimentation, and other sources. The dynamic model was validated by comparing the simulation results with expected dynamic behavior obtained from experiments. In an effort to understand the effect of segment number, the blade response for one, two, three, four, and seven blade segments was studied.

Blade property selection is critical for capturing realistic blade motion. The blade segment properties can be specified individually to allow for variation in properties with radius. The segment properties depend on the number of segments being used. The mass and geometric properties were obtained approximately from data provided by a helicopter manufacturer. The blade weight is distributed such that the inboard-most meter of the blade is significantly heavier per unit length than the rest of the rotor. This section was selected to be the first segment, and the remaining rotor was divided evenly into the remaining segments. The mass

Table 1 Segment properties for blade with four segments

No.	Fraction of blade length (%)	Fraction of blade mass (%)	Rotational spring stiffness (Nm/rad)	Rotational damping coefficient (N ms/rad)
1	13	46	320,000	4000
2	29	20	320,000	4000
3	29	16	320,000	4000
4	29	18	320,000	4000

moments of inertia were calculated by considering the blade segment as a thin rod. The properties for three, four, and seven segments were selected in this manner; the two segment model was created by combining the first and second, and third and fourth segments from the four segment model.

In the nonrotating planar case, the rotor blade motion due to ship motion is not sufficient to lift the blade off the droop stops. Thus the root flap rotational spring was given a constant stiffness and the blade behavior was studied assuming that semi-rigid and articulated blades behave similarly.

The damping and stiffness of the connecting rotational springs were obtained using the results from an experiment conducted on a full scale helicopter with articulated blades. The data were gathered by deflecting the tip of the blade down using a rope, releasing the rope, and video taping the response. The blades did not lift off the droop stops. This type of test is referred to as a "pull test." The stiffness and damping properties were selected by comparing the digitized experimental data with the simulation results. A blade "drop test" was simulated, in which the blades were allowed to settle after being released from horizontal. The experimental data suggest a blade structural damping ratio of about 0.05 based on the measured logarithmic decrement. This value falls well within the expected range.

Although stiffness and damping can vary with radius, they were assumed constant in this study. Detailed blade-specific data can be included when specific helicopter studies are conducted and the properties of each helicopter become available.

A summary of the chosen properties for a four-segment blade is given in Table 1. The representative blade length was taken to be approximately 8.5 m, and the representative blade weight was taken to be approximately 200 kg.

Figure 7 shows the experimental and simulated results for a helicopter blade "pull test." Note that the test used to tune the

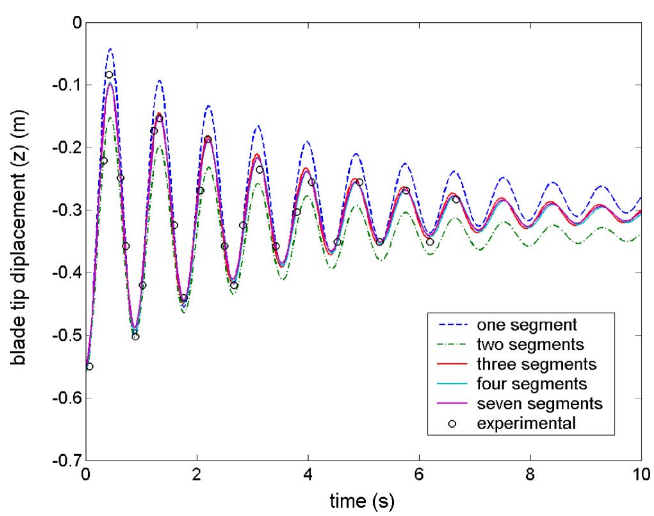


Fig. 7 Blade tip deflection in free vibration for different blade segmentation

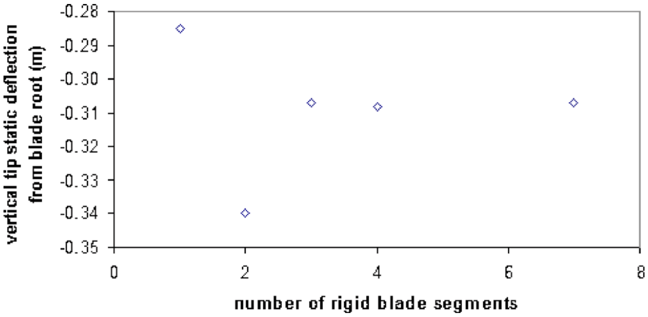


Fig. 8 Blade static deflection for experimental natural frequency with blade segmentation

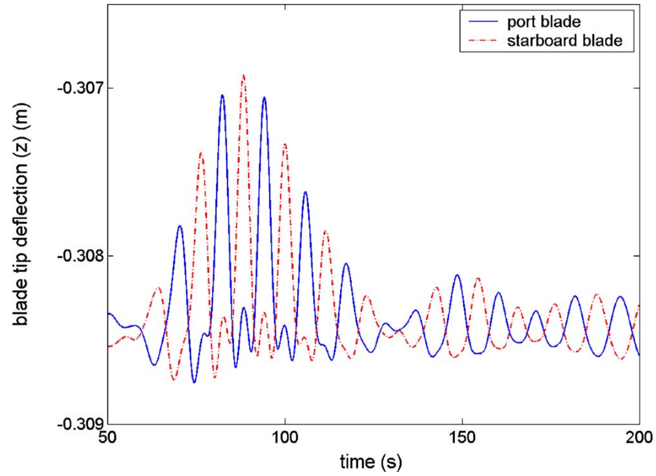


Fig. 10 Blade tip deflection due to sway in the roll plane

blade properties was a "drop test." The results converge well with the experimental response for three or more blade segments.

The experimental data and the simulated data do not agree exactly. This can be attributed to a number of things, including some vertical measurement error induced by the method used. Since the numerical model includes helicopter suspension stiffness and damping but not friction, a model of the blades alone was found to agree more closely with the experimental data.

Convergence is also demonstrated by the static deflection of the blades for varying numbers of segments, as shown in Fig. 8.

According to the static and dynamic convergence tests, three blade segments are sufficient to provide convergence on the tip static deflection for this relatively stiff sample blade. A four-segment model was used to complete further simulations.

The suspension response was also validated using a drop test. Figure 9 shows the results of dropping the helicopter from zero suspension and blade deflection with both rigid and flexible blades. With the blades rigid, the suspension response is characterized by a single overshoot and a settling time of about 1 s. These are typical values. With the blades flexible, the natural frequency of the blades contributes to suspension motion beyond 1 s. The vertical suspension stiffness and damping values are approximately 1,160,000 N/m and 122,800 Ns/m respectively; the helicopter mass was approximated as 13,000 kg. The damping ratio required to achieve the approximate required suspension response is 0.5.

Results

Simulations were run with the model in both the roll plane and the pitch plane in order to determine the effect of ship motion on blade deflection. Although ship motion alone cannot cause blade sailing, it may be a significant contributor to the motion.

The blade deflections resulting from each separate motion direction were examined, by separating out the individual motions at the flight deck and applying them independently. This is an important point because roll isolated at the ship center of mass will result in both roll and sway at the flight deck. The individual motions were then applied simultaneously. In the roll plane, Figs. 10–12 show the blade deflections due to each motion separately.

The effect of sway is very small, inducing only a few millimeters of blade tip deflection. Due to sway only, the blades move out of phase with one another. The ship heave and roll motions induce somewhat more significant blade deflections; about 4 cm each. The blades move in phase when subjected to heave only. The blades move out of phase when subjected to roll. An interesting result of blade deflection under roll is that the value of the mean of oscillation amplitude is increased during severe roll due to centripetal effects from the flight deck rotation.

The blade deflection due to all three motions in the roll plane is shown in Fig. 13.

The port blade deflections are smaller than the starboard deflections because the roll and heave motions are more in phase on the port side of the model. The total blade deflection confirms that blade sailing could potentially be influenced by ship motion.

In the pitch plane, Figs. 14 and 15 show the blade deflections due to surge and pitch separately. The deflections due to heave

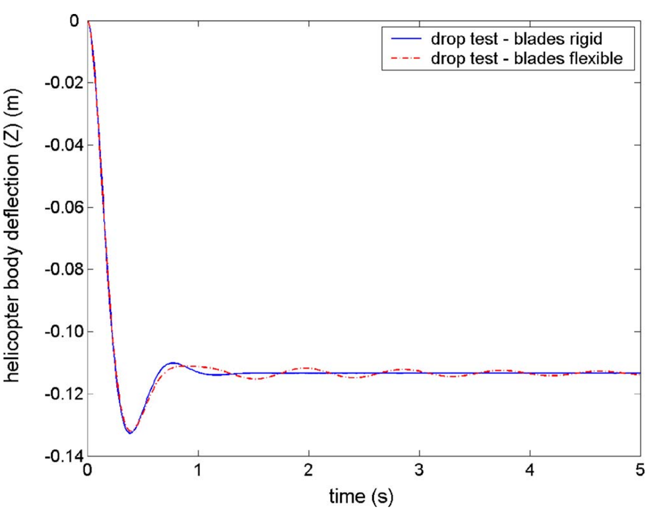


Fig. 9 Suspension drop test with blades rigid and flexible

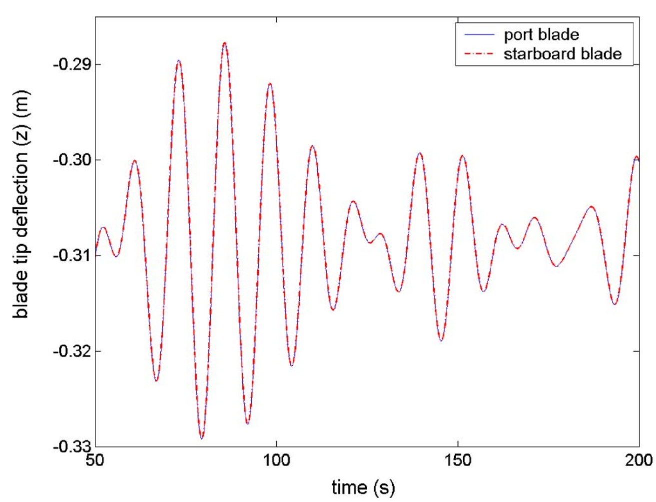


Fig. 11 Blade tip deflection due to heave in the roll plane

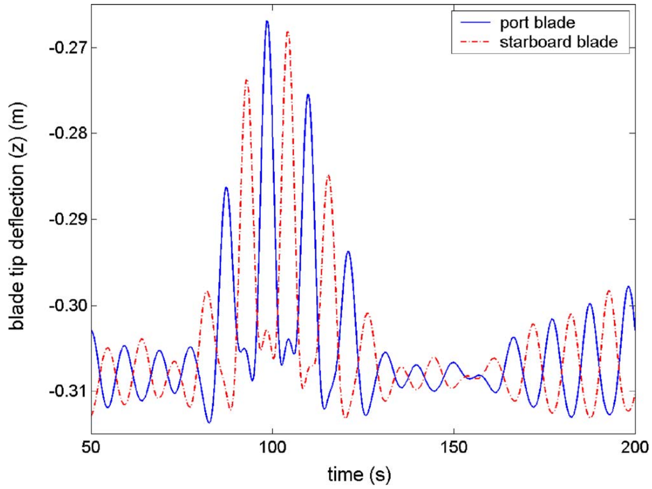


Fig. 12 Blade tip deflection due to roll in the roll plane

only in the pitch plane are identical to the heave-only motions in the roll plane; this was verified through simulation.

Surge and pitch do not induce much blade deflection when applied independently. This is supported by the fact that the blade

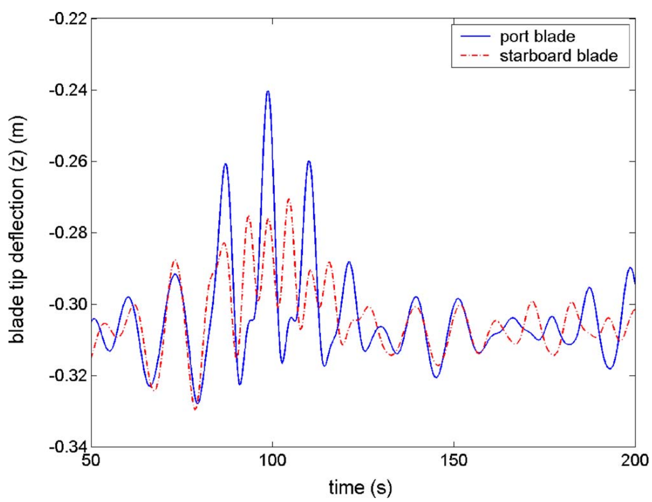


Fig. 13 Blade tip deflection in the roll plane

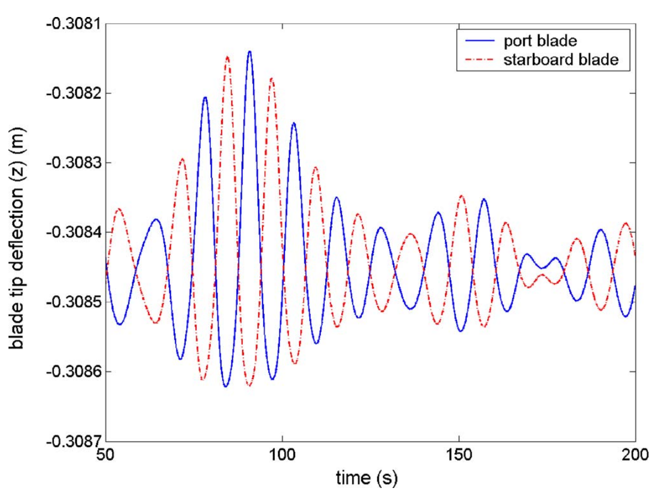


Fig. 14 Blade tip deflection due to surge in the pitch plane

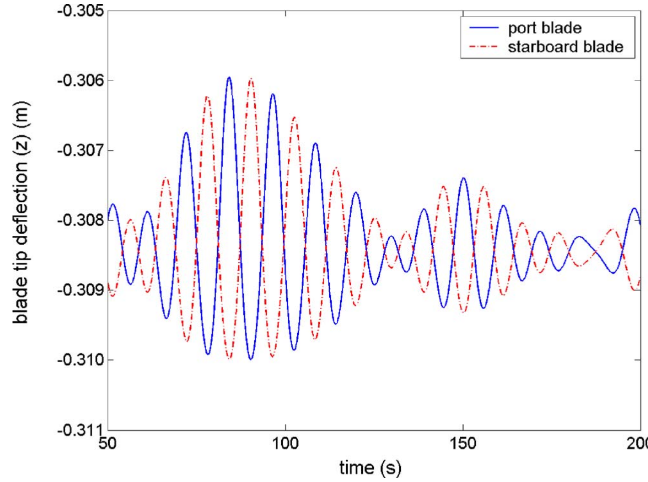


Fig. 15 Blade tip deflection due to pitch in the pitch plane

deflection in the pitch plane due to all three motions, as shown in Fig. 16, closely resembles motion due to heave only.

Conclusions

Based on the results from this study, the following conclusions can be drawn:

1. An elastic blade and landing gear model have been validated for further studies of blade sailing;
2. The equations of motion for the model with an arbitrary number of rigid blade segments has been derived and implemented;
3. A rotor blade composed of rigid beam segments can be used to approximate the behavior of a continuous flexible beam representative of a rotor blade provided enough segments are used; and
4. Ship motion has an impact on the motion of nonrotating blades.

This paper establishes the fundamental dynamic approach required for the study of blade sailing, and the dynamic equations can be easily advanced in order to study the phenomenon. In addition to aerodynamic effects, which are already being addressed in the research, three-dimensional dynamics, including variable blade rotation, torsional flexibility, lead/lag flexibility,

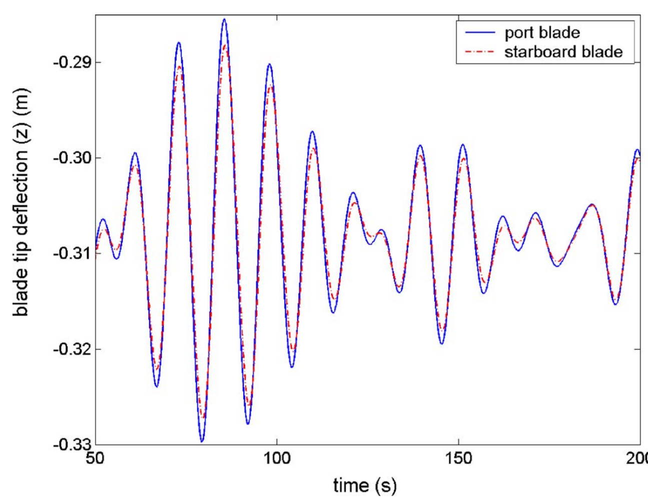


Fig. 16 Blade tip deflection in the pitch plane

and structural coupling shall be included. Once the aerodynamic and dynamic models have been refined, a validation wind tunnel test using a scaled rotor shall be completed.

Acknowledgment

For donations of time and information in support of this work, special thanks are due to: Major Chris Brown, from 424 Transport and Rescue Canadian Forces Squadron in 8th Wing from Trenton, ON, Canada, for facilitating helicopter field testing; and Professor Angelo Mingarelli, from Carleton University in Ottawa, ON, Canada, for mathematical consultation. The project is supported both financially and actively by Carleton University and the Department of Mechanical and Aerospace Engineering, the Natural Sciences and Engineering Research Council of Canada (NSERC), the National Research Council Institute for Aerospace Research (NRC-IAR), and Defence Research and Development Canada (DRDC).

Nomenclature

A = ship motion amplitude
$[\mathbf{A}]$ = arbitrary matrix
C = helicopter suspension attachment point
D = ship displacement
F = applied force
F = applied force due to ship motion
G = differentiation quantity
$[\mathbf{I}]$ = identity matrix
J = helicopter mass moment of inertia or (subscripts) blade segment moment of inertia
M = applied moment due to ship motion
$[\mathbf{M}]$ = model mass matrix
Q = applied force
$\{\mathbf{Q}\}$ = force vector
$[\mathbf{R}]$ = rotational matrix
S = sample equation
T = kinetic energy
U = potential energy
V = ship velocity
X, Y, Z = global inertial coordinates fixed in space or distance in that direction
a = geometric constant
b = geometric constant
$\{\mathbf{b}\}$ = arbitrary vector
c = Suspension or blade segment damping coefficient
d = length of ith segment of nth blade
g = acceleration due to gravity
k = suspension or segment spring stiffness
$l_{1,2,r,cg}$ = helicopter dimensions
m = mass of the helicopter body or (with subscripts) blade segment
n_1 = number of segments in port blade
n_2 = number of segments in starboard blade
n_{dof} = number of model degrees of freedom
q = model generalized coordinate
$\{\mathbf{q}\}$ = model generalized coordinate
t = time
v = geometric constant
w = geometric constant
x, y, z = local coordinates fixed to the helicopter or distance in that direction
$\{\mathbf{x}\}$ = arbitrary degree of freedom vector
α = equation of motion coefficient
β = equation of motion coefficient
γ = equation of motion coefficient
ϵ = equation of motion coefficient
θ = model generalized coordinate - helicopter orientation
ϕ = model generalized coordinate - relative blade segment angle
ω = ship motion frequency
ψ = ship motion phase angle

Subscripts

C = quantity at point C on helicopter
Y = in the Y direction
Z = in the Z direction
i = embedded sum index
j = embedded sum index
k = embedded sum index
i,n = of the ith blade segment of nth blade
k,j = of the kth ship dof and the jth frequency component
p = generalized coordinate index
ship = related to ship
y = in the y direction
z = in the z direction
θ = in the θ direction

Appendix

Some geometric constant are defined as

$$v = \sqrt{l_{cg}^2 + 2l_{cg}l_r + l_r^2 + l_1^2} \tag{A1}$$

$$a = \arctan\left(\frac{l_{cg} + l_r}{l_1}\right) \tag{A2}$$

$$w = \sqrt{l_{cg}^2 + 2l_{cg}l_r + l_r^2 + l_2^2} \tag{A3}$$

$$b = \arctan\left(\frac{l_{cg} + l_r}{l_2}\right) \tag{A4}$$

Since they depend on predefined properties of the helicopter, defining them in this way simplifies the subsequent equations.

The kinetic energy contributions are

$$T_1 = \tfrac{1}{2}m(\dot{Y}_C - l_{cg}(\cos(\theta))\dot{\theta})^2 + \tfrac{1}{2}m(\dot{Z}_C - l_{cg}(\sin(\theta))\dot{\theta})^2 \tag{A5}$$

$$T_2 = \tfrac{1}{2}J\dot{\theta}^2 \tag{A6}$$

$$\begin{aligned} T_3 = \tfrac{1}{2}\sum_{k=1}^{n_1} m_{k,1}\Bigg(\Bigg(&-\sum_{i=1}^{k}\Bigg(\sin\Bigg(\theta + \sum_{j=1}^{i}\phi_{j,1}\Bigg)\Bigg)\Bigg(\dot{\theta} + \sum_{j=1}^{i}\dot{\phi}_{j,1}\Bigg)d_{i,1} \\ &- v\dot{\theta}\sin(\theta + a) + \dot{Y}_C + \frac{1}{2}\Bigg(\sin\Bigg(\theta + \sum_{j=1}^{k}\phi_{j,1}\Bigg)\Bigg) \\ &\times\Bigg(\dot{\theta} + \sum_{j=1}^{k}\dot{\phi}_{j,1}\Bigg)d_{k,1}\Bigg)^2 + \Bigg(\sum_{i=1}^{k}\Bigg(\cos\Bigg(\theta + \sum_{j=1}^{i}\phi_{j,1}\Bigg)\Bigg) \\ &\times\Bigg(\dot{\theta} + \sum_{j=1}^{i}\dot{\phi}_{j,1}\Bigg)d_{i,1} + v\dot{\theta}\cos(\theta + a) + \dot{Z}_C \\ &- \frac{1}{2}\Bigg(\cos\Bigg(\theta + \sum_{j=1}^{k}\phi_{j,1}\Bigg)\Bigg)\Bigg(\dot{\theta} + \sum_{j=1}^{k}\dot{\phi}_{j,1}\Bigg)d_{k,1}\Bigg)^2\Bigg) \end{aligned} \tag{A7}$$

$$T_4 = \frac{1}{2}\sum_{k=1}^{n_1} J_{k,1}\left(\dot{\theta} + \sum_{j=1}^{k}\dot{\phi}_{j,1}\right)^2 \tag{A8}$$

$$T_5 = \frac{1}{2}\sum_{k=1}^{n_2} m_{k,2}\left(\left(\sum_{i=1}^{k}\left(\sin\left(\theta+\sum_{j=1}^{i}\phi_{j,2}\right)\right)\left(\dot{\theta}+\sum_{j=1}^{i}\dot{\phi}_{j,2}\right)d_{i,2}\right.\right.$$
$$- w\dot{\theta}\sin(b-\theta) + \dot{Y}_C - \frac{1}{2}\left(\sin\left(\theta+\sum_{j=1}^{k}\phi_{j,2}\right)\right)$$
$$\times\left(\dot{\theta}+\sum_{j=1}^{k}\dot{\phi}_{j,2}\right)d_{k,2}\Bigg)^2 + \left(-\sum_{i=1}^{k}\left(\cos\left(\theta+\sum_{j=1}^{i}\phi_{j,2}\right)\right)\right.$$
$$\times\left(\dot{\theta}+\sum_{j=1}^{i}\dot{\phi}_{j,2}\right)d_{i,2} - w\dot{\theta}\cos(b-\theta) + \dot{Z}_C$$
$$\left.\left. + \frac{1}{2}\left(\cos\left(\theta+\sum_{j=1}^{k}\phi_{j,2}\right)\right)\left(\dot{\theta}+\sum_{j=1}^{k}\dot{\phi}_{j,2}\right)d_{k,2}\right)^2\right) \tag{A9}$$

$$T_6 = \frac{1}{2}\sum_{k=1}^{n_2} J_{k,2}\left(\dot{\theta}+\sum_{j=1}^{k}\dot{\phi}_{j,2}\right)^2 \tag{A10}$$

The potential energy contributions are

$$U_1 = \tfrac{1}{2}k_Y Y_C^2 + \tfrac{1}{2}k_\theta \theta^2 + \tfrac{1}{2}k_Z Z_C^2 \tag{A11}$$

$$U_2 = gm(l_{cg}\cos(\theta) + Z_C) \tag{A12}$$

$$U_3 = \sum_{k=1}^{n_1}\frac{1}{2}k_{k,1}\phi_{k,1}^2 \tag{A13}$$

$$U_4 = \sum_{k=1}^{n_1} m_{k,1} g\left(Z_C + \sum_{i=1}^{k}\left(\sin\left(\theta+\sum_{j=1}^{i}\phi_{j,1}\right)\right)d_{i,1} + v\,\sin(\theta + a)\right.$$
$$\left. - \tfrac{1}{2}\left(\sin\left(\theta+\sum_{j=1}^{k}\phi_{j,1}\right)\right)d_{k,1}\right) \tag{A14}$$

$$U_5 = \sum_{k=1}^{n_2}\frac{1}{2}k_{k,2}\phi_{k,2}^2 \tag{A15}$$

$$U_6 = \sum_{k=1}^{n_2} m_{k,2} g\left(Z_C - \sum_{i=1}^{k}\left(\sin\left(\theta+\sum_{j=1}^{i}\phi_{j,2}\right)\right)d_{i,2} + w\,\sin(b-\theta)\right.$$
$$\left. + \frac{1}{2}\left(\sin\left(\theta+\sum_{j=1}^{k}\phi_{j,2}\right)\right)d_{k,2}\right) \tag{A16}$$

References

[1] Newman, S. J., 1999, "The Phenomenon of Helicopter Rotor Blade Sailing," Proceedings of the Institution of Mechanical Engineers Part G, **213**(G), pp. 347–363.

[2] Keller, J. A., 2001, "Analysis and Control of the Transient Aeroelastic Response of Rotors During Shipboard Engagement and Disengagement Operations," Ph.D. thesis, The Pennsylvania State University, University Park, PA.

[3] Newman, S. J., 1990, "A Theoretical Model for Predicting the Blade Sailing Behaviour of a Semi-Rigid Rotor Helicopter," Vertica, **14**(4), pp. 531–544.

[4] Smith, E. C., Keller, J. A., and Kang, H., 1998, "Recent Developments in the Analytical Investigation of Shipboard Rotorcraft Engage and Disengage Operations," *Proceedings of 15th NATO RTO Meeting on Fluid Dynamics Problems of Vehicles Operating Near or In the Air-Sea Interface*, Amsterdam, The Netherlands, October, Research and Technology Organization, Applied Vehicle Technology.

[5] Geyer, W. P., Smith, E. C., and Keller, J. A., 1996, "Validation and Application of a Treatment Aeroelastic Analysis for Shipboard Engage/Disengage Operations," *Proceedings of American Helicopter Society 52th Annual Forum*, Washington, D.C., June, American Helicopter Society.

[6] Geyer, W. P. Jr., Smith, E. C., and Keller, J. A., 1998, "Aeroelastic Analysis of Transient Blade Dynamics During Shipboard Engage/Disengage Operations," J. Aircr., **35**(3), 445–453.

[7] Keller, J. A., and Smith, E. C., 1999, "Analysis and Control of the Transient Shipboard Engagement Behavior of Rotor Systems," *Proceedings of American Helicopter Society 55th Annual Forum*, Montreal, QC, May, American Helicopter Society.

[8] Geyer, W. P. Jr., and Smith, E. C., 1995, "Aeroelastic Analysis of Transient Blade Dynamics During Shipboard Engage/Disengage Operations," *Proceedings of the 2nd International AHS Aeromechanics Specialists Meeting*, Alexandria, VA, October, American Helicopter Society, pp. 8-91–8-114.

[9] Keller, J. A., and Smith, E. C., 2003, "Active Control of Gimballed Rotors Using Swashplate Actuation During Shipboard Engagement Operations," J. Aircr., **40**(4), pp. 726–733.

[10] Bottasso, C. L., and Bauchau, O. A., 2001, "Multibody Modeling of Engage and Disengage Operations of Helicopter Rotors," J. Am. Helicopter Soc., **46**(4), pp. 290–300.

[11] Kang, H., and He, C., 2004, "Modeling and Simulation of Rotor Engagement and Disengagement During Shipboard Operations," *Annual Forum Proceedings of the American Helicopter Society*, Baltimore, MD, June 7–10, Vol. 1, pp. 315–324.

[12] Newman, S. J., 1992, "The Application of a Theoretical Blade Sailing Model to Predict the Behaviour of Articulated Helicopter Rotors, Aeronaut. J., June/July, pp. 233–239.

[13] Hurst, D. W., and Newman, S. J., 1988, "Wind Tunnel Measurements of a Ship Induced Turbulence and the Prediction of Helicopter Rotor Blade Response," Vertica, **12**(3), pp. 267–278.

[14] Newman, S. J., 1995, "The Verification of a Theoretical Helicopter Rotor Blade Sailing Method by Means of Windtunnel Testing," Aeronaut. J., Feb., pp. 41–51.

[15] Keller, J. A., and Smith, E. C., 1999, "Experimental and Theoretical Correlation of Helicopter Rotor Blade-Droop Stop Impacts," J. Aircr., **36**(2), pp. 443–450.

[16] Langlois, R. G., and Anderson, R. J., 2005, "A Tutorial Presentation of Alternative Solutions to the Flexible Beam on a Rigid Cart Problem," CSME Transactions, **29**(3), pp. 357–374.

[17] Huston, R. L., 1991, "Computer Methods in Flexible Multibody Dynamics," Int. J. Numer. Methods Eng., **32**(8), pp. 1657–1668.

[18] Ginsberg, J. H., 1998, *Advanced Engineering Dynamics*, Cambridge University Press, Cambridge, UK.

[19] Thomson, W. T., and Dahleh, M. D., 1998, *Theory of Vibration With Applications*, 5th ed., Pearson Education, Upper Saddle River, NJ.

[20] Heck, A., 2003, *Introduction to Maple*, 3rd ed., Springer, New York.

[21] Val Healey, J., 1987, "The Prospects for Simulating the Helicopter/Ship Interface," Nav. Eng. J., March, pp. 45–63.

[22] McTaggart, K. A., 1997, "Shipmo7: An Updated Strip Theory Program for Predicting Ship Motions and Sea Loads in Waves," Defence Research Establishment Atlantic, DREA, Technical Memorandum No. 96/243.

[23] Lloyd, A. R. J. M., 1998, *Seakeeping: Ship Behaviour in Rough Weather*, rev. ed., Gosport, UK.

[24] Lewis, E. V., ed., 1989, *Principles of Naval Architecture*, 2nd ed., The Society of Naval Architects and Marine Engineers, Jersey City, NJ.

[25] Phillion, R. H., and Langlois, R. G., 2004, "Avoiding Repetition in Ship Motion Simulation," *Proceedings of the 2004 CSME Forum*, London, Ontario, Canada, June, Canadian Society for Mechanical Engineers.

[26] Langlois, R. G., and Sopher, G., 2001, "Importance of Random Frequency Spacing in Ship Motion," *Proceedings of the 18th Canadian Congress of Applied Mechanics*, St. John's, Newfoundland, Canada, June.

[27] Wall, A. S., Zan, S. J., Langlois, R. G., and Afagh, F. F., 2005, "A Numerical Model for Studying Helicopter Blade Sailing in an Unsteady Airwake, Volume RTO-MP-AVT-123," *Flow-Induced Unsteady Loads and the Impact on Military Vehicles, Proceedings*, Budapest, Hungary, April, North American Treaty Organization.

Tarun Kant
Professor
e-mail: tkant@civil.iitb.ac.in

Sandeep S. Pendhari
Research Scholar
e-mail: spendhari@iitb.ac.in

Yogesh M. Desai
Professor
e-mail: desai@civil.iitb.ac.in

Department of Civil Engineering,
Indian Institute of Technology Bombay,
Powai, Mumbai-400 076,
India

A Novel Finite-Element–Numerical-Integration Model for Composite Laminates Supported on Opposite Edges

An attempt is made here to devise a new methodology for an integrated stress analysis of laminated composite plates wherein both in-plane and transverse stresses are evaluated simultaneously. The method is based on the governing three-dimensional (3D) partial differential equations (PDEs) of elasticity. A systematic procedure is developed for a case when one of the two in-plane dimensions of the laminate is considered infinitely long (y direction) with no changes in loading and boundary conditions in that direction. The laminate could then be considered in a two-dimensional (2D) state of plane strain in x-z plane. It is here that the governing 2D PDEs are transformed into a coupled system of first-order ordinary differential equations (ODEs) in transverse z direction by introducing partial discretization in the finite inplane direction x. The mathematical model thus reduces to solution of a boundary value problem (BVP) in the transverse z direction in ODEs. This BVP is then transformed into a set of initial value problems (IVPs) so as to use the available efficient and effective numerical integrators for them. Through thickness displacement and stress fields at the finite element discrete nodes are observed to be in excellent agreement with the elasticity solution. A few new results for cross-ply laminates under clamped support conditions are also presented for future reference and also to show the generality of the formulation. [DOI: 10.1115/1.2722770]

Keywords: plane-strain, partial discretization, laminate, boundary value problem, finite element method, numerical integration method

1 Introduction

Composite materials possess ideal engineering properties and therefore these materials are used in many engineering fields. A three-dimensional (3D) elasticity solution of laminated composite beams or plates or shells is extremely complex. Pagano [1–3], Srinivas and Rao [4], and Srinivas et al. [5] have given flexure, vibration, and buckling response of simply-supported rectangular plates and laminates by analytically solving the governing boundary value problem (BVP) defined by 3D partial differential equations (PDEs). However, these solutions lack generality. Their solutions have been used, over the last three decades, as benchmark solutions by researchers involved in developing general numerical techniques and also by those concerned with the range of applicability of the approximate two-dimensional (2D) plate/shell and one-dimensional (1D) beam/arch theories [6–23]. Accurate estimation of interlaminar stresses is a major concern in the design of laminated composites to avoid delamination. In the available approaches [24], the in-plane stresses are first computed in the first phase of any general laminate analysis. The transverse interlaminar stresses are then estimated by integrating the 3D elasticity equilibrium equations in the second post-processing phase, but serious computational and analytical problems are associated with this second post-processing phase involving accuracy and inconsistency of mathematical model itself.

Taking a cue from the foregoing development, an attempt is made here to extend the strategy of transforming the governing system of PDEs to a system of ODEs for elastostatic problems whose behavior is mathematically formulated as a two-point BVP governed by a set of linear first-order ordinary differential equations (ODEs)

$$\frac{d}{dz}\mathbf{y}(z) = \mathbf{A}(z)\mathbf{y}(z) + \mathbf{p}(z) \tag{1}$$

in the domain $z_1 < z < z_2$.

BVP in ODEs, not only describe one-dimensional (1D) elastostatic problems exactly but also 2D and 3D problems approximately whose behavior is governed by a system of PDEs. Conceptualizing a finite element (FE) discretization in the lamina plane, a set of implicit first-order ODEs is obtained. The solution vector $\mathbf{y}(z)$ of which consists of a set of primary dependent variables (stress components and the corresponding displacements on the lamina plane) whose number equals the order of the PDE system times the number of discrete FE mesh nodes. Availability of efficient, accurate, and, above all, proven robust ODE numerical integrators for IVPs helps in obtaining the set of primary variables at all nodal points through the thickness. Ingenuity lies here in transforming the BVP into a set of initial value problems (IVPs) [25]. Furthermore, the secondary set of dependent variables over the entire nodal set is simply computed by substitution of the values of the primary variables on the right hand side of algebraic expressions, node by node.

2 Partial Discretization Formulation

A laminate supported on two opposite edges $x=0$, a, and loaded transversely by distributed load, which is independent of y is considered. The dimension of the laminate in the y direction is infinite. The thickness h is composed of a number of isotropic and/or orthotropic layers bonded together and whose principal material

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received May 3, 2006; final manuscript received January 3, 2007. Review conducted by Sanjay Govindjee.

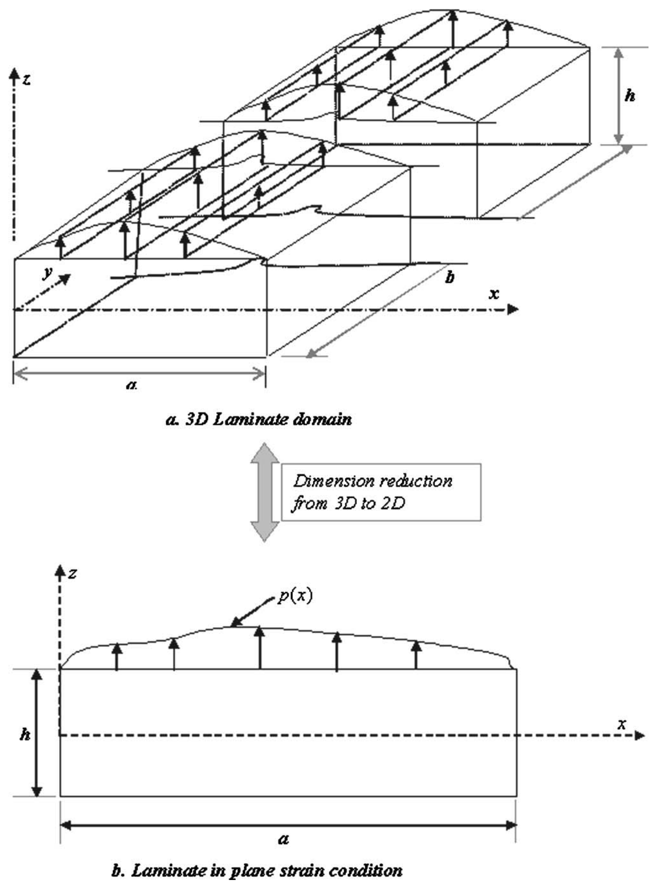


Fig. 1 Laminate subjected to transverse loading

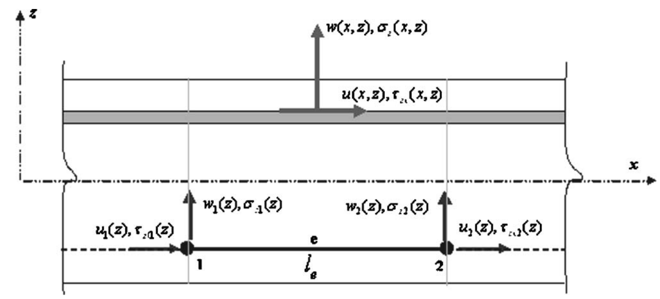


Fig. 2 Linear finite element with dependent variables

directions are coincident with the geometrical coordinate axis. Under such a condition, the laminate is in a 2D state of plane strain in x-z plane (Fig. 1).

The 2D differential equations of equilibrium are

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{zx}}{\partial z} + B_x = 0$$

$$\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \sigma_z}{\partial z} + B_z = 0 \tag{2}$$

where B_x and B_z are the body forces per unit volume in x and z directions, respectively.

The material constitute relations for each layer can be written as

$$\begin{Bmatrix} \sigma_x \\ \sigma_z \\ \tau_{zx} \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & 0 \\ 0 & 0 & C_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_z \\ \gamma_{zx} \end{Bmatrix} \tag{3}$$

The stiffness coefficients C_{ij} are the elastic constants derived by setting $\varepsilon_y = \gamma_{xy} = \gamma_{yz} = 0$ in the 3D material stiffness matrix and are given in the Appendix. The general linear strain-displacement relations in 2D can be written as,

$$\varepsilon_x = \frac{\partial u}{\partial x}, \quad \varepsilon_z = \frac{\partial w}{\partial z}, \quad \text{and} \quad \gamma_{zx} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \tag{4}$$

Equations (2)–(4) have eight unknowns, u, w, ε_x, ε_z, γ_{zx}, σ_x, σ_z, and τ_{zx}. It is to be noted that continuity of transverse stresses and the displacement fields (Fig. 2) are the essential requirements for the accurate analysis of layered components [1–3]. These conditions are naturally enforced in the present formulation. Through a simple algebraic manipulation of the above three sets of Eqs. (2)–(4), a system of PDEs involving four dependent variables $u, w, \tau_{zx}, \sigma_z$ are obtained as follows:

$$\frac{\partial u}{\partial z} = \frac{\tau_{zx}}{C_{33}} - \frac{\partial w}{\partial x}$$

$$\frac{\partial w}{\partial z} = \frac{1}{C_{22}}\left[\sigma_z - C_{21}\frac{\partial u}{\partial x}\right]$$

$$\frac{\partial \tau_{zx}}{\partial z} = \left[-C_{11} + \left(\frac{C_{12}C_{21}}{C_{22}}\right)\right]\frac{\partial^2 u}{\partial x^2} - \frac{C_{12}}{C_{22}}\frac{\partial \sigma_z}{\partial x} - B_x$$

$$\frac{\partial \sigma_z}{\partial z} = -\frac{\partial \tau_{zx}}{\partial x} - B_z \tag{5}$$

This set of dependent variables is called a "primary set," which is naturally defined at a plane z=a constant, and the secondary de-

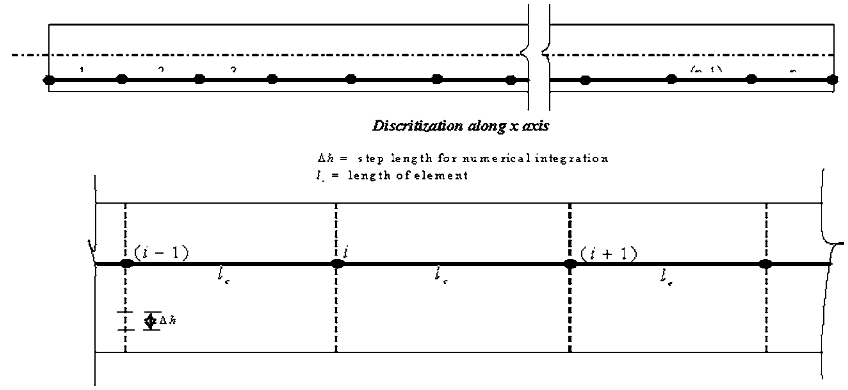


Fig. 3 Linear elements (concept of partial discritization)

Table 1 Boundary conditions (BCs)

Group	Edge	BCs on displacement field	BCs on stress field
A	$x=0$ and a	$w=0$	—
	$x=a/2$	$u=0$	$\tau_{zx}=0$
	$z=h/2$	—	$\sigma_z=p(x)$; $\tau_{zx}=0$
	$z=-h/2$	—	$\sigma_z=0$; $\tau_{zx}=0$
B	$x=0$ and a	$w=0$ and $u=0$	—
	$z=h/2$	—	$\sigma_z=p(x)$; $\tau_{zx}=0$
	$z=-h/2$	—	$\sigma_z=0$; $\tau_{zx}=0$

pendent variable σ_x can simply be expressed as a function of the primary set of variables as follows:

$$\sigma_x=\left(C_{11}-\frac{C_{12}C_{21}}{C_{22}}\right)\frac{\partial u}{\partial x}+\frac{C_{12}}{C_{22}}\sigma_z \tag{6}$$

It is noted that the primary set of variables (u,w,τ_{zx},σ_z) is a function of independent coordinates x and z. It is proposed to carry out FE discretization in only the x direction such that the discrete dependent vector $\mathbf{y}(z)$ will be a only function of independent coordinate z, and a system of coupled discrete first-order ODEs connecting all FE nodes results. This new formulation is described below, first with reference to a two-noded linear element in the x direction with mixed set of primary variables as nodal degrees of freedom (Fig. 2).

The approximate variation of displacements field over the element domain along the longitudinal axis x can be written as

$$u \simeq \hat{u}(x,z)=u_1(z)N_1(x)+u_2(z)N_2(x)$$

$$w \simeq \hat{w}(x,z)=w_1(z)N_1(x)+w_2(z)N_2(x) \tag{7}$$

and from the basic relations of theory of elasticity it can be shown that

$$\tau_{zx} \simeq \hat{\tau}_{zx}(x,z)=\tau_{zx1}(z)N_1(x)+\tau_{zx2}(z)N_2(x)$$

$$\sigma_z \simeq \hat{\sigma}_z(x,z)=\sigma_{z1}(z)N_1(x)+\sigma_{z2}(z)N_2(x) \tag{8}$$

where $N_1=1-(x/l_e)$ and $N_1=x/l_e$.

Substituting Eqs. (7) and (8) into Eq. (5), the domain residuals are obtained as

$$\frac{\partial \hat{u}(x,z)}{\partial z}+\frac{\partial \hat{w}(x,z)}{\partial x}-\frac{\hat{\tau}_{zx}(x,z)}{C_{33}}=R_{1D}(x)$$

$$\frac{\partial \hat{w}(x,z)}{\partial z}+\frac{C_{21}}{C_{22}}\frac{\partial \hat{u}(x,z)}{\partial x}-\frac{\hat{\sigma}_z(x,z)}{C_{22}}=R_{2D}(x)$$

$$\frac{\partial \hat{\tau}_{zx}(x,z)}{\partial z}+\left(C_{11}-\frac{C_{12}C_{21}}{C_{22}}\right)\frac{\partial^2 \hat{u}(x,z)}{\partial x^2}+\frac{C_{12}}{C_{22}}\frac{\partial \hat{\sigma}_z(x,z)}{\partial x}+\hat{B}_x(x,z)$$
$$=R_{3D}(x)$$

$$\frac{\partial \hat{\sigma}_z(x,z)}{\partial z}+\frac{\partial \hat{\tau}_{zx}(x,z)}{\partial x}+\hat{B}_z(x,z)=R_{4D}(x) \tag{9}$$

The strong Bubnov-Galerkin weighted residual statements [26] can then be written with the help of Eq. (9) as follows:

$$\int_0^{l_e} N_i(x)\left(\frac{\partial \hat{u}(x,z)}{\partial z}+\frac{\partial \hat{w}(x,z)}{\partial x}-\frac{\hat{\tau}_{zx}(x,z)}{C_{33}}\right)dx=0 \tag{10}$$

$$\int_0^{l_e} N_i(x)\left(\frac{\partial \hat{w}(x,z)}{\partial z}+\frac{C_{21}}{C_{22}}\frac{\partial \hat{u}(x,z)}{\partial x}-\frac{\hat{\sigma}_z(x,z)}{C_{22}}\right)dx=0 \tag{11}$$

$$\int_0^{l_e} N_i(x)\left[\frac{\partial \hat{\tau}_{zx}(x,z)}{\partial z}+\left(C_{11}-\frac{C_{12}C_{21}}{C_{22}}\right)\frac{\partial^2 \hat{u}(x,z)}{\partial x^2}+\frac{C_{12}}{C_{22}}\frac{\partial \hat{\sigma}_z(x,z)}{\partial x}\right.$$
$$\left.+\hat{B}_x(x,z)\right]dx=0 \tag{12}$$

$$\int_0^{l_e} N_i(x)\left(\frac{\partial \hat{\sigma}_z(x,z)}{\partial z}+\frac{\partial \hat{\tau}_{zx}(x,z)}{\partial x}+\hat{B}_z(x,z)\right)dx=0 \tag{13}$$

Equation (12), which contains a second-order derivative of $\hat{u}$, is replaced by its weak form with the help of integration by parts as follows:

$$\int_0^{l_e} N_i(x)\frac{\partial \hat{\tau}_{zx}(x,z)}{\partial z}dx-\int_0^{l_e}\frac{dN_i(x)}{dx}\left(C_{11}-\frac{C_{12}C_{21}}{C_{22}}\right)\frac{\partial \hat{u}(x,z)}{\partial x}dx$$
$$+\int_0^{l_e} N_i(x)\frac{C_{12}}{C_{22}}\frac{\partial \hat{\sigma}_z(x,z)}{\partial x}dx+\left[N_i(x)\left(C_{11}\right.\right.$$
$$\left.\left.-\frac{C_{12}C_{21}}{C_{22}}\right)\frac{\partial \hat{u}(x,z)}{\partial x}\right]_0^{l_e}+\int_0^{l_e} N_i(x)\hat{B}_x(x,z)dx=0 \tag{14}$$

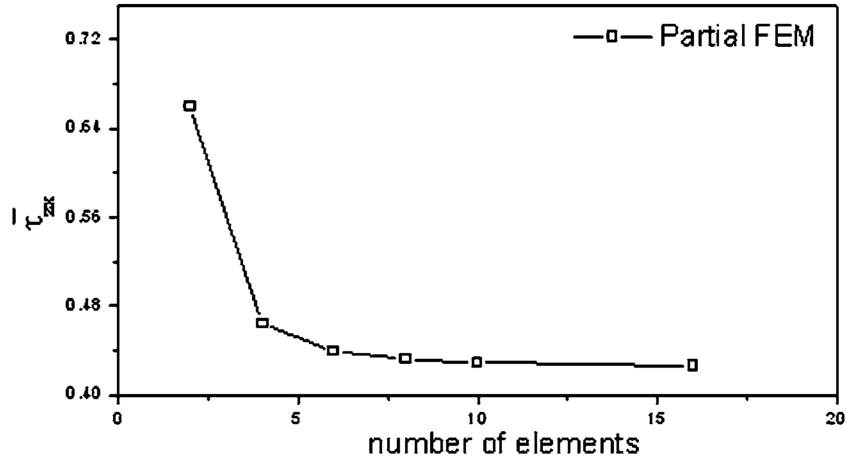
On substitution for approximate functions from Eqs. (7) and (8), the following eight semi-discrete equations are obtained:

$$\begin{bmatrix} A_{11}^e & 0 & 0 & 0 & A_{15}^e & 0 & 0 & 0 \\ 0 & A_{22}^e & 0 & 0 & 0 & A_{26}^e & 0 & 0 \\ 0 & 0 & A_{33}^e & 0 & 0 & 0 & A_{37}^e & 0 \\ 0 & 0 & 0 & A_{44}^e & 0 & 0 & 0 & A_{48}^e \\ A_{51}^e & 0 & 0 & 0 & A_{55}^e & 0 & 0 & 0 \\ 0 & A_{62}^e & 0 & 0 & 0 & A_{66}^e & 0 & 0 \\ 0 & 0 & A_{73}^e & 0 & 0 & 0 & A_{77}^e & 0 \\ 0 & 0 & 0 & A_{84}^e & 0 & 0 & 0 & A_{88}^e \end{bmatrix}\frac{d}{dz}\begin{Bmatrix} u_1^e(z) \\ w_1^e(z) \\ \tau_{zx1}^e(z) \\ \sigma_{z1}^e(z) \\ u_2^e(z) \\ w_2^e(z) \\ \tau_{zx2}^e(z) \\ \sigma_{z2}^e(z) \end{Bmatrix}$$

$$=\begin{bmatrix} 0 & B_{12}^e & B_{13}^e & 0 & 0 & B_{16}^e & B_{17}^e & 0 \\ B_{21}^e & 0 & 0 & B_{24}^e & B_{25}^e & 0 & 0 & B_{28}^e \\ B_{31}^e & 0 & 0 & B_{34}^e & B_{35}^e & 0 & 0 & B_{38}^e \\ 0 & 0 & B_{43}^e & 0 & 0 & 0 & B_{47}^e & 0 \\ 0 & B_{52}^e & B_{53}^e & 0 & 0 & B_{56}^e & B_{57}^e & 0 \\ B_{61}^e & 0 & 0 & B_{64}^e & B_{65}^e & 0 & 0 & B_{68}^e \\ B_{71}^e & 0 & 0 & B_{74}^e & B_{75}^e & 0 & 0 & B_{78}^e \\ 0 & 0 & B_{83}^e & 0 & 0 & 0 & B_{87}^e & 0 \end{bmatrix}\begin{Bmatrix} u_1^e(z) \\ w_1^e(z) \\ \tau_{zx1}^e(z) \\ \sigma_{z1}^e(z) \\ u_2^e(z) \\ w_2^e(z) \\ \tau_{zx2}^e(z) \\ \sigma_{z2}^e(z) \end{Bmatrix}$$

$$+\begin{Bmatrix} 0 \\ 0 \\ p_3^e \\ p_4^e \\ 0 \\ 0 \\ p_7^e \\ p_8^e \end{Bmatrix}$$

which can be written in a compact form as



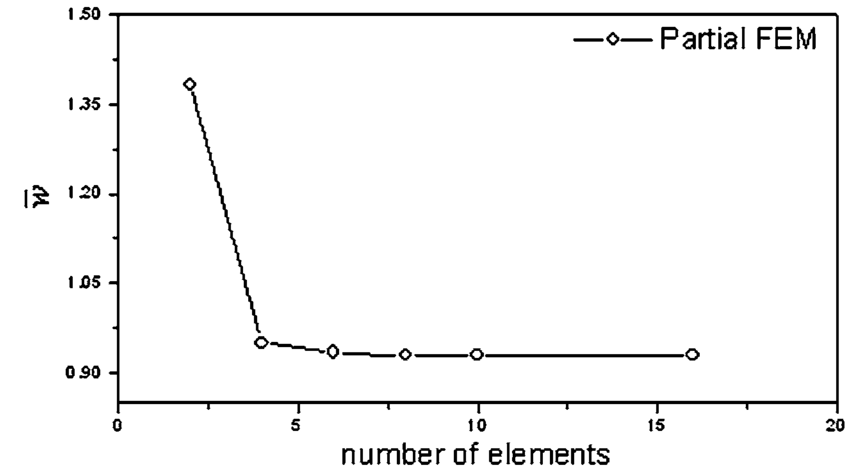


Fig. 4 Convergence of (*a*) maximum transverse shear stress ($\overline{\tau_{zx}}$) and (*b*) mid-plane transverse displacement ($\bar{w}$) with number of elements for a 0 deg/90 deg/0 deg laminate under cylindrical bending

$$A^e(x)\frac{d}{dz}\boldsymbol{y}^e(z) = \mathbf{B}^e(x,z)\boldsymbol{y}^e(z) + \boldsymbol{p}^e(x,z) \tag{15}$$

The elements of matrices $\mathbf{A}^e(x)$, $\mathbf{B}^e(x,z)$ and vector $\mathbf{p}^e(x,z)$ are given in the Appendix. When the total x dimension is discretized with n two-noded elements (Fig. 3), then the semi-discrete system of equation for the entire domain turns out to be

$$\sum_{k=1}^{n} \mathbf{A}^e(x)\frac{d}{dz}\boldsymbol{y}^e(z) = \sum_{k=1}^{n} \mathbf{B}^e(x,z)\boldsymbol{y}^e(z) + \sum_{k=1}^{n} \boldsymbol{p}^e(x,z)$$

or

$$\mathbf{A}(x)\frac{d}{dz}\boldsymbol{y}(z) = \mathbf{B}(x,z)\boldsymbol{y}(z) + \boldsymbol{p}(x,z) \tag{16}$$

Multiplication of Eq. (16) by $[\mathbf{A}(x)]^{-1}$ on both sides results in

$$\frac{d}{dz}\boldsymbol{y}(z) = \mathbf{C}(x,z)\boldsymbol{y}(z) + \boldsymbol{f}(x,z) \tag{17}$$

where $\mathbf{C}(x,z)=[\mathbf{A}(x)]^{-1}\mathbf{B}(x,z)$ and $\boldsymbol{f}(x,z)=[\mathbf{A}(x)]^{-1}\mathbf{p}(x,z)$.

Equation (17) defines the governing equations of a two-point BVP in ODEs in the domain $-(h/2)<z<(h/2)$. $\boldsymbol{y}(z)$ is an m-dimensional (m=number of nodes$\times$4) vector of dependent variables, $\mathbf{C}(x,z)$ is an $m\times m$ coefficient matrix (which is a function of element geometry along x and material properties variation both in the x and z directions), and $f(x,z)$ is an m-dimensional vector of nonhomogeneous (loading) terms. Any $m/2$ elements of $\boldsymbol{y}(z)$ are prescribed at the two ends, $z=-(h/2)$ and $h/2$ as boundary conditions. It is clearly seen that mixed and/or nonhomogeneous boundary conditions are easily admitted in this formulation. The basic approach to the numerical integration of the BVP defined by Eq. (17) is to transform the given BVP into a set of IVPs—one particular (nonhomogeneous) and $m/2$ complimentary (homogeneous). Clearly, the reason behind this is the availability of a number of successful and well-tested algorithms for numerical solution of IVPs in ODEs. The solution of the original BVP defined by Eq. (17) is obtained by forming a linear combination of one nonhomogeneous and $m/2$ homogeneous solutions so as to satisfy the boundary conditions at $z=h/2$. This gives rise to a system of $m/2$ linear algebraic equations, the solution of which determines the unknown $m/2$ components of the vector of initial values $\boldsymbol{y}(z)$. Then a final numerical integration of Eq. (17) with completely known initial vector of dependent variables $\boldsymbol{y}(z)$ produces the desired results. It is intended here to extend the applicability of this procedure, which is documented by Kant and Ramesh [25].

3 Numerical Studies

A two-noded linear element with mixed (displacements/stresses) degrees of freedom is employed in the present numerical

Table 2 Comparison of normalized transverse displacement ($\bar{w}$), in-plane normal stress ($\overline{\sigma_x}$), and transverse shear stress ($\overline{\tau_{zx}}$) of two-layered (0 deg/90 deg) unsymmetric laminates under cylindrical bending

Aspect ratio	Source	Stresses/displacement $\overline{\sigma_x}$ $(a/2, h/2)$	$\overline{\sigma_x}$ $(a/2, -h/2)$	$\overline{\tau_{zx}}$ (0,max)	$\bar{w}$ $(a/2, 0)$
4	Partial FEM	0.2325	−1.8142	0.6983	4.6826
		(−3.0037)	(−3.3355)	(2.6157)	(−0.2705)
	Pagano [1][a]	0.2397	−1.8768	0.6805	4.6953
	Enhblom and Ochoa [9]	0.1864	−1.7371	N/A	N/A
	Lu and Liu [13]	0.2232	−1.8750	N/A	4.7773
10	Partial FEM	0.1952	−1.7403	0.7343	2.9503
		(−1.5633)	(−1.4162)	(1.0319)	(−0.1185)
	Pagano [1][a]	0.1983	−1.7653	0.7268	2.9538
	Lu and Liu [13]	0.2000	−1.7500	N/A	3.0000
20	Partial FEM	0.1890	−1.7241	0.7432	2.6980
		(−1.3570)	(−1.4405)	(1.1432)	(−0.1739)
	Pagano [1][a]	0.1916	−1.7493	0.7348	2.7027
50	Partial FEM	0.1866	−1.7196	0.7465	2.6267
		(−1.6341)	(−1.4500)	(1.2752)	(−0.2127)
	Pagano [1][a]	0.1897	−1.7449	0.7371	2.6323

[a]The analytical solution given in this paper is programed by the present authors and numerical results for various aspect ratios, not available in the original paper are obtained and presented here.
N/A=results are not available.

study involving both validation and solution of new problems. A computer code incorporating the present methodology was developed in FORTRAN-90. The accuracy of the proposed new formulation for layered composites is established by comparison of the present numerical results with that of elasticity solution [1] and also with others. In all the examples, the layer elastic coefficients are those of a unidirectional graphite/epoxy composite

$$\mathbf{E_L} = 25 \times 10^6 \text{ psi}; \quad \mathbf{E_T} = 10^6 \text{ psi}; \quad \mathbf{G_{LT}} = 0.5 \times 10^6 \text{ psi}$$

$$\mathbf{G_{TT}} = 0.2 \times 10^6 \text{ psi}; \quad \boldsymbol{\nu}_{\mathbf{LT}} = \boldsymbol{\nu}_{\mathbf{TT}} = 0.25$$

where subscripts **L** and **T** refer to the fiber direction and transverse direction perpendicular to fiber direction.

Two support conditions on opposite edges considered here are tabulated in Table 1. All laminates are subjected to sinusoidal transverse load on their top surface. The intensity of sinusoidal loading can be expressed as

$$p(x) = p_0 \sin \frac{\pi x}{a} \tag{18}$$

where p_0 represents the peak intensity of load.

The dependent quantities are nondimensionalized in the following manner:

Table 3 Comparison of normalized transverse displacement ($\bar{w}$), in-plane normal stress ($\overline{\sigma_x}$), and transverse shear stress ($\overline{\tau_{zx}}$) of three-layered (0 deg/90 deg/0 deg) symmetric laminates under cylindrical bending

Aspect ratio	Source	Stresses/displacement $\overline{\sigma_x}$ $(a/2, h/2)$	$\overline{\sigma_x}$ $(a/2, -h/2)$	$\overline{\tau_{zx}}$ (0,max)	$\bar{w}$ $(a/2, 0)$
4	Partial FEM	1.1211	−1.0782	0.4149	2.9134
		(−4.6278)	(−4.7105)	(4.8522)	(0.9074)
	Pagano [1][a]	1.1755	−1.1315	0.3957	2.8872
	Spilker [8]	N/A	N/A	0.3909	2.8410
	Enhblom and Ochoa [9]	0.6256	−0.6318	0.4434	N/A
10	Partial FEM	0.7216	−0.7211	0.4285	0.9308
		(−2.049)	(−2.0644)	(1.0851)	(−0.0859)
	Pagano [1][a]	0.7367	−0.7363	0.4239	0.9316
	Spilker [8]	N/A	N/A	0.4529	0.9312
	Enhblom and Ochoa [9]	0.6373	−0.6373	0.4459	N/A
20	Partial FEM	0.6439	−0.6440	0.4431	0.6152
		(−2.1280)	(−2.1425)	(1.3031)	(−0.3240)
	Pagano [1][a]	0.6579	−0.6581	0.4374	0.6172
50	Partial FEM	0.6211	−0.6211	0.4483	0.5246
		(−2.1581)	(−2.1581)	(1.5402)	(−0.4554)
	Pagano [1][a]	0.6348	−0.6348	0.4415	0.5270

[a]The analytical solution given in this paper is programmed by the present authors and numerical results for various aspect ratios, not available in the original paper are obtained and presented here.
N/A=results are not available.

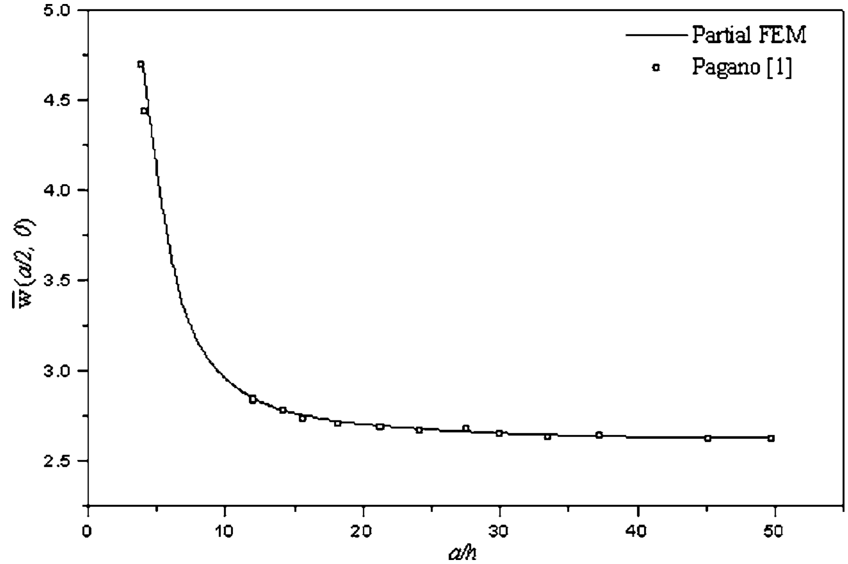


Fig. 5 Variation of normalized transverse displacement ($\bar{w}$) with respect to a/h ratios of 0 deg/90 deg unsymmetric laminates under cylindrical bending

$$\bar{z} = \frac{z}{h}; \quad \bar{u} = \frac{E_2 u(0,z)}{h p_0}; \quad \bar{w} = \frac{100 E_2 h^3 w(a/2,0)}{p_0 a^4}$$

$$\overline{\sigma_x} = \frac{h^2 \sigma_x(a/2,0)}{p_0 a^2}; \quad \overline{\sigma_z} = \frac{\sigma_z(a/2,z)}{p_0}; \quad \overline{\tau_{zx}} = \frac{h \tau_{xz}(0,z)}{p_0 a} \tag{19}$$

in which a bar over a variable defines its nondimensionalized value and the percentage error between present and elasticity solution [1] is calculated as

$$\%\ \text{error} = \frac{\text{Present analysis} - \text{Elasticity solution}}{\text{Elasticity solution}} \times 100 \tag{20}$$

and these are presented in parentheses in Table 1.

A convergence study on number of elements along the x direction and number of steps required for numerical integration in thickness direction is performed first. The method was found to yield converged solution for a laminates in-plane strain with 12–16 elements in the x direction and with 16–20 steps in the thickness, z direction. Convergence plot of midplane transverse displacement ($\bar{w}$) and maximum transverse shear stress ($\overline{\tau_{zx}}$) with the number of elements in the x direction are shown graphically for the symmetric (0 deg/90 deg/0 deg) laminates in Fig. 4 for a/h ratio of 10.

Group A. The examples considered in this group are selected to establish the accuracy of stress predictions through the thickness by the present method. A two-layer unsymmetric (0 deg/90 deg) and a three-layer symmetric (0 deg/90 deg/0 deg) cross-ply square laminates, simply supported on opposite edges in the x direction are considered for this purpose. Boundary conditions are specified in Table 1. The results obtained through present analysis are compared to the 3D elasticity solution given by Pagano [1] and also with available results in the literature [8,9,13] for cylindrical bending. Numerical results for a/h ratios of 4, 10, 20, and 50 are given in Tables 2 and 3 for both configurations. The variation of midplane transverse displacement $\bar{w}(a/2,0)$ with different a/h ratios is shown in Fig. 5. Through thickness variation of normalized in-plane normal stress ($\overline{\sigma_x}$), inplane displacement ($\bar{u}$), transverse shear stress ($\overline{\tau_{xz}}$) and transverse normal stress ($\overline{\sigma_z}$) for a/h ratio of 4 are presented in Figs. 6 and 7 for 0 deg/90 deg unsymmetric laminate and 0 deg/90 deg/0 deg symmetric laminate, respectively. Moreover, through thickness variation of transverse displacement ($\bar{w}$) is depicted in Fig. 8. Excellent agreement is seen between the present and the elasticity solution.

Group B. The examples considered under this group are an extension of group A for clamped end conditions to show the ability of the present formulation to handle problems with general boundary conditions and high stress gradients. The lamination schemes, material properties and geometrical details are kept same as group A. Boundary conditions are specified in Table 1. Numerical results for normal in-plane stress ($\overline{\sigma_x}$) at top and bottom, transverse shear stress ($\overline{\tau_{xz}}$), and transverse displacement ($\bar{w}$) at midplane with different a/h ratios are presented in Table 4 for both 0 deg/90 deg unsymmetric and 0 deg/90 deg/0 deg symmetric laminates. The normal in-plane stress ($\overline{\sigma_x}$) and transverse shear stress ($\overline{\tau_{xz}}$) variations through thickness of laminate with a/h ratio of 4 are shown graphically in Figs. 9 and 10 for 0 deg/90 deg unsymmetric and 0 deg/90 deg/0 deg symmetric laminates, respectively. These results should serve as benchmark solutions for future investigation.

4 Concluding Remarks

A novel partial discretization method with mixed degrees of freedom has been proposed in this paper. It ensures the fundamental elasticity relationship between stress, strain, and displacement fields within the elastic continuum and implicitly maintains the continuity of displacements and transverse stresses at the laminate interface. It is first of its kind of a mixed partial FE model that is based on the solution of a two-point BVP through the thickness of laminates. Good agreement of the results with the elasticity solution suggests that the method is extremely accurate. Generality of the method is proven by incorporation of clamped edge conditions at $x=0$, a. The most significant advantage of the present formulation lies in the fact that both displacement and transverse interlaminar stresses are simultaneously evaluated at the finite element node with the same degree of accuracy through a numerical integration process and thus eliminating the post-processing module that is required in other analytical models for calculation of transverse stresses from in-plane stresses.

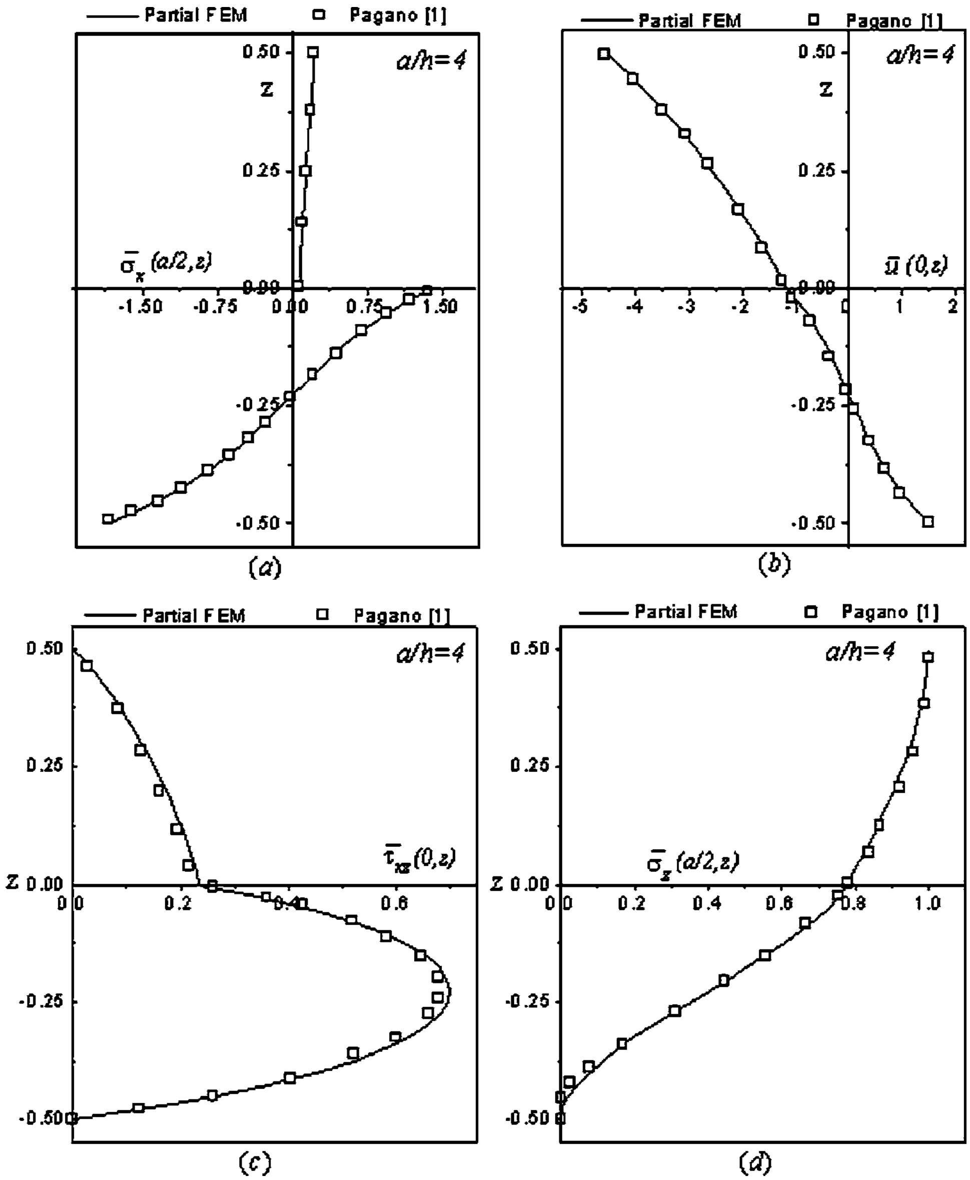


Fig. 6 Variation of normalized (*a*) in-plane normal stress $\overline{\sigma_x}$, (*b*) in-plane displacement $\bar{u}$, (*c*) transverse shear stress $\overline{\tau_{zx}}$, and (*d*) transverse normal stress $\overline{\sigma_z}$ through thickness of 0 deg/90 deg unsymmetric laminate under cylindrical bending

Acknowledgment

Partial support of USIF Indo-US Collaborative Sponsored Research Project No. IND104 (95IU001) is gratefully acknowledged.

Appendix

The stiffness coefficients C_{ij}

$$C_{11} = \frac{E_1(1 - \upsilon_{23}\upsilon_{32})}{\Delta}$$

$$C_{12} = C_{21} = \frac{E_1(\upsilon_{21} + \upsilon_{31}\upsilon_{23})}{\Delta}$$

$$C_{22} = \frac{E_2(1 - \upsilon_{13}\upsilon_{31})}{\Delta}$$

$$C_{33} = G_{xz}$$

were $\Delta = (1 - \upsilon_{12}\upsilon_{21} - \upsilon_{13}\upsilon_{31} - \upsilon_{23}\upsilon_{32} - 2\upsilon_{12}\upsilon_{31}\upsilon_{23})$, E_i=Young's moduli of lamina in the material principle direction ($i=1,2,3$), and υ_{ij}=generalized Poisson's ratios of lamina ($i,j=1,2,3$).

Elements of matrix $A^e(x)$ are

$$A^e_{11} = A^e_{22} = A^e_{33} = A^e_{44} = A^e_{55} = A^e_{66} = A^e_{77} = A^e_{88} = \int_0^{l_e} N_1(x)N_1(x)dx$$

$$= \frac{l_e}{3}$$

$$A^e_{15} = A^e_{26} = A^e_{37} = A^e_{48} = A^e_{51} = A^e_{62} = A^e_{73} = A^e_{84} = \int_0^{l_e} N_1(x)N_2(x)dx$$

$$= \frac{l_e}{6}$$

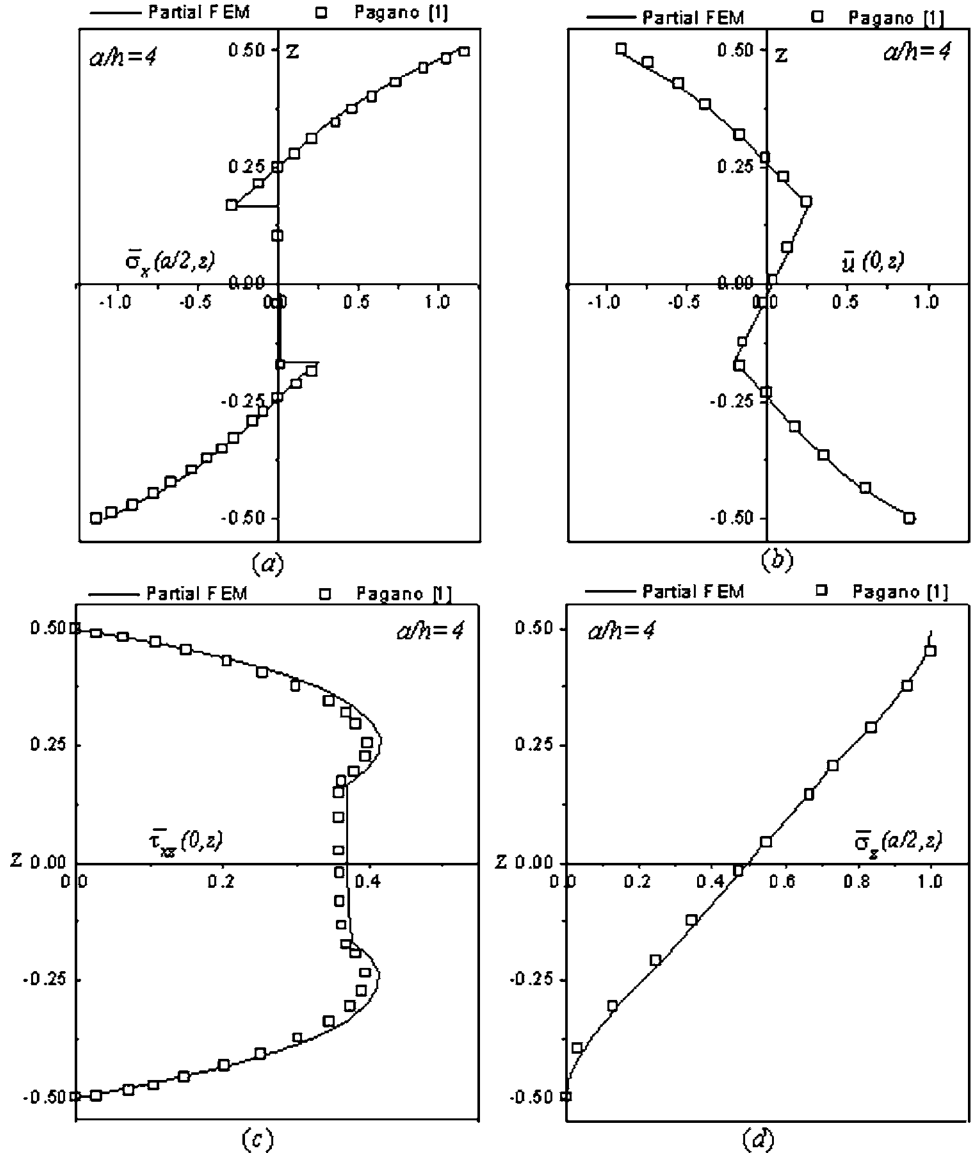


Fig. 7 Variation of normalized (*a*) inplane normal stress $\overline{\sigma_x}$, (*b*) inplane displacement $\bar{u}$, (*c*) transverse shear stress $\overline{\tau_{zx}}$, and (*d*) transverse normal stress $\overline{\sigma_z}$ through thickness of 0 deg/90 deg/0 deg symmetric laminate under cylindrical bending

Elements of matrix $B^e(x,z)$ are

$$B_{12}^e = -\int_0^{l_e} N_1(x)\frac{dN_1(x)}{dx}dx = \frac{1}{2}$$

$$B_{13}^e = \frac{1}{C_{33}}\int_0^{l_e} N_1(x)N_1(x)dx = \frac{l_e}{3C_{33}}$$

$$B_{16}^e = -\int_0^{l_e} N_1(x)\frac{dN_2(x)}{dx}dx = -\frac{1}{2}$$

$$B_{17}^e = \frac{1}{C_{33}}\int_0^{l_e} N_1(x)N_2(x)dx = \frac{l_e}{6C_{33}}$$

$$B_{21}^e = -\frac{C_{21}}{C_{22}}\int_0^{l_e} N_1(x)\frac{dN_1(x)}{dx}dx = \frac{C_{31}}{2C_{22}}$$

$$B_{24}^e = \frac{1}{C_{22}}\int_0^{l_e} N_1(x)N_1(x)dx = \frac{l_e}{3C_{22}}$$

$$B_{25}^e = -\frac{C_{21}}{C_{22}}\int_0^{l_e} N_1(x)\frac{dN_2(x)}{dx}dx = -\frac{C_{21}}{2C_{22}}$$

$$B_{28}^e = \frac{1}{C_{22}}\int_0^{l_e} N_1(x)N_2(x)dx = \frac{l_e}{6C_{22}}$$

$$B_{31}^e = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\int_0^{l_e} \frac{dN_1(x)}{dx}\frac{dN_1(x)}{dx}dx = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{1}{l_e}$$

$$B_{34}^e = -\frac{C_{12}}{C_{22}}\int_0^{l_e} N_1\frac{dN_1(x)}{dx}dx = \frac{C_{12}}{2C_{22}}$$

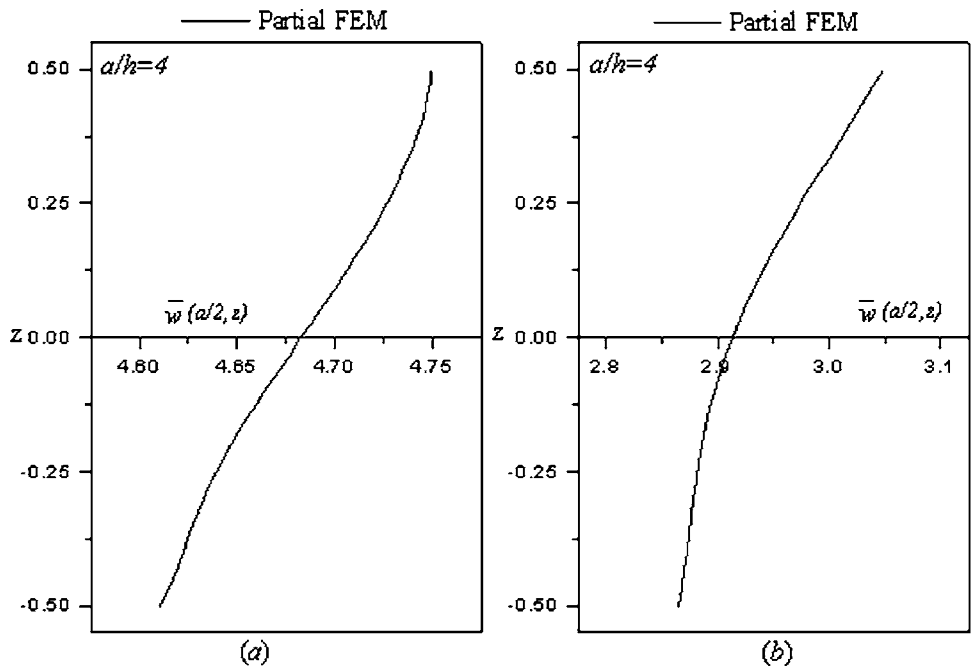


Fig. 8 Variation of normalized transverse displacement $\bar{w}$ through thickness of (*a*) 0 deg/90 deg unsymmetric and (*b*) 0 deg/90 deg/0 deg symmetric laminates under cylindrical bending

$$B_{35}^e = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\int_0^{l_e} \frac{dN_1(x)}{dx}\frac{dN_2(x)}{dx}dx = -\left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{1}{l_e}$$

$$B_{38}^e = -\frac{C_{12}}{C_{22}}\int_0^{l_e} N_1(x)\frac{dN_2(x)}{dx}dx = -\frac{C_{12}}{2C_{22}}$$

$$B_{43}^e = -\int_0^{l_e} N_1(x)\frac{dN_1(x)}{dx}dx = \frac{1}{2}$$

$$B_{47}^e = -\int_0^{l_e} N_1(x)\frac{dN_2(x)}{dx}dx = -\frac{1}{2}$$

$$B_{52}^e = -\int_0^{l_e} N_2(x)\frac{dN_1(x)}{dx}dx = \frac{1}{2}$$

$$B_{53}^e = \frac{1}{C_{33}}\int_0^{l_e} N_2(x)N_1(x)dx = \frac{l_e}{6C_{33}}$$

$$B_{56}^e = -\int_0^{l_e} N_2(x)\frac{dN_2(x)}{dx}dx = -\frac{1}{2}$$

$$B_{57}^e = \frac{1}{C_{33}}\int_0^{l_e} N_2(x)N_2(x)dx = \frac{l_e}{3C_{33}}$$

$$B_{61}^e = -\frac{C_{21}}{C_{22}}\int_0^{l_e} N_2(x)\frac{dN_1(x)}{dx}dx = \frac{C_{21}}{2C_{22}}$$

$$B_{64}^e = \frac{1}{C_{22}}\int_0^{l_e} N_2(x)N_1(x)dx = \frac{l_e}{6C_{22}}$$

$$B_{65}^e = -\frac{C_{21}}{C_{22}}\int_0^{l_e} N_2(x)\frac{dN_2}{dx}(x)dx = -\frac{C_{21}}{2C_{22}}$$

$$B_{68}^e = \frac{1}{C_{22}}\int_0^{l_e} N_2(x)N_2(x)dx = \frac{l_e}{3C_{22}}$$

Table 4 Normalized transverse displacement ($\bar{w}$), in-plane normal stress ($\overline{\sigma_x}$), and transverse shear stress ($\overline{\tau_{zx}}$) of laminates under clamped support condition in-plane strain condition

		Stresses/displacement							
		0 deg/90 deg unsymmetric laminate				0 deg/90 deg/0 deg symmetric laminate			
Aspect ratio	Source	$\overline{\sigma_x}$ $(a/2,h/2)$	$\overline{\sigma_x}$ $(a/2,-h/2)$	$\overline{\tau_{zx}}$ (0,max)	$\bar{w}$ $(a/2,0)$	$\overline{\sigma_x}$ $(a/2,h/2)$	$\overline{\sigma_x}$ $(a/2,-h/2)$	$\overline{\tau_{zx}}$ (0,max)	$\bar{w}$ $(a/2,0)$
4	Partial FEM	0.1241	−0.8041	1.4232	2.7930	0.5511	−0.5166	0.8833	2.0886
10	Partial FEM	0.0909	−0.6697	2.8418	1.0623	0.3320	−0.3304	1.5079	0.4898
20	Partial FEM	0.0739	−0.5956	2.6678	0.6792	0.2428	−0.2424	1.5624	0.2007
50	Partial FEM	0.0649	−0.5746	2.3424	0.5225	0.2065	−0.2064	1.6142	0.1083

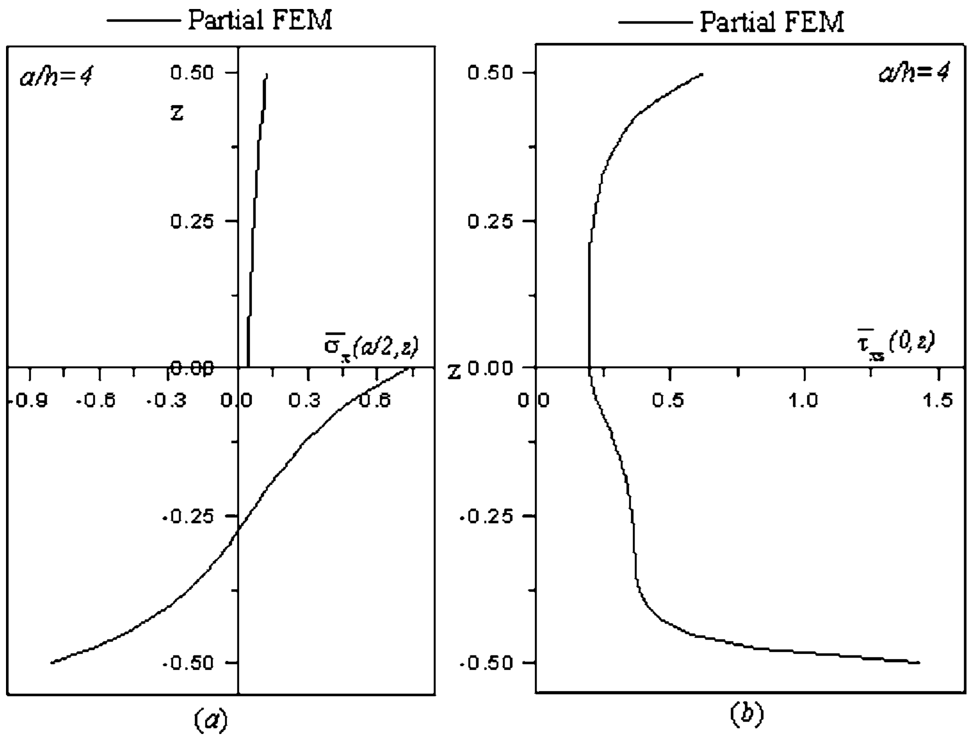


Fig. 9 Variation of normalized (*a*) in-plane normal stress $\overline{\sigma_x}$ and (*b*) transverse shear stress $\overline{\tau_{zx}}$ through thickness of 0 deg/90 deg unsymmetric laminate for clamped supported boundary conditions

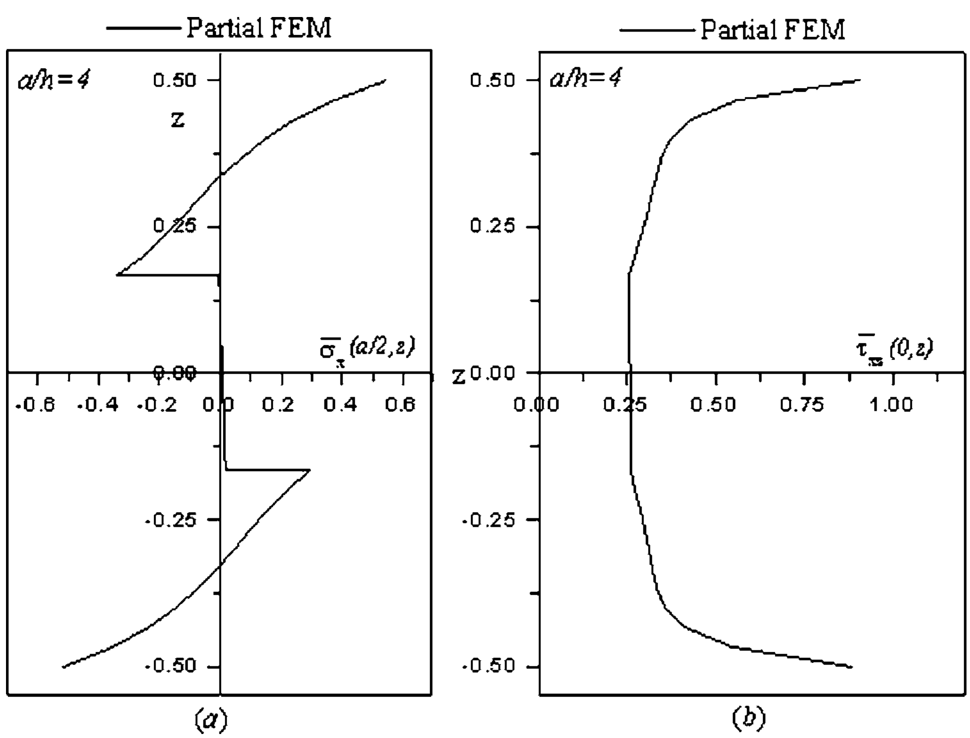


Fig. 10 Variation of normalized (*a*) in-plane normal stress $\overline{\sigma_x}$ and (*b*) transverse shear stress $\overline{\tau_{zx}}$ through thickness of 0 deg/90 deg/0 deg symmetric laminate for clamped supported boundary conditions

$$B_{71}^e = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\int_0^{l_e} \frac{dN_2(x)}{dx}\frac{dN_1(x)}{dx}dx = -\left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{1}{l_e}$$

$$B_{74}^e = -\frac{C_{12}}{C_{22}}\int_0^{l_e} N_2(x)\frac{dN_1(x)}{dx}dx = \frac{C_{12}}{2C_{22}}$$

$$B_{75}^e = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\int_0^{l_e} \frac{dN_2(x)}{dx}\frac{dN_2(x)}{dx}dx = \left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{1}{l_e}$$

$$B_{78}^e = -\frac{C_{12}}{C_{22}}\int_{l_1}^{l_2} N_2(x)\frac{dN_2(x)}{dx}dx = -\frac{C_{12}}{2C_{22}}$$

$$B_{83}^e = -\int_0^{l_e} N_2(x)\frac{dN_1(x)}{dx}dx = \frac{1}{2}$$

$$B_{87}^e = -\int_0^{l_e} N_2(x)\frac{dN_2(x)}{dx}dx = -\frac{1}{2}$$

Elements of vector $p^e(x,z)$ are

$$p_3^e = -\int_0^{l_e} N_1(x)\hat{B}_x(x,z)dx - \left[N_1(x)\left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{d\hat{u}(x,z)}{dx}\right]_0^{l_e}$$

$$p_4^e = -\int_0^{l_e} N_1(x)\hat{B}_z(x,z)dx$$

$$p_7^e = -\int_0^{l_e} N_2(x)\hat{B}_x(x,z)dx - \left[N_2(x)\left(C_{11} - \frac{C_{12}C_{21}}{C_{22}}\right)\frac{d\hat{u}(x,z)}{dx}\right]_0^{l_e}$$

$$p_8^e = -\int_0^{l_e} N_2(x)\hat{B}_z(x,z)dx$$

References

[1] Pagano, N. J., 1969, "Exact Solution for Composite Laminates in Cylindrical Bending," J. Compos. Mater., **3**, pp. 398–411.

[2] Pagano, N. J., 1970, "Exact Solution for Rectangular Bidirectional Composites and Sandwich Plates," J. Compos. Mater., **4**, pp. 20–34.

[3] Pagano, N. J., 1974, "On the Calculation of Interlaminar Normal Stress in Composite Laminates," J. Compos. Mater., **8**, pp. 65–81.

[4] Srinivas, S., and Rao, A. K., 1970, "Bending, Vibration and Buckling of Simply Supported Thick Orthotropic Rectangular Plates and Laminates," Int. J. Solids Struct., **6**, pp. 1463–1481.

[5] Srinivas, S., Joga Rao, C. V., and Rao, A. K., 1970, "An Exact Analysis for Vibration of Simply Supported Homogeneous and Laminated Thick Rectangular Plates," J. Sound Vib., **12**, pp. 187–199.

[6] Pandya, B. N., and Kant, T., 1988, "Higher Order Shear Deformation Theories for Flexure of Sandwich Plates: Finite Element Evaluations," Int. J. Solids Struct., **24**, pp. 1267–1286.

[7] Lo, K. H., Christensen, R. M., and Wu, E. M., 1977, "A High-Order Theory of Plate Deformation—Part 2: Laminated Plate," ASME J. Appl. Mech., **44**, pp. 669–676.

[8] Spilker, R. L., 1982, "Hybrid-Stress Eight-Node Elements for This and Thick Multilayered Laminated Plates," Int. J. Numer. Methods Eng., **18**, pp. 801–828.

[9] Engblom, J. J., and Ochoa, O., 1985, "Through-the-Thickness Stress Predictions for Laminated Plates of Advanced Composite Materials," Int. J. Numer. Methods Eng., **21**, pp. 1759–1776.

[10] Liou, W., and Sun, C. T., 1987, "A Three-Dimensional Hybrid Stress Isoparametric Element for the Analysis of Laminated Composite Plates," Compos. Struct., **25**, pp. 241–249.

[11] Sciuva, M. D., 1987, "An Improved Shear Deformation Theory for Moderately Thick Multilayered Anisotropic Shells and Plates," ASME J. Appl. Mech., **54**, pp. 589–596.

[12] Noor, A. K., and Burton, W. S., 1989, "Stress and Free Vibration Analyses of Multilayered Composite Plates," Compos. Struct., **11**, pp. 183–204.

[13] Lu, X., and Liu, D., 1992, "An Interlaminar Shear Stress Continuity Theory for Both Thin and Thick Composite Laminates," ASME J. Appl. Mech., **59**, pp. 502–509.

[14] Barbero, E. J., 1992, "A 3-D Finite Element for Laminated Composites With 2-D Kinematic Constrains," Compos. Struct., **45**, pp. 263–271.

[15] Chyou, H. A., Sandhu, R. S., and Butalia, T. S., 1995, "Variational Formulation and Finite Element Implementation of Pagano's Theory of Laminated Plates," Mech. Compos. Mater. Struct., **2**, pp. 111–137.

[16] Pai, P. F., 1995, "A New Look at the Shear Correction Factors and Warping Functions of Anisotropic Laminates," Int. J. Solids Struct., **32**, pp. 2295–2313.

[17] Yong, Y. K., and Cho, Y., 1995, "Higher-Order Partial Hybrid Stress, Finite Element Formulation for Laminated Plate and Shell Analysis," Compos. Struct., **57**, pp. 817–827.

[18] Qi, Y., and Knight, N. F., 1996, "A Refined First-Order Shear Deformation Theory and its Justification by Plane-Strain Bending Problem of Laminated Plates," Int. J. Solids Struct., **33**, pp. 49–64.

[19] Aitharaju, V. R., and Averill, R. C., 1999, "C^0 Zigzag Kinematic Displacement Models for the Analysis of Laminated Composites," Mech. Compos. Mater. Struct., **6**, pp. 31–56.

[20] Carrera, E., 2000, "A Priori vs. a Posteriori Evaluation of Transverse Stresses in Multilayered Orthotropic Plates," Compos. Struct., **48**, pp. 245–260.

[21] Alfano, G., Auricchio, F., Rosati, L., and Sacco, E., 2001, "MITC Finite Elements for Laminated Composite Plates," Int. J. Numer. Methods Eng., **50**, pp. 707–738.

[22] Kant, T., and Swaminathan, K., 2001, "Analytical Solutions for Vibration of Laminated Composite and Sandwich Plates Based on a Higher Order Refined Theory," Compos. Struct., **53**, pp. 73–85.

[23] Kant, T., and Swaminathan, K., 2002, "Analytical Solutions for the Static Analysis of Laminated Composite and Sandwich Plates Based on a Higher Order Refined Theory," Compos. Struct., **56**, pp. 329–344.

[24] Kant, T., and Swaminathan, K., 2000, "Estimation of Transverse/Interlaminar Stresses in Laminated Composites—A Selective Review and Survey of Current Developments," Compos. Struct., **49**, pp. 65–75.

[25] Kant, T., and Ramesh, C. K., 1981, "Numerical Integration of Linear Boundary Value Problems in Solid Mechanics by Segmentation Method," Int. J. Numer. Methods Eng., **17**, pp. 1233–1256.

[26] Zienkiewicz, O. C., and Taylor, R. L., 2005, *The Finite Element Method for Solids and Structural Mechanics*, 6th ed., Elsevier Butterworth Heinemann, Oxford.

Fatigue Modeling for Elastic Materials With Statistically Distributed Defects

Ilya I. Kudish
Science & Mathematics Department,
Kettering University,
1700 W. Third Avenue,
Flint, MI 48504
e-mail: ikudish@kettering.edu

*The paper is devoted to formulation and analysis of a new model of structural fatigue that is a direct extension of the model of contact fatigue developed by Kudish (2000, STLE Tribol. Trans., **43**, pp. 711–721). The model is different from other published models of structural fatigue (Collins, J. A., 1993, Failure of Materials and Mechanical Design: Analysis, Prediction, Prevention, 2nd ed., Wiley, New York) in a number of aspects such as statistical approach to material defects, stress analysis, etc. The model is based on fracture mechanics and fatigue crack propagation. The model takes into account local stress distribution, initial statistical distribution of defects versus their size, crack location, and orientation, and material fatigue resistance parameters. The assumptions used for the new model derivation are stated clearly and their validity is discussed. The model considers the kinetics of crack distribution by taking into account the fact that the crack distribution varies with the number of applied loading cycles due to crack growth. A qualitative and quantitative parametric analysis of the model is performed. Some analytical formulas for fatigue life as a function of the initial defect distribution, material fatigue resistance, and stress state are obtained. Examples of application of the model to predicting fatigue of beam bending and torsion and contact fatigue for tapered bearings is presented.* [DOI: 10.1115/1.2722771]

Keywords: fatigue, statistical analysis, crack propagation, life prediction methods, torsion, bending, rolling contact

1 Model Development

We will consider a three-dimensional model assuming that the stress state of the material is given as a function of the number of loading cycles N and the coordinates (x,y,z). The coordinates (x,y,z) are introduced in such a way that the x- and y-axes are directed along the material surface, the z axis is directed perpendicular to the material surface.

1.1 Initial Statistical Defect Distribution. Experiments have demonstrated that structural fatigue in metals is due to the presence of defects such as nonmetallic inclusions, carbides, etc. (see [1,2]). In contact fatigue studies, Dudragne et al. [3] and Murakami et al. [4] (see Tables 1 and 2 in [4]) have stated that in their tests all contact fatigue failures are initiated at subsurface nonmetallic inclusions. Spektor et al. [5] have experimentally showed that the main source of fatigue cracks in bearings is oxide inclusions. In clean steels, fatigue cracks may originate at carbide sites or at grain boundaries. In summary, fatigue cracks originate at different material defects which, in most cases, for the purpose of fatigue life prediction can be treated as small fatigue cracks (see review by Kudish and Burris [6]).

In accordance with practice, we will assume that the material defects are far from each other and do not interact. Suppose there is a characteristic size L_σ in the material that is determined by the typical variations of the material stresses and surface geometry. Let us also assume that there is a size L_f in material such that $L_d \ll L_f \ll L_\sigma$, where L_d is the typical distance between material defects. In other words, we will assume that the defect population in any of such volumes L_f^3 is large enough to ensure an adequate statistical representation of the phenomenon. By doing so, we assume that any parameter variations on the scale of L_f are indistinguishable for the purposes of the fatigue analysis and that any volume L_f^3 can be represented by its center; i.e., a point (x,y,z). Therefore, we can assume that there is an initial statistical defect distribution in the material. For the purpose of further analysis, we will replace each defect by a penny-shaped crack with a diameter approximately equal to the diameter of the defect (see further and [6]) or by a semi-circular shaped surface crack. The usage of penny-shaped subsurface and semi-circular shaped surface cracks is advantageous to our analysis due to the fact that in the accepted approximation the cracks maintain their shape. The orientation of these cracks will be considered later. The initial statistical distribution is described by the probabilistic density function $f(0,x,y,z,l_0)$, such that $f(0,x,y,z,l_0)dl_0dxdydz$ is the number of defects with the radii between l_0 and l_0+dl_0 in the volume $dxdydz$. The defect distribution $f(0,x,y,z,l_0)$ represents the local characteristic of the material defectiveness. The model can be developed for any particular initial distribution $f(0,x,y,z,l_0)$. Some experimental studies [7] show that a distribution of nonmetallic inclusions versus their size resembles a log-normal one. The model analysis can be simplified (see below) for a log-normal initial defect distribution $f(0,x,y,z,l_0)$ versus the defect initial radius l

$$f(0,x,y,z,l_0) = 0, \quad l_0 \leq 0,$$
$$f(0,x,y,z,l_0) = \frac{\rho(0,x,y,z)}{(2\pi)^{1/2}\sigma_{\ln} l_0} \exp\left[-\frac{1}{2}\left(\frac{\ln(l_0) - \mu_{\ln}}{\sigma_{\ln}}\right)^2\right], \quad l_0 > 0 \tag{1}$$

where $\rho(0,x,y,z)$ is the initial crack volume density while $\mu_{\ln}$ and $\sigma_{\ln}$ are the mean and standard deviation of crack radii.

1.2 Fracture Mechanics and Stress Intensity Factors. In experimental studies of steels, including bearing ones, it is established that the crack initiation stage is much shorter than the crack propagation stage (see [8–12]). Soon after cycling loading starts, small fatigue cracks are initiated near material oxide inclusions (see [6]). In the case of very small oxides, fatigue cracks may

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 30, 2006; final manuscript received January 4, 2007. Review conducted by Martin Ostoja-Starzewski.

initiate near carbides or other structural defects.

The fatigue process in material starts as accumulation of dislocations near defects which rapidly grow into microcracks (see [6]) as most of the plastic deformations of the material occur during the first few loading cycles. As the linear fracture mechanics (LEFM) is a natural extension of the dislocation theory, LEFM can be effectively used to describe the mechanisms controlling fatigue. Data on stresses near voids and inclusions ([13,14]) indicate that, even when a crack size is small in comparison with the adjacent inclusion/void, the stress intensity factors at the crack tips/edges can be reasonably well approximated by the stress intensity factors for cracks of combined size of the inclusion/void and the adjacent small cracks. Therefore, for the purpose of stress intensity factor analysis, such combined structures of inclusions and cracks can be replaced by equivalent cracks of slightly larger size. One can assume that in $f(0,x,y,z,l_0)$, the variable l_0 represents crack radius equal to about 1.1–1.2 of the inclusion radius.

The next major assumption employed in the model is that LEFM for quasi-brittle elastic materials is applicable to the analysis of fatigue cracks. To verify the above assumption, the radius r_p of the plasticity zone at a crack tip has to be much smaller than the crack radius l. For a plane stress state we have

$$\frac{2r_p}{l} \ll 1, \quad r_p = \frac{1}{2\pi}\left(\frac{k_1}{\sigma_p}\right)^2 \tag{2}$$

where k_1 is the normal stress intensity factor at the crack edge and σ_p is the material yield stress.

In most cases of structural fatigue, the assumptions of LEFM for quasi-brittle elastic materials hold well during most of the material fatigue life while fatigue cracks are small. A verification of inequality (2) for cracks under typical conditions for contact fatigue is given by Kudish in [15]. Therefore, for small cracks the assumption $2r_p/l \ll 1$ holds and the methods of quasi-brittle fracture mechanics can be applied. For larger cracks with the stress intensity factor $k_1 \approx k_f$ (k_f=material fracture toughness) the latter inequality can get violated. However, as it is shown below, the crack propagation rate is much higher for larger cracks than for smaller ones and, therefore, during almost the entire fatigue process cracks remain small and their behavior can be considered based on LEFM for quasi-brittle materials.

Let us assume that the material contains a number of cracks that are modeled by penny-shaped cuts. According to LEFM, the stress intensity factors at the edge of a single crack can be represented in the form

$$k_1 = F_1(\varphi,\psi)\sigma_1(\pi l)^{1/2}, \quad k_2 = F_2(\varphi,\psi)\sigma_1(\pi l)^{1/2},$$
$$k_3 = F_3(\varphi,\psi)\sigma_1(\pi l)^{1/2} \tag{3}$$

where σ_1 is the maximum of the local principal tensile stress, l is the crack radius, $F_1(\varphi,\psi)$, $F_2(\varphi,\psi)$, and $F_3(\varphi,\psi)$ are certain functions of the crack orientation angles φ and ψ with respect to the xy-plane and some other stresses and geometric parameters describing the particular crack location. It is well known (see [16–19]) that for "small cracks" functions $F_1(\varphi,\psi)$, $F_2(\varphi,\psi)$, and $F_3(\varphi,\psi)$ are practically independent of l. The notion of "small cracks" includes surface (semi-circular shaped) cracks (for which F_1, F_2, and F_3 are independent of l) and cracks that are far from the material surface in comparison with their radius l.

1.3 Direction of Fatigue Crack Propagation. The resultant stress field in material is formed by stresses $\sigma_x(N,x,y,z)$, $\sigma_y(N,x,y,z)$, $\sigma_z(N,x,y,z)$, $\tau_{xz}(N,x,y,z)$, $\tau_{xy}(N,x,y,z)$, and $\tau_{zy}(N,x,y,z)$. There are regions in the material subjected to tensile stress and other regions subjected to compressive stress. Conceptually, there is no difference between the phenomena of structural and contact fatigue as the local response of the loaded material to the same stress in both cases is the same. What is different between these two cases is the stress distribution. As long as the stress levels do not exceed the limits of applicability of the quasi-brittle linear fracture mechanics when plastic zones at crack edges are small, the rest of the material behaves like an elastic solid. The actual stress distributions in cases of structural and contact fatigue are taken into proper account. In both cases there are zones with tensile stresses in material. In contact interactions where compressive stress is usually dominant there are still regions in material subjected to tensile stress caused by contact frictional stress. For contact fatigue and even relatively high compressive residual stress the existence of zones in material with tensile stress is shown by Kudish in [19,20].

It is widely accepted that fatigue cracks are initiated by shear stresses. Experimental and theoretical studies suggest (see [6]) that soon after initiation, fatigue cracks propagate in the direction perpendicular to the local maximum tensile (principal) stress. Therefore, we will assume that fatigue is caused by propagation of penny-shaped subsurface and semi-circular shaped surface cracks under the action of principal maximum tensile stresses. On a plane perpendicular to a principal stress, the shear stresses are equal to zero, i.e., the shear stress intensity factors $k_2=k_3=0$. Therefore, to find the plane of fatigue crack propagation (i.e., the orientation angles α and β) it is necessary to find the direction of the maximum principal tensile stress. The latter is equivalent to the solution of the equations

$$k_2(N,\alpha,\beta,l,x,y,z) = 0, \quad k_3(N,\alpha,\beta,l,x,y,z) = 0 \tag{4}$$

It is important to remember that for the most part of their lives, fatigue cracks remain small. Equations (4) and the fact that for small cracks, $k_{20}=k_2 l^{-1/2}$ and $k_{30}=k_3 l^{-1/2}$ are independent from N and l lead to the conclusion that for cyclic loading with constant amplitude, cracks maintain their shape and the angles α and β characterizing the plane of fatigue crack growth are independent from the number of loading cycles N and the crack radius l. Therefore, α and β are functions of only crack location; i.e., $\alpha = \alpha(x,y,z)$ and $\beta=\beta(x,y,z)$. For most stress fields (excluding stress fields with special symmetry), at any point (x,y,z), it is possible to find several sets of solutions (α_m,β_m) to Eqs. (4). The crack propagation angles α and β are determined by one of these sets of angles (α_n,β_n) for which the normal stress intensity factor $k_1(N,l,x,y,z)$ is maximal.

1.4 Crack Propagation Calculations. Propagation of a fatigue crack subjected to only normal tensile stress can be described by the initial-value problem

$$\frac{dl}{dN} = F[l, \max_T(k_1), k_{\text{th}}, k_f], \quad l|_{N=0} = l_0 \tag{5}$$

where F is a given function that may depend on the parameters of the material microstructure, and k_{th} and k_f are the material stress intensity threshold and fracture toughness, respectively. In Eq. (5) the maximum is taken over a loading cycle T. Equation (5) should be solved at every material point of the stressed volume V at which $\max_T(k_1) > k_{\text{th}}$. A typical graph of crack propagation rate dl/dN versus l is schematically presented in Fig. 1. It is clear from the graph that there are three distinct stages of crack development: (a) growth of small cracks, (b) propagation of well developed cracks, and (c) explosive and, usually, unstable growth of large cracks. The phase of small crack growth is the slowest one and it represents the main portion of the entire crack life. That usually causes confusion about the duration of small crack initiation and propagation phases. The next phase, propagation of well developed cracks, usually takes significantly less time than the phase of small crack growth. Finally, the explosive crack growth takes almost no time.

A number of crack propagation equations of type (5) are analyzed by Yarema [21]. Any of these equations can be used in the model to describe fatigue crack propagation. The fatigue threshold k_{th} was introduced solely to reflect the situations in which for sufficiently small applied stresses, fatigue failure is not observed.

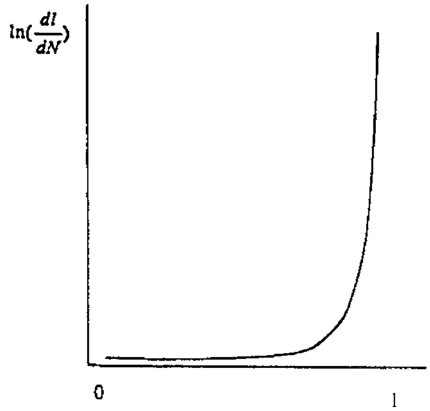


Fig. 1 Schematic graph of crack propagation rate *dl/dN* versus crack radius/size *l*

There is no physical mechanism behind the fatigue threshold. The mechanism of fatigue retardation can be provided by residual stress. The sufficiently high compressive residual stress may arrest fatigue crack growth which is consistent with the purpose of fatigue threshold introduction. Therefore, fatigue threshold k_{th} is an artificial substitution for compressive residual stress. Simultaneous consideration of compressive residual stress and fatigue threshold represents "double dipping" as both of them serve the same purpose—arresting growth of fatigue cracks. The fatigue threshold is not a material constant but it is a function of numerous parameters, among which are the parameters of material plastic and heat treatment and some kind of averaged level of residual stress distribution. The inadequacy of the notion of fatigue threshold in application to bearing steels was experimentally demonstrated by Shimizu in recent papers [22,23]. He showed experimentally that in bearing steels, the material fatigue threshold does not exist, i.e., $k_{\mathrm{th}}=0$, while for structural steels it seems to be positive. Summarizing the above discussion, the simplest and consistent way to consider propagation of fatigue cracks is to assume that $k_{\mathrm{th}}=0$. By taking into account the residual stress we avoid the above-mentioned effect of double dipping. For $k_{\mathrm{th}}=0$, Eq. (5) can be used in the Paris' form

$$\frac{dl}{dN}=g_0(\max_T k_l)^n, \quad l|_{N=0}=l_0 \tag{6}$$

where g_0 and n are the parameters of material fatigue resistance and l_0 is the crack initial radius.

Assuming that the amplitude of cyclic loading is constant and taking into account the fact that for small cracks $k_{10}=k_1 l^{-1/2}$ is independent from l, the solution of the initial-value problem (6) can be obtained in the form

$$l=l_0\left\{1-N\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n/l_0^{(2-n)/2}\right\}^{2/(2-n)}, \quad n>2 \tag{7}$$

It is necessary to note that in material

$$\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n/l_0^{(2-n)/2} \gg 1 \tag{8}$$

Based on Eq. (7), one can formally determine the aforementioned phases of crack development. Namely, the phase of small crack growth is described by the relation

$$N \gg l_0^{(2-n)/2}\left\{\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n\right\}^{-1} \tag{9}$$

the phase of well developed crack propagation is represented by

$$N \approx l_0^{(2-n)/2}\left\{\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n\right\}^{-1} \tag{10}$$

and, finally, the phase of crack explosive growth is determined by

$$N \cong l_0^{(2-n)/2}\left\{\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n\right\}^{-1} \tag{11}$$

In Eq. (10), N is of the same order of magnitude as the expression on the right-hand side but is not necessarily numerically close to it; in Eq. (11), N is approximately equal to the expression on the right-hand side. It is clear from Eqs. (8)–(11) that it takes many loading cycles for a small crack to grow into a well developed crack. Equations (8), (10), and (11) show that for a well developed crack, it takes fewer loading cycles to grow to a critical crack with radius $l_k=(k_f/k_{10})^2$ for which $k_1=k_f$. According to Eq. (11), the phase of explosive crack growth takes just a few loading cycles.

Based on this analysis, it is clear that the number of loading cycles needed for a crack to reach its critical radius is almost independent from the material fracture toughness k_f (see below). For further analysis it is necessary to determine the crack initial radius l_{0k} which after N loading cycles reaches the critical size of l_k. Equation (7) provides the solution

$$l_{0k}=\left\{l_k^{(2-n)/2}+N\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n\right\}^{2/(2-n)}, \quad l_k=(k_f/k_{10})^2 \tag{12}$$

where l_{0k} depends on N, x, y, and z. Obviously, for $n>2$ and fixed x, y, and z, the value of l_{0k} is a decreasing function of N.

It is important to keep in mind that $K=\max_T(k_{10})$ is a function of x, y, and z. Obviously, if $K(x_1,y_1,z_1)<K(x_2,y_2,z_2)$, then $l_k(N,x_1,y_1,z_1)=[k_f/K(x_1,y_1,z_1)]^2>[k_f/K(x_2,y_2,z_2)]^2=l_k(N,x_2,y_2,z_2)$. For $n>2$ and $N>0$ from Eq. (12), it follows that

$$l_{0k}(N,x_1,y_1,z_1)>l_{0k}(N,x_2,y_2,z_2) \quad \text{if } K(x_1,y_1,z_1)<K(x_2,y_2,z_2) \tag{13}$$

Thus, $l_{0k}(N,x,y,z)$ is minimal for maximal $k_{10}(x,y,z)$, which, in turn, occurs where the tensile stress reaches its maximum.

2 Crack Statistics

To describe crack statistics after the crack initiation phase is over, it is necessary to make certain assumptions. The simplest assumptions of this kind are: the existing cracks do not heal and new cracks are not created. In other words, the number of cracks in any material volume remains constant in time. This leads to the equation

$$f(N,x,y,z,l)dl=f(0,x,y,z,l_0)dl_0 \tag{14}$$

which being solved for the crack distribution function $f(N,x,y,z,l)$ gives

$$f(N,x,y,z,l)=f(0,x,y,z,l_0)\frac{dl_0}{dl} \tag{15}$$

where l_0 and dl_0/dl as functions of N and l can be obtained from the solution of Eq. (7). To give a simple illustration of the crack distribution $f(N,x,y,z,l)$ behavior; let us use Paris' law (6) for crack propagation. From Eq. (7) one then gets

$$l_0=\left\{l^{(2-n)/2}+N\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n\right\}^{2/(2-n)},$$
$$\frac{dl_0}{dl}=\left\{1+N\left(\frac{n}{2}-1\right)g_0[\max_T k_{10}]^n l^{(n-2)/2}\right\}^{n/(n-2)} \tag{16}$$

Equations (15) and (16) lead to the expression for the crack distribution function f after N loading cycles

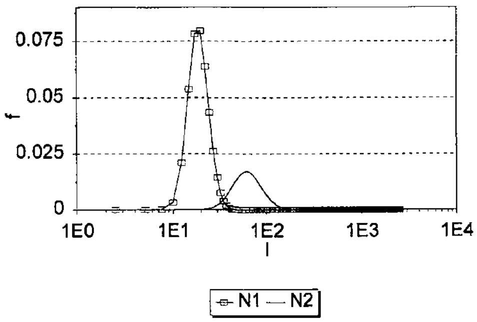


Fig. 2 Schematic view of a crack distribution evolution with number of loading cycles, $N_1 < N_2$

$$f(N,x,y,z,l) = \frac{f[0,x,y,z,l_0(N,l)]}{\left[1 + N\left(\frac{n}{2} - 1\right)g_0(\max_T k_{10})^n l^{(n-2)/2}\right]^{n/(n-2)}} \quad (17)$$

where $l_0(N,l)$ is determined by the first of Eqs. (16). If $N_2 > N_1 \geqslant 0$ then $f(N_2,x,y,z,l)$ can be obtained from $f(N_1,x,y,z,l)$ by a corresponding stretching of this distribution along the l- and f-axes. Obviously, for $N = N_2 > N_1$, the crack distribution is wider than the one for $N = N_1$. A schematic view of such a distribution evolution with N is given in Fig. 2. It is worth noting that in spite of the appearance caused by the logarithmic l-axis, the area under the curves is conserved (see Eq. (14)).

A number of important conclusions can be made based on Eq. (17). Namely, the crack distribution function $f(N,x,y,z,l)$ depends on the initial crack distribution $f(0,x,y,z,l_0)$ and changes with the number of applied loading cycles N in such a way that the crack volume density $\rho(N,x,y,z)$ remains constant. Therefore, it is safe to assume that the crack distribution $f(N,x,y,z,l)$ changes with the number of loading cycles N. However, all fatigue models that take into account material defect distribution implicitly assume that the defect distribution does not change with the number of loading cycles.

2.1 Local Fatigue Damage Accumulation. It is clear that if at a certain point (x,y,z) after N loading cycles radii of all cracks $l \leqslant l_k$, then there is no damage at this point and the local survival probability $p(N,x,y,z) = 1$. On the other hand, if at this point after N loading cycles radii of all cracks $l \geqslant l_k$, then all cracks reached the critical size, the material at this point is completely damaged, and $p = 0$. It is reasonable to assume that the material local survival probability $p(N,x,y,z)$ is a certain monotonic measure of the portion of cracks with radius l below the critical radius l_k. Therefore, $p(N,x,y,z)$ can be represented by the following expressions

$$p(N,x,y,z) = \frac{1}{\rho}\int_0^{l_k} f(N,x,y,z,l)dl \quad \text{if } f(0,x,y,z,l_0) > 0;$$

$$p(N,x,y,z) = 1 \text{ otherwise} \quad (18)$$

$$\rho = \rho(N,x,y,z) = \int_0^{\infty} f(N,x,y,z,l)dl, \quad \rho(N,x,y,z) = \rho(0,x,y,z)$$

Equations (18) determine the material local (at (x,y,z)) survival probability after N loading cycles as a ratio of the number of fatigue cracks with radius below l_k to the total number of cracks at the point (x,y,z). Obviously, $p(N,x,y,z)$ is a monotonically decreasing function of N because fatigue crack radii l tend to grow with N.

To calculate $p(N,x,y,z)$ from Eqs. (18), one can use the function f determined by Eq. (17). However, it is more convenient to modify Eqs. (18) by using Eq. (14) and changing in Eqs. (18) the integration variable from l to l_0

$$p(N,x,y,z) = \frac{1}{\rho}\int_0^{l_{0k}} f(0,x,y,z,l_0)dl_0 \quad \text{if } f(0,x,y,z,l_0) > 0;$$

$$p(N,x,y,z) = 1 \text{ otherwise} \quad (19)$$

where l_{0k} is determined by Eq. (12) and ρ is the initial volume density of cracks. Thus, to every material point (x,y,z) is assigned a certain local survival probability: $0 \leqslant p(N,x,y,z) \leqslant 1$.

Equations (19) demonstrate that the material local survival probability $p(N,x,y,z)$ is controlled mainly by the initial crack distribution $f(0,x,y,z,l_0)$, material fatigue resistance parameters g_0 and n, and applied stresses. Moreover, it easily follows from Eqs. (19) that the material local survival probability $p(N,x,y,z)$ is a decreasing function of N.

2.2 Survival Probability of Material as a Whole. To come up with the survival probability $P(N)$ of the material as a whole, we will assume that the material fails as soon as it fails at least at one point. Let $p_i(N) = p(N,x_i,y_i,z_i)$, $i = 1,2,\ldots,N_c$ (N_c = total number of points in the material stressed volume V populated with cracks). Based on the above assumption the survival probability $P(N)$ is then equal to

$$P(N) = \prod_{i=1}^{N_c} p_i(N) \quad (20)$$

while the probability of failure is $1 - P(N)$. Obviously, $P(N)$ from Eq. (20) satisfies inequalities

$$[p_m(N)]^{N_c} \leqslant P(N) \leqslant p_m(N) \quad p_m(N) = \min_V p(N,x,y,z) \quad (21)$$

where the minimum is taken over the (stressed) volume of the solid.

The right inequality in (21) shows that the survival probability $P(N)$ is never greater than $p_m(N)$. Moreover, the material survival probability $P(N)$ is close to the local survival probability of the "most dangerous" defect. The reason for that is the high value of the power n in Paris' Eq. (6) for fatigue crack growth. Usually, n varies between 6.67 and 9. The first failure is created by the cracks from a small volume element with the least favorable conditions; i.e., with the smallest survival probability $p_m(N)$ (see Eq. (21)). Therefore, in most cases, at relatively early stages of the fatigue process

$$P(N) = p_m(N) \quad (22)$$

The detailed substantiation of the assumption that the first failure is created by the cracks with the smallest survival probability $p_m(N)$ is given by Kudish in [15]. This situation is also illustrated by the example in Fig. 3, where cracks are randomly distributed over an elastic material and are subjected to a uniform cycling tensile stress field. In Fig. 3, the values of the normal stress intensity factor are shown at different time moments (L_0 and k_0 are the values of a crack characteristic size and normal stress intensity factor k_1, the maximum of k_1 is taken with respect to the number of cycles N). These graphs clearly show that effectively, only the cracks with initially the largest sizes, and, therefore, with the initially greater values of k_1, propagate. Moreover, they propagate much faster than all other cracks. As a result of that, the cracks with the initially larger values of k_1 reach their critical size way ahead of all other cracks. This event determines the time and place where fatigue initially starts. Therefore, for high values of n, the survival probability $P(N)$ of the material as a whole is a local fatigue characteristic; i.e., it is determined not by the entire material stressed volume but by the single (maybe several) material

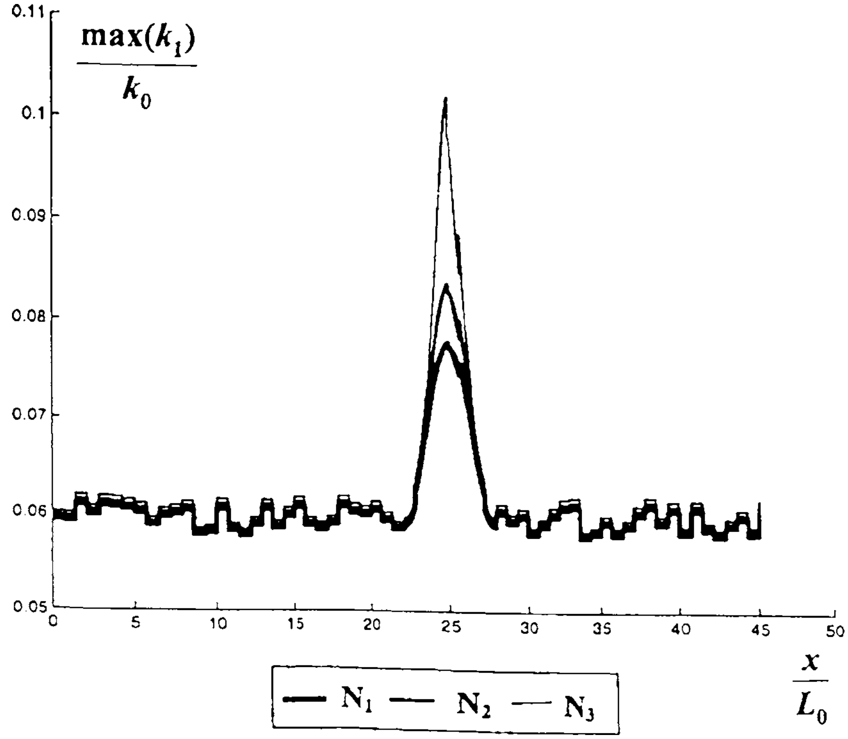


Fig. 3 Illustration of nonuniform over material volume growth of the stress intensity factor k_1 caused by crack growth under cycling loading for the number of loading cycles $N_1=0$, N_2, N_3; $0<N_2<N_3$. The data are obtained for the power $n=9$.

defect with the initially highest values of the stress intensity factor k_1. The higher the power n, the more accurate the above assumption.

This simple analysis leads to the following conclusion. At the relatively early stages of the fatigue process (which are of most interest for practical applications), the survival probability of any volume of the material is equal to 1 except for the small vicinity of the location of the most rapidly growing cracks. Thus, at the early stages of the fatigue process, the survival probability $P(N)$ of the material as a whole is determined by the local survival probability (see Eq. (20)) at the point with the maximum stress intensity factor k_1; i.e. $P(N)=p_m(N)$. This coincides with the intuitive understanding of the fatigue process and with typical observations that fatigue damage takes place first at the most stressed points of the solid.

If the initial crack distribution is taken in the log-normal form of Eq. (1), then according to Eqs. (19), the expression for $p_m(N)$ is relatively simple and it is given by

$$p_m(N) = \frac{1}{2}\min_V\left\{1 + \operatorname{erf}\left[\frac{\ln l_{0k}(N,x,y,z) - \mu_{\ln}}{\sqrt{2}\sigma_{\ln}}\right]\right\}$$

where erf(·) is the error integral. Monotonicity of the function erf and the fact that $\mu_{\ln}$ and $\sigma_{\ln}$ are constants lead to the expression for $p_m(N)$

$$p_m(N) = \frac{1}{2}\left\{1 + \operatorname{erf}\left[\frac{\ln[\min_V l_{0k}(N,x,y,z)] - \mu_{\ln}}{\sqrt{2}\sigma_{\ln}}\right]\right\} \qquad (23)$$

According to Eq. (12), for $n>2$, the value of $\min_V l_{0k}(N,x,y,z)$ is reached at the point(s) where $\max_V k_{10}$ is reached.

To determine the material fatigue life N_* for the given survival probability P_*, it is necessary to solve the equation

$$p_m(N_*) = P_* \qquad (24)$$

for N_*. Finally, the fatigue model is reduced to Eqs. (4), (12), and (22)–(24) for the fatigue life N_* of material as a whole. A detailed analysis of solutions of Eq. (24) is given in subsequent sections.

3 Model Analysis

First, let us consider the fatigue model behavior in some simple cases. If $f(0,x,y,z,l_0)$ is a uniform crack distribution across the material volume, then based on Eqs. (19) and inequality (13), it can be shown that $p(N,x,y,z)$ reaches its minimum at the points where k_{10} and the principal tensile stress reach their maximum values. This means that the material local failure probability $(1-p)$ reaches its maximum at the points with maximal tensile stress. Therefore, for a uniform initial crack distribution $f(0,x,y,z,l_0)=f(0,0,0,l_0)$, the survival probability $P(N)$ from Eq. (22) is determined by the material local survival probability at the points at which the maximal tensile stress is attained.

However, the latter conclusion is not necessarily correct if the initial crack distribution $f(0,x,y,z,l_0)$ is not uniform across the material. Suppose $K=\max_V k_{10}(N,x,y,z)$ is maximal at (x_m,y_m,z_m) and at the initial time moment $N=0$, at some point (x_*,y_*,z_*), there exist cracks larger than the ones at the point (x_m,y_m,z_m); i.e., $\int_0^{l_k} f dl_0|_{(x_*,y_*,z_*)} < \int_0^{l_k} f dl_0|_{(x_m,y_m,z_m)}$. After a certain number of loading cycles $(N>0)$ the material damage at the point (x_*,y_*,z_*) may advance to a greater extent than at the point (x_m,y_m,z_m) where l_{0k} reaches its maximum value. Therefore, fatigue failure may occur at the point (x_*,y_*,z_*) instead of the point (x_m,y_m,z_m) and in a material with non-uniform initial defect distribution the material weakest point is not necessarily the material most stressed point.

In [24], it was shown that in models based on the stressed volume considerations, (see, for example, the Lundberg-Palmgren model for contact fatigue) the dependence of the survival probability on the stressed volume is exponential. That contradicts experimental studies (see [25]), which show that there is a relatively weak dependence of fatigue life on the material stressed volume. In the model presented in this paper, the stressed volume plays no explicit role. However, implicitly it does as the initial crack distribution $f(0,x,y,z,l_0)$ depends on the material volume. In a larger material volume, there is a greater chance to find initial defects of greater size than in a smaller one. These larger defects represent a potential source of fatigue damage and may cause a decrease in fatigue life.

Now, let us establish the relationship between the mean $\mu_{\ln}$ and standard deviation $\sigma_{\ln}$ of the initial log-normal crack distribution and the regular initial mean μ and standard deviation σ

$$\mu=\exp[\mu_{\ln}+0.5\sigma_{\ln}^2],\quad \sigma=\mu\sqrt{\exp(\sigma_{\ln}^2)-1},$$
$$\mu_{\ln}=\ln\frac{\mu^2}{\sqrt{\mu^2+\sigma^2}},\quad \sigma_{\ln}=\sqrt{\ln\left[1+\left(\frac{\sigma}{\mu}\right)^2\right]} \tag{25}$$

Suppose the material failure occurs with the probability $1-P(N)$ at a particular point (x,y,z). By that we determine the point (x,y,z) where in Eq. (23) the minimum over the material volume V is attained. Therefore, at this point in Eq. (23) the operation of minimum over the material volume V can be dropped. Solving Eqs. (23) and (24), one gets

$$N_*=\left[\left(\frac{n}{2}-1\right)g_0(\max_T k_{10})^n\right]^{-1}\times\left(\exp\left\{\left(1-\frac{n}{2}\right)[\mu_{\ln}+\sqrt{2}\sigma_{\ln}\operatorname{erf}^{-1}(2P_*-1)]\right\}-\frac{2-n}{l_{k^2}}\right) \tag{26}$$

where $\operatorname{erf}^1(\cdot)$ is the inverse function to the error integral $\operatorname{erf}(\cdot)$. Let us simplify this equation for the case of a material initially free of damage; i.e., $P(0)=1$. Discounting the very tail of the initial crack distribution, one gets $\max_V(l_0)\leqslant l_k$. Thus, for well developed cracks (see Eq. (10)) and, in many cases, even for small cracks (see Eq. (9)), the second term in Eq. (12) for l_{0k} dominates the first one. This means that the dependence of l_{0k} on l_k and k_f can be neglected and Eq. (26) can be approximated by

$$N_*=\left[\left(\frac{n}{2}-1\right)g_0(\max_T k_{10})^n\right]^{-1}\times\exp\left\{\left(1-\frac{n}{2}\right)[\mu_{\ln}+\sqrt{2}\sigma_{\ln}\operatorname{erf}^{-1}(2P_*-1)]\right\} \tag{27}$$

It follows from Eqs. (3) that k_{10} is proportional to the maximum tensile stress σ_1. Making use of Eqs. (25) and (27), one arrives at a simple analytical formula (see also [15])

$$N_*=\frac{C_0}{(n-2)g_0\sigma_1^n}g(\mu,\sigma) \tag{28}$$

$$g(\mu,\sigma)=\left(\frac{\sqrt{\mu^2+\sigma^2}}{\mu^2}\right)^{(n/2)-1}\times\exp\left\{\left(1-\frac{n}{2}\right)\sqrt{2\ln\left[1+\left(\frac{\sigma}{\mu}\right)^2\right]}\operatorname{erf}^{-1}(2P_*-1)\right\} \tag{29}$$

where the constant C_0 depends on the details of the solid geometry and the stress state, i.e., on the ratios of σ_x/σ_z, σ_y/σ_z, and σ_{xz}/σ_z, and the function $g(\mu,\sigma)$ is completely determined by the parameters of the initial defect distribution, fatigue resistance parameter n, and survival probability P_*. Finally, assuming that $\sigma\ll\mu$ Eqs. (28) and (29) lead to the formula

$$N_*=\frac{C_0}{(n-2)g_0\sigma_1^n\mu^{(n/2)-1}}\exp\left[\left(1-\frac{n}{2}\right)\frac{\sqrt{2}\sigma}{\mu}\operatorname{erf}^{-1}(2P_*-1)\right] \tag{30}$$

Equation (30) demonstrates the intuitively obvious fact that the fatigue life N is inversely proportional to the parameter g_0; i.e., for materials with lower crack propagation rate fatigue life is longer and vice versa. Equation (30) exhibits a commonly accepted and widely used behavior of the structural and contact fatigue (in the latter case, σ_1 should be replaced by the maximum Hertzian pressure p_H) life N_* versus stress σ_1. From the experimental data on fatigue crack propagation in steels, it is well known that $20/3\leqslant n\leqslant 9$. Keeping in mind that usually $\sigma\ll\mu$, for these values of n the fatigue life N_* is practically inversely proportional to a positive power of the mean crack size; i.e., $\mu^{(n/2)-1}$. Therefore, the fatigue life N_* is a decreasing function of the initial mean crack (defect) size μ. This conclusion is valid for any material survival probability P_* and is supported by the experimental data discussed in [1,2,24]. If $P_*>0.5$ (see Eqs. (24) and (30)) then $\operatorname{erf}^1(2P_*-1)>0$ and (keeping in mind that $n>2$) the fatigue life N_* is a decreasing function of the initial standard deviation σ of crack sizes. Similarly, if $P_*<0.5$, then N_* is an increasing function of the initial standard deviation σ of crack sizes. For $P_*=0.5$, according to Eq. (30) the fatigue life N_* is independent from σ while according to Eqs. (28) and (29) it is a slowly increasing function of σ.

4 Application of the Model to Torsion and Bending Fatigue

We will assume that in the beam material the defect distribution is space-wise uniform and follows Eq. (1). To use the available formulas for torsion and bending loadings, we will also assume that the residual stress is zero.

First, let us consider a beam made of an elastic material which is directed along the y-axis and has elliptical cross section (a and b are the ellipse semi-axes, $b<a$). The beam is under action of a torque M_y about the y-axis applied to its ends. The side surfaces of the beam are free of stresses. It can then be shown (see [26], p. 398) that

$$\tau_{xy}=-\frac{2G\gamma a^2}{a^2+b^2}z,\quad \tau_{zy}=\frac{2G\gamma b^2}{a^2+b^2}x,\quad \sigma_x=\sigma_y=\sigma_z=\tau_{xz}=0 \tag{31}$$

where G is the material shear elastic modulus, $G=E/[2(1+\nu)]$, and γ is a dimensionless constant. By introducing the principal stresses σ_1, σ_2, and σ_3 that satisfy the equation $\sigma^3-(\tau_{xy}^2+\tau_{zy}^2)\sigma=0$, we obtain that

$$\sigma_1=-\sqrt{\tau_{xy}^2+\tau_{zy}^2},\quad \sigma_2=0,\quad \sigma_3=\sqrt{\tau_{xy}^2+\tau_{zy}^2} \tag{32}$$

For the case of $a>b$, the maximum principal tensile stress $\sigma_1=-2G\gamma a^2b/(a^2+b^2)$ is reached at the surface of the points $(0,y,\pm b)$ and depending on the sign of M_y, it acts in one of the directions described by the directional cosines

$$\cos(\eta,x)=\mp\frac{\sqrt{2}}{2},\quad \cos(\eta,y)=\pm\frac{\sqrt{2}}{2},\quad \cos(\eta,z)=0 \tag{33}$$

where η is the direction along one of the principal stress axes. For the considered case, the moments of inertia of the beam elliptic cross section about the x- and y-axes, I_x and I_z, as well as the moment of torsion M_y applied to the beam are as follows (see [26], pp. 395, 399) $I_x=\pi ab^3/4$, $I_z=\pi a^3b/4$, $C=4I_xI_z/(I_x+I_z)$, and $M_y=G\gamma C$. Keeping in mind that according to Hasebe and Inohara [16] and Isida [17], the stress intensity factor k_1 for an edge crack of radius l and inclined to the surface of a half-plane at the angle of $\pi/4$ (see Eq. (33)) is $k_1=0.705|\sigma_1|\sqrt{\pi l}$, we obtain $k_{10}=(1.41/\sqrt{\pi})|M_y|/ab^2$. By substituting the expression for k_{10} into Eq. (27), we obtain the fatigue life of a beam under torsion

$$N_* = \frac{2}{(n-2)g_0}\left(\frac{1.257ab^2}{|M_y|}\right)^n g(\mu,\sigma) \tag{34}$$

where $g(\mu,\sigma)$ is determined by Eq. (29).

Now, let us consider bending fatigue of a beam/console made of an elastic material with elliptic cross section (a and b are the ellipse's semi-axes) and length L. The beam is directed along the y-axis and it is under the action of a bending force P_x directed along the x-axis and applied to its free end. The side surfaces of the beam are free of stresses. The other end of the beam at $y=0$ is fixed. It can then be shown (see [26]) that

$$\sigma_x = \sigma_z = 0, \quad \sigma_y = -\frac{P_x}{I_z}x(l-y), \quad \tau_{xz} = 0,$$

$$\tau_{xy} = \frac{P_x}{2(1+\nu)I_z}\frac{2(1+\nu)a^2+b^2}{3a^2+b^2}\left[a^2 - x^2 - \frac{(1-2\nu)a^2z^2}{2(1+\nu)a^2+b^2}\right], \tag{35}$$

$$\tau_{zy} = -\frac{P_x}{(1+\nu)I_z}\frac{(1+\nu)a^2+\nu b^2}{3a^2+b^2}xz$$

where I_z is the moment of inertia of the beam cross section about the z-axis. Again, by introducing the principal stresses that satisfy the equation $\sigma^3 - \sigma_z\sigma^2 - (\tau_{xy}^2 + \tau_{zy}^2)\sigma = 0$, we find that

$$\sigma_1 = \frac{1}{2}[\sigma_y - \sqrt{\sigma_y^2 + 4(\tau_{xy}^2 + \tau_{zy}^2)}], \quad \sigma_2 = 0,$$

$$\sigma_3 = \frac{1}{3}[\sigma_y + \sqrt{\sigma_y^2 + 4(\tau_{xy}^2 + \tau_{zy}^2)}] \tag{36}$$

The tensile principal stress σ_1 reaches its maximum $4|P_x|L/(\pi a^2 b)$ at the surface of the beam at one of the points $(\pm a, 0, 0)$ (depending on the sign of the load P_x) and is acting along the y-axis—the axis of the beam. Based on Eq. (36) and the solution for the surface crack inclined to the surface of a half-space at the angle of $\pi/2$ (see [16,17]), we obtain $k_{l0} = (4.484/\sqrt{\pi})|P_x|L/a^2b$. By substituting the expression for k_{10} into Eq. (27), we obtain the bending fatigue life of a beam

$$N_* = \frac{2}{(n-2)g_0}\left(\frac{0.395a^2b}{|P_x|L}\right)^n g(\mu,\sigma) \tag{37}$$

where $g(\mu,\sigma)$ is determined from Eq. (29).

In cases of torsion and bending, fatigue life is independent of the elastic characteristic of the beam material (see Eqs. (34) and (37)) and it is dependent on fatigue parameters of the beam material (n and g_0), the initial defect distribution (μ and σ), the geometry of the beam cross section (a and b) and its length L, and the applied loading (P_x or M_y).

It is important to realize that in the presence of a subsurface compressive residual stress, the actual mechanism of fatigue may vary. Instead of fatigue cracks being nucleated at the beam surface, the cracks leading to fatigue failure may be nucleated somewhere beneath the beam surface and then propagate at a rate slower than the one for similar surface cracks until they reach the surface. A similar scenario may be realized when the initial defect distribution is not space-wise uniform, i.e., fast fatigue crack growth may start at a different point at the surface or beneath the surface of the beam.

5 Application of the Model to Contact Fatigue

Let us consider an infinite rigid cylinder of radius R parallel to the y-axis that rolls and slides over an elastic half-space $z \leq 0$ along the x-axis. The material of the half-space contains an initial statistical space-wise uniform distribution of cracks that satisfies Eq. (1). The pressure applied to the surface of the half-space follows the Hertzian distribution $q(x) = p_H[1-(x/a_H)^2]^{1/2}$ for $|x| \leq a_H$, where p_H and a_H are the Hertzian maximum pressure and contact half-width, respectively, $p_H = (\mathrm{P}E'/\pi R)^{1/2}$ and $a_H = 2(PR/\pi E')^{1/2}$ (P is the normal load applied to a unit cylinder length, E' is the effective elastic modulus: $E' = E/(1-\nu^2)$). There is also a frictional stress $\tau = -\lambda q$ in the contact (λ is the friction coefficient) and a subsurface residual stress q^0 that acts along the x-axis. Kudish [15] showed that under normal lubrication conditions and observed in practice near surface high compressive and deeper beneath the surface small tensile residual stresses contact fatigue (pitting) is initiated at some subsurface defect. Taking into account that the problem is essentially two dimensional and using the assumption that cracks are small in comparison with the distance to the half-space/plane surface and to other cracks, one obtains asymptotic formulas for the normal k_1 and shear k_2 stress intensity factors (see Kudish [19])

$$k_1 = l^{1/2}[Z^r\theta(Z^r) + q^0\sin^2\alpha], \quad k_2 = l^{1/2}[Z^i - 0.5q^0\sin 2\alpha]$$

$$Z = \frac{1}{\pi}\int_{-a_H}^{a_H}[q(t)\overline{D_0(t)} + \tau(t)\overline{G_0(t)}]dt, \quad \tau = -\lambda q,$$

$$Z^r = \mathrm{Re}(Z), \quad Z^i = \mathrm{Im}(Z) \tag{38}$$

$$D_0(t) = \frac{i}{2}\left[-\frac{1}{t-X} + \frac{1}{t-\bar{X}} - \frac{e^{-2i\alpha}(\bar{X}-X)}{(t-\bar{X})^2}\right],$$

$$G_0(t) = \frac{1}{2}\left[\frac{1}{t-X} + \frac{1-e^{-3i\alpha}}{t-\bar{X}} - \frac{e^{-2i\alpha}(t-X)}{(t-\bar{X})^2}\right], \quad X = x + iz$$

where i is the imaginary unit ($i^2=-1$), $\theta(x)$ is a step function: $\theta(x)=0$, $x\leq 0$, $\theta(x)=1$, $x>0$, and uppercase variables stand for the operation of complex conjugation. It is important to mention that according to Eqs. (38) the quantities of $k_{10} = k_1 l^{1/2}$ and $k_{20} = k_2 l^{-1/2}$ are functions of x and z and are independent from crack half-length l. Using Eqs. (38), Kudish [19] (see also Figs. 7 and 8 in [27]) has shown that for relatively high compressive residual stresses ($q^0 = -0.01p_H$) there are surface and subsurface zones of tensile resultant stress in material. In Tallian, et al. [20], it is shown that for contact interactions, the shear stress intensity factor k_2 is insensitive to variations of the friction coefficient λ, while the normal stress intensity factor k_1 is strongly dependent on λ. Moreover, Mode II cracks are unstable and ultimately propagate according to the Mode I mechanism. That substantiates the assumption that both contact and structural fatigue mechanisms are controlled by tensile stresses.

It is also important to mention that the residual stress q^0 is as essential as pressure q and frictional stress $\tau = -\lambda q$. Depending on the residual stress behavior, the actual mechanism of contact fatigue may vary. For example, in most practical cases, the residual stress is highly compressive right near the contact surface; deeper beneath the surface it changes to a low tensile one. In such cases, for sufficiently smooth lubricated contacts fatigue is of a subsurface nature (see [28]). However, when the residual stress is tensile right near the contact surface, fatigue may be caused by surface (instead of subsurface) defects. In the latter case the lubricant presence may play a detrimental role (see [28]).

For contact interactions, the principal stresses satisfy a full cubic equation. They will be found numerically by determining the direction and the location of the point at which the normal stress intensity factor k_1 reaches its maximum.

Let us analyze the model of contact fatigue in detail based on Eq. (27). It is important to notice that for contact fatigue the stress intensity factors k_1 and k_2 as well as the fatigue life N depend not only on the material fatigue parameters but also on its elastic parameters E and ν through the maximum Hertzian stress p_H (see above).

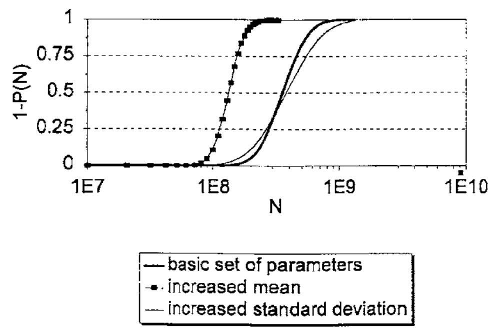


Fig. 4 Graphs of pitting probability $1-P(N)$ calculated for the basic set of parameters with $\mu=49.41$ μm, $\sigma=7.61$ μm ($\mu_{ln}=3.888+\ln(\mu\text{m})$, $\sigma_{ln}=0.1531$), for the same set of parameters with changed initial value of crack mean half lengths $\mu=74.12$ μm ($\mu_{ln}=4.300+\ln(\mu\text{m})$, $\sigma_{ln}=0.1024$), and for the same set of parameters with changed initial value of crack standard deviation $\sigma=11.423$ μm ($\mu_{ln}=3.874+\ln(\mu\text{m})$, $\sigma_{ln}=0.2282$)

By the basic set of model parameters typical for bearing testing, we denote the following: maximum Hertzian pressure $p_H=2$ GPa, contact region half-width in the direction of motion $a_H=0.249$ mm, friction coefficient $\lambda=0.002$, residual stress varying from high compressive value of $q^0=-237.9$ MPa ($q^0=-0.11695p_H$) on the surface to very low tensile value $q^0=0.035$ MPa ($q^0=0.175\times10^{-4}p_H$) at the depth of 400 μm below it, fracture toughness k_f varies between 15 MPa m$^{1/2}$ and 95 MPa m$^{1/2}$ for -400 $\mu\text{m}\leq z\leq0$ μm, $g_0=8.863$ MPa^{-n} m$^{1-n/2}$ $\cdot$cycle^{-1}, $n=6.67$, mean of crack initial half-lengths $\mu=49.41$ μm ($\mu_{ln}=3.888+\ln(\mu\text{m})$), and crack half-length initial standard deviation $\sigma=7.61$ μm ($\sigma_{ln}=0.1531$). Numerical results show that the fatigue life is practically independent from the material fracture toughness k_f, which supports the assumption used for derivation of Eqs. (27)–(29). To illustrate the dependence of contact fatigue life N on some of the model parameters, just one parameter from the basic set of parameters will be varied at a time and graphs of the pitting probability $1-P(N)$ for the basic and modified sets of parameters will be compared. Figure 4 shows that as the crack mean μ and standard deviation σ increase the contact fatigue life N decreases. Similarly, N decreases as the tensile residual stress occurring under the material surface and/or the friction coefficient λ increase (see Figs. 5 and 6). In addition, the fatigue life N does not change when the magnitude of the compressive residual stress is increased/decreased by 20% of its basic value while the tensile portion of the residual stress distribution remains the same. This is in agreement with the fact that tensile stresses control fatigue. Contrary to that, when the compressive residual stress becomes small enough, the frictional stress may supersede it and create new/expanded zones with tensile stresses that potentially may cause/accelerate fatigue failure.

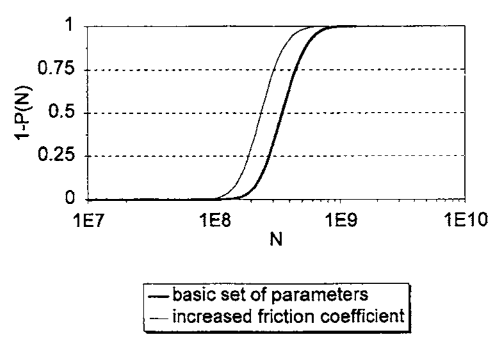


Fig. 5 Graphs of pitting probability $1-P(N)$ calculated for the basic set of parameters with $\lambda=0.002$ and for the same set of parameters with changed friction coefficient $\lambda=0.004$

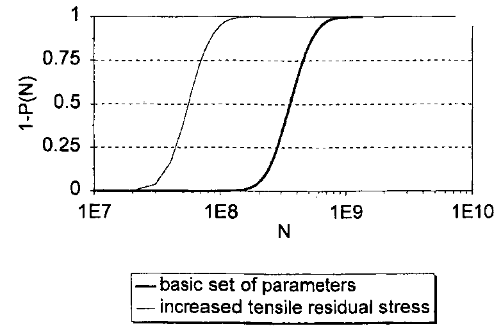


Fig. 6 Graphs of pitting probability $1-P(N)$ calculated for the basic set of parameters and for the same set of parameters with changed profile of residual stress q^0 in such a way that at points where q^0 is compressive, its magnitude is unchanged and at points where q^0 is tensile, its magnitude is doubled

The numerical results lead to the conclusion that $\ln(N)$ is practically a linear function of $\ln(\mu)$ (see Eq. (30)). This behavior of N versus μ is similar to the fatigue life-defect size relationship obtained experimentally for bearings by Stover and Kolarik II [29]. This supports the validity of the approach used for the developing of the new fatigue model.

Let us consider an example of the further validation of the new fatigue model for tapered roller bearings based on a series of approximate fatigue life calculations. We will assume that bearing fatigue life can be closely approximated by taking into account only the most loaded contact. The following parameters have been used for calculations: $p_H=2.12$ GPa, $a_H=0.265$ mm, $\lambda=0.002$, $g_0=6.009$ MPa^{-n} m$^{1-n/2}\cdot$cycle^{-1}, $n=6.67$, the residual stress varied from $q^0=-237.9$ MPa on the surface to $q^0=0.035$ MPa at the depth of 400 μm below the surface, and the fracture toughness k_f varied between 15 MPa m$^{1/2}$ and 95 MPa$\cdot$m$^{1/2}$. The crack (inclusion) initial mean half-length μ varied between $\mu=49.41$ μm and 244.25 μm, and the crack length initial standard deviation varied between $\sigma=7.61$ μm and 37.61 μm. The results for the fatigue life $N_{15.9}$ (for $P(N_{15.9})=P_*=0.159$) calculations are given in the Table 1 and they practically coincide with the experimental data obtained by The Timken Co. (see Fig. 19 in [30]). In the present paper this graph is given as Fig. 7. In Fig. 7, the fatigue life N is given as a function of the cumulative inclusion length (sum of all inclusion lengths over a cubic inch of steel) and in the present paper the fatigue life is calculated as a function of the mean inclusion half-length μ. However, the numerical data for the fatigue life can be brought in perfect compliance with the experimental data by choosing the right measure of steel cleanliness and the proper values for g_0 and n.

Based on the results from the new model, bearing fatigue life can be significantly improved for steels with the same cumulative defect size but smaller mean half-length μ (see Fig. 4). For bearing steels with small percentage of failures (survival probability

Table 1 Tapered bearing fatigue life $N_{15.9}$ versus the initial material inclusion size mean μ and standard deviation σ

μ, μm	σ, μm	$N_{15.9}$, cycles
49.41	7.61	2.5×10^8
73.13	11.26	1.0×10^8
98.42	15.16	5.0×10^7
147.11	22.66	2.0×10^7
244.25	37.62	6.0×10^6

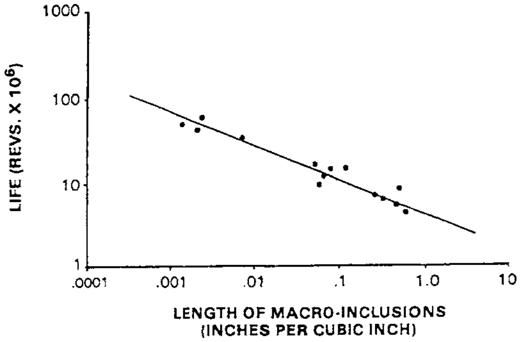


Fig. 7 Bearing life-inclusion versus cumulative length correlation (Fig. 19 from [30] is reprinted with permission of the Iron and Steel Society)

$P>0.5$) with the same cumulative defect size, fatigue life can be improved significantly if the width of the initial defect distribution is reduced; i.e., when the standard deviation σ is made smaller (see Fig. 4). Figures 5 and 6 show that elevated tensile residual stress is more detrimental to fatigue life then higher friction.

6 Conclusions

A new statistical fatigue model that takes into account the most important parameters of the fatigue phenomenon (such as acting stresses, initial statistical defect distribution, orientation of fatigue crack propagation, and material fatigue resistance) is derived and analyzed. The model allows studying the effects of material cleanliness, applied stresses, and material fatigue resistance on fatigue life as single or composite entities. Some analytical results for torsion and bending fatigue of beams and its validation by the experimentally obtained fatigue life data for tapered bearings are presented.

References

[1] Bowles, C. Q., and Schijve, J., 1973, "The Role of Inclusions in Fatigue Crack Initiation in an Aluminum Alloy," Int. J. Fract., **9**, pp. 171–179.
[2] Broek, D., 1986, *Elementary Fracture Mechanics*, 4th ed., Martinus Nijhoff Publishers, Boston, pp. 51–55.
[3] Dudragne, G., Fougeres, R., and Theolier, M., 1981, "Analysis Method for Both Internal Stresses and Microstructural Effect Under Pure Rolling Fatigue Conditions," ASME J. Lubr. Technol., **103**(4), pp. 521–525.
[4] Murakami, Y., Kodama, S., and Konuma, S., 1989, "Quantitative Evaluation of Effects of Non-Metallic Inclusions on Fatigue Strength of High Strength Steels. I: Basic Fatigue Mechanism and Evaluation of Correlation Between the Fatigue Fracture Stress and the Size and Location of Non-Metallic Inclusions," Int. J. Fatigue, **11**(5), pp. 291–298.
[5] Spektor, A. G., Zelbet, B. M., and Kiseleva, S. A., 1980, *Structure and Properties of Bearing Steels*, "Metallurgia" Publishing, Moscow, pp. 74–88.
[6] Kudish, I. I., and Burris, K. W., 2000, "Modern State of Experimentation and Modeling in Contact Fatigue Phenomenon. Part I. Contact Fatigue Versus Normal and Tangential Contact and Residual Stresses. Nonmetallic Inclusions and Lubricant Contamination. Crack Initiation and Crack Propagation. Surface and Subsurface Cracks," STLE Tribol. Trans., **43**(2), pp. 187–196.
[7] Bokman, M. A., Pshenichnov, Yu. P., and Pershtein, E. M., 1984, "The Microcrack and Non-Metallic Inclusion Distribution in Alloy D16 After a Plastic Strain," Plant Laboratory, Moscow, **11**, pp. 71–74.
[8] Wu, H. C., and Yang, S. S., 1988, "On the Influence of Strain-Path in Multiaxial Fatigue Failure," ASME J. Eng. Mater. Technol., **109**, pp. 107–113.
[9] Nisitani, H., and Goto, M., 1984, "Effect of Stress Ratio on the Propagation of Small Crack of Plain Specimens Under High and Low Stress Amplitudes," Trans. Jpn. Soc. Mech. Eng., Ser. A, **50**(453) pp., 1090–1096.
[10] Shao, E., Huang, X., Wang, C., Zhu, Y., and Cheng, Q., 1988, "A Method of Detecting Rolling Contact Crack Initiation and the Establishment of Crack Propagation Curves," STLE Tribol. Trans., **31**(1), pp. 6–11.
[11] Clarke, T. M., Miller, G. R., Keer, L. M., and Cheng, H. S., 1985, "The Role of Near-Surface Inclusions in the Pitting of Gears," ASLE Trans., **28**(1), pp. 111–116.
[12] Nélias, D., Dumont, M. L., Champiot, F., Vincent, A., Girodin, D., Fougéres, R., and Flamand, L., 1999, "Role of Inclusions, Surface Roughness and Operating Conditions on Rolling Contact Fatigue," ASME J. Tribol., **121**(1), pp. 240–251.
[13] Murakami, Y., ed., 1987, *Stress Intensity Factors Handbook*, Vol. 1, Pergamon, Oxford, UK, pp. 239–240, 244–248.
[14] Newman, J. C., Jr., 1971, "An Improved Method of Collocation for the Stress Analysis of Cracked Plates With Various Shaped Boundaries," NASA TN D-6376, NASA, Washington, D.C., pp. 1–45.
[15] Kudish, I. I., 2000, "A New Statistical Model of Contact Fatigue," STLE Tribol. Trans., **43**(4), pp. 711–721.
[16] Hasebe, N., and Inohara, S., 1980, "Stress Analysis of a Semi-Infinite Plate With an Oblique Edge Crack," Ing.-Arch., **49**, pp. 51–62.
[17] Isida, M., 1979, "Tension of a Half Plane Containing Array Cracks, Branched Cracks and Cracks Emanating From Sharp Notches," Trans. Jpn. Soc. Mech. Eng., Ser. A, **45**(392), pp. 306–317.
[18] Isida, M., 1966, "Stress Intensity Factors for the Tension of an Eccentrically Cracked Strip," ASME J. Appl. Mech., **33**, pp. 674–675.
[19] Kudish, I. I., 1987, "Contact Problem of the Theory of Elasticity for Prestressed Bodies with Cracks," J. Appl. Mech. Tech. Phys., **28**(2), pp. 295–303.
[20] Tallian, T., Hoeprich, M., and Kudish, I. I., 2001, "Author's Closure," STLE Tribol. Trans., **44**(2), pp. 153–155.
[21] Yarema, S. Ya., 1981, "Methodology of Determining the Characteristics of the Resistance to Crack Development (Crack Resistance) of Materials in Cyclic Loading," Sov. Mater. Sci., **17**(4), pp. 371–380.
[22] Shimizu, S., 2000, "P-S-N Curves Model for Rolling Contact Machine Elements," *Proceedings of the Intern. Tribol. Conf.*, Nagasaki, **3**, pp. 1767–1772.
[23] Shimizu, S., 2002, "Fatigue Limit Concept in Life Prediction Model for Rolling Contact Machine Elements," STLE Tribol. Trans., **45**(1), pp. 39–46.
[24] Kudish, I. I., and Burris, K. W., 2000, "Modern State of Experimentation and Modeling in Contact Fatigue Phenomenon. Part II. Analysis of the Existing Statistical Mathematical Models of Bearing and Gear Fatigue Life. New Statistical Model of Contact Fatigue," STLE Tribol. Trans., **43**(2), pp. 293–301.
[25] Romaniv, O. N., Yarema, S. Ya., Nikiforchin, G. N., Makhutov, N. A., and Stadnik, M. M., 1990, "Fracture Mechanics and Strength of Materials," *Fatigue and Cyclic Crack Resistance of Construction Materials* Vol. 4, Naukova Dumka, Kiev, USSR, pp. 354–358.
[26] Lurye, A. I., 1970, *Theory of Elasticity*, Nauka, Moscow, USSR.
[27] Kudish, I. I., and Burris, K. W., 2004, "Modeling of Surface and Subsurface Crack Behavior Under Contact Load in the Presence of Lubricant," Int. J. Fract., **125**(1–2), pp. 125–147.
[28] Kudish, I. I., 2002, "Lubricant-Crack Interaction, Origin of Pitting, and Fatigue of Drivers and Followers," STLE Tribol. Trans., **45**(4), pp. 583–594.
[29] Stover, J. D., and Kolarik, R. V., II, 1987, "The Evaluation of Improvements in Bearing Steel Quality Using an Ultrasonic Macro-Inclusion Detection Method," The Timken Company Technical Note, January, pp. 1–12.
[30] Stover, J. D., Kolarik, R. V., II, and Keener, D. M., 1990, "The Detection of Aluminum Oxide Stringers in Steel Using an Ultrasonic Measuring Method," *Proceedings of the 31st Mech. Working and Steel Proc. Conference*, Chicago, October 22–25, Iron and Steel Soc., Inc., pp. 431–440.

Size Effect of Cohesive Delamination Fracture Triggered by Sandwich Skin Wrinkling

Zdeněk Bažant
Northwestern University,
CEE, 2145 Sheridan Road,
Evanston, IL 60208

Peter Grassl[1]
University of Glasgow,
Glasgow, United Kingdom

Because the observed size effect follows neither the strength theory nor the linear elastic fracture mechanics, the delamination fracture of laminate-foam sandwiches under uniform bending moment is treated by the cohesive crack model. Both two-dimensional geometrically nonlinear finite element analysis and one-dimensional representation of skin (or facesheet) as a beam on elastic-softening foundation are used. The use of the latter is made possible by realizing that the effective elastic foundation stiffness depends on the ratio of the critical wavelength of periodic skin wrinkles to the foam core thickness, and a simple description of the transition from shortwave to longwave wrinkling is obtained by asymptotic matching. Good agreement between both approaches is achieved. Skin imperfections (considered proportional to the the first eigenmode of wrinkling), are shown to lead to strong size dependence of the nominal strength. For large imperfections, the strength reduction due to size effect can reach 50%. Dents from impact, though not the same as imperfections, might be expected to cause as a similar size effect. Using proper dimensionless variables, numerical simulations of cohesive delamination fracture covering the entire practical range are performed. Their fitting, heeding the shortwave and longwave asymptotics, leads to an approximate imperfection-dependent size effect law of asymptotic matching type. Strong size effect on postpeak energy absorption, important for impact analysis, is also demonstrated. Finally, discrepancies among various existing formulas for critical stress at periodic elastic wrinkling are explained by their applicability to different special cases in the shortwave-longwave transition.
[DOI: 10.1115/1.2722778]

Dedicated to Professor Franz Ziegler on the occasion of his 70th birthday

1 Introduction

A major question in extrapolating small-scale laboratory tests to full-scale sandwich structures is the size effect. Delamination of the skin (or facesheet) is often triggered by wrinkling instability, which has generally been considered to be free of size effect [1–5]. The absence of size effect has been inferred from the fact that the critical stress for buckling generally exhibits no size effect. However, this inference is valid only for the symmetry-breaking bifurcation of equilibrium path in perfect structures [6]. In actual sandwich structures, the geometrical shape of the skin is always geometrically imperfect, at least to some degree, due to imprecise manufacturing. Dents from impacts represent severe imperfections, usually accompanied by preexisting delaminations.

Buckling of imperfect quasibrittle structures generally leads to snapthrough instability which typically exhibits size effect on the nominal strength (e.g., [7] Chap. 13). The size effect is understood as the dependence of the dimensionless nominal strength of structure on its characteristic dimension (considered here as the skin thickness) when geometrically similar structures are compared (i.e., when all the structural dimensions are varied in proportion to the chosen characteristic dimension) [6,7]. The objective of this paper is to verify that this indeed happens for buckling driven delamination, to quantify the size effect, and to determine its intensity. A secondary objective is to assess the size effect on the postpeak energy absorption, important for judging survival under blast or dynamic impact.

Delamination in sandwiches and laminate composites has traditionally been analyzed by strength theory (either elastoplasticity or elasticity with strength limit) [8–14]. In this classical theory, there is no size effect.

Linear elastic fracture mechanics (LEFM) was applied to the analogous problem of delamination of micrometer-range metallic films from their substrate [15,16] and was also used in ([7], p. 770). For sandwich structures, however, LEFM now appears as unrealistic because, according to recent experiments [17,18], the size effect in typical laboratory tests is about half as strong as expected for LEFM. Therefore, the structure is quasibrittle [19], which means that the size of the fracture process zone (FPZ) cannot be considered to be negligible compared to the cross-sectional dimension of normal-size sandwich structures. Thus, delamination fracture should be simulated by the cohesive crack model rather than LEFM. This model has already been used in some recent numerical simulations [20,21] and will be adopted here.

Since buckling driven delamination is difficult to control in experiments, it is not surprising that only few experimental studies have been reported (e.g., [22,23]), and that none provides comprehensive insight. Thus, the present study will rely on numerical simulations using geometrically nonlinear finite element analysis as well as the softening foundation model, which is an adaptation of Winkler elastic foundation. Dimensionless variables will be used to cover the entire practical range. This goal will also necessitate clarifying confusion that still exists even for elastic skin wrinkling. It will be seen that different existing wrinkling formulas apply to different special cases, such a shortwave and longwave wrinkling.

This study deals exclusively with the deterministic size effect [6,7,24]. The Weibull-type statistical size effect on the mean struc-

[1]Formerly of Northwestern University, Evanston, IL.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received May 16, 2006; final manuscript received January 24, 2007. Review conducted by Robert M. McMeeking.

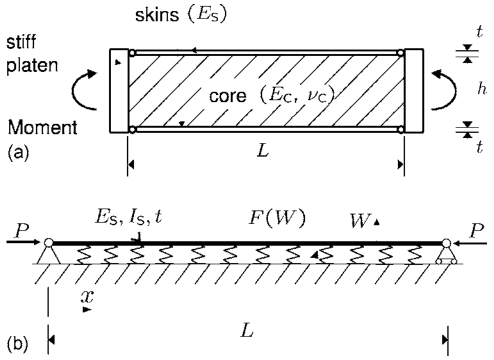


Fig. 1 (*a*) The geometry of a typical sandwich beam subjected to pure bending and (*b*) the beam subjected to an axial compression force *P* supported by a softening foundation

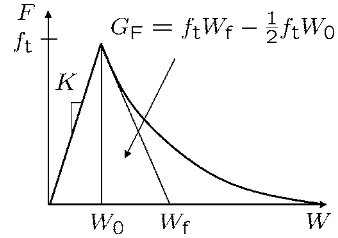


Fig. 2 Force-displacement relation of the softening foundation

tural strength may, of course, also occur but must, in principle, be negligible when the location of failure initiation is fixed, either by mechanics or by defects, such as notches or dents. The present study is also limited to two-dimensional analysis of sandwich beams subjected to pure bending. Sandwich structures subjected to compression with or without bending are expected to lead to interaction of overall buckling and wrinkling (see Appendix), which is beyond the scope of this paper. So is the three-dimensional wrinkling, leading to two-dimensional delamination blisters, for which a similar, but probably weaker, size effects may be expected.

2 Softening Foundation Model

The analysis of delamination in sandwich structures subjected to pure bending, as shown in Fig. 1(*a*), can be simplified by modeling the skin as an axially compressed beam supported by a softening foundation consisting of independent continuously distributed nonlinear springs (Fig. 1(*b*)). For the mathematically analogous problem of a foundation with bilinear elastic-plastic hardening response, the solution is available [25]. Here, the problem is solved for bilinear elastic-softening response, in which the softening represents gradual decohesion due to a cohesive crack under the beam. The differential equation of the problem reads

$$E_s I_s \frac{d^4 W}{dX^4} + P\frac{d^2 W}{dX^2} + F = -P\frac{d^2 W^\circ}{dX^2} \tag{1}$$

where E_s=Young's modulus of the skin, $I_s = t^3/12$=moment of inertia (per unit width) of the cross section of the skin of thickness t, P=axial force in the beam (per unit width), X=coordinate in the axial direction, and $W(X)$=deflection (lateral displacement) of the skin, additional to the initial deflection W°. Furthermore, F is the distributed lateral force (traction), defined as

$$F = \begin{cases} KW & \text{if } W \leqslant W_0 \\ KW_0 e^{-(W-W_0)/(W_f - W_0)} & \text{if } W > W_0 \end{cases} \tag{2}$$

where K is the foundation modulus (i.e., the spring stiffness of the foundation per unit length), W_0 is the displacement at which the tensile strength f_t is reached (Fig. 2) and W_f controls the fracture energy G_F of the cohesive crack, which lies in the core very near the skin,

$$G_F = f_t\left(W_f - \frac{W_0}{2}\right) \tag{3}$$

G_F represents the total area under the stress-displacement curve in Fig. 2 (and not the area under the postpeak part of that curve, for unloading follows the elastic stiffness). The distributed spring stiffness K (per unit length of beam) may be interpreted as

$$K = \frac{E_c}{h_{eq}} \tag{4}$$

where E_c is Young's modulus of the sandwich core and h_{eq} represents the equivalent (or effective) depth of the foundation.

First, we consider the case of shortwave wrinkling of compressed skin, which is not affected by the opposite skin, and leave the case of interacting skins for later consideration. In this case, by contrast to many previous studies, h_{eq} cannot be considered as constant. Rather, it depends on the stress field in the core below the skin (Fig. 3) and represents the thickness of a uniformly stressed strip of core material that gives the same foundation stiffness as the actual, nonuniformly stressed, core (and has a negligible shear modulus).

3 Elastic Shortwave Wrinkling and Equivalent Foundation Depth

Consider that the wavelength $L_{cr} \ll h$ (h=core thickness). In that case, and approximately if $L_{cr} < h$, the core may be regarded as an infinite half-space. The reason is that the alternating tractions applied on the core by the periodically wrinkled skin (Fig. 3(*b*)) are self-equilibrated over a segment of length $2L_{cr}$ where L_{cr} is the half wavelength of skin buckling (Fig. 3(*c*)). Therefore, according to the St. Venant principle, the stresses caused by periodic wrinkling must exponentially decay to nearly zero over a distance from the skin roughly equal to $2L_{cr}$. Therefore, it must be possible to write

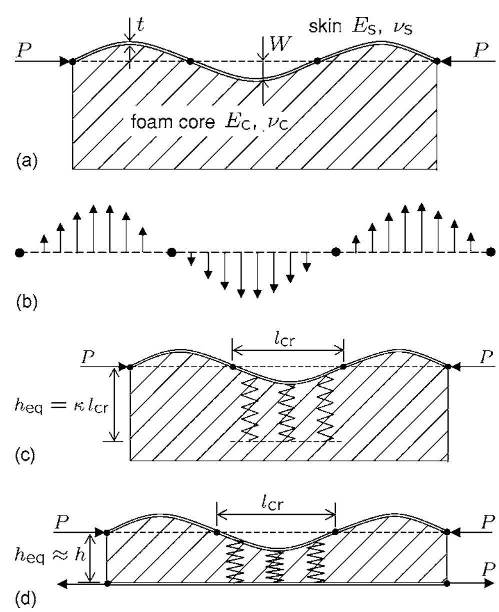


Fig. 3 (*a*) The deflection of the top skin, (*b*) equilibriated stress acting on the foam, (*c*) equivalent height for shortwave, and (*d*) longwave wrinkling

$$h_{eq_0} = \kappa L_{cr} \tag{5}$$

where κ is some constant and subscript 0 refers to the limit case $L_{cr}/h \to 0$. For periodic skin buckling, the solution of the homogeneous differential equation for a beam on elastic foundation ([7], p. 316) yields

$$L_{cr} = \pi\left(\frac{E_s I_s}{K}\right)^{1/4} = \pi\left(\frac{\kappa L_{cr} E_s I_s}{E_c'}\right)^{1/4}, \quad I_s = \frac{t^3}{12} \tag{6}$$

where $I_s = t^3/12$ = central moment of inertia of the skin cross section (per unit width b). Solving (6) for L_{cr} provides

$$L_{cr} = t\left(\frac{\pi^4 \kappa E_s}{12 E_c'}\right)^{1/3} \tag{7}$$

where E_c' = effective Young's (elastic) modulus of the core; for plane stress, $E_c' = E_c$, and for plane strain, $E_c' = E_c/(1-\nu_c^2)$ where ν_c = Poisson ratio of the core. In this expression, ν_c accounts for the out-of-plane effect of Poisson ratio. Note that the in-plane effect of Poisson ratio, manifested in the effect of shear modulus $G_c = E_c'/2(1+\nu_c)$ of the core on its resistance to skin wrinkling, is known to be negligible in beam bending.

Thus, the critical axial compressive force in the skin at bifurcation is (per unit width b)

$$P_{cr_0} = 2\sqrt{K E_s I_s} = k_1 (E_c'^2 E_s)^{1/3} t \quad \text{where } k_1 = \left(\frac{2}{3\kappa^2 \pi^2}\right)^{1/3} \tag{8}$$

Note that this expression for P_{cr} has the same form as that derived in [3] by solving the elastic boundary value problem problem under certain simplifications. The present derivation is far shorter, but it does not yield the value of κ. Comparison of the two expressions indicates dependence on ν_c

$$\kappa = \alpha\sqrt{1+\nu_c} \tag{9}$$

The solution in [3] is matched if $\alpha = 0.43$. Here, however, $\alpha = 0.53$ is used, as determined from a single finite element analysis of P_{cr}.

A similar expression, namely, $P_{cr} = 0.85 t (E_s k_t^2)^{1/3}$ (where k_t is a function of E_c), was proposed in [26] with k_t taking into account the influence of orthotropic core.

4 Elastic Moment-Induced Longwave Wrinkling

Consider now that the critical wavelength $L_{cr} \gg h$ (Fig. 3(*d*)) and that the sandwich beam is subjected to bending moment only (i.e., with no axial force). Then the opposite skin is under tension and may be approximated as a rigid base, with no deflection. The transverse compressive stress in the core is now almost uniform, and

$$h_{eq_\infty} = h \tag{10}$$

i.e., the foundation stiffness $K = E_c'/h_{eq}$ is constant (independent of the critical wavelength). The critical axial compressive force in the skin at bifurcation for periodic skin buckling (with no delamination) is ([7], p. 316)

$$P_{cr_\infty} = 2\sqrt{K E_s I_s} = \sqrt{\frac{E_c' E_s t^3}{3h}} \tag{11}$$

which is the same as reported in [4]; subscript ∞ refers to the limit case $L_{cr}/h \to \infty$, for which the solution is exact.

The hypothesis of the opposite skin being rigid is justified if the skin is sufficiently thick or subjected to sufficient tension, or both. Similar to shortwave wrinkling, the longwave wrinkling is resisted primarily by transverse normal stresses in the core, while the shear stresses in the core (which dominate global buckling) play a minor role.

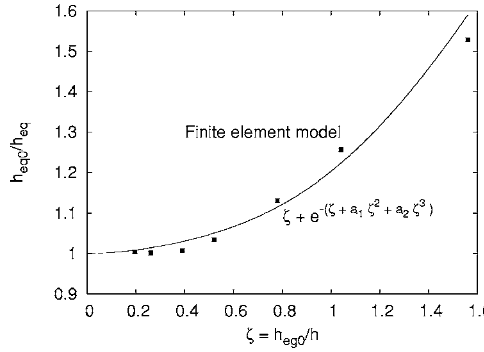


Fig. 4 Evolution of the equivalent height h_{eq} for the transition from shortwave to longwave wrinkling

5 Asymptotic Matching of Elastic Shortwave-to-Longwave Transition

In general, the equivalent height h_{eq} for both shortwave wrinkling in (5) and longwave wrinkling in (10) is subjected to the upper bound

$$h_{eq} = \min(h_{eq_0}, h_{eq_\infty}) \tag{12}$$

In reality, the transition between shortwave wrinkling and longwave wrinkling will not be abrupt but smoothly distributed over a certain range of the dimensionless variable

$$\zeta = \frac{h_{eq_0}}{h} \tag{13}$$

The shortwave bound $h_{eq} = h_{eq_0}$ must be tangentially approached for $\zeta \to 0$, and the longwave bound $h_{eq} = h_{eq_\infty}$ must be an asymptote for $\zeta \to \infty$. A smooth transition meeting these asymptotic conditions may simply be described by the function $h_{eq_0}/h_{eq} = \zeta + e^{-\zeta}$, which, however, has no free parameters to adjust according to finite element results. A more general expression that has such parameters, a_1 and a_2, and meets all the asymptotic conditions is

$$\frac{h_{eq_0}}{h_{eq}} = \zeta + e^{-(\zeta + a_1\zeta^2 + a_2\zeta^3)} \tag{14}$$

where $a_1 = 0.24$ and $a_2 = 0.36$, as obtained by fitting numerical results with the Marquardt-Levenberg algorithm for nonlinear least-squares optimization. The exponential decay in the expression for h_{eq_0}/h_{eq} is favored by the fact that, according to St. Venant principle, the self-equilibrated tractions applied in a localized disturbance (such the wavelength of the skin) are known to decay with the distance from the disturbance exponentially. Comparisons of Eq. (14) to the results of finite element simulations are shown in Fig. 4.

Usually the end of a sandwich beam has either a laterally supported compressed skin or a zero bending moment (and thus no compression in the skin). In the rare case of an end with laterally sliding compressed skin ([7], p. 318) a semi-infinite skin wrinkles nonperiodically, as an exponentially decaying modulated sinusoid, and then both (8) and (11) must be divided by 2, with no change to the rest of analysis.

6 Formulation in Dimensionless Variables

The solution may generally be expressed as a relation among seven dimensional variables: $E_s I_s, K, P, W_0, W^\circ, W_f, x$ which involve two independent dimensions, force and length. According to the Vashy-Buckingham theorem of dimensional analysis, the number of dimensionless variables governing the problem is $7-2=5$. They may be chosen the same as in a previous study of plastic bilinearly hardening foundation [25]

$$x = X\left(\frac{E_s I_s}{K}\right)^{-1/4}, \quad \lambda = \frac{1}{2}P(KE_s I_s)^{-1/2} \tag{15}$$

$$w = \frac{W}{W_0}, \quad w^{\circ} = \frac{W^{\circ}}{W_0}, \quad w_f = \frac{W_f}{W_0} \tag{16}$$

Substituting Eqs. (15), (16), and (2) into (1), yields the dimensionless differential equation

$$\frac{d^4 w}{dx^4} + 2\lambda\frac{d^2 w}{dx^2} + w = -2\lambda\frac{d^2 w^{\circ}}{dx^2} \quad \text{if } w \leqslant 1 \tag{17a}$$

$$\frac{d^4 w}{dx^4} + 2\lambda\frac{d^2 w}{dx^2} + e^{(w-1)/(w_f-1)} = -2\lambda\frac{d^2 w^{\circ}}{dx^2} \quad \text{if } w > 1 \tag{17b}$$

where w=dimensionless deflection. For a perfect beam ($w^{\circ}=0$), the first eigenmode of buckling at bifurcation is determined from (17a) as $w=\sin x$, and the corresponding load at bifurcation results in $\lambda=1$.

A generic imperfection of skin may be expressed as a linear combination (or infinite series) of all the eigenmodes of elastic skin wrinkling. Similar to other buckling problems, the first eigenmode may be expected to have the dominant influence for loads near the first critical load [7]. Therefore, the imperfection δ of the skin is chosen to be proportional to the aforementioned displacement profile $w=\sin x$ of perfect skin at first bifurcation, i.e., $w^{\circ}=\delta\sin x$. The solution of (17) for the elastic case ($w_{\max}\leqslant 1$), with the aforementioned imperfection, is

$$w(x) = \frac{\lambda\delta}{1-\lambda}\sin x \tag{18}$$

This solution will be used in Sec. 9 for deriving the size effect law. The size effect is understood as the dependence of dimensionless nominal strength λ on skin thin thickness t when all the structural dimensions vary in proportion to t, i.e., h/t=const. For buckling failures with material failure criteria expressed solely in terms of stresses and strains, the size effect is nil [6], i.e., λ is independent of t.

The dimensionless variables x, w, w°, and λ are size independent. However, ensuring constant fracture energy requires that the dimensionless parameter w_f be considered size dependent, as obtained by inserting (3) into (16),

$$w_f = \frac{G_F E_c}{f_t^2 h_{eq}} + \frac{1}{2} \tag{19}$$

(note that this size dependence is analogous to the dependence of fracture energy on the mesh size in the crack band model [7,27]). Parameters G_F, E_c, and f_t are material properties independent of the structure size, whereas h_{eq} is proportional to the structure size. Thus, the size dependence of w_f can be characterized as

$$w_f = \frac{1}{\xi} + \frac{1}{2} \tag{20}$$

where

$$\xi = \frac{h_{eq}}{l_0}, \quad l_0 = \frac{E_c G_F}{f_t^2} \tag{21}$$

ξ is dimensionless and l_0 is known as Irwin's characteristic material length. For cracks in bulk, l_0 characterizes the fracture process zone length but not for delamination cracks (see [28,29] for opening and [30] for shear mode). What matters here is that l_0 represents a length parameter formed solely from basic material constants.

The initial boundary value problem solver of the commercial package MATLAB is used for numerical solution of the ordinary nonlinear differential equation (17). The amplitude of the initial displacement field is slightly increased in the middle of the beam, in order to control the location of the delamination growth. This increase is chosen to be so small that its effect on the magnitude of the load carried by the skin be imperceptible. The results are later compared to those of the finite element simulations in Sec. 8.

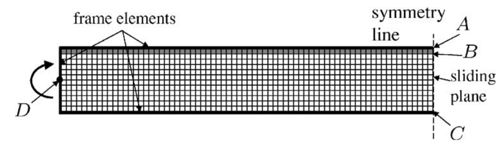


Fig. 5 Finite element mesh

To simplify analysis, only one half of the beam is modeled and symmetric deformation is assumed. Even though the actual growth of delamination blister must be expected to be nonsymmetric (one-sided) ([7], Chap. 12) the assumption of symmetry should be satisfactory because asymmetric growth of delamination fracture should produce a deflection curve symmetric with respect to a moving center of the blister.

When the delamination blister grows, the equivalent core depth below the blister (though not elsewhere) increases, which decreases the core stiffness. However, in view of satisfactory agreement with the finite element results, this effect appears to be minor and is not considered here.

7 Geometrically Nonlinear Finite Element Analysis

To determine parameter α in (17) and to validate the simplified modeling of delamination by the softening foundation model, a geometrically nonlinear finite element program (FEAP, procured from R. Taylor, Berkeley, CA) is used. A sandwich beam, depicted in Fig. 1(a), is considered and is modeled using the finite element mesh in Fig. 5. The skins are represented by beam elements taking into account large displacements and large rotations. For the core, plane stress finite elements based on a linearized small displacement formulation are used. The core is treated as isotropic, and for the skin, only the longitudinal elastic modulus E_s needs to be considered since the transverse and shear moduli of laminate skin are immaterial for bending and axial deformation.

The beam is considered to be subjected to a uniform bending moment M. However, as long as the core thickness h is large enough for the stresses from wrinkling to decay to nearly zero over the core thickness, the only loading that matters is the axial force in the skin, which is $P=M/(h+t)$. Whether this force is produced by moment alone, or a combination of bending moment and axial force, is immaterial.

An elastic stress-strain relation is used for all the elements of the core except a narrow band of elements under the skin (marked gray in Fig. 5). It is known that the delamination fracture occurs within the core very near the interface with the skin, but not within the interface. Therefore, perfect bond between the skins and the core is enforced. Transverse softening of the aforementioned band, which can be regarded as distributed microcracking, simulates delamination. In the softening band, the stress-strain law is elastic in the prepeak, and the postpeak response follows the isotropic damage model proposed in [27], which is defined as

$$\boldsymbol{\sigma} = (1-\omega)\mathbf{D}_c : \boldsymbol{\varepsilon} = (1-\omega)\bar{\boldsymbol{\sigma}} \tag{22}$$

Here, $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ are the stress and strain tensors in the core, $\bar{\boldsymbol{\sigma}}$ is the effective stress tensor, ω is the damage variable, and $\mathbf{D}_c$ is the isotropic elastic stiffness tensor of the core, which is based on the Young's modulus E_c and the Poisson's ratio ν_c. The damage variable ω is a function of history variable γ, which is defined as the maximum equivalent strain $\tilde{\varepsilon}$ reached in the history of the material: $\gamma(t)=\max\,\tilde{\varepsilon}(\tau)$ for $\tau\leqslant t$. The equivalent strain is defined as

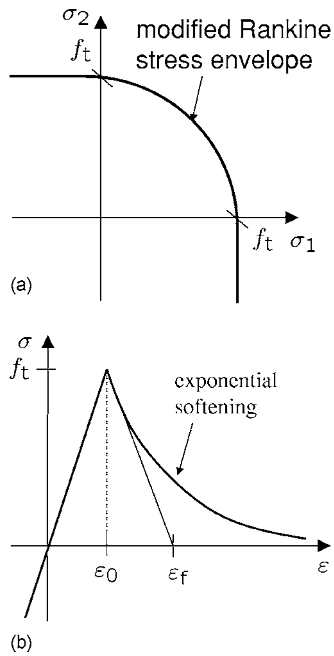


Fig. 6 Initial stress envelope represented in the principal stress space obtained with the damage loading function

$$\tilde{\varepsilon} = \frac{\sqrt{\langle \bar{\boldsymbol{\sigma}}_c \rangle : \langle \bar{\boldsymbol{\sigma}}_c \rangle}}{E_c} \tag{23}$$

where $\langle x \rangle = \max(x, 0)$. This definition corresponds to a Rankine-type strength envelope with a smooth round-off in the sectors of two positive principal stresses, as shown in Fig. 6(*a*).

The damage variable ω is related to the history variable γ as

$$\omega = g(\gamma) = \begin{cases} 0 & \text{if } \gamma \leqslant \varepsilon_0 \\ 1 - \dfrac{\varepsilon_0}{\gamma} \exp\left(-\dfrac{\gamma - \varepsilon_0}{\varepsilon_f - \varepsilon_0}\right) & \text{if } \gamma \geqslant \varepsilon_0 \end{cases} \tag{24}$$

where $\varepsilon_0 = f_t / E_c$, f_t is the tensile strength of the core. The parameter ε_f is related to the fracture energy G_F as

$$\varepsilon_f = \frac{G_F}{E_c \varepsilon_0 h_e} + \frac{1}{2}\varepsilon_0 \tag{25}$$

where h_e is the depth of the element row (softening band) adjacent to the skin (Fig. 5). This damage law results in an exponential stress-strain curve in uniaxial tension, as presented in Fig. 6(*b*). The inelastic strains determined by the isotropic damage model are fully reversible, i.e., the secant stiffness points toward the origin (this reversibility would, of course, be unrealistic if crack unloading were not absent from the present simulations).

As before, only one-half of the beam is modeled (Fig. 5). The loading moment M is applied at point D and assumed to be transferred by a rigid loading platen into the upper and lower skins. The structure is restrained in longitudinal direction at point A. The loading is controlled by prescribing the displacement of point B. The same initial displacement, i.e., $\overset{\circ}{w} = \delta \sin x$, is prescribed for the upper skin. The imperfection amplitude δ at the middle of the beam at point A is slightly increased in the same way as for the softening foundation model, to control the place where the delamination begins.

8 Results and Comparison of Softening Foundation to Finite Elements

The effect of the structure size on the relation between the load parameter λ and the mid-point displacement $w_a = w(l/2)$ is shown, for three imperfection amplitudes $\delta = 0.1, 1, 2$, in Fig. 7 (in which, $l = L(E_s I_s / K)^{-1/4}$, where the beam length L is chosen to be $17L_{cr}$.) As one can see, the results of the softening foundation model are in reasonable approximate agreement with the more accurate finite element results. The comparison shows that the size has a strong effect on the postpeak part of the load-displacement relation. The larger the size, the less energy is dissipated in relation to the energy dissipated by delaminating the entire skin. The deflection curves of the upper skins obtained from the softening foundation model for a constant imperfection $\delta = 0.1$, and for the sizes $\xi = 1$ and $\xi = 0.05$, are shown in Fig. 8, respectively. In accordance with the load-displacement curves in Fig. 7, the lateral displacement w_a for the small structure size is greater than for the large structure size. The overall deflection pattern is similar. However, a closer examination of the size effect on the evolution of the diagram of load versus blister length b (which is the normalized length in the middle portion of the beam in which $w > 1$) reveals a size effect on the nominal strength; see Fig. 9. The larger the size, the smaller is λ_{max} (i.e., it varies in proportion to the nominal strength, for the

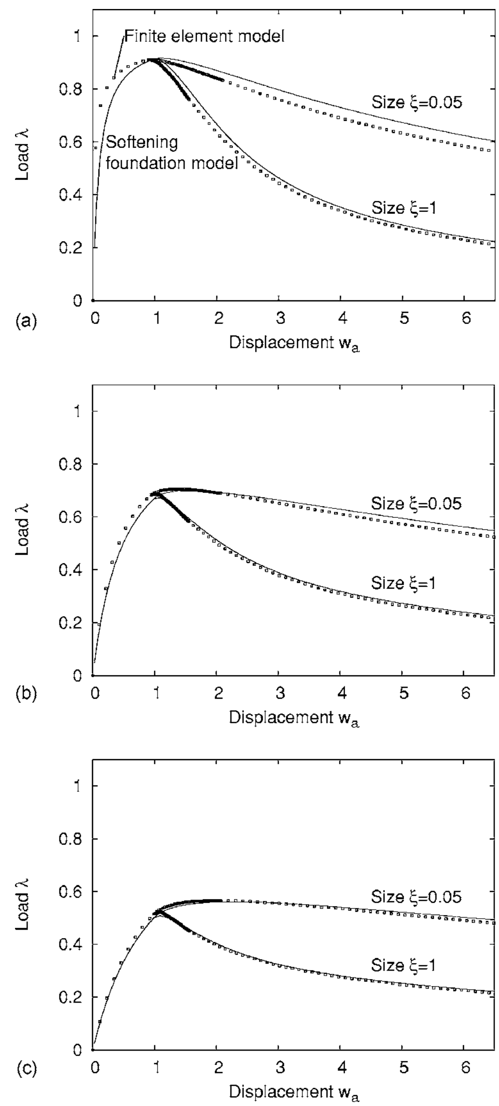


Fig. 7 Load λ versus the midpoint displacement w_a obtained with the softening foundation model and the finite element model for the imperfections: (*a*) δ=0.1, (*b*) δ=0.5, and (*c*) δ=1 for two sizes (ξ=1 and ξ=0.05)

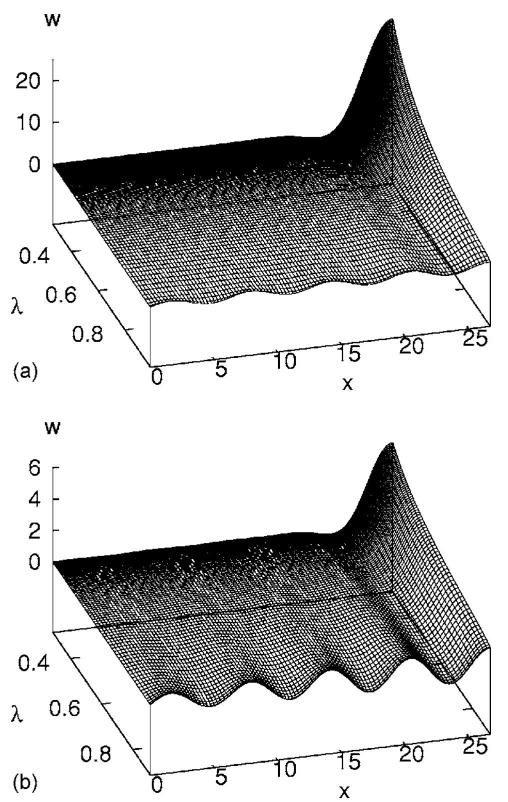


Fig. 8 Deflection of the upper skin for the postpeak regime for imperfection δ=0.1 and for the sizes: (*a*) ξ=0.05 and (*b*) ξ=1 obtained from the softening foundation model

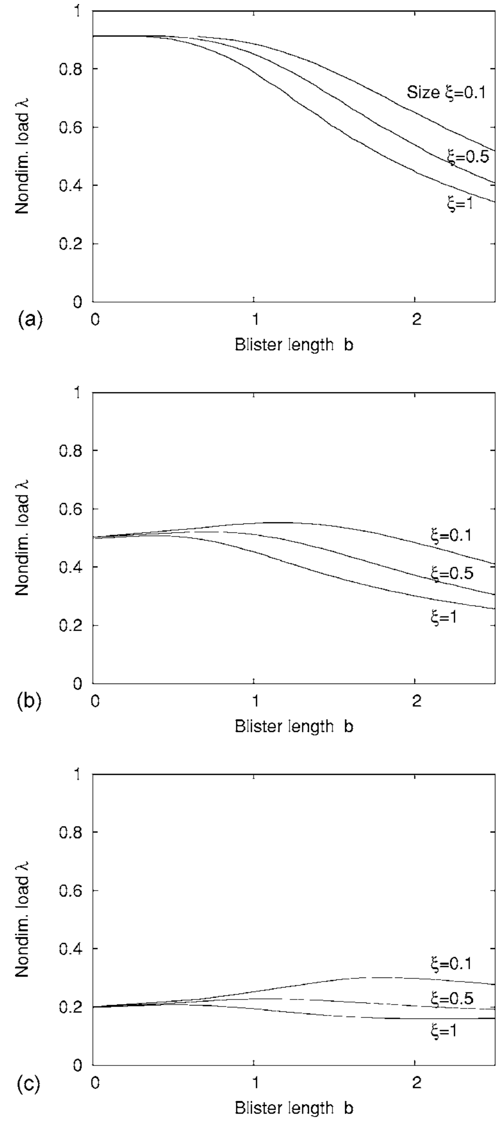


Fig. 9 Load λ versus the blister length (mid beam region in which $w>1$) of the upper skin for the imperfections: (*a*) δ=0.1, (*b*) δ=1, and (*c*) δ=4 and for the sizes ξ=1, ξ=0.5, and ξ=0.1 obtained from the softening foundation model

loading case presented in Fig. 1(*a*)). Furthermore, note that the size effect intensity depends strongly on the imperfection amplitude. A law for this size effect is proposed next.

9 Size Effect Law for Imperfection Sensitive Wrinkling

The size effect on the dimensionless nominal strength, $\lambda_N = \lambda_{\max}$, shown in Fig. 10, has a form similar to the size effect law for crack initiation in quasibrittle structures, which reads [6,19,24,31,32] $\lambda_N = \lambda_\infty[1+1/(k+\xi)]$, where λ_∞ has the meaning of nominal strength of infinitely large structure. This law, however, is not directly applicable since imperfections are seen in Fig. 10 to influence the size effect. Therefore, a generalized law of the form

$$\lambda_N(\delta,\xi) = \lambda_\infty(\delta)\left[1 + \frac{1}{k(\delta) + a\xi^b}\right], \quad k(\delta) = c\delta^{-d} \qquad (26)$$

is proposed here, with constants a,b,c,d and parameters λ_∞ and k, depending on the imperfection amplitude δ. For large sizes ($\xi \to \infty$), the nominal strength is decided by initiation of cohesive crack ($w=1$), and in that case (18) leads to

$$\lambda_N(\delta,\infty) = \lambda_\infty = 1/(1+\delta) \qquad (27)$$

Note that here the large-size limit does not correspond to LEFM, which is the case for type 2 size effect [24], seen in specimens with notches or large stress-free cracks. Rather, in the absence of preexisting delamination crack, we see a particular case of type 1 size effect [24] because the geometry is positive [6,19], causing failure to occur at crack initiation.

For small sizes ($\xi \to 0$), the nominal strength in (26) turns into

$$\lambda_N(\delta,0) = \lambda_\infty(\delta)\left[1 + \frac{1}{k(\delta)}\right] \qquad (28)$$

Parameters a,b,c,d in (26) are determined as optimal fits of numerical results using the Marquardt-Levenberg algorithm for nonlinear least-squares optimization. First, the parameters c and d are determined from the fit of the results for the smallest size ($\xi = 0.001$) in Fig. 10, for varying imperfections. Then the parameters a and b in (26) are fitted for the largest imperfection ($\delta = 6$) and varying size. The optimum values are $a=9.94$, $b=1.2$, $c=6.82$, and $d=1.21$. The size effect law in (26) using these parameters is compared to the results of the softening foundation model in Fig. 10. The approximation is seen to be satisfactory.

To elucidate the typical values of ξ and δ for small laboratory specimens, consider the material properties $E_c=200$ MPa, $\nu_c = 0.25$, $E_s=150$ GPa, $f_t=1$ MPa, $G_F=750$ N/m. Furthermore, let the skin thickness be $t=0.001$ m and let the beam height be so great that the assumption of shortwave wrinkling (Sec. 3) is valid. Equations (9), (7), and (5) give the equivalent height as $h_{eq} = 0.0091$ m. According to (21), the dimensionless size of the beam, is $\xi=0.061$. Furthermore, according to (16), the dimensionless amplitude $\delta=6$ corresponds to an imperfection amplitude of 27% of the skin thickness t. The nominal strength of this beam

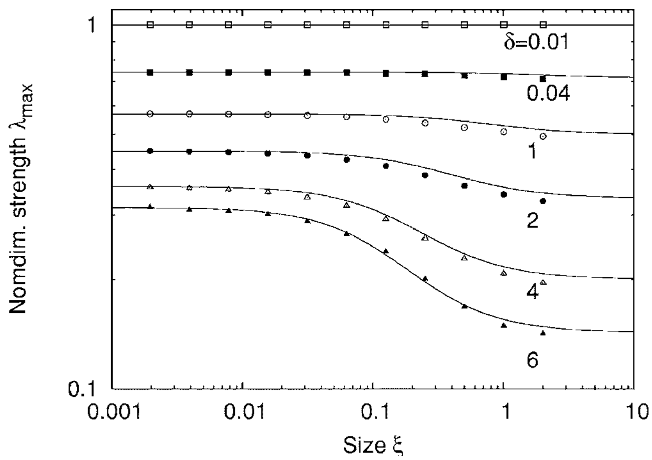


Fig. 10 Comparison of the size effect law in Eq. (26) with the nominal strength-size curves obtained from the softening foundation model for different imperfections

falls into the transitional range of the size effect law in Fig. 10. Therefore, up-scaling leads to a considerable reduction of the nominal strength.

10 Conclusions

1. In view of recent experiments revealing a size effect deviating from both LEFM and strength theory, the delamination fracture of laminate-foam sandwich structures must be treated as a cohesive crack with a softening stress-separation relation characterized by both fracture energy and tensile strength. In contrast to LEFM, no preexisting interface flaw needs to be considered.
2. The skin (or facesheet) can be treated as a beam on elastic-softening foundation, provided that the equivalent (or effective) core depth h_{eq} for which the hypothesis of uniform transverse stress gives the correct foundation stiffness is considered to depend on the critical wavelength L_{cr} of skin wrinkles; h_{eq}=core thickness h for the asymptotic case of longwave wrinkling ($L_{cr}/h \to \infty$), while (because of St. Venant principle) $h_{eq} \propto L_{cr}$ for the asymptotic case of shortwave wrinkling ($L_{cr}/h \to 0$).
3. A properly formulated softening foundation model for the skin in sandwich beams subjected to bending moment can give good agreement with geometrically nonlinear finite element analysis of delamination fracture triggered by wrinkling.
4. Although the nominal strength of sandwich structures failing by wrinkling-induced delamination fracture is size independent when there is no imperfection, it becomes strongly size dependent with increasing imperfection. A size effect causing strength reduction by 50% is possible for larger imperfections. Dents from impact may be expected to have a similar effect, even though they are not merely geometrical imperfections (because of being usually accompanied by initial delamination).
5. Introduction of proper dimensionless variables makes it possible to cover with numerical simulations the entire practical range, and fitting the dimensionless numerical results for cohesive delamination fracture with a formula of correct shortwave and longwave asymptotics allows constructing an approximate size effect law for nominal strength of imperfect sandwich beams subjected to uniform bending moment.
6. There is also a strong size effect on postpeak energy absorption by a sandwich structure, both in the presence and absence of imperfections. This is important for impact and blast resistance.
7. The distinction between shortwave and longwave periodic wrinkling permits clarifying discrepancies among various existing formulas for elastic wrinkling, e.g., Hoff and Mautner's formula [3] applies to the limit of shortwave wrinkling, Heath's formula [4] to the limit of longwave wrinkling.

Appendix: Comments on Transition to Global Buckling and "Naive" Optimization

The case where the opposite skins carry equal (or almost equal) axial force tends to produce skin wrinkling in which both skins deflect in the same direction. This is a different buckling mode—the global buckling of a sandwich beam in which the transverse normal stresses are negligible and the shear resistance of the core is paramount. The critical axial stress in the skins at bifurcation is given by Engesser's formula ([7], p. 32, 736)

$$\sigma_{cr_g} = \left(\frac{1}{2bt}\right)[F_{cr0}^{-1} + (Gbh)^{-1}]^{-1} \tag{A1}$$

Here, G=elastic shear modulus of the core, $F_{cr0}=(\pi^2/L_{ef}^2)R$ =Euler critical axial load corresponding to negligible shear deformation of the core, L_{ef}=effective buckling length of sandwich column, and R=bending stiffness$\approx b[t^3+3t(h+t)^2]E_s/6$ (note that Haringx's formula does not apply to sandwich buckling [33,34]).

A unified condition for bifurcation stress σ in the skin, applicable to both skin wrinkling and global buckling (distinguished now by subscripts w and g) may be written as $(\sigma-\sigma_w)(\sigma-\sigma_g)$ =0, where σ_w, σ_g=maximum stresses for wrinkling of imperfect skin or buckling of imperfect column, each considered alone. However, according to Koiter ([7], Sec. 4.6) proximity of two critical loads generally enhances imperfection sensitivity. Thus, when σ_w and σ_g are nearly equal, the maximum stress σ for combined wrinkling and buckling gets reduced. This could be approximately captured by the following relation:

$$(\sigma - \sigma_w{}^p)(\sigma - \sigma_g{}^r) = r^{p+q} \tag{A2}$$

where p,q are positive empirical constants (probably of the order of 1) and r is a measure of failure stress reduction due to mode interaction. If one term of the product in this equation tends to ∞, the other must tend to 0, i.e., the equation implies the individual critical stresses σ_w and σ_g. Equation (A2) also has the necessary property that the reduction of failure stress σ is maximum when the two critical stresses coincide. These features agree with the well-known fact that the so-called naive optimum designs, in which one or more critical stresses coincide, ought to be avoided. However, verification, calibration or possible modification of Eq. (A2) is beyond the scope of this study.

References

[1] Reissner, M. E., 1937, "On the Theory of Beams Resting on a Yielding Foundation," Proc. Natl. Acad. Sci. U.S.A., **23**, pp. 328–333.
[2] Gough, G. S., Elam, C. F., and de Bruyne, N. A., 1940, "The Stabilisation of a Thin Sheet by a Continuous Supporting Medium," J. R. Aeronaut. Soc., **44**, pp. 12–43.
[3] Hoff, N. J., and Mautner, S. E., 1945, "The Buckling of Sandwich-Type Panels," J. Aeronaut. Sci., **12**, pp. 285–297.
[4] Heath, W. G., 1960, "Sandwich Construction, Part 2: The Optimum Design of Flat Sandwich Panels," Aircr. Eng., **32**, pp. 230–235.
[5] Niu, K., and Talreja, R., 1999, "Modeling of Wrinkling in Sandwich Panels Under Compression," J. Eng. Mech., **125**(8), pp. 875–883.
[6] Bažant, Z. P., 2002, *Scaling of Structural Strength*, Hermes-Penton, London (and 2nd ed., Elsevier, New York, 2005).
[7] Bažant, Z. P., and Cedolin, L., 1991, *Stability of Structures: Elastic, Inelastic, Fracture and Damage Theories*, Oxford University Press, London.
[8] Sallam, S., and Simitses, G. J., 1985, "Delamination Buckling and Growth of Flat, Cross-Ply Laminates," Compos. Struct., **4**, pp. 361–381.
[9] Yin, W. L., Sallam, S., and Simitses, G. J., 1986, "Ultimate Axial Load Capacity of a Delaminated Beam-Plate," AIAA J., **34**, pp. 123–128.
[10] Daniel, I. M., and Ishai, O., 1994, *Engineering Mechanics of Composite Materials*, Oxford University Press, New York.
[11] Wadee, M. A., and Blackmore, A., 2001, "Delamination From Localized Instabilities in Compression Sandwich Panels," J. Mech. Phys. Solids, **49**, pp. 1281–1299.

[12] Wadee, M. A., 2002, "Localized Buckling in Sandwich Struts With Pre-Existing Delaminations and Geometrical Imperfections," J. Mech. Phys. Solids, **50**, pp. 1767–1787.
[13] Frostig, Y., and Thomsen, O. T., 2005, "Non-Linear Behavior of Delamination Unidirectional Sandwich Panels With Partial Contact and a Transversely Flexible Core," Int. J. Non-Linear Mech., **40**, pp. 633–651.
[14] Aviles, F., and Carlsson, L. A., 2005, "Elastic Foundation Analysis of Local Face Buckling in Debonded Sandwich Columns," Mech. Mater., **37**, pp. 1026–1034.
[15] Hutchinson, J. W., and Suo, Z., 1992, "Mixed Mode Cracking in Layered Materials," Adv. Appl. Mech., **29**, pp. 63–191.
[16] Jensen, H. M., and Sheinman, I., 2002, "Numerical Analysis of Buckling-Driven Delamination," Int. J. Solids Struct., **39**, pp. 3373–3386.
[17] Bažant, Z. P., Zhou, Y., and Daniel, I. M., 2006, "Size Effect on Strength of Laminate-Foam Sandwich Plates," ASME J. Eng. Mater. Technol., **128**, pp. 366–374.
[18] Bayldon, J., Bažant, Z. P., Daniel, I. M., and Yu, Q., 2006, "Size Effect on Compressive Strength of Laminate-Foam Sandwich Plates," ASME J. Eng. Mater. Technol., **128**, pp. 169–174.
[19] Bažant, Z. P., and Planas, J., 1998, *Fracture and Size Effect in Concrete and Other Quasibrittle Materials*, CRC Press, Boca Raton.
[20] Remmers, J. J. C., and de Borst, R., 2001, "Delamination Buckling of Fibre-Metal Laminates," Compos. Sci. Technol., **61**, pp. 2207–2213.
[21] Han, T., Ural, A., Chen, C., Zehnder, A. T., Ingraffea, A. R., and Billington, S. L., 2002, "Delamination Buckling and Propagation Analysis of Honeycomb Panels Using a Cohesive Element Approach," Int. J. Fract., **115**, pp. 101–123.
[22] Gdoutos, E. E., Daniel, I. M., and Wang, K.-A., 2003, "Compression Facing Wrinkling of Composite Sandwich Structures," Mech. Mater., **35**, pp. 511–522.
[23] Vadakke, V., and Carlsson, L. A., 2004, "Experimental Investigation of Compression Failure of Sandwich Speciments With Face/Core Debond," Composites, Part B, **35**, pp. 583–590.
[24] Bažant, Z. P., 2004, "Scaling Theory for Quasibrittle Structural Failure," Proc. Natl. Acad. Sci. U.S.A., **101**, pp. 13400–13407.
[25] Tvergaard, V., and Needleman, A., 1980, "On the Localization of Buckling Patterns," ASME J. Appl. Mech., **47**, pp. 613–619.
[26] Vonach, W. K., and Rammerstorfer, F. G., 2000, "The Effects of In-Plane Core Stiffness on the Wrinkling Behavior of Thick Sandwiches," Acta Mater., **141**, pp. 1–10.
[27] Bažant, Z. P., and Oh, B.-H., 1983, "Crack Band Theory for Fracture of Concrete," Mater. Struct., **16**, pp. 155–177.
[28] Suo, Z., Bao, G., and Fan, B., 1992, "Delamination R-Curve Phenomena Due to Damage," J. Mech. Phys. Solids, **40**, pp. 1–16.
[29] Bao, G., and Suo, Z., 1992, "Remarks on Crack-Bridging Concepts," Appl. Mech. Rev., **45**, pp. 355–366.
[30] Massabó, R., and Cox, B. N., 1999, "Concepts for Bridged Mode II Delamination Cracks," J. Mech. Phys. Solids, **47**(6), pp. 1265–1300.
[31] Bažant, Z. P., and Yavari, A., 2004, "Is the Cause of Size Effect on Structural Strength Fractal or Energetic-Stastical," Eng. Fract. Mech., **72**, pp. 1–31.
[32] Bažant, Z. P., 1997, "Scaling of Quasi-Brittle Fracture: Asymptotic Analysis," Int. J. Fract., **83**, pp. 19–40.
[33] Bažant, Z. P., and Beghini, A., 2005, "Which Formulation Allows Using a Constant Shear Modulus for Small-Strain Buckling of Soft-Core Sandwich Structures," ASME J. Appl. Mech., **72**, pp. 785–787.
[34] Bažant, Z. P., and Beghini, A., 2006, "Stability and Finite Strain of Homogenized Structures Soft in Shear: Sandwich or Fiber Composites, and Layered Bodies," Int. J. Solids Struct., **43**, pp. 1571–1593.

Reynolds-Stress Modeling of Three-Dimensional Secondary Flows With Emphasis on Turbulent Diffusion Closure

I. Vallet
Institut d'Alembert,
Université Pierre et Marie Curie,
75005 Paris, France
e-mail: vallet@ccr.jussieu.fr

*The purpose of this paper is to assess the importance of the explicit dependence of turbulent diffusion on the gradients of mean-velocity modeling in second moment closures on three-dimensional (3D) detached and secondary flows prediction. Following recent theoretical work of Younis, Gatski, and Speziale, 2000, [Proc. Royal Society Lon. A, **456**, pp. 909–920], we propose a triple-velocity correlation model, including the effects of the spatial gradients of mean velocity. A model for both the slow and rapid parts of the pressure-diffusion term was also developed and added to a wall-normal-free Reynolds-stress model. The present model is validated against 3D detached and secondary flows. Further developments, especially on the echo terms (which should appear in the formulation of pressure-velocity correlation), are discussed.*
[DOI: 10.1115/1.2722780]

1 Introduction

In second-moment closures, the turbulent diffusion d_{ij}^{T} appearing in the transport equation of the Reynolds-stresses due to velocity fluctuations d_{ij}^{u} is generally modeled with the Daly and Harlow [1] proposal (DH) or the Hanjalić and Launder [2] model (HL), whereas the turbulent diffusion due to pressure fluctuations d_{ij}^{p} is neglected. The model proposed by Lumley [3] was obtained by considering weakly anisotropic and inhomogeneous flow and contains the HL closure [3] as a particular case. Note that the HL model adopts the tensorial form previously proposed by Hirt [4], but with a reoptimized coefficient.

Numerous previous assessments for the triple-velocity correlation [5–8] based on a priori comparisons to experimental or direct numerical simulation (DNS) data have shown that the HL and the Lumley [3] models give the best overall results, in one-dimensional (1D) plane channel flow [9,10] and in three-dimensional (3D) boundary layers [7]. Nevertheless, in general, these three closures (DH, HL, and Lumley) give quite similar results. Taking into account that all of these models are bilinear in the Reynolds stresses and their gradients, Gatski [11] did not consider this conclusion surprising. Furthermore, recent theoretical work [12] suggested the explicit dependence on spatial gradients of mean velocity of the triple-velocity correlation.

Concerning the pressure-diffusion term, there exist few studies and most of them take into account only the slow part of the pressure-velocity correlation. Indeed, the pressure-velocity correlation is often neglected not only because of the lack of experimental or DNS data, but also because this term does not seem important in plane channel flow [9]. As early as 1969, Hirt [4] proposed a function of the Reynolds stresses gradients and therefore modeled only the slow part of the pressure-velocity correlation. The only theoretical closure for the slow part was established in 1978 by Lumley [3] for weakly inhomogeneous flows. This expression was however used in inhomogeneous flows by several authors [13–16]. In 1996, Demuren et al. [15] used an elliptic relaxation approach with the Lumley [3] closure for the slow part of the pressure-velocity correlation and proposed a rapid-part closure for the pressure-diffusion term. More recently, Suga [17] developed a rapid-part pressure-velocity correlation model and used the DH proposal to model both the triple-velocity correlation and the slow-part pressure-diffusion terms.

In the present work, we develop a turbulent-diffusion closure, including mean-velocity gradients both for the triple-velocity correlation and for the pressure-diffusion term. This complete model was added to the low-Reynolds-number second-moment closure developed by Gerolymos and Vallet [18] (GV-RSM), where a unit vector pointing in the turbulence inhomogeneity direction [19] is used to obtain a Reynolds-Stress model (RSM) completely independent of wall topology, which can be used to simulate arbitrary 3D complex geometries.

2 Turbulence Modeling

2.1 Governing Equation. The exact transport equations for the Favre-Reynolds-averaged Reynolds stresses in an inertial frame of reference are

$$\underbrace{\frac{\partial\bar{\rho}\widetilde{u_i''u_j''}}{\partial t}+\frac{\partial(\bar{\rho}\widetilde{u_i''u_j''}\tilde{u}_\ell)}{\partial x_\ell}}_{\text{convection } C_{ij}}=\underbrace{\frac{\partial}{\partial x_\ell}\left(-\bar{\rho}\widetilde{u_i''u_j''u_\ell''}-\overline{p'u_j''}\delta_{i\ell}-\overline{p'u_i''}\delta_{j\ell}+\overline{u_i''\tau_{j\ell}'}+\overline{u_j''\tau_{i\ell}'}\right)}_{\text{diffusion } d_{ij}}$$
$$\underbrace{-\left(\bar{\rho}\widetilde{u_i''u_\ell''}\frac{\partial\tilde{u}_j}{\partial x_\ell}+\bar{\rho}\widetilde{u_j''u_\ell''}\frac{\partial\tilde{u}_i}{\partial x_\ell}\right)}_{\text{production } P_{ij}}+\underbrace{\overline{p'\left(\frac{\partial u_i''}{\partial x_j}+\frac{\partial u_j''}{\partial x_i}-\frac{2}{3}\frac{\partial u_k''}{\partial x_k}\delta_{ij}\right)}}_{\text{redistribution } \phi_{ij}}+\underbrace{\frac{2}{3}\overline{p'\frac{\partial u_k''}{\partial x_k}}\delta_{ij}}_{\text{pressure-dilatation } \phi_p}$$
$$\underbrace{-\left(\overline{\tau_{j\ell}'\frac{\partial u_i''}{\partial x_\ell}}+\overline{\tau_{i\ell}'\frac{\partial u_j''}{\partial x_\ell}}\right)}_{\text{dissipation } \bar{\rho}\varepsilon_{ij}}+\underbrace{\left(-\overline{u_i''}\frac{\partial\bar{p}}{\partial x_j}-\overline{u_j''}\frac{\partial\bar{p}}{\partial x_i}+\overline{u_i''}\frac{\partial\bar{\tau}_{j\ell}}{\partial x_\ell}+\overline{u_j''}\frac{\partial\bar{\tau}_{i\ell}}{\partial x_\ell}\right)}_{\text{density fluctuation effects } K_{ij}} \tag{1}$$

Convection C_{ij}, and production due to the turbulence interaction with the mean-flow gradients P_{ij}, are exact terms, whereas the diffusion $d_{ij}=d_{ij}^{\mathrm{T}}+d_{ij}^{\mu}$ due to turbulent transport $d_{ij}^{\mathrm{T}}=d_{ij}^{u}+d_{ij}^{p}$ and to molecular viscosity d_{ij}^{μ}, the pressure-strain redistribution ϕ_{ij}, and the dissipation $\bar{\rho}\varepsilon_{ij}$ terms require modeling. In all the models used, direct compressibility effects K_{ij} and pressure dilatation ϕ_p were neglected

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 17, 2006; final manuscript received January 18, 2007. Review conducted by Bassam A. Younis.

$$\rho' \ll \bar{\rho}; \quad \overline{u_i''} \cong 0; \quad K_{ij} \cong 0; \quad \phi_p \cong 0 \tag{2}$$

although K_{ij} and ϕ_p have little influence in the subsonic flows studied in the present work, the redistribution and the turbulent-diffusion terms are of prime importance. With these approximations (Eq. (2)), Reynolds and Favre averages are approximately equal

$$\tilde{u}_i \cong \bar{u}_i; \quad \widetilde{u_i'' u_j''} \cong \overline{u_i' u_j'} \tag{3}$$

and for this reason, we will use simple Reynolds averages hereafter.

For all the models studied in the present work, the turbulence length scale was determined by adopting the Launder and Sharma [20] modified dissipation-rate ε^* equation, with a modified diffusion term, where a tensorial diffusion coefficient is used [2,21].

$$\frac{\partial \bar{\rho}\varepsilon^*}{\partial t} + \frac{\partial(\overline{u_\ell}\bar{\rho}\varepsilon^*)}{\partial x_\ell} = \underbrace{\frac{\partial}{\partial x_\ell}\left[C_\varepsilon \frac{\mathrm{k}}{\varepsilon^*}\bar{\rho}\overline{u_m' u_\ell'}\frac{\partial \varepsilon^*}{\partial x_m} + \bar{\mu}\frac{\partial \varepsilon^*}{\partial x_\ell}\right]}_{d_\varepsilon^{\mathrm{T}} + d_\varepsilon^{\mu}} + C_{\varepsilon 1} P_{\mathrm{k}} \frac{\varepsilon^*}{\mathrm{k}}$$
$$- C_{\varepsilon 2}\bar{\rho}\frac{\varepsilon^{*2}}{\mathrm{k}} + \frac{2\bar{\mu}\mu_{\mathrm{T}}}{\bar{\rho}}(\nabla^2 \bar{\vec{V}})^2$$
$$C_\varepsilon = 0.18; \quad C_{\varepsilon 1} = 1.44; \quad C_{\varepsilon 2} = 1.92(1 - 0.3\mathrm{e}^{-\mathrm{Re}_\mathrm{T}^{*2}}) \tag{4}$$

2.1.1 Gerolymos-Vallet Wall-Normal-Free RSM (GV-RSM). In complex flows, such as a large recirculation zone, but even on a flat plate where the normal to the wall and the distance from the wall can be easily determined, the inhomogeneous part of the flow is not confined only in the immediate vicinity of the solid wall. Consequently, the use of conventional echo terms [22,23], with damping functions, where geometric parameters appear explicitly, are inappropriate. Instead, it is preferable to split the redistributive term into an homogeneous and an inhomogeneous part following the proposal of Craft and Launder [24].

Gerolymos and Vallet [18] and Gerolymos et al. [19] introduced a unit vector pointing in the direction of inhomogeneity of the turbulent field $\vec{e}_{\mathrm{I}}$ to replace the geometric wall normals

$$\vec{e}_{\mathrm{I}} = e_{\mathrm{I}_i}\vec{e}_i = \frac{\mathrm{grad}\left\{\dfrac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-\mathrm{Re}_\mathrm{T}/30}]}{1 + 2\sqrt{A_2} + A^{16}}\right\}}{\left\|\mathrm{grad}\left\{\dfrac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-\mathrm{Re}_\mathrm{T}/30}]}{1 + 2\sqrt{A_2} + A^{16}}\right\}\right\|}; \quad \ell_{\mathrm{T}} = \frac{\mathrm{k}^{3/2}}{\varepsilon}; \tag{5}$$

$$\mathrm{Re}_\mathrm{T} = \mathrm{k}^2(\breve{\nu}\varepsilon)^{-1} \tag{5}$$

where ℓ_{T} is the turbulence length scale and Re_T is the turbulent Reynolds number.

The model for pressure-strain and dissipation terms (which adopts, for the homogeneous part the return-to-isotropy model of Rotta [25] and the isotropization of production model of Naot et al. [26], with reoptimized coefficients functions) reads in an inertial frame of reference

$$\phi_{ij} - \bar{\rho}\varepsilon_{ij} = - C_1^{\mathrm{H}}\bar{\rho}\varepsilon a_{ij} - C_2^{\mathrm{H}}\left(P_{ij} - \frac{1}{3}\delta_{ij}P_{mm}\right) + C_1^{\mathrm{I}}\frac{\varepsilon}{\mathrm{k}}\left[\bar{\rho}\overline{u_n' u_m'}e_{\mathrm{I}_n}e_{\mathrm{I}_m}\delta_{ij} - \frac{3}{2}\bar{\rho}\overline{u_n' u_i'}e_{\mathrm{I}_n}e_{\mathrm{I}_j} - \frac{3}{2}\bar{\rho}\overline{u_n' u_j'}e_{\mathrm{I}_n}e_{\mathrm{I}_i}\right] + C_2^{\mathrm{I}}\left[\phi_{nm2}^{\mathrm{H}}e_{\mathrm{I}_n}e_{\mathrm{I}_m}\delta_{ij} - \frac{3}{2}\phi_{in2}^{\mathrm{H}}e_{\mathrm{I}_n}e_{\mathrm{I}_j} - \frac{3}{2}\phi_{jn2}^{\mathrm{H}}e_{\mathrm{I}_n}e_{\mathrm{I}_i}\right] - \frac{2}{3}\delta_{ij}\bar{\rho}\varepsilon \tag{6}$$

with

$$C_1^{\mathrm{H}} = 1 + 2.58AA_2^{1/4}[1 - \mathrm{e}^{-(\mathrm{Re}_\mathrm{T}/150)^2}] \tag{7}$$

$$A_2 = a_{ik}a_{ki}; \quad A_3 = a_{ik}a_{kj}a_{ji}; \quad A = \left[1 - \frac{9}{8}(A_2 - A_3)\right];$$
$$a_{ij} = \frac{\overline{u_i' u_j'}}{\mathrm{k}} - \frac{2}{3}\delta_{ij} \tag{8}$$

where the function C_1^{H} proposed by Launder and Shima [27] contains the anisotropy part of the dissipation tensor ε_{ij} by using a dependence on the anisotropy tensor invariants A_2 and A_3 as suggested by Lumley [3].

Another feature of this model is the form of the functional dependence of the rapid pressure-strain coefficient C_2^{H} on the flatness parameter A (Eqs. (9) and (8)).

$$C_2^{\mathrm{H}} = \min[1, 0.75 + 1.3\max[0, A - 0.55]] \times A^{[\max(0.25, 0.5 - 1.3\max[0, A - 0.55])]}\left[1 - \max\left(0, 1 - \frac{\mathrm{Re}_\mathrm{T}}{50}\right)\right] \tag{9}$$

$$C_1^{\mathrm{I}} = 0.83\left[1 - \frac{2}{3}(C_1^{\mathrm{H}} - 1)\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-(\mathrm{Re}_\mathrm{T}/30)}]}{1 + 2A_2^{0.8}}\right\}\right\|$$
$$C_2^{\mathrm{I}} = \max\left[\frac{2}{3} - \frac{1}{6C_2^{\mathrm{H}}}, 0\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-(\mathrm{Re}_\mathrm{T}/30)}]}{1 + 1.8A_2^{\max(0.6, A)}}\right\}\right\| \tag{10}$$

where the inhomogeneous part coefficients C_1^{I} and C_2^{I} vanish in homogeneous turbulence and are completely independent of wall topology.

The pressure-diffusion term was neglected and the tensorially symmetric expression for the triple-velocity moment introduced by Hirt [4], but with the C_s coefficient proposed by Hanjalić and Launder [2] (Eq. (11)) was used to model the turbulent-diffusion term d_{ij}^{T}

$$d_{ij}^{\mathrm{T}} = \frac{\partial}{\partial x_\ell}(-\bar{\rho}\overline{u_i' u_j' u_\ell'}) = \frac{\partial}{\partial x_\ell}\left[C_s\frac{\mathrm{k}}{\varepsilon}\left(\bar{\rho}\overline{u_i' u_m'}\frac{\partial \overline{u_j' u_\ell'}}{\partial x_m} + \bar{\rho}\overline{u_j' u_m'}\frac{\partial \overline{u_\ell' u_i'}}{\partial x_m} + \bar{\rho}\overline{u_\ell' u_m'}\frac{\partial \overline{u_i' u_j'}}{\partial x_m}\right)\right]; \quad C_s = 0.11 \tag{11}$$

2.1.2 Wall-Normal-Free Launder-Shima-Sharma RSM (WNF-LSS RSM). The previous relation for the unit vector (Eq. (5)) developed by Gerolymos and Vallet [18] can be used in any existing RSM to replace the geometric unit-normals. Following this idea, a wall-normal-free version of the Launder and Shima second-moment closure [27,28], using the Launder and Sharma [20] modified dissipation-rate ε^* equation (Eq. (4)), was developed by Gerolymos et al. [19]. This model retains the $C_2^{\mathrm{H}}(A)$ form proposed by Launder and Shima [27] (Eq. (13)) and the Daly and Harlow turbulent-diffusion model d_{ij}^{T} (DH; Eq. (14)), which take into account a part of pressure diffusion as suggested by Lumley [29] and Launder [30], and confirmed for 2D supersonic flow by Sauret and Vallet [16]. The geometric-normals were replaced by the inhomogeneity-direction-indicator $\vec{e}_{\mathrm{I}}$ (Eq. (5)). The functions C_1^{I} and C_2^{I} were then optimized to obtain the correct plane-channel-flow mean velocity and Reynolds-stress profiles [19]. The model can be summarized as

$$C_1^{\mathrm{H}} = 1 + 2.58AA_2^{1/4}[1 - \mathrm{e}^{-(\mathrm{Re}_\mathrm{T}/150)^2}] \tag{12}$$

$$C_2^{\mathrm{H}} = 0.75\sqrt{A} \tag{13}$$

$$d_{ij}^{\mathrm{T}} = \frac{\partial}{\partial x_\ell}\left[C_s\frac{\mathrm{k}}{\varepsilon}\left(\bar{\rho}\overline{u_\ell' u_m'}\frac{\partial \overline{u_i' u_j'}}{\partial x_m}\right)\right]; \quad C_s = 0.22 \tag{14}$$

$$C_1^{\mathrm{I}} = 0.90\left[1 - \frac{2}{3}(C_1^{\mathrm{H}} - 1)\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-(\mathrm{Re}_\mathrm{T}/30)}]}{1 + 1.8A_2^{0.8}}\right\}\right\|$$

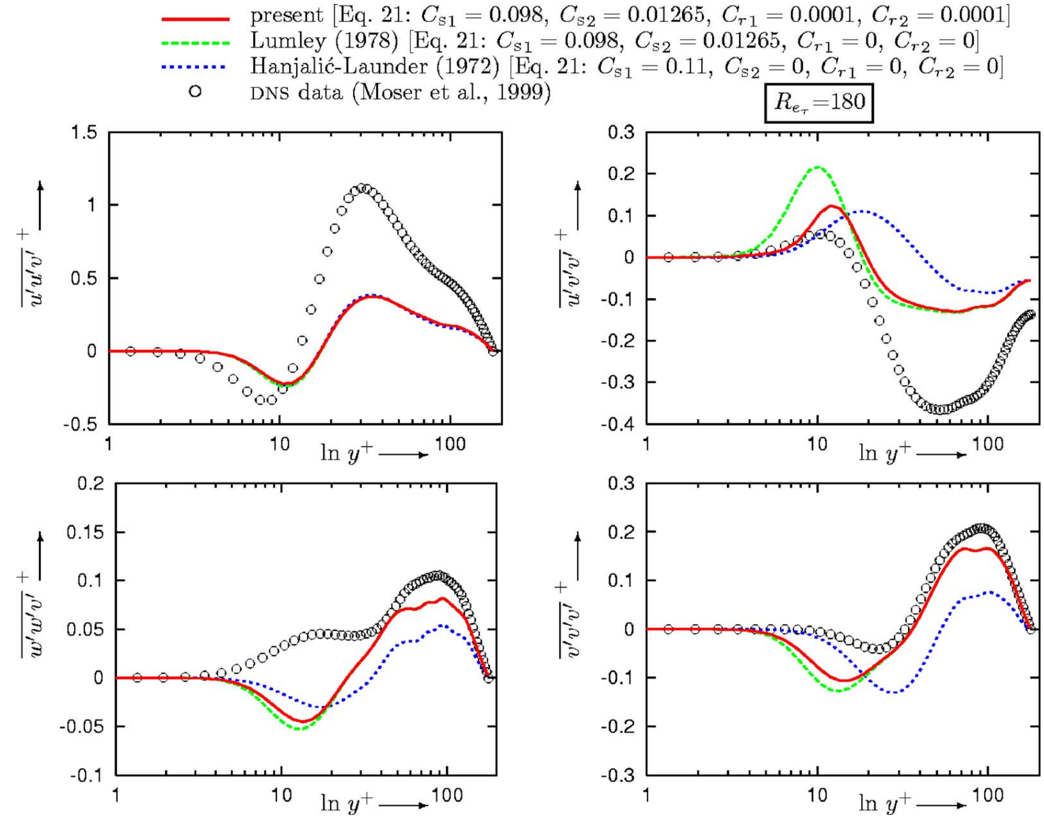


Fig. 1 A priori comparison of three triple-velocity correlation closures $\overline{u_i'u_j'u_\ell'}^+=\overline{u_i'u_j'u_\ell'}/u_\tau^3$ with the DNS data of Moser et al. [10] for fully developed plane channel flow (Re$_\tau$=180; $y^+=yu_\tau/\bar{\nu}$)

$$C_2^{\mathrm{I}} = \max\left[\frac{2}{3} - \frac{1}{6C_2^{\mathrm{H}}}, 0\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1-\mathrm{e}^{-(\mathrm{Re_T}/30)}]}{1+1.8A_2^{\max(0.6,A)}}\right\}\right\| \tag{15}$$

where the coefficients C_1^{H}, C_2^{H}, C_1^{I}, C_2^{I} are used in (Eq. (6)).

2.2 Triple-Velocity Correlations. The turbulent diffusion term d_{ij}^{T} present in the exact transport equations for the Favre-Reynolds-averaged Reynolds stresses is due to velocity fluctuations d_{ij}^{u} (divergence of the triple-velocity moments) and to pressure fluctuations d_{ij}^{p} (divergence of the pressure-velocity correlations)

$$d_{ij}^{\mathrm{T}} = d_{ij}^{u} + d_{ij}^{p} = \frac{\partial}{\partial x_\ell}(-\bar{\rho}\overline{u_i'u_j'u_\ell'}) + \frac{\partial}{\partial x_\ell}(-\overline{p'u_j'}\delta_{i\ell} - \overline{p'u_i'}\delta_{j\ell}) \tag{16}$$

In a previous work [16], we have used the formulation suggested by Lumley [3] for modeling the triple-velocity correlation term $-\overline{u_i'u_j'u_\ell'}$ with the coefficients proposed by Schwarz and Bradshaw [7]

$$-\overline{u_i'u_j'u_\ell'} = C_{s1}\frac{\mathrm{k}}{\varepsilon}G_{ij\ell} + C_{s2}\frac{\mathrm{k}}{\varepsilon}[G_{imm}\delta_{j\ell} + G_{jmm}\delta_{i\ell} + G_{\ell mm}\delta_{ij}] \tag{17}$$

$$C_{s1} = 0.098; \quad C_{s2} = 0.01265;$$

$$G_{ij\ell} = \overline{u_i'u_k'}\frac{\partial\overline{u_j'u_\ell'}}{\partial x_k} + \overline{u_j'u_k'}\frac{\partial\overline{u_i'u_\ell'}}{\partial x_k} + \overline{u_k'u_\ell'}\frac{\partial\overline{u_i'u_j'}}{\partial x_k} \tag{18}$$

Nonetheless, it should be noted that the previous models do not explicitly contain mean-flow velocity gradients. Younis et al. [12] have noted this theoretical drawback in the triple-velocity correlation closures and suggested that the triple-velocity correlation model should be of the following form:

$$\overline{u_i'u_j'u_\ell'} = F_{ij\ell}\left[\overline{u_k'u_m'}, \varepsilon, \frac{\partial\overline{u_k'u_m'}}{\partial x_n}, \bar{S}_{mn}, \bar{W}_{mn}\right]$$

$$\bar{S}_{mn} = \frac{1}{2}\left[\frac{\partial\bar{u}_m}{\partial x_n} + \frac{\partial\bar{u}_n}{\partial x_m}\right]; \quad \bar{W}_{mn} = \frac{1}{2}\left[\frac{\partial\bar{u}_m}{\partial x_n} - \frac{\partial\bar{u}_n}{\partial x_m}\right] \tag{19}$$

where $\bar{S}_{mn}$ and $\bar{W}_{mn}$ are respectively the mean flow rate-of-strain and vorticity tensors. Younis et al. [12] have suggested taking into account the dependence of triple-velocity correlations on mean-flow gradients. Such that the final model corresponded to the HL model [2] plus terms bilinear in $\bar{S}_{mn}$ and $\partial\overline{u_k'u_m'}/\partial x_n$. It should be noted the Hanjalić and Launder [2] model (Eq. (11)) corresponds to the first term of the Lumley [3] model (Eq. (17)) with $C_{s1}=0.11$ and $C_{s2}=0$.

The model proposed by Younis et al. [12] is

$$-\overline{u_i'u_j'u_\ell'} = C_{s1}\frac{\mathrm{k}}{\varepsilon}\left[\overline{u_i'u_k'}\frac{\partial\overline{u_j'u_\ell'}}{\partial x_k} + \overline{u_j'u_k'}\frac{\partial\overline{u_i'u_\ell'}}{\partial x_k} + \overline{u_k'u_\ell'}\frac{\partial\overline{u_i'u_j'}}{\partial x_k}\right] + C_{r1}\frac{\mathrm{k}^3}{\varepsilon^2}\left[\bar{S}_{ki}\frac{\partial\overline{u_j'u_\ell'}}{\partial x_k} + \bar{S}_{kj}\frac{\partial\overline{u_\ell'u_i'}}{\partial x_k} + \bar{S}_{k\ell}\frac{\partial\overline{u_i'u_j'}}{\partial x_k}\right]$$

$$\bar{S}_{ij} = \frac{1}{2}\left[\frac{\partial\bar{u}_i}{\partial x_j} + \frac{\partial\bar{u}_j}{\partial x_i}\right]; \quad C_{s1} = 0.11; \quad C_{r1} = \text{not given} \tag{20}$$

It is quite straightforward to evaluate the triple-velocity correlation models by conducting a priori tests for fully developed plane channel flow and comparing to available DNS data. However, in

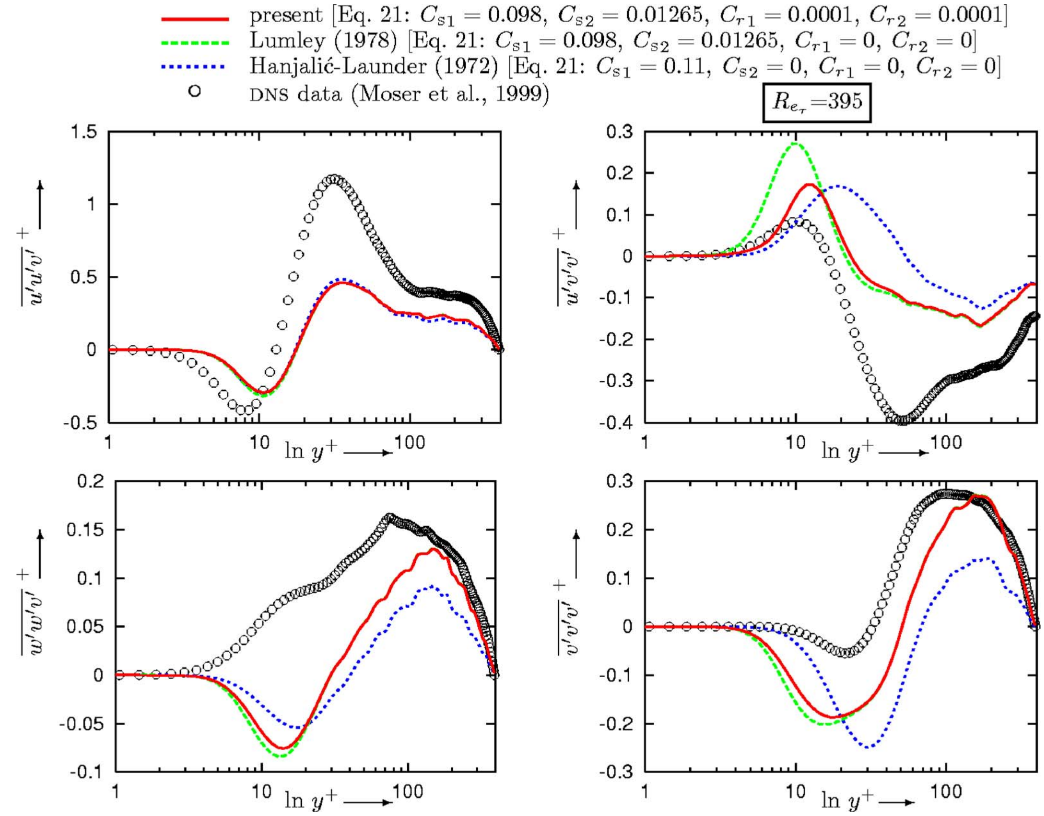


Fig. 2 A priori comparison of three triple velocity correlation closures $\overline{u_i' u_j' u_\ell'}^{+} = \overline{u_i' u_j' u_\ell'} / u_\tau^3$ with the DNS data of Moser et al. [10] for fully developed plane channel flow ($Re_\tau = 395$; $y^+ = y u_\tau / \bar{\nu}$)

fully developed plane channel flow, where $\bar{v}=\bar{w}=0$ and the only nonzero component of the mean flow rate-of-strain tensor is $\bar{S}_{xy} = \bar{S}_{yx} = (1/2)(d\bar{u}/dy)$, so that the additional term proposed by Younis et al. [12] to the HL model only influences the correlations $\overline{u'u'v'}$ and $\overline{u'v'v'}$, which include the fluctuating velocity component u'. Furthermore, this additional term influences the flow mainly close to the wall where the mean flow velocity gradients are important. Considering that the HL model is quite satisfactory close to the wall for the prediction of the $\overline{u'u'v'}$ and $\overline{u'v'v'}$ correlations, the model suggested by Younis et al. [12] gives results very close to the HL model for this test case.

In the present work, we have explored the possibility of developing a model that corresponds to the Lumley [3] triple-velocity correlations model plus a term containing $\bar{S}_{ij}$ in order to have a model available in 3D complex flows. We propose the following form:

$$-\overline{u_i' u_j' u_\ell'} = \underbrace{C_{s1}\frac{\mathrm{k}}{\varepsilon}G_{ij\ell} + C_{s2}\frac{\mathrm{k}}{\varepsilon}[G_{imm}\delta_{j\ell} + G_{jmm}\delta_{i\ell} + G_{\ell mm}\delta_{ij}]}_{-\overline{u_i' u_j' u_\ell'}^{(1)}}$$

$$+ \underbrace{C_{r1}\frac{\mathrm{k}^3}{\varepsilon^2}G^r_{ij\ell} + C_{r2}\frac{\mathrm{k}^3}{\varepsilon^2}[G^r_{imm}\delta_{j\ell} + G^r_{jmm}\delta_{i\ell} + G^r_{\ell mm}\delta_{ij}]}_{-\overline{u_i' u_j' u_\ell'}^{(2)}}$$

$$G_{ij\ell} = \overline{u_i' u_k'}\frac{\partial \overline{u_j' u_\ell'}}{\partial x_k} + \overline{u_j' u_k'}\frac{\partial \overline{u_i' u_\ell'}}{\partial x_k} + \overline{u_k' u_\ell'}\frac{\partial \overline{u_i' u_j'}}{\partial x_k};$$

$$G^r_{ij\ell} = \bar{S}_{ki}\frac{\partial \overline{u_j' u_\ell'}}{\partial x_k} + \bar{S}_{kj}\frac{\partial \overline{u_\ell' u_i'}}{\partial x_k} + \bar{S}_{k\ell}\frac{\partial \overline{u_i' u_j'}}{\partial x_k}$$

$$\bar{S}_{ij} = \frac{1}{2}\left[\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i}\right] \tag{21}$$

with the following coefficients, which were optimized by a priori assessments for fully developed plane channel flows (Figs. 1–3),

$$C_{s1} = 0.098;\quad C_{s2} = 0.01265;\quad C_{r1} = 0.0001;\quad C_{r2} = 0.0001$$

2.3 Pressure-Velocity Correlations

2.3.1 Theory. The redistribution tensor is the most important modeled term. The starting point for developing closures for ϕ_{ij} is, under the assumption of incompressible flow, the Poisson equation for the fluctuating pressure [31]

$$\frac{1}{\bar{\rho}}\nabla^2 p' = \frac{1}{\bar{\rho}}\nabla^2(p_r' + p_s') = \underbrace{-2\frac{\partial u_k'}{\partial x_\ell}\frac{\partial \bar{u}_\ell}{\partial x_k}}_{\text{rapid part}} \underbrace{-\frac{\partial^2}{\partial x_\ell \partial x_k}(u_k' u_\ell' - \overline{u_k' u_\ell'})}_{\text{slow part}} \tag{22}$$

equation (22), obtained for $\bar{\rho}$=const from the fluctuating momentum equation, introduces the idea that solenoidal pressure fluctuations, associated with the fluctuating velocity field, are generated by two separate mechanisms: (i) by the interaction of velocity fluctuations with mean-velocity gradients (called fast or rapid pressure fluctuations because they interact immediately with an imposed mean-velocity gradient) and (ii) by the turbulence-turbulence interaction (called slow pressure fluctuations). This

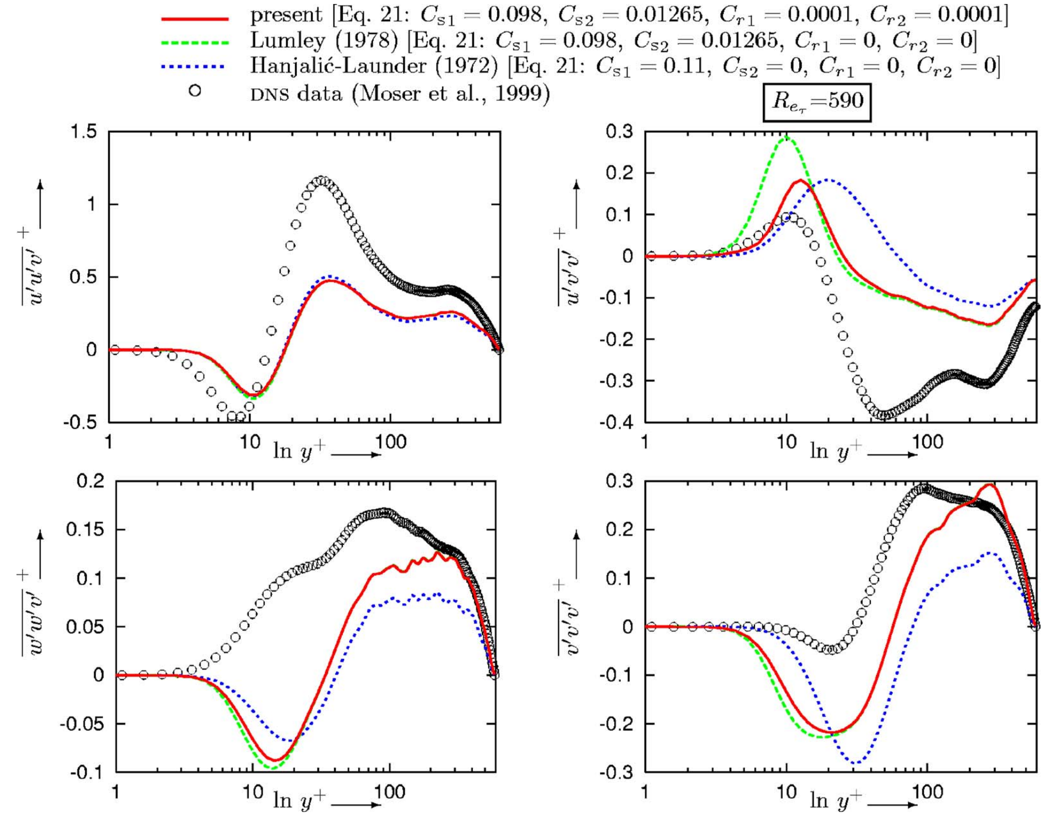


Fig. 3 A priori comparison of three triple velocity correlation closures $\overline{u_i'u_j'u_\ell'}^+=\overline{u_i'u_j'u_\ell'}/u_\tau^3$ with the DNS data of Moser et al. [10] for fully developed plane channel flow ($Re_\tau=590$; $y^+=yu_\tau/\bar\nu$)

idea of distinguishing between pressure fluctuations associated with mean-flow gradients (p_r') and pressure fluctuations that are associated with turbulence-turbulence interactions only (p_s') is applied in general for all the correlations that contain the fluctuating pressure.

The Poisson-equation for pressure (Eq. (22)) can be solved using space integrals and surface integrals [31,32]

$$\frac{1}{\bar\rho}p'(\vec{x}) = \frac{1}{\bar\rho}(p_r' + p_s') = \frac{1}{2\pi}\int\int\int_{\mathfrak{V}}\left(\frac{\partial \bar{\mathfrak{u}}_\ell}{\partial \mathfrak{x}_k}\frac{\partial \mathfrak{u}_k'}{\partial \mathfrak{x}_\ell}\right)\frac{d\mathfrak{v}(\vec{\mathfrak{x}})}{|\vec{\mathfrak{x}}-\vec{x}|}$$
$$+\frac{1}{4\pi}\int\int\int_{\mathfrak{V}}\left(\frac{\partial^2 \mathfrak{u}_k'\mathfrak{u}_\ell'}{\partial \mathfrak{x}_\ell \partial \mathfrak{x}_k} - \frac{\partial^2 \overline{\mathfrak{u}_k'\mathfrak{u}_\ell'}}{\partial x_\ell \partial \mathfrak{x}_k}\right)\frac{d\mathfrak{v}(\vec{x})}{|\vec{\mathfrak{x}}-\vec{x}|}$$
$$+\frac{1}{4\pi\bar\rho}\int\int_{\partial\mathfrak{V}}\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\frac{\partial \mathfrak{p}'}{\partial \mathfrak{n}} - \mathfrak{p}'\frac{\partial}{\partial \mathfrak{n}}\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\right)\right)d\mathfrak{S} \tag{23}$$

where p, u are the pressure and velocity at point $\vec{x}$, and the volume and surface integrals are taken over all other points $\vec{\mathfrak{x}}$ where the pressure and velocity are $\mathfrak{p}$ and $\mathfrak{u}$. Note that if Eq. (23) is premultiplied by a function of $\vec{x}$, this function can be entered into the integrals which are over $\vec{\mathfrak{x}}$. Obviously, in the case of unbounded flow, where $\partial\mathfrak{V}$ is very far away (at infinity; $|\vec{\mathfrak{x}}-\vec{x}|\to\infty$), only the volume integrals remain, the surface integral going to zero. On the other hand, for flow near-solid boundaries, the surface integral indicates that the unsteady pressure field reacts to the presence of the wall (surface integral; $\mathfrak{n}$ is the normal distance from the wall). Terms related to the surface integral are usually called wall-echo terms, since for an infinite plane solid boundary they can be related to reflection from the wall (method of images [32–34]). Obviously, the reflection term (surface integral) contains contributions from (and to) both rapid and slow pressure.

By multiplying the integral equation for the fluctuating pressure by $2S_{ij}'$ (again in the context of incompressible flow, where $S_{ii}=S_{ii}'=0$), and averaging, an integral formula is obtained for ϕ_{ij}

$$\phi_{ij}(\vec{x}) = \phi_{ij\mathrm{r}} + \phi_{ij\mathrm{s}} \equiv \phi_{ij2} + \phi_{ij1} = \frac{1}{\bar\rho}\left[\overline{p'\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)}\right]$$
$$= \frac{1}{2\pi}\int\int\int_{\mathfrak{V}}\left(\frac{\partial \bar{\mathfrak{u}}_\ell}{\partial \mathfrak{x}_k}\overline{\frac{\partial \mathfrak{u}_k'}{\partial \mathfrak{x}_\ell}\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)}\right)\frac{d\mathfrak{v}(\vec{\mathfrak{x}})}{|\vec{\mathfrak{x}}-\vec{x}|}$$
$$+\frac{1}{4\pi}\int\int\int_{\mathfrak{V}}\left[\overline{\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)\left(\frac{\partial^2 \mathfrak{u}_k'\mathfrak{u}_\ell'}{\partial \mathfrak{x}_\ell \partial \mathfrak{x}_k} - \frac{\partial^2 \overline{\mathfrak{u}_k'\mathfrak{u}_\ell'}}{\partial x_\ell \partial \mathfrak{x}_k}\right)}\right]$$
$$\times\frac{d\mathfrak{v}(\vec{\mathfrak{x}})}{|\vec{\mathfrak{x}}-\vec{x}|} + \frac{1}{4\pi\bar\rho}\int\int_{\partial\mathfrak{V}}$$
$$\times\left[\overline{\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\frac{\partial \mathfrak{p}'}{\partial \mathfrak{n}} - \mathfrak{p}'\frac{\partial}{\partial \mathfrak{n}}\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\right)\right)}\right]d\mathfrak{S} \tag{24}$$

The last term is obviously associated with reflection from the wall [8,23,31] (also known as wall-blockage or echo term). It is often modeled using the normal-to-the-wall-direction and the distance from the wall, although in the present work we focus on wall-normal-free models.

In the same way as far the redistribution term ϕ_{ij}, the pressure-velocity correlation, which plays a part in the turbulent-diffusion process (Eq. (16)), can be split into a slow part (designated by the superscript 1) and a rapid part (designated by the superscript 2) plus an echo term (designated by the superscript w)

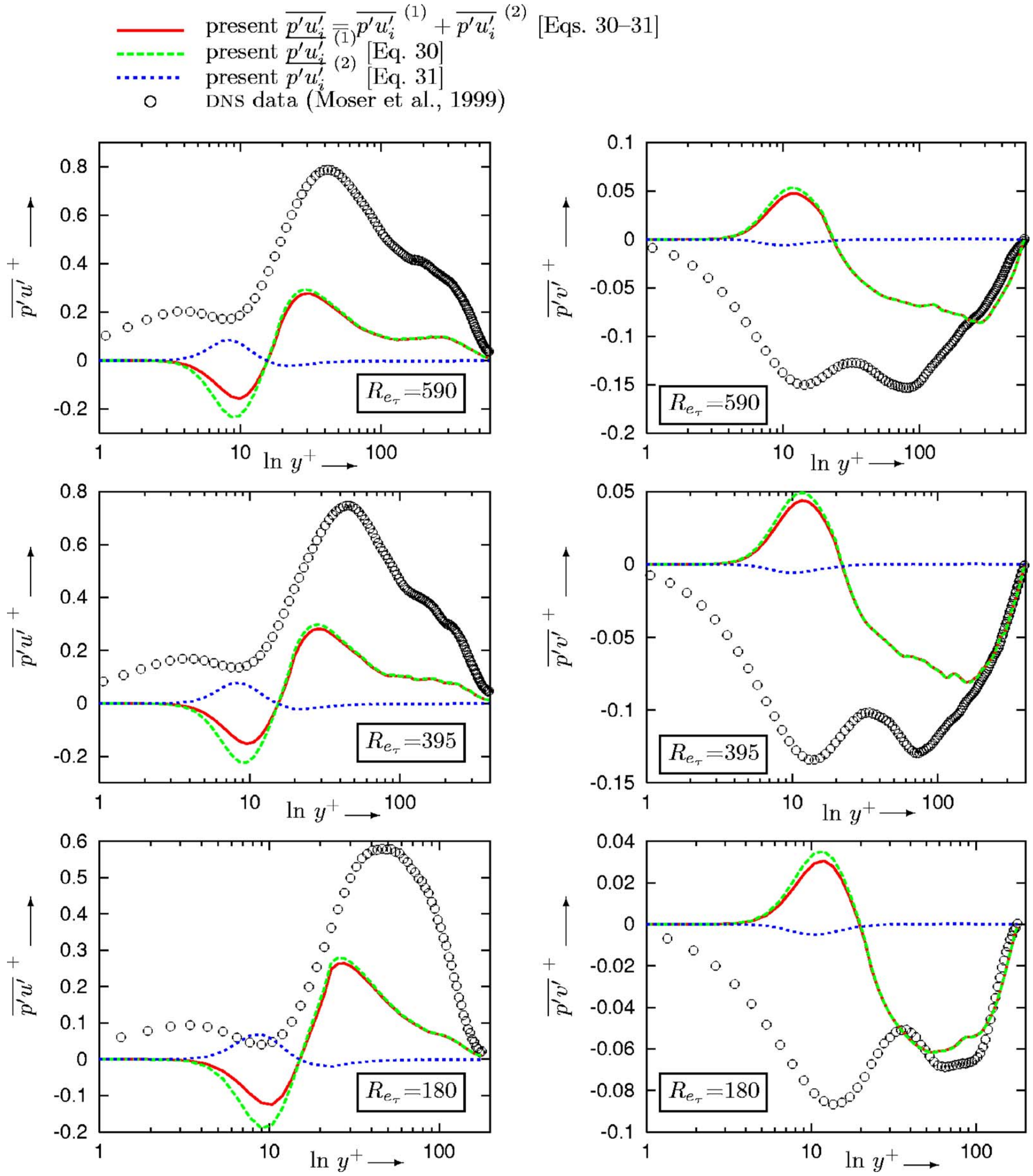


Fig. 4 A priori comparison of pressure-velocity correlation closures $\overline{p'u_i'} = \overline{p'u_i'}^{(1)} + \overline{p'u_i'}^{(2)}$ (Eqs. (30) and (31)) with the DNS data of Moser et al. [10] for fully developed plane channel flow (Re_τ=180, 395, 590); note that $\overline{p'u'}^{+}$ is one-order of magnitude larger than $\overline{p'v'}^{+}$ ($\overline{p'u_i'}^{+} = \overline{p'u_i'}/(\bar{\rho}u_\tau^3)$; $y^+ = yu_\tau/\bar{\nu}$)

$$\frac{1}{\bar\rho}\overline{p'u_i'}(\vec{x}) = \frac{1}{\bar\rho}\overline{p'u_i'} = \underbrace{\frac{1}{2\pi}\int\int\int_{\mathfrak{V}}\left(\frac{\partial\bar{\mathfrak{u}}_l}{\partial\mathfrak{x}_k}\overline{\frac{\partial\mathfrak{u}_k'}{\partial\mathfrak{x}_l}u_i'}\right)\frac{d\mathfrak{v}(\vec{\mathfrak{x}})}{|\vec{\mathfrak{x}}-\vec{x}|}}_{\frac{1}{\bar\rho}\overline{p'u_i'}^{(2)}\equiv\frac{1}{\bar\rho}\overline{p'u_i'}^{(r)}}$$
$$+\underbrace{\frac{1}{4\pi}\int\int\int_{\mathfrak{V}}\left[\overline{u_i'\left(\frac{\partial^2\mathfrak{u}_k'\mathfrak{u}_l'}{\partial\mathfrak{x}_l\partial\mathfrak{x}_k}-\frac{\partial^2\overline{\mathfrak{u}_k'\mathfrak{u}_l'}}{\partial x_l\partial\mathfrak{x}_k}\right)}\right]\frac{d\mathfrak{v}(\vec{\mathfrak{x}})}{|\vec{\mathfrak{x}}-\vec{x}|}}_{1/\bar\rho\overline{p'u_i'}^{(1)}\equiv\frac{1}{\bar\rho}\overline{p'u_i'}^{(s)}}$$
$$+\underbrace{\frac{1}{4\pi\bar\rho}\int\int_{\partial\mathfrak{V}}\left[\overline{u_i'\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\frac{\partial\mathfrak{p}'}{\partial\mathfrak{n}}-\mathfrak{p}'\frac{\partial}{\partial\mathfrak{n}}\left(\frac{1}{|\vec{\mathfrak{x}}-\vec{x}|}\right)\right)}\right]d\mathfrak{S}}_{1/\bar\rho\overline{p'u_i'}^{(w)}} \tag{25}$$

2.3.2 Previous Works. The simplest way to take into account the pressure-diffusion process d_{ij}^{p} is to use an asymmetric turbulent-diffusion closure (since the exact turbulent-diffusion term is not symmetric; Eq. (1)). The Daly and Harlow proposal [29,30] (Eq. (14)) is such a model. However, the DH model is not mathematically correct for 3D complex flows [29]. Furthermore, by not using mean-velocity gradients in the expression, the DH closure neglects the rapid part of the pressure-diffusion term (Eq. (25)).

In 1969, Hirt [4] proposed a symmetric triple-velocity correlation closure (Eq. (11) with C_s=0.33, which is the tensorial form later used by Hanjalić and Launder [2] in the HL model) and an explicit model for the slow part of pressure-velocity correlation

$$\frac{1}{\bar\rho}\overline{p'u_i'}^{(1)} = -\frac{\mathrm{k}^2}{\varepsilon}\frac{\partial\overline{u_i'u_p'}}{\partial x_p} \tag{26}$$

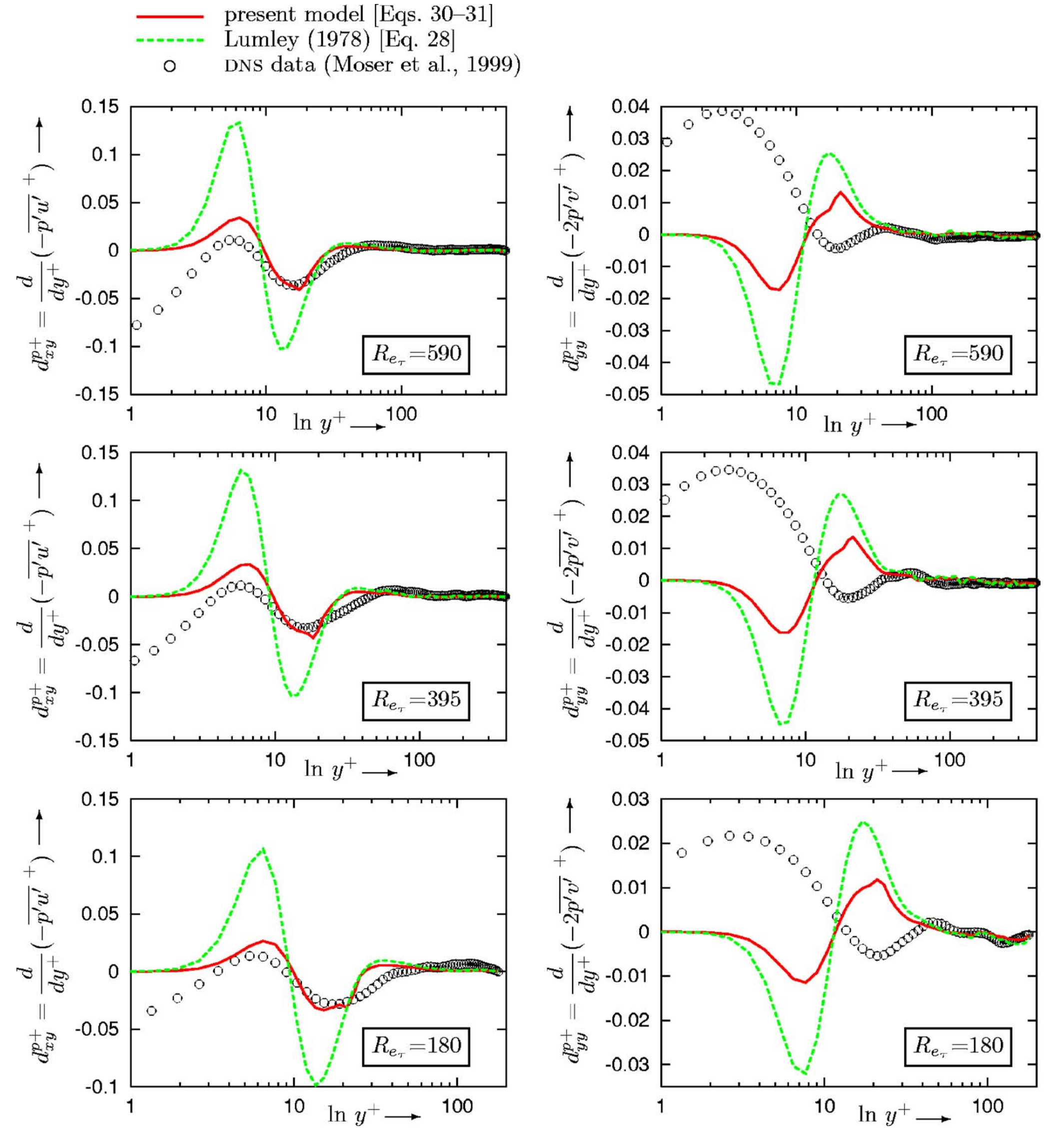


Fig. 5 A priori comparison of pressure-diffusion closures $d^{p+}_{ij}=\partial(-\overline{p'u'_i}^{+}\delta_{jy}-\overline{p'u'_j}^{+}\delta_{iy})/\partial y^+$ ($d^{p+}_{ij}=d^{p}_{ij}/(u_\tau^4/\bar{\nu})$; $y^+=yu_\tau/\bar{\nu}$) with the DNS data of Moser et al. [10] for fully developed plane channel flow (Re_τ=180, 395, 590); note that d^{p+}_{xy} is one-order of magnitude larger than d^{p+}_{yy}

In 1978, Lumley [3] used Fourier transforms to obtain a homogeneous expression for the slow part whose validity is limited to weakly inhomogeneous flows

$$\frac{1}{\bar{\rho}}\overline{p'u_i'}^{(1)} = -\frac{1}{5}\overline{u_i'u_p'u_p'} \tag{27}$$

where the triple-velocity correlation is modeled by the Lumley [3] closure (Eq. (17)). Thus, following the Lumley proposal, the pressure-diffusion term is modeled as $(\overline{p'u_i'}^{(2)}\cong 0)$

$$d^p_{ij} = d^{p(1)}_{ij} = \frac{\partial}{\partial x_\ell}(-\overline{p'u_j'}^{(1)}\delta_{i\ell} - \overline{p'u_i'}^{(1)}\delta_{j\ell}) = \frac{\partial}{\partial x_\ell}(0.2\bar{\rho}\overline{u_j'u_p'u_p'}\delta_{i\ell} + 0.2\bar{\rho}\overline{u_i'u_p'u_p'}\delta_{j\ell}) \tag{28}$$

This theoretical formulation, in its original or modified forms, was used by several authors [13–16].

Recent DNS databases over a backward-facing step [35] or behind a rectangular trailing edge [36] have shown the importance of the turbulent-pressure diffusion term in recirculating flows. In 1996, Demuren et al. [15] developed a model for the rapid part of the pressure-diffusion term

$$d^{p(2)}_{ij} = \frac{\partial}{\partial x_\ell}(-\overline{p'u_j'}^{(2)}\delta_{i\ell} - \overline{p'u_i'}^{(2)}\delta_{j\ell}) = C_r\left(\overline{u_j'u_m'}\frac{\partial \bar{u}_m}{\partial x_\ell}\delta_{i\ell} + \overline{u_i'u_m'}\frac{\partial \bar{u}_m}{\partial x_\ell}\delta_{j\ell}\right); \quad C_r \in [0.1,0.3] \tag{29}$$

whereas the slow part was modeled by using a nonlocal elliptic relaxation approach with the Lumley [3] model (Eq. (27)) for the local source term. They used the Mellor and Herring [37] proposal (which is a simplified version of the HL model) to calculate the triple-velocity correlation, which led to an underprediction of the pressure-diffusion slow part $\overline{p'u_i'}^{(1)}$ in plane mixing layer [38].

In 2004, Suga [17] proposed a linear function of the mean-velocity gradient, the Reynolds stresses, and the length-scale vector to model the rapid part of the pressure-velocity correlation $\overline{p'u_i'}^{(2)}$.

2.3.3 *Present Closure*. We have developed, based on the work of Lumley [3] (Eq. (21)), a near-wall model for the slow part of the pressure-velocity correlation

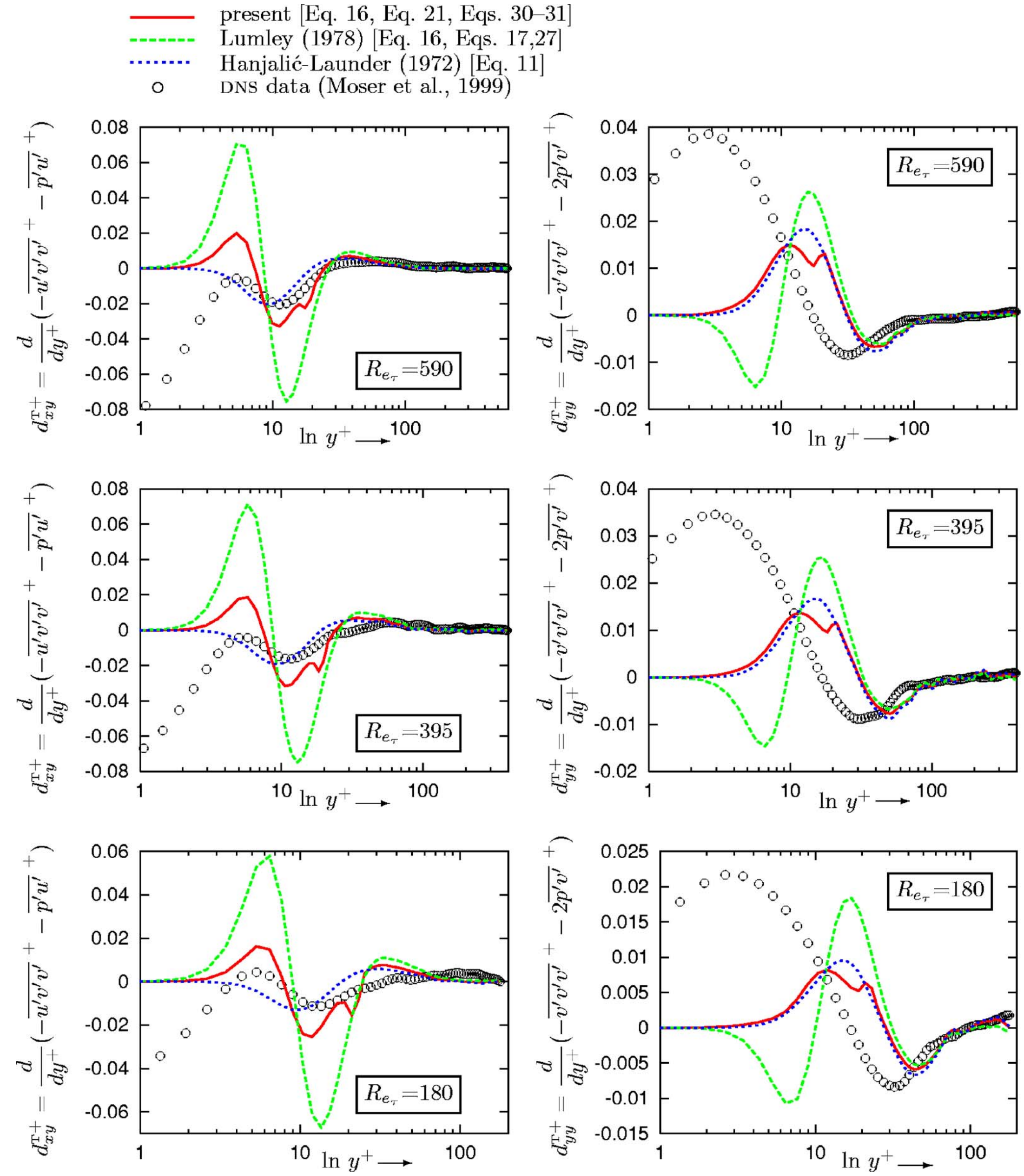


Fig. 6 A priori comparison of turbulent-diffusion closures $d^T_{ij} = d^u_{ij} + d^p_{ij}$ ($d^{T+}_{ij} = d^T_{ij}/(u_\tau^4/\bar{\nu})$; $y^+ = yu_\tau/\bar{\nu}$) with the DNS data of Moser et al. [10] for fully developed plane channel flow (Re_τ=180, 395, 590)

$$\frac{1}{\rho}\overline{p'u_i'}^{(1)} = -C_{SP}\overline{u_i'u_p'u_p'}^{(1)} = -C_{SP}\left\{-C_{S1}\frac{k}{\varepsilon}G_{imm} - C_{S2}\frac{k}{\varepsilon}[G_{imm}\delta_{pp} + 2G_{pmm}\delta_{ip}]\right\}$$

$$G_{ij\ell} = \overline{u_i'u_k'}\frac{\partial\overline{u_j'u_\ell'}}{\partial x_k} + \overline{u_j'u_k'}\frac{\partial\overline{u_i'u_\ell'}}{\partial x_k} + \overline{u_k'u_\ell'}\frac{\partial\overline{u_i'u_j'}}{\partial x_k} \quad C_{S1} = 0.098;$$

$$C_{S2} = 0.01265; \quad C_{SP} = 0.085[1 + \min(0.5, A^{\max[0.25, 2(1-6A)]})] \tag{30}$$

where the coefficient C_{SP} of the pressure diffusion model was modified to account for near-wall effects by using a function of the flatness parameter of Lumley [3] and of the turbulent Reynolds number (Eq. (30)). Indeed, the original value proposed by Lumley [3] C_{SP}=0.2 (Eq. (28)), which means that the pressure-diffusion contribution equals to −20% of the triple-velocity correlation, was determined from mathematical considerations for homogeneous flows, and in consequence this value is too high close to the wall (cf. Sec. 3.3, Fig. 5). Furthermore, the coefficient C_{SP} should not be zero near the wall because the pressure-diffusion term is important in this zone in detached flows [35]. The proposed coefficient C_{SP} value is 0.1275, except close to the wall where it is sharply damped to a value of 0.085 (Eq. (30)). This coefficient was developed by Sauret and Vallet [16] for the slow part of the pressure diffusion and was validated in several complex 3D flows. We used the same near-wall coefficient function C_{SP} to develop a rapid part closure

$$\frac{1}{\rho}\overline{p'u_i'}^{(2)} = -C_{SP}\left\{-(2C_{r1} + 2C_{r2})\frac{k^3}{\varepsilon^2}\left[\bar{S}_{ik}\frac{\partial k}{\partial x_k} + \bar{S}_{pk}\frac{\partial\overline{u_i'u_p'}}{\partial x_k}\right]\right\}$$

$$C_{r1} = 0.0001; \quad C_{r2} = 0.0001; \quad C_{SP} = 0.085[1 + \min(0.5, A^{\max[0.25, 2(1-6A)]})] \tag{31}$$

We also used the recalibrated coefficient C_1^{I} and C_2^{I} proposed by Sauret and Vallet [16]

$$C_1^{\mathrm{I}} = 0.83\left[1 - \frac{2}{3}(C_1^{\mathrm{H}} - 1)\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-(\mathrm{Re}_{\mathrm{T}}/30)}]}{1 + 2.05A_2^{0.8}}\right\}\right\|$$

$$C_2^{\mathrm{I}} = \max\left[\frac{2}{3} - \frac{1}{6C_2^{\mathrm{H}}}, 0\right]\left\|\mathrm{grad}\left\{\frac{\ell_{\mathrm{T}}[1 - \mathrm{e}^{-(\mathrm{Re}_{\mathrm{T}}/30)}]}{1 + 1.5A_2^{\max(0.6,A)}}\right\}\right\| \quad (32)$$

It should be noted that the present model does not take into account the echo part of the pressure-velocity correlation ($\overline{p'u_i'}^{(w)} \cong 0$; Eq. (25)). The echo part is expected to be important in the near-wall region and is the subject of ongoing research. The omission of this term should be kept in mind when comparing the model to data very close to the wall, as will be done next.

3 A Priori Assessments of Turbulent Diffusion Models in Plane Channel Flow

3.1 Triple-Velocity Correlations Models. The present model (Eq. (21)) for the triple-velocity moments was evaluated a priori for fully developed plane channel flows at three different Reynolds numbers ($\mathrm{Re}_\tau = 180, 395, 590$) [10]. We have compared (Figs. 1–3) the present model, the HL model and the Lumley model with DNS data [10].

Unlike the HL model, the Lumley proposal does not predict the correct asymptotic behavior of the $\overline{u'v'v'}$ component. In consequence, the present model containing a dependence on $\bar{S}_{ij}$ is able to adjust the asymptotic behavior and still preserve the comparatively good prediction of the $\overline{u'v'v'}$ component away from the wall. The present model slightly improves the $\overline{w'w'v'}$ and $\overline{v'v'v'}$ components without changing the $\overline{u'u'v'}$ component for the three Reynolds numbers Re_τ tested. It can be seen that this model gives the best overall results.

3.2 Pressure-Velocity Correlation Models. The previous study (Sec. 3.1) highlights the potential improvement in triple-velocity correlation modeling by including the effects of the mean flow rate-of-strain tensor in the closure. The importance of these terms suggests that by taking them into account the pressure diffusion model (Eq. (31)) will potentially improve the prediction of 3D and secondary flows, where mean flow velocity gradients are important. To this purpose, we have pursued the investigation by adding a rapid-part closure ($\overline{p'u_i'}^{(2)}$; Eq. (31)) to the near-wall slow-part pressure diffusion model (Eq. (30)).

We present an a priori assessment of the slow and the rapid parts of the pressure-velocity correlation model (Fig. 4) for fully developed plane channel flows [10] at three different Reynolds numbers (Re_τ=180, 395, 590).

The inclusion of the rapid part $\overline{p'u_i'}^{(2)}$ improves the pressure-velocity correlation prediction, especially for the $\overline{p'u'}$ profiles close to the wall.

Considering that the slow part, which is a direct function of the triple-velocity correlation closure, is dominant especially for the $\overline{p'v'}$ profiles, the present model predicts the correct shape for $\overline{p'u'}$ (not for $\overline{p'v'}$ near the wall), but underestimates the magnitude of the correlation.

For example, for Re_τ=180, the $\overline{p'v'}^+$ profile prediction is in agreement with DNS data for $y^+ > 30$, but fails to reproduce the correct sign of the peak near the wall ($y^+ < 20$).

This is partly attributed to the unsatisfactory triple-velocity-correlation prediction near the wall (Fig. 1), since in the model

$$\frac{1}{\rho}\overline{p'v'}^{(1)} = -C_{\mathrm{SP}}(\overline{u'u'v'} + \overline{v'v'v'} + \overline{w'w'v'}) \quad (33)$$

and certainly to the absence of the echo term, which should be important near the wall for the $\overline{v'v'}$ component.

3.3 Pressure-Diffusion and Turbulent-Diffusion Models. We have compared the Lumley proposal (Eq. (28)) and the present model for the pressure-diffusion term, which is the divergence of the pressure-velocity correlation (Fig. 5).

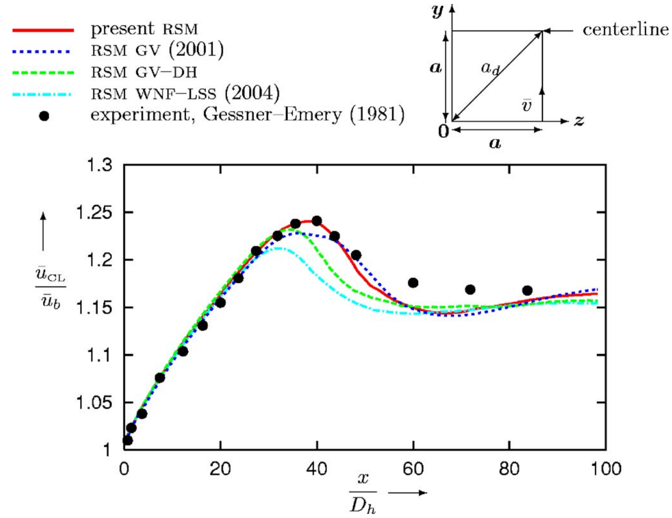


Fig. 7 Comparison of measured [39] streamwise evolution of centerline velocity ($y=z=a$) with computations using four different RSMs for developing flow in a square duct ($\mathrm{Re_B}$=250,000, T_{u_i}=1%, ℓ_{T_i}=50 mm, $\delta_{y_i}=\delta_{z_i}$=0.1 mm, $\Delta y_w^+=\Delta z_w^+<0.5$; 18×10^6 points grid discretizing 1/4 of the square duct)

This comparison is very important because it is term (d_{ij}^p) and not the pressure-velocity correlation $\overline{p'u_i'}$ that appears in the Reynolds-stress transport equation (Eq. (1)). The pressure-diffusion d_{xy}^{p+} of the $\overline{u'v'}$ component is in good agreement with DNS data except for the viscous sublayer zone close to the wall. The Lumley model (Eq. (28)) overpredicts both peaks (with an erroneous sign for the $\overline{v'v'}$ component), especially the one close to the wall, indicating that the slow-part coefficient of 0.20 is too high and that the rapid-part effect is significant enough to be taken into account. However, the turbulent diffusion both for the $\overline{u'v'}$ and the $\overline{v'v'}$ components does not vanish at the wall contrary to the triple-velocity correlation, and the addition of an echo term is required.

Finally, to conclude this a priori assessment, we compared the present model to the turbulent-diffusion closures d_{ij}^{T} proposed by Lumley [3] and Hanjalić and Launder [2] (Fig. 6).

The Lumley model gives the worst prediction near the wall for the $\overline{u'v'}$ and $\overline{v'v'}$ components, mainly due to the too high value of 1/5 for the slow-part coefficient of the pressure-velocity closure (Eq. (27)). Furthermore, the Lumley proposal neglects the rapid-part both of the triple-velocity and the pressure-velocity correlations.

Although, the Hanjalić and Launder [2] model does not take into account the pressure-diffusion term d_{ij}^p, it gives the best overall prediction for the turbulent-diffusion d_{xy}^{T} of the $\overline{u'v'}$ component. The present model, which uses a near-wall coefficient function C_{SP} both for the slow and the rapid parts of the pressure-velocity correlation closure, improves on the Lumley model prediction. Note also that the present model captures the near-wall peak ($y^+ \sim 7$) in the d_{xy}^{T} prediction. However, all three of these models fail near the wall ($y^+ < 10$), indicating the necessity to develop new pressure-diffusion closures with the correct normal-to-the-wall gradient. An improvement of the triple-velocity correlation is also necessary to give satisfactory prediction of the turbulent diffusion of the $\overline{v'v'}$ component.

4 Validation

The present model was then validated against two 3D configurations: (i) developing turbulent flow in a square duct [39] and (ii) high subsonic flow in an aircraft engine inlet S-duct [40]. Care

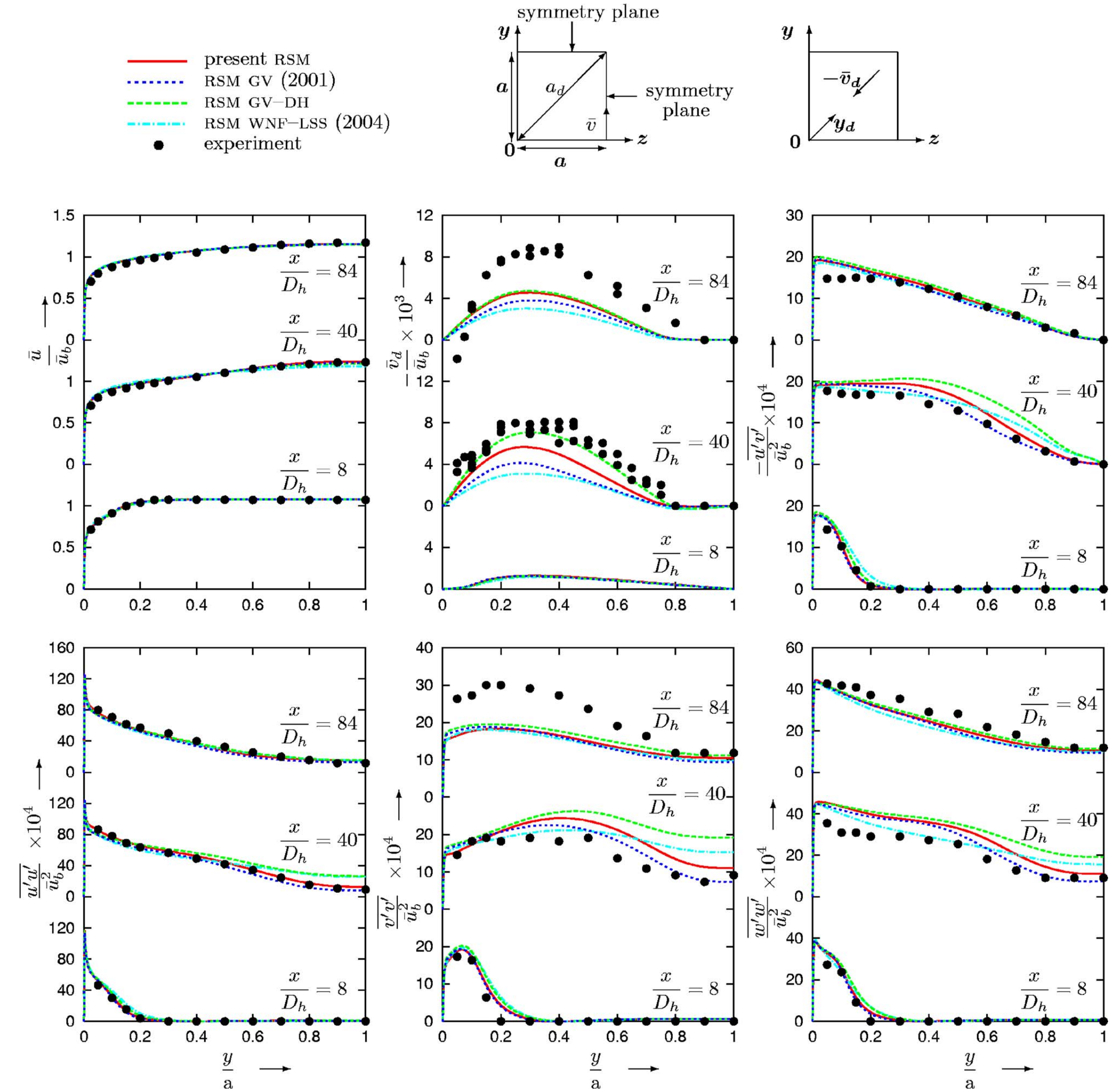


Fig. 8 Comparison of measurements [39] at the symmetry-plane ($z=a$, along y) and grid-converged computations using the four wall-normal-free RSMs at various axial locations x/D_h, for developing turbulent flow in a square duct ($\mathrm{Re_B}=250{,}000$, $T_{u_i}=1\%$, $\ell_{\mathrm{T}_i}=50$ mm, $\delta_{y_i}=\delta_{z_i}=0.1$ mm, $\Delta y_w^+=\Delta z_w^+<0.5$; 18×10^6 points grids discretizing 1/4 of the square duct)

was taken to compare to previous wall-normal-free RSM variants, which use simpler models for d_{ij}^{T}. The computations were performed using the RSM-3D implicit upwind compressible flow solver [41,42].

4.1 Developing Turbulent Flow in a Square Duct. This test case, studied experimentally by Gessner and Emery [39], is very interesting for evaluating the capacity of a model to correctly predict anisotropy-driven secondary flows, where stream-wise vorticity is important. The experimental configuration [39,43], consists of a square duct (height $L_y=L_z=2a=0.254$ m, length $L_x=2a\times100=25$ m). The flow is quasi-incompressible with bulk Reynolds number $\mathrm{Re}_B=250{,}000$ ($\mathrm{Re_B}=\bar{u}_{\mathrm{B}}D_h\nu^{-1}$, where $\bar{u}_{\mathrm{B}}$ is the bulk velocity, $D_h=2a$ is the hydraulic diameter of the duct and ν is the kinematic viscosity). The numerical computations were performed at atmospheric total inlet conditions ($T_{t_i}=288$ K, $p_{t_i}=101{,}325$ Pa) with a turbulence intensity, $T_{u_i}=1\%$, and a turbulence length scale, $\ell_{\mathrm{T}_i}=50$ mm. The outflow pressure was adjusted to obtain the correct $\mathrm{Re_B}$ ($p_o=0.995\times p_{t_i}$, corresponding to an inlet Mach number at centerline $M_{\mathrm{CL}_i}=0.0516$), and the inflow boundary layer was adjusted to a value of $\delta_i=0.1$ mm to obtain a close fit to the experimental centerline velocity in the region $x/D_h\in[0,10]$ (Fig. 7).

This a well-known test case and has been computed by Gessner and Emery [39] using their ARSM closure, and more recently by So and Yuan [43] and Gerolymos et al. [19] using wall-normal-free RSM (with an inflow boundary-layer thickness $\delta_{y_i}=\delta_{z_i}=0.5$ mm too high). Preliminary tests showed that very fine grids were needed to obtain grid-converged results. The results presented were obtained on a 18×10^6 points grid ($N_i\times N_j\times N_k$

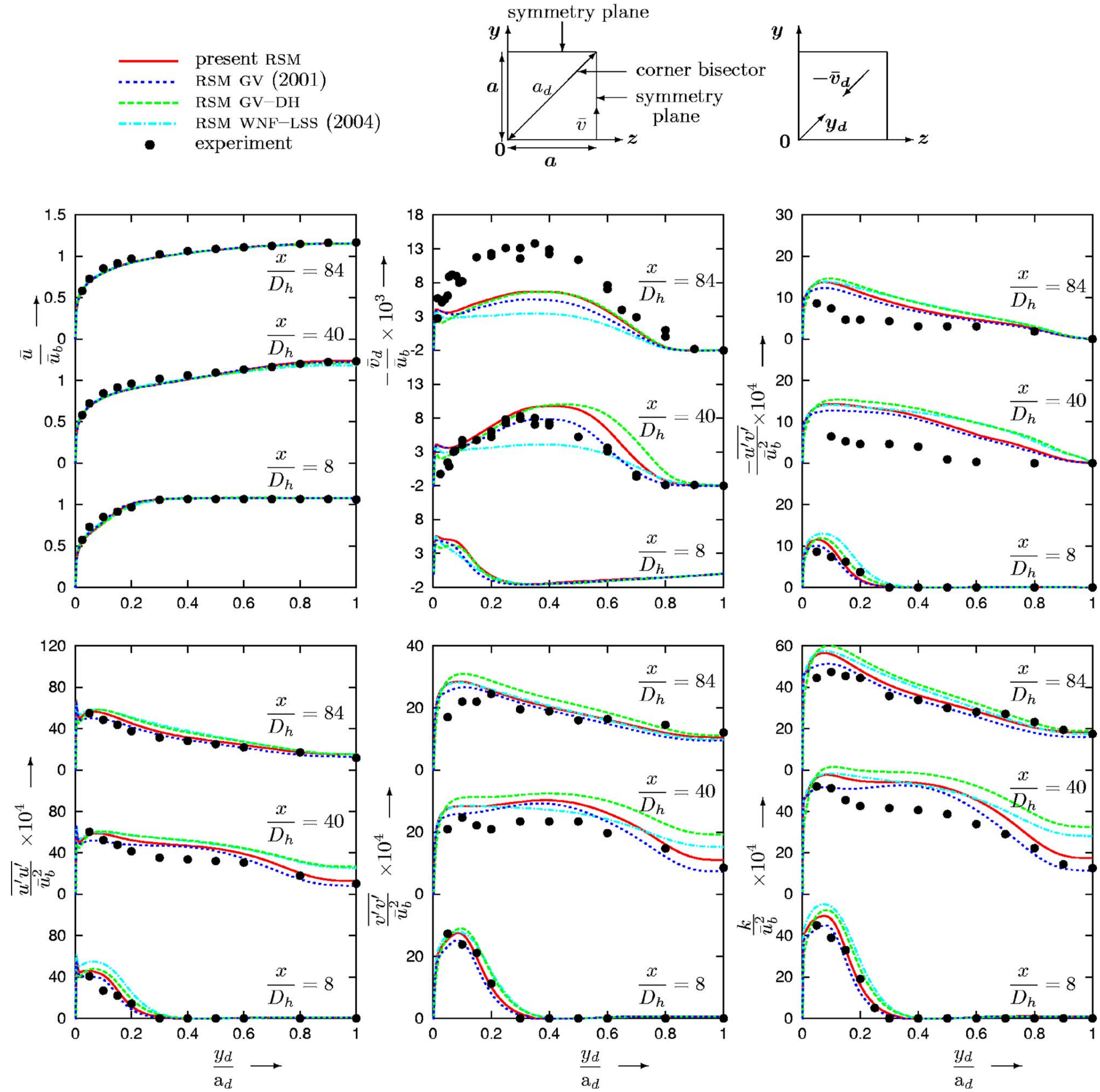


Fig. 9 Comparison of measurements [39] at the corner-bisector (along y_d) and grid-converged computations using the four wall-normal-free RSMs at various axial locations x/D_h, for developing turbulent flow in a square duct ($\mathrm{Re_B}=250{,}000$, $T_{u_i}=1\%$, $\ell_{\mathrm{T}_i}=50$ mm, $\delta_{y_i}=\delta_{z_i}=0.1$ mm, $\Delta y_w^+=\Delta z_w^+<0.5$; 18×10^6 point grids discretizing 1/4 of the square duct)

$=N_x\times N_y\times N_z=801\times149\times149$) discretizing 1/4 of the duct with symmetry conditions at the y- and z-wise symmetry planes. In the y and z directions, the grid was stretched geometrically (65% of the $N_y=N_z$ points were stretched with ratio $r_y=r_z=1.067$, and the remaining 35% were equidistributed in the centerline region).

Previous work on this configuration [19] has highlighted the importance of the turbulent diffusion in correctly predicting the developing 3D turbulent boundary layer entrainment, and as a result, the peak in the centerline velocity $\bar{u}_{\mathrm{CL}}$ at $x\sim40\,D_h$ (Fig. 7). In order to assess the relative importance of two modeling parameters: (i) redistribution coefficient C_2^{H} and (ii) turbulent-diffusion, we compare four RSM variants, all based on quasi-linear isotropization of production and return to isotropy models for the redistribution term ϕ_{ij} (Eq. (6)).

1. Wall-normal-free version of the Launder-Shima-Sharma Reynolds-stress model [19] (WNF-LSS RSM), which has a baseline $C_2^{\mathrm{H}}=0.75\sqrt{A}$ and the DH model for the turbulent-diffusion d_{ij}^{T} (Sec. 2.1.2).
2. Wall-normal-free Reynolds-stress model of Gerolymos and Vallet [18] (GV-RSM) which uses an optimized $C_2^{\mathrm{H}}(A)$ (Eq. (9)), the HL closure for the triple-velocity correlation and neglects the pressure diffusion d_{ij}^{p} (Sec. 2.1.1).
3. GV-DH test-model, which uses $C_2^{\mathrm{H}}(A)$ of GV-RSM (Eq. (9)) and the DH model for the turbulent-diffusion d_{ij}^{T}. This is a test model [19], *not recommended for practical use*, which was developed for the single purpose of testing the influence of the turbulent-diffusion closure.
4. Present model, which uses the $C_2^{\mathrm{H}}(A)$ of GV (Eq. (9)) and

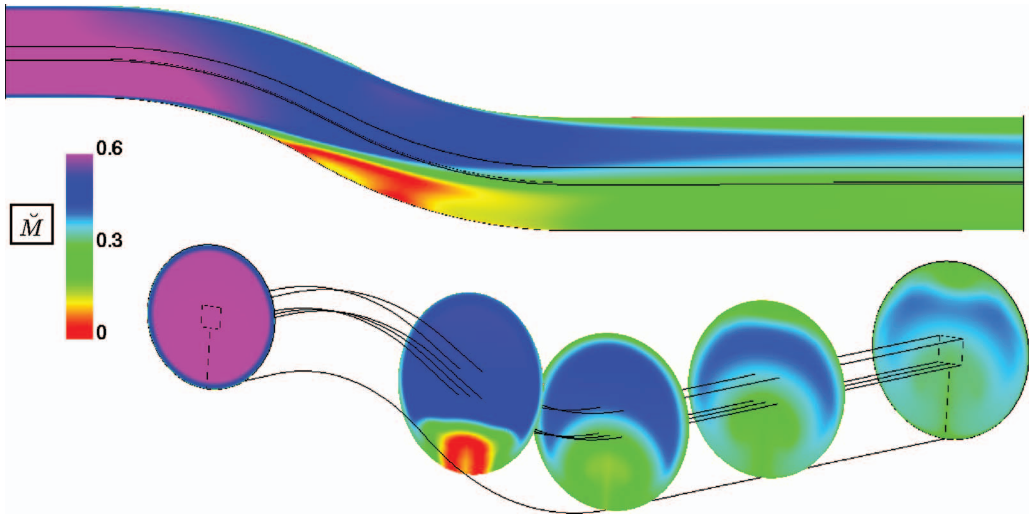


Fig. 10 Iso Mach-number in the S-duct of Wellborn et al. [40], computed with the GV-RSM (Re_{CL} $=2.6\times10^6$, 3.8×10^6 point grids)

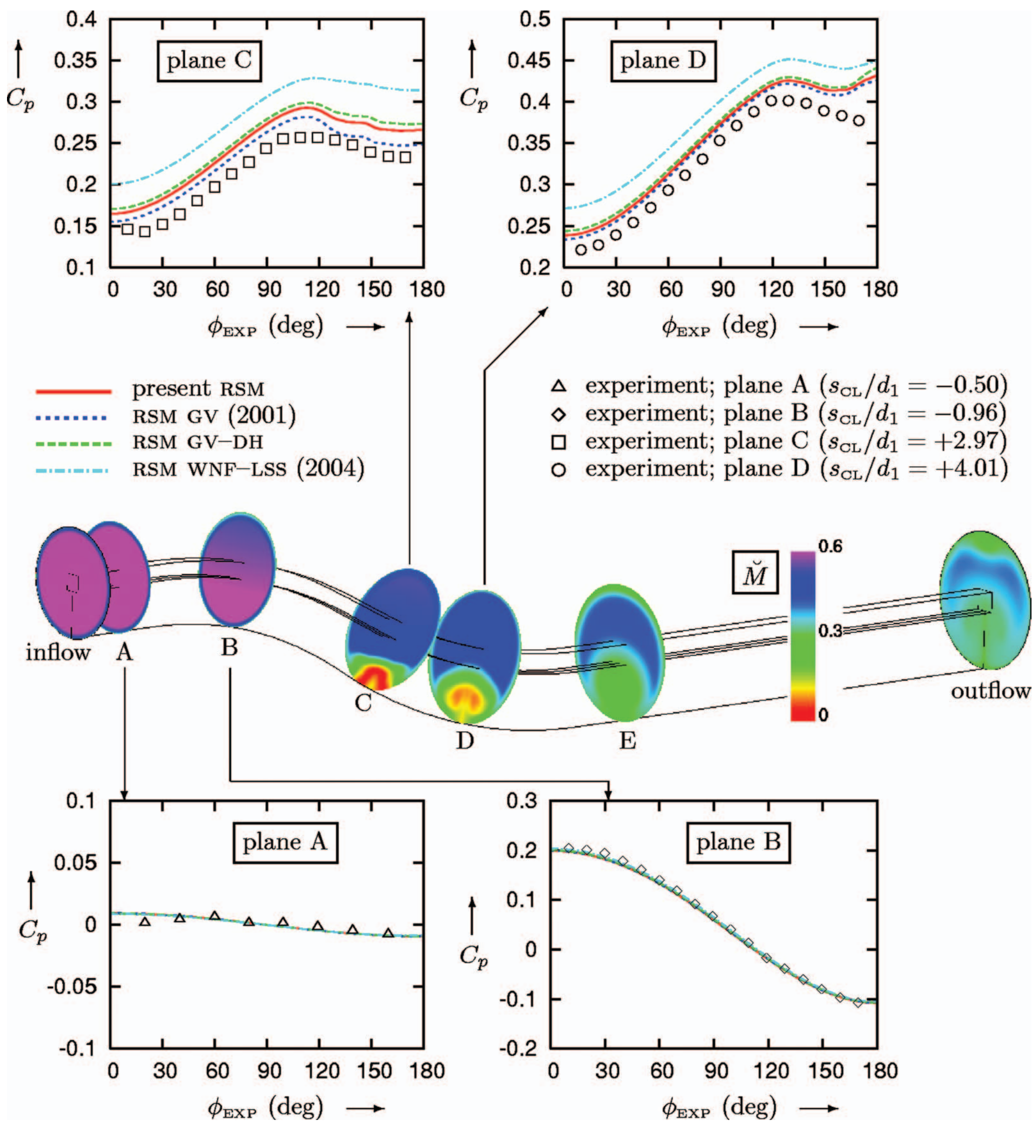


Fig. 11 Comparison of computed and measured pressure coefficient C_p along the circumferential ϕ_{EXP} direction at four planes normal to the duct centerline, using four Reynolds-stress models ($Re_{CL}=2.6\times10^6$, 3.8×10^6 point grids)

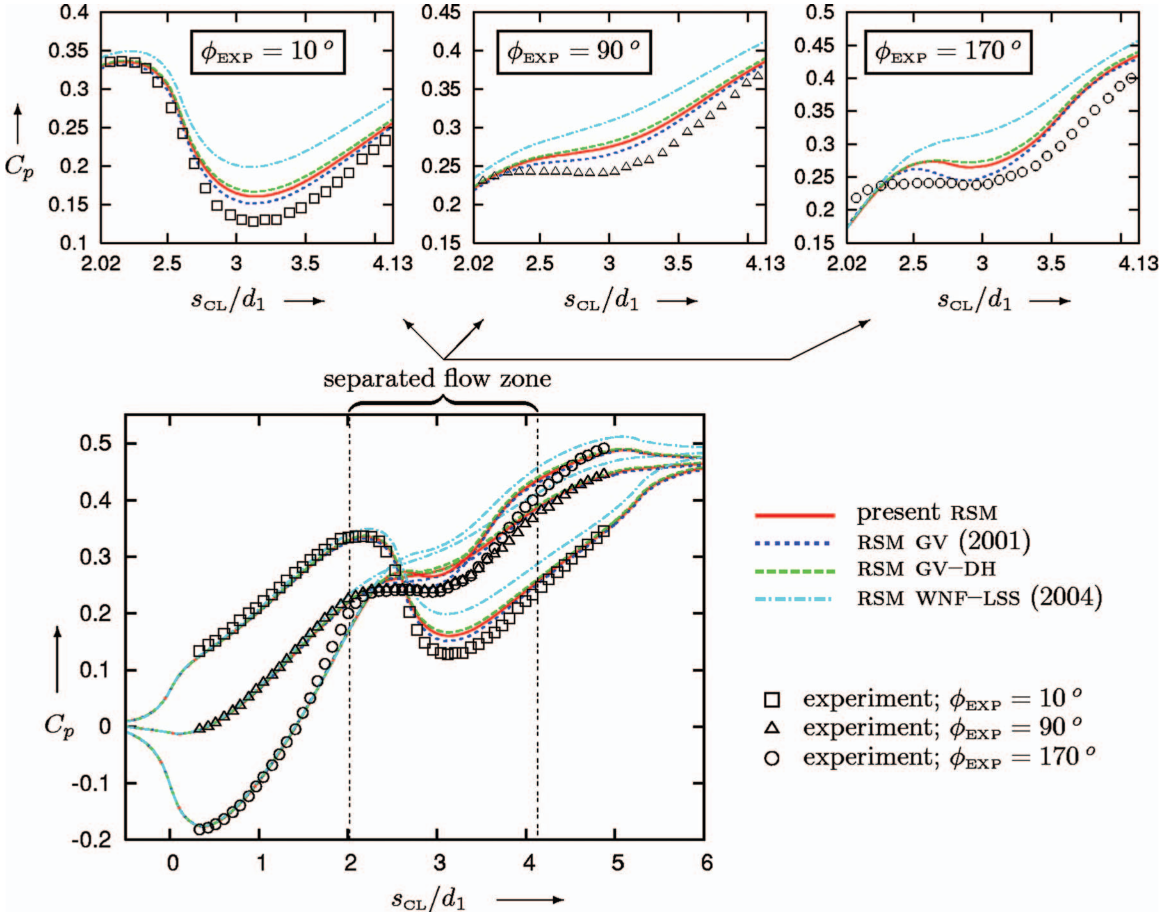


Fig. 12 Comparison of computed and measured pressure coefficient C_p along the centerline s_{CL}/d_1 directions at three circumferential angles ϕ_{EXP} using four Reynolds-stress models ($Re_{CL}=2.6\times10^6$, 3.8×10^6 point grid), with a zoom in the experimental separated flow region between $s_{CL}/d_1=2.02$ and $s_{CL}/d_1=4.13$

includes the new model for the turbulent diffusion d_{ij}^u (Eq. (21)) and d_{ij}^p (Eqs. (30) and (31)), containing the influence of the mean-velocity gradients.

The important contribution of the redistribution coefficient C_2^{H} is clearly seen (Fig. 7) by comparing the WNF-LSS model and the three other models, which use the $C_2^{\mathrm{H}}(A)$ of GV (Eq. (9)). The new model noticeably improves the prediction of this flow, particularly in the peak region $x/D_h \sim 40$ compared to GV-RSM prediction (Fig. 7).

Detailed comparison of velocity and Reynolds-stresses profiles (Figs. 8 and 9) shows that the use of the particular form of the rapid redistribution coefficient C_2^{H}, and enhances the ability of the model to predict the correct turbulence structure. The inclusion of the influence of the mean-velocity gradients in the turbulent-diffusion term is mainly felt near the wall and at the centerline.

4.2 Diffusing 3D S-Duct. The second configuration computed is a diffusing S-duct of circular cross section (Fig. 10), studied experimentally by Wellborn et al. [40]. This configuration was meshed using 3.8×10^6 points grid, consisting of two structured grid blocks [44]. The inflow ($x=-0.2$ m) and outflow ($x=+2$ m) stations are of circular cross section with radii $r_1=10.21$ cm at inflow, and $r_2=12.57$ cm at outflow. Inflow conditions (at $x=-0.2$ m, $s_{\mathrm{CL}}/d_1=-9.8$) were adjusted to obtain the experimentally measured conditions at plane A located at $s_{\mathrm{CL}}/d_1=-0.50$, s_{CL} being the centerline curvilinear coordinate along the duct centerline and $d_1=2r_1=0.2042$ m the inlet section diameter. All the computations were obtained using the following experimental inflow conditions:

$$T_{t_i}=296.4\ \mathrm{K};\quad p_{t_i}=111{,}330\ \mathrm{Pa}$$

$$T_{u_i}=0.65\ \%\ ;\quad \ell_{\mathrm{T}_i}=50\ \mathrm{mm};\quad \delta_{i_{95}}=7.10\ \mathrm{mm};\quad \Pi_i=0.6 \tag{34}$$

where T_{t_i} is the inlet total temperature, p_{t_i} is the inlet total pressure, $T_{u_i}=\sqrt{2/3\mathrm{k}_i}V_i^{-1}$ is the inlet freestream turbulence intensity, $\ell_{\mathrm{T}_i}=\mathrm{k}_i^{2/3}\varepsilon_i^{-1}$ is the inlet freestream turbulence-length scale, $\delta_{i_{95}}$ is the inlet boundary-layer thickness, which corresponds to 95% of the centerline velocity V_{CL}, and Π_i is Coles wake parameter used to define the inflow profile applied as boundary condition [45]. The outflow static pressure $p_o=98{,}100$ Pa was adjusted to obtain the experimental static pressure coefficient $C_p=(p-p_{\mathrm{CL}})/(p_{t_{\mathrm{CL}}}-p_{\mathrm{CL}})=0.466$ at $s/d_1=8.46$.

Comparison of computed and measured results (Figs. 11 and 12), and Table 1 using four RSMs, indicate that the GV-RSM [18] gives quite satisfactory results as do all of the three models which use the modified coefficient $C_2^{\mathrm{H}}(A)$ of the GV-RSM. The turbulent-diffusion term does not substantially influence the results for this configuration. On the other hand, the standard Launder and Shima model ($C_2^{\mathrm{H}}=0.75\sqrt{A}$) substantially underpredicts separation.

Table 1 Plane A ($s/d_1=-0.50$) boundary-layer parameters

	Experiment	Present RSM	GV-RSM	GV-DH RSM	WNF-LSS RSM
M_{CL}	0.6	0.6	0.6	0.61	0.62
$\text{Re}_{\text{CL}}\times 10^{-6}$	2.6	2.61	2.61	2.62	2.65
$\delta_{95}\times 10^{3}$ m	7.1	8.3	8.4	8.4	8.3
$\delta^*_{k_{axi}}/r_1\times 10^2$	1.46	1.42	1.43	1.42	1.36
$\theta_{k_{axi}}/r_1\times 10^2$	1.06	1.03	1.04	1.03	0.99
$H_{k_{axi}}$	1.377	1.375	1.377	1.372	1.363

M_{CL}=centerline Mach-number; $\text{Re}_{\text{CL}}=V_{\text{CL}}d_1\check{\nu}^{-1}$=Reynolds number; $\delta^*_{k_{axi}}$=kinematic displacement-thickness; $\theta_{k_{axi}}$=kinematic momentum thickness; $H_{k_{axi}}=\delta^*_{k_{axi}}\theta^{-1}_{k_{axi}}$=kinematic boundary-layer shape-factor, axisymetric integrate boundary-layer thicknesses defined by Fujii and Okiishi [46].

5 Conclusions

The present work examined the effect of taking into account the mean-velocity gradient on the models for the turbulent-diffusion term in second-moment closures. A model for the triple-velocity correlation and the pressure-diffusion term including mean-velocity gradients was developed and examined both a priori and a posteriori for (i) simple plane channel flows, (ii) developing 3D turbulent flow in a square-duct, and (iii) flow in an aircraft engine inlet S-duct.

Taking into account the mean velocity gradients on the models for the triple-velocity correlation for fully developed plane channel flow (Re_τ=180, 180, 396, and 590) improves the prediction of the asymptotic behavior near the wall and gives the best overall results compared to the HL and Lumley proposals. The explicit modeling of the rapid part of the pressure-velocity correlation gives better prediction of the pressure-diffusion term than the Lumley model. However, the model proposed for the rapid part of the pressure-diffusion term is a function of triple-velocity correlation which vanish at the wall contrary to the pressure-velocity correlation. Further developments, particularly for the wall-echo effect in the pressure-diffusion term, are necessary to predict the correct wall value of the turbulent-diffusion term.

A priori assessments for fully developed plane channel flow are very useful for developing modeling expressions for unclosed terms. Nevertheless, this type of assessment is not sufficient for evaluating models where all terms in the transport equations are coupled, and validations in 3D complex flows are necessary. The complete model proposed for the turbulent-diffusion term, which includes the mean-velocity gradients, improves the prediction of the secondary flow in a square duct. On the other hand, for the S-duct, where large separation is present, the redistribution process seems more important than turbulent diffusion.

Acknowledgment

The computations presented in this work were performed at the Institut pour le Développement des Ressources en Informatique Scientifique (IDRIS), where computer ressources were made available by the Comité Scientifique.

References

[1] Daly, B. J., and Harlow, F. H., 1970, "Transport Equations in Turbulence," Phys. Fluids, **13**, pp. 2634–2649.
[2] Hanjalić, K., and Launder, B. E., 1972, "A Reynolds Stress Model of Turbulence and its Application to Thin Shear Flows," J. Fluid Mech., **52**, pp. 609–638.
[3] Lumley, J. L., 1978, "Computational Modeling of Turbulent Flows," Adv. Appl. Mech., **18**, pp. 123–176.
[4] Hirt, C. W., 1969, "Generalized Turbulence Transport Equations," Int Sem. of Int. Center for Heat and Mass Transfer, Herceg Novi, Yugoslavia.
[5] Cormack, D. E., Leal, L. G., and Seinfeld, J. H., 1978, "An Evaluation of Mean Reynolds Stress Turbulence Models: The Triple Velocity Correlation," ASME J. Fluids Eng., **100**, pp. 47–54.
[6] Amano, R. S., and Goel, P., 1986, "Triple-Velocity Products in a Channel With Backward Facing Step," AIAA J., **24**, pp. 1040–1043.
[7] Schwarz, W. R., and Bradshaw, P., 1994, "Term-by-Term Tests of Stress-Transport Turbulence Models in a 3-D Boundary Layer," Phys. Fluids, **6**, pp. 986–998.
[8] Hanjalić, K., 1994, "Advanced Turbulence Closure Models: A View of Current Status and Future Prospects," Int. J. Heat Fluid Flow, **15**, pp. 178–203.
[9] Kim, J., Moin, P., and Moser, R., 1987, "Turbulence Statistics in Fully Developed Channel Flow at Low-Reynolds-Number," J. Fluid Mech., **177**, pp. 133–166.
[10] Moser, R. D., Kim, J., and Mansour, N. N., 1999, "Direct Numerical Simulation of Turbulent Channel Flow up to Re_τ=590," Phys. Fluids, **11**, pp. 943–945.
[11] Gatski, T. B., 2004, "Constitutive Equations for Turbulent Flows," Theor. Comput. Fluid Dyn., **18**, pp. 345–369.
[12] Younis, B. A., Gatski, T. B., and Speziale, C. G., 2000, "Towards a Rational Model for the Triple Velocity Correlations of Turbulence," Proc. R. Soc. London, Ser. A, **456**, pp. 909–920.
[13] Fu, S., 1993, "Modelling of the Pressure-Velocity Correlation in Turbulence Diffusion," Comput. Fluids, **22**, pp. 199–205.
[14] Straatman, A. G., 1999, "A Modified Model for Diffusion in Second-Moment Turbulence Closures," ASME J. Fluids Eng., **121**, pp. 747–756.
[15] Demuren, A. O., Rogers, M. M., Durbin, P., and Lele, S. K., 1996, "On Modelling Pressure Diffusion in Non-homogeneous Shear Flows," *Studying Turbulence Using Numerical Simulation Databases-VI, Proceeding of the CTR Summer Program 1996*, Center for Turbulence Research, NASA Ames/Stanford University.
[16] Sauret, E., and Vallet, I., 2007, "Near-Wall Turbulent Pressure Diffusion Modeling and Influence in 3-D Secondary Flows," ASME J. Fluids Eng., to be published.
[17] Suga, K., 2004, "Modeling the Rapid Part of the Pressure-Diffusion Process in the Reynolds Stress Transport Equation," ASME J. Fluids Eng., **126**, pp. 634–641.
[18] Gerolymos, G. A., and Vallet, I., 2001, "Wall-Normal-Free Near-Wall Reynolds-Stress Closure for 3-D Compressible Separated Flows," AIAA J., **39**, pp. 1833–1842.
[19] Gerolymos, G. A., Sauret, E., and Vallet, I., 2004, "Contribution to the Single-Point-Closure Reynolds-Stress Modelling of Inhomogeneous Flow," Theor. Comput. Fluid Dyn., **17**, pp. 407–431.
[20] Launder, B. E., and Sharma, B. I., 1974, "Application of the Energy Dissipation Model of Turbulence to the Calculation of Flows Near a Spinning Disk," Lett. Heat Mass Transfer, **1**, pp. 131–138.
[21] Launder, B. E., Tselepidakis, D. P., and Younis, B. A., 1987, "A Second-Moment Closure Study of Rotating Channel Flow," J. Fluid Mech., **183**, pp. 63–75.
[22] Shir, C. C., 1973, "A Preliminary Numerical Study of Atmospheric Turbulent Flows in the Idealized Planetary Boundary Layer," J. Atmos. Sci., **30**, pp. 1327–1339.
[23] Gibson, M. M., and Launder, B. E., 1978, "Ground Effects on Pressure Fluctuations in the Atmospheric Boundary-Layer," J. Fluid Mech., **86**, pp. 491–511.
[24] Craft, T. J., and Launder, B., 1996, "A Reynolds-Stress Model Designed for Complex Geometries," Int. J. Heat Fluid Flow, **17**, pp. 245–254.
[25] Rotta, J., 1951, "Statistische Theorie Nichthomogener Turbulenz—1. Mitteilung," Chem. Eng. Prog., **129**, pp. 547–572.
[26] Naot, D., Shavit, A., and Wolfshtein, M., 1970, "Interactions Between Components of the Turbulent Velocity Correlation Tensor Due to Pressure Fluctuations," Isr. J. Technol., **8**, pp. 259–269.
[27] Launder, B. E., and Shima, N., 1989, "2-Moment Closure for the Near-Wall Sublayer: Development and Application," AIAA J., **27**, pp. 1319–1325.
[28] Gerolymos, G. A., and Vallet, I., 1997, "Near-Wall Reynolds-Stress 3-D Transonic Flows Computation," AIAA J., **35**, pp. 228–236.
[29] Lumley, J. L., 1979, "Second Order Modeling of Turbulent Flows," *Prediction Methods for Turbulent Flows*, VKI Lecture Series, von Kármán Institute for Fluid Dynamics, 72 Chaussée de Waterloo, 1640 Rhode-sur-Génèse.
[30] Launder, B. E., Reynolds, W. C., Rodi, W., Mathieu, J., and Jeandel, D., 1984, *Turbulence Models and Their Applications II*, No. 56 in Collection de la Direction des Etudes et Recherches d'Electricité de France, Eyrolles, Paris.
[31] Chou, P. Y., 1945, "On Velocity Correlations and the Solutions of the Equations of Turbulent fluctuations," Q. Appl. Math., **3**, pp. 38–54.
[32] Jovanović, J., 2004, *The Statistical Dynamics of Turbulence*, Springer, New York.

[33] Piquet, J., 1999, *Turbulent Flows—Models and Physics*, Springer-Verlag, Berlin.
[34] Pope, S. B., 2000, *Turbulent Flows*, Cambridge University Press, Cambridge.
[35] Le, H., Moin, P., and Kim, J., 1997, "Direct Numerical Simulation of Turbulent Flow Over a Backward-Facing Step," J. Fluid Mech., **330**, pp. 349–374.
[36] Yao, Y. F., Thomas, T. G., and Sandham, N. D., 2001, "Direct Numerical Simulation of Turbulent Flow Over a Rectangular Trailing Edge," Theor. Comput. Fluid Dyn., **14**, pp. 337–358.
[37] Mellor, G. L., and Herring, H. J., 1973, "A Survey of the Mean Turbulent Field Closure Models," AIAA J., **11**, pp. 590–599.
[38] Rogers, M. M., and Moser, R. D., 1994, "Direct Simulation of a Self-Similar Turbulent Mixing Layer," Phys. Fluids, **6**, pp. 903–923.
[39] Gessner, F. B., and Emery, A. F., 1981, "The Numerical Prediction of Developing Turbulent Flow in Rectangular Ducts," ASME J. Fluids Eng., **103**, pp. 445–455.
[40] Wellborn, S. R., Reichert, B. A., and Okiishi, T. H., 1994, "Study of the Compressible Flow in a Diffusing S-Duct," J. Propul. Power, **10**, pp. 668–675.
[41] Gerolymos, G. A., and Vallet, I., 2005, "Mean-Flow-Multigrid for Implicit Reynolds-Stress-Model Computations," AIAA J., **43**, pp. 1887–1898.
[42] Gerolymos, G. A., and Vallet, I., 2007, "Advances in the Numerical Computation of Complex Flows Using Reynolds-Stress Models," AIAA Computational Fluid Dynamics Conference, Jun 25–28, Miami.
[43] So, R. M. C., and Yuan, S. P., 1999, "A Geometry Independent Near-Wall Reynolds-Stress Closure," Int. J. Eng. Sci., **37**, pp. 33–57.
[44] Gerolymos, G. A., and Vallet, I., 2007, "Reynolds-stress Model Computations of Aircraft Engine Inlet *s*-Duct Flows," AIAA Fluid Dynamics Conference, Jun 25–28, Miami.
[45] Gerolymos, G. A., Sauret, E., and Vallet, I., 2004, "Influence of Inflow-Turbulence in Shock-Wave/Turbulent-Boundary-Layer Interaction Computations," AIAA J., **42**, pp. 1101–1106.
[46] Fujii, S., and Okiishi, T. H., 1972, "Curved Diffusing Annulus Turbulent Boundary-Layer Development," J. Aircr., **9**, pp. 97–98.

J. K. Jiang
Graduate Student

Y. R. Iwai
Professor

Department of Mechanical Engineering,
Fukui University,
Fukui 910-8507, Japan

H. Su
Associate Professor
Department of Mechanical Engineering,
Xi'an Polytechnic University,
P.R.C.

Minimizing and Restricting Vibrations in High-Speed Cam-Follower Systems Over a Range of Speeds

Over a range of cam speeds, the problems of minimizing and restricting vibrations in high-speed cam-follower systems over a range of speeds are formulated as constrained optimization problems. A universal Hermite cam displacement is suggested. The combination of these methods can form a general design environment. In this environment, the designer can arbitrarily select one to formulate as the objective function from quantifiable cam properties such as the residual vibration, primary vibration, pressure angle, radius of curvature, and contact stress, while the rest of quantifiable cam properties can be formulated as the design constraints. [DOI: 10.1115/1.2723812]

1 Introduction

The Polydyne approach [1,2] has been a successful method in eliminating residual vibrations at one or multiple design cam speeds. However, in many cases, the working speed of cams is not one or several fixed speeds but must be a range of speeds. To reduce residual vibrations over a range of speeds, Kwakernaak and Smit [3] introduced a linear programming and a quadratic optimization procedure. Kanzaki and Itao [2] reduced residual vibrations over a wider range of speeds by selectively extinguishing residual vibrations at adjacent design speeds. Wiederrich and Roth [4] used a Lagrange multiplier technique to minimize the output acceleration. Fabien, Longman, and Freudenstein [5] applied an optimal control theory to the cam design. Chew and Chuang [6] adopted a generalized Lagrange multiplier technique to minimize residual vibrations.

On the other hand, the traditional cam design process includes the specification of motion curves for the follower and the subsequent calculation of the cam profile. Once the profile is generated, the radius of curvature, pressure angles, and other properties of the profile are checked for feasibility, and the process is repeated until a feasible design is generated. Although there are such standard motion curves as 4-5-6-7 polynomial for choosing, they cannot always satisfy the actual needs. The universal motion curves like the B-splines [8] have made up for lack of standard curves in quantity to a great degree, but the case that a feasible design may not be obtained without several repetitions is not improved, let alone a best design under the given conditions. In order to make every design that is not only a feasible but also the best one, an optimization method should be introduced. There have been some successful examples [2–7] that optimization technology is applied to cam design. Nevertheless, the majority of them are aimed at the specified conditions.

This paper reports the research results of minimizing or restricting residual vibrations and primary vibrations over a range of speeds. Minimizing vibrations and restricting vibrations are the different patterns to control vibrations. Minimizing vibrations is to obtain the cam profiles with the lowest vibration over a range of speeds, while restricting vibrations is to obtain the cam profiles with vibrations within the specified value over a range of speeds. Except for the method of minimizing residual vibrations, the rest of methods have never been reported. In these methods, the control of vibrations is direct and other quantifiable cam properties can be controlled directly. Also, this paper suggests a universal cam displacement represented by Hermite curves, and describes the use of it in optimal cam design. The used optimization method is the Complex algorithm [9], while the Simplex algorithm [10] is used when there is no constraint.

2 Dynamic Response of the Follower System

2.1 System Model. In this research, a linear, single-DOF lumped parameter model of a high-speed cam-follower system is used. This model is introduced by Kanzaki and Itao [2] and has been adopted in other researches [5,6,11]. This model consists of two springs, one mass and one dashpot. The output mass m models the mass of the follower; the stiffnesses k_s and k_f represent the stiffnesses of the return spring and the follower, respectively; and the damping coefficient c models the viscous friction within the system. The equation of motion of the model is often given in normalized coordinates,

$$\ddot{Y}(\tau) + 2\zeta(2\pi\lambda)\dot{Y}(\tau) + (2\pi\lambda)^2 Y(\tau) = (2\pi\lambda)^2 Y_c(\tau) \tag{1}$$

where the dimensionless displacement of the output mass m is $Y = y/h$, and the dimensionless displacement of the cam is $Y_c = y_c/h_c$. h and h_c are the maximum displacements of the output mass and cam, respectively. The dimensionless output speed is $\dot{Y} = dY(\tau)/d\tau$, and the dimensionless output acceleration is $\ddot{Y} = d^2Y(\tau)/d\tau^2$. The dimensionless time is $\tau = t/t_1$, and t_1 is the rise time of the cam. The damping ratio is $\zeta = c/(2m\omega_0)$, the natural frequency, $\omega_0 = \sqrt{(k_s+k_f)/m}$, the period, $t_0 = 2\pi/\omega_0$, and the speed ratio, $\lambda = t_1/t_0$. Since the dimensionless time is inversely proportional to the speed ratio, a decrease in the speed ratio denotes an increase in cam speed.

In the following section, a method for computing vibrational responses will be presented.

2.2 Residual Vibration and Primary Vibration. During the rise or fall of the output motion, a vibrational response exists because of the dynamics of the system. This will be referred to as the primary vibration. The vibrational response may persist into the dwells and will be called residual vibration.

For calculating vibrational responses, we can substitute the dynamic displacement deviation, $U = Y - Y_c$ into Eq. (1) to get a vibrational response equation

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received March 22, 2005; final manuscript received February 4, 2007. Review conducted by N. Sri Namachchivaya.

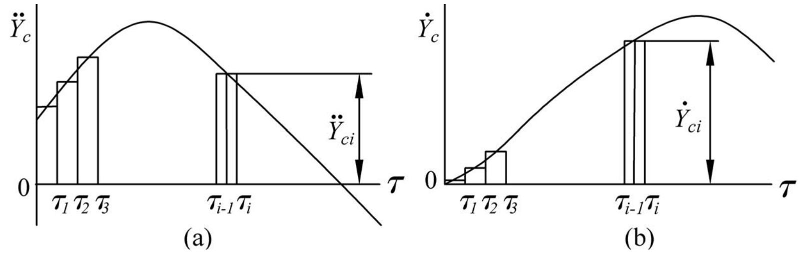


Fig. 1 Approximate substitution of motion curves with step curves (*a*) acceleration curve (*b*) velocity curve

$$\ddot{U}(\tau) + 2\zeta a_1 \dot{U}(\tau) + a_1^2 U(\tau) = -\ddot{Y}_c(\tau) - 2\zeta a_1 \dot{Y}_c(\tau) \tag{2}$$

where $a_1 = 2\pi\lambda$. It is obvious from Eq. (2) that the vibrational response is governed by the cam velocity and acceleration if the parameters for mass, springs and damper are fixed. The solution of Eq. (2) yields the primary vibration at any speed ratio λ.

It is often difficult to solve the dynamic deviation U when the functions $\ddot{Y}_c$ and $\dot{Y}_c$ are complicated. In order to obtain a common method of calculating dynamic responses, which is effective not only for the cam acceleration and velocity represented by simple functions like the polynomial, but also for the cam acceleration and velocity represented by parametric curves like B-spline, we have used a numerical method to solve Eq. (2).

The general solution of the homogeneous equation of Eq. (2) is

$$U_h(\tau) = e^{-\zeta a_1 \tau}[c_1 \cos(a_2\tau) + c_2 \sin(a_2\tau)] \tag{3}$$

where $a_2 = a_1\sqrt{1-\zeta^2}$. In order to obtain the particular solution of Eq. (2), we disperse the functions of the cam acceleration and velocity. As shown in Fig. 1, we use ladder curves of a certain tiny time interval $\Delta\tau_i$ to replace the cam acceleration and velocity curves, respectively. In every time interval, it can be approximately considered that $\ddot{Y}_{ci}$ and $\dot{Y}_{ci}$ are all constant and they are respectively equal to the cam acceleration and velocity of the midpoints of time interval $\Delta\tau_i = \tau_i - \tau_{i-1}$.

Let the particular solution of Eq. (2) be

$$U_p(\tau) = A(\text{const.}) \tag{4}$$

then

$$\dot{U}_p = \ddot{U}_p = 0 \tag{5}$$

Substitute Eqs. (4) and (5) into Eq. (2), then A is given by

$$A = -(\ddot{Y}_c + 2\zeta a_1 \dot{Y}_c)/a_1^2 \tag{6}$$

so that the general solution of Eq. (2), which is the sum of homogeneous and particular solution, is

$$U(\tau) = e^{-\zeta a_1 \tau}[c_1 \cos(a_2\tau) + c_2 \sin(a_2\tau)] + A \tag{7}$$

From Eq. (7), we can obtain

$$\dot{U}(\tau) = e^{-\zeta a_1 \tau}\{c_1[-\zeta a_1 \cos(a_2\tau) - a_2 \sin(a_2\tau)] + c_2[-\zeta a_1 \sin(a_2\tau) + a_2 \cos(a_2\tau)]\} \tag{8}$$

Supposing U_{i-1} and $\dot{U}_{i-1}$ corresponding to time τ_{i-1} are known, since $\ddot{Y}_{ci}$ and $\dot{Y}_{ci}$ begin to excite cam systems at time τ_{i-1}, the exciting time corresponding to τ_{i-1} is $\tau=0$. The boundary conditions of Eqs. (7) and (8) can then be shown to be

$$U(\tau)|_{\tau=0} = U_{i-1}$$

$$\dot{U}(\tau)|_{\tau=0} = \dot{U}_{i-1} \tag{9}$$

Substituting the boundary conditions into Eqs. (7) and (8), c_1 and c_2 are determined, i.e.,

$$c_1 = U_{i-1} + (\ddot{Y}_{ci} + 2\zeta a_1 \dot{Y}_{ci})/a_1^2$$

$$c_2 = (\dot{U}_{i-1} + \zeta a_1 c_1)/a_2 \tag{10}$$

Substitution of c_1 and c_2 into Eqs. (7) and (8) leads to a numerical solution of Eq. (2),

$$U_i = [a_4 \cos(a_2\Delta\tau) + a_5 \sin(a_2\Delta\tau)]a_6 - a_3$$

$$\dot{U}_i = [(a_2 a_5 - a_1 a_4 \zeta)\cos(a_2\Delta\tau) - (a_2 a_4 + a_1 a_5 \zeta)\sin(a_2\Delta\tau)]a_6 \tag{11}$$

where $a_3 = (\ddot{Y}_{ci} + 2a_1\zeta\dot{Y}_{ci})/a_1^2$, $a_4 = a_3 + U_{i-1}$, $a_5 = (\dot{U}_{i-1} + a_1 a_4 \zeta)/a_2$, $a_6 = e^{-a_1\zeta\Delta\tau}$.

When Eq. (11) is used, calculating the responses of motions should begin with the known points of motions. For this reason, the starting point of the first time interval is usually selected at the beginning of the rise. At this point, the initial displacements are $Y_{c0}=0$, $Y_0=0$, the initial velocities $\dot{Y}_{c0}=0$, $\dot{Y}_0=0$, so the initial displacement deviation is $U_0=0$ and its derivative $\dot{U}_0=0$. Substituting U_0 and $\dot{U}_0$ into Eq. (11), the displacement deviation U_1 and its derivative $\dot{U}_1$ at the ending of first time interval can be obtained. Again, taking U_1 and $\dot{U}_1$ as theinitial displacement deviation and its derivative of the second time interval, and using the same method, the displacement deviation U_2 and its derivative $\dot{U}_2$ at the ending of second time interval can be obtained. Repeating the above calculation process can obtain the displacement deviations, $U_1, U_2, U_3, \ldots$ corresponding to each time interval, whereby obtaining the output responses of the whole motion process $0 \le \tau \le 1$.

As to the residual vibration, when $\tau > 1$, $\ddot{Y}_c(\tau)=0$, $\dot{Y}_c(\tau)=0$, so that the right-hand side of Eq. (2) disappears. Therefore, from the homogeneous solution of Eq. (2), the amplitude of the residual vibrations at any speed ratio λ can be found as

$$A_1 = \sqrt{U(1)^2 + \{[\dot{U}(1) + a_1\zeta U(1)]/a_2\}^2} \tag{12}$$

3 Criterion for Residual Vibrations

3.1 Previous Work. To reduce residual vibrations over a range of speeds, some optimality criteria have been proposed. Wiederrich and Roth [4] suggested the control of acceleration; Kwakernaak and Smit [3] suggested the control of acceleration as well as the control of velocity, acceleration, and jerk; Kanzaki and Itao [2] suggested the control of cam speeds of zero residual vibration; Fabien, Longman and Freudenstein [5] suggested the control of jerk and follower spring force; while Chew and Chuang [6] suggested to directly control the residual vibration. Apart from Kanzaki and Itao as well as Chew and Chuang, the others suggested operating velocity, acceleration, jerk, and spring force that are indirectly effecting residual vibration, and the methods suggested by whom are all indirect ones.

Since the method suggested by Kanzaki and Itao cannot make residual vibrations over a range of speeds be minimized, now we will make an analysis on the method suggested by Chew and Chuang in the effectiveness of reducing residual vibrations, to show why we will still propose a new way of restraining residual vibrations.

In order to minimize the amplitude A_1 of residual vibrations over a speed range $[\lambda_1, \lambda_2]$, Chew and Chuang used the following functional:

$$J = \int_{\Omega_1}^{\Omega_2} A_1/\Omega^w d\Omega \tag{13}$$

as the objective function, where $\Omega = 2\pi\lambda$, $\Omega_1 = 2\pi\lambda_1$, $\Omega_2 = 2\pi\lambda_2$; w is a weighting factor. The intention of introducing the weight is to apply a more heavy penalty to the residual vibrations at the higher cam speeds than that at the lower speeds over a range of speeds. Chew and Chuang consider that the residual vibrations at low speeds are generally low and great attention should be paid to the high residual vibrations at the high speeds.

This criterion is very smart. But, since the residual vibrations at each speed over the specified range of speeds are all penalized, the effect of penalizing high residual vibrations is lowered. Therefore, it is difficult that the cams with the least residual vibrations over a specified range of speeds are obtained by this criterion.

In following sections, we will suggest a new objective function. If the word "punishment" is still used to describe the feature of the function vividly, this function can only punish the maximum residual vibration over a range of speeds. To advance this function, a preparatory work yet remains to be done. We will start with it.

3.2 Control Variables. The concept of control variables is introduced by using a polynomial displacement function. The polynomial cam displacement of degree n and the cam velocity and acceleration can be expressed as

$$Y_c(\tau) = \sum_{i=0}^{n} C_i \tau^i \ (0 \le \tau \le 1) \tag{14}$$

$$\dot{Y}_c(\tau) = \sum_{i=1}^{n} i C_i \tau^{i-1} \ (0 \le \tau \le 1) \tag{15}$$

$$\ddot{Y}_c(\tau) = \sum_{i=2}^{n} i(i-1) C_i \tau^{i-2} \ (0 \le \tau \le 1) \tag{16}$$

For the minimal boundary conditions,

$$Y_c(0) = 0 \quad Y_c(1) = 1 \quad \dot{Y}_c(0) = 0 \quad \dot{Y}_c(1) = 0 \tag{17}$$

the polynomial coefficients or the relationship among coefficients can be determined in terms of Eqs. (14) and (15) as the following:

$$C_0 = 0 \quad C_1 = 0 \ C_2 = 1 - \sum_{i=3}^{n} C_i$$

$$C_n = -1/(n-2)\left[2 + \sum_{i=3}^{n-1} (i-2) C_i\right] \tag{18}$$

Known from Eq. (18), among the $n+1$ coefficients of $C_0, C_1, \ldots, C_n$ in addition to $C_0 = C_1 = 0$, there are $n-1$ nonzero coefficients of $C_2, C_3, \ldots, C_n$. It can also be known from the latter two items of Eq. (18) that among the nonzero coefficients, there are only $n-3$ coefficients of $C_3, C_4, \ldots, C_{n-1}$, being the independent variables. When the degree n is specified, the cam displacement and its derivatives are only decided by the $n-3$ coefficients. Therefore, the $n-3$ coefficients are called the control variable.

If the functions of cam displacement and its derivatives are controlled by n control variables $x_i, i=1,2,\ldots,n$, these functions can be expressed as the function of a vector X of the control variable, i.e., $Y_c(X), \dot{Y}_c(X), \ddot{Y}_c(X), X = [x_1 \ x_2 \ \cdots \ x_n]^T$. We will also express other functions as the functions of vector X when necessary.

3.3 Maximum Residual Vibration. It has been known that the vibrational response is only governed by the cam velocity and acceleration, if the parameters for mass, springs, and damper are given. As to the cam velocity and acceleration, they are all the functions of the control variable. It is known from Eq. (2) that the vibration response is related to the speed ratio λ. Accordingly, the amplitude A_1 of residual vibrations can be expressed as

$$A_1 = g[\dot{Y}_c(X), \ddot{Y}_c(X), \lambda] \tag{19}$$

When control variables X are given, the cam velocity and acceleration are determined uniquely. At this time, it can always calculate the amplitudes of residual vibrations for each given value of the speed ratio λ ($\lambda > 0$) by the recursion formula (11) and Eq. (12). Hence, for an arbitrary specified speed range $[\lambda_h, \lambda_l]$, the maximum amplitude of residual vibrations can not only be solved but also be expressed by

$$f_r(X) = \max_{\lambda_h \le \lambda \le \lambda_l} g[\dot{Y}_c(X), \ddot{Y}_c(X), \lambda] \tag{20}$$

If it is the sole purpose to minimize residual vibrations over a range of speeds, Eq. (20) as an objective function can obviously be considered. However, we still hope that while the residual vibration is reduced, other cam properties can also be controlled simultaneously.

4 Optimization Models and Solution Methods

4.1 Minimization of Residual Vibration. Although several methods for reducing residual vibrations over a range of speeds have been proposed, the problem of controlling other cam properties cannot be well solved. Kwakernaak and Smit [3] only took the control over the maximal cam velocity, acceleration and jerk into account. As compared to Polydyne cam design, Fabien, Longman, and Freudenstein [5] improved the maximum contact stress, contact force, and energy loss, while Kanzaki and Itao [2], Wiederrich and Roth [4], and Chew and Chuang [6] did not take the problem into account.

In order to achieve the purpose of controlling various cam properties, we take the control variables of the cam displacement and its derivatives as the design variables, the function shown by Eq. (20) as the objective function, and the requirements of design, manufacturing, etc. for cams as the design constraints, that is, the problem is expressed as

$$f_r(X^*) = \min_X f_r(X) \tag{21}$$

X is subject to the constraints

$$u_i(X) = 0 \quad i = 1,2,\ldots,p \tag{22}$$

$$v_j(X) \le 0 \quad j = 1,2,\ldots,q \tag{23}$$

where X^* represents the optimal solution of the constrained optimization problem.

4.2 Restriction of Residual Vibration. In many cases, the property desired to be as good as possible is not the residual vibration but one of other cam properties. As to the residual vibration, it is only required not to surpass a specified value. Equations (21)–(23) cannot solve this problem, and other literature did not propose any method to solve this problem either.

For this problem, we still take the control variable X of the cam displacement and its derivatives as the design variables. As to the requirements of residual vibrations over a range of speeds no more than a specified value, we let Eq. (20) satisfy $f_r(X) \le [RV]$, and let this equation be one of design constraints, i.e.,

$$\min \text{ or } \max\ F(X);\ X = [x_1\ \ x_2\ \ \cdots\ \ x_n]^T \tag{24}$$

X is subject to the constraints

$$f_r(X) - [RV] \le 0 \tag{25}$$

$$u_i(X) = 0 \quad i = 1,2, \ldots ,p \tag{26}$$

$$v_j(X) \le 0 \quad j = 1,2, \ldots ,q \tag{27}$$

where $[RV]$ is the allowable value of the residual vibration over a specified range of speeds.

When the model is used to design cams, the designer should select one from the concerned cam properties (excluding residual vibration), which is most expected to be as good as possible, to formulate it as the objective function, while the rest of the cam properties including residual vibration are formulated as the design constraints.

Minimization and Restriction of Primary Vibration. There is no literature to raise a method to solve the problems of how to minimize and restrict primary vibration over a range of speeds. Since the aforesaid solutions to the residual vibration are also suitable for the primary vibration, both the problems of controlling primary vibration are also solved incidentally and even the procedure is very similar, i.e., first set up a equation corresponding to Eq. (20) for primary vibrations, and then, set up two constrained optimization models corresponding to Eqs. (21)–(23) and (24)–(27) with the equation. The above-mentioned process will not be described in detail because of the limit of the paper.

4.3 Methods of Solving Models. Since this research is to provide a universal method for dynamic cam design, a numerical method has been used to solve the vibrational response equation, thus making the solution become a recursion formula; again, since the problems of controlling vibrations over a range of speeds are formulated as constrained optimization problems, a direct search method of constrained optimization problems need to be adopted to seek for the solutions. The complex method is adopted in this research. The process of searching optimal solution with this method has been detailed in literature [9]. Because of this, we can no longer describe the solution process. When the models degenerate into an unconstrained optimization problem, the Simplex method [10] can be adopted.

5 Hermite Cam Displacement

5.1 Choice of Curve. Many curves have been used for representing the motion curve of cams and there is a more systematic introduction found in [8]. As viewed from the literature on the optimization of dynamic cam design, however, there are a few curves found in use. One reason is that some methods are only suitable for the case of simple curves such as polynomial to represent the motion curves of cams. Another reason is that the parameters of some curves cannot be determined automatically by optimization program via calculation, and must be specified by the designer in term of his experience. For instance, it can be seen from the literature [7] that the number of control points and the abscissas of all inner knot points of B-spline need to be determined by the designer.

Several dynamic cam design technologies of cams have been introduced previously, which are not subject to the restriction of the complexity of cam displacement functions, whereby creating the prerequisite conditions for adopting various curves to express the cam displacement. Hermite curves have two advantages of flexibility as B-spline and convenience of doing without determining parameters artificially. In the following sections, we will use it to set up a universal cam displacement for the models shown by Eqs. (21)–(23) and (24)–(27).

5.2 Hermite Representation of Cam Displacement. One parametric Hermite curve segment of degree $2k+1$ is determined by position vectors $\boldsymbol{Q}_0, \boldsymbol{Q}_1$, which decide the end points of the curve, and the derivative vectors $\dot{\boldsymbol{Q}}_0, \dot{\boldsymbol{Q}}_1, \ddot{\boldsymbol{Q}}_0, \ddot{\boldsymbol{Q}}_1, \ldots, \boldsymbol{Q}_0^{(k)}, \boldsymbol{Q}_1^{(k)}$, which decide on the curve shapes. Once the position vectors are fixed, the derivative vectors are the parameters controlling the curve shapes. The value k decides on the flexibility of the curve so that, a larger k has better flexibility of the curve, and at the same time the parameters controlling the curve shapes are larger.

Generally the cam displacement consists of the rise, fall, and dwell. Since it is the most convenient that the dwell is represented by a straight-line segment, Hermite representation of the cam displacement is only the rise and fall represented by the curves. In following sections, we will use a 7th-degree $(k=3)$ Hermite curve segment to represent the rise or fall of the cam displacement, considering the flexibility of the curve and the convenience of operating.

5.3 Motion Equations. One 7th-degree Hermite curve can be expressed as

$$\boldsymbol{P}(u) = H\boldsymbol{Q}\ \ (0 \le u \le 1) \tag{28}$$

where

$$H = [H_{0,0}\ \ H_{0,1}\ \ H_{1,0}\ \ H_{1,1}\ \ H_{2,0}\ \ H_{2,1}\ \ H_{3,0}\ \ H_{3,1}]$$

$$\boldsymbol{Q} = [\boldsymbol{Q}_0\ \ \boldsymbol{Q}_1\ \ \dot{\boldsymbol{Q}}_0\ \ \dot{\boldsymbol{Q}}_1\ \ \ddot{\boldsymbol{Q}}_0\ \ \ddot{\boldsymbol{Q}}_1\ \ \dddot{\boldsymbol{Q}}_0\ \ \dddot{\boldsymbol{Q}}_1]^T$$

$$H_{0,0} = 20u^7 - 70u^6 + 84u^5 - 35u^4 + 1$$

$$H_{0,1} = -20u^7 + 70u^6 - 84u^5 + 35u^4$$

$$H_{1,0} = 10u^7 - 36u^6 + 45u^5 - 20u^4 + u$$

$$H_{1,1} = 10u^7 - 34u^6 + 39u^5 - 15u^4$$

$$H_{2,0} = 2u^7 - 15u^6/2 + 10u^5 - 5u^4 + u/2$$

$$H_{2,1} = -2u^7 + 13u^6/2 - 7u^5 - 5u^4/2$$

$$H_{3,0} = u^7/6 - 2u^6/3 + u^5 - 2u^4/3 + u^3/6$$

$$H_{3,1} = u^7/6 - u^6/2 + u^5/2 - u^4/6$$

If $\boldsymbol{Q}_0 = \{s_0, \varphi_0\}$, $\boldsymbol{Q}_1 = \{s_1, \varphi_1\}$, $\dot{\boldsymbol{Q}}_0 = \{s_2, \varphi_2\}$, $\dot{\boldsymbol{Q}}_1 = \{s_3, \varphi_3\}$, $\ddot{\boldsymbol{Q}}_0 = \{s_4, \varphi_4\}$, $\ddot{\boldsymbol{Q}}_1 = \{s_5, \varphi_5\}$, $\dddot{\boldsymbol{Q}}_0 = \{s_6, \varphi_6\}$, $\dddot{\boldsymbol{Q}}_1 = \{s_7, \varphi_7\}$, the parametric equation of the cam displacement can then be expressed as

$$\begin{cases} S(u) = HQ_s \\ \varphi(u) = HQ_\varphi \end{cases} \tag{29}$$

where

$$Q_s = [s_0\ \ s_1\ \ s_2\ \ s_3\ \ s_4\ \ s_5\ \ s_6\ \ s_7]^T$$

$$Q_\varphi = [\varphi_0\ \ \varphi_1\ \ \varphi_2\ \ \varphi_3\ \ \varphi_4\ \ \varphi_5\ \ \varphi_6\ \ \varphi_7]^T$$

Differentiate $S(u)$ with respect to $\varphi(u)$ to get the cam velocity, acceleration, and jerk functions. The results are

$$\begin{cases} dS/d\varphi = \dot{S}/\dot{\varphi} \\ \varphi = HQ_\varphi \end{cases} \tag{30}$$

$$\begin{cases} d^2S/d\varphi^2 = (\dot{\varphi}\ddot{S} - \dot{S}\ddot{\varphi})/\dot{\varphi}^3 \\ \varphi = HQ_\varphi \end{cases} \tag{31}$$

$$\begin{cases} d^3S/d\varphi^3 = [\dot{\varphi}(\dot{\varphi}\dddot{S} - \dot{S}\dddot{\varphi}) - 3\ddot{\varphi}(\dot{\varphi}\ddot{S} - \dot{S}\ddot{\varphi})]/\dot{\varphi}^5 \\ \varphi = HQ_\varphi \end{cases} \tag{32}$$

where

$$\dot{S}=\dot{H}Q_s\ \ddot{S}=\ddot{H}Q_s\ \dddot{S}=\dddot{H}Q_s\ \dot{\varphi}=\dot{H}Q_\varphi\ \ddot{\varphi}=\ddot{H}Q_\varphi\ \dddot{\varphi}=\dddot{H}Q_\varphi$$

$$\dot{H}=[\dot{H}_{0,0}\ \ \dot{H}_{0,1}\ \ \dot{H}_{1,0}\ \ \dot{H}_{1,1}\ \ \dot{H}_{2,0}\ \ \dot{H}_{2,1}\ \ \dot{H}_{3,0}\ \ \dot{H}_{3,1}]$$

$$\ddot{H}=[\ddot{H}_{0,0}\ \ \ddot{H}_{0,1}\ \ \ddot{H}_{1,0}\ \ \ddot{H}_{1,1}\ \ \ddot{H}_{2,0}\ \ \ddot{H}_{2,1}\ \ \ddot{H}_{3,0}\ \ \ddot{H}_{3,1}]$$

$$\dddot{H}=[\dddot{H}_{0,0}\ \ \dddot{H}_{0,1}\ \ \dddot{H}_{1,0}\ \ \dddot{H}_{1,1}\ \ \dddot{H}_{2,0}\ \ \dddot{H}_{2,1}\ \ \dddot{H}_{3,0}\ \ \dddot{H}_{3,1}]$$

When $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ express the starting and ending points of the rise, Eqs. (29)–(32) are the equations of motion of the rise. When $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ express the starting and ending points of the fall, Eqs. (29)–(32) are the equations of motion of the fall. In other words, all the equations are universal to the rise and the fall. Only when the 16 unknowns $s_0, s_1, \ldots, s_7$ and $\varphi_0, \varphi_2, \ldots, \varphi_7$ are given, can Eqs. (29) to (32) be decided. Since we have used one Hermite curve segment to represent the rise and the fall, respectively, 32 unknowns should be decided. For a concrete design problem, the rise angle, fall angle and lift are always known, i.e., the position vectors at the endpoints are always known, whereby there are only 24 unknowns unspecified.

5.4 Continuity Conditions and Control Variables. To design cams with better properties, it is necessary to let Eq. (29) satisfy some boundary conditions. A 7th-degree Hermite cam displacement can achieve the continuity of velocity, acceleration, and jerk at the boundary of the rise or fall. Therefore, this section will discuss the conditions corresponding to the continuity of three kinds. Owing to the deduction process of the conditions of the rise and fall being similar, we will only deduce the conditions of the rise here.

At the start of the rise, $u=0$, and then substitute it into Eqs. (29)–(32), thus obtaining

$$S(0)=s_0 \tag{33}$$

$$\varphi(0)=\varphi_0 \tag{34}$$

$$dS(0)/d\varphi(0)=s_2/\varphi_2 \tag{35}$$

$$d^2S(0)/d\varphi^2(0)=[\varphi_2 s_4 - s_2\varphi_4]/\varphi_2^3 \tag{36}$$

$$d^3S(0)/d\varphi^3(0)=[\varphi_2(\varphi_2 s_6-\varphi_6 s_2)-3\varphi_1(\varphi_2 s_4-\varphi_4 s_2)]/\varphi_2^5 \tag{37}$$

Let Eqs. (33)–(37) equal to zero respectively, whereby obtaining

$$s_0=0\ \ \varphi_0=0\ \ s_2=0\ \ \varphi_2\neq 0\ \ \varphi_2 s_6-\varphi_6 s_2=0\ \ \varphi_2 s_4 - s_2\varphi_4=0 \tag{38}$$

From Eq. (38), the continuity conditions of jerk at the beginning of the rise can be obtained,

$$s_0=\varphi_0=s_2=s_4=s_6=0 \quad \varphi_2\neq 0 \tag{39}$$

If the acceleration continuity at the beginning of the rise is expected, there is $d^3S(0)/d\varphi^3(0)\neq 0$, from which we can obtain $s_6\neq 0$, whereby the continuity conditions of acceleration at the beginning of the rise is

$$s_0=\varphi_0=s_2=s_4=0\ \ \varphi_2\neq 0\ \ s_6\neq 0 \tag{40}$$

If at the beginning of the rise the velocity continuity is expected, there are $d^3S(0)/d\varphi^3(0)\neq 0$ and $d^2S(0)/d\varphi^2(0)\neq 0$, from which we can obtain $s_4\neq 0$ and $s_6\neq 0$, whereby the continuity conditions of velocity at the beginning of the rise is

$$s_0=\varphi_0=s_2=0 \quad \varphi_2\neq 0 \quad s_4\neq 0 \quad s_6\neq 0 \tag{41}$$

As for the continuity conditions at the end of the rise, attention is paid to $u=1$ and to the coordinate of the rise ending point (h,Φ) (Φ is the rise angle), so that it is easy to obtain the continuity conditions of jerk, acceleration, and velocity in the following:

$$s_1=h \quad s_3=s_5=s_7=0 \quad \varphi_1=\Phi \quad \varphi_3\neq 0 \tag{42}$$

$$s_1=h \quad s_3=s_5=0 \quad \varphi_1=\Phi \quad \varphi_3\neq 0 \quad s_7\neq 0 \tag{43}$$

$$s_1=h \quad s_3=0 \quad \varphi_1=\Phi \quad \varphi_3\neq 0 \quad s_5\neq 0 \quad s_7\neq 0 \tag{44}$$

It can be seen from the comparison of Eq. (39)–(44) that the differences of three continuity conditions lie in whether s_4, s_6, s_5, and s_7 equal zero or not. We suggest that the velocity continuity conditions be used in the case of having no special causes, that is, take s_4, s_5, s_6, s_7 as the design variables, and let the optimization program decide whether they need zero or not. When the velocity continuity conditions are used, it is known from Eqs. (41) and (44) that the ordinate values of derivative vectors $\dot{\boldsymbol{Q}}_0$ and $\dot{\boldsymbol{Q}}_1$ must be specified as zero, i.e., $s_2=0$ and $s_3=0$, so that there are 10 unknowns of the displacement function of the rise part, and then it can be inferred that the number of the unknowns of the fall part are also 10. Therefore, the total of the unknowns is 20. They are $\varphi_2, \varphi_3, \ldots, \varphi_7$ and s_4, s_5, s_6, s_7. These 20 unknowns are just the control variables of the 7th-degree Hermite cam displacement.

Table 1 Cam profile coefficients

No.	Speed range	Coefficients in ascending order
1	$1\le\lambda\le 2$	0.0, 0.0, 157.14, −2404.18, 14722.08, −46518.73, 82826.13, −83945.26, 45210.74, −10046.92[a]
2	$1\le\lambda\le 2$	0.0, 0.0, 155.94, −2375.63, 14518.55, −45838.44, 81599.52, −82715.88, 44565.36, −9908.42
3	$1.0499\le\lambda\le 1.9436$	0.0, 0.0, 155.95, −2384.63, 14601.14, −46138.82, 82157.45, −83276.62, 44855.61, −9969.08
4	$1\le\lambda\le 10$	0.0, 0.0, 10.77, −58.17, 169.99, −286.35, 309.91, −242.42, 133.19, −35.92.

[a]Reference [6].

6 Examples

Now we present two examples to show the effect of new methods in reducing residual vibrations and the capacity of satisfying various design requirements, as well as the effectiveness of complicated cam displacement functions. The first one compares the results produced by the methods suggested in this paper and literature [6]. The second one simulates a cam design with many design requirements to be satisfied.

6.1 Comparing Results Produced by the Existing Method. In order to make the results produced by the method suggested in this paper be comparable to the results produced by the method advanced by literature [6], we have adopted the following measures: to begin with, since Chew and Chuang did not take into account the problem of controlling other cam properties so that we have to adopt the degraded form of the model shown by Eqs. (21)–(23), i.e., to eliminate the design constraints (22) and (23); secondly, Chew and Chuang did not take into account the viscous damping of the system. Therefore, we specify parameter $\zeta=0$; finally, in corresponding to Chew and Chuang' doings, we also adopt the polynomial cam displacement, i.e., using Eq. (14) and specifying degree $n=9$ and boundary condition (17).

First, we see a case of minimizing residual vibrations over a narrower speed range $1\le\lambda\le 2$. Over this range of speeds, two optimized cam displacement functions are obtained by the methods suggested in this paper and the literature [6]. Their coefficients are listed in Table 1 (see Nos. 1 and 2). It can be seen from Table 1 that the different coefficients are obtained due to a different objective function for the same displacement functions and the same range of speeds. Figure 2(*a*) illustrates the differences in the residual vibration spectrum for the two different sets of coefficients. This figure indicates that the maximum residual vibration of the displacement function [6] is located at $\lambda=2$ of the low speed end. In comparison, the residual vibration of new displacement function over that range of speeds is reduced by about 40%.

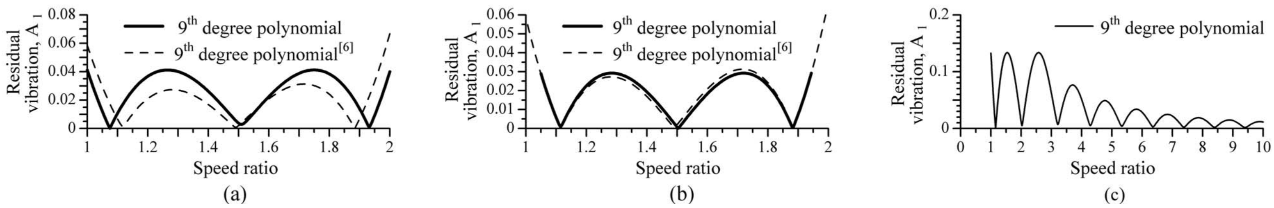


Fig. 2 Comparison of residual vibration characteristics (*a*) λ=1–2; (*b*) λ=1.0499–1.9436; and (*c*) λ=1–10

It still indicates that the objective function (13) cannot restrain the residual vibrations well enough at both the high and the low speed end.

Of course, we have noted that the cam displacement function [6] has lower residual vibrations over a smaller range of speeds. For this matter, we respecify a smaller speed range $1.0499 \le \lambda \le 1.9436$ and over this speed range we obtain a new optimized cam displacement function. Its coefficients are listed in Table 1 (see No. 3). Figure 2(*b*) indicates that the residual vibration of the new optimized function over this smaller range of speeds is still reduced by about 6.6%.

Now, we see a case of a wider speed range $1 \le \lambda \le 10$. Over this range of speeds, an optimized cam displacement function is obtained by the method suggested in this paper. Its coefficients are also listed in Table 1 (see No. 4). Figure 2(*c*) shows the residual vibration of this optimized function. Although Figs. 3(*b*), 4(*b*), and 6(*b*) in the literature [6] show the spectra of several optimized functions over this range of speeds, the coefficients of these optimized functions were not given. For this reason, we can only make a visual comparison. After the comparison, it can be known that the maximum residual vibration of this new optimized function over the range of speeds is about 0.13, obviously being smaller than the residual vibration of all above-mentioned optimized functions in the literature [6] at $\lambda=1$ of the high speed end.

We have known via the above comparison that the optimal criteria of both the methods are the residual vibration, but the ways of operating the residual vibration are different, leading to the different results. As viewed from the effects, operating the maximum residual vibration over a range of speed is the better way obviously.

In addition, it can be known from the above examples that when the speed range is changed from $1 \le \lambda \le 2$ into $1.0499 \le \lambda \le 1.9436$, the speed range is only reduced by 10.63%, but the residual vibration is reduced about 30%. This case just illustrates that serious attention should be paid to the speed range in cam design.

6.2 Cam Design for Intake Valve-Gear. The design parameters for a disk cam with a translating roller follower in the intake valve-gear of an internal combustion engine are shown in Table 2. The design indexes of the primary vibration, residual vibration, pressure angle, etc. are listed in Table 3.

Since the design includes the requirements of restricting residual vibrations and primary vibrations, any existing methods cannot be used to design this cam. However, the model shown by Eqs. (24)–(27) is just applicable to this case. According to the design requirements, this design problem can be stated as follows:

$$\min f(X) = -\int_0^{\Phi_1+\Phi_2} (\text{follower displacement})d\varphi \qquad (45)$$

subject to

$$f_r(X) - [\mathrm{RV}] \le 0 \qquad (46)$$

$$f_p(X) - [\mathrm{PV}] \le 0 \qquad (47)$$

$$\alpha_{\max}(X) - [\alpha] \le 0 \qquad (48)$$

$$\alpha'_{\max}(X) - [\alpha'] \le 0 \qquad (49)$$

$$NR_{\max}(X) + [\mathrm{NR}] \le 0 \qquad (50)$$

$$-F_{\min}(X) \le 0 \qquad (51)$$

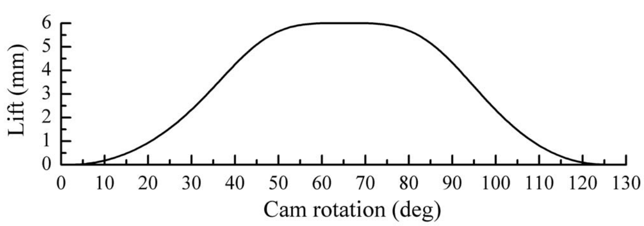


Fig. 3 Follower displacement curve

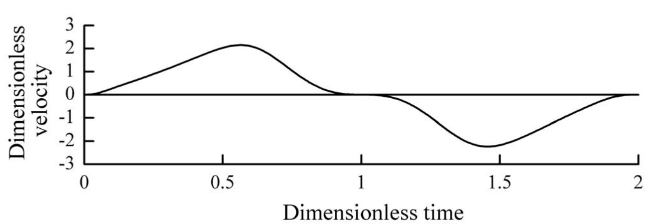


Fig. 4 Follower velocity curve

Table 2 Design parameters

Prime circle radius	22 mm
Roller radius	10 mm
Rocker arm ratio	1.0
Preload force of valve spring	330 N
Equivalent spring constant of valve spring	60 N/mm
Equivalent spring constant of follower	3500 N/mm
Equivalent mass of follower at output	0.17 kg
Speed range of cam rotation	1600–2400 rpm
Contact width between cam and roller	10 mm
Follower offset	0.0
Rise of cam	6 mm
Rise angle and return angle	65 deg
Elastic modulus	200×10^3 N/mm^2
Poisons ratio	0.298

Table 3 Permissible values

[PV]	[RV]	$[\alpha]$	$[\alpha']$	[NR]	$[\sigma_H]$
0.2 mm	0.06 mm	35 deg	70 deg	−250 mm	800 N/mm^2

Table 4 Optimal solution (φ_i (i=1–7), deg; s_i, (i=4–7), mm)

Variables	Rise	Return
φ_2	46.494	84.876
φ_3	78.160	72.199
φ_4	−258.626	−346.684
φ_5	286.366	427.331
φ_6	8805.370	4546.238
φ_7	4392.870	8199.268
s_4	0.418	−0.766
s_5	−0.465	0.004
s_6	2.016	−30.973
s_7	27.381	−3.998

$$\sigma_{H\max}(X) - [\sigma_H] \leq 0 \tag{52}$$

where Φ_1 and Φ_2 are the rise and fall angles, respectively, φ is the camshaft angle, f_r and f_p are the maximum residual vibration and maximum primary vibration over the specified speed range, respectively, $\alpha_{\max}$ and $\alpha'_{\max}$ are the maximum pressure angles of the rise and fall of the cam profile, respectively, $NR_{\max}$ is the maximum negative radius of the curvature of the cam profile, $F_{\min}$ and $\sigma_{H\max}$ are the minimum contact force and maximum contact stress between the follower and cam at the maximum cam speed, respectively.

In this model, the objective function (45) is the integral of the displacement, which is to make the flow area become maximization. Equations (46) and (47) are to make the motion of the valve head smooth. Equation (48) is to ensure the cam to have a better driving efficiency. Equation (49) is to avoid self-lock of the cam mechanism. Equation (50) is to make the cam be ground by a grinding wheel with a radius no bigger than $|[NR]|$. Equation (51) is to make the follower not be separated from the cam. Equation (52) restricts the maximum contact stress between the cam and follower. Except Eqs. (46) and (47), other equations in this model can be found in the literature [7].

The cam displacement specified for this design is the 7th-degree Hermite function with the rise and fall represented by one Hermite curve segment, respectively. Also, the velocity continuity is specified for the starting point of the rise, the ending point of the fall and the connection point of the rise and fall.

Design results are presented in Table 4 and in Figs. 3–11. Table 4 shows the optimal solution. Figures 3–11 show the cam displacement, velocity, acceleration, primary vibration, residual vibration, pressure angle, radius of curvature, contact force, and contact stress.

The following observations can be made from the design results. To begin with, it can be seen from Figs. 6–11 that the cam designed satisfies all the design requirements, and also from Figs. 6–11 that among Eqs. (46)–(52), only Eqs. (46) and (51) in this design are playing a real constraint or known as the tight constraints.

It can be seen from Table 4 that the values of s_4 and s_5 of the rise and fall do not equal zero, indicating that there exists the finite accelerations at the start of the rise, the end of the fall and the interboundary of the rise and fall, for the values of s_4 and s_5 are so small that we cannot see the existence of accelerations at

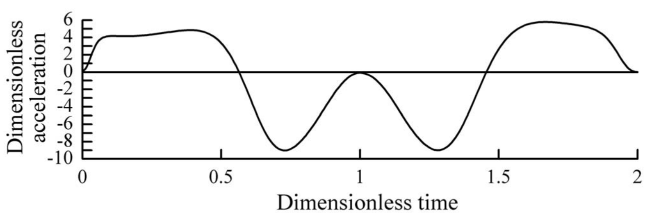


Fig. 5 Follower acceleration curve

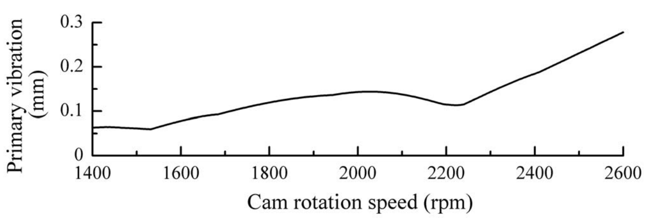


Fig. 6 Primary vibration characteristics

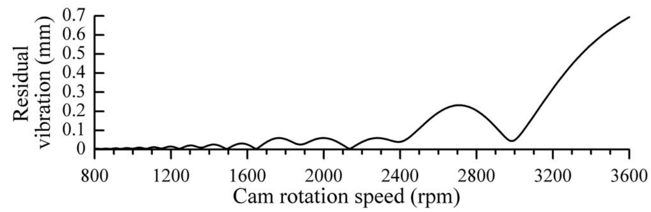


Fig. 7 Residual vibration characteristics

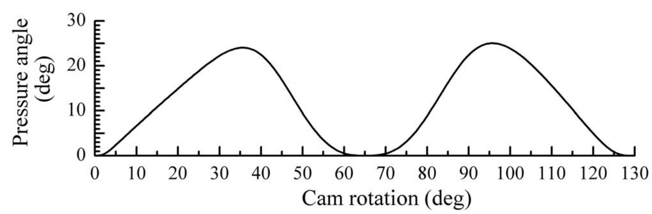


Fig. 8 Pressure angle of cam profile

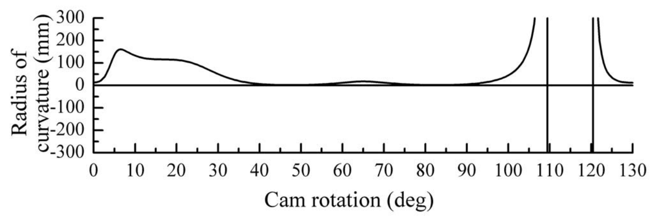


Fig. 9 Radius of curvature of cam profile

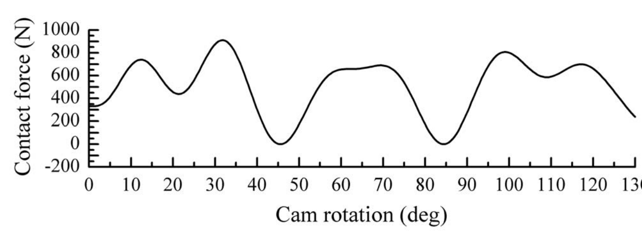


Fig. 10 Cam-follower contact force

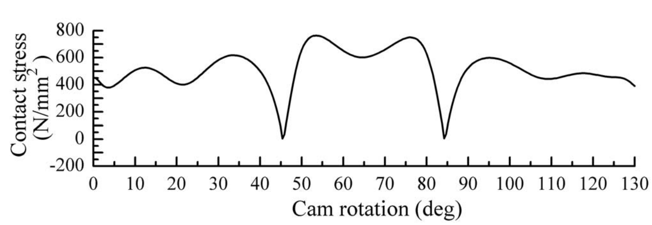


Fig. 11 Cam-follower contact stress

the three places in Fig. 5.

Figure 7 shows that the heights of three waviness within the specified range of speeds are equal to or nearly equal to the allowed values of the residual vibration, which illustrates that the residual vibration is restricted below the specified values effectively.

Figure 10 shows that the minimum contact force between the follower and cam is always larger than or equal to zero, whereby indicating the follower cannot separate from the cam. The follower and cam is in the critical state of contact at the camshaft angles of 45 deg and 85 deg.

Finally, it can be seen from Fig. 11 that the maximum contact stresses between the follower and cam, which are nearly equal, appear at the camshaft angle of 53 deg on the rise, and 77 deg on the fall. This means that the deactivation of the profile surface will appear earliest at these two places.

This example shows that the two functions of restricting residual and primary vibrations can be combined with the functions of restricting other cam properties; the technologies of controlling vibrations suggested in this paper are effective for complicated cam displacement functions; Hermite cam displacement adopted to the design cam is also convenient.

7 Conclusions

This investigation has presented a method for minimizing or restricting residual and primary vibrations over a range of speeds, respectively, and has shown that operating the maximum residual vibration over a range of speeds is more effective for reducing residual vibrations. All the methods can control various quantifiable cam properties, and can be combined into one optimization model to meet such different needs as both minimizing residual vibrations and restricting primary vibrations, both restricting residual vibrations and restricting primary vibrations, etc. With such features and the universal Hermite cam displacement, this investigation has developed an environment for the design of high-speed cams. In this environment, not only can any one of quantifiable cam properties be formulated as an objective function and others be formulated as design constraints, but also the designer does need not to make great efforts on choosing motion curves, and the design result can be generated computationally and optimally by the cam optimization program.

References

[1] Stoddart, D. A., 1953, "Polydyne Cam Design," Mach. Des., **25**, pp. 121–135; pp. 146–154; pp. 149–164.

[2] Kanzaki, K., and Itao, K., 1972, "Polydyne Cam Mechanisms for Typehead Positioning," ASME J. Eng. Ind., **108**, pp. 250–254.

[3] Kwakernaak, H., and Smit, J., 1968, "Minimum Vibration Cam Profiles," J. Mech. Eng. Sci., **10**(3), pp. 219–227.

[4] Wiederrich, J. L., and Roth, R., 1975, "Dynamic Synthesis of Cams Using Finite Trigonometric Series," ASME J. Eng. Ind., **97**, pp. 287–293.

[5] Fabien, B. C., Longman, R. W., and Freudenstein, F., 1994, "The Design of High-Speed Dwell-Rise-Dwell Cams Using Linear Quadratic Optimal Control Theory," ASME J. Mech. Des., **116**, pp. 867–874.

[6] Chew, M., and Chuang, C. H., 1995, "Minimizing Residual Vibrations in High-Speed Cam-Follower Systems Over a Range of Speeds," ASME J. Mech. Des., **117**, pp. 166–172.

[7] Sandgren, E., and West, R. L., 1989, "Shape Optimization of Cam Profiles Using a B-Spline Representation," ASME J. Mech., Transm., Autom. Des., **111**, pp. 195–201.

[8] Norton, R. L., 2002, *Cam Design and Manufacturing Handbook*, Industrial Press, New York.

[9] Box, M. J., 1965, "A New Method of Constrained Optimization and a Comparison with Other Methods," Comput. J., **8**, pp. 42–52.

[10] Spendley, W., Hext, G. R., and Himsworth, F. R., 1962, "Sequential Application of Simplex Designs in Optimization and Evolutionary Operation," Technometrics, **4**, pp. 441–461.

[11] Srinivasan, L. N., and Ge, Q. Jeffrey, 1998, "Designing Dynamically Compensated and Robust Cam Profiles with Bernstein-Bézier Harmonic Curves," ASME J. Mech. Des., **120**, pp. 40–45.

On the Analytic Solution of Magnetohydrodynamic Flow of a Second Grade Fluid Over a Shrinking Sheet

T. Hayat

Z. Abbas[1]
e-mail: za-qau@yahoo.com

Department of Mathematics,
Quaid-I-Azam University 45320,
Islamabad 44000, Pakistan

M. Sajid
Theoretical Plasma Physics Division,
PINSTECH,
P.O. Nilore,
Islamabad 44000, Pakistan

In this study, we derive an analytical solution describing the magnetohydrodynamic boundary layer flow of a second grade fluid over a shrinking sheet. Both exact and series solutions have been determined. For the series solution, the governing nonlinear problem is solved using the homotopy analysis method. The convergence of the obtained solution is analyzed explicitly. Graphical results have been presented and discussed for the pertinent parameters. [DOI: 10.1115/1.2723820]

Keywords: second grade fluid, MHD fluid flow, shrinking sheet, analytic solution

1 Introduction

Because of their use in different technological processes, flows of non-Newtonian fluids have attracted considerable attention from the researchers. Due to the complexity of fluids, several models of non-Newtonian fluids have been proposed. The departure of viscoelastic behavior of non-Newtonian fluids from the Navier–Stokes equations manifests itself in a variety of ways: non-Newtonian viscosity (shear thinning or shear thickening), stress relaxation, nonlinear creeping, development of normal stress differences, and yield stress. Although the governing equations for viscoelastic fluids are in general much more complicated, nonlinear, and higher order than the Navier–Stokes equations, even then several investigators [1–10] have recently engaged in studying the flows of such fluids.

Boundary layer theory has been successfully applied to non-Newtonian fluids of various models. It is well known that boundary layer on continuous surface is an important type of flow occurring in a number of technical problems. Examples may be found in continuous casting, glass fiber production, metal extrusion, hot rolling, textiles, and wire drawing. This type of flow has been considered first by Sakiadis [11,12]. Following his works an increasing number of papers analyzing various aspects have been published. Some recent attempts on boundary layer flows of Newtonian and non-Newtonian fluids past a stretching surface have been made by several investigators (see Refs. [13–22] and references therein). But literature on the shrinking flow problem is very scarce. According to our information, only two such attempts have been made by Wang [23] and Miklavcic and Wang [24]. In Ref. [23], he studied unsteady shrinking film solution, whereas in Ref. [24], they proved the existence and uniqueness for steady viscous flow caused by a shrinking sheet for a specific value of the suction parameter.

So far no attempt has been made to study the boundary flow of a non-Newtonian fluid over a shrinking sheet. Therefore, the focus of the present endeavor is threefold. First we consider the second grade fluid (a simplest subclass of viscoelastic fluid). Second, the magnetohydrodynamic (MHD) influence on the shrinking flow is determined. Third, analytic solution by the homotopy analysis method (HAM) [25,26] is given. HAM has already been used successfully for several nonlinear problems [27–42].

The presentation proceeds as follows. In Sec. 2 the boundary layer problem for MHD shrinking flow is formulated. Analytic solutions are provided in Sec. 3. In Sec. 4, the convergence of the HAM solution is presented. This section also includes the comparison between exact and HAM solutions. Section 5 consists of results and discussion. The concluding remarks are given in Sec. 6.

2 Mathematical Formulation

Let us consider an incompressible second grade fluid past a horizontal shrinking sheet at $y=0$. The x and y axes are taken along and perpendicular to the sheet, respectively, as shown in Fig. 1. The flow is confined to $y>0$. A constant magnetic field of strength $\mathbf{B}_0$ acts in the direction of the y axis. The induced magnetic field is negligible, which is a valid assumption on a laboratory scale. This assumption is justified when the magnetic Reynolds number is small [43–49]. Since no external electric field is applied and the effect of polarization of the ionized fluid is negligible, we can assume that the electric field $\mathbf{E}=\mathbf{0}$. The boundary layer equations governing the MHD flow are

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

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=\nu\frac{\partial^2 u}{\partial y^2}+\frac{\alpha_1}{\rho}\left(u\frac{\partial^3 u}{\partial x\,\partial y^2}+\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial y^2}+\frac{\partial u}{\partial y}\frac{\partial^2 v}{\partial y^2}+v\frac{\partial^3 u}{\partial y^3}\right)-\frac{\sigma B_0^2}{\rho}u \tag{2}$$

where ν is the kinematic viscosity; σ is the electrical conductivity; ρ is the fluid density; α_1 is the second grade/viscoelastic parameter; and u and v are x and y components of velocity.

The boundary conditions of the system are

$$u=-ax, \quad v=-V \quad \text{at } y=0 \tag{3}$$

$$u\to 0, \quad \frac{\partial u}{\partial y}\to 0 \quad \text{as } y\to\infty \tag{4}$$

where $a>0$ is the shrinking constant; and V (>0) is the suction velocity. In order to solve the problem completely in unbounded domains, it is possible to augment the boundary conditions by assuming certain asymptotic structures for the solutions at infinity. Here, the second condition in Eq. (4) is the augmented condition discussed by Grag and Rajagopal [50]. Later Vajravelu and Roper

[1]Corresponding author.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received November 12, 2006; final manuscript received January 30, 2007. Review conducted by Nesreen Ghaddar.

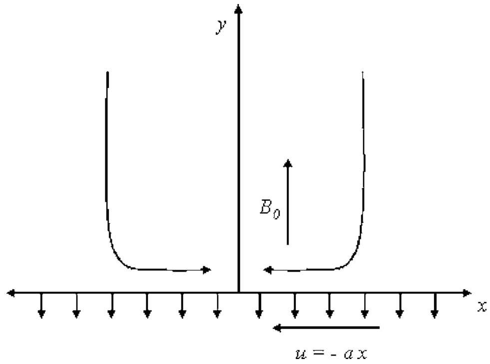


Fig. 1 Flow model for the problem

[51], Cortell [15], and others have used this condition for flow problems over a stretching sheet.

The formulation of the boundary value problem is now completed. In order to solve this problem, it is convenient to nondimensionalize the governing equations and conditions. This can be accomplished by using the following transformations

$$u = axf'(\eta), \quad v = -\sqrt{a\nu}f(\eta), \quad \eta = \sqrt{\frac{a}{\nu}}y \tag{5}$$

The Eq. (1) is now identically satisfied and after performing the mathematical operations, the resulting dimensionless problem can be written as

$$f''' - M^2 f' - f'^2 + ff'' + K(2f'f''' - f''^2 - ff^{iv}) = 0 \tag{6}$$

$$f = S, \quad f' = -1 \quad \text{at } \eta = 0$$

$$f' \to 0, \quad f'' \to 0 \quad \text{as } \eta \to \infty \tag{7}$$

where $S = V/\sqrt{a\nu}$, $M^2 = \sigma B_0^2/\rho a$, and $K = a\alpha_1/\nu\rho \neq 0$ and a prime indicates differentiation with respect to η.

The exact solution of Eqs. (6) and (7) is of the form

$$f(\eta) = S - \frac{1}{\beta}(1 - e^{-\beta\eta}) \tag{8}$$

Substituting Eq. (8) into Eq. (6) we get the following cubic equation in β as

$$\beta^3 KS - (K-1)\beta^2 - S\beta - M^2 + 1 = 0 \tag{9}$$

which has one real and two complex roots. The real root is given by

$$\beta = \frac{1}{6KS}\left(\frac{2(K-1) + 2^{4/3}((1-K)^2 + 3KS^2)}{A} + 2^{2/3}A\right)$$

in which

$$A = \left[\begin{array}{c} -2(1-K)^3 - 9(1-K)KS^2 - 27K^2(1-M^2)S^2 \\ +\sqrt{-4((1-K)^2 + 3KS^2)^3 + (2(1-K)^3 + 9(1-K)KS^2 + 27K^2(1-M^2)S^2)^2} \end{array}\right]^{1/3}$$

and the solution Eq. (8) is valid for all nonzero values of K. We have just shown the values of f in Table 3 for K up to 20 in order to compare it with HAM. However, we can select $K > 20$. Also, the existing solution Eq. (8) is unique.

The shear stress $\tilde{\tau}_w$ at the surface is defined as

$$\tilde{\tau}_w\Big|_{y=0} = \left[\mu\frac{\partial u}{\partial y} + \alpha_1\left(u\frac{\partial^2 u}{\partial x\,\partial y} + v\frac{\partial^2 u}{\partial y^2} + 2\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}\right)\right]\Bigg|_{y=0} \tag{10}$$

The above equation in dimensionless form becomes

$$\tau_w = \frac{\tilde{\tau}_w}{a^{3/2}x\sqrt{\mu\rho}} = [f'' + K(3f'f'' - ff''')]|_{\eta=0} \tag{11}$$

The problem consisting of Eqs. (6) and (7) can also be solved analytically by using HAM in the next section.

3 Analytic Solution by HAM

For the HAM solution, the initial approximation f_0 of f and the auxiliary linear operator $\mathcal{L}$ can be written as

$$f_0(\eta) = S - 1 + e^{-\eta} \tag{12}$$

$$\mathcal{L}(f) = f''' - f' \tag{13}$$

where

$$\mathcal{L}[C_1\eta + C_2 e^{\eta} + C_3 e^{-\eta}] = 0 \tag{14}$$

and C_i $(i=1-3)$ are arbitrary constants. Let $\hbar$ be an auxiliary nonzero parameter and $p \in [0,1]$ is the embedding parameter then we get the following problems:

zeroth-order deformation problem

$$(1-p)\mathcal{L}[\hat{f}(\eta,p) - f_0(\eta)] = p\hbar\mathcal{N}[\hat{f}(\eta,p)] \tag{15}$$

$$\hat{f}(0,p) = S, \quad \hat{f}'(0,p) = -1, \quad \hat{f}'(\infty,p) = 0 \tag{16}$$

$$\begin{aligned}\mathcal{N}[\hat{f}(\eta,p)] = {} & \frac{\partial^3\hat{f}(\eta,p)}{\partial\eta^3} - M^2\frac{\partial\hat{f}(\eta,p)}{\partial\eta} - \left(\frac{\partial\hat{f}(\eta,p)}{\partial\eta}\right)^2 \\ & + \hat{f}(\eta,p)\frac{\partial^2\hat{f}(\eta,p)}{\partial\eta^2} + \alpha\left\{2\frac{\partial\hat{f}(\eta;p)}{\partial\eta}\frac{\partial^3\hat{f}(\eta;p)}{\partial\eta^3}\right. \\ & \left. - \left(\frac{\partial^2\hat{f}(\eta;p)}{\partial\eta^2}\right)^2 - \hat{f}(\eta;p)\frac{\partial^4\hat{f}(\eta;p)}{\partial\eta^4}\right\}\end{aligned} \tag{17}$$

***m*th-order deformation problem**

$$\mathcal{L}[f_m(\eta) - \chi_m f_{m-1}(\eta)] = \hbar\mathcal{R}_m(\eta) \tag{18}$$

$$f_m(0) = f'_m(0) = f'_m(\infty) = 0 \tag{19}$$

$$\begin{aligned}\mathcal{R}_m(\eta) = {} & f'''_{m-1}(\eta) - M^2 f'_{m-1}(\eta) \\ & + \sum_{k=0}^{m-1}\left[\begin{array}{c} f_{m-1-k}f''_k - f'_{m-1-k}f'_k \\ +\alpha(2f'_{m-1-k}f''_k - f''_{m-1-k}f''_k - f_{m-1-k}f^{iv}_k) \end{array}\right]\end{aligned} \tag{20}$$

$$\chi_m = \begin{cases} 0, & m \leqslant 1 \\ 1, & m > 1 \end{cases} \tag{21}$$

The mth-order deformation problem up to first few order of approximations can be solved through symbolic software MATHEMATICA. The solution can be written as

$$f_m(\eta) = \sum_{n=0}^{m+1} \sum_{q=0}^{m+1-n} a_{m,n}^q \eta^q e^{-n\eta}, \quad m \geqslant 0 \tag{22}$$

where

$$a_{m,0}^0 = \chi_m \chi_{m+2} a_{m-1,0}^0 - \sum_{q=0}^{m} \Psi_{m,1}^q \mu_{1,1}^q - \sum_{n=2}^{m+1} \left[\begin{array}{c} (n-1)\Psi_{m,n}^0 \mu_{n,0}^0 \\ + \sum_{q=1}^{m+1-n} \Psi_{m,n}^q ((n-1)\mu_{n,0}^q - \mu_{n,1}^q) \end{array} \right] \tag{23}$$

$$a_{m,0}^k = \chi_m \chi_{m+1-k} a_{m-1,0}^k, \quad 1 \leqslant k \leqslant m+1 \tag{24}$$

$$a_{m,1}^0 = \chi_m \chi_{m+1} a_{m-1,1}^0 + \sum_{q=0}^{m} \Psi_{m,1}^q \mu_{1,1}^q + \sum_{n=2}^{m+1} \left\{ n\Psi_{m,n}^0 \mu_{n,0}^0 + \sum_{q=1}^{m+1-n} \Psi_{m,n}^q (n\mu_{n,0}^q - \mu_{n,1}^q) \right\} \tag{25}$$

$$a_{m,1}^k = \chi_m \chi_{m-k+1} a_{m-1,1}^k + \sum_{q=k-1}^{m} \Psi_{m,1}^q \mu_{1,k}^q, \quad 1 \leqslant k \leqslant m+1 \tag{26}$$

$$a_{m,n}^k = \chi_m \chi_{m+2-n-k} a_{m-1,n}^k + \sum_{q=k}^{m+1-n} \Psi_{m,n}^q \mu_{n,k}^q, \quad 2 \leqslant n \leqslant m+1, \quad 0 \leqslant k \leqslant m+1-n \tag{27}$$

$$\mu_{1,k}^q = \sum_{p=0}^{q+1-k} \frac{q!}{k!\,2^{q+1-k-p}}, \quad q \geqslant 0, \quad 1 \leqslant k \leqslant q+1 \tag{28}$$

$$\mu_{n,k}^q = \sum_{r=0}^{q-k} \sum_{p=0}^{q-k-r} \frac{q!}{k!\,(n-1)^{q+1-k-r-p} n^{r+1} (n+1)^{p+1}}, \quad 0 \leqslant k \leqslant q+1-n, \quad q \geqslant 0, \quad n \geqslant 2 \tag{29}$$

$$\Psi_{m,n}^q = \hbar \left[\begin{array}{c} \chi_{m+2-n-q} d_{m-1,n}^q - M^2 b_{m-1,n}^q + \beta_{m,n}^q - \gamma_{m,n}^q \\ + K(2\delta_{m,n}^q - \Delta_{m,n}^q - \Lambda_{m,n}^q) \end{array} \right] \tag{30}$$

$$\beta_{m,n}^q = \sum_{k=0}^{m-1} \sum_{i=\max\{0,n-m+k\}}^{\min\{n,k+1\}} \sum_{j=\max\{0,q-m+k+n-i\}}^{\min\{q,k+1-i\}} c_{k,i}^j a_{m-1-k,n-i}^{q-j}$$

$$\gamma_{m,n}^q = \sum_{k=0}^{m-1} \sum_{i=\max\{0,n-m+k\}}^{\min\{n,k+1\}} \sum_{j=\max\{0,q-m+k+n-i\}}^{\min\{q,k+1-i\}} b_{k,i}^j b_{m-1-k,n-i}^{q-j}$$

$$\delta_{m,n}^q = \sum_{k=0}^{m-1} \sum_{i=\max\{0,n-m+k\}}^{\min\{n,k+1\}} \sum_{j=\max\{0,q-m+k+n-i\}}^{\min\{q,k+1-i\}} d_{k,i}^j b_{m-1-k,n-i}^{q-j}$$

$$\Delta_{m,n}^q = \sum_{k=0}^{m-1} \sum_{i=\max\{0,n-m+k\}}^{\min\{n,k+1\}} \sum_{j=\max\{0,q-m+k+n-i\}}^{\min\{q,k+1-i\}} c_{k,i}^j c_{m-1-k,n-i}^{q-j}$$

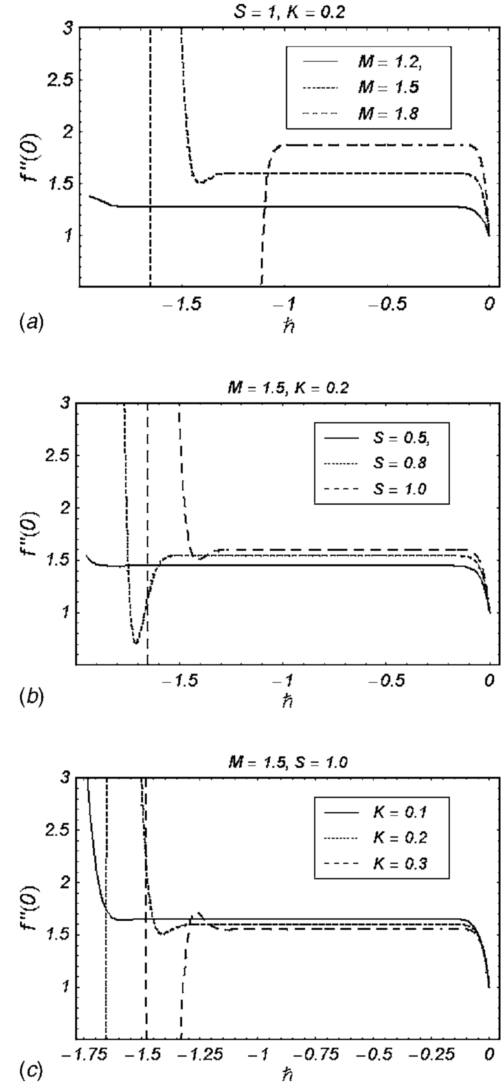


Fig. 2 $\hbar$ **curves for 30th order of approximation**

$$\Lambda_{m,n}^q = \sum_{k=0}^{m-1} \sum_{i=\max\{0,n-m+k\}}^{\min\{n,k+1\}} \sum_{j=\max\{0,q-m+k+n-i\}}^{\min\{q,k+1-i\}} e_{k,i}^j a_{m-1-k,n-i}^{q-j}$$

$$b_{m,n}^k = (k+1)a_{m,n}^{k+1} - na_{m,n}^k, \quad c_{m,n}^k = (k+1)b_{m,n}^{k+1} - nb_{m,n}^k$$

$$d_{m,n}^k = (k+1)c_{m,n}^{k+1} - nc_{m,n}^k, \quad e_{m,n}^k = (k+1)d_{m,n}^{k+1} - nd_{m,n}^k \tag{31}$$

$$a_{0,0}^0 = s - 1, \quad a_{0,1}^0 = 1 \tag{32}$$

For the detailed procedure of deriving the above relations the reader is referred to Ref. [37]. Therefore, the totally explicitly analytical solution is

$$f(\eta) = \sum_{m=0}^{\infty} f_m(\eta) = \lim_{M \to \infty} \left[\sum_{m=0}^{M} a_{m,0}^0 + \sum_{n=1}^{M+1} e^{-n\eta} \left(\sum_{m=n-1}^{M} \sum_{k=0}^{m+1-n} a_{m,n}^k \eta^k \right) \right] \tag{33}$$

4 Convergence of the Analytic Solution

The explicit, analytic expression given by Eq. (33) contains the auxiliary parameter $\hbar$, which gives the convergence region and rate of approximation for the homotopy analysis method. The auxiliary parameter $\hbar$ depends upon the physical parameters of the flow problem. In Fig. 2 the $\hbar$ curves are plotted for different

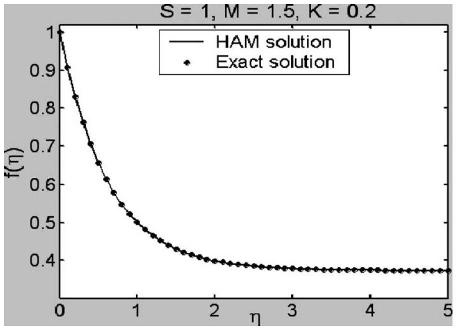


Fig. 3 Comparison of HAM and exact solutions

values of M, S, and K at the 30th order of approximation. Figure 2(a) gives the admissible range of $\hbar$ for different values of M ($=1.2, 1.5, 1.8$) keeping S and K fixed. The variations of S and K for the range of $\hbar$ can be seen in Figs. 2(b) and 2(c), respectively. Figure 2 indicates the range for the admissible values of the parameter $\hbar$ which is $-1.8 \leqslant \hbar < -0.1$. The series Eq. (33) converges in the whole region of η when $\hbar = -0.8$. It is further noted from Fig. 2 that the interval for admissible values of $\hbar$ increases by increasing order of approximation. Figure 3 gives a comparison between the HAM solution Eq. (33) and the exact solution given in Eq. (8). Table 1 is displayed to show the convergence of the HAM solution with increasing order of approximation. Table 2 gives a comparison between the HAM solution and shows a good agreement with the exact solution.

5 Results and Discussion

The main results of interest here are the influence of suction velocity S, the Hartman number M, and the viscoelastic parameter K on the velocity profiles f and f'. In order to analyze these important characteristics of the problem, we plot Figs. 4–9. The variation of velocity and the shear stress at the wall is also tabulated in Tables 3 and 4, respectively.

Figures 4 and 5 represent the variations of f and f' for various values of suction parameter S. It is noted from these figures that the magnitude of both f and f' decreases when the suction parameter S increases and this decrement is larger in the case of f. Moreover, the thickness of the boundary layer decreases with the increase in S. This is in keeping with the fact that suction causes reduction in the boundary layer thickness.

In order to illustrate the influence of Hartman number M on f and f', we prepared Figs. 6 and 7, respectively. As expected, increasing the magnitude of M reduces f and f'. This is due to the effect of magnetic force against the flow direction. It can be seen that with the increase of M, the magnitude of f increases more rapidly when compared with f'. Figures 6 and 7 further depict that there is a decrease in the thickness of the boundary layer due to an increase in M.

The flow dependence of a second grade fluid on the material parameter K can be clearly observed from Figs. 8 and 9. From these figures it can be determined that increasing K decreases f and f'. These figures also indicate that large values of K cause f and f' to become flatter. It may also be noted from Fig. 9 that the boundary layer thickness decreases when K increases. Also, the magnitude of f is larger than f' when K increases.

Table 2 Comparison of HAM and exact solutions for f when $S=1$, $M=1$, $K=0.3$, and $\hbar=-0.8$

η	HAM solution, Eq. (33)	Exact solution, Eq. (8)
0.0	1.0	1.0
0.2	0.818731	0.818731
0.5	0.606531	0.606531
1.0	0.367879	0.367879
2.0	0.135335	0.135335
3.0	0.0497871	0.0497871
5.0	0.00673795	0.00673795
10.0	0.0000453999	0.0000453999

Table 1 shows $f''(0)$. It is interesting to note here that the values of $f''(0)$ correspond to the exact solution at the 15th order of approximation. Table 2 provides a comparison of f for HAM and exact solutions for various values of η while keeping other parameters fixed. An excellent agreement is observed here.

Tables 3 and 4 have been made just to see the influence of K, M, and S on f and the shear stress at the wall (τ_w), respectively. Table 3 shows the variations of K, M, and S on f for both HAM and exact solutions given by Eqs. (33) and (8), respectively. It is found that f decreases as the viscoelastic parameter K increases, and increases for large values of M and S. Table 4 elucidates the variation of K, M, and S on the shear stress at the wall τ_w. It is observed that the magnitude of the shear stress first decreases and after $K=0.5$, it increases for large values of $K(\leqslant 10)$. The shear stress τ_w increases when both M and S increases. It is further noted that the agreement between HAM and exact solution is quite good.

6 Concluding Remarks

In this paper, the MHD second grade fluid flow due to a shrinking sheet is considered. The series solution is obtained and the convergence is shown. The effects of the sundry parameters are discussed through graphs. A comparison between HAM and exact solutions is given. This kind of analytic solution for MHD flow of a second grade fluid over a shrinking sheet is presented for the first time in the literature. The following observations have been made:

- The magnitude of f and f' is decreased by increasing S, M, and K, respectively;
- The boundary layer thickness is decreased as S, M, and K increases;
- The magnitude of f is larger when compared with f';
- The HAM results for MHD viscous fluid can be obtained by setting $K=0$;
- The HAM results are identical to the exact solution (Tables 1 and 2).

Acknowledgment

The authors wish to express their cordial thanks to the anonymous referees for their valuable suggestions for the improvement of the quality of this work. The financial support from Higher Education Commission (HEC) is also acknowledged.

Table 1 Values of $f''(0)$ for HAM and exact solutions when $S=1$, $M=1.5$, $K=0.2$, and $\hbar=-0.8$

Order of approximation	1	5	10	15	20	25	Exact solution
$f''(0)$	1.5	1.5879	1.59416	1.59443	1.59443	1.59443	1.59443

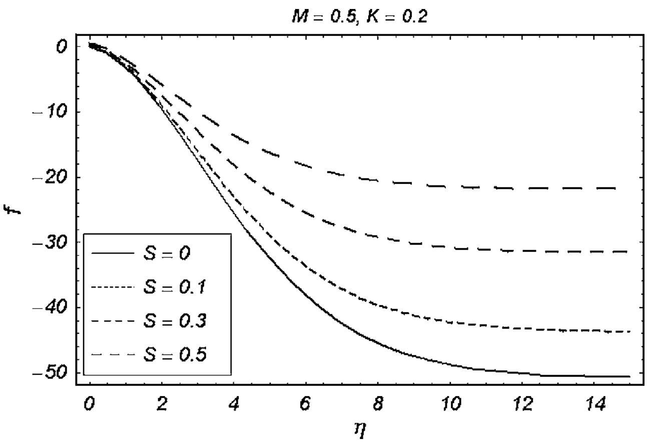


Fig. 4 Effects of suction *S* on *f* at $\hbar = -0.8$

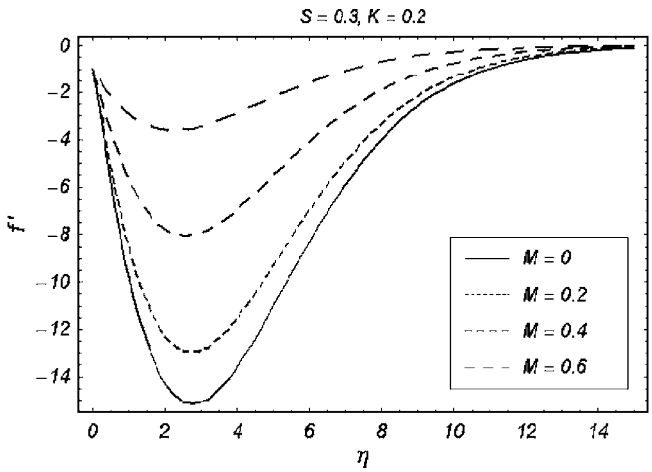


Fig. 7 Effects of Hartman number *M* on *f*′ at $\hbar = -0.8$

Nomenclature

u, v = velocities in x, y direction
x, y = spatial coordinates
V = suction velocity
a = the shrinking constant
B_0 = applied magnetic field
p = embedding parameter
$\mathcal{L}$ = auxiliary linear operator
$\mathcal{N}$ = nonlinear operator
$\hbar$ = nonzero auxiliary parameter
f = real function
η = similarity variable
μ = dynamic viscosity
ν = kinematic viscosity
ρ = density
α_1 = material moduli
σ = electrical conductivity

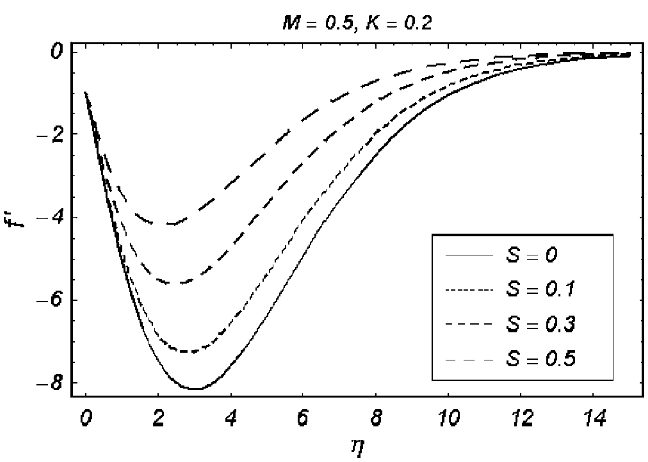


Fig. 5 Effects of suction *S* on *f*′ at $\hbar = -0.8$

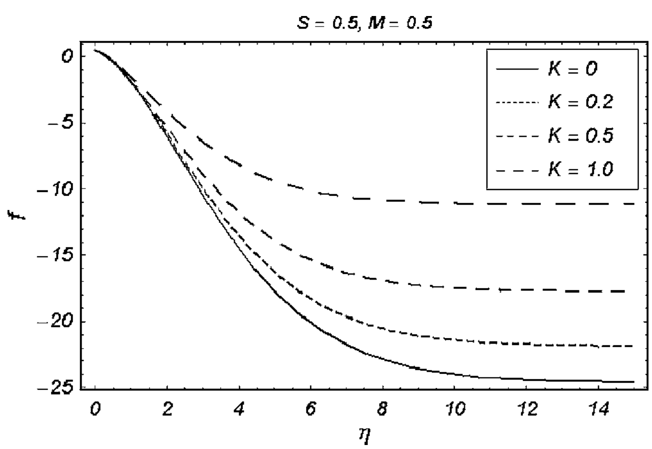


Fig. 8 Effects of second grade parameter *K* on *f* at $\hbar = -0.8$

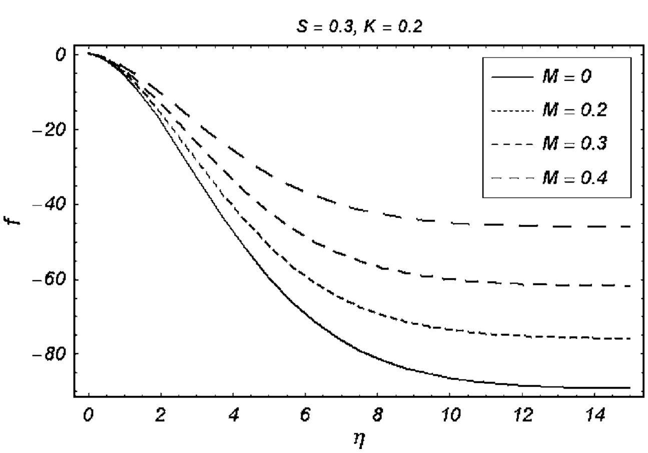


Fig. 6 Effects of Hartman number *M* on *f* at $\hbar = -0.8$

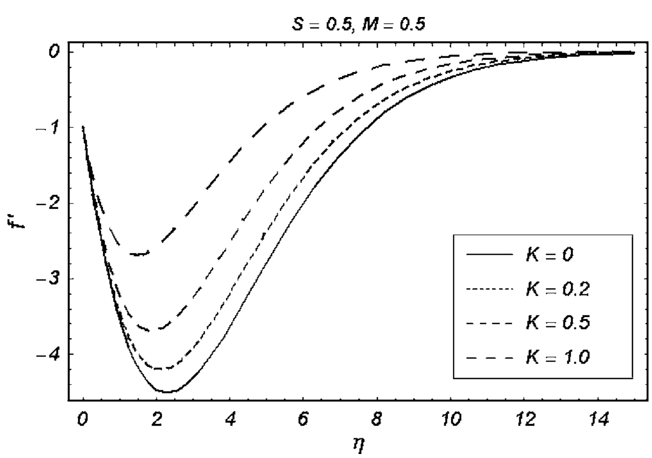


Fig. 9 Effects of second grade parameter *K* on *f*′ at $\hbar = -0.8$

Table 3 Comparison of HAM and exact solutions for f at $\eta = 0.2$

K	M	S	HAM solution, Eq. (33)	Exact solution, Eq. (8)
0.1	1.5	1.0	0.829654	0.829654
0.2			0.828747	0.828753
0.5			0.826977	0.826944
0.7			0.826439	0.826136
1.0			0.825786	0.825734
2.0			0.823493	0.823478
5.0			0.821473	0.821484
10.0			0.820394	0.820393
20.0			0.819669	0.819669
0.5	1.0		0.818731	0.818731
	1.2		0.822586	0.822585
	1.5		0.826977	0.826944
	2.0		0.832040	0.832758
	3.0		0.842011	0.842021
	4.0		0.849647	0.849604
	5.0		0.856786	0.856129
	10.0		- --	0.879903
	20.0		- --	0.907408
	1.5	0.0	−0.071811	−0.071812
		0.2	0.027922	0.027922
		0.5	0.327404	0.327402
		1.0	0.826977	0.826944
		1.5	1.326691	1.326690
		2.0	1.826531	1.826530
		3.0	2.826341	2.826340
		5.0	4.826151	4.826150
		10.0	9.825981	9.825990

Table 4 Comparison of HAM and exact solutions for shear stress τ_w at the wall $\eta=0$

K	M	S	HAM solution, Eq. (33)	Exact solution, Eq. (8)
0.1	1.5	1.0	1.427490	1.427450
0.2			1.146130	1.146210
0.5			0.358841	0.358816
1.0			−0.855190	−0.855289
1.5			−2.007020	−2.007130
2.0			−3.120720	−3.123660
3.0			−5.594256	−5.594730
5.0			−9.502730	−9.509060
10.0			−19.752400	−19.768000
0.5	1.0		0.0	0.0
	1.2		0.136592	0.136510
	1.5		0.358841	0.358816
	2.0		0.780263	0.780140
	3.0		1.794175	1.794080
	5.0		4.366281	4.366020
	1.5	0.1	−0.658248	−0.658203
		0.2	−0.533468	−0.533490
		0.5	−0.184517	−0.184580
		1.0	0.358844	0.358816
		1.5	0.881981	0.882960
		2.0	1.396990	1.398320
		3.0	2.416510	2.416960
		5.0	4.435196	4.435010

S = dimensionless suction velocity, $S=V/\sqrt{a\nu}$
M^2 = the Hartman number, $M^2=\sigma B_0^2/\rho a$
K = dimensionless viscoelastic or second grade parameter, $K=a\alpha_1/\nu\rho$
τ_w = dimensionless shear stress at the wall
β = real function
A = constant

References

[1] Khan, M., Hayat, T., and Asghar, S., 2006, "Exact Solution for MHD Flow of a Generalized Oldroyd-B Fluid with Modified Darcy's Law," Int. J. Eng. Sci., **44**, pp. 333–339.
[2] Fetecau, C., and Fetecau, C., 2005, "On Some Axial Couette Flows of a Non-Newtonian Fluid," ZAMP, **56**, pp. 1098–1106.
[3] Fetecau, C., and Fetecau, C., 2006, "Starting Solutions for the Motion of a Second Grade Fluid due to Longitudinal and Torsional Oscillations of a Circular Cylinder," Int. J. Eng. Sci., **44**, pp. 788–796.
[4] Fetecau, C., and Fetecau, C., 2005, "Decay of a Potential Vortex in an Oldroyd-B Fluid," Int. J. Eng. Sci., **43**, pp. 430–351.
[5] Fetecau, C., and Fetecau, C., 2005, "Starting Solutions for Some Unsteady Unidirectional Flows of a Second Grade Fluid," Int. J. Eng. Sci., **43**, pp. 781–789.
[6] Hayat, T., and Kara, A. H., 2006, "Couette Flow of a Third Grade Fluid with Variable Magnetic Field," Math. Comput. Modell., **43**, pp. 132–137.
[7] Hayat, T., 2005, "Oscillatory Solution in Rotating Flow of a Johnson–Segalman Fluid," Z. Angew. Math. Mech., **85**, pp. 449–456.
[8] Tan, W. C., and Masuoka, T., 2005, "Stokes First Problem for an Oldroyd-B Fluid in a Porous Half Space," Phys. Fluids, **17**, pp. 023101.
[9] Chen, C. I., Chen, C. K., and Yang, Y. T., 2004, "Unsteady Unidirectional Flow of an Oldroyd-B Fluid in a Circular Duct with Different Given Volume Flow Rate Conditions," Heat Mass Transfer, **40**, pp. 203–209.
[10] Tan, W. C., and Masuoka, T., 2005, "Stokes First Poroblem for Second Grade Fluid in a Porous Half Space," Int. J. Non-Linear Mech., **40**, pp. 515–522.
[11] Sakiadis, B. C., 1961, "Boundary Layer Behaviour on Continuous Solid Surfaces," AIChE J., **7**, pp. 26–28.
[12] Sakiadis, B. C., 1961, "Boundary Layer Behaviour on Continuous Solid Surfaces: II. The Boundary Layer on a Continuous Flat Surface," AIChE J., **17**, pp. 221–225.
[13] Cortell, R., 2006, "Effects of Viscous Dissipation and Work Done by Deformation on the MHD Flow and Heat Transfer of a Viscoelastic Fluid over a Stretching Sheet," Phys. Lett. A, **357**, pp. 298–305.
[14] Cortell, R., 2005, "Flow and Heat Transfer of a Fluid Through a Porous Medium over a Stretching Surface with Internal Heat Generation/Absorption and Suction/Blowing," Fluid Dyn. Res., **37**, pp. 231–245.
[15] Cortell, R., 2006, "A Note on Flow and Heat Transfer of a Viscoelastic Fluid over a Stretching Sheet," Int. J. Non-Linear Mech., **41**, pp. 78–85.
[16] Sadeghy, K., Najafi, A. H., and Saffaripour, M., 2005, "Sakiadis Flow of an Upper-Convected Maxwell Fluid," Int. J. Non-Linear Mech., **40**, pp. 1220–1228.
[17] Hayat, T., Abbas, Z., and Sajid, M., 2006, "Series Solution for the Upper-Convected Maxwell Fluid over a Porous Stretching Plate," Phys. Lett. A, **358**, pp. 396–403.
[18] Hayat, T., and Sajid, M., 2007, "Analytic Solution for Axisymmetric Flow and Heat Transfer of a Second Grade Fluid Past a Stretching Sheet, Int. J. Heat Mass Transfer, **50**, pp. 75–84.
[19] Liao, S., 2003, "On the Analytic Solution of Magnetohydrodynamic Flows of Non-Newtonian Fluids over a Stretching Sheet," J. Fluid Mech., **488**, pp. 189–212.
[20] Xu, H., 2005, "An Explicit Analytic Solution for Convective Heat Transfer in an Electrically Conducting Fluid at a Stretching Surface with Uniform Free Stream," Int. J. Eng. Sci., **43**, pp. 859–874.
[21] Liu, I. C., 2005, "Flow and Heat Transfer of an Electrically Conducting Fluid of a Second Grade in a Porous Medium over a Stretching Sheet Subject to a Transverse Magnetic Field," Int. J. Non-Linear Mech., **40**, pp. 465–474.
[22] Yürüsoy, M., 2006, "Unsteady Boundary Layer Flow of Power-Law Fluid on Stretching Sheet Surface," Int. J. Eng. Sci., **44**, pp. 325–332.
[23] Wang, C. Y., 1990, "Liquid Film on an Unsteady Stretching Sheet," Q. Appl. Math., **48**, pp. 601–610.
[24] Miklavcic, M., and Wang, C. Y., 2006, "Viscous Flow due to a Shrinking Sheet," Q. Appl. Math., **64**, pp. 283–290.
[25] Liao, S. J., 2003, *Beyond Perturbation: Introduction to Homotopy Analysis Method*, Chapman & Hall/CRC, Boca Raton, FL.
[26] Liao, S. J., 2004, "On the Homotopy Analysis Method for Nonlinear Problems," Appl. Math. Comput., **147**, pp. 499–513.
[27] Liao, S. J., 1999, "A Uniformly Valid Analytic Solution of 2D Viscous Flow Past a Semi-infinite Flat Plate," J. Fluid Mech., **385**, pp. 101–128.
[28] Liao, S. J., and Cheung, K. F., 2003, "Homotopy Analysis of Nonlinear Progressive Waves in Deep Water," J. Eng. Math., **45**, pp. 105–116.
[29] Liao, S. J., 2006, "An Analytic Solution of Unsteady Boundary-Layer Flows Caused by an Impulsively Stretching Plate," Commun. Nonlinear Sci. Numer. Simul., **11**, pp. 326–339.
[30] Liao, S. J., 2005, "Comparison between the Homotopy Analysis Method and Homotopy Perturbation Method," Appl. Math. Comput., **169**, pp. 1186–1194.
[31] Abbasbandy, S., 2006, "The Application of Homotopy Analysis Method to Nonlinear Equations Arising in Heat Transfer," Phys. Lett. A, **360**, pp. 109–113.
[32] Tan, Y., and Abbasbandy, S., 2007, "Homotopy Analysis Method for Quadratic Recati Differential Equation," Commun. Nonlinear Sci. Numer. Simul., in press.
[33] Zhu, S. P., 2006, "An Exact and Explicit Solution for the Valuation of American Put Options," Quant. Finance, **6**, pp. 229–242.
[34] Zhu, S. P., 2006, "A Closed-Form Analytical Solution for the Valuation of Convertible Bonds with Constant Dividend Yield," ANZIAM J., **47**, pp. 477–494.
[35] Wu, Y., Wang, C., and Liao, S. J., 2005, "Solving Solitary Waves with Discontinuity by Means of the Homotopy Analysis Method," Chaos, Solitons Fractals, **26**, pp. 177–185.
[36] Hayat, T., Khan, M., and Ayub, M., 2004, "On the Explicit Analytic Solutions of an Oldroyd 6-Constant Fluid," Int. J. Eng. Sci., **42**, pp. 1235–135.
[37] Hayat, T., Khan, M., and Asghar, S., 2004, "Homotopy Analysis of MHD

Flows of an Oldroyd 8-Constant Fluid," Acta Mech., **168**, pp. 213–232.
[38] Sajid, M., Hayat, T., and Asghar, S., 2006, "On the Analytic Solution of Steady Flow of a Fourth Grade Fluid," Phys. Lett. A, **355**, pp. 18–24.
[39] Abbas, Z., Sajid, M., and Hayat, T., 2006, "MHD Boundary Layer Flow of an Upper-Convected Maxwell Fluid in Porous Channel," Theor. Comput. Fluid Dyn., **20**, pp. 229–238.
[40] Hayat, T., Abbas, Z., Sajid, M., and Asghar, S., 2007, "The Influence of Thermal Radiation on MHD Flow of a Second Grade Fluid," Int. J. Heat Mass Transfer, **50**, pp. 931–941.
[41] Sajid, M., Hayat, T., and Asghar, S., 2007, "Comparison of the HAM and HPM Solutions of Thin Film Flows of Non-Newtonian Fluids on a Moving Belt," Nonlinear Dyn., in press.
[42] Hayat, T., and Sajid, M., 2007, "On Analytic Solution for Thin Film Flow of a Fourth Grade Fluid Down a Vertical Cylinder," Phys. Lett. A, **361**, pp. 316–322.
[43] Shercliff, J. A., 1965, *A Text Book of Magnetohydrodynamics*, Pergamon, Elmsford, New York.
[44] Nanousis, N. D., 1999, "Theoretical Magnetohydrodynamic Analysis of Mixed Convection Boundary Layer Flow Over a Wedge with Uniform Suction or Injection," Acta Mech., **138**, pp. 21–30.
[45] Vajravelu, K., and Rivera, J., 2003, "Hydromagnetic Flow at an Oscillating Plate," Int. J. Non-Linear Mech., **38**, pp. 305–312.
[46] Hayat, T., Zumurad, M., Asghar, S., and Siddiqui, A. M., 2003, "Magnetohydrodynamic Flow due to Non-coaxial Rotations of a Porous Oscillating Disk and a Fluid at Infinity," Int. J. Eng. Sci., **41**, pp. 1177–1196.
[47] Tokis, J. N., 1978, "Hydromagnetic Unsteady Flow due to an Unsteady Plate," Astrophys. Space Sci., **58**, pp. 167–174.
[48] Pop, I., Kumari, M., and Nath, G., 1994, "Conjugate MHD Flow Past a Flat Plate," Acta Mech., **106**, pp. 215–220.
[49] Abel, S., Veena, P. H., Rajagopal, K., and Pravin, V. K., 2004, "Non-Newtonian Magnetohydrodynamic Flow over a Stretching Surface with Heat and Mass Transfer, Int. J. Non-Linear Mech., **39**, pp. 1067–1078.
[50] Garg, V. K., and Rajagopal, K. R., 1991, "Flow of a Non-Newtonian Fluid Past a Wedge," Acta Mech., **88**, pp. 113–123.
[51] Vajravelu, K., and Roper, T., 1999, "Flow and Heat Transfer in a Second Grade Fluid over a Stretching Sheet," Int. J. Non-Linear Mech., **34**, pp. 1031–1036.

Plane Contact of Hot Flat-Ended Punch and Thermoelastic Half-Space Involving Finite Friction

Volodymyr Pauk
Faculty of Civil and Environmental Engineering,
Technical University of Kielce,
Kielce 25-314, Poland
e-mail: pauk@tu.kielce.pl

Plane normal contact of a rigid flat-ended hot punch and a thermoelastic half-space is considered under the assumption of finite friction between contacting surfaces. The problem is treated with the boundary integral methods. We consider independently normal and tangential contact problems. The rise of slip and stick regions under the punch is studied. The nonlinear equations for the unknown stick region size are obtained and solved numerically. [DOI: 10.1115/1.2723821]

Keywords: contact problem, finite friction, thermoelasticity, stick, slip

1 Introduction

When a rigid flat-ended punch is pressed symmetrically into an elastic half-space the common contact area consists of the central zone of the stick contact and two zones situated symmetrically near punch corners, in which contacting surfaces slip. The distribution of these zones is unknown and depends on the friction coefficient and the Poisson ratio. This formulation of the normal contact problem was proposed by Galin [1], and was solved by Spence [2] as an eigenvalue problem. In this paper we study a similar contact problem assuming that the punch is hot. The plane contact problem involving a hot punch was considered first by Borodachev [3]. The effect of thermal stresses in contact problems is very important. These stresses can lead to a separation between contacting surfaces. The plane thermoelastic contact involving the separation in contact zone was considered by Comninou et al. [4].

In our paper the effects due to the finite friction and thermal deformation will be treated simultaneously. To simplify the problem we use the Spence [2] assumption because the friction force has no effect on the vertical displacements of the half-space boundary. This assumption permits us to solve the normal problem independently of the tangential one. After this, having the distribution of normal traction, we can solve the tangential problem. The main feature of the thermal contact is that there are two different situations—full contact and separation, which implies two regimes of the slip arising in the tangential problem.

The contact problem involving the stick–slip transition is a subject of considerable interest in investigations of the fretting fatigue. The experimental data due to this phenomenon are widely presented in special literature. The studies involving thermal effects on the fretting fatigue are rather rare. The reports of experimental investigations into this topic are presented in the Refs. [5–8].

Our paper is organized as follows. In Sec. 2, the normal contact of the hot punch resting on the elastic half-space is recalled. We give the solution of this problem for two kinds of punch temperature and we study a full contact as well as a separation arising for some level of this temperature.

Section 3 presents the problem of tangential contact in the case of low temperature for which the normal contact remains over the whole width of the punch. It is shown that some slipping must take place near the contact area edges. After this, the problem is solved under the assumption that both stick and slip regions are distributed within the common contact area.

A similar study for high punch temperature involving some separation in the contact zone is presented in Sec. 4.

2 Normal Contact

Let us first consider the plane problem of stationary thermoelasticity for the half-space with the following boundary conditions

$$\sigma_{yy}(x,0) = \begin{cases} -p(x), & |x| \leqslant a \\ 0, & |x| > a \end{cases} \tag{1}$$

$$\tau_{xy}(x,0) = \begin{cases} q(x), & |x| \leqslant a \\ 0, & |x| > a \end{cases} \tag{2}$$

$$T(x,0) = \begin{cases} T_0(x), & |x| \leqslant a \\ 0, & |x| > a \end{cases} \tag{3}$$

where $\sigma_{yy}(x,y)$, $\tau_{xy}(x,y)$, and $T(x,y)$ are stress and temperature fields in the half-space, and the normal $p(x)$, tangential $q(x)$ tractions, as well as the contact temperature $T_0(x)$ are assumed here to be prescribed.

The vertical $v(x)$ and horizontal $u(x)$ displacements of the half-space surface, which are solutions of the boundary problem considered, can be obtained in the forms [3,9]

$$\frac{\partial v}{\partial x} = \frac{1-2\nu}{2\mu}q(x) + \frac{1-\nu}{\pi\mu}\int_{-a}^{a}\frac{p(s)}{s-x}ds - \frac{(1+\nu)\alpha}{\pi}\int_{-a}^{a}\frac{T_0(s)}{s-x}ds, \quad |x| < \infty \tag{4}$$

$$\frac{\partial u}{\partial x} = -\frac{1-2\nu}{2\mu}p(x) + \frac{1-\nu}{\pi\mu}\int_{-a}^{a}\frac{q(s)}{s-x}ds + (1+\nu)\alpha T_0(x), \quad |x| < \infty \tag{5}$$

where ν, μ, and α are, respectively, the Poisson ratio, shear modulus, and thermal expansion coefficient of the half-space material.

In further analysis we will assume the following, namely, the first term on the right side of Eq. (4) will be neglected. This means that the tangential traction has no effect on the vertical displacements of the half-space surface.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received June 30, 2005; final manuscript received February 4, 2007. Review conducted by Matthew R. Begley.

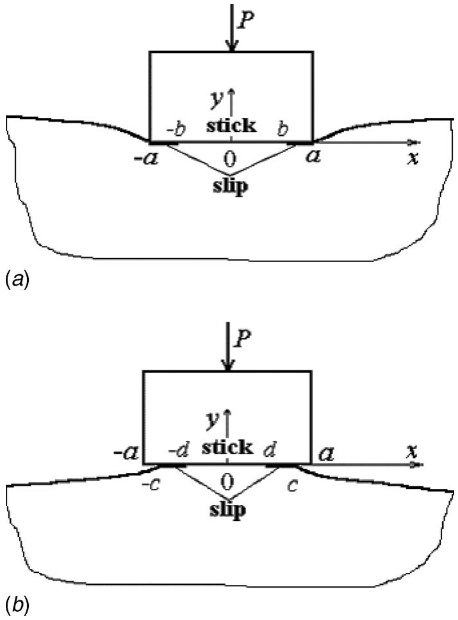


Fig. 1 (*a*) Perfect contact and (*b*) contact with separation

Now we consider the normal indention of the rigid flat-ended punch into the thermoelastic half-space (Fig. 1). It is assumed that the punch has temperature $T_0(x)$ and the temperature of the half-space surface outside the contact path is equal to zero. The problem is considered to be planar and stationary.

Satisfying the boundary condition of the normal problem

$$\frac{\partial v}{\partial x} = 0, \quad |x| \leqslant a \tag{6}$$

with the help of Eq. (4) we obtain the following integral equation

$$\frac{1-\nu}{\pi\mu}\int_{-a}^{a}\frac{p(s)-\gamma T_0(s)}{s-x}ds = 0, \quad |x| \leqslant a \tag{7}$$

where we have introduced the parameter $\gamma=(1+\nu)/(1-\nu)\alpha\mu$.

The integral Eq. (7) is of the Cauchy type and its solution satisfying the equilibrium condition

$$\int_{-a}^{a} p(s)ds = P \tag{8}$$

has the following form [10]

$$p(x) = \gamma T_0(x) + \frac{P-P_0}{\pi\sqrt{a^2-x^2}}, \quad |x| \leqslant a \tag{9}$$

where

$$P_0 = \gamma\int_{-a}^{a} T_0(s)ds \tag{10}$$

Further analysis will deal with two forms of the punch temperature:

(i) constant temperature

$$T_0(x) = T_0 = \text{const}, \quad |x| \leqslant a \tag{11}$$

(ii) elliptic distribution of the temperature

$$T_0(x) = T_0\sqrt{1-\frac{x^2}{a^2}} = \text{const}, \quad |x| \leqslant a \tag{12}$$

In the first case the contact pressure Eq. (9) can be calculated as

$$p(x) = \gamma T_0 + \frac{P-2a\gamma T_0}{\pi\sqrt{a^2-x^2}}, \quad |x| \leqslant a \tag{13}$$

When the punch temperature has the form Eq. (12) we find

$$p(x) = \gamma T_0\frac{\sqrt{a^2-x^2}}{a} + \frac{P-\frac{\pi}{2}a\gamma T_0}{\pi\sqrt{a^2-x^2}}, \quad |x| \leqslant a \tag{14}$$

Simple analysis of the formulas (13) and (14) permits us to conclude

$$\text{(a) If} \quad -\frac{P}{(\pi-2)a\gamma} \leqslant T_0 \leqslant \frac{P}{2a\gamma} \tag{15}$$

in the case of constant temperature and

$$-\frac{2P}{\pi a\gamma} \leqslant T_0 \leqslant \frac{2P}{\pi a\gamma} \tag{16}$$

in the case of elliptic temperature, where the normal pressure is positive for $|x| \leqslant a$ and we have the full contact between the punch and half-space (Fig. 1(*a*))

$$\text{(b) If} \quad T_0 > \frac{P}{2a\gamma} \quad \text{or} \quad T_0 > \frac{2P}{\pi a\gamma} \tag{17}$$

in the case of constant or elliptical temperatures, the normal pressure is negative near the points $x=\pm a$ which means that the separation arises in the vicinity of the punch corners (Fig. 1(*b*)). This situation will be considered later

$$\text{(c) If} \quad T_0 < -\frac{P}{(\pi-2)a\gamma} \quad \text{or} \quad T_0 < -\frac{2P}{\pi a\gamma} \tag{18}$$

in the case of constant or elliptical temperatures, the separation arises in the central points of the contact area. This case is called the "cool punch problem" [4] and effects of finite friction in this case will be presented separately.

To consider the separation problem (Fig. 1(*b*)) we need the solution of the similar contact problem for the parabolic punch. It can be obtained from the integral equation similar to Eq. (7) with the right side equal to x/R, where R is the punch radius. The bounded contact pressure under the parabolic punch of the constant temperature can be obtained in the form

$$p(x) = \gamma T_0 + \frac{2[P-2a\gamma T_0]}{\pi a^2}\sqrt{a^2-x^2}, \quad |x| \leqslant a \tag{19}$$

where the contact size a satisfies the equation

$$\frac{a^2}{R} = \frac{2(1-\nu)}{\pi\mu}[P-2a\gamma T_0] \tag{20}$$

When the contact temperature is given by Eq. (12) then the normal pressure has the form

$$p(x) = \frac{2P}{\pi a^2}\sqrt{a^2-x^2}, \quad |x| \leqslant a \tag{21}$$

where the contact size a satisfies the equation

$$\frac{a^2}{R} = \frac{2(1-\nu)}{\pi\mu}\left[P-\frac{\pi}{2}a\gamma T_0\right] \tag{22}$$

The contact size c in the separation problem can be obtained directly from formulas (20) and (22) as the limiting solution for $R\to\infty$ and replacing a by c. We obtain

$$c = \frac{P}{2\gamma T_0} \tag{23}$$

for the constant punch temperature and

$$c = \frac{2P}{\pi\gamma T_0} \tag{24}$$

for the elliptic one.

Corresponding distributions of the normal traction can be obtained from Eqs. (19), (23), (21), and (24) in the forms

$$p(x) = \gamma T_0, \quad |x| \leqslant c \tag{25}$$

and

$$p(x) = \gamma T_0 \sqrt{1 - \frac{x^2}{c^2}}, \quad |x| \leqslant c \tag{26}$$

Summarizing this section we have distribution of the normal pressure as well as the separation contact size.

3 Tangential Contact: The Case of Full Contact

Let us now consider tangential contact in the case of full contact, i.e., conditions (15) and (16) take place, and normal contact occurs over the path $(-a,a)$ (Fig. 1(a)).

First we assume the adhesive contact over this path. The boundary conditions of this problem are

$$\frac{\partial u}{\partial x} = 0, \quad |x| \leqslant a \tag{27}$$

$$|q(x)| < fp(x), \quad |x| \leqslant a \tag{28}$$

where f is the friction coefficient.

Satisfying condition (27) with help of Eq. (5) we arrive at the following integral equation

$$\frac{1-\nu}{\pi\mu}\int_{-a}^{a}\frac{q(s)}{s-x}ds = -\gamma T_0(x) + \frac{1-2\nu}{2(1-\nu)}p(x), \quad |x| \leqslant a \tag{29}$$

which should be solved together with the equilibrium condition

$$\int_{-a}^{a} q(s)ds = 0 \tag{30}$$

The solution of the system Eqs. (29) and (30) depends on the distribution of the punch temperature and has the form [6]

$$q(x) = -\frac{\gamma T_0}{2(1-\nu)}\frac{x}{\sqrt{a^2-x^2}} - \frac{1-2\nu}{2(1-\nu)}\frac{P-2a\gamma T_0}{\pi^2\sqrt{a^2-x^2}}\ln\left|\frac{a+x}{a-x}\right|, \quad |x| \leqslant a \tag{31}$$

when the temperature is constant, and

$$q(x) = -\frac{\gamma T_0}{2(1-\nu)\pi}\left\{\frac{2x}{\sqrt{a^2-x^2}} + \frac{\sqrt{a^2-x^2}}{a}\ln\left|\frac{a+x}{a-x}\right|\right\} - \frac{1-2\nu}{2(1-\nu)}\frac{P-\frac{\pi}{2}\gamma T_0}{\pi^2\sqrt{a^2-x^2}}\ln\left|\frac{a+x}{a-x}\right|, \quad |x| \leqslant a \tag{32}$$

if the function $T_0(x)$ is elliptic given by Eq. (12).

Analyzing the ratio $q(x)/p(x)$ we conclude that it tends to infinity for $x \to \pm a$, which makes impossible the satisfaction of the boundary condition (28) for finite values of the friction coefficient. Similar to the pure elastic problem [1,2], the contact area in the problem considered is also divided (see Fig. 1(a)) into a central region $(-b,b)$ with the stick conditions

$$\frac{\partial u}{\partial x} = 0, \quad |x| \leqslant b \tag{33}$$

$$|q(x)| < fp(x), \quad |x| \leqslant b \tag{34}$$

and lateral zones $(-a,-b)$, (b,a) where the slip conditions

$$\frac{\partial u}{\partial x} > 0, \quad b < |x| \leqslant a \tag{35}$$

$$|q(x)| = fp(x), \quad b < |x| \leqslant a \tag{36}$$

must be satisfied. The stick area size b is unknown.

Using the presentation of Eq. (5) for horizontal displacements, the boundary condition (33) gives the integral equation

$$\frac{1-\nu}{\pi\mu}\int_{-a}^{a}\frac{q(s)}{s-x}ds = -\gamma T_0(x) + \frac{1-2\nu}{2(1-\nu)}p(x), \quad |x| \leqslant b \tag{37}$$

Since this equation is written in the stick zone and the unknown function $q(x)$ is defined in the whole contact region, it is impossible to obtain the solution directly from Eq. (37). For this reason we will present the tangential traction in the following form

$$q(x) = fp(x)\mathrm{sgn}(x) - \begin{cases} q_0(x), & |x| \leqslant b \\ 0, & b < |x| \leqslant a \end{cases} \tag{38}$$

which satisfies automatically boundary conditions (34) and (36). Here $q_0(x)$ is a new unknown function called the corrective traction.

Substituting the superposition Eq. (38) into Eq. (37) after some calculations we obtain the integral equation for the corrective traction

$$\frac{1}{\pi}\int_{-b}^{b}\frac{q_0(s)}{s-x}ds = g(x), \quad |x| \leqslant b \tag{39}$$

where the right side has the form

$$g(x) = -\frac{\gamma T_0}{2(1-\nu)} + \frac{f\gamma T_0}{\pi}\int_0^a\left\{\frac{1}{s+x}+\frac{1}{s-x}\right\}ds + \frac{1-2\nu}{2(1-\nu)}\frac{P-2\gamma T_0}{\sqrt{a^2-x^2}} + f\frac{P-2\gamma T_0}{\pi^2}\int_0^a\left\{\frac{1}{s+x} + \frac{1}{s-x}\right\}\frac{ds}{\sqrt{a^2-s^2}}ds, \quad |x| \leqslant b \tag{40}$$

in the case of the constant punch temperature and when it is elliptic Eq. (12) we find

$$g(x) = -\frac{\gamma T_0}{2(1-\nu)}\frac{\sqrt{a^2-x^2}}{a} + \frac{f\gamma T_0}{\pi}\int_0^a\left\{\frac{1}{s+x}+\frac{1}{s-x}\right\}\frac{\sqrt{a^2-s^2}}{a}ds + \frac{1-2\nu}{2(1-\nu)}\frac{P-\frac{\pi}{2}\gamma T_0}{\sqrt{a^2-x^2}} + f\frac{P-\frac{\pi}{2}\gamma T_0}{\pi^2}\int_0^a\left\{\frac{1}{s+x} + \frac{1}{s-x}\right\}\frac{ds}{\sqrt{a^2-s^2}}ds, \quad |x| \leqslant b \tag{41}$$

The equilibrium condition (30) with the superposition Eq. (38) reads

$$\int_{-b}^{b} q_0(s)ds = 0 \tag{42}$$

The system Eqs. (39) and (42) has the solution

$$q_0(x) = \frac{1}{\pi\sqrt{b^2-x^2}}\int_{-b}^{b}\frac{\sqrt{b^2-s^2}}{x-s}g(s)ds = \frac{\varphi(x)}{\pi\sqrt{b^2-x^2}}, \quad |x| \leqslant b \tag{43}$$

where the function $\varphi(x)$ depends on the punch temperature and, involving formulas (40) and (41), can be obtained in the form

$$\varphi(x) = \frac{\pi\gamma T_0 x}{2(1-\nu)} - \frac{1-2\nu}{2(1-\nu)}[P - 2\gamma T_0]I_1(x) - f\gamma T_0 I_2(x) - f[P - 2\gamma T_0]I_3(x), \quad |x| \leq b \tag{44}$$

in the case of the constant temperature Eq. (11) and

$$\varphi(x) = \frac{\pi\gamma T_0 I_0(x)}{2(1-\nu)} - \frac{1-2\nu}{2(1-\nu)}\left[P - \frac{\pi}{2}\gamma T_0\right]I_1(x) - f\gamma T_0 I_4(x) - f\left[P - \frac{\pi}{2}\gamma T_0\right]I_3(x), \quad |x| \leq b \tag{45}$$

in the case of elliptic distribution Eq. (12). Here the following integrals

$$I_0(x) = \frac{1}{\pi a}\int_{-b}^{b} \frac{\sqrt{b^2-s^2}\sqrt{a^2-s^2}}{x-s}ds, \quad |x| \leq b < a \tag{46}$$

$$I_1(x) = \frac{1}{\pi}\int_{-b}^{b} \frac{\sqrt{b^2-s^2}}{\sqrt{a^2-s^2}}\frac{ds}{x-s}, \quad |x| \leq b < a \tag{47}$$

$$I_2(x) = -\int_{b}^{a} \sqrt{s^2-b^2}\left\{\frac{1}{x+s} + \frac{1}{x-s}\right\}ds, \quad |x| \leq b < a \tag{48}$$

$$I_3(x) = \frac{1}{\pi}\int_{b}^{a} \frac{\sqrt{s^2-b^2}}{\sqrt{a^2-s^2}}\left\{\frac{1}{x+s} + \frac{1}{x-s}\right\}ds, \quad |x| \leq b < a \tag{49}$$

$$I_4(x) = -\frac{1}{a}\int_{b}^{a} \sqrt{a^2-s^2}\sqrt{s^2-b^2}\left\{\frac{1}{x+s} + \frac{1}{x-s}\right\}ds, \quad |x| \leq b < a \tag{50}$$

are introduced.

In these formulae and in the subsequence $K(\cdot)$, $E(\cdot)$ are the complete elliptic integrals of the first and second kinds, respectively [11].

To provide the continuity of the tangential traction the corrective traction must be bounded in the points separating stick and slip zones, i.e.,

$$q_0(\pm b) = 0 \tag{51}$$

This condition with the aid of formulas (43)–(50) gives the following equation for the dimensionless stick size $\beta = b/a$:

(i) for the constant punch temperature Eq. (11)

$$(1-\Theta)\left[\frac{1-2\nu}{2(1-\nu)f}K(\beta) - K(\sqrt{1-\beta^2})\right] + \frac{\pi\Theta}{2}\ln\frac{\beta}{1+\sqrt{1-\beta^2}} + \frac{\pi^2\Theta}{8(1-\nu)f} = 0 \tag{52}$$

where we have introduced the normalized punch temperture

$$\Theta = \frac{2a\gamma T_0}{P} \tag{53}$$

and

(ii) for the elliptic distribution Eq. (12)

$$(1-\Theta)\left[\frac{1-2\nu}{2(1-\nu)f}K(\beta) - K(\sqrt{1-\beta^2})\right] - 2\Theta[K(\sqrt{1-\beta^2}) - E(\sqrt{1-\beta^2})] + \frac{\Theta}{(1-\nu)f}E(\beta) = 0 \tag{54}$$

where now

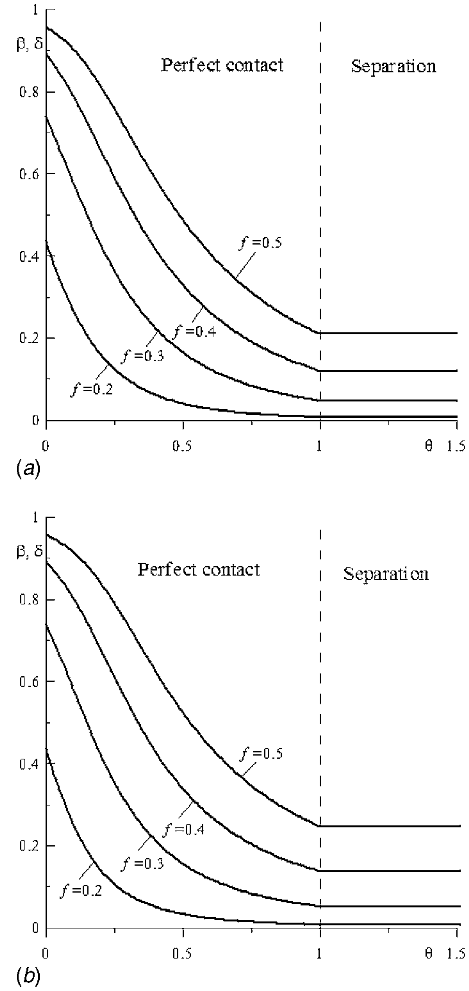


Fig. 2 (*a*) Stick size as a function of the punch temperature in the case of constant temperature and (*b*) stick size as a function of the punch temperature in the case of elliptic temperature

$$\Theta = \frac{\pi a\gamma T_0}{2P} \tag{55}$$

The well-known Spence result [2] for the stick area size in the case of the isothermal contact

$$\frac{1-2\nu}{2(1-\nu)f}K(\beta) = K(\sqrt{1-\beta^2}) \tag{56}$$

may be obtained directly from formulas (52) or (54) setting $\Theta = 0$.

To find unknown stick size we have nonlinear Eqs. (52) or (55). They were solved numerically by the Newton method. Results of calculations for $0 \leq \Theta \leq 1$ are presented in Figs. 2(*a*) and 2(*b*) for the constant and elliptic punch temperature, respectively. The problem also depends on the friction coefficient and the Poisson ratio. For calculations we have taken $\nu = 0.3$ and four values of f. We see that the growth of the punch temperature leads to decreasing the stick size $\beta = b/a$. The difference between results for the constant and elliptic distributions of the punch temperature is not great for small values of the friction coefficient, but for $f = 0.5$ this difference reaches 25% in the limiting point $\Theta = 1$.

4 Tangential Contact: The Case of the Contact With Separation

Let us now assume that the punch temperature satisfies condition (17) and the separation between the punch and the half-space occurs in the region $(-a,-c)$ and (c,a), Fig. 1(b). The new contact size c is given by Eqs. (23) or (24) and the corresponding normal pressure by Eqs. (25) and (26). Considering the tangential problem, similarly to the case of the perfect contact, we assume first that the contact is adhesive in the whole contact path, i.e., the boundary conditions Eqs. (27), (28), are satisfied in each point of the region $(-c,c)$. Similarly to Eqs. (31) and (32) we can find the distribution of the tangential traction providing the adhesive contact in the forms

$$q(x) = -\frac{\gamma T_0}{2(1-\nu)} \frac{x}{\sqrt{c^2-x^2}}, \quad |x| \leq c \tag{57}$$

for the constant punch temperature Eq. (11) and

$$q(x) = -\frac{\gamma T_0}{2(1-\nu)} \frac{c}{\sqrt{c^2-x^2}} \left[\frac{2x}{c} + \left(1 - \frac{x^2}{c^2}\right) \ln \left| \frac{c+x}{c-x} \right| \right], \quad |x| \leq c \tag{58}$$

in the case of elliptic temperature Eq. (12).

In both cases the ratio $q(x)/p(x)$ tends to infinity for $x \to \pm c$, what implies the existence of a central region $(-d,d)$ with the stick conditions Eqs. (33) and (34) and lateral zones $(-c,-d)$, (d,c) with the slip conditions Eqs. (35) and (36). The new stick size d is unknown.

Further analysis of the contact with separation is similar to that performed for the perfect contact. The solution for the corrective traction $q_0(x)$ defined in the stick area $(-d,d)$ has the form

$$q_0(x) = \frac{fP\varphi_0(x)}{\pi\sqrt{d^2-x^2}}, \quad |x| \leq d \tag{59}$$

where

$$\varphi_0(x) = \frac{\pi x}{4(1-\nu)fc} - \frac{1}{2c} J_2(x), \quad |x| \leq d \tag{60}$$

in the case of the constant temperature Eq. (11) and

$$\varphi_0(x) = \frac{2}{\pi c} \left[\frac{\pi}{2(1-\nu)f} J_0(x) - J_4(x) \right], \quad |x| \leq d \tag{61}$$

in the case of elliptic temperature distribution Eq. (12). Here the following integrals

$$J_0(x) = \frac{1}{\pi c} \int_{-d}^{d} \frac{\sqrt{d^2-s^2}\sqrt{c^2-s^2}}{x-s} ds, \quad |x| \leq d < c \tag{62}$$

$$J_2(x) = -\int_{d}^{c} \sqrt{s^2-d^2} \left\{ \frac{1}{x+s} + \frac{1}{x-s} \right\} ds, \quad |x| \leq d < c \tag{63}$$

$$J_4(x) = -\frac{1}{c} \int_{d}^{c} \sqrt{c^2-s^2}\sqrt{s^2-d^2} \left\{ \frac{1}{x+s} + \frac{1}{x-s} \right\} ds, \quad |x| \leq d < c \tag{64}$$

are introduced.

Satisfying, with aid of the above formulae, the physical condition

$$q_0(\pm d) = 0 \tag{65}$$

we obtain the equation for the dimensionless stick size $\delta = d/c$

$$\ln \frac{\delta}{1+\sqrt{1-\delta^2}} = -\frac{\pi}{4(1-\nu)f} \tag{66}$$

for the constant punch temperature Eq. (11) and

$$\frac{K(\sqrt{1-\delta^2}) - E(\sqrt{1-\delta^2})}{E(\delta)} = -\frac{1}{2(1-\nu)f} \tag{67}$$

in the case of elliptic temperature Eq. (12).

The first of these equations can be solved in the explicit form

$$\delta = \frac{2}{1 + \exp\left(\frac{\pi}{4(1-\nu)f}\right)} \tag{68}$$

and the second one should be solved numerically with the help of the Newton method.

These solutions are presented in Figs. 2(a) and 2(b) in the range $\Theta \geq 1$ for $\nu = 0.3$ and four values of the friction coefficient. It is important to notice that the temperature has no effects on the stick size in the case of the contact with separation.

5 Conclusions

The contact problem for the rigid hot punch and the thermoelastic half-space was considered. The finite friction under the punch is involved. It was shown that the boundaries of contacting bodies slip near the punch edges while the central region of the common contact area is in the stick condition. The equations for the stick area size are derived for two cases of the punch temperature. It was shown that the temperature has the greatest effects on the stick–slip distribution under the punch. Obtained results can find an application in the investigation of the contact fatigue and fretting.

Nomenclature

Oxy	=	coordinate system
a	=	punch half-width
b	=	stick size in the case of the perfect contact
c	=	separated contact size
d	=	stick size in the case of the contact with separation
f	=	friction coefficient
$p(x)$	=	normal traction
$q(x)$	=	tangential traction
$T(x,y)$	=	temperature field
$u(x,y)$	=	horizontal displacements
$v(x,y)$	=	vertical displacements
α	=	thermal expansion coefficient
$\beta = b/a$	=	normalized stick size in the case of the perfect contact
$\delta = d/c$	=	normalized stick size in the case of the contact with separation
ν	=	Poisson ratio
μ	=	shear modulus
$\gamma = (1+\nu)/(1-\nu)\alpha\mu$	=	thermoelastic parameter
$\sigma_{yy}(x,y), \tau_{xy}(x,y)$	=	stresses components

References

[1] Galin, L. A., 1945, "The Punch Indentation With Friction and Adhesion," J. Appl. Math. Mech., **9**, pp. 413–424.

[2] Spence, D. A., 1973, "An Eigenvalue Problem for Elastic Contact With Finite Friction," Proc. Cambridge Philos. Soc., **73**, pp. 249–268.

[3] Borodachev, N. M., 1963, "Plane Thermoelastic Problem on the Punch Inden-

tation," Eng. J., **3**, pp. 234–240.
[4] Comninou, M., Dundurs, J., and Barber, J. R., 1981, "Planar Hertz Contact With Heat Conduction," ASME J. Appl. Mech., **48**, pp. 549–554.
[5] Sproles, E. S., and Duquette, D. J., 1978, "Interface Temperature Measurements in the Fretting of a Medium Carbon Steel," Wear, **47**, pp. 387–396.
[6] Attia, M. H., and D'Silva, N. S., 1985, "Effect of Mode of Motion and Process Parameters on the Prediction of Temperature Rise in Fretting Wear," Wear, **106**, pp. 203–211.
[7] Szolwinski, M. P., Harish, G., Farris, T. N., and Sakagami, T., 1999, "In-Situ Measurement of Near-Surface Fretting Contact Temperatures in an Aluminum Alloy," ASME J. Tribol., **121**, pp. 11–19.
[8] Podgornik, B., Kalin, M., Vižintin, J., and Vodopivec, F., 2001, "Microstructural Changes and Contact Temperature During Fretting in Steel-Steel Contact," ASME J. Tribol., **123**, pp. 670–675.
[9] Johnson, K. L., 1987, *Contact Mechanics*, Cambridge University Press, Cambridge, UK.
[10] Muskhelishvili, N. I., 1953, *Singular Integral Equations*, Noordhoff International Publishing, Groningen, The Netherlands
[11] Abramowitz, M., and Stegun, I., 1970, *Handbook of Mathematical Functions*, Dover, New York.

Martin Karlsson
Student
e-mail: karmar@ltu.se

Jan-Olov Aidanpää
Associate Professor

Division of Computer Aided Design,
Polhem Laboratory,
Luleaa University of Technology,
Luleaa, Norrbotten 97187, Sweden

Richard Perers
Student

Mats Leijon
Professor

Department of Electricity,
Aangstroem Laboratory,
Uppsala University,
Uppsala 751 21, Sweden

Rotor Dynamic Analysis of an Eccentric Hydropower Generator With Damper Winding for Reactive Load

Asymmetry in the magnetic circuit, around the air gap circumference, in a hydroelectric generator will give rise to a unbalanced magnetic pull. In this paper, a hydropower rotor system is modeled and the influence of electro-mechanical forces due to overexcitation is analyzed. The active power has been kept constant and the rotor excitation has been changed in order to vary the output of reactive power. The electromagnetic field is solved with the finite element method. Two electromagnetic models are compared: one with and one without damper winding. The mechanical model of the generator consists of a four degrees of freedom rigid disk connected to an elastic shaft supported by two bearings with linear properties. It has been found that the unbalanced magnetic pull slightly increases for reactive loads resulting in a decrease of natural frequencies and an increase of unbalance response. When the damper winding is included, the magnetic pull will decrease compared to the model without damper winding, and the pull force has two components: one radial and one tangential. The tangential component can influence the stability of the mechanical system for a range of design parameters.
[DOI: 10.1115/1.2723822]

1 Introduction

Asymmetry of the magnetic circuit in electrical machines can lead to vibrations. An off-centered rotor in a generator results in asymmetry in the air gap. The rotor will be affected by forces due to the asymmetrical magnetic field around the air gap. To determine the forces due to the asymmetrical magnetic field, the magnetic flux density must be determined in the whole air gap region around the rotor. The determination of magnetic forces has been carried out for more than a century.

Early papers by Behrend [1], Gray [2], and Robinson [3] suggested linear equations for magnetic pull. Covo [4] and Ohishi et al. [5] considered the saturation of the magnetic core to improve the magnetic pull equations. Früchtenicht et al. [6] observed the existence of two electromechanical force components (one radial and one tangential) when the rotor performs a circular motion. An analytical model for vibrations in induction motors was presented by Belmans et al. [7,8]. Smith et al. [9] derived analytical equations for unbalanced magnetic pull in induction motors using the air gap permeance approach, including stator and rotor magnetomotive force (MMF) harmonics. Arrkio et al. [10] solved the magnetic field for cage induction motors using time-stepping finite element analysis to numerically determine a linearized electromagnetic force for rotor dynamic analysis. Tenhunen et al. [11–14] have used the same technique to develop the electromechanical force calculations for rotor-dynamical analysis.

Most papers published regarding unbalanced magnetic pull concern asynchronous motors. However, in the area of synchronous generators, there are only a few papers published. Gustavsson et al. suggested a linear model [15] and a nonlinear model [16] for the radial magnetic pull in a hydropower synchronous generator. In these models, the distance between the generator rotor rim and spider hub were also considered. Karlsson et al. [17] presented a linear model for the radial force due to the discrete shape of the rotor and stator. Lundström et al. [18] presented a continuous radial force model due to the rotor and stator shape. Normally, only radial magnetic forces are considered in dynamical models of synchronous machines. Lundström et al. [19] showed the existence of a tangential magnetic force component in synchronous machines due to the damping winding and used a linearized model to analyze the rotor dynamical behavior of these forces. Burakov et al. [20] used the same techniques as Arrkio et al., and Tenhunen et al. [10,14] to determine a low-order parametric force model for a salient-pole synchronous machine. Burakov et al. [21] continued the work and carried out analysis of the influence of parallel connections in the stator winding. Tampion et al. [22] analyzed stator vibrations of a turbo generator stator core due to a reactive load and showed the effect of the reactive load on the vibrations of the stator core. Karlsson et al. [23] concluded that variation of active load, for a synchronous generator, only slightly influenced the electro-mechanical interaction of rotor systems. In this paper, a rotor dynamical analysis for different reactive loads at constant active load for a hydropower generator rotor with eccentricity is carried out.

2 Modeling and Simulation

The mechanical and electromagnetic systems are modeled separately. Initially, simulations are carried out for the electromagnetic field to obtain a force due to static eccentricity. These forces are then used in the mechanical calculations.

2.1 Brief Theory of Unbalanced Magnetic Pull. In the case of static eccentricity the rotor center is displaced relative to the stator center with a distance e (see Fig. 1). In the rotor frame (subscript r in Fig. 1) the air gap length is a function of both angular position, θ, and time, t, according to

$$\delta(\theta,t) = \delta_0(1 - \varepsilon \sin(\theta + \Omega t)) \tag{1}$$

where δ_0 is the nominal air gap length; Ω is the rotational frequency of the rotor; and ε is the relative eccentricity defined as the

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received February 19, 2006; final manuscript received February 8, 2007. Review conducted by Marc P. Mignolet.

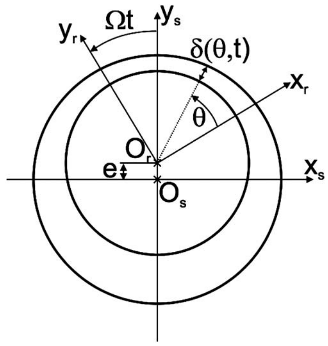


Fig. 1 Static rotor eccentricity

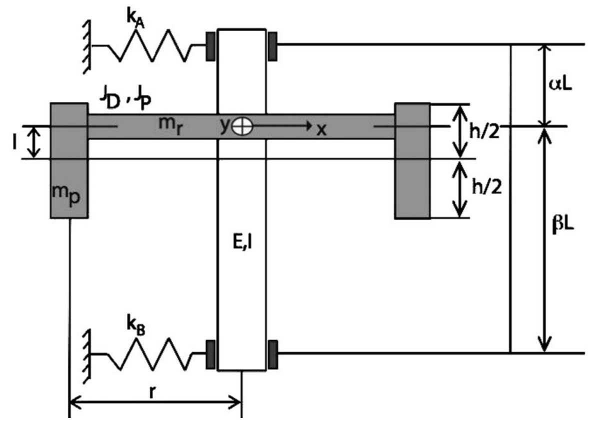


Fig. 2 Rotor geometry

ratio between the absolute eccentricity divided by the nominal air gap length.

In the rotor coordinate system the fundamental magnetomotive force (MMF) component can be expressed as

$$F(\theta) = F_1 \cos(p\theta) \tag{2}$$

where p is the number of pole pairs. For small eccentricities the varying air gap permeance can be expressed as

$$\Lambda(\theta,t) = \frac{\mu_0}{\delta(\theta,t)} = \frac{\mu_0}{\delta_0(1 - \varepsilon \sin(\theta + \Omega t))} \approx \frac{\mu_0}{\delta_0}(1 + \varepsilon \sin(\theta + \Omega t)) \tag{3}$$

where μ_0 is the relative permeability of free space. The air gap flux density of a machine with an eccentric rotor can be expressed as the product of Eqs. (2) and (3) according to

$$B = B_p \sin(p\theta) + B_{p+1} \sin((p+1)\theta + \Omega t) - B_{p-1} \sin((p-1)\theta - \Omega t) \tag{4}$$

where $B_p = F_1/\delta_0$ and $B_{p\pm 1=}\varepsilon B_p/2$. The static eccentricity thus gives rise to two parasitic waves, with pole pair numbers $p \pm 1$, moving relative to the rotor. These parasitic waves will induce currents with frequency Ω in the damper winding. The damper currents do rise to a secondary set of air gap flux density waves, which opposes its origin according to the Lenz law.

2.2 Mechanical System. The mechanical rotor system is modeled as a Stodola–Green rigid disk, connected to a uniform elastic and massless shaft supported by two linear–elastic bearings with bearing stiffnesses k_A and k_B. The shaft has length, L, Young's modulus, E, and area moment of inertia, I. The position of the rotor rim geometrical center (and center of gravity) is in position αL, where α is a nondimensional number $(\alpha < 1)$. The disk represents the physical properties of a generator with the masses of rotor rim, m_r, and rotor poles, m_p, polar moment of inertia, J_P, and diametrical moment of inertia, J_D. The disk spins with an angular velocity, Ω, and has an unbalance, u, in the lateral directions. The geometrical center and center of gravity of the poles are displaced a distance l from the center of the rotor rim. The mechanical model is illustrated in Fig. 2. The disk can translate in two directions, x and y, and rotate around the same. Hence, the rotor system has four degrees of freedom. With matrix notation, the equations of motion are

$$\bar{M}\ddot{\vec{x}} + (\Omega \bar{G} + \bar{C})\dot{\vec{x}} + \bar{K}\vec{x} = \vec{F} \tag{5}$$

where the mass and moments of inertia matrix, gyroscopic matrix, and stiffness matrix are

$$\bar{M} = \begin{pmatrix} m & 0 & 0 & 0 \\ 0 & m & 0 & 0 \\ 0 & 0 & J & 0 \\ 0 & 0 & 0 & J \end{pmatrix}, \quad \bar{G} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & J_P \\ 0 & 0 & -J_P & 0 \end{pmatrix}$$

$$\bar{K} = \begin{pmatrix} k_{11} & 0 & 0 & k_{12} \\ 0 & k_{11} & -k_{12} & 0 \\ 0 & -k_{12} & k_{22} & 0 \\ k_{12} & 0 & 0 & k_{22} \end{pmatrix} \tag{6}$$

where $m = m_r + m_p$ and $J = J_D + m_p l^2$. The stiffness matrix is derived by the inverse of the flexibility matrix

$$\bar{\Theta} = \begin{pmatrix} \Upsilon & 0 & 0 & -\Phi \\ 0 & \Upsilon & \Phi & 0 \\ 0 & \Phi & \Psi & 0 \\ -\Phi & 0 & 0 & \Psi \end{pmatrix} \tag{7}$$

where the coefficients

$$\Upsilon = \frac{L^3\alpha^2\beta^2}{3EI} + \frac{\alpha^2}{k_B} + \frac{\beta^2}{k_A} \tag{8}$$

$$\Phi = \frac{L^2\alpha\beta(\alpha - \beta)}{3EI} - \frac{1}{L}\left(\frac{\alpha}{k_B} - \frac{\beta}{k_A}\right) \tag{9}$$

$$\Psi = \frac{L(1 - 3\alpha\beta)}{3EI} + \frac{1}{L^2}\left(\frac{1}{k_B} + \frac{1}{k_A}\right) \tag{10}$$

are determined using beam theory and $\beta = 1 - \alpha$. The damping matrix due to bearing damping is constructed in the same manner as the bearing part of the stiffness matrix (e.g., c_i is used instead of k_i in Eqs. (8)–(10)), which result in the matrix

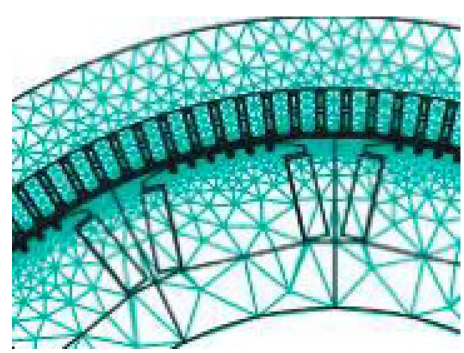


Fig. 3 Part of the computational mesh for the electric field

$$\bar{C} = \begin{pmatrix} c_{11} & 0 & 0 & c_{12} \\ 0 & c_{11} & -c_{12} & 0 \\ 0 & -c_{12} & c_{22} & 0 \\ c_{12} & 0 & 0 & c_{22} \end{pmatrix} \tag{11}$$

The displacement vector is

$$\vec{x} = (x \ y \ \varphi_x \ \varphi_y)^T \tag{12}$$

where x and y are lateral displacements and φ_x and φ_y are rotations around the x and y axis. The unbalanced force vector is

$$\vec{F} = (mu\Omega^2 \cos(\Omega t) mu\Omega^2 \sin(\Omega t) \ 0 \ 0)^T \tag{13}$$

where t is the time.

2.3 Computing the Electromagnetic Field. The electromagnetic field in the generator is solved for a two-dimensional axial cross section. A nonlinear magnetic material with a single-valued magnetization curve for the rotor and stator is used. Coil end impedance is approximately included by circuit equations. The magnetic field is solved through a time-stepping finite element technique for a rotating field. The field winding was supplied from a constant current source. The magnitude of the current was obtained from the stationary model. In the time-varying solution the magnetic field is viewed from a fixed coordinate system in the stator and from a rotating system in the rotor. The time dependence of the rotating field is taken care of by time depending boundary conditions via a line placed in the middle of the air gap between the rotor and the stator. Circuit equations connect the field solution to the coil end impedance and the external load. The voltages, currents, and electromagnetic torque are obtained from combined field and circuit equations simultaneously solved with the time-stepping finite element technique. Additional circuit equations are added to describe the damper winding network. The damper winding forms a continuous cage similar to the squirrel cage in an induction machine except for the fact that the damper bars are embedded only in the salient poles.

Due to the asymmetrical geometry created by the eccentric rotor, the whole generator must be modeled, resulting in extensive computations. A typical mesh contains about 85,000 elements. Figure 3 shows the finite element mesh of the generator, which is more detailed in the areas of special interest, such as the air gap, the stator teeth, and in the damper winding. The mesh is coarser in areas of less importance, such as the stator yoke and the rotor rim, to save computational time.

The two-dimensional model enables the magnetic vector potential to be expressed as

$$\vec{A} = A_z(x,y,t)\vec{z} \tag{14}$$

The magnetic vector potential is solved for every node in the domain according to

$$\sigma \frac{\partial A_z}{\partial t} = \nabla \left(\frac{1}{\mu_0 \mu_r} \nabla A_z \right) - \sigma \frac{\partial V}{\partial z} \tag{15}$$

where σ is the conductivity; $\mu_0\mu_r$ is the magnetic permeability; and $\partial V/\partial z$ is the applied potential. The applied potential is a source term that connects external circuits to the field equations. The magnetic flux density is calculated from the magnetic potential according to

$$\vec{B} = \mathrm{rot}(\vec{A}) \tag{16}$$

Figure 4 shows the magnetic field lines in a synchronous generator obtained from the finite element calculations.

In order to compute the electromagnetic forces, acting between the rotor and the stator, numerically both energy methods or Maxwell's stress tensor can be applied. The methods give the same result for a sufficiently accurate mesh. The forces according to Maxwells stress tensor can be computed as a surface integral in the air gap, where the integration surface is a band placed in the middle of the air gap between the rotor and the stator in the two-dimensional cross section. The surface integral is defined as

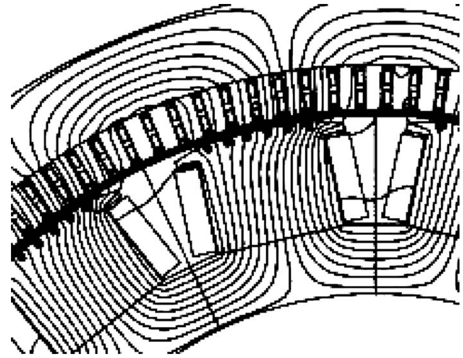


Fig. 4 Magnetic field around one pole obtained from calculations

$$\vec{F} = \frac{1}{D} \int_{S_d} \left[\frac{1}{\mu_0} B_r B_\varphi \vec{\varphi} + \frac{1}{2\mu_0} (B_r^2 - B_\varphi^2) \vec{r} \right] dS \tag{17}$$

where B_r and B_φ are the radial and tangential components of the magnetic flux density; D is the width of the air gap band; and S_d is the cross-sectional area of the air gap band. This method has shown to give an accurate result [24].

Coulomb [25] presented a method to calculate the forces based on the principle of virtual work. The electromagnetic force is defined using the magnetic coenergy

$$W_c = \int_V \left(\int_0^H \vec{B} \, d\vec{H} \right) dV \tag{18}$$

In the two-dimensional model the volume integral becomes a surface integral over the air gap region. The forces are calculated as the derivatives of the coenergy in the air gap, according to

Table 1 Numerical data of Porjus U8

	Item	Value	Unit
m_r	Mass rotorrim	21,000	kg
m_p	Mass rotorpoles	9000	kg
J_D	Diametric moment of inertia	11,406	kg m^2
J_P	Polar moment of inertia	20,000	kg m^2
E	Young's modulus	200	GPa
I	Area moment of inertia	0.0635	m^4
k_A	Bearing stiffness	500	MN/m
k_B	Bearing stiffness	500	MN/m
L	Length of the shaft	3.6	m
κ_1	Damping parameter	0.001	s
κ_2	Damping parameter	3	1/s
Ω	Rotational speed	22.44	rad/s
u	Unbalance	0.28	mm
S	Apparent power	11	MVA
P	Real power	9	MW
V	Voltages	10	kV
$cos(\phi)$	Power factor	0.9	—
n	Number of poles	14	—
f_{el}	Grid frequency	50	Hz
D_{inner}	Stator inner diameter	2.53	m
D_{outer}	Stator outer diameter	3.10	m
h	Rotor length	0.75	m
nd	Number of damper bars per pole	6	—

$$F_x = \frac{\partial W_c}{\partial x}, \quad F_y = \frac{\partial W_c}{\partial y} \tag{19}$$

The forces used in this paper have been calculated by both Maxwell stress tensor and Coulombs method, giving the same result with sufficient accuracy.

2.4 Interaction With the Mechanical System. If the rotor is displaced with a small distance e in the y direction, a linearized force due to the magnetic field can be written as

$$F_y = k_{M,r} e \tag{20}$$

$$F_x = k_{M,t} e \tag{21}$$

where $k_{M,r}$ and $k_{M,t}$ are the radial and tangential magnetic stiffnesses obtained from the electromagnetic calculations. The radial magnetic stiffness is the stiffness in the eccentricity direction, and the tangential magnetic stiffness is the stiffness in the perpendicular direction of the eccentricity due to the magnetic pull. The electromechanical forces and moments, $\overrightarrow{F_M}$, on the rotor will be dependent on the rotor displacement, x and y, the inclination of the rotor, φ_x and φ_y, and the distance, l, between rotor spider hub and the geometrical center of the rotor rim. These forces and moments are obtained by integrating the radial and tangential force distribution over the rotor height, h

$$\overrightarrow{F_M} = \begin{pmatrix} \int_{l-h/2}^{l+h/2} \left(\frac{k_{M,r}x}{h} + \frac{k_{M,t}y}{h} + \frac{k_{M,r}\varphi_y}{h}\zeta - \frac{k_{M,t}\varphi_x}{h}\zeta \right) d\zeta \\ \int_{l-h/2}^{l+h/2} \left(\frac{k_{M,r}y}{h} - \frac{k_{M,t}x}{h} - \frac{k_{M,r}\varphi_x}{h}\zeta - \frac{k_{M,t}\varphi_y}{h}\zeta \right) d\zeta \\ \int_{l-h/2}^{l+h/2} \left(-\frac{k_{M,r}y}{h} + \frac{k_{M,t}x}{h} + \frac{k_{M,r}\varphi_x}{h}\zeta + \frac{k_{M,t}\varphi_y}{h}\zeta \right) \zeta d\zeta \\ \int_{l-h/2}^{l+h/2} \left(\frac{k_{M,r}x}{h} + \frac{k_{M,t}y}{h} + \frac{k_{M,r}\varphi_y}{h}\zeta - \frac{k_{M,t}\varphi_x}{h}\zeta \right) \zeta d\zeta \end{pmatrix}$$

$$= \begin{pmatrix} k_{M,r} & 0 & 0 & lk_{M,r} \\ 0 & k_{M,r} & -lk_{M,r} & 0 \\ 0 & -lk_{M,r} & \Gamma k_{M,r} & 0 \\ lk_{M,r} & 0 & 0 & \Gamma k_{M,r} \end{pmatrix} \begin{pmatrix} x \\ y \\ \varphi_x \\ \varphi_y \end{pmatrix} + \begin{pmatrix} 0 & k_{M,t} & -lk_{M,t} & 0 \\ -k_{M,t} & 0 & 0 & -lk_{M,t} \\ lk_{M,t} & 0 & 0 & \Gamma k_{M,t} \\ 0 & lk_{M,t} & -\Gamma k_{mM,t} & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ \varphi_x \\ \varphi_y \end{pmatrix}$$

$$= (\bar{K}_{m,r} + \bar{K}_{m,t})\vec{x} \tag{22}$$

where

$$\Gamma = \frac{1}{3h}\left(\left(\frac{h}{2} + l \right)^3 + \left(\frac{h}{2} - l \right)^3 \right) \tag{23}$$

The matrices in Eq. (22) is the magnetic stiffness matrices. Adding Eq. (5) to Eq. (22) to derive the equations of motion for the electro-mechanical rotor system

$$\bar{M}\ddot{\vec{x}} + (\Omega\bar{G} + \bar{C})\dot{\vec{x}} + (\bar{K} - \bar{K}_{m,r} - \bar{K}_{m,t})\vec{x} = \vec{F} \tag{24}$$

2.5 Natural Frequencies and Stability. Equation (24) is transferred to a system of first-order equations

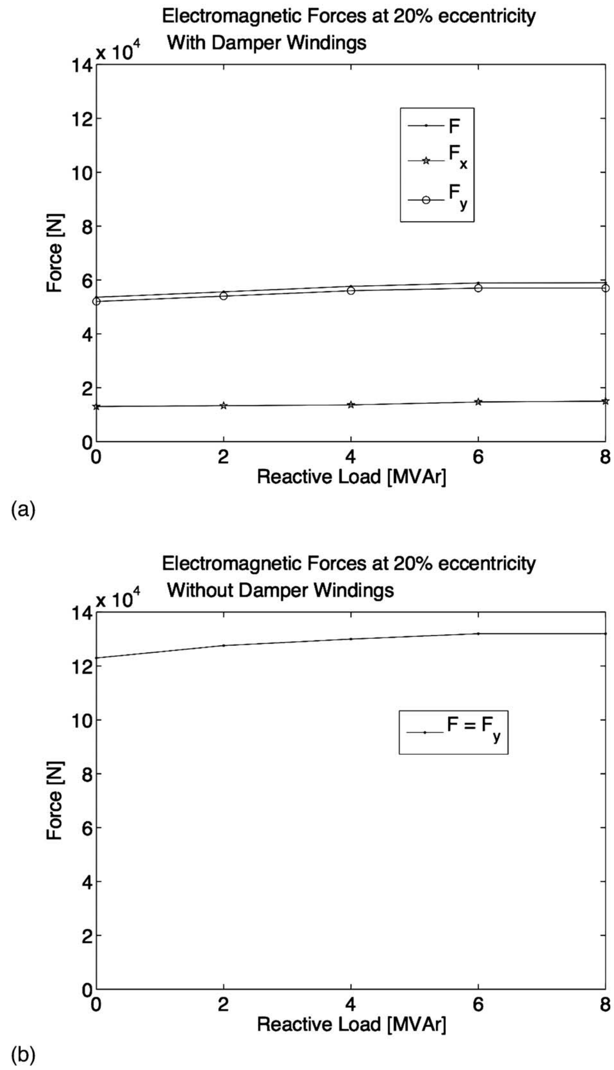


Fig. 5 The radial F_y, tangential F_x, and resulting F force on the rotor for 9 MW active power and varying reactive power, solved with (*a*) and without (*b*) damper winding. (Note that there will not be a tangential contribution to the total unbalance magnetic pull for the calculation without damper.)

$$\begin{pmatrix} -\bar{I} & \bar{N} \\ \bar{N} & \bar{M} \end{pmatrix} \begin{pmatrix} \dot{\vec{x}} \\ \ddot{\vec{x}} \end{pmatrix} + \begin{pmatrix} \bar{N} & \bar{I} \\ \bar{K} - \overline{K_m} & \bar{C} + \Omega\bar{G} \end{pmatrix} \begin{pmatrix} \vec{x} \\ \dot{\vec{x}} \end{pmatrix} = \begin{pmatrix} \vec{B} \\ \vec{F} \end{pmatrix} \tag{25}$$

where $\bar{N}$ and $\vec{B}$ are zero valued matrix vectors, respectively. With the matrix notation, Eq. (25) becomes

$$-\bar{S}\dot{\vec{y}} + \bar{R}\vec{y} = \vec{H} \tag{26}$$

The homogeneous solution to Eq. (26) is

$$\vec{y}_h(t) = \sum_{i=1}^{n} D_i \vec{Y}_i e^{\lambda_i t} \tag{27}$$

where $\bar{Y}_i$ are the eigenvectors corresponding to the eigenvalues λ_i; and D_i are constants determined by initial conditions. The eigenvalues are complex and can be expressed as

$$\lambda_i = \sigma_i + i\omega_{n,i} \tag{28}$$

where σ_i is the decay rate and $\omega_{n,i}$ is the damped eigenfrequency. The system is stable if all eigenvalues have a negative decay rate.

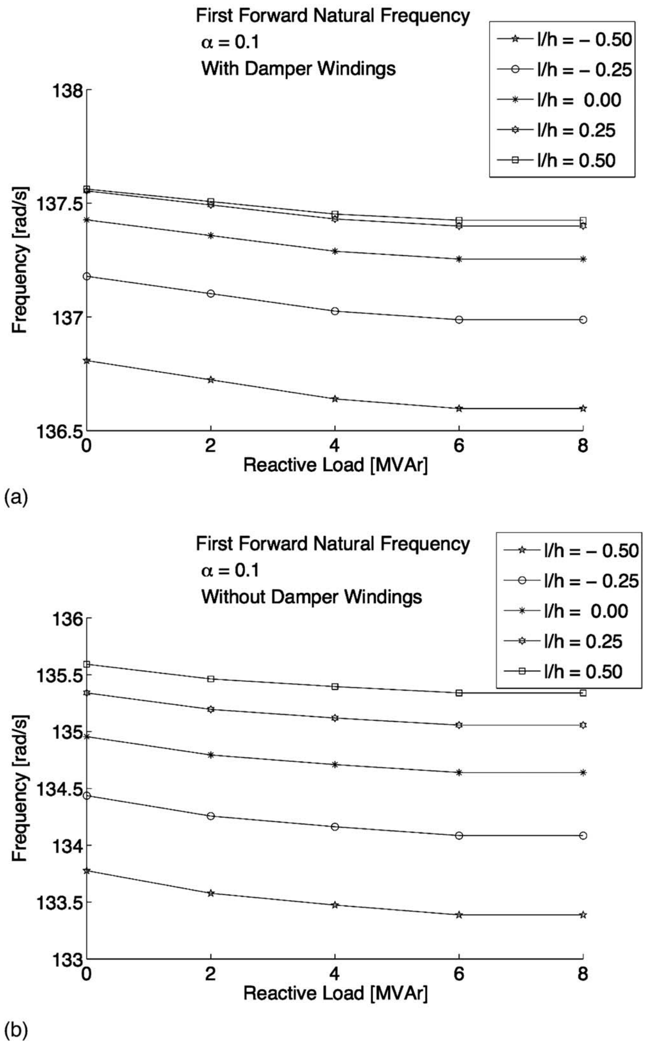


Fig. 6 First forward natural frequency when α=0.1 and *l*/*h*= −0.50, −0.25, 0.00, 0.25, 0.50, for the rotor system for 9 MW active power, varying reactive power and electrical model with (*a*) and without (*b*) damper winding

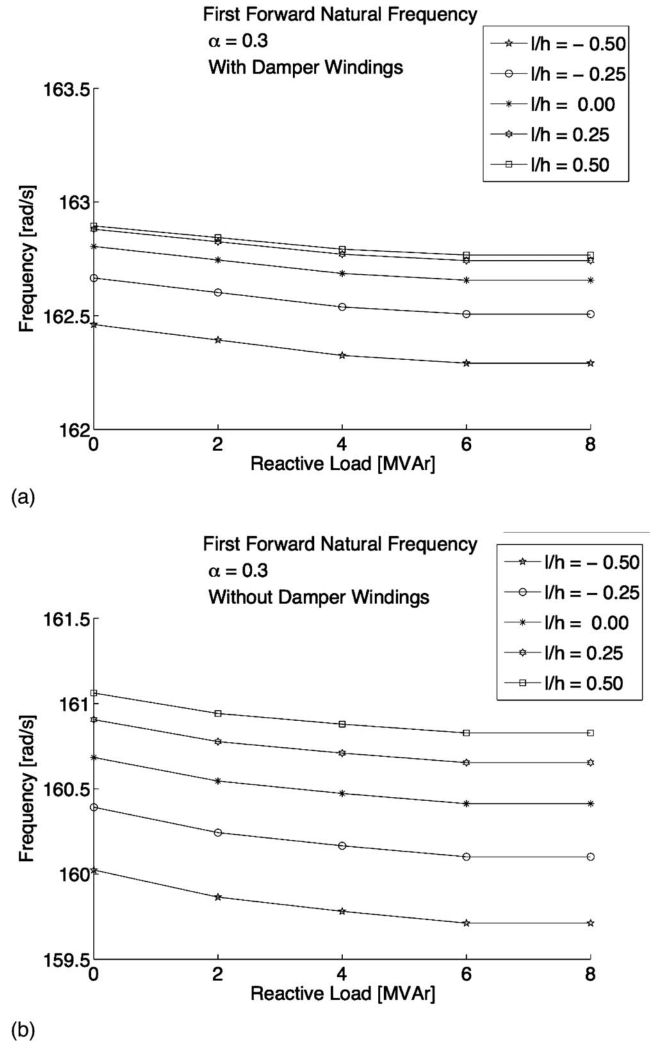


Fig. 7 First forward natural frequency when α=0.3 and *l*/*h*= −0.50, −0.25, 0.00, 0.25, 0.50, for the rotor system for 9 MW active power, varying reactive power and electrical model with (*a*) and without (*b*) damper winding

2.6 Unbalanced Response.

For the case of unbalanced excitation, the particular solution to Eq. (26) becomes

$$\vec{y}_p(t) = \vec{a}\sin(\Omega t) + \vec{b}\cos(\Omega t) \tag{29}$$

with

$$\vec{a} = [\Omega^2 \overline{S}\overline{R}^{-1}\bar{S} + \bar{R}]^{-1}[\overrightarrow{H_s} - \Omega S R^{-1}\overrightarrow{H_c}] \tag{30}$$

$$\vec{b} = \bar{R}^{-1}[\overrightarrow{H_c} + \Omega\bar{S}\vec{a}] \tag{31}$$

$$\overrightarrow{H_s} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & m_G u \Omega^2 & 0 & 0 \end{pmatrix}^T \tag{32}$$

$$\overrightarrow{H_c} = \begin{pmatrix} 0 & 0 & 0 & 0 & m_G u \Omega^2 & 0 & 0 & 0 \end{pmatrix}^T \tag{33}$$

The unbalanced response is analyzed to investigate how the reactive load affects the amplitude of the vibrations.

3 Results

The simulations were carried out for the research and education unit Porjus U8 hydropower generator, located at Porjus International Hydropower Centre, the Municipality of Jokkmokk, in northern Sweden. Numerical data are presented in Table 1.

Computations of the UMP have been carried out for 20% static eccentricity. The rotors have been displaced in the y direction according to Fig. 1. The real power has been kept constant to 9 MW and the rotor excitation has been changed in order to vary the output of reactive power. Calculations have been done with and without the damper winding. The results are presented in Fig. 5.

In the absence of damper winding, only one force component exists and it points in the y direction toward the smallest air gap. This force component is referred to as the radial force component. If the damper winding is taken into account a force component perpendicular to the eccentricity direction in the x direction occurs. This force component is referred to as the tangential force component. The magnitude of the magnetic pull force is reduced and the direction is shifted from the short air gap, to the direction towards the rotation, when the damper winding is considered. The UMP is reduced because of the damper cage reaction, which will counteract the parasitic eccentricity flux harmonics and thus equalize the flux distribution around the air gap circumference according to the reasoning in Sec. 2.1. The unbalanced force created by the eccentricity is almost constant in time, except for a small ripple due to the salient poles representing only about 1% of the steady-state force.

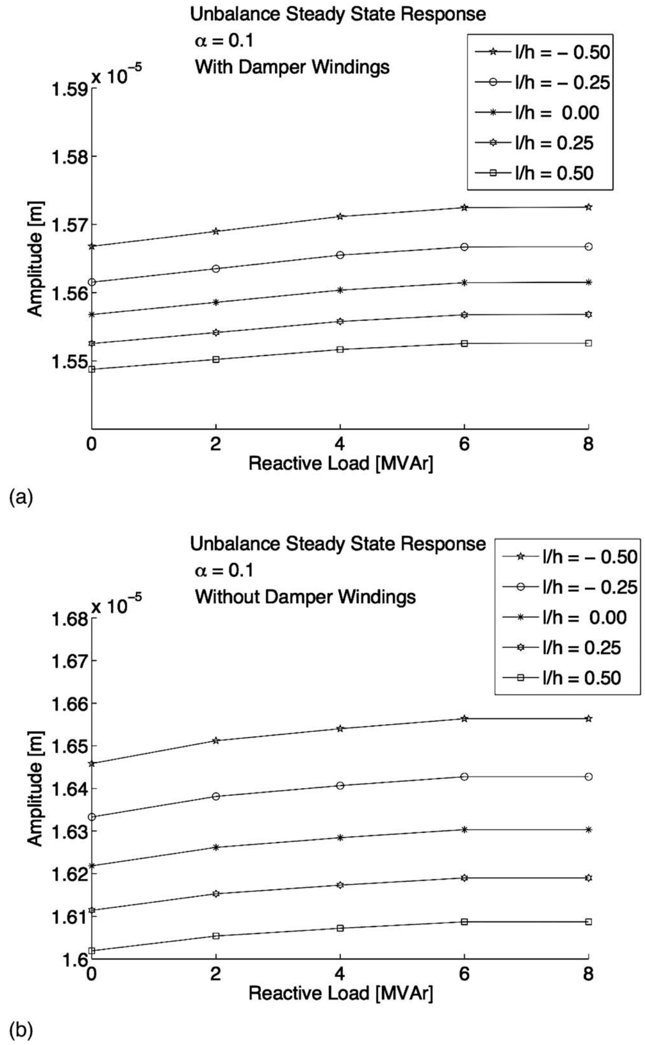


Fig. 8 Unbalance response when α=0.1 and *l*/*h*=−0.50, −0.25, 0.00, 0.25, 0.50, for the rotor system for 9 MW active power, varying reactive power and electrical model with (*a*) and without (*b*) damper winding

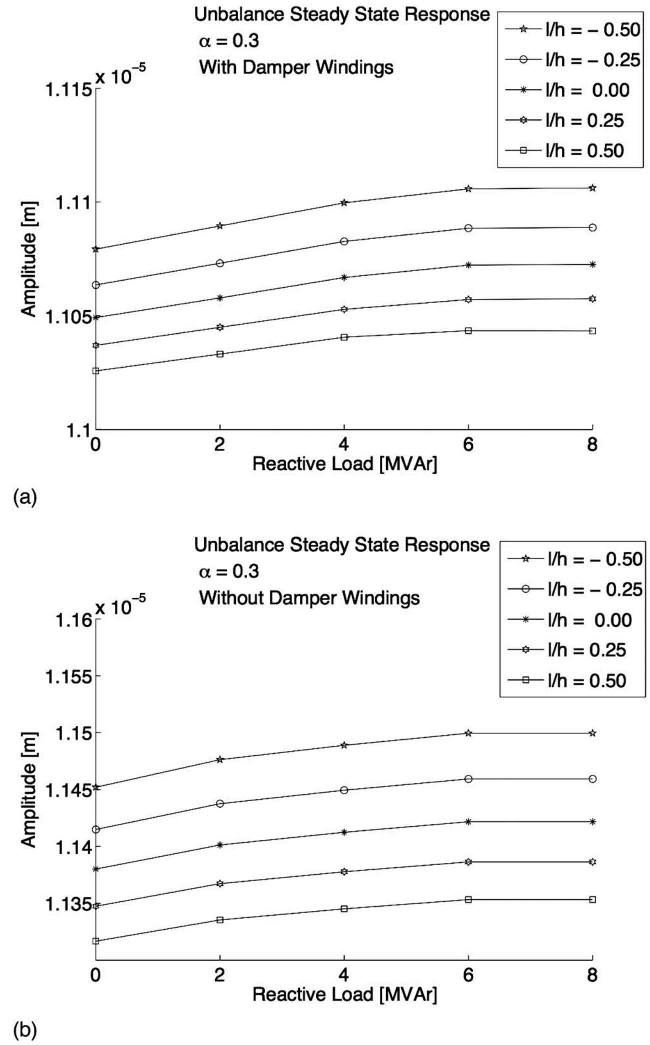


Fig. 9 Unbalance response when α=0.3 and *l*/*h*=−0.50, −0.25, 0.00, 0.25, 0.50, for the rotor system for 9 MW active power, varying reactive power and electrical model with (*a*) and without (*b*) damper winding

The UMP slightly increases when the rotor excitation is increased. The magnitude of the fundamental magnetic flux component, the largest contributor to the unbalanced force, increases with rotor excitation since a larger fundamental is needed to obtain the rated voltage due to the internal voltage drop over the leakage reactance. The increase of the UMP force decays for large rotor excitations due to saturation.

In Figs. 6 and 7, the change of the first forward natural frequencies are presented for different reactive loads. In Figs. 8 and 9, the change of the steady-state response is presented for different reactive loads. The Porjus U8 is stiffer than most commercial hydropower units. To predict how the reactive load will influence a more general case, a stability analysis is carried out for a variation in bearing stiffness. The ratio between bearing stiffness $k_{A,\gamma}$ and $k_{B,\gamma}$ and the nominal bearing stiffness k_A and k_B used is introduced as

$$\gamma = \frac{k_{A,\gamma}}{k_A}, \quad k_{B,\gamma} = k_{A,\gamma} \tag{34}$$

Figures 10 and 11 present the stability regions for the rotor system with different bearing properties and different reactive loads and Figs. 12 and 13 present the stability regions for the rotor system with a variation in mechanical parameters.

4 Discussion

According to Figs. 6–9, one can see that the first forward natural frequencies decrease and the unbalanced responses increase for an increase in reactive load. For the case Porjus U8, the influence of reactive load is relatively small due to a stiff rotor system. For a weaker rotor system the influence of reactive power on the response and natural frequencies will be stronger. According to Figs. 10–13, one can see that the stability region, for simulations with damper winding, decreases with an increase of reactive power, and for simulations without damper winding only a small influence of the reactive load is noted. The radial electromechanical component will result in the negative stiffness matrix and the tangential component will result in the skew-symmetric electro-mechanical matrix. Both these magnetic stiffness matrices will influence the stability of the rotor system. The tangential electro-mechanical force component only exists in the case with damper winding and has a strong influence on the stability of the rotor system.

The simulations are carried out for a generator with a relatively stiff rotor-bearing system. The design is unique for each hydropower generator so there exist generators with a range of stiffnesses of the rotor-bearing system. The tendency of behaviors for

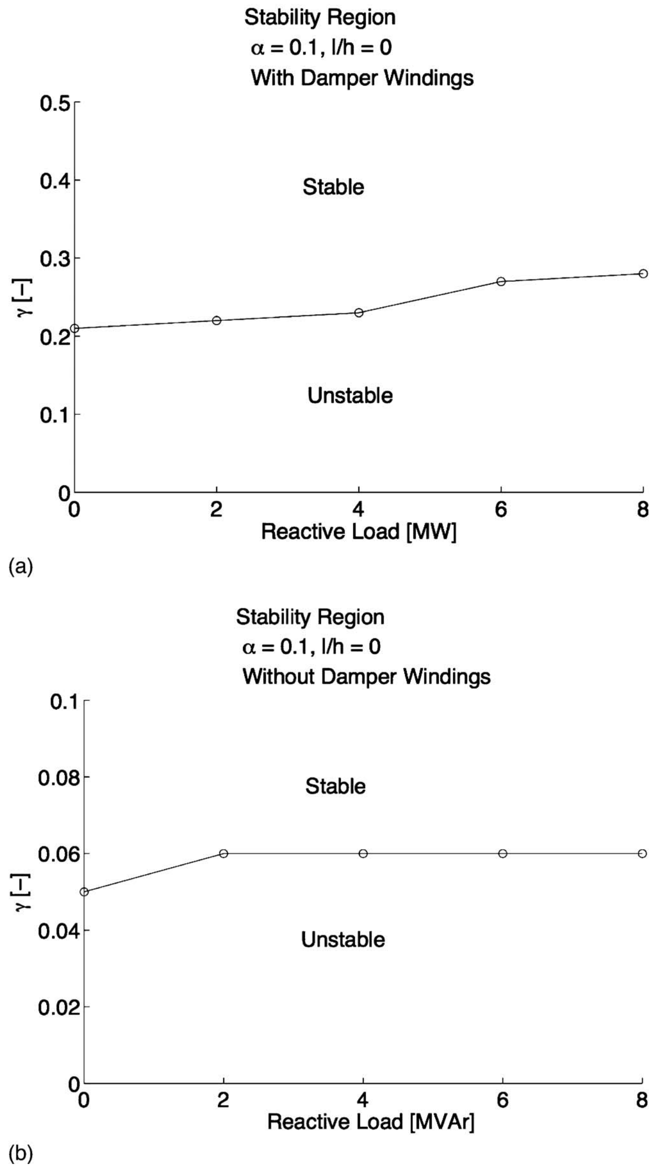


Fig. 10 Stability region of the rotor when α=0.1, *l*/*h*=0, and variation of γ, for simulations with (*a*) and without (*b*) damper windings

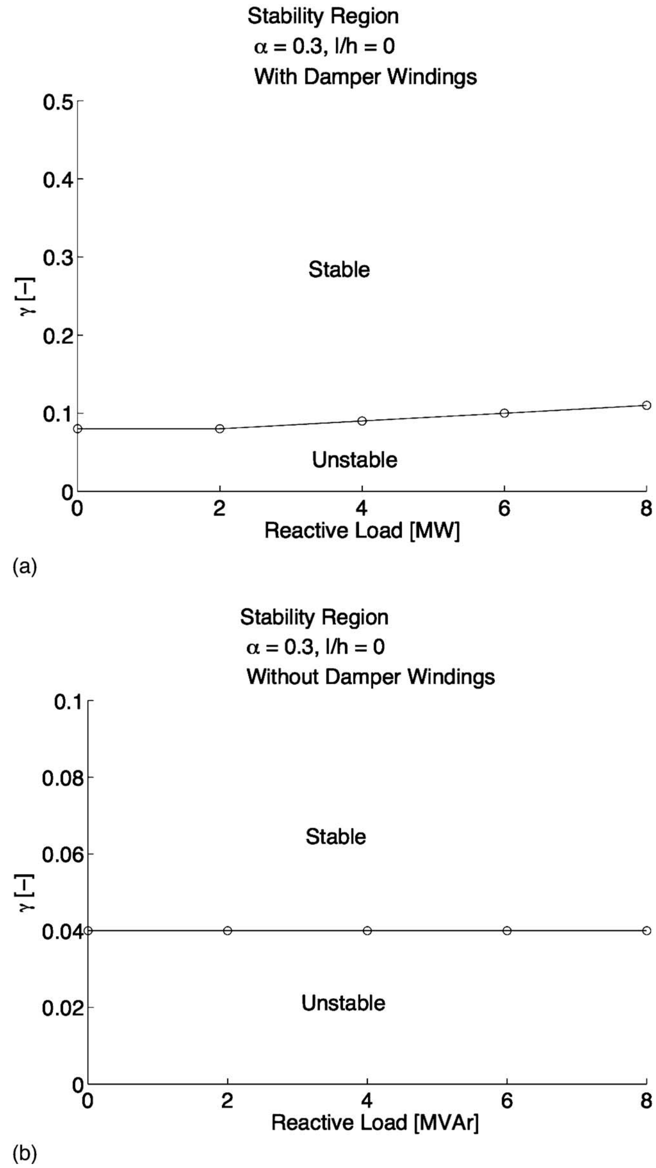


Fig. 11 Stability region of the rotor when α=0.3, *l*/*h*=0, and variation of γ, for simulations with (*a*) and without (*b*) damper windings

different machines is therefore presented by decreasing the bearing stiffness in order to analyze how reactive load will affect a hydroelectric rotor system in general.

In this paper the electro-mechanical interaction is modeled as a linear displacement dependent force. According to the results in Ref. [20] the electro-mechanical force will be dependent on the whirling frequency. The radial component of the electro-mechanical force will increase near the synchronous whirling frequency and the tangential component will decrease. An explanation for this behavior is when the rotor is whirling asynchronously, the damper windings want to force the whirling to synchronously speed. During synchronous whirling there will not be any current in the damper windings and the result is the same UMP as in simulations without damper windings. The computational cost to carry out similar calculations as in Ref. [20] on a 11 MVA 14-pole generator at different loads limited the study in this paper to only be concerned with static eccentricity in the electro-mechanical model. The model does not consider parallel paths of the stator winding, where equalizing currents in the parallel paths have a damping effect of the UMP in electrical machines. In Ref. [26] it is shown for an induction machine that parallel connections in the stator winding have about the same influence on the UMP as the rotor cage, depending on the rotor slot opening width. The rotor system is modeled as a massless shaft, the bearings are assumed linear, and the generator is assumed rigid. The turbine and turbine bearing are not included in the model. The reason for the simplifications used in this paper is that the model is used to predict the tendency of an electro-mechanical phenomenon rather than to simulate a specific unit in detail.

5 Conclusions

In this paper, it has been numerically shown that a change in reactive load will change the electro-mechanical force, thereby influencing the natural frequencies, unbalanced response, and the stability of the system. This implies that changes in the reactive load of a synchronous generator can change the dynamical behavior of the machine.

In this paper, it has been observed that for the case with damper winding compared to the case without damper winding,

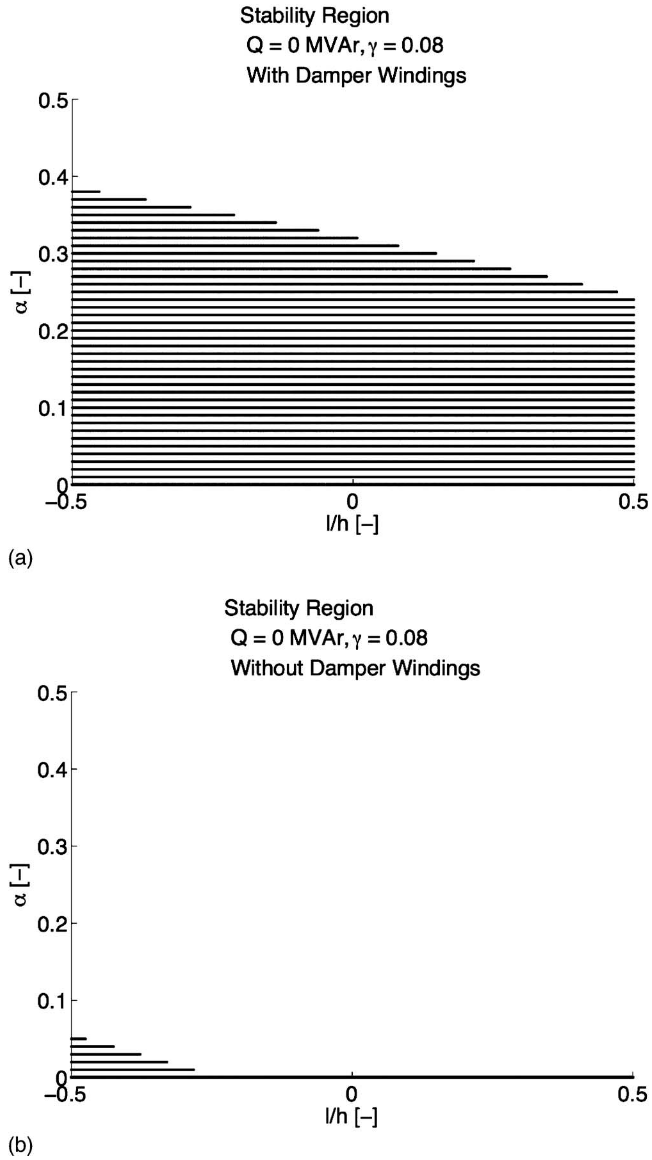


Fig. 12 Stability region of the rotor when Q=0 MVAr, γ=0.08, variation of α and l/h. For simulation with (a) and without (b) damper windings. The stable region is white and the unstable is black.

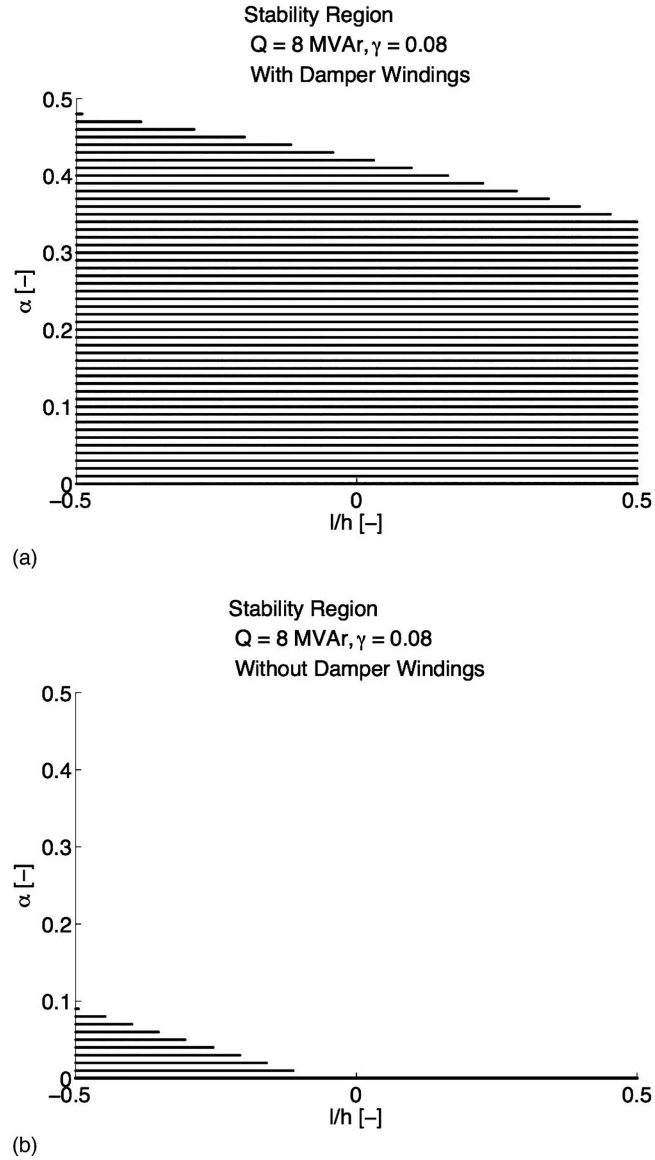


Fig. 13 Stability region of the rotor when Q=8 MVAr, γ=0.08, variation of α and l/h for simulation with (a) and without (b) damper windings. The stable region is white and the unstable is black.

- The first natural frequency increases;
- The unbalance response decreases; and
- The stability region decreases.

One can conclude that the machine will be operating under better conditions, but the stability of the machine might be affected. These observations are practical and it is useful for plant owners and the power utility industry to plan operations and maintenance. In planning production of reactive power one should consider whether the machine is suitable due to the dynamical robustness.

Acknowledgment

Porjus International Hydropower Centre is hereby acknowledged for support regarding data for the Porjus U8 Unit. Elforsk AB and The Swedish Authority of Energy are financing this research via the Elektra program. Dr. Karl-Erik Karlsson and Dr. Arne Wolfbrandt at Uppsala University are acknowledged for developing the in-house code for electromagnetic calculations.

References

[1] Behrend, B. A., 1900, “On the Mechanical Force in Dynamos Caused by Magnetic Attraction,” J. Am. Inst. Electr. Eng., **17**, p. 617.
[2] Gray, A., 1926, “*Electrical Machine Design*, McGraw–Hill, New York, pp. 498–500.
[3] Robinson, R. C., 1943, “The Calculation of Unbalanced Magnetic Pull in Synchronous and Induction Motors,” Electron. Eng. (U.K.), **62**, pp. 620–624.
[4] Covo, A., 1954, “Unbalanced Magnetic Pull in Induction Motors With Eccentric Rotors,” Trans. Am. Inst. Electr. Eng., **73**(III), pp. 1421–1425.
[5] Ohishi, H., Sakabe, S., Tsumagari, K., and Yamashita, K., 1987, “Radial Magnetic Pull in Salient Poles Machines,” IEEE Trans. Energy Convers., EC-2(3).
[6] Früchtenicht, J., Jordan, H., and Seinsch, H. O., 1982, “Exzentrizitätsfelder als Ursache von Laufinstabilitäten bei Asynchronmaschinen. Teil I und II,” Arch. Elektrotech. (Berlin), **65**, pp. 271–292.
[7] Belmans, R., Geysen, W., Jordan, H., and Vandenput, A., 1982, “Unbalance Magnetic Pull and Homopolar Flux in Three Phase Induction Motors With Eccentric Rotors,” *Proceedings, International Conference on Electrical Machines—Design and Application*, July 13–15, Budapest, pp. 916–921.
[8] Belmans, R., Vandenput, A., and Geysen, W., 1987, “Influence of Unbalanced Magnetic Pull on the Radial Stability of Flexible-shaft Induction Machines,” IEE Proc.: Electr. Power Appl., **134**(2), pp. 101–109.
[9] Smith, A. C., and Dorell, D. G., 1996, “Calculation and Measurement of Unbalanced Magnetic Pull in Cage Induction Motors With Eccentric Rotors.

Part 1: Analytical Model," IEE Proc.: Electr. Power Appl., **143**(3), pp. 202–210.
[10] Arrkio, A., Antila, M., Simon, A., and Lantto, E., 2000, "Electromagnetic Force on a Whirling Cage Rotor," IEE Proc.: Electr. Power Appl., **147**(5), pp. 353–360.
[11] Tenhunen, A., and Arrkio, A., 2001, "Modelling of Induction Machines With Skewed Rotor Slots," IEE Proc.: Electr. Power Appl., **148**(1), pp. 45–50.
[12] Tenhunen, A., Benedetti, T., Holopainen, T. P., and Arrkio, A., 2003, "Electromagnetic Force of the Cage Rotor in Conical Whirling Motion," IEE Proc.: Electr. Power Appl., **150**(5), pp. 563–568.
[13] Tenhunen, A., Holopainen, T. P., and Arrkio, A., 2003, "Impulse Method to Calculate the Frequency Response of the Electromagnetic Forces on Whirling Cage Rotors," IEE Proc.: Electr. Power Appl., **150**(6), pp. 752–756.
[14] Tenhunen, A., Holopainen, T. P., and Arrkio, A., 2004, "Effects of Saturation on the Forces of Induction Motors With Whirling Cage Rotor," IEEE Trans. Magn., **40**(2), pp. 766–769.
[15] Gustavsson, R. K., and Aidanpää, J.-O., 2004, "The Influence of Magnetic Pull on the Stability of Generator Rotors," *Proceedings 10th of International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Honolulu, HI, March 7–17.
[16] Gustavsson, R. K., and Aidanpää, J.-O., 2006, "The Influence of Non-linear Magnetic Pull on Hydropower Generator Rotors," J. Sound Vib., **297**(3–5), pp. 551–562.
[17] Karlsson, M., and Aidanpää, J.-O., 2005, "Dynamic Behavior in a Hydropower Rotor System Due to the Influence of Generator Shape and Fuid Dynamics," *Proceeding of PWR2005*, ASME Power, Chicago, IL, April 5–7, pp. 905–913.
[18] Lundström, N., and Aidanpää, J.-O., 2007, "Dynamic Consequences of Electromagnetic Pull due to Deviations in the Generator Shape," J. Sound Vib., **301**(1–2), pp. 207–225.
[19] Lundström, L., Gustavsson, R., Aidanpää, J.-O., Dahlbäck, N., and Leijon, M., 2007, "Influence on the Stability of Generator Rotors due to Radial and Tangential Magnetic Pull Force," IET Electric Power Applications, **1**(1), pp. 1–8.
[20] Burakov, A., and Arkkio, A., 2006, "Low-Order Parametric Force Model for a Salient-pole Synchronous Machine With Eccentric Rotor," Electr. Eng., **89**(2), pp. 127–136.
[21] Burakov, A., and Arkkio, A., 2006, "Low-Order Parametric Force Model for a Eccentric-Rotor Electrical Machine With Parallel Connections in Stator Winding," IEE Proc.: Electr. Power Appl., **153**(4), pp. 592–600.
[22] Tampion, A. A., Stoll, R. L., and Sykulski, J. K., 1991, "Variation of Turbogenerator Stator Core Vibration With Load," IEE Proc.-Commun., **138**(5), pp. 389–400.
[23] Karlsson, M., Perers, R., Gustavsson, R., Aidanpää, J.-O., Karlsson, T., and Leijon, M., 2006, "Rotor Dynamical Analysis of a Hydropower Generator for Active Power," *Proceedings of International Symposium on Water Resources and Renewable Energy Development in Asia*, November 30–December 1, Bangkok.
[24] Arkkio, A., 1987, "Analysis of Induction Motors Based on the Numerical Solution of the Magnetic Feld and Circuit Equations," Acta Polytech. Scand., **59**, pp. 1–97.
[25] Coulomb, J. L., 1989, "Methodology for the Determination of Global Electromechanical Quantities From a Fnite Element Analysis and its Application to the Evaluation of Magnetic Forces, Tourques and Stiffness," IEEE Trans. Magn., **19**(6), pp. 2514–2519.
[26] Tenhunen, A., Holopainen, T. P., and Arkkio, A., 2003, "Effects of Equalizing Currents on Electromagnetic Forces of Whirling Cage Rotor," *Proceedings of IEMDC 03*, Vol. 1, Madison, WI, June 1–4, pp. 257–263.

Stability of the Boiling Two-Phase Flow of a Magnetic Fluid

Jun Ishimoto
Associate Professor
Institute of Fluid Science,
Tohoku University,
Sendai 980-8577, Japan
e-mail: ishimotojun@ieee.org

Elucidation of magnetic stabilization of boiling two-phase flow by utilizing the magnetization of the fluid is proposed herein. The effect of magnetic field on the stability of the boiling two-phase pipe flow of the magnetic fluid under a nonuniform magnetic field is investigated both theoretically and experimentally. First, governing equations of boiling two-phase flow based on the unsteady thermal nonequilibrium two-fluid model are presented and analytically solved using a linearization method. The analytical results on stabilization are then inspected experimentally using an experimental apparatus composed of a small test loop. Results of the analytical study on the void waves, show that the stabilization of two-phase flow can be obtained by practical use of the magnetic body force acting on the fluid and by applying the appropriate superficial gas-phase velocity. Those results also show that magnetic stabilization is obtained because the two-phase magnetic body force enhances the diffusion effect of the void waves. It is experimentally clarified that the two-phase flow state can be stabilized and homogenized by magnetization of the fluid and that vapor bubbles can be minutely produced by effective use of the magnetic body force. The axial magnetic field is more effective for stabilization and homogenization of the two-phase magnetic fluid flow than the transverse magnetic field.
[DOI: 10.1115/1.2723825]

Keywords: magnetic fluid, multiphase flow, stability, boiling, magnetohydrodynamics, liquid metal MHD

1 Introduction

Precise investigation of the stability of a two-phase flow of a magnetic fluid [1,2] is very interesting and important not only as the basic study on hydrodynamics of magnetic fluids, but also for finding solutions to problems related to the development of practical engineering applications of magnetic fluids, such as a new fluid driving system using cavitating flow of magnetic fluid, gas-liquid two-phase flow or boiling two-phase flow of magnetic fluid, which have been proposed by one of the authors [3,4].

The boiling two-phase flow system has the advantage of not requiring high-speed flow for producing the two-phase flow state. Furthermore, with this system, as opposed to the cavitating flow system, a closed flow circulation loop with low fluid velocity can be achieved. The principle of such a fluid driving system is schematically depicted in Fig. 1. In this boiling system, the flow is heated in the region of the negative magnetic field gradient, and boiling nucleation is induced at a point downstream. Furthermore, the flow is additionally accelerated not only by the pumping effect of the vapor bubbles, but also by the rise of magnetic pressure induced by the unbalance of magnetic body forces in the single- and two-phase flow regions under a nonuniform magnetic field. The decrease of magnetic body force in the two-phase region is caused by (1) a decrease of apparent magnetization due to vapor bubble inclusion, and (2) a decrease of magnetization due to the temperature increase. In order to utilize the temperature sensitivity in magnetization of the fluid more effectively, the working fluid herein assumed is hexane-based temperature sensitive magnetic fluid with dispersed Mn–Zn ferrite particles. Therefore, this system is regarded as an energy conversion system by which thermal energy can be converted to fluid kinetic (driving) energy.

The idea of using a two-phase flow system originated from the two-phase liquid-metal MHD power generation system proposed and developed by Petrick and Branover [5]. Subsequent to their proposal, we reported the results of a theoretical study which demonstrated the possibility of using an electrically conducting magnetic fluid [6–10] as a working fluid in a boiling two-phase liquid-metal MHD (LMMHD) power generation system [11]. Our results indicated that a better driving force or pressure rise than that of the conventional system could be obtained by using an electrically conducting magnetic fluid as the working fluid due to the advantage of the practical application of fluid magnetization. These previous studies indicate that high performance of a power generation system is possible by the application of an electrically conducting magnetic fluid to the working fluid in the two-phase LMMHD [12,13] power generation system. It was also predicted that stabilization of two-phase flow is closely related to the driving performance as well as to the development of such energy conversion systems. Thus, it is necessary to precisely analyze the effect of a magnetic field on the stabilization of the boiling two-phase magnetic fluid flow. Furthermore, presented stability analysis does not only contribute to the development of the two-phase LMMHD power generation systems, but also contribute to the stability improvement for any kind of fluid driving system or fluid acceleration system which utilizes the boiling two-phase flow method.

The idea of magnetic stabilization of two-phase flow originated from research on the magnetically stabilized bed (MSB) [14–18]. Rosensweig [1,2] analyzed the hydrodynamic stability in the state of uniform fluidization using equations of motion in conjunction with expressions for magnetic stress. In spite of the progress in research on MSB, significant results on the magnetic stabilization of the boiling two-phase flow of magnetic fluid have not been obtained to date. In the application of boiling two-phase flow of magnetic fluid to an actual energy conversion apparatus, it is important to analyze the two-phase flow stability so as to improve the total performance of a fluid circulation system using multiphase flow. However, there have been no precise investigations of the two-phase stabilization problem. One of the difficulties in such a study is that the basic equations for theoretical analysis of the stability of two-phase flow have not been established. Further-

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received February 24, 2006; final manuscript received February 5, 2007. Review conducted by Bassam A. Younis.

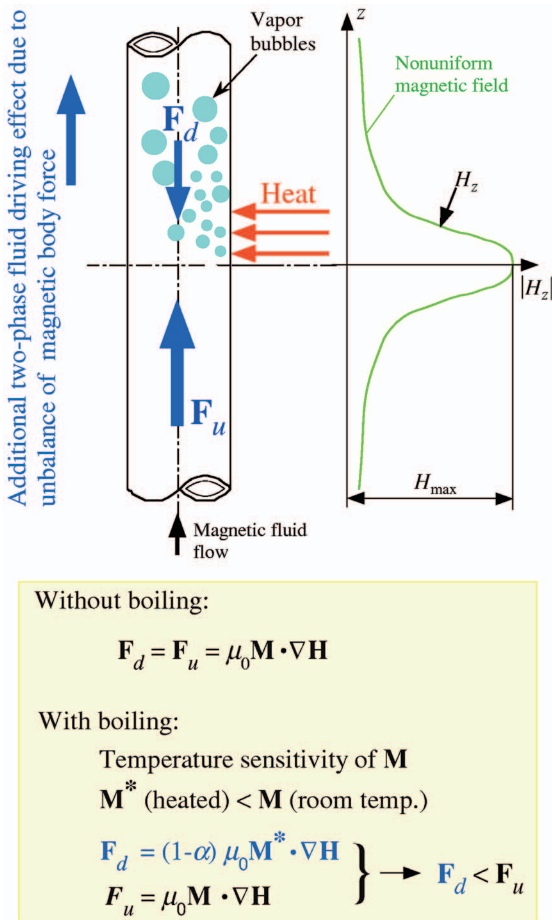


Fig. 1 Principle of two-phase energy conversion system using boiling two-phase flow of magnetic fluid. Magnetic body force $F_u=\mu_0 M\cdot\nabla H=F_d$ in the case without boiling, and $F_d=(1-\alpha)\mu_0 M^*\cdot\nabla H<F_u$ in the case with boiling.

more, experimental confirmation due to the opaqueness of the fluid and the strong effect of applied magnetic fields on the measuring devices is difficult [3].

To overcome these difficulties, we construct a new model for theoretical analysis of boiling two-phase magnetic fluid flow which is based on the unsteady thermal nonequilibrium two-fluid model [19,20]. We also develop a special measurement technique to investigate the stability of the two-phase flow using a high-response pressure transducer and flow visualization technique with an image processor.

In the present paper, theoretical and experimental studies are made to clarify the effects of a nonuniform magnetic field on the stability of the boiling two-phase pipe flow of magnetic fluid. To clarify the effect of magnetic field on the stability of such flow, governing equations of boiling two-phase flow based on the unsteady thermal nonequilibrium two-fluid model [19,20], which takes the effect of a two-phase magnetic body force acting on the boiling magnetic fluid flow state into consideration, are proposed and analytically solved by using the linearization method [1,14]. Additionally, the analytical results on magnetic stabilization are inspected experimentally by using an experimental apparatus composed of a small test loop.

2 Theoretical Study

To clarify the stability of the two-phase flow, it is important to investigate the mechanism of the generation of an unstable flow state. In the case of boiling two-phase flow, the main factor for the generation of the unstable flow state is considered to the rapid change and transition of the flow pattern. We mainly analyzed the stability and the dispersion relation of the void fraction perturbation because the transition of the flow state is closely related to the stability, growth, and decay of the void fraction perturbation in the boiling two-phase flow.

2.1 Governing Equations. The following assumptions are employed to formulate the governing equations:

1. The state of the two-phase flow is one-dimensional unsteady laminar flow.
2. The magnitude of small disturbance waves is sufficiently large compared to the bubble diameter.
3. The thermal nonequilibrium between gas and liquid phases is considered.
4. The direction of the magnetization vector is instantaneously tracked to that of the magnetic field vector.

Under the above conditions, the governing equations of the boiling two-phase magnetic fluid flow, taking into account the effect of nonuniform magnetic field based on the unsteady two-fluid model, are derived as follows.

The mass conservation equation for the gas and liquid phases is

$$\frac{\partial}{\partial t}(\rho_k\alpha_k)+\frac{\partial}{\partial z}(\rho_k\alpha_k v_k)=\Gamma_k \tag{1}$$

where the subscript k denotes the gas phase ($k=g$) or liquid phase ($k=l$). t is the time, α_g and α_l are the gas- and liquid-phase volume fractions, respectively, ρ_g and ρ_l are the gas and liquid-phase densities, respectively. The relationship $(\alpha_g+\alpha_l=1)$ is assumed. Γ_g and Γ_l are the gas and liquid-phase generation densities, respectively.

The combined equation of motion for the total gas and liquid phase is

$$\frac{\partial}{\partial t}[\rho_g\alpha_g v_g+\rho_l(1-\alpha)v_l]+\frac{\partial}{\partial z}[\rho_g\alpha v_g^2+\rho_l(1-\alpha)v_l^2] = -\frac{\partial p_l}{\partial z}-(1-\alpha)\rho_l g+F_{mTz}+\Gamma_g(v_g^{(i)}-v_l^{(i)}) - \frac{32[\eta_g\alpha v_g+\eta_l(1-\alpha)v_l]}{D^2} \tag{2}$$

where F_{mTz} represents the magnetic body force term in two-phase flow which is precisely explained in the next section. The superscript (i) denotes the interface between the gas and liquid phases. To consider the effects of slip and radial expansion of the bubbles, the equation of motion for the gas phase is here replaced with the translational motion of a single bubble [21].

The equation of motion for the gas phase is

$$\frac{4}{3}\pi\bar{R}^3\rho_g\left(\frac{\partial v_g}{\partial t}+v_g\frac{\partial v_g}{\partial z}\right)=-\frac{4}{3}\pi\bar{R}^3\frac{\partial p_l}{\partial z}-\frac{4}{3}\pi\bar{R}^3\rho_g g-F_D-F_{\mathrm{VM}} \tag{3}$$

where $\bar{R}$ is the mean radius of bubbles defined by the following Eq. (9), F_D is the drag force, and F_{VM} is the virtual mass force considering the expansion of a bubble [21,22].

The energy equation for the gas phase is

$$\frac{\partial}{\partial t}(\rho_g\alpha v_g^2)+\frac{\partial}{\partial z}\left[\rho_g\alpha v_g\left(h_g+\frac{v_g^2}{2}\right)\right]=-\rho_g\alpha v_g g+\Gamma_g h_g^{(i)}+q_g^{(i)}a^{(\mathrm{i})} \tag{4}$$

The energy equation for the liquid phases is

$$\frac{\partial}{\partial t}[\rho_l(1-\alpha)v_l^2] + \frac{\partial}{\partial z}\left[\rho_l(1-\alpha)v_l\left(h_l + \frac{v_l^2}{2}\right)\right]$$
$$= -\rho_l(1-\alpha)v_l g + \Gamma_l h_l^{(i)} + q_l^{(i)} a^{(\mathrm{i})}$$
$$-(1-\alpha)\mu_0 T_l\left(\frac{\partial M_l}{\partial T_l}\right)_{H_T} v_l \frac{\partial H_T}{\partial z} + Q_w \tag{5}$$

where the fourth term on the rhs is the magnetocaloric effect, Q_w is the heat transfer rate per unit volume. $h_g^{(i)}$ and $h_l^{(i)}$ are the enthalpy of the gas-phase and that of the liquid-phase at the interface, respectively. $a^{(\mathrm{i})}$ is the gas-liquid interfacial area concentration per unit volume. $\Gamma_g h_g^{(i)}$ and $\Gamma_l h_l^{(i)}$ are the interfacial energy transfer terms due to the liquid-vapor phase change. $q_g^{(i)}$ and $q_l^{(i)}$ are the heat transfer terms of mutual interaction between the vapor and liquid interface.

It is assumed that energy transfer is caused by the heat transfer between an isothermal spherical bubble and the surrounding liquid. Assuming a spherical bubble with equivalent radius $\bar{R}$, the expression of $a^{(\mathrm{i})}$ is obtained by the following equation:

$$a^{(\mathrm{i})} = \frac{3\alpha}{\bar{R}} \tag{6}$$

Assuming that the vapor gas phase follows the ideal gas law, the equation of state for gas-phase is derived as

$$p_g = \mathfrak{R} T_g \rho_g \tag{7}$$

where $\mathfrak{R}$ is the gas constant.

The equation of the expansion and contraction of a gas-bubble [22] is as follows:

$$\bar{R}\left(\frac{\partial^2\bar{R}}{\partial t^2} + \frac{\partial v_g}{\partial t}\frac{\partial \bar{R}}{\partial z} + 2v_g\frac{\partial^2\bar{R}}{\partial t\partial z} + v_g\frac{\partial v_g}{\partial z}\frac{\partial\bar{R}}{\partial z} + v_g^2\frac{\partial^2\bar{R}}{\partial z^2}\right)$$
$$+\frac{3}{2}\left(\frac{\partial\bar{R}}{\partial t} + v_g\frac{\partial\bar{R}}{\partial z}\right)^2$$
$$= \frac{1}{\rho_l}\left[p_g - p_l - \frac{2\gamma_l}{\bar{R}} - \frac{4\eta_l}{\bar{R}}\left(\frac{\partial\bar{R}}{\partial t} + v_g\frac{\partial\bar{R}}{\partial z}\right)\right] \tag{8}$$

where pressure p_l in Eqs. (3) and (8) includes the effect of the two-phase magnetic body force $\mathbf{F}_{mT}$ which will be appear in Eq. (17) in the next section. Therefore, the gas-phase equations implicitly include the effect of two-phase magnetic body force.

The mass conservation equation for single gas bubble [22] is

$$\frac{4}{3}\pi\left[\frac{\partial}{\partial t}(\rho_g\bar{R}^3) + v_g\frac{\partial}{\partial z}(\rho_g\bar{R}^3)\right] = \frac{\partial}{\partial t}\left(\frac{\sum\Gamma_g}{\sum N_g}\right) + v_g\frac{\partial}{\partial z}\left(\frac{\sum\Gamma_g}{\sum N_g}\right) \tag{9}$$

where N_g is the number density of the generated vapor bubbles.

Maxwell's equation in two-phase flow [1] is

$$\nabla\cdot\mathbf{B}_T = 0$$

$$\nabla\times\mathbf{H}_T = \mathbf{0}$$

$$\mathbf{B}_T = \mu_0(\mathbf{H}_T + \mathbf{M}_T) \tag{10}$$

where $\mathbf{B}$ is the vector of magnetic flux density, $\mathbf{H}$ is the vector of magnetic field, and $\mathbf{M}$ is the vector of magnetization. The subscript T denotes two-phase flow. $\mathbf{B}_T$, $\mathbf{H}_T$, and $\mathbf{M}_T$ are defined by following expressions:

$$\mathbf{B}_T = \alpha\mathbf{B}_g + (1-\alpha)\mathbf{B}_l$$

$$\mathbf{H}_T = \alpha\mathbf{H}_g + (1-\alpha)\mathbf{H}_l$$

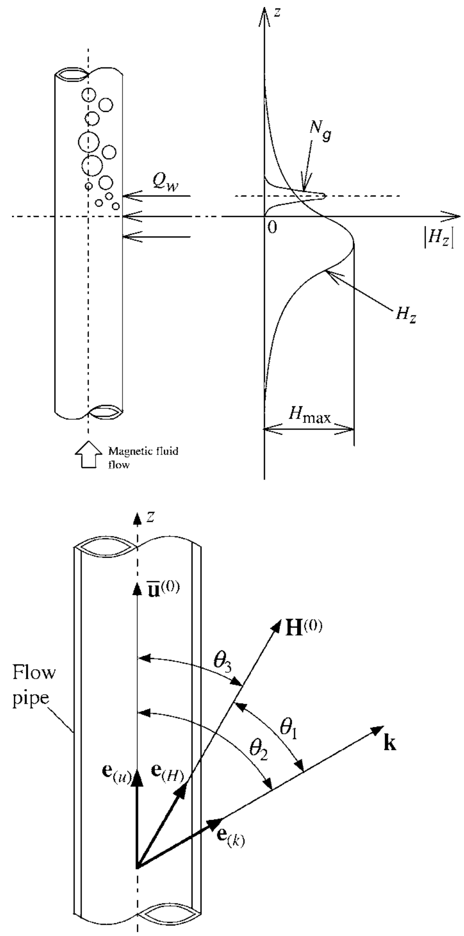


Fig. 2 Schematic of theoretical system (analytical model and nomenclature for stability analysis)

$$\mathbf{M}_T = (1-\alpha)\mathbf{M}_l \tag{11}$$

where in the two-phase flow state, we applied the simplest assumption on $\mathbf{M}$, namely, that the magnetization in the two-phase flow $\mathbf{M}_T$ is only influenced by the liquid-phase volume fraction $(1-\alpha)$ because magnitude of the magnetization or magnetic susceptibility in gas-phase is much smaller than that in the liquid-phase.

2.2 Method for Linear Stability Analysis. Figure 2 schematically shows that the system used in the theoretical analysis referring to the experimental system in the next section. The present analysis extends an analytical method developed by Anderson and Jackson [14] for nonmagnetized fluidized beds and that by Rosensweig [1,2] for magnetized fluidized beds. For the sake of simplicity, regarding the analyzed system, it is assumed that the direction of the mainstream is upward, opposite the gravitational force. A disturbance wave propagates through the system in the specified direction by its wave vector $\mathbf{k}$, which is oriented at an angle θ_2 relative to the direction of flow. A nonuniform applied magnetic field $\mathbf{H}$ is imposed on the system at an arbitrary orientation, specified by angle θ_1 relative to the direction of wave propagation. $\mathbf{e}_{(H)}$, $\mathbf{e}_{(u)}$, and $\mathbf{e}_{(k)}$ are the unit vectors in the direction of magnetic, stream, and wave number, respectively. Unknown variables in the basic equations are written as the sum of the value in the equilibrium state and a small perturbation, and they are derived as follows:

$$\begin{cases} \alpha = \alpha^{(0)} + \tilde{\alpha}, \quad p_l = p_l^{(0)} + \tilde{p}_l \\ p_g = p_g^{(0)} + \tilde{p}_g, \quad v_l = v_l^{(0)} + \tilde{v}_l \\ v_g = v_g^{(0)} + \tilde{v}_g, \quad T_l = T_l^{(0)} + \tilde{T}_l \\ T_g = T_g^{(0)} + \tilde{T}_g, \quad \rho_g = \rho_g^{(0)} + \tilde{\rho}_g \\ \mathbf{H}_T = \mathbf{H}_T^{(0)} + \tilde{\mathbf{H}}_T = \mathbf{e}_{(H)} H_T^{(0)} + \tilde{\mathbf{H}}_T \\ \mathbf{M}_T = \mathbf{M}_T^{(0)} + \tilde{\mathbf{M}}_T = \mathbf{e}_{(H)} M_T^{(0)} + \tilde{\mathbf{M}}_T \end{cases} \tag{12}$$

where superscripts (0) and ~ denote the equilibrium state and perturbation from the equilibrium state, respectively. The variables, including the small perturbation, are substituted into the basic Eqs. (1)–(11). The basic equations are linearized by neglecting the terms of greater than second order in small perturbation. In this manner, a set of linear partial differential equations for the perturbation is obtained. Next, the void fraction perturbation and the magnetic field perturbation are assumed to be represented by the plane wave solution which is derived as follows:

$$\begin{cases} \tilde{\alpha} = \hat{\tilde{\alpha}} W \\ \tilde{\mathbf{H}}_T = \hat{\tilde{\mathbf{H}}}_T W \end{cases} \tag{13}$$

where

$$W = \exp(\omega_c t)\exp(i\mathbf{k} \cdot \mathbf{x}) \tag{14}$$

with $\hat{\tilde{\alpha}}$ being the amplitude of the void fraction perturbation, $\hat{\tilde{\mathbf{H}}}$ is the constant vector of the magnetic field, $\mathbf{k}$ is the wave number vector of the plane wave disturbance, and $\mathbf{x}$ is the position vector. ω_c denotes the complex frequency and is defined as

$$\omega_c = \omega_r - i\omega_i \tag{15}$$

where the real part ω_r of ω_c determines the rate of growth or decay of the wave with time. If ω_r is positive, the disturbance grows and the flow state is unstable, whereas if ω_r is negative, the disturbance decays and the flow state is stable. Thus, ω_r is termed the growth factor. Next, in order to linearize the magnetic body force term considering the effect of nonuniform magnetic field, the magnetic field distribution H_T is derived as the following equation. In introducing H_T, we referred to the analytical solution of the magnetic field distribution of a Helmholtz coil [23] and the measurement results of magnetic field of the electromagnet used in the present experimental study,

$$\begin{cases} H_T = |\mathbf{H}_T^{(0)} + \tilde{\mathbf{H}}_T| C_{(H)} = (H_T^{(0)} + \tilde{\mathbf{H}}_T \cdot \mathbf{e}_{(H)}) C_{(H)} \\ C_{(H)} = \exp\left\{ -\left[a_{(H)} \left(\dfrac{z + b_{(H)}}{D} \right) \right]^2 \right\} \\ H_T^{(0)} \approx H_{\max} \end{cases}$$

$$a_{(H)} = \begin{cases} 0.1389 & \text{(in the axial magnetic field)} \\ 0.370 & \text{(in the transverse magnetic field)} \end{cases}$$

$$b_{(H)} = \begin{cases} 0.021 & \text{(in the axial magnetic field)} \\ 0 & \text{(in the transverse magnetic field)} \end{cases} \tag{16}$$

where $a_{(H)}$ and $b_{(H)}$ in (16) are the empirical coefficients concerned with the magnetic field profiles of the electromagnet and the relative position between maximum field strength and heating area in the present experimental apparatus. According to Eq. (16) and linearized Maxwell's equations, the linearized magnetic body force term in two-phase flow, taking into account the effects of nonuniform magnetic field, is derived as the following equation:

$$\begin{aligned} \mathbf{F}_{mT} = \mu_0 \mathbf{M}_T \cdot \nabla \mathbf{H}_T = \mu_0 (1 - \alpha^{(0)}) M_l^{(0)} [H_T^{(0)} \nabla C_{(H)} \\ + C_{(M)} C_{(H)} M_l^{(0)} \nabla \tilde{\alpha} + C_{(M)} M_l^{(0)} \tilde{\alpha} \nabla C_{(H)}] \end{aligned} \tag{17}$$

where μ_0 is the permeability in a vacuum, $C_{(M)}$ is defined as

$$C_{(M)} = \frac{\cos^2 \theta_1}{1 + (1 - \alpha^{(0)})(1 - \cos^2 \theta_1)\chi^{(0)} + \hat{\chi}(1 - \alpha^{(0)})\cos^2 \theta_1} \tag{18}$$

and θ_1 is defined as

$$\theta_1 = \cos^{-1}(\mathbf{e}_{(H)} \cdot \mathbf{e}_{(k)}) \tag{19}$$

where $\mathbf{e}_{(H)}$ and $\mathbf{e}_{(k)}$ are the unit vector in the direction of $\mathbf{H}$ and $\mathbf{k}$, respectively. The tangent susceptibility $\hat{\chi}$ and chord susceptibility $\chi^{(0)}$ are defined as

$$\hat{\chi} = \left(\frac{\partial M_l^{(0)}}{\partial H_T} \right)_{H_T^{(0)}}$$

$$\chi^{(0)} = \frac{M^{(0)}}{H_T^{(0)}} \tag{20}$$

It is now possible to formulate a partial-differential equation in the void fraction perturbation $\tilde{\alpha}$. Equation (17) is substituted into the linearized equation of motion, and the divergence of the linearized basic equations are computed. After this analysis, by eliminating the unknown perturbation values except for $\tilde{\alpha}$, the partial-differential equation in the void fraction perturbation $\tilde{\alpha}$ is derived as

$$A_1 \frac{\partial^2 \tilde{\alpha}}{\partial t^2} + C_1 \frac{\partial \tilde{\alpha}}{\partial z} + D_1 \frac{\partial \tilde{\alpha}}{\partial t} + (E_1 + F_1) \frac{\partial^2 \tilde{\alpha}}{\partial z^2} = 0 \tag{21}$$

where the coefficients A_1, C_1, D_1, E_1, and F_1 do not directly include the perturbation values and are expressed as follows:

$$A_1 = \frac{\rho_g^{(0)}}{\rho_l} + 1 - \frac{\rho_g^{(0)} v_g^{(0)}}{\rho_l} \tag{22}$$

$$C_1 = \frac{g}{\alpha^{(0)}} C_{\text{cof}} \tag{23}$$

$$\begin{aligned} C_{\text{cof}} = & -\alpha^{(0)} + \frac{\Gamma_g}{\rho_l} \left[\frac{v_g^{(0)}}{g} + \frac{\alpha^{(0)} v_l^{(0)}}{(1 - \alpha^{(0)}) g} \right] - \frac{32 \eta_g \alpha^{(0)} v_l^{(0)}}{\rho_l g D^2 (1 - \alpha^{(0)})} \\ & + \frac{3 C_D \pi}{4 \pi \bar{R}^{(0)}} \left[\left(\frac{v_g^{(0)}}{g} + \frac{\alpha^{(0)} v_l^{(0)}}{(1 - \alpha^{(0)}) g} \right) (v_g^{(0)} - v_l^{(0)}) \right] + c_{m1} \\ & + c_\rho \left\{ \frac{\Gamma_l}{\rho_l} \left(-\frac{\Gamma_g}{\rho_g^{(0)} g} + \frac{2 \alpha^{(0)} v_g^{(0)}}{\rho_g^{(0)} g} c_R + \frac{\alpha^{(0)} \Gamma_l}{(1 - \alpha^{(0)}) \rho_l} \right) \right. \\ & + \frac{32}{\rho_l D^2} \left[\frac{\eta_g \alpha^{(0)^2}}{g} \left(\frac{\Gamma_g}{\rho_g^{(0)} \alpha^{(0)}} - \frac{2 v_g^{(0)}}{\rho_g^{(0)}} c_R \right) + \frac{\eta_l \alpha^{(0)} \Gamma_l}{\rho_l g} \right] \\ & - \frac{3 C_D \pi \alpha^{(0)}}{4 \pi \bar{R}^{(0)} g} \left[\left(\frac{\Gamma_g}{\rho_g^{(0)} \alpha^{(0)}} - \frac{2 v_g^{(0)}}{\rho_g^{(0)}} c_R \right) (v_g^{(0)} - v_l^{(0)}) \right. \\ & \left. - \frac{v_g^{(0)} \alpha^{(0)} \Gamma_l}{(1 - \alpha^{(0)}) \rho_l} \right] - \frac{\alpha^{(0)}}{\rho_l} c_R - \frac{3 \alpha^{(0)} C_D}{4 \bar{R}^{(0)^2} g} (v_g^{(0)^2} \\ & - 2 v_g^{(0)} v_l^{(0)} + v_l^{(0)^2}) \left. \frac{\partial \tilde{\bar{R}}}{\partial z} \right|_{z = z_l} \\ & \left. - \frac{3 C_{VM} \alpha^{(0)}}{\bar{R}^{(0)} g} (v_g^{(0)^2} - v_l^{(0)} v_g^{(0)}) \left. \frac{\partial^2 \tilde{\bar{R}}}{\partial z^2} \right|_{z = z_l} + c_{m2} \right\} \end{aligned} \tag{24}$$

$$c_\rho = 1 \Bigg/ \left(\frac{\Gamma_g}{2 \rho_g^{(0)} v_g^{(0)}} - \frac{\alpha^{(0)}}{\rho_g^{(0)}} \left. \frac{\partial \tilde{\rho}_g}{\partial z} \right|_{z = z_l} \right) \tag{25}$$

$$c_R = \frac{3}{4\pi\bar{R}^{(0)^3}}\frac{\partial}{\partial z}\left(\frac{\sum \Gamma_g}{\sum N_g}\right)\Bigg|_{z=z_l} - \frac{3\rho_g^{(0)}}{\bar{R}^{(0)}}\frac{\partial\tilde{\bar{R}}}{\partial z}\Bigg|_{z=z_l} \tag{26}$$

$$c_{m1} = -\frac{\alpha^{(0)}}{\rho_l g}\mu_0(1-\alpha^{(0)})C_{(M)}M_l^{(0)^2}\frac{\partial C_{(H)}}{\partial z} \tag{27}$$

$$c_{m2} = -\frac{\alpha^{(0)}}{\rho_l g}\left[\mu_0(1-\alpha^{(0)})M_l^{(0)}H_T^{(0)}\frac{\partial^2 C_{(H)}}{\partial z^2} + \mu_0(1 - \alpha^{(0)})\frac{\partial^2 C_{(H)}}{\partial z^2}C_{(M)}C_\alpha M_l^{(0)^2}\right] \tag{28}$$

$$D_1 = \frac{\Gamma_g g}{\rho_l \bar{u}_g^{(0)}}\left[\frac{v_g^{(0)}}{g} + \frac{\alpha^{(0)}v_l^{(0)}}{(1-\alpha^{(0)})g}\right] + \frac{32\alpha^{(0)}v_g^{(0)}}{D^2\rho_l g}\left(-\eta_g + \frac{\eta_g v_g^{(0)}}{1-\alpha^{(0)}} + \eta_l\right) + \frac{3\mathrm{C_D}\pi}{4\pi\bar{R}^{(0)}}\left[(v_g^{(0)} - v_l^{(0)})\left(\frac{v_g^{(0)}}{g} + \frac{\alpha^{(0)}v_g^{(0)}}{(1-\alpha^{(0)})g}\right)\right] \tag{29}$$

$$E_1 = -\frac{\rho_g^{(0)}v_g^{(0)^3}}{\rho_l} + \frac{4\rho_g^{(0)}v_g^{(0)^2}}{\rho_l\alpha^{(0)}} + 2v_l^{(0)^2}\left(\frac{1}{\alpha^{(0)}} + \frac{\alpha^{(0)}}{1-\alpha^{(0)}}\right) + 2\mathrm{C_{VM}}\left(\frac{v_g^{(0)^2}}{\alpha^{(0)}} + \frac{v_l^{(0)^2}}{1-\alpha^{(0)}}\right) \tag{30}$$

$$F_1 = -\mu_0 C_{(M)}(1-\alpha^{(0)})\frac{M_l^{(0)^2}}{\rho_l}C_{(H)} \tag{31}$$

where z_l denotes the distance from the position of $z=0$ where the boiling starts and vapor bubbles initially appear. $\bar{u}_g^{(0)}$ denotes the superficial gas-phase velocity and is expressed by

$$\bar{u}_g^{(0)} = \alpha^{(0)}v_g^{(0)} \tag{32}$$

Focusing on the second order space differential term in Eq. (21), the linearized two-phase magnetic body force term F_1 is included in this differential term. Thus, the two-phase magnetic body force enhances the diffusion effect of the void waves.

Substituting the plane-wave expression of Eq. (13) for $\tilde{\alpha}$ into Eq. (17) and assuming $\tilde{\alpha}>0$ results in the following complex quadratic algebraic equation for ω_c,

$$\hat{\tilde{\alpha}}[A_2\omega_c^2 + D_2\omega_c + (E_2 + F_2 + iB_2D_2C_2)] = 0 \tag{33}$$

where

$$A_2 = \frac{\alpha^{(0)}}{1-\alpha^{(0)}}A_1$$

$$B_2 = k\cos(\theta_1)\frac{g}{1-\alpha^{(0)}}\cdot\frac{1}{D_2}$$

$$C_2 = C_{\mathrm{cof}}$$

$$D_2 = \frac{\alpha^{(0)}}{1-\alpha^{(0)}}D_1$$

$$E_2 = e_2\cdot k^2$$

$$e_2 = -\frac{\alpha^{(0)}}{1-\alpha^{(0)}}E_1$$

$$F_2 = \mu_0 C_{(M)}\frac{\alpha^{(0)}}{\rho_l}M_l^{(0)^2}C_{(H)}\cdot k^2 \tag{34}$$

where θ_2 is the angle between the direction of flow and the wave vector $\mathbf{k}$, as indicated in Fig. 1, and defined as

$$\theta_2 = \cos^{-1}(\mathbf{e}_{(u)}\cdot\mathbf{e}_{(k)}) \tag{35}$$

In Eq. (33), A_2, C_2, and D_2 depend on the properties of the unperturbed system, whereas B_2, E_2, and F_2 depend on the magnitude of the wave vector k.

Equation (33) is quadratic in ω_c, and hence the real and imaginary parts of its roots can be readily obtained. Accordingly, Eq. (33) is expressed as

$$(G_1\omega_c + 1)^2 = 1 - G_2 - iG_3 \tag{36}$$

where

$$G_1 = \frac{2A_2}{D_2}$$

$$G_2 = \frac{4A_2(E_2+F_2)}{D_2^2}$$

$$G_3 = \frac{4A_2B_2C_2}{D_2} \tag{37}$$

The real and imaginary parts of the two roots of Eq. (36) can now be obtained with the use of complex-variable algebra and are derived as

$$\omega_r = \frac{1}{G_1}\left[-1 \pm \sqrt{\frac{w_1 + (1-G_2)}{2}}\right] \tag{38}$$

$$\omega_i = \pm\frac{1}{G_1}\sqrt{\frac{w_1 - (1-G_2)}{2}} \tag{39}$$

$$w_1 = [(1-G_2)^2 + G_3^2]^{1/2} \tag{40}$$

In Eq. (38), the choice of the negative sign before the radical corresponds to waves that decay and hence need not be further considered here. Magnetic stabilization is achieved if $\omega_r<0$ and neutral stability is achieved when $\omega_r=0$. Choosing a positive sign before the radical in Eq. (38), the condition of stable flow state is derived as

$$G_2 > \frac{1}{4}G_3^2 \tag{41}$$

Substituting G_2 and G_3 into Eq. (41) and then A_2, C_2, B_2, E_2, e_2, and F_2 in the resulting expression gives the following criterion of stability. The stability criterion describes a transition from the stabilized flow state to an unstable flow state. The criterion is given in terms of two dimensionless parameters:

$$N_m\cdot N_v\begin{cases} >1 & \text{(unstable)}\\ =1 & \text{(neutrally stable)}\\ <1 & \text{(stable)}\end{cases} \tag{42}$$

where N_m represents the ratio of kinetic energy to magnetic energy of the two-phase magnetic fluid and N_v is the modulus of void fraction which includes the effect of the tangent and chord susceptibility $\hat{\chi}$ and $\chi^{(0)}$, respectively. They are defined as

$$N_m = \frac{\rho_l\bar{u}_g^{(0)^2}}{\mu_0 M_z^{(0)^2}} \tag{43}$$

$$N_v = \left[\frac{1}{16}\left(\frac{4A_2C_2}{D_2}\right)^2\left(\frac{\bar{u}_g^{(0)}}{\alpha^{(0)}}\right)^2\frac{g^2}{(1-\alpha^{(0)})\bar{u}_g^{(0)^4}A_1C_{(H)}} - \frac{e_2}{\alpha^{(0)}\bar{u}_g^{(0)^2}C_{(H)}}\right]\times[1 + (1-\alpha^{(0)})\chi^{(0)} - (1-\alpha^{(0)})(\chi^{(0)} - \hat{\chi})\cos^2(\theta_2 - \theta_3)]\frac{\cos^2\theta_2}{\cos^2(\theta_2-\theta_3)} \tag{44}$$

where $\theta_3=\theta_2-\theta_1$ is the angle between the direction of the flow and the applied field. If $\theta_2-\theta_3=\pi/2$, namely, if the wave is perpendicular to the magnetic field and $\theta_2 \neq \pi/2$, namely, the wave is other than transverse to the flow, N_v is infinite and thus stabilization of the two-phase flow is impossible. Since disturbance waves of all orientations can be present, an oblique field cannot stabilize the two-phase flow. Thus, stabilization is achieved when the applied magnetic field is axial ($\theta_3=0$). When the field is axially oriented ($\theta_3=0$) and $\chi^{(0)} > \hat{\chi}$, N_v is greatest when $\theta_2=\pi/2$, and N_v decreases with θ_2 to 0. Here, we assumed that $\theta_1=\theta_2=0$, namely, both applied magnetic field and disturbance waves are in the axial direction and also in parallel to the flow direction because such condition is most suitable for stabilization of two-phase flow. Substituting of $\theta_1=\theta_2=0$ into Eq. (44), the following equation is obtained:

$$N_v = \left[\frac{1}{16}\left(\frac{4A_2C_2}{D_2}\right)^2 \left(\frac{\bar{u}_g^{(0)}}{\alpha^{(0)}}\right)^2 \frac{g^2}{(1-\alpha^{(0)})\bar{u}_g^{(0)^4} A_1 C_{(H)}} - \frac{e_2}{\alpha^{(0)}\bar{u}_g^{(0)^2} C_{(H)}} \right] \times [1 + (1-\alpha^{(0)})\hat{\chi}] \tag{45}$$

Next, we derive the superficial gas-phase velocity in the equilibrium state $\bar{u}_g^{(\mathrm{m})}$ for a representative value, taking into account the dependence of the magnetic field on the superficial gas-phase velocity $\bar{u}_g^{(0)}$. Here, we assume that the system is in an uniform equilibrium state, and thus unknown variables in Eq. (2) and Eq. (3) are expressed as

$$\alpha = \alpha^{(0)} \quad p_l = p_l^{(0)} \quad p_g = p_g^{(0)} \quad \bar{R} = \bar{R}^{(0)}$$

$$v_l = v_l^{(0)} \quad v_g = v_g^{(0)} \quad \rho_g = \rho_g^{(0)}$$

$$\mathbf{H} = \mathbf{H}_T^{(0)} = \mathbf{e}_{(H)} H_T^{(0)}$$

$$\mathbf{M} = \mathbf{M}_T^{(0)} = \mathbf{e}_{(H)} M_T^{(0)} \tag{46}$$

where by substituting these values into Eqs. (2) and (3), and additionally by eliminating the pressure gradient terms, the initial equilibrium gas-phase velocity $v_g^{(\mathrm{m})}$ is derived as the following equations:

$$v_g^{(\mathrm{m})} = \frac{-L_2 + \sqrt{L_2^2 - 4L_1L_3}}{2L_1}$$

$$L_1 = \frac{3\rho_l C_\mathrm{D}}{8\bar{R}^{(0)}}$$

$$L_2 = \Gamma_g - \frac{3\rho_l C_\mathrm{D} v_l^{(0)}}{4\bar{R}^{(0)}} - \frac{32\eta_g \alpha^{(\mathrm{m})}}{D^2}$$

$$L_3 = \rho_g^{(0)} g - \frac{3\rho_l C_\mathrm{D} v_l^{(0)^2}}{8\bar{R}^{(0)}} - (1-\alpha^{(\mathrm{m})})\rho_l g - \Gamma_g v_l^{(0)} - \frac{32\eta_l(1-\alpha^{(\mathrm{m})})v_l^{(0)}}{D^2} \tag{47}$$

where $\alpha^{(\mathrm{m})}$ denotes the void fraction in the initial equilibrium state. The expression for $v_g^{(0)}$ is derived by changing $\alpha^{(\mathrm{m})}$ to $\alpha^{(0)}$ in Eq. (47). Also, the superficial gas-phase velocity $\bar{u}_g^{(\mathrm{m})}$ in an initial equilibrium state is defined as $\bar{u}_g^{(\mathrm{m})} = \alpha^{(\mathrm{m})} v_g^{(\mathrm{m})}$, and the normalized superficial gas-phase velocity $\bar{u}_g^*$ is derived as

$$\bar{u}_g^* = \frac{\bar{u}_g^{(0)}}{\bar{u}_g^{(\mathrm{m})}} = \frac{\alpha^{(0)} v_g^{(0)}}{\alpha^{(\mathrm{m})} v_g^{(\mathrm{m})}} \tag{48}$$

The expression for N_v of Eq. (45) combined with the neutral stability criterion $N_m N_v=1$ of Eq. (42) and the definition of N_m from Eq. (43) give the normalized magnetization parameter M_{np}^* which appears as the abscissa in Fig. 2,

$$M_{\mathrm{np}}^* = \left(\frac{\mu_0 M_l^{(0)^2}}{\rho_l \bar{u}_g^{(\mathrm{m})^2}}\right)^{1/2} = N_v^{1/2} \frac{\bar{u}_g^{(0)}}{\bar{u}_g^{(\mathrm{m})}} \tag{49}$$

Next, we investigate the theoretical prediction regarding the effect of a magnetic field on the dispersion relation of the void waves in the stable and unstable regions of the system. To accomplish this task in a general way, the following expressions are substituted into the Eq. (38) of growth factor ω_r:

$$\begin{cases} G_2 = \left[\dfrac{k}{g/(\alpha^* \bar{u}_g^{(0)^2})}\right]^2 \cdot \dfrac{1}{N_m N_v} \\ G_3 = \dfrac{2k}{g/(\alpha^* \bar{u}_g^{(0)^2})} \\ \alpha^* = \dfrac{2C_2 \cos\theta_2}{\alpha^{(0)}} \end{cases} \tag{50}$$

As a result, the equation of dispersion relation, including the effect of $N_m N_v$ for void waves, is derived as

$$\omega_r = \frac{1}{G_1}\left\{ -1 \pm \frac{1}{\sqrt{2}} \left[\left[\left(1 - \left(\frac{k}{g/\alpha^* \bar{u}_g^{(0)^2}}\right)^2 \cdot \frac{1}{N_m N_v}\right)^2 + 4\left(\frac{k}{g/\alpha^* \bar{u}_g^{(0)^2}}\right)^2 \right]^{1/2} + \left(1 - \left(\frac{k}{g/\alpha^* \bar{u}_g^{(0)^2}}\right)^2 \cdot \frac{1}{N_m N_v}\right) \right]^{1/2} \right\} \tag{51}$$

In Eq. (39), the imaginary part ω_i of ω_c determines the phase velocity V_p of the void waves which represents the propagation velocity of the waves and is derived as

$$V_p = \frac{\omega_i}{|\mathbf{k}|} \tag{52}$$

By combination of Eqs. (39), (50), and (52), the normalized phase velocity V_p^* is derived as

$$V_p^* = \frac{V_p}{\bar{u}_g^{(0)}} = \frac{\alpha^*}{2\sqrt{2}k^*}\left\{ \left[\left(1 - \frac{k^{*2}}{N_m N_v}\right)^2 + 4k^{*2} \right]^{1/2} - \left(1 - \frac{k^{*2}}{N_m N_v}\right) \right\}^{1/2} \tag{53}$$

where k^* denotes the normalized wave number and is defined as

$$k^* = \frac{k}{g/(\alpha^* \bar{u}_g^{(0)^2})} \tag{54}$$

Taking into account the above-mentioned theoretical analysis, it is now possible to precisely investigate the effect of the magnetic field on the stability and dispersion relation of the void waves in the boiling two-phase flow of magnetic fluid.

2.3 Conditions for Analysis. To construct the numerical conditions, we refer to the present experimental study on the boiling two-phase flow condition of magnetic fluid which will be presented in the next section. Furthermore, the physical conditions used in the theoretical analysis are the same as those of the experimental study. As a practical example, we use the fluid properties of hexane-based temperature sensitive magnetic fluid with manganese-zinc ferrite particles of 50 weight concentration. The conditions for analysis are summarized in Table 1.

3 Experimental Study

To confirm the validity of the analytical results and to obtain fundamental data for performance evaluation which can be applied to actual two-phase systems, the experimental study on the stability of boiling two-phase pipe flow of magnetic fluid under an axial and a transverse nonuniform magnetic fields is conducted with a flow apparatus. The experimental apparatus is constructed as a test loop as shown in Fig. 3. This is composed of a loop tube,

Table 1 Numerical and experimental conditions

Pressure in the equilibrium state	$p_l^{(0)}$	112.8	kPa
Equivalent bubble radius	$\bar{R}$	0.4	mm
Gas constant	$\mathfrak{R}$	96.5	J/(kg K)
Temperature in the equilibrium state	$T_g^{(0)}=T_l^{(0)}$	283	K
Velocity in the equilibrium state	$v_l^{(0)}$	0.01	m/s
Heat input per unit volume	Q_w	6.7×10^7	W/m^3
Void fraction in the initial equilibrium state	$\alpha^{(m)}$	0.0013	
Liquid-phase density	ρ_l	1386	kg/m^3
Gas-phase density in the equilibrium state	$\rho_g^{(0)}$	1.205	kg/m^3
Surface tension	γ_l	0.0213	N/m

the inner diameter D and the total length l_t of which are $D=8.0$ mm and $l_t=3.3$ m, respectively. The pipe is filled with hexane-based temperature sensitive magnetic fluid with manganese-zinc ferrite particles. The low evaporation temperature of hexane is suitable for use as working fluid in an energy conversion system using boiling two-phase flow of magnetic fluid.

It is possible to measure the distribution of the liquid-phase pressure as well as the temperature and the mean velocity of the flow by using this apparatus. The liquid-phase pressure differences are measured using semiconductor-type pressure difference transducers (2) and (3). The effect of magnetic field on the unsteady pressure fluctuation along the z coordinate being measured by pressure difference transducers (2). The absolute pressure is measured with a semiconductor-type pressure transducer (8). The measured signals are stored and averaged in a data logger (11) controlled by a computer (10) through a GPIB interface. The flow rate is automatically controlled by a needle valve (23) and a pump (24), and measured by a the venturi tube (13). Also, the total measurement is conducted during the steady flow state. The distributions of the liquid-phase pressure, temperature and the mean velocity are measured along the axial direction of the pipe. Furthermore, the ultrasonic wave echo system (16) and (17) with a video tape recorder (18) and image processor (19) is applied to measure the superficial gas-phase velocity and void fraction in two-phase magnetic fluid flow. The ultrasonic wave probe (16) is closely attached to the flow pipe with a gel pad. The ultrasonic wave echo visualizes unsteady behavior of the gas phase in the magnetic fluid.

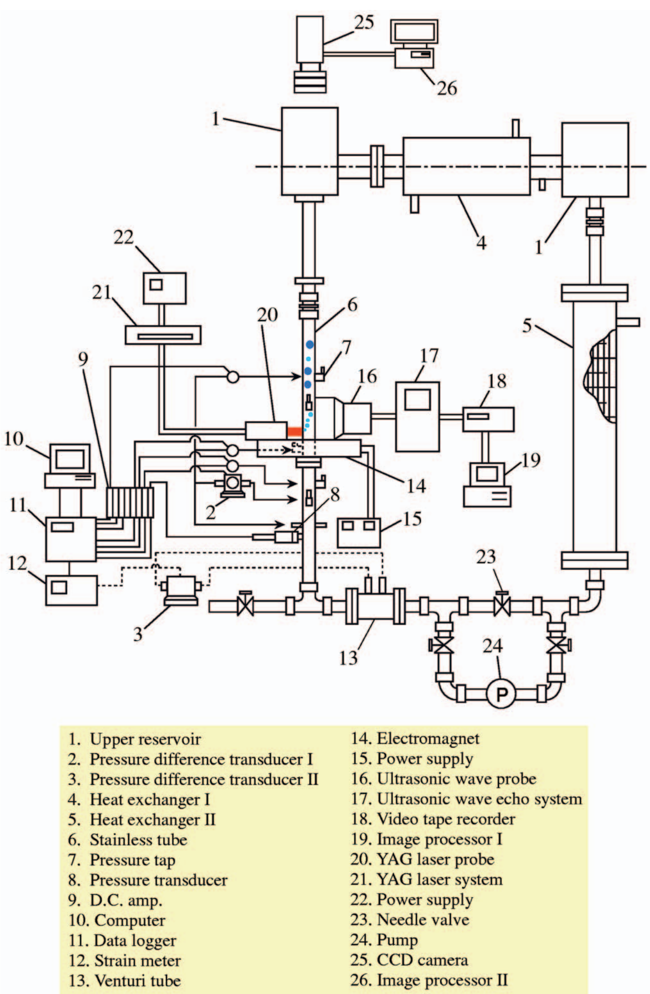


Fig. 3 Schematic of total experimental apparatus

The magnetic fluid is heated with constant heat flux passing through the point of maximum magnetic field strength in the downstream region by applying a YAG laser (20) to produce a boiling two-phase flow. Clockwise circulation flow is induced in the loop by the buoyancy force and the nonequilibrium magnetic driving force in the two-phase flow. The nonequilibrium magnetic force is based on the decrease of apparent magnetization due to the vapor bubbles production with an increase of the void fraction α and the temperature increase. After the generated vapor bubbles are condensed by the heat exchanger (4) and (5), the flow becomes single-phase again. Here, the YAG laser system (21) with a wave length of 1064 nm, a beam diameter of 4 mm, and a maximum power of 50 W is used to heat the magnetic fluid.

The nonuniform axial magnetic field is applied to magnetic fluid by the electromagnet (14) with dc power supply (15). Furthermore, for another experimental case, the transverse magnetic field is applied by a permanent magnet for comparison with the case of the axial magnetic field. The flow pipe is made of quartz glass to prevent absorption of the laser power. The shapes of the vapor bubbles on the free surface before they flow into the heat exchanger are captured by a CCD camera (25) and the contour of bubbles are specified by using an image processor (26).

The measurement conditions are the same as those for theoretical analysis previously summarized in Table 1.

4 Results and Discussion

Figure 4 shows a stability diagram for the dependence of the normalized superficial gas-phase velocity $\bar{u}_g^*$ on the normalized magnetization parameter M_{np}^* of the fluid. In this figure, the curve of $N_m/N_v=1$ represents the neutral stability line which separates the state of unstable flow from the region of the stable two-phase flow. Additionally, z_l^* denotes the non-dimensional parameter which is defined as

$$z_l^* = z_l/D \tag{55}$$

where z_l denotes the distance from the position of $z=0$ as we have previously mentioned. The horizontal line $\bar{u}_g^*=1$ represents the theoretical equilibrium solution for the initial boiling condition. $\bar{u}_g^*$

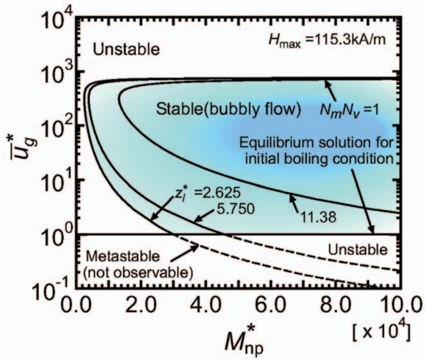


Fig. 4 Stability diagram for the dependence of the normalized superficial gas-phase velocity on the normalized magnetization parameter

is obtained by applying $\alpha^{(0)}=\alpha^{(m)}$ in Eq. (48). A stabilized regime exists in the inner portion of the shaded region between the line of the initial boiling condition and the curve of $\bar{u}_g^*$. It is found that the stabilized regime increases with the normalized magnetization parameter M_{np}^*. Also the regime increases as z_l^* approaches the region where the magnetic field strength is great ($z_l^* \to 0$) due to the effect of the magnetization of the fluid. Namely, the extension of the stable region is caused by (1) the increase of the effect of magnetization of the fluid, and (2) the approach of the magnetization to saturation magnetization with a decrease in magnetic susceptibility $\tilde{\chi}$ due to the strong magnetic field. In Eq. (21), the linearized two-phase magnetic body force term F_1 is included in the second order space differential term. Therefore, it may be reasonable to say that magnetic stabilization is obtained because the two-phase magnetic body force enhances the diffusion effect of void waves. According to this result, the stabilization of two-phase flow is obtained by practical use of the magnetic body force acting on the fluid and by applying an appropriate superficial gas-phase velocity.

Figure 5 shows the effect of the wave number k on the growth factor ω_r of the wave for different N_mN_v, namely, the dispersion relation of the void waves. In the case of $N_mN_v<1$, in which the effect of magnetization is stronger, it is found that the flow state easily becomes more stable state with the longer wavelength of disturbances (with the smaller wave number of disturbance). The shortest wavelength (largest wave number) disturbances reach an asymptotic growth rate that is constant at a given value of N_mN_v. The negative values of ω_r show that the system exhibits a stable state for the wave mode because the wave decays with time. The curve of $N_mN_v=\infty$ represents the universal instability of nonmagnetized two-phase flow, and as a result, the growth rate of the wave is independent of the magnitude of the disturbance when $N_mN_v=\infty$.

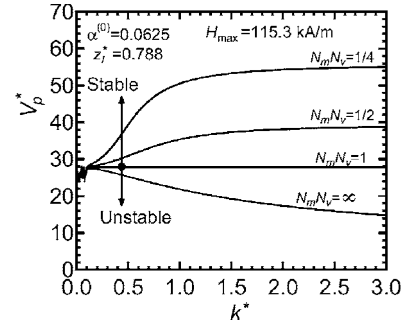


Fig. 6 Effect of normalized wave number on normalized phase velocity

Figure 6 shows the effect of the normalized wave number k^* on the normalized phase velocity $V_p^*(=V_p/\bar{u}_g^{(0)})$ for different N_mN_v. In the case of $N_mN_v<1$ of the stable flow state, it is found that the wave propagation speed V_p is much faster than the superficial gas-phase velocity $\bar{u}_g^{(0)}$. Also, the phase velocity increases with the effect of magnetic field, indicating that the faster decaying mode propagates more rapidly through the two-phase flow.

Figure 7 shows an application of the stability diagram to the experimental data. The experimental data are measured at the point where the production rate of the vapor bubbles is maximum. To obtain the superficial gas-phase velocity $\bar{u}_g^{(0)}$, the gas-phase velocity and the void fraction are measured by flow visualization measurement using the ultrasonic wave echo system with image processing. Especially, as in the case of stronger magnetic field, the measurement points shift to the region where the flow state is more stable. The tendency of measurement results are in reasonable agreement with the analytical results.

Figure 8 shows the effect of the axial magnetic field on the unsteady pressure fluctuation along the z coordinate. The pressure fluctuation is expressed by the nondimensional unsteady value $\Delta P_{lT(\mathrm{fl})}^*$, which is defined as

$$\Delta P_{lT(\mathrm{fl})}^* = \frac{\Delta p_{lT} - \Delta p_{lT(\mathrm{av})}}{\Delta p_{lT(\mathrm{av})}} \tag{56}$$

where Δp_{lT} is the cross-sectional mean effective driving pressure from which the influence on the prudence of the liquid-phase fluid can be deducted. $\Delta p_{lT(\mathrm{av})}$ denotes the average value of the pressure difference in the sequence of data sampling time. The normalized pressure fluctuation $\Delta P_{lT(\mathrm{fl})}^*$ decreases with the increase in the maximum magnetic field strength H_{max}. The magnitude of $\Delta P_{lT(\mathrm{fl})}^*$ is further suppressed when approaching a stronger magnetic field region ($z_l^* \to 0$) due to the strong effect of magnetic

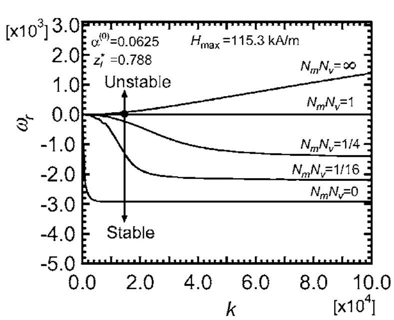


Fig. 5 Effect of wave number on growth factor

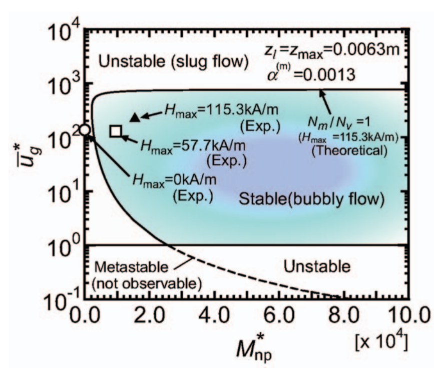


Fig. 7 Application of stability diagram to experimental data

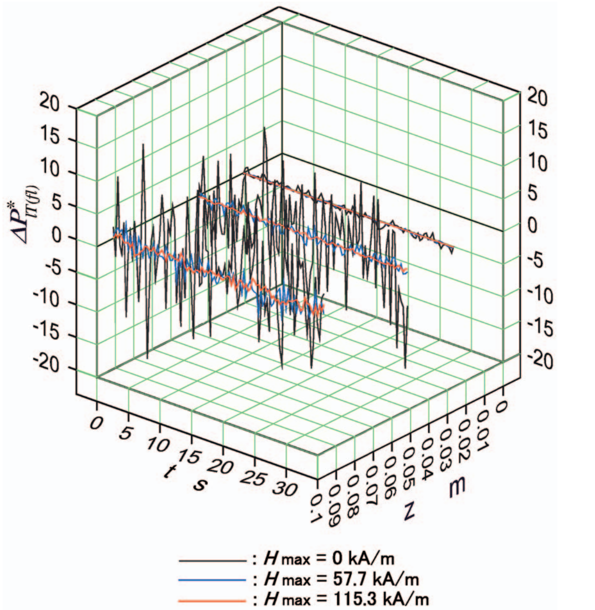


Fig. 8 Effect of axial magnetic field on unsteady pressure fluctuation

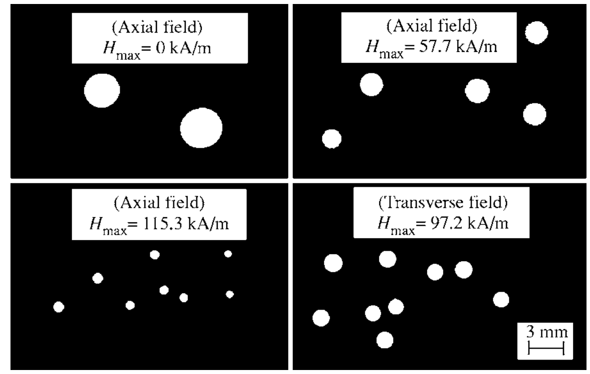


Fig. 10 Effect of magnetic field on the cross-sectional area of a vapor bubble on the free surface

body force. As a result, the two-phase flow state can be stabilized due to the magnetic body force.

Figure 9 shows the effect of the direction of the magnetic field on the unsteady pressure fluctuation $\Delta P^*_{lT(\mathrm{fl})}$. It is found that the magnitude of the pressure fluctuation in the nonmagnetic field is suppressed by either the axial or the transverse magnetic field. However, the magnitude of $\Delta P^*_{lT(\mathrm{fl})}$ in the axial field exhibits a smaller value than that in the transverse magnetic field. As a result, the axial magnetic field is more effective for stabilization of the two-phase magnetic fluid flow than of the transverse magnetic field. This tendency of the measurement results is in reasonable agreement with the analytical results.

Figure 10 shows the effect of the axial magnetic field on the aspects of the cross section of the vapor bubbles on the free surface of the upper reservoir (1) in the experimental apparatus (in Fig. 3). The bubbles are in the process of condensation. The white circles and the black background denote the cross-sectional areas of the bubbles and the magnetic fluid, respectively. The diameter of a vapor bubble and its cross-sectional area decrease with an increase in the magnetic field strength. According to this result, it is clarified that the vapor bubbles become minute and that the two-phase flow state is homogenized by the magnetic force. Furthermore, focusing on the direction of the magnetic field under almost the same maximum magnetic field strength conditions (H_{max}=115.3 kA/m, in the axial magnetic field; and H_{max} =97.2 kA/m, in the transverse magnetic field), it is found the bubble diameter in the case of the axial field becomes smaller than that in the case of the transverse field. According to this result, it is clarified that the axial magnetic field is more capable of diminishing the size of the generated vapor bubbles than in the transverse field. Therefore, the axial magnetic field is effective for homogenization of the boiling two-phase flow with magnetization of the fluid.

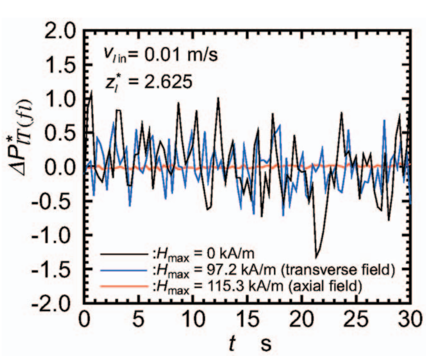


Fig. 9 Effect of the direction of magnetic field on unsteady pressure fluctuation

5 Conclusions

1. According to analytical study on void waves, it was found that stabilization of two-phase flow can be obtained by practical use of the magnetic body force acting on the fluid and by applying an appropriate superficial gas-phase velocity. Furthermore, it was found that magnetic stabilization is obtained because the two-phase magnetic body force enhances the diffusion effect of the void waves.
2. It was experimentally clarified that the two-phase flow state can be stabilized and homogenized by practical use of the magnetization of the fluid and that vapor bubbles can be minutely produced by effective use of the magnetic body force. Furthermore, the measurement results were found to reasonably agree with the analytical results.
3. It was both experimentally and theoretically clarified that the axial magnetic field is more effective than the transverse magnetic field for stabilization and homogenization of the two-phase magnetic fluid flow.

Acknowledgment

The authors would like to thank Emeritus Prof. Shinichi Kamiyama and Professor Hideya Nishiyama (Institute of Fluid Science, Tohoku University, Japan) for their helpful discussions. This research was supported by a Grant-in-Aid for Scientific Research (Contract No. 17686015) from the Ministry of Education, Science and Culture, Japan, and also by TEPCO Research Foundation, Japan.

Nomenclature

$a^{(i)}$ = interfacial area concentration
B = strength of magnetic flux
$\mathbf{B}$ = magnetic flux vector
C_D = drag coefficient
C_{VM} = virtual mass coefficient
c_p = specific heat
D = inner diameter of flow pipe
$\mathbf{e}$ = unit vector
g = gravitational acceleration

H = strength of magnetic field
$\mathbf{H}$ = vector of magnetic field
h = enthalpy
i = imaginary unit
k = wave number
$\mathbf{k}$ = wave number vector
l_1, l_2 = length of flow pipe
M = strength of magnetization
$\mathbf{M}$ = vector of magnetization
N_g = number density of bubbles
p = absolute pressure
q = heat flux
$\bar{R}$ = mean bubble radius
$\mathfrak{R}$ = gas constant
T = absolute temperature
t = time
v = velocity
z = axial coordinate
α = void fraction
Γ = phase generation density
γ = surface tension
η = dynamic viscosity
μ_0 = permeability in vacuum
ρ = density
ω_c = complex angular frequency

Subscripts

$(\ )_g$ = gas-phase
$(\ )_{\text{in}}$ = inlet of applied magnetic field region
$(\ )_l$ = liquid-phase
$(\ )_{\max}$ = maximum value
$(\ )_T$ = two-phase flow
$(\ )_z$ = component in the z-direction
$(\ )_0$ = position for boiling starts ($z=0$)

Superscripts

$(\ )^{(i)}$ = interface
$(\ )^{(m)}$ = initial equilibrium state
$\widetilde{(\ )}$ = perturbation
$(\ )^{(0)}$ = equilibrium state

References

[1] Rosensweig, R. E., 1985, *Ferrohydrodynamics*, Cambridge University Press, NY.
[2] Rosensweig, R. E., and Ciprios, G., 1991, "Magnetic Liquid Stabilization of Fluidization in a Bed of Nonmagnetic Spheres," Powder Technol., **64**, pp. 115–123.
[3] Ishimoto, J., Okubo, M., Nishiyama, H., and Kamiyama, S., 1995, "Basic Study on an Energy Conversion System Using Gas-Liquid Two-Phase Flows of Magnetic Fluid (Analysis on the Mechanism of Pressure Rise)," JSME Int. J., Ser. B, **39**, pp. 72–79.
[4] Ishimoto, J., 2004, "Numerical Prediction of Cavitating MHD Flow of Electrically Conducting Magnetic Fluid in a Converging-Diverging Nozzle," ASME J. Appl. Mech., **71**(6), pp. 825–838.
[5] Petrick, M., and Branover, H., 1985, "Liquid Metal MHD Power Generation—Its Evolution and Status," Prog. Astronaut. Aeronaut., **100**, pp. 371–400.
[6] Fedonenko, A. I., and Smirnov, V. I., 1983, "Particle Interaction and Clumping an Electrically Conducting Magnetic Fluid," J. Magn. Magn. Mater., **19**, pp. 388–391.
[7] Charles, S. W., and Popplewell, J., 1980, "Progress in the Development of Ferromagnetic Liquids," IEEE Trans. Magn., **MAG-16**, pp. 172–177.
[8] Shepherd, P. G., and Popplewell, J., 1971, "Ferrofluids Containing Ni–Fe Alloy Particles," Philos. Mag., **23**, pp. 239–242.
[9] Alekseev, V. A., 1991, "Structural Transformations in an Electrically Conducting Ferrocolloid," Magnetohydrodynamics (N.Y.), **27**, pp. 18–22.
[10] Alekseev, V. A., Veprik, I. Y., Minukov, S. G., and Fedonenko, A. I., 1990, "Influence of Microstructure on Physico-Mechanical Properties of Liquid Metal-Based Magnetic Colloids," J. Magn. Magn. Mater., **85**, pp. 133–136.
[11] Okubo, M., Ishimoto, J., Nishiyama, H., and Kamiyama, S., 1993, "Analytical Study on Two-Phase MHD Flow of Electrically Conducting Magnetic Fluid," Magnetohydrodynamics (N.Y.), **29**, pp. 291–297.
[12] Eckert, S., Gerbeth, G., and Lielausis, O., 2000, "The Behaviour of Gas Bubbles in a Turbulent Liquid Metal Magnetohydrodynamic Flow, Part I: Dispersion in Quasi-Two-Dimensional Magnetohydrodynamic Turbulence," Int. J. Multiphase Flow, **26**, pp. 45–66.
[13] Eckert, S., Gerbeth, G., and Lielausis, O., 2000, "The Behaviour of Gas Bubbles in a Turbulent Liquid Metal Magnetohydrodynamic Flow, Part II: Magnetic Field Influence on the Slip Ratio," Int. J. Multiphase Flow, **26**, pp. 67–82.
[14] Anderson, T. B., and Jackson, R., 1968, "Fluid Mechanical Description of Fluidized Beds: Stability of State of Uniform Fluidization," Ind. Eng. Chem. Fundam., **7**(1), pp. 12–21.
[15] Liu, Y. A., Hamby, R. K., and Colberg, R. D., 1991, "Fundamental and Practical Developments of Magnetofluidized Beds: A Review," Powder Technol., **64**, pp. 3–41.
[16] Hou, Y. Y., and Williams, R. A., 2002, "Magnetic Stabilisation of a Liquid Fluidised Bed," Powder Technol., **124**(3), pp. 287–294.
[17] Z. Al-Qodah, M. A.-B., and Al-Hassan, M., 2001, "Hydro-Thermal Behavior of Magnetically Stabilized Fluidized Beds," Powder Technol., **115**(1), pp. 58–67.
[18] Ganzha, V. L., and Saxena, S. C., 2000, "Hydrodynamic Behavior of Magnetically Stabilized Fluidized Beds of Magnetic Particles," Powder Technol., **107**, pp. 31–35.
[19] Kataoka, I., and Serizawa, A., 1989, "Basic Equations of Turbulence in Gas-Liquid Two-Phase Flow," Int. J. Multiphase Flow, **15**, pp. 843–855.
[20] Tomiyama, A., and Shimada, N., 2001, "A Numerical Method for Bubbly Flow Simulation Based on a Multi-Fluid Model," ASME J. Pressure Vessel Technol., **123**, pp. 510–516.
[21] Clift, R., Grace, J. R., and Weber, M. E., 1978, *Bubbles, Drops, and Particles*, Academic, San Diego, CA.
[22] Brennen, C. E., 2005, *Fundamentals of Multiphase Flow*, Cambridge University Press, NY.
[23] Otis, D. R., 1966, "Computation and Measurement of Hall Potentials and Flow-Field Perturbations in Magnetogasdynamic Flow of an Axisymmetric Free Jet," J. Fluid Mech., **24**, pp. 41–63.

Measurement of the Total Energy Release Rate for Cracks in PZT Under Combined Mechanical and Electrical Loading

H. Jelitto

F. Felten[1]

Institute of Advanced Ceramics,
Hamburg University of Technology,
Denickestrasse 15,
D-21073 Hamburg, Germany

M. V. Swain
Biomaterials,
Faculty of Dentistry,
University of Sydney,
Sydney 2010, Australia

H. Balke
Technische Universität Dresden,
Institut für Festkörpermechanik,
D-01062 Dresden, Germany

G. A. Schneider
Institute of Advanced Ceramics,
Hamburg University of Technology,
Denickestrasse 15,
D-21073 Hamburg, Germany

Four-point-bending V-notched specimens of lead zirconate titanate (PZT) poled parallel to the long axis are fractured under conditions of controlled crack growth in a custom-made device. In addition to the mechanical loading electric fields, up to 500 V/mm *are applied parallel and anti-parallel to the poling direction, i.e., perpendicular to the crack surface. To determine the different contributions to the total energy release rate, the mechanical and the piezoelectric compliance, as well as the electrical capacitance of the sample, are recorded continuously using small signal modulation/demodulation techniques. This allows for the calculation of the mechanical, the piezoelectric, and the electrical part of the total energy release rate due to linear processes. The sum of these linear contributions during controlled crack growth is attributed to the intrinsic toughness of the material. The nonlinear part of the total energy release rate is mostly associated to domain switching leading to a switching zone around the crack tip. The measured force-displacement curve, together with the modulation technique, enables us to determine this mechanical nonlinear contribution to the overall toughness of PZT. The intrinsic material toughness is only slightly dependent on the applied electric field (10% effect), which can be explained by screening charges or electrical breakdown in the crack interior. The part of the toughness due to inelastic processes increases from negative to positive electric fields by up to 100%. For the corresponding nonlinear electric energy change during crack growth, only a rough estimate is performed.*
[DOI: 10.1115/1.2744027]

Keywords: PZT, piezoelectric, fracture criterion, compliance, energy release rate

1 Introduction

Piezoelectric ceramics are commonly used as sensors and actuators in versatile, technical fields, such as the automotive industry, medical technology, metrology, and sonar applications [1]. In this context, the reliability is of particular importance, as such ceramics are susceptible to brittle fracture [2]. Because of their piezoelectric properties, in most applications these ceramics are used under combined mechanical and electrical loads. A large number of theoretical papers have been published concerning the influence of electric fields on cracks in piezoelectric materials [3–7]. Most approaches predict strong effects in retarding crack growth when an electric field is applied perpendicular to a nonconducting crack. On the other hand, experimental work has shown that the theoretically predicted effects are greatly overestimated and even partially contradictory to the experimental results. It seems evident that the fracture toughness in poled ferroelectric ceramics is larger for crack growth parallel to the polarization direction than perpendicular to it, which is related to ferroelastic domain switching. Moreover, concerning the influence of an additional electric field, different effects were reported that have not lead to a consistent understanding until now [8–13]. However, the evaluation of these experiments is difficult because of the nonlinear ferroic behavior, and also appropriate assumptions of the electrical boundary conditions of the crack are necessary to describe fracture in piezoelectric materials. The theoretical description of nonconducting cracks is often based on the assumption of complete impermeability. This is a major simplification, because electrical discharge and crack geometry effects such as bridging and branching elevates the permittivity interior to the crack significantly, as found by Schneider et al. [14]. Hence, finding an adequate fracture criterion that takes both mechanical and electrical loads into account is still one of the most challenging issues.

A critical value of the total energy release rate is used for one of the potential fracture criteria, as it is based on thermodynamic considerations [6,7]. Thus, a crack will start to propagate when a critical value $\mathcal{G}_c$ is reached, which is related to the surface energy of the material and the energy dissipation in the process zone. In several theoretical papers, the total, as well as the crack tip energy release rate, has been calculated for different geometries and under different assumptions regarding the electrical boundary conditions [3,5,7,8,15–17]. However, there is still a lack of experimental data to evaluate the theoretical predictions. Beside preliminary results by ourselves [18], no experimental approach has been published that directly enables simultaneous measurement of all substantial contributions to the total energy release rate directly from macroscopic properties without assuming specific electrical boundary conditions of the crack.

Against this background, a four-point-bending experiment with poled PZT specimen is performed under mechanical loads and applied electric fields from −500 to +500 V/mm, i.e., anti-parallel and parallel to the poling direction, respectively. As will be described in Sec. 3, the experimental setup used enables simultaneous in situ determination of the three linear components of the total energy release rate, i.e., the mechanical, the electric, and the piezoelectric part, as a function of the crack length and for differ-

[1]Currently at Robert Bosch GmbH, Corporate Sector Research and Aerospace Engineering, Postfach 106050, D-70049 Stuttgart, Germany.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received April 15, 2004; final manuscript received March 19, 2007. Review conducted by Robert M. McMeeking.

ent applied electric fields. Concerning the mechanical energies, the linear elastic and nonlinear contributions to the energy release rate were separated experimentally.

2 Theoretical Basis

In the following, the theoretical framework used to determine the energy release rate and the toughness is described. We apply Griffith's energy balance to our PZT material, which is ferroelectric and ferroelastic. This means that if a combination of stress and electric field is inside the switching surface, the material behaves completely linearly piezoelectric. If the electric field and mechanical stress are outside this region, domain switching takes place that leads to a change of the remanent polarization and remanent strain in these volumes. As will be seen in the following, it is not necessary for this investigation to know the switching criteria and constitutive equations in order to evaluate the energy changes in the PZT ceramic. We mainly use the experimentally determined linear part of the potential energy Π [3]:

$$\mathrm{d}\Pi = -\Delta \mathrm{d}F - Q\mathrm{d}V - \mathcal{G}\mathrm{d}A \tag{1}$$

Here, A is the crack surface area and Δ, F, Q, and V are displacement, force, charge, and voltage, respectively. In the following "linear part of the energy" or "linear energy release rate," respectively, means that the energy contributions, being partly quadratic in F or V, are attributed to the "linear" part of the constitutive equations. Equation (1) defines $\mathcal{G}$ as the crack driving force or energy release rate. Since the investigated PZT ceramic is ferroelectric and ferroelastic, the displacement Δ has a linear elastic part Δ_l and a remanent part Δ_r coming from the ferroelastic state and including inelastic as well as residual elastic deformation:

$$\Delta = \Delta_l + \Delta_r \tag{2}$$

Similarly, we have linear dielectric behavior leading to charges Q_l and remanent polarizations leading to charges Q_r:

$$Q = Q_l + Q_r \tag{3}$$

By using the linear terms of Eqs. (2) and (3), we get the linear part of $\mathrm{d}\Pi$:

$$\mathrm{d}\Pi_l = -\Delta_l\mathrm{d}F - Q_l\mathrm{d}V - \mathcal{G}_l\mathrm{d}A \tag{4}$$

In the case of stable steady state crack growth, a fracture criterion can be formulated, where $\mathcal{G}_l$ reaches the critical value $\mathcal{G}_{lc}$. (The subscript "c" means "critical.") Sakai and Bradt [19] gave different methods for separating the energies due to linear and nonlinear processes during crack advance. We use one of them to evaluate our experimental data. They also stated that the change in energy due to linear processes during crack advance can be associated with the breaking of the atomic bonds, meaning the intrinsic part $\mathcal{G}_c^{\mathrm{intr}}$ of the toughness. Using this assumption and generalizing it tentatively on the piezoelectric case, we obtain directly from Eq. (4) with F and V being constant:

$$\mathcal{G}_c^{\mathrm{intr}} = \mathcal{G}_{lc} = 2\gamma_s = -\left(\frac{\partial \Pi_l}{\partial A}\right)_{F,V}^{c} \tag{5}$$

with γ_s being the effective surface energy. Strictly speaking, the assumption that the linear processes correspond to the intrinsic toughness has not been proven, but it seems reasonable and therefore it is used here. The domain switching area also contains elastic residual stresses, which can influence the intrinsic energy release rate; these stresses, in principle, would not be measured with our compliance method, described below. On the other hand, the specimen compliance is dependent only on the crack length (assuming a constant elastic modulus, see further below) and is not dependent on how much the crack advance is influenced by residual stresses. The energy from residual stresses would be newly created during crack growth. The difference between this created energy and the amount contributing to crack advance remains as residual stress energy in the crack wake. It contributes to the overall toughness and would be constant along the crack in the case of steady state crack growth. Thus, it is necessary to assume steady state conditions. This issue has been addressed before by Rose and Swain [20].

In the process zone, the remanent polarization and the remanent strain change because of the high stress and electric field. Thus, beside $\mathcal{G}_c^{\mathrm{intr}}$, other processes exist, connected with domain switching, and leading to a toughness increase $\Delta\mathcal{G}_c$ as well as to a heightened overall toughness $\mathcal{G}_c$. Crack propagation takes place if the total energy release rate $\mathcal{G}$ becomes equal to a critical value, i.e., $\mathcal{G}=\mathcal{G}_c$, with

$$\mathcal{G}_c = \mathcal{G}_c^{\mathrm{intr}} + \Delta\mathcal{G}_c \tag{6}$$

Here the term $\Delta\mathcal{G}_c$ is not further specified. A detailed description of the segmentation into different energy contributions for the presence of a switching zone in ferroelectrics is given by Kreher [21].

The linear part of Π in Eq. (1) can be expressed with the compliances for linear piezoelectric materials as proposed by Suo [22]:

$$\Pi_l(V,F,A) = -\frac{1}{2}C_e^F V^2 - \frac{1}{2}C_m^V F^2 - C_p VF \tag{7}$$

where C_e^F, C_m^V, and C_p are the electric capacitance, the mechanical compliance, and the piezoelectric compliance, respectively. The superscript "F" means constant force and "V" means constant voltage. The electric charge Q_l and the displacement Δ_l are given by

$$Q_l = (-\partial\Pi_l/\partial V)_{A,F} = C_e^F V + C_p F \tag{8}$$

$$\Delta_l = (-\partial\Pi_l/\partial F)_{A,V} = C_m^V F + C_p V \tag{9}$$

The energy release rate due to the linear processes therefore is

$$G_l^{\mathrm{tot}} = -\left(\frac{\partial \Pi_l}{\partial A}\right)_{F,V} = G_e^F + G_m^V + G_p \tag{10}$$

where

$$G_e^F = \frac{V^2}{2}\frac{\partial C_e^F}{\partial A} \tag{11}$$

$$G_m^V = \frac{F^2}{2}\frac{\partial C_m^V}{\partial A} \tag{12}$$

and

$$G_p = FV\frac{\partial C_p}{\partial A} \tag{13}$$

are the linear electric, mechanical, and piezoelectric components of the energy release rate, respectively. From Eqs. (8) and (9) we readily find:

$$C_e^F = \left(\frac{\partial Q_l}{\partial V}\right)_{A,F} \tag{14}$$

$$C_m^V = \left(\frac{\partial \Delta_l}{\partial F}\right)_{A,V} \tag{15}$$

and

$$C_p = \left(\frac{\partial Q_l}{\partial F}\right)_{A,V} = \left(\frac{\partial \Delta_l}{\partial V}\right)_{A,F} \tag{16}$$

Since $C_p^F = C_p^V$ and therefore $G_p^F = G_p^V$, the superscripts of C_p and G_p are omitted. In order to determine the linear contributions to the total energy release rate, C_e^F, C_m^V, and C_p must be measured for different crack lengths. This enables calculating the derivatives of the compliances with respect to the crack surface area. Hence, the experiment must be performed under conditions of stable crack growth.

We are aware of the fact that the quantities G_e, G_m, and G_p are not invariant to the transformation of the variables, as e.g., a change from the variables F and V to the variables F and Q. But although the different energy parts vary quantitatively, the sum of the components, which is $\mathcal{G}_c^{\mathrm{intr}}$ at crack growth, remains the same. It is mentioned here that in the literature, as for example by Park and Sun [8], the linear release rate from the closure integral is decomposed into a mechanical and an electrical part, representing an invariant formulation. On the other hand, we use Eqs. (11) to (13) because the given energy components can be measured best in this composition. Thus e.g., the mechanical part G_m^V in this paper differs from G_m in Ref. [8]. (The only exception is at zero electric field, since then we have $G_m^V = G_l^{\mathrm{tot}}$.) To avoid confusion, we add superscript "V," "F,"or "(Δ)" to the quantities to indicate the given boundary conditions.

For the correct data evaluation, the capacitance needs to be measured at constant force F, represented by C_e^F. In our experimental device, which is displacement controlled, the capacitance is measured at constant displacement Δ, denoted by C_e^{Δ}. From thermodynamic considerations, the following relation can be derived:

$$C_e^F = C_{\mathrm{e}}^{\Delta}\left(1 + \frac{C_P^2}{C_e^{\Delta} C_m^V}\right) \tag{17}$$

Thus, we get C_e^F directly from the measured quantities. For our bending experiment in the range of crack lengths below 2.5 mm, the quotient in the brackets on the right-hand side of Eq. (17), a measure for the difference between C_e^{Δ} and C_e^F, is below 0.05%. However, the conversion is taken into account, because C_e^F can be calculated exactly.

In the following all energy release rates refer to the critical state of controlled crack growth. Thus, the superscript "c" for "critical" is not always given.

Since in our experiments the crack length is measured during controlled crack growth, the mechanical stress intensity factor K_I can be determined. In the Griffith crack solution, with the polarization vector acting perpendicular to the crack surface, K_I is independent of any external electric load [3]. Fully coupled piezoelectric finite element (FE) calculations for the given geometry, including combined electromechanical loading, also reveal that K_I is not dependent on the electric field [23]. The corresponding equation, taken from [23], is:

$$K_{\mathrm{I}} = \sqrt{w}\sigma_a k_{I\sigma}(\alpha) \tag{18}$$

with $\sigma_a=(3/2)(F/b)(s_o-s_i)/w^2$ being the bending stress of an uncracked sample loaded by the force F; s_o and s_i are the outer and inner support distances, respectively, b the thickness, and w the height of the specimen. Concerning K_I, we get identical shape functions $k_{I\sigma}$ for the permeable and the impermeable crack:

$$k_{\mathrm{I}\sigma} = \frac{\sqrt{\pi\alpha}}{(1-\alpha)^{3/2}}(1.07 - 1.89\alpha + 2.14\alpha^2 - 0.95\alpha^3) \tag{19}$$

Here, $\alpha=a/w$ is the normalized crack length with a being the absolute crack length. The formula is valid for $0.05<\alpha<0.95$. The shape function $k_{I\sigma}$, based on the FE calculations, is approximated by a least-squares fit [23]. We evaluate K_I during stable crack growth and denote it as K_{I}^C in order to distinguish it from the mechanical fracture toughness K_{IC}.

3 Experimental Procedure

3.1 Preparation of Specimens. Commercially available morphotropic PZT ceramic bars (PIC151, PI-Ceramic, Lederhose, Germany) of dimensions $3\times4\times28$ mm^3 are poled in the longitudinal direction using an electric field of 1.7 kV/mm (47.5 kV) for 15 min at room temperature. Before poling they were polished on the long side to enable microscopic determination of the crack extension. After poling the notch is cut with a diamond saw blade of 120 μm thickness and sharpened with a razor blade as specified by the single edge V-notch beam method (SEVNB) [24,25]. The notch depth is always about 1.0 mm and the tip radius around 10 μm.

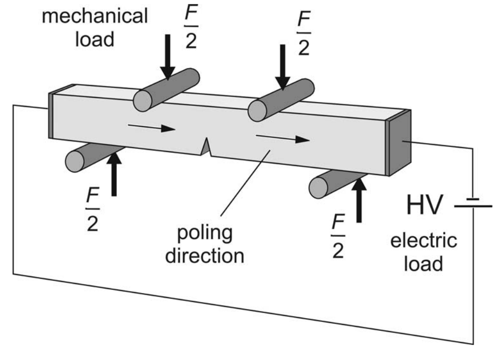


Fig. 1 Schematic geometrical four-point-bending setup with electromechanical load

3.2 Experimental Setup. Fracture experiments are performed under conditions of stable crack growth in a four-point-bending device. The supports used (Fraunhofer-Institut für Werkstoffmechanik, Freiburg, Germany) have support distances of 10 mm and 20 mm and a roller diameter of 5 mm. Ceramic rollers are used to achieve electrical insulation. The principal experimental setup is given in Fig. 1.

The supports are mounted in a very rigid metal frame (Fig. 2) as similarly done before by Fett et al. [26]. The dead weight of the upper support is compensated by weights connected via a rotatable wheel, which is dynamically decoupled using a spring. A steel plate is pre-stressed in the frame to reduce the thread slackness of the main screw (1) (Fig. 2), and thus increases the stiffness of the device. The displacement at the load points on the sample is changed manually using a hand wheel and a helical gear unit (E040B, ZAE Antriebs Systeme, Hamburg, Germany). This allows a precise and instantaneous control of the displacement with the precision of a few nanometers. The load accuracy is around 0.01 N. Due to its high stiffness, a quartz sensor (quartz dynamic load cell 9212, Kistler Instrumente GmbH) of high sensitivity (−11.3 pC/N) is used to measure the force. Quartz sensors exhibit an electrical drift, which must be taken into account. Since the overall drift is linear, this effect can be compensated.

The compliances are determined by a small modulation of the displacement. Therefore, a thin piezo-actuator is placed in the load line and excited by a low ac voltage of 5 Hz frequency. Simultaneously, the displacement of the upper supports with a modulation amplitude of about 30 nm is measured using an inductive position encoder (Fig. 2). In order to apply the theory correctly, the displacement is measured at the points where the mechanical load is induced. This is at the upper support rollers, from where the displacement is transferred via a movable linkage to the position encoder. In addition, we assume that the small displacement amplitude of 30 nm only leads to a linear response of the PZT sample, which is substantiated further below. It means that we measure $d\Delta_l$ and dQ_l for different crack lengths (but each measurement at constant crack length).

Optionally, the specimen can be loaded with high voltage up to 14 kV, i.e., an electric field up to 500 V/mm. In parallel, the capacitance and the electric current are measured using custom-made electronics (schematic diagram in [27]). The capacitance measurement has been calibrated with different capacitors, previously characterized with a precision LCR meter (Hewlett-Packard 4284A). Thus, an accuracy of about ±0.02 pF is determined for

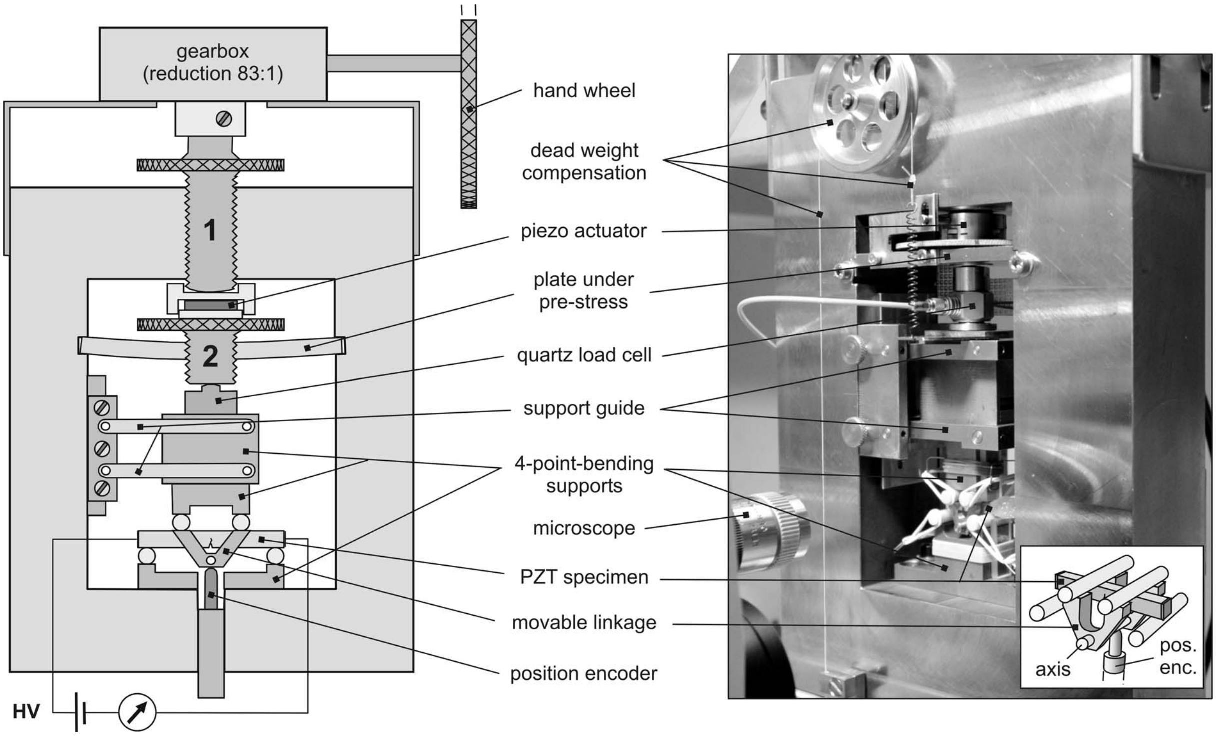


Fig. 2 Schematic drawing and photograph of the displacement controlled four-point-bending device. The insertion in the photograph shows the mechanical arrangement to measure the displacement of the upper support. The two V-shaped rods are movably connected by an axis and transfer the displacement of the upper support rollers to the position encoder. The electrical insulation at the high voltage side of the specimen is achieved by a coating with thermoplastics.

the measurement of capacitance changes. The ac signal (10 kHz, amplitude 1.5 V) necessary for this purpose is negligible compared to the applied electrical load of several kV.

According to Eq. (16), two equivalent procedures are possible to access the piezoelectric compliance. Varying the mechanical load under constant voltage, the change of the charge in the sample can be determined, i.e., $C_p^V=(\partial Q_l/\partial F)_{A,V}$. Another method is to alternate the voltage under constant force, which yields a change of the displacement, i.e., $C_p^F=(\partial\Delta_l/\partial V)_{A,F}$. Since the high voltage is kept constant by the power supply, the first option can be realized directly with the modulation technique and is used for all measurements presented in this paper. The second option was implemented only for test purposes. Generally, during the measurement of C_e^F, C_m^V, and C_p, the crack length remains unchanged.

For the evaluation of the compliances from the modulated mechanical load, the amplitudes of the force (dF), the linear displacement (dΔ_l), and the electric current (dI) must be correlated. Therefore the signals are recorded using a PC with a built-in AD converter. The amplitudes are determined by fitting sine functions to the measured data containing ten periods, which means 2 s measuring time per data point. Actually, amplitude and phase are fitted, whereas the frequency of 5 Hz is fixed. The evaluation procedure is similar to the formalism used in lock-in techniques. The compliances are calculated afterwards according to Eqs. (15) and (16) using the amplitudes of the fitted functions. The charge dQ_l necessary to calculate the piezoelectric compliance is determined by analytical integration of the sine function, which corresponds to the current signal. The accuracy of the current signal is ±0.02 pA. Furthermore, K_I^C at crack growth is calculated according to Eqs. (18) and (19) using the applied load and the crack length measured with an optical microscope (Wild M3Z) (Fig. 2).

3.3 Measured and Derived Quantities. Since the experiment is relatively complex, the measured and derived quantities are listed in Table 1 as a basis for further analysis. The time t is needed for the linear drift correction concerning the quartz sensor. $\mathcal{G}_c^{\mathrm{intr}}$ is derived solely from measured macroscopic quantities. The data are recorded simultaneously so that an entire set of parameters for all crack lengths is acquired with a single bending bar. As mentioned before, each measurement of dF, dΔ_l, and dQ_l along crack advance is performed at constant crack length.

The measuring technique in combination with stable crack advance in piezo ceramics is quite new. Therefore, the main steps of the data evaluation, including the correction due to the finite compliance of the four-point-bending device, and due to the compression of the sample at the contact points are summarized in the Appendix.

Table 1 Measured and derived physical quantities. They are used in the equations given before and determined for every crack length. The numbers in brackets indicate the equations, which define the quantities. The amplitudes of the sine functions due to the tiny 5 Hz modulation are dF, dΔ_l, and dQ_l, and lead to C_m^V and C_p. The energy release rate $\mathcal{G}_c^{\mathrm{mech}}$ also includes remanent processes and is defined further in the text.

Measured	Calculated (Eq.)
F	C_m^V (15)
dF	C_p (16)
Δ	C_e^F (17)
dΔ_l	K_I^C (18)
dQ_l	G_m^V (12)
V	G_p (13)
C_e^Δ	G_e^F (11)
a	$\mathcal{G}_c^{\mathrm{intr}}$ (5) and (10)
t	$\mathcal{G}_c^{\mathrm{mech}}$ (27)

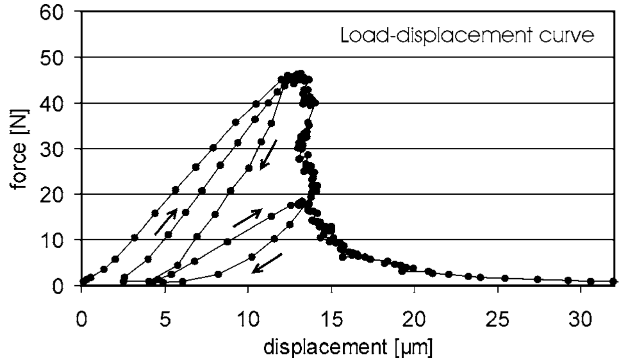


Fig. 3 Raw data of the load-displacement diagram for poled PZT. Two (out of six) unloading cycles are shown. The measurement is performed with the 5 Hz modulation.

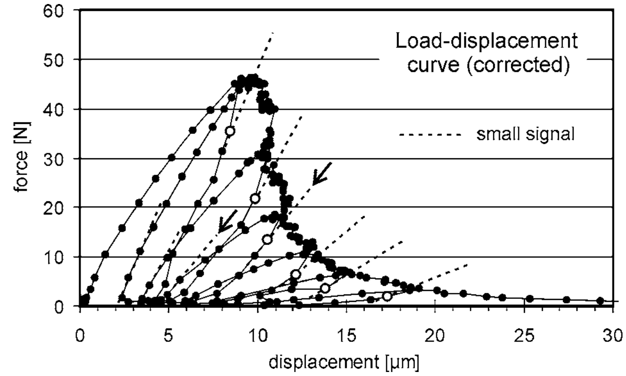


Fig. 4 Load-displacement curves of Fig. 3, including all six unloading cycles. The dotted "small signal" lines and the displacement are corrected according to Eqs. (A1) and (A5) in the Appendix . The starting point on the displacement axis is shifted arbitrarily to the origin (as also shown in Fig. 3).

4 Results and Discussion

4.1 Force-Displacement Curves and Their Interpretation. The raw data of a representative load-displacement curve of poled PZT without applied electric field (both electrodes of the specimen are connected with ground potential), including two unloading cycles, are given in Fig. 3. Already, the very first loading before crack growth leads to a remanent displacement of 2.5 μm after unloading. This effect is typical for all tests and for all applied electric fields. We interpret this inelastic behavior of the sample as the creation of a frontal process zone (see also Fig. 6). It can be seen that the unloading and the loading lines in Fig. 3 form closed hysteresis loops, implying that irreversible processes occur. The second unloading/loading cycle after substantial crack propagation of 1.5 mm shows the same irreversible hysteresis loop behavior. This is taken as switching of the ferroelastic domains around the crack tip during unloading and reloading. After a given crack advance of 1.5 mm and after complete unloading, an additional remanent displacement of about 1.5 μm is observed.

The areas of the "unloading hysteresis loops" become two to three times narrower if the 5 Hz modulation is not used. Obviously, the small mechanical modulation with an amplitude of 30 nm facilitates domain wall movements. But in both cases, i.e., with and without 5 Hz modulation, the inelastic remanent displacements as well as the slopes of the dashed loading lines, especially directly after reloading at $F=0$ N (arrows pointing up) are the same. In addition, the shape and the area of the main load-displacement curve are unchanged by the 5 Hz modulation.

The measured displacement in Fig. 3 must be corrected due to the device compliance and to the compression of the sample. For details, see the Appendix (Fig. 18 and Eq. (A5)). The corrected diagram of Fig. 3 is shown in Fig. 4 with all six unloading/loading cycles. The open circles in Fig. 4 represent the places of slight unloading, where the small signal compliance C_m^V is measured. The slopes of the dotted lines, corresponding to $1/C_m^V$, run nearly tangentially to the unloading lines.

The loading lines are slightly curved (Fig. 4), implying that the specimen becomes softer with increasing force. This seems reasonable because of increasing domain switching. On the other hand, the slopes of the small signal compliance values (dotted lines) fit well to the unloading/loading cycles when reloading starts at $F=0$ N. The two arrows in Fig. 4 indicate two parallel dotted lines representing the small signal compliance. Additionally, when measuring the small signal compliance along the data points of an unloading/loading cycle, its value is almost constant within approximately 5% (except at forces less than about 5 N). It follows that the small signal compliance is independent of actual load conditions. (Note that the modulation amplitude of about 30 nm for measuring C_m^V is less than the *thickness* of the dotted lines.)

As a consequence of these results, we assume that the measured mechanical small signal compliance C_m^V represents solely the linear mechanical response of the ceramic. It means that under small signal modulation, no domain switching occurs because of the tiny modulation amplitude and the short time constant (5 Hz). If we vary the small signal amplitude as well as the frequency of 5 Hz by a factor of 2, C_m^V is unchanged within the experimental uncertainty. With a variation of about 1% C_m^V represents a minimum constant compliance value, implying that we measure below a certain threshold, where no domain switching occurs any more. It seems that even in the general case of large scale yielding, the modulation technique measures solely the linear response of the system.

The energy contributions during complete unloading of the specimen and during crack advance are shown schematically in the load-displacement diagrams in Figs. 5(*i*) and 5(*ii*). Here, we address the pure mechanical case, which turns out to be reasonable, when the electrical influence is examined further below. At point (c) (Fig. 5(*i*)) the elastically stored energy is given by the area (B). When unloading the sample completely from point (c) to (d), the area (C) denotes the energy, which is regained mechanically by domain back-switching. If loading again the specimen to point (c), the area (D) is a measure for the hysteresis energy, converted into thermal motion. The parallel lines 1 and 3 denote the (inverse) small signal compliance, corresponding to the linear elastic material response.

During the experiments for the evaluation of the critical energy release rates, the specimens were never unloaded, in order not to disturb the measurement. However, due to the small signal modulation technique, the total energy release rate can be separated into the energy due to linear elastic processes (area (E)) and the remanent energy (area (F)) as proposed by Sakai and Bradt [19] (see Fig. 5(*ii*)). Referring to that reference, we assume also that area (E) corresponds solely to the creation of a new crack surface. Area (F) includes energies from remanent inelastic processes as well as energies due to elastic residual stresses in the domain switching area, which influence the crack advance. Rose and Swain [20] denote the overall energy (E)+(F) per newly created crack surface as specific incremental "work of fracture." Note that the displacement increment $d\Delta_l$ in Fig. 5(*ii*) due to linear processes at crack advance is not identical to the quantity $d\Delta_l$ in Table 1, which is measured at constant crack length. The two quantities ($d\Delta_l$ and $d\Delta_r$), given in Fig. 5(*ii*), are not used further in this paper and are given only for clarification with respect to Δ_l and Δ_r in Eq. (2).

The hysteresis processes during unloading and loading can be explained by a process zone, which is partly reducing its size during unloading and is increasing again during loading. In Fig. 6,

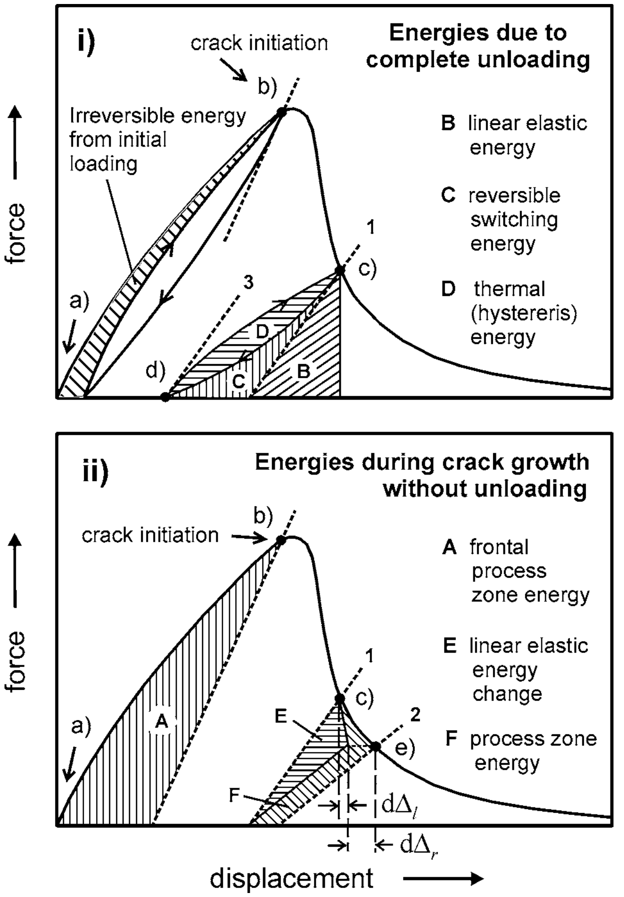


Fig. 5 Schematic load-displacement diagrams for crack advance in ferroelectric PZT. (i) Energies during a complete unloading cycle at crack initiation and after a certain crack advance, (ii) energies during crack growth from (c) to (e) without unloading.

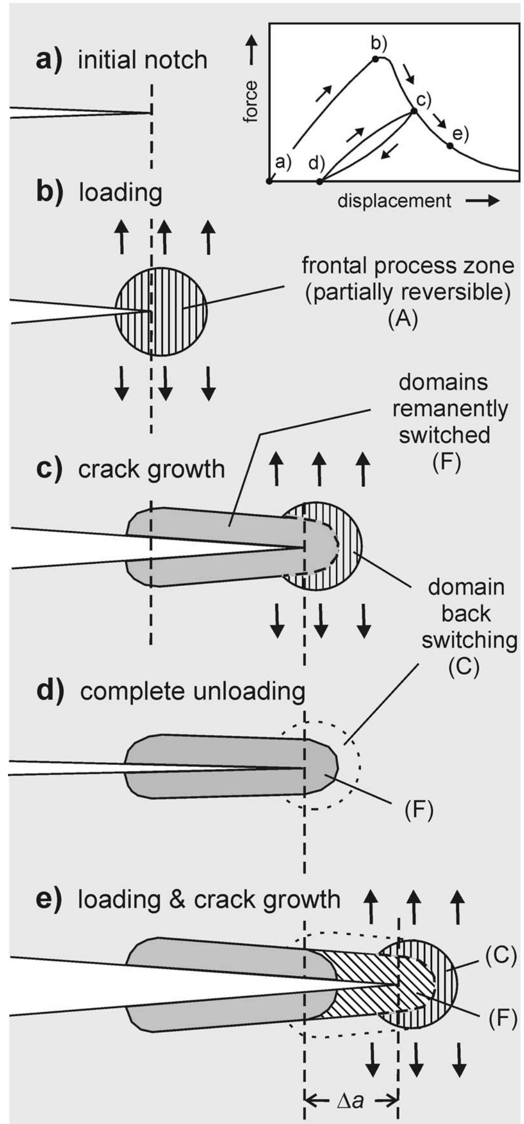


Fig. 6 Schematic crack and process zone area. (*a*) Initial notch. (*b*) The specimen is loaded to a value just before the crack starts to grow, which creates the frontal process zone. (*c*) The crack has grown. (*d*) The specimen is completely unloaded. (*e*) The crack is loaded again and has grown by an amount Δa. The diagonally hatched area (F) in (*e*) shows the process zone area of remanently switched domains, corresponding to the crack extension Δa. The vertical arrows indicate tensile stress. The capital letters (A), (C), and (F) correlate to the corresponding areas in Fig. 5. The panels (*a*) to (*e*) correspond to the points (a) to (e) in Fig. 5.

the crack and the process zone are shown schematically. The charts 6(*a*) to 6(*e*) in Fig. 6 correspond to the points (a) to (e) in the inset force-displacement diagram and to Fig. 5. The very first loading from (a) to (b) in Fig. 5(*i*) shows some nonlinearity that we interpret as the creation of a frontal process zone. By unloading the sample from points (c) to (d) (Fig. 5(*i*)), the process zone around the crack tip decreases by the vertically hatched area (C) in Fig. 6(*c*). Reloading from points (d) to (c) increases the process zone again and leads to the nonlinear load-displacement curve. Therefore, the process zone height in the crack wake is drawn smaller than the process zone at the crack tip. Due to this reversible process, we assume partial domain back-switching in the wake. When the crack grows by an increment Δa from (c) to (e) (Figs. 5(*ii*), 6), the fully developed process zone is shifted under stationary conditions along the distance Δa. The diagonally hatched area (F) in Fig. 6(*e*) indicates the area of remanent domain switching belonging to Δa (energy (F) in Fig. 5(*ii*)).

With this interpretation of the load-displacement curve at zero electric field, the intrinsic toughness during stable steady state crack growth as defined in Eq. (5) is

$$\mathcal{G}_c^{\mathrm{intr}} = G_m^V = \frac{\mathrm{area(E)}}{\Delta A} \tag{20}$$

The toughness part, including domain switching, during steady state crack growth is

$$\Delta\mathcal{G}_{\mathrm{c}} = \frac{\mathrm{area}(F)}{\Delta A} \tag{21}$$

The interpretation of the force-displacement curve given above is also applied for mechanical loadings with a constant applied electric field. Under constant voltage, the F-Δ curve includes the piezoelectric displacement Δ_p, which can be calculated according to $d\Delta_p = V dC_p$. The compliance $C_p(A)$ is a function of the crack length and the measured value varies roughly between ± 10 pm/V for the essential crack extensions (see Figs. 8(*b*) and 9(*b*)). Accordingly, for maximum applied voltages of 14 kV, the piezoelectric displacement varies between ± 0.14 μm, which is small enough to be neglected.

Since the Q-V curve during crack extension is not measured,

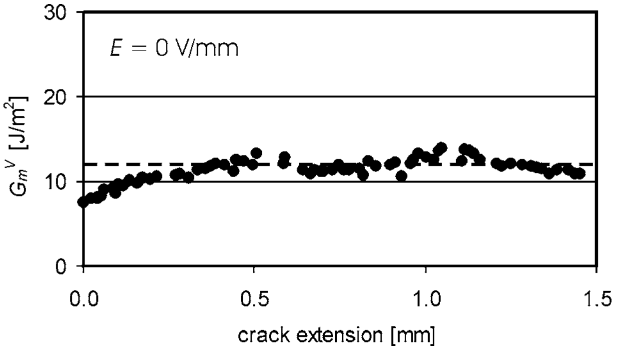


Fig. 7 Intrinsic toughness for poled PZT without applied electric field. The compliance curve for calculating the energy release rate is fitted for crack extensions between 0 mm and 1.5 mm. The dashed line represents 12 J/m^2.

we only have information about the mechanical remanent energy part. With our experimental setup, it was not possible to determine the total $\Delta\mathcal{G}_c$, because we do not know the whole electrical energy change due to irreversible processes. (A rough estimate is given further below.) On the other hand, we have the complete information to calculate $\mathcal{G}_c^{\text{intr}}$ by evaluating the compliances C_e^F, C_m^V, and C_p during stable crack growth and entering them into Eq. (10).

4.2 Critical Energy Release Rate. As the small signal modulation measures the linear elastic response, we may apply the approach of Suo using Eqs. (10) to (13). For zero electric field, the resulting intrinsic toughness $\mathcal{G}_c^{\text{intr}}=G_m^V(a)$, given in Fig. 7, is approximately 12 J/m^2.

With the setup in Fig. 1, the electric loads between −14 kV and 14 kV had been applied before the specimens were loaded mechanically. Furthermore, for all electric loads including zero field, the specimens were not unloaded in between, as said before.

As expected, the measured mechanical compliance C_m^V at electric fields of −500, −250, 0, 250, and 500 V/mm reveals a monotonic increase with respect to the crack length (Fig. 8(*a*)). The curves look similar, irrespective of the applied electric field. The piezoelectric compliance curves in Fig. 8(*b*) are also similar except one at −500 V/mm, which proceeds significantly higher than the other curves. The strong electric field opposite to the poling direction together with the high mechanical stress probably leads to large scale domain reorientation processes. Although the applied field is definitely below the coercive field of approximately E_c=850 V/mm, it is probable that the original piezoelectric state of the PZT ceramic is already disturbed.

The calibration of the current and charge measurement, respectively, was verified in a uniaxial compression test with a cubic poled PZT PIC151 specimen. The zero crossing of C_p can be understood qualitatively as follows. For an ideal bending bar, in principle, the piezoelectric compliance is zero. The charges generated in the compression and in the tension zone compensate each other. In the present case, the bending bar is single edge notched, which leads to asymmetric behavior. This asymmetry probably changes its characteristic, as the crack proceeds through the specimen, which could explain the tendency in Fig. 8(*b*). Both quantities C_m^V and C_p are corrected according to Eqs. (A1) and (A2).

The electrical capacitance is not measured absolutely, but only its change is acquired with the precision mentioned above. Hence, the curves shown in Fig. 8(*c*) are shifted vertically to fit the arbitrarily chosen mutual value of 8.5 pF at a crack length of a =1.5 mm. (It is a typical capacitance of the used specimens with such crack length.) For crack lengths smaller than 3 mm, the capacitance curve is nearly linear and for longer cracks, it declines strongly.

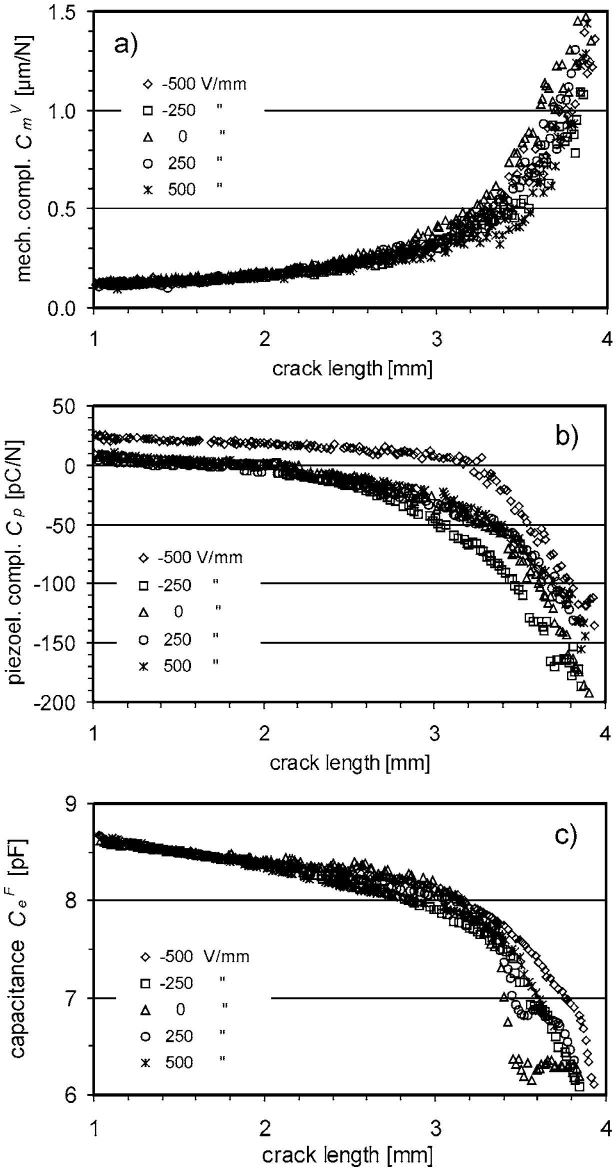


Fig. 8 (*a*) Mechanical and (*b*) piezoelectric compliance as well as (*c*) capacitance as a function of the crack length for different electric fields and corrected according to Eqs. (A1) and (A2). For a better comparison the crack length is used, being the sum of notch depth (between 0.98 mm and 1.07 mm) and crack extension.

For calculating the derivatives of the compliances, suitable analytical functions are fitted with respect to the crack length. The functions and the motivation for their choice are given in the Appendix (Eqs. (A6) to (A8)). Figure 9 presents the curves for E=500 V/mm. The quantities are fitted between 0 mm and 1.8 mm crack extension corresponding to 1.0 mm and 2.8 mm of total crack length.

Note that the piezoelectric compliance ($\partial Q_l/\partial F$) in Figs. 8(*b*) and 9(*b*) exhibits an unsteady behavior like a step at the zerocrossing. With increasing crack length, the decreasing charge amplitude dQ_l as a function of time (sine function) does not continuously pass zero, which would be equivalent to a sudden phase shift of 180 deg. Instead, the charge amplitude dQ_l passes a minimum at a small positive value, and simultaneously the phase is shifted slowly from 0 deg to 180 deg. Since, in the calculation of C_p (Eq. (16)), this continuous phase shift of the charge signal with respect to the force signal is not included, the sign of C_p is changed when

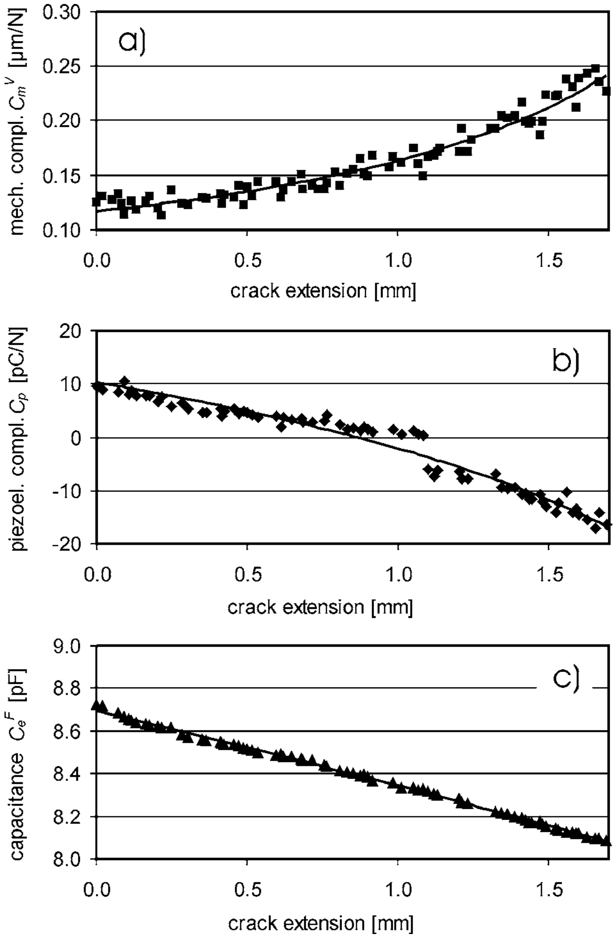


Fig. 9 (*a*) and (*b*) Compliances and (*c*) capacitance as a function of the crack extension for an electric field of 500 V/mm and corrected as described before. The fitted analytical functions are used to differentiate the experimental data with respect to the crack surface area. (For the discontinuity of C_p, see text and compare with Fig. 8(*b*)).

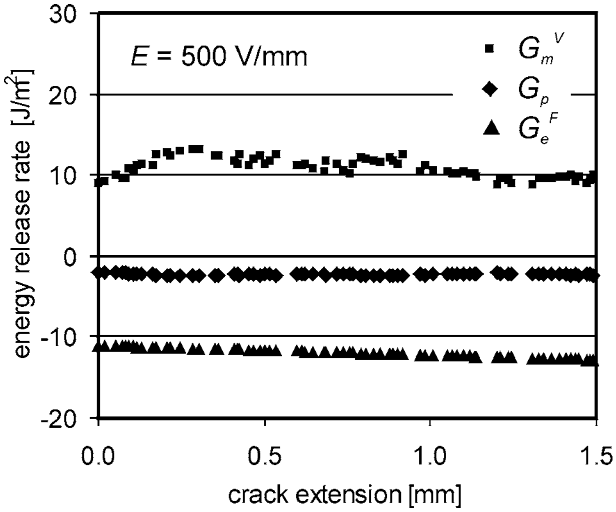


Fig. 10 Measured energy release rates G_m^V, G_e^F, and G_p for an electric field of 500 V/mm during stable crack advance

the measured phase shift becomes more than 90 deg. Thus, this drop of a few pC/N has a technical reason. It cannot be considered by the given theoretical approach and is bypassed, while fitting a smooth curve along this step in the C_p diagram. Fixing the phase at a certain value would yield a smooth zero crossing of C_p, but yields other difficulties while calculating C_p. In Fig. 8(*b*), it can be seen that this step is comparatively small with regard to the whole C_p curve.

The derivatives of the fitted compliance and capacitance curves are calculated analytically and multiplied by the measured force F and electrical load V according to Eqs. (11) to (13). Thus, for the linear processes we obtain the mechanical, the piezoelectric, and the electric energy release rates, which are shown in Fig. 10 for an electric field of +500 V/mm.

The evaluation of the mechanical energy release rate G_m^V during the first 200 μm crack extension that shows an increase with increasing crack length has a strong uncertainty. It stems from the increase of the force F at the beginning during the first 200 μm crack advance. Looking at the expression of G_m^V (Eq. (12)), the initial increase of the term F^2 can, in principle, be compensated by a corresponding variation of $\partial C_m^V / \partial A$. Nevertheless, $C_m^V(a)$ is fitted by a hyperbola of only three free parameters (Eq. (A6)), which does not allow us to fit short range variations.

The electric and the piezoelectric components are negative. For crack extensions from 0.5 mm to 1.5 mm, G_m^V, G_e^F, and G_p are almost constant as for all other applied electric loads. This indicates that the crack extension is in a steady state regime, which confirms the assumptions for the formulation of $\mathcal{G}_c$. At crack extensions above 2 mm, i.e., at crack lengths larger than 3 mm, the mechanical and the piezoelectric energy release rates approach zero smoothly. The electric part in that region, on the other hand, decreases dramatically down to about -150 J/m^2. We assume that at these large crack lengths the process zone touches the specimen's "back side." Additionally, the correction of C_m^V according to Eq. (A1) gets a strong influence on the result, since the device compliance C_{m01} is relatively large for small loads (compare Fig. 17(*a*)). Therefore, the range of crack lengths above 2.5 mm is not further analyzed.

Assuming a steady state crack growth between 0.5 mm and 1.5 mm crack extension, we average the data in this interval in order to minimize the experimental scatter and to display these medium values versus the applied electric field (Fig. 11).

Since the electrical field concentration at the crack tip is geometry dependent, it would be more appropriate to display the results as a function of the intensity factors K_{I} and K_{IV}. However, we would have to postulate certain boundary conditions in order to calculate an electrical intensity factor K_{IV}. Even if it has been experimentally shown that for the PZT used the apparent relative permittivity in the crack is not 1 but around 40 [14], the electric charge distribution on the crack surfaces is unknown. To avoid unproved conditions, we simply display the measured components of the energy release rate as a function of the (known) electric far field. The original notch depths are always within 0.98 mm and 1.07 mm and the other geometrical features are nearly identical for all specimens, so that the comparison in Fig. 11 is reasonable. The experimental data of the energy release rate contributions are presented without any theoretical assumption.

The measured energy release rate G_m^V at crack propagation is roughly constant over the measured range of applied electric fields. At zero electric field we have $G_m^V = 11.5$ J/m^2 (averaged for two samples). This agrees very well with the value of 12 J/m^2 found by Heyer et al. [28] for the same material but for a conducting crack at $K_E = 0$ kV/m$^{1/2}$, meaning pure mechanical loading.

The measured G_e^F values during controlled crack propagation are negative as expected [22] and exhibit a parabolic shape with respect to the electric load. The reason is as follows. As shown in Figs. 8(*c*) and 9(*c*), C_e^F and thus also $\partial C_e^F / \partial a$ are almost independent of the applied electric field. Hence, the part of the energy release rate given in Eq. (11) depends mainly on the square of the

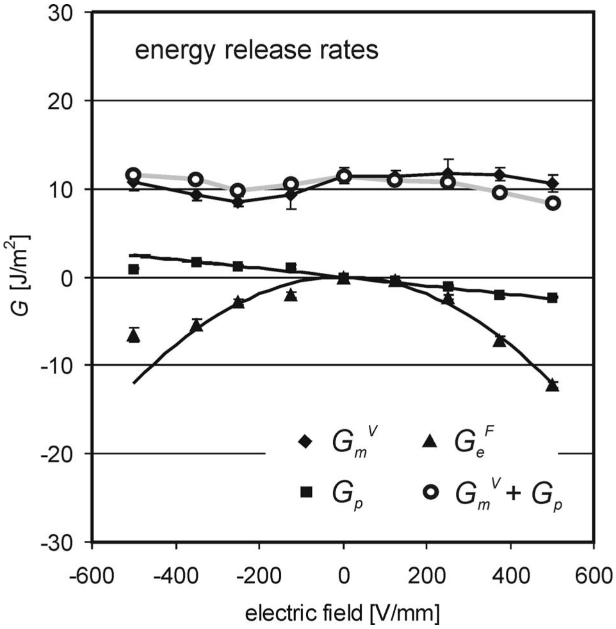


Fig. 11 Intrinsic energy release rates averaged between 0.5 mm and 1.5 mm crack extension. The linear and parabolic functions shown are valid for the idealized case, i.e., that the derivatives of C_p and C_e^F are independent of the electric field (see text). The data points are measured with only one sample for each electric field, except for the one at zero field, where two samples are averaged. The open circles connected by the gray line are the sum of all three contributions, where additionally G_e^F has been set to zero. Thus, actually it is the sum of G_m^V and G_p.

voltage V^2, which means that G_e^F is proportional to E^2 in this range of crack lengths. The piezoelectric part G_p during controlled crack propagation is small and shows a linear behavior, which can be explained by an analogous argument.

At applied fields of −500 V/mm, the piezoelectric and the electric energy release rates deviate from the main trend lines. We assume that this is due to the beginning large scale depolarization in the sample, as explained before in the context of Fig. 8(*b*).

Concerning the linear processes, Fig. 11 shows the three components of the total energy release rate available at crack growth. The sum of them, which we define as $\mathcal{G}_c^{\text{intr}}$ (Eq. (5)) or G_l^{tot} in the critical state (Eq. 10), respectively, is not constant, as we would expect. It even becomes negative for 500 V/mm. This is not compatible with the interpretation of $\mathcal{G}_c^{\text{intr}}$ as the energy necessary to break the atomic bonds. From the experimental results, it is appealing to take only the mechanical part G_m^V as the criterion for crack growth, because it is almost constant for all applied electric fields. A similar conclusion has already been suggested by Park and Sun [8], but the physical argument that justifies this approach is missing.

If we assume that free charges completely screen the remanent as well as the dielectrically induced polarization charges on the crack surfaces, the electric field would penetrate the crack cavity without any disturbance. As a consequence, there would not be any change in the measured capacitance as a function of crack length and G_e^F would be zero.

On the other hand, our measurements, which show a change in the small signal capacitance with crack length obviously contradict this assumption. A possible solution for this dilemma is that during the 10 kHz modulation of the voltage with an amplitude of 1.5 V biased by the high applied voltage of several kV, the screening charges cannot follow. The consequence is that the small signal modulation measurement (10 kHz) detects a purely linear response of our sample. Especially for the modulated signal and only for this, no screening charges are created at the crack surface. This would explain the measured change in capacitance with crack advance even though under the quasistatic conditions of crack growth, the polarization charges would be completely screened.

This explanation is in so far very probable because unscreened remanent polarization charges would lead to never measured extremely high crack growth retarding effects. This was shown theoretically in [16]. In almost all of the literature, it is implicitly assumed that the remanent polarization charges at the crack surfaces are screened. Taking this idea as serious, the induced polarization charges should also be screened, which is consistent with the above given approach. Sources for screening charges could be water surface layers or already existing ions, and produced ions and electrons from dielectric breakdown events in the crack cavity atmosphere.

A model assumption for breakdown in the interior of the crack is that the electric field cannot rise above a critical level. Thus, free charge is created and deposited on the crack faces sufficient to keep the electric field in the crack to the breakdown level [17]. As internal sources for screening, a small electric conductivity of the PZT could create space charges or domain switching localized at the crack surface and could average out the remanent polarization. A similar structure was detected during domain wall movement [29]. Both effects (screening and breakdown) would create an electric dc current during crack advance, which is not covered in the intrinsic energy release rate and would certainly influence the energy balance.

The measured change of the piezoelectric compliance is obtained by the slow 5 Hz modulation signal, which is applied mechanically. The charge response on the electrodes is due to the state of the sample with the assumed unchanged screening charges on the crack surface. Therefore, this electric signal should give the correct physical result.

For the evaluation of the linear energy release rate, we have to sum up the mechanical, piezoelectric, and electric part as given in Eq. (10), but now we set $G_e^F=0$, as explained above. The result for the total critical linear energy release rate with values between about 9 J/m^2 and 12 J/m^2 is shown in Fig. 11. This would imply that Eqs. (5) and (10) represent a theoretical approach, which practically must be modified in the case of electric loads, because then the physical situation is much more complex.

Although the resulting curve (open circles) is nearly constant, a slight dependency on the electric load seems to exist. Neglecting this dependency, we get an average value for G_m^V as well as for $G_m^V+G_p$ of about (10.5 ± 1.0) J/m^2. We assume that the small changes of up to 2 J/m^2 are significant and due to electric field effects of e.g., not completely screened polarization charges.

As far as we know, this is the first time that the intrinsic energy release rate was measured. In our opinion, care must be taken when comparing this result with other experiments such as those of Park and Sun [8], Fu and Zhang [12], or our own [23], which show different tendencies. In all the experiments performed before, it has not been distinguished between the linear part of the energy release rate and the energy release rate including domain switching effects.

4.3 Measured Fracture Resistance Curves at Different Applied Electric Fields. Even though the evaluation of the critical stress intensity factor during controlled crack growth is not the central issue of this article, it is easily possible to evaluate the measured crack growth data according to Eq. (18). The corresponding *R*-curves are shown in Fig. 12. These results must be treated carefully because they cannot be regarded as classical *R*-curves. The associated K_{IV}-curve would be needed additionally in order to give the full evaluation of the data which could be taken to identify a generalized mixed mode fracture criterion. (Therefore, the index "c" in K_{I}^c is written in the upper position.)

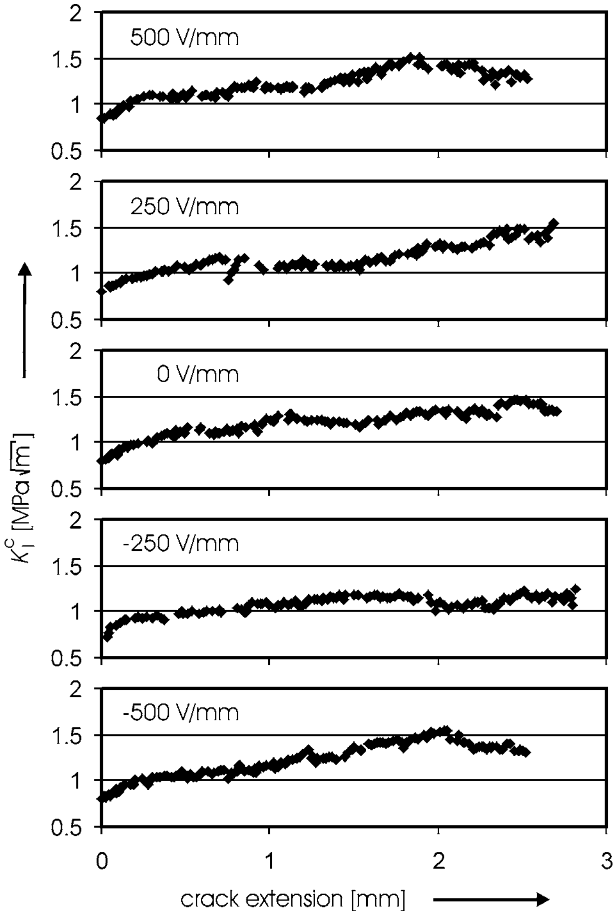


Fig. 12 Fracture resistance curves for external electric fields between −500 and 500 V/mm. The initial increase and also the main level are similar for all fields with the exception of −250 V/mm. Here, the curve proceeds slightly lower than the other ones.

Beside this it can be stated that under all applied electric fields, the fracture resistance curve is rising to a plateau value after about 0.3 mm crack advance.

From Fig. 12, we see that the *R*-curve at $E=-250$ V/mm proceeds significantly lower than the other curves, which is reflected by the tendency in Fig. 11. The corresponding mechanical energy release rate G_m is about 2 J/m^2 or 3 J/m^2 smaller than the values at the other electric fields. The slight unsteady increase of the *R*-curves differs from one to the other specimen, even at the same electric field. This is probably due to the individual crack path and residual stresses from poling in each specimen (more information is given further below).

4.4 Implications for the Stress Intensity Factor Applying Irwin's Equation. The crack closure integral leads to the Irwin matrix **H** (Suo et al. [3]), which establishes the link between the linear (crack tip) energy release rate and the crack tip stress intensity factor. The crack tip energy release rate is given by [3]

$$G_l^{\text{tip}} = \frac{1}{4}\mathbf{K}\cdot\mathbf{H}\cdot\mathbf{K}^T \tag{22}$$

with

$$\mathbf{K} = (K_{\text{II}}, K_{\text{I}}, K_{\text{III}}, K_{\text{IV}}) \tag{23}$$

and K_{I}, K_{II}, and K_{III} being the mechanical stress intensity factors and K_{IV} being the electric intensity factor at the crack tip. The symmetric Irwin matrix was calculated by Kemmer [30] for the material parameters of PZT PIC151 and since K_{II} and K_{III} are zero in four-point bending, we are left with

$$\begin{pmatrix} H_{22} & H_{24} \\ H_{42} & H_{44} \end{pmatrix} = \begin{pmatrix} 45.1\times10^{-12}\ \text{m}^2/\text{N} & 39.7\times10^{-3}\ \text{m}^2/\text{As} \\ 39.7\times10^{-3}\ \text{m}^2/\text{As} & -123\times10^{6}\ \text{V}^2/\text{N} \end{pmatrix} \tag{24}$$

Let $\Delta\varphi$ be the electric potential jump between the opposite crack faces. If we assume a permeable crack, then the condition $\Delta\varphi=0$ is equivalent to the following equation [3,23]:

$$K_{\text{IV}} = -\frac{H_{24}}{H_{44}}K_{\text{I}} \tag{25}$$

Inserting the expression of K_{IV} into Eq. (22) and using the quantities of Eq. (24), we get

$$\frac{K_{\text{I}}^2}{G_l^{\text{tip}}} = 4\left(H_{22} - \frac{H_{24}^2}{H_{44}}\right)^{-1} = 69\ \text{GPa} \tag{26}$$

During controlled crack growth, the crack tip energy release rate is physically this part of the released energy in the sample which is used to break the atomic bonds and therefore, in our terminology, the intrinsic toughness $\mathcal{G}_c^{\text{intr}}$. The experimentally measured plateau value of the intrinsic toughness for zero electric field is $\mathcal{G}_c^{\text{intr}}=G_m^V=11.5$ J/m^2.

The crack tip toughness is close to the starting value of the measured *R*-curves. Probably it is even a little bit less because already the frontal zone may lead to crack tip shielding. But since there is no other information available, we identify—as an attempt—the starting value of the *R*-curve with the intrinsic fracture toughness (without domain switching), being aware that this can be seen only as a rough estimate. For the measured K_{I}^C-curve without electric field (Fig. 12), we get $K^{\text{intr}}=0.82$ MPa$\sqrt{\text{m}}$ (averaged for two samples). Using the intrinsic toughness and intrinsic fracture toughness, we get $(K^{\text{intr}})^2/\mathcal{G}_c^{\text{intr}}=58$ GPa, which is in good agreement with the theoretical result of Eq. (26).

Furthermore, the FE calculations reveal that at zero electric field, K_{IV} for the permeable and impermeable crack are nearly the same within a few percent. Thus, if we assume the semipermeable crack being a "linear superposition" of permeable and impermeable crack [31], the above equations should be reasonable for all degrees of semipermeability.

4.5 Mechanical Inelastic Energies. Beside the intrinsic critical energy release rate $\mathcal{G}_c^{\text{intr}}$, we evaluate now the mechanical part of the total critical energy release rate $\mathcal{G}_c$ (Eq. (6)), including both intrinsic energies and all other remanent mechanical processes due to domain switching (area (E)+(F) in Fig. 5(*ii*)). This "energy release rate" $\mathcal{G}_c^{\text{mech}}$ is calculated from the load-displacement diagram as given by Sakai and Bradt [19]. Thus, we determine geometrically the area (E)+(F) (Fig. 5(*ii*)) and divide it by the corresponding new crack surface area d*A*. If we indicate the points (c) and (e) in Fig. 5(*ii*) with the indices "1" and "2" and taking into account that the slopes of the lines 1 and 2, are the inverse mechanical compliances $(1/C_m^V)$, we get:

$$\mathcal{G}_c^{\text{mech}} = \frac{C_{m1}^V F_1^2 + (F_1+F_2)(\Delta_2-\Delta_1) - C_{m2}^V F_2^2}{2\text{d}A} \tag{27}$$

According to Rose and Swain [20], this corresponds to the work of fracture, as said before. The equation cannot be applied directly to each pair of adjoining experimental data points because of the experimental scatter. Instead, we use an averaging procedure over several data points. In Eq. (27), F and Δ are averaged over each ten adjoining data points, C_m^V is taken from the analytical fit, and the result is averaged again as before. The work of fracture $\mathcal{G}_c^{\text{mech}}$, including the intrinsic toughness and also remanent processes, is given in Fig. 13 for applied electric fields of −500, 0, and

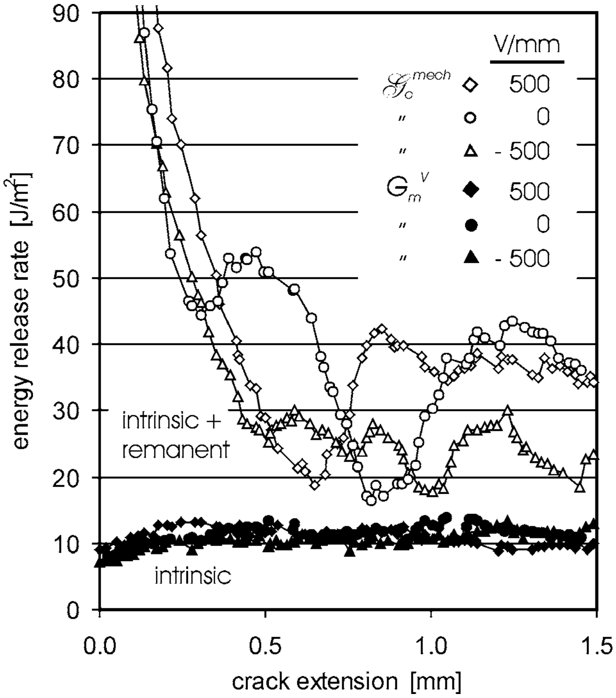


Fig. 13 Critical mechanical energy release rates, especially linear contribution (black points) and mechanical work of fracture (open points) for different applied electric fields (−500, 0, and +500 V/mm). The latter energy release rate comprises energies due to linear elastic processes as well as remanent energies.

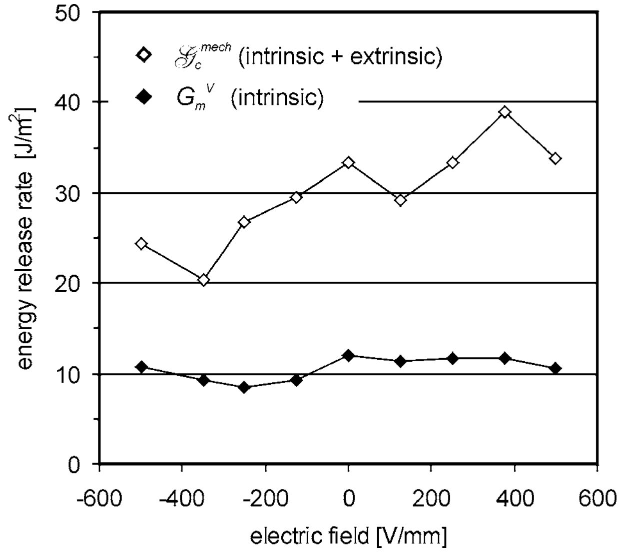


Fig. 14 Comparison of the "intrinsic" mechanical energy release rate G_m^V (lower curve, as in Fig. 11) with $\mathcal{G}_c^{\mathrm{mech}}$ (upper curve) including remanent switching processes during crack advance. When calculating the standard deviation for the data points between 0.5 mm and 1.5 mm crack extension (Fig. 13), which gives a rough estimate for the error of the data, we get for the lower curve around ±1 J/mm² and for the upper curve about ±10 J/mm² (the latter number seems a little bit high.)

500 V/mm.

The high initial values of $\mathcal{G}_c^{\mathrm{mech}}$ are based on energies from creating the frontal process zone. This effect was observed also by Rose and Swain [20] for partially stabilized zirconia. There are strong variations of $\mathcal{G}_c^{\mathrm{mech}}$ for crack extensions above 0.5 mm, which are different for each specimen, even if the same electric field is applied. Thus, it seems that $\mathcal{G}_c^{\mathrm{mech}}$ is quite sensitive on the individual crack path. This would include e.g., crack bifurcation, double cracks, and crack bridging. It means that these fluctuations are only qualitatively significant, but not their individual shapes. (Beside this we have to admit that the variations are possibly influenced by the evaluation procedure.)

On the other hand, the critical energy release rate G_m^V does not exhibit such variations, implying that the small signal compliance method is quite insensitive to the individual crack path, crack bridging, etc. For crack extensions above 0.5 mm, $\mathcal{G}_c^{\mathrm{mech}}$ is about two to three times larger than $\mathcal{G}_c^{\mathrm{intr}}$. Thus, in our four-point-bending configuration, remanent switching processes, being the difference $\mathcal{G}_c^{\mathrm{mech}}-\mathcal{G}_c^{\mathrm{intr}}$, are up to twice as large as $\mathcal{G}_c^{\mathrm{intr}}$.

From Fig. 14, showing medium mechanical energy release rates (averaged between 0.5 mm and 1.5 mm crack extension) for different electric fields, we get the same result. For example, during the measurement at $E=375$ V/mm, much more crack branching and bridging was observed than in the other measurements, due to the properties of the individual specimen. This can be seen by the increased values in the upper curve (Fig. 14) at this particular electric field, but not in the lower curve for $\mathcal{G}_c^{\mathrm{intr}}$. Looking at the general trend, the contribution to the energy release rate from remanent switching processes increases from −500 V/mm to 500 V/mm by about a factor of 2.

From load-displacement curves in which we unloaded the specimen completely just at the moment of crack initiation, we can evaluate the energy due to domain switching in the frontal zone. The energy stored is about $W_F=120$ μJ for our sample geometry (measured at zero external electric field).

There may be some concerns if so much energy is put only into the frontal zone or whether large scale deformation takes place. Kounga Njiwa et al. [32] used a liquid crystal display technique to monitor the electric surface potentials in front of the crack tip and estimated a switching zone size of 600 μm to 800 μm. Hackemann and Pfeiffer [33] applied a spatially focused X-ray diffraction technique and determined about 290 μm switching zone size. Both experiments were performed with PZT PIC151 and demonstrate a substantial switching zone. These experiments show that no large scale domain switching processes take place. A simple estimate confirms these experimental findings. If the remanent polarization of 0.32 C/m² [34] times 2 and multiplied by the coercive field of 0.85 kV/mm is taken as a measure for the energy density change during domain switching [35], we get about $w_s=0.0005$ J/mm³. If we roughly approximate the frontal zone by a circle of area S going through the sample of thickness $t=3$ mm, S is calculated to be $S=W_F/w_S t$. For the linear dimension (diameter) of the switching zone size, this finally gives $2\sqrt{S/\pi}\approx 0.3$ mm, which corresponds well to the experimental results of [32,33].

In principle, the compliance change during crack extension is also affected by the growth of the process zone because the linear piezoelectric tensor is not isotropic. A polarization change may change the dielectric and elastic properties by approximately 10%. The volume change of an assumed remanent process zone of width 1 mm during a measurable crack extension dA of 50 μm $\times 3$ mm (sample thickness) is 0.15 mm³. Compared with the loaded sample volume of about 240 mm³, this leads to a negligible compliance change of less than 0.1%.

Thus, by evaluating the linear part of the energy release rate together with the load-displacement curve as described before, we can separate the linear (crack tip) energy release rate from switching zone processes. If in previous experiments these energy release rates were not separated, the detailed interpretation of the results sometimes seems difficult.

4.6 Electric Dissipative Energies. If $\mathcal{G}_c^{\text{elec}}$ is the electric energy release rate, including intrinsic as well as dissipative energies, we get for the toughness

$$\mathcal{G}_c = \mathcal{G}_c^{\text{mech}} + \mathcal{G}_c^{\text{elec}} \tag{28}$$

Whereas $\mathcal{G}_c^{\text{mech}}$ is determined from the load-displacement diagram using the measured mechanical compliance, the term $\mathcal{G}_c^{\text{elec}}$ can be derived in principle from a corresponding charge-voltage diagram using the measured capacitance.

A remark concerning the loading path must be added. The experiments with electric load are started by increasing the voltage up to a constant value before the sample is loaded mechanically. During crack growth, the voltage is kept at a fixed level by the power supply. With respect to the load-displacement diagram, an extension from the pure mechanical to the electromechanical case can be achieved, if we vary the electric load proportionally to the mechanical load. We would get a corresponding diagram with a generalized load and a generalized displacement. Unfortunately, in our case the situation is more complicated. On the other hand, the results show that the influence of the electric load is weak and therefore, we neglect this problem. The following should be seen as a rough estimate concerning nonlinear electrical processes.

As mentioned before, $\mathcal{G}_c^{\text{elec}}$ could not be measured, since the long-run charge measurement was disturbed by the capture of electromagnetic radio frequencies. Note that within the measured quantities in Table 1, only the charge modulation $\mathrm{d}Q_l$ is measured at constant crack length, whereas the accumulated charge Q during crack advance and in between is missing. Under electrical aspects, up to now we determined the state of the sample just before and after each crack advance.

In order to still get information about electrical processes during actual crack growth, the charge was measured during crack advance once with a poled PZT specimen at an electric field of 250 V/mm. Each crack advance takes about 1 or 2 s, so that electromagnetic disturbances during this time have only little influence. The electric current is automatically integrated during each measurement and divided afterwards by the area of the created new crack surface. The average charge generated immediately per new crack surface area was about $(410\pm200)\ \mu\mathrm{C/m^2}$ measured in the range of crack extensions between 0.5 mm and 1.5 mm. This phenomenon is clearly no piezoelectric effect, as could be seen in "short-time Q-F diagrams."

After stopping the crack the electric current does not stop immediately but decreases slowly. Such a delayed "time effect" in the behavior of PZT was observed on other occasions. However, it does not influence the measurement of the intrinsic toughness.

The measured charges $(410\ \mu\mathrm{C/m^2})$ if attributed to crack surface charges are less than 0.2% of the remanent polarization of $0.32\ \mathrm{C/m^2}$ for poled PZT PIC151. This is a strong indication that the remanent crack surface polarization charges are balanced by internal processes such as local domain switching or space charges. On the other hand, the dielectric polarization charge $P = \varepsilon_0\chi E$ for our material parameters gives about $2000\ \mu\mathrm{C/m^2}$, which is still higher than but much closer to the measured values. In summary, these results support the idea of screening crack surface charges or breakdown effects, but are not sufficient to demonstrate them, and additional, more refined measurements are necessary.

Concerning the total energy balance, a future experiment for measuring the total energy release rate including all electric dissipative processes is technically possible and might help to answer the remaining questions. However, beside this we hope that the present experimental data provide a useful basis for further discussions.

5 Summary

The intrinsic toughness as well as the extrinsic nonlinear domain switching toughening in polarized PZT PIC151 is measured. During stable crack growth, the mechanical and the piezoelectrical compliance as well as the electrical capacitance of the PZT sample is measured simultaneously as a function of the crack length using modulation/demodulation techniques. Computation of the derivatives of these generalized compliances with respect to the crack surface area allows for the calculation of the mechanical, the piezoelectric, and the electrical part of the energy release rate due to linear processes. To our knowledge, this is the first experimental result where these linear parts are well separated from nonlinear, especially domain switching contributions, although Rose and Swain [20] as well as Sakai and Bradt [19] had presented the basic principle before.

This novel experimental technique, based on the small signal modulation and stable crack growth, is easily applicable also to other ceramics with process zones in order to separate the intrinsic toughness from extrinsic nonlinear contributions.

With the assumption of screened polarization charges—for example due to electric breakdown—and by neglecting the slight electric field dependency, an average intrinsic toughness of $(10.5\pm1.0)\ \mathrm{J/m^2}$ for electric fields between -500 V/mm and 500 V/mm is determined. It seems probable that the slight field dependency of the measured intrinsic toughness is due to small contributions of the linear electric part of the energy release rate.

The information of the mechanical compliance and the load-displacement curve allows for the evaluation of mechanical energy changes related to domain switching in the vicinity of the crack. Additionally, a frontal process zone is identified. From the corresponding energy, a frontal zone size of the order of 0.3 mm can be estimated consistently to other experimental results. The mechanical toughening effect of the process zone is increasing from negative to positive external electric fields.

Acknowledgment

The support of this work by the German Research Foundation (DFG) is gratefully acknowledged. Furthermore, we thank T. Scholz as well as D. Schmidt from our group for their effective support regarding the realization of the experimental setup. We are indebted to C. Häusler from the Technical University Dresden for the idea of equal support distances in four-point bending, providing a method for the correct data preparation. We also thank P. Neumeister from the Technical University Dresden, as well as H. Keßler from Framatome ANP, Erlangen, and R. McMeeking from the University of California, Santa Barbara for valuable hints and discussions.

Appendix: Data Processing

Evaluation of C_m^V and C_p

The mechanical compliance is measured at constant voltage. To simplify matters, the superscript "V" is omitted. The finite compliance of the experimental device leads to corrections of the measured C_m and C_p values. The essential mechanical arrangement is given in Fig. 15, where $\mathrm{d}F$ and $\mathrm{d}\Delta$ are the measured 5 Hz modulation amplitudes of force and displacement, respectively. Thus, the measured mechanical (mm) compliance, simply, is $C_{mm} = \mathrm{d}\Delta/\mathrm{d}F$. To obtain the actual bending compliance of the specimen (C_m), C_{mm} must be decomposed according to Fig. 15.

Obviously, the compliance of the mechanical support of the specimen directly increases the overall compliance as described by the series connection. In addition, the specimen is compressed slightly by the support rollers. Both aspects are included in C_{m01}. Mainly because of the circular shape of the support rollers, C_{m01} is dependent on the applied load F. It is measured by using equal support distances of 20 mm each, as shown in Fig. 16(*b*). Taking the arrangements in Fig. 16, the subtraction (a)−(b) for each applied mechanical load F yields pure bending of the specimen. Concerning Fig. 15, the case in Fig. 16(*b*) means a parallel connection of C_{m01} and C_{m02} only.

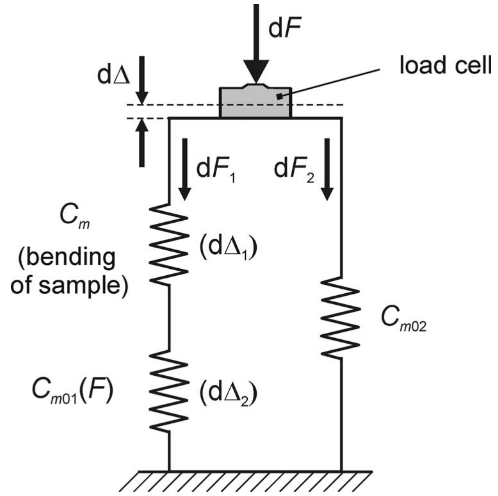


Fig. 15 Schematic arrangement of mechanical compliances of the specimen and the equipment with $C_{m01} \ll C_m \ll C_{m02}$. The parameters $d\Delta_1$ as well as $d\Delta_2$ are the displacements, belonging to C_m and C_{m01}. The latter one represents the compliance of the mechanical frame plus the compression compliance of the specimen and C_{m02} corresponds to small elastic deformations in the mounting and to inertia effects of the upper movable support.

Due to the dynamic modulation measurement, additional effects must be taken into account. Since the modulation amplitude is very small, frictional forces pin the rotatable wheel of the dead-weight compensation (Fig. 2, right), which leads to a parallel connection of the decoupling spring. A similar effect could exist in the support guide. Together with the influence of the inertia of the slightly oscillating upper support mass, these effects can be summarized using an apparent compliance C_{m02} connected in parallel. This quantity can be determined directly by performing a measurement without a specimen, simulating an "infinite" specimen compliance. We get $C_{m02}=(16\pm2)$ μm/N for both 10/20 and 20/20 mm supports. The compliance C_{m01} depends on the mechanical load F, whereas C_{m02} does not. Since the frequency of 5 Hz is much smaller than the resonance frequency of the oscillating mass, corresponding effects can be neglected.

With regard to Fig. 15 we have to consider at first C_{m02} for both arrangements (a) and (b) and then subtract the resulting compliances from each other. Thus, the pure bending compliance of the specimen is

$$C_m = \frac{C_{mm}}{(1 - C_{mm}/C_{m02})} - \frac{C_{m01}}{(1 - C_{m01}/C_{m02})} \tag{A1}$$

The correction of the piezoelectric compliance is similar. From the measured quantities using the 10/20 mm support we have $C_{pm} = dQ/dF$ and the analog value for the 20/20 mm support is C_{p01}. On the basis of Figs. 15 and 16 with $dF=dF_1+dF_2$ and $d\Delta = d\Delta_1 + d\Delta_2$ (see Fig. 15) and replacing the displacements $d\Delta$, $d\Delta_1$, and $d\Delta_2$ by corresponding charge quantities, we obtain a similar expression for the piezoelectric compliance with respect to pure bending load:

$$C_p = \frac{C_{pm}}{(1 - C_{mm}/C_{m02})} - \frac{C_{p01}}{(1 - C_{m01}/C_{m02})} \tag{A2}$$

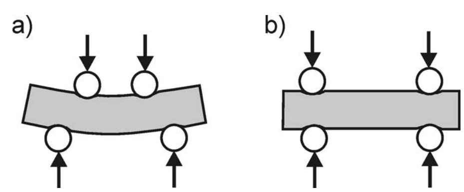


Fig. 16 Schematic drawing of the mechanical load configuration: (*a*) bending and compression of the specimen as well as loading of the device, (*b*) compression of the specimen and loading of the device. The difference (a)−(b) yields the pure bending compliance of the specimen.

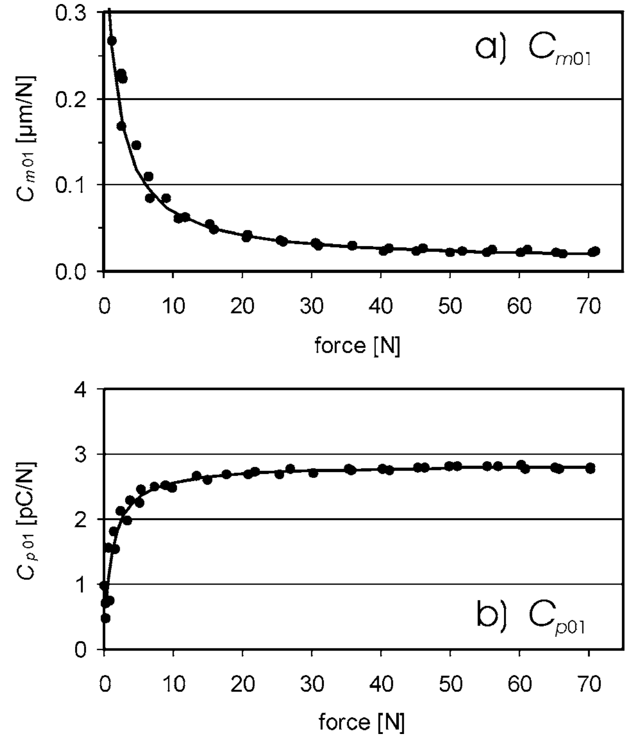


Fig. 17 (*a*) Mechanical and (*b*) piezoelectric compliance measured with the 20/20 mm supports (Fig. 16(*b*)).

Note that a quantity C_{p02} does not exist, since without a specimen, no electric current can be measured. The denominators in Eqs. (A1) and (A2) are identical, correspondingly. The correction term C_{p01}, measured with the arrangement in Fig. 16(*b*), is slightly force dependent and of the order of 2 or 3 pC/N. The main difference between the mechanical and the piezoelectric case is that C_{m01} consists of *two* parts, i.e., the device compliance and the compression compliance of the sample, and that C_{p01} consists only of the piezoelectric compliance due to the compression of the sample. The associated corrections are approximately 5% to 30%, because mostly $C_{m01} \ll C_m \ll C_{m02}$ in the range of the analyzed crack lengths. To apply Eqs. (A1) and (A2), the compliances $C_{m01}(F)$ and $C_{p01}(F)$ are fitted by appropriate analytical functions (see Eqs. (A3) and (A4)). The corrected values of C_m and C_p are used in Eq. (17) for calculating C_e^F. The capacitance itself does not need any correction.

The mechanical and the piezoelectric compliance C_{m01} and C_{p01} are given in Fig. 17 as a function of the applied mechanical load. At relevant forces of 10 N to 50 N, the mechanical correction term C_{m01} still decreases by a factor of 3 (Fig. 17(*a*)), whereas the piezoelectric term C_{p01} is almost constant around 2.7 pC/N (Fig. 17(*b*)). The latter value varies for different PZT samples by about ±10%. The functions fitted to the data are

$$C_{m01} = \left(\frac{0.6854}{F[\mathrm{N}] + 1.5896} + 0.01034\right)\left[\frac{\mu\mathrm{m}}{\mathrm{N}}\right] \tag{A3}$$

and

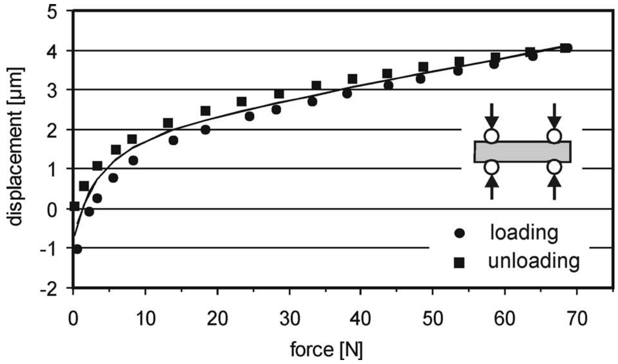


Fig. 18 Displacement as a function of the force, measured with equal support distances as in Fig. 16(*b*). As described before, this effect corresponds to the compliance of the mechanical support and to the compression of the sample. The zero position on the displacement axis is arbitrary. The slightly different paths for loading and unloading are fitted by the same curve (Eq. (A5)).

$$C_{p01} = \left(\frac{-3.1857}{F[\mathrm{N}] + 1.2039} + 2.8452 \right) \left[\frac{\mathrm{pC}}{\mathrm{N}} \right] \tag{A4}$$

Note that in Fig. 17, the compliance measurements are not dependent on whether the force increases or decreases. This emphasizes once more that the small signal compliance measurements are not influenced by the actual domain configuration and by switching processes.

Displacement Correction

The reasons leading to C_{m01} also yield an additional displacement, which has to be subtracted from the measured displacement. As before, this displacement is force dependent and determined directly by using the 20/20 mm supports. The data shown in Fig. 18 are fitted by the following analytical function:

$$\Delta_0(F) = \left(0.03399F[\mathrm{N}] + 1.800 - \frac{125.47}{(F[\mathrm{N}] + 7.00)^2} \right) [\mu\mathrm{m}] \tag{A5}$$

This function, describing the finite stiffness of the device and the compression of the sample, is subtracted from the measured displacement. The quantity C_{m02} originates from dynamical effects and therefore does not influence the quasistatic load-displacement measurement.

In principle, the data in Fig. 18 must be fitted separately for the loading and the unloading path. Nevertheless, this would make the data evaluation even more complicated. As this correction is relevant only for the load-displacement diagrams and not for the small signal modulation measurements, we use only the single curve in Fig. 18.

Analytical Functions Fitted to the Measured Compliances. The best results for the mechanical compliance are obtained using a hyperbolic function with $1/a$ characteristic, namely,

$$C_m(a) = \frac{p_1}{a - p_2} + p_3 \tag{A6}$$

where p_1, p_2, and p_3 are free parameters and a is the crack extension. For the fit of the piezoelectric compliance data, a logarithmic function is used:

$$C_p(a) = p_1 \ln\left(\frac{p_2 - a}{mm} \right) + p_3 \tag{A7}$$

and the capacitance is fitted by a second-order polynomial:

$$C_e(a) = p_1 a^2 + p_2 a + p_3 \tag{A8}$$

These functions are not chosen accidentally. In the case of using polynomial functions for C_m and C_p, the corresponding derivatives $\partial C_m / \partial a$ and $\partial C_p / \partial a$ and energy release rates exhibit strong variations depending on the order of the polynomial used, because the derivatives of C_m and C_p are rather small for short crack lengths. To avoid this, smoother functions should be used, which lead to more constant energy release rates. Generally, if we plot the applied force F as a function of the crack length, we obtain a nearly linear decrease with respect to the crack extension a. Hence, $\partial C_m / \partial a$ should have a $1/a^2$ characteristic to compensate the F^2 influence in Eq. (12). Integration of $\partial C_m / \partial a$ leads to the $1/a$ characteristic, which is finally used for C_m in Eq. (A6). The argument for C_p is analogous. For C_e^F, a polynomial of second order is sufficient. It is important that the choice of these functions has no physical intention. However, these functions require fewer free parameters than polynomials and fit very well to the experimental data.

The results are also sensitive to the range of crack extensions where these functions are fitted. To avoid an influence of the process zone, when it reaches the specimens's "back side," the compliances and the capacitance are mainly fitted in the range between 0 mm and 1.6 mm crack extension.

References

[1] Haertling, G. H., 1999, "Ferroelectric Ceramics: History and Technology," J. Am. Ceram. Soc., **82**(4), pp. 797–818.

[2] Freiman, S. W., and Pohanka, R. C., 1989, "Review of Mechanically Related Failures of Ceramic Capacitors," J. Am. Ceram. Soc., **72**(12), pp. 2258–2263.

[3] Suo, Z., Kuo, C.-M., Barnett, D. M., and Willis, J. R., 1992, "Fracture Mechanics for Piezoelectric Ceramics," J. Mech. Phys. Solids, **40**(4), pp. 739–765.

[4] McMeeking, R. M., 1989, "Electrostrictive Stresses Near Crack-Like Flaws," ZAMP, **40**, pp. 615–627.

[5] McMeeking, R. M., 2001, "Towards a Fracture Mechanics for Brittle Piezoelectric and Dielectric Materials," Int. J. Fract., **108**(1), pp. 25–41.

[6] Guiu, F., Algueró, M., and Reece, M. J., 2003, "Crack Extension Force and Rate of Mechanical Work of Fracture in Linear Dielectrics and Piezoelectrics," Philos. Mag., **83**(7), pp. 873–888.

[7] Zhang, T. Y., Zhao, M., and Tong, P., 2002, "Fracture of Piezoelectric Ceramics," Adv. Appl. Mech., **38**, pp. 147–289.

[8] Park, S., and Sun, C.-T., 1995, "Fracture Criteria for Piezoelectric Ceramics," J. Am. Ceram. Soc., **78**(6), pp. 1475–1480.

[9] Tobin, A. G., and Pak, Y. E., 1993, "Effect of Electric Fields on Fracture Behavior of PZT Ceramics," Proc. SPIE, **1916**, pp. 78–86.

[10] Wang, H., and Singh, R. N., 1997, "Crack Propagation in Piezoelectric Ceramics: Effects of Applied Electric Fields," J. Appl. Phys., **81**(11), pp. 7471–7479.

[11] Lynch, C. S., 1998, "Fracture of Ferroelectric and Relaxor Electro-Ceramics: Influence of Electric Field," Acta Mater., **46**(2), pp. 599–608.

[12] Fu, R., and Zhang, T. Y., 2000, "Effects of an Electric Field on the Fracture Toughness of Poled Lead Zirconate Titanate Ceramics," J. Am. Ceram. Soc., **83**(5), pp. 1215–1218.

[13] Schneider, G. A., and Heyer, V., 1999, "Influence of the Electric Field on Vickers Indentation Crack Growth in $BaTiO_3$," J. Eur. Ceram. Soc., **19**, pp. 1299–1306.

[14] Schneider, G. A., Felten, F., and McMeeking, R. M., 2003, "The Electrical Potential Difference Across Cracks in PZT Measured by Kelvin Probe Microscopy and the Implications for Fracture," Acta Mater., **51**, pp. 2235–2241.

[15] Balke, H., Kemmer, G., and Drescher, J., 1997, "Some Remarks on Fracture Mechanics of Piezoelectric Solids," *Proceedings of the International Conference and Exhibition of Micro Materials'97*, B. Michel and T. Winkler, eds., pp. 398–401.

[16] Haug, A., and McMeeking, R., 2006, "Cracks With Surface Charge in Poled Ferroelectrics," Eur. J. Mech. A/Solids, **25**, pp. 24–41.

[17] Landis, C. M., 2004, "Energetically Consistent Boundary Conditions for Electromechanical Fracture," Int. J. Solids Struct., **41**, pp. 6291–6315.

[18] Jelitto, H., Felten, F., Häusler, C., Kessler, H., Balke, H., and Schneider, G. A., 2005, "Measurement of Energy Release Rates for Cracks in PZT Under Electromechanical Loads," Electroceramics 2004, J. Eur. Ceram. Soc., **25**, pp. 2817–2820.

[19] Sakai, M., and Bradt, R. C., 1986, "Graphical Methods for Determining the Nonlinear Fracture Parameters of Silica and Graphite Refractory Composites," *Fourth International Symposium on the Fracture Mechanics of Ceramics*, VPI, Chicago, June 19–21, 1985, Plenum Press, New York, Vol. 7, pp. 127–142.

[20] Rose, L. R. F., and Swain, M. V., 1986, "Two R-Curves for Partially Stabilized Zirkonia," J. Am. Ceram. Soc., **69**(3), pp. 203–207.

[21] Kreher, W. S., 2002, "Influence of the Domain Switching Zones on the Fracture Toughness of Ferroelectrics," J. Mech. Phys. Solids, **50**, pp. 1029–1050.

[22] Suo, Z., 1991, "Mechanics Concepts for Failure in Ferroelectric Ceramics,"

Smart Structures and Materials, ASME 1991, AD-Vol. 24/AMD-Vol. 123, pp. 1–6.
[23] Jelitto, H., Keßler, H., Schneider, G. A., and Balke, H., 2004, "Fracture Behavior of Poled Piezoelectric PZT Under Mechanical and Electrical Loads," J. Eur. Ceram. Soc., **25**(5), pp. 749–757.
[24] Kübler, J., 1998, "Bestimmung der Bruchzähigkeit keramischer Werkstoffe mit der SEVNB Methode: Resultate eines VAMAS/ESIS Ringversuches," in *Proceedings of the Werkstoffwoche*, EMPA, Dubendorf, Switzerland.
[25] Kübler, J., 2001, "Fracture Toughness of Ceramics Using the SEVNB Method: From a Preliminary Study to a Standard Test Method," in *Fracture Resistance Testing of Monolithic and Composite Brittle Materials, ASTM STP 1409*, J. A. Salem, M. G. Jenkins, and G. D. Quinn, eds., American Society for Testing and Materials, West Conshohocken, PA.
[26] Fett, T., Munz, D., and Thun, G., 1995, "Evaluation of Bridging Parameters in Aluminas From *R*-Curves by Use of the Fracture Mechanical Weight Function," J. Am. Ceram. Soc., **78**(4), pp. 949–951.
[27] Jelitto, H., Felten, F., and Schneider, G. A., 2005, "Experimenteller Aufbau zur Messung der Energiefreisetzungsrate für Risswachstum in PZT unter elektromechanischer Last," DVM-Bericht 237, 37. Tagung des DVM-Arbeitskreises Bruchvorgänge, Technische Sicherheit, Zuverlässigkeit und Lebensdauer, pp. 365–372.
[28] Heyer, V., Schneider, G. A., Balke, H., Drescher, J., and Bahr, H.-A., 1998, "A Fracture Criterion for Conducting Cracks in Homogeneously Poled Piezoelectric PZT-PIC151 Ceramics," Acta Mater., **46**(18), pp. 6615–6622.
[29] Muñoz-Saldaña, J., Schneider, G. A., and Eng, L. M., 2001, "Stress Induced Movement of Ferroelastic Domain Walls in $BaTiO_3$ Single Crystals Evaluated by Scanning Force Microscopy," Surf. Sci., **480**, pp. L402–L410.
[30] Kemmer, G., 2000, "Berechnung von elektromagnetischen Intensitätsparametern bei Rissen in Piezokeramiken," Fortschritt-Berichte VDI, Reihe **18**, Nr. 261, VDI Verlag, Düsseldorf (in German), p. 33.
[31] Kessler, H., Balke, H., Jelitto, H., and Schneider, G. A., 2004, "An Approximation for Electrically Semipermeable Edge Cracks and its Application to Fracture Analysis of PZT," Proc. Appl. Math. Mech., **4**, pp. 282–283.
[32] Kounga Njiwa, A. B., Lupascu, D. C., and Rödel, J., 2004, "Crack Tip Switching Zone in Ferroelectric Ferroelastic Materials," Acta Mater., **52**, pp. 4919–4927.
[33] Hackemann, S., and Pfeiffer, W., 2003, "Domain Switching in Process Zones of PZT: Characterization by Microdiffraction and Fracture Mechanical Methods," J. Eur. Ceram. Soc., **23**, pp. 141–151.
[34] Kolleck, A., 2000, "Einfluß der ferroelastischen Domänenschaltprozesse auf die Bruchzähigkeit und Bruchfestigkeit von $BaTiO_3$ und PZT," Fortschritt-Berichte VDI, Reihe **5**, Nr. 614, VDI Verlag, Düsseldorf (in German), pp. 159–160.
[35] Hwang, S. C., Lynch, C. S., and McMeeking, R. M., 1995, "Ferroelectric/Ferroelastic Interactions and a Polarization Switching Model," Acta Metall. Mater., **43**(5), pp. 2073–2084.

Steady-State Dynamic Response of a Kirchhoff's Slab on Viscoelastic Kelvin's Foundation to Moving Harmonic Loads

Lu Sun
Transportation College,
Southeast University,
Nanjing 210096, China;
Department of Civil Engineering,
The Catholic University of America,
Washington, DC 20064
e-mail: e-mail: sunl@cua.edu

In this paper, fast Fourier transform and complex analysis are used to analyze the dynamic response of slabs on a viscoelastic foundation caused by a moving harmonic load. Critical speed and resonance frequency of the slab to a moving harmonic load are obtained analytically. It is proved that there exists a bifurcation in critical speed. One branch of critical speed increases as load frequency increases, while the other branch of critical speed decreases as load frequency increases. There are two critical speeds when the load frequency is low, but only one critical speed exists when the load frequency is high. A parametric study is also performed to study the effect of load speed, load frequency, material properties of the slab and the damping coefficient on dynamic response. It is found that the damping coefficient has significant influence on dynamic response. For small damping, the maximum response of the slab increases with increased load speed and frequency. However, for large damping, the maximum response of the slab decreases with increased load speed and frequency. [DOI: 10.1115/1.2744033]

Keywords: slab, moving load, critical speed, complex analysis, fast Fourier transform, algebraic equation

1 Introduction

The subject of pavement response to moving loads is of great importance for pavement design and transportation infrastructure management (Yoder and Witczak [1], Haas et al. [2], Sun and Deng [3], Sun [4], and Sun et al. [5–7]). Highway and airfield pavements need to provide sufficient structural capacity to carry vehicular loads, (Monismith et al. [8], Sun and Deng [3], and Sun [4]). Existing pavement design methods are based on the response of pavements to static loads. The importance of dynamic pavement response to moving load increases due to the trend of high-speed surface transportation in recent years (Sun [9–12], Kim and Roesset [13], Kononov and Dieterman [14], Chen and Huang [15,16], Clouteau et al. [17], Shamalta and Metrikine [18], and Sun et al. [5–7]).

Bush [19], Scullion et al. [20], and Uzan and Lytton [21] used measured pavement dynamic response information to investigate pavement nondestructive evaluation. Salawu and Williams [22] studied the full-scale force-vibration test before and after structural repairs on bridges. Sun and Deng [3] and Sun and Kennedy [23] studied the effects of pavement surface roughness and vehicle suspension system on dynamic pavement loads. With the recent development of a rolling weight deflectometer (Lee et al. [24]), it is now possible to efficiently collect continuous deflection data of highway and airfield pavements at a highway speed for transportation infrastructure management purposes. The vast amount of data needs to be interpreted in order to assess highway and airport conditions. This application essentially falls into pavement nondestructive testing and evaluation. Being able to characterize the forward dynamic response of pavement to moving loads is a critical key component for ensuring the success of the development of nondestructive evaluation algorithms. Pavement response to a moving load becomes valuable for these purposes.

The majority of highway and airport pavements are made of asphalt cement or concrete cement. The former is often named the flexible pavement, while the latter the rigid pavement. Pavement mechanics uses multilayered systems to characterize flexible pavement and thin slabs on a Winkler elastic foundation to characterize rigid pavements. In this paper the case of rigid pavements is considered, and the case of flexible pavement was addressed elsewhere (Sun and Greenberg [25]). Kenney [26] studied the steady-state response of a moving load on a beam on elastic foundation. Finite element procedures and integral transforms have also been developed to carry out the response of a thin slab to dynamic loads with applications in pavement design and nondestructive evaluation (Taheri [27], Kukreti et al. [28], Zaghloul et al. [29], Kim and Roesset [13], Sun [10,12,30,31], and Sun et al. [5–7]). Deshun [32] applied the vibrational principle to solve the vibration of thick slabs. The vertical vibration of an elastic slab on a fluid-saturated porous half-space subjected to a harmonic load was investigated by Bo [33], in which the Hankel transform is used to convert the governing equation to the Fredholm integral equation of the second order and a numerical calculation can then be carried out. In existing studies, pavement foundation is often treated as an elastic Winkler foundation. Due to the presence of damping in subgrade material, it is more appropriate to use a viscoelastic foundation model.

One-dimensional structures, such as beams, are studied most extensively in many of the existing studies (Sun [9,10,30]). In this paper, we apply Fourier transform and complex analysis to investigate a two-dimensional structure. Specifically, the dynamic response of an infinite slab on viscoelastic foundation subject to moving constant loads and moving harmonic loads is studied. A general dynamic response of slab to arbitrary dynamic loads is obtained, which contains quiescence harmonic loads, a moving constant load, and moving harmonic loads as special cases. The critical speed and resonance frequency of the slab is derived analytically, which has not been seen in the literature. Another original aspect of the paper is the development of a fast Fourier trans-

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received May 8, 2005; final manuscript received April 27, 2007. Review conducted by Ellen M. Arruda.

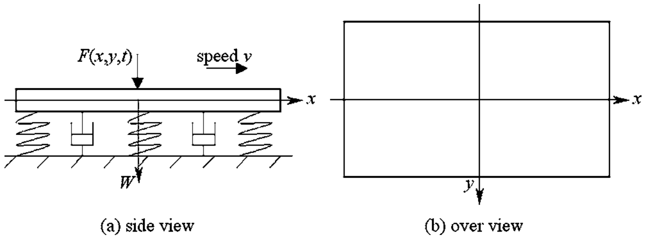


Fig. 1 An infinite slab on a viscoelastic foundation subjected to a moving load

form (FFT)-based efficient computation schema and a parametric study for numerically evaluating the dynamic response of a slab under moving loads.

This paper is organized as follows. In Sec. 2, the governing equation of a thin slab on a viscoelastic foundation is established. In Sec. 3, the Fourier transform is applied to develop the solution of pavement response under various types of dynamic loads with a special emphasis on moving loads. In Sec. 4, critical speed and resonance frequency are derived. In Sec. 5, the dynamic coefficient is defined to quantify the effect of a moving load. In Sec. 6, computational implementation of a slab response to a moving load is formulated and the parametric study is conducted. In Sec. 7, concluding remarks are made.

2 The Governing Equation and General Responses

Figure 1 shows a thin slab in an orthogonal x-y-z coordinate system. Denote the displacement of the slab in the z direction as $W(x,y,t)$. Three assumptions are made to simplify the mathematical model of a thin slab. These assumptions are: (1) the strain component ε_z in the perpendicular direction of the slab is sufficiently small such that it can be ignored; (2) the stress components τ_{zx}, τ_{zy}, and σ_z are far less than the other stress components; therefore, the deformation caused by τ_{zx}, τ_{zy}, and σ_z can be negligible; and (3) the displacement parallel to the horizontal direction of the slab is zero.

Let the reaction force from the viscoelastic foundation to the slab be given by $q(x,y,t)=KW(x,y,t)+C\partial(x,y,t)/\partial t$ (Kenney [26], Zaghloul et al. [29], and Sun [12,34]). Here, K is the modulus of the subgrade reaction and is assumed to be constant to represent linear elasticity of the subgrade, and C is the damping coefficient of the foundation. The well-known governing equation for a Kirchhoff's slab is (Shamalta and Metrikine [18] and Sun [11,34])

$$D\nabla^2\nabla^2 W(x,y,t) + KW(x,y,t) + C\frac{\partial W(x,y,t)}{\partial t} + \rho h \frac{\partial^2}{\partial t^2}W(x,y,t) = F(x,y,t) \quad (1)$$

where $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2$ is the Laplace operator, h is the thickness of the slab, ρ is the density of the slab, $D=Eh^3/[12(1-\mu^2)]$ is the stiffness of the slab; E and μ are the Young's modulus of elasticity and Poisson ratio of the slab, respectively.

For the sake of seeking a steady-state solution, it is assumed that the load be applied at time $t=-\infty$ such that all transient effects vanish. Equation (1) is a linear partial differential equation, and can be solved using the approach of integral transformation. We define the three-dimensional Fourier transform pair

$$\tilde{f}(\boldsymbol{\xi}) = \mathbf{F}[f(\mathbf{x})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\mathbf{x})\exp(-i\boldsymbol{\xi}x)d\mathbf{x} \quad (2a)$$

$$f(\mathbf{x}) = \mathbf{F}^{-1}[\tilde{f}(\boldsymbol{\xi})] = (2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{f}(\boldsymbol{\xi})\exp(i\boldsymbol{\xi}x)d\boldsymbol{\xi} \quad (2b)$$

where $\mathbf{x}=(x,y,t)$, $\boldsymbol{\xi}=(\xi,\eta,\omega)$, $\mathbf{F}[\cdot]$ and $\mathbf{F}^{-1}[\cdot]$ are Fourier transform and its inversion, respectively. Since only the steady-state solution is of interest, Fourier transform can be applied to both sides of (1):

$$D(\xi^2+\eta^2)^2\tilde{W}(\boldsymbol{\xi}) + K\tilde{W}(\boldsymbol{\xi}) + iC\omega\tilde{W}(\boldsymbol{\xi}) - \rho h\omega^2\tilde{W}(\boldsymbol{\xi}) = \tilde{F}(\boldsymbol{\xi}) \quad (3)$$

where $\tilde{F}(\boldsymbol{\xi})$ and $\tilde{W}(\boldsymbol{\xi})$ are the Fourier transform of $F(\mathbf{x})$ and $W(\mathbf{x})$, respectively. In the derivation of (3), the following property of the Fourier transform is used:

$$\mathbf{F}[f^{(n)}(t)] = (i\omega)^n\mathbf{F}[f(t)] \quad (4)$$

Equation (4) is an algebraic equation and the displacement response in the transformed domain can be easily obtained by rearranging the terms:

$$\tilde{W}(\boldsymbol{\xi}) = \tilde{F}(\boldsymbol{\xi})[D(\xi^2+\eta^2)^2 + K + iC\omega - \rho h\omega^2]^{-1} \quad (5)$$

The dynamic response $W(\mathbf{x})$ in the original time-space domain can be constructed as the inverse Fourier transform of (5):

$$W(\mathbf{x}) = (2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(i\boldsymbol{\xi}x)\tilde{F}(\boldsymbol{\xi})[D(\xi^2+\eta^2)^2 + K + iC\omega - \rho h\omega^2]^{-1}d\boldsymbol{\xi} \quad (6)$$

Equation (6) is the dynamic displacement response of a thin slab resting on a viscoelastic foundation subject to a general dynamic load. Several special types of dynamic loads will be studied in detail in the following section.

3 Response to a Special Type of Dynamic Load

3.1 Response to a Static Load. For a static concentrated load, it can be expressed as

$$F_s(\mathbf{x}) = P\delta(x)\delta(y) \quad (7)$$

Here, $\delta(\cdot)$ is the Dirac-delta function defined by

$$\int_{-\infty}^{\infty}\delta(x-x_0)f(x)dx = f(x_0) \quad (8)$$

The Fourier transform of $F_s(\mathbf{x})$ is given by

$$\tilde{F}_s(\boldsymbol{\xi}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\delta(x)\delta(y)\exp(-i\boldsymbol{\xi}x)d\mathbf{x} = 2\pi P\delta(\omega) \quad (9)$$

Substituting (9) into (6) yields static displacement field under point load (7)

$$W_{\text{static}}(\mathbf{x}) = (2\pi)^{-2}\bar{P}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[i(\xi x+\eta y)]}{(\xi^2+\eta^2)^2+\bar{K}}d\xi d\eta \quad (10)$$

where $\bar{K}=K/D$. We define the coordinate transform $x=r\cos\theta$, $y=r\sin\theta$, $\xi=\zeta\cos\psi$ and $\eta=\zeta\sin\psi$. Since $\sin\theta\cos\psi+\cos\theta\sin\psi=\sin(\theta+\psi)$, applying this relation and the Euler formula $\exp(i\varsigma)=\cos\varsigma+i\sin\varsigma$ to (10) gives

$$W_{\text{static}}(\mathbf{x}) = (2\pi)^{-2}\bar{P}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\cos[r\zeta\sin(\theta+\psi)]}{\zeta^4+\bar{K}}\zeta d\zeta d\psi \quad (11)$$

It is known that the Bessel function can be expressed as $J_0(z)=(2\pi)^{-1}\int_0^{2\pi}\cos(z\cos\phi)d\phi$ (Watson [35]), in which $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. Adopting this representation, (15) can be rewritten as

$$W_{\text{static}}(\mathbf{x}) = (2\pi)^{-1}\bar{P}\int_{-\infty}^{\infty}\frac{J_0(r\zeta)}{\zeta^4+\bar{K}}\zeta d\zeta \tag{12}$$

This is exactly identical to the static solution given in Zhu et al. [36].

3.2 Response to a Quiescent Harmonic Load. A quiescent harmonic point load can be expressed as

$$F_{qh}(\mathbf{x}) = P\delta(x)\delta(y)e^{i\Omega t} \tag{13}$$

The Fourier transform of $F_{qh}(\mathbf{x})$ is given by

$$\tilde{F}_{qh}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\delta(x)\delta(y)e^{i\Omega t}\exp(-i\boldsymbol{\xi}x)d\mathbf{x} = 2\pi P\delta(\omega-\Omega) \tag{14}$$

Substituting (14) into (6) yields the dynamic displacement field under quiescent harmonic point load (13)

$$W(\mathbf{x}) = (2\pi)^{-2}P\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[i(\xi x+\eta y)]}{D(\xi^2+\eta^2)^2+K+iC\Omega-\rho h\Omega^2}d\xi d\eta \tag{15}$$

3.3 Response to a Moving Constant Load. A moving constant point load denoted by $F_{\text{mc}}(\mathbf{x})$ can be expressed as

$$F_{\text{mc}}(\mathbf{x}) = P\delta(x-vt)\delta(y) \tag{16}$$

The Fourier transform of the moving load $\tilde{F}_{\text{mc}}(\boldsymbol{\xi})$ can be obtained by taking a Fourier transform on both sides of (16):

$$\tilde{F}_{\text{mc}}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\delta(x-vt)\delta(y)\exp(-i\boldsymbol{\xi}x)d\mathbf{x}$$

$$= P\int_{-\infty}^{\infty}\exp[-i(\xi v+\omega)t]dt = 2\pi P\delta(\xi v+\omega) \tag{17}$$

Substituting (17) into (6) yields the dynamic displacement field under moving point load (16)

$$W_{\text{mc}}(\mathbf{x}) = (2\pi)^{-2}P\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[i(\xi x+\eta y)]}{D(\xi^2+\eta^2)^2+K-iCv\xi-\rho hv^2\xi^2}d\xi d\eta \tag{18}$$

3.4 Response to a Moving Harmonic Load. A moving harmonic point load denoted by $F_{\text{mh}}(\mathbf{x})$ can be expressed as

$$F_{\text{mh}}(\mathbf{x}) = P\delta(x-vt)\delta(y)e^{i\Omega t} \tag{19}$$

The Fourier transform $\tilde{F}_{\text{mh}}(\boldsymbol{\xi})$ of (19) is

$$\tilde{F}_{\text{mh}}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\delta(x-vt)\delta(y)e^{i\Omega t}\exp(-i\boldsymbol{\xi}x)d\mathbf{x}$$

$$= P\int_{-\infty}^{\infty}\exp[-i(\xi v+\omega-\Omega)t]dt = 2\pi P\delta(\xi v+\omega-\Omega) \tag{20}$$

Here, the property of the Dirac-delta function (9) is utilized for the evaluation of the above integrals. Substituting (10) into the general solution (6) gives the dynamic pavement displacement response $W_{\text{mh}}(\mathbf{x})$ corresponding to the moving harmonic point load (19)

$$W_{\text{mh}}(\mathbf{x}) = (2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(i\boldsymbol{\xi}x)2\pi P\delta(\xi v+\omega-\Omega)[D(\xi^2+\eta^2)^2+K+iC\omega-\rho h\omega^2]^{-1}d\boldsymbol{\xi}$$

$$= \frac{\bar{P}}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp\{i[\xi x+\eta y+(\Omega-\xi v)t]\}}{(\xi^2+\eta^2)^2+\bar{K}+i(\Omega-\xi v)\bar{C}-\bar{m}(\Omega-\xi v)^2}d\xi d\eta \tag{21}$$

Clearly, a constant moving load can be treated as a special case of a moving harmonic load by setting frequency $\Omega=0$, while a quiescent harmonic load can be treated as a special case of a moving harmonic load as load speed $v=0$. In other words, Eqs. (10), (15), and (18) are special cases of (21). Without loss of generality, in the following analysis, only (21) will be used for further studies.

Knowing the vertical velocity and vertical acceleration of slab is also paramount because these dynamic quantities can be conveniently measured using geophones (velocity transducers) and accelerometers. The first and second partial derivatives of displacement with respect to time are, respectively, the vertical velocity and acceleration responses of the slab

$$V_{\text{mh}}(\mathbf{x}) = \frac{\partial}{\partial t}W_{\text{mh}}(\mathbf{x})$$

$$= \frac{i\bar{P}}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{(\Omega-\xi v)\exp\{i[\xi x+\eta y+(\Omega-\xi v)t]\}}{(\xi^2+\eta^2)^2+\bar{K}+i(\Omega-\xi v)\bar{C}-\bar{m}(\Omega-\xi v)^2}d\xi d\eta \tag{22}$$

$$A_{\text{mh}}(\mathbf{x}) = \frac{\partial^2}{\partial t^2}W_{\text{mh}}(\mathbf{x})$$

$$= \frac{-\bar{P}}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{(\Omega-\xi v)^2\exp\{i[\xi x+\eta y+(\Omega-\xi v)t]\}}{(\xi^2+\eta^2)^2+\bar{K}+i(\Omega-\xi v)\bar{C}-\bar{m}(\Omega-\xi v)^2}d\xi d\eta \tag{23}$$

where $\bar{P}=P/D$, $\bar{K}=K/D$, $\bar{C}=C/D$, and $\bar{m}=\rho h/D$. It is noted that in the derivation of (22) and (23), we used the exchangeability of the partial differential operators and the double integration.

4 Critical Speed and Resonance Frequency

The dynamic displacement response of a slab to a moving harmonic load is governed by the zeros of the denominator of the integrand in (21). In one-dimensional structures such as beams, it has been both experimentally observed and theoretically proven that there exists a critical speed and resonance frequency (Kenney [26] and Sun [9,10,30]) at which the amplitude of displacement response of the structure becomes infinity. In two-dimensional structures such as slabs, critical speed and resonance frequency

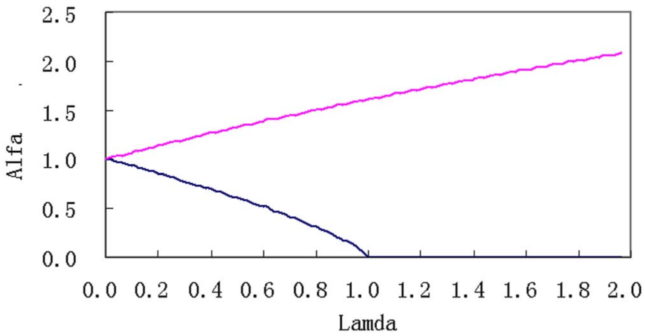


Fig. 2 Critical speed versus moving load frequency

also exist, under which resonance occurs. Chen and Huang [15,37] used the condition of the determinant of dynamic stiffness matrix being zero to specify critical speed and resonance frequency of the Timoshenko and Bernoulli-Euler beams. Dieterman and Metrikine [38] took advantage of kinematic invariant to identify critical speed and resonance frequency of an elastic layer. In this section, complex analysis will be used to specify critical speed and resonance frequency for a Kirchhoff slab under a moving harmonic load.

When damping presents in a dynamic system, it dissipates vibration energy. As a result, resonance cannot occur in dynamic systems with damping or other energy absorbing mechanisms. To study resonance, we need to investigate zeros of the denominator of the integrant in (21) without including the damping term, which corresponds to roots of the following equation:

$$(\xi^2+\eta^2)^2+\bar{K}-\bar{m}(\Omega-\upsilon\xi)^2=0 \tag{24}$$

Both ξ and η are variables for integration in (21). Since Eq. (21) involves double integration, the integration with respect to η is evaluated first. To this end, we represent η in terms of ξ by solving (24)

$$\eta_j(\xi)=(\pm)\{-\xi^2(\pm)[\bar{m}(\Omega-\upsilon\xi)^2-\bar{K}]^{1/2}\}^{1/2},\quad j=1,\ldots,4 \tag{25}$$

It should be noted that ξ is real valued in (25).

When applying the complex analysis to evaluate the integral with respect to η, the four roots of (25) become four poles in the complex η plane. We construct a closed contour to surround these poles. For $\mathrm{Im}(\eta)\geqslant 0$, we select the closed contour in the upper half η plane, while for $\mathrm{Im}(\eta)<0$, the lower half η plane. Due to the length limitation of the paper, only the case of $\mathrm{Im}(\eta)\geqslant 0$ is presented here and the case of $\mathrm{Im}(\eta)<0$ can be derived in analogy. Applying the theorem of residue, Eq. (21) can be further simplified to (ignoring the damping of the system) (Sun [11,12])

$$W_{\mathrm{mh}}(\mathbf{x})=(8\pi)^{-1}i\bar{P}\exp(i\Omega t)\times\sum_{\mathrm{Im}(\eta_j)>0}\int_{-\infty}^{\infty}\frac{\exp[i\xi(x-\upsilon t)]\exp[i\eta_j(\xi)y]}{\eta_j(\xi)[\eta_j^2(\xi)+\xi^2]}d\xi \tag{26}$$

According to (25), $\eta_j^2(\xi)+\xi^2=0$ is equivalent to $\bar{K}-\bar{m}(\Omega-\upsilon\xi)^2=0$ or $\{\xi-[\Omega+(\bar{K}/\bar{m})^{1/2}]/\upsilon\}\{\xi-[\Omega-(\bar{K}/\bar{m})^{1/2}]/\upsilon\}=0$. This corresponds to two first-order poles in a complex ξ plane. We can apply the theorem of residue again to evaluate integral (26), which requires the poles specified by $\eta_j(\xi)=0$ to be known. Note that ξ has to be real valued; we only need to identify two real zeros of $\eta_j(\xi)=0$. Substituting $\eta_j(\xi)=0$ into (25), it follows

$$g(\xi)=\xi^4-\bar{m}\upsilon^2\xi^2+2\bar{m}\upsilon\Omega\xi-\bar{m}\Omega^2+\bar{K}=0 \tag{27}$$

In this paper, (27) is called the characteristic equation of a Kirchhoff slab and $g(\xi)$ is called the characteristic function.

The purpose of this section is not about the analytical evaluation of (26), which can be found in Sun [11,12], but about the investigation of critical speed and resonance frequency, which correspond to high-order poles or zeros of characteristic function $g(\xi)$. As complex roots of (27) must appear as a conjugate pair, the number of real roots of (27) can only be two or four. A duplicated real root or a second-order pole requires the first derivative of $g(\xi)$ to be zero, whereas a fourth-order pole requires up to the third derivative of $g(\xi)$ to be zero. The first, second, and third derivatives of characteristic function $g(\xi)$ are listed below

$$g'(\xi)=4\xi^3-2\bar{m}\upsilon^2\xi+2\bar{m}\upsilon\Omega=0 \tag{28}$$

$$g''(\xi)=12\xi^2-2\bar{m}\upsilon^2=0 \tag{29}$$

$$g'''(\xi)=24\xi=0 \tag{30}$$

Clearly, simultaneous satisfaction of Eqs. (27)–(30) ensures the existence of a fourth-order pole. From (30) it follows that the fourth-order pole is $\xi=0$, which can only occur when $\upsilon=0$ (from (28) and (29)) and $\Omega=(\bar{K}/\bar{m})^{1/2}$ (from (27)). Define natural frequency as $\Omega_0=(\bar{K}/\bar{m})^{1/2}$. The above derivation suggests that slab resonance occurs when the load is a position-fixed harmonic load with frequency Ω_0.

Slab resonance may also occur when $\upsilon\neq 0$ (a moving harmonic load), which corresponds to a second-order pole and requires simultaneous satisfaction of (27) and (28). Mathematical manipulation $(27)\times 4-(28)\times\xi$ leads to

$$-2\bar{m}\upsilon^2\xi^2+6\bar{m}\upsilon\Omega\xi+4(\bar{K}-\bar{m}\Omega^2)=0 \tag{31}$$

The two roots of the above equation are

$$\xi=\frac{3\Omega}{2\upsilon}\pm\frac{(8\bar{m}\bar{K}+\bar{m}^2\Omega^2)^{1/2}}{2\bar{m}\upsilon} \tag{32}$$

Mathematical manipulation $(31)\times(3\Omega+\upsilon\xi)+(28)\times\bar{m}\upsilon^3/2$ leads to

$$(\bar{m}^2\upsilon^4-4\bar{K}-14\bar{m}\Omega^2)\upsilon\xi-[\bar{m}^2\upsilon^4-12(\bar{K}-\bar{m}\Omega^2)]\Omega=0 \tag{33}$$

The real root of (33) is

$$\xi=\frac{\Omega(\bar{m}^2\upsilon^4+12\bar{K}-12\bar{m}\Omega^2)}{\upsilon(\bar{m}^2\upsilon^4-4\bar{K}-14\bar{m}\Omega^2)} \tag{34}$$

The second-order poles given by (32) and (34) should be identical as they all satisfy Eqs. (27) and (28).

It is noted that $\upsilon\neq 0$ in (32) and (34). Equalizing these two equations leads to the condition at which slab resonance occurs:

$$\frac{3\Omega}{2\upsilon}\pm\frac{(8\bar{m}\bar{K}+\bar{m}^2\Omega^2)^{1/2}}{2\bar{m}\upsilon}=\frac{\Omega(\bar{m}^2\upsilon^4+12\bar{K}-12\bar{m}\Omega^2)}{\upsilon(\bar{m}^2\upsilon^4-4\bar{K}-14\bar{m}\Omega^2)} \tag{35}$$

Let $\beta=[1+8\bar{K}/(\bar{m}\Omega^2)]^{1/2}$. The above equation yields the relationship between frequency-dependent critical speed υ_Ω and moving load frequency

$$\upsilon_\Omega^4=\frac{(14\beta-18)\bar{m}\Omega^2+(4\beta-36)\bar{K}}{(\beta-1)\bar{m}^2} \tag{36}$$

Substituting $\Omega=0$ into (36) gives the critical speed $\upsilon_0^4=4\bar{K}/\bar{m}^2$ for a moving constant load, which is consistent with the result of Sun

Table 1 Default values for parametric study

Parameter	Default value	Range of values
P	2.5×10^3 N	2.5×10^3 N
E	2.8×10^{10} N/m^2	$(2.8,3.2)\times 10^{10}$ N/m^2
K	2.3×10^8 N/m^3	$(1,10)\times 10^8$ N/m^3
ρ	2.3×10^3 kg/m^3	$(1.75,22.4)\times 10^3$ kg/m^3
h	0.25 m	(0.1,0.4) m
μ	0.25	0.25
C	1×10^7 N s/m^3	$(0,10)\times 10^7$ N s/m^3
v	30 m/s	(0,50) m/s
Ω	0 Hz	(0,100) Hz

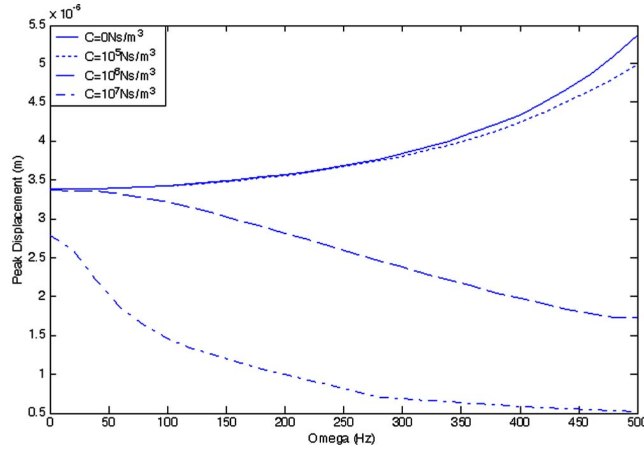


Fig. 4 Peak displacement of a slab to a moving load with different load frequencies (v=30 m/s)

[11]. Equation (36) can also be expressed in terms of v_0^4 and natural frequency $\Omega_0=(\bar{K}/\bar{m})^{1/2}$. To this end, we define the ratio $\lambda=\Omega/\Omega_0$ and therefore $\beta=(1+8/\lambda^2)^{1/2}$. From (36) we have

$$v_\Omega = \left[\frac{(14\beta-18)\lambda^2+(4\beta-36)}{4(\beta-1)}\right]^{\frac{1}{4}} v_0 = \alpha_\lambda v_0 \qquad (37)$$

where α_λ is a dimensionless coefficient representing the factor in front of critical speed v_0. Figure 2 shows a bifurcation relationship between the ratio λ and the coefficient α_λ.

According to (37), it is clear that there is only one critical speed for certain frequency range, while for some other frequency range there might exist two critical speeds. These frequency ranges depend on parameters $\bar{K}$ and $\bar{m}$. The existence of two possible critical speeds is consistent with the result obtained by Chen and Huang [37] for a Timoshenko beam and a Bernoulli-Euler beam and by Shamalta and Metrikine [18] for an elastic layer. Indeed, the relationship (36) describing the critical speed and resonance frequency is also valid for a Bernoulli-Euler beam whose parameters are specified as $\bar{K}=K/(EI)$ and $\bar{m}=m/(EI)$ with I being the moment of inertia. The critical speed for a moving constant load is also $v_0=\sqrt[4]{4\bar{K}/\bar{m}^2}$ for a Bernoulli-Euler beam (Sun [9]).

5 Dynamic Coefficient

It is of great interest to investigate the maximum displacement response of a slab subject to a moving constant load. Let $\max(W_{\text{static}}(\mathbf{x}))$ be the maximum displacement of a slab when the moving load is a static load. It can be obtained from (10) by setting $x=y=0$

$$\max[W_{\text{static}}(\mathbf{x})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\bar{P}}{(2\pi)^2[(\xi^2+\eta^2)^2+\bar{K}]} d\xi d\eta \qquad (38)$$

Let $\max(W_{\text{mc}}(\mathbf{x}))$ be the maximum displacement of a slab on a viscoelastic foundation subject to a moving constant load. Define the dynamic coefficient as the ratio

$$\text{dynamic coefficient} = \max[W_{\text{mc}}(\mathbf{x})]/\max[W_{\text{static}}(\mathbf{x})] \qquad (39)$$

Although the complexity of (39) does not permit analytical treatment of the dynamic coefficient, which has to be done numerically, and will be evaluated in the next section, it is, however, possible to approximate the dynamic coefficient based on fundamental physical principles. Here, we confine our attention to the subsonic case where the speed of moving load is lower than the critical speed v_0 of a moving constant load (Sun [11,12]) as vehicle speed on highways typically falls into such a subsonic speed range.

Because of the damping effect, the maximum displacement $\max(W_{\text{mc}}(\mathbf{x}))$ of a slab on a viscoelastic foundation should occur behind the moving load. The maximum displacement, by definition, should be greater than the dynamic displacement of a slab at ($x=vt$, $y=0$), the point beneath the moving load:

$$\max[W_{\text{mc}}(\mathbf{x})] \geq W_{\text{mc}}(\mathbf{x}|_{x=vt,y=0}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\bar{P}}{(2\pi)^2[(\xi^2+\eta^2)^2+\bar{K}-i\xi v\bar{C}-\bar{m}v^2\xi^2]} d\xi d\eta \qquad (40)$$

The equality is satisfied when the damping coefficient $C=0$.

Let $\max(W_m(\mathbf{x}))$ be the maximum displacement of a slab on an elastic Winkler foundation (i.e., $C=0$) subject to the same moving

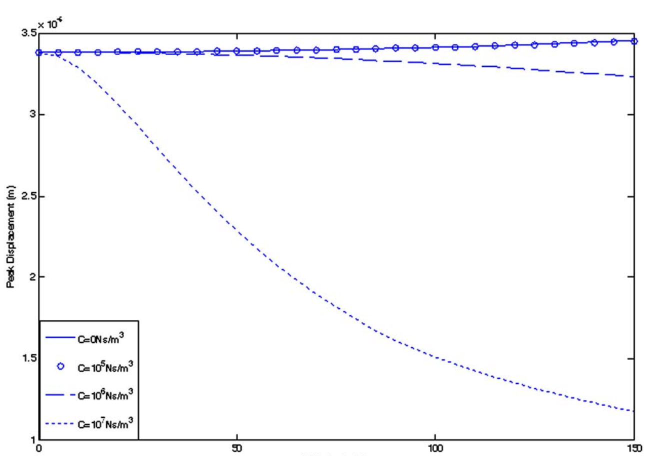


Fig. 3 Peak displacement of a slab to a moving constant load at different speeds (Ω=0 Hz)

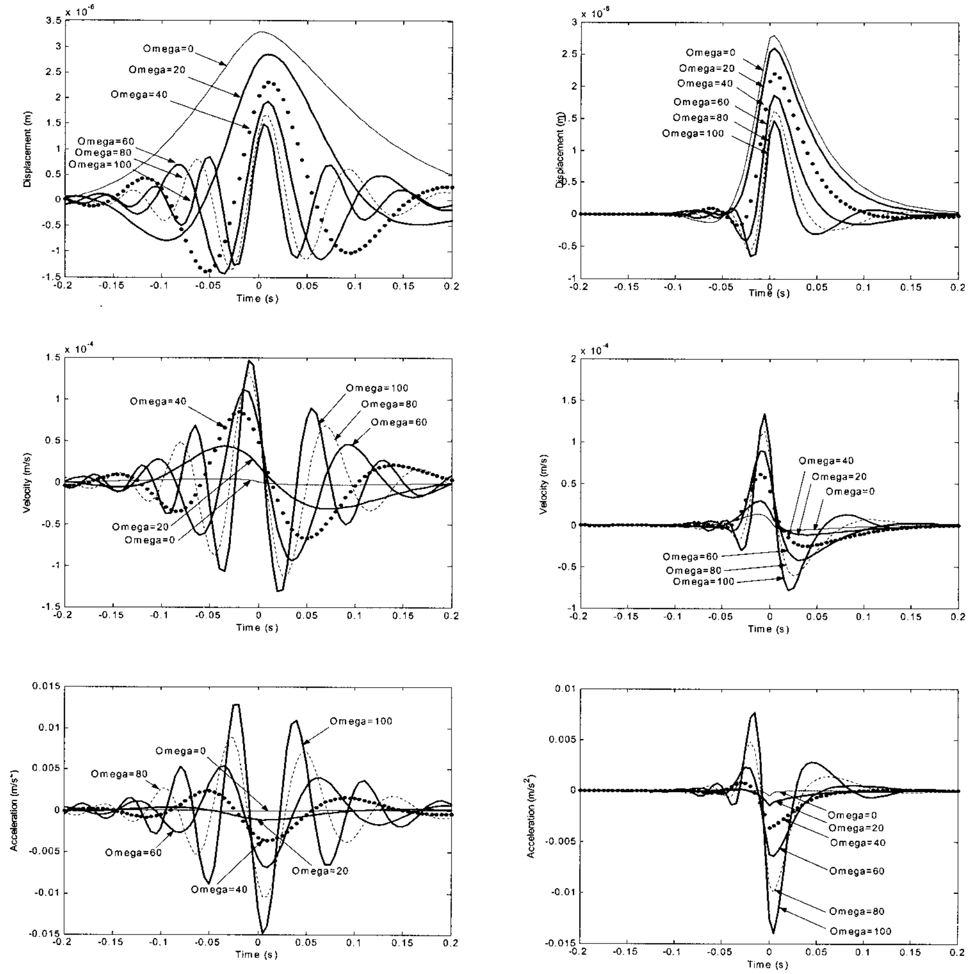


Fig. 5 Dynamic responses of a slab to a moving load with large damping $C=10^7$ N s/m^3 and different load frequencies (left column: $v=10$ m/s; right column: $v=30$ m/s)

constant load. The necessary condition for a maximum displacement to appear for this case is when the first-order partial derivative of the displacement with respect to time t becomes zero, i.e., $\partial W_{\mathrm{mc}}(\mathbf{x})/\partial t|_{C=0}=0$, or equivalently,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\xi\exp\{i[\xi(x-vt)+\eta y]\}}{D(\xi^2+\eta^2)^2+K-\rho h v^2\xi^2}d\xi d\eta=0$$

When $x-vt=0$ and $y=0$, the numerator and the denominator of the integrand are an odd function and an even function of ξ, respectively. A nontrivial solution of this integral equation is $x=vt$ and $y=0$. This suggests that the maximum displacement, given by (41), appears beneath the moving load and travels with the load at the same speed. In other words,

$$\max[W_m(\mathbf{x})]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\bar{P}}{(2\pi)^2[(\xi^2+\eta^2)^2+\bar{K}-\bar{m}v^2\xi^2]}d\xi d\eta \tag{41}$$

Under subsonic conditions, one could argue that because the damping effect will absorb some energy, reducing the magnitude of slab vibration, the maximum displacement of a slab on viscoelastic foundation ($C>0$) will be less than that of the same slab on an elastic Winkler foundation ($C=0$). This leads to $\max(W_{\mathrm{mc}}(\mathbf{x}))\leqslant\max(W_m(\mathbf{x}))$. The equality sign is satisfied when the damping coefficient $C=0$. Combining this inequality and (40), the dynamic coefficient defined by (39) can be limited to the following range:

$$\frac{W_{\mathrm{mc}}(\mathbf{x}|_{x=vt,y=0})}{\max[W_{\mathrm{static}}(\mathbf{x})]}\leqslant \text{dynamic coefficient}=\frac{\max[W_{\mathrm{mc}}(\mathbf{x})]}{\max[W_{\mathrm{static}}(\mathbf{x})]}\leqslant\frac{\max[W_m(\mathbf{x})]}{\max[W_{\mathrm{static}}(\mathbf{x})]} \tag{42}$$

6 Numerical Computation

6.1 Formulation. In the degenerate situation where a slab degrades to a beam, analytical treatment for obtaining a closed-form solution can be carried out using complex analysis (Henrici [39] and Sun [9,10,30]). However, due to the complexity of the slab characteristic equation of the current problem, analytical treatment will be very involved if not impossible. To better understand the displacement response of a rigid pavement to a moving load, it is of help to observe the dynamic response numerically. In what follows, numerical methods will be formulated for computation.

Let $\xi'=2\pi\xi$ and $\eta'=2\pi\eta$, and define the Galilean transform

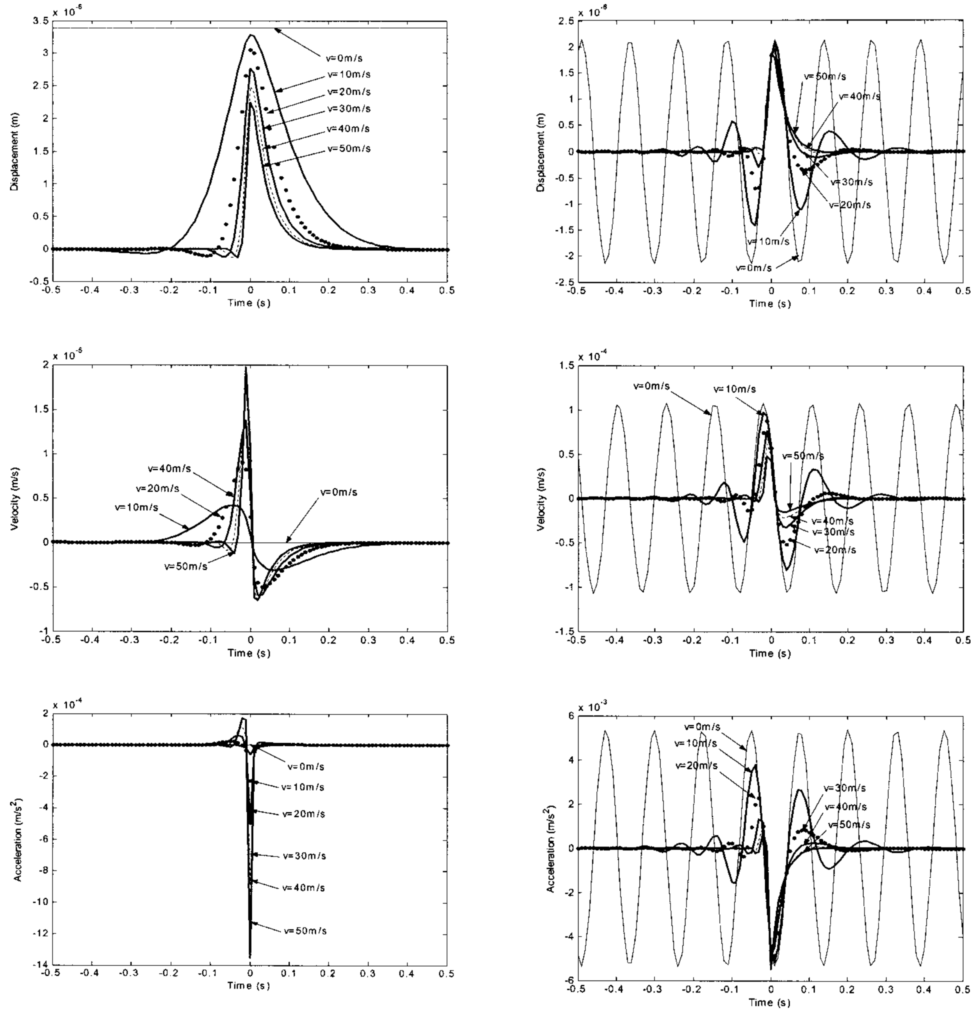


Fig. 6 Dynamic responses of a slab to a moving load at different speeds with large damping $C=10^7$ N s/m^3 (left column: $\Omega=0$ Hz; right column: $\Omega=50$ Hz)

$x'=x-vt$, $y=y'$, and $t=t'$. This transformation creates a moving coordinate system $\mathbf{x}'=(x',y',t')$ whose origin travels with the moving load at the same speed. The displacement response (13) can be rewritten as

$$W_{\mathrm{mh}}(\mathbf{x}')=\bar{P}e^{i\Omega t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[i2\pi(\xi'x'+\eta'y')]}{(2\pi)^4(\xi'^2+\eta'^2)^2+\bar{K}+i(\Omega-2\pi\xi'v)\bar{C}-\bar{m}(\Omega-2\pi\xi'v)^2}d\xi'd\eta' \tag{43}$$

Let the integrand be $F(\xi',\eta')$. This function can then be decomposed into the real and imaginary parts as

$$F(\xi',\eta')=\alpha(\xi',\eta')+i\beta(\xi',\eta') \tag{44}$$

$$\alpha(\xi',\eta')=\frac{(2\pi)^4(\xi'^2+\eta'^2)^2+\bar{K}-\bar{m}(2\pi\xi'v-\Omega)^2}{[(2\pi)^4(\xi'^2+\eta'^2)^2+\bar{K}-\bar{m}(\Omega-2\pi\xi'v)^2]^2+[(\Omega-2\pi\xi'v)\bar{C}]^2} \tag{45a}$$

$$\beta(\xi',\eta')=\frac{-(2\pi\xi'v-\Omega)\bar{C}}{[(2\pi)^4(\xi'^2+\eta'^2)^2+\bar{K}-\bar{m}(\Omega-2\pi\xi'v)^2]^2+[(\Omega-2\pi\xi'v)\bar{C}]^2} \tag{45b}$$

Given that $\bar{K}>0$ and $\bar{C}>0$, the denominator of the integrand is nonzero since $(2\pi)^4(\xi'^2+\eta'^2)^2+\bar{K}-\bar{m}(\Omega-2\pi\xi'v)^2$ and $(\Omega-2\pi\xi'v)\bar{C}$ cannot be zero simultaneously. Equation (43) can now be written as

$$W_{\mathrm{mh}}(\mathbf{x}')=\bar{P}e^{i\Omega t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\xi',\eta')\exp[i2\pi(\xi'x'+\eta'y')]d\xi'd\eta' \tag{46}$$

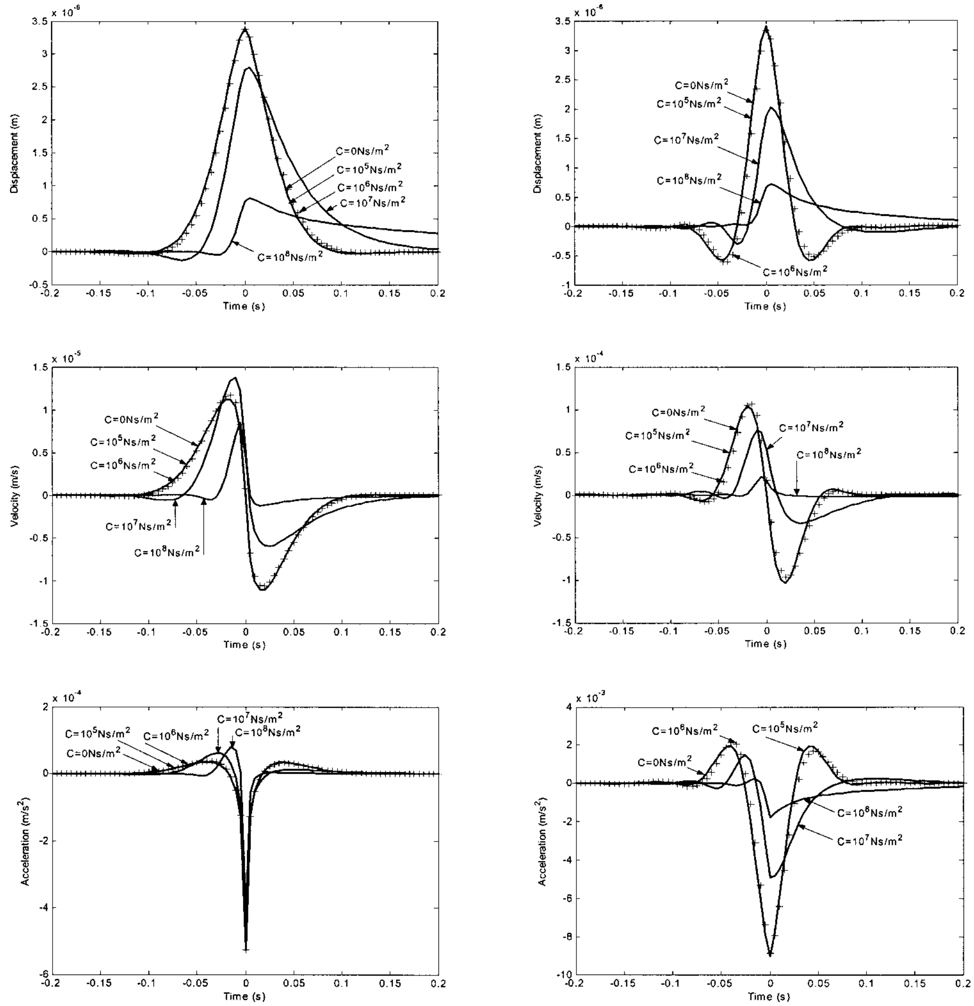


Fig. 7 Dynamic responses of a slab to a moving load at speed v=30 m/s with different damping coefficients (left column: Ω=0 Hz; right column: Ω=50 Hz)

Discretization of (34) by letting $d\xi' = \Delta\xi'$ and $d\eta' = \Delta\eta'$ leads to

$$
\begin{aligned}
W_{\mathrm{mh}}(x_k', y_l') &= \bar{P}e^{i\Omega t}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(\xi_m', \eta_n')\exp[i2\pi(\xi_m' x_k' + \eta_n' y_l')]\Delta\xi'\Delta\eta' \\
&= \bar{P}e^{i\Omega t}\Delta\xi'\Delta\eta'\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(m\Delta\xi', n\Delta\eta') \\
&\quad\times\exp[i2\pi(mk\Delta\xi'\Delta x' + nl\Delta\eta'\Delta y')]
\end{aligned}
\tag{47}
$$

where $x_k' = k\Delta x'$, $y_l' = l\Delta y'$, $k=0,\ldots,M$ and $l=0,\ldots,N$. The summation (47) is exactly the definition of an inverse fast Fourier transform (IFFT), and therefore, can be efficiently evaluated. To do so, follow the definition of a discrete fast Fourier transform pair:

$$
F(m,n) = \mathrm{FFT}\{f(k,l)\} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(k,l)e^{-i2\pi nk/M}e^{-i2\pi kn/N} \tag{48}
$$

$$
f(k,l) = \mathrm{IFFT}\{F(m,n)\} = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(m,n)e^{\frac{i2\pi mk}{M}}e^{\frac{i2\pi nl}{N}} \tag{49}
$$

Equation (47) can now be written as

$$
W_{\mathrm{mh}}(x_k', y_l') = \bar{P}e^{i\Omega t}\,\mathrm{IFFT}[F(\xi_m', \eta_n')]/\Delta x'\Delta y' \tag{50}
$$

where $\xi_m' = m\Delta\xi'$, $\eta_n' = n\Delta\eta'$, $\Delta\xi'\Delta x' = 1/M$, $\Delta\eta'\Delta y' = 1/N$, and IFFT [] is the inverse FFT. The benefit of (50) is that it permits more efficient computation via the use of FFT.

It should be noted that the resulting dynamic responses are indeed presented in the Galilean coordinate system, a moving coordinate system. If one is interested in knowing the dynamic response in a fixed coordinate system, all responses in a Galilean coordinate system must be converted to the fixed coordinate system. It should also be noted that Eq. (23) is not applicable for computation of the slab acceleration under the load at $x=vt$ and $y=0$, since the integral in (23) is divergent. The order of differentiation and integration may be changed only if the integral is uniformly convergent, which is not the case for the integral in Eq. (22). Indeed, the computation of acceleration of plate under the load at $x=vt$ and $y=0$ is not directly based on (23). Rather, the

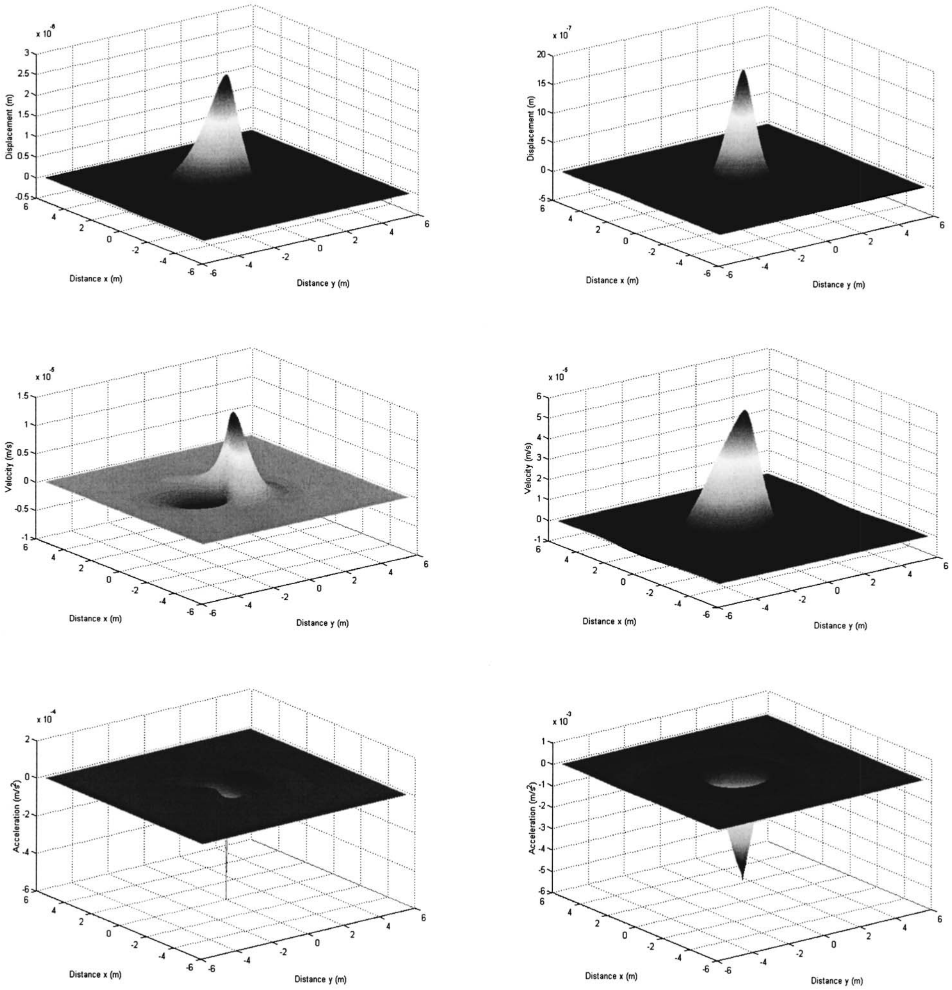


Fig. 8 3-D dynamic responses of a slab to a moving load at speed $v=30$ m/s with large damping $C=10^7$ N s/m^3 (left column: $\Omega=0$ Hz; right column: $\Omega=50$ Hz)

acceleration is obtained by numerically computing the derivative of vertical velocity using the differencing method, while the vertical velocity can be computed numerically using (22) or as a derivative of dynamic displacement.

6.2 Parametric Studies. In this section, a parametric study is conducted to uncover the spatial and temporal evolution of a slab's is response to a moving load. Table 1 lists default values and ranges of parameters used in the numerical computation. These parameters are typical material and structural properties of highway and airport pavements.

Figure 3 shows peak displacement responses of a slab to a moving load with four different damping coefficients: $C=0$, 10^5, 10^6, and 10^7 N s/m^3, respectively. The observation point is set at the origin of the coordinate system $x=y=0$. It can be seen that damping has a significant effect on peak displacement. When damping coefficients are small (e.g., less than 10^5 N s/m^3), peak displacement increases as the speed increases. When damping coefficients are large (e.g., greater than 10^6 N s/m^3), peak displacement decreases as speed increases.

Figure 4 shows peak displacement of a slab to a moving harmonic load with four damping coefficients: $C=0$, 10^5, 10^6, and 10^7 N s/m^3, respectively. The observation point is set at the origin of the coordinate system $x=y=0$ and load speed $v=30$ m/s. For small damping coefficients (e.g., less than 10^5N s/m^3), peak displacement increases as frequencies increase. For large damping coefficients (e.g., greater than 10^6 N s/m^3), peak displacement decreases as the frequencies increase.

Figure 5 shows the displacement, velocity, and acceleration responses of a slab to a point load moving at speed $v=10$ m/s and $v=30$ m/s at six different load frequencies:$\Omega=0$, 20, 40, 60, 80, and 100 Hz, respectively. Other parameters are given in Table 1. The observation point is set at the origin of the coordinate system $x=y=0$. The shape of the displacement response to a moving constant load ($\Omega=0$ Hz) is significantly different from that of a moving harmonic load ($\Omega\neq 0$ Hz). Only one peak is observed in the former, while there are several peaks in the latter, reflecting oscillation patterns of responses. For a moving constant load, the displacement response increases as the load approaches the observation point, reaches its maximum at the observation point, and decays from the maximum to zero as the load leaves the observation point. The shapes of dynamic displacement response are asymmetric with respect to time $t=0$ due to the damping effect.

In Fig. 5, slab responses to moving loads at speed $v=10$ m/s and at $v=30$ m/s are mainly during $(-0.3,0.3)$ s and $(-0.1,0.1)$ s, respectively. These are equivalent to a longitudinal $(-3,3)$ m vicinity of the observation point along the traveling direction. The maximum displacement responses of a slab decrease as load frequencies increase. However, the maximum velocity and acceleration responses show significantly different patterns. The higher the load frequency, the larger the peak velocity and acceleration will be. The dynamic response of a slab under a moving load at speed $v=30$ m/s shows less of an oscillation pattern than its counterpart at speed $v=10$ m/s. It should be noted that these results are obtained for large damping $C=10^7$ N s/m^3.

Figure 6 shows the displacement, velocity, and acceleration responses of a slab to a moving load with frequencies $\Omega=0$ Hz and $\Omega=50$ Hz at six different load speeds: $v=0$, 10, 20, 30, 40, and 50 m/s, respectively. The damping coefficient is also large at $C=10^7$ N s/m^3. The observation point is set at the origin of the coordinate system; i.e., $x=y=0$. For a static load ($\Omega=0$), both velocity and acceleration responses are zero, whereas the displacement response remains a constant about 3.4×10^{-6} m, which is greater than the dynamic displacement when the load is moving. As the load speed increases, the maximum displacement decreases while the maximum velocity and acceleration increase. Furthermore, as the load speed increases, the time duration of response gets shortened. For instance, when the load speed is $v=10$ m/s, the time duration of response is approximate ($-0.4, 0.4$) s, but approximately ($-0.1, 0.2$) s when the load speed is $v=30$ m/s.

Because of the damping effect, it is observed from the left-top diagram in Fig. 6 that for a moving load, it takes the dynamic displacement less time (about 0.02 s) to reach its maximum from zero than to vanish (about 0.3 s) from the maximum. The maximum displacement appears after the load passes through the observation point ($t=0$). When reflected in space, this time delay suggests that the maximum displacement occurs behind the moving load. The maximum acceleration seems to appear at $t=0$ in time and $x=0$ in space, indicating that acceleration seems to be less influenced by the damping effect. In the vicinity of the moving load, a positive response is observed, while outside this vicinity, a negative response exists. A similar conclusion can be drawn for a moving harmonic load with $\Omega=50$ Hz.

Figure 7 shows the dynamic responses of a slab with different damping coefficients at two load frequencies: $\Omega=0$ Hz and $\Omega=50$ Hz. The load speed is $v=30$ m/s and the observation point is at the origin of the coordinate system, i.e., $x=y=0$. Five levels of the damping coefficient are studied: $C=0$, 10^5, 10^6, 10^7, and 10^8 N s/m^3. Because of the damping effect, the shape of the displacement response is asymmetric, as compared to slabs with elastic foundation (no damping). It can be seen from Fig. 7 that both positive and negative displacements are observed in the vicinity of time $t=0$, indicating that the slab experiences both compressive and tensile stresses as the load passes over. For a constant moving load, the maximum displacement response of a slab appears after time $t=0$ due to the damping effect. The time lag between the maximum displacement and time $t=0$ increases as the damping coefficient increases. The amplitude of response decreases as the damping coefficient increases, which applies to both moving constant loads and moving harmonic loads. As the damping coefficient is small $C\leqslant10^5$ N s/m^2, there is almost no distinguishable difference between the response of a slab with and without damping.

It is of interest to investigate the dynamic response when viewed from a moving coordinate system specified by $x'=x-vt$ and $y'=y$. This moving coordinate system travels at the same speed as the moving load and the origin of the moving coordinate system passes through the origin of the fixed coordinate system at time $t=0$. Figure 8 shows three-dimensional displacement, velocity, and acceleration responses of a slab in the moving coordinate system to a single moving point load at load speed $v=30$ m/s and damping coefficient $C=10^7$ N s/m^3 with two load frequencies $\Omega=0$ Hz and $\Omega=50$ Hz, respectively. Other parameters used in the creation of Fig. 8 are listed in Table 1.

To study the effect of various parameters on the dynamic coefficient of a slab, sensitivity analysis is conducted by varying one

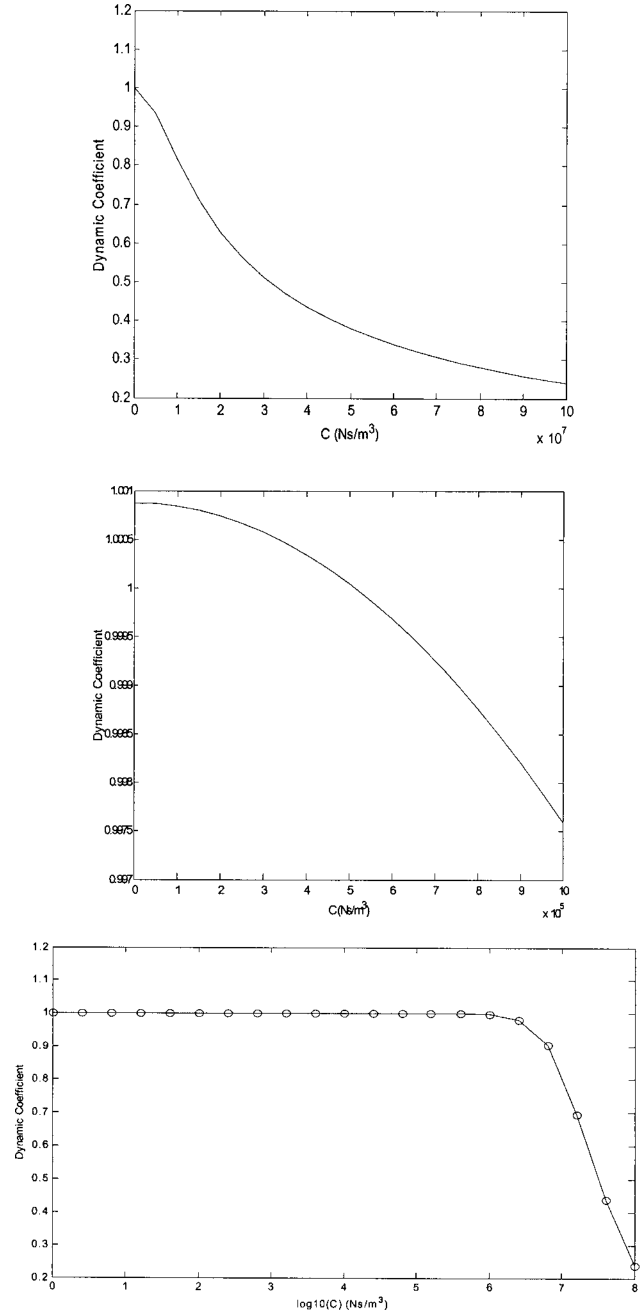


Fig. 9 Effect of damping coefficient on dynamic coefficient

parameter at a time while keeping other parameters in Table 1 unchanged. Figure 9 shows the effect of damping on the dynamic coefficient. A general trend is that when the damping coefficient increases, the dynamic coefficient decreases. However, when the damping is very small, the dynamic coefficient exceeds unity slightly since the peak displacement response increases as the speed increases.

Because the damping coefficient affects the dynamic coefficient considerably, two default values of the damping coefficient are used. One is small damping $C=1\times10^5$ N s/m^3 and the other is large damping $C=1\times10^7$ N s/m^3. Figures 10 and 11 summarize the effects of parameters on the dynamic coefficient obtained from the numerical computation. Other parameters are listed in Table 1. At a small damping level, from Fig. 10 it can be seen that the dynamic coefficient decreases as Young's modulus E, slab thickness h, and modulus of subgrade reaction K increase. The dy-

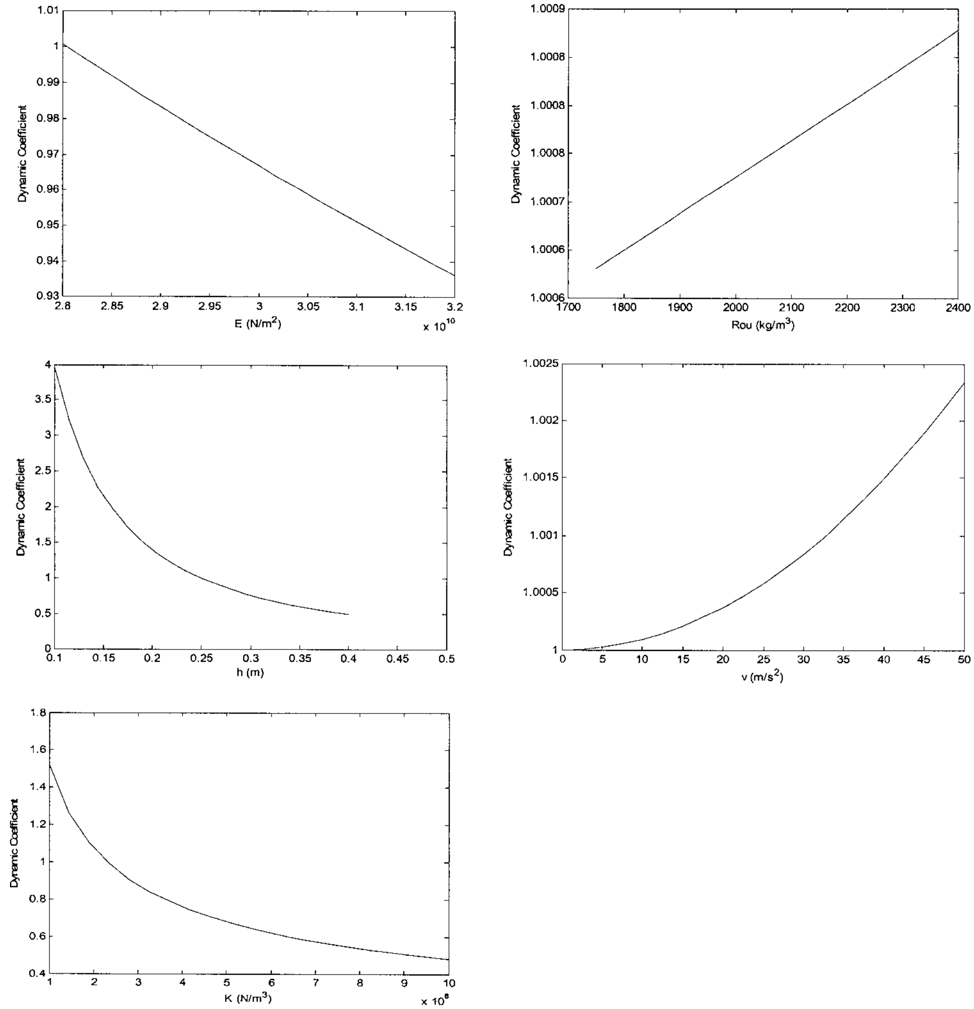


Fig. 10 Dynamic coefficient as a function of various parameters ($C=1\times10^5$ N s/m^3)

namic coefficient increases as the density of the slab ρ and load speed v increase. At a large damping level, from Fig. 11 it can be seen that the dynamic coefficient decreases as Young's modulus E, slab thickness h, modulus of subgrade reaction K, and load speed v increase. The dynamic coefficient increases as the density of the slab ρ increases. A noticeable feature is that at these two damping levels, the dynamic coefficient seems to be linearly related to Young's modulus E and density of the slab ρ, and nonlinearly related to other parameters. For a thin slab, the dynamic coefficient may exceed unity significantly.

7 Concluding Remarks

In this paper, integral transform and numerical analysis are used to analyze the dynamic response of a slab caused by a moving load with constant speed. The following concluding remarks can be made from the analysis. The result of this paper can be used for vehicle weigh-in-motion and highway infrastructure health monitoring using continuous measurement of pavement deflection.

- Critical speed and resonance frequency are related to each other via Eq. (37). There exists a bifurcation in critical speed. One branch of critical speed increases as load frequency increases, while the other branch of critical speed decreases as load frequency increases. There are two critical speeds when the load frequency is low, but only one critical speed exists when the load frequency is high.
- A slab's response to a moving load is only appreciable within a small vicinity of the moving load. For the parameters used in this study, it is approximately within the $(-3,3)$ m neighborhood of the observation point along the traveling direction. As the load speed increases, the time duration of the nonzero slab response decreases.
- When damping is small, peak displacement increases as load speed or load frequency increases. However, when damping is large, peak displacement decreases as load speed or load frequency increases.
- Dynamic displacement of a slab with damping subject to a moving load is asymmetric with a longer right tail, as shown in the left-top plot of Fig. 5. The response of a slab under a moving load at speed $v=30$ m/s shows less oscillations than that of a slab under a moving load at speed $v=10$ m/s.
- The damping coefficient has a significant affect on the shape of peak displacement of a slab. The peak displacement of a slab with small (large) damping coefficient increases (decreases) as the load frequency and the load speed increase.
- Because of the damping effect, the maximum dynamic displacement appears after the load passes through the observation point ($t=0$). When reflected in space, this time lag suggests that the maximum displacement occurs behind the moving load. The maximum acceleration seems to appear at $t=0$ in time and $x=0$ in space.
- When the damping coefficient is small, i.e., $C\leqslant10^5$ N s/m^2, there is almost no distinguishable difference

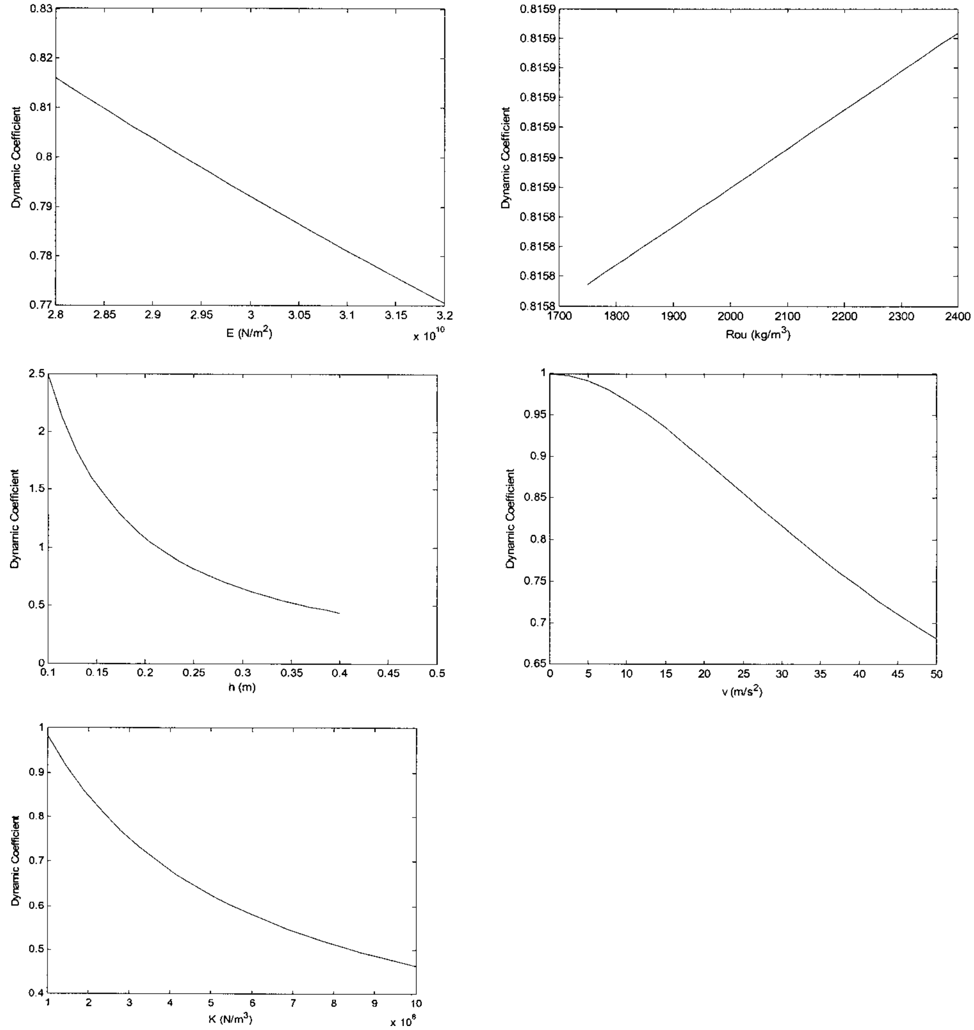


Fig. 11 Dynamic coefficient as a function of various parameters ($C = 1\times 10^7$ N s/m^3)

between the response of a slab with and without damping, though there is a very slight increase in the dynamic coefficient. When the damping coefficient is not very small, the dynamic coefficient decreases as the damping coefficient increases.

- At a small damping level, the dynamic coefficient decreases as Young's modulus E, slab thickness h, and modulus of subgrade reaction K increase, and increases as the density of the slab ρ and load speed v increase. At a large damping level, the dynamic coefficient decreases as Young's modulus E, slab thickness h, modulus of subgrade reaction K, and load speed v increases, and increases as density of the slab ρ increases. A linear relationship exists between the dynamic coefficient and Young's modulus E or density of the slab ρ.

Acknowledgment

This research is sponsored in part by the National Science Foundation (NSF) under Grant No. CMS-0408390 and an NSF Career award with Dr. P.N. Balaguru, Dr. Win Aung, and Dr. Douglas Foutch as program directors, by the American Chemical Society Petroleum Research Foundation under Grant No. PRF-44468-G9 with Dr. R.E. Siatkowski as program manager, and by the Changjiang Scholarship of Ministry of Education of China, to which the author is very grateful.

References

[1] Yoder, E. J., and Witczak, M. W., 1975, *Principles of Pavement Design*, John Wiley & Sons, New York.

[2] Haas, R., Hudson, W. R., and Zniewski, J., 1994, *Modern Pavement Management*, Krieger Publishing Company, Malabar, FL.

[3] Sun, L., and Deng, X., 1998, "Predicting Vertical Dynamic Loads Caused by Vehicle-Pavement Interaction," J. Transp. Eng., **126**, pp. 470–478.

[4] Sun, L., 2002, "Optimum Design of 'Road-Friendly' Vehicle Suspension Systems Subject to Rough Road Surface," Appl. Math. Model., **26**, pp. 635–652.

[5] Sun, L., Zhang, Z., and Ruth, J., 2001, "Modeling Indirect Statistics of Surface Roughness," J. Transp. Eng., **127**, pp. 105–111.

[6] Sun, L., Hudson, W. R., and Zhang, Z., 2003, "Empirical-Mechanistic Method Based Stochastic Modeling of Fatigue Damage to Predict Flexible Pavement Fatigue Cracking for Transportation Infrastructure Management," J. Transp. Eng., **129**, pp. 109–117.

[7] Sun, L., Kenis, W., and Wang, W., 2006, "Stochastic Spatial Excitation Induced by a Distributed Contact With Homogenous Gaussian Random Fields," J. Eng. Mech., **132**, pp. 714–722.

[8] Monismith, C. L., Sousa, J., and Lysmer, J., 1988, "Modern Pavement Design Technology Including Dynamic Load Conditions," SAE Technical Paper, No. 881856.

[9] Sun, L., 2001, "Closed-Form Representation of Beam Response to Moving Line Loads," ASME J. Appl. Mech., **68**, pp. 348–350.

[10] Sun, L., 2002, "A Closed-Form Solution of Beam on Viscoelastic Subgrade Subjected to Moving Loads," Comput. Struct., **80**, pp. 1–8.

[11] Sun, L., 2005, "Analytical Dynamic Displacement Response of Rigid Pavements to Moving Concentrated and Line Loads," Int. J. Solids Struct., **43**, pp. 4370–4383.

[12] Sun, L., 2005, "Dynamics of Plate Generated by Moving Harmonic Loads," ASME J. Appl. Mech., **72**, pp. 772–777.

[13] Kim, S. M., and Roesset, J. M., 1996, "Dynamic Response of Pavement Sys-

tems to Moving Loads," Center for Transportation Research, The University of Texas at Austin, Research Report No. 1422-2.
[14] Kononov, A. V., and Dieterman, H. A., 1999, "A Uniformly Moving Constant Load Along a Winkler Supported Strip," Eur. J. Mech. A/Solids, **18**(4), pp. 731–743.
[15] Chen, Y. H., and Huang, Y. H., 2000, "Dynamic Stiffness of Infinite Timoshenko Beam on Viscoelastic Foundation in Moving Coordinate," Int. J. Numer. Methods Eng., **48**, pp. 1–18.
[16] Chen, Y. H., Huang, Y. H., and Shih, C. T., 2001, "Response of an Infinite Timoshenko Beam on a Viscoelastic Foundation to a Harmonic Moving Load," J. Sound Vib., **241**(5), pp. 809–824.
[17] Clouteau, D., Degrande, G., and Lombaert, G., 2001, "Numerical Modelling of Traffic Induced Vibrations," Meccanica, **36**(4), pp. 401–420.
[18] Shamalta, M., and Metrikine, A. V., 2003, "Analytical Study of the Dynamic Response of an Embedded Railway Track to a Moving Load," Arch. Appl. Mech., **73**(1-2), pp. 131–146.
[19] Bush, A. J., 1980, "Nondestructive Testing for Light Aircraft Pavements, Phase II," Development of the nondestructive evaluation methodology, Final Report FAA No. RD-80-9, Federal Aviation Administration.
[20] Scullion, T., Uzan, J., and Paredes, M., 1990, "MODULUS: A Microcomputer-based Backcalculation System," Transp. Res. Rec., No. 1260.
[21] Uzan, J., and Lytton, R., 1990, "Analysis of Pressure Distribution Under Falling Weight Deflectometer Loading," J. Transp. Eng., **116**, pp. 246–250.
[22] Salawu, O. S., and Williams, C., 1995, "Full-Scale Force-Vibration Test Conducted Before and After Structural Repairs on Bridge," J. Struct. Eng., **121**(2), pp. 161–173.
[23] Sun, L., and Kennedy, T. W., 2002, "Spectral Analysis and Parametric Study of Stochastic Pavement Loads," J. Eng. Mech., **128**, pp. 318–327.
[24] Lee, J. L. Y., Stokoe, K. H. II, Murphy, M. R., and Rozycki, D. K., 2002, "Application of the Rolling Dynamic Deflectometer to Project-Level Pavement Studies," 81st Annual Meeting, Transportation Research Board, Washington, DC.
[25] Sun, L., and Greenberg, B., 2000, "Dynamic Response of Linear Systems to Moving Stochastic Sources," J. Sound Vib., **229**(4), pp. 957–972.
[26] Kenney, J. T., 1954, "Steady-State Vibrations of Beam on Elastic Foundation for Moving Load," ASME J. Appl. Mech., **20**, pp. 359–364.
[27] Taheri, M. R., 1986, "Dynamic Response of Slabs to Moving Loads," Ph.D. thesis, Purdue University, West Lafayette, IN.
[28] Kukreti, A. R., Taheri, M., and Ledesma, R. H., 1992, "Dynamic Analysis of Rigid Airport Pavements With Discontinuities," J. Transp. Eng., **118**(3), pp. 341–360.
[29] Zaghloul, S. M., White, T. D., Drnevich, V. P., and Coree, B., 1994, "Dynamic Analysis of FWD Loading and Pavement Response Using a Three-Dimensional Dynamic Finite Element Program," Transp. Res. Board, Washington, DC.
[30] Sun, L., 2003, "An Explicit Representation of Steady State Response of a Beam Resting on an Elastic Foundation to Moving Harmonic Line Loads," Int. J. Numer. Analyt. Meth. Geomech., **27**, pp. 69–84.
[31] Sun, L., 2001, "Computer Simulation and Field Measurement of Dynamic Pavement Loading," Math. Comput. Simul., **56**, pp. 297–313.
[32] Deshun, Z., 1999, "A Dynamic Model for Thick Elastic Slabs," J. Sound Vib., **221**(2), pp. 187–203.
[33] Bo, J., 1999, "The Vertical Vibration of an Elastic Circular Slab on a Fluid-Saturated Porous Half Space," Int. J. Eng. Sci., **37**, pp. 379–393.
[34] Sun, L., 2003, "Dynamic Response of Kirchhoff Slab on a Viscoelastic Foundation to Harmonic Circular Loads," ASME J. Appl. Mech., **70**, pp. 595–600.
[35] Watson, G. N., 1966, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, London.
[36] Zhu, Z., Wang, B., and Guo, D., 1984, *Pavement Mechanics*, Renming Jiaotong Pub. Inc., Beijing, China.
[37] Chen, Y. H., and Huang, Y. H., 2003, "Dynamic Characteristics of Infinite and Finite Railways to Moving Loads," J. Eng. Mech., **129**(9), pp. 987–995.
[38] Dieterman, H. A., and Metrikine, A., 1997, "Critical Velocities of a Harmonic Load Moving Uniformly Along an Elastic Layer," ASME J. Appl. Mech., **64**(3), pp. 596–600.
[39] Henrici, P., 1974, *Applied and Computational Complex Analysis*, Vol. 1, John Wiley & Sons, New York.

Extension of Stoney's Formula to Arbitrary Temperature Distributions in Thin Film/Substrate Systems

Y. Huang
Department of Mechanical and Industrial Engineering,
University of Illinois,
Urbana, IL 61801

A. J. Rosakis
Graduate Aeronautical Laboratory,
California Institute of Technology,
Pasadena, CA 91125

Current methodologies used for the inference of thin film stress through curvature measurements are strictly restricted to stress and curvature states that are assumed to remain uniform over the entire film/substrate system. By considering a circular thin film/substrate system subject to nonuniform and nonaxisymmetric temperature distributions, we derive relations between the film stresses and temperature, and between the plate system's curvatures and the temperature. These relations featured a "local" part that involves a direct dependence of the stress or curvature components on the temperature at the same point, and a "nonlocal" part that reflects the effect of temperature of other points on the location of scrutiny. Most notably, we also derive relations between the polar components of the film stress and those of system curvatures which allow for the experimental inference of such stresses from full-field curvature measurements in the presence of arbitrary nonuniformities. These relations also feature a "nonlocal" dependence on curvatures making full-field measurements of curvature a necessity for the correct inference of stress. Finally, it is shown that the interfacial shear tractions between the film and the substrate are related to the gradients of the first curvature invariant and can also be inferred experimentally. [DOI: 10.1115/1.2744035]

Keywords: nonuniform film temperatures and stresses, nonuniform substrate curvatures, stress-curvature relations, nonlocal effects, interfacial shears

1 Introduction

Substrates formed of suitable solid-state materials may be used as platforms to support various thin film structures. Integrated electronic circuits, integrated optical devices and optoelectronic circuits, microelectromechanical systems deposited on wafers, three-dimensional electronic circuits, systems-on-a-chip structures, lithographic reticles, and flat panel display systems are examples of such thin film structures integrated on various types of plate substrates.

The above-described thin film structures on substrates are often made from a multiplicity of fabrication and processing steps (e.g., sequential film deposition, thermal anneal, and etch steps) and often experience stresses caused by each of these steps. Examples of known phenomena and processes that build up stresses in thin films include, but are not limited to, lattice mismatch, chemical reaction, doping by, e.g., diffusion or implantation, rapid deposition by evaporation or sputtering, and of course thermal treatment (e.g., various thermal anneal steps). The film stress build-up associated with each of these steps often produces undesirable damage (e.g., cracking, interface delamination) that may be detrimental to the manufacturing process because of its cumulative effect on process "yield" [1]. Known problems associated with thermal excursions, in particular, include stress-induced film cracking and film/substrate delamination resulting during uncontrolled wafer cooling that follows the many anneal steps.

The intimate relation between stress-induced failures and process yield loss makes the identification of the origins of stress build-up, the accurate measurement and analysis of stresses, and the acquisition of information on the spatial distribution of stresses a crucial step in designing and controlling processing steps and in ultimately improving reliability and manufacturing yield.

Stress changes in thin films following discrete process steps or occurring during thermal excursions may be calculated in principle from changes in the film/substrate systems curvatures or "bow" based on analytical correlations between such quantities. Early attempts to provide such correlations are well documented [2]. Various formulations have been developed for this purpose and most of these are essentially extensions of Stoney's approximate plate analysis [3].

Stoney used a plate system composed of a stress bearing thin film of thickness h_f, deposited on a relatively thick substrate of thickness h_s, and derived a simple relation between the curvature (κ) of the system and the stress ($\sigma^{(f)}$) of the film as follows:

$$\sigma^{(f)} = \frac{E_s h_s^2 \kappa}{6h_f(1-\nu_s)} \tag{1.1}$$

In the above, the subscripts "f" and "s" denote the thin film and substrate, respectively, and E and ν are the Young's modulus and Poisson's ratio, respectively. Equation (1.1) is called the Stoney formula, and it has been extensively used in the literature to infer film stress changes from experimental measurement of system curvature changes [2].

Stoney's formula was derived for an isotropic "thin" solid film of uniform thickness deposited on a much "thicker" plate substrate based on a number of assumptions. Stoney's assumptions include the following: (1) Both the film thickness h_f and the substrate thickness h_s are uniform and $h_f \ll h_s \ll R$, where R represents the characteristic length in the lateral direction (e.g., system radius R shown in Fig. 1); (2) the strains and rotations of the plate system are infinitesimal; (3) both the film and substrate are homogeneous, isotropic, and linearly elastic; (4) the film stress states are in-plane isotropic or equibiaxial (two equal stress components in

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received June 23, 2005; final manuscript received February 9, 2006. Review conducted by Huajian Gao.

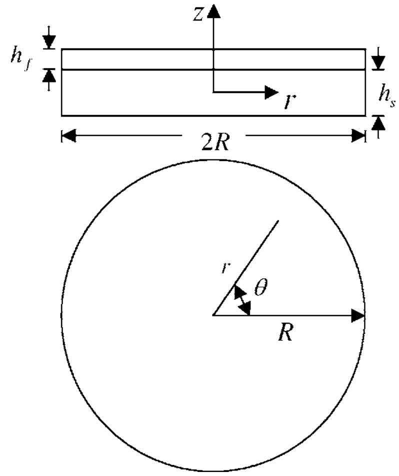


Fig. 1 A schematic diagram of the thin film/substrate system, showing the cylindrical coordinates (r, θ, z)

any two, mutually orthogonal in-plane directions) while the out-of-plane direct stress and all shear stresses vanish; (5) the system's curvature components are equibiaxial (two equal direct curvatures) while the twist curvature vanishes in all directions; and (6) all surviving stress and curvature components are spatially constant over the plate system's surface, a situation which is often violated in practice.

The assumption of equibiaxial ($\kappa_{xx}=\kappa_{yy}=\kappa,\ \kappa_{xy}=\kappa_{yx}=0$) and spatially constant curvature (κ independent of position) is equivalent to assuming that the plate system would deform spherically under the action of the film stress. If this assumption were to be true, a rigorous application of Stoney's formula would indeed furnish a single film stress value. This value represents the common magnitude of each of the two direct stresses in any two, mutually orthogonal directions (i.e., $\sigma_{xx}=\sigma_{yy}=\sigma^{(f)},\ \sigma_{xy}=\sigma_{yx}=0,\ \sigma^{(f)}$ independent of position). This is the uniform stress for the entire film and it is derived from the measurement of a single uniform curvature value that fully characterizes the system provided the deformation is indeed spherical.

Despite the explicitly stated assumptions of spatial stress and curvature uniformity, the Stoney formula is often, arbitrarily, applied to cases of practical interest where these assumptions are violated. This is typically done by applying Stoney's formula pointwise, and thus extracting a local value of stress from a local measurement of the curvature of the system. This approach of inferring film stress clearly violates the uniformity assumptions of the analysis and, as such, its accuracy as an approximation is expected to deteriorate as the levels of curvature nonuniformity become more severe.

Following the initial formulation by Stoney, various researchers have derived a number of extensions to relax some of the other assumptions (other than the assumption of uniformity) made by Stoney's analysis. Such extensions of the initial formulation include relaxation of the assumption of equibiaxiality as well as the assumption of small deformations/deflections. A biaxial form of Stoney, appropriate for anisotropic film stresses, including different stress values at two different directions and nonzero, in-plane shear stresses, was derived by relaxing the assumption of curvature equibiaxiality [2]. Related analyses treating discontinuous films in the form of bare periodic lines [4] or composite films with periodic line structures (e.g., bare or encapsulated periodic lines) have also been derived [5–7]. These latter analyses have also removed the assumption of equibiaxiality and have allowed the existence of three independent curvature and stress components in the form of two, nonequal, direct components and one shear or twist component. However, the uniformity assumption of all of these quantities over the entire plate system was retained. In addition to the above, single, multiple, and graded films and substrates have been treated in various "large" deformation analyses [8–11]. These analyses have removed both the restrictions of an equibiaxial curvature state as well as the assumption of infinitesimal deformations. They have allowed for the prediction of kinematically nonlinear behavior and bifurcations in curvature states. These bifurcations are transformations from an initially equibiaxial to a subsequently biaxial curvature state that may be induced by an increase in film stress beyond a critical level. This critical level is intimately related to the systems aspect ratio, i.e., the ratio of in-plane to thickness dimension and the elastic stiffness. These analyses also retain the assumption of spatial curvature and stress uniformity across the system. However, they allow for deformations to evolve from an initially spherical shape to an energetically favored shape (e.g., ellipsoidal, cylindrical, or saddle shapes) which features three different, still spatially constant, curvature components [12].

None of the above-discussed extensions of Stoney's methodology has relaxed the most restrictive of Stoney's original assumption of spatial uniformity, which does not allow either film stress or curvature components to vary across the plate surface. This crucial assumption is often violated in practice since film stresses and the associated system curvatures are nonuniformly distributed over the plate area. Huang and Rosakis [13] and Huang et al. [14] have recently made progress to remove the two restrictive assumptions of the Stoney analysis relating to spatial uniformity and equibiaxiality. They have studied the cases of thin film/substrate systems subject to nonuniform but axisymmetric temperature distribution $T(r)$ and misfit strain $\varepsilon_m(r)$, respectively. Their results show that the relations between film stresses and substrate curvatures feature not only a "local" part that involves a direct dependence of stresses on curvatures at the same point, but also a "nonlocal" part which reflects the effect of curvatures at other points on the location of scrutiny. The "nonlocal" effect comes into play in the axisymmetric analysis via the average curvature in the thin film.

The main purpose of the present paper is to remove the two restrictive assumptions of the Stoney analysis relating to spatial uniformity and equibiaxiality for the general case of a thin film/substrate system subject to arbitrary temperature distribution $T(r,\theta)$ whose presence will create a nonaxisymmetric stress and curvature field as well as arbitrarily large stress and curvature gradients. Such a nonuniform temperature field may arise in the processing or application of the thin film/substrate system. Our goal is to relate film stresses and system curvatures to the temperature distribution and to ultimately derive a relation between the film stresses and the system curvatures for general nonaxisymmetric temperature distributions. Such a relation would allow for the accurate experimental inference of film stress from full-field and real-time curvature measurements that may occur during or after thermal processing. The full-field curvature measurements (e.g., [15]), together with the present study, provide the stress field in the film.

2 Governing Equations

A thin film deposited on a substrate is subject to arbitrary temperature distribution $T(r,\theta)$, where r and θ are the polar coordinates (Fig. 1). The thin film and substrate are circular in the lateral direction and have a radius R.

The thin-film thickness h_f is much less than the substrate thickness h_s, and both are much less than R; i.e., $h_f \ll h_s \ll R$. The Young's modulus, Poisson's ratio, and coefficient of thermal expansion of the film and substrate are denoted by E_f, ν_f, α_f, E_s, ν_s, and α_s, respectively. The substrate is modeled as a plate since it can be subjected to bending, and $h_s \ll R$. The thin film is modeled as a membrane which cannot be subject to bending due to its small thickness $h_f \ll h_s$.

Let $u_r^{(f)}$ and $u_\theta^{(f)}$ denote the displacements in the radial (r) and circumferential (θ) directions. The strains in the thin film are

$$\varepsilon_{rr} = \frac{\partial u_r^{(f)}}{\partial r}, \quad \varepsilon_{\theta\theta} = \frac{u_r^{(f)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(f)}}{\partial \theta}$$

and

$$\gamma_{r\theta} = \frac{1}{r}\frac{\partial u_r^{(f)}}{\partial \theta} + \frac{\partial u_\theta^{(f)}}{\partial r} - \frac{u_\theta^{(f)}}{r}$$

The stresses in the thin film can be obtained from the linear thermo-elastic constitutive model as

$$\sigma_{rr} = \frac{E_f}{1-\nu_f^2}\left[\frac{\partial u_r^{(f)}}{\partial r} + \nu_f\left(\frac{u_r^{(f)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(f)}}{\partial \theta}\right) - (1+\nu_f)\alpha_f T\right]$$

$$\sigma_{\theta\theta} = \frac{E_f}{1-\nu_f^2}\left[\nu_f\frac{\partial u_r^{(f)}}{\partial r} + \frac{u_r^{(f)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(f)}}{\partial \theta} - (1+\nu_f)\alpha_f T\right] \quad (2.1)$$

$$\sigma_{r\theta} = \frac{E_f}{2(1+\nu_f)}\left(\frac{1}{r}\frac{\partial u_r^{(f)}}{\partial \theta} + \frac{\partial u_\theta^{(f)}}{\partial r} - \frac{u_\theta^{(f)}}{r}\right)$$

The membrane forces in the thin film are

$$N_r^{(f)} = h_f\sigma_{rr} \quad N_\theta^{(f)} = h_f\sigma_{\theta\theta} \quad N_{r\theta}^{(f)} = h_f\sigma_{r\theta} \quad (2.2)$$

It is recalled that, for uniform temperature distribution T =constant, the normal and shear stresses across the thin film/substrate interface vanish except near the free edge $r=R$; i.e., $\sigma_{zz}=\sigma_{rz}=\sigma_{r\theta}=0$ at $z=h_s/2$ and $r<R$. For nonuniform temperature distribution $T=T(r,\theta)$, the shear stress σ_{rz} and $\sigma_{\theta z}$ at the interface may not vanish anymore, and are denoted by τ_r and τ_θ, respectively. It is important to note that the normal stress traction σ_{zz} still vanishes (except near the free edge $r=R$) because the thin film cannot be subject to bending. The equilibrium equations for the thin film, accounting for the effect of interface shear stresses τ_r and τ_θ, become

$$\frac{\partial N_r^{(f)}}{\partial r} + \frac{N_r^{(f)} - N_\theta^{(f)}}{r} + \frac{1}{r}\frac{\partial N_{r\theta}^{(f)}}{\partial \theta} - \tau_r = 0$$

$$\frac{\partial N_{r\theta}^{(f)}}{\partial r} + \frac{2}{r}N_{r\theta}^{(f)} + \frac{1}{r}\frac{\partial N_\theta^{(f)}}{\partial \theta} - \tau_\theta = 0 \quad (2.3)$$

The substitution of Eqs. (2.1)–(2.3) yields the following governing equations for $u_r^{(f)}$, $u_\theta^{(f)}$, τ_r and τ_θ

$$\frac{\partial^2 u_r^{(f)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_r^{(f)}}{\partial r} - \frac{u_r^{(f)}}{r^2} + \frac{1-\nu_f}{2}\frac{1}{r^2}\frac{\partial^2 u_r^{(f)}}{\partial \theta^2} + \frac{1+\nu_f}{2}\frac{1}{r}\frac{\partial^2 u_\theta^{(f)}}{\partial r\,\partial \theta} - \frac{3-\nu_f}{2}\frac{1}{r^2}\frac{\partial u_\theta^{(f)}}{\partial \theta} = \frac{1-\nu_f^2}{E_f h_f}\tau_r + (1+\nu_f)\alpha_f\frac{\partial T}{\partial r}$$

$$\frac{1+\nu_f}{2}\frac{1}{r}\frac{\partial^2 u_r^{(f)}}{\partial r\,\partial \theta} + \frac{3-\nu_f}{2}\frac{1}{r^2}\frac{\partial u_r^{(f)}}{\partial \theta} + \frac{1-\nu_f}{2}\left(\frac{\partial^2 u_\theta^{(f)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_\theta^{(f)}}{\partial r} - \frac{u_\theta^{(f)}}{r^2}\right) + \frac{1}{r^2}\frac{\partial^2 u_\theta^{(f)}}{\partial \theta^2} = \frac{1-\nu_f^2}{E_f h_f}\tau_\theta + (1+\nu_f)\alpha_f\frac{1}{r}\frac{\partial T}{\partial \theta} \quad (2.4)$$

Let $u_r^{(s)}$ and $u_\theta^{(s)}$ denote the displacements in the radial (r) and circumferential (θ) directions, respectively, at the neutral axis (z =0) of the substrate, and w the displacement in the normal (z) direction. It is important to consider w since the substrate can be subject to bending and is modeled as a plate. The strains in the substrate are given by

$$\varepsilon_{rr} = \frac{\partial u_r^{(s)}}{\partial r} - z\frac{\partial^2 w}{\partial r^2}$$

$$\varepsilon_{\theta\theta} = \frac{u_r^{(s)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial \theta} - z\left(\frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2}\right) \quad (2.5)$$

$$\gamma_{r\theta} = \frac{1}{r}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r} - 2z\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w}{\partial \theta}\right)$$

The stresses in the substrate can then be obtained from the linear thermo-elastic constitutive model as

$$\sigma_{rr} = \frac{E_s}{1-\nu_s^2}\left\{\frac{\partial u_r^{(s)}}{\partial r} + \nu_s\left(\frac{u_r^{(s)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial \theta}\right) - z\left[\frac{\partial^2 w}{\partial r^2} + \nu_s\left(\frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2}\right)\right] - (1+\nu_s)\alpha_s T\right\}$$

$$\sigma_{\theta\theta} = \frac{E_s}{1-\nu_s^2}\left[\nu_s\frac{\partial u_r^{(s)}}{\partial r} + \frac{u_r^{(s)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial \theta} - z\left(\nu_s\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2}\right) - (1+\nu_s)\alpha_s T\right] \quad (2.6)$$

$$\sigma_{r\theta} = \frac{E_s}{2(1+\nu_s)}\left[\frac{1}{r}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r} - 2z\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w}{\partial \theta}\right)\right]$$

The forces and bending moments in the substrate are

$$N_r^{(s)} = \int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} \sigma_{rr}dz = \frac{E_s h_s}{1-\nu_s^2}\left[\frac{\partial u_r^{(s)}}{\partial r} + \nu_s\left(\frac{u_r^{(s)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial \theta}\right) - (1+\nu_s)\alpha_s T\right]$$

$$N_\theta^{(s)} = \int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} \sigma_{\theta\theta}dz = \frac{E_s h_s}{1-\nu_s^2}\left[\nu_s\frac{\partial u_r^{(s)}}{\partial r} + \frac{u_r^{(s)}}{r} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial \theta} - (1+\nu_s)\alpha_s T\right] \quad (2.7)$$

$$N_{r\theta}^{(s)} = \int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} \sigma_{r\theta}dz = \frac{E_s h_s}{2(1+\nu_s)}\left(\frac{1}{r}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r}\right)$$

$$M_r = -\int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} z\sigma_{rr}dz = \frac{E_s h_s^3}{12(1-\nu_s^2)}\left[\frac{\partial^2 w}{\partial r^2} + \nu_s\left(\frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2}\right)\right]$$

$$M_\theta = -\int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} z\sigma_{\theta\theta}dz = \frac{E_s h_s^3}{12(1-\nu_s^2)}\left(\nu_s\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2}\right) \quad (2.8)$$

$$M_{r\theta} = -\int_{-\frac{h_s}{2}}^{\frac{h_s}{2}} z\sigma_{r\theta}dz = \frac{E_s h_s^3}{12(1+\nu_s)}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w}{\partial \theta}\right)$$

The shear stresses τ_r and τ_θ at the thin film/substrate interface are equivalent to the distributed forces τ_r in the radial direction and τ_θ in the circumferential direction, and bending moments $(h_s/2)\tau_r$ and $(h_s/2)\tau_\theta$ applied at the neutral axis (z=0) of the substrate. The in-plane force equilibrium equations of the substrate then become

$$\frac{\partial N_r^{(s)}}{\partial r} + \frac{N_r^{(s)} - N_\theta^{(s)}}{r} + \frac{1}{r}\frac{\partial N_{r\theta}^{(s)}}{\partial \theta} + \tau_r = 0$$

$$\frac{\partial N_{r\theta}^{(s)}}{\partial r} + \frac{2}{r}N_{r\theta}^{(s)} + \frac{1}{r}\frac{\partial N_{\theta\theta}^{(s)}}{\partial \theta} + \tau_\theta = 0 \tag{2.9}$$

The out-of-plane moment and force equilibrium equations are given by

$$\begin{gathered}\frac{\partial M_r}{\partial r} + \frac{M_r - M_\theta}{r} + \frac{1}{r}\frac{\partial M_{r\theta}}{\partial \theta} + Q_r - \frac{h_s}{2}\tau_r = 0\\ \frac{\partial M_{r\theta}}{\partial r} + \frac{2}{r}M_{r\theta} + \frac{1}{r}\frac{\partial M_\theta}{\partial \theta} + Q_\theta - \frac{h_s}{2}\tau_\theta = 0\end{gathered} \tag{2.10}$$

$$\frac{\partial Q_r}{\partial r} + \frac{Q_r}{r} + \frac{1}{r}\frac{\partial Q_\theta}{\partial \theta} = 0 \tag{2.11}$$

where Q_r and Q_θ are the shear forces normal to the neutral axis. The substitution of Eq. (2.7) into Eq. (2.9) yields the following governing equations for $u_r^{(s)}$, $u_\theta^{(s)}$, and τ:

$$\begin{gathered}\frac{\partial^2 u_r^{(s)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_r^{(s)}}{\partial r} - \frac{u_r^{(s)}}{r^2} + \frac{1-\nu_s}{2}\frac{1}{r^2}\frac{\partial^2 u_r^{(s)}}{\partial \theta^2} + \frac{1+\nu_s}{2}\frac{1}{r}\frac{\partial^2 u_\theta^{(s)}}{\partial r\,\partial \theta}\\ - \frac{3-\nu_s}{2}\frac{1}{r^2}\frac{\partial u_\theta^{(s)}}{\partial \theta} = -\frac{1-\nu_s^2}{E_s h_s}\tau_r + (1+\nu_s)\alpha_s\frac{\partial T}{\partial r}\\ \frac{1+\nu_s}{2}\frac{1}{r}\frac{\partial^2 u_r^{(s)}}{\partial r\,\partial \theta} + \frac{3-\nu_s}{2}\frac{1}{r^2}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{1-\nu_s}{2}\left(\frac{\partial^2 u_\theta^{(s)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r^2}\right)\\ + \frac{1}{r^2}\frac{\partial^2 u_\theta^{(s)}}{\partial \theta^2} = -\frac{1-\nu_s^2}{E_s h_s}\tau_\theta + (1+\nu_s)\alpha_s\frac{1}{r}\frac{\partial T}{\partial \theta}\end{gathered} \tag{2.12}$$

Elimination of Q_r and Q_θ from Eqs. (2.10) and (2.11), in conjunction with Eq. (2.8), gives the following governing equation for w (and τ)

$$\nabla^2(\nabla^2 w) = \frac{6(1-\nu_s^2)}{E_s h_s^2}\left(\frac{\partial \tau_r}{\partial r} + \frac{\tau_r}{r} + \frac{1}{r}\frac{\partial \tau_\theta}{\partial \theta}\right) \tag{2.13}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}$$

The continuity of displacements across the thin film/substrate interface requires

$$u_r^{(f)} = u_r^{(s)} - \frac{h_s}{2}\frac{\partial w}{\partial r}, \quad u_\theta^{(f)} = u_\theta^{(s)} - \frac{h_s}{2}\frac{1}{r}\frac{\partial w}{\partial \theta} \tag{2.14}$$

Equations (2.4) and (2.12)–(2.14) constitute seven ordinary differential equations for seven variables, namely $u_r^{(f)}$, $u_\theta^{(f)}$, $u_r^{(s)}$, $u_\theta^{(s)}$, w, τ_r, and τ_θ. We discuss below how to decouple these seven equations under the limit $h_f/h_s \ll 1$ such that we can solve $u_r^{(s)}$, $u_\theta^{(s)}$ first, then $u_r^{(f)}$ and $u_\theta^{(f)}$, followed by τ_r and τ_θ, and finally w.

(i) Elimination of τ_r and τ_θ from force equilibrium equations (2.4) for the thin film and (2.12) for the substrate yields two equations for $u_r^{(f)}$, $u_\theta^{(f)}$, $u_r^{(s)}$, and $u_\theta^{(s)}$. For $h_f/h_s \ll 1$, $u_r^{(f)}$ and $u_\theta^{(f)}$ disappear in these two equations, which become the following governing equations for $u_r^{(s)}$ and $u_\theta^{(s)}$ only:

$$\begin{gathered}\frac{\partial^2 u_r^{(s)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_r^{(s)}}{\partial r} - \frac{u_r^{(s)}}{r^2} + \frac{1-\nu_s}{2}\frac{1}{r^2}\frac{\partial^2 u_r^{(s)}}{\partial \theta^2} + \frac{1+\nu_s}{2}\frac{1}{r}\frac{\partial^2 u_\theta^{(s)}}{\partial r\,\partial \theta}\\ - \frac{3-\nu_s}{2}\frac{1}{r^2}\frac{\partial u_\theta^{(s)}}{\partial \theta} = (1+\nu_s)\alpha_s\frac{\partial T}{\partial r} + O\left(\frac{h_f}{h_s}\right)\\ \frac{1+\nu_s}{2}\frac{1}{r}\frac{\partial^2 u_r^{(s)}}{\partial r\,\partial \theta} + \frac{3-\nu_s}{2}\frac{1}{r^2}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{1-\nu_s}{2}\left(\frac{\partial^2 u_\theta^{(s)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r^2}\right)\\ + \frac{1}{r^2}\frac{\partial^2 u_\theta^{(s)}}{\partial \theta^2} = (1+\nu_s)\alpha_s\frac{1}{r}\frac{\partial T}{\partial \theta} + O\left(\frac{h_f}{h_s}\right)\end{gathered} \tag{2.15}$$

(ii) Elimination of $u_r^{(f)}$ and $u_\theta^{(f)}$ from the continuity condition (2.14) and equilibrium equation (2.4) for the thin film gives τ_r and τ_θ in terms of $u_r^{(s)}$ and w. Their substitution into the moment equilibrium equation (2.13) yields the governing equation for the normal displacement w, from which it can be shown that w is on the order of h_f/h_s, i.e.,

$$w = O\left(\frac{h_f}{h_s}\right) \tag{2.16}$$

Equation (2.16) and the continuity condition (2.14) then give the displacements $u_r^{(f)}$ and $u_\theta^{(f)}$ in the thin film as

$$u_r^{(f)} = u_r^{(s)} + O\left(\frac{h_f}{h_s}\right), \quad u_\theta^{(f)} = u_\theta^{(s)} + O\left(\frac{h_f}{h_s}\right) \tag{2.17}$$

(iii) The equilibrium equation (2.4) for the thin film gives the interface shear stresses in terms of $u_r^{(s)}$ and $u_\theta^{(s)}$ as

$$\begin{gathered}\tau_r = \frac{E_f h_f}{1-\nu_f^2}\left\{\frac{\nu_s - \nu_f}{2}\left(\frac{1}{r^2}\frac{\partial^2 u_r^{(s)}}{\partial \theta^2} - \frac{1}{r}\frac{\partial^2 u_\theta^{(s)}}{\partial r\,\partial \theta} - \frac{1}{r^2}\frac{\partial u_\theta^{(s)}}{\partial \theta}\right) + [(1+\nu_s)\alpha_s - (1+\nu_f)\alpha_f]\frac{\partial T}{\partial r} + O\left(\frac{h_f}{h_s}\right)\right\}\\ \tau_\theta = \frac{E_f h_f}{1-\nu_f^2}\left\{\frac{\nu_s - \nu_f}{2}\left(-\frac{1}{r}\frac{\partial^2 u_r^{(s)}}{\partial r\,\partial \theta} + \frac{1}{r^2}\frac{\partial u_r^{(s)}}{\partial \theta} + \frac{\partial^2 u_\theta^{(s)}}{\partial r^2} + \frac{1}{r}\frac{\partial u_\theta^{(s)}}{\partial r} - \frac{u_\theta^{(s)}}{r^2}\right) + [(1+\nu_s)\alpha_s - (1+\nu_f)\alpha_f]\frac{1}{r}\frac{\partial T}{\partial \theta} + O\left(\frac{h_f}{h_s}\right)\right\}\end{gathered} \tag{2.18}$$

where Eq. (2.15) has been used.

(iv) The displacement w is determined from the moment equilibrium equation (2.13) by eliminating τ_r and τ_θ using Eq. (2.18). It can be verified that the resulting w is indeed on the order of h_f/h_s as suggested in Eq. (2.16).

We expand the arbitrary nonuniform temperature distribution $T(r,\theta)$ to the Fourier series,

$$T(r,\theta) = \sum_{n=0}^{\infty} T_c^{(n)}(r)\cos n\theta + \sum_{n=0}^{\infty} T_s^{(n)}(r)\sin n\theta \tag{2.19}$$

where

$$T_c^{(0)}(r) = \frac{1}{2\pi}\int_0^{2\pi} T(r,\theta)d\theta, \quad T_c^{(n)}(r) = \frac{1}{\pi}\int_0^{2\pi} T(r,\theta)\cos n\theta d\theta \quad (n \ge 1)$$

and

$$T_s^{(n)}(r) = \frac{1}{\pi}\int_0^{2\pi} T(r,\theta)\sin n\theta d\theta \ (n \ge 1)$$

Without losing generality, we focus on the $\cos n\theta$ term here. The corresponding displacements and interface shear stresses can be expressed as

$$\begin{gathered}u_r^{(s)} = u_r^{(sn)}(r)\cos n\theta, \quad u_\theta^{(s)} = u_\theta^{(sn)}(r)\sin n\theta, \quad w = w^{(n)}(r)\cos n\theta\\ \tau_r = \tau_r^{(n)}(r)\cos n\theta, \quad \tau_\theta = \tau_\theta^{(n)}(r)\sin n\theta\end{gathered} \tag{2.20}$$

Equation (2.15) then gives two ordinary differential equations for $u_r^{(sn)}$ and $u_\theta^{(sn)}$, which have the general solution

$$
\begin{Bmatrix} u_r^{(sn)} \\ u_\theta^{(sn)} \end{Bmatrix} = \begin{Bmatrix} 1-\nu_s - \frac{1+\nu_s}{2}n \\ \frac{1+\nu_s}{2}n + 2 \end{Bmatrix} \left[A_0 r^{n+1} + \frac{\alpha_s}{4(n+1)} \frac{1+\nu_s}{1-\nu_s} r T_c^{(n)} \right]
+ \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \frac{\alpha_s}{4(n+1)} \frac{1+\nu_s}{1-\nu_s} \left\{ -\left[1-\nu_s - \frac{n}{2}(1+\nu_s) \right] r T_c^{(n)} + 2(1-\nu_s)(n+1)\frac{1}{r^{n+1}} \int_0^r \eta^{1+n} T_c^{(n)}(\eta) d\eta \right\}
- \begin{Bmatrix} 1-\nu_s + \frac{1+\nu_s}{2}n \\ \frac{1+\nu_s}{2}n - 2 \end{Bmatrix} \frac{\alpha_s}{4(n-1)} \frac{1+\nu_s}{1-\nu_s} r T_c^{(n)}
+ \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} D_0 r^{n-1} - \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \frac{\alpha_s}{4(n-1)} \frac{1+\nu_s}{1-\nu_s} \times \begin{Bmatrix} \left[1-\nu_s + \frac{n}{2}(1+\nu_s) \right] r T_c^{(n)} \\ -2(1-\nu_s)(n-1) r^{n-1} \int_r^R \eta^{1-n} T_c^{(n)}(\eta) d\eta \end{Bmatrix}
+ O\left(\frac{h_f}{h_s}\right) \tag{2.21}
$$

where we have used the condition that the displacements are finite at the center $r=0$, and A_0 and D_0 are constants to be determined.

The interface shear stresses are obtained from Eq. (2.18) as

$$
\begin{aligned}
\tau_r^{(n)} &= \frac{E_f h_f}{1-\nu_f^2} \left\{ [(1+\nu_s)\alpha_s - (1+\nu_f)\alpha_f] \frac{dT_c^{(n)}}{dr} - 2(\nu_s - \nu_f) n(n+1) A_0 r^{n-1} + O\left(\frac{h_f}{h_s}\right) \right\} \\
\tau_\theta^{(n)} &= \frac{E_f h_f}{1-\nu_f^2} \left\{ -[(1+\nu_s)\alpha_s - (1+\nu_f)\alpha_f] \frac{n}{r} T_c^{(n)} + 2(\nu_s - \nu_f) n(n+1) A_0 r^{n-1} + O\left(\frac{h_f}{h_s}\right) \right\}
\end{aligned} \tag{2.22}
$$

The normal displacement w is determined from Eq. (2.13) as

$$
w^{(n)} = A_1 r^{n+2} + B_1 r^n - \frac{3}{n} \frac{1-\nu_s^2}{E_s h_s^2} \frac{E_f h_f}{1-\nu_f^2} [(1+\nu_s)\alpha_s - (1+\nu_f)\alpha_f] \times \left[r^n \int_r^R \eta^{1-n} T_c^{(n)}(\eta) d\eta + r^{-n} \int_0^r \eta^{n+1} T_c^{(n)}(\eta) d\eta \right] + O\left(\frac{h_f^2}{h_s^2}\right) \tag{2.23}
$$

where we have used the condition that the displacement w is finite at the center $r=0$, and A_1 and B_1 are constants to be determined.

3 Boundary Conditions

The first two boundary conditions at the free edge $r=R$ require that the net forces vanish:

$$
N_r^{(f)} + N_r^{(s)} = 0 \quad \text{and} \quad N_{r\theta}^{(f)} + N_{r\theta}^{(s)} = 0 \quad \text{at } r=R \tag{3.1}
$$

which give A_0 and D_0 as

$$
A_0 = \frac{\alpha_s}{R^{2n+2}} \int_0^R \eta^{n+1} T_c^{(n)}(\eta) d\eta + O\left(\frac{h_f}{h_s}\right)
$$

$$
D_0 = -\frac{n+1}{2R^{2n}} (1+\nu_s)\alpha_s \int_0^R \eta^{n+1} T_c^{(n)}(\eta) d\eta + O\left(\frac{h_f}{h_s}\right) \tag{3.2}
$$

under the limit $h_f/h_s \ll 1$. The other two boundary conditions at the free edge $r=R$ are the vanishing of net moments, i.e.,

$$
M_r - \frac{h_s}{2} N_r^{(f)} = 0 \text{ and } Q_r - \frac{1}{r}\frac{\partial}{\partial\theta}\left(M_{r\theta} - \frac{h_s}{2} N_{r\theta}^{(f)} \right) = 0 \text{ at } r=R \tag{3.3}
$$

which give A_1 and B_1 as

$$
\begin{aligned}
A_1 &= 3 \frac{E_f h_f}{1-\nu_f^2} \frac{1-\nu_s^2}{E_s h_s^2} \left[(1+\nu_f) \frac{1-\nu_s}{3+\nu_s} (\alpha_s - \alpha_f) - (\nu_s - \nu_f)\alpha_s \right] \frac{1}{R^{2n+2}} \int_0^R \eta^{n+1} T_c^{(n)}(\eta) d\eta + O\left(\frac{h_f^2}{h_s^2}\right) \\
B_1 &= -\frac{n+1}{n} R^2 A_1 + O\left(\frac{h_f^2}{h_s^2}\right)
\end{aligned} \tag{3.4}
$$

It is important to point out that the boundary conditions can also be established from the variational principle (e.g., [11]). The total potential energy in the thin film/substrate system with the free edge at $r=R$ is

$$
\Pi = \int_0^R r dr \int_0^{2\pi} d\theta \int_{-\frac{h_s}{2}}^{\frac{h_s}{2}+h_f} U dz \tag{3.5}
$$

where U is the strain energy density which gives $\partial U/\partial \varepsilon_{rr} = \sigma_{rr}$, $\partial U/\partial \varepsilon_{\theta\theta} = \sigma_{\theta\theta}$, and $\partial U/\partial \gamma_{r\theta} = \sigma_{r\theta}$. For constitutive relations in Eqs. (2.1) and (2.6), we obtain

$$
U = \frac{E}{2(1-\nu^2)} \left[\varepsilon_{rr}^2 + \varepsilon_{\theta\theta}^2 + 2\nu \varepsilon_{rr}\varepsilon_{\theta\theta} + \frac{1-\nu}{2}\gamma_{r\theta}^2 - 2(1+\nu)\alpha T(\varepsilon_{rr} + \varepsilon_{\theta\theta}) \right] \tag{3.6}
$$

where E, ν, and α take their corresponding values in the thin film (i.e., E_f, ν_f, and α_f for $h_s/2 + h_f \geq z \geq h_s/2$) and in the substrate (i.e., E_s, ν_s, and α_s for $h_s/2 \geq z \geq -h_s/2$). For the displacement fields in Sec. 2 and the associated strain fields, the potential energy Π in Eq. (3.5) becomes a quadratic function of parameters A_0, D_0, A_1, and B_1. The principle of minimum potential energy requires

$$
\frac{\partial \Pi}{\partial A_0} = 0 \quad \frac{\partial \Pi}{\partial D_0} = 0 \quad \frac{\partial \Pi}{\partial A_1} = 0 \quad \frac{\partial \Pi}{\partial B_1} = 0 \tag{3.7}
$$

It can be shown that, as expected in the limit $h_f/h_s \ll 1$, the above four conditions in Eq. (3.7) are equivalent to the vanishing of net forces in Eq. (3.1) and net moments in Eq. (3.3).

4 Thin Film Stresses and Substrate Curvatures

We provide the general solution that includes both cosine and sine terms in this section. The substrate curvatures are

$$
\kappa_{rr} = \frac{\partial^2 w}{\partial r^2} \quad \kappa_{\theta\theta} = \frac{1}{r}\frac{\partial w}{\partial r} + \frac{1}{r^2}\frac{\partial^2 w}{\partial \theta^2} \quad \kappa_{r\theta} = \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w}{\partial \theta} \right) \tag{4.1}
$$

The sum of substrate curvatures is related to the temperature by

$$\kappa_{rr}+\kappa_{\theta\theta}=12\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s}{E_s h_s^2}\left\{(\alpha_s-\alpha_f)T+\left[\frac{(1+\nu_s)^2}{2(1+\nu_f)}-1\right]\alpha_s(T-\overline{T})+\frac{(1-\nu_s)}{2}\alpha_f(T-\overline{T})+(1+\nu_s)\left[\frac{1-\nu_s}{3+\nu_s}(\alpha_s-\alpha_f)-\frac{\nu_s-\nu_f}{1+\nu_f}\alpha_s\right]\sum_{n=1}^{\infty}(n+1)\frac{r^n}{R^{2n+2}}\left[\cos n\theta\int_0^R \eta^{n+1}T_c^{(n)}(\eta)d\eta+\sin n\theta\int_0^R \eta^{n+1}T_s^{(n)}(\eta)d\eta\right]\right\} \tag{4.2}$$

where $\overline{T}=(1/\pi R^2)\int\int_A T(\eta,\varphi)dA$ is the average temperature over the entire area A of the thin film, $dA=\eta d\eta d\varphi$, and $\overline{T}$ is also related to $T_c^{(0)}$ by $\overline{T}=(2/R^2)\int_0^R \eta T_c^{(0)}(\eta)d\eta$. The difference between two curvatures $(\kappa_{rr}-\kappa_{\theta\theta})$ and the twist $\kappa_{r\theta}$ are given by

$$\kappa_{rr}-\kappa_{\theta\theta}=6\frac{E_f h_f}{1-\nu_f^2}\frac{1-\nu_s^2}{E_s h_s^2}[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f]\times\left[T-\frac{2}{r^2}\int_0^r \eta T_c^{(0)}d\eta-\sum_{n=1}^{\infty}\frac{n+1}{r^{n+2}}\left(\cos n\theta\int_0^r \eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^r \eta^{n+1}T_s^{(n)}d\eta\right)-\sum_{n=1}^{\infty}(n-1)r^{n-2}\left(\cos n\theta\int_r^R \eta^{1-n}T_c^{(n)}d\eta+\sin n\theta\int_r^R \eta^{1-n}T_s^{(n)}d\eta\right)\right]+6\frac{E_f h_f}{1-\nu_f^2}\frac{1-\nu_s^2}{E_s h_s^2}\frac{1}{3+\nu_s}\times\{(1-\nu_s)[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f]-4(\nu_s-\nu_f)\alpha_s\}\sum_{n=1}^{\infty}\frac{n+1}{R^{n+2}}\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right]\times\left(\cos n\theta\int_0^R \eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^R \eta^{n+1}T_s^{(n)}d\eta\right) \tag{4.3}$$

$$\kappa_{r\theta}=3\frac{E_f h_f}{1-\nu_f^2}\frac{1-\nu_s^2}{E_s h_s^2}[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f]\times\left[-\sum_{n=1}^{\infty}\frac{n+1}{r^{n+2}}\left(\sin n\theta\int_0^r \eta^{n+1}T_c^{(n)}d\eta-\cos n\theta\int_0^r \eta^{n+1}T_s^{(n)}d\eta\right)+\sum_{n=1}^{\infty}(n-1)r^{n-2}\left(\sin n\theta\int_r^R \eta^{1-n}T_c^{(n)}d\eta-\cos n\theta\int_r^R \eta^{1-n}T_s^{(n)}d\eta\right)\right]-3\frac{E_f h_f}{1-\nu_f^2}\frac{1-\nu_s^2}{E_s h_s^2}\frac{1}{3+\nu_s}\{(1-\nu_s)\times[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f]-4(\nu_s-\nu_f)\alpha_s\}\sum_{n=1}^{\infty}\frac{n+1}{R^{n+2}}\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right]\times\left(\sin n\theta\int_0^R \eta^{n+1}T_c^{(n)}d\eta-\cos n\theta\int_0^R \eta^{n+1}T_s^{(n)}d\eta\right) \tag{4.4}$$

The stresses in the thin film are obtained from Eq. (2.1). Specifically, the sum of stresses $\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}$ is related to the temperature by

$$\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}=\frac{E_f}{1-\nu_f}\left[2(\alpha_s-\alpha_f)T-(1-\nu_s)\alpha_s(T-\overline{T})+2(1-\nu_s)\alpha_s\sum_{n=1}^{\infty}\frac{n+1}{R^{2n+2}}r^n\left(\cos n\theta\int_0^R \eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^R \eta^{n+1}T_s^{(n)}d\eta\right)\right] \tag{4.5}$$

The difference between stresses, $\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}$, and shear stress $\sigma_{r\theta}^{(f)}$ are given by

$$\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}=\frac{E_f}{1+\nu_f}(1+\nu_s)\alpha_s\left\{T-\frac{2}{r^2}\int_0^r \eta T_c^{(0)}d\eta-\sum_{n=1}^{\infty}\frac{n+1}{r^{n+2}}\left(\cos n\theta\int_0^r \eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^r \eta^{n+1}T_s^{(n)}d\eta\right)-\sum_{n=1}^{\infty}(n-1)r^{n-2}\left(\cos n\theta\int_r^R \eta^{1-n}T_c^{(n)}d\eta+\sin n\theta\int_r^R \eta^{1-n}T_s^{(n)}d\eta\right)-\sum_{n=1}^{\infty}\frac{n+1}{R^{n+2}}\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right]\left(\cos n\theta\int_0^R \eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^R \eta^{n+1}T_s^{(n)}d\eta\right)\right\} \tag{4.6}$$

$$\sigma_{r\theta}^{(f)}=\frac{E_f}{2(1+\nu_f)}(1+\nu_s)\alpha_s\left\{-\sum_{n=1}^{\infty}\frac{n+1}{r^{n+2}}\left(\sin n\theta\int_0^r \eta^{n+1}T_c^{(n)}d\eta-\cos n\theta\int_0^r \eta^{n+1}T_s^{(n)}d\eta\right)+\sum_{n=1}^{\infty}(n-1)r^{n-2}\left(\sin n\theta\int_r^R \eta^{1-n}T_c^{(n)}d\eta-\cos n\theta\int_r^R \eta^{1-n}T_s^{(n)}d\eta\right)+\sum_{n=1}^{\infty}\frac{n+1}{R^{n+2}}\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right]\left(\sin n\theta\int_0^R \eta^{n+1}T_c^{(n)}d\eta-\cos n\theta\int_0^R \eta^{n+1}T_s^{(n)}d\eta\right)\right\} \tag{4.7}$$

The interface shear stresses τ_r and τ_θ are related to the temperature by

$$\tau_r=\frac{E_f h_f}{1-\nu_f^2}\left[\left[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f\right]\frac{\partial T}{\partial r}-2(\nu_s-\nu_f)\alpha_s\sum_{n=1}^{\infty}n(n+1)\frac{r^{n-1}}{R^{2n+2}}\left(\cos n\theta\int_0^R\eta^{n+1}T_c^{(n)}d\eta+\sin n\theta\int_0^R\eta^{n+1}T_s^{(n)}d\eta\right)\right] \tag{4.8}$$

$$\tau_\theta=\frac{E_f h_f}{1-\nu_f^2}\left\{\left[(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f\right]\frac{1}{r}\frac{\partial T}{\partial \theta}+2(\nu_s-\nu_f)\alpha_s\sum_{n=1}^{\infty}n(n+1)\frac{r^{n-1}}{R^{2n+2}}\left(\sin n\theta\int_0^R\eta^{n+1}T_c^{(n)}d\eta-\cos n\theta\int_0^R\eta^{n+1}T_s^{(n)}d\eta\right)\right\} \tag{4.9}$$

For uniform temperature distribution T=constant, the curvatures in the substrate obtained from Eqs. (4.2)–(4.4) become

$$\kappa=\kappa_{rr}=\kappa_{\theta\theta}=6\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s}{E_s h_s^2}(\alpha_s-\alpha_f)T$$

The stresses in the thin film obtained from Eqs. (4.5)–(4.7) become

$$\sigma^{(f)}=\sigma_{rr}^{(f)}=\sigma_{\theta\theta}^{(f)}=\frac{E_f}{1-\nu_f}(\alpha_s-\alpha_f)T$$

For this special case only, both stress and curvature states become equibiaxial. The elimination of temperature T from the above two equations yields a simple relation $\sigma^{(f)}=(E_s h_s^2/6(1-\nu_s)h_f)\kappa$, which is exactly the Stoney formula in Eq. (1.1), and it has been used to estimate the thin-film stress $\sigma^{(f)}$ from the substrate curvature κ, if the temperature, stress, and curvature are all constant and if the plate system shape is spherical. In the following, we extend such a relation for arbitrary nonaxisymmetric temperature distribution.

5 Extension of Stoney Formula for Nonaxisymmetric Temperature Distribution

The stresses and curvatures are all given in terms of temperature in the previous section. We extend the Stoney formula for arbitrary nonuniform and nonaxisymmetric temperature distribution in this section by establishing the direct relation between the thinfilm stresses and substrate curvatures.

We first define the coefficients C_n and S_n related to the substrate curvatures by

$$C_n=\frac{1}{\pi R^2}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\left(\frac{\eta}{R}\right)^n\cos n\varphi dA$$
$$S_n=\frac{1}{\pi R^2}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\left(\frac{\eta}{R}\right)^n\sin n\varphi dA \tag{5.1}$$

where the integration is over the entire area A of the thin film, and $dA=\eta d\eta d\varphi$. Since both the substrate curvatures and film stresses depend on the temperature T, elimination of temperature gives the film stress in terms of substrate curvatures by

$$\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}=\frac{E_s h_s^2}{1-\nu_s}\frac{1-\nu_f}{6h_f}\frac{\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\left\{\kappa_{rr}-\kappa_{\theta\theta}-\sum_{n=1}^{\infty}(n+1)\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right](C_n\cos n\theta+S_n\sin n\theta)\right\} \tag{5.2}$$

$$\sigma_{r\theta}^{(f)}=\frac{E_s h_s^2}{1-\nu_s}\frac{1-\nu_f}{6h_f}\frac{\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\left\{\kappa_{r\theta}+\frac{1}{2}\sum_{n=1}^{\infty}(n+1)\times\left[n\left(\frac{r}{R}\right)^n-(n-1)\left(\frac{r}{R}\right)^{n-2}\right](C_n\sin n\theta-S_n\cos n\theta)\right\} \tag{5.3}$$

$$\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}=\frac{E_s h_s^2}{6h_f(1-\nu_s)}\left\{\kappa_{rr}+\kappa_{\theta\theta}+\left[\frac{1-\nu_s}{1+\nu_s}-\frac{(1-\nu_f)\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\right](\kappa_{rr}+\kappa_{\theta\theta}-\overline{\kappa_{rr}+\kappa_{\theta\theta}})-\left[\frac{1-\nu_s}{1+\nu_s}-\frac{2(1-\nu_f)\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\right]\sum_{n=1}^{\infty}(n+1)\times\left(\frac{r}{R}\right)^n(C_n\cos n\theta+S_n\sin n\theta)\right\} \tag{5.4}$$

where $\overline{\kappa_{rr}+\kappa_{\theta\theta}}=C_0=(1/\pi R^2)\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})dA$ is the average curvature over entire area A of the thin film. Equations (5.2)–(5.4) provide direct relations between individual film stresses and substrate curvatures. It is important to note that stresses at a point in the thin film depend not only on curvatures at the same point (local dependence), but also on the curvatures in the entire substrate (nonlocal dependence) via the coefficients C_n and S_n.

The interface shear stresses τ_r and τ_θ can also be directly related to substrate curvatures via

$$\tau_r=\frac{E_s h_s^2}{6(1-\nu_s^2)}\left[\frac{\partial}{\partial r}(\kappa_{rr}+\kappa_{\theta\theta})-\frac{1-\nu_s}{2R}\sum_{n=1}^{\infty}n(n+1)(C_n\cos n\theta+S_n\sin n\theta)\left(\frac{r}{R}\right)^{n-1}\right] \tag{5.5}$$

$$\tau_\theta=\frac{E_s h_s^2}{6(1-\nu_s^2)}\left[\frac{1}{r}\frac{\partial}{\partial \theta}(\kappa_{rr}+\kappa_{\theta\theta})+\frac{1-\nu_s}{2R}\sum_{n=1}^{\infty}n(n+1)(C_n\sin n\theta-S_n\cos n\theta)\left(\frac{r}{R}\right)^{n-1}\right] \tag{5.6}$$

This provides a way to estimate the interface shear stresses from the gradients of substrate curvatures. It also displays a nonlocal dependence via the coefficients C_n and S_n.

Since interfacial shear stresses are responsible for promoting system failures through delamination of the thin film from the substrate, Eqs. (5.5) and (5.6) have particular significance. They show that such stresses are related to the gradients of $\kappa_{rr}+\kappa_{\theta\theta}$ and not to its magnitude, as might have been expected of a local, Stoney-like formulation. The implementation value of Eqs. (5.5) and (5.6) is that it provides an easy way of inferring these special interfacial shear stresses once the full-field curvature information is available. As a result, the methodology also provides a way to evaluate the risk of and to mitigate such important forms of failure. It should be noted that for the special case of spatially constant curvatures, the interfacial shear stresses vanish as is the case for all Stoney-like formulations described in Sec. 1.

It can be shown that the relations between the film stresses and substrate curvatures given in the form of infinite series in Eqs. (5.2)–(5.4) can be equivalently expressed in the form of integration as

$$\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}=\frac{E_s h_s^2}{1-\nu_s}\frac{1-\nu_f}{6h_f}\frac{\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\times\left\{\kappa_{rr}-\kappa_{\theta\theta}-\frac{1}{\pi R^2}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\times\frac{\frac{\eta}{R}F_{\text{minus}}\left(\frac{r}{R},\frac{\eta}{R},\varphi-\theta\right)}{\left[1-2\frac{\eta r}{R^2}\cos(\varphi-\theta)+\frac{\eta^2 r^2}{R^4}\right]^3}dA\right\} \tag{5.7}$$

$$\sigma_{r\theta}^{(f)}=\frac{E_s h_s^2}{1-\nu_s}\frac{1-\nu_f}{6h_f}\frac{\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\left\{\kappa_{r\theta}-\frac{1}{2}\frac{1}{\pi R^2}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\frac{\frac{\eta}{R}F_{\text{shear}}\left(\frac{r}{R},\frac{\eta}{R},\varphi-\theta\right)}{\left[1-2\frac{\eta r}{R^2}\cos(\varphi-\theta)+\frac{\eta^2 r^2}{R^4}\right]^3}dA\right\} \tag{5.8}$$

$$\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}=\frac{E_s h_s^2}{6h_f(1-\nu_s)}\left\{\kappa_{rr}+\kappa_{\theta\theta}+\left[\frac{1-\nu_s}{1+\nu_s}-\frac{(1-\nu_f)\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\right](\kappa_{rr}+\kappa_{\theta\theta}-\overline{\kappa_{rr}+\kappa_{\theta\theta}})-\left[\frac{1-\nu_s}{1+\nu_s}-\frac{2(1-\nu_f)\alpha_s}{(1+\nu_s)\alpha_s-(1+\nu_f)\alpha_f}\right]\frac{r}{\pi R^3}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\frac{\frac{\eta}{R}F_{\text{plus}}\left(\frac{r}{R},\frac{\eta}{R},\varphi-\theta\right)}{\left[1-2\frac{\eta r}{R^2}\cos(\varphi-\theta)+\frac{\eta^2 r^2}{R^4}\right]^4}dA\right\} \tag{5.9}$$

where functions F_{minus}, F_{shear}, and F_{plus} are given by

$$\begin{aligned}F_{\text{minus}}(r_1,\eta_1,\varphi_1)&=-r_1^2\eta_1(6+9\eta_1^2+r_1^2\eta_1^4)+r_1(2+9\eta_1^2+6r_1^2\eta_1^2+6r_1^2\eta_1^4)\cos\varphi_1-\eta_1(3+3r_1^2\eta_1^2+2r_1^4\eta_1^2)\cos 2\varphi_1\\&\quad+r_1\eta_1^2\cos 3\varphi_1\\F_{\text{shear}}(r_1,\eta_1,\varphi_1)&=r_1(2+9\eta_1^2-6r_1^2\eta_1^2)\sin\varphi_1-\eta_1(3+3r_1^2\eta_1^2-2r_1^4\eta_1^2)\sin 2\varphi_1+r_1\eta_1^2\sin 3\varphi_1\\F_{\text{plus}}(r_1,\eta_1,\varphi_1)&=2(1+2r_1^2\eta_1^2)\cos\varphi_1-r_1\eta_1\cos 2\varphi_1-r_1\eta_1(4+r_1^2\eta_1^2)\end{aligned} \tag{5.10}$$

The interface shear stresses can also be related to substrate curvatures via integrals as

$$\tau_r=\frac{E_s h_s^2}{6(1-\nu_s^2)}\left\{\frac{\partial}{\partial r}(\kappa_{rr}+\kappa_{\theta\theta})-\frac{1-\nu_s}{\pi R^3}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\frac{\frac{\eta}{R}F_{\text{radial}}\left(\frac{r}{R},\frac{\eta}{R},\varphi-\theta\right)}{\left[1-2\frac{\eta r}{R^2}\cos(\varphi-\theta)+\frac{\eta^2 r^2}{R^4}\right]^3}dA\right\} \tag{5.11}$$

$$\tau_\theta=\frac{E_s h_s^2}{6(1-\nu_s^2)}\left\{\frac{1}{r}\frac{\partial}{\partial\theta}(\kappa_{rr}+\kappa_{\theta\theta})-\frac{1-\nu_s}{\pi R^3}\int\int_A(\kappa_{rr}+\kappa_{\theta\theta})\frac{\frac{\eta}{R}F_{\text{circumferential}}\left(\frac{r}{R},\frac{\eta}{R},\varphi-\theta\right)}{\left[1-2\frac{\eta r}{R^2}\cos(\varphi-\theta)+\frac{\eta^2 r^2}{R^4}\right]^3}dA\right\} \tag{5.12}$$

where

$$\begin{aligned}F_{\text{radial}}(r_1,\eta_1,\varphi_1)&=(1+3r_1^2\eta_1^2)\cos\varphi_1-r_1\eta_1(3+r_1^2\eta_1^2\cos 2\varphi_1)\\F_{\text{circumferential}}(r_1,\eta_1,\varphi_1)&=(1-3r_1^2\eta_1^2)\sin\varphi_1+r_1^3\eta_1^3\sin 2\varphi_1\end{aligned} \tag{5.13}$$

Finally it should be noted that Eq. (5.4) also reduces to Stoney's result for the case of spatial curvature uniformity. Indeed for this case, Eq. (5.4) reduces to:

$$\sigma_{rr}+\sigma_{\theta\theta}=\frac{E_s h_s^2}{6(1-\nu_s)h_f}(\kappa_{rr}+\kappa_{\theta\theta}) \tag{5.14}$$

If in addition the curvature state is equibiaxial ($\kappa_{rr}=\kappa_{\theta\theta}$), as assumed by Stoney, Eq. (1.1) is recovered while relation (5.2) furnishes $\sigma_{rr}=\sigma_{\theta\theta}$ (stress equibiaxiality) as a special case.

6 Discussion and Conclusions

Unlike Stoney's original analysis and its extensions discussed in Sec. 1, the present analysis, together with Huang and Rosakis [13] and Huang et al. [14] for the special case of axisymmetry, show that the dependence of film stresses on substrate curvatures is not generally "local." Here the stress components at a point on the film will, in general, depend on both the local value of the curvature components (at the same point) and on the value of curvatures of all other points on the plate system (nonlocal dependence). The more pronounced the curvature nonuniformities are, the more important such nonlocal effects become in accurately determining film stresses from curvature measurements. This demonstrates that analyses methods based on Stoney's approach and its various extensions cannot handle the nonlocality of the stress/curvature dependence and may result in substantial stress prediction errors if such analyses are applied locally in cases where spatial variations of system curvatures and stresses are present.

The presence of nonlocal contributions in such relations also has implications regarding the nature of diagnostic methods needed to perform wafer-level film stress measurements. Notably, the existence of nonlocal terms necessitates the use of full-field methods capable of measuring curvature components over the entire surface of the plate system (or wafer). Furthermore, measurement of all independent components of the curvature field is necessary. This is because the stress state at a point depends on curvature contributions (from κ_{rr}, $\kappa_{\theta\theta}$, and $\kappa_{r\theta}$) from the entire plate surface.

Regarding the curvature-temperature (Eqs. (4.2)–(4.4)) and stress-temperature (Eqs. (4.5)–(4.7)) relations, the following points are noteworthy. These relations also generally feature a dependence of local temperature $T(r,\theta)$ which is "Stoney-like" as well as a "nonlocal" contribution from the temperature of other points on the plate system. Furthermore, the stress and curvature states are always nonequibiaxial (i.e., $\sigma_{rr}^{(f)} \neq \sigma_{\theta\theta}^{(f)}$ and $\kappa_{rr} \neq \kappa_{\theta\theta}$) in the presence of temperature nonuniformities. Only if T=constant these states become equibiaxial, the "nonlocal" contributions vanish, and Stoney's original results are recovered as a special case.

Finally, it should be noted that the existence of nonuniformities also results in the establishment of shear stresses along the film/substrate interface. These stresses are in general related to the derivatives of the first curvature invariant $\kappa_{rr}+\kappa_{\theta\theta}$ (Eqs. (5.11) and (5.12)). In terms of temperature, these interfacial shear stresses are also related to the gradients of the temperature distribution $T(r,\theta)$. The occurrence of such stresses is ultimately related to spatial nonuniformities, and as a result, such stresses vanish for the special case of uniform $\kappa_{rr}+\kappa_{\theta\theta}$ or T considered by Stoney and its various extensions. Since film delamination is a commonly encountered form of failure during wafer manufacturing, the ability to estimate the level and distribution of such stresses from wafer-level metrology might prove to be invaluable in enhancing the reliability of such systems.

References

[1] *The National Technology Road Map for Semiconductor Technology*, 2003, Semiconductor Industry Association, San Jose, CA.

[2] Freund, L. B., and Suresh, S., 2004, *Thin Film Materials: Stress, Defect Formation and Surface Evolution*, Cambridge University Press, Cambridge, U.K.

[3] Stoney, G. G., 1909, "The Tension of Metallic Films Deposited by Electrolysis," Proc. R. Soc. London, **82**, pp. 172–175.

[4] Wikstrom, A., Gudmundson, P., and Suresh, S., 1999, "Thermoelastic Analysis of Periodic Thin Lines Deposited on a Substrate," J. Mech. Phys. Solids, **47**, pp. 1113–1130.

[5] Shen, Y. L., Suresh, S., and Blech, I. A., 1996, "Stresses, Curvatures, and Shape Changes Arising from Patterned Lines on Silicon Wafers," J. Appl. Phys., **80**, pp. 1388–1398.

[6] Wikstrom, A., Gudmundson, P., and Suresh, S., 1999, "Analysis of Average Thermal Stresses in Passivated Metal Interconnects," J. Appl. Phys., **86**, pp. 6088–6095.

[7] Park, T. S., and Suresh, S., 2000, "Effects of Line and Passivation Geometry on Curvature Evolution During Processing and Thermal Cycling in Copper Interconnect Lines," Acta Mater., **48**, pp. 3169–3175.

[8] Masters, C. B., and Salamon, N. J., 1993, "Geometrically Nonlinear Stress-Deflection Relations for Thin Film/Substrate Systems," Int. J. Eng. Sci., **31**, pp. 915–925.

[9] Salamon, N. J., and Masters, C. B., 1995, "Bifurcation in Isotropic Thin Film/Substrate Plates," Int. J. Solids Struct., **32**, pp. 473–481.

[10] Finot, M., Blech, I. A., Suresh, S., and Fijimoto, H., 1997, "Large Deformation and Geometric Instability of Substrates With Thin-Film Deposits," J. Appl. Phys., **81**, pp. 3457–3464.

[11] Freund, L. B., 2000, "Substrate Curvature Due to Thin Film Mismatch Strain in the Nonlinear Deformation Range," J. Mech. Phys. Solids, **48**, pp. 1159–1174.

[12] Lee, H., Rosakis, A. J., and Freund, L. B., 2001, "Full Field Optical Measurement of Curvatures in Ultra-Thin Film/Substrate Systems in the Range of Geometrically Nonlinear Deformations," J. Appl. Phys., **89**, pp. 6116–6129.

[13] Huang, Y., and Rosakis, A. J., 2005, "Extension of Stoney's Formula to Non-Uniform Temperature Distributions in Thin Film/Substrate Systems. The Case of Radial Symmetry," J. Mech. Phys. Solids, **53**, pp. 2483–2500.

[14] Huang, Y., Ngo, D., and Rosakis, A. J., 2005, "Non-Uniform, Axisymmetric Misfit Strain in Thin Films Bonded on Plate Substrates/Substrate Systems: The Relation Between Non-uniform Film Stresses and System Curvatures," Acta Mech. Sin., **21**, pp. 362–370.

[15] Brown, M. A., Park, T.-S., Rosakis, A. J., Ustundag, E., Huang, Y., Tamura, N., and Valek, B., 2006, "A Comparison of X-Ray Microdiffraction and Coherent Gradient Sensing in Measuring Discontinuous Curvatures in Thin Film-Substrate Systems," ASME J. Appl. Mech., **73**, pp. 723–729.

Optimal Shape of a Rotating Rod With Unsymmetrical Boundary Conditions

Teodor M. Atanackovic
Professor
Faculty of Technical Sciences,
University of Novi Sad,
Trg D. Obradovica 6,
21000 Novi Sad, Serbia

Governing equations of a compressed rotating rod with clamped–elastically clamped (hinged with a torsional spring) boundary conditions is derived. It is shown that the multiplicity of an eigenvalue of this system can be at most equal to two. The optimality conditions, via Pontryagin's maximum principle, are derived in the case of bimodal optimization. When these conditions are used the problem of determining the optimal cross-sectional area function is reduced to the solution of a nonlinear boundary value problem. The problem treated here generalizes our earlier results presented in Atanackovic, 1997, Stability Theory of Elastic Rods, World Scientific, River Edge, NJ. The optimal shape of a rod is determined by numerical integration for several values of parameters. [DOI: 10.1115/1.2744041]

1 Introduction

Consider an elastic rod BC of length L, clamped at end B and simply supported at the end C. The change of angle of the rod axis at C is constrained by a torsional spring with a spring constant c (see Fig. 1). Suppose that the rod has circular cross section, that its axis is straight, and that it rotates with the constant angular velocity ω about its axis. Let $x-B-y$ be the rectangular Cartesian coordinate system with the axis x oriented along the rod axis in the undeformed state. At the end C the rod is loaded by a concentrated force $\mathbf{F}$ having constant intensity F and the action line parallel to the x axis. Let Π be a plane defined by the system $x-B-y$ that rotates with the angular velocity ω about x axis. At a certain velocity the rod loses stability (buckles) so that it is bent under the action of centrifugal forces and force $\mathbf{F}$. If the rod is bent it will assume a relative (with respect to the rotating plane Π), equilibrium configuration (see Fig. 1). The problem of determining the rotation speed ω and the force F at which the rod axis changes from a straight to bent shape is indeed an old one and its solution has importance in mechanical engineering. When the cross-sectional area $A(S)$ is known and $F=0$ it has been treated in Refs. [1–3]. The case $\omega\neq$, $F\neq 0$ was analyzed in Refs. [4–6].

In Ref. [7] we formulated and solved the problem of determining optimal shape (the function $A(S)$ for the rotating rod stable against buckling and having the smallest mass) when the end C is free and when with $F=0$. In Ref. [8] the clamped-free boundary conditions for the case $\omega\neq$, $F\neq 0$ was treated. Finally in Ref. [12] the case $\omega\neq$, $F\neq 0$ with both ends clamped was analyzed. The case of both ends clamped significantly changed the procedure, based on Pontryagin's principle, which was used in Refs. [7,8] since in that case we had to deal with bimodal optimization. In formulating the optimality conditions in Ref. [8] we used, implicitly, the symmetry with respect to the point $S=L/2$ that the problem has when both ends are clamped.

Our intention in this paper is to generalize the procedure of Refs. [10,7,12] to include the rod shown in Fig. 1. Thus, we shall derive new optimality condition for rotating the compressed rod via Pontryagin's principle. Our conditions will be derived without any assumptions about symmetry of the solutions (such as $A(S)=A(L-S)$). Then we shall solve the corresponding system of equations and determine the optimal shape of the rod rotating with given angular velocity ω and being compressed at the end C with the force F. As a special case we shall recover the optimality condition of Seyranian [13] (see also Refs. [15,16]) corresponding to the correct solution of the problem of Keller [14].

The major characteristics of our approach is the use of Pontryagin's principle in determining optimality conditions. It is known that in the case of bimodal optimization (as is the case here) the eigenvalue of the spectral problem describing stability boundary is not a differentiable functional (usually taken in the form of Rayleigh's quotient) of the buckling modes. Thus, in principle, the standard version of Pontryagin's principle is not applicable to the problems where eigenvalues are minimized (maximized). In order to avoid this problem, we will take another approach. Namely, we shall assume that the eigenvalue (in our case eigenvalue pair) is fixed and *given*. We shall minimize the volume of the column. By writing the cross-sectional area function in a suitable form we will have a functional (see Eq. (30)) that is differentiable with respect to a certain set of parameters [17].

2 Mathematical Model

We describe the rod axis, i.e., the line (a plane curve in Π) joining centroids of circular cross sections, by the functions $x(S)$ and $y(S)$. The angle between the tangent to the rod axis and Bx axis is denoted by $\theta(S)$. Thus we obtain the nonlinear differential equations of the rod in the following form

$$\frac{dH}{dS}=0, \quad \frac{dV}{dS}=-\rho A\omega^2 y, \quad \frac{dM}{dS}=-V\cos\theta+H\sin\theta$$

$$\frac{d\theta}{dS}=\frac{M}{EI}, \quad \frac{dx}{dS}=\cos\theta, \quad \frac{dy}{dS}=\sin\theta \tag{1}$$

where H, V are components of the contact force at an arbitrary cross section S (a resultant force representing the influence of the part $[0,S)$ of the rod on the part $[S,L]$) along the x and y axis, respectively; M stands for the contact couple; and where ρ denotes the mass density of the rod. The boundary conditions corresponding to Eq. (1) read

$$H(0)=H(L)=-F, \quad x(0)=0, \quad y(0)=y(L)=0$$

$$\theta(0)=0, \quad M(L)+c\theta(L)=0 \tag{2}$$

The volume of the rod is

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received December 2, 2005; final manuscript received February 23, 2007. Review conducted by Oliver M. O'Reilly.

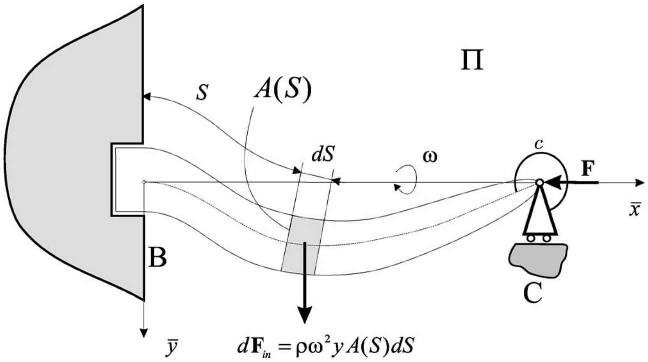


Fig. 1 Coordinate system and loading configuration

$$W = \int_0^L A(S)dS \tag{3}$$

We assume that the cross section of the rod is circular, so that

$$I = \alpha A^2 \tag{4}$$

where $\alpha = (1/4\pi)$ is a constant. By using the dimensionless variables and parameters

$$t = \frac{S}{L}, \quad a = \frac{A}{L^2}, \quad y = \frac{\bar{y}}{L}, \quad w = \frac{W}{L^3}, \quad k = \frac{c}{E\alpha L^3}$$

$$m = \frac{M}{E\alpha L^3}, \quad v = \frac{V}{E\alpha L^2}, \quad \lambda_1 = \frac{\rho\omega^2 L^2}{E\alpha}, \quad \lambda_2 = \frac{F_0}{E\alpha L^2} \tag{5}$$

the relevant differential equations and the boundary conditions, after linearization, become

$$\dot{v} = -\lambda_1 a \eta, \quad \dot{m} = -v - \lambda_2 \theta, \quad \dot{\eta} = \theta, \quad \dot{\theta} = \frac{m}{a^2} \tag{6}$$

and

$$\eta(0) = \eta(1) = 0, \quad \theta(0) = 0 \quad m(1) + k\theta(1) = 0 \tag{7}$$

where the dot over the variable represents the derivative with respect to dimensionless arc length t. The dimensionless volume is

$$w = \int_0^1 a(t)dt \tag{8}$$

Note that for the case of the nonrotating rod, simply supported at C, i.e., $\lambda_1 = 0$, $k = 0$ the system Eq. (6) reduces to the case treated in Ref. [14] and when $\lambda_1 = 0$, $k \neq 0$ to the system treated in Ref. [16].

In the optimization procedure we need multiplicity of the eigenvalue pair (λ_1, λ_2) in Eqs. (6) and (7) for a given $a(t)$. We prove that the maximal multiplicity is equal to two. Thus suppose that for given $a(t)$ and (λ_1, λ_2) there are three linearly independent solutions ν_i, m_i, η_i, θ_i, $i = 1, 2, 3$ of Eqs. (6) and (7). Then

$$\mathcal{N} = c_1\nu_1 + c_2\nu_2 + c_3\nu_3, \quad \mathcal{M} = c_1 m_1 + c_2 m_2 + c_3 m_3$$

$$\mathcal{Y} = c_1\eta_1 + c_2\eta_2 + c_3\eta_3, \quad \mathcal{T} = c_1\theta_1 + c_2\theta_2 + c_3\theta_3 \tag{9}$$

for arbitrary c_i, $i = 1, 2, 3$ satisfies Eqs. (6) and (7), i.e.

$$\dot{\mathcal{N}} = -\lambda_1 a \mathcal{Y}, \quad \dot{\mathcal{M}} = -\mathcal{N} - \lambda_2 \mathcal{T}, \quad \dot{\mathcal{Y}} = \mathcal{T}, \quad \dot{\mathcal{T}} = \frac{\mathcal{M}}{a^2} \tag{10}$$

Note that from Eq. $(7)_{1,3}$

$$\mathcal{Y}(0) = \mathcal{T}(0) = 0 \tag{11}$$

Since $\mathcal{N}(0) = c_1\nu_1(0) + c_2\nu_2(0) + c_3\nu_3(0)$, $\mathcal{M}(0) = c_1 m_1(0) + c_2 m_2(0) + c_3 m_3(0)$ we can always choose c_i so that

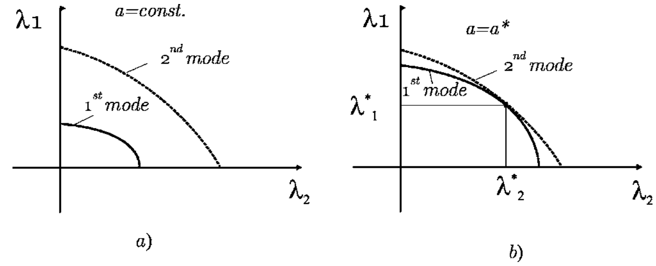


Fig. 2 Interaction curves for: (*a*) constant cross section; and (*b*) optimal cross section

$$\mathcal{N}(0) = \mathcal{M}(0) = 0 \tag{12}$$

System (10) with the initial conditions Eqs. (11) and (12) has the unique solution $\mathcal{N} = \mathcal{M} = \mathcal{Y} = \mathcal{T} = 0$. Thus ν_i, m_i, η_i, θ_i, $i = 1, 2, 3$ are linearly dependent.

If $a(t)$ is known the values of $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ for which Eqs. (6) and (7) have a nontrivial solution, define a set of curves $\mathcal{C}_n$, $n = 1, 2, \ldots$, called the interaction curves. The interaction curves for first two modes, for the case of a rod with constant cross section, are shown in Fig. 2(a). For the case of a variable cross section the interaction curves may approach each other and intersect. This could happen for the optimal cross section and we show the case when the interaction curves corresponding to the first and second mode touch each other in Fig. 2(b). In this case there are two buckling modes $\eta_1(t)$ and $\eta_2(t)$ corresponding to $(\lambda_1, \lambda_2) = (\lambda_1^*, \lambda_2^*)$.

In the optimization procedure that follows we shall allow for the situation shown in Fig. 2(b).

3 The Optimization Problem

The problem of determining the optimal shape of the rod, may be stated as: *Given* $(\lambda_1^*, \lambda_2^*)$ *find* $a^*(t)$ *such that* $(\lambda_1^*, \lambda_2^*)$ *belongs to the lowest interaction curve defined by the system Eqs.* (6) and (7) *and, at the same time, the volume of the rod* $w^* = \int_0^1 a^*(t)dt$ *is minimal.*

Mathematically, this problem may be stated as: given $(\lambda_1^*, \lambda_2^*)$ find $a(t)$ such that

$$w = \int_0^1 a(t)dt \tag{13}$$

is minimized subject to the constraints (6) and (7). The cross-sectional area $a(t)$ will be treated as the control variable.[1]

We use Pontryagin's principle to determine $a^*(t)$ as shown in Refs. [9,11]. Let $x_1 = y$, $x_2 = \nu$, $x_3 = \theta$, $x_4 = m$. Then the system Eqs. (6) and (7) becomes

$$\dot{x}_1 = x_3, \quad \dot{x}_2 = -\lambda_1 a x_1, \quad \dot{x}_3 = \frac{x_4}{a^2}, \quad \dot{x}_4 = -x_2 - \lambda_2 x_3 \tag{14}$$

and

$$x_1(0) = 0, \quad x_3(0) = 0, \quad x_1(1) = 0, \quad x_4(1) + kx_3(1) = 0 \tag{15}$$

Suppose that for given $a(t) = \tilde{a}(t)$ and given (λ_1, λ_2) there are two (since the maximal multiplicity of an eigenvalue pair is equal to two) linearly independent solutions of the system Eqs. (14) and (15) that we denote by $(\bar{x}_1, \ldots, \bar{x}_4)$ and $(\hat{x}_1, \ldots, \hat{x}_4)$. Thus

$$\dot{\bar{x}}_1 = \bar{x}_3, \quad \dot{\hat{x}}_1 = \hat{x}_3$$

$$\dot{\bar{x}}_2 = -\lambda_1 \tilde{a} \bar{x}_1, \quad \dot{\hat{x}}_2 = -\lambda_1 \tilde{a} \hat{x}$$

[1] Since the pair $(\lambda_1^*, \lambda_2^*)$ is given we shall not differentiate it with respect to $a(t)$.

$$\dot{\bar{x}}_3=\frac{\bar{x}_4}{\tilde{a}^2},\quad \dot{\hat{x}}_3=\frac{\hat{x}_4}{\tilde{a}^2}$$

$$\dot{\bar{x}}_4=-\bar{x}_2-\lambda_1\bar{x}_3,\quad \dot{\hat{x}}_4=-\hat{x}_2-\lambda_1\hat{x}_3 \tag{16}$$

satisfying

$$\bar{x}_1(0)=0,\quad \hat{x}_1(0)=0$$

$$\bar{x}_1(1)=0,\quad \hat{x}_1(1)=0$$

$$\bar{x}_3(0)=0,\quad \hat{x}_3(0)=0$$

$$\bar{x}_4(1)+k\bar{x}_3(1)=0,\quad \hat{x}_4(1)+k\hat{x}_3(1)=0 \tag{17}$$

To determine the minimum of Eq. (13) subject to Eqs. (16) and (17), we form Pontryagin's function $\mathcal{H}$, as (see Ref. [9])

$$\mathcal{H}=\tilde{a}+\bar{p}_1\bar{x}_3-\lambda_1\bar{p}_2\tilde{a}\bar{x}_1+\bar{p}_3\frac{\bar{x}_4}{\tilde{a}^2}-\bar{p}_4(\bar{x}_2+\lambda_2\bar{x}_3)+\hat{p}_1\hat{x}_3-\lambda_1\hat{p}_2\tilde{a}\hat{x}_1$$
$$+\hat{p}_3\frac{\hat{x}_4}{\tilde{a}^2}-\hat{p}_4(\hat{x}_2+\lambda_2\hat{x}_3) \tag{18}$$

The co-state variables $\bar{p}_i,\hat{p}_i,i=1,\ldots,4$ satisfy

$$\dot{\bar{p}}_1=-\frac{\partial H}{\partial\bar{x}_1}=\lambda_1\bar{p}_2\tilde{a},\quad \dot{\hat{p}}_1=-\frac{\partial H}{\partial\hat{x}_1}=\lambda_1\hat{x}\tilde{a}$$

$$\dot{\bar{p}}_2=-\frac{\partial H}{\partial\bar{x}_2}=\bar{p}_4,\quad \dot{\hat{p}}_2=-\frac{\partial H}{\partial\hat{x}_2}=\hat{p}_4$$

$$\dot{\bar{p}}_3=-\frac{\partial H}{\partial\bar{x}_3}=-\bar{p}_1+\lambda_2\bar{p}_4,\quad \dot{\hat{p}}_3=-\frac{\partial H}{\partial\hat{x}_3}=-\hat{p}_1+\lambda_2\hat{p}_4$$

$$\dot{\bar{p}}_4=-\frac{\partial H}{\partial\bar{x}_4}=-\frac{\bar{p}_3}{\tilde{a}^2},\quad \dot{\hat{p}}_4=-\frac{\partial H}{\partial\hat{x}_4}=-\frac{\hat{p}_3}{\tilde{a}^2} \tag{19}$$

subject to

$$\bar{p}_2(1)=0,\quad \hat{p}_2(1)=0$$

$$\bar{p}_2(0)=0,\quad \hat{p}_2(0)=0$$

$$\bar{p}_4(0)=0,\quad \hat{p}_4(0)=0$$

$$\bar{p}_3(1)-k\bar{p}_4(0)=0,\quad \hat{p}_3(1)-k\hat{p}_4(1)=0 \tag{20}$$

The optimality condition, $\mathcal{H}_{\min\, a\in U}$, leads to (see Refs. [18–20])

$$\frac{\partial\mathcal{H}}{\partial a}=1-2\bar{p}_3\frac{\bar{x}_4}{\tilde{a}^3}-2\hat{p}_3\frac{\hat{x}_4}{\tilde{a}^3}-\lambda_1(\bar{p}_2\bar{x}_1+{}_1\hat{p}_2\tilde{a}\hat{x}_1)=0 \tag{21}$$

so that

$$\tilde{a}^3=2\frac{\bar{p}_3\bar{x}_4+\hat{p}_3\hat{x}_4}{1+\lambda_1(\bar{p}_2\bar{x}_1+\hat{p}_2\hat{x}_1)} \tag{22}$$

In our earlier works we used the special identification of state $\bar{x}_i,\hat{x}_i,i=1,\ldots,4$ and co-state $\bar{p}_i,\hat{p}_i,i=1,\ldots,4$ variables. Namely it is easy to see that solutions to Eqs. (16) and (17) and Eqs. (19) and (20) are related as

$$\bar{p}_1=\bar{x}_2,\quad \bar{p}_2=-\bar{x}_1,\quad \bar{p}_3=\bar{x}_4,\quad \bar{p}_4=-\bar{x}_3$$

$$\hat{p}_1=\hat{x}_2,\quad \hat{p}_2=-\hat{x}_1,\quad \hat{p}_3=\hat{x}_4,\quad \hat{p}_4=-\hat{x}_3 \tag{23}$$

The identification of state and co-state variables Eq. (23) is of central importance. In principle we may take $\bar{p}_i,\hat{p}_i,i=1,\ldots 4$ as linear combination of $\bar{x}_i,\hat{x}_i,i=1,\ldots 4$. From Eq. (23) we conclude that

$$\bar{p}_1=\beta_{11}\bar{x}_2+\beta_{12}\hat{x}_2,\quad \bar{p}_2=-\beta_{11}\bar{x}_1-\beta_{12}\hat{x}_1$$

$$\bar{p}_3=\beta_{11}\bar{x}_4+\beta_{12}\hat{x}_4,\quad \bar{p}_4=-\beta_{11}\bar{x}_3-\beta_{12}\hat{x}_3$$

$$\hat{p}_1=\beta_{21}\bar{x}_2+\beta_{22}\hat{x}_2,\quad \hat{p}_2=-\beta_{21}\bar{x}_1-\beta_{22}\hat{x}_1$$

$$\hat{p}_3=\beta_{21}\bar{x}_4+\beta_{22}\hat{x}_4,\quad \hat{p}_4=-\beta_{21}\bar{x}_4-\beta_{22}\hat{x}_4 \tag{24}$$

where $\beta_{ij},i,j=1,2$ are any real constants, satisfying Eqs. (19) and (20). With Eq. (24) the optimality condition Eq. (22) becomes

$$\tilde{a}^3=2\frac{\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2}{1+\lambda_1[\beta_{11}(\bar{x}_1)^2+2\gamma_{12}\bar{x}_1\hat{x}_1+\beta_{22}(\hat{x}_1)^2]} \tag{25}$$

where $\gamma_{12}=(\beta_{12}+\beta_{21})/2$. The optimality condition Eq. (25) generalizes the earlier results. For example in the case when the rod is not rotating, i.e., $\lambda_1=0$, Eq. (25) reduces to (the constant 2 is included in β_{11}, β_{22}, and γ_{12})

$$\tilde{a}^3=\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2 \tag{26}$$

earlier obtained by Seyranian [13,16]. Since $\tilde{a}(t)\geqslant 0$ the coefficients β_{11}, γ_{12}, and β_{22} must satisfy the positive definite condition, that is

$$\beta_{11}\beta_{22}\geqslant\tfrac{1}{4}(\gamma_{11})^2 \tag{27}$$

By using Eq. (25) in Eq. (16) we obtain

$$\dot{\bar{x}}_1=\bar{x}_3$$

$$\dot{\bar{x}}_2=-\lambda_1\bar{x}_1\left[2\frac{\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2}{1+\lambda_1[\beta_{11}(\bar{x}_1)^2+2\gamma_{12}\bar{x}_1\hat{x}_1+\beta_{22}(\hat{x}_1)^2]}\right]^{1/3}$$

$$\dot{\bar{x}}_3=\frac{\bar{x}_4}{\left[2\frac{\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2}{1+\lambda_1[\beta_{11}(\bar{x}_1)^2+2\gamma_{12}\bar{x}_1\hat{x}_1+\beta_{22}(\hat{x}_1)^2]}\right]^{2/3}}$$

$$\dot{\bar{x}}_4=-\bar{x}_2-\lambda_2\bar{x}_3$$

$$\dot{\hat{x}}_1=\hat{x}_3$$

$$\dot{\hat{x}}_2=-\lambda_1\hat{x}_1\left[2\frac{\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2}{1+\lambda_1[\beta_{11}(\bar{x}_1)^2+2\gamma_{12}\bar{x}_1\hat{x}_1+\beta_{22}(\hat{x}_1)^2]}\right]^{1/3}$$

$$\dot{\hat{x}}_3=\frac{\hat{x}_4}{\left[\frac{\beta_{11}(\bar{x}_4)^2+2\gamma_{12}\bar{x}_4\hat{x}_4+\beta_{22}(\hat{x}_4)^2}{1+\lambda_1[\beta_{11}(\bar{x}_1)^2+2\gamma_{12}\bar{x}_1\hat{x}_1+\beta_{22}(\hat{x}_1)^2]}\right]^{2/3}}$$

$$\dot{\hat{x}}_4=-\hat{x}_2-\lambda_2\hat{x}_3 \tag{28}$$

subject to Eq. (17). Note that with Eqs. (24) and (25) the Pontryagin's function $\mathcal{H}$ is not explicitly dependant on t so that it is a first integral of the system Eq. (27).

We how comment on the constants β_{11}, β_{22}, and γ_{12}. They are subject to Eq. (27), otherwise they are arbitrary. Without loss of generality we may take $\beta_{11}=\beta_{22}=1$ (just redefine $\bar{x}_i,\hat{x}_i,i=1,\ldots,4$ as $\bar{X}_i=\bar{x}_i\sqrt{\beta_{11}}$, $\hat{X}_i=\hat{x}_i\sqrt{\beta_{22}}$). Then, Eq. (27) becomes (with new constant γ_{12} defined as $\gamma_{12}\sqrt{\beta_{11}\beta_{22}}$)

$$1\geqslant(\gamma_{12})^2 \tag{29}$$

To fix γ_{12} we use Eq. (26) in Eq. (13) and to minimize the resulting expression with respect to γ_{12}, i.e.

Table 1 Optimal volume as a function of the spring constant

k	1	0.8	0.6	0.4	0.2	0.1
$w^*(k)$	0.926928	0.938855	0.951899	0.966237	0.982131	0.990798

$$w^* = \min_{\gamma_{12}} \int_0^1 \tilde{a}(t,\gamma_{12})dt$$

$$= \min_{\gamma_{12}} \int_0^1 \left\{ 2\frac{(\bar{x}_4)^2 + 2\gamma_{12}\bar{x}_4\hat{x}_4 + (\hat{x}_4)^2}{1 + \lambda_1[(\bar{x}_1)^2 + 2\gamma_{12}\bar{x}_1\hat{x}_1 + (\hat{x}_1)^2]} \right\}^{1/3} dt \quad (30)$$

From Eq. (30) we obtain

$$\frac{\partial w^*}{\partial \gamma_{12}} = \int_0^1 \frac{\partial \tilde{a}(t,\gamma_{12})}{\partial \gamma_{12}} dt = 0 \quad (31)$$

or by using Eq. (26) to evaluate $\partial\tilde{a}(t,\gamma_{12})/\partial\gamma_{12}$, we obtain

$$\int_0^1 \frac{2}{3\tilde{a}^2} \left\{ \frac{2\bar{x}_4\hat{x}_4}{1 + \lambda_1[(\bar{x}_1)^2 + 2\gamma_{12}\bar{x}_1\hat{x}_1 + (\hat{x}_1)^2]} - \frac{2\lambda_1\bar{x}_1\hat{x}_1[(\bar{x}_4)^2 + 2\gamma_{12}\bar{x}_4\hat{x}_4 + (\hat{x}_4)^2]}{(1 + \lambda_1[(\bar{x}_1)^2 + 2\gamma_{12}\bar{x}_1\hat{x}_1 + (\hat{x}_1)^2])^2} \right\} dt = 0 \quad (32)$$

In the special case when $\lambda_1=0$, the condition (32) leads to

$$\int_0^1 \frac{\bar{x}_4\hat{x}_4}{\tilde{a}^2} dt = 0 \quad (33)$$

The orthogonality condition of the type Eq. (33) was used "for the sake of convenience" (see Eq. (4.1) in Refs. [16,21] for $c_2 \to 0$). Note also that Eqs. (29) and (32) are of the same form as conditions (66) and (67) of Ref. [21].

In the examples of the following section we shall show that, depending on the eigenfunctions, we may have $\gamma_{12}=0$ satisfying Eq. (32). Sometimes, however, condition (32) leads to $\gamma_{12} \neq 0$. It is always the case for the symmetric boundary conditions if $\bar{x}_i, i=1,\ldots,4$ is symmetric and $\hat{x}_i, i=1,\ldots,4$ is antisymmetric mode.

4 Numerical Results

(i) We start with the benchmark example of Ref. [14]. This corresponds to $\lambda_1=0$ in Eq. (28). For the case $k=0$ the second mode is not twice the continuously differentiable function (the displacement is continuous and the slope is discontinuous function). This is a consequence of the fact that the point $t=0.2895$ at which the cross-sectional area is zero, may be considered as an internal hinge. Thus we assume a value of k, and determine $\tilde{a}(t,k)$, $w^*(t,k)$. Then we let $k \to 0$. In Table 1, we show $w^*(t,k)$ for several values of k.

The buckling modes for the optimal cross section for $\lambda_2=27.22$, $k=0.1$ are shown in Fig. 3. When the value $\lambda_2=27.22$ (determined in Ref. [14]) is used we obtained the value of volume as $w^*=0.990798$. Thus, bimodal analysis confirms that the solution of Ref. [14] is optimal. This was confirmed in Refs. [22,16]. In numerical analysis of this example, in solving Eqs. (28) and (17), we get $\gamma_{12}=0$.

The optimal cross-sectional area is shown in Fig. 4. It is in agreement with the results of Ref. [16] where (in our notation) $k=4$ was used. The characteristic values of the cross-sectional area are: $a_{\min}=a(0.28973)=7.24\times 10^{-4}$, $a_{\max}=a(0.6428)=1.354089$; and $a(0)=1.32703$.

(ii) Next, we consider the case $\lambda_1=30$, $\lambda_2=37.5$, $k=4$. This time we used condition (32) to determine γ_{12}. The resulting value was $\gamma_{12}=0.21504$. First and second buckling modes are shown in Fig. 5, while the optimal cross-sectional area is shown in Fig. 6.

(iii) Finally we consider the case when the stiffness of the torsional spring is increasing. Thus we assume $\lambda_1=40$, $\lambda_2=40$, k

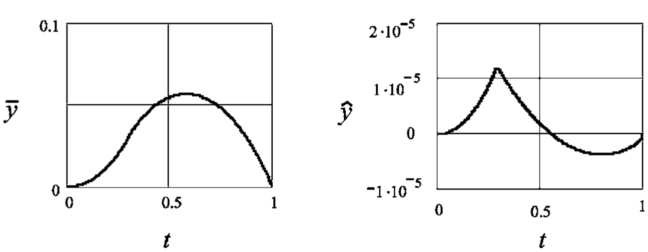


Fig. 3 Buckling modes corresponding to optimal cross-sectional area for $\lambda_1=0$, $\lambda_2=27.22$, $k=0.1$

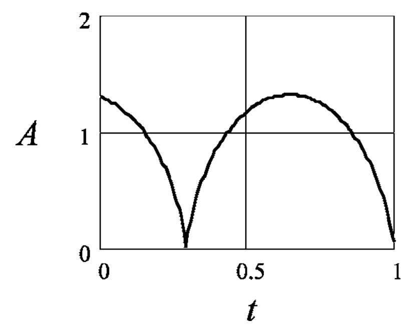


Fig. 4 Optimal cross-sectional area for $\lambda_1=0$, $\lambda_2=27.22$, $k=0.1$

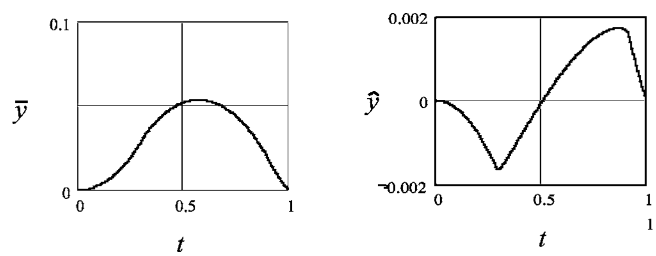


Fig. 5 Buckling modes corresponding to optimal cross-sectional area for $\lambda_1=30$, $\lambda_2=37.5$, $k=4$

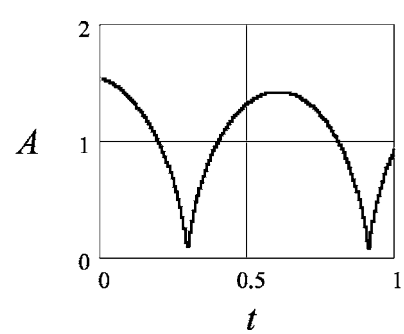


Fig. 6 Optimal cross-sectional area for $\lambda_1=30$, $\lambda_2=37.5$, $k=4$

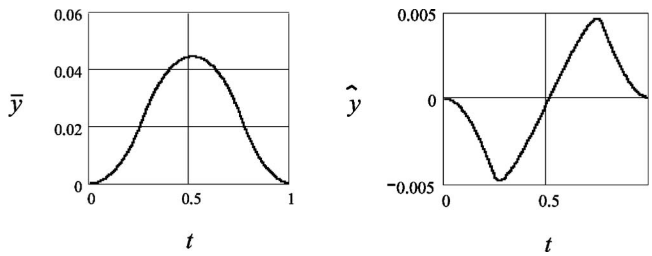


Fig. 7 Buckling modes corresponding to optimal cross-sectional area for $\lambda_1=40$, $\lambda_2=40$, $k=100$

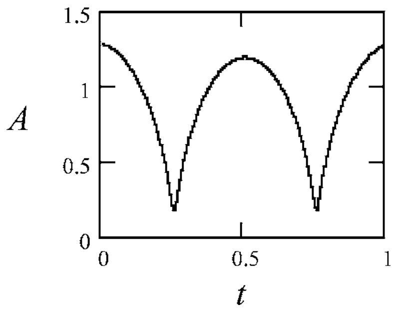


Fig. 8 Optimal cross-sectional area for $\lambda_1=40$, $\lambda_2=40$, $k=100$

$=100$. Again we force a solution of Eqs. (28) and (17) to satisfy Eq. (32) and obtained $\gamma_{12}=-0.235145$. The modes are shown in Fig. 7 and the cross-sectional area in Fig. 8.

The characteristic values of the cross–sectional area are: $A(0)=1.28163$, $A(1)=1.271813$, and $A_{\min}=A(0.75855)=0.17464$.

5 Conclusion

We studied the problem of determining the optimal shape against buckling of a rotating compressed rod with clamped–elastically clamped (hinged with torsional spring) boundary conditions. We formulate the optimality conditions by using Pontriyagin's maximum principle that includes the optimality condition of Refs. [13,16], as a special case when the rod is not rotating.

In our formulation the constant γ_{12} that contains the mixed term in the optimal cross section (see Eq. (26)) satisfies the inequality Eq. (29), and condition (32). The orthogonality condition of this type was proposed in Ref. [13]. In some examples the conditions (29) and (32) were satisfied, with $\gamma_{12}=0$.

In the example (i) we treated the case of Keller formulated in Ref. [14], i.e., $\lambda_1=0$. Our analysis shows that, although obtained by unimodal analysis, the solution presented in Ref. [14] is correct. Thus we confirm the conclusion earlier stated in Refs. [22,16]. In example (ii) we considered both parameters of equal importance and we imposed the orthogonality condition (32). Finally in example (iii) the torsional spring is considered to be stiff (large value of the constant k). The optimal shape tends to become symmetric $(a(t)=a(1-t))$ as $k\to\infty$, thus approaching the solution presented in Ref. [12] corresponding to a rod with clamped ends.

Acknowledgment

This research was supported by the Ministry of Science Grant No. 144019A.

References

[1] Stodola, A., 1906, *Steam Turbines*, D. Van Nostrand Company, New York.
[2] Bazely, N., and Zwahlen, B., 1968, "Remarks on the Bifurcation of Solutions of a Nonlinear Eigenvalue Problem," Arch. Ration. Mech. Anal., **28**, 51–58.
[3] Parter, S. V., 1970, "Nonlinear Eigenvalue Problems for Some Fourth Order Equations: I Maximal Solutions," SIAM J. Math. Anal., **1**, pp. 437–478.
[4] Atanackovic, T. M., 1986, "Buckling of Rotating Compressed Rods," Acta Mech., **60**, pp. 49–66.
[5] Atanackovic, T. M., 1997, "On the Rotating Rod With Variable Cross Section," Arch. Appl. Mech., **67**, pp. 447–456.
[6] Atanackovic, T. M., 1997, *Stability Theory of Elastic Rods*, World Scientific, River Edge, N.J.
[7] Atanackovic, T. M., 2001, "Optimal Shape of a Rotating Rod," J. Appl. Mech., **68**, pp. 860–864.
[8] Atanackovic, T. M., 2004, "On the Optimal Shape of a Compressed Rotating Rod," Meccanica, **39**, pp. 147–157.
[9] Spasic, D. T., and Atanackovic, T. M., 2004, "Bimodal Optimization of a Compressed Rotating Rod," Acta Mech., **173**, pp. 77–87.
[10] Leavy, R., 1990, "Buckling Optimization of Beams and Plates on Elastic Foundation," J. Eng. Mech., **116**, pp. 18–34.
[11] Seyranian, A. P., 1984, "On the Lagrange Problem," Mech. Solids, **19**, pp. 100–111.
[12] Seyranian, A. P., 1995, "New Solutions to Lagrange's Problem," Phys. Dokl., **40**, pp. 251–253.
[13] Seyranian, A. P., and Privalova, O. G., 2003, "The Lagrange Problem on Optimal Column: Old and New Results," Struct. Multidiscip. Optim., **25**, pp. 393–410.
[14] Tadjbakhsh, I., and Keller, J. B., 1962, "Strongest Columns and Isoperimetric Inequalities for Eigenvalues," J. Appl. Mech., **29**, pp. 159–164.
[15] Cox, S. J., and Overton, M. L., 1992, "On the Optimal Design of Columns Against Buckling," SIAM J. Math. Anal., **23**, pp. 287–325.
[16] Błachut, J., and Życzkowski, M., 1984, "Bimodal Optimal Design of Clamped-Clamped Columns Under Creep Conditions," Int. J. Solids Struct., **20**, pp. 571–577.
[17] Atanackovic, T. M., 2006, "Optimal Shape of Column With Own Weight: Bi and Single Modal Optimization," Meccanica, **41**, pp. 173–196.
[18] Vujanovic, B. D., and Atanackovic, T. M., 2004, *An Introduction to Modern Variational Techniques in Mechanics and Engineering*, Birkhäuser, Boston.
[19] Alekseev, V. M., Tihomirov, V. M., and Fomin, S. V., 1979, *Optimal Control*, Nauka, Moscow (in Russian).
[20] Sage, A. P., and White, C. C., 1977, *Optimum System Control*, Prentice–Hall, Englewood Cliffs, NJ.
[21] Seyranian, A. P., Lund, E., and Olhoff, N., 1994, "Multiple Eigenvalues in Structural Optimization Problems," Struct. Optim., **8**, pp. 207–227.
[22] Kirmser, P. G., and Hu, K.-K., 1993, "The Shape of the Ideal Column Reconsidered," Math. Intell., **15**, pp. 211–170.

The Bridged Crack Model for the Analysis of Brittle Matrix Fibrous Composites Under Repeated Bending Loading

Alberto Carpinteri[1]
e-mail: alberto.carpinteri@polito.it

Simone Puzzi

Department of Structural Engineering and Geotechnics,
Politecnico di Torino,
Corso Duca degli Abruzzi 24,
10129 Torino, Italy

In this paper, we present a fracture-mechanics based model, the so-called bridged crack model (Carpinteri, A., 1981, "A Fracture Mechanics Model for Reinforced Concrete Collapse," Proc. of IABSE Colloquium on Advanced Mechanics of Reinforced Concrete, Delft, I.A.B.S.E., Zürich, pp. 17–30; Carpinteri, A., 1984, "Stability of Fracturing Process in R.C. Beams," J. Struct. Engng. (A.S.C.E.), ***110****, pp. 544–558) for the analysis of brittle matrix composites with discontinuous ductile reinforcements under the condition of repeated bending loading. In particular, we address the case of composites with very high number of reinforcements (i.e., fiber-reinforced composites, rather than conventionally reinforced concrete). With this aim, we propose a new iterative procedure and compare it to the algorithm recently proposed by Carpinteri, Spagnoli, and Vantadori (2004, "A Fracture Mechanics Model for a Composite Beam with Multiple Reinforcements Under Cyclic Bending," Int. J. Solids Struct.,* ***41****, pp. 5499–5515), showing the advantages in terms of computational efficiency. Furthermore, we analyze the combined effects of crack length, brittleness number, and fiber number on the cyclic behavior of the composite beam, showing the conditions enhancing the energy dissipation in the composite system. Eventually, we analyze crack propagation and propose, consistently with the model premises, a fracture-mechanics-based crack propagation criterion that allows one to simulate cyclic bending tests under the fixed grip condition.*
[DOI: 10.1115/1.2744042]

Keywords: bridged crack, brittle matrix composite, repeated loading

Introduction

Independently of the matrix and fiber materials and in spite of different physical and mechanical properties, fibrous composites present a common feature, namely, the *bridging* or the reinforcing action exerted by the fibers spread into the matrix. This action affects the global structural response of the composite component mainly in the post-cracking phase, resulting in an increase of several mechanical properties: strength, stiffness, toughness, ductility, crack resistance, and fatigue strength. The cracking process is controlled by the reinforcements, which act across micro- and macrocracks. For this reason, the bridging mechanism is central also in all mechanical models for the analysis of the composite material response, especially if the matrix is brittle.

Two fracture-mechanics-based approaches have been extensively used in the last 20 years for modeling the constitutive behavior of fibrous composites: the bridged crack model [1,2] and the cohesive crack model. Both model types, in accordance with the ones proposed by Barenblatt [4] for the analysis of brittle heterogeneous materials and by Dugdale [5] for the analysis of ductile materials, replace the bridging zone by a fictitious crack and represent the bridging actions by a closing traction distribution $\sigma(w)$ (cohesive law), where w is the crack opening, or by a series of concentrated loads. In other words, the fracture process zone (FPZ) is substituted by a discrete crack and the localized closing tractions, either continuous or discontinuous, on the crack faces, represent the bridging mechanisms active in the FPZ.

The fundamental difference between these two model types is in the assumed stress field in the crack tip vicinity; for the bridged crack model, it is a singular stress field, whereas for the cohesive crack model, it is finite and limited by the tensile strength. As a result, also, the crack propagation conditions are different: In the case of the bridged crack model, crack propagates when the stress-intensity factor at the crack tip attains the critical value K_{IC}, which is a measure of the matrix toughness. On the other hand, the cohesive crack model assumes that crack propagation occurs when the stress-intensity factor is equal to zero (i.e., when the stress at the crack tip equals the composite strength). In the first case, two factors affect the global toughness of the composite: The first is the matrix toughness, represented by the critical value of the stress-intensity factor, which is assumed to be a material property, and the second is the reinforcing phase toughening mechanism, governed by the properties of the reinforcements and by their interaction with the matrix. In the cohesive crack, only a global toughening mechanism of the composite is defined; it is represented by the shielding effect due to the cohesive tractions. As a consequence, the matrix toughening, explicitly represented in the former case by the matrix toughness, is in this latter case merged with the toughening mechanism produced by the secondary phase through the cohesive law.

The cohesive crack model has been extensively used with the aim at describing concrete and fiber-reinforced cementitious composites (among others, see the papers by Hillerborg et al. [6], Petersson [7], Hillerborg [8], Carpinteri [9–11], Shah [12], Cotterell et al. [13], Li and Liang [14], Wecharatana and Shah [15], and Visalvanich and Naaman [16]). Nevertheless, the bridged crack model, which seems to be more suitable to represent the discontinuous nature of the reinforcing actions, has been used even more extensively, both in the case of cementitious and ceramic composites. Among models specifically developed for the case of con-

[1]Corresponding Author.
Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received January 16, 2006; final manuscript received March 7, 2007. Review conducted by Matthew R. Begley.

crete, we could mention those of Carpinteri [1,2], Ballarini et al. [17], Mai [18], Jenq and Shah [19], Foote et al. [20], Zhang and Li [21], and Ruiz [22]. Other papers dealt preferably with ceramic composites, in which reinforcements could be fibers, ductile particles or even grain of the matrix itself, as in the case of SiC-SiC composites. This class includes the models of Erdogan and Joseph [23], Mai [24], Cox [25], Marshall et al. [26], Marshall and Cox [27], Budiansky et al. [28], Cox and Marshall [29,30], and Ballarini and Muju [31].

Both the cohesive and the bridged crack models lead to a nonlinear problem because in general, the closing tractions depend on the crack face openings, and vice versa. Several different approaches are taken in the aforementioned models, in order to resolve the nonlinear problem. A first approach due to Barenblatt [4], utilizes the superposition of the stress-intensity factors, from which a singular, nonlinear, integral equation is defined. The equation is then resolved directly in particular cases or through iterative procedures. Several researchers also followed a simplified path, by considering the planar crack face hypothesis. In this case, the solution is very simple and the unknowns are reduced to only one, namely, the FPZ depth. It should be noted that this approach does not take into account the compatibility and leads to, solely, an equilibrium solution. The same problem has been evidenced in other approaches, which are not directly referred to the planar crack face hypothesis. On the other hand, the bridged crack model proposed by Carpinteri [1] and co-workers takes into account both equilibrium and compatibility. This latter model has been deeply investigated for the case of monotonic loading; the interested reader is referred to [1,2,32–38]. In particular, Carpinteri and Massabò [35] unified the bridged crack model and the cohesive one in a single formulation, demonstrating their practical equivalence. More recently, Carpinteri et al. [36] presented an improved version in which the reinforcements at two different length scales are represented. In fact, the actions exerted by the larger fibers (or bars) at the macroscale are represented through concentrated bridging loads, whereas those of the microscopic fibers at the microscale are modeled through a cohesive bridging law; this enhancement allows one to model composites with a nonlinear matrix.

Regarding cyclic loading and elastoplastic shakedown, this problem was originally addressed by Carpinteri and Carpinteri [39] and Carpinteri [40] in the case of a single reinforcement. Recently, Carpinteri and Puzzi [41,42] analyzed the case of two and any number of reinforcements. Similar developments were published by Carpinteri et al. [3], who also extended the model to beams with a "T" cross section and prestressed [43].

In this paper, we review the fundamental issues in repeated loading of reinforced beams and, with the aim at modeling composites with a very high number of reinforcements, we propose a new solution procedure. Furthermore, we will analyze the combined effects of crack length, brittleness number, and fiber number on the cyclic behavior of the composite beam; eventually, we propose a crack propagation criterion based on fracture mechanics, which allows one to simulate cyclic bending tests not only under dead load conditions, but also under fixed grip conditions.

Geometry and Constitutive Equations

Let us consider a composite beam subjected to bending with an edge crack of length a, whose faces are bridged by intact reinforcements (either fibers or bars), as shown in Fig. 1. The model focuses onto the cracked cross section and considers a portion of the beam of vanishing length Δl, centered on the crack, subjected to the bending moment M. Let b and h be the section thickness and height, respectively, N the total number of discrete reinforcing elements, and n the number of them acting across the crack wake. The normalized crack depth $\xi = a/h$ and the normalized coordinate $\zeta = z/h$ are defined, z being the coordinate related to the bottom of the cross section. The generic position of the ith reinforcement is described by the coordinate c_i, and its action is represented by an

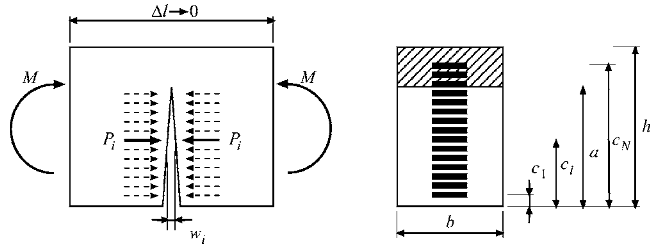


Fig. 1 Model geometry

indeterminate force P_i, while the crack opening in correspondence of its position is given by w_i. The solution of the above problem consists in the determination of the n unknown reinforcement actions $P_i, i=1,\ldots,n$ and of the n unknown crack openings.

Regarding materials, the beam matrix is considered to be elastic brittle, and smeared damage is not considered. The bridging law linking the closing action P_i of the ith fiber with the corresponding crack opening w_i is assumed to be rigid-perfectly plastic, so that it could represent both the fiber yielding or the matrix-fiber slippage (or pullout). In fact, the maximum bridging traction is defined by the ultimate force $P_{Pi} = A_i \sigma_y$, A_i being the single reinforcement cross-sectional area, and σ_y the minimum between the reinforcement yield strength and the slippage stress. This assumption is valid for all brittle matrix composites with ductile reinforcements, such as metal-toughened ceramics and the large majority of fiber-reinforced cementitious composites. On the other hand, it is not apt to describe the traction law of materials reinforced with brittle fibers.

To resolve the statically indeterminate problem of the reinforcement action, compatibility displacement conditions at the cracked cross section are introduced. The crack opening in correspondence to the ith fiber can be computed by superposition, by adding the contribution w_{iM} due to the external bending moment M to those due the n redundant reaction forces P_i (w_{ij}, i, $j=1,\ldots,n$),

$$w_i = w_{iM} + \sum_{j=1}^{n} w_{ij} = \lambda_{iM} M - \sum_{j=1}^{n} \lambda_{ij} P_j \tag{1}$$

where λ_{iM} and λ_{ij} are the local compliances, i.e., the crack opening displacements at the ith fiber level, due to a unit bending moment $M=1$ and a unit closing action $P_j=1$, respectively. The localized rotation also may be computed by superposition

$$\phi = \phi_M + \sum_{i=1}^{n} \phi_i = \lambda_{MM} M - \sum_{i=1}^{n} \lambda_{Mi} P_i \tag{2}$$

where λ_{Mi} is the local rotation due to a pair of unit opposite opening forces applied at c_i; according to Betti's theorem, it is equal to the compliance λ_{iM} that appears in Eq. (1). λ_{MM} represents the rotation due to the action of a unit bending moment M. The energy balance for the determination of the generic local compliance of a cracked element subjected to a generic loading condition can be found in the papers by Bosco and Carpinteri [32] and Carpinteri and Massabò [35]. The final expressions of the compliances are

$$\lambda_{iM} = \frac{2}{hbE} \int_{\zeta_i}^{\xi} Y_P(\xi_i, \xi) Y_M(\xi) \mathrm{d}\xi$$

$$\lambda_{ij} = \lambda_{ji} = \frac{2}{bE} \int_{\max[\zeta_i, \zeta_j]}^{\xi} Y_P(\xi_i, \xi) Y_P(\varsigma_j, \xi) \mathrm{d}\xi$$

$$\lambda_{MM}=\frac{2}{h^2bE}\int_0^{\xi}Y_M(\xi)^2\mathrm{d}\xi \tag{3}$$

where the shape functions Y_M and Y_{Pi} are given in different stress intensity factors (SIFs) handbooks [44,45].

The foregoing theoretical scheme allows us to describe the indeterminate problem of the n unknown forces in a matrix form. Let $\{w\}=\{w_1,\ldots,w_n\}^T$ be the vector of the crack face displacements (openings) in correspondence to the n fibers, with numbering starting from the bottom fiber. Correspondingly, the vector $\{P\}=\{P_1,\ldots,P_n\}^T$ of the indeterminate fiber actions, is defined. The local compliances due to the external bending moment M are collected into the vector $\{\lambda_M\}=\{\lambda_{1M},\ldots,\lambda_{nM}\}^T$, whereas those due to the bridging tractions P_i are collected in a matrix $[\lambda]$, of dimensions $(n\times n)$, the generic element ij of which is the local compliance λ_{ij}. Recalling the expression of the ith crack opening w_i, Eq. (1), it is possible to express the vector $\{w\}$ through the following equation:

$$\{w\}=\{\lambda_M\}M-[\lambda]\{P\} \tag{4}$$

Considering a beam loaded by the external bending moment M, before the onset of yielding (or slippage between the reinforcement and the matrix), the bridging actions keep the crack locally closed. The compatibility condition is therefore expressed by the following linear matrix system of n equations: $\{w\}=\{0\}$. The system solution leads to the computation of the vector $\{P\}$ of the unknown forces exerted by the reinforcements; they are functions of the applied bending moment M, as follows (it can be easily shown that the matrix $[\lambda]$ is symmetric and positively definite):

$$\{P\}=[\lambda]^{-1}\{\lambda_M\}M \tag{5}$$

After the plastic limit of the most loaded fiber (say, the ith) has been reached, the value of the ith opening w_i becomes greater than zero and a priori unknown. This means that the ith equation of the n relations (5) does not hold any longer. In this new phase, the statically indeterminate problem is solved by imposing a zero crack opening in correspondence of the $(n-1)$ reinforcements, which still are in their elastic regime, through Eq. (4). By doing so, a linear system is obtained, of rank $(n-1)$, since the bridging action P_i is already known and equal to its plastic limit: $P_i=P_{Pi}$. After the bridging tractions have been computed, it is a simple task to determine the amount of the crack opening w_i by introducing their values into the ith equation of the system (4).

A similar procedure can be followed to resolve the indeterminate problem at a subsequent stage, when two or more reinforcements are in the plastic (or slippage) stage. It can be easily observed that, for any given load, the number of compatibility conditions is equal to the number of static unknowns. In other words, when a fiber reaches its maximum attainable load, which corresponds to plastic flow or sliding, the true unknown becomes the corresponding opening displacement. As a result, the indeterminacy degree of the initially statically indeterminate problem is reduced by one.

To write this procedure mathematically in a general form, supposing that n_f fibers are in a condition of plastic flow, while the remaining n_c still behave elastically, the problem is partially statically indeterminate and partially determinate, the true unknowns being n_f displacements $\{w_f\}$ and n_c reactions $\{P_c\}$ (the subscripts f and c refer to free and constrained, respectively). Partitioning Eq. (4) in order to separate the static variables from the kinematic ones, the following system is obtained:

$$\begin{Bmatrix} w_f \\ w_c \end{Bmatrix}=\begin{Bmatrix} \lambda_{M,f} \\ \lambda_{M,c} \end{Bmatrix}M-\begin{bmatrix} \lambda_{ff} & \lambda_{fc} \\ \lambda_{cf} & \lambda_{cc} \end{bmatrix}\begin{Bmatrix} P_f \\ P_c \end{Bmatrix} \tag{6}$$

Considering that $\{w_c\}=\{0\}$, from the lower part of Eq. (6) the n_c statically indeterminate bridging actions can be determined

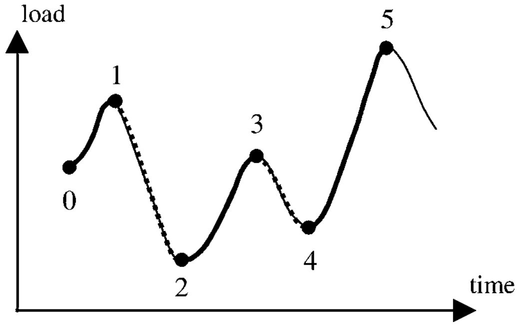


Fig. 2 Decomposition of the load history into monotonic parts

$$\{P_c\}=[\lambda_{cc}]^{-1}\{\{\lambda_{M,c}\}M-[\lambda_{cf}]\{P_f\}\} \tag{7}$$

Introducing Eq. (7) into the upper part of Eq. (6), and considering again that $P_{fi}=P_{Pi}$, it is straightforward to compute the displacements $\{w_f\}$,

$$\begin{aligned}\{w_f\}&=(\{\lambda_{M,f}\}-[\lambda_{fc}][\lambda_{cc}]^{-1}\{\lambda_{M,c}\})M\\&\quad+([\lambda_{fc}][\lambda_{cc}]^{-1}[\lambda_{cf}]-[\lambda_{ff}])\{P_f\}\end{aligned} \tag{8}$$

For the case of repeated loading, the generalization of these formulas is obvious: it is only necessary to consider each single monotonic part of the loading process, with the load either increasing or decreasing, as starting from an initial configuration (here after indicated by the subscript zero), as shown in Fig. 2. Therefore, Eq. (4) can be rewritten in its incremental form,

$$\{w\}-\{w_0\}=\{\lambda_M\}(M-M_0)-[\lambda](\{P\}-\{P_0\}) \tag{9}$$

As in the previous case, the system of the n equations (9) presents $2n$ unknowns: the displacements $\{w\}$ and the reactions $\{P\}$ of the n reinforcements crossing the crack wake. The problem can be solved on the basis of the already described compatibility condition: The ith component of the left-hand side term of Eq. (9) should be zero till the inverse limit force of the ith reinforcement is attained: $P_i=-P_{Pi}$. Unless this condition is reached, the n indeterminate reactions $\{P\}$ are expressed by

$$\{P\}-\{P_0\}=[\lambda]^{-1}\{\lambda_M\}(M-M_0) \tag{10}$$

while the displacements are known: $\{w\}=\{w_0\}$. Equation (10) is a generalization of Eq. (5); following the same reasoning, Eqs. (6)–(8) could also be generalized as follows:

$$\begin{aligned}\begin{Bmatrix} w_f \\ w_c \end{Bmatrix}&=\begin{Bmatrix} w_{0,f} \\ w_{0,c} \end{Bmatrix}+\begin{Bmatrix} \lambda_{M,f} \\ \lambda_{M,c} \end{Bmatrix}(M-M_0)\\&\quad-\begin{bmatrix} \lambda_{ff} & \lambda_{fc} \\ \lambda_{cf} & \lambda_{cc} \end{bmatrix}\left(\begin{Bmatrix} P_f \\ P_c \end{Bmatrix}-\begin{Bmatrix} P_{0,f} \\ P_{0,c} \end{Bmatrix}\right)\end{aligned} \tag{11}$$

with the following note: $P_{fi}=g\ P_{Pi}$, $i=1,\ldots,n_f$, where g is a flag variable, equal to 1 if the load is increasing, equal to -1 if, on the contrary, the load is decreasing. Considering that $w_c=w_{0,c}$, from the lower part of Eq. (11), we obtain

$$\{P_c\}=\{P_{0,c}\}+[\lambda_{cc}]^{-1}\{\{\lambda_{M,c}\}(M-M_0)-[\lambda_{cf}](\{P_f\}-\{P_{0,f}\})\} \tag{12}$$

As in the previous case, it is finally possible to compute the free crack opening displacements $\{w_f\}$,

$$\begin{aligned}\{w_f\}&=\{w_{0,f}\}+(\{\lambda_{M,f}\}-[\lambda_{fc}][\lambda_{cc}]^{-1}\{\lambda_{M,c}\})(M-M_0)\\&\quad+([\lambda_{fc}][\lambda_{cc}]^{-1}[\lambda_{cf}]-[\lambda_{ff}])(\{P_f\}-\{P_{0,f}\})\end{aligned} \tag{13}$$

The last three equations are the kernel of the algorithm in the case of repeated or cyclic loading.

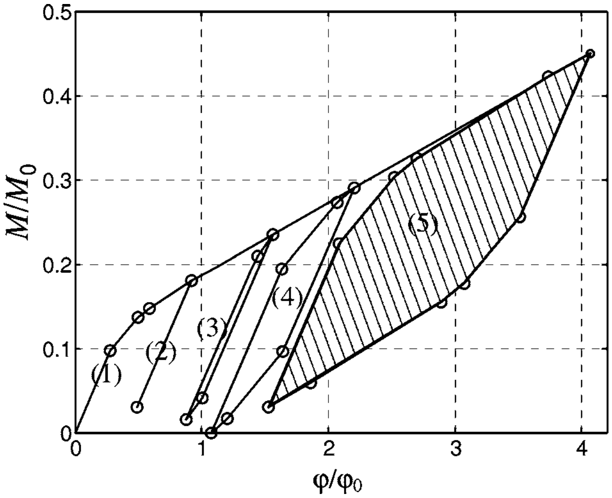


Fig. 3 Typical moment-rotation response, with evidence of elastic (2) and plastic (3–5) shakedown

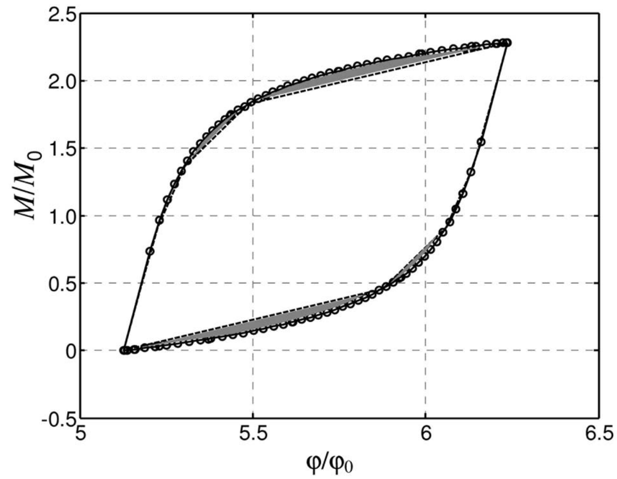


Fig. 4 Moment versus rotation relation for a specimen with n=50 reinforcements: complete loading-unloading cycle. The continuous line with circles represents the outcome of the exact algorithm proposed in [3], whereas the dotted one reports results of the iterative procedure.

Numerical Procedure

Two possibilities are given for computing the static and kinematic unknowns for an assigned load M; the simplest way, proposed by Carpinteri et al. [3], is to consider the fact that the system behavior is piecewise linear. Starting from the initial condition, in which all fibers behave elastically ($n_c=n$, $n_f=0$), the authors consider that the most solicited reinforcement (say, the jth) is the first to reach its plastic limit and compute the load factor, which exactly gives $P_j=P_{Pj}$. After this first step, they set $P_j=P_{Pj}$ ($n_c=m-1$, $n_f=1$) and then compute the other reinforcements actions from Eq. (12); again, the system response is linear, till the most solicited reinforcement (say, the kth) reaches its plastic limit. Again, the load factor is computed, which gives $P_k=P_{Pk}$, and then the procedure is iterated up to the maximum bending moment M. This algorithm, although efficient, has the disadvantage that, in the case of a high number of fibers, may require a high number of numerical steps for describing a complete loading-unloading cycle.

To overcome this limitation, the following iterative procedure is proposed, in which the true unknowns are $2n+1$, namely, the crack openings $\{w\}$, the reinforcements indeterminate actions $\{P\}$ and the number n_f of plasticized (or slipped) reinforcing elements. The procedure is summarized in the following flowchart:

1. initialize n_f (and $n_c=n-n_f$).
2. compute bridging actions $\{P_c\}$, (Eq. (10) if $n_f=0$, Eq. (12) otherwise).
3. loop entering condition if there are stresses outside the allowed range:

$$P_i > P_{Pi} \text{ or } P_i < -P_{Pi}.$$

   a. Stresses exceeding or nearby the maximum are set to the maximum:

$$P_i = \min(P_i, P_{Pi}).$$

   b. Stresses below or nearby the minimum are set to the minimum:

$$P_i = \max(P_i, -P_{Pi}).$$

   c. Update n_f (and n_c), if necessary.
   d. Loop exit condition: if n_f has been changed in step c., return to step 2, otherwise exit the loop.

4. Compute the crack openings $\{w_f\}$, Eq. (13).

The present model is able to capture the flexural behavior of fiber-reinforced materials, with their hardening and the elastic-plastic shakedown above certain load thresholds [3,39–43], each fiber yielding results in a slight decrease of the overall system stiffness. A synthetic example is reported in Fig. 3 for the case of four fibers: elastic, as well plastic, shakedown is clearly visible.

The advantage of the described procedure is that it allows one to compute a complete loading-unloading cycle without the necessity to compute the values of plastic (or shakedown) moments; this point is very important if the model is used in fatigue calculations, where a large number of cycles should be simulated. An example is reported in Fig. 4, where a complete loading unloading cycle is represented in terms of bending moment versus localized rotation (note that both quantities are expressed in nondimensional or normalized form; the normalization factors are $M_0=K_{Ic}bh^{3/2}$ and $\varphi_0=K_{IC}/(E\sqrt{h})$). The continuous line corresponds to the procedure by Carpinteri et al. [3], the circles being the values of the moment at which reinforcements attain their limits, either in tension or in compression; the dotted line corresponds to the present iterative procedure, with the complete cycle approximated by means of only ten points. The drawback of this procedure is represented by a small error in the evaluation of the dissipated energy per cycle, which is slightly underestimated, as could be seen in the graph of Fig. 4 (the gray-shaded area corresponds to the error, which is <5%).

Crack Propagation and Energy Dissipation Capability

Thus far, the algorithm considers only the system behavior at a fixed crack length a; nevertheless, it is clear that, in real problems, the crack may propagate. There are several ways of including crack propagation in the bridged crack model; most of the papers available in the literature consider directly an empirical crack propagation criterion according to the Paris and Erdogan law [46], as for instance, in Matsumoto, and Li [47] and Carpinteri et al. [3,43]. This is probably because these papers are focused on the fatigue modeling of the fibrous composite. On the contrary, in those papers in which the focus is on the constitutive flexural behavior of the composite beam, crack propagation is introduced—consistently with the model premises—on the basis of linear elastic fracture mechanics (LEFM).

In virtue of the superposition principle, the total stress intensity factor K_I is the sum of two contributions: that due to the externally applied bending moment M and that of the n forces applied on the crack surfaces, due to the reinforcements, both defined in [32]. Equating the total stress intensity factor K_I to its critical value K_{IC}, the fracture propagation moment M_F is obtained. Its value is given by the following relation:

$$\frac{M_F}{bh^{3/2}K_{IC}} = \frac{1}{Y_M(\xi)}\left[1 + N_P \sum_{i=1}^{n} \alpha_i \frac{\rho_i}{\rho} Y_{Pi}\left(\frac{c_i}{h}, \xi\right)\right] \tag{14}$$

where Y_m and Y_{Pi}, $i=1,\ldots,n$, are the shape functions already defined, $\xi = a/h$ is the relative crack depth, and α_i is the ratio of the actual value of the force carried by the ith reinforcement to its limit value at plastic flow. The brittleness number N_P that appears in Eq. (14) is defined as [1]

$$N_P = \rho \frac{\sigma_y h^{1/2}}{K_{IC}} \tag{15}$$

where ρ is the volume fraction of the fibers. This parameter is the fundamental quantity governing the system behavior: the higher N_P, the more ductile the system behavior results to be [1,2,32–35].

As evidenced first by Carpinteri and Carpinteri [39] for the case of one fiber, by Carpinteri and Puzzi for the case of two fibers [41] and by Carpinteri et al. [3] for the case of three or more fibers, the value of fracture propagation moment M_F could be higher or lower than M_{SD} (and even M_P), depending on the values of ξ and N_P. In other words, by varying the fundamental parameters that describe the composite beam, it is possible to pass from a ductile to a brittle behavior, in which fracture propagation precedes the onset of shakedown and, in some cases, even the onset of fiber slippage. In these latter cases, it is evident that the composite is not able to dissipate any energy if subjected to repeated or cyclic loading.

The easiest effect to ascertain is that of crack length ξ: typically, the most brittle behavior is found when ξ is higher, whereas short to medium crack lengths usually lead to the presence of elastic or plastic shakedown. However, the combined effect of all parameters must be investigated, since the above statement is valid only in the case of a medium to high value of the brittleness number N_P: on the contrary, if N_P is very low (which corresponds to the case of lightly reinforced beams), shakedown appears at almost any crack length (see examples for one and three fibers in [39] and [3], respectively).

In order to gain a more complete information, we performed a detailed analysis by varying the number of fibers n and the brittleness number N_P at a constant crack length. By doing so, we could evaluate the effect of an increase in both the fiber number and the brittleness number N_P. Results are summarized in Fig. 5, where, by varying the fiber number n from 1 up to 150 and the brittleness number N_P from 0.01 up to 2.00, five distinct zones are obtained. The considered crack depth is $\xi = 0.3$, and the fibers are equally spaced. In the first zone, marked by (A), the combination of a very low fiber number and a sufficiently high brittleness number provides the most brittle system response, with unstable fracture preceding the fiber yielding (or slippage). In the second zone, marked by (B), at least one fiber undergoes yielding, but the fracture moment always precedes the onset of plastic shakedown. In the third one, marked by (C), all fibers undergo yielding, but plastic shakedown is again ruled out by unstable crack propagation. By further decreasing the brittleness number, or by further increasing the fiber number, a more ductile system behavior could be obtained: In zone (D), in fact, plastic shakedown of one or more fibers occurs, therefore providing a more ductile system response, with greater capacity of energy dissipation under repeated and cyclic loading condition. Eventually, the most favorable condition is met only in the case of a very low brittleness number; in this case, which corresponds to the zone marked by (E) in Fig. 5, all reinforcements undergo plastic shakedown and the energy absorption capacity is maximum. In Fig. 5(b), a zoom on the lower values of n is reported in order to clearly observe the bounds of the zone (A). It can be noted that the upper bound is almost horizontal; therefore, the most brittle behavior may occur only in the case of a beam with a very low number of fibers, namely, only in the cases of $n=1$ or 2.

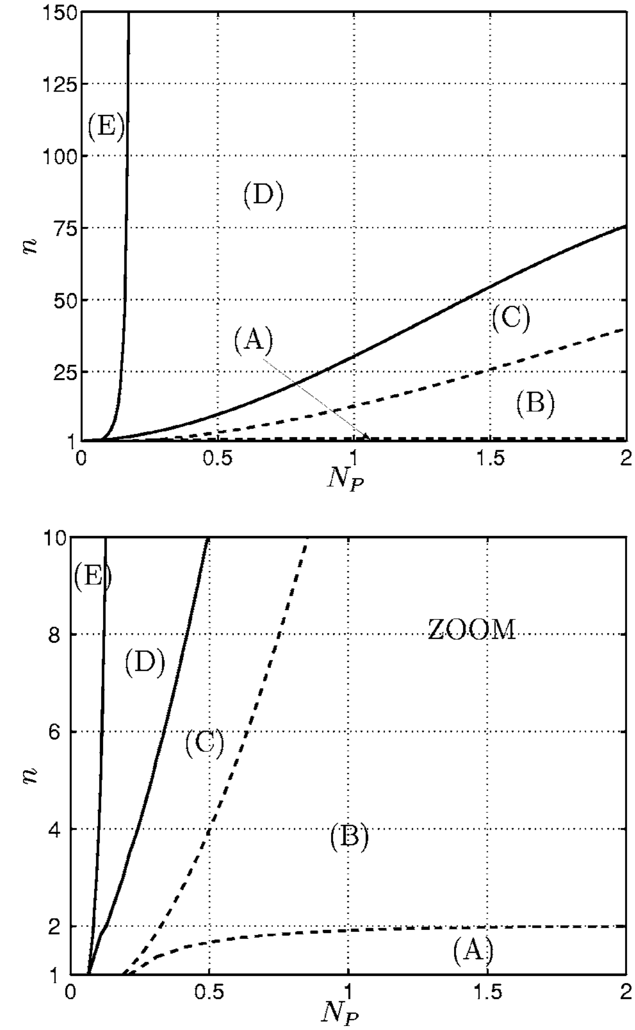


Fig. 5 Effect of fiber number n and brittleness number N_P on the system response type and energy dissipation: (A) very brittle to (E) very ductile. Graph (*a*) presents the whole diagram and (*b*) shows a zoomed view of the portion near the axes origin.

Thus, the following conclusions could be drawn from the obtained graphs: In order to have great energy dissipation and probably longer fatigue life, the brittleness number should be not too high, while an increase of the fiber number is, in general, beneficial. Eventually, it could be remarked that, if the crack length is varied, then a similar diagram is obtained in which all curves shift toward the upper-left direction if the crack length is increased (or, conversely, towards the lower-right direction if the crack length is decreased). In other words, as previously observed, longer cracks produce a more brittle behavior. If we introduce a third axis for representing the crack length, the curves of Fig. 5 expand into manifolds, which give a complete information about the combined effects of ξ, N_P and n on the energy dissipation capability of the reinforced composite beam.

It is evident that the proposed crack propagation criterion, based on the LEFM solution for a crack crossed by fibers, results in a threshold function for the external load: if the load remains below it, then the crack does not propagate; otherwise, if the load overcomes it, then the crack starts propagating. The crack propagation stability has been addressed by Carpinteri and co-workers in several papers [2,32–35]. It has been shown, in the case of monotonic loading that the crack propagation may be catastrophic or not, depending on the brittleness number. In the former case, which corresponds to structures badly reinforced (low values of N_P), the crack propagation is not stable and phenomena of instability (both of the snap-back and/or the snap-through type) are possible. In the latter case, which corresponds to more heavily reinforced composite beams (with higher values of N_P), the crack propagation stops after a certain amount, thanks to the bridging actions, which avoid the crack to prosecute.

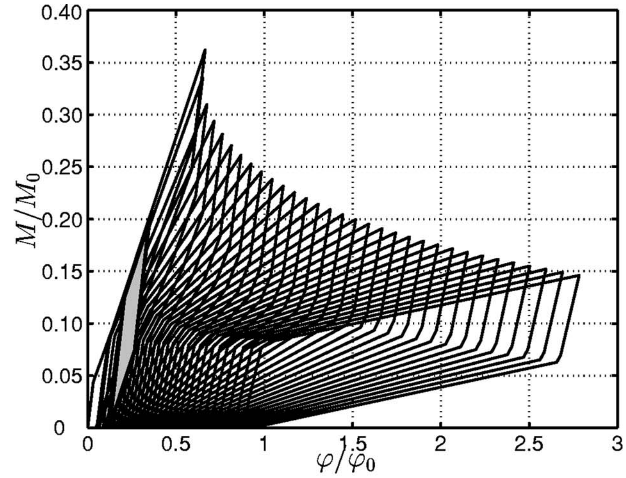
In the case of cyclic loading, the same behavior may happen and the equality $M=M_F$, i.e., external load equal to the fracture propagation moment, may not bring to brittle failure, as was the case, for instance, in the paper by Carpinteri et al. [3]. This could obviously be the case, but not in general. There is no particular reason, if cyclic loading is considered, for obtaining a constitutive composite behavior different from that obtained in the case of monotonic loading, at least if the loading frequency is not too high. Furthermore, there is an additional reason for considering that the function $M_F(\xi)$ is not always a decreasing function of ξ, in particular, if fiber-reinforced composites are modeled: In this case, in fact, crack propagation involves crossing of new intact fibers, which exert their actions across the crack wake. In this case, the sum in Eq. (14) changes its upper limit n, which is increased. The physical resulting effect is a possible increase of M_F as the crack propagates.

In order to take into consideration the above remarks, the numerical algorithm with crack propagation may be constructed as follows: if the external load overcomes the threshold value M_F, then the crack length is increased; as a consequence, all compliances, displacements, reinforcement reactions, and M_F have to be updated. The crack length is increased until one of the following occurs: either failure of the composite beam is achieved or the value of the fracture propagating moment overcomes the external loading value: $M<M_F$. In the latter case, crack propagation stops and the algorithm prosecutes as shown before. The algorithm is summarized in the following chart, in which internal procedure 2.a is the iterative procedure previously outlined:

1. Data initialization and computation of initial compliances
2. Loop for any given load $M^{(k)}$ (the apex (k) indicates the kth iteration):
   a. Iterative procedure for the computation of $P_i^{(k)}$, $w_i^{(k)}$ $(i=1,\ldots,n)$, $\varphi^{(k)}$, and $n_c^{(k)}$, $n_f^{(k)}$
   b. Computation of the fracture propagation moment $M_F^{(k)}$, Eq. (14)
   c. Crack propagation condition: $M^{(k)}>M_F^{(k)}$
      i. Update crack depth $a^{(k)}$
      ii. Update number of fibers $n^{(k)}$ across the crack and compliances
      iii. Return to step 2.a
   d. Check for load inversion: If load is inverted, then update values of variables $\{P_0\}$, $\{w_0\}$, φ_0.

This algorithm has the advantage that it allows one to simulate not only the dead load condition, but also the fixed grip condition, which is sometimes used also in the case of repeated loading [48]. In the latter case, it is possible to reproduce the complete softening behavior of the structure, with the envelope of the maxima of the cycles being coincident with the moment-rotation diagram obtained from the bridged crack model in the case of monotonic loading.

Two examples are reported in Fig. 6; they refer to a composite beam with $n=2$ reinforcements equally spaced in an initial crack of length $\xi=0.3$. The system response is represented in terms of moment-rotation relation, expressed, as before, through normalized (or nondimensional) quantities. The response of a beam characterized by $N_P=0.05$ is reported in Fig. 6(a), whereas in Fig. 6(b) that of a beam with $N_P=0.15$ is presented. Both diagrams clearly evidence the softening branch; as expected, the beam with higher brittleness number (Fig. 6(b)) exhibits higher load levels (higher cracking resistance) and also higher plastic deformations. Nevertheless, the beam in Fig. 6(b) exhibits plastic shakedown of only one reinforcement, while the beam 6(a) of both of them. In fact, if we look at Fig. 5(b), we could note that the beam in Fig. 5(b) is contained in the region marked by (D), while the beam 5(a) lies in region (E). Furthermore, as the crack propagates, the beam in Fig. 5(b) is no longer able to dissipate energy, since the shakedown disappears, being preceded by crack propagation. This effect is due to the fact that the lines in Fig. 5 shift towards the upper-left corner as the crack propagates, as already remarked. On the contrary, the beam in Fig. 6(a) continues to display hysteretic cycles as the crack propagates. As a result, the more ductile composite beam (with higher N_P) does not dissipate a larger amount of energy. This fact is even more clear if we compare the beams under the same external bending load, $M/M_0=0.2$, as shown in Fig. 6. The beam in Fig. 6(a) describes a hysteretic cycle, therefore dissipating energy whereas the beam in Fig. 6(b) is not able to dissipate any energy, presenting only elastic shakedown.



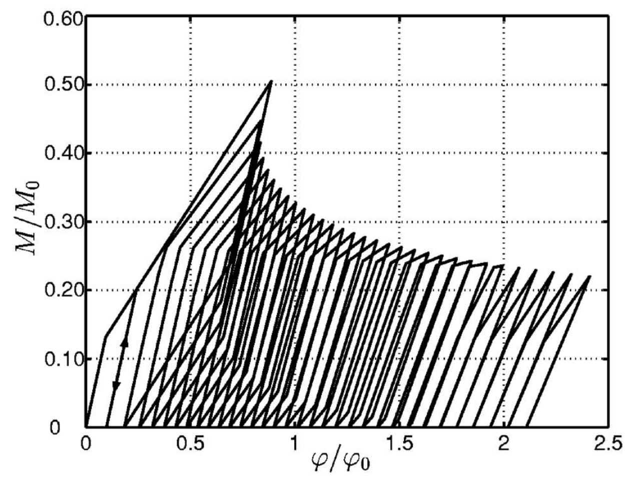


Fig. 6 Repeated loading (fixed grip condition) of a composite beam with two reinforcements and initial crack length $\xi=0.30$; $N_P=0.05$ and $N_P=0.15$ in (a) and (b), respectively. Large hysteresis loops are visible only in (b).

The above example clearly evidences that, when repeated loading is considered, the brittleness number is not enough to characterize the beam behavior. In fact, the beam with higher N_P does not necessarily result in being the more ductile, with reference to hysteretic energy dissipation. It is also evident the relevance of the diagram reported in Fig. 5. This diagram allows to evaluate, for instance, if an increase in n is beneficial or not to the beam, in terms of the increase in energy dissipation. If we consider, for instance, a beam with $N_P=0.05$, initial crack length $\xi=0.3$ and $n=1$, an increase to $n=3$ leads to a greater amount of energy dissipation and to a longer fatigue life (as in the example reported by Carpinteri et al. [3]). In fact, the increase in n determines a transition from zone (C) to zone (D), where plastic shakedown occurs; see Fig. 5(b). If we consider a larger brittleness number, for instance, $N_P=0.5$, the same increase in fiber number does not lead to any improvement of the reinforced beam, since in the latter case the beam remains within zone (C), where no shakedown occurs. As a consequence, no energy dissipation occurs and the increase in n is not effective.

Conclusions

In this paper, we presented some issues in the modeling of brittle matrix composites with discontinuous reinforcements under the condition of repeated bending loading. In particular, addressing the case of composites with high number of reinforcements, we proposed a new iterative procedure applied to the bridged crack model [1,2]. Furthermore, we analyzed the combined effects of crack length, brittleness number, and fiber number on the cyclic behavior of the composite beam, drawing interesting considerations about hysteretic energy dissipation in the composite beam. Eventually, we analyzed crack propagation by modeling it with a fracture-mechanics-based criterion and showed examples of simulations of repeated bending tests under fixed grip conditions.

Acknowledgment

Support by the European Community is gratefully acknowledged by the authors. Thanks are also due to the Italian Ministry of Education, University and Scientific Research (MIUR).

Nomenclature

A_i = cross-sectional area of the ith reinforcement
E = Young's modulus of the matrix
K_{IC} = critical value of the Stress-Intensity Factor
M = bending moment
M_F = crack propagation moment
M_P = plastic bending moment
M_{SD} = plastic shake-down moment
M_0 = value of bending moment at load inversion
N = total number of reinforcements
N_P = brittleness number
$\{P\}$ = vector of the reinforcement actions
$\{P_0\}$ = vector of the reinforcement actions at load inversion
P_i = ith reinforcement action
P_{Pi} = limit value of the ith reinforcement action
Y_M = shape function for the stress intensity factor
Y_{Pi} = shape function for the stress intensity factor
a = crack length
b = section thickness
c_i = position of the ith reinforcement
g = flag variable (=1 loading, =−1 unloading)
h = section height
n = number of reinforcements across the crack wake
n_c = number of non-plasticized reinforcements
n_f = number of plasticized reinforcements
$\{w\}$ = vector of the crack openings
$\{w_0\}$ = vector of the crack openings at load inversion
w_c = free crack opening
w_f = constrained crack opening
w_i = total crack opening at the level of the ith reinforcement
w_{iM} = crack opening at the level of the ith reinforcement, due to the bending moment M
w_{ij} = crack opening at the level of the ith reinforcement, due to a pair of forces P_j
α_i = ratio of the actual value of the force carried by the ith reinforcement to its limit value
Δl = cracked section length
$[\lambda]$ = matrix of extensional-extensional compliances
$\{\lambda_{\mathrm{M}}\}$ = vector of extensional-bending compliances
λ_{iM} = extensional-bending local compliance of the cracked beam
λ_{ij} = extensional-extensional local compliance of the cracked beam
λ_{MM} = bending-bending local compliance of the cracked beam
ϕ = total local rotation of the cracked beam
ϕ_i = local rotation due to a pair of forces P_i
ϕ_M = local rotation due to the bending moment M
ρ = total reinforcement percentage
ρ_i = reinforcement percentage of the ith reinforcement
σ_y = reinforcement yield strength
ξ = normalized crack length
ζ = normalized coordinate related to the bottom of the cross section
ζ_i = normalized position of the ith fiber

References

[1] Carpinteri, A., 1981, "A Fracture Mechanics Model for Reinforced Concrete Collapse," *Proc. of the IABSE Colloquium on Advanced Mechanics of Reinforced Concrete*, Delft, I.A.B.S.E., Zürich, pp. 17–30.
[2] Carpinteri, A., 1984, "Stability of Fracturing Process in R.C. Beams," J. Struct. Eng., **110**, pp. 544–558.
[3] Carpinteri, An., Spagnoli, A., and Vantadori, S., 2004, "A Fracture Mechanics Model for a Composite Beam With Multiple Reinforcements Under Cyclic Bending," Int. J. Solids Struct., **41**, pp. 5499–5515.
[4] Barenblatt, G. I., 1962, "The Mathematical Theory of Equilibrium Cracks in Brittle Fracture," Adv. Appl. Mech., **7**, pp. 55–129.
[5] Dugdale, D. S., 1960, "Yielding of Steel Sheets Containing Slits," J. Mech. Phys. Solids, **8**, pp. 100–104.
[6] Hillerborg, A., Modeer, M., and Petersson, P. E., 1976, "Analysis of Crack Formation and Crack Growth in Concrete by Means of Fracture Mechanics and Finite Elements," Cem. Concr. Res., **6**, pp. 773–782.
[7] Petersson, P. E., 1981, "Crack Growth and Development of Fracture Zones in Plain Concrete and Similar Materials," Ph.D. thesis, Lund. Institute of Technology, Lund, Denmark.
[8] Hillerborg, A., 1980, "Analysis of Fracture by Means of the Fictitious Crack Model, Particularly for Fibre Reinforced Concrete," Int. J. Cem. Comp., **2**, pp. 177–184.
[9] Carpinteri, A., 1989, "Cusp Catastrophe Interpretation of Fracture Instability," J. Mech. Phys. Solids, **37**, pp. 567–582.
[10] Carpinteri, A., 1985, "Interpretation of the Griffith Instability as a Bifurcation of the Global Equilibrium," *Application of Fracture Mechanics to Cementitious Composites (Proc. of a NATO Advanced Research Workshop, Evanston, IL, 1984)*, S. P. Shah, ed., Martinus Nijhoff, Dordrecht, pp. 287–316.
[11] Carpinteri, A., ed., 1999, *Nonlinear Crack Models for Nonmetallic Materials*, Kluwer, Dordrecht.
[12] Shah, S. P., 1988, "Fracture Toughness of Cement-Based Materials," Mater. Struct., **21**, pp. 145–150.
[13] Cotterell, B., Paravasivam, P., and Lam, K. Y., 1992, "Modelling the Fracture of Cementitious Composites," Mater. Struct., **25**, pp. 14–20.
[14] Li, V. C., and Liang, E., 1986, "Fracture Processes in Concrete and Fiber Reinforced Cementitious Composites," J. Eng. Mech., **112**, pp. 566–586.
[15] Wecharatana, M., and Shah, S. P., 1983 "A Model for Prediction of Fracture Resistance of Fiber Reinforced Concrete," Cem. Concr. Res., **13**, pp. 819–829.
[16] Visalvanich, K., and Naaman, A. E., 1983, "Fracture Model for Fiber Reinforced Concrete," ACI J. **80**, pp. 128–138.
[17] Ballarini, R., Shah, S. P., and Keer, L. M., 1984, "Crack Growth in Cement Based Composites," Eng. Fract. Mech., **20**, pp. 433–445.
[18] Mai, Y. W., 1985, "Fracture Measurements for Cementitious Composites," *Application of Fracture Mechanics to Cementitious Composites (Proc. of a NATO Advanced Research Workshop*, Evanston, IL, 1984), S. P. Shah, ed., Martinus Nijhoff, Dordrecht, pp. 399–429.
[19] Jenq, Y. S., and Shah, S. P., 1985, "A Two-Parameter Fracture Model for

Concrete," J. Eng. Mech., **111**, pp. 1227–1241.
[20] Foote, R. M. L., Mai, Y. W., and Cotterell, B., 1986, "Crack Growth Resistance Curves in Strain-Softening Materials," J. Mech. Phys. Solids, **6**, pp. 593–607.
[21] Zhang, J., and Li, V. C., 2004, "Simulation of Crack Propagation in Fiber-Reinforced Concrete by Fracture Mechanics," Cem. Concr. Res., **34**, pp. 333–339.
[22] Ruiz, G., 2001, "Propagation of a Cohesive Crack Crossing a Reinforcement Layer," Int. J. Fract., **111**, pp. 265–282.
[23] Erdogan, F., and Joseph, P. F., 1987, "Toughening of Ceramics Through Crack Bridging by Ductile Particles," J. Mech. Phys. Solids, **35**, pp. 262–270.
[24] Mai, Y. W., 1991, "Fracture and Fatigue of Non-transformable Ceramics: The Role of Crack-Interface Bridging," *Fracture Processes in Concrete, Rock and Ceramics*, J. G. M. Van Mier, ed., E.&F.N. Spon, London, pp. 3–26.
[25] Cox, B. N., 1991, "Extrinsic Factors in the Mechanics of Bridged Cracks," Acta Metall. Mater., **39**, pp. 1189–1201.
[26] Marshall, D. B., Cox, B. N., and Evans, A. G., 1985, "The Mechanics of Matrix Cracking in Brittle Matrix Fiber Composites," Acta Metall., **33**, pp. 2013–2021.
[27] Marshall, D. B., and Cox, B. N., 1987, "Tensile Fracture of Brittle-Matrix Composites: Influence of Fiber Strength," Acta Metall. **35**, pp. 2607–2619.
[28] Budiansky, B., Hutchinson, J. W., and Evans, A. G., 1986, "Matrix Fracture in Fiber-Reinforced Ceramics," J. Mech. Phys. Solids, **34**, pp. 167–189.
[29] Cox, B. N., and Marshall, D. B., 1991, "Crack Bridging in the Fatigue of Fibrous Composites," Fatigue Fract. Eng. Mater. Struct., **14**, pp. 847–861.
[30] Cox, B. N., and Marshall, D. B., 1991, "Stable and Unstable Solutions for Bridged Cracks in Various Specimens," Acta Metall. Mater., **39**, pp. 579–589.
[31] Ballarini, R., and Muju, S., 1991, "Stability Analysis of Bridged Cracks in Brittle Matrix Composites," ASME Paper No. 91-GT-094.
[32] Bosco, C., and Carpinteri, A., 1992, "Fracture Behavior of Beam Cracked Across Reinforcement," Theor. Appl. Fract. Mech., **17**, pp. 61–68.
[33] Carpinteri, A., and Bosco, C., 1995, "Discontinuous Constitutive Response of Brittle Matrix Fibrous Composites," J. Mech. Phys. Solids, **43**, pp. 261–274.
[34] Carpinteri, A., and Massabò, R., 1996, "Bridged Versus Cohesive Crack in the Flexural Behavior of Brittle-Matrix Composites," Int. J. Fract., **81**, pp. 125–145.
[35] Carpinteri, A., and Massabò, R., 1997, "Continuous vs Discontinuous Bridged-Crack Model for Fiber-Reinforced Materials in Flexure," Int. J. Solids Struct., **34**, pp. 2321–2338.
[36] Carpinteri, A., Ferro, G., and Ventura, G., 2003, "Bending Behaviour Simulations of Fibre Reinforced Beams With Rebars by Using the Bridged Crack Model," *Computational Modelling of Concrete Structures, Proc. of Euro-C 2003 Conference*, St. Johann im Pongau, Austria, N. Bićcanićc, R. de Borst, H. Mang, and G. Meschke, eds., Balkema, pp. 707–716.
[37] Carpinteri, A., Ferro, G., and Ventura, G., 2003, "Size Effects on Flexural Response of Reinforced Concrete Elements with a Nonlinear Matrix," Eng. Fract. Mech., **70**, pp. 995–1013.
[38] Carpinteri, A., Ferro, G., and Ventura, G., 2004, "Double Brittle-to-Ductile Transition in Bending of Fibre Reinforced Concrete Beams With Rebars," Int. J. Numer. Analyt. Meth. Geomech., **28**, pp. 737–756.
[39] Carpinteri, A., and Carpinteri, An., 1984, "Hysteretic Behavior of RC Beams," J. Struct. Eng., **110**, pp. 2073–2084.
[40] Carpinteri, An., 1992, "Reinforced Concrete Beam Behavior Under Cyclic Loadings," *Applications of Fracture Mechanics to Reinforced Concrete*, A. Carpinteri, ed., Elsevier, New York, pp. 547–578.
[41] Carpinteri, A., and Puzzi, S., 2003, "Hysteretic Flexural Behaviour of Brittle Matrix Fibrous Composites: The Case of Two Fibers," *Proc. of 16th National (Italian) Congress of Theoretical and Applied Mechanics (AIMETA)*, A.I.M.E.T.A., Ferrara, CD-ROM, File 102.
[42] Carpinteri, A., and Puzzi, S., 2004, "The Bridged Crack Model for the Analysis of FRC Elements Under Repeated Bending Loading," *Fibre-Reinforced Concretes, Proc. of 6th RILEM Symposium on Fibre Reinforced Concrete*, Varenna, Italy, R.I.L.E.M. Publications S.a.r.l., Bagneux, M. Di Prisco, R. Felicetti, and G. A. Plizzari, eds., pp. 767–776.
[43] Carpinteri, An., Spagnoli, A., and Vantadori, S., 2005, "Mechanical Damage of Ordinary or Prestressed Reinforced Concrete Beams Under Cyclic Bending," Eng. Fract. Mech., **72**, pp. 1313–1328.
[44] Tada, H., Paris, P., and Irwin, G., 1985, *The Stress Analysis of Cracks Handbook*, Paris Productions, St. Louis, MO (and Del Research Corporation, 1963).
[45] Murakami, Y., 1987, *Stress Intensity Factors Handbook*, Pergamon Press, Oxford.
[46] Paris, P. C., and Erdogan, F., 1963, "A Critical Analysis of Crack Propagation Laws," ASME J. Basic Eng., **85**, pp. 528–534.
[47] Matsumoto, T., and Li, V. C., 1999, "Fatigue Life Analysis of Fiber Reinforced Concrete With a Fracture Mechanics Based Model," CRC Crit. Rev. Solid State Mater. Sci., **21**, pp. 249–261.
[48] Reis, J. M. L., and Ferreira, A. J. M., 2003, "Fracture Behaviour of Glass Fiber Reinforced Polymer Concrete," Polym. Test., **22**, pp. 149–153.

Exact Solutions for Free-Vibration Analysis of Rectangular Plates Using Bessel Functions

Jiu Hui Wu

A. Q. Liu

School of Electrical & Electronic Engineering,
Nanyang Technological University,
Singapore 639798

H. L. Chen

Institute of Vibration and Noise Control,
Xi'an Jiaotong University,
Xi'an 710049, China

A novel Bessel function method is proposed to obtain the exact solutions for the free-vibration analysis of rectangular thin plates with three edge conditions: (i) fully simply supported; (ii) fully clamped, and (iii) two opposite edges simply supported and the other two edges clamped. Because Bessel functions satisfy the biharmonic differential equation of solid thin plate, the basic idea of the method is to superpose different Bessel functions to satisfy the edge conditions such that the governing differential equation and the boundary conditions of the thin plate are exactly satisfied. It is shown that the proposed method provides simple, direct, and highly accurate solutions for this family of problems. Examples are demonstrated by calculating the natural frequencies and the vibration modes for a square plate with all edges simply supported and clamped.
[DOI: 10.1115/1.2744043]

1 Introduction

The free-vibration analysis of a rectangular plate is of interest in the field of mechanics, civil, and aerospace engineering. Back in 1823, by using a double trigonometric series, Navier obtained the exact solution of bending of a rectangular plate with all edges simply supported [1]. In 1899, by using a single Fourier series, Levy developed a method for solving the rectangular plate bending problems with two opposite edges simply supported and the two remaining opposite edges with arbitrary conditions of supports [2]. In 1934, Way obtained the exact solutions for the large deflection analysis of a clamped circular plate [3]. For the free-vibration analysis of rectangular plates, accurate analytical results were presented for the cases having two opposite sides simply supported, whereas the other cases with the possible combinations of clamped, simply supported, and free edge conditions were analyzed by using the Ritz method by Leissa in 1973 [4]. In addition, the method of superposition was proposed by Gorman to examine free-vibration analysis of cantilever plates in 1976 [5] and that of rectangular plates with a combination of clamped and simply-supported edge conditions in 1977 [6].

More recently, many papers on the vibration analysis of rectangular plates have been published [7–10]. The free-vibration analysis of isotropic and anisotropic rectangular thin plates subjected to general boundary conditions was conducted by using a modified Ritz method by Narita in 2000 [10]. For centuries, however, an exact solution for a fully clamped rectangular plate has not yet been obtained, and it is currently considered that an exact solution is not achievable for the rectangular plate problem of this type.

In this paper, a Bessel function method is proposed to obtain an exact solution for the vibration problems of a rectangular plate by superposing different Bessel functions to satisfy three edge conditions: (i) fully simply supported, (ii) fully clamped, and (iii) two opposite edges simply supported and the other two edges clamped. By employing the proposed method, the exact solutions of the natural frequencies and mode shapes can be obtained for the rectangular thin plate with the aforementioned edge conditions. This new method provides simple, direct, and highly accurate solutions for this family of problems.

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received March 14, 2007; final manuscript received April 23, 2005. Review conducted by Oliver M. O'Reilly.

2 Thin Plate Theory

The free harmonic vibration of a thin plate with a constant thickness h is governed by the differential equation

$$D\nabla^4 W - \omega^2 \rho h W = 0 \tag{1}$$

where $W(x,y)$ is a typical mode, ∇^4 is the biharmonic differential operator (i.e., $\nabla^4=\nabla^2\nabla^2$, $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2$ in Cartesian coordinates), $D=Eh^3/12(1-\nu^2)$ is the bending rigidity with E and ν being the Young's modulus and the Poisson's ratio, respectively, ω is the natural frequency, and ρ is the mass density.

For a finite solid circular plate, the nth vibration mode of Eq. (1) in polar coordinates is [11]

$$W_n(r,\theta) = [A_n J_n(kr) + B_n I_n(kr)] \begin{matrix}\cos\\ \sin\end{matrix} (n\theta) \tag{2}$$

where A_n and B_n are constants to be determined, J_n and I_n are the Bessel function and the modified Bessel function of the first kind of order n, respectively, and $k^4=\omega^2\rho h/D$.

Thus, in Cartesian coordinates, Eq. (2) can be converted into

$$W_n(x,y) = [A_n J_n(k\sqrt{x^2+y^2}) + B_n I_n(k\sqrt{x^2+y^2})] \begin{matrix}\cos\\ \sin\end{matrix} \left[n \operatorname{atan}\left(\frac{y}{x}\right)\right] \tag{3}$$

According to [12], there exist the addition formulas,

$$J_n(R) \begin{matrix}\cos\\ \sin\end{matrix}\, n\psi = \sum_{m=-\infty}^{\infty} J_{n-m}(x) J_m(y) \begin{matrix}\cos\\ \sin\end{matrix}\, m\varphi \tag{4a}$$

and

$$I_n(R) \begin{matrix}\cos\\ \sin\end{matrix}\, n\psi = \sum_{m=-\infty}^{\infty} I_{n-m}(x) I_m(y) \begin{matrix}\cos\\ \sin\end{matrix}\, m\varphi \tag{4b}$$

where ψ is defined as $R\cos\psi=x-y\cos\varphi$, $R\sin\psi=y\sin\varphi$, and when y approaches 0, ψ approaches 0.

To satisfy $\psi=\operatorname{atan}(y/x)$, $\varphi=\pi/2$ is selected. Then, Eq. (3) becomes

$$W_n(x,y) = \sum_{m=-\infty}^{\infty} [A_n J_{n-m}(kx) J_m(ky) + B_n I_{n-m}(kx) I_m(ky)] \begin{matrix} \cos \\ \sin \end{matrix} \left(\frac{m\pi}{2}\right) \tag{5}$$

Equation (5) is the general solution for the vibration modes of solid rectangular plates, where A_n and B_n can be determined by use of the orthogonal characteristic of these vibration modes. It is noted that $J_{n-m}(kx)J_m(ky)$ and $I_{n-m}(kx)I_m(ky)$ are the core parts of the solution, which can be further adopted to construct the free-vibration solutions of a rectangular thin plate with different edge conditions. By superposing these Bessel functions to satisfy the edge conditions of a rectangular plate, the exact solution for the free vibration can be obtained. We call this approach the Bessel function method. It will be applied to analyze the free vibration of a rectangular plate with three edge conditions in the sequel.

3 Free-Vibration Analysis of a Rectangular Plate Under Different Edge Conditions

For a rectangular plate with edge lengths a and b, there are eight boundary conditions for every case. Three cases are discussed below: (i) fully simply supported, (ii) fully clamped, and (iii) two opposite edges simply supported and the other two edges clamped.

3.1 Fully Simply Supported Rectangular Plate. In this case, the boundary conditions are

$$\begin{aligned} &W|_{x=0} = 0, \quad W|_{x=a} = 0, \quad \left.\frac{\partial^2 W}{\partial x^2}\right|_{x=0} = 0, \quad \left.\frac{\partial^2 W}{\partial x^2}\right|_{x=a} = 0 \\ &W|_{y=0} = 0, \quad W|_{y=b} = 0, \quad \left.\frac{\partial^2 W}{\partial y^2}\right|_{y=0} = 0, \quad \left.\frac{\partial^2 W}{\partial y^2}\right|_{y=b} = 0 \end{aligned} \tag{6}$$

In order to satisfy all the above edge conditions, the vibration mode function can be constructed as

$$\begin{aligned} W_{n,m} = {} & (A_n\{J_{n-m}(kx) + J_{n-m}[k(a-x)]\}\{J_m(ky) + J_m[k(b-y)]\} \\ & + B_n\{I_{n-m}(kx) + I_{n-m}[k(a-x)]\}\{I_m(ky) \\ & + I_m[k(b-y)]\}) \sin\frac{m\pi}{2} \cos\frac{n\pi}{2} \end{aligned} \tag{7}$$

where the mode subscripts m and n are odd and even numbers for nontrivial solutions, respectively. Based on the following properties of the special functions:

$$J_m(0) = I_m(0) = \begin{cases} 1 & m=0 \\ 0 & m \neq 0 \end{cases} \tag{8a}$$

$$J''_m(0) = \begin{cases} \dfrac{-1}{2} & m=0 \\ \dfrac{1}{4} & m = \pm 2 \\ 0 & \text{others} \end{cases} \tag{8b}$$

and

$$I''_m(0) = \begin{cases} \dfrac{1}{2} & m=0 \\ \dfrac{1}{4} & m = \pm 2 \\ 0 & \text{others} \end{cases} \tag{8c}$$

for the case of $a=b$, Eq. (7) satisfies all the edge conditions provided that $m=n/2$ and

$$\det \begin{vmatrix} J_{n-m}(ka) & I_{n-m}(ka) \\ J''_m(kb) & I''_m(kb) \end{vmatrix} = 0 \tag{9}$$

Equation (9) is the frequency equation for the fully simply supported square plate.

For a fully simply supported rectangular plate, the well-known exact solution for the free-vibration analysis was obtained by Navier [1]. In the Navier-type solution, the mode functions and natural frequencies are [4,13]

$$W_{m,n} = A_{m,n} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b}, \quad (m,n = 1,2,\ldots) \tag{10a}$$

$$k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \tag{10b}$$

The free-vibration solutions expressed in Eqs. (7) and (9) by the proposed Bessel function method are different from the Navier-type solutions due to the different derivation processes. In the former method, using the Bessel functions in Eq. (5) that satisfy the governing equation (1), and Eq. (7) is constructed to satisfy the edge conditions in Eq. (6). In the Navier-type solutions, the double Fourier sine series in Eq. (10a) was constructed to satisfy the edge conditions, whereas Eq. (10b) was obtained from Eq. (1). Table 1 compares the first ten nondimensional natural frequencies ka of a square plate with an edge length of a obtained from Eqs. (9) and (10b), applying the Bessel function method and the Navier-type solution, respectively. It is noted that the two kinds of solutions are different in the sense that the Navier-type natural frequencies are relatively sparse. Because the frequency equations are both derived exactly from the mode function expansions, considering the convergence of Eq. (7), the natural frequency solutions from Eq. (9) are exact and a complement to the Navier-type solutions.

3.2 Fully Clamped Rectangular Plate. A fully clamped rectangular plate has the boundary conditions

Table 1 Comparison of the nondimensional natural frequencies using two different methods for a fully simply supported square plate

Order	Present theory in Eq. (9)	Navier-type solution shown in Eq. (10b)
1	3.6744	4.4429
2	6.2931	7.0248
3	6.9380	8.8858
4	8.7100	9.9346
5	9.7066	11.3272
6	10.1215	12.9531
7	11.0385	13.3286
8	12.2961	14.0496
9	12.9751	15.7080
10	13.2846	17.7715

Table 2 Comparison of the nondimensional natural frequencies using three different methods for a fully clamped square plate

Order	Present theory in Eq. (14)	Rayleigh-Ritz method [14]	Finite element method [15]
1	5.9057	5.9992	5.9540
2	8.3466	8.5680	8.4870
3	9.1969	10.4053	10.1833
4	10.6870	11.4734	11.3759
5	11.8367	11.5000	11.4140
6	12.4022	12.8511	

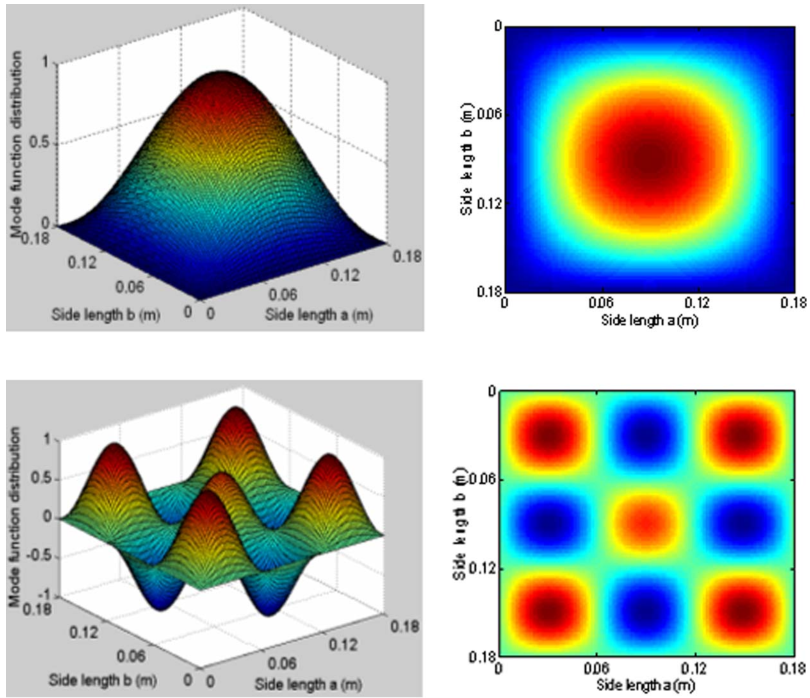


Fig. 1 Vibration mode functions with $n=2$ and $m=1$ for a fully simply supported plate at different nondimensional natural frequencies: (*a*) $ka=3.6744$ and (*b*) $ka=10.1215$

$$W|_{x=0}=0, \quad W|_{x=a}=0, \quad \left.\frac{\partial W}{\partial x}\right|_{x=0}=0, \quad \left.\frac{\partial W}{\partial x}\right|_{x=a}=0$$

$$W|_{y=0}=0, \quad W|_{y=b}=0, \quad \left.\frac{\partial W}{\partial y}\right|_{y=0}=0, \quad \left.\frac{\partial W}{\partial y}\right|_{y=b}=0 \tag{11}$$

The vibration mode function in this case can be constructed as

$$\begin{aligned} W_{n,m} = & (A_n\{J_{n-m}(kx)+J_{n-m}[k(a-x)]\}\{J_m(ky)+J_m[k(b-y)]\} \\ & + B_n\{I_{n-m}(kx)+I_{n-m}[k(a-x)]\}\{I_m(ky) \\ & + I_m[k(b-y)]\}) \cos\frac{m\pi}{2}\cos\frac{n\pi}{2} \end{aligned} \tag{12}$$

where m and n are even numbers. For the fully clamped square plate ($a=b$), since

$$J'_m(0)=\begin{cases} \frac{1}{2} & m=1 \\ \frac{-1}{2} & m=-1 \\ 0 & \text{others} \end{cases} \tag{13}$$

all the above conditions can be satisfied with $m=n/2$, $m\neq 0$, and

$$\det\begin{vmatrix} J_{n-m}(ka) & I_{n-m}(ka) \\ J'_m(kb) & I'_m(kb) \end{vmatrix}=0 \tag{14}$$

Equation (14) is the frequency equation for the fully clamped plate.

In order to verify Eq. (14), we compared the nondimensional frequencies ka of the fully clamped square plate obtained from Eq. (14), the Rayleigh-Ritz method [14], and the finite element method [15]. The comparison is summarized in Table 2. As shown in Table 2, the first two natural frequencies derived from Eq. (14) employing the Bessel function method are very close to those from the Rayleigh-Ritz method and the finite element method, and the discrepancy among them is $<3\%$. It is also known that the results obtained by the Rayleigh-Ritz method constitute upper bounds for the natural frequencies. As shown in Table 2, the first six natural frequencies from the Rayleigh-Ritz method are almost all higher than the first six exact eigenfrequencies from Eq. (14). Therefore, the frequency equation of the Bessel function method for the fully clamped plate is verified.

3.3 Rectangular Plate With Two Opposite Edges Simply Supported and the Other Two Edges Clamped.

The rectangular plate in this case has the boundary conditions

$$W|_{x=0}=0, \quad W|_{x=a}=0, \quad \left.\frac{\partial^2 W}{\partial x^2}\right|_{x=0}=0, \quad \left.\frac{\partial^2 W}{\partial x^2}\right|_{x=a}=0$$

$$W|_{y=0}=0, \quad W|_{y=b}=0, \quad \left.\frac{\partial W}{\partial y}\right|_{y=0}=0, \quad \left.\frac{\partial W}{\partial y}\right|_{y=b}=0 \tag{15}$$

The vibration mode function in this case can be constructed from

$$\begin{aligned} W_{n,m} = & (A_n\{J_{n-m}(kx)+J_{n-m}[k(a-x)]-J_{n-m}(ka)-J_{n-m}(0)\} \\ & \times\{J_m(ky)+J_m[k(b-y)]-J_m(kb)-J_m(0)\}+B_n\{I_{n-m}(kx) \\ & + I_{n-m}[k(a-x)]-I_{n-m}(ka)-I_{n-m}(0)\}\{I_m(ky)+I_m[k(b-y)] \\ & - I_m(kb)-I_m(0)\}) \exp\left(\frac{m\pi i}{2}\right) \end{aligned} \tag{16}$$

where n is an even number. For the case of $a=b$, all the boundary conditions are satisfied when $m=n/2$, $m\neq 0, \pm 1, \pm 2$, and

$$\det\begin{vmatrix} J''_{n-m}(ka) & I''_{n-m}(ka) \\ J'_m(kb) & I'_m(kb) \end{vmatrix}=0 \tag{17}$$

Equation (17) represents the frequency equation for this case.

Therefore, the vibration mode functions and frequency equations of the rectangular plate with different edge conditions: (i) fully simply supported, (ii) fully clamped, and (iii) two oppo-

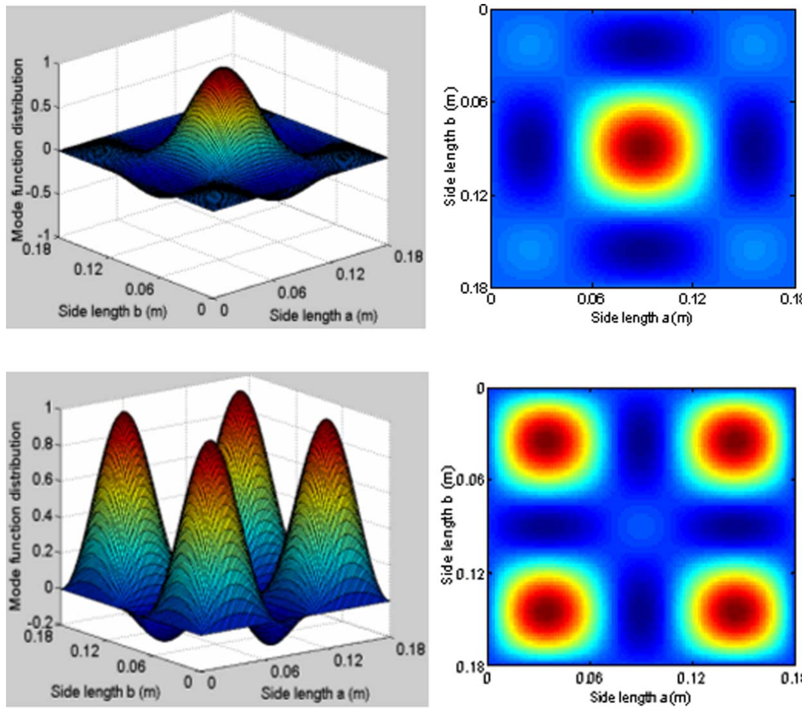


Fig. 2 Vibration mode functions with *n*=6 and *m*=3 for a fully simply supported plate at different nondimensional natural frequencies: (*a*) *ka*=9.7066 and (*b*) *ka*=12.9751

site edges simply supported and the other two edges clamped, have been derived as Eqs. (7) and (9), Eqs. (12) and (14), and Eqs. (16) and (17), respectively. Based on the derived equations above, the modes shapes and natural frequencies can be obtained.

4 Numerical Illustrations

In this section, the different vibration mode functions of the square plate with two different edge conditions: (i) fully simply

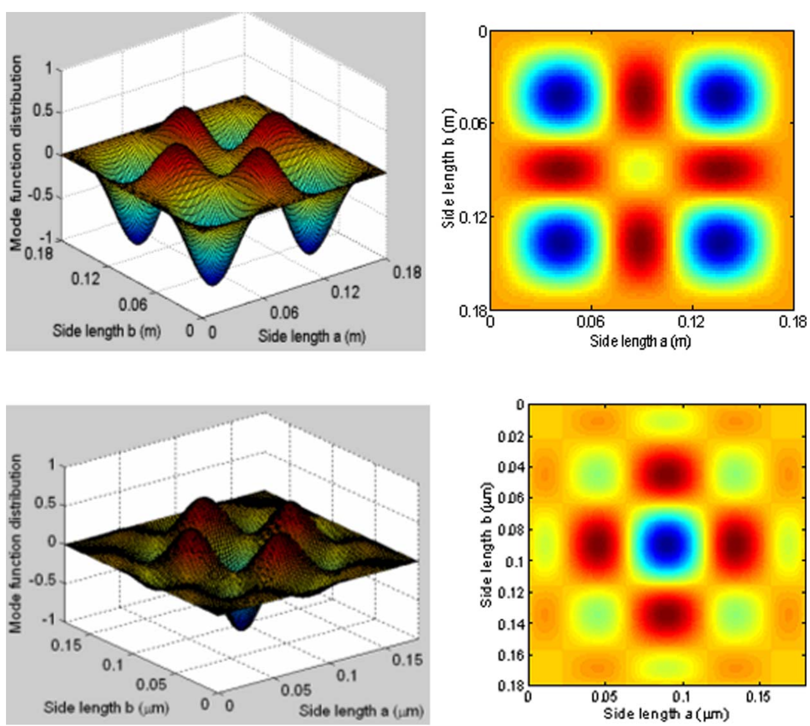


Fig. 3 Vibration mode functions with *n*=4 and *m*=2 for a fully clamped square plate at different nondimensional natural frequencies: (*a*) *ka*=12.4022 and (*b*) *ka*=15.5795

supported and (ii) fully clamped, are calculated with an edge length of $a=0.18$ m. Figures 1(a) and 1(b) show the vibration mode functions distribution and the corresponding contours for a fully simply supported rectangular plate with $n=2$ and $m=1$ at a nondimensional natural frequency of 3.6744 and 10.1215, respectively. As shown in Fig. 1(a), it can be seen that, at the lowest natural frequency $ka=3.6744$, the boundary conditions in Eq. (6) are satisfied from the mode distribution and Eq. (7) is validated. In the contour shown at the right-hand side, the only one peak is observed at the center of the square plate. In Fig. 1(b), it is observed that the number of peaks increases as the order of the vibration mode becomes higher. The corresponding contour at the right-hand side shows the distribution of the peaks on the plate under investigation. Figures 2(a) and 2(b) show the vibration mode functions distribution and the corresponding contours for a fully simply supported rectangular plate with $n=6$ and $m=3$ at a nondimensional natural frequency of 9.7066 and 12.9751, respectively. As shown in Fig. 2, when n and m vary, the boundary conditions in Eq. (6) are also satisfied and Eq. (7) is validated. Figures 3(a) and 3(b) show the vibration mode functions and the corresponding contours for a fully clamped rectangular plate with $n=4$ and $m=2$ at a nondimensional natural frequency of 12.4022 and 15.5795, respectively. As can be seen in Fig.3, the boundary conditions are satisfied by employing the Bessel function method and the derived equations are validated.

5 Conclusions

A novel Bessel function method is presented and used to obtain the exact solutions for the free-vibration analysis of a rectangular plate with three different edge conditions: (i) fully simply supported, (ii) fully clamped, and (iii) two opposite edges simply supported and the other two edges clamped. This proposed method provides the exact solutions for the natural frequencies and mode shapes of a rectangular plate. Because of the high accuracy provided by the proposed method, it can be used to verify other free-vibration analyses and to evaluate the precision of commercial software. The direct exact solutions obtained for the most fundamental structural element employing the proposed method will serve as a base of and provide an insight into the analysis of complex structures.

Acknowledgment

The authors would express our great gratitude to Professor Oliver M. O'Reilly, Associate Editor of the *Journal of Applied Mechanics*, and the anonymous reviewers for their useful suggestions.

References

[1] Navier, C. L. M. H., 1823, "Extrait des recherches sur la flexion des plans elastiques," Bull. Sci. Soc. Philomarhique de Paris, **5**, pp. 95–102.
[2] Levy, M., 1899, "Sur L'equilibrie Elastique D'une Plaque Rectangulaire," C. R. Acad. Sci., **129**, pp. 535–539.
[3] Way, S., 1934, "Bending of Circular Plates With Large Deflections," ASME J. Appl. Mech., **56**, pp. 627–636.
[4] Leissa, A. W., 1973, "The Free Vibration of Rectangular Plates," J. Sound Vib., **31**, pp. 257–293.
[5] Gorman, D. J., 1976, "Free Vibration Analysis of Cantilever Plates by the Method of Super-Position," J. Sound Vib., **49**, pp. 453–467.
[6] Gorman, D. J., 1977, "Free-Vibration Analysis of Rectangular Plates With Clamped-Simply Supported Edge Conditions by the Method of Superposition," ASME J. Appl. Mech., **44**, pp. 743–749.
[7] Pan, E., 2001, "Exact Solution for Simply Supported and Multilayered Magneto-Electro-Elastic Plates," ASME J. Appl. Mech., **68**, pp. 608–618.
[8] Cheng, Z.-Q., and Reddy, J. N., 2003, "Green's Functions for Infinite and Semi-Infinite Anisotropic Thin Plates," ASME J. Appl. Mech., **70**, pp. 260–267.
[9] Tong, P., and Huang, W., 2002, "Large Deflection of Thin Plates in Pressure Sensor Applications," ASME J. Appl. Mech., **69**, pp. 785–789.
[10] Narita, Y., 2000, "Combinations for the Free-Vibration Behaviors of Anisotropic Rectangular Plates Under General Edge Conditions," ASME J. Appl. Mech., **67**, pp. 568–573.
[11] Ventsel, E., and Krauthammer, T., 2001, *Thin Plates and Shells Theory, Analysis, and Applications*, Marcel Dekker, New York, pp. 284–285.
[12] Wang, Z. X., and Guo, D. R., 1989, *Special Functions*, World Scientific, Singapore, pp. 345–455.
[13] Zheng, Z., 1980, *Mechanical Vibration*, Mechanical Industry Press, Beijing.
[14] Young, D., 1950, "Vibration of Rectangular Plates by the Ritz Method," ASME J. Appl. Mech., **17**, pp. 448–453.
[15] Kerboua, Y., Lakis, A. A., Thomasb, M., and Marcouiller, L., 2007, "Hybrid Method for Vibration Analysis of Rectangular Plates," Nucl. Eng. Des., **237**, pp. 791–801.

Airflow-Housing-Induced Resonances of Rotating Optical Disks

R. M. C. Mestrom
e-mail: r.m.c.mestrom@tue.nl

R. H. B. Fey

H. Nijmeijer

Department of Mechanical Engineering,
Dynamics and Control Group,
P.O. Box 513,
Eindhoven University of Technology,
5600 MB Eindhoven,
the Netherlands

P. M. R. Wortelboer

W. Aerts

Emerging Technologies and Systems,
Philips Optical Storage,
P.O. Box 80002,
5600 JB Eindhoven,
the Netherlands

Numerous excitation sources for disk vibrations are present in optical drives. For increasing rotation speeds, airflow-housing-induced vibrations have become more and more important. Currently, drives are designed in which rotation speeds are so high that critical speed resonances may show up. The presence of these resonances depends on the layout of the inner housing geometry of the drive. The influence of the drive inner housing geometry is investigated systematically by means of a numerical-experimental approach. An analytical model is derived, containing disk dynamics and the geometry-induced pressure distribution acting as the excitation mechanism on the disk. The Reynolds' lubrication equation is used as a first approach for the modeling of the pressure distribution. The model is numerically implemented using an approach based on a combination of finite element and finite difference techniques. An idealized, drive-like environment serves as the experimental setup. This setup resembles the situation in the numerical model, in order to be able to verify the numerical model. Wedge-like airflow disturbances are used in order to obtain insight into the influence of drive inner geometry on the critical speed resonances of optical disks. A disk tilt measurement method is designed that yields a global view of the disk deformation. By means of two newly proposed types of plots, numerical and experimental results can be compared in a straightforward way. A qualitative match between the numerical and experimental results is obtained. The numerical and experimental methods presented provide insight into airflow-housing-induced vibrations in optical drives. Additionally, reduction of some critical speed resonances is found to be possible for certain drive inner geometry configurations. [DOI: 10.1115/1.2745356]

Keywords: rotating disk dynamics, disk modes, airflow-induced vibration, critical speed resonance, drive inner housing geometry

1 Introduction

In optical data storage, polycarbonate disks are used as the medium for data recording. Since these disks are relatively thin, they are weak in the transverse (out-of-plane) direction, which makes them susceptible to transverse vibrations. Numerous excitation mechanisms for these vibrations are present in optical drives; for instance, imbalance, shocks, and suspension flexibility. Due to increasing demands on the read-out speeds of the disk, airflow-induced vibrations have become more and more important due to higher rotation speeds. For example, an optical pickup unit (OPU) operating close to the rotating disk acts as a local excitation due to the air that has to flow through the narrow opening between the OPU and the disk. As a result, disk modes are always excited to a certain extent. Currently, rotation speeds in optical storage devices are such that critical speed resonances of optical disks can show up, resulting in large tilt in the disk in both the radial and tangential/circumferential directions. This phenomenon was discovered during the writing process of DVDs, which is performed using a constant data rate. During writing, the rotation frequency decreases with radial position on the disk. At certain radial positions, the number of writing errors exceeded the maximum allowable value. Rotation frequencies at which this occurred have been found to correspond with critical speeds of the disk.

In the field of rotating disk dynamics, extensive research has been performed. The study on the vibrations of spinning disks was first reported in Lamb and Southwell [1] and Southwell [2]. Since then, much research on this topic has been performed. Among the more recent investigations has been d'Angelo [3], who has shown that clamping of the disk is important in terms of stiction and slip. Furthermore, the clamping radius is found to have a large influence on the natural frequencies. Chen and Bogy [4] have investigated the influence of system parameters on the eigenvalues of the system. Malhotra et al. [5] have derived the equations of motion for a rotating disk, containing both bending and membrane (stretch) effects. Several methods are described to approximate the response of the disk. Chung et al. [6] have analyzed the free vibrations of a spinning disk. Using a Galerkin method, approximations for the natural frequencies, mode shapes, and critical speeds of a freely spinning disk have been obtained. Furthermore, their dependency on the rotation frequency (due to stretching of the disk) has been investigated. Jia [7] has derived the equations of motion for a spinning disk, using energy considerations (Hamilton's principle). The effect of centrifugal flattening is explained, which causes a decrease of the initial disk warpage as the rotation speed increases. Lee et al. [8] have numerically predicted critical speeds of optical disks and have compared these with experimental results, obtaining a good match. Furthermore, the effect of flutter (self-excited, aerodynamically induced disk vibrations) has been discussed and is measured in the experiments. Chung et al. [9,10] have analyzed the dynamics of a rotating disk with angular acceleration and eccentricity, respectively. Stationary in-plane motion is assumed, whereas transverse motion is assumed to be dynamic. The transverse motion is obtained by using Galerkin techniques. Eccentricity is found to result in distortion of mode shapes and an increase in critical speed for the lowest mode. Heo and Chung [11] have performed a similar analysis for angular misalignment (rigid-body tilt). This has been found to result in a

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received April 24, 2006; final manuscript received March 16, 2007. Review conducted by Igor Mezic.

"beating" phenomenon both for the in-plane and transverse disk response. Lee et al. [12] have experimentally determined that critical speeds for spinning disks are higher in vacuum situations, caused by lack of added mass effects of the air surrounding the disk.

Interaction between a rotating disk and a stationary load system or airflow has been reported in the following references. Pelech and Shapiro [13] have considered a flexible disk rotating on a gas film close to a wall. The air film height is extremely small with respect to the disk radius, resulting in very small Reynolds numbers. Benson and Bogy [14] have discussed the steady deflection of a very flexible spinning disk due to a stationary fixed transverse load. Transverse stiffness has to be included in the analysis as membrane theory has proven to be unable to solve this problem. Adams [15] has analyzed the interaction between a flexible disk and a read/write head by assuming a steady, axisymmetric fluid flow in the interaction region. The disk and fluid equations are solved separately and then combined. Licari and King [16] have described the development of a numerical model (finite element method (FEM) combined with finite differences) to simulate the interaction between a magnetic recording head and a rotating flexible disk. The Reynolds equation is used for modeling the head-disk interaction. Carpino and Domoto [17] also have investigated a rotating disk near a flat plate. The Reynolds equation is used to model the incompressible and laminar airflow. The solution to the coupled partial differential equations is found from an axisymmetric part and a linearized nonaxisymmetric part. The total solution is found by combining these two parts. Adams [18] has considered an elastic disk subjected to a point load and rotating close to a stationary baseplate. Four different airflow models have been considered, including the Reynolds equation. For each disk mode, the airflow is accounted for by a stiffness and damping parameter. Kim et al. [19] have considered a disk, rotating in a cartridge. Computational fluid dynamics (CFD) calculations show that objects, present in the airflow, considerably affect the velocity field of the airflow in the cartridge. The pressure distribution acting on the disk has been found to be closely related to the inner shape of the cartridge. Tatewaki et al. [20] have performed numerical simulations of unsteady airflow in hard disk drives. The pressure-time series obtained in this way has been applied on a FEM model of the disks. The presence of the read/write head is found to have a large influence on the response. Naganathan et al. [21] and Bajaj et al. [22] have performed a numerical study of a flexible disk, rotating near a rigid wall. Based on Pinkus and Lund [23], the full Navier-Stokes equations for the airflow are simplified and the Reynolds equation for cylindrical geometry is obtained. The coupled disk-fluid partial differential equations are discretized using finite differences and are solved directly. Self-excited vibrations are found to occur due to coupling between dynamics of the disk and hydrodynamics of the fluid.

Changes in the inner housing geometry of drives have been found to influence the presence of critical speed resonances, due to changes in airflow. In this paper, the influence of drive inner housing geometry on the critical speed resonances in optical disks is investigated both numerically and experimentally in a systematic way. For this purpose, a measurement setup is designed, which consists of an idealized drive containing simplified inner housing geometry. Furthermore, a new method is developed to measure disk deformation in the experimental setup. Moreover, a numerical model is developed, which describes the effect of airflow-housing-induced excitation on a rotating optical disk. Finally, new methods for presenting both the numerical and experimental results are proposed by means of two new types of plots: so-called avalanche plots and maximum absolute tilt plots. As an additional result of the systematic investigation of the influence of inner housing geometry, some reduction of critical speed resonances is found to be possible for certain inner housing geometries.

The outline for this paper is as follows. In Sec. 2, some theoretical background on disk dynamics will be given. A model for the disk and the airflow will be presented in Sec. 3, together with the underlying assumptions. Section 4 will describe the numerical implementation of this model. Section 5 will provide a description of the experimental setup and the measurement method. The interpretation of the results and the comparison between the simulations and experiments will be made in Sec. 6. Finally, in Sec. 7, conclusions will be presented.

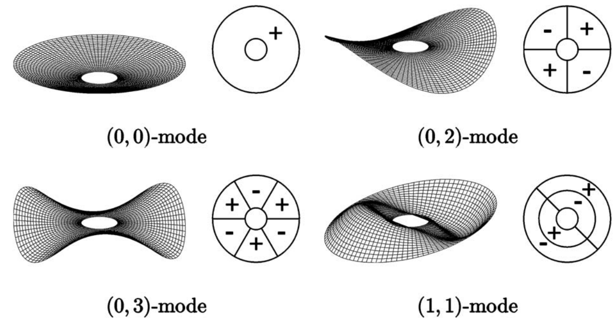


Fig. 1 Some examples of disk modes

2 Theoretical Background

The dynamics of disks are characterized by a number of vibration forms, also called disk modes, and their corresponding natural frequencies. As a disk is a continuum, infinitely many disk modes exist, with distinct natural frequencies. Only modes with low natural frequency (say, below 1000 Hz) are of practical importance for this research. As a result, only transverse (out-of-plane) disk modes are considered because in-plane modes have high natural frequencies (>1500 Hz). The transverse disk modes are denoted by (m,n), where m is the number of nodal circles and n is the number of nodal diameters. Some examples of disk modes are given in Fig. 1. The $(0,0)$ mode is also known as the umbrella mode and the $(0,2)$ mode is called a saddle mode.

For a rotating disk, two additional effects show up:

- First, rotation of the disk causes a build-up of radial stress in the disk. Since this results in stretching of the midplane of the disk, this is called the stretch effect. Midplane stretching causes an increased disk stiffness. Therefore, the natural frequency of each mode will increase with the rotation frequency Ω.
- Second, the rotating disk is observed by a non-corotating (Earth-fixed) observer (for instance, the lense of the OPU). As a result, each (m,n) mode $(n>0)$ splits into a forward traveling and backward traveling wave (see [1,2]). This is called mode splitting.

The stretch effect and mode splitting are depicted in Fig. 2, a so-called Campbell plot, for an (m,n) mode $(n>0)$. Here, the stretch effect is seen for both the body-fixed and the Earth-fixed observer. In the former case, the curve has a positive slope for increasing rotation frequency Ω. In the latter case, mode splitting results in a forward and backward traveling wave.

For a certain rotation frequency, the so-called critical speed, the backward traveling wave reaches zero natural frequency. At the critical speed, a constant (non-time-varying), Earth-fixed excitation can bring the disk into a critical speed resonance (resonance with zero natural frequency). The disk attains a stationary (nonrotating) deformed shape, which is dominated by the mode shape that has its critical speed at this rotation frequency.

In an optical drive, the inner geometry forms an excitation source. As the disk rotates, it generates airflow in the drive. Due to airflow over the drive inner geometry, pressure differences are generated in the drive. These pressure differences depend on the

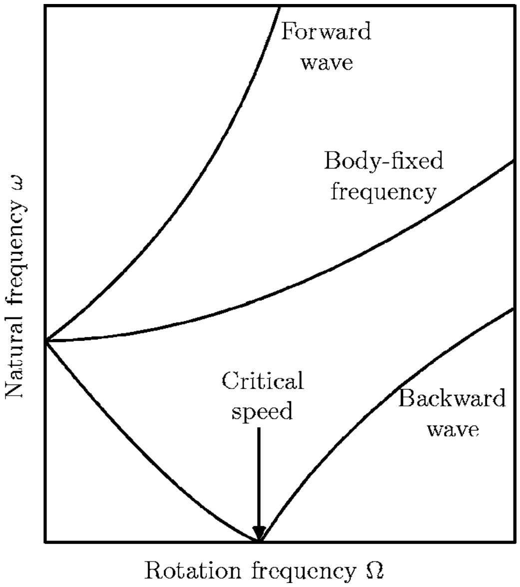


Fig. 2 Campbell plot: natural frequencies in the body-fixed and Earth-fixed frames as a function of the rotation frequency

airflow, but for a constant rotation frequency Ω, they will be Earth fixed and more or less of constant nature. Namely, due to instationarities in the airflow and the disk vibrations, they will fluctuate around some constant value. As a result, a more or less constant pressure distribution acts on the disk. Since this excitation is both Earth fixed and constant, it is a mechanism for exciting critical speed resonances.

3 Analytical Model

3.1 Disk Model. Consider a circular annular disk, clamped at a radius r_c and free at its outer radius r_o, and rotating with angular velocity Ω (see Fig. 3). The geometry of the disk is described by radial and circumferential coordinates r and ϕ, respectively. It rotates at a distance $s(r,\phi,t)$ above a rigid, fixed baseplate, and a transverse pressure distribution $p(r,\phi,t)$ is assumed to act on it. The transverse deflection of the disk is $w(r,\phi,t)$.

The assumptions underlying the analytical model for the disk can be summarized as follows:

(1) the disk is made of an isotropic, homogeneous material;
(2) the material is linearly elastic, such that Hooke's law holds;
(3) the density ρ, Young's modulus E, and the Poisson ratio ν are constant throughout the disk and over time;
(4) the disk is flat and has constant thickness h; hence, it contains no imbalance or warpage;

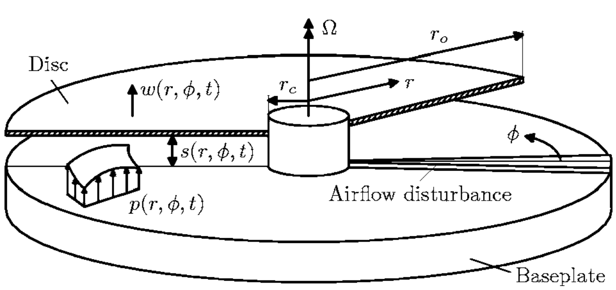


Fig. 3 Schematic representation of a flexible disk rotating above a rigid, fixed baseplate

(5) the disk is thin ($h \ll r_o$), such that the Kirchoff plate theory (see [24]) holds; this theory contains the following assumptions:
  (a) Straight lines, perpendicular to the midsurface (transverse normals) before deformation remain straight after deformation;
  (b) Transverse normals do not experience elongation (they are assumed to be inextensible);
  (c) Transverse normals rotate such that they remain perpendicular to the midsurface after deformation;
(6) in-plane displacement is assumed to be axisymmetric, stationary, and much smaller than the transverse displacement w;
(7) the disk rotates with a constant angular velocity Ω; hence, $\dot{\Omega}=0$;
(8) rotatory (in-plane) inertia is neglected;
(9) thermal effects are not taken into account.

With these assumptions, a linear model for the transverse deflection of the disk can be obtained (see for example [5,6,9,10,21,22]). The equation for the transverse deflection w, containing both membrane and bending stiffness, is given for an Earth-fixed observer by:

$$\rho h\left(\frac{\partial^2 w}{\partial t^2}+2\Omega\frac{\partial^2 w}{\partial t\,\partial\phi}+\Omega^2\frac{\partial^2 w}{\partial\phi^2}\right)+D\nabla^4 w-\frac{\partial}{r\,\partial r}\left(rq_{rr}\frac{\partial w}{\partial r}\right)-\frac{\partial}{r\,\partial\phi}\left(q_{\phi\phi}\frac{\partial w}{r\,\partial\phi}\right)=p \tag{1}$$

where D is the bending rigidity of the disk and ∇^4 is the biharmonic operator:

$$D=\frac{Eh^3}{12(1-\nu^2)}\qquad \nabla^4=\left(\frac{\partial^2}{\partial r^2}+\frac{\partial}{r\,\partial r}+\frac{\partial^2}{r^2\,\partial\phi^2}\right)^2 \tag{2}$$

and q_{rr} and $q_{\phi\phi}$ are, respectively, the radial and tangential internal forces per unit length due to the centrifugal action of rotation:

$$\begin{aligned} q_{rr}&=-\frac{\rho\Omega^2 h}{8}\left[(3+\nu)r^2-C_1+\frac{C_2}{r^2}\right]\\ q_{\phi\phi}&=-\frac{\rho\Omega^2 h}{8}\left[(1+3\nu)r^2-C_1-\frac{C_2}{r^2}\right]\end{aligned} \tag{3}$$

with the constants C_1 and C_2 given by:

$$\begin{aligned} C_1&=(1+\nu)\frac{(3+\nu)r_o^4+(1-\nu)r_c^4}{(1+\nu)r_o^2+(1-\nu)r_c^2}\\ C_2&=(1-\nu)r_c^2r_o^2\frac{(1+\nu)r_c^2-(3+\nu)r_o^2}{(1+\nu)r_o^2+(1-\nu)r_c^2}\end{aligned} \tag{4}$$

The disk is clamped at the clamping radius, requiring zero transverse displacement and slope at r_c, and free at the outer radius, yielding zero edge reaction and zero bending moment at radius r_o. This results in the following four boundary conditions associated with Eq. (1):

$$w=0\quad \frac{\partial w}{\partial r}=0\quad \text{at } r=r_c$$

$$m_{rr}=0-D\frac{\partial}{\partial r}(\nabla^2 w)+\frac{\partial m_{r\phi}}{r\,\partial\phi}=0\quad \text{at } r=r_o \tag{5}$$

where

$$m_{rr} = -D\left[\frac{\partial^2 w}{\partial r^2} + \nu\left(\frac{\partial w}{r\,\partial r} + \frac{\partial^2 w}{r^2\,\partial \phi^2}\right)\right]$$

$$m_{r\phi} = -(1-\nu)D\left(\frac{\partial^2 w}{r\,\partial r\,\partial \phi} - \frac{\partial w}{r^2\,\partial \phi}\right) \tag{6}$$

3.2 Airflow Model. The excitation mechanism acting on the disk is provided by the transverse pressure distribution $p = p(r,\phi,t)$ on the right-hand side of Eq. (1). This pressure distribution is calculated from a model for the airflow over/through the drive inner geometry. As the goal of this research is to obtain general trends for geometry influence, the airflow model should preferably be simple and intuitive. Consequently, the use of highly detailed models, which can only be calculated by time-consuming CFD methods, is not considered. Namely, the interest does not lie in the detailed flow patterns and velocity field of the air in the drive, but in the global excitation mechanism (pressure distribution) caused by the airflow.

An airflow model that satisfies these requirements is a lubrication approximation (the Reynolds equation) for the air films below and above the disk [21,22]. The Reynolds equation is often used in problems where there is no real lubrication, but where mainly pressure calculations are important. Therefore, it provides a common first step as a model for the airflow to obtain an estimate of the pressure distribution resulting from the geometry. Turbulence and convective terms, describing fluid inertance effects, are neglected in this approach. Inclusion of these terms would result in vortices in the fluid and, due to fluid viscosity, in higher pressure differences. Hence, without these terms, a lower estimate of the pressure distribution will be obtained.

The airflow between the disk and the baseplate (see Fig. 3) is modeled by the Reynolds equation for circular geometry, resulting from simplification of the Navier-Stokes equations. The following simplifications are used:

(1) the fluid is Newtonian;
(2) the flow is laminar and incompressible;
(3) the fluid has constant viscosity μ and density ρ_a;
(4) thermal effects are negligible;
(5) all fluid internal forces, except for the centrifugal force, are negligible compared to the viscous forces;
(6) the fluid film thickness is small compared to the diameter of the disk; pressure variations across the film (in thickness direction) are neglected.

The Reynolds lubrication equation is derived by using the continuity equation, together with the simplified Navier-Stokes equations. A detailed explanation and derivation can be found in [23] for a general case, and in [21] for a flexible disk rotating over a rigid baseplate. For an Earth-fixed description, this results in the following lubrication equation:

$$\begin{aligned}\frac{\partial}{\partial r}\left(\frac{rs^3}{\mu}\frac{\partial p}{\partial r}\right) + \frac{\partial}{r\,\partial \phi}\left(\frac{s^3}{\mu}\frac{\partial p}{\partial \phi}\right) &= 6r\Omega\frac{\partial s}{\partial \phi} + 12r\frac{\partial s}{\partial t} + \frac{3\rho_a\Omega^2}{10\mu}\frac{\partial}{\partial r}(r^2 s^3) \\ &\quad - \frac{\rho_a}{10\mu^3}\frac{\partial}{\partial r}\left[\mu\Omega s^5\frac{\partial p}{\partial \phi} - \frac{3}{28}\frac{s^7}{r^2}\left(\frac{\partial p}{\partial \phi}\right)^2\right]\end{aligned} \tag{7}$$

where $s = s(r,\phi,t)$ is the film thickness depending on the geometry of the baseplate and the disk shape (see Fig. 3). Centrifugal effects in the fluid are included in the last two groups of terms on the right-hand side. The first group represents the effects induced by shear, whereas the second group contains the effects caused by the circumferential pressure gradient.

3.3 Combined Model. As will become clear in Sec. 5, the disk rotates in a cylindrical cavity, which means that rigid baseplates are located above as well as below the disk. Additionally, walls are present in the radial direction. This is depicted in Fig. 4, where s_u and s_l denote the nominal gap heights above and below the disk, respectively. $p_u(r,\phi,t)$ and $p_l(r,\phi,t)$ are the pressures resulting from airflow over airflow disturbances in the upper and lower cavities, respectively. For both cavities, Eq. (7) has to be solved and their pressure difference acts as an excitation on the disk (Eq. (1)). The presence of walls in the radial direction is accounted for by specifying boundary conditions at points p_1 through p_4 in Fig. 4. At p_1, zero ambient pressure is assumed, whereas zero pressure gradient $\partial p/\partial r = 0$ is assumed at p_2. Furthermore, airflow from the lower to the upper cavity is assumed not to take place. Therefore, pressure p_3 is assumed to equal p_2. Zero pressure gradient is once more assumed at p_4.

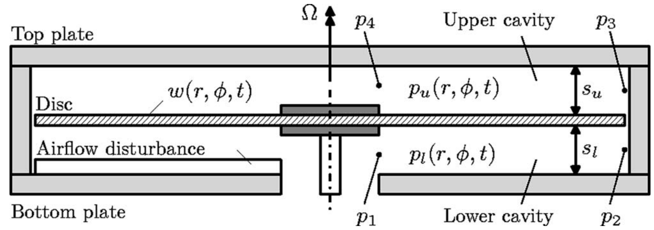


Fig. 4 Schematic situation for the combined model

4 Numerical Model

The model for the disk and the airflow in the cavities above and below it consists of two distinct parts, i.e., Eqs. (1) and (7), between which the interaction takes place. As a result, it forms a multiphysics problem, similar to problems found in hard disk research (see [20]), where interaction takes place between the spinning disks, the read-write head, and airflow in the drive. Here, the two parts of the problem are calculated separately, and coupled afterwards, in an iterative way. In order to perform calculations for a variety of inner housing geometry layouts, a mesh-like description of both the disk and the pressure field is preferred. Therefore, the disk part of the analytical model (Eq. (1)) is approximated by FEM techniques and the Reynolds equation (Eq. (7)) is numerically implemented using finite difference techniques.

For the disk, the FEM package ANSYS [25] is used. Elastic, four-node shell elements are used to mesh the disk. The node grid contains 25 equidistant nodes in radial direction and 100 equidistant nodes in circumferential direction. The nodes at the clamping radius r_c are fully constrained in order to approximate the clamping condition of the disk. In the FEM model, for each rotation frequency, two subsequent analyses are performed. First, a static analysis is performed, in which the disk is prestressed due to midplane stretching as a result of the centrifugal load caused by rotation. Next, a modal analysis (Block-Lanczos solution procedure) of this prestressed state is performed, in which the natural frequencies and the transverse mode shapes for a number of disk modes are calculated. This corresponds to a modal solution of Eq. (1) in body-fixed coordinates. In this way, body-fixed natural frequencies are obtained (see Fig. 2). After some initial calculations, a quantity of 26 modes is considered to be sufficient to describe the disk dynamics for this research. This corresponds with the (0,0)–(0,8) mode and the (1,0)–(1,4) mode, covering a body-fixed frequency range of up to 2500 Hz. In this way, a description of the disk in terms of modal coordinates is obtained, which is exported to the numerical programming package Matlab® [26]. Additionally, modal damping is added to each mode separately, such that transients will decay during time simulations. The dimensionless damping coefficients are set to 0.005 for each mode, which will become clear from Sec. 6.1.

The lubrication equation (Eq. (7)) is implemented in Matlab® using finite difference discretization. As this discretization scheme is the same as used in [21,22], it is not discussed in detail here. However, problems are experienced with the discretization of the mixed derivative term and the nonlinear term (terms in [···] in Eq. (7)). These terms cannot be discretized by means of finite differences in a stable way, as found from a von Neumann stability analysis (see for example, [27]). Therefore, these terms are omitted in further analysis (as is also the case in [23]). As a result, the Reynolds equation without these terms is considered to be a first step in the modeling of the airflow below and above the disk.

The total numerical model consists of a combination of the implementations of the disk equations and the Reynolds equation. These two implementations are coupled iteratively (which will be explained below) and integrated in time to determine the dynamic response. As the modal description of the disk is valid for a body-fixed coordinate system, the disk in the model does not rotate physically, but contains a prestressed state, corresponding to a certain rotation frequency. Consequently, the pressure distribution has to rotate over the disk in order to obtain the same dynamic situation as for a rotating disk and an Earth-fixed load. Therefore, the mesh for the pressure distribution has to rotate relative to the disk mesh. To avoid computationally costly interpolation, time steps in the algorithm are coupled to the mesh size and the rotation frequency. This suggests fixed time steps and, therefore, second-order implicit time integration (trapezoid method, see [28]) is implemented.

The two-sided interaction between the disk model and the pressure equation is implemented as follows. After a number of complete revolutions of the pressure distribution over the nonrotating disk, the film thicknesses of the fluid films above and below the disk are calculated. Evaluation after each time increment would be computationally too costly. After some simulations with the complete model, it is found that the calculated disk deflection is too small to influence the pressure distributions above and below the disk significantly. As a result, the pressure distribution is calculated only at the beginning of each simulation (based on an undeformed disk) and no further coupling between the disk equations and the pressure equations takes place.

Parameter values for the numerical model are found from [29] for the polycarbonate disk, and from [30] for the airflow part (ambient temperature 20°C) and equal: $E=2.7\times10^9$ Pa, $\nu=0.33$, $\rho=1200$ kg/m^3, $h=1.2$ mm, $r_c=12.75$ mm, $r_o=60$ mm, $\rho_a=1.23$ kg/m^3, and $\mu=18.6\times10^{-6}$ Pa s.

5 Experimental Approach

5.1 Experimental Setup. In order to be able to gain insight into the general influence of inner housing geometry on critical speed resonances, no real optical drive is used, since this forms a very complicated system. Instead, an idealized drive is considered, which consists of an aluminum base, to which the drive mechanism is rigidly mounted. The stator part of the motor is mounted to the bottom plate of the housing and a disk is clamped, by means of a magnetic clamper, to the turntable, fixed to the rotor part of the motor. The disk is a polycarbonate CD, with a highly reflective coating on both sides, which is beneficial for the measurement method. Several of the described parts can be seen in Fig. 5.

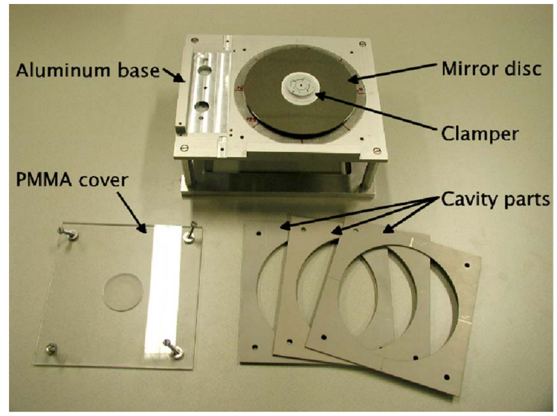


Fig. 5 Picture with parts of the idealized drive

As a normal optical drive forms a more or less air-tight unit, the idealized drive should also be more or less air tight. Therefore, around the disk, a cylindrical cavity is present, built from aluminum cavity parts. The cavity is covered by a transparent polymethyl methacrylate (PMMA) top plate, to enable measurement of disk tilt (see Sec. 5.2). The only hole in the setup is a hole in the bottom plate, through which the turntable enters the cavity. A schematic overview of the idealized drive is given in Fig. 6. This situation is identical to the analytical situation in Fig. 4. The dimensions denoted in Fig. 6 are: $r_h=15$ mm, $r_w=60.5$ mm, $s_u=3.3$ mm, and $s_l=5.5$ mm. This means that the cavities above and below the disk do not have the same height, which is also the case in a real optical drive. As an excitation source for critical speed resonances, simple wedge-like geometrical airflow disturbances are placed on the baseplate below the disk. These disturbances have a radial dimension ranging from r_h to r_o, a tangential dimension of 10 deg, and a height of 2 mm.

5.2 Measurement Method. In order to be able to measure radial and tangential tilt of the disk (and, indirectly, its transverse deflection), a measurement method is designed, which yields a global view of the disk deformation. The method is a projection method, in which a pattern, reflected in a disk, is measured by means of a camera (see Fig. 7(*a*)). As the disk deforms, a change of the measured pattern can be observed due to a local change of the angle of the reflective disk (see Fig. 7(*b*)). As a result, small variations in the disk shape can result in large variations in the pattern reflected on the disk. The pattern, consisting of white dots on a black background, permits measurement analysis by means of image processing software. Examples of a reference and "deformed" camera frame can be seen in Fig. 8. Actually, averaging is performed over 100 high speed camera frames, captured at

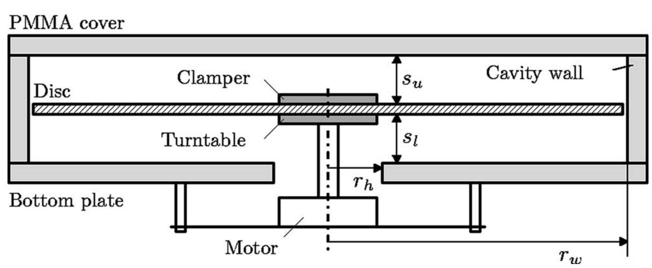


Fig. 6 Schematic overview of the idealized drive

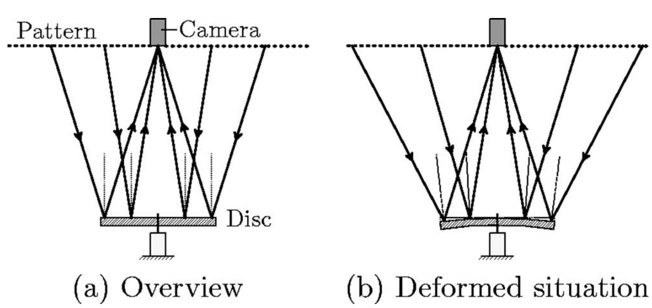
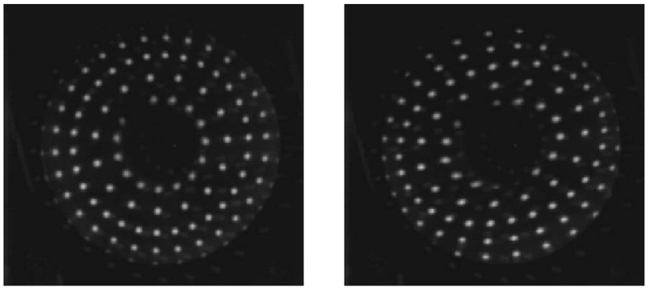


Fig. 7 Schematic overview of the projection method



(a) Reference frame (b) ‘Deformed’ frame

Fig. 8 Examples of the camera view

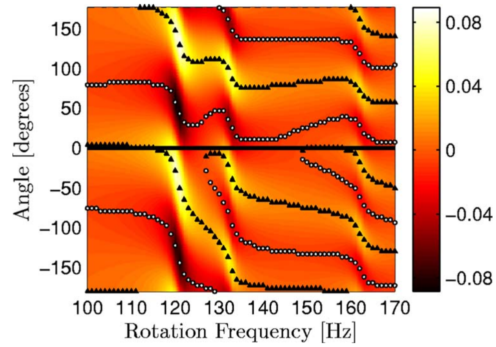


Fig. 10 Example of an avalanche plot: $\alpha_t(\phi,f)$ [mrad] (maxima: ▲; minima: ∘)

1000 fps, in order to reduce the effect of disk imperfections and reflection variations. The situation in Fig. 8(*b*) corresponds to a saddle mode of the disk.

Using image processing routines written in Matlab®, local radial and tangential tilt values of the disk can be calculated from the camera frames. Next, the disk shape that approximately matched these tilt values is determined by least squares techniques using the radial and tangential tilt of the mode shapes from the FEM model of the disk (see Sec. 4). Even with only 26 modes included, residues of the fit are always lower than 10%, indicating that a nice match is obtained. Some residues will always be present, as the disk used in the experimental setup will contain effects that cannot properly be described by a finite number of mode shapes, like, for instance, warpage. The residue of 10% translates to approximately ±1 mrad accuracy in the measured tilt values.

6 Results

In order to gain insight into the influence of drive inner housing geometry on the presence of critical speed resonances in optical drives, first some initial results are discussed, mainly based on the simulation model. Next, simulations and experiments are compared.

6.1 Exploratory Simulations and Presentation of Results. In order to determine at which rotation frequencies critical speed resonances of the polycarbonate disk can occur, a Campbell plot is constructed from the natural frequencies obtained from the FEM calculations of the disk. For rotation frequencies below 200 Hz, the natural frequencies of the forward and backward traveling waves are depicted in Fig. 9.

The maximum rotation frequency in current optical drives

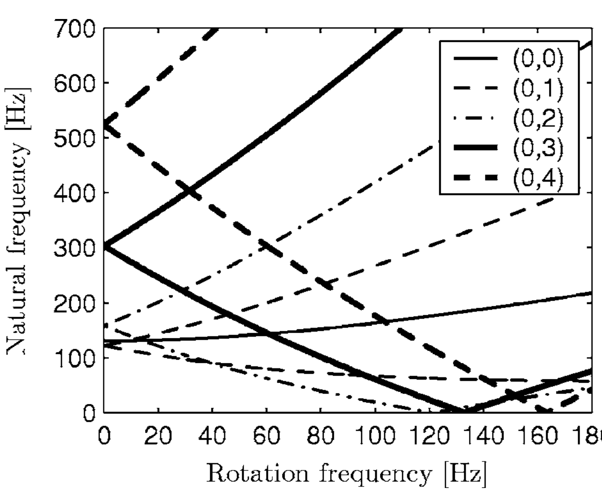


Fig. 9 Campbell plot (for an Earth-fixed observer)

equals about 160 Hz. Therefore, three modes are of practical interest, namely, the $(0,2)$, $(0,3)$, and $(0,4)$ modes, which have their critical speeds at 120.9, 133.1, and 163.4 Hz, respectively, as can be seen in Fig. 9. Below a rotation frequency of 100 Hz, no critical speeds are found. Therefore, the rotation frequency range of interest is set to $100 \leq f \leq 170$ Hz, with a resolution of $\Delta f = 1$ Hz.

From both the numerical model and the measurements, a steady-state Earth-fixed disk shape $w(r,\phi)$ can be obtained for each rotation frequency. In the simulation model, a time simulation is performed for each rotation frequency until transients have disappeared. In the measurements, the disk is spun to the desired rotation frequency. After approximately 10 s, the steady state is reached and a series camera frame is captured and averaged. By comparing this frame to a reference frame (undeformed disk) the disk shape can be reconstructed (see Sec. 5.2).

The results of a measurement or simulation series consist of the steady-state Earth-fixed disk shape $w(r,\phi)$ for the range of rotation frequencies f considered. This will be denoted as $w(r,\phi,f)$. Each steady-state disk shape contains contributions of all the mode shapes included in the model/fit. However, in the frequency range considered, the (m,n) modes with $m>0$ will hardly be excited by wedge-like airflow disturbances. Namely, the excitation does not match mode shapes with nodal circles. As a result, the maximum transverse deflection and the maximum radial and tangential tilt of each steady-state disk shape $w(r,\phi,f)$ are present at the outer rim of the disk. This means that the deflection at the outer rim r_o of the disk contains characteristic information of the disk shape. Hence, three-dimensional information on the disk shape can be reduced to a two-dimensional representation by considering the transverse deflection at the outer rim versus circumferential position on the disk; this implicates a reduction from $w(r_o,\phi,f) \mapsto w(\phi,f)$. For the radial and tangential tilt, similar representations are obtained, denoted by $\alpha_r(\phi,f)$ and $\alpha_t(\phi,f)$, respectively.

Next, a new way of representing the characteristics of a simulation series is proposed, consisting of a plot with the steady-state circumferential profile at the outer rim of the disk versus the rotation frequency. This profile could be the transverse deflection $w(\phi,f)$, radial tilt $\alpha_r(\phi,f)$, or tangential tilt $\alpha_t(\phi,f)$. Additionally, the circumferential location of local maxima and minima of the profile is indicated. The name *avalanche plot* is proposed as a name for this new type of plot.

An example of a tangential tilt avalanche plot $(\alpha_t(\phi,f))$ is shown in Fig. 10. The inner housing geometry for this simulation is shown in Fig. 11(*a*). The resulting pressure difference $\Delta p_{\text{dist}} = p_{\text{dist},l} - p_{\text{dist},u}$ is shown in Fig. 11(*b*) for $f=120$ Hz, where $p_{\text{dist},l}$ and $p_{\text{dist},u}$ denote the pressures in the lower and upper cavity, respectively.

The airflow disturbance is located at $\phi=0$ deg in the lower

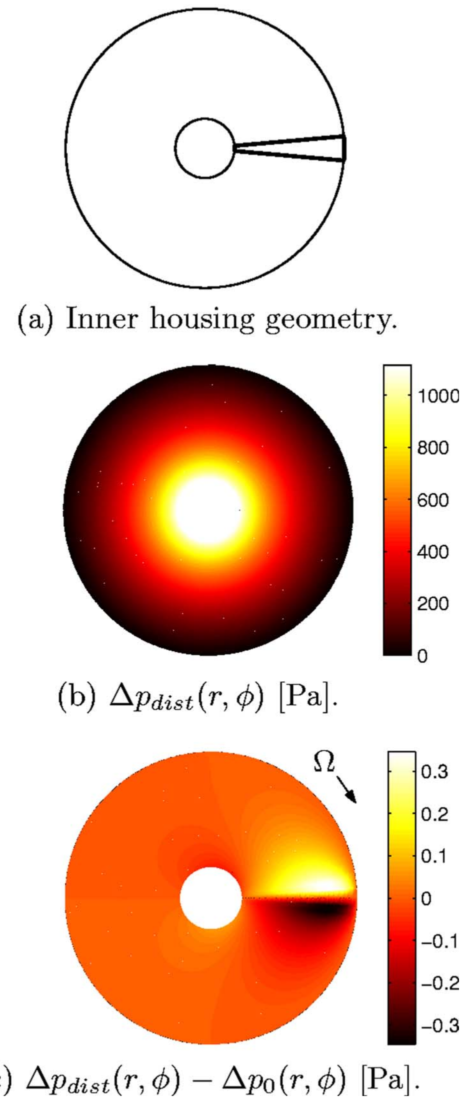


Fig. 11 Airflow disturbance configuration (*a*) and pressure difference (*b*) at a rotation frequency of f=120 Hz. The pressure difference between the situation with and without disturbance is depicted in (*c*).

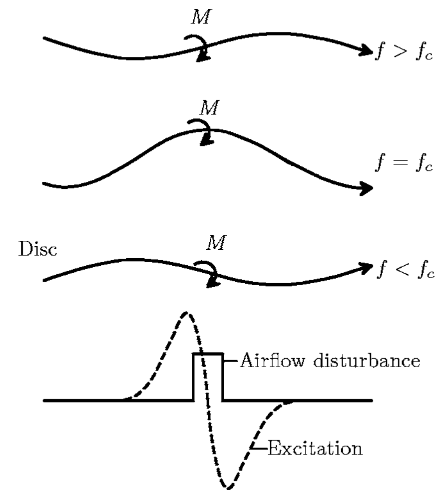


Fig. 12 Circumferential movement of the disk shape around the critical speed f_c

cavity and the disk rotates in a clockwise direction. Due to the large values of the pressure, the effect of the airflow disturbance cannot clearly be distinguished. Therefore, in Fig. 11(*c*), the pressure difference between a simulation with and without airflow disturbance is depicted, in which it can be seen that the airflow disturbance has a local influence on the pressure field.

Local maxima and minima of $\alpha_t(\phi,f)$ are indicated by black triangles and white circles in Fig. 10, respectively. Furthermore, the circumferential location of the airflow disturbance is indicated by a black solid line. From the figure, it can be seen that near three rotation frequencies, large tangential tilts occur. These frequencies correspond to the critical speeds from Fig. 9 and equal approximately 121, 133, and 163 Hz. From the number of local maxima and minima at these frequencies, it can be seen that they belong to the $(0,2)$, $(0,3)$, and $(0,4)$ modes, respectively. Note that new local maxima/minima appear between two critical speeds (around 125 Hz and 150 Hz in Fig. 10). They originate from a single nucleus and split up in a new pair of local minima and maxima. In this case, the nucleus is located at a circumferential position near the airflow disturbance.

Near the critical speeds, a movement of the local maxima and minima in circumferential direction can be observed. This means that the steady-state disk shape rotates in the ϕ direction relative to the airflow disturbance, located at ϕ=0 deg. To find an explanation for this rotation, a small part of the avalanche plot can be considered. Namely, consider rotation frequencies ranging from 155 Hz to 170 Hz, in which the critical speed of the $(0,4)$ mode (f_c=163 Hz) is contained. Below the critical speed, a local tangential tilt maximum is located approximately at the position of the airflow disturbance. Above the critical speed, a minimum is located at this position. This is the same effect as the 180 deg phase change of a single mass-spring mechanical system around its natural frequency. In order to understand this, Fig. 12 is considered.

The pressure difference, resulting from airflow over inner housing geometry, can be considered to act on the disk as a moment M. To see this, consider Fig. 11(*c*) where, in the clockwise direction, the excitation on the disk contains a maximum, followed by a minimum. For $f<f_c$, deformation and moment are in phase in a static way (stiffness determined, as for an ordinary single mass-spring mechanical system), as is depicted in Fig. 12. The out-of-phase situation ($f>f_c$) contains a disk slope of opposite sign; hence, the phase change in disk tilt equals 180 deg (see Fig. 12). At the critical speed $f=f_c$, the transverse displacement is largest and the tangential tilt at the location of the disturbance is near zero. For the critical speed resonance considered here (the $(0,4)$ mode), a change of sign of the disk slope corresponds with a rotation of the disk shape of 45 deg in the circumferential direction. This equals the circumferential movement of the local maxima and minima between 155 Hz and 170 Hz in Fig. 10.

Around the critical speeds of the $(0,2)$ and $(0,3)$ modes, similar rotations of the disk shape take place, with circumferential movements of 90 deg and 60 deg, respectively. However, since these critical speeds are located close to each other, some interference effects take place, making it difficult to see the rotations properly in the ϕ direction.

The size of the rotation frequency intervals over which the movement of the local maxima and minima takes place in the simulation model depends on the modal damping values in the numerical model. The damping values are adjusted such that the rotation frequency intervals in the simulations have similar length as the ones in the experiments. In this way, a dimensionless damping coefficient of 0.005, or 0.5%, is estimated for all modes. These damping values are used for all simulations in the remainder of this paper.

The information in the avalanche plot of Fig. 10 can be reduced one step further by calculating the maximum absolute radial or

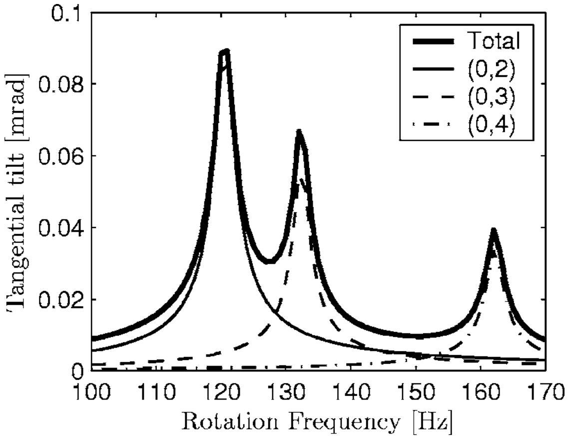


Fig. 13 Maximum absolute tangential tilt $\alpha_{t,\max}(f)$ versus rotation frequency. Individual contributions of the $(0,2)$, $(0,3)$, and $(0,4)$ mode are also indicated

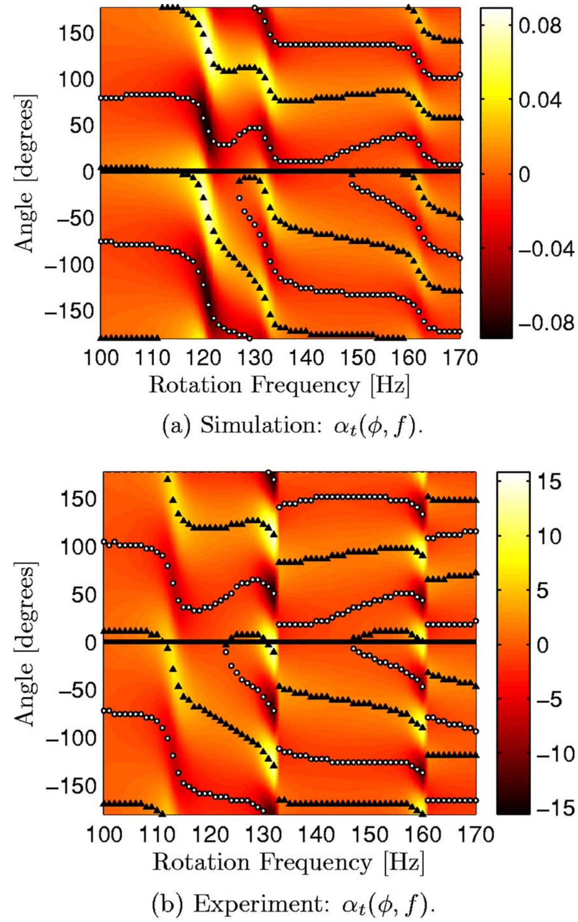


Fig. 14 Simulation and experimental avalanche plot for the tangential tilt $\alpha_t(\phi, f)$ of the reference configuration

tangential tilt on the outer rim as a function of the rotation frequency, corresponding with a reduction from $\alpha_r(\phi,f) \mapsto \alpha_{r,\max}(f)$ or $\alpha_t(\phi,f) \mapsto \alpha_{t,\max}(f)$. Figure 13 gives an example for the tangential tilt. This type of plot will be called a maximum absolute tilt plot and is especially useful for comparing the effect of different inner housing geometries (see also Sec. 6.2). Additionally, in Fig. 13, the contribution of three individual modes is indicated, where it can be seen that, at the critical speeds, the response is dominated by the mode that is in critical speed resonance.

6.2 Simulations versus Experiments. In order to compare the simulation model with the experiments, a series of inner housing geometries with different airflow disturbance configurations is defined, which will result in illustrative responses with respect to some critical speed resonances. In a real drive housing, one airflow disturbance is always present: the OPU, which is always located close to the disk. Therefore, critical speed resonances will always be excited to a certain extent. This situation is approximated by using a single wedge-like airflow disturbance positioned at a 0 deg circumferential position. In this way, a reference situation is obtained, showing a certain response with respect to critical speed resonances. This situation will be discussed in detail in Sec. 6.2.1. In order to gain more insight into the effect of inner housing geometry on the presence of critical speed resonances, a second airflow disturbance will be placed next to the reference disturbance. This will be discussed in Sec. 6.2.2.

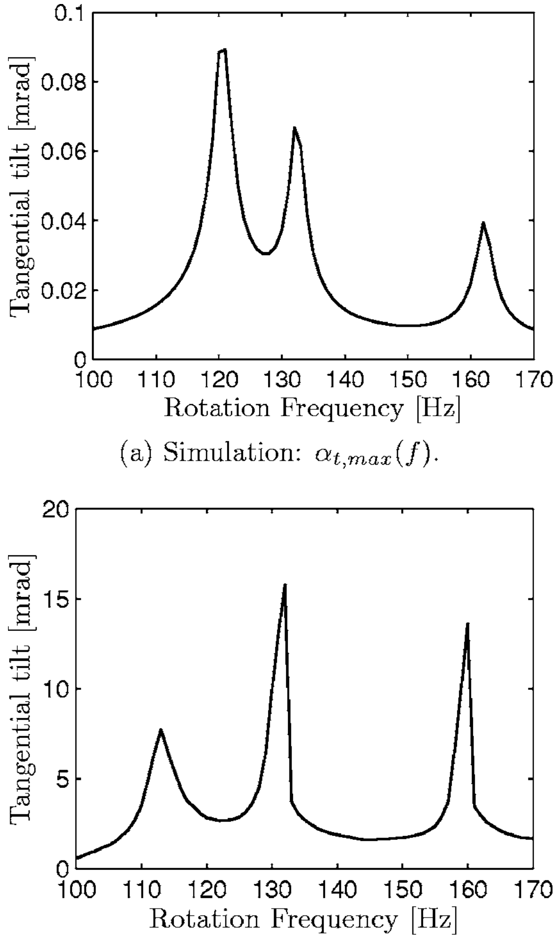
6.2.1 Single Airflow Disturbance. Simulation and experimental avalanche plots of the tangential tilt $\alpha_t(\phi,f)$ of the reference situation are depicted in Fig. 14. For the sake of convenience, the simulation avalanche plot from Fig. 10 is repeated here in Fig. 14(*a*). From Figs. 14(*a*) and 14(*b*), it can clearly be seen that three frequency regions are present in which critical speed resonances show up, corresponding with the critical speeds of the $(0,2)$, $(0,3)$, and $(0,4)$ mode, respectively. Note that the circumferential movements of the local maxima and minima in the experiment near the critical speeds of the $(0,3)$ and the $(0,4)$ mode are not very smooth (at $f=132$ Hz and 160 Hz), which may be caused by temperature effects in the disk. After all, due to the presence of the critical speed resonance, the disk deforms relatively a lot, resulting in high energy dissipation. As a result, the disk is believed to warm up, causing the related natural frequencies and critical speeds to decrease. In measurements, this is observed as a transient critical speed resonance. For instance, only during the first few seconds of a measurement, the critical speed resonance is present after which it disappears. As only the steady-state deformation is measured, this results in nonsmooth transitions in Fig. 14(*b*). For the $(0,2)$ resonance, the transverse disk deflection is believed not to be large enough for the disk to dissipate enough energy to warm up significantly. In order to verify this hypothesis, experiments at higher temperatures should be conducted.

The critical speeds for both the simulations $f_{c,\text{sim}}$ and the experiments $f_{c,\text{exp}}$ are summarized in Table 1, together with their relative difference $((f_{c,\text{exp}} - f_{c,\text{sim}})/f_{c,\text{sim}}) \times 100\%$. From Table 1 it can be seen that there is a mismatch between the critical speeds in the simulation and the experiment. Only a small mismatch is present between the critical speeds of the $(0,3)$ and $(0,4)$ mode. A larger discrepancy is present for the $(0,2)$ mode. This may be caused by a mismatch between the experimental clamping condition of the disk and the assumed boundary condition at the inner radius (zero displacement) in the finite element model. The clamping condition has influence on the natural frequencies of all modes, but for $(0,n)$ modes with $n \geq 3$, the effect is very small (see, for example, [31] for analytical solutions).

Furthermore, the magnitude of the tilt in the simulations is much too small (compare the scaling of the experimental and simulation avalanche plots). The experimental tilt is two orders of

Table 1 Critical speeds for the experiments, simulations, and relative difference

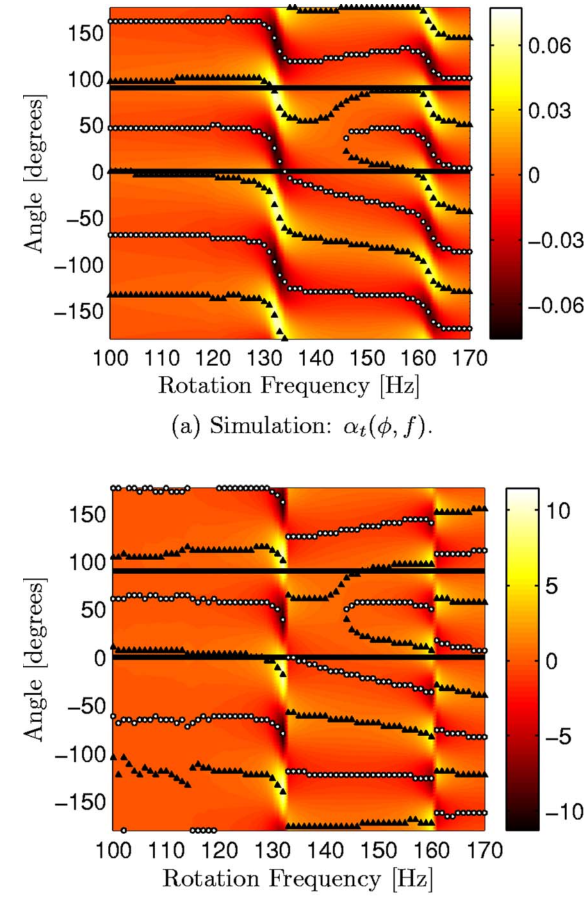
Mode	Experiment (Hz)	Simulation (Hz)	Relative difference (%)
(0,2)	113	121	−6.6
(0,3)	132	133	−0.8
(0,4)	160	163	−1.8



(a) Simulation: $\alpha_{t,max}(f)$.

(b) Experiment: $\alpha_{t,max}(f)$.

Fig. 15 Maximum absolute tangential tilt versus rotation frequency for the reference configuration



(a) Simulation: $\alpha_t(\phi, f)$.

(b) Experiment: $\alpha_t(\phi, f)$.

Fig. 16 Simulation and experimental avalanche plots for the configuration with two airflow disturbances, 90 deg apart

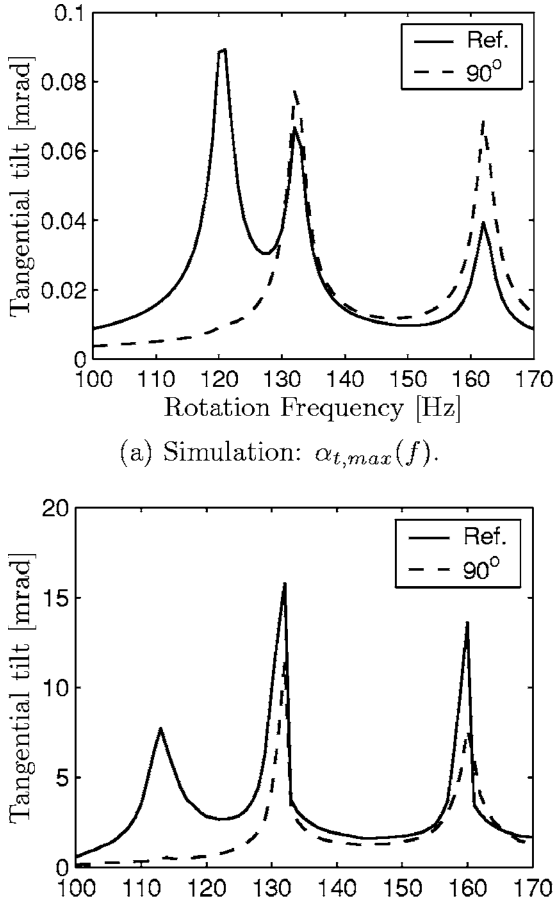
magnitude larger than the tilt in the simulations. However, a *qualitative* match is found between the experimental and simulation results. The circumferential locations of the local maxima and minima of the radial and tangential tilt match approximately. This means that the explanation for the rotation in ϕ direction of the disk shape around the critical speed (see Sec. 6.1) is correct and that the pressure distribution calculated in the model corresponds with the excitation in the experimental setup in a qualitative sense.

To address the effect of other inner housing geometries (Sec. 6.2.2), it is more straightforward to consider maximum absolute tilt plots for tangential tilt ($\alpha_{t,\max}(f)$). These are depicted in Fig. 15 for both the simulation and the experiment with one airflow disturbance. From this figure, another difference between the simulations and measurements can be seen. Namely, in the experiments the $(0,3)$ and $(0,4)$ modes are much more dominant relative to the $(0,2)$ mode than in the simulation. This possibly means that a larger mismatch is present for higher rotation frequencies, which could indicate a frequency-dependent mismatch. Possible causes are nonlinear and turbulent effects of the airflow in the drive, which are not included in the current simulation model. Furthermore, mixture between the air in the upper and lower cavity has not been taken into account yet. This is worthwhile investigating in future research, together with the effect of different airflow disturbance heights.

6.2.2 Two Airflow Disturbances. In an attempt to gain more insight into the airflow-housing-induced excitation of critical speed resonances, the influence of a second airflow disturbance, placed in the lower cavity, is investigated. As an additional result, reduction of critical speed resonances is found to be possible. Several different inner housing geometries are considered, in which the tangential distance between the two airflow disturbances is varied from 30 deg to 90 deg. The effect of the added airflow disturbance is judged by comparing maximum absolute tilt plots. Avalanche plots are used to assess the match between the simulation model and the experiments. Since maximum absolute tilt plots and avalanche plots have similar characteristics for radial and tangential tilt, only the plots for tangential tilt are depicted.

One configuration will be discussed in detail now, to clarify how the aforementioned comparison is carried out. In this configuration, two airflow disturbances of 2 mm height and 10 deg angular width are present below the disk. The angle between the two disturbances equals 90 deg. Tangential tilt avalanche plots for the simulation and the experiment are depicted in Figs. 16(*a*) and 16(*b*), respectively. Comparison of these two figures shows a qualitative match. Namely, in both the experiment and the simulation, the critical speed resonance of the $(0,2)$ mode is suppressed. Instead, from 100–130 Hz, the disk has a $(0,3)$ shape, since three local maxima/minima are found in this range. No new maxima/minima appear around 125 Hz, which normally would be the case (see Fig. 14). At $f \approx 145$ Hz, a new local maximum/minimum couple appears at 45 deg, in between the two airflow disturbances.

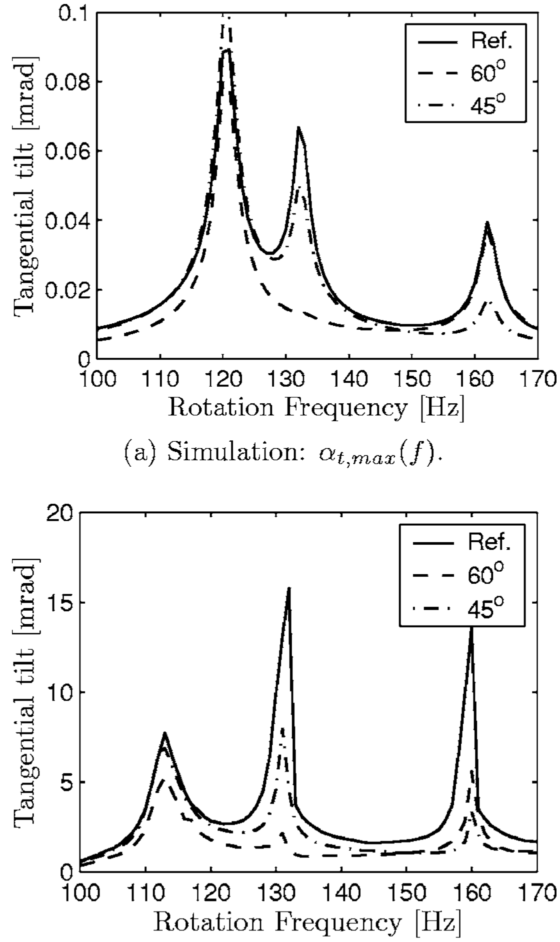
The suppression of the critical speed resonance of the $(0,2)$ mode is also clearly observed in Fig. 17, where the maximum absolute tangential tilt is plotted versus the rotation frequency for both the experiments and the simulations. This suppression is



(a) Simulation: $\alpha_{t,max}(f)$.

(b) Experiment: $\alpha_{t,max}(f)$.

Fig. 17 Maximum absolute tangential tilt versus rotation frequency (solid line: reference; dashed line: two disturbances, 90 deg apart



(a) Simulation: $\alpha_{t,max}(f)$.

(b) Experiment: $\alpha_{t,max}(f)$.

Fig. 18 Maximum absolute tangential tilt versus rotation frequency (solid line: reference; dashed line: two disturbances, 60 deg apart; dashed-dotted line: two disturbances, 45 deg apart

caused by the pressure distribution resulting from the geometry. From Sec. 6.1, it has become clear that the pressure resulting from a single disturbance causes a specific orientation of a mode in critical speed resonance in circumferential direction. However, in the case of two disturbances that are 90 deg apart, their individual orientation demands are in antiphase for the $(0,2)$ mode. For instance, both individual disturbances would cause a similar tangential disk slope at their location, but this is not possible for the $(0,2)$ mode, since the shape of the $(0,2)$ mode repeats itself every 180 deg in circumferential direction. This means that similar (say, positive) tangential slopes are 180 deg apart, whereas the disturbances request 90 deg spacing between positive tangential slopes. As a result, this disturbance configuration hardly excites the $(0,2)$ mode. However, this would be the ideal excitation configuration for the $(0,4)$ mode.

In the maximum absolute tilt plot of the simulation (Fig. 17(*a*)), it can be seen that the tilt at the critical speeds of the $(0,3)$ and $(0,4)$ modes increases for the configuration with two disturbances 90 deg apart, compared to the single airflow disturbance configuration. The $(0,4)$ mode shows the largest relative increase, due to the fact that the excitation from the two airflow disturbances matches the disk shape of this mode better. In the experiment (Fig. 17(*b*)), this effect is not seen. An explanation for this has not been found yet.

In analogy to the suppression of the $(0,2)$ mode, the $(0,3)$ and $(0,4)$ modes can also be suppressed by a certain airflow disturbance configuration. Two airflow disturbances with 60 deg spacing in between hardly excite the $(0,3)$ mode, whereas the $(0,4)$ mode is suppressed by two disturbances located 45 deg apart. This is shown in Fig. 18 for both the simulations and the experiments.

For the situation with two disturbances located 60 deg apart, the $(0,3)$ mode is not suppressed completely in both the simulation and the experiment, which could indicate that the airflow disturbances influence each other's pressure distribution.

The most ineffective excitation for the $(0,4)$ mode contains two disturbances 45 deg apart. However, for both the experiment and the simulation, again, the resonance of this mode is not completely suppressed, which once more indicates that the two airflow disturbances influence each other. Furthermore, in the experiment, the presence of the resonance of the $(0,3)$ mode is also reduced considerably.

From the comparison of the experiments with the simulations, it becomes clear that the model only matches with the experiments in a qualitative sense. In order to obtain a quantitative match, the effect of turbulence and nonlinear effects in the airflow should also be taken into account. Additionally, direct coupling between the two parts in the numerical model might be necessary once the disk transverse deflection is no longer too small to influence the pressure distributions above and below the disk, as is the case in the current implementation. For the geometry configuration with a single airflow disturbance, a good qualitative match is obtained. When two disturbances are used, trends in the experiments are also found from the simulation model. However, when the distance between the disturbances becomes smaller, the qualitative match is less apparent. This is most probably due to the numerical model for the airflow that does not match the reality accurately enough. Again, turbulence and nonlinear effects should probably no longer be neglected for these cases. Furthermore, mixture between the air above and below the disk takes place in the experimental setup, which is also not taken into account in the current model.

Inner housing geometries considered in this research only address the effect of airflow disturbances located in the same (bottom) cavity. Alternatively, it would be worthwhile to investigate the effect on critical speed resonances of two airflow disturbances placed opposite to each other, one in the upper and one in the lower cavity. Due to limitations in the measurement method, however, this has not been investigated yet.

7 Conclusions and Recommendations

In order to gain insight into the influence of drive inner housing geometry on critical speed disk resonances in an optical drive, a mixed numerical-experimental approach has been used. First, theory on disk dynamics, transverse disk modes, and critical speeds is explained and conditions for the occurrence of critical speed disk resonances are stated.

Next, for the numerical analysis, a model has been derived, consisting of two parts: one part describing the disk dynamics and the other part describing the excitation mechanism due to airflow in a drive-like environment. Effects such as, for instance, acoustics, thermal effects, shocks, and suspension flexibility are not taken into account in the numerical model. The disk dynamics are described by a truncated number of mode shapes of a radially prestressed (due to rotation speed) finite element disk model. In order to obtain the airflow induced pressure distribution that serves as the transverse excitation in the disk model, the Reynolds lubrication equation has been implemented using finite difference techniques.

From an experimental point of view, a drive-like environment has been created, which resembles the situation in the numerical model. Wedge-like airflow disturbances are used in the drive's housing. In order to measure disk tilt (both radial and tangential), a measurement method has been designed. This is an indirect method, in which a pattern, reflected in a mirror coated disk, is measured. In this way, a global view of the disk deformation is obtained. By making use of image processing software developed in the programming package Matlab®, and a least-squares fit using the mode shapes from the modal analysis, the transverse disk deflection can be reconstructed from the measurements.

Results from both the numerical and the experimental analyses can be presented by means of two newly proposed types of plots: an avalanche plot and a maximum absolute tilt plot. These plots enable comparison of the simulation results with the experimental results in a straightforward way. Additionally, they enable comparison of the influence of different airflow disturbance configurations on the occurrence of critical speed resonances.

From the results of the numerical-experimental approach, it has been found that the numerical simulations match the experimental results in a qualitative sense. Similar orientations of disk shapes in ϕ direction are found at critical speed disk resonances. The circumferential locations of points with local maximum and minimum tilt values also correspond. However, tilt amplitudes in the numerical model are two orders of magnitude smaller than tilt amplitudes in the experiments. Additionally, a slight mismatch for the critical speeds is found, probably caused by inaccurate modeling of the boundary conditions at the clamping radius of the disk.

In drives that are currently developed, critical speeds of the $(0,2)$, $(0,3)$, and $(0,4)$ modes can be present in the velocity profile. Critical speed resonances of these modes are always excited to a certain extent in practice (due to the OPU). A reference inner housing geometry, containing a single airflow disturbance at a circumferential position of 0 deg, provides insight into the excitation mechanism of critical speed resonances. As an additional result of the simulations and experiments for inner housing geometries containing two airflow disturbances, reduction of critical speed resonances of the $(0,2)$, $(0,3)$, and $(0,4)$ modes has been found to be possible. This is achieved by placing additional airflow disturbances at angles of 90, 60, and 45 deg circumferential spacing, relative to the reference disturbance at $\phi=0$ deg, respectively.

The main contributions of the research presented in this paper can be summarized as follows:

- A suitable and very powerful measurement method for disk tilt has been designed. A global view of the tilt is obtained, and by use of the modal analysis from the numerical model, the disk shape can be reconstructed;
- A numerical model has been developed whose results match with the experimental results in a qualitative sense;
- A systematic approach has provided insight into the effect of inner housing geometry on the presence of critical speed resonances;
- Both the experimental measurement method and the numerical model can be used as a design tool for the development of future generations of drives.

Recommendations for further work are the following. The airflow model should be improved by including the effect of the nonlinear terms in the lubrication equation or even by inclusion of turbulence effects. Additionally, an improved airflow model should allow mixture between air above and below the disk. Furthermore, direct coupling between the disk and the airflow model should be implemented once the disk deflection is found to influence the pressure distribution significantly.

Acknowledgment

Valuable discussions with Hans Kuerten on the modeling of the airflow are thankfully acknowledged.

References

[1] Lamb, H., and Southwell, R. V., 1921, "The Vibrations of a Spinning Disk," Proc. R. Soc. London, Ser. A, **99**, pp. 272–280.
[2] Southwell, R. V., 1922, "On the Free Transverse Vibrations of a Uniform Circular Disk Clamped at its Center and on the Effects of Rotation," Proc. R. Soc. London, Ser. A, **101**, pp. 133–153.
[3] d'Angelo, C., 1991, "Vibration and Aeroelastic Stability of a Disk Rotating in a Fluid," Ph.D. thesis, University of California, Berkeley (UCB), CA.
[4] Chen, J.-S., and Bogy, D. B., 1992, "Effects of Load Parameters on the Natural Frequencies and Stability of a Flexible Spinning Disk With a Stationary Load System," ASME J. Appl. Mech., **59**(2), pp. S230–S235.
[5] Malhotra, N., Sri Namachchivaya, N., and Whalen, T. M., 1995, "Finite Amplitude Dynamics of a Flexible Spinning Disk," *Proceedings of the 1995 Design Engineering Technical Conferences: Vibration of Nonlinear, Random, and Time-Varying Systems*, Boston, MA, Vol. 3A, pp. 239–250.
[6] Chung, J., Kang, N.-C., and Lee, J. M., 1996, "A Study on Free Vibration of a Spinning Disk," KSME Int. J., **10**(2), pp. 138–145.
[7] Jia, H. S., 2000, "Analysis of Transverse Runout in Rotating Flexible Disks by Using Galerkin's Method," Int. J. Mech. Sci., **42**, pp. 237–248.
[8] Lee, S.-Y., Kim, J.-D., and Kim, S., 2002, "Critical and Flutter Speeds of Optical Disks," Microsyst. Technol., **8**, pp. 206–211.
[9] Chung, J., Oh, J.-E., and Yoo, H. H., 2000, "Non-Linear Vibration of a Flexible Spinning Disc With Angular Acceleration," J. Sound Vib., **231**(2), pp. 375–391.
[10] Chung, J., Heo, J. W., and Han, C. S., 2003, "Natural Frequencies of a Flexible Spinning Disk Misaligned With the Axis of Rotation," J. Sound Vib., **240**(4), pp. 763–775.
[11] Heo, J. W., and Chung, J., 2004, "Vibration Analysis of a Flexible Rotating Disk With Angular Misalignment," J. Sound Vib., **274**(3–5), pp. 821–841.
[12] Lee, S.-Y., Yoon, D.-W., and Park, K., 2003, "Aerodynamic Effect on Natural Frequency and Flutter Instability in Rotating Optical Disks," Microsyst. Technol., **9**, pp. 369–374.
[13] Pelech, I., and Shapiro, A. H., 1964, "Flexible Disk Rotating on a Gas Film Next to a Wall," ASME J. Appl. Mech., **31**, pp. 577–584.
[14] Benson, R. C., and Bogy, D. B., 1978, "Deflection of a Very Flexible Spinning

Disk Due to a Stationary Transverse Load," ASME J. Appl. Mech., **45**, pp. 636–642.
[15] Adams, G. G., 1980, "Analysis of the Flexible Disk-/Head Interface," ASME J. Appl. Mech., **102**, pp. 86–90.
[16] Licari, J. P., and King, F. K., 1981, "Elastohydrodynamic Analysis of Head to Flexible Disk Interface Phenomena," ASME J. Appl. Mech., **48**, pp. 763–768.
[17] Carpino, M., and Domoto, G. A., 1988, "Investigation of a Flexible Disk Rotating Near a Rigid Surface," ASME J. Tribol., **110**, pp. 664–669.
[18] Adams, G. G., 1993, "The Point-Load Solution and Simulation of a Flexible Disk Using Various Disk-to-Baseplate Air-Flow Models," Tribol. Trans., **36**(3), pp. 470–476.
[19] Kim, S., Han, G., and Son, H., 1998, "A Study of Characteristics of Disk Vibration and Rotating Airflow in Magneto Optical Disk Drives," IEEE Trans. Consum. Electron., **44**(3), pp. 601–605.
[20] Tatewaki, M., Tsuda, N., and Maruyama, T., 2001, "A Numerical Simulation of Unsteady Airflow in HDDs," Fujitsu Sci. Tech. J., **37**(2), pp. 227–235.
[21] Naganathan, G., Ramadhyani, S., and Bajaj, A. K., 2003, "Numerical Simulations of Flutter Instability of a Flexible Disk Rotating Close to a Rigid Wall," J. Vib. Control, **9**, pp. 95–118.
[22] Bajaj, A. K., Naganathan, G., and Ramadhyani, S., 2003, "On the Forced Harmonic Response of a Flexible Disc Rotating Near a Rigid Wall," J. Sound Vib., **263**, pp. 451–466.
[23] Pinkus, O., and Lund, J. W., 1981, "Centrifugal Effects in Thrust Bearings and Seals Under Laminar Conditions," ASME J. Lubr. Technol., **103**, pp. 126–136.
[24] Reddy, J. N., 1999, *Theory and Analysis of Elastic Plates*, Taylor & Francis, Philadelphia, PA.
[25] ANSYS Release 7.0; ANSYS, Inc.
[26] Matlab® Release 13, Version 6.5, The Mathworks, Inc.
[27] Roache, P. J., 1972, *Computational Fluid Dynamics*, Hermosa Publishers, Albuquerque, NM.
[28] Heath, M. T., 2002, *Scientific Computing: An Introductory Survey*, McGraw-Hill, New York.
[29] Willems, K., 2003, "Spinning Disc Dynamics of Optical Discs: A Numerical and Experimental Approach," MS thesis, Eindhoven University of Technology, Eindhoven, the Netherlands.
[30] Janna, W. S., 2000, *Engineering Heat Transfer*, 2nd ed., CRC Press, Boca Raton, FL.
[31] Leissa, A. W., 1969, "Vibration of Plates," NASA-SP-160.

Myoung-Gyu Lee[1]
Department of Materials Science and Engineering,
2041 College Road,
Ohio State University,
Columbus, OH 43210

Daeyong Kim
Research and Development Division for Hyundai Motor Company and Kia Motors Corporation,
772-1, Jangduk-Dong,
Whasung-Si, Gyunggi-Do, 445-706, Korea

R. H. Wagoner
Department of Materials Science and Engineering,
2041 College Road,
Ohio State University,
Columbus, OH 43210

Kwansoo Chung[2]
School of Material Science and Engineering,
Seoul National University,
56-1, Shinlim-Dong,
Kwanak-Ku, Seoul 151-742, Korea
e-mail: kchung@snu.ac.kr

Semi-Analytic Hybrid Method to Predict Springback in the 2D Draw Bend Test

A simplified numerical procedure to predict springback in a 2D draw bend test was developed based on the hybrid method which superposes bending effects onto membrane solutions. In particular, the procedure was applied for springback analysis of a specially designed draw bend test with directly controllable restraining forces. As a semi-analytical method, the new approach was especially useful to analyze the effects of various process and material parameters on springback. The model can accommodate general anisotropic yield functions along with nonlinear isotropic-kinematic hardening under the plane strain condition. For sensitivity analysis, process effects such as the amount of bending curvature, normalized back force and friction, as well as material property effects such as hardening behavior including the Bauschinger effect and yield surface shapes were studied. Also, for validation purposes, the new procedure was applied for the springback analysis of the dual-phase high strength steel and results were compared with experiments. [DOI: 10.1115/1.2745390]

Keywords: springback, sheet metal forming, draw-bend test, plane strain analysis, anisotropic yield surface, nonlinear kinematic hardening, Bauschinger effect, dual-phase high strength steel

1 Introduction

Springback, as mainly the elastically driven shape change of formed parts, is one of drawback, which needs to be minimized or compensated for with proper prediction in the design stage of forming processes for quality end products. Springback is however affected by the complex combination of bending, unbending, and stretching imposed on parts during forming processes. Therefore, the proper understanding of the effects of process parameters as well as material properties on springback is so useful to effectively design forming processes. There are three approaches in analyzing springback: the analytical method, the semianalytical method without involving major finite element analysis, and the finite element analysis. The classical analytical approach assumes simple tool description and material properties. Examples of this approach include the springback analysis of pure bending with elasto-perfect plasticity [1], plane strain pure bending [2], plane strain bending [3,4], plane stress bending [5] with additional tensile force, and biaxial elastic-plastic pure bending of a rectangular plate [6,7]. The analysis of process effects was mainly performed for 2D draw forming, which involves the die corner and sidewall curl regions [8,9]. More recently, Jeunechamps et al. developed a closed form method to predict springback in creep age-forming and investigated the effects of geometric parameters on springback of aluminum plates [10]. Since the analytical method involves simplifications of process conditions and material properties, the method is not so effective in the accurate analysis of springback for real forming. However, it provides a useful basis for the qualitative understanding of process and property effects on springback.

With the rapid development of computational power and solution techniques, the finite element method (FEM) has been widely utilized to predict and understand the springback. Examples of the FEM springback analysis of 2D formed parts include the cantilever beam analysis by Kawaguchi et al. [11], the 2D draw bending benchmark problem by Mattiasson et al. [12] and He and Wagoner [13]. The advantages of the FEM method over analytical and semianalytical methods are its capabilities to model complicated tool descriptions and realistic constitutive behavior. Li et al. [14] investigated the springback of draw bend tests including parametric studies on numerical and physical parameters such as the size of meshes, the number of integration points, tool radius, and restraining forces. Geng and Wagoner [15] analyzed springback angles and the role of anticlastic curvature especially with large (restraining) back forces using a series of simulated draw bending tests in conjunction with an anisotropic hardening rule and four different yield functions. Chung et al. [16] and Lee et al. [17,18] evaluated the springback of a modified automotive part by implementing the modified combined isotropic-kinematic hardening and nonquadratic anisotropic yield function. In terms of the finite element technique, in order to improve the accuracy of the elastic-plastic stress distribution in springback prediction, higher order finite elements with enhanced assumed strain method have been developed to remove volumetric and shear lockings [19,20].

Although FEM is useful and considered as one of the most accurate numerical tools for the analysis of springback, the high computational cost and highly sensitive nature of predicted results to numerical parameters are difficulties that have yet to be overcome. Computational cost to prepare meshes and connectivity for sheets and tools during preprocessing and to generate proper results during postprocessing are much larger than those of analytic or semianalytical methods. Regarding the high sensitivity of numerical parameters on springback, the effects of element types and the integration scheme were reported [21,22]. The sensitivity of springback was also ascertained by widely scattered simulation results from various FE codes submitted to the Numisheet benchmark problem [23]. Even the same FE code fails to give consistent results when different numerical parameters were utilized for the

[1]Current address: Eco-Materials Research Center, Korea Institute of Machinery and Materials 66, Sangnam, Changwon, Kyungnam 641-101, Korea.
[2]Corresponding author.
Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received September 22, 2006; final manuscript received April 10, 2007. Review conducted by Antoinette Maniatty.

same problem. In an effort to reduce computational cost without sacrificing that much of the accuracy in solutions, the hybrid membrane/shell method was developed for the springback analysis of 2D draw bending [24,25] and 3D parts [26].

In the present paper, a semianalytical analysis method was developed based on the hybrid method [25], particularly for the springback analysis of a specially designed 2D draw bend test with direct controllable restraining back forces [27,28]. For this purpose, a simplified plane strain formulation based on elastic-plasticity with nonlinear isotropic-kinematic hardening and the nonquadratic anisotropic yield function was derived. To represent the anisotropy of rolled sheets, the yield function Yld2000-2d [29] was utilized. As for the hardening law, the nonlinear kinematic hardening law based on the Chaboche model [30,31], which has been successfully applied to FE analysis to predict springback [15,16,32–35], was adopted.

Utilizing the semianalytical method developed here, sensitivity tests were performed to understand process and material property effects on springback in the particular 2D draw bend test. As for the process effect, simulations were carried out for various bending curvatures, restraining back forces, and friction coefficients. As for the material property effect, various values of material parameters were tried out for isotropic-kinematic hardening and the yield function. Finally, the predicted shapes and magnitudes of springback were verified against experimental measurements for the dual-phase high strength steel.

2 Draw Bend Test

The draw bend test considered in the work is schematically shown in Fig. 1(*a*). The draw bend test has the following advantages over the typical U-channel bend test being widely used for the study of forming and springback: (1) direct control of applied force so that springback can be assessed for explicitly known restraining (back) forces, (2) a single tool without bend at the bottom of the channel which would have made the analysis complicated, and (3) longer draw distance attainable. Therefore, with a simpler setup and directly controllable tensile forces, springback can be systematically measured and analyzed under a range of controlled laboratory conditions but applicable to industrial practice.

Detailed procedures of the draw bend test are well documented elsewhere [27,28] and the test is briefly reviewed here. As shown in Fig. 1(*a*), the upper grip provides a constant restraining back force, while the lower grip is displaced at a constant speed, thus drawing the sheet strip over the cylindrical tool with a radius chosen. The sheet sample undergoes tensile loading, bending over the cylindrical tool and unbending at the exit of the cylindrical tool. After the strip travels a prescribed distance, the grips are instantly removed and the sheet is allowed to springback. Figure 1(*b*) shows a typical schematic view of the blank sheet before and after springback. In order to measure the magnitude of springback after the removal of tools, two main regions are considered. The springback in region II in Fig. 1(*b*) represents the change of curvature from the curvature of the cylindrical tool to that of the relaxed sheet. The angle change $\Delta\theta_1$ is

$$\Delta\theta_1 = \theta_1 - \frac{\pi}{2} = \frac{\pi}{2}\left(\frac{r}{r'} - 1\right) \tag{1}$$

where r and r' are radii of the curvature of region II before and after springback, respectively. On the other hand, the springback in region III represents the change of curvature from zero to a finite value. By assuming constant curvature of region III,

$$\Delta\theta_2 = \theta_2 = \frac{127\ \text{mm}}{r''} \tag{2}$$

where r'' is the radius of curvature of region III after springback. Therefore, in general, the angle change in region II is negative, while it is positive in region III. It is difficult to measure the springback angle with great accuracy especially in region II because the magnitude of angle change is quite small compared to that of the other region and variation is small with respect to the change of conditions. Therefore, following the suggestion made previously [14,27], the sum of the two angle changes, $\Delta\theta = \Delta\theta_1 + \Delta\theta_2$, is adopted as the measurement of springback.

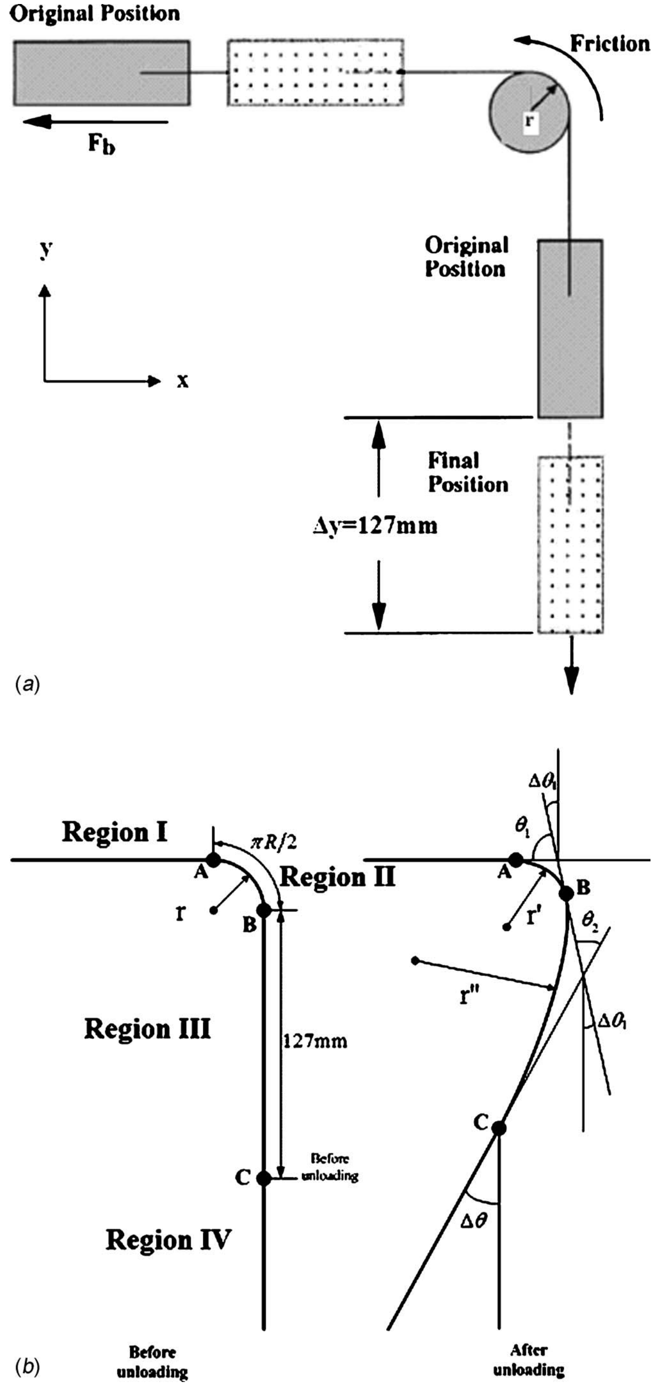


Fig. 1 (*a*) Geometry of the draw bend test with (*b*) deformed shape before and after springback

The three main control parameters which influence the magnitude of springback are restraining force, cylindrical tool size, and friction. For the restraining force, the back force normalized by the tension required to yield the sheet in the absence of bending is used. Since the ratio of the cylindrical tool radius (r) over the thickness of the sheet (t) is known as an important parameter (r/t ratio) from the previous analytical study, the test can be performed with various sizes of cylindrical tools. The friction coefficient is hard to measure accurately in spite of introducing well characterized lubricants. Therefore, various Coulomb friction coefficients are utilized to analyze the sensitivity of springback to the condition of friction between the tool and the blank sheet.

3 Hybrid Method

In the hybrid scheme, the solutions of bending and unbending are superposed onto membrane solutions. The draw bend test adopted in this paper has simple tool geometries in addition to the direct application of restraining back forces so that solutions can be derived in a straight manner without the aid of finite element calculation unlikely as done in the previous works [24,25]. As for the constitutive behavior of materials, the combined isotropic-kinematic hardening based on nonlinear kinematic hardening is applied [16]. Therefore, the numerical scheme is inevitably needed to calculate solutions, which satisfy the equilibrium condition at all material points. Details of the way the hybrid method was applied for the semi-analytic analysis are illustrated here.

3.1 Constitutive Equations for a Thin Shell under the Plane Strain Condition. Under the plane strain assumption for a thin shell, strain components have the following relationship:

$$d\varepsilon_2^e = -d\varepsilon_2^p \ (\neq 0) \tag{3}$$

where superscripts e and p represent elastic and plastic components, respectively. The subscript 2 shows the component of strain along which the strain is constrained (the width direction of a thin shell). When the plane stress condition is further imposed for a thin sheet, the stress state is generally not proportional under the elasto-plastic formulation as schematically shown in Fig. 2(a) [18]. As a way to achieve proportionality in loading for the simple analysis here, the loading in Fig. 2(a) was approximated to that in Fig. 2(b); within the yield surface, the stress is proportional under the condition, $d\varepsilon_2 = d\varepsilon_2^e = 0$, and as soon as the stress reaches the yield surface, the stress is proportional under the condition, $d\varepsilon_2^p = 0$.

By the linear elasticity,

$$d\sigma_{\text{ps}} = E^{\text{ps}}(d\varepsilon_{\text{ps}} - d\varepsilon_{\text{ps}}^p) \tag{4}$$

where $E^{\text{ps}} = E/1-\nu^2$. Here, the ps means the components of the plane strain deformation, while E and ν are Young's modulus and Poisson ratio of elasticity, respectively. By applying the modified plastic work equivalence principle for the isotropic-kinematic hardening law [16],

$$\bar{\sigma}_{\text{iso}} d\bar{\varepsilon} = |\sigma_{\text{ps}} - \alpha_{\text{ps}}| \cdot |d\varepsilon_{\text{ps}}^p| \tag{5}$$

where the upper bar is used for equivalent (or effective) values and subscript iso means isotropic hardening (therefore, $\bar{\sigma}_{\text{iso}}$ represents the current size of the expanding yield surfaće), while α is the back stress which represents the translation of the initial yield surface. When the yield function is specified, the ratio β between $|\sigma_{\text{ps}} - \alpha_{\text{ps}}|$ and $\bar{\sigma}_{\text{iso}}$ is determined under the plane strain condition ($d\varepsilon_2^p = 0$). Therefore, Eq. (5) leads to

$$|\sigma_{\text{ps}} - \alpha_{\text{ps}}| = \beta\bar{\sigma}_{\text{iso}}$$

$$|d\varepsilon_{\text{ps}}^p| = \frac{1}{\beta} d\bar{\varepsilon} \tag{6}$$

Note that the ratio β depends on the shape of yield surface and it is $2/\sqrt{3}$ for the von Mises yield function.

The evolution of the back stress under the plane strain condition for the combined isotropic-kinematic hardening law based on the modified Armstrong-Frederick model [36] becomes

$$d\alpha_{\text{ps}} = C_1 |d\varepsilon_{\text{ps}}^p| \frac{\beta^2(\sigma_{\text{ps}} - \alpha_{\text{ps}})}{|\sigma_{\text{ps}} - \alpha_{\text{ps}}|} - C_2 \alpha_{\text{ps}} \beta |d\varepsilon_{\text{ps}}^p| \tag{7}$$

where C_1 and C_2 are material parameters measured at the reference state, which is the uniaxial tension state in this work.

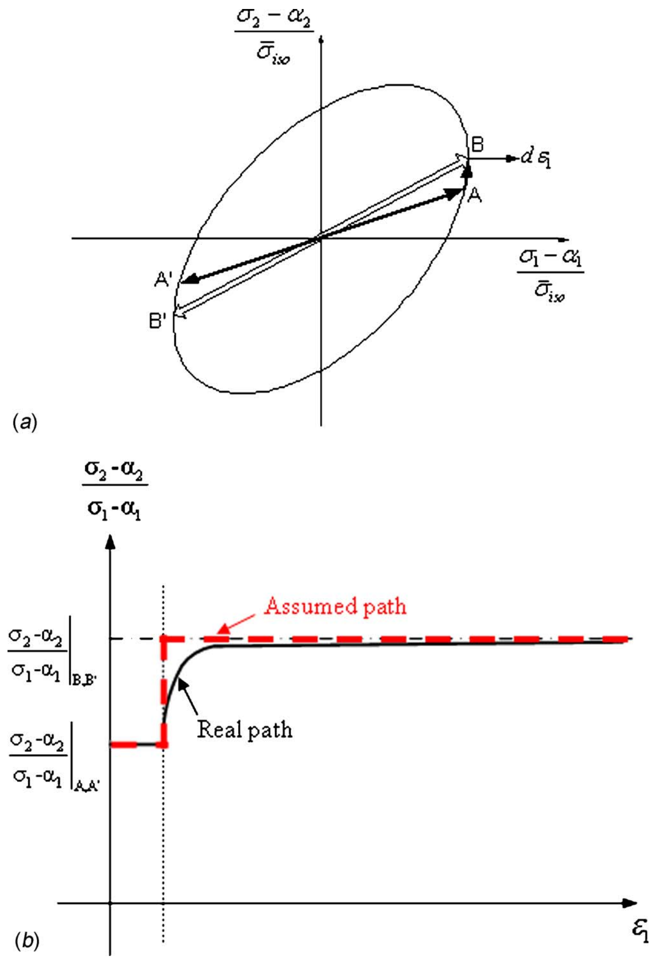


Fig. 2 (***a***) **Stress state of a thin sheet in the plane-stress stress field under the plane strain deformation in elasto-plasticity and** (***b***) **rapid development of the proportional stress state during the initial plastic deformation**

The hardening parameters $\bar{\sigma}_{\text{iso}}(\bar{\varepsilon})$ for isotropic hardening and C_1 and C_2 for kinematic hardening are provided by performing tension/compression uniaxial tests (compression tests with the various amounts of tensile prestrains). A specially designed device to prevent buckling when the specimen is compressed has been utilized, which is well documented in the previous literature [37]. From the measured data, initially the isotropic hardening data is separated from the total hardening data. Then, the two kinematic hardening parameters which are assumed constants in this work are obtained by the fitting the equation $\bar{\alpha} = (C_1/C_2)(1-e^{-c_2\bar{\varepsilon}})^2$ for the rest portion of the hardening. The detailed characterization method for the yield stress surface and hardening data is referred to the previous work [17].

When the strain increment $d\varepsilon_{\text{ps}}$ is prescribed, the updated stresses satisfy the following consistency condition: the newly updated stress and back stress in the stress field satisfied the yield stress size evolution condition specified by the isotropic hardening data. Therefore, the consistency condition for the plane strain condition becomes

[2] This is the solution of the ordinary differential equation, $d\bar{\alpha} = C_1 - C_2\bar{\alpha}$, which is the relationship at the reference state (or the uniaxial tension test condition).

$$\bar{\sigma}_{\rm iso} - \bar{\sigma}_{\rm iso}\left(\beta \int |d\varepsilon_{\rm ps}^{p}|\right) = \frac{|\sigma_{\rm ps} - \alpha_{\rm ps}|}{\beta} - \bar{\sigma}_{\rm iso}\left(\beta \int |d\varepsilon_{\rm ps}^{p}|\right) = 0 \tag{8}$$

where $|\sigma_{\rm ps} - \alpha_{\rm ps}|/\beta$ is the value obtained from the stress field and $\bar{\sigma}_{\rm iso}(\beta\int|d\varepsilon_{\rm ps}^{p}|)$ is the value obtained from the experimentally measured isotropic hardening data.

When discretized for the numerical formulation with the discrete time increment Δt, Eq. (8) becomes the following nonlinear equation for the discrete equivalent strain increment $\Delta\bar{\varepsilon}$, considering the discretized forms of Eqs. (4), (6), and (7),

$$G_{n+1} = \bar{\sigma}_{\rm iso} - \bar{\sigma}_{\rm iso}(\Delta\bar{\varepsilon}) = \frac{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|}{\beta} - \bar{\sigma}_{\rm iso}(\Delta\bar{\varepsilon}) = 0 \tag{9}$$

where

$$\sigma_{{\rm ps},n+1} = \sigma_{{\rm ps},n+1}^{T} - \Delta\bar{\varepsilon} \cdot E^{\rm ps} \cdot \frac{1}{\beta} \cdot \frac{\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}}{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|} \tag{10}$$

$$\alpha_{{\rm ps},n+1} = \alpha_{{\rm ps},n} + C_1\Delta\bar{\varepsilon} \cdot \frac{\beta(\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1})}{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|} - C_2\alpha_{{\rm ps},n+1}\Delta\bar{\varepsilon} \tag{11}$$

Note that Eq. (11) is equivalent to Eq. (7), but here the discrete equivalent plastic strain increment $\Delta\bar{\varepsilon}$ is used. In Eq. (10), $\sigma_{{\rm ps},n+1}^{T}$ is a trial stress at the time step $n+1$, which is initially assumed to be elastic for a prescribed discrete strain increment $\Delta\varepsilon_{\rm ps}$; i.e.,

$$\sigma_{{\rm ps},n+1}^{T} = \sigma_{{\rm ps},n} + E^{\rm ps} \cdot \Delta\varepsilon_{\rm ps} \tag{12}$$

If the stress is outside the yield criterion, the current step is considered as elastoplastic and solution for Eq. (9) is performed based on the Newton-Rhapson scheme. When the predictor-corrector scheme is applied to solve Eq. (9) in which $\sigma_{{\rm ps},n+1}$ and $\alpha_{{\rm ps},n+1}$ on the right-hand side of Eqs. (10) and (11) are considered known as they are corrected during iteration, the linearized form of Eq. (9) is, for the kth iteration,

$$G^{(k)} + \left(\frac{\partial G}{\partial \Delta\bar{\varepsilon}}\right)^{(k)} \cdot \delta(\Delta\bar{\varepsilon})^{(k+1)} = 0 \quad \text{or} \quad \delta(\Delta\bar{\varepsilon})^{(k+1)} = -\frac{G^{(k)}}{\left(\dfrac{\partial G}{\partial \Delta\bar{\varepsilon}}\right)^{(k)}} \tag{13}$$

where

$$\frac{\partial G}{\partial \Delta\bar{\varepsilon}} = \frac{\partial G}{\partial(\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1})}\left(\frac{\partial\sigma_{{\rm ps},n+1}}{\partial\Delta\bar{\varepsilon}} - \frac{\partial\alpha_{{\rm ps},n+1}}{\partial\Delta\bar{\varepsilon}}\right) + \frac{\partial G}{\partial\bar{\sigma}_{\rm iso}}\frac{\partial\bar{\sigma}_{\rm iso}}{\partial\Delta\bar{\varepsilon}} \tag{14}$$

$$\frac{\partial G}{\partial(\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1})} = \frac{1}{\beta} \cdot \frac{\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}}{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|} \tag{15}$$

$$\frac{\partial\sigma_{{\rm ps},n+1}}{\partial\Delta\bar{\varepsilon}} = -E^{\rm ps} \cdot \frac{1}{\beta} \cdot \frac{\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}}{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|} \tag{16}$$

$$\frac{\partial\alpha_{{\rm ps},n+1}}{\partial\Delta\bar{\varepsilon}} = C_1 \cdot \frac{\beta(\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1})}{|\sigma_{{\rm ps},n+1} - \alpha_{{\rm ps},n+1}|} - C_2\alpha_{{\rm ps},n+1} \tag{17}$$

$$\frac{\partial G}{\partial\bar{\sigma}_{\rm iso}} = -1 \tag{18}$$

The detailed numerical procedure for the general plane stress case is documented elsewhere [16].

3.2 Resultant Force and Bending Moment. The through-thickness tangential strain distribution during bending, unbending with/without tension is calculated by sectioning the thickness into N (even number) initially equal-sized layers for each element having an initial area A_0. In Fig. 3, the current thickness (therefore, also the radius of curvature) of each layer is calculated by applying the following recursive equations from the middle layer under the incompressibility assumption,

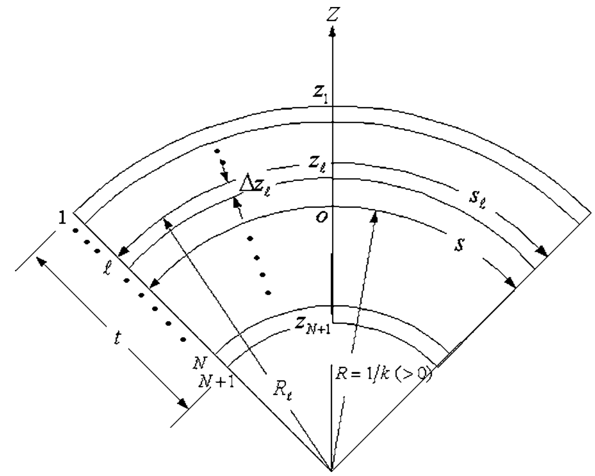


Fig. 3 Definition of N layers through the thickness

$$\frac{A_0}{N} = \frac{1}{2}\frac{S}{R}(R_{l-1}^2 - R_l^2) = \frac{1}{2}\frac{S}{R}((R_l + \Delta z_{l-1})^2 - R_l^2) \tag{19}$$

for layers above the center lines and

$$\frac{A_0}{N} = \frac{1}{2}\frac{S}{R}(R_l^2 - R_{l+1}^2) = \frac{1}{2}\frac{S}{R}(R_l^2 - (R_l - \Delta z_l)^2) \tag{20}$$

for layers below the center line. Solutions of the above equations lead to

$$\Delta z_{\ell-1} = \sqrt{R_\ell^2 + 2\frac{R}{S}\frac{A_0}{N}} - R_\ell \quad \text{(for layers above the central line)} \tag{21}$$

$$\Delta z_\ell = R_\ell - \sqrt{R_\ell^2 - 2\frac{R}{S}\frac{A_0}{N}} \quad \text{(for layers below the central line)} \tag{22}$$

where R $(=1/\kappa$, κ is curvature) and S are the current radius of curvature and length of the centerline. Note that R is the radius of curvature at the current center line which is iteratively obtained from the tool radius r and length of center line S. Equations (21) and (22) correspond to the positive bending in which the outer fiber in Fig. 3 is stretched, while the inner fiber is compressed. For the negative bending Eq. (22) holds for $\Delta z_{\ell-1}$ above the central line and Eq. (21) holds for Δz_ℓ below the central line.

The discrete tangential strain increment of each layer is calculated by considering the current length

$$\Delta\varepsilon_{{\rm ps},l} = \ln\left(\frac{S_l}{{}^0S_l}\right) \quad \text{with} \quad S_\ell = (1 + \kappa z_\ell)S \tag{23}$$

where ${}^0S_\ell$ is the length of the ℓth layer at the previous time step.

Applying the constitutive equations, the stress update algorithm and kinematics for the algorithm explained in the previous section, the following resultant tangential force and the bending moment are obtained for a cross section:

$$T = \sum_{\ell=1}^{N} \langle\sigma(z_\ell)\rangle \cdot \Delta z_\ell \cdot W \quad \text{where} \quad \langle\sigma(z_\ell)\rangle = \frac{\sigma(z_\ell) + \sigma(z_{\ell+1})}{2} \tag{24}$$

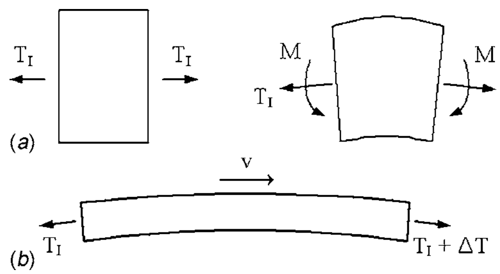


Fig. 4 (*a*) An element in region I under the tension and in region II under tension and bending moment, (*b*) tensile force increase at the entrance of region II

$$M = \sum_{\ell=1}^{N} \langle \sigma(z_\ell) \rangle \cdot \langle z_\ell \rangle \Delta z_\ell \cdot W \quad \text{where} \quad \langle z_\ell \rangle = \frac{z_\ell + z_{\ell+1}}{2} \tag{25}$$

where W is the width of the thin sheet. Note that the resultant force and the bending moment are determined by the current radius of curvature and length of the centerline, R and S.

3.3 Solution Procedure for the Draw Bend Test. In the simulation of the bending test shown in Fig. 1, the radius of curvature for the centerline R is predetermined for each element considering the position of the element along the tool geometry (the region along the sample where the element is located) in Fig. 1. Therefore, the equilibrium condition for the bending moment is not considered for the forming analysis, while the membrane strain (or the length of the central line in each element S) is calculated considering the force equilibrium or, equivalently, by equating the tensile force distribution prescribed along the sample with the resultant force in Eq. (24). The tensile force distribution is determined in terms of material properties, cylindrical tool geometry and friction, as discussed here.

The deformed blank sheet during the draw bend test is divided into four regions as shown in Fig. 1(*b*). In region I where the sheet is stretched by the prescribed restraining back force F_b, a constant tensile force due to the explicitly prescribed back force is assumed in this region. Note that there is no friction involved in the present draw bend test.[3] Therefore, the tensile force at region I is

$$T_{\mathrm{I}} = F_b \tag{26}$$

As for the tensile force in region II, the force instantaneously increases at the entrance of region II (at A in Fig. 1(*b*)) to overcome the resistance associated with bending, which each material element undergoes in addition to tension, when it enters this region. The instantaneous force increment at A can be derived by equating the work rate required to pull the sheet over the tool surface with the plastic work increase rate associated with additional bending, as similarly done in the previous works for the draw bead model by Stoughton [38] and Marciniak and Duncan [39].

Consider a material element with length ${}^0S$ under tension T_{I} in region I as shown in Fig. 4(*a*). Now, the material is assumed to take up a new central line length S and bending as soon as it enters region II under the tension T_{I} and the bending moment. Since R and T_{I} are prescribed, the new S is determined from Eq. (24) along with the stress distribution after extension and bending (the bending moment can also be obtained from Eq. (25) but its value is not needed). Therefore, the plastic work increment rate associated with the material deformation becomes

$$\frac{dW}{dt} = \left[\iint (\sigma \cdot d\varepsilon) dV \Big|_{\text{under tension } T_{\mathrm{I}} \text{ and bending}} - \iint (\sigma \cdot d\varepsilon) dV \Big|_{\text{under tension } T_{\mathrm{I}}} \right] \cdot \left(\frac{v}{{}^0S} \right) \tag{27}$$

where v is the velocity of material movement and S^0/v is the time duration of the movement. The increment of the tensile force ΔT is obtained by equating Eq. (27) with the work increment rate by external forces applied on the strip as shown in Fig. 4(*b*), $\Delta T \cdot v$. Therefore, ΔT is approximated as

$$\Delta T_{\mathrm{I-II}} = \left[\iint (\sigma \cdot d\varepsilon) dz \Big|_{\text{under tension } T_{\mathrm{I}} \text{ and bending}} - \iint (\sigma \cdot d\varepsilon) dz \Big|_{\text{under tension } T_{\mathrm{I}}} \right] \cdot W \tag{28}$$

When the Coulomb friction is considered between the cylindrical tool and the material in region II, the tensile force in region II becomes

$$T_{\mathrm{II}} = (T_{\mathrm{I}} + \Delta T_{\mathrm{I-II}}) \cdot \exp\{\mu \cdot \theta\} \tag{29}$$

where μ and θ are the friction coefficient and the clockwise angle along the cylindrical tool, respectively.

As for the tensile force in region III, the force instantaneously increases at the entrance of region III (B in Fig. 1(*b*)) to overcome the resistance associated with complete unbending of the material. Therefore, as similarly done for $\Delta T_{\mathrm{I-II}}$,

$$\Delta T_{\mathrm{II-III}} = \left[\iint (\sigma \cdot d\varepsilon) dz \Big|_{\text{under tension } T_{\mathrm{II}}(\theta=\pi/2)} - \iint (\sigma \cdot d\varepsilon) dz \Big|_{\text{under tension } T_{\mathrm{II}}(\theta=\pi/2) \text{ and bending}} \right] \cdot W \tag{30}$$

and

$$T_{\mathrm{III}}(=T_{\mathrm{IV}}) = T_{\mathrm{II}}\left(\theta = \frac{\pi}{2}\right) + \Delta T_{\mathrm{II-III}} \tag{31}$$

The tensile force instantaneously increases at the entrance of region III (at B in Fig. 1(*b*)) and the force remains constant for regions III and IV.

Note that the tensile forces in the whole regions are numerically predetermined before analysis considering material properties and tool geometry. Equating Eq. (24) with the prescribed tensile force distribution leads to simultaneous nonlinear equations for the membrane strain distribution. Here, the bisection method was utilized to find solutions iteratively in which solutions were achieved within a small number of iterations for most cases.

3.4 Deformed Shape After Springback. Springback simulation after forming is to find the curvature and length distributions of central lines, which are solved from the following equilibrium conditions after unloading:

$$T = M = 0 \tag{32}$$

Once the curvature and the length of the central lines are obtained, the deformed shape is calculated from the following geometrical relationship. The coordinates of nodal points are expressed as

$$\Delta x_i = S_i \cos\left(\sum_{j=1}^{i} \kappa_j S_j \right) \quad \text{and} \quad \Delta y_i = S_i \sin\left(\sum_{j=1}^{i} \kappa_j S_j \right) \tag{33}$$

where Δx_i and Δy_i are the increment of coordinates and S_i denotes the length of the centerline of the ith element.

[3] In the conventional U channel draw bend test, friction by the blank holder or the draw bead should be considered, which makes the analysis a little more complicated.

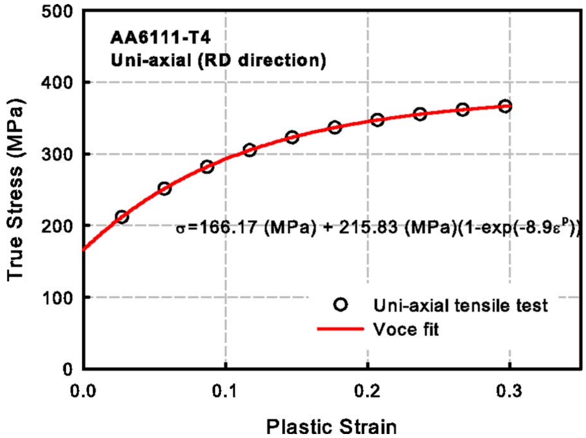


Fig. 5 Stress-strain curve with the Voce fit for sensitivity tests

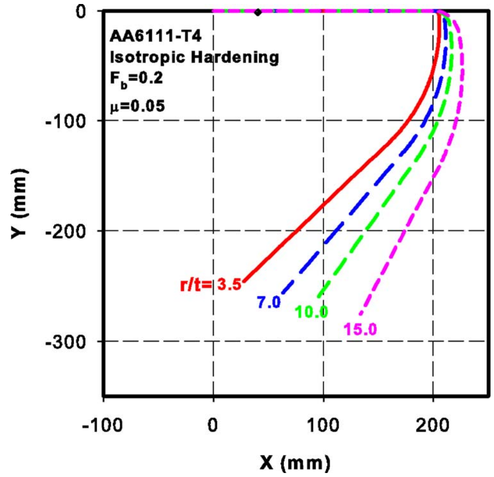


Fig. 6 Final shapes for the variation of radius of tool-to-thickness

4 Sensitivity Tests

Numerical sensitivity tests were performed for the draw bend test to investigate the effect of process and material parameters on springback. As for process parameters, the r/t ratio, back force and friction have been studied. Regarding material parameters, different hardening models have been compared with/without the Bauschinger effect. The effect of the nonquadratic yield surface shape, which represents crystal structures, on springback was also investigated. Material aspects such as Young's modulus, yield strength, and thickness were omitted here as many studies have been done on these [15,18].

For this study, the constitutive parameters of the aluminum alloy AA6111-T4 obtained from the previous report [40] were considered as reference values. Figure 5 shows the uniaxial stress-strain curve of the test material with the Voce fit. Sensitivity tests on the material were carried out by varying a specific parameter, while fixing other parameters. The reference values are as follows: isotropic yield function (by setting all of the constants unity in Barlat's Yld2000-2d), yield potential exponent is 8 to represent the fcc crystal structure (therefore, the plane strain parameter β in Eq. (6) is 1.1089). Young's modulus and Poisson ratio were 70 GPa and 0.33, respectively. As for the hardening rule, the isotropic hardening law was utilized for all the simulations of process effects, while kinematic hardening was also considered for the simulations of material effects. The reference values for the process parameters are $F_b=0.2$, $r/t=3.5$ and friction coefficient $\mu=0.05$.

The sheet thickness was divided into 20 layers through the thickness and 200 increments were used for the bending and unbending steps. Finer elements were used for region II and region III. In this particular sensitivity test, the characteristic element length for regions II and III is 1 mm, and is 3 mm for other regions. The initial rectangular strip with 508 mm × 50 mm and 1.51 mm thickness was considered to be displaced 127 mm by the lower grip.

4.1 Process Effects. Figures 6 and 7 show the variation of predicted deformed geometry after springback and springback angle with respect to the r/t ratio, respectively. The figures show that the springback decreases (almost) linearly as the r/t ratio increases. The moment-curvature curves of material points on the sidewall region (region III) were calculated to examine detailed material behavior. For the larger curvature (or less r/t ratio), the moment is unloaded from the larger (absolute) value when the material is unloaded from bended or bended-unbended (straightened) regions. Therefore, the magnitude of the curvature change during unloading (the amount of springback) is proportional to the magnitude of moment before unloading as shown in Fig. 8. The variation of the overall shape is dominated by the sidewall curl in region III which is much larger than that in region II in the draw bend test.

Similar analysis was performed for various back forces in Figs. 9–11. The result shows that the magnitude of springback decreases as back force increases. The analysis of moment-curvature

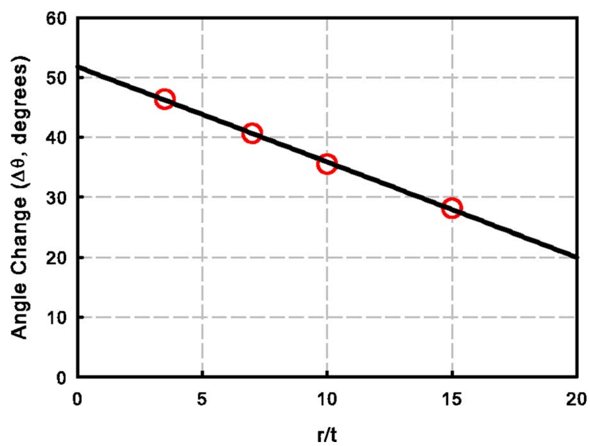


Fig. 7 Variation of the springback angle ($\Delta\theta$) with the radius of tool-to-sheet thickness (r/t)

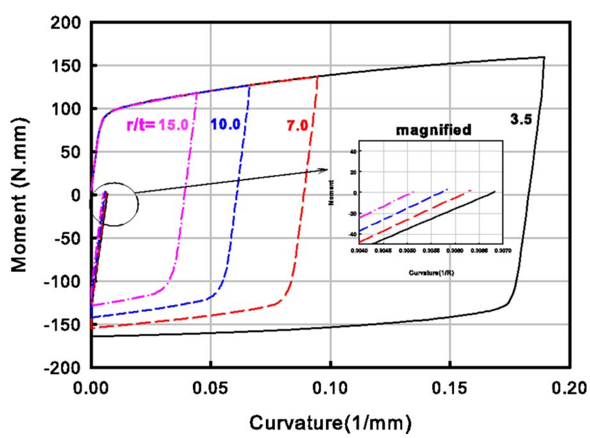


Fig. 8 Effect of the radius of tool-to-thickness (r/t) on moment-curvature curves

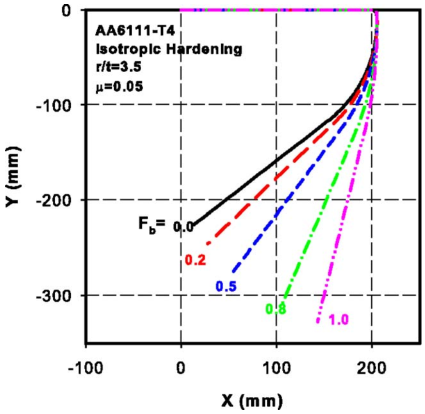


Fig. 9 Final shapes for the variation of normalized back force (F_b)

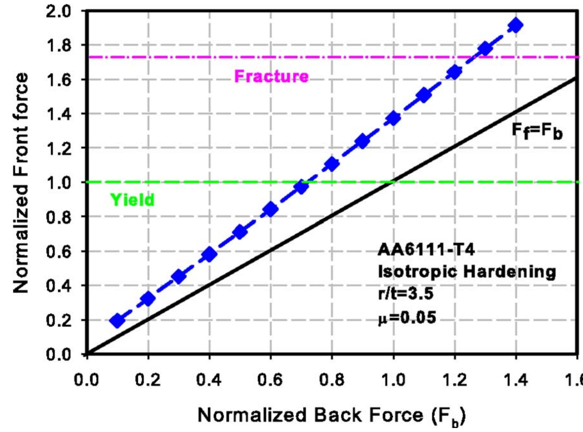


Fig. 12 Normalized back force versus normalized front force

curves in Fig. 11 shows that the moment at the instance of unloading drops significantly as the back force increases. In Fig. 10, the springback angle generally decreases linearly but it is notable that approximately two linear regions exist. The change of slopes occurs between the normalized forces 0.5 and 0.8. The front force corresponding to T_{III} or T_{IV} which is increased as the material passes the cylindrical tool is calculated. As shown in Fig. 12, the normalized tension at region III (or region IV) becomes greater than unity when the back force falls between 0.6 and 0.8. Therefore, the sudden change of slope might be caused by the transition from the partially plastic to fully plastic response through the thickness of the material element. This sudden change of springback when the material passes over the elastic region was also discussed by comparing measurements and FE analysis predictions for aluminum alloy 6022-T4 [15]. It was observed that the secondary curvature appeared for the large back force, called the anticlastic curvature, which persists throughout unloading, thus presenting a greater effective cross-sectional moment of inertia resisting the principal springback. The current simple plane strain analysis cannot show the anticlastic secondary curvature although it shows the change of slope.

Figures 13–15 show the predicted dependence of the springback angle with friction. As friction increases, springback decreases since tangential force increases in regions III and IV when the material passes through the cylindrical tool, but friction has a smaller effect than other process parameters.

4.2 Material Effects. The proper material model is important for the accurate prediction of springback. In general, the continuum constitutive equation of plasticity consists of the boundary of the elastic region (by the yield function) and its evolution (by the hardening law). Here, two main aspects of material description were considered by investigating the effect of the hardening law and the yield function shape.

For the hardening model, the nonlinear isotropic-kinematic hardening model was considered to represent the common feature

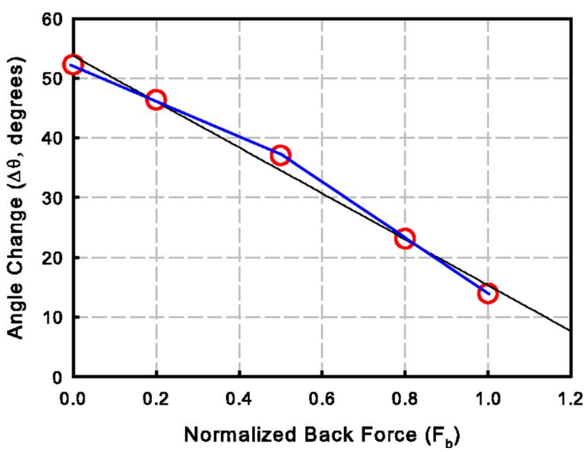


Fig. 10 Variation of the springback angle with the normalized back force (F_b)

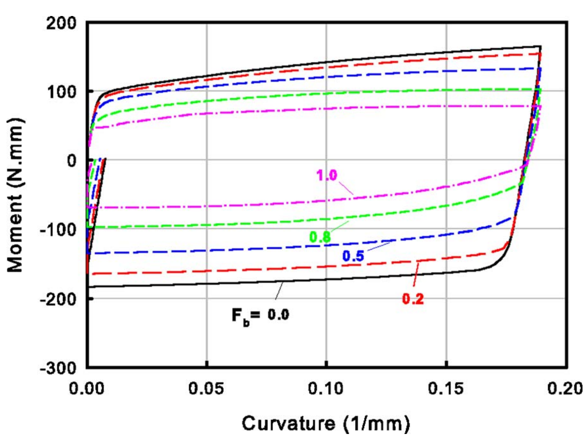


Fig. 11 Effect of normalized back force (F_b) on moment-curvature curves

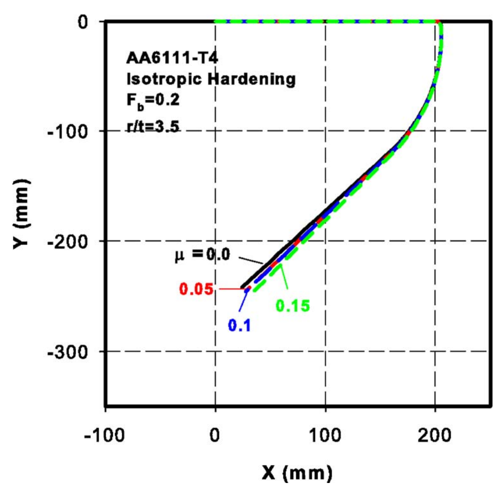


Fig. 13 Final shapes for the variation of friction coefficient

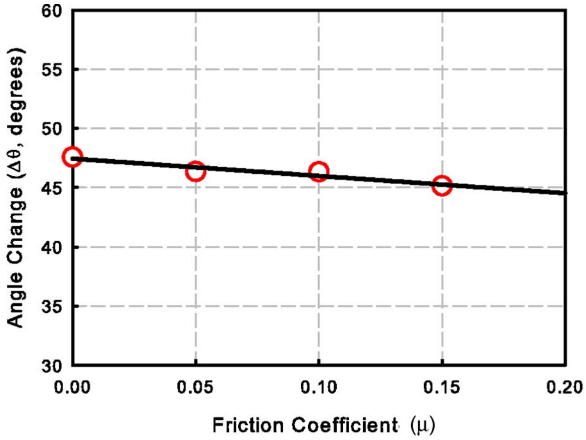


Fig. 14 Variation of the springback angle with friction coefficient

of stress-strain response for loading, unloading, and reverse loading. For the reference hardening curve, which is usually the uniaxial tensile curve, two parts can be defined,

$$\bar{\sigma} = \bar{\sigma}_{\text{iso}} + \bar{\sigma}_{\text{kine}} = \{\bar{\sigma}_0 + \zeta q(1 - \exp(-b\bar{\varepsilon}))\} + (1-\zeta)q(1 - \exp(-b\bar{\varepsilon})) \tag{34}$$

where q and b are material parameters. Constant ζ determines the ratio between the size change and translation of the yield surface. For $\zeta=1$, the yield surface only expands without translation, while it translates without size change for $\zeta=0$.

For the evolution of back stress, the Armstrong-Frederic-type nonlinear kinematic hardening rule [36] becomes, for the uniaxial tensile case,

$$d\alpha = C_1 d\varepsilon^p - C_2 \alpha d\varepsilon^p \tag{35}$$

Integrating Eq. (35) and comparing it with the kinematic part of hardening in Eq. (34) gives

$$C_2 = b \text{ and } C_1 = (1-\zeta)qC_2 \tag{36}$$

Figures 16–18 show deformed shapes and springback angles for different hardening parameters ζ for low back force ($F_b=0.2$) and high back force ($F_b=0.8$), respectively. In general, larger springback is predicted when the material approaches pure isotropic, $\zeta=1$. However, when the large back force is applied, the difference becomes insignificant as shown in Fig. 16(*b*). Note that the springback of pure kinematic hardening for large back force is similar to that of isotropic hardening. The moment-curvature curves of a sidewall curl element are compared in Fig. 18, which shows that the magnitude of moment at the end of the full unbending (up to zero curvature) after bending becomes lower as the material approaches pure kinematic with more severe Bauschinger effect. During the full unloading (at the zero curvature), the material behavior is generally elastic. However, for kinematic hardening, full unloading may incur elastic-plastic deformation especially if the stress before unloading is large enough, while the Bauschinger effect is severe. The magnified figure in Fig. 18(*b*) shows that unloading curves are linear in most hardening cases except for pure kinematic hardening, in which the elastic-plastic transition induces larger springback.

The dependence of springback on the shape of yield surface has been investigated. The yield surface can represent from the von Mises yield surface ($m=2$) to the Tresca yield surface (m =infinity) by varying the exponent value. In Fig. 19, three yield surfaces with different exponents are schematically shown. It is commonly recommended that exponent 6 for bcc and 8 for fcc metals. Three exponent values 2, 6, and 8 were chosen to show the effect of yield surface shape on the magnitude of springback. The corresponding plane strain parameter β values are 1.1547, 1.1167, and 1.1089, respectively. In Figs. 20–22 show that springback increases as the exponent decreases. As the exponent decreases, β increases in Eq. (6), which leads to the increase of stress and moment in the moment-curvature curve as shown in Fig. 22 for the same uniaxial hardening behavior, therefore, inducing the increase of springback.

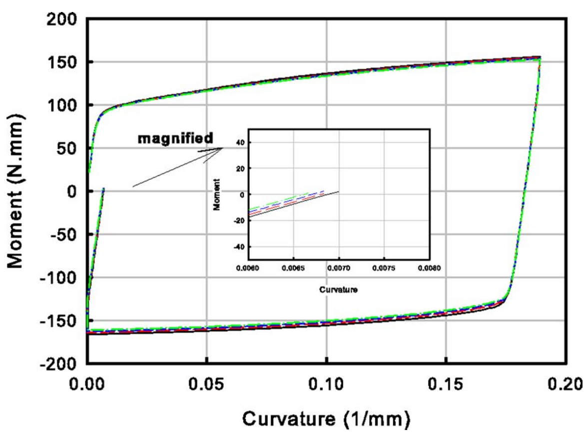


Fig. 15 Effect of friction on moment-curvature curves

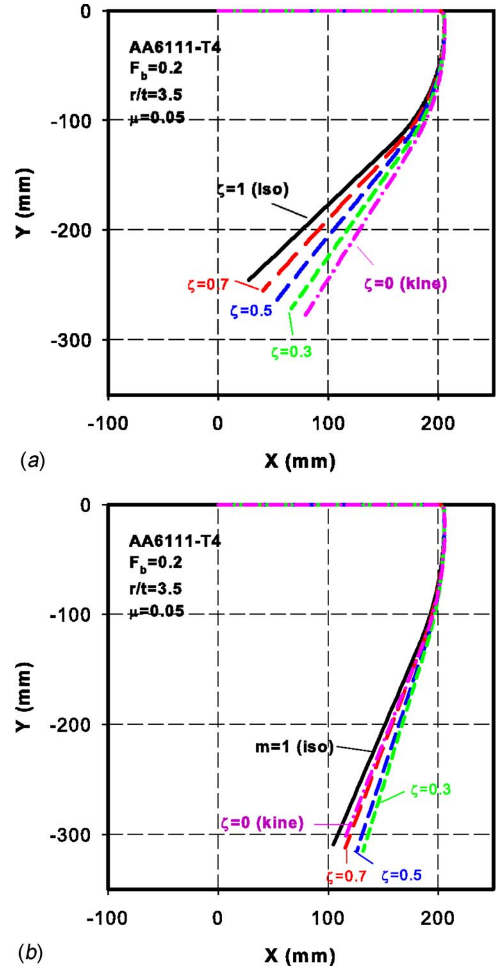


Fig. 16 Final shapes for the variation of hardening laws: (*a*) low back force, (*b*) high back force

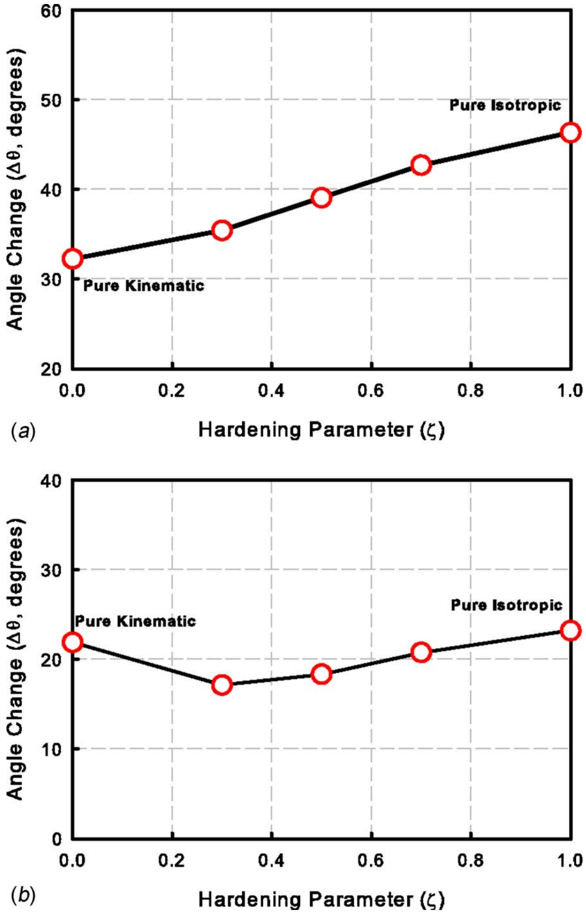


Fig. 17 Variation of the springback angle with hardening laws: (*a*) low back force, (*b*) high back force

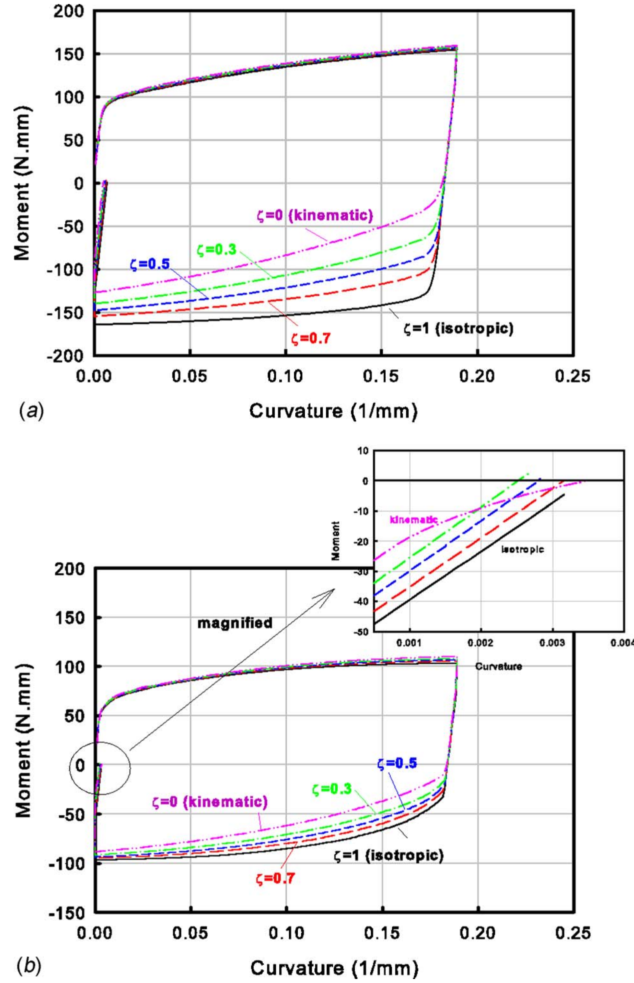


Fig. 18 Effect of hardening laws on moment-curvature curves: (*a*) low back force, (*b*) high back force

5 Verification

In order to verify the prediction capability of the present semi-analytic hybrid method, calculated springback results were compared with experimental measurements. A dual-phase high strength steel (DP-Steel) was selected as a sample material because significant springback is attainable compared to the mild steel. The tensile curve of the DP-Steel in the rolling direction is shown in Fig. 23 with its fit by the Voce equation. Isotropic elasticity with Young's modulus, 200 GPa and Poisson ratio, 0.3 were used. For plastic hardening, combined isotropic-kinematic hardening was applied along with the mixed hardening parameter $\zeta = 0.7$ in Eq. (30), which was obtained in the previous study [17] to represent the Bauschinger effect. As for the anisotropic yield function, Yld2000-2d with the exponent 6 for the bcc crystal structure was used and the anisotropic parameters were obtained from the previous work [41].

Two different sizes of the cylindrical tool (therefore, two different r/t ratios) with various normalized back forces were considered. For all cases, friction was controlled by applying the standard industrial lubricant MP404 to cylindrical tools for measurement. For simulations, the friction coefficient was assumed to be $\mu = 0.1$ based on the previous work with the same lubricant and sheet material [18] and negligible effect of friction from the sensitivity tests in the previous section. The same sheet strips and draw distance were used as done in the previous section.

Figure 24 shows the calculated and measured deformed shapes of sheet strips after springback with various back forces in the case of $r/t = 11.28$. The measured springback shows that the magnitude decreases as the back force increases, which is consistent with the result of the sensitivity test. In the figure, the combined isotropic-kinematic hardening predicts the springback for $F_b = 0.2 - 0.8$ although some deviations are observed for $F_b = 1.1 - 1.3$. Figure 25 shows deformed shapes after springback for smaller die radius, $r/t = 4.8$. For the two different r/t ratios, good prediction capability is shown especially in the range of low con-

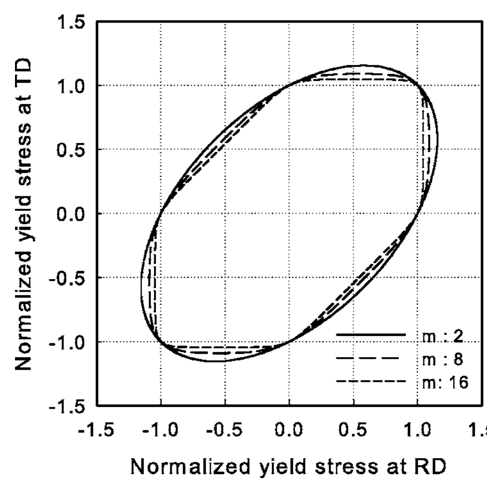


Fig. 19 Schematic shape of the Yld2000-2d surface with three different exponents

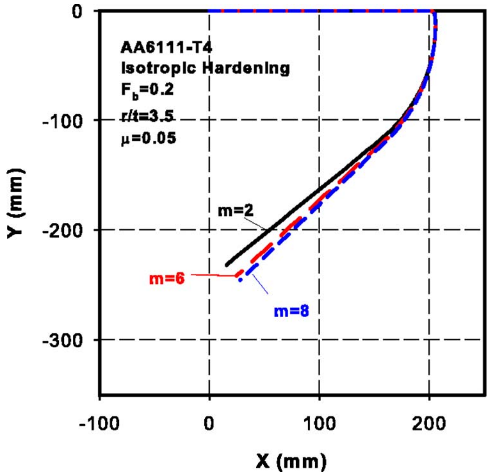


Fig. 20 Final shapes for the variation of the yield surface exponent

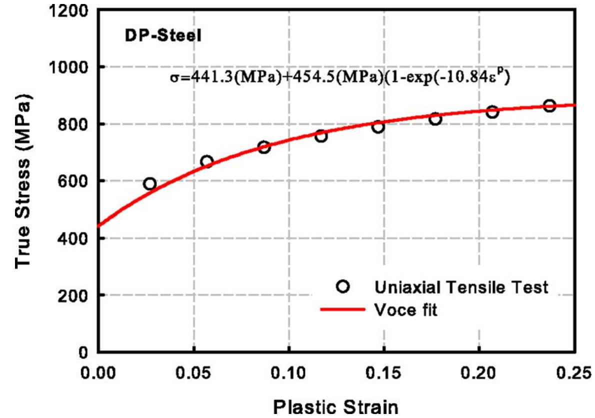


Fig. 23 Measured stress-strain curve and the Voce fit of DP-Steel

straint forces as shown in Figs. 24 and 25. The sudden reduction in measured springback angle is observed for $F_b=0.8-1.1$, while simulated springback angles change almost linearly in the whole test region as shown in Figs. 26 and 27.

Wang et al. [42] and Li et al. [14] analyzed the sudden decrease of the springback angle under the particular condition by introducing the concept of persistent anticlastic curvature (or secondary curvature). According to their observation in the present draw bend test, anticlastic curvature is developed in the unbending process during forming and it persists after springback when the applied sheet tension exceeds a critical value near yielding. They concluded that for the current draw bend test the springback process is closer to plane stress rather than plane strain because of the persistent anticlastic curvature, which cannot be accounted for by the current plane strain analysis. The secondary curvatures can be reproduced by adopting higher order finite element analysis utilizing 3D continuum or shell elements [14].

The comparisons of calculated predictions with experimental measurements show that the developed numerical method can be used as an effective tool to predict the springback with reasonable accuracy even if fine resolution of material properties and process, which might be achieved by more costly numerical methods such as FEM, may not be attainable. Also, high numerical efficiency can be achieved. For example, the CPU time of standard Pentium 2.8 GHz processor with the current hybrid method is approximately 65 s, while over 480 s with commercial finite element program ABAQUS/Standard under similar conditions.

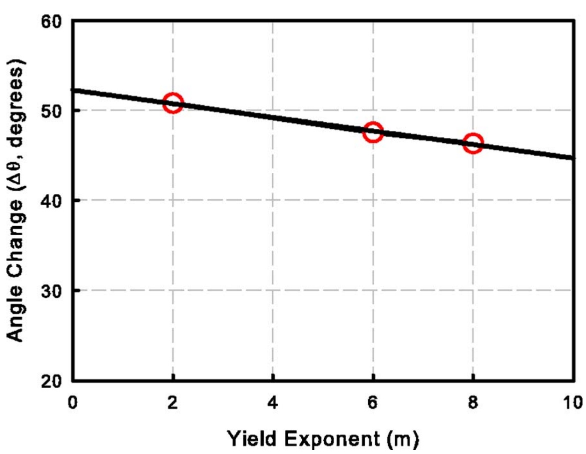


Fig. 21 Variation of the springback angle with the yield surface exponent

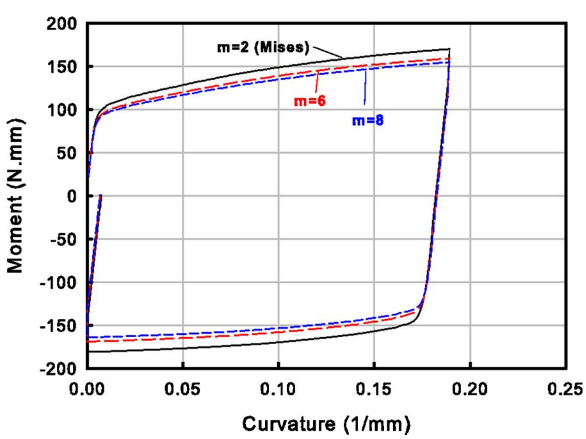


Fig. 22 Effect of yield surface exponent on moment-curvature curves

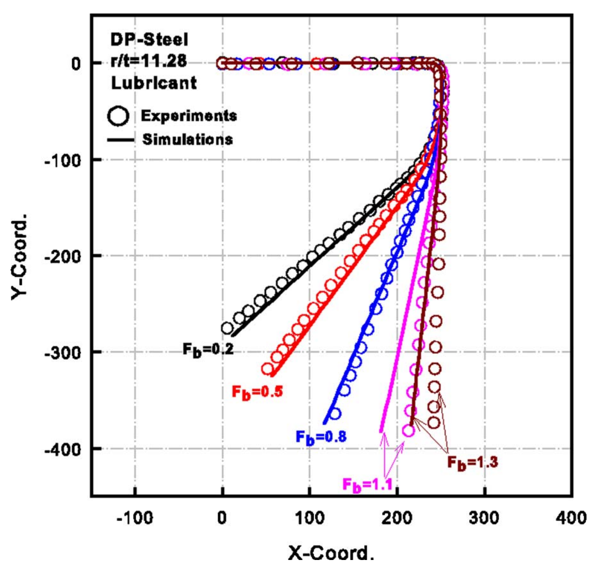


Fig. 24 Comparison between predicted and measured deformed shapes after springback with various normalized back forces for $r/t=11.28$

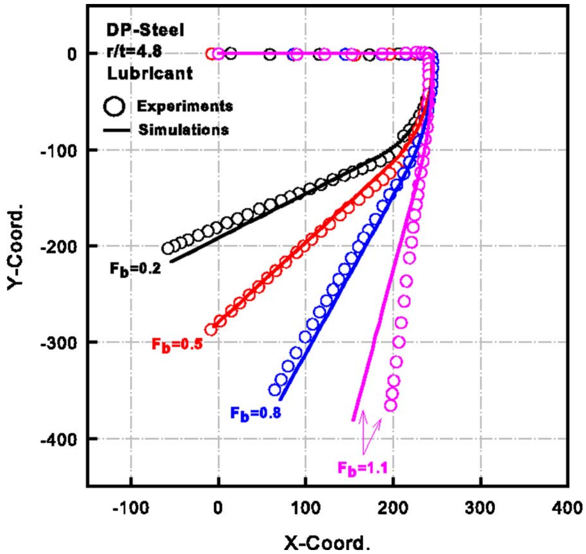


Fig. 25 Comparison between predicted and measured deformed shapes after springback with variant of normalized back force for r/t=4.8

6 Summary

In the present paper, a semianalytic hybrid method has been derived for the simulation of springback in a draw bend test under the plane strain condition. The current simple approach is a fast solution-generating method, which is numerically effective in considering various important processes and material parameters. Sensitivity tests for the effects of process and material parameters on springback were investigated using this simple method. Regarding the effect of process parameters, springback decreases as the r/t ratio, constraining back force and friction between sheets and tools increase. As for the effect of material properties, springback increases as the Bauschinger effect decreases from pure isotropic hardening to pure kinematic hardening However, the elastic-plastic unloading may increase the springback at the sidewall region for pure kinematic hardening especially when restraining back force becomes large. For the yield surface shape, springback increases as the sharpness of yield surface corners decreases. The prediction capability of the developed numerical scheme was verified by comparing the deformed shapes of a dual-phase steel sheet after springback. The results showed reasonably good agreements with measured results although small deviation existed for large back force.

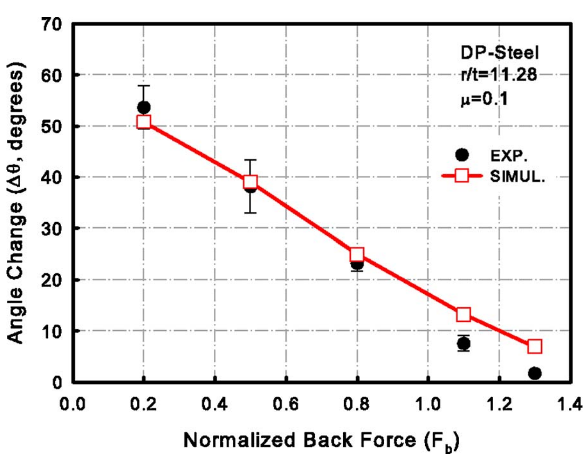


Fig. 26 Magnitude of the springback angle with normalized back force (r/t=11.28)

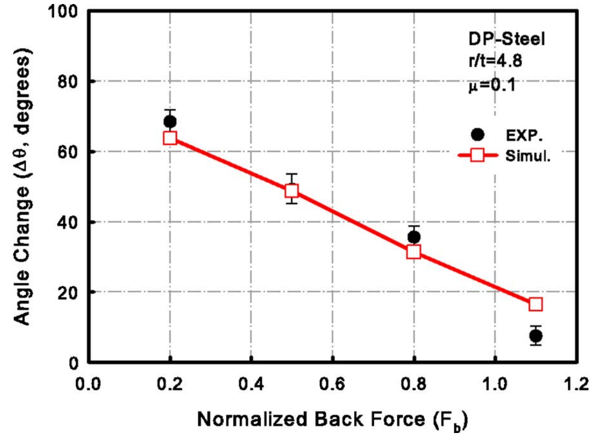


Fig. 27 Magnitude of the springback angle with normalized back force (r/t=4.8)

Acknowledgment

This work was supported by the SRC/ERC Program of MOST/KOSEF (R11-2005-065) and by the National Science Foundation (NSF) (DMI-0355429). MGL also appreciates the partial support from the Center for Advanced Materials Processing of the 21st Century Frontier R&D Program by MOST.

References

[1] Gardiner, F. J., 1957, "The Springback of Metals," Trans. ASME, **79**(1), pp. 1–9.
[2] Queener, C. A., and De Angelis, R. J., 1968, "Elastic Springback and Residual Stresses in Sheet Formed by Bending," Trans. Inst. Electron., Inf. Commun. Eng. D-II, **61**, pp. 757–768.
[3] Baba, A., 1964, "Effect of Tensile Force in Stretch-Forming Process on the Springback," Bull. JSME, **7**, pp. 834–843.
[4] Wang, C., Kinzel, G., and Altan, T., 1993, "Mathematical Modeling of Plane-Strain Bending of Sheet and Plate," J. Mater. Process. Technol., **39**, pp. 279–304.
[5] Chan, K. C., and Wang, S. H., 1999, "Theoretical Analysis of Springback in Bending of Integrated Circuit Leadframes," J. Mater. Process. Technol., **91**(13), pp. 111–115.
[6] Johnson, W., and Yu, T. X., 1981, "Springback After the Biaxial Elastic-Plastic Pure Bending of a Rectangular Plate," Int. J. Mech. Sci., **23**(10), pp. 619–30.
[7] Yu, T. X., and Johnson, W., 1982, "Influence of Axial Force on the Elastic-Plastic Bending and Springback of a Beam," J. Mech. Work. Technol., **6**(1), pp. 5–21.
[8] Sudo, C., Kojima, M., and Matsuoka, T., 1974, "Some Investigations on Elastic Recovery of Press Formed Part," *Proceedings of the 8th Biennial Congress of IDDRG*, p. 192.
[9] Thompson, N. E., and Ellen, C. H., 1985, "A Simple Theory for Side-Wall Curl," Journal of Applied Metalworking, **4**(1), pp. 39–42.
[10] Jeunechamps, P. P., Ho, K. C., Lin, J., Ponthot, J. P., Dean, T. A., 2006, "A Closed Form Technique to Predict Springback in Creep Age-Forming," Int. J. Mech. Sci., **48**, pp. 621–629.
[11] Kawaguchi, T., Imatani, S., and Yamaguchi, K., 1994, "An Elasto-Viscoplastic Finite Element Analysis of Sheet Metal Bending Process," J. Jpn. Soc. Technol. Plast., **35**, pp. 125–130.
[12] Mattiasson, K., Strange, A., Thilderkvist, P., and Samuelsson, A., 1995, "Springback in Sheet Metal Forming," 5th International Conference on Numerical Methods in Industrial Forming Process, New York, pp. 115–124.
[13] He, N., and Wagoner, R. H., 1996, "Springback Simulation in Sheet Metal Forming," *Proceedings of Numisheet'96*, J. K. Lee, G. L. Kinzel, and R. H. Wagoner, eds., The Ohio State University, pp. 308–315.
[14] Li, K. P., Carden, W. P., and Wagoner, R. H., 2001, "Simulation of Springback," Int. J. Mech. Sci., **44**(1), pp. 103–122.
[15] Geng, L., and Wagoner, R. H., 2002, "Role of Plastic Anisotropy and Its Evolution on Springback," Int. J. Mech. Sci., **44**(1), pp. 123–148.
[16] Chung, K., Lee, M. G., Kim, D., Kim, C., Wenner, M. L., Barlat, F., 2005, "Springback Evaluation of Automotive Sheets Based on Isotropic-Kinematic Hardening Laws and Non-Quadratic Anisotropic Yield Functions—Part I: Theory and Formulation," Int. J. Plast., **21**(5), pp. 861–882.
[17] Lee, M. G., Kim, D., Kim, C., Wagoner, R. H., Wenner, M. L., and Chung, K.,

2005, "Springback Evaluation of Automotive Sheets Based on Isotropic-Kinematic Hardening Laws and Non-Quadratic Anisotropic Yield Functions—Part II: Characterizations of Material Properties," Int. J. Plast., **21**(5), pp. 883–914.
[18] Lee, M. G., Kim, D., Kim, C., Wenner, M. L., and Chung, K., 2005, "Springback Evaluation of Automotive Sheets Based on Isotropic-Kinematic Hardening Laws and Nonquadratic Anisotropic Yield Functions—Part III: Applications," Int. J. Plast., **21**(5), pp. 915–353.
[19] Bui, Q. V., Papeleux, L., and Ponthot, J. P., 2004, "Numerical Simulation of Springback Using Enhanced Assumed Strain Elements," J. Mater. Process. Technol., **153–154**, pp. 314–318.
[20] Simo, J. C., and Amero, F., 1992, "Geometrically Nonlinear Enhanced Strain Mixed Methods and the Method of Compatible Modes," Int. J. Numer. Methods Eng., **33**, pp. 1413–1449.
[21] Wenner, M. L., 1983, "On the Work Hardening and Springback in Plane Strain Draw Forming," Journal of Applied Metalworking, **2**(4), pp. 342–349.
[22] Zhang, J. T., and Lee., D., 1995, "Development of New Model for Plane Strain Bending and Springback Analysis," J. Mater. Eng. Perform., **4**, pp. 291–300.
[23] "NUMISHEET'93 Benchmark Problem," 1993, *Proceedings of the 2nd International Conference on Numerical Simulation of 3D Sheet Metal Forming Processes*, A. Makinouchi, E. Nakamachi, E. Onate, and R. H. Wagoner, eds., Isehara, Japan.
[24] Pouboghrat, F., and Chu, E., 1995, "Springback in Plane Strain Stretch/Draw Sheet Forming," Int. J. Mech. Sci., **36**(3), pp. 327–341.
[25] Pouboghrat, F., Chung, K., and Richmond, O., 1998, "A Hybrid Membrane/Shell Method for Rapid Estimation of Springback in Anisotropic Sheet Metals," ASME J. Appl. Mech., **65**, pp. 671–684.
[26] Yoon, J. W., Pouboghrat, F., Chung, K., Yang, D., 2002, "Springback Prediction for Sheet Metal Forming Process Using a 3D Hybrid Membrane/Shell Method," Int. J. Mech. Sci., **44**(10), pp. 2133–2153.
[27] Carden, W. D., Geng, L. M., Matlock, D. K., and Wagoner, R. H., 2002, "Measurement of Springback," Int. J. Mech. Sci., **44**(1), pp. 79–101.
[28] Wang, J. F., Wagoner, R. H., Carden, W. D., Matlock, D. K., and Barlat, F., 2004, "Creep and Anelasticity in the Springback of Aluminum," Int. J. Plast., **20**(12), pp. 2209–2232.
[29] Barlat, F., Brem, J. C., Yoon, J. W., Chung, K., Dick, R. E., Choi, S. H., Pouboghrat, F., Chu, E., and Lege, D. J., 2003, "Plane Stress Yield Function for Aluminum Alloy Sheets," Int. J. Plast., **19**(9), pp. 1297–1319.
[30] Chaboche, J. L., 1986, "Time Dependent Constitutive Theories for Cyclic Plasticity," Int. J. Plast., **2**, pp. 149–188.
[31] Chun, B. K., Jinn, J. T., and Lee, J. K., 2002, "Modeling the Bauschinger Effect for Sheet Metals—Part I: Theory," Int. J. Plast., **18**(5), pp. 571–595.
[32] Ohno, N., and Wang, J. D., 1993, "Nonlinear Hardening Rules with Critical State of Dynamics Recovery—Part 1: Formulation and Basic Features for Ratcheting Behavior," Int. J. Plast., **9**(3), pp. 375–390.
[33] Ohno, N., and Wang, J. D., 1993, "Nonlinear Hardening Rules with Critical State of Dynamics Recovery—Part 2: Applications to Experiments of Ratcheting Behavior," Int. J. Plast., **9**(3), pp. 391–403.
[34] Ohno, N., and Katchi, Y., 1986, "A Constitutive Model of Cyclic Plasticity for Nonlinear Hardening Materials," ASME J. Appl. Mech., **53**(2), pp. 395–403.
[35] Bower, A. F., 1989, "Cyclic Hardening Properties of Hard-Drawn Copper and Rail Steel," J. Mech. Phys. Solids, **37**, pp. 455–470.
[36] Armstrong, P. J., and Frederick, C. O., 1966, "A Mathematical Representation of the Multiaxial Bauschinger Effect," G.E.G.B. Report No. RD/B/N 731.
[37] Boger, R. K., Wagoner, R. H., Barlat, F., Lee, M. G., and Chung, K., 2005, "Continuous, Large Strain, Tension/Compression Testing of Sheet Material," Int. J. Plast., **21**, pp. 2319–2343.
[38] Stoughton, T. B., 1988, "Model of Drawbead Forces in Sheet Metal Forming," *Proceedings of the 15th IDDRG, Dearborn*, USA and Toronto, Canada, pp. 205–215.
[39] Marciniak, Z., and Duncan, J., 1992, *Mechanics of Sheet Metal Forming*, Edward Arnolds, London, UK.
[40] Wagoner, R. H., Gan, W., and Kovacs, W., 2005, "Characterization of Friction-Stir Welds for Tailor-Welded Blank Applications," First year report for GM R&D Center.
[41] Lee, W., Kim, J., Ryou, H., Kim, D., Kim, C., Wenner, M. L., Chung, K., 2006, "Numerical Sheet Forming Simulation of Friction Stir Welded TWB Automotive Sheets," Final Report for GM R&D Center.
[42] Wang, J. F., Wagoner, R. H., Matlock, D. K., and Barlat, F., 2005, "Anticlastic Curvature in Draw-Bend Springback," Int. J. Solids Struct., **42**(12), pp. 1287–1307.

On the Stoney Formula for a Thin Film/Substrate System With Nonuniform Substrate Thickness

X. Feng

Y. Huang

Department of Mechanical and Industrial Engineering,
University of Illinois,
Urbana, IL 61801

A. J. Rosakis

Graduate Aeronautical Laboratory,
California Institute of Technology,
Pasadena, CA 91125

*Current methodologies used for the inference of thin film stress through system curvature measurements are strictly restricted to stress and curvature states which are assumed to remain uniform over the entire film/substrate system. Recently Huang, Rosakis, and coworkers [Acta Mech. Sinica, **21**, pp. 362–370 (2005); J. Mech. Phys. Solids, **53**, 2483–2500 (2005); Thin Solid Films, **515**, pp. 2220–2229 (2006); J. Appl. Mech., in press; J. Mech. Mater. Struct., in press] established methods for the film/substrate system subject to nonuniform misfit strain and temperature changes. The film stresses were found to depend nonlocally on system curvatures (i.e., depend on the full-field curvatures). These methods, however, all assume uniform substrate thickness, which is sometimes violated in the thin film/substrate system. Using the perturbation analysis, we extend the methods to nonuniform substrate thickness for the thin film/substrate system subject to nonuniform misfit strain.* [DOI: 10.1115/1.2745392]

Keywords: thin films, nonuniform misfit strain, nonuniform substrate thickness, nonlocal stress-curvature relations, interfacial shears

1 Introduction

Stoney [1] used a plate system composed of a stress bearing thin film, of uniform thickness h_f, deposited on a relatively thick substrate, of uniform thickness h_s, and derived a simple relation between the curvature, κ, of the system and the stress, $\sigma^{(f)}$, of the film as follows:

$$\sigma^{(f)} = \frac{E_s h_s^2 \kappa}{6h_f(1-\nu_s)} \tag{1}$$

In the above the subscripts f and s denote the thin film and substrate, respectively, and E and ν are the Young's modulus and Poisson's ratio, respectively. Equation (1) is called the Stoney formula, and it has been extensively used in the literature to infer film stress changes from experimental measurement of system curvature changes [2].

Stoney's formula involves the following assumptions:

(i) Both the film thickness h_f and substrate thickness h_s are uniform, the film and substrate have the same radius R, and $h_f \ll h_s \ll R$;
(ii) The strains and rotations of the plate system are infinitesimal;
(iii) Both the film and substrate are homogeneous, isotropic, and linearly elastic;
(iv) The film stress states are in-plane isotropic or equibiaxial (two equal stress components in any two, mutually orthogonal in-plane directions) while the out-of-plane direct stress and all shear stresses vanish;
(v) The system's curvature components are equibiaxial (two equal direct curvatures) while the twist curvature vanishes in all directions; and
(vi) All surviving stress and curvature components are spatially constant over the plate system's surface, a situation which is often violated in practice.

Despite the explicitly stated assumptions, the Stoney formula is often arbitrarily applied to cases of practical interest where these assumptions are violated. This is typically done by applying Stoney's formula pointwise and thus extracting a local value of stress from a local measurement of the system curvature. This approach of inferring film stress clearly violates the uniformity assumptions of the analysis and, as such, its accuracy as an approximation is expected to deteriorate as the levels of curvature nonuniformity become more severe.

Following the initial formulation by Stoney, a number of extensions have been derived to relax some assumptions. Such extensions of the initial formulation include relaxation of the assumption of equibiaxiality as well as the assumption of small deformations/deflections. A biaxial form of Stoney formula (with different direct stress values and nonzero in-plane shear stress) was derived by relaxing the assumption (v) of curvature equibiaxiality [2]. Related analyses treating discontinuous films in the form of bare periodic lines [3] or composite films with periodic line structures (e.g., bare or encapsulated periodic lines) have also been derived [4–6]. These latter analyses have removed assumptions (iv) and (v) of equibiaxiality and have allowed the existence of three independent curvature and stress components in the form of two, nonequal, direct components and one shear or twist component. However, the uniformity assumption (vi) of all of these quantities over the entire plate system was retained. In addition to the above, single, multiple and graded films and substrates have been treated in various "large" deformation analyses [7–10]. These analyses have removed both the restrictions of an equibiaxial curvature state as well as the assumption (ii) of infinitesimal deformations. They have allowed for the prediction of kinematically nonlinear behavior and bifurcations in curvature states that have also been observed experimentally [11,12]. These bifurcations are transformations from an initially equibiaxial to a subsequently biaxial curvature state that may be induced by an increase in film stress beyond a critical level. This critical level is intimately related to the systems aspect ratio, i.e., the ratio of in-plane to thickness dimension and the elastic stiffness. These analyses also retain the assumption (vi) of spatial curvature and stress uniformity across the system. However, they allow for deformations to evolve from an initially spherical shape to an energetically

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received October 18, 2006; final manuscript received January 14, 2007. Review conducted by Robert M. McMeeking.

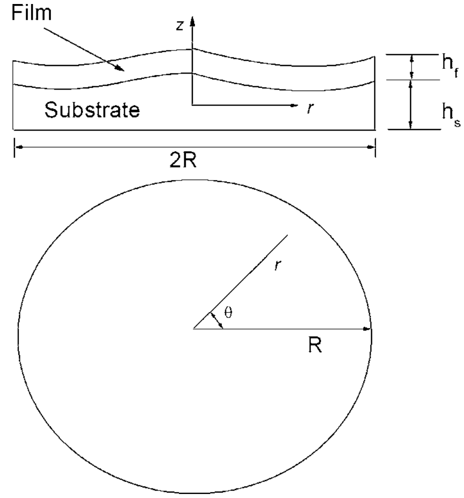


Fig. 1 A schematic diagram of a thin film/substrate system with the cylindrical coordinates (r, θ, z)

favored shape (e.g., ellipsoidal, cylindrical or saddle shapes) that features three different, still spatially constant, curvature components [11,12].

The above-discussed extensions of Stoney's methodology have not relaxed the most restrictive of Stoney's original assumption (vi) of spatial uniformity which does not allow either film stress and curvature components to vary across the plate surface. This crucial assumption is often violated in practice since film stresses and the associated system curvatures are nonuniformly distributed over the plate area. Recently Huang et al. [13] and Huang and Rosakis [14] relaxed the assumption (vi) [and also (iv) and (v)] to study the thin film/substrate system subject to non-uniform, axisymmetric misfit strain (in thin film) and temperature change (in both thin film and substrate), respectively, while Ngo et al. [15] studied the thin film/substrate system subject to arbitrarily nonuniform (e.g., nonaxisymmetric) misfit strain and temperature. The most important result is that *the film stresses depend nonlocally on the substrate curvatures*, i.e., they depend on curvatures of the entire substrate. The relations between film stresses and substrate curvatures are established for arbitrarily nonuniform misfit strain and temperature change, and such relations degenerate to Stoney's formula for uniform, equibiaxial stresses and curvatures.

Feng et al. [16] relaxed part of the assumption (i) to study the thin film and substrate of different radii, i.e., the thin film has a smaller radius than the substrate. Ngo et al. [15] further relaxed the assumption (i) for arbitrarily nonuniform thickness of the thin film. The main purpose of the present paper is to relax the remaining portion in assumption (i), i.e., the uniform thickness of the substrate. To do so we consider the case of thin film/substrate system with nonuniform substrate thickness subject to nonuniform misfit strain field in the thin film. Our goal is to relate film stresses and system curvatures to the misfit strain distribution, and to ultimately derive a relation between the film stresses and the system curvatures that would allow for the accurate experimental inference of film stress from full-field and real-time curvature measurements.

2 Governing Equations and Boundary Conditions

Consider a thin film of uniform thickness h_f which is deposited on a circular substrate of thickness h_s and radius R (Fig. 1). The substrate thickness is nonuniform, but is assumed to be axisymmetric $h_s = h_s(r)$ for simplicity, where r and θ are the polar coordinates. The film is very thin, $h_f \ll h_s$, such that it is modeled as a membrane, and is subject to nonuniform misfit strain ε_m. Here the misfit strain is also assumed to be axisymmetric $\varepsilon_m = \varepsilon_m(r)$ for simplicity. The substrate is modeled as a plate since $h_s \ll R$. The Young's modulus and Poisson's ratio of the film and substrate are denoted by E_f, ν_f, E_s, and ν_s, respectively.

Let u_f and u_s denote the displacements in the radial direction in the thin film and substrate, respectively. The in-plane membrane strains are obtained from $\varepsilon_{\alpha\beta} = (u_{\alpha,\beta} + u_{\beta,\alpha})/2$ for infinitesimal deformation and rotation, where $\alpha, \beta = r, \theta$. The linear elastic constitutive model, together with the vanishing out-of-plane stress $\sigma_{zz} = 0$, give the in-plane stresses as

$$\sigma_{\alpha\beta} = \frac{E}{1-\nu^2}[(1-\nu)\varepsilon_{\alpha\beta} + \nu\varepsilon_{\kappa\kappa}\delta_{\alpha\beta} - (1+\nu)\varepsilon_m\delta_{\alpha\beta}],$$

where $E, \nu = E_f, \nu_f$ in the thin film and E_s, ν_s in the substrate, and the misfit strain ε^m is only in the thin film. The nonvanishing axial forces in the thin film and substrate are

$$N_r = \frac{Eh}{1-\nu^2}\left[\frac{du_r}{dr} + \nu\frac{u_r}{r} - (1+\nu)\varepsilon_m\right]$$
$$N_\theta = \frac{Eh}{1-\nu^2}\left[\nu\frac{du_r}{dr} + \frac{u_r}{r} - (1+\nu)\varepsilon_m\right] \tag{2}$$

where $h = h_f$ in the thin film and $h_s(r)$ in the substrate, and once again the misfit strain ε_m is only in the thin film.

Let w denote the lateral displacement in the normal (z) direction. The curvatures are given by $\kappa_{\alpha\beta} = w_{,\alpha\beta}$. The bending moments in the substrates are

$$M_r = \frac{E_s h_s^3}{12(1-\nu_s^2)}\left(\frac{d^2w}{dr^2} + \nu_s\frac{1}{r}\frac{dw}{dr}\right)$$
$$M_\theta = \frac{E_s h_s^3}{12(1-\nu_s^2)}\left(\nu_s\frac{d^2w}{dr^2} + \frac{1}{r}\frac{dw}{dr}\right) \tag{3}$$

For nonuniform misfit strain distribution $\varepsilon_m = \varepsilon_m(r)$, the shear stress along the radial direction at the film/substrate interface does not vanish, and is denoted by τ. The in-plane force equilibrium equations for the thin film and substrate, accounting for the effect of interface shear stress τ, becomes

$$\frac{dN_r}{dr} + \frac{N_r - N_\theta}{r} \mp \tau = 0 \tag{4}$$

where the minus sign in front of the interface shear stress is for the thin film, and the plus sign is for the substrate. The moment and out-of-plane force equilibrium equations for the substrate are

$$\frac{dM_r}{dr} + \frac{M_r - M_\theta}{r} + Q - \frac{h_s}{2}\tau = 0 \tag{5}$$

$$\frac{dQ}{dr} + \frac{Q}{r} = 0 \tag{6}$$

where Q is the shear force normal to the neutral axis. Equation (6), together with the requirement of finite Q at $r=0$, gives $Q=0$.

The substitution of Eq. (2) into (4) yields the governing equations for u and τ,

$$\frac{d^2u_f}{dr^2} + \frac{1}{r}\frac{du_f}{dr} - \frac{u_f}{r^2} = \frac{1-\nu_f^2}{E_f h_f}\tau + (1+\nu_f)\frac{d\varepsilon_m}{dr} \tag{7}$$

$$\frac{d}{dr}\left[h_s\left(\frac{du_s}{dr} + \frac{u_s}{r}\right)\right] - (1-\nu_s)\frac{dh_s}{dr}\frac{u_s}{r} = -\frac{1-\nu_s^2}{E_s}\tau \tag{8}$$

Equations (3), (5), and (6) give the governing equation for w and τ,

$$\frac{d}{dr}\left[h_s^3\left(\frac{d^2w}{dr^2} + \frac{1}{r}\frac{dw}{dr}\right)\right] - (1-\nu_s)\frac{1}{r}\frac{dh_s^3}{dr}\frac{dw}{dr} = \frac{6(1-\nu_s^2)}{E_s}h_s\tau \tag{9}$$

The continuity of displacements across the film/substrate interface requires

$$u_f = u_s - \frac{h_s}{2}\frac{dw}{dr} \tag{10}$$

Equations (7)–(10) constitute four ordinary differential equations (ODEs) for u_f, u_s, w, and τ. The ODEs are linear, but have nonconstant coefficients.

The boundary conditions at the free edge $r=R$ require that the net forces and net moments vanish,

$$N_r^{(f)} + N_r^{(s)} = 0 \tag{11}$$

$$M_r - \frac{h_s}{2}N_r^{(f)} = 0 \tag{12}$$

where the superscripts f and s denote the film and substrate, respectively.

3 Perturbation Method for Small Variation of Substrate Thickness

In the following we assume small variation of substrate thickness

$$h_s = h_{s0} + \Delta h_s = h_{s0} + \beta h_{s1} \tag{13}$$

where h_{s0} (=constant) is the average substrate thickness, and $\Delta h_s(r)$ is the substrate thickness variation which satisfies $|\Delta h_s| \ll h_{s0}$; $\Delta h_s(r)$ is also written as βh_{s1} in (13), where $0 < \beta \ll 1$ is a small, positive constant, and $h_{s1}=h_{s1}(r)$ is on the same order as h_{s0}.

We use the perturbation method to solve the ODEs analytically for $\beta \ll 1$. Two possible scenarios are considered separately in the following:

(i) The substrate thickness variation Δh_s is on the same order as the thin film thickness h_f, i.e., $\Delta h_s \sim h_f$. This is represented by $\beta = h_f/h_{s0}$ ($\ll 1$). For this case the film stresses and system curvatures are identical to their counterparts for a constant substrate thickness h_{s0}. This is because the Stoney formula (1), as well as all its extensions, holds only for thin films, $h_f \ll h_s$. As compared to unity (one), terms that are on the order of $O(h_f/h_s)$ are always neglected. In this case the difference between the film stresses (or system curvatures, $\cdots$) for nonuniform substrate thickness h_s and those for constant thickness h_{s0} is on the order of $O(\Delta h_s/h_{s0})$ (as compared to unity), which is the same as $O(h_f/h_s)$ since $\Delta h_s \sim h_f$, and is therefore negligible.

(ii) The substrate thickness variation Δh_s is much larger than the thin film thickness h_f, i.e., $|\Delta h_s| \gg h_f$. This is represented by $h_f/h_{s0} \ll \beta$ ($\ll 1$). In the following we focus on this case and use the perturbation method (for $\beta \ll 1$) to obtain the analytical solution.

Elimination of τ from (7) and (8) yields an equation for u_f and u_s. For $h_f/h_{s0} \ll 1$, u_f disappears in this equation, which becomes the governing equation for u_s,

$$\frac{d}{dr}\left[h_s\left(\frac{du_s}{dr} + \frac{u_s}{r}\right)\right] - (1-\nu_s)\frac{dh_s}{dr}\frac{u_s}{r} = \frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s^2}{E_s}\frac{d\varepsilon_m}{dr} \tag{14}$$

The above equation, together with (8), gives the interface shear stress

$$\tau = -\frac{E_f h_f}{1-\nu_f}\frac{d\varepsilon_m}{dr} \tag{15}$$

This is a remarkable result that holds regardless of the substrate thickness and boundary conditions at the edge $r=R$. Therefore, the interface shear stress is proportional to the gradient of misfit strain. For uniform misfit strain $\varepsilon_m(r)$=constant, the interface shear stress vanishes (even for nonuniform substrate thickness).

We use the perturbation method to write u_s as

$$u_s = u_{s0} + \beta u_{s1} \tag{16}$$

where $\beta \ll 1$, u_{s0} is the solution for a constant substrate thickness h_{s0}, and is given by Huang et al. [13]

$$u_{s0} = \frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s^2}{E_s h_{s0}}\left[\frac{1}{r}\int_0^r \eta\varepsilon_m(\eta)d\eta + \frac{1-\nu_s}{1+\nu_s}\frac{\overline{\varepsilon_m}}{2}r\right] \tag{17}$$

and

$$\overline{\varepsilon_m} = \frac{2}{\pi R^2}\int_0^R \eta\varepsilon_m d\eta$$

is the average misfit strain in the thin film; u_{s1} in (16) is on the same order as u_{s0}. In the following we use u' to denote du/dr. The substitution of (16) and (17) into (14) and the neglect of $O(\beta^2)$ terms give the following linear ODE with constant coefficients for u_{s1},

$$\left(u_{s1}' + \frac{u_{s1}}{r}\right)' = (1-\nu_s)\frac{h_{s1}'}{h_{s0}}\frac{u_{s0}}{r} - \left[\frac{h_{s1}}{h_{s0}}\left(u_{s0}' + \frac{u_{s0}}{r}\right)\right]' \tag{18}$$

Its general solution is

$$u_{s1}(r) = -\frac{h_{s1}}{h_{s0}}u_{s0} + \frac{1}{2r}\int_0^r \eta\left[1+\nu_s+(1-\nu_s)\frac{r^2}{\eta^2}\right]\frac{h_{s1}'(\eta)}{h_{s0}}u_{s0}(\eta)d\eta + \frac{A}{2}r \tag{19}$$

where the constant A is to be determined. The total substrate displacement is then given by

$$u_s(r) = \left(2-\frac{h_s}{h_{s0}}\right)u_{s0} + \frac{1}{2r}\int_0^r \eta\left[1+\nu_s+(1-\nu_s)\frac{r^2}{\eta^2}\right] \times \frac{h_s'(\eta)}{h_{s0}}u_{s0}(\eta)d\eta + \frac{\beta A}{2}r \tag{20}$$

The substitution of (15) into (9) yields the governing equation for the displacement w',

$$\left[h_s^3\left(w'' + \frac{w'}{r}\right)\right]' - (1-\nu_s)(h_s^3)'\frac{w'}{r} = -\frac{6E_f h_f}{1-\nu_f}\frac{1-\nu_s^2}{E_s}h_s\varepsilon_m' \tag{21}$$

Its perturbation solution can be written as

$$w' = w_0' + \beta w_1' \tag{22}$$

where w_0' is the solution for a constant substrate thickness h_{s0}, and is given by Huang et al. [13]

$$w_0' = -6\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s^2}{E_s h_{s0}^2}\left[\frac{1}{r}\int_0^r \eta\varepsilon_m(\eta)d\eta + \frac{1-\nu_s}{1+\nu_s}\frac{\overline{\varepsilon_m}}{2}r\right] \tag{23}$$

and once again

$$\overline{\varepsilon_m} = \frac{2}{\pi R^2}\int_0^R \eta\varepsilon_m d\eta$$

is the average misfit strain in the thin film; w_1' in (22) is on the same order as w_0'. Equations (21)–(23) give the following linear ODE with constant coefficients for w_1'

$$\left(w_1''+\frac{w_1'}{r}\right)'=-\frac{6E_fh_f}{1-\nu_f}\frac{1-\nu_s^2}{E_sh_{s0}^2}\frac{h_{s1}}{h_{s0}}\varepsilon_m'-3\left[\frac{h_{s1}}{h_{s0}}\left(w_0''+\frac{w_0'}{r}\right)\right]'+3(1-\nu_s)\frac{h_{s1}'}{h_{s0}}\frac{w_0'}{r} \tag{24}$$

Its general solution is

$$w_1'=-3\frac{h_{s1}}{h_{s0}}w_0'+\frac{3}{2r}\int_0^r\eta\left[1+\nu_s+(1-\nu_s)\frac{r^2}{\eta^2}\right]\frac{h_{s1}'(\eta)}{h_{s0}}w_0'(\eta)d\eta+\frac{B}{2}r+3\frac{E_fh_f}{1-\nu_f}\frac{1-\nu_s^2}{E_sh_{s0}^2}\frac{1}{r}\int_0^r\frac{d}{d\eta}\left[(r^2-\eta^2)\frac{h_{s1}(\eta)}{h_{s0}}\right]\varepsilon_m(\eta)d\eta \tag{25}$$

where the constant B is to be determined. The complete solution for w' is obtained from (22) as

$$w'=\left(4-3\frac{h_s}{h_{s0}}\right)w_0'+\frac{3}{2r}\int_0^r\eta\left[1+\nu_s+(1-\nu_s)\frac{r^2}{\eta^2}\right]\frac{h_s'(\eta)}{h_{s0}}w_0'(\eta)d\eta+\frac{\beta B}{2}r+3\frac{E_fh_f}{1-\nu_f}\frac{1-\nu_s^2}{E_sh_{s0}^2}\frac{1}{r}\int_0^r\frac{d}{d\eta}\left\{(r^2-\eta^2)\left[\frac{h_s(\eta)}{h_{s0}}-1\right]\right\}\times\varepsilon_m(\eta)d\eta \tag{26}$$

The displacement u_f in the thin film is then obtained from u_s in (20) and w' in (26) via (10).

The constants A and B, or equivalently, βA and βB, are determined from the boundary conditions (11) and (12) as

$$\beta A=-\frac{1-\nu_s}{R^2}\int_0^R\frac{R^2-\eta^2}{\eta}\frac{h_s'(\eta)}{h_{s0}}u_{s0}(\eta)d\eta \tag{27}$$

$$\beta B=-\frac{3(1-\nu_s)}{R^2}\int_0^R\frac{R^2-\eta^2}{\eta}\frac{h_s'(\eta)}{h_{s0}}w_0'(\eta)d\eta-6\frac{E_fh_f}{1-\nu_f}\frac{1-\nu_s}{E_sh_{s0}^2}\frac{1}{R^2}\times\int_0^R\frac{d}{d\eta}\left\{\left[(1+\nu_s)R^2+(1-\nu_s)\eta^2\right]\left[\frac{h_s(\eta)}{h_{s0}}-1\right]\right\}\varepsilon_m(\eta)d\eta \tag{28}$$

4 Thin-Film Stresses and System Curvatures

The system curvatures $\kappa_{rr}=d^2w/dr^2$ and $\kappa_{\theta\theta}=(1/r)(dw/dr)$ are obtained from (26). Their sum $\kappa_\Sigma\equiv\kappa_{rr}+\kappa_{\theta\theta}$ is given in terms of the misfit strain by

$$\kappa_\Sigma=-6\frac{E_fh_f}{1-\nu_f}\frac{1-\nu_s^2}{E_sh_{s0}^2}\left\{\begin{aligned}&\left(3-2\frac{h_s}{h_{s0}}\right)\varepsilon_m+\left[4-3\frac{h_s}{h_{s0}}+\frac{3(1-\nu_s)}{2}\frac{h_s-h_s(0)}{h_{s0}}\right]\frac{1-\nu_s}{1+\nu_s}\overline{\varepsilon_m}\\&+\int_0^r\left[\frac{3(1-\nu_s)}{\eta^2}\int_0^\eta\rho\varepsilon_m(\rho)d\rho-\varepsilon_m(\eta)\right]\frac{h_s'(\eta)}{h_{s0}}d\eta\end{aligned}\right\} \tag{29}$$

where

$$\overline{\varepsilon_m}=\frac{2}{\pi R^2}\int_0^R\eta\varepsilon_m d\eta$$

is the average misfit strain in the thin film. The difference of system curvatures $\kappa_\Delta\equiv\kappa_{rr}-\kappa_{\theta\theta}$ is given by

$$\kappa_\Delta=-6\frac{E_fh_f}{1-\nu_f}\frac{1-\nu_s^2}{E_sh_{s0}^2}\left\{\begin{aligned}&\left(4-3\frac{h_s}{h_{s0}}\right)\left(\varepsilon_m-\frac{2}{r^2}\int_0^r\eta\varepsilon_m(\eta)d\eta\right)\\&+\left(\frac{h_s}{h_{s0}}-1\right)\varepsilon_m-\frac{2}{r^2}\int_0^r\eta\left[\frac{h_s(\eta)}{h_{s0}}-1\right]\varepsilon_m(\eta)d\eta\\&-\frac{1}{r^2}\int_0^r\eta^2\left[\varepsilon_m(\eta)+\frac{3(1+\nu_s)}{\eta^2}\int_0^\eta\rho\varepsilon_m(\rho)d\rho+\frac{3(1-\nu_s)}{2}\overline{\varepsilon_m}\right]\frac{h_s'(\eta)}{h_{s0}}d\eta\end{aligned}\right\} \tag{30}$$

The thin film stresses are obtained from the constitutive relations

$$\sigma_{rr}^{(f)}=\frac{E_f}{1-\nu_f^2}\left[u_f'+\nu_f\frac{u_f}{r}-(1+\nu_f)\varepsilon_m\right]$$

and

$$\sigma_{\theta\theta}^{(f)}=\frac{E_f}{1-\nu_f^2}\left[\nu_fu_f'+\frac{u_f}{r}-(1+\nu_f)\varepsilon_m\right]$$

where u_f is given in (10). The sum of thin film stresses, up to the $O(\beta^2)$ accuracy (as compared to unity), is related to the misfit strain by

$$\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}=\frac{E_f}{1-\nu_f}(-2\varepsilon_m) \tag{31}$$

The difference of thin film stresses $\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}$ is on the order of $O((E_f^2/E_s)\varepsilon_m(h_f/h_{s0}))$, which is very small as compared to $\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}$. Therefore only its leading term is presented

$$\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}=4E_f\frac{E_fh_f}{1-\nu_f^2}\frac{1-\nu_s^2}{E_sh_{s0}}\left[\varepsilon_m-\frac{2}{r^2}\int_0^r\eta\varepsilon_m(\eta)d\eta\right] \tag{32}$$

4.1 Special Case: Uniform Misfit Strain. For uniform misfit strain distribution ε_m=constant (and nonuniform substrate thickness), the interface shear stress in (15) vanishes. The thin film stresses become constant and equibiaxial, and are given by

$$\sigma_{rr}^{(f)} = \sigma_{\theta\theta}^{(f)} = \frac{E_f}{1-\nu_f}(-\varepsilon_m) \tag{33}$$

The curvatures in (29) and (30) become

$$\kappa_\Sigma = -12\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s}{E_s h_{s0}^2}\left\{1 - \frac{5-\nu_s}{2}\left(\frac{h_s}{h_{s0}} - 1\right) + (1-2\nu_s)\frac{h_s - h_s(0)}{h_{s0}}\right\}\varepsilon_m$$

$$\kappa_\Delta = 18\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s^2}{E_s h_{s0}^2}\left\{\frac{h_s}{h_{s0}} - \frac{2}{r^2}\int_0^r \eta\frac{h_s(\eta)}{h_{s0}}d\eta\right\}\varepsilon_m \tag{34}$$

which are neither constant nor equibiaxial for varying substrate thickness.

Figure 2 shows a substrate with a step change in thickness; a uniform thickness h in the outer region ($r > R_{\text{in}}$) and a slightly different value $h - \Delta h$ in the inner region ($r < R_{\text{in}}$), where $|\Delta h| \ll h$. The average thickness becomes $h_{s0} = h - \Delta h\left(R_{\text{in}}^2/R^2\right)$. The curvature in the circumferential direction is

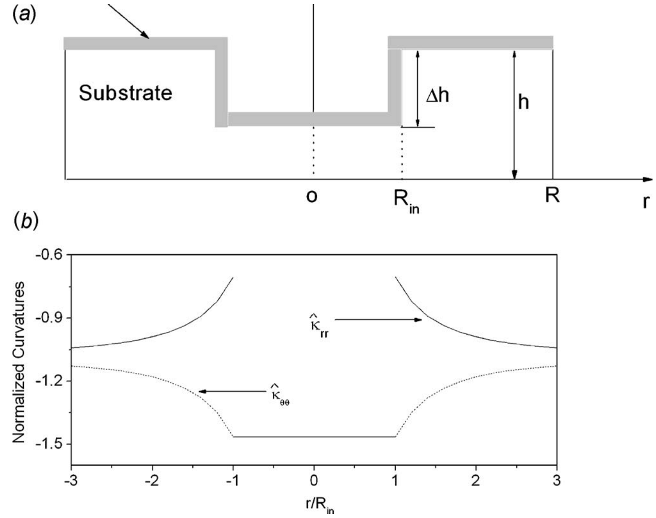


Fig. 2 (***a***) **A schematic diagram of a thin film/substrate system with a step change in substrate thickness. (*b*) The normalized system curvatures $\hat{\kappa}_{rr} = \kappa_{rr}/\kappa_0$ and $\hat{\kappa}_{\theta\theta} = \kappa_{\theta\theta}/\kappa_0$, where $\kappa_0 = 6(E_f h_f/1-\nu_f)/(1-\nu_s/E_s h^2)\varepsilon_m$, $\Delta h/2h = 0.1$, $\nu_s = 0.27$, and $R_{\text{in}} = R/3$.**

$$\kappa_{\theta\theta} = -6\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s}{E_s h^2}\varepsilon_m \begin{cases} 1 + \dfrac{\Delta h}{2h}\left[5 - \nu_s - (1-\nu_s)\dfrac{R_{\text{in}}^2}{R^2}\right] & \text{for } r < R_{\text{in}} \\ 1 + \dfrac{\Delta h}{2h}\left[5 - \nu_s - (1-\nu_s)\dfrac{R_{\text{in}}^2}{R^2} - 3(1+\nu_s)\left(1 - \dfrac{R_{\text{in}}^2}{r^2}\right)\right] & \text{for } r > R_{\text{in}} \end{cases} \tag{35}$$

which is a constant in the inner region, and is continuous across $r = R_{\text{in}}$. The curvature in the radial direction κ_{rr} is the same constant as $\kappa_{\theta\theta}$ in the inner region; however, it is discontinuous across $r = R_{\text{in}}$, and is given by

$$\kappa_{rr} = -6\frac{E_f h_f}{1-\nu_f}\frac{1-\nu_s}{E_s h^2}\varepsilon_m \begin{cases} 1 + \dfrac{\Delta h}{2h}\left[5 - \nu_s - (1-\nu_s)\dfrac{R_{\text{in}}^2}{R^2}\right] & \text{for } r < R_{\text{in}} \\ 1 + \dfrac{\Delta h}{2h}\left[5 - \nu_s - (1-\nu_s t)\dfrac{R_{\text{in}}^2}{R^2} - 3(1+\nu_s)\left(1 + \dfrac{R_{\text{in}}^2}{r^2}\right)\right] & \text{for } r > R_{\text{in}} \end{cases} \tag{36}$$

The continuous $\kappa_{\theta\theta}$ and discontinuous κ_{rr} are illustrated in Fig. 2. Similar discontinuity in κ_{rr} has been observed for varying thin film thickness [17,18].

It should be pointed out that the results in this section hold for discontinuous substrate thickness. This is because the film stresses in (31) and (32) depend only on the misfit strain and are independent of substrate thickness. The system curvatures in (29) and (30) involve the derivative of substrate thickness h_s', which is not well defined for a discontinuous h_s. However, it appears only in the integration such that (29) and (30) still hold.

In the following, we extend the Stoney formula for arbitrary nonuniform misfit strain distribution and nonuniform substrate thickness.

5 Extension of Stoney Formula for Nonuniform Misfit Strain Distribution and Nonuniform Substrate Thickness

In this section we extend the Stoney formula for arbitrary nonuniform misfit strain distribution and nonuniform substrate thickness by establishing the direct relation between the thin-film stresses and substrate curvatures. We invert the misfit strain from (29) as

$$\varepsilon_m = -\frac{1-\nu_f}{6E_f h_f}\frac{E_s}{1-\nu_s^2}\left\{\begin{aligned} &h_s^2\kappa_\Sigma - \frac{1-\nu_s}{2}\overline{h_s^2\kappa_\Sigma} \\ &+ \frac{1}{2}\int_r^R \left[(1-3\nu_s)\kappa_\Sigma(\eta) - 3(1-\nu_s)\kappa_\Delta(\eta)\right]h_s^2(\eta)\frac{h_s'(\eta)}{h_{s0}}d\eta \\ &- \frac{1-\nu_s}{R^2}\int_0^R \eta^2\left[\kappa_\Sigma(\eta) - \kappa_\Delta(\eta)\right]h_s^2(\eta)\frac{h_s'(\eta)}{h_{s0}}d\eta \end{aligned}\right\} \tag{37}$$

where

$$\overline{h_s^2\kappa_\Sigma} = \frac{2}{R^2}\int_0^R \eta h_s^2\kappa_\Sigma d\eta$$

is the average of $h_s^2\kappa_\Sigma$, and we have used (30) in establishing (37).

The thin film stresses are obtained by substituting (37) into (31) and (32) as

$$\sigma_{rr}^{(f)}+\sigma_{\theta\theta}^{(f)}=\frac{E_s}{3(1-\nu_s^2)h_f}\left\{\begin{array}{l} h_s^2\kappa_\Sigma-\frac{1-\nu_s}{2}\overline{h_s^2\kappa_\Sigma} \\ +\frac{1}{2}\int_r^R[(1-3\nu_s)\kappa_\Sigma(\eta)-3(1-\nu_s)\kappa_\Delta(\eta)]h_s^2(\eta)\frac{h_s'(\eta)}{h_{s0}}d\eta \\ -\frac{1-\nu_s}{R^2}\int_0^R\eta^2[\kappa_\Sigma(\eta)-\kappa_\Delta(\eta)]h_s^2(\eta)\frac{h_s'(\eta)}{h_{s0}}d\eta \end{array}\right\} \tag{38}$$

$$\sigma_{rr}^{(f)}-\sigma_{\theta\theta}^{(f)}=-\frac{2E_fh_{s0}}{3(1+\nu_f)}\kappa_\Delta \tag{39}$$

Equations (38) and (39) provide direct relations between film stresses and system curvatures. The system curvatures in (38) always appear together with the square of substrate thickness, i.e., $h_s^2\kappa_\Sigma$ and $h_s^2\kappa_\Delta$. It is important to note that stresses at a point in the thin film depend not only on curvatures at the same point (local dependence), but also on curvatures in the entire substrate (nonlocal dependence) via the term $\overline{h_s^2\kappa_\Sigma}$ and the integrals in (38). For uniform substrate thickness, (38) and (39) degenerate to Huang et al. [13]

The interface shear stress τ can also be directly related to system curvatures via (15) and (37)

$$\tau=\frac{E_s}{6(1-\nu_s^2)}\left\{\frac{d}{dr}(h_s^2\kappa_\Sigma)-\frac{1}{2}[(1-3\nu_s)h_s^2\kappa_\Sigma-3(1-\nu_s)h_s^2\kappa_\Delta]\frac{h_s'}{h_{s0}}\right\} \tag{40}$$

Equation (40) provides a way to determine the interface shear stresses from the gradients of system curvatures once the full-field curvature information is available. Since the interfacial shear stress is responsible for promoting system failures through delamination of the thin film from the substrate, Eq. (40) has a particular significance. It shows that such stress is related to the gradient of $\kappa_{rr}+\kappa_{\theta\theta}$, as well as to the magnitude of $\kappa_{rr}+\kappa_{\theta\theta}$ and $\kappa_{rr}-\kappa_{\theta\theta}$ for nonuniform substrate thickness.

In summary, (38)–(40) provide a simple way to determine the thin film stresses and interface shear stress from the nonuniform misfit strain in the thin film and nonuniform substrate thickness.

References

[1] Stoney, G. G., 1909, "The Tension of Metallic Films Deposited by Electrolysis," Proc. R. Soc. London, Ser. A, **82**, pp. 172–175.

[2] Freund, L. B., and Suresh, S., 2004, *Thin Film Materials; Stress, Defect Formation and Surface Evolution*, Cambridge University Press, Cambridge, U.K.

[3] Wikstrom, A., Gudmundson, P., and Suresh, S., 1999, "Thermoelastic Analysis of Periodic Thin Lines Deposited on a Substrate," J. Mech. Phys. Solids, **47**, pp. 1113–1130.

[4] Wikstrom, A., Gudmundson, P., and Suresh, S., 1999, "Analysis of Average Thermal Stresses in Passivated Metal Interconnects," J. Appl. Phys., **86**, pp. 6088–6095.

[5] Shen, Y. L., Suresh, S., and Blech, I. A., 1996, "Stresses, Curvatures, and Shape Changes Arising From Patterned Lines on Silicon Wafers," J. Appl. Phys., **80**, pp. 1388–1398.

[6] Park, T. S., and Suresh, S., 2000, "Effects of Line and Passivation Geometry on Curvature Evolution During Processing and Thermal Cycling in Copper Interconnect Lines," Acta Mater., **48**, pp. 3169–3175.

[7] Masters, C. B., and Salamon, N. J., 1993, "Geometrically Nonlinear Stress–Deflection Relations for Thin Film/Substrate Systems," Int. J. Eng. Sci., **31**, pp. 915–925.

[8] Salamon, N. J., and Masters, C. B., 1995, "Bifurcation in Isotropic Thin Film/Substrate Plates," Int. J. Solids Struct., **32**, pp. 473–481.

[9] Finot, M., Blech, I. A., Suresh, S., and Fijimoto, H., 1997, "Large Deformation and Geometric Instability of Substrates With Thin-Film Deposits," J. Appl. Phys., **81**, pp. 3457–3464.

[10] Freund, L. B., 2000, "Substrate Curvature Due to Thin Film Mismatch Strain in the Nonlinear Deformation Range," J. Mech. Phys. Solids, **48**, p. 1159.

[11] Lee, H., Rosakis, A. J., and Freund, L. B., 2001, "Full Field Optical Measurement of Curvatures in Ultra-Thin Film/Substrate Systems in the Range of Geometrically Nonlinear Deformations," J. Appl. Phys., **89**, pp. 6116–6129.

[12] Park, T. S., and Suresh, S., 2000, "Effects of Line and Passivation Geometry on Curvature Evolution During Processing and Thermal Cycling in Copper Interconnect Lines," Acta Mater., **48**, pp. 3169–3175.

[13] Huang, Y., Ngo, D., and Rosakis, A. J., 2005, "Non-Uniform, Axisymmetric Misfit Strain: In Thin Films Bonded on Plate Substrates/Substrate Systems: The Relation Between Non-Uniform Film Stresses and System Curvatures," Acta Mech. Sin., **21**, pp. 362–370.

[14] Huang, Y., and Rosakis, A. J., 2005, "Extension of Stoney's Formula to Non-Uniform Temperature Distributions in Thin Film/Substrate Systems. The Case of Radial Symmetry," J. Mech. Phys. Solids, **53**, pp. 2483–2500.

[15] Ngo, D., Huang, Y., Rosakis, A. J., and Feng, X., 2006, "Spatially Non-Uniform, Isotropic Misfit Strain in Thin Films Bonded on Plate Substrates: The Relation Between Non-Uniform Film Stresses and System Curvatures," Thin Solid Films, **515**, pp. 2220–2229.

[16] Feng, X., Huang, Y., Jiang, H., Ngo, D., and Rosakis, A. J., 2006, "The Effect of Thin Film/Substrate Radii on the Stoney Formula for Thin Film/Substrate Subjected to Non-Uniform Axisymmetric Misfit Strain and Temperature," J. Mech. Mater. Struct., in press.

[17] Brown, M., Rosakis, A. J., Feng, X., Feng, X., Huang, Y., and Üstündag, E., 2006, "Thin Film/Substrate Systems Featuring Arbitrary Film Thickness and Misfit Strain Distributions. Part II: Experimental Validation of the Non-Local Stress/Curvature Relations," Int. J. Solids Struct., **44**, pp. 1755–1767.

[18] Brown, M. A., Park, T. S., Rosakis, A. J., Ustundag, E., Huang, Y., Tamura, N., and Valek, B., 2006, "A Comparison of X-Ray Microdiffraction and Coherent Gradient Sensing in Measuring Discontinuous Curvatures in Thin Film: Substrate Systems," ASME J. Appl. Mech., **73**, pp. 723–729.

Journal of Applied Mechanics

Technical Briefs

On Some Benchmark Results for the Interaction of a Crack With a Circular Inclusion

A.-V. Phan
Department of Mechanical Engineering,
University of South Alabama,
Mobile, AL 36688

L. J. Gray

T. Kaplan

Computer Science and Mathematics Division,
Oak Ridge National Laboratory,
Oak Ridge, TN 37831

*Stress intensity factor calculations for crack-inclusion interaction problems are presented. The problems considered include the benchmark problems first discussed by Helsing and Jonsson (2002, ASME J. Appl. Mech, **69**, pp. 88–90), and subsequently by Wang, Mogilevskaya, and Crouch (2003, ASME J. Appl. Mech., **70**, pp. 619–621). The numerical results are obtained using the symmetric-Galerkin boundary element method in conjunction with an improved quarter-point element for evaluating the stress intensity factors by means of the displacement correlation technique. The converged results confirm the accuracy of the previous simulations and demonstrate that accurate solutions for these interaction problems can be obtained with numerical methods that are applicable in three dimensions.* [DOI: 10.1115/1.2722773]

1 Introduction

The interaction between an arbitrary crack and a circular inclusion is a subject of important interest, and thus has attracted a great deal of contributions from various research groups. However, some numerical benchmark results on the subject have recently been questioned by Helsing and Jonsson (HJ); they showed in Ref. [1] that their converged results for the stress intensity factors (SIFs) K_{I} and K_{II} differ from values published in the literature. Following their challenge to the computational mechanics community to confirm or disprove their findings, Wang, Mogilevskaya and Crouch (WMC) [2] have produced results obtained using a Galerkin boundary integral (GBI) method and a complex variables boundary element method (CVBEM). The WMC solutions agree with the HJ results, even though in one case it is not clear precisely how well: for the problem of a circular arc crack interacting with a circular inclusion (Fig. 1(*b*)), the HJ results were only shown graphically.

Although the WMC work has substantially confirmed the HJ results, we believe some further discussion is warranted. In our opinion, it is important to verify, for eventual applications, that these problems can also be accurately solved using "standard" numerical techniques, i.e., techiques that are directly applicable in three dimensions. Note that the integral equation elasticity formulations in HJ and WMC rely on complex variable methods and are therefore decidedly two-dimensional algorithms. Moreover, some of the numerical approximations employed in their analyses, e.g., global function approximations rather than local element interpolations, are not routinely used in three dimensions. This is *not* to say that these are not good methods—quite the contrary, we think these approaches are remarkably accurate—just not directly extendable to three dimensions. Thus, in addition to confirming the HJ and WMC solutions, the purpose of this paper is to establish that a general algorithm can successfully solve these types of problems.

For the problem of a straight crack interacting with a circular inclusion (Fig. 1(*a*)), the WMC calculations did not agree with (a likely misprint in) the HJ result, and in this note we confirm the WMC answer. In addition, by tabulating the results for the circular crack problem shown in Fig. 1(*b*), we can confirm the WMC calculations.

2 Problem Descriptions and Method of Solution

We consider the two problems discussed in Ref. [1], shown in Figs. 1(*a*) and 1(*b*). The straight crack problem (Fig. 1(*a*)) was initially studied by Erdogan, Gupta and Ratwani (EGR) [3], while problem (Fig. 1(*b*)) involves a circular arc crack and was studied by Cheeseman and Santare [4]. Plane strain is assumed. For both problems, the shear moduli of the matrix and the inclusion are respectively $G_1=1$ and $G_2=23$, and the corresponding Poisson's ratios are $\nu_1=0.35$ and $\nu_2=0.3$.

The calculations presented herein utilize a (more or less standard) quadratic element symmetric-Galerkin boundary integral analysis [5]. However, the stress intensity factor computation is based upon the modified quarter point (MQP) element [6]. The basic idea of the MQP is to modify the quadratic shape functions at the crack tip (adding an appropriate cubic term) in order that the crack opening displacement satisfy a known constraint: the term that is linear in distance to the tip must vanish [7]. This element has been shown to yield highly accurate SIF values (even by means of the simple displacement correlation technique) for standard crack problems, and this work establishes that this carries over to the crack/inclusion problems considered herein. The MQP approach of modifying the crack-tip shape functions extends directly to three dimensions; alternatively, the constraint on the tip displacement can be incorporated directly by employing an appro-

Contributed by the Applied Mechanics Division of ASME for publication in the Journal of Applied Mechanics. Manuscript received October 4, 2006; final manuscript received January 2, 2007. Review conducted by Subrata Mukherjee.

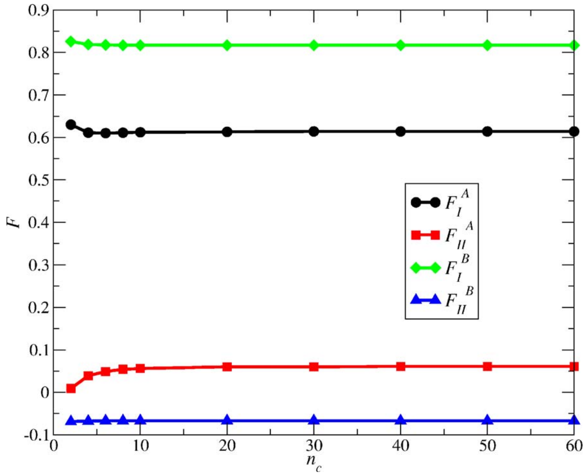


Fig. 1 Crack-circular inclusion interaction under remote stress: (*a*) Straight crack and (*b*) circular arc crack

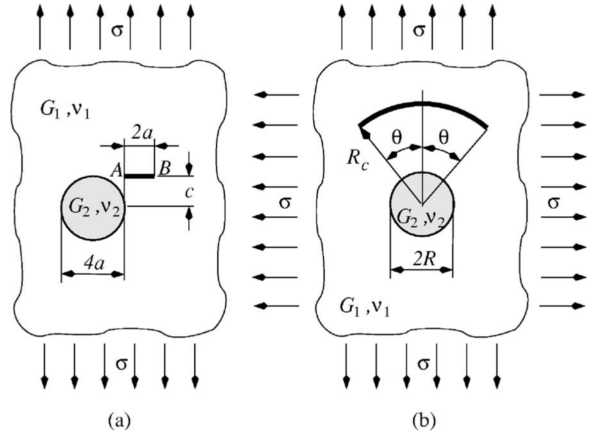


Fig. 2 Convergence of *F* for *c*/*a*=1.0

priate expansion, as in [8]. The highly accurate results obtained by these authors would indicate that the MQP in 3D will also be very successful.

3 Results and Discussions

For both problems, the straight/arc crack and the matrix-inclusion interface are discretized into uniform elements. All the results reported in this section are convergent with respect to mesh refinement. For example, Fig. 2 shows the convergence of the normalized SIFs F_{I} and F_{II}, defined as $F_{\mathrm{I}}=K_{\mathrm{I}}/(\sigma\sqrt{\pi a})$ and $F_{\mathrm{II}}=K_{\mathrm{II}}/(\sigma\sqrt{\pi a})$, for the straight crack problem (Fig. 1(*a*)) with $c/a=1.0$. Here, the matrix-inclusion interface is meshed using $n_i=68$ elements. It can be seen that the solution for the SIFs at both crack tips A and B converges quickly as the number of crack elements n_c approaches 10.

3.1 Straight Crack Interacting With Circular Inclusion. Our symmetric-Galerkin boundary integral results using standard (SQP) and modified quarter-point elements for the normalized SIFs at crack tips A and B are presented in Tables 1 and 2, respectively, together with the WMC, HJ, and EGR results. Here, it is important to observe that the additional accuracy provided by the MQP is essential in matching the results in HJ and WMC. Finally, the significant differences in the numerical methods provide additional confirmation of the correctness of the SIF results.

As expected, only coarse meshes are needed to obtain the converged solution when a crack tip is not very close to the inclusion boundary. For example, in case $c/a=8$, $n_c=8$ and $n_i=28$ are required to reach the convergence. Overall, our MQP solutions agree very well with HJ and WMC, and confirm WMC's belief that there is a misprint in the F_{II} HJ result for the case $c/a=3$. Our result (−0.003) is much closer to the WMC value (−0.004) than the HJ (−0.035).

To further demonstrate the accuracy and robustness of the nu-

Table 1 Normalized stress intensity factors at crack tip *A* ($F=K/(\sigma\sqrt{\pi a})$)

	F_I^A					F_{II}^A				
c/a	SQP	MQP	WMC	HJ	EGR	SQP	MQP	WMC	HJ	EGR
0.3	0.202	0.234	0.236	0.235	0.225	0.089	0.073	0.074	0.073	0.072
0.5	0.338	0.348	0.347	0.347	0.341	0.113	0.102	0.102	0.102	0.101
1.0	0.616	0.614	0.613	0.613	0.613	0.068	0.061	0.061	0.061	0.057
1.5	0.761	0.756	0.755	0.755	0.763	0.015	0.012	0.012	0.012	−0.007
2.0	0.836	0.830	0.830	0.830	0.845	0.020	0.018	0.018	0.018	−0.021
3.0	0.944	0.937	0.936	0.936	0.953	0.068	0.067	0.067	0.067	−0.001
4.0	1.010	1.003	1.003	1.003	1.014	0.080	0.079	0.079	0.079	0.002
8.0	1.049	1.043	1.043	1.043	1.043	0.032	0.032	0.032	0.032	−0.026

Table 2 Normalized stress intensity factors at crack tip *B* ($F=K/(\sigma\sqrt{\pi a})$)

	F_I^B					F_{II}^B				
c/a	SQP	MQP	WMC	HJ	EGR	SQP	MQP	WMC	HJ	EGR
0.3	0.803	0.790	0.790	0.790	0.784	−0.022	−0.022	−0.023	−0.023	−0.004
0.5	0.803	0.797	0.797	0.797	0.792	−0.037	−0.037	−0.037	−0.037	−0.006
1.0	0.823	0.817	0.817	0.817	0.817	−0.067	−0.067	−0.067	−0.067	−0.005
1.5	0.839	0.833	0.833	0.833	0.839	−0.074	−0.074	−0.074	−0.074	0.008
2.0	0.855	0.850	0.850	0.850	0.860	−0.058	−0.057	−0.058	−0.058	0.034
3.0	0.903	0.898	0.897	0.897	0.905	−0.004	−0.003	−0.004	−0.035	0.089
4.0	0.953	0.948	0.947	0.947	0.951	0.032	0.032	0.032	0.032	0.117
8.0	1.028	1.022	1.022	1.022	1.020	0.032	0.032	0.032	0.032	0.088

Table 3 Normalized stress intensity factors for the case where the inclusion is of the same material as the matrix (analytical solution: $F_I = K_I/(\sigma\sqrt{\pi a}) = 1$ and $F_{II} = K_{II}/(\sigma\sqrt{\pi a}) = 0$)

c/a	0.3	0.5	1.0	1.5	2.0	3.0	4.0	8.0
F_I^A	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
F_I^B	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
F_{II}^A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F_{II}^B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4 Normalized stress intensity factors at the crack tips ($F = K/K_o$)

	θ=30 deg				θ=45 deg				θ=75 deg			
	F_I		F_{II}		F_I		F_{II}		F_I		F_{II}	
$\frac{R_c}{R}$	MQP	WMC	MQP	WMC	MQP	WMC	MQP	WMC	MQP	WMC	MQP	WMC
1.1	0.919	0.919	1.494	1.494	0.928	0.928	1.393	1.393	0.955	0.955	1.363	1.363
1.2	0.944	0.944	1.353	1.353	0.962	0.962	1.284	1.285	1.009	1.009	1.281	1.281
1.5	0.961	0.961	1.202	1.202	0.990	0.990	1.159	1.159	1.056	1.056	1.160	1.160
2.0	0.972	0.972	1.103	1.104	0.992	0.992	1.092	1.092	1.059	1.059	1.083	1.084
3.0	0.986	0.986	1.039	1.040	0.992	0.992	1.042	1.043	1.039	1.039	1.036	1.037
4.0	0.992	0.992	1.120	1.021	0.995	0.994	1.024	1.024	1.025	1.025	1.020	1.021
5.0	0.995	0.995	1.012	1.013	0.996	0.996	1.015	1.016	1.017	1.017	1.013	1.014
6.0	0.996	0.997	1.008	1.009	0.997	0.997	1.010	1.011	1.012	1.012	1.009	1.010

merical techniques, the above calculation is modified so that the inclusion and matrix are of the same material. In this case, there is a known analytical result [9]: $K_I^A = K_I^B = \sigma\sqrt{\pi a}$ and $K_{II}^A = K_{II}^B = 0$, and the converged MQP numbers in Table 3 are virtually identical to this solution. A minor numerical error for F_I^A in case $c/a = 0.3$ is expected as this crack tip is very close to the inclusion boundary.

3.2 Circular Arc Crack Interacting With Circular Inclusion. For this problem, the SIFs are respectively normalized by the SIFs K_{Io} and K_{IIo} in the absence of the inclusion which are given by [9]

$$\begin{Bmatrix} K_{Io} \\ K_{IIo} \end{Bmatrix} = \frac{\sigma\sqrt{\pi R_c \sin\theta}}{1 + \left(\sin\frac{\theta}{2}\right)^2} \begin{Bmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} \end{Bmatrix} \quad (1)$$

Our converged results using the MQP element are listed along with the WMC's solution in Table 4 where a very good agreement can be seen. As mentioned earlier, it is important to numerically confirm the WMC results for this problem as the corresponding HJ results are only available graphically.

4 Conclusions

It has been established that a standard Galerkin boundary integral algorithm, together with a modified quarter-point crack-tip element, is capable of accurately solving the benchmark problems discussed by Helsing and Jonsson [1], and also by Wang et al. [2]. The accuracy is confirmed not only through agreement with these previous results but also by observed convergence with mesh refinement. This algorithm should therefore be equally successful for three-dimensional applications.

Acknowledgment

This research was supported in part by the U.S. Army Research Office Grant No. W911 NF-04-1-0257, and by the Applied Mathematical Sciences Research Program of the Office of Mathematical, Information, and Computational Sciences, the U.S. Department of Energy under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC.

References

[1] Helsing, J., and Jonsson, A., 2002, "On the Accuracy of Benchmark Tables and Graphical Results in the Applied Mechanics Literature," ASME J. Appl. Mech., **69**, pp. 88–90.
[2] Wang, J., Mogilevskaya, S. G., and Crouch, S. L., 2003, "Benchmark Results for the Problem of Interaction Between a Crack and a Circular Inclusion," ASME J. Appl. Mech., **70**, pp. 619–621.
[3] Erdogan, F., Gupta, G. D., and Ratwani, M., 1974, "Interaction Between a Circular Inclusion and an Arbitrarily Oriented Crack," ASME J. Appl. Mech., **41**, pp. 1007–1013.
[4] Cheeseman, B. A., and Santare, M. H., 2000, "The Interaction of a Curved Crack With a Circular Inclusion," Int. J. Fract., **103**, pp. 259–277.
[5] Bonnet, M., 1995, *Boundary Integral Equation Methods for Solids and Fluids*, Wiley, New York.
[6] Gray, L. J., Phan, A.-V., Paulino, G. H., and Kaplan, T., 2003, "Improved Quarter-Point Crack Tip Element," Eng. Fract. Mech., **70**, pp. 269–283.
[7] Gray, L. J., and Paulino, G. H., 1998, "Crack Tip Interpolation, Revisited," SIAM J. Appl. Math., **58**, pp. 428–455.
[8] Li, S., Mear, M. E., and Xiao, L., 1998, "Symmetric Weak Form Integral Equation Method for Three-Dimensional Fracture Analysis," Comput. Methods Appl. Mech. Eng., **151**, pp. 435–459.
[9] Tada, H., Paris, P., and Irwin, G., 1973, *The Stress Analysis of Cracks Handbook*, Del Research Corporation, Hellertown, PA.

Thermoelastic Interaction Between Singularities and Interfaces in an Anisotropic Trimaterial

Seung Tae Choi
Mem. ASME
Micro Device and Systems Lab,
Samsung Advanced Institute of Technology,
P.O. Box 111,
Suwon 440-600, Republic of Korea
e-mail: stchoi@kaist.ac.kr

The method of analytic continuation and Schwarz-Neumann's alternating technique were applied to the thermoelastic interaction problems of singularities and interfaces in an anisotropic "trimaterial," which denotes an infinite body composed of three dissimilar materials bonded along two parallel interfaces. It was assumed that the linear thermoelastic materials are under general plane deformations in which the plane of deformation is perpendicular to the planes of the two parallel interfaces. The author then showed that by alternately applying the method of analytic continuation across two parallel interfaces the solution for the thermoelastic singularities in an anisotropic trimaterial can be obtained in a series form from a solution for the same singularities in a homogeneous anisotropic medium.
[DOI: 10.1115/1.2722779]

1 Introduction

Thin-film and layered structures are technologically important in electronics and optoelectronics. In these structures, defects (such as dislocations) are inevitable and affect the performance of systems [1]. From a mechanical point of view, dislocations are treated as singularities, and analysis of an elastic field near a singularity is important for understanding the behavior of structures. In addition to the intrinsic physical significance, the elastic field of a dislocation can serve as a kernel function in singular integral equations to simulate cracks. However, the elastic field near singularities is not easy to obtain because of the difficulty of satisfying the boundary conditions at free surfaces or interfaces or both. Recently, Choi and Earmme [2] obtained a solution of an elastic singularity in an anisotropic trimaterial by using the method of analytic continuation [3] and Schwarz-Neumann's alternating technique [4]. In this study, by alternately applying the method of analytic continuation across two parallel interfaces, the author shows that the solution for thermoelastic singularities in an anisotropic trimaterial can also be obtained in a series form from a solution for the same singularities in a homogeneous anisotropic medium. For conciseness, the notations in [2] are employed here and the reader is referred to [2] for a more detailed explanation or explicit forms.

2 Anisotropic Thermoelasticity

To briefly review anisotropic thermoelasticity, let us consider a generalized two-dimensional heat conduction and deformation in which temperature T and the displacement u_j depend only on x_1 and x_2. The constitutive equation for a linear thermoelastic material is

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received July 11, 2006; final manuscript received January 29, 2007. Review conducted by Zhigang Suo.

$$h_i = -\kappa_{ij}\frac{\partial T}{\partial x_j}, \quad \sigma_{ij} = C_{ijkm}\frac{\partial u_k}{\partial x_m} - \beta_{ij}T, \quad (i,j=1,2,3) \tag{1}$$

where h_i is the heat flux, κ_{ij} is the coefficient of heat conduction, and β_{ij} is the stress-temperature coefficient. The conservation of energy and equation of equilibrium are expressed, respectively, as follows:

$$\kappa_{ij}\frac{\partial^2 T}{\partial x_i \partial x_j} = 0, \quad C_{ijkm}\frac{\partial^2 u_k}{\partial x_j \partial x_m} - \beta_{ij}\frac{\partial T}{\partial x_j} = 0 \tag{2}$$

A general solution for the temperature, corresponding heat flux, displacement, and corresponding stress, which satisfy Eq. (2), may be written as follows [5–7]:

$$T = 2\,\mathrm{Re}[\chi'(z_\tau)] \tag{3}$$

$$h_i = -2\,\mathrm{Re}[(\kappa_{i1} + \tau\kappa_{i2})\chi''(z_\tau)] \tag{4}$$

$$u_i = 2\,\mathrm{Re}[A_{ij}f_j(z_{\underline{j}}) + c_i\chi(z_\tau)] \tag{5}$$

$$\sigma_{1i} = -2\,\mathrm{Re}[L_{ij}\mu_{\underline{j}}f'_j(z_{\underline{j}}) + d_i\tau\chi'(z_\tau)] \tag{6}$$

$$\sigma_{2i} = 2\,\mathrm{Re}[L_{ij}f'_j(z_{\underline{j}}) + d_i\chi'(z_\tau)] \tag{7}$$

In these equations, the function $\chi(z_\tau)$ is an analytic function of the complex variable $z_\tau = x_1 + \tau x_2$, and τ is the eigenvalue with a positive imaginary part of the equation $\kappa_{22}\tau^2 + 2\kappa_{12}\tau + \kappa_{11} = 0$. Column vector $\mathbf{c}$ is the eigenvector of the following sextic equation:

$$[C_{i1k1} + \tau(C_{i1k2} + C_{i2k1}) + \tau^2 C_{i2k2}]\mathbf{c}_k = \beta_{i1} + \tau\beta_{i2} \tag{8}$$

And column vector $\mathbf{d}$ is given by

$$d_i = (C_{i2k1} + \tau C_{i2k2})c_k - \beta_{2i} \tag{9}$$

The dimensionless bimaterial constant γ^{ab} is defined as follows:

$$\gamma^{ab} \equiv \frac{\kappa^b - \kappa^a}{\kappa^b + \kappa^a} \tag{10}$$

where $\kappa = \sqrt{\kappa_{11}\kappa_{22} - \kappa_{12}^2}$ and the indices a and b stand for materials a and b, respectively. For mathematical simplicity, the two bimaterial vectors are defined as follows:

$$\mathbf{e}^{ab} \equiv (\mathbf{L}^a)^{-1}(\mathbf{B}^a + \bar{\mathbf{B}}^b)^{-1}\{\bar{\mathbf{B}}^b[\mathbf{d}^b - \mathbf{d}^a + \gamma^{ab}(\bar{\mathbf{d}}^b - \mathbf{d}^a)] + i[\mathbf{c}^b - \mathbf{c}^a + \gamma^{ab}(\bar{\mathbf{c}}^b - \mathbf{c}^a)]\} \tag{11}$$

$$\mathbf{g}^{ab} \equiv (\bar{\mathbf{L}}^b)^{-1}(\mathbf{B}^a + \bar{\mathbf{B}}^b)^{-1}\{-\mathbf{B}^a[\mathbf{d}^b - \mathbf{d}^a + \gamma^{ab}(\bar{\mathbf{d}}^b - \mathbf{d}^a)] + i[\mathbf{c}^b - \mathbf{c}^a + \gamma^{ab}(\bar{\mathbf{c}}^b - \mathbf{c}^a)]\} \tag{12}$$

Of particular importance is the fact that the following factors (namely, the derivative of temperature, the heat flux, the derivative of displacement, and the traction) must be continuous across a perfectly bonded interface where $x_2 = 0$,

$$\frac{\partial T}{\partial x_1}(x_1) = \chi''(x_1) + \bar{\chi}''(x_1) \tag{13}$$

$$h_2 = -i\kappa\chi''(x_1) + i\kappa\bar{\chi}''(x_1) \tag{14}$$

$$\frac{\partial u_i}{\partial x_1}(x_1) = A_{ij}f'_j(x_1) + \bar{A}_{ij}\bar{f}'_j(x_1) + c_i\chi'(x_1) + \bar{c}_i\bar{\chi}'(x_1) \tag{15}$$

$$\sigma_{2i}(x_1) = L_{ij}f'_j(x_1) + \bar{L}_{ij}\bar{f}'_j(x_1) + d_i\chi'(x_1) + \bar{d}_i\bar{\chi}'(x_1) \tag{16}$$

3 Thermoelastic Singularity in an Anisotropic Bimaterial

Section 2 explained that a general solution of a generalized two-dimensional heat conduction and deformation in anisotropic thermoelasticity can be expressed by the analytic functions $\chi(z)$

and $f_j(z)$. Note also that while the function $\chi(z)$, which is responsible for the thermal field, can be solved independently of the elastic field, the function $f_j(z)$, which is responsible for the elastic field, depends on the thermal field, namely, $\chi(z)$. The elastic field should therefore be calculated after the thermal field is known. Once the solution $\chi^0(z)$ and $f_j^0(z)$ of a singularity in a homogeneous medium has been examined, the solution can then be used as a building block for the same singularity in a bimaterial and a trimaterial. The solution $\chi^0(z)$ for the temperature dislocation or the heat source at (x_1^0, x_2^0) in an infinite homogeneous medium is given as follows:

$$\chi^{0'}(z_\tau) = \left(\frac{T_0}{4\pi i} - \frac{Q_0}{4\pi\kappa}\right)\ln(z_\tau - s_\tau) \tag{17}$$

where $s_\tau = x_1^0 + \tau x_2^0$, T_0 is the magnitude of the temperature dislocation, and Q_0 is the intensity of the heat source [7]. The solution $f_j^0(z)$ for the line force or the elastic dislocation at (x_1^0, x_2^0) in an infinite homogeneous medium is given in Eq. (14) of Choi and Earmme [2].

Figure 1 of Choi and Earmme [2] shows anisotropic bimaterial bonded along the x_1 axis. In consideration of this type of anisotropic bimaterial, the author used the method of analytic continuation to construct a bimaterial solution for thermoelastic singularities in terms of a homogeneous solution for the same singularities. First, a singularity located in the lower half-space is treated, in which the elastic constants of material b are implied in $\chi^0(z)$ and $f_j^0(z)$. When the continuity conditions of the quantities given in Eqs. (13)–(16) are applied and the analytic continuation arguments are used, the solutions are expressed in terms of $\chi^0(z)$ and $f_j^0(z)$ as follows:

$$\chi(z_\tau) = \begin{cases} (1+\gamma^{ab})\chi^0(z_\tau^a), & \text{in } S_a \\ \chi^0(z_\tau^b) + \gamma^{ab}\bar{\chi}^0(z_\tau^b), & \text{in } S_b \end{cases} \tag{18}$$

$$f_i(z_{\underline{i}}) = \begin{cases} U_{ij}^{ab} f_j^0(z_{\underline{i}}^a) + e_i^{ab}\chi^0(z_{\underline{i}}^a), & \text{in } S_a \\ f_i^0(z_{\underline{i}}^b) + \bar{V}_{ij}^{ab}\bar{f}_j^0(z_{\underline{i}}^b) + \bar{g}_i^{ab}\bar{\chi}^0(z_{\underline{i}}^b), & \text{in } S_b \end{cases} \tag{19}$$

When a similar procedure is used for a singularity located in the upper half-space, the solution is easily found to be

$$\chi(z_\tau) = \begin{cases} \chi^0(z_\tau^a) + \gamma^{ba}\bar{\chi}^0(z_\tau^a), & \text{in } S_a \\ (1+\gamma^{ba})\chi^0(z_\tau^b), & \text{in } S_b \end{cases} \tag{20}$$

$$f_i(z_{\underline{i}}) = \begin{cases} f_i^0(z_{\underline{i}}^a) + \bar{V}_{ij}^{ba}\bar{f}_j^0(z_{\underline{i}}^a) + \bar{g}_i^{ba}\bar{\chi}^0(z_{\underline{i}}^a), & \text{in } S_a \\ U_{ij}^{ba} f_j^0(z_{\underline{i}}^b) + e_i^{ba}\chi^0(z_{\underline{i}}^b), & \text{in } S_b \end{cases} \tag{21}$$

where the material constants involved in $\chi^0(z)$ and $f_j^0(z)$ are for material a.

4 Thermoelastic Singularity in an Anisotropic Trimaterial

The alternating technique together with the results of Sec. 3 can be used to analyze thermoelastic singularities in a trimaterial with two parallel interfaces shown in Fig. 2 of Choi and Earmme [2]. Because of the difficulty of simultaneously satisfying the continuity conditions along the two interfaces, the method of analytic continuation should be alternately applied to the two interfaces. Consider a trimaterial with a singularity as shown in Fig. 2 of Choi and Earmme [2], where materials a, b, and c occupy regions $S_a: x_2 \geqslant h$, $S_b: h \geqslant x_2 \geqslant 0$, and $S_c: x_2 \leqslant 0$, respectively, and the materials are perfectly bonded along the two parallel interfaces, $\Gamma: x_2 = 0$ and $\Gamma_*: x_2 = h$. By alternately applying the method of analytic continuation across the two parallel interfaces, the solution for thermoelastic singularities in an anisotropic trimaterial can also be obtained in a series form as

$$\chi(z_\tau) = \begin{cases} (1+\gamma^{ab})\sum_{n=1}^{\infty}\chi^n(z_\tau^a - \tau^a h + \tau^b h), & \text{in } S_a \\ \sum_{n=1}^{\infty}[\chi^n(z_\tau^b) + \gamma^{ab}\bar{\chi}^n(z_\tau^b - \tau^b h + \bar{\tau}^b h)], & \text{in } S_b \\ \chi^0(z_\tau^c) + \gamma^{bc}\bar{\chi}^0(z_\tau^c) + (1+\gamma^{cb})\gamma^{ab}\sum_{n=1}^{\infty}\bar{\chi}^n(z_\tau^c - \tau^b h + \bar{\tau}^b h), & \text{in } S_c \end{cases} \tag{22}$$

$$f_i(z_{\underline{i}}) = \begin{cases} \sum_{n=1}^{\infty}[U_{ij}^{ab} f_j^n(z_{\underline{i}}^a - \mu_{\underline{i}}^a h + \mu_{\underline{j}}^b h) + e_i^{ab}\chi^n(z_{\underline{i}}^a - \mu_{\underline{i}}^a h + \tau^b h)], & \text{in } S_a \\ \sum_{n=1}^{\infty}[f_i^n(z_{\underline{i}}^b) + \bar{V}_{ij}^{ab}\bar{f}_j^n(z_{\underline{i}}^b - \mu_{\underline{i}}^b h + \bar{\mu}_{\underline{j}}^b h) + \bar{g}_i^{ab}\bar{\chi}^n(z_{\underline{i}}^b - \mu_{\underline{i}}^b h + \bar{\tau}^b h)], & \text{in } S_b \\ f_i^0(z_{\underline{i}}^c) + \bar{V}_{ij}^{bc}\bar{f}_j^0(z_{\underline{i}}^c) + \bar{g}_i^{bc}\bar{\chi}^0(z_{\underline{i}}^c) + \sum_{n=1}^{\infty}\{U_{ij}^{cb}[\bar{V}_{jk}^{ab}\bar{f}_k^n(z_{\underline{i}}^c - \mu_{\underline{j}}^b h + \bar{\mu}_{\underline{k}}^b h) + \bar{g}_j^{ab}\bar{\chi}^n(z_{\underline{i}}^c - \mu_{\underline{j}}^b h + \bar{\tau}^b h)] + e_i^{cb}\gamma^{ab}\bar{\chi}^n(z_{\underline{i}}^c - \tau^b h + \bar{\tau}^b h)\}, & \text{in } S_c \end{cases} \tag{23}$$

where the recurrence formula for $\chi^n(z)$ and $f_i^n(z)$ are

$$\chi^{n+1}(z) = (\gamma^{ab}\gamma^{cb})^n(1+\gamma^{bc})\chi^0(z + \tau^b hn - \bar{\tau}^b hn), \quad n = 0, 1, 2, \ldots \tag{24}$$

$$f_i^{n+1}(z) = \begin{cases} U_{ij}^{bc} f_j^0(z) + e_i^{bc}\chi^0(z), & n = 0 \\ \bar{V}_{ij}^{cb}[V_{jk}^{ab} f_k^n(z - \bar{\mu}_{\underline{j}}^b h + \mu_{\underline{k}}^b h) + g_j^{ab}\chi^n(z - \bar{\mu}_{\underline{j}}^b h + \tau^b h)] + \bar{g}_i^{cb}\gamma^{ab}\chi^n(z - \bar{\tau}^b h + \tau^b h), & n = 1, 2, 3, \ldots \end{cases} \tag{25}$$

In Eqs. (24) and (25), the thermoelastic constants of material c are implied in $\chi^0(z)$ and $f_i^0(z)$. Equations (22)–(25) give the complete solution for the singularity located in region S_c.

It is worth noting that the purely thermal field expressed by $\chi(z)$ has analogy with anti-plane elastic deformation of anisotropic media. Thus, replacing U_{ij}^{ab}, V_{ij}^{ab}, $f_i(z_{\underline{i}})$, z_i, and μ_i in Eqs. 26 and 27 of Choi and Earmme [2] with $(1+\gamma^{ab})$, γ^{cb}, $\chi(z)$, z_τ, and τ, respectively, directly gives Eqs. (22) and (24) in this paper, which verifies the correctness of the thermal field presented in this paper. The elastic field given by Eq. (23) together with Eqs. (24) and (25) can be decomposed into two parts: the purely elastic field, the same as Eqs. (26) and (27) of Choi and Earmme [2], and the thermoelastic field, expressed in terms of $\chi^n(z)$ $(n=0,1,2,\ldots)$. By referring to Eqs. (19) and (21) in this paper and Eq. (24) of Choi and Earmme [2], one can easily check that Eqs. (23) and (25) satisfy the displacement gradient and traction continuities along two parallel interfaces, Γ and Γ_*, which guarantees the thermoelastic field is correct.

The other case, for a singularity in region S_b, has the following solution when the same procedure is used as in the case of a singularity in region S_c,

$$\chi(z_\tau)=\begin{cases}(1+\gamma^{ab})\displaystyle\sum_{n=1}^{\infty}\chi^n(z_\tau^a-\tau^a h+\tau^b h), & \text{in } S_a\\ \displaystyle\sum_{n=1}^{\infty}[\chi^n(z_\tau^b)+\gamma^{ab}\bar{\chi}^n(z_\tau^b-\tau^b h+\bar{\tau}^b h)], & \text{in } S_b\\ (1+\gamma^{cb})\left[\chi^0(z_\tau^c)+\gamma^{ab}\displaystyle\sum_{n=1}^{\infty}\bar{\chi}^n(z_\tau^c-\tau^b h+\bar{\tau}^b h)\right], & \text{in } S_c\end{cases} \tag{26}$$

$$f_i(z_{\underline{i}})=\begin{cases}\displaystyle\sum_{n=1}^{\infty}[U_{ij}^{ab}f_j^n(z_{\underline{i}}^a-\mu_{\underline{i}}^a h+\mu_{\underline{j}}^b h)+e_i^{ab}\chi^n(z_{\underline{i}}^a-\mu_{\underline{i}}^a h+\tau^b h)], & \text{in } S_a\\ \displaystyle\sum_{n=1}^{\infty}[f_i^n(z_{\underline{i}}^b)+\bar{V}_{ij}^{ab}\bar{f}_j^n(z_{\underline{i}}^b-\mu_{\underline{i}}^b h+\bar{\mu}_{\underline{j}}^b h)+\bar{g}_i^{ab}\bar{\chi}^n(z_{\underline{i}}^b-\mu_{\underline{i}}^b h+\bar{\tau}^b h)], & \text{in } S_b\\ U_{ij}^{cb}f_j^0(z_{\underline{i}}^c)+e_i^{cb}\chi^0(z_{\underline{i}}^c)+\displaystyle\sum_{n=1}^{\infty}\{U_{ij}^{cb}[\bar{V}_{jk}^{ab}\bar{f}_k^n(z_{\underline{i}}^c-\mu_{\underline{j}}^b h+\bar{\mu}_{\underline{k}}^b h)+\bar{g}_j^{ab}\bar{\chi}^n(z_{\underline{i}}^c-\mu_{\underline{j}}^b h+\bar{\tau}^b h)]+e_i^{cb}\gamma^{ab}\bar{\chi}^n(z_{\underline{i}}^c-\tau^b h+\bar{\tau}^b h)\}, & \text{in } S_c\end{cases} \tag{27}$$

where the recurrence formula for $\chi^n(z)$ and $f_i^n(z)$ are

$$\chi^{n+1}(z)=(\gamma^{ab}\gamma^{cb})^n[\chi^0(z+\tau^b hn-\bar{\tau}^b hn)+\gamma^{cb}\bar{\chi}^0(z+\tau^b hn-\bar{\tau}^b hn)], \quad n=0,1,2,\ldots \tag{28}$$

$$f_i^{n+1}(z)=\begin{cases}f_i^0(z)+\bar{V}_{ij}^{cb}\bar{f}_j^0(z)+\bar{g}_i^{cb}\bar{\chi}^0(z), & n=0\\ \bar{V}_{ij}^{cb}[V_{jk}^{ab}f_k^n(z-\bar{\mu}_{\underline{j}}^b h+\mu_{\underline{k}}^b h)+g_j^{ab}\chi^n(z-\bar{\mu}_{\underline{j}}^b h+\tau^b h)]+\bar{g}_i^{cb}\gamma^{ab}\chi^n(z-\bar{\tau}^b h+\tau^b h), & n=1,2,3,\ldots\end{cases} \tag{29}$$

In Eqs. (28) and (29), the thermoelastic constants involved in $\chi^0(z)$ and $f_i^0(z)$ are for material b.

The rate of convergence of the obtained solutions depends on the bimaterial constants γ^{ab} and γ^{cb} and the bimaterial matrices $\mathbf{T}^{ab}$ and $\mathbf{T}^{cb}$, which represent a mismatch of the thermoelastic constants of the two constituent materials. Knowing that the absolute value of γ^{ab} (or γ^{cb}) defined in Eq. (10) is less than or equal to unity, and using the same argument as given in Sec. 6.1 of Choi and Earmme [2], one verifies that a sufficient condition for the convergence of the series solutions is satisfied. The smaller the difference in the thermoelastic constants of two adjacent materials, a and b (or c and b), the smaller the norm of γ^{ab} (or γ^{cb}) and the norm of $\mathbf{T}^{ab}$ (or $\mathbf{T}^{cb}$). The convergence rate consequently becomes more rapid. Furthermore, since the ordinates of the image singularities are linearly proportional to thickness h of material b, the thickness of the material b also affects the rate of convergence; that is, as h gets larger, the series solution is more rapidly convergent. Thus, the sum of the first three or four terms provides a good approximation for most combinations of materials [2]. Even if materials a or c or both are rigid or nonexistent, the solutions still remain valid. These limiting cases are discussed by Choi and Earmme [2]. For another limiting case in which two adjacent materials, say materials a and b, are identical, the series solution for a trimaterial reduces to a bimaterial solution.

5 Conclusion

Schwarz-Neumann's alternating technique and the method of analytic continuation were used to study the thermoelastic singularity in an anisotropic trimaterial. The results show that a homogeneous solution for thermoelastic singularity serves as a base for deriving a trimaterial solution for the same singularity in a series form. The convergence rate of the series solution depends on the material combinations and the thickness of the middle material. The smaller the mismatch of thermoelastic constants of adjacent materials, the more rapid the convergence rate is. As two adjacent materials degenerate to become a homogeneous material, the trimaterial solution reduces to a bimaterial solution. The trimaterial solution studied here can be applied to a variety of thermoelastic problems for singularities as well as cracks (with continuous distribution of dislocations) in a bimaterial (including a half-plane problem), a finite thin film on semi-infinite substrate, a finite strip of thin film, and so on. In fact, the merit of this trimaterial solution is its wide applicability to bimaterial problems in addition to the trimaterial problem per se.

References

[1] Tu, K. N., Mayer, J. W., and Feldman, L. C., 1992, *Electronic Thin Film Science for Electrical Engineers and Materials Scientists*, Macmillan, New York, pp. 157–189.

[2] Choi, S. T., and Earmme, Y. Y., 2002, "Elastic Study on Singularities Interacting With Interfaces Using Alternating Technique: Part I. Anisotropic Trimaterial," Int. J. Solids Struct., **39**, pp. 943–957.

[3] Suo, Z., 1990, "Singularities, Interfaces and Cracks in Dissimilar Anisotropic Media," Proc. R. Soc. London, Ser. A, **A427**, pp. 331–358.

[4] Sokolnikoff, I. S., 1956, *Mathematical Theory of Elasticity*, McGraw-Hill, New York, pp. 318–326.

[5] Eshelby, J. D., Read, W. T., and Shockley, W., 1953, "Anisotropic Elasticity With Applications to Dislocation Theory," Acta Metall., **1**, pp. 251–259.

[6] Stroh, A. N., 1958, "Dislocations and Cracks in Anisotropic Elasticity," Philos. Mag., **3**, pp. 625–646.

[7] Ting, T. C. T., 1996, *Anisotropic Elasticity: Theory and Applications*, Oxford University Press, London, pp. 506–511.